UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE CIÊNCIA E TECNOLOGIA DE ALIMENTOS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA DE ALIMENTOS

Transcriptoma Global e de Genes Alvo da Rota de Biossíntese de Glicoalcalóides em Novas Variedades de *Solanum tuberosum* L.

Roberta Fogliatto Mariot

Porto Alegre

2015

Roberta Fogliatto Mariot

TRANSCRIPTOMA GLOBAL E DE GENES ALVO DA ROTA DE BIOSSÍNTESE DE GLICOALCALÓIDES EM NOVAS VARIEDADES DE *SOLANUM TUBEROSUM* L.

Tese apresentada como requisito parcial para obtenção do titulo de Doutor em Ciência e Tecnologia de Alimentos pelo Programa de Pós-graduação em Ciência e Tecnologia de Alimentos do Instituto de Ciência e Tecnologia de Alimentos da Universidade Federal do Rio Grande do Sul.
Orientador: Jeverson Frazzon.

Porto Alegre

2015

CIP - Catalogação na Publicação

Mariot, Roberta Fogliatto. Transcriptoma global e de genes alvo da rota de biossíntese de glicoalcalóides em novas variedades de solanum tuberosum l. / Roberta Fogliatto Mariot. -2015.
104 f.

Orientador: Jeverson Frazzon.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Instituto de Ciência e Tecnologia de Alimentos, Programa de Pós-Graduação em Ciência e Tecnologia de Alimentos, Porto Alegre, BR-RS, 2015.

1. Glicoalcalóides totais. 2. Solanum tuberosum.3. transcriptômica. I. Frazzon, Jeverson , orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

Tese de Doutorado

Transcriptoma Global e de Genes Alvo da Rota de Biossíntese de Glicoalcalóides em Novas Variedades de *Solanum tuberosum* L.

Roberta Fogliatto Mariot

Aprovada em: ____/____/_____.

Jeverson Frazzon (Orientador)
Doutor em Ciências Biológicas: Bioquímica
(ICTA/UFRGS)

Ana Carolina Maisonnave Arisi
Doutora em Bioquímica
(Centro de Ciências Agrárias/UFSC)

Maurício Reis Bogo
Doutor em Genética e Biologia Molecular
(Faculdade de Biociências/PUCRS)

Ana Paula Körbes
Doutora em Matemática e Ciências Naturais
(PPGBM/UFRGS)

Vitor Manfroi
Doutor em Ciência e Tecnologia Agroindustrial
(ICTA/UFRGS)

RESUMO

Batatas (*Solanum tuberosum*) juntamente com arroz, trigo e milho, são as culturas mais consumida no mundo. A inclusão de novas variedades de *S. tuberosum* para consumo humano e sob a forma de ração necessita avaliação quanto sua segurança. Normas internacionais citam glicoalcalóides (GA) como tóxicos-chave em avaliação de novas variedades de batatas. Portanto, o objetivo do trabalho foi propor um modelo de avaliação da segurança de novas variedades de batata, incluindo geneticamente modificados, baseado no transcriptoma e análise comparativa de variedades com histórico de segurança. Adicionalmente, pretendeu-se determinar genes referência para análise transcricional de tubérculos. E ainda, caracterizar os genes *glycoalkaloid metabolism* (*GAME*s), que fazem parte da biossíntese de GA e relacionar sua expressão com o total de glicoalcalóides (TGA) em diferentes genótipos de tubérculos de batatas. Para tanto, construiu-se um banco de dados com *RNAseq* de 90 tubérculos de batatas com grande variabilidade genética e natural. O banco de dados também foi utilizado para ranquear candidatos a genes normalizadores, baseado no intervalo interquartil (IQR), que, posteriormente, tiveram sua estabilidade calculada pelo: *geNorm*, *NormFinder* e *BestKeeper*. O conteúdo de TGA foi medido em cromatografia líquida de alta eficiência com espectrômetro de massa (CLAE-EM) e a expressão dos *GAME*s foi determinada por reação de transcriptase reversa quantitativa seguida da reação em cadeia da polimerase (RT-qPCR). Para análise da região promotora utilizou-se: *Plant Pan – Plant Promoter Analysis Navigator*. Através dos resultados obtidos e analisados, podemos afirmar que o modelo, quando devidamente validado e finalizado, pode ser considerado bastante promissor. Os genes *C2*, *SEC3* e *CUL3A* foram considerados como excelentes normalizadores para tubérculos de batata. A expressão dos *GAME*s foi maior em tubérculos com maior conteúdo de TGA, e com exceção do *GAME7*, os genes *GAME4*, *GAME6*, *GAME8ab*, *GAME11* e *GAME12* demonstraram ser estatisticamente ($\alpha = 0,05$) mais expressos em amostras com maior teor de TGA e menos em amostras com menor TGA, confirmando a relação destes genes com a produção de glicoalcalóides. Na análise de fatores de transcrição

dos *GAME*s muitos elementos *cis* relacionados a estresse biótico e abiótico, e regulação por luz foram encontrados, além de muitas cópias desses elementos *cis* nas regiões promotoras de todos os *GAME*s. Os resultados obtidos no presente trabalho são importantes para um melhor entendimento da formação e regulação dos GA em tubérculos de batatas, auxiliando na predição e prevenção de formação deste composto.

Palavras-chave: Glicoalcalóides totais, *Solanum tuberosum*, transcriptômica.

ABSTRACT

Potato (*Solanum tuberosum*) is the third most important food crop in the world after rice and wheat in terms of human consumption. The introduction of new varieties of *S. tuberosum* for food or feed requires the checking of not only substances of pro-nutritional functioning on the human body, but should examine the content of toxic compounds, such as glycoalkaloids (GA), as an international norm recommendation for new potatoes varieties evaluation, including genetic modified. For this reason, the goal of the present work was build a food safety assessment model based on transcriptomics, and comparative with history of safe human consumption. Besides, we intended to determined reference genes for transcriptional analysis of potato tuber, and characterized glycoalkaloid metabolism (*GAME*) genes that partake of GA biosynthesis pathway. The correlation between the expression of *GAME* genes and total glycoalkaloids (TGA) content in different genotypes of potato tubers was evaluated. To do so, a database was build with *RNASeq* data from 90 potato tubers with large genotype and natural variability. These database were also used to ranking the candidates to be reference genes based on Interquartile Range (IQR). The stability of candidates for reference genes were evaluated by *geNorm*, *NormFinder* and *BestKeeper*. The content of TGA were measured on high performance liquid chromatography acoplated on mass spectrometry (LC-MS). Later, Reverse Transcriptase quantitative Polymerase Chain Reactions (RT-qPCR) determined the expression of GAME genes. For the promoter region analysis of *GAME*s were used “Plant Pan – Plant Promoter Analysis Navigator”. It is possible affirmed that the food safety assessment model after a properly validation will be promising. The best reference genes for transcriptional analysis studies of potato tubers were *C2*, *SEC3* and *CUL3A*. The expression of *GAME4*, *GAME6*, *GAME7*, *GAME8ab*, *GAME11*, and *GAME12* were higher in tubers with high TGA content, and, exception for *GAME7*, the other *GAME* genes demonstrate statically ($\alpha = 0,05$) higher expression on samples with high TGA and lower on samples with low TGA content, confirming that there is a relation between *GAME* genes and GA biosynthesis. On transcription factors analysis of *GAME* genes several cis-

elements related to abiotic and biotic stress, and light regulated were found, also many copies of these factors in all *GAME* promoter region. All these results are very helpful to better understand GA formation and regulation on potato tubers, helping to predict and prevent the formation of these toxic.

Keywords: Total glycoalkaloids, transcriptomics, *Solanum tuberosum*

LISTA DE SIGLAS

CLAE-EM: Cromatografia Líquida de Alta Eficiência e Detecção em Espectrômetro de Massa

CUL3A: ATCUL3/ATCUL3A/CUL3/CUL3A

GA : Glicoalcalóides

GAME4 :*Glycoalkaloid Metabolism 4*

GAME6 :*Glycoalkaloid Metabolism 6*

GAME7 :*Glycoalkaloid Metabolism 7*

GAME8a*: Glycoalkaloid Metabolism 8a*

GAME8b :*Glycoalkaloid Metabolism 8b*

GAME11 :*Glycoalkaloid Metabolism 11*

GAME12 :*Glycoalkaloid Metabolism 12*

IQR : Intervalo Interquartil (*Interquartile Range*)

NGS : Sequenciamento de Última Geração *(next-generation sequencing*)

RT-qPCR : Transcriptase Reversa quantificação de Reações em Cadeia da Polimerase (*Reverse transcriptase quantitative polymerase chain reaction*)

SEC3 :*Exocyst complex component sec3*

SGT1 : *Solanum tuberosum 3 UDP-galactose: solanidine galactosyltransferase*

SGT2 : *Solanum tuberosum UDP-glucose: solanidine glucosyltransferase*

SGT3 : Solanum tuberosum *rhamnose: beta-solanine / beta-chaconine rhamnosyltransferase S. tuberosum Solanum tuberosum*

TGA : Total de Glicoalcalóides

SUMÁRIO

1. INTRODUÇÃO

Nas últimas décadas, apesar de todo o desenvolvimento tecnológico e científico, aumentou o número de alertas sobre perigos presentes em alimentos, por exemplo, dioxinas, contaminação por Salmonela, sorotipo de *Escherichia coli* O157:H7, entre outros. Todos estes fatos criaram uma desconfiança entre os consumidores. Por isso, pesquisas relacionadas com segurança e qualidade dos alimentos se tornaram prioridades (KUIPER et al., 2001). Kuiper e seus colaboradores (2001) salientam que a produção de alimentos atualmente é complexa e, perigos intrínsecos, podem ser adicionados, como por exemplo, alimentos de origem microbiana, animal e vegetal podem conter toxinas naturais do próprio alimento e estar contaminados por patógenos e produtos químicos, inclusão de ingredientes e novos alimentos antes não utilizados para consumo humano ou animal sob a forma de rações. Desta forma, para garantir uma alimentação segura para a população, toda a cadeia de produção de alimentos, do campo a mesa, deve ser devidamente regulamentada e fiscalizada.

Com a introdução de novos alimentos, através do aumento no número de tecnologias disponíveis para melhoramento vegetal, introduzindo características desejadas no genoma das plantas, podem ocorrer alguns casos relacionados com potenciais riscos à segurança alimentar. No entanto, a avaliação de risco dessas novas variedades ainda não é um procedimento padrão para sua liberação no mercado. Diversos debates têm surgido em torno deste assunto, sendo levantadas diversas discussões sobre a necessidade de desenvolver metodologias que garantam a segurança alimentar dessas novas plantas disponibilizadas aos consumidores. Neste contexto, uma das recomendações foi de investigar o uso das tecnologias ômicas como ferramentas para identificar alterações não intencionais na fisiologia das plantas. Assim, diversos institutos de pesquisa na Europa têm investigado o uso da proteômica, transcriptômica e metabolômica para detecção destes efeitos inesperados (VAN DIJK et al., 2009).

Os glicoalcalóides (GA) são metabólitos secundários de plantas que, acima de determinadas concentrações, podem ser tóxicos tanto para bactérias e fungos como para insetos e até seres humanos. Os principais GA encontrados nas batatas domésticas (*Solanum tuberosum*) são α-solanina e α-chaconina. Estas são substâncias que contribuem no sabor de batatas, mas em altas concentrações podem causar amargor e apresentam alta toxicidade. Sua principal função está relacionada à resposta ao estresse, como produção de fitoalexinas para proteção contra o ataque de patógenos. Aumentos no teor de α-solanina e α-chaconina em batata foram relacionados com aumentada síntese de clorofila, que causa esverdeamento da casca; embora esses processos sejam independentes, ambos são ativados pela luz (FRIEDMAN, 2006; NEMA et al., 2008).

2. OBJETIVOS

Em virtude do acima exposto, os objetivos desse trabalho foram:

- Determinar genes referência para análise transcriptômica de 8 amostras de tubérculos de batatas.
- Caracterizar os genes *glycoalkaloid metabolism* (*GAME*s) que fazem parte da biossíntese de glicoalcalóides em batatas.
- Relacionar a expressão dos genes *GAME*s com o conteúdo total de glicoalcalóides (α-chaconina e α-solanina) em diferentes genótipos de 8 amostras de tubérculos de batatas
- Propor um modelo de avaliação de segurança de novas variedades de *S. tuberosum* baseado em transcriptoma global e análise comparativa a variedades com histórico de segurança.

3. REFERÊNCIAL TEÓRICO

3.1 BATATA (*SOLANUM TUBEROSUM*)

Batatas são consumidas em todo o mundo e, atrás apenas do arroz, trigo e milho, são a cultura mais consumida no mundo (FAO, 2008). O cultivo da batata originou-se na América do Sul a mais de 10.000 anos (CAMIRE; KUBOW; DONNELY, 2009) e nos últimos cinco séculos, o cultivo dos tubérculos adaptados se expandiu para o continente europeu, africano e asiático. Batatas atualmente são cultivadas em mais de 160 países (CAMIRE; KUBOW; DONNELY, 2009), e a média global per capita de consumo de batatas é de 33 kg/ano (FAO, 2008).

Por ano, mais de 300 milhões de toneladas de batatas são produzidas em todo o mundo, conforme dados extraídos da base de dados da Organização das Nações Unidas para a Alimentação e a Agricultura (FAOSTATE DATABASE, 2013), e mais de 1 bilhão de pessoas têm a batata com a base da alimentação. No Brasil, mais de 3 milhões de toneladas de batatas produzidas anualmente, em torno de 1% de toda a produção mundial, sendo o Rio Grande do Sul um dos Estados brasileiros de maior produção, atrás somente de Minas Gerais (PEREIRA et al., 2005).

Conforme Liedl e seus colaboradores (1987), dentre as culturas, a batata é segunda em relação a conteúdo de proteína/hectare, perdendo apenas para soja. Dentre as proteínas da batata, a de maior quantidade é a patatina, considerada uma das mais equilibradas proteínas vegetais conhecidas. Burlingame e seus colaboradores (2009) referem-se aos tubérculos como uma importante fonte de carboidratos, proteínas, antioxidantes e vitaminas para dieta e também como sistema de propagação vegetal.

A batata é um membro da família *Solanaceae*, possuindo mais de 3.000 espécies. Possuindo um dos mais ricos recursos genéticos dentre as plantas cultiváveis, com 190 espécies selvagens e primitivas do gênero *Solanum*

(BARRELL et al., 2013; JACOBS et al., 2011; SPOONER; HIJMANS, 2001). As espécies de tubérculos *Solanum* são amplamente distribuídas nas Américas. Muitas espécies selvagens, que podem ter sido cruzadas com as espécies cultivadas de batatas, proporcionaram um amplo espectro de resistências a doenças e pragas, tolerâncias a estresse abiótico como geadas, secas, entre outros (BARRELL et al., 2013; PLAISTED; HOOPES,1989).

Algumas décadas atrás, a batata comercial possuía uma base genética limitada, em virtude de ser uma planta original do centro da América do Sul e recém introduzida na Europa. Além disso, sua baixa propagação vegetativa torna *S. tuberosum* uma planta altamente heterozigota, poliploide, e os seus cultivares, portanto, muitas vezes limitadamente férteis, tornando a sua seleção uma tarefa desafiadora. Sabe-se ainda, que os esquemas de melhoramento tradicionais envolvem relativamente longos ciclos reprodutivos, geneticamente complexos e os tubérculos de batata possuem uma sensibilidade à endogamia. Entretanto, ao longo das últimas cinco décadas, mediante aplicações de biotecnologia vegetal, que facilitaram os cruzamentos entre espécies, ampliou-se assim o *pool* genético das plantas cultivadas atualmente (BARRELL et al., 2013). A biotecnologia vegetal tradicionalmente englobou a aplicação de células e tecidos para o melhoramento de variedades. E. a partir de meados da década de 80, o desenvolvimento e a aplicação da transgenia, tornou a atividade de pesquisa relacionada à biotecnologia vegetal dominante (BARRELL et al. 2013).

Muitos genótipos de batata são altamente sensíveis em culturas celulares e geram oportunidades de aplicação de biotecnologia a melhoramento das batatas. A recente publicação do sequenciamento completo do genoma da batata (THE POTATO SEQUENCING CONSORTIUM, 2011) tem apresentado novas oportunidades para biotecnologia de *S. tuberosum*.

3.2 GLICOALCALÓIDES (GA)

Em virtude do avanço da biotecnologia vegetal, novos alimentos vêm sendo criados, entrando no mercado tanto nas prateleiras do varejo quanto na alimentação animal. Em virtude disso, o *OECD´s Task Force for Safety of Novel Foods and Feeds*, no ano de 1999, decidiu criar documentos chamados *consensus documents*, com o objetivo de focar o trabalho de desenvolvimento científico na área de segurança destes novos alimentos. Estes documentos contêm informações sobre nutrientes, anti-nutrientes e tóxicos relevantes que devem ser utilizadas como base no desenvolvimento de avaliações de segurança regulatórias para um alimento/ração animal específico. O *consensus document* que aborda considerações da composição dos tubérculos de batata, identificando os nutrientes, antinutrientes e tóxicos para avaliação da segurança de alimentos para novas variedades destas plantas, cita o GA como o tóxico-chave a ser considerado na avaliação da segurança de novas variedades, segundo o OECD (2002).

Batatas naturalmente contêm muitos tipos de alcalóides, mas o mais importante grupo de alcalóides em batatas comerciais são os GA, compostos principalmente por 2 tipos: α-chaconina (solanidina-glicose-raminose-raminose) e α-solanina (solanidina-galactose-glicose-raminose) (NEMA et al., 2008; FRIEDMAN, 2006; OECD, 2002).

Segundo, Smith e seus colaboradores (1996), os GA não estão uniformemente distribuídos no interior dos tubérculos, mas estão presentes em concentrações mais elevadas na sua parte mais externa. Em virtude disto, o tamanho do tubérculo é importante para avaliação do nível de GA. Grandes variações imprevisíveis deste tóxico em batatas podem surgir em virtude de muitos fatores, como cultivar, local do plantio, estação do ano, estresse da planta, prática de cultivo, entre outros. OECD (2002), citando Smith e seus colaboradores (1996) como referência, determinaram que o limite de segurança aceito para o conteúdo de GA totais em tubérculos para o consumo humano é de 200 mg/kg peso fresco. Entretanto, a concentração total de GA nem sempre é essencial para definir a sua toxicidade a humanos e animais. Sabe-se que a

α-chaconina é três vezes mais tóxica que a α-solanina, portanto é desejável que a razão (α-chaconina/ α-solanina) entre as duas seja a menor possível (FRIEDMAN, 2006). Além disso, existe um efeito sinérgico quando estes 2 compostos estão presentes no mesmo tecido, mesmo que a toxicidade da α-chaconina seja significantemente maior que a α-solanina (NEMA et al., 2008).

Importante salientar que estes tóxicos não são desativados durante o processo de cocção das batatas, pois a α-chaconina/ α-solanina são moléculas termoestáveis. Portanto, uma vez que os tubérculos possuem em sua composição altos teores de GA (NEMA et al. 2008), não existe uma forma de prevenção da intoxicação tanto de seres humanos quanto de animais. Em virtude disto, a avaliação de novos cultivares de *S. tuberosum* em relação a este perigo é de suma importância para a segurança de alimentos.

A intoxicação por GA pode causar diversos sintomas, desde desordens gastrointestinais, até confusão, alucinação, paralisia parcial, convulsões, coma até levar a morte, dependendo da susceptibilidade humana e da quantidade do tóxico presente (OECD, 2002; SMITH et al., 1996). A suscetibilidade humana ao envenenamento por GA é alta e muito variável: doses orais entre 1 – 5 mg/kg de peso corpóreo são severamente tóxicos, já doses entre 3 – 6 mg/kg de peso corpóreo podem ser letais, conforme Morris e Lee (1984).

O OECD (2002) menciona ainda a suscetibilidade de suínos em relação aos GA em batatas, foram determinadas que concentrações de 150 mg de GA totais/kg de peso fresco não provocaram nenhum sintoma e/ou diminuição do crescimento desses animais. Resultados similares foram encontrados em relação ao gado, onde não foram encontrados riscos de intoxicação dos animais quando porções de batatas com os brotamentos devidamente removidos foram incorporadas na ração, as quais continham concentrações que variavam de 2000 a 5000 mg de GA total/peso fresco.

Como mencionado anteriormente, os GA são particularmente concentrados na região exterior dos tubérculos de batata, entretanto, em

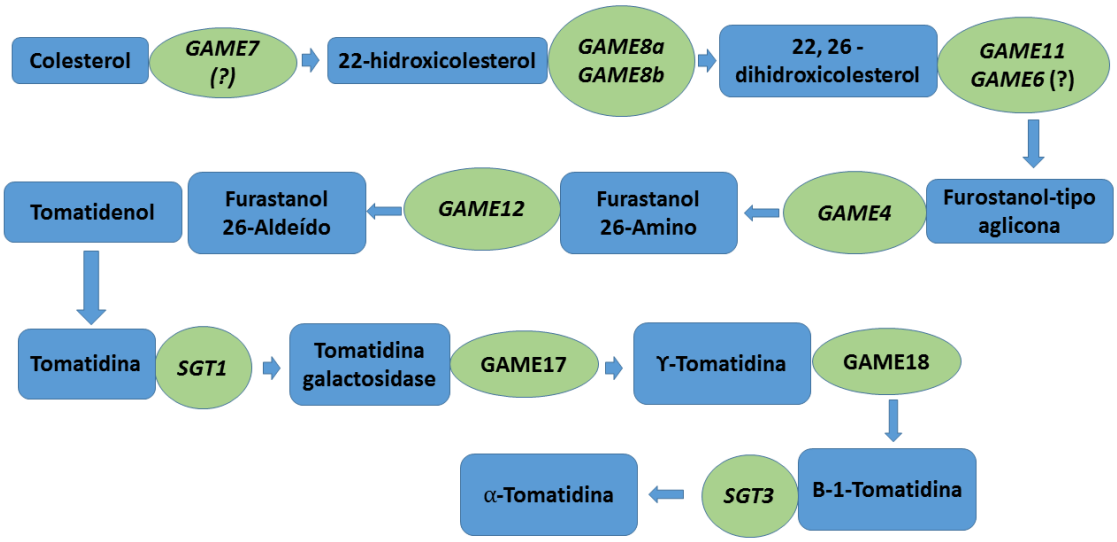
tubérculos de batata esverdeados e/ou com brotamentos as concentrações também são altas no interior do tubérculo. Contudo, em todos os casos, o descascamento reduz consideravelmente o conteúdo de GA totais do tubérculo (NEMA et al., 2008; FRIEDMAN, 2006). Além disso, o pós-colheita é relatado por influenciar a formação de GA em tubérculos durante o seu armazenamento; por exemplo, danos físicos, germinação, altas temperaturas e exposição à luz podem incrementar o conteúdo de GA significantemente. E os métodos adotados por produtores e industrializadores deste produto para atrasar a formação de GA incluem armazenamento das batatas na escuridão, tratamentos com produtos químicos, irradiação e atmosferas modificadas (MACHADO; TOLEDO; GARCIA, 2007; SENGUL; KELES; KELES, 2004).

Apesar dos GA serem considerados tóxicos para humanos, trazendo riscos a saúde e perdas para indústria, estes compostos possuem um papel essencial no desenvolvimento das plantas da família das *Solanaceae*, podendo ser considerados antimicrobianos, inseticida e, também com propriedades fungicidas, protegendo a planta de ataques de insetos, pestes e herbívoros (NEMA et al.; 2008; FRIEDMAN, 2006). Além disso, nos últimos 15 anos, alguns estudos sugerem que os GA também podem possuir efeitos benéficos para os consumidores destes compostos, dependendo da dose e também da condição de uso, como efeitos contra células humanas cancerígenas (LEE et al. 2004; FRIEDMAN et al. 2005) e no sistema imune em ratos (GUBAREV et al., 1998).

Apesar da importância dos GAs, estudos relacionados a rota metabólica de produção e regulação destes compostos eram bastante escassos. Os GAs são produzidos em todas as partes da planta, como folhas, caule, tubérculos, e a biossíntese ocorre através da via do colesterol (FRIEDMAN, 2006; GINZBERG et al., 2012). Entretanto, muitas etapas entre o colesterol até a formação das moléculas de α-chaconine and α-solanine eram desconhecidas. No ano de 2013 Itkin e seus colaboradores publicaram um importante trabalho sobre biossíntese de GAs em tomates e batatas, e sugeriram uma rota

metabólica de biossíntese de GA em plantas da família das *Solanaceae* (Figura 1).

Figura 1. Rota metabólica de biossíntese de glicoalcalóides para plantas da família das *Solanaceae* adaptado de ITKIN et al. (2013).



Antes deste trabalho, apenas os genes que codificavam as enzimas das etapas finais da rota de biossíntese entre o colesterol ao GA foram descritos. São eles: solanidine galactosyltransferase (*SGT1*), solanidine glucosyltransferase (*SGT2*) (PAJEROWSKA et al., 2005; MCCUE et al., 2006; KRITS et al., 2007; GINZBERG et al., 2012) e β-solanine/β-chaconine rhamnosyl transferase (*SGT3*) (MCCUE et al., 2007), haviam sido descritas.

3.3 TRANSCRIPTÔMICA E AS TECNOLOGIAS ÔMICAS

Há uma década, pesquisas com expressão gênica eram reservadas apenas para genética humana e para organismos modelos. Mesmo para estes

sistemas, microarranjos e análises em série de expressão gênica eram as únicas ferramentas disponíveis para estudar transcriptoma e padrões globais de expressão. Já para organismos não-modelos, estudo de expressão gênica eram ainda mais restritivas, somente pequenas escalas de genes candidatos usando RT-qPCR ou hibridização de espécies relacionadas em microarranjos eram utilizadas (NAURIN et al. 2008). Todavia, com o rápido desenvolvimento dos sequenciamentos de última geração (*next-generation sequencing* – NGS) (WOLF, 2013) e suas ferramentas de análise, nos últimos anos, esta situação mudou drasticamente. Análises do genoma (genômica) e transcriptoma global (transcriptômica) tornaram-se uma opção realística para organismos não-modelo, como por exemplo *S. tuberosum*, mesmo em pequenos laboratórios (WOLF, 2013; ELLEGREN et al., 2012; LAMICHHANEY et al., 2012), e em breve será uma prática padrão em biologia molecular.

Conforme visto acima, os avanços dos NGS e diversos outros instrumentos analíticos, vêm possibilitando a análise de genes, transcritos, proteínas (proteômica) e metabólitos (metabolômica) (FUKUSHIMA; KANAYA; NISHIDA, 2014; STITT, 2013; LUCAS et al., 2011; LEI et al., 2011). Estas tecnologias ômicas não são somente plataformas onde se monitoram o inventário das células, mas também permitem a oportunidade de avaliar o comportamento celular de diversas perspectivas e, portanto aumentam nosso entendimento destes sistemas (DHONDT; WUYTS; INZE, 2013; KROUK et al., 2010; SAITO; MATSUDA, 2010).

Plantas são consideradas fontes primordiais de alimentos, energia e compostos; e o campo de desenvolvimento de sistemas biológicos, baseado em tecnologias ômicas, têm fornecido informações de como estes produtos são sintetizados, com o objetivo de melhor compreender a relação entre genótipo e fenótipo destes organismos (FUKUSHIMA; KANAYA; NISHIDA, 2014; WECKWERTH, 2011; BENFEY; MITCHELL-OLDS, 2008; KELL, 2002). No entanto, a análise de dados destas diversas perspectivas é de suma importância para que o entendimento destas relações seja interpretado corretamente. As principais abordagens eficazes e eficientes para analisar

dados de tecnologias ômicas são análises de rede e análises de rotas metabólicas conforme Fukushima, Kanaya e Nishida (2014), Carter e seus colaboradores (2013) e Ramanan e seus colaboradores (2012).

Os capítulos 4, 5 e 6 referem-se aos artigos científicos publicados e que serão submetidos para publicação em revistas da área. Segue uma pequena descrição de cada um deles:

Capítulo 4. Artigo científico pulicado na revista PLos One, que trata da determinação de genes de referência para a análise transcricional de tubérculos de batatas.

4. ARTIGO CIENTÍFICO 1

4.1 CARTA DE ACEITE “PLos One”

PONE-D-14-42876R2
Selection of reference genes for transcriptional analysis of edible tubers of potato (Solanum tuberosum L.)

Dear Dr. Frazzon:

I am pleased to inform you that your manuscript has been deemed suitable for publication in PLOS ONE. Congratulations! Your manuscript is now with our production department.

If you or your institution will be preparing press materials for this manuscript, you must inform our press team in advance. Your manuscript will remain under strict press embargo until 2 pm Eastern Time on the date of publication. For more information please contact onepress@plos.org.

For any other questions or concerns, please email plosone@plos.org.

Thank you for submitting your work to PLOS ONE.

With kind regards,

PLOS ONE Editorial Office Staff
on behalf of

Prof. Zhilong Bie
Academic Editor
PLOS ONE

Selection of reference genes for transcriptional analysis of edible tubers of potato (*Solanum tuberosum* L.)

Roberta Fogliatto Mariot[1¶], Luisa Abruzzi de Oliveira[2¶], Marleen M. Voorhuijzen[3], Martijn Staats[3], Ronald C.B. Hutten[4], Jeroen P. Van Dijk[3], Esther Kok[3] and Jeverson Frazzon[1*]

[1] Food Science Department, Food Science and Technology Institute, Federal University of Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil

[2] Genetics Department, Federal University of Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil

[3] RIKILT, Wageningen UniversityR, Wageningen, The Netherlands

[4] Plant Breeding, Wageningen UR, Wageningen, The Netherlands

*Corresponding author

E-mail: jeverson.frazzon@ufrgs.br (JF)

¶These authors contributed equally to this work.

Abstract

Potato (*Solanum tuberosum*) yield has increased dramatically over the last 50 years and this has been achieved by a combination of improved agronomy and biotechnology efforts. Gene studies are taking place to improve new qualities and develop new cultivars. Reverse transcriptase quantitative polymerase chain reaction (RT-qPCR) is a bench-marking analytical tool for gene expression analysis, but its accuracy is highly dependent on a reliable normalization strategy of an invariant reference genes. For this reason, the goal of this work was to select and validate reference genes for transcriptional analysis of edible tubers of potato. To do so, RT-qPCR primers were designed for ten genes with relatively stable expression in potato tubers as observed in *RNA-Seq* experiments. Primers were designed across exon boundaries to avoid genomic DNA contamination. Differences were observed in the ranking of candidate genes identified by *geNorm, NormFinder* and *BestKeeper* algorithms. The ranks determined by *geNorm* and *NormFinder* were very similar and for all samples the most stable candidates were *C2*, exocyst complex component sec3 (*SEC3*) and *ATCUL3/ATCUL3A/CUL3/CUL3A* (*CUL3A*). According to *BestKeeper*, the importin alpha and ubiquitin-associated/ts-n genes were the most stable. Three genes were selected as reference genes for potato edible tubers in RT-qPCR studies. The first one, called *C2*, was selected in common by *NormFinder* and *geNorm*, the second one is *SEC3*, selected by *NormFinder*, and the third one is *CUL3A*, selected by *geNorm*. Appropriate reference genes identified in this work will help to improve the accuracy of gene expression quantification analyses by taking into account differences that may be observed in RNA quality or reverse transcription efficiency across the samples.

Introduction

A wide range of biological processes leads to changes in mRNA transcription levels, and these variations are important to ensure timely cellular responses. Based on this, mRNA transcriptional profiling has become a popular research field in functional genomics studies, as it can be used to evaluate complex regulatory gene networks [1-3]. Reverse transcriptase quantitative polymerase chain reaction (RT-qPCR) has been commonly used to analyze gene expression in different organisms and under numerous conditions, since it permits specific and reproducible quantification of nucleic acids [1, 4]. However, the stability of the expressed housekeeping gene is a fundamental factor in the appropriate standard normalization of the data, which is usually normalized to more than one reference gene to avoid differences in complementary DNA (cDNA) quantity, purity, RNA stability, and enzymatic efficiency of cDNA synthesis and subsequent PCR amplifications [5-6]. The assortment of an appropriate reference gene is an absolute requirement to minimize non-biological variation between samples and achieve precise results [7]; hence, the selection of suitable reference genes is crucial to RT-qPCR analysis. The ideal reference gene would be stably expressed through all examined samples [8-9].

Many reference genes have already been identified for several crops under different treatments and conditions, particularly for model plants [10]. However, the expression of putative reference genes differs across individual sets of organs and experimental conditions [7, 11]. In this context, several free software packages such as *geNorm* [12], *NormFinder* [9] or *BestKeeper* [13], may be used in order to the best internal controls from a group of candidate normalization genes for a specific set of biological samples.

The goal of this study was to examine by RT-qPCR the stability of ten putative reference genes selected from *RNAseq* experiments. We have focused the investigation of control genes by evaluating the expression variability of 10 genes with relatively high stability levels in potato tubers.

Material and Methods

Ethics statements

The field experiments in both years (in this case normal yield trials) were performed on a trial field in the proximity of Wageningen (GPS coordinates: 51.95230, 5.63490) owned by Wageningen UR. No specific permission was required to carry out these potato trials.

Field experimental design

Eight potato edible tubers from four distinct genotypes, experimental lines, obtained in duplicates, one grown in 2011 and the other in 2012, with a post-harvest storage time of 13 and 28 days, and cultivated at Plant Breeding Sciences - Wageningen University and Research Center (WUR) – Wageningen, The Netherlands.

The varieties HZ 94 DTA 11 and RH00-386-2 are diploid, and the varieties RH4X-029-2 and RH4X-036-11 are tetraploid potato breeding clones. Although, all 4 clones have a wild potato species clone as a grandparent, they are all considered and treated as “normal” potatoes (*Solanum tuberosum*).

All potato samples are listed and detailed on Table 1.

Table 1. Field information of the eight potato samples used in this study for experimental validation of candidate reference genes.

Sample ID	Varieties	Parents	Grand parents	Year of Harvest	Time Post-harvest (days)
HZ-2	HZ94DTA11	RH90-012-2 x RH89-039-16	RH87-217-34 **x** TAR 24717-4 (S. tarijense)	2011	13
HZ94-2	HZ94DTA11		BC 1034 **x** SUH 2293	2012	28
RH00-2	RH00-386-2	RH97-649-11 x 96-2039-10	IVP92-057-17 **x** SPG 15458-B18 (S. spegazzinii)	2011	13
RH386-1	RH00-386-2		RH89-050-25 **x** RH89-035-38	2012	28
RH-029-2	RH4X-029-2	M 94-110-2 x FRIESLANDER	93-71-3 (S. hougasii) **x** W 72-38-720	2011	13
RH29-2	RH4X-029-2		GLORIA **x** 74 A 3	2012	28
RH036-1	RH4X-036-11	M 94-125-1 x FRESCO	BILDTSTAR **x** 93-114-5 (S. fendleri)	2011	13
RH36-1	RH4X-036-11		CEB 60-15-28 **x** PROVITA	2012	28

Samples preparation

Four average sized tubers were selected; of these, opposite eights were pooled to minimize variation effects in the tuber. Potato tubers were washed in water at room temperature dried with paper and chopped using a food processor into 1 cm^3 cubes. Potato cubes were immediately frozen in liquid N_2 to avoid tuber oxidation, packed in plastic bags and stored in an ultra-freezer at -80°C. Samples were sent to ZIRBUS Technology, Tiel, The Netherlands, for lyophilisation, milling and vacuum packaging. Potato powder was stored at room temperature until use.

RNA isolation and quality assessment

RNA was isolated from 0.5 g of each freeze-dried sample, according to the hexadecyltrimethylammonium bromide (CTAB) buffer lysis method, followed by chloroform/isoamyl alcohol extraction and overnight precipitation with lithium chloride (LiCl) proposed by van Dijk et al. (2009) [14], with some modifications, as follows. Lysis was performed with the extraction buffer pre-warmed to 60ºC before use; the chloroform/isoamyl alcohol extraction was repeated three times before the LiCl precipitation; and the final precipitation with 96% ethanol was performed with the tubes kept on ice and then centrifuged at 4ºC for 15 min at 14,000 g. Total RNA isolated was dissolved in 100 μL of 10 mM Tris (pH 7,0) and warmed to 65ºC for 10 min. Total RNA was stored at -80°C until use.

RNA purity and concentration were assessed by absorbance measurements using a Nanodrop 1000 instrument (Thermo Fisher Scientific, NanoDrop Technologies Wilmington, DE, USA). For integrity evaluation, 1 μg of RNA was migrated by electrophoresis (10 min at 80 V and 50 min at 100 V) in denaturing agarose gel (1%

agarose, 5% formamide, 1X TBE) stained with ethidium bromide. Gels were visualized in Gel Doc XR+ Systems (Bio-Rad Laboratories, Life Technologies Corporation, Carlsbad, CA, USA) and analyzed using Quantity One 1-D (Bio-Rad Laboratories).

Candidate gene selection and primer design

Candidate potato reference genes with stable expression levels in tubers were selected from a large collection of *RNAseq* profiles generated for 90 potato tubers grown under diverse range of growth conditions, locations, and growth year. Ten potato genes with more than 50 counts per million reads and with lowest interquartile range (IQR) were selected using R version 3.01 [15] for further evaluation with RT-qPCR (Table 2). Information about candidate genes was determined using Ensembl Plant Database (http://plants.ensembl.org/index.html).

Table 2. Candidate potato reference genes with more than 50 counts per million reads (highest expression), lowest inter quartile ranges (IQRs) and known functions, used for experimental validation.

Gene	IQR	Gene Code	Location	Transcript Code	Forward/Reverse primer	Amplicon (bp)	PCR efficiency (%)	r^2
eukaryotic translation initiation factor 3 subunit	12.84	PGSC0003DMG400009231	11:9004475-9012212	PGSC0003DMT400023872	3'GCGAAGATCCCAGTGAACAA5' 5'CAGCATCTTCACCAGCACTTA3'	123	93.6	0.998
dead-box atp-dependent rna helicase 39*	17.25	PGSC0003DMG400023195	12:54853693-54861561	PGSC0003DMT400059671 PGSC0003DMT400059672	3'TATGGGTGCCAAAGGGAAAG5' 5'CGTCTACTGAGAGAGACTCCAA3'	116	86	0.998
3-oxoacyl-(acyl-carrier protein) reductase	17.61	PGSC0003DMG401026981	6:52692660-52698742	PGSC0003DMT400069374	3'AGTTGAAGCTCCGGTTGTTATT5' 5'GTTCACAAGGACCTTACAACCA3'	100	96.9	0.998
importin subunit alpha	17.66	PGSC0003DMG400007289	6:100326-106781	PGSC0003DMT400018802 PGSC0003DMT400018803	3'ACCTCGATAAGAAGCTGGAGA5' 5'AGTTTCCGGAACTGTGTTGT3'	100	96	0.996
exocyst complex component sec3	17.75	PGSC0003DMG402015451	12:56757079-56759688	PGSC0003DMT400039945	3'GGAGCAGTATATCCAAGGACAA5' 5'AGGAACATTGTAGTGACAAACTTAG3'	75	90.3	0.995
ATCUL3/ATCUL3A/CUL3/CUL3A	17.80	PGSC0003DMG400001321	2:46264503-46268790	PGSC0003DMT400003337 PGSC0003DMT400003338 PGSC0003DMT400003339	3'GAGGACCGGTGAAGTGATAAAC5' 5'TCAGCCGAGACATCAAGAAAC3'	120	90	0.994
ubiquitin-associated/ts-nTS-N domain-containing protein	19.03	PGSC0003DMG402005949	6:54204271-54209252	PGSC0003DMT400015247 PGSC0003DMT400015248	3'TGAGAAGGCTGAAGAGACTTTG5' 5'GTAAGTTCTGGGTGGTGGTATT3'	131	101	0.996
C2	19.90	PGSC0003DMG400023712	10:57539858-57542161	PGSC0003DMT400060959	3'GGCCACTCAGATTGTCTCTATG5' 5'AGCTTTGCTTCTCCTCATACTC3'	118	88,1	0.998
dck/dgk-like deoxyribonucleoside kinase	20.49	PGSC0003DMG400009278	11:13898057-13903160	PGSC0003DMT400023985	3'GATATTGAAGCAAAGAGGCAGTATG5' 5'GATTGCCCTTAGGCTGTTCT3'	112	95.5	0.997
2-isopropylmalate synthase b	21.25	PGSC0003DMG400016337	6:39132151-39142170	PGSC0003DMT400042133	3'AAAGTGGCATCCATCAGGA5' 5'GACAATACCAGATTTATTAGCACGA3'	104	104.5	0.998

*Gene with alternative-splicing isoforms.

Solanum tuberosum genes, cDNA sequences, and exon-intron-exon junctions were also obtained from Ensembl Plant Database. All primers were designed using the Primer Quest tool from IDT DNA (http://www.idtdna.com/primerquest/Home/Index) with melting temperatures between 58°C and 62°C, GC contents from 45 to 65% and amplicon lengths ranging from 75 to 150 bp (Table 2). The Oligo Analyzer software from IDT DNA was also used to infer primer secondary structures (http://www.idtdna.com/analyzer/applications/oligoanalyzer/).

Since 4 out of the 10 candidate genes display alternative splicing (see Table 2), BLAST searches were performed, in order to design oligonucleotides complementary to a region of homology between the different transcripts of a given gene.

To determine PCR efficiencies, standard curves were constructed with four points in five-fold dilutions starting from a 1/5 cDNA concentration (1:5, 1:25, 1:125 and 1:625), according to Perini, et al. (2014) [16] and strongly suggested by Bustin, et al. (2009) [17]. Reaction efficiencies (E) and correlation coefficients (r^2) were estimated using *StepOne Software v.2.3* (*Life Technologies*), based on the slopes of the plots and the Cps (crossing points) versus log input of cDNA. E and r^2 values for each reaction performed are also presented in Table 2.

Complementary DNA synthesis

Each RNA sample was converted into cDNA in triplicates, as recommended by Bustin, et al. (2009) [17]. One microgram of total RNA was used for synthesis according to the manufacturer´s protocol, using the *iScript* cDNA Synthesis Kit (BIORAD). Specificity of the primers was checked for the 24 resulting cDNAs by end-

point PCR followed by electrophoresis in agarose gel and melting curve analysis. The cDNA samples were stored at -20°C until use.

Quantitative PCR (qPCR)

qPCR chain reactions were carried out in a *StepOne* Plus Real Time PCR System (Life Technologies) using SYBR Green (BIORAD; 1:10,000 dilution) for monitoring double strand DNA synthesis during qPCR. Reactions were performed in a 20 µL final volume with 10 µL of diluted cDNA (1:50), 0.2 µM of each primer, 0.1 mM of dNTPs, 0.25 units of Platinum Taq DNA Polymerase (Life Technologies) 1X Buffer Solution, and 1.5 mM of $MgCl_2$. Each cDNA was analyzed in four technical replicates, and negative controls were included. PCR cycling conditions were as follows: 94°C for 5 min, 40 cycles at 94°C for 15 seconds, 60°C for 10 seconds, 72°C for 15 seconds and 60°C for 35 seconds, and a final melting curve between 50 and 99°C (Δ0.3°C/s).

Gene expression stability analyses

All results from RT-qPCR were compared using *NormFinder* [9], *geNorm* – v. 3.5 [12] software and *BestKeeper* an Excel-based program [13].

The *NormFinder* algorithm ranks candidate genes based on their stability of expression and determines the best pair of genes for using as endogenous controls for the samples. *geNorm* calculates the average expression stability (M-value), defining the mean variation of a certain gene in relation to the other candidate genes. Following, *geNorm* determines the best number of reference genes through the pairwise variation estimation (V). Vandesompele et al. (2002) [12] suggested a V cut-off value of 0.15, below which the inclusion of an additional reference gene would not be required. Finally, *BestKeeper* estimates the reference genes with the greatest expression stability

by assessing a *BestKeeper* Index specific for each sample, which is calculated as the geometric mean of the Cp values of its candidate housekeeping genes [13].

Results

RT-qPCR analysis of candidate reference genes

In order to select a reliable set of reference genes for gene expression studies in potato edible tubers, RT-qPCR assays were performed for 10 candidate housekeeping genes. The correlation coefficients (r^2) for all resulting amplification curves were higher than 0.99, and all 10 primer pairs allowed amplification efficiencies (E) between 86 and 104.5% (Table 2). Considering the optimal PCR efficiency as 100%, which allows duplication of the whole target cDNA at each PCR cycle during the exponential phase, the observed efficiency values were considered acceptable; hence, the amplification products of each reaction were comparable to each other.

Primers for elongation factor 1 alpha, *18S rRNA* [18, 19], and actin [20] genes were initially included in the data set; however, they were discarded from the analysis due to unexpected amplification products.

Next, Cp values [21] were used to analyze the steady state mRNA levels of each candidate gene in eight different potato samples, showing a relative wide range of Cp values (Figure 1). In all tested samples, the lowest mean Cp value was observed for the gene eukaryotic translation initiation factor 3 subunit, followed by exocyst complex component sec3 (*SEC3*).

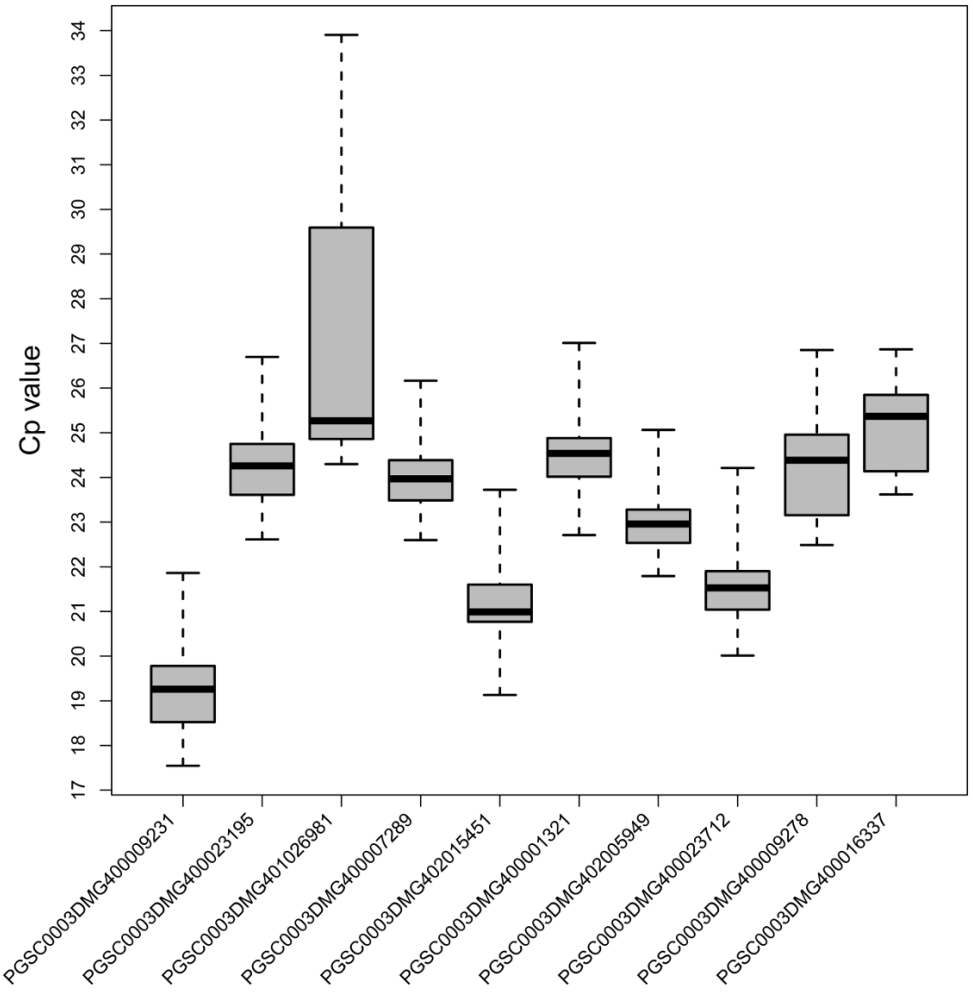


Figure 1. Expression profiles of the tested reference genes in raw Cp values for all 8 samples. Expression data are displayed as raw Cp values for each reference gene across all samples. The line denotes the median and boxes indicate the 25/75 percentiles.

Analyses of reference genes stability via *geNorm*, *NormFinder* and *BestKeeper* algorithms

Three different algorithms, *NormFinder*, *geNorm* and *BestKeeper*, were applied in computational assessment of gene expression stability in order to minimize potential biases intrinsic to each software.

NormFinder uses a mathematical modelling that allows an estimation of gene expression based in a variation of reference genes and in a subgroup of sample sets, considering the best genes those with the lowest stability value, with minimal intra and inter group variation, and indicate the best combination of gene pairs groups and subgroups [9]. Table 3 presents the ranking of the candidate reference genes according to their stability value for the samples, as determined by *NormFinder*. This algorithm identified *C2*, followed by *SEC3*, as the most stably expressed genes in all 8 different samples.

Table 3. Ranking of candidate reference genes according to the estimated values of stability of expression, as calculated by the *NormFinder* algorithm and M value calculated using *geNorm* estimated M–values, for the candidate reference genes.

Gene	**Gene Code**	**Rank by *NormFinder***	**Stability by *NormFinder***	**M-value by *geNorm***
C2	PGSC0003DMG400023712	1	0,010	0.647
exocyst complex component sec3	PGSC0003DMG402015451	2	0,015	0.716
ATCUL3/ATCUL3A/CUL3/CUL3 A	PGSC0003DMG400001321	3	0,016	0.658
dead-box atp-dependent rna helicase 39	PGSC0003DMG400023195	4	0,018	0.736
ubiquitin-associated /ts-n domain-containing protein	PGSC0003DMG402005949	5	0,019	0.756
importin subunit alpha	PGSC0003DMG400007289	6	0,019	0.709
dck/dgk-like deoxyribonucleoside kinase	PGSC0003DMG400009278	7	0,021	0.829
2-isopropylmalate synthase b	PGSC0003DMG400016337	8	0,021	0.763
eukaryotic translation initiation factor 3 subunit	PGSC0003DMG400009231	9	0,024	0.734
3-oxoacyl-(acyl-carrier protein) reductase	PGSC0003DMG401026981	10	0,084	3.222
-	-	Best combination of 2 genes (*SEC3* and *C2*)	0,010	-

*Lowest M value by *geNorm*.

Table 3 describes the ranking of candidate genes as assessed by *geNorm*. Also, pairwise variations (V) were calculated for obtaining the optimal number of normalization factors and the use of 2 primer pairs were definitively enough for this dataset. Figure 2 shows the M-values and pairwise variation (V) calculated by *geNorm* for all candidates and their best partners for the potato samples. The most stable candidate gene was *C2*, followed by *ATCUL3/ATCUL3A/CUL3/CUL3A* (*CUL3A*), with M-values above 0.7 for both. In agreement, the best gene pair consisted also of the *C2* and *CUL3A* (see Figure 2). Additionally, the V-values were below the established 0.15 threshold suggested by Vandesompele et al. (2002) [12], corroborating that the inclusion of an additional gene is not required for data normalization.

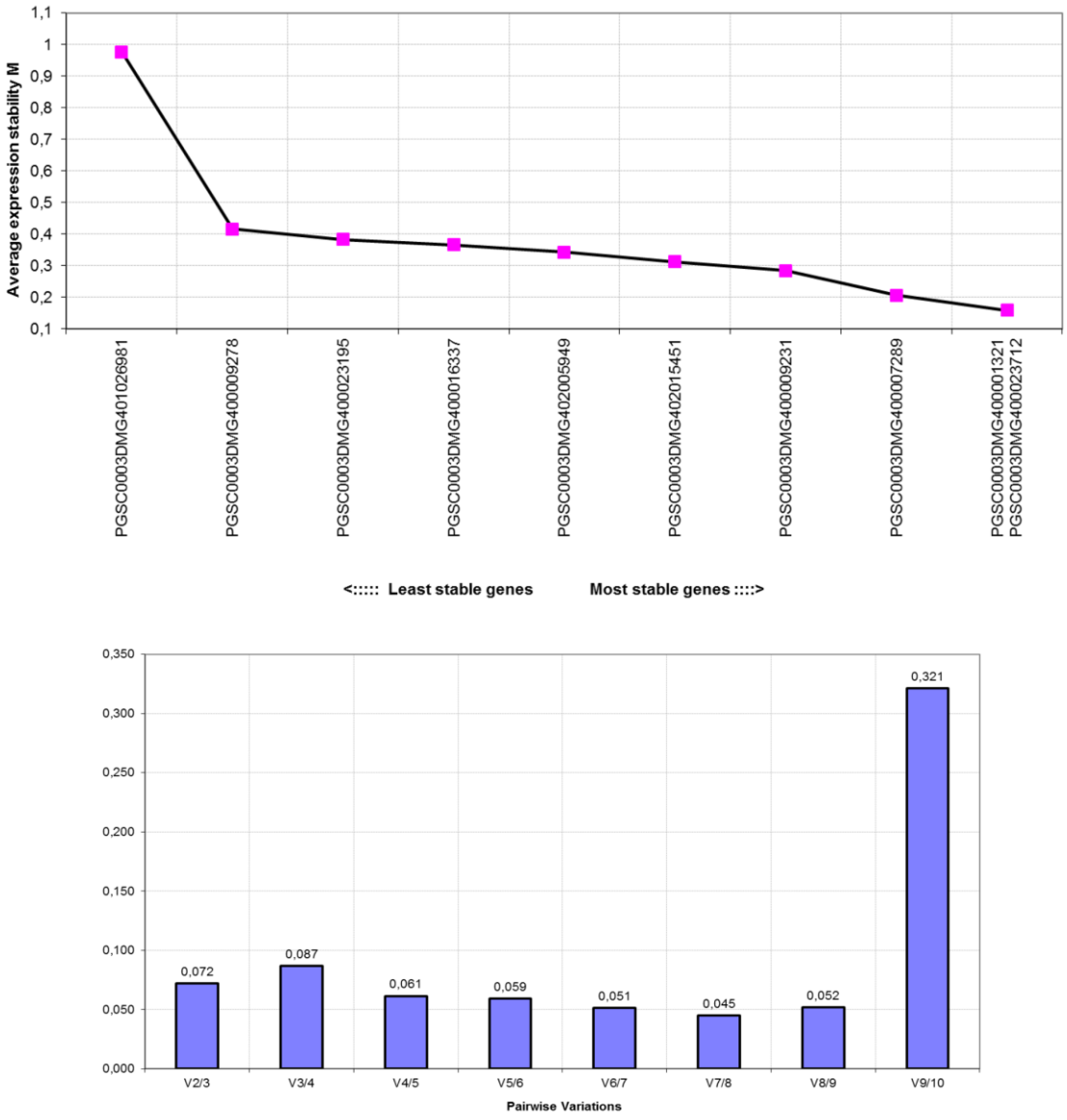


Figure 2. Average values of stability of gene expression for the selected reference genes assessed by *geNorm*. The plots indicate expression profiles and the determination of the optimal number of control genes for the eight samples**.**

According to the *BestKeeper* algorithm, the importin subunit alpha and ubiquitin-associated/ts-n domain-containing protein genes were the most stably expressed ones in *S. tuberosum* edible tubers across all eight samples, with a standard deviation (SD) of 0.66 for both candidates (Table 4). Only 3-oxoacyl-(acyl-carrier protein) reductase and

dck/dgk-like deoxyribonucleoside kinase (less stable) were considered to be inconsistent for *BestKeeper* quality parameters ([±Cp] > 1.00), with SDs 1.01 and 3.05, respectively.

Table 4. Descriptive statistics of candidate reference gene expression patterns, as measured by BestKeeper.

Gene	Gene Code	Geometric Mean [CP]	Arithmetic Mean [CP]	min [CP]	max [CP]	Standard Deviation [± CP][a]	Coefficient of Variation [% CP]	min [x-fold]	max [x-fold]	Standard Deviation [± x-fold][a]
eukaryotic translation initiation factor 3 subunit	PGSC0003DMG400009231	19.33	19.37	17.55	21.86	0.87	4.48	-3.45	5.77	1.83
dead-box ATP-dependent RNA helicase 39	PGSC0003DMG400023195	24.14	24.17	22.61	26.70	0.85	3.52	-2.89	5.87	1.80
3-oxoacyl-(acyl-carrier protein) reductase	PGSC0003DMG401026981	26.93	27.14	24.30	33.90	3.05[a]	11.24	-6.19	126.03	8.29
importin subunit alpha	PGSC0003DMG400007289	24.08	24.10	22.60	26.16	0.66[b]	2.72	-2.80	4.24	1.58
exocyst complex component sec3	PGSC0003DMG402015451	21.17	21.20	19.13	23.73	0.82	3.87	-4.12	5.86	1.77
ATCUL3/ATCUL3A/CUL3/CUL3A	PGSC0003DMG400001321	24.60	24.62	22.71	27.01	0.71	2.88	-3.71	5.32	1.63
ubiquitin-associated /ts-n domain-containing protein	PGSC0003DMG402005949	23.04	23.06	21.79	25.06	0.66[b]	2.87	-2.38	4.06	1.58
C2	PGSC0003DMG400023712	21.65	21.67	20.01	24.21	0.72	3.34	-3.11	5.91	1.65
dCK/dgk-like deoxyribonucleoside kinase	PGSC0003DMG400009278	24.31	24.34	22.49	26.85	1.01[a]	4.13	-3.54	5.83	2.01
2-isopropylmalate synthase b	PGSC0003DMG400016337	25.16	25.17	23.62	26.87	0.81	3.20	-2.90	3.28	1.75

[a]Genes with standard deviations [±Cp] > 1.00 are considered to have inconsistent expression patterns (3-oxoacyl-(Acyl-carrier protein) reductase and dck/dgk-like deoxyribonucleoside kinase).

[b]Based on the standard deviations (SDs), genes can be ranked from most stably (lowest SD, importin subunit alpha and ubiquitin-associated /ts-n domain-containing protein) to least stably (highest SD, 3-oxoacyl-(acyl-carrier protein) reductase) expressed.

Discussion

Recently, the quantification of RNA transcripts has become increasingly rapid and precise due to advances in gene quantification strategies. Associated with that, newly identified reference genes showing more stable expression patterns than traditional normalization genes have been reported by analyzing microarray and transcriptome sequencing data [22-23], and these high throughput techniques might be excellent potential sources of good candidate reference genes, as showed in the present work.

The accuracy of RT-qPCR results is highly dependent on a reliable normalization strategy that employs an invariant (i.e. stably expressed) reference gene [24-25]. For example, Nicot et al. (2005) and Lopez-Pardo, Ruiz de Galarreta and Ritter (2013) [18, 19] already performed this analysis testing several reference genes, including the elongation factor 1 alpha, with successfully results. Different of our data, the analysis was based on candidates chosen from the literature, not on gene expression experiments, such as microarray or *RNAseq*. In addition, Nicot et al. (2005) [18] did not use samples derived from edible tubers, but samples from a pool of all parts of the potato plant, both under biotic and abiotic stresses, without any distinction between different plant organs. Still, Lopez-Pardo, Ruiz de Galarreta and Ritter (2013) [19] used potato edible tubers as samples, but specifically under cold stress.

It has become clear that no single gene is constitutively expressed in all cell types and under all experimental conditions. For instance, the expression of the so-called ‘housekeeping’ genes, although constant under some experimental conditions, can vary considerably in other cases, implying that the stability of the proposed control

gene has to be tested before each new experiment [7, 10, 11, 16, 26-28]. Normalization with multiple reference genes is becoming a common practice and the gold standard for the technique, but reports that identify such genes in plant investigations are still limited [7, 16, 18, 28-41].

In the present work we evaluated by RT-qPCR 10 reference genes displaying relatively stable expression in edible tubers. Results obtained by *geNorm* and *NormFinder* were very similar to each other and more different than those obtained by *BestKeeper*. While the *geNorm* and *NormFinder* algorithms correct for inter-sample variations, *BestKeeper* does not regard differences in RNA quality or cDNA conversion efficiency across samples, which might influence the distinct findings observed here. Differently from the pairwise approach used by *geNorm*, *NormFinder* selects the top rank candidates with minimal variation rather than correlated expression, which is less influenced by co-regulated genes. Moreover, *NormFinder* takes into consideration systematic differences between sample subgroups [9, 12-13]. Hence, it is expected that the comparison of these three algorithms, as performed here, might provide a more reliable set of reference genes under specific experimental conditions. In this sense, our study provides evidence for the use of certain genes as normalizers in gene expression experiments for potato edible tubers, which is essential for obtaining accurate and reliable gene expression data profiles.

From our analysis three genes called *SEC3*, *CUL3A* and *C2* were selected as the best normalizers in gene expression of potato edible tubers. However, for each set of samples a validation are needed, and the best reference gene may be different, this could be observed in this present work that the rank for ten candidates to be reference genes for our 8 samples were not exactly the same order of the rank as *RNAseq* database.

The gene *SEC3* as well as *SEC5*, *SEC6*, *SEC8*, *SEC10*, *SEC15*, *EXO70*, and *EXO84* genes are part of an evolutionarily conserved octameric protein complex of secretory vesicles [42-43]. The *Arabidopsis* genome encodes single or multiple isoforms of all exocyst subunits [44], and homologous structural models of plant exocyst subunits indicate well conserved rod-like structural features, including putative phosphatidylinositol phosphate binding sites on SEC3 and EXO70 subunits. Through interaction with RAB and RHO GTPases, these proteins are known to be crucial for the proper targeting of the exocyst to membranes [45].

The *CUL3* gene is a constituent of ubiquitin ligase complexes [40]. In *Arabidopsis*, both CUL3A and CUL3B proteins interact with the RING-H2 finger protein RBX1 and with several members of plant BTB domain proteins [46-47], suggesting that they form similar CUL3-based E3 complexes. However, *cul3a* loss-of-function mutants are viable and fertile, exhibiting only slightly delayed flowering and reduced sensitivity to far red light [46]. This viability might be attributed to functional redundancy between the two *CUL3* genes in *Arabidopsis*, since disruption of both genes causes embryo lethality, indicating that CUL3 plays important roles during early steps of plant development [48-49]. Indeed, CUL3 seems to regulate the ethylene-independent distal root patterning and primary root growth by a novel ethylene-dependent pathway, thus implicating CUL3 in the division and organization of the root stem cell niche and columella root cap cells [50].

Finally, the gene that is referred to in the EnsemblPlants database and hence in this paper as *C2* is actually coding for a yet uncharacterized protein, designated M1C6S3_SOLTU in the UniProt database (http://www.uniprot.org/uniprot/M1C6S3). In this entry it is mentioned that the protein contains three C2 domains. The C2 domain

polypeptide is one of the most prevalent eukaryotic lipid-binding domains used in diverse functional contexts. This structural domain helps target proteins to cell membranes, and its typical version (PKC-C2) has a beta-sandwich conformation composed of 8 β-strands that co-ordinate two or three calcium ions, which bind in a cavity formed by the first and final loops of the domain on the membrane binding face [50-51].

Conclusions

Transcriptome data such as those obtained from microarray and *RNAseq* experiments provide an excellent resource of selecting candidate RT-qPCR reference genes. Here, through bioinformatics and experimental data, we show the selection and validation of ten putative reference genes for RT-qPCR studies in potato samples. The *C2*, *SEC3*, and *CUL3A* genes were found to be the most stable and suitable normalizers for potato edible tubers expression studies. In summary, these findings provide useful tools for the normalization of RT-qPCR experiments and will enable more accurate and reliable gene expression studies related to functional genomics in potato.

Acknowledgments

We would like to thanks M.Sc. Eduardo Preusser de Mattos for critical english review of the manuscript.

References

1. Mackay IM, Arden KE, Nitsche A. Real-time PCR in virology. Nucleic Acids Res. 2002; 30: 1292–1305.

2. Radoni A, Thulke S, Mackay IM, Landt O, Siegert W, Nitsche A. Guideline to reference gene selection for quantitative real-time PCR. Biochem Biophys Res Commun 2004; 313: 856–862.

3. Obrero A, Die JV, Roman B, Gomez P, Nadal S, González-Verdejo CI. Selection of reference genes for gene expression studies in Zucchini (*Cucurbita pepo*) using qPCR. J Agric Food Chem 2011; 59: 5402–5411.

4. Derveaux S, Vandersompele J, Hellemans J. How to do successful gene expression analysis using real-time PCR. Methods 2010; 50:227–230.

5. Pfaffl MW A new mathematical model for relative quantification in real time RT-PCR. Nucleic Acids Res 2001; 29: 2002-2007.

6. Chang E, Shi S, Liu J, Cheng T, Xue L, Yang X, et al. Selection of reference genes for quantitative gene expression studies in *Platycladus orientalis* (*Cupressaceae*) using real-time PCR. PLoS One 2012; 7: e33278.

7. Gutierrez L, Mauriat M, Pelloux J, Bellini C, Wuytswinkel UV. Towards a systematic validation of references in real time RT PCR. Plant Cell 2008; 20:1734–1735.

8. Bustin SA. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. J Mol Endocrinol 2002; 29:23–39.

9. Andersen CL, Jensen JL, Ørntoft TF Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res 2004; 64:5245–5250.

10. Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR. Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. Plant Physiol 2005; 139: 5–17.

11. Hruz T, Wyss M, Docquier M, Pfaffl MW, Masanetz S, Borghi L, et al. RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization. BMC Genomics 2011; 12: 156. doi: 10.1186/1471-2164-12-156.

12. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT–PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002; 3:1–12.

13. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: *BestKeeper* – Excel-based tool using pair-wise correlations. Biotechnol Lett 2004; 26: 509–515.

14. van Dijk JP, Cankar K, Scheffer SJ, Beenen HG, Shepher LVT, Stewart D, et al. Transcriptome analysis of potato tubers effects of different agricultural practices. J. Agric Food Chem 2009; 57: 1612–1623.

15. R Core Team [2014] R: A Language and Environment for Statistical Computing, http://www.R-project.org. Accessed 2013 Dec 3.

16. Perini P, Pasquali G, Pinheiro MM, de Oliveira PR D, Revers LF. Reference genes for transcriptional analysis of flowering and fruit ripening stages in apple (*Malus* X *domestica* Borkh.). Mol Breed 2014; 34: 829-842.

17. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem 2009; 55: 611-622.

18. Nicot N, Hausman JF, Hoffmann L, Evers D Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. J Exp Bot 2005; 56: 2907–2914.

19. Lopez-Pardo R, Ruiz de Galarreta J, Ritter E. Selection of housekeeping genes for qRT-PCR analysis in potato tubers under cold stress. Mol Breed 2013; 31: 39-45.

20. Venkatesh J, Yu JW, Park SW. Genome-wide analysis and expression profiling of the *Solanum tuberosum* aquaporins. Plant Physiol Biochem 2013; 73: 392-404.

21. Walker NJ A Technique whose time has come. Science 2002; 296: 557-559.

22. Demidenko NV, Logacheva MD, Penin AA. Selection and validation of reference genes for quantitative real-time PCR in buckwheat (*Fagopyrum esculentum*) based on transcriptome sequence data. PLoS One 2011; 6: e19434.

23. Park S, Kim Y, Huh J, Lee S, Kim S, Kim S-U, et al. Selection of new appropriate reference genes for *RT-qPCR* analysis via transcriptome sequencing of Cynomolgus Monkeys (*Macaca fascicularis*). PLoS One 2013, 8: e60758.

24. Kong Q, Yuan J, Gao L, Zhao S, Jiang W, Huang Y, et al. Identification of suitable reference genes for gene expression normalization in qRT-PCR analysis in watermelon. PLoS One 2014; 9: e90612.

25. McMillan M, Pereg L. Evaluation of reference genes for gene expression analysis using quantitative RT-PCR in *Azospirillum brasilense*. PLoS One 2014; 9: e98162.

26. Volkov RA, Panchuk II, Schöffl F. Heat-stress dependency and developmental modulation of gene expression: the potential of house-keeping genes as internal standards in mRNA expression profiling using real-time RT–PCR. J Exp Bot 2003; 54: 2343-2349.

27. Guénin S, Mauriat M, Pelloux J, Van Wuytswinkel O, Bellini C, Gutierrez L. Normalization of qRT–PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. J Exp Bot 2009; 60: 487–493.

28. Oliveira LA, Breton MC, Bastolla FM, Camargo SS, Margis R, Frazzon J, et al. Reference genes for the normalization of gene expression in *Eucalyptus* species. Plant Cell Physiol 2012; 53: 405–422.

29. Coker JS, Davies E Selection of candidate housekeeping controls in tomato plants using EST data. Biotechniques 2003; 35:740–748.

30. Brunner AM, Yakovlev IA, Strauss SH. Validating internal controls for quantitative plant gene expression studies. BMC Plant Biol 2004; 4:14–20.

31. Exposito-Rodrıguez M, Borges AA, Borges-Pérez A, Pérez JÁ Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. BMC Plant Biol 2008; 8:131–142.

32. Hu R, Fan C, Li H, Zhang Q, Fu YF Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT–PCR. BMC Mol Biol 2009; 10:93. doi:10.1186/1471-2199-10-93.

33. Løvdal T, Lillo C. Reference gene selection for quantitative real-time PCR normalization in tomato subjected to nitrogen, cold, and light stress. Anal Biochem 2009; 387: 238–242.

34. Paolacci AR, Tanzarella OA, Porceddu E, Ciaffi M. Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. BMC Mol Biol 2009; 10:11. doi:10.1186/1471-2199-10-11.

35. de Almeida MR, Ruedell CM, Ricachenevsky FK, Sperotto RA, Pasquali G, Fett-Neto, AG. Reference gene selection for quantitative reverse transcription–polymerase chain reaction normalization during *in vitro* adventitious rooting in *Eucalyptus globulus* Labill. BMC Mol Biol 2010; 11: 73. doi:10.1186/1471-2199-11-73.

36. Hong SM, Bahn SC, Lyu A, Jung HS, Ahn JH Identification and testing of superior reference genes for a starting pool of transcript normalization in *Arabidopsis*. Plant Cell Physiol 2010; 51: 1694–1706.

37. Kulcheski FR, Marcelino-Guimarães FC, Nepomuceno AL, Abdelnoor RV, Margis R. The use of microRNAs as reference genes for quantitative polymerase chain reaction in soybean. Anal Biochem 2010; 406: 185–192.

38. Maroufi A, Van Bockstaele E, De Loose M. Validation of reference genes for gene expression analysis in chicory (*Cichorium intybus*) using quantitative real-time PCR. BMC Mol Biol 2010; 11: 15. doi:10.1186/1471-2199-11-15.

39. Schmidt GW, Delaney SK. Stable internal reference genes for normalization of real-time RT–PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress. Mol Genet Genomics 2010; 283: 233–241.

40. Wang KLC, Yoshida H, Lurin C, Ecker JR. Regulation of ethylene gas biosynthesis by the *Arabidopsis* ETO1 protein. Nature 2004; 428: 945-950.

41. Hoenemann C, Hohe A. Selection of reference genes for normalization of quantitative real-time PCR in cell cultures of *Cyclamen persicum*. Electron J Biotechnol 2011; 14: 8. doi: 10.2225/vol14-issue1-fulltext-8.

42. TerBush DR, Maurice T, Roth D, Novick P. The exocyst is a multiprotein complex required for exocytosis in *Saccharomyces cerevisiae*. EMBO J 1996; 15:6483-6494.

43. Hsu SC, Ting AE, Hazuka CD, Davanger S, Kenny JW, Kee Y, Scheller RH. The mammalian brain rsec6/8 complex. Neuron 1996; 17:1209-1219.

44. Synek L, Schlager N, Elias M, Quentin M, Hauser MT, Zarsky V. AtEXO70A1, a member of a family of putative exocyst subunits specifically expanded in land plants, is important for polar growth and plant development. Plant J 2006; 48: 54–72.

45. Zărský V, Cvrčková F, Potockú M, Hála M. Exocytosis and cell polarity in plants — exocyst and recycling domains. New Phytol 2009; 183:255-272.

46. Dieterle M, Thomann A, Renou J-P, Parmentier Y, Cognat V, Lemonnier G, et al. Molecular and functional characterisation of *Arabidopsis* Cullin 3A. Plant J. 2005; 41: 386–399.

47. Weber H, Bernhardt A, Dieterle M, Hano P, Mutlu A. Arabidopsis AtCUL3a and AtCUL3b form complexes with members of the BTB/POZMATH protein family. Plant Physiol 2005; 137: 83–93.

48. Thomann A, Brukhin V, Dieterle M, Gheyeselinck J, Vantard M, Grossniklaus U, Genschik P. *Arabidopsis CUL3A* and *CUL3B* genes are essential for normal embryogenesis. Plant J 2005; 43: 437–448.

49. Figueroa P, Gusmaroli G, Serino G, Habashi J, Ma L, Shen Y, et al. Arabidopsis has two redundant Cullin3 proteins that are essential for embryo development and that interact with RBX1 and BTB proteins to form multisubunit E3 Ubiquitin Ligase complexes in vivo. Plant Cell 2005; 17: 1180–1195.

50. Zhang D, Aravind L. Identification of novel families and classification of the C2 domain superfamily elucidate the origin and evolution of membrane targeting activities in eukaryotes. Gene 2010; 469: 18–30.

51. Zhang D, Aravind L. Novel transglutaminase-like peptidase and C2 domains elucidate the structure, biogenesis and evolution of the ciliary compartment. Cell Cycle 2012; 11: 3861–75.

Capítulo 5. Artigo científico que será submetido a uma revista da área, onde o objetivo foi caracterizar os genes *glycoalkaloid metabolism* (*GAME*s) que fazem parte da biossíntese de glicoalcalóides em batatas. Pretendeu-se ainda, relacionar a expressão dos GAMEs com o conteúdo total de glicoalcalóides (α-chaconina e α-solanina) em diferentes genótipos de 8 amostras de tubérculos de batatas.

5. ARTIGO CIENTÍFICO 2

Characterization and transcriptional profile of glycoalkaloid metabolism (*GAME*) genes in *Solanum tuberosum* L. genotypes with contrasting total glycoalkaloid content

Roberta Fogliatto Mariot[1], Luisa Abruzzi de Oliveira[2], Marleen M. Voorhuijzen[3], Martijn Staats[3], Ronald C.B. Hutten[4], Jeroen P. van Dijk[3], Esther Kok[3] and Jeverson Frazzon[1*]

[1]Federal University of Rio Grande do Sul – Food Science Institute - ICTA/UFRGS, Brazil

[2]Federal University of Rio Grande do Sul – Pos-Graduate Program in Cellular and Molecular Biology – PPGBCM, Brazil

[3] RIKILT Food Safety Institute - Wageningen University, The Netherlands

[4]Plant Breeding, Wageningen UR, Wageningen, The Netherlands

*Corresponding author

E-mail: jeverson.frazzon@ufrgs.br (JF)

ABSTRACT

Before commercial release, new potato (*Solanum tuberosum*) varieties must be evaluated for content of toxic compounds such as glycoalkaloids (GAs), which are potent poisons. There are more than 80 GAs in cultivated potato, the most common (90%) being α-chaconine and α-solanine. Since GAs are thermoresistant, a safe upper limit of 200 mg GA/g fresh weight has been established for commercial potato tubers. GA biosynthesis proceeds via the cholesterol pathway, but the steps from cholesterol to α-chaconine and α-solanine are unknown but role has been proposed for glycoalkaloid metabolism (*GAME*) genes in the biosynthesis of GAs. Therefore, the goal of this study was to evaluate the relationship between total glycoalkaloid content (TGA) and the expression of *GAME* genes in potato tubers. In addition, promoter regions of *GAME* genes were investigated. Eight potato tubers were used, TGA was measured by HPLC-MS, and RT-qPCR were performed to determine the relative expression of *GAME* genes. The samples were divided into groups with high and low TGA content. We searched for *cis*-elements in the 2000 base pairs (bp) upstream of the transcription start site of each *GAME* gene using the PlantPAN database. There was a relationship between TGA content and the relative expression of *GAME* genes in potato tubers. The analysis of the putative promoter regions of the *GAME* genes showed the presence of several *cis*-elements related to the response of the plant to biotic and abiotic stresses and light. The many copies of these *cis*-elements confirmed that TGA content was related to these factors. These findings provide an important step towards understanding of TGA regulation and variation in potato tubers.

Key-words: *Solanum tuberosum*, Glycoalkaloids, *GAME* genes, gene expression.

INTRODUCTION

Potato (*Solanum tuberosum* L.) is the third most important human food crop in the world after rice and wheat. The introduction of new varieties of potatoes for consumption or food processing requires analysis not only for pro-nutritional substances, but also for toxic compounds such as glycoalkaloids (GAs). More than 80 different GAs have been identified in various potato species. The two major GAs in cultivated potatoes are α-chaconine (ca. 60%) and α-solanine (ca. 40%) (Friedman, Rayburn & Bantle, 1991; Blankemeyer et al., 1992; Haase, 2010), which usually comprise more than 90% of the total GA content (Friedman & McDonald, 1997). These are steroidal glycosides derived from aglycone solanidine that vary in their trisaccharide sugar content, and are important natural toxic components of potato tubers (Ostry, Ruprich, & Skarkova, 2010).

GAs are potent poisons with a lethal dose of 3–5 mg kg^{-1} body weight, similar to strychnine and arsenic (JECFA, 1992). The symptoms of GA poisoning include gastrointestinal disorders, confusion, hallucinations, partial paralysis, convulsions, coma, and even death (Smith, Roddick & Jones, 1996). As GAs are not destroyed during cooking or frying (Maga,1994), an upper limit of 20 mg per 100 g fresh weight of potato tubers has been established for safe release of new commercial cultivars (Valkonen et al., 1996). The greatest concentration of total glycoalkaloids (TGA) is in the skin and just beneath the surface (up to 1.5 mm deep), and within the eyes and damaged areas (Ostry et al., 2010; Friedman, 2006). TGA content varies greatly among individual tubers of the same variety of potato (Leri et al., 2011).

Despite the significance of GAs, studies on their biosynthetic pathway and the factors that regulate GA levels were scarce until couples years ago. It is known that GAs are produced in all parts of the potato plant including leaves, roots, tubers, and sprouts, and that biosynthesis proceeds via the cholesterol pathway in the following steps: acetate (C2) → mevalonate (C6) → isopentenyl pyrophosphate (C5) → squalene (C30) → cholesterol (C27) (Friedman, 2006; Ginzberg et al., 2012). However, many steps between cholesterol and formation of α-chaconine and α-solanine were unknown. Up to 2013, only enzymes in the final steps of the pathway from cholesterol to GA, such as solanidine galactosyltransferase (sgt1), solanidine glucosyltransferase (sgt2) (Pajerowska et al., 2005; McCue et al., 2006; Krits et al., 2007; Ginzberg et al., 2012) and β-solanine/β-chaconine rhamnosyl transferase (sgt3) (McCue et al., 2007) had been described. Itkin and collaborators, in an elegant work, described the glycoalkaloid metabolism (*GAME*) genes and proposed a GA biosynthesis pathway, summarized in Figure 1. Research has been more focused on tomato (*Solanum licopersicum*) rather than potato, and just few aspects are still not elucidated (for more information, please see Itkin et al. 2013).

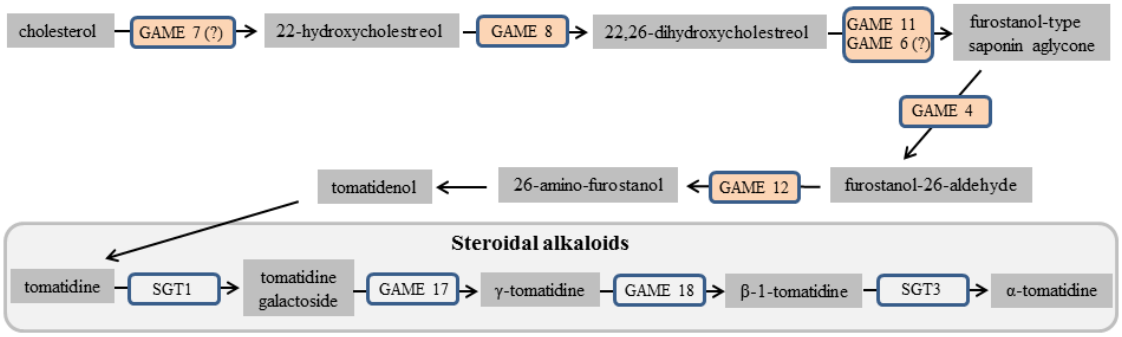


Figure 1: TGA biosynthetic pathway in Solanaceae plants. Suggested biosynthetic pathway from cholesterol adapted from Itkin et al. (2013). Glycoalkaloid metabolism (*GAME*) genes analyzed and involved in steroidal alkaloids biosynthesis are named inside orange boxes.

The goal of this study was to evaluate a possible relationship between TGA content and the expression of *GAME* genes in edible tubers of potatoes with different TGA content. In addition, we analyzed the putative promoter regions of the *GAME* genes to check the presence of *cis*-elements related to the response of the plant to biotic and abiotic stresses, since unpredictable variations in TGA levels can be triggered by light exposure and stress factors.

MATERIAL AND METHODS

Database search and in silico characterization

In order to identify the Glycoalkaloid Metabolism (*GAME*) genes, location on potato genome, gene and cDNA sequences and lengths, exon-intron-exons junctions, transcripts and all information needed for *Solanum tuberosum* was obtained on Ensembl Plant Database (http://plants.ensembl.org/index.html).

The putative promoter region from the 2,000-base pairs (bp) upstream of the transcription start site of each *GAME* gene was used to search for putative cis-elements. The analysis was performed using the Plant Pan – Plant Promoter Analysis Navigator (http://plantpan.mbc.nctu.edu.tw/index.php - Chang et al., 2008), and the *cis*-elements identified were classified based on their putative biological functions. The structures of the seven potato GAME-encoding genes selected to the functional analysis and their alternative transcripts were analyzed.

Ethics statements

The field experiments in both years (in this case normal yield trials) were performed on a trial field in the proximity of Wageningen (GPS coordinates: 51.95230, 5.63490) owned by Wageningen UR. No specific permission was required to carry out these potato trials.

Field experimental design

Eight potato edible tubers from four distinct genotypes, experimental lines, obtained in duplicates, one grown in 2011 and the other in 2012, with a post-harvest storage time of 13 and 28 days, and cultivated at Plant Breeding Sciences - Wageningen University and Research Center (WUR) – Wageningen, The Netherlands.

The varieties HZ 94 DTA 11 and RH00-386-2 are diploid, and the varieties RH4X-029-2 and RH4X-036-11 are tetraploid potato breeding clones. Although, all 4 clones have a wild potato species clone as a grandparent, they are all considered and treated as “normal” potatoes (*Solanum tuberosum*).

All potato samples are listed and detailed on Table 1.

Table 1. Field information and total glycoalkaloid content of the eight potato samples used in this study.

Sample ID	Varieties	Parents	Grand parents	Year of Harvest	Time Post-harvest (days)	TGA Content (µg/g of fresh weight)*
HZ-2	HZ94DTA11	RH90-012-2 x RH89-039-16	RH87-217-34 **x** TAR 24717-4 (*S. tarijense*)	2011	13	1030,6
HZ94-2	HZ94DTA11		BC 1034 **x** SUH 2293	2012	28	1804,3
RH00-2	RH00-386-2	RH97-649-11 x 96-2039-10	IVP92-057-17 **x** SPG 15458-B18 (*S. spegazzinii*)	2011	13	1346,5
RH386-1	RH00-386-2		RH89-050-25 **x** RH89-035-38	2012	28	1758,8
RH-029-2	RH4X-029-2	M 94-110-2 x FRIESLANDER	93-71-3 (*S. hougasii*) **x** W 72-38-720	2011	13	100,5
RH29-2	RH4X-029-2		GLORIA **x** 74 A 3	2012	28	361,3
RH036-1	RH4X-036-11	M 94-125-1 x FRESCO	BILDTSTAR **x** 93-114-5 (*S. fendleri*)	2011	13	286,5
RH36-1	RH4X-036-11		CEB 60-15-28 **x** PROVITA	2012	28	267,6

*sum of α-chaconina and α-solanina.

Samples preparation

Four average sized tubers were selected; of these, opposite eights were pooled to minimize variation effects in the tuber. Potato tubers were washed in water at room temperature dried with paper and chopped using a food processor into 1 cm^3 cubes. Potato cubes were immediately frozen in liquid N_2 to avoid tuber oxidation, packed in plastic bags and stored in an ultra-freezer at -80°C. Samples were sent to ZIRBUS Technology, Tiel, The Netherlands, for lyophilisation, milling and vacuum packaging. Potato powder was stored at room temperature until use.

TGA Analysis

The quantification of α-chaconine and α-solanine was performed using 40 ± 0.5 mg of each freeze-dried potato sample in duplicate, followed for addition extraction solution (acetonitrile/water/formic acid 50:50/0.2), homogenized with rotatory tumbler during 60 minutes and centrifuged at 4,000g during 15 minutes. After separation by centrifugation, 1 mL from supernatant were transferred to 2 different tubes and kept at mines 20ºC until analysis.

In the day of analysis, 50 µL from supernatant were transferred to a 4 new tubes already filled with a Acetonitrile and Water solution (50:50), called Sample Solution, followed quantity: tube 1 with 950 µL, tube 2 with 925 µL, tube 3 with 900 µL and tube 4. Each tube were analysed with Liquid Chromatography- Mass Spectrometry (LC-MS), using spiking solutions to compare the results.

RNA Isolation and Quality Assessment

RNA was isolated from 0.5 g of each freeze-dried sample, according to the hexadecyltrimethylammonium bromide (CTAB) buffer lysis method, followed by chloroform/isoamyl alcohol extraction and overnight precipitation with lithium chloride (LiCl) proposed by van Dijk et al. (2009), with some modifications, as follows. Lysis was performed with the extraction buffer pre-warmed to 60ºC before use; the chloroform/isoamyl alcohol extraction was repeated three times before the LiCl precipitation; and the final precipitation with 96% ethanol was performed with the tubes kept on ice and then centrifuged at 4ºC for 15 min at 14,000 g. Total RNA isolated was dissolved in 100 μL of 10 mM Tris (pH 7,0) and warmed to 65ºC for 10 min. Total RNA was stored at -80°C until use.

RNA purity and concentration were assessed by absorbance measurements using a NanoDrop™ Lite Spectrophotometer. For integrity evaluation, 1 μg of RNA was migrated by electrophoresis (50 min at 90 V) in denaturing agarose gel (1.5% agarose, 1X TBE) stained with ethidium bromide.

Complementary DNA Synthesis

For each RNA sample, cDNA were prepared in duplicate, recommendation by Bustin, et al, 2009. An initial treatment with the enzyme DNase was carried out to eliminate the contamination of genomic DNA molecules. For this, 1 μg the amount of total RNA from each sample was taken in a total of 10 μL of a solution containing RQ1 RNase-Free DNase (Promega), 10X reaction buffer [400mM Tris-HCl (pH 8), 100 mM MgSO4 and 10mM CaCl2] (Promega) and Diethyl pyrocarbonate (DEPC) (Sigma-

Aldrich) treated water. After incubation at 37 °C for 30 min, 1 µL Stop Solution [20mM EGTA (pH 8)] (Promega) was also added, with further heating at 65°C for 15 min. Initiating the synthesis of complementary DNA (cDNA), the second step consisted of adding 1 µL Oligo(dT) 15 Primer with a concentration of 0.5 µg/µl that hybridizes to the poly(A) tail of mRNA (Promega), followed for incubation at 70 °C for 10 min, and the third last step was characterized by the addition of a 25 µL reaction mix with 5 µL 5X reaction buffer [50 mM Tris-HCl (pH 8.3 @25°C), 75 mM KCl, 3 mM MgCl2 and 10 mM DTT], 1 µL 200 U M-MLV Reverse Transcriptase both from Promega, 1.25 µL of each dATP/dCTP/dGTP/dTTP 10 mM (Invitrogen), 0.65 of RNaseOUT™ Recombinant Ribonuclease Inhibitor (Invitrogen) and DEPC (Sigma-Aldrich) treated water to final volume, followed for heating at 40°C for 60 min. The cDNA were stored at -20°C until use.

PCR end point experiments (MyCycler™ Thermal Cycler, Bio-Rad, Hercules, CA) were performed for all samples and for all primers pair, including samples treated only with the enzyme DNase, to verify whether cDNA synthesis was realized correctly and to confirm the expected amplicon length for each primer pair.

Primers Design and Tests

All primers were design using Primer Quest tool from IDT DNA (http://www.idtdna.com/primerquest/Home/Index) with melting temperatures between 58°C and 62°C, GC content 45-65% and amplicon lengths 75-150 bp. In order to verify secondary structures of the primers, Oligo Analyzer software from IDT DNA were used (http://www.idtdna.com/analyzer/applications/oligoanalyzer/).

GAME6, *GAME7* and *GAME11*, and the reference *CUL3A* genes have alternative splicing, see on Table 2, blasts were done to check the common region between different transcripts to design the oligonucleotides. *GAME8a* and *GAME8b* genes are copies of each other, they are located on the same chromosome, for this reason, we designed one oligonucleotide for both genes, consequently the final relative expression were the sum of expression of the genes *GAME8a* and *GAME8b*.

All primers designed, and their respective amplicon for *GAME*s and reference genes used in this work for *S. tuberosum* edible tubers are described in Table 2.

In order to evaluate the PCR efficiency, a standard curve was constructed with four points in a fivefold dilution starting from a 1/5 cDNA concentration (1:5, 1:25, 1:125 and 1:625) in agreement with Perini, et al (2014) and suggested by Bustin, et al (2009). PCR efficiency and correlation coefficient (R^2) were calculated using *StepOne Software v.2.3* (Applied Biosystems), based on $E = 10^{-1/slope}$, slope of the plot, Cp (crossing point) versus log input of cDNA. E and R^2 for each PCR used in this present work are also presented in Table 2.

Table 2. glycoalkaloids metabolism genes, respective enzyme function, and their references genes: gene code, location on potato genome, transcripts (alternative splicing) code, primer sequences, amplicon length, PCR efficiency and correlation coefficient (R^2).

Gene	Enzyme Function in Glycoalkaloid Pathway	Gene Code	Location	Transcript Code	Forward/Reverse primer	Amplicom (bp)	PCR efficiency (%)	R^2
GAME4 (glycoalkaloid metabolism 4)	Furostanol-type aglycone oxidized to 26-aldehyde	PGSC0003DMG400024274	12: 5,853,429-5,858,832	PGSC0003DMT400062367	5' GATCAAATGTTTGTGGAGATTACTGC 3' 5' TTCGAGCCTTGAGTCCCTTAT 3'	100	89.0	0.995
GAME6 (glycoalkaloid metabolism 6)*	Oxidized at C22 and closure of the E-ring result in furostanol-type aglycone	PGSC0003DMG400011750	7: 41,884,815-41,887,155	PGSC0003DMT400030674	5' GCACGAGGTTCTGAGGTTAT 3' 5' CTGTATCGCGATGGATTGTTTG 3'	134	99.8	0.997
				PGSC0003DMT400030673				
				PGSC0003DMT400030672				
				PGSC0003DMT400030671				
GAME7 (glycoalkaloid metabolism 7)*	Cholesterol is hydroxylated at C22	PGSC0003DMG402012386	7: 52,651,099-52,659,776	PGSC0003DMT400032233	5' GACACAGGCCAGAGAAGAAG 3' 5' GGGATATAGCCGTAATGACTCG 3'	118	102.6	0.996
				PGSC0003DMT400032242				
				PGSC0003DMT400032232				
				PGSC0003DMT400032240				
				PGSC0003DMT400032234				
				PGSC0003DMT400032241				
GAME8a (glycoalkaloid metabolism 8a)#	Cholesterol is hydroxylated at C26	PGSC0003DMG400026594	6: 44,439,443-44,441,537	PGSC0003DMT400068388	5' ATAATGAACATGATCTTCCAAGAGGTG 3' 5' GTTCTTTACGGTGCTTCGCATAA 3'	75	91.0	0.995
GAME8b (glycoalkaloid metabolism 8b)#	Cholesterol is hydroxylated at C26	PGSC0003DMG400026586	6: 44,479,607-44,481,751	PGSC0003DMT400068373				
GAME11 (glycoalkaloid metabolism 11)*	22,26-dihydroxycholesterol is hydroxylated at C16	PGSC0003DMG400011751	7: 41,845,573-41,847,191	PGSC0003DMT400030676	5' CCTCCAAGATATAGGGAGGTTATTG 3' 5' AAACCTAGCCCTTCAGCTAAC 3'	95	98.7	0.998
				PGSC0003DMT400030677				
GAME12 (glycoalkaloid metabolism 12)	26-aldehyde substrate for transamination catalyzed (nucleophilic attack of the amino-nitrogen at C22)	PGSC0003DMG400024281	12: 5,776,516-5,780,741	PGSC0003DMT400062385	5' TGGGCTTCCTTCACTACATC 3' 5' CTGGCAAGTGATTGTTCCAATA 3'	92	105.6	0.996
C2[1]	NA	PGSC0003DMG400023712	10:57539858-57542161	PGSC0003DMT400060959	3'GGCCACTCAGATTGTCTCTATG5' 5'AGCTTTGCTTCTCCTCATACTC3'	118	88,1	0.998
SEC3 (exocyst complex component sec3)[1]	NA	PGSC0003DMG402015451	12:56757079-56759688	PGSC0003DMT400039945	3'GGAGCAGTATATCCAAGGACAA 5' 5'AGGAACATTGTAGTGACAAACTTAG 3'	75	90.3	0.995

CUL3A **(**ATCUL3/ATCUL3A/ CUL3/CUL3A)*[1]	NA	PGSC0003DMG400001321	2:46264503-46268790	PGSC0003DMT400003337 PGSC0003DMT400003338 PGSC0003DMT400003339	3' GAGGACCGGTGAAGTGATAAAC 5' 5' TCAGCCGAGACATCAAGAAAC 3'	120	90	0.994

*gene with alternative splicing; #same sequence and primer; different code and location on potato genome; [1]reference gene; NA- Not applied.

Quantitative PCR

All polymerase chain reactions were carried out on Step One Plus Real Time PCR System (Applied Biosystems). In order to monitor the double strand DNA synthesis during the qPCR running SYBR Green (BIORAD; 1:10,000 dilution) was used.

Reactions were performed in 20 µL volumes with 10 µL of diluted cDNA (1:100), 0.2 µM of primer par, 0.1 mM of dNTP, 0.25 units of TAQ Polimerase Platinum®Taq DNA Polymerase (Applied Biosystems), 1X Buffer Solution (Applied Biosystems), and 1.5 mM of $MgCl_2$ (Applied Biosystems). Each cDNA was analyzed in four technical replicates and negative controls were included for all primers. Following cycling were conducted: 94°C for 5 min, 40 times amplification cycles of 94°C for 15 seconds, 60°C for 10 seconds, 72°C for 15 seconds and 60°C for 35 seconds, and final melting curve between 50 and 99°C, delta is 0.3°C/s.

Data Analysis

Data were analyzed by transforming raw Cp, in Log 10 (raw Cp) values and finally into relative quantities using the $E^{-\Delta\Delta Cp}$ method (Livak & Schmittgen 2001). Therefore, all data were expressed relative to the expression of the most highly-expressed genes identified in a previous study (Mariot et al., 2015) were used as reference genes for data normalization. The three genes used were: *C2*, exocyst complex component sec3 (*SEC3*) and *CUL3A*.

$E^{-\Delta\Delta Cp}$ (difference between Ct value of the studied genes and Ct value of the each constitutive gene, generated in total 6 observations for each sample) values of sample (lowest TGA content – Sample ID: HZ94-2) and other 7 different potato

samples, for all *GAME* genes (*GAME4*, *GAME6*, *GAME7*, *GAME8ab*, *GAME11* and *GAME12*). For statistically significant difference ($P<0.05$) for sample HZ94-2, compared to other seven samples, the gene was considered to be more transcribed ($E^{-\Delta\Delta Ct}>1$) or less transcribed ($E^{-\Delta\Delta Ct}<1$).

Analysis of the variance by ANOVA, using SAS software (version 9.3), at a significance level of 5%, were done in order to verify whether there was statistically significant difference between samples with high and low TGA content. Further, *t* test, by SAS software (version 9.3), were performed in order to verify this set of samples are significantly different from each other.

RESULTS AND DISCUSSION

Characterization of exon–intron organization and promoter regions of GAME genes

The structure of *GAME* genes are shown in Figure 2 (Hu et al., 2014). *GAME* gene codes, isoform/transcript codes, location on the potato genome and their function in the TGA biosynthesis pathway are listed in Table 2.

GAME4 is located on chromosome 12, has one isoform of 1614 bp length, with eight exons and a length of 7804 bp. In contrast, *GAME6* has four isoforms, the first with only one exon, the second and the third with three different exons, and the last with five exons. *GAME6* is located on chromosome 7, and has a length of 4741 bp. *GAME7* is also located on chromosome 7, and has six isoforms: the first with five exons, the second with three exons, the third with only one exon, the fourth with five exons, the

fifth with two exons and the sixth with four exons. *GAME7* gene is longer than the other *GAME* genes, at 11078 bp. *GAME8a* and *GAME8b* genes are copies on the same chromosome; both genes are located on chromosome 6, and have one isoform with five exons. Although copies, *GAME8a* and *GAME8b* genes have differences in promoter regions and, they have a small difference in the gene length, 4495 bp and 4545 bp, respectively. *GAME11*, located on chromosome 7, is the shortest gene of the *GAME* genes, at 4019 bp, with two isoforms both with five exons. Finally, *GAME12* is located on chromosome 12, has one isoform with 11 exons and is more than 6626 bp long.

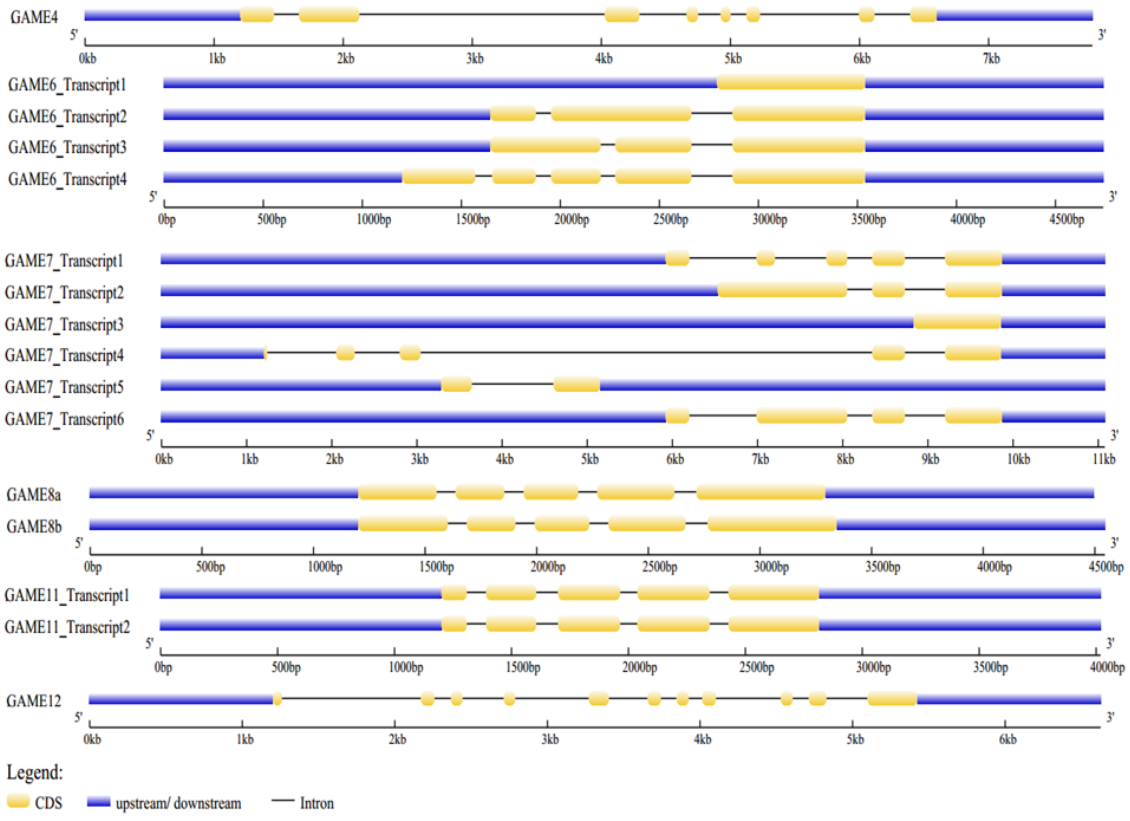


Figure 2: Exon–intron structure of *GAME* genes. The yellow boxes represent the exons, the lines connecting them represent the introns and blue boxes represent upstream and downstream sequences. Alternative transcripts are also shown. (Hu et al., 2014).

Large and often unpredictable variations in TGA levels can arise from differences in variety, locality, season, cultivation practice, and stress factors (OECD, 2002), resistance to viral and bacterial diseases (Austin et al., 1988; Pehu et al., 1990; Rokka et al., 1994), insect deterrence (Sanford et al., 1992, 1997), defense response against fungal pathogens (Fewell & Roddick, 1997; Percival, Karin & Dixon, 1998), harvest and postharvest treatments, such as drought (Bejarano et al., 2000), high temperature (Lafta & Lorenzen, 2000), light exposure of tubers (Dale et al., 1993; Percival, Karin & Dixon, 1994) and wounding (Bergenstråhle et al., 1992; Choi et al., 1994).

Analysis of the putative promoter regions (Chang et al., 2008) of the *GAME* genes identified enrichment of putative *cis*-elements that are associated with the response of the plant to abiotic and biotic stresses. Several regulatory *cis*-elements that are known to be responsive to a variety of stress factors and light were found in *GAME* gene promoters (Table 3). Three different *cis*-elements specific to *S. tuberosum* were also identified. Two light-regulated *cis*-elements were identified on all *GAME* genes: GATABOX contained up to 31 copies on *GAME12* in the promoter region; and GT1CONSENSUS contained up to 28 copies on *GAME6* in the promoter region. Some *cis*-elements related to stress were also found on all *GAME* genes. Three were responses to dehydration and water stress: MYB2CONSENSUSAT contained up to five copies on *GAME12* in the promoter region; MYBCORE presented four copies on *GAME6* and *GAME11*; and MYCCONSENSUSAT contained up to 26 copies on *GAME8a* (Table 3). A further *cis*-element common to all *GAME* genes was WBOXATNPR1 for disease response with up to seven copies in the *GAME8b* promoter region (Table 3). These results may indicate that these *GAME* genes respond to these stress conditions.

Table 3. Number of copies and biological function of the putative cis-elements related to stress and light regulated that were identified in the *GAME* genes promoters

Cis-element	**Function**	***GAME4***	***GAME6***	***GAME7***	***GAME8a***	***GAME8b***	***GAME11***	***GAME12***
ABRE-like	Dehydration response and low temperature response - water and thermo stress	1					1	1
ABREATCONSENSUS	Stress regulated response	1						1
ABRELATERD1	Dehydration response	4		2		1	1	2
ABREZMRAB28	Water stress response and cold tolerance							2
ACGTATERD1	Response to dehydration	12		6	6	4	4	4
ASF1MOTIFCAMV	Abiotic and biotic stress	1	1		3	2		1
CACGTGMOTIF	Defense related	8						4
CBF2	Water stress response							2
EMBP1TAEM	Various stress treatment	1						1
GATABOX	Light regulated	15	24	19	25	24	22	31
GBF1/2/3	Hypoxia							2
GT1CONSENSUS	Light regulated	26	28	22	23	26	19	18
HDZIP2ATATHB2	Light regulated				1	1		1
HSEs	Heat stress					2		
IBOX	Light regulated			2	2		2	2
LTRECOREATCOR15	Cold responsive - Thermo stress	1	1	2	1			1
MYB1AT	Dehydration response - water stress	3	2		4	2	15	3
MYB1LEPR	Defence-related						1	2
MYB2AT	Water Stress			1			3	2
MYB2CONSENSUSAT	Dehydration response - water stress	2	2	1	2	1	3	5
MYBCORE	Water stress	3	4	2	3	2	4	1
MYBST1	Unknown function - Specific for *S. tuberosum*	1	2	4	2		1	4
MYCATERD1	Dehydration - water stress		4	1	2	2		
MYCATRD22	Water stress – Dehydration			1	2	2		
MYCCONSENSUSAT	Dehydration response - water stress	14	12	1	26	14	12	2
P1BS	Phosphate starvation			4	4	4		4
SORLIP1AT	Light regulated	5			2		1	22
SORLIP2AT	Light regulated				1			
SREATMSD	Injury			1	1			
SURE2STPAT21	Unknown function - Specific for *S. tuberosum*	1				1		
TAAAGSTKST1	Unknown function - Specific for *S. tuberosum*	8	8	6	7	7	6	5
TBOXATGAPB	Light regulated	2	2	1		1	1	3
UP1ATMSD	Injury response	1			1			
WBOXATNPR1	Disease response	1	3	4	1	7	5	5

Analysis of a 2,000-bp region upstream of the transcription start site of each gene was performed using Plant Pan – Plant Promoter Analysis Navigator database (Chang et al., 2008).

TGA content

The eight potato samples from four different genotypes were divided into two groups according to their glycoalkaloid content. The first group was composed of genotypes HZ94DTA11 and RH00-386-2 which presented high TGA content with a ranged from 1030.6 to 1758.8 µg/g fresh weight. The second group comprised samples with low TGA content (RH4X-029-2 and RH4X-036-11 genotypes) with a ranged from 361.3 to 100.5 µg/g fresh weight in this group (Table 1). In all potato samples analysed the TGA content increases with increasing time post-harvest (13 and 28 days), in agreement with Fitzpatrick et al. (1977), Griffiths et al. (1997) and Friedman & McDonald (1999). The differences in the values showed in our experiments could be related to year of harvest (2011 and 2012), even being the location the same in both years. From four genotypes analysed here, the TGA content increased with time after harvest in three genotypes (HZ94DTA11, RH00-386-2 and RH4X-029-2), only the TGA content of genotype RH4X-036-11 did not change over time (Table 1).

Expression profile of GAME genes

The efficiencies (E) of primer pairs designed for all genes were evaluated using a standard curve with serial dilutions of S. tuberosum edible tuber cDNA. The correlation coefficient (R^2) for all resulting amplification curves was greater than 0.99, and all six primer pairs allowed amplification efficiencies between 88% and 105.6% (Table 2). Considering that the optimal PCR efficiency is 100%, when the whole target cDNA would be duplicated at every PCR cycle during the exponential phase, the

efficiency values obtained were therefore considered acceptable. Hence, the amplification products of each reaction were comparable.

The expression of GAME4, GAME6, GAME7, GAME8a, GAME8b, GAME11 and GAME12 was analyzed by RT-qPCR to verify differences in the expression pattern in edible potato tuber samples with different TGA content. To normalize the expression profile of GAME genes in different potato samples, the lowest TGA content sample was used to normalize the transcript levels in other samples for each gene.

As listed on Table 4, all GAME genes had higher relative expression level in samples with higher TGA content ($\alpha = 0.05$; t-test), as expected, except for GAME7. This exception may be explained by the fact that the enzyme encoded by the GAME7 gene catalyzes hydroxylation of cholesterol, the first step of the TGA pathway (Figure 1).

Table 4. Relative Expression of *GAME* genes in potato tubers with high and low content of TGA.

Gene	**Relative Expression**		**p**
	High TGA Samples	**Low TGA Samples**	
GAME4	1,027±0,017	1,001±0,005	<0,0001*
GAME6	1,027±0,023	0,988±0,019	<0,0001*
GAME7	0,987±0,026	0,976±0,041	0,2957
GAME8a* and *GAME8b	1,039±0,020	0,989±0,021	<0,0001*
GAME11	1,019±0,016	0,985±0,021	<0,0001*
GAME12	1,003±0,014	0,982±0,034	0,0058*

±SD standard deviation; n = 24 each group, 48 in total; * $p < 0.05$ (T test).

Glycoalkaloids are cited as toxic in the "Consensus Document on Compositional Considerations for New Varieties of Potatoes" (OECD, 2002) and information on GA content is needed for use during the regulatory assessment of new potato varieties. Due

to the significance of GAs for food safety, knowledge of expression and regulation of *GAME* genes that participate in the GA biosynthetic pathway in potato tubers may provide an important insight into GA prediction and control.

CONCLUSION

An analysis of the putative promoter regions of *GAME* genes found the presence of several *cis*-elements related to the response of potato plants to biotic and abiotic stresses and light, as well many copies of *cis*-elements on their promoter regions that confirmed that unpredictable variations in TGA levels could be related to these stressors. There was a relationship between TGA content and the expression of *GAME* genes in potato tubers. These findings provide an important step towards understanding of TGA regulation and variation in potato tubers.

REFERENCES

Austin S, Lojkowska E, Ehlenfeldt MK, Kelman A, Helgeson JP (1988) Fertile interspecic somatic hybrids of Solanum: a novel source of resistance to *Erwinia soft* rot. Phytopathology 78: 1216–1220.

Bejarano L, Mignolet E, Devaux A, Espinola N, Carrasco E, Larondelle Y (2000) Glycoalkaloids in potato tubers: the effect of variety and drought stress on the alpha-solanine and alpha-chaconine contents of potatoes. J Sci Food Agric 80: 2096–2100.

Bergenstråhle A, Tillberg E, Jonsson L (1992) Regulation of glycoalkaloid accumulation in potato tuber disks. J Plant Physiol 140: 269–275.

Blankemeyer JT, Stringer BK, Rayburn JR, Bantle, JA, Friedman M (1992) Effect of potato glycoalkaloids, α-chaconine and α-solanine on membrane potential of frog embryos. Journal of Agricultural and Food Chemistry 40: 2022–2025.

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, et al. (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem 55: 611-622.

Chang WC, Lee TY, Huang HD, Huang HY, Pan RL (2008) PlantPAN: Plant Promoter Analysis Navigator, for identifying combinatorial *cis*-regulatory elements with distance constraint in plant gene group. BMC Genomics 9: 561.

Choi D, Bostock RM, Avdiushko S, Hildebrand DF (1994) Lipid-derived signals that discriminate wound- and pathogen-responsive isoprenoid pathways in plants: methyl jasmonate and the fungal elicitor arachidonic acid induce different 3-hydroxy-3-methylglutaryl-coenzyme A reductase genes and antimicrobial isoprenoids in *Solanum tuberosum* L. Proc Nat Acad Sci USA 91: 2329– 2333.

Dale MFB, GriYths DW, Bain H, Todd D (1993) Glycoalkaloid increase in *Solanum tuberosum* on exposure to light. Ann Appl Biol 123: 411–418.

Fewell AM, Roddick JG (1997) Potato glycoalkaloid impairment of fungal development. Mycol Res 101: 597–603.

Fitzpatrick TJ, Herb SF, Osman SF, McDermott JA (1977) Potato glycoalkaloids: increases and variations of ratios in aged slices over prolonged storage. Am Potato J 54: 539-544.

Friedman M (2006) Reviews: Potato Glycoalkaloids and Metabolites: Roles in the Plant and in the Diet. J Agric Food Chem 54: 8655-8681.

Friedman M, Dao L (1992). Distribution of glycoalkaloids in potato plants and commercial potato products. Journal of Agricultural and Food Chemistry 40: 419–423.

Friedman M, McDonald GM (1997). Potato glycoalkaloids: Chemistry, analysis, safety and plant physiology. Critical Reviews in Plant Sciences 16: 55–132.

Friedman M, McDonald GM (1999) Postharvest changes in glycoalkaloid content of potatoes. Adv Exp Med Biol 459: 121-43.

Friedman M, Rayburn JR, Bantle JA (1991) Developmental toxicology of potato alkaloids in the frog embryo tetragenesis assay-Xenopus (FETAX). Food and Chemical Toxicology 29: 537–547.

Ginzberg I, Thippeswamy M, Fogelman E, Demirel U, Mweetwa AM, Tokuhisa J, Veilleux RE (2012) Induction of potato steroidal glycoalkaloid biosynthetic pathway by overexpression of cDNA encoding primary metabolism HMG-CoA reductase and squalene synthase. Planta 235: 1341–1353.

Griffiths DW, Bain H, Finlay M, Dale B (1997) The effect of low-temperature storage on the glycoalkaloid content of potato (*Solanum tuberosum*) tubers. J. Sci Food Agric 74: 301-307.

Haase NU (2010) Glycoalkaloid concentration in potato tubers related to storage and consumer offer. Potato Research 53: 297–307.

Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G (2014) GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics doi: 10.1093/bioinformatics/btu817.

Itkin M, Heinig U, Tzfadia O, Bhide AJ, Shinde B, et al. (2013) Biosynthesis of Antinutritional Alkaloids in *Solanaceous* Crops Is Mediated by Clustered Genes. Sciencexpress 341: 175-179.

JECFA. (1992). Evaluation of certain food additivies and naturally occurring toxicants. Thirty-ninth report of the Joint FAO/WHO Expert Committee on Food Additives. WHO Technical Report Series, no. 828. (pp. 31–33). Geneva: WHO.

Krits P, Fogelman E, Ginzberg I (2007) Potato steroidal glycoalkaloid levels and the expression of key isoprenoid metabolic genes. Planta 227: 143–150.

Lafta AM, Lorenzen JH (2000) Infuence of high temperature and reduced irradiance on glycoalkaloid levels in potato leaves. J Am Soc Hortic Sci 125: 563–566.

Leri F, Innocenti M, Andrenalli L, Vecchio V, Mulinacci N (2011) Rapid HPLC/DAD/MS method to determine phenolic acids, glycoalkaloids and anthocyanins in pigmented potatoes (*Solanum tuberosum* L.) and correlations with variety and geographical origin. Food Chemistry 125: 750–759.

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta deltaC(T)) method. Methods 25: 402–408.

Maga JA (1994) Glycoalkaloids in Solanaceae. Food Rev Int 10: 385–418.

Mariot RF, de Oliveira LA, Voorhuijzen MM, Staats M, Hutten RCB, et al. (2015) Selection of reference genes for transcriptional analysis of edible tubers of potato (*Solanum tuberosum* L.). PLos One doi: 10.1371/journal.pone.0120854. *In press*. Accepted on Mar 8 2015.

McCue KF, Allen PV, Shepherd LVT, Blake A, Maccree MM, Rockhold DR, et al. (2007) Potato glycosterol rhamnosyltransferase, the terminal step in triose side-chain biosynthesis. Phytochemistry 68: 327–334.

McCue KF, Allen PV, Shepherd LVT, Blake A, Whitworth J, Maccree MM, Rockhold DR, et al. (2006) The primary in vivo steroidal alkaloid glucosyltransferase from potato. Phytochemistry 67: 1590–1597.

Organisation For Economic Co-Operation And Development (2002). Environment Directorate Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology. Consensus Document on Compositional Considerations for New Varieties Of Potatoes: Key Food And Feed Nutrients, Anti-Nutrients And Toxicants. Series on the Safety of Novel Foods and Feeds, 4: 1-26.

Ostry V, Ruprich J, Skarkova J (2010) Glycoalkaloids in potato tubers: The effect of peeling and cooking in salted water. Acta Alimentaria 39: 130–135.

Pajerowska KM, Parker JE, Gebhardt C (2005) Potato Homologs of *Arabidopsis thaliana* Genes Functional in Defense Signaling—Identification, Genetic Mapping, and Molecular Cloning. Mol Plant Microbe Interact 18: 1107–1119.

Pehu E, Gibson RW, Jones MGK, Karp A (1990) Studies on the genetic basis of resistance to potato leaf roll virus, potato virus Y and potato virus X in *Solanum brevidens* using somatic hybrids of Solanum brevidens and *Solanum tuberosum*. Plant Sci 69: 95–101.

Percival GC, Karim MS, Dixon GR (1998) Influence of light enhanced glycoalkaloids on resistance of potato tubers to *Fusarium sulphureum* and *Fusarium solani* var *coeruleum*. Plant Pathol 47: 665–670.

Perini P, Pasquali G, Pinheiro MM, de Oliveira PRD, Revers LF (2014) Reference genes for transcriptional analysis of flowering and fruit ripening stages in apple (*Malus* X *domestica* Borkh.). Mol Breed 34: 829-842.

Rokka V-M, Xu Y-S, Kankila J, Kuusela A, Pulli S, Pehu E (1994) Identication of somatic hybrids of dihaploid *Solanum tuberosum* lines and *S. brevidens* by species speciWc RAPD patterns and assessment of disease resistance of the hybrids. Euphytica 80: 207.

Sanford LL, Deahl KL, Sinden SL, Ladd TL (1992) Glycoalkaloid contents in tubers from *Solanum tuberosum* populations selected for potato leafhopper resistance. Am Potato J 69: 693–703.

Sanford LL, Kobayashi RS, Deahl KL, Sinden SL (1997) Diploid and tetraploid *Solanum chacoense* genotypes that synthesize leptine glycoalkaloids and deter feeding by Colorado potato beetle. Am Potato J 74: 15–21.

Smith DB, Roddick JG, Jones JL (1996) Potato glycoalkaloids: some unanswered questions. Trends Food Sci Technol 7: 126–131.

Valkonen JPT, Keskitalo M, Vasara T, Pietilä L (1996) Potato glycoalkaloids: a burden or a blessing? Crit Rev Plant Sci 15: 1–20.

van Dijk JP, Cankar K, Scheffer SJ, Beenen HG, Shepher LVT, et al. (2009) Transcriptome Analysis of potato tubers effects of different agricultural practices. J Agric Food Chem 57: 1612–1623.

Capítulo 6. Artigo científico está sendo elaborado em colaboração com o grupo de pesquisa do RIKILT Food Safety Institute de Wageningen University e será submetido a uma revista da área. Neste capítulo, o objetivo foi propor um modelo de avaliação de segurança de novas variedades de *S. tuberosum* baseado em transcriptoma global e análise comparativa a variedades com histórico de segurança.

6. DESENVOLVIMENTO DE FERRAMENTAS ÔMICAS PARA AVALIAÇÃO DA SEGURANÇA DE NOVOS ALIMENTOS

6.1 INTRODUÇÃO

Este estudo é parte do projeto da cooperação internacional CAPES/Wageningen, intitulado “Desenvolvimento de ferramentas ômicas para avaliação da segurança de alimentos”, sob coordenação da Professora Dra. Ana Carolina Maisonnave Arisi, professora associada da Universidade Federal de Santa Catarina (UFSC). Este projeto faz parte de um projeto maior intitulado “*Development of classification models for food safety assessment using gene expression profiling*”, o qual é realizado na Universidade de Wageningen (WUR), no Instituto Rikilt de Segurança dos Alimentos. A WUR é uma universidade pública holandesa e possui institutos de pesquisa em agricultura do Ministério da Agricultura da Holanda. O foco da pesquisa cientifica é em ciências biológicas, social e comercial nas áreas de ciências agrárias e recursos naturais. No campo do conhecimento de ciências agrárias, a WUR é conhecida mundialmente pela sua excelência. A universidade pode ser considerada internacional, pois possui mais de 6.000 alunos de mais de 105 diferentes países. O RIKILT Food Safety Institute, parte da WUR, é um centro de pesquisa inovação independente em segurança e qualidade de alimentos. Além disso, é referência nacional e da União Europeia em análise de alimentos, especializado em detectar e identificar substâncias em alimentos e

ração animal, e determinação da funcionalidade e efeito dessas substâncias detectadas. *Novel Food and Agrichain*, grupo de pesquisa do RIKILT, coordenado pela Dra. Esther Kok, trabalha no desenvolvimento de modelos de classificação para acesso de risco de segurança de alimentos usando perfil de expressão gênica por NGS, com o intuito de se detectar e caracterizar o perfil de RNAs em diversas variedades e condições de cultivo de batata, com diferentes concentrações de glicoalcalóides. A tecnologia NGS foi escolhida devido ao seu alto poder descritivo e, principalmente, para embasar e fazer a mudança das análises de transcriptoma realizadas no Rikilt de microarranjos para tecnologia de NGS, como ferramenta para avaliação de segurança de alimentos. Durante o período do doutoramento sanduíche, foram ampliadas o número de amostras a serem utilizadas para a construção do modelo de classificação de risco de segurança alimentar analisados por NGS.

O objetivo deste projeto é avançar no campo da segurança de alimentos de novas culturas agrícolas através da construção de uma ferramenta multivariada de decisão, utilizando-se dados de análise transcriptômica (RNA-seq) de batatas contendo diferentes níveis de glicoalcalóides, abrangendo variabilidade genética e de condições ambientais. Comparar os perfis de expressão gênica das batatas 'inseguras' (com alto conteúdo de glicoalcalóides) com o banco de dados de referência. Assim pretende-se desenvolver um modelo, através de análise estatística multivariada, capaz de predizer se um alimento poderá ser considerado seguro.

6.2 MATERIAL E MÉTODOS

6.2.1 Variedades de batatas

Para a construção do banco de dados, um total de 90 amostras de batatas foram utilizadas, sendo elas com 21 diferentes genótipos, cultivados em diferentes condições, tipos de solo e países, presença/ausência de *Phytophtora infestans*, bem como colhidos em diferentes anos. A descrição de

cada um dos estudos realizados com tubérculos de batatas encontra-se aabaixo:

Estudo BligthMOP: 4 amostras, genótipos Lady Balfour e Sante cultivadas na fazenda experimental de Nafferton no Reino Unido e colhidas em 2004, objetivando a gestão da praga *Phytophtora infestans* na agricultura orgânica de batatas (VAN DIJK et al., 2009).

Estudo QLIFs: 3 amostras do genótipo Santé foram cultivadas sob adubação orgânica e convencional e cada um combinado com proteção de cultivo orgânico ou convencional, dando quatro tratamentos diferentes, sendo colhidas em 2005 (VAN DIJK et al, 2012).

Estudo One-class classifier "microarranjo": 11 amostras, genótipos Biogold, Fontane, Innovator, Lady Rosetta e Maris Piper. Os tubérculos foram cultivados no Plant Research International (PRI), Wageningen University and Research Center (WUR), Países Baixos, e colhidos em 2010. Cada genótipo foi plantado em dois lotes contendo diferentes solos; um lote foi cultivado somente em argila e o outro somente em areia, com duas parcelas individuais de cada variedade, e em cada parcela, quatro plantas individuais. A exceção foi Maris Piper que teve somente uma parcela cultivada em cada solo, e cada parcela contendo três plantas. Foi obtido um lote total de 18 amostras. Os tubérculos foram armazenados por 7 dias ao abrigo da luz e em temperatura ambiente (VAN DIJK et al, 2014).

Estudo One-class classifier "next generation sequencing": 44 amostras, genótipos Bintije e Nicola e 10 variedades cruzadas com espécies selvagens (25% "sangue" selvagem), com expectativa de altos índices de glicoalcalóides para algumas delas, todas em duplicatas: *S. berthaultii, S. spegazzinii, S. stoloniferum, S. hougasii, S. brachycarpum, S. tarijense, S. fendleri, S. pampasense, S. venturi, S. microdontum*. Cinco amostras de Bintije e Nicola foram analisadas em diferentes tempos após colheitas com 1, 2, 4, 6 e 8

semanas até a liofilização. Estas séries para Bintije e Nicola, bem como os híbridos de *S. pampasense, S. venturi*, também foram cultivadas em campo contendo a praga *Phytophtora infestans*. Os tubérculos foram cultivados no PRI, na WUR e colhidos em 2011.

Estudo One-class classifier "next generation sequencing": 28 amostras, genótipos Bintje, Nicola, Innovator e Lady Rosetta e 10 variedades experimentais, todas em duplicatas. Todas as amostras foram analisadas em 28 dias após a colheita. Os tubérculos foram cultivados no PRI, na WUR e colhidos em 2012.

Detalhes sobre as 90 amostras de tubérculos de batatas utilizados estão listados na Tabela 1 a seguir:

Tabela 1. Informações de 90 amostras de tubérculos de batatas utilizadas para construir o banco de dados de *RNASeq*.

Genótipo	Código da Amostra	Armazenamento pós-colheita (dias)	*Phytophthora infestans*	Tipo de variedade	Ano de colheita	Local de cultivo
94-2031-01	94-1	13	ausência	experimental	2011	Wageningen
94-2031-01	94-2	13	ausência	experimental	2011	Wageningen
94-2031-01	94-20-1	28	ausência	experimental	2012	Wageningen
94-2031-01	94-20-1-2	28	ausência	experimental	2012	Wageningen
Biogold	BZ1	7	ausência	comercial	2010	Wageningen
Biogold	BK2	7	ausência	comercial	2010	Wageningen
Sante	BM274	28	ausência	comercial	2004	Reino Unido
Sante	BM282	28	ausência	comercial	2004	Reino Unido
Sante	QL1158	32	ausência	comercial	2005	Reino Unido
Sante	QL1159	32	ausência	comercial	2005	Reino Unido
Sante	QL1160	32	ausência	comercial	2005	Reino Unido
LadyBalfour	BM276	28	ausência	comercial	2004	Reino Unido
LadyBalfour	BM287	28	ausência	comercial	2004	Reino Unido

Bintje	B-OBP-1-t4	41	ausência	comercial	2011	Wageningen
Bintje	B-OBP-2-t3	27	ausência	comercial	2011	Wageningen
Bintje	B-OBP-2-t5	55	ausência	comercial	2011	Wageningen
Bintje	B-OBP-2t1	6	ausência	comercial	2011	Wageningen
Bintje	B-OBPt2-2	13	ausência	comercial	2011	Wageningen
Bintje	BP-1t1	10	presença	comercial	2011	Wageningen
Bintje	BP-1-t3	31	presença	comercial	2011	Wageningen
Bintje	BP-1-t4	41	presença	comercial	2011	Wageningen
Bintje	BP-2t2	17	presença	comercial	2011	Wageningen
Bintje	BP-2-t5	55	presença	comercial	2011	Wageningen
Bintje	B-1	28	ausência	comercial	2012	Wageningen
Bintje	B-2	28	ausência	comercial	2012	Wageningen
Fontane	FK2	7	ausência	comercial	2010	Wageningen
Fontane	FZ1	7	ausência	comercial	2010	Wageningen
HZ94DTA11	HZ-1	13	ausência	experimental	2011	Wageningen
HZ94DTA11	HZ-2	13	ausência	experimental	2011	Wageningen
HZ94DTA11	HZ94-1	28	ausência	experimental	2012	Wageningen
HZ94DTA11	HZ94-2	28	ausência	experimental	2012	Wageningen
Innovator	IZ2	7	ausência	comercial	2010	Wageningen
Innovator	IK1	7	ausência	comercial	2010	Wageningen
Innovator	IN-1	28	ausência	comercial	2012	Wageningen
Innovator	IN-2	28	ausência	comercial	2012	Wageningen
IVP06-161-16	IP-1	10	presença	experimental	2011	Wageningen
IVP06-161-16	IP-2	10	presença	experimental	2011	Wageningen
IVP06-161-16	IVP06-1	13	ausência	experimental	2011	Wageningen
IVP06-161-16	IVP06-2	13	ausência	experimental	2011	Wageningen

IVP06-161-16	IV-161-1	28	ausência	experimental	2012	Wageningen
IVP06-161-16	IV-161-2	28	ausência	experimental	2012	Wageningen
IVP4X-154-4	IVP4X-1	13	ausência	experimental	2011	Wageningen
IVP4X-154-4	IVP4X-2	13	ausência	experimental	2011	Wageningen
LadyRosetta	LK1	7	ausência	comercial	2010	Wageningen
LadyRosetta	LK2	7	ausência	comercial	2010	Wageningen
LadyRosetta	LZ2	7	ausência	comercial	2010	Wageningen
LadyRosetta	LR-1	28	ausência	comercial	2012	Wageningen
LadyRosetta	LR-2	28	ausência	comercial	2012	Wageningen
MarisPiper	MK	7	ausência	comercial	2010	Wageningen
MarisPiper	MZ	7	ausência	comercial	2010	Wageningen
Nicola	N-OBP-1-t4	41	ausência	comercial	2011	Wageningen
Nicola	N-OBP-1-t5	55	ausência	comercial	2011	Wageningen
Nicola	N-OBP-2-t3	27	ausência	comercial	2011	Wageningen
Nicola	N-OBP-1t1	6	ausência	comercial	2011	Wageningen
Nicola	N-OBPt2-2	13	ausência	comercial	2011	Wageningen
Nicola	NP-1t2	17	presença	comercial	2011	Wageningen
Nicola	NP-1-t3	31	presença	comercial	2011	Wageningen
Nicola	NP-1-t5	55	presença	comercial	2011	Wageningen
Nicola	NP-2-t1	10	presença	comercial	2011	Wageningen
Nicola	NP-2-t4	41	presença	comercial	2011	Wageningen
Nicola	Nic-1	28	ausência	comercial	2012	Wageningen
Nicola	Nic-2	28	ausência	comercial	2012	Wageningen
RH00-386-2	RH00-1	13	ausência	experimental	2011	Wageningen
RH00-386-2	RH00-2	13	ausência	experimental	2011	Wageningen
RH00-386-2	RH386-1	28	ausência	experimental	2012	Wageningen

RH00-386-2	RH386-2	28	ausência	experimental	2012	Wageningen
RH4X-029-2	RH-029-1	13	ausência	experimental	2011	Wageningen
RH4X-029-2	RH-029-2	13	ausência	experimental	2011	Wageningen
RH4X-029-2	RH29-1Replicate	28	ausência	experimental	2012	Wageningen
RH4X-029-2	RH29-2Replicate	28	ausência	experimental	2012	Wageningen
RH4X-036-11	RH036-1	13	ausência	experimental	2011	Wageningen
RH4X-036-11	RH036-2	13	ausência	experimental	2011	Wageningen
RH4X-036-11	RH36-1	28	ausência	experimental	2012	Wageningen
RH4X-036-11	RH36-2	28	ausência	experimental	2012	Wageningen
RH4X-054-3	RH-054-1	13	ausência	experimental	2011	Wageningen
RH4X-054-3	RH-054-2	13	ausência	experimental	2011	Wageningen
RH4X-054-3	RH4X-1	28	ausência	experimental	2012	Wageningen
RH4X-054-3	RH4X-2	28	ausência	experimental	2012	Wageningen
RH4X-753-3	RH753-1	13	ausência	experimental	2011	Wageningen
RH4X-753-3	RH753-2	13	ausência	experimental	2011	Wageningen
RH4X-753-3	RP-1	10	presença	experimental	2011	Wageningen
RH4X-753-3	RP-2	10	presença	experimental	2011	Wageningen
RH90-038-21	RH90-1	13	ausência	experimental	2011	Wageningen
RH90-038-21	RH90-2	13	ausência	experimental	2011	Wageningen
RH90-038-21	RH90-1Replicate	28	ausência	experimental	2012	Wageningen
RH90-038-21	RH90-2Replicate	28	ausência	experimental	2012	Wageningen
IVP4X-159-6	IV159-1	28	ausência	experimental	2012	Wageningen
IVP4X-159-6	IV159-2	28	ausência	experimental	2012	Wageningen
RH4X-753-3	RH753-1Replicate	28	ausência	experimental	2012	Wageningen
RH4X-753-3	RH753-2Replicate	28	ausência	experimental	2012	Wageningen

Fonte: Autor (2015).

6.2.2 Preparo das Amostras

Cada tubérculo foi lavado com água corrente e seco com papel, após o corte em um multiprocessador, as batatas foram imediatamente imersas em nitrogênio líquido para o seu congelamento. A desidratação por liofilização, moagem e empacotamento a vácuo das amostras de batata foram realizadas pela empresa ZIRBUS Technology, Tiel, Reino da Holanda. As amostras foram armazenadas a -20°C.

6.2.3 Extração de glicoalcalóides totais

A quantificação de α-solanina e α-chaconina foram feitas a partir de 40 ± 0,5 mg das batatas liofilizadas e moídas em duplicata (2 tubos por amostra), seguida por adição de solução de extração (acetonitrila/água/ácido fórmico 50:50/0.2), homogeneização por tambor rotatório durante 60 minutos e centrifugação dos tubos a 4000 g por 15 min.

Após a separação por centrifugação, 1mL do sobrenadante foram transferidos para 2 diferentes tubos e mantidos a -20ºC até o dia da análise. No dia da análise, 50µL do sobrenadante de cada um dos tubos foram transferidos a 4 diferentes tubos contendo a “solução de amostras” (Acetonitrila e Água (50:50)), nas seguintes quantidades:

- Tubo 1: 950 µL Solução de amostra
- Tubo 2: 925 µL Solução de amostra
- Tubo 3: 900 µL Solução de amostra
- Tubo 4: 850 µL Solução de amostra

Cada um dos tubos foi analisado em Cromatografia Liquida de Alta Eficiência acoplada a Espectrometria de Massa (CLAE-MS), utilizando soluções padrões (contendo quantidades já conhecidas de α-solanina e α-chaconina) para comparação dos resultados.

6.2.4 Isolamento de RNA total

O RNA das amostras foi isolado a partir de 0,5 g de batatas liofilizadas trituradas, conforme descrito por Lehesranta et al. (2007), baseado no método CTAB com consecutiva extração clorofórmio/álcool isoamílico seguida por precipitação com LiCl. As seguintes modificações no método foram feitas: o tampão de extração foi pré-aquecido a 60°C antes de ser utilizado, a extração clorofórmio/álcool isoamílico foi repetida três vezes antes da precipitação com LiCl e a precipitação final com etanol 96% foi feita com os tubos mantidos em gelo e centrifugação a 4°C por 15 min a 14000g. O RNA foi dissolvido em 100 µL de Tris 10 mM (pH 7) e aquecido a 65°C por 10 min para a sua solubilização.

A concentração e pureza dos RNAs isolados foram avaliados pela medição dos picos de absorbância a 230, 260 e 280 nm (razão A260/A280 entre 1,8 e 2,0 e A260/A230 em torno de 2,0) em Nanodrop 1000 Instrument (NanoDrop Technologies). Para avaliação da degradação, 1 µg de RNA foi migrado por eletroforese (60 min, 80 V) em gel de agarose desnaturante (1% agarose, 5% formamida, TBE 1X) corado com brometo de etídeo. Os géis foram visualizados em Fotodocumentador Bio-Rad e analisados com o auxílio do software Quantity One 1-D (Bio-Rad). Amostras com a relação quantidade de rRNA 18S e rRNA total acima de 40% foram consideradas adequadas para as análises de sequenciamento de próxima geração. A avaliação da pureza e qualidade do RNA foi realizada em todas as amostras.

6.2.5 Sequenciamento do RNA mensageiro (mRNA) e Transcriptoma

Após a verificação da qualidade dos RNAs, 20 microgramas dos RNAs foram enviados para a empresa BASECLEAR (http://www.baseclear.com/dna-sequencing/next-gen-sequencing/transcriptome-analysis/), situada em Leiden na Holanda. A partir do RNA total, o isolamento do mRNA foi realizado através de um oligo dT, e um DNA complementar (cDNA) foi gerada através de uma transcriptase reversa reação de cadeia da polimerase (RT PCR), para a construção das bibliotecas. As bibliotecas de cDNA foram clivadas em

sequencias de 51 pares de base (pb) e sequenciadas (*Illumina HiSeq2000 paired-end sequencing*).

6.2.6 Análise de dados dos RNA-seq

As sequencias de nucleotídeos oriundas do sequenciamento do mRNA, descrita acima, foi analisada em etapas, utilizando as seguintes ferramentas da bioinformática:

Etapa I) **FASTX-Toolkit** – é uma coleção de ferramentas para leitura de pequenos fragmentos (*short-reads - FASTA/FASTQ files preprocessing*). RNA-seq geralmente produz pequenos fragmentos de leitura, que neste presente projeto foram ajustados e lidos por esta ferramenta. Esta análise foi realizada em 2 etapas, na primeira os *reads* foram lidos, e diversos gráficos foram gerados indicando a qualidade do sequenciamento do mRNA. Na segunda etapa foi realizada uma filtragem dos reads com baixa qualidade (inferior a 28%) e remoção dos adaptadores utilizados no sequenciamento da tecnologia ILLUMINA. Para as etapas seguintes, somente as sequencias com qualidade suficiente foram utilizadas.

Etapa II) Botwie – o principal objetivo desta ferramenta é achar em que parte do genoma o *read* ocorre, alinhamento das sequencias. É considerada uma ferramenta rápida inclusive para grandes genomas, como a *S. tuberosum*. O genoma da *S. tuberosum* Grupo phureja, mais atual, depositado no *Potato Genome Sequence Consortium*, foi utilizado para alinhar as sequencias obtida das 90 amostras, além disso, foi possível construir um novo referencial utilizando as informações das sequencias analisadas, incluindo os novos genes encontrados.

Etapa III) TopHat – é uma ferramenta utilizada para mapear os *splicings* entre os pequenos fragmentos de leitura do RNA-seq. A partir do alinhamento dos *reads* gerados no *RNAseq* com o genoma referência, realizado na Etapa II, o TopHat identificou os *splicings* entre os *exons* das 90 amostras. O

mapeamento dos *splicings* utilizou o genoma de referência da batata *Solanum tuberosum* Grupo phureja (genes referências: mitocôndria, nuclear e de cloroplasto).

Etapa IV) Cufflinks – esta ferramenta "monta" os transcritos e, torna possível a determinação da regulação diferencial no RNA-seq, permitindo o controle de qualidade do pós-alinhamento.

Etapa V) HTSeq-count – este software é usado para diversas análises em RNA-seq. Neste presente projeto ele foi utilizado para contabilizar quantos *reads* alinharam com cada gene, gerando o nível de transcrição. No caso de determinação de expressão no RNA-seq, os genes são considerados união dos exons, onde cada exon pode ser considerado como um gene, por exemplo, no caso de checar *splicings* alternativos.

6.3 RESULTADOS PARCIAIS

Os resultados obtidos para as 28 amostras de batatas colhidas em 2012 e processadas durante o estágio doutoral serão detalhados a seguir.

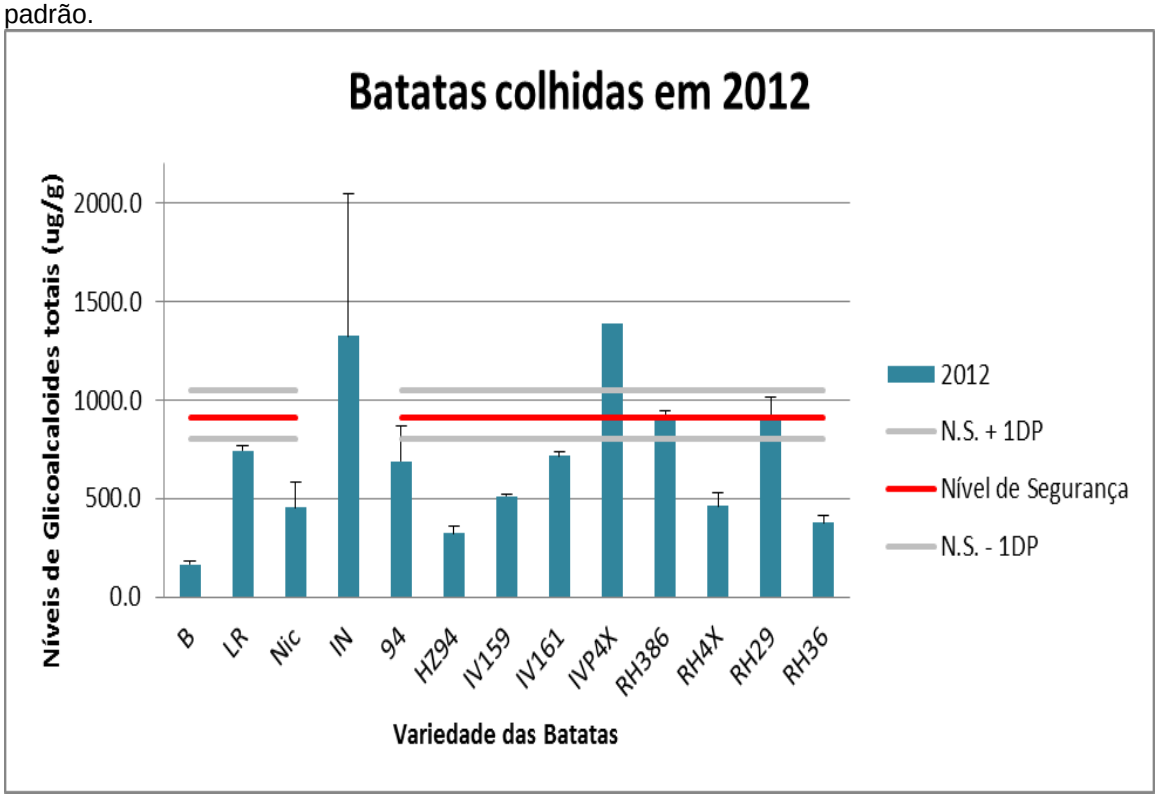
6.3.1 Glicoalcalóides Totais

As áreas dos picos detectadas pelo espectro de massas após a separação por cromatografia das amostras de batata foram comparadas com as áreas obtidas de análises de soluções dos padrões puros de α-solanina 200µg/ml e α-chaconina 800µg/ml dissolvidos em metanol, adquiridos da Sigma Chemical Co. (St. Louis, EUA). Para tanto, foram utilizadas curvas de calibração externa em seis níveis de concentrações (0/0, 25/100, 50/200, 100/400, 250/1000, 500/2000 ng/ml de alfa-solanina e alfa-chaconina

respectivamente) para a quantificação dos compostos de interesse α-solanina e α-chaconina, conforme mostrado na Figura 2.

Duas amostras evidenciaram altos níveis de glicoalcalóides, da variedade Innovator e da experimental IVP4X. Outras duas amostras mostraram índices elevados de glicoalcalóides acima do limite permitido para o consumo humano, as experimentais: RH386 e RH29 (Figura 2).

Figura 2. Distribuição da concentração dos níveis de glicoalcalóides totais (α-solanina + α-chaconina) para cada variedade e tipo de batata analisada. A faixa vermelha indica o limite seguro dos níveis de glicoalcalóides para consumo humano e as faixas cinzas indicam o desvio padrão.



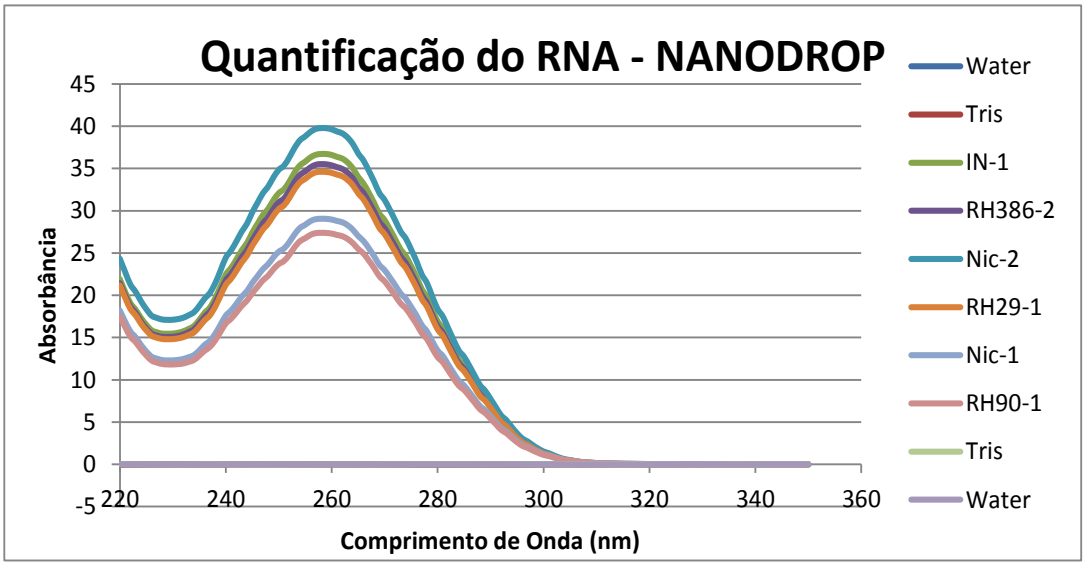
Fonte: Autor (2015).

6.3.2 Extração de RNAs e controle de qualidade

A quantificação dos RNAs foi feita através da leitura da absorbância a 260 e 280 nm e sua pureza verificada pela relação entre as absorbâncias a 260/230 nm e 260/280 nm no aparelho Nanodrop 1000 Instrument (NanoDrop Technologies). A qualidade dos RNAs também foi confirmada através da

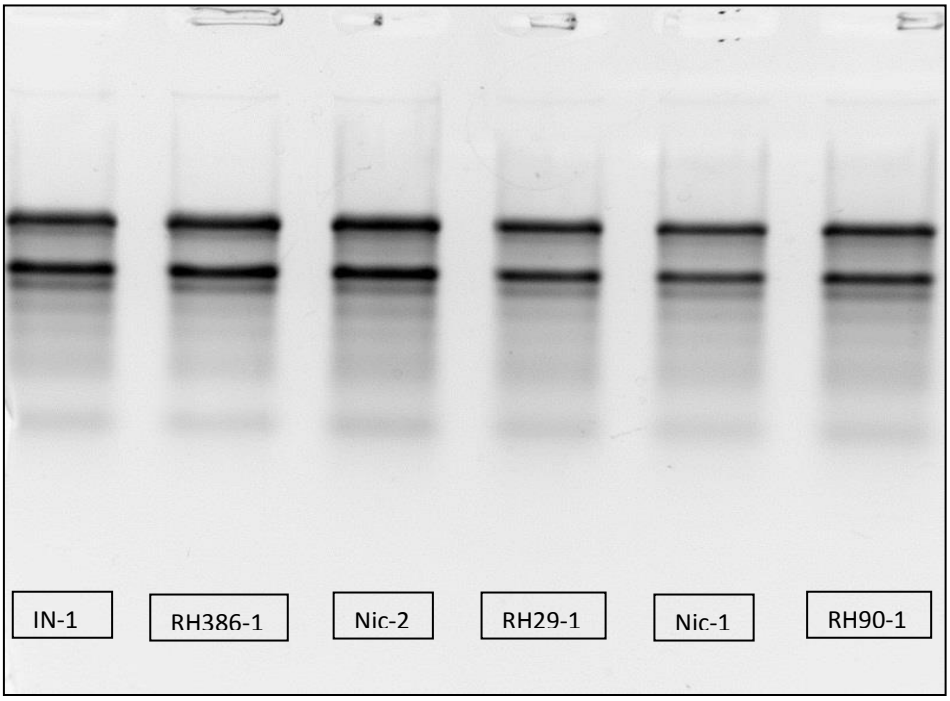
verificação da suavidade das curvas de densidade ótica (Figura 3). Todas as amostras apresentaram pico esperado a 260 nm. Os RNAs das amostras de batatas isolados pelo método CTAB modificado tiveram sua integridade verificada através da migração em gel desnaturante (1% agarose contendo 5% formamida) (Figura 4). Além da inspeção visual, foi feita a quantificação da intensidade das bandas 28S (superior) e 18S (inferior) e do rastro na região de menor massa. A relação 28S/18S próxima de 2 (ou seja, 50% de intensidade da banda superior) indica baixa degradação do RNA, o que identifica RNA de boa qualidade.

Figura 3. Curvas de absorbância de RNAs extraídos de algumas das amostras de batatas a partir do método CTAB com modificações.



Fonte: Autor (2015).

Figura 4. Gel de agarose 1% contendo 5% formamida para verificação da integridade dos RNas das amostras de batatas liofilizadas. A figura mostra RNAs de apenas algumas das amostras de batatas utilizadas neste estudo. Entretanto, a qualidade dos RNAs das outras amostras foram muito similares aos observados nesta figura (dado não mostrado).



Fonte: Autor (2015).

6.3.3 Sequenciamento do mRNA

Um experimento típico de *RNASeq* é caracterizado primeiramente pela conversão dos RNAs em uma biblioteca de fragmentos de cDNA, a qual pode então ser sequenciada pelo aparelho Illumina Hiseq 2000. Começando com uma amostra de RNA total, o protocolo para a construção de uma biblioteca para RNA-seq pela Illumina inclui os seguintes passos: isolamento de mRNA com cauda poli- A, fragmentação do RNA, trancrição reversa para formar cDNA utilizando primers randômicos, ligação de adaptadores, seleção por tamanho em gel e enriquecimento por PCR. A biblioteca de cDNA resultante é colocada em uma das oito raias da célula de fluxo do sequenciador. Fragmentos individuais de cDNA se ligam a superfície da raia e passam pelo passo de amplificação, onde serão convertidos em grupos de fita dupla de DNA. A célula

de fluxo é então colocada no sequenciador, onde cada grupo é sequenciado em paralelo. Especificamente, a cada ciclo, os quatro nucleotídeos marcados fluorescentemente são adicionados e os sinais emitidos em cada grupo é gravado. Para cada célula de fluxo, esse processo é repetido por um certo número de ciclos. As intensidades de fluorescência são então convertidas em bases denominadas. O número de ciclos determina o comprimento dos reads e o número de grupos determina o número de reads.

A qualidade das sequencias, antes e depois da filtragem, foi avaliada através do FASTX-Toolkit, conforme descrito anteriormente.

Até o presente momento o *pipeline* de análise de dados das sequencias do RNA seq não foram finalizadas. Após sua finalização, através dos resultados obtidos, será construído um banco de dados, incluindo batatas seguras (com baixo teor de glicoalcalóides) e inseguras (com alto teor de glicoalcalóides). O SIMCA será utilizado para agrupar as batatas seguras e inseguras, obtendo o modelo de segurança de novas variedades de batatas. Assim, quando uma nova variedade de *S. tuberosum* até mesmo geneticamente modificada, estiver sendo testada, as mesmas etapas (isolamento do RNA e seu sequenciamento) deverão ser realizadas, incluindo esta nova amostra será possível ver em que grupo irá encaixar-se: segura ou insegura.

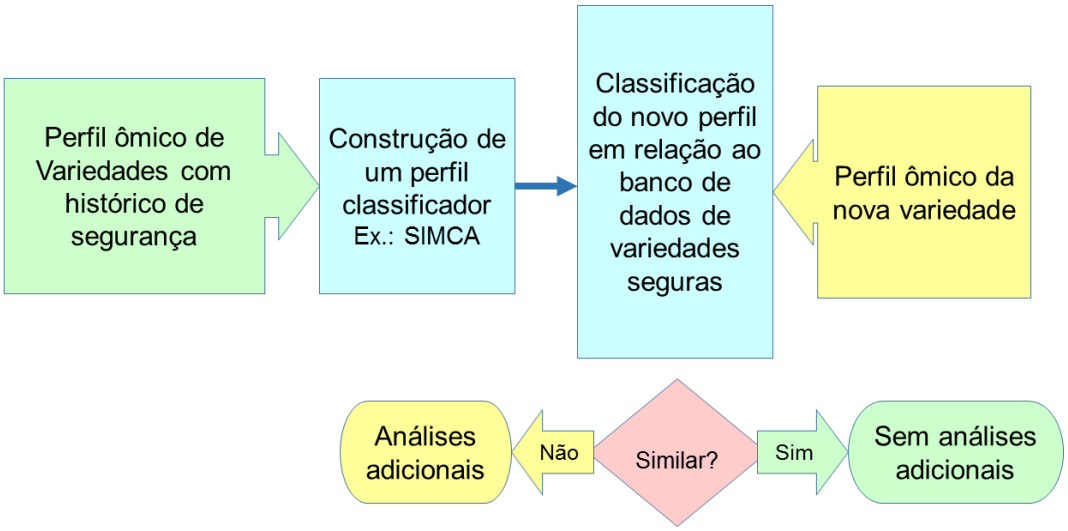
6.4 PERSPECTIVAS

Para a construção do modelo de segurança de alimentos para efeitos não-intencionais de novas variedades de batatas, todas as 90 amostras de batatas serão incluídas no banco de dados. Ainda, será utilizada a ferramenta estatística para dados multivariáveis, Soft Independent Modeling of Class Analogy (SIMCA) para a construção de um modelo que poderá ser chamado de classificador de validação cruzada. Os softwares GeneMaths XT2.12 e Pirouette 4.0 serão utilizados para aplicação do SIMCA.

Através deste modelo será possível classificar as amostras em conjuntos de dados baseados em referências: batatas como seguras ou inseguras, conforme o conteúdo de glicoalcalóides nas batatas estudadas.

A seguir a Figura 5 ilustra o modelo de avaliação de segurança de novas variedades de plantas baseado em análise comparativa e utilizando tecnologias ômicas.

Figura 5. Esquema mostrando o resumo de um novo conceito em avaliação de segurança de novas variedades de alimentos baseado em análise comparativa utilizando tecnologias ômicas.



Fonte: Baseado em VAN DIJK et al. (2014).

O modelo de avaliação de segurança será uma importante ferramenta auxiliar para predizer se uma nova variedade pode ser considerada segura ou não. Caso tenha sucesso com a batata, este modelo poderá ser expandido para outras plantas comestíveis, como por exemplo: tomate, arroz, soja, entre outros

7. CONSIDERAÇÕES FINAIS

A utilização de tecnologias ômicas baseado em análises comparativas com alimentos com histórico de segurança e que reduzem a utilização de animais experimentais são uma tendência na área de segurança de novas variedades e de organismos geneticamente modificados. Por este motivo, o modelo de avaliação da segurança de novas variedades de batatas baseado em dados de transcriptômica (*RNAseq*), é bastante promissor, ainda que necessite de validações e ajustes. Adicionalmente, o modelo, posteriormente a sua devida validação, poderá ser ampliado para outras variedades de alimentos, como soja, arroz, tomate, mandioca, sorgo, mamão, entre outros.

A escolha de genes referências estáveis é determinante na obtenção de resultados legítimos em estudos de expressão gênica relativa. Além disso, sabe-se que dados de transcriptoma obtidos, por exemplo, de experimentos de microarranjos e *RNASeq* demonstraram ser uma excelente fonte de seleção destes genes referência. Entretanto, uma validação dos genes selecionados por dados de transcriptoma em RT-qPCR, utilizando ainda softwares, como o *NormFinder*, *geNorm* e o *BestKeeper,* para determinar a estabilidade da expressão dos genes para o grupo de amostras específico a ser estudado é essencial para confiabilidade dos resultados finais obtidos.

Para o presente estudo, dados de transcriptoma de 90 amostras de tubérculo de batatas, incluindo as oito amostras experimentais (com alto e baixo teor de TGA), foram utilizadas para determinar um ranking de estabilidade da expressão de genes candidatos a normalizadores baseado no cálculo do intervalo interquartil (*Interquartile Range* – IQR). Os dez genes mais homogeneamente expressos (com menor IQR) dentre as 90 amostras de batatas foram validados para as oito amostras em RT-qPCR, e sua estabilidade foi determinada através de diferentes softwares: *NormFinder*, *geNorm* and *BestKeeper*, para evitar tendências. Os genes *C2*, *SEC3* e *CUL3A* demonstraram ser estáveis e adequados para estudos de expressão gênica em

tubérculos comestíveis de batata nas condições das oito amostras testadas e validadas. Estes genes puderam ser utilizados para determinação de expressão de genes alvos avaliados em tubérculos de batatas neste presente trabalho.

A rota metabólica e a regulação da biossíntese de GA em batatas ainda são pouco conhecidas. Até o ano de 2013, apenas os genes *Solanum tuberosum 3 UDP-galactose: solanidine galactosyltransferase* (*SGT1)*, *Solanum tuberosum UDP-glucose: solanidine glucosyltransferase* (*SGT2*) e *Solanum tuberosum rhamnose: beta-solanine / beta-chaconine rhamnosyltransferase* (*SGT3*) que codificam enzimas responsáveis pela adição de açúcares no final da rota metabólica (a partir do colesterol até a formação de α-chaconina e α-solanina) de biossíntese de GA eram conhecidos e investigados. No entanto, um estudo realizado em 2013 propôs uma rota metabólica para plantas da família das Solanaceae, incluindo a *S. tuberosum* L., e descrevendo outros genes que codificam as enzimas catalisadoras de reações desta rota: os genes *GAME*s. Em batatas os genes identificados foram os seguintes: *GAME4*, *GAME6, GAME7, GAME8a, GAME8b, GAME11* e o *GAME12.* E, juntamente com o *SGT1*, *SGT2* e *SGT3*, os genes *GAME*s codificam as principais enzimas que participam da biossíntese da α-chaconina e α-solanina a partir de moléculas de colesterol em batatas.

Um dos capítulos deste trabalho foi investigar a relação entre os genes *GAME*s e o conteúdo de GA em tubérculo de batatas com diferentes níveis de TGA. Através dos resultados obtidos, conclui-se que existe relação entre o conteúdo de TGA e a expressão dos genes *GAME*s em tubérculos de batatas. Todos os genes *GAME*s expressaram mais em tubérculos com maiores níveis de TGA. Com exceção do *GAME7*, os genes *GAME4*, *GAME6*, *GAME8ab*, *GAME11* e *GAME12* demonstraram ser significativamente ($\alpha = 0,05$) mais expressos em amostras com maior teor de TGA, e menos expressos em amostras com menor teor de glicoalcalóides confirmando a relação destes genes com a produção de glicoalcalóides em tubérculos de batata.

Para uma maior caracterização dos genes *GAME*s foi realizada uma análise de elementos *cis* da região promotora dos genes: *GAME4*, *GAME6*, *GAME7*, *GAME8ab*, *GAME11* e *GAME12*. Todos os genes analisados apresentaram uma série de fatores de transcrição relacionados à resposta de plantas a estresse biótico e abiótico, e regulação por luz, além de cópias destes elementos *cis*. Estes resultados confirmam a literatura de que diferentes níveis de TGA em tubérculos de batatas são influenciados por estes fatores, além é claro de variabilidade genética, local de cultivo, estação do ano, entre outros.

Os resultados encontrados no presente trabalho são importantes para um melhor entendimento da formação e regulação dos GA em tubérculos de batatas, e ainda podem auxiliar na predição e prevenção de formação deste tóxico neste alimento amplamente consumido.

Como perspectivas da pesquisa, pretende-se ainda determinar a expressão dos genes *SGT1* e *SGT3* nas oito amostras de batatas, realizadas através do RT-qPCR, e, ainda, estes genes serão incluídos na análise de região promotora.

8. REFERÊNCIAS

BARRELL, P.J. et al. Applications of biotechnology and genomics in potato improvement – Review Article. **Plant Biotechnology Journal**, v.11, p. 907-920, 2013.

BENFEY, P.N.; MITCHELL-OLDS, T. From genotype to phenotype: systems biology meets natural variation. **Science**, v. 320, p. 495–497, 2008.

BURLINGAME, B.; MOUILLE, B.; CHARRONDIERE, R. Nutrients, bioactive non-nutrients and anti-nutrients in potatoes. **Journal of Food Composition and Analysis,** v. 22, p. 494–502, 2009.

CAMIRE, M.E.; KUBOW, S.; DONNELLY, D.J. Potatoes and human health. **Critical Reviews in Food Science and Nutrition**, v. 49, p. 823-840, 2009.

Carter, H.; Hofree, M.; Ideker, T. Genotype to phenotype via network analysis. **Current Opinion in Genetics & Development**, v. 23, p. 611–621, 2013.

DHONDT, S.; WUYTS, N.; INZE, D. Cell to whole plant phenotyping: the best is yet to come. **Trends in Plant Science**, v. 18, p. 428–439, 2013.

ELLEGREN, H.; SMEDS, L.; BURRI, R. The genomic landscape of species divergence in *Ficedula flycatchers*. **Nature**, v. 491, p. 756–760, 2012.

FAO. Plant Production and Protection Division . Rome : Food and Agriculture Organization of the United Nations, 2008. **The Potato** . Disponível em: <ftp://ftp.fao.org/docrep/fao/011/i0500e/i0500e02.pdf>. Acesso em 6 dez. 2014.

FAOSTAT DATABASE. Rome : Food and Agriculture Organization of the United Nations, Statistical Division, 2013. Disponível em: <http://faostat3.fao.org/browse/Q/QC/E> . Acesso em : 6 dez. 2014.

FRIEDMAN, M. et al. Anticarcinogenic effects of glycoalkaloids from potatoes against human cervical, liver, lymphoma, and stomach cancer cells. **Journal of Agricultural and Food Chemistry**, v.53, p. 6162-6169, 2005.

FRIEDMAN, M. Potato glycoalkaloids and metabolites: roles in the plant and in the diet – Reviews. **Journal of Agricultural and Food Chemistry**, v. 54, p. 8655-8681, 2006.

FUKUSHIMA, A.; KANAYA, S.; NISHIDA, K. Integrated network analysis and effective tools in plant systems biology. **Frontiers in Plant Science**, v. 5, nov. 2014. Disponível em:<http://journal.frontiersin.org/article/10.3389/fpls.2014.00598/full>. Acesso em: dez. 2014.

Ginzberg, I. et al. Induction of potato steroidal glycoalkaloid biosynthetic pathway by overexpression of cDNA encoding primary metabolism HMG-CoA reductase and squalene synthase. **Planta**, v. 235, p. 1341–1353, 2012.

GUBAREV, M.I. et al. Plant-derived glycoalkaloids protect mice against lethal infection with *Salmonella typhimurium*. **Phytotherapy Research**, v. 12, p. 79-88, 1998.

JACOBS, M.M.J. et al. What's in a name; Genetic structure in Solanum section Petota studied using population-genetic tools. **BMC Evolutionary Biology**, v.11, p. 42, 2011.

KELL, D.B. Genotype-phenotype mapping: genes as computer programs. **Trends in Genetics**, v. 18, p. 555–559, 2002.

Krits, P.; Fogelman, E.; Ginzberg, I. Potato steroidal glycoalkaloid levels and the expression of key isoprenoid metabolic genes. **Planta**, v.227, p. 143–150, 2007.

KROUK, G. et al. Nitrate signaling: adaptation to fluctuating environments. **Current Opinion in Plant Biology**, v. 13, p. 266–273, 2010.

KUIPER, H.A.; Kleter, G.A.; Noteborn, H.P.; Kok, E.J. Assessment of the food safety issues related to genetically modifed foods. **The Plant Journal,** v. 27, p. 503-528, 2001.

LAMICHHANEY, S. et al. Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. **Proceedings of the National Academy of Sciences**, v. 109, p. 19345–19350, 2012.

LEE, K.R. et al. Glycoalkaloids and metabolites inhibit the growth of human colon (HT29) and liver (HepG2) cancer cells. **Journal of Agricultural and Food Chemistry**, v. 52, p. 2832-2839, 2004.

LEHESRANTA, S.J. et al. Effects of agricultural production systems and their components on protein profiles of potato tubers. **Proteomics**, v. 7, p. 597–604, 2007.

LEI, Z.; HUHMAN, D.V.; SUMNER, L.W. Mass spectrometry strategies in metabolomics. **The Journal of Biological Chemistry**, v. 286, p. 25435–25442, 2011.

LIEDL, B.E.; KOSIER, T.; DESBOROUGH, S.L. HPLC isolation and nutritional value of a major tuber protein. **American Journal of Potato Research**, v. 64, p. 545–557, 1987.

LUCAS, M.; LAPLAZE, L.; BENNETT, M.J. Plant systems biology: net- work matters. **Plant, Cell & Environment**, v. 34, p. 535–553, 2011.

MACHADO, R.M.D.; TOLEDO, M.C.F.; GARCIA, L.C. Effect of light and temperature on the formation of glycoalkaloids in potato tubers. **Food Control**, v. 18, p. 503–508, 2007.

McCue, K.F.; Allen, P.V.; Shepherd, L.V.T.; Blake, A.; Maccree, M.M.; Rockhold, D.R. et al. Potato glycosterol rhamnosyltransferase, the terminal step in triose side-chain biosynthesis. **Phytochemistry**, v.68, p. 327–334, 2007.

McCue, K.F.; Allen, P.V.; Shepherd, L.V.T.; Blake, A.; Whitworth, J.; Maccree, M.M.; Rockhold, D.R. et al. The primary in vivo steroidal alkaloid glucosyltransferase from potato. **Phytochemistry**, v. 67, p. 1590–1597, 2006.

MORRIS, S.C.; LEE, T.H. The toxicity and teratogenicity of Solanaceae glycoalkaloids, particularly those of the potato (*Solanum tuberosum*): a review. **Food Technology in Australia**, v.36, p. 118-124, 1984.

NAURIN, S. et al. A microarray for large-scale genomic and transcriptional analyses of the zebra finch (*Taeniopygia guttata*) and other passerines. **Molecular Ecology Resources**, v.8, p. 275–281, 2008.

NEMA, P.K. et al. K. Review: Potato glycoalkaloids: formation and strategies for mitigation. **Journal of the Science of Food and Agriculture**, v. 88, p. 1869–1881, 2008.

OECD. Joint Meeting of the Chemicals Committee, Working Party on Chemicals, Pesticides and Biotechnology. **Consensus document on compositional considerations for new varieties of potatoes**: key food and feed nutrients, anti-nutrients and toxicants. 2002. 26p. (Series on the Safety of Novel Foods and Feeds, 4). Disponível em: < http://www.oecd.org/science/biotrack/latestdocuments/3/>. Acesso em: 5 jan. 2015.

Pajerowska, K.M.; Parker, J.E.; Gebhardt, C. Potato Homologs of Arabidopsis thaliana Genes Functional in Defense Signaling—Identification, Genetic Mapping, and Molecular Cloning. **Molecular Plant-Microbe Interactions**, v. 18, p. 1107–1119, 2005.

PEREIRA, A.S.et al. **Produção de Batata no Rio Grande do Sul**. Pelotas: EMBRAPA Clima Temperado, 2005. 14 p. (EMBRAPA Clima Temperado, Circular Técnica**,** 48). Disponível em :<http://www.cpact.embrapa.br>. Acesso em: 10 jan. 2015.

PLAISTED, R.L.; HOOPES, R.W. The past record and future prospects for the use of exotic potato germplasm. **American Potato Journal**, v. 66, p. 603–627, 1989.

RAMANAN, V.K. et al. Pathway analysis of genomic data: concepts, methods, and prospects for future development. **Trends in Genetics**, v. 28, p. 323–332, 2012.

SAITO, K.; MATSUDA, F. Metabolomics for functional genomics, systems biology, and biotechnology. **Annual Review of Plant Biology**, v. 61, p. 463–489, 2010.

SENGUL, M.; KELES, F.; KELES, M.S. The effect of storage conditions (temperature, light, time) and variety on the glycoalkaloid content of potato tubers and sprouts. **Food Control**, v. 15, p. 281–286, 2004.

SMITH, D.B.; RODDICK, J.G.; JONES, J.L. Potato glycoalkaloids: Some unanswered questions. **Trends in Food Science & Technology**, v. 7, p. 126-131, 1996.

SPOONER, D.M.; HIJMANS, R.J. Potato systematics and germplasm collecting, 1989–2000. **American Journal of Potato Research**, v. 78, p. 237–268, 2001.

STITT, M. Systems integration of plant metabolism: means, motive and opportunity. **Current Opinion in Plant Biology**, v. 16, p. 381–388, 2013.

THE POTATO GENOME SEQUENCING CONSORTIUM. Genome sequence and analysis of the tuber crop potato. **Nature**, v. 475, p. 189–195, 2011.

VAN DIJK, J.P. et al. Transcriptome analysis of potato tubers effects of different agricultural practices. **Journal of Agriculture and Food Chemistry**, v. 57, p. 1612–1623, 2009.

VAN DIJK, J.P. et al. The identification and interpretation of differences in the transcriptomes of organically and conventionally grown potato tubers. **Journal of Agriculture Food Chemistry**, v. 60, n. 9, p.2090-2101, mar. 2012.
VAN DIJK J.P. et al. Safety assessment of plant varieties using transcriptomics profiling and a one-class classifier. **Regulatory Toxicology and Pharmacology**, v. 70, p. 297-303, 2014.

WECKWERTH, W. Green systems biology – From single genomes, proteomes and metabolomes to ecosystems research and biotechnology. **Journal of Proteomics**, v. 75, p. 284–305, 2011.

WOLF, J.B.W. Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. **Molecular Ecology Resources**, v. 13, p. 559–572, 2013.