

Article

# Altered Steroidal Glycoalkaloid (SGA) Biosynthesis in Diploid Potatoes as Induced by EMS-Mutagenesis

Ashok Somalraju <sup>1</sup>, Kaushik Ghose <sup>2</sup>, David Main <sup>1</sup>, Jason McCallum <sup>1</sup>,  
Benoit Bizimungu <sup>3</sup>, Bourlaye Fofana <sup>1,\*</sup>

<sup>1</sup> Charlottetown Research and Development Centre, Agriculture and Agri-Food Canada, Charlottetown, Prince Edward Island, C1A 4N6, Canada

<sup>2</sup> Department of Plant and Soil Science, Texas Tech University, Lubbock, TX 79409-2122, USA

<sup>3</sup> Fredericton Research and Development Centre, Agriculture and Agri-Food Canada, Fredericton, New Brunswick, E3B 4Z7, Canada

\* Correspondence: Bourlaye Fofana, Email: bourlaye.fofana@canada.ca; Tel.: +1-902-370-1454.

## ABSTRACT

Steroidal glycoalkaloids (SGAs) can be toxic to humans at more than 100 mg/100 g dry weight of potato tubers. The objective of the current study was to characterize phenotypically and genotypically a subset of 1750 ethyl methane sulfonate (EMS)-mutagenized diploid potato clones previously reported in our group for altered SGA production. The study reports on a wide range of SGA profiles in 246 EMS-mutant lines, of which 14% showed lower SGA content than the wild types and commercial varieties. An Ampliseq gene panel sequencing of 9 key SGA biosynthetic genes from 87 EMS-treated lines showing varied SGA profiles revealed 61 unique functional SNP mutations in 56 unique EMS-treated individual lines. Mutational frequencies in the target genes ranged from 1/16 kb (*SGT2*) to 1/341 kb (*GAME7*), with an average of 1/47 kb. Among these mutations, mutations were detected in *GAME7*, *GAME6*, *GAME11*, *GAME4CH6*, *GAME4CH12*, and *SGT3* genes of low SGA EMS-treated lines, genes deemed essential for steroidal aglycone hydroxylation, oxidation, and solanidine glycosylation. Subsequent comparative transcriptomic analysis of a low SGA mutant line and a high SGA wild type line showed significant downregulation of UDP-glycosyltransferases and cytochrome P450s expression in the low SGA EMS-mutant line. Using EMS-mutagenesis, this study is the first to show evidence of an effective alteration of SGA production in diploid potato tubers and paves the way for more functional analysis of this mutant population as well as diploid mutant potato cultivar development.

## Open Access

Received: 04 June 2020

Accepted: 16 October 2020

Published: 29 October 2020

Copyright © 2020 by the author(s). Licensee Hapres, London, United Kingdom. This is an open access article distributed under the terms and conditions of [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

**KEYWORDS:** diploid; EMS; *GAME* genes; glycoalkaloids; mutations; potato; SGAs; RNA transcriptome

## ABBREVIATIONS

SGA, Steroidal glycoalkaloid; EMS, Ethyl methane sulfonate; GAME, Glycoalkaloid metabolism; SGT1, Solanidine galactosyltransferase; SGT2, Solanidine glucosyltransferase; SGT3, Rhamnosyltransferase; KASP, Competitive allele specific PCR

## INTRODUCTION

Potato (*Solanum tuberosum* L) is an important tuber crop of the *Solanaceae* family. It is the most important vegetable crop in Canada, with a total production of 4.8 million tons and a total export estimated at \$1.72 billion in the 2016–2017 cropping season [1]. Potatoes are a rich source for carbohydrates, vitamins, minerals, micronutrients and dietary fiber. Despite its dietary and nutritional importance, potato products may contain anti-nutritional factors such as steroidal glycoalkaloids (SGAs) [2,3] as well as acrylamide-forming factors including asparagine [4] that is synthesized by *asparagine synthetase 1* and 2 (*AS1* and *AS2*) [5,6]. Because of its wide cultivation and consumption worldwide, management of anti-nutritional factors in potato products is considered as a priority for producers and food industries [6].

SGAs are naturally occurring secondary plant metabolites produced by solanaceous crops [7]. They constitute an elaborate class of compounds in the plant defense mechanism [8] but are toxic to humans at higher concentrations [9]. The acceptable level of SGAs in potato tubers for human consumption is 20 mg/100 g fresh weight (FW) or 100 mg/100 g dry weight (DW), a level established as the upper limit for the release of commercial potato cultivars by USDA [10,11]. SGA poisoning is reported to cause various health conditions, including gastrointestinal disorders, hallucinations, partial paralysis, coma, and even death [12]. During food processing, SGAs are stable and not destroyed by cooking [9]. Several factors including drought [13], high temperature [14], tuber exposure to light, and wounding are reported to increase SGA content in potato tubers [15,16]. SGA content can vary depending on potato varieties and plant tissues during development, with the highest SGA levels observed in the floral and sprouting tissues [17]. A recent study by Krits et al. [18] suggested that SGA biosynthesis occurs mainly in the innermost phelloderm cell layers. To date, as many as 80 different SGAs have been identified in different potato varieties [19]. Alpha-solanine and  $\alpha$ -chaconine constitute the major SGAs in potato, accounting for 95% of the total SGAs encountered in cultivated potatoes [20,21]. The ratio of  $\alpha$ -solanine to  $\alpha$ -chaconine in cultivated potatoes can widely vary depending on the anatomy or cultivar [22]. Structurally, potato glycoalkaloids are nitrogen-containing steroids with a 27 carbon ( $C_{27}$ ) solanidine (aglycone) backbone to which carbohydrate moieties are attached at the 3-OH position [23–25]. In potatoes, trisaccharides solatriose and chacotriose are

attached to the aglycone solanidine at the 3-OH position to form  $\alpha$ -solanine and  $\alpha$ -chaconine [9,23,26].

SGA biosynthesis can be divided into three major steps: the first step occurs in the cytosol and starts with the condensation of acetyl CoA into mevalonic acid, HMG-CoA, and C5 isoprenoid units. This step is part of the primary metabolism in the mevalonate/isoprenoid pathway, and is catalyzed by a family of enzymes including 3-hydroxy-3-methylglutaryl coenzyme-A synthases and 3-hydroxy-3-methylglutaryl coenzyme-A reductases (*HMGR*) [27–29]. The C5 isoprenoid units are further condensed to form squalene and 2,3-oxidosqualene from farnesyl diphosphate, and the reactions are mediated by squalene synthase (*SQS*) and squalene epoxide (*SQE*) [18,29]. At the end of this step in primary metabolisms, lanosterol synthase (*LAS*) and cycloartenol synthase (*CAS*) convert 2,3-oxidosqualene into lanosterol and cycloartenol [28,29]. Cycloartenol leads to various phytosterols including cholesterol, sitosterol, stigmasterol and campesterol, of which cholesterol is the widely accepted precursor for glycoalkaloid formation [30]. Synthesis of cholesterol from cycloartenol is reported to be mediated by sterol side chain reductase 2 (*SSR2*) and  $\Delta(7)$ -sterol-C5-desaturase (*C5-SD*) genes [31,32]. The second step in SGA biosynthesis is the conversion of cholesterol to solanidine. It occurs in the cytosol [19,33] and is considered as secondary metabolism. It is catalyzed by a group of *glycoalkaloid metabolism* (*GAME*) genes namely *GAME4*, *GAME6*, *GAME7*, *GAME8*, *GAME11*, *GAME12* located on chromosomes 6, 7 or 12 [24,29,34]. *GAME7* and *GAME8* genes are involved in the hydroxylation of cholesterol at C22 and C26 positions, resulting in 22,26-dihydroxycholesterol formation [34]. This is followed by hydroxylations and oxidations at the C16 and C22 positions by *GAME11* and *GAME6*, forming a furostanol-type aglycone intermediate [19,34,35]. This intermediate is further oxidized and transaminated by *GAME4* and *GAME12* genes, leading to the aglycone which acts as the precursor in the formation of SGAs [34]. The third step of glycoalkaloid metabolism involves the glycosylation of solanidine at the 3-OH position by the addition of sugar moieties and is catalyzed by a set of glycosyltransferases, namely solanidine galactosyltransferase (*SGT1*), solanidine glucosyltransferase (*SGT2*), and rhamnosyltransferase (*SGT3*), leading to the formation of  $\alpha$ -solanine and  $\alpha$ -chaconine [19,36].

The majority of the genes involved in the second and third steps of SGA biosynthesis are located on chromosomes 7 and 12 [24,34]. Among these genes, chromosome 7 harbours most of the genes including *GAME6*, *GAME7*, *GAME11*, *SGT1*, and *SGT3*, whereas chromosome 6 carries *GAME8a*, *GAME8b* and chromosome 12 contains *GAME4*, *GAME12* [37]. In potato, two *GAME4* have been reported in Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>), with one on chromosome 6 referred to as *GAME4CH6* (PGSC0003DMG400010882), and one on chromosome 12, referred to as *GAME4CH12* (PGSC0003DMG400024274), both of which are predicted with cytochrome monooxygenase oxidation

activities. *GAME9* is a transcription factor reported to be involved in the transcriptional upregulation of multiple SGA biosynthetic genes [25,32,38]. Cárdenas et al. [32] demonstrated that *GAME9* regulates the expression of SGAs in both tomato and potato. Functional studies of SGTs have shown that downregulation of *SGT1* gene in potato reduced the accumulation of  $\alpha$ -solanine and increased  $\alpha$ -chaconine content, thereby with no overall change in total glycoalkaloids [28]. Conversely, down-regulation of *SGT2* lowered  $\alpha$ -chaconine content and elevated  $\alpha$ -solanine levels [36]. Moreover, Itkin et al. [34] showed that RNAi-mediated *GAME4* gene silencing substantially reduces SGA content in potato leaves and tubers and the level remains unchanged even after exposure to light. RNAi-induced gene silencing of the *St16DOX* (2-oxoglutarate-dependent dioxygenase) gene in transgenic potato lines was also shown to significantly reduce SGAs without affecting tuber yield [24]. Approaches including transcription activator-like effector nucleases (TALENs) [31] and transgenic approaches have also been used successfully in developing low SGA producing potato lines [39].

While genetic engineering approaches are exciting among the scientific community for improving potato tuber characteristics, transgenic potato acceptability by consumers is a challenge. Given their role in plant defense, a complete elimination of SGAs from the plant is not ideal. However, due to the negative anti-nutritional effect of SGAs on human health, the use of consumer-accepted approaches to reducing tuber SGA content in potato cultivars is a prime goal for all breeding programs. The traditional breeding methods to breed-out anti-nutritional factors are labour intensive and time consuming, whereas transgenic approaches are of concern to the public. Moreover, cultivated potato (*S. tuberosum* L.) is an auto tetraploid crop species with a narrow and highly heterozygous genetic background [40], making its genome complex and difficult to work with, especially for curating small recessive and deleterious mutations from breeding lines [41]. Because potato is highly susceptible to in-breeding depression, it is usually propagated clonally as heterotic F<sub>1</sub> hybrids [42,43]. Moreover, widespread self-incompatibility in diploid potato hampers its self-fertilization [43]. Recently, as a proof-of-concept, we have developed and morphologically characterized an EMS-mutagenized pre-breeding diploid potato population, for use in diploid and tetraploid potato breeding [44]. The objective of the current study is to characterize this population at the genetic and phenotypic levels for the identification of lines carrying EMS-induced mutations in SGA biosynthesis genes and displaying altered SGA content in the tubers. Using a panel of SGA biosynthesis genes in a reverse genetics approach and SGA phenotyping, we identified mutant lines with an altered SGA content and showed that mutations in genes including *SGT3* and *GAME4CH6* may be sufficient to lower the SGA levels in potato tubers.



## MATERIALS AND METHODS

### Plant Materials

The plant materials used in this study were derived from a EMS-mutant population previously developed and described by our group and in which a large phenotypic variation in terms of plant architecture, leaf and stem color, flower color, tuber color, tuber shape and size, as well as yield potential were observed in EMS-treated lines compared to the non-EMS treated controls [44].

Leaf samples were collected from 1776 EMS-treated and 242 non-EMS treated seedlings grown in a greenhouse and were stored at  $-80^{\circ}\text{C}$  for further processing. To maximize throughput and minimize the sequencing efforts, a pooling strategy with 3 to 10 samples per pool was adopted as a first step for pre-screening using four targeted genes *SGT1*, *SGT2*, *AS1*, and *AS2*. To achieve this, a 6.7 mm diameter leaf disk was sampled from 1,750 of the 1776 EMS-treated, and 140 of the 242 non-EMS treated lines (representing 20 non-EMS treated lines for each of the 7 crosses) and arranged in  $2 \times 96$  well plates containing glass beads. Leaf samples were freeze-dried for pooled DNA extraction. In a second step, based on the data from step 1 screening, single 6.7 mm diameter leaf disks were sampled from individual plants of 364 EMS-treated lines, focusing preferentially on lines that revealed mutations in more than one target gene in the pools, along with 20 non-EMS treated lines (2–3 lines per cross). The samples were arranged individually in  $4 \times 96$  well plates containing glass beads. Leaf samples were freeze-dried for DNA extraction. After the first pre-screening, lines showing mutations only in *AS1* and/or *AS2* were dropped from the study, and 46 other EMS-treated lines were selected based on a set of desired agronomic traits including tuber forming ability in the field, tuber availability, resistance to light-induced tuber greening, or resistance to Roundup herbicide (data not shown).

### Reverse Genetics

#### *DNA extraction and targeted gene amplification*

Dried leaf materials in the 96 well plates were disrupted and homogenized in a TissueLyser II (Qiagen, Hilden, Germany) at 30 Hz for 1 min and genomic DNA extraction was performed on a KingFisher® 96 robotic system (Thermo Fisher Scientific, Carlsbad, CA, USA) using the Sbeadex Maxi Plant Kit (LGC, Berlin, Germany) according to the manufacturer specification. Four target genes were used to pre-screen the population as a first step (Supplementary Table S1). To enhance the specificity and robustness of the amplification using multiplex PCR, a nested PCR approach was used. The first round of multiplex PCR (PCR I) was run to pre-amplify *SGT1*, *SGT2*, *AS1*, and *AS2* genes simultaneously under long range PCR conditions (Supplementary Table S2). The reaction mixture contained 50 ng of genomic DNA, 5 pM of each primer, 4  $\mu\text{L}$  of  $5\times$  MyTaq buffer (Bioline, Luckenwalde, Germany), 2  $\mu\text{L}$  of BioStab II PCR

optimizer (Sigma Aldrich, Oakville, CA, USA), 1 mM ATP, 20 ng of *E.coli* SSB (MCLAB, South San Francisco, CA, USA), 20 ng Tth-RecA (MCLAB, South San Francisco, CA, USA), 1.2 U of MyTaq (Bioline, Luckenwalde, Germany), 0.4 U of HerculanaseII Fusion (Agilent, Waldbronn, Germany), in a total reaction volume of 20  $\mu$ L. Cycling conditions consisted of 2 min at 96 °C for pre-denaturation, followed by 20 cycles of 15 s at 96 °C for denaturation, 15 s at 60 °C for annealing, 5 min at 68 °C for extension, and 8 °C for holding. Then, the reaction products from this pre-amplification multiplex PCR I were digested by mixing with 20  $\mu$ L of 1 $\times$  Exonuclease buffer containing 2 U of ExoI (NEB, Frankfurt, Germany), incubated at 37 °C for 30 min, and inactivated for 5 min at 95 °C. The digestion products were purified from Long-range PCR primers and diluted.

In a second round of Multiplex PCR (PCR II), 24 amplicons covering all the 4 genes were generated by simultaneously amplifying 2–3 regions of the genes of interest in a total of 10 multiplex PCRs using 48 primers designed in 4 different versions, each elongated at the 5' end by an inline barcode (Supplementary Table S3). The reaction mixture was as in multiplex PCR I, except that it contained instead 1  $\mu$ L of the purified Multiplex PCR I products as template, 1 U of MyTaq (Bioline, Luckenwalde, Germany), and 0.2 U of HerculanaseII Fusion (Agilent, Waldbronn, Germany) in a 20  $\mu$ L reaction volume. Cycling conditions were 1 min at 96 °C for pre-denaturation, followed by 20 cycles of 15 s at 96 °C for denaturation, 15 s at 55 °C for annealing, 90 s at 70 °C for extension, and 8 °C for holding. To purify the amplicons, one 96-well plate of amplicon pools was generated by combining 5  $\mu$ L per well from all 4 Multiplex PCR II plates and mixed. A half volume (120  $\mu$ L) of the pooled amplicon was purified using Agencourt AMPure XP beads (Beckman Coulter, Krefeld, Germany). In short, 80  $\mu$ L of AMPure beads were added to each well, mixed by pipetting, and incubated for 10 min at room temperature. The sample was incubated on a magnet for 5 min to collect magnetic beads, and the supernatant was removed. The beads were washed twice with 200  $\mu$ L of 75% ethanol, air dried, and the DNA was eluted from the beads using 20  $\mu$ L of 5 mM Tris buffer (pH 9).

#### *Library construction and sequencing*

Using the purified amplicons, a total of 96 Illumina libraries with different adaptor indices were constructed using the Ovation Rapid DR Multiplex System 1–96 (NuGEN Technologies, Inc., Redwood, CA, USA) following the kit manual. The libraries were purified, pooled, and fragments in the 300–600 bp range were gel-purified to remove any non-specific amplicons. The final library was sequenced as 300 bp paired-end reads using an Illumina MiSeq V3 Chemistry (Illumina, San Diego, CA, USA).

### *Bioinformatics analysis of sequence reads and EMS-induced SNP discovery*

After the sequencing run, reads were pre-processed by demultiplexing the reads using Illumina's CASAVA v1.8.2 data analysis software (Illumina Inc.). The barcodes were sorted and trimmed, and individual FASTQ files were generated for each sample. To confirm the reference identity of the wild type target sequences, reads from the 3 wild type plants in each cross were aligned to the genomic reference sequence for each target gene (Supplementary Table S1) using Bowtie2 v2.2.3 [45] with default parameters in “local” alignment mode. Next, a consensus sequence for each target gene was generated from the non EMS-treated controls for each cross and was aligned with its respective reference sequences obtained from phytozome (Supplementary Table S1). The alignments were coordinate-sorted and saved in a standard Binary Sequence Alignment/Map (BAM) format. Subsequently, SAMtools mpileup followed by fixed cutoff variant calling and genotyping (<http://samtools.sourceforge.net/mpileup.shtml>) was used to call SNPs. To detect EMS-induced SNPs in the EMS-treated lines, reads from all the 1750 (first round sequencing) or 364 (second round sequencing) EMS-treated lines were aligned using Bowtie2 v2.2.3 as described earlier with both the phytozome reference and the cross-specific consensus sequences, and raw and filtered SNPs were called. SNPs causing codon changes were identified using SnpEff 4.0e (<http://snpeff.sourceforge.net/>). Exon positions described for each target gene in Phytozome were used to set up a SnpEff database for automatic annotation of VCF files following SnpEff & SnpSift manual [46] and an allele table was constructed for all samples, and mutation rates, mutational effects of missense and nonsense mutations were determined as previously reported [47,48]. To validate some of the detected SNPs, Kompetitive Allele Specific PCR (KASP, LGC Genomics, Beverly, MA, USA) assay was performed to target 22 SNP mutations (Supplementary Table S4) in 49 potato lines shown to carry SNPs mutation in *AS1*, *AS2*, *SGT1*, or *SGT2*. The KASP assays were performed and analyzed as described by Fofana et al. [48].

### **SGA Phenotypic Characterization in Mutant Lines**

#### *Plant materials*

After generating and characterizing the potato mutant population at the morphological levels [44], it was of interest to determine whether glycoalkaloid biosynthesis is altered in the pre-screened and reduced population. Among the 364 EMS-most putative mutants detected after the pre-screening sequencing and the 70 non-treated lines, a total of 246 EMS-treated lines and 21 non-treated control lines across the 7 crosses were selected for total SGA quantification using UPLC. The selection of these EMS-treated lines was based on the observed phenotypic characteristics in the field, tuber yield, tuber availability, presence of positive mutations in *SGT1* and/or *SGT2* from the second pre-screening sequencing data. Three

commercial varieties Atlantic, Yukon Gold and Russet Burbank were also analyzed and used as references.

#### *SGA extraction and quantitation*

A modified Quick, Easy, Cheap, Effective, Rugged, and Safe (QuEChERS) method [49] was employed for glycoalkaloid extraction. Three tubers per line, each representing a biological replicate, were used. Tuber periderm tissue samples were used for total SGA quantification. To avoid tissue sampling on one side of the tuber surface, four cores were cut at random from different sides of each tuber using a 1.5 cm diameter cork borer, and the 2 mm thick periderm was separated from the core, pooled and lyophilized. The lyophilized samples (79–259 mg) were blended using a polytron (Kinematica, Luzern, Switzerland) at 30,000 rpm in a 20 mL mixture of water: acetonitrile (5:15), acidified with 1% formic acid. Samples were incubated in a sonic bath at 37 °C for 1 h before the addition of 3 g of QuEChERS salts (2.0 g  $\text{MgSO}_4$ , 0.7 g sodium citrate dihydrate ( $\text{C}_6\text{H}_5\text{Na}_3\text{O}_7 \cdot 2\text{H}_2\text{O}$ ), and 0.3 g NaCl). The samples were then vigorously mixed by vortex and subsequently briefly centrifuged at 1200 rpm, to promote phase separation, with the upper acetonitrile layer containing the extracted glycoalkaloids. The upper acetonitrile organic phase was collected, further diluted with acetonitrile as required for calibration curve linearity, and SGA content was analyzed and quantified by UPLC-MS on a Waters Acquity H-Class UPLC and Tandem Quadrupole (TQD) MS system (Waters, Milford, MA, USA). Briefly, a Waters Acquity BEH-amide column (2.1 × 150 mm, 1.7  $\mu\text{m}$  particle size) was used for separations, at a flow rate of 0.500 mL/min and column temperature of 45 °C, using the following binary gradient:  $t_0$ , B = 90%;  $t_6$  min, B = 65%;  $t_7$  min, B = 50%,  $t_8$  min, B = 50% (isocratic);  $t_9$  min, B = 90%;  $t_{10}$  min, B = 90% (isocratic), where A consisted of 10 mM ammonium formate ( $\text{NH}_4\text{HCOO}$ ) with 0.1% formic acid in water, and B was acetonitrile, modified from Nikolić et al. [50]. Using this BEH-amide column method,  $\alpha$ -chaconine and  $\alpha$ -solanine were fully baseline separated with 1 min between peaks, and with all glycoalkaloid peaks eluting by 5.5 min. Electrospray ionization mass spectrometric analysis (ESI-MS) was conducted in positive ion mode (PIM), with 3.9 kV capillary voltage, 90 V cone voltage, 3 V extractor, 150 °C source temperature, nitrogen desolvation gas temperature of 430 °C and flow rate of 500 L/h, and cone gas flow for 3 L/h. A combination of PIM-ESI mass scans (120–1200 amu), a selective ion recording (SIR) for pseudomolecular ions  $[\text{M} + \text{H}]^+$  of  $\alpha$ -chaconine ( $m/z$  853),  $\alpha$ -solanine ( $m/z$  869), and solanidine ( $m/z$  399) (Supplementary Figure S1), and multiple-reaction monitoring (MRM) for  $\alpha$ -chaconine (853/707) and  $\alpha$ -solanine (869/707) were used to identify peaks, using authentic standards for  $\alpha$ -chaconine,  $\alpha$ -solanine, and solanidine (ChromaDex, Irvine, CA, USA). For SIR scans, linear  $y = mx + b$  calibration curves of integrated areas under the curve were constructed (MassLynx, Waters, Mississauga, ON, Canada), relating total ion counts (TIC) vs concentration, achieving  $r^2 = 0.996$ , 0.995, and 0.990 for  $\alpha$ -

chaconine,  $\alpha$ -solanine, and solanidine, respectively (Supplementary Figure S1). The linear detector responses ranged from 4.1 to 0.01  $\mu\text{g/mL}$ . The SGA content was expressed as mean  $\pm$  standard deviation  $\text{mg}/100\text{ g}$  dry weight from three biological and technical replicates.

### Targeted Gene Panel Ampliseq Sequencing of SGA Biosynthetic Genes and Mutation Validation

The glycoalkaloid metabolism (*GAME*) gene clusters involved in the core SGA biosynthetic pathway have been reported using a comparative co-expression analysis of tomato and potato [34] and the core pathway starts prior to the glycosylation steps of the aglycone. Similar to previous reports [48,51], targeted Ampliseq sequencing was performed on 9 genes including *SGT1* and *SGT2* using the Ion PGM™ system (Thermo Fisher Scientific, Waltham, MA, USA) to determine whether EMS treatment induced mutations in any other SGA biosynthetic genes. The details of the 9 target gene structures are shown in Supplementary Figure S2. The plant materials used in this Ampliseq experiment consisted of 87 EMS-treated and 15 non-treated lines, of which 41 out of 87 EMS-treated lines were selected based on the presence of functional mutations in at least one or both of the SGT target genes after the first (pooled samples) and second (individual samples) Miseq amplicon sequencing screening. The remaining 46 EMS-treated lines were selected based on the agronomic traits including tuber forming ability in the field, tuber availability, resistance to light induced tuber greening, or resistance to Roundup herbicide (data not shown). DNA extraction from each of the 102 lines (87 EMS-treated and 15 non-treated) was performed using Mag-Bind® Plant DNA Plus 96 Kit (Omega Bio-tek, Norcross, GA, USA) as per the manufacturer's guidelines. The extracted DNA was quantified using a Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific, Carlsbad, CA, USA) as per the manufacturer's specifications.

#### Custom Ion AmpliSeq panel design

A custom gene panel consisting of nine SGA biosynthesis genes was designed using Ion AmpliSeq Designer (IAD) software v.5.2 (Thermo Fisher Scientific, Carlsbad, CA, USA) as previously reported [48]. The gene panel included three SGT biosynthetic genes (*SGT1*, *SGT2* and *SGT3*) and 6 *GAME* genes (*GAME4CH6*, *GAME4CH12*, *GAME6*, *GAME7*, *GAME11* and *GAME12*) reported by Itkin et al. [34] and was used as targets for Ampliseq sequencing. The custom Ampliseq panel was designed in such a way that it covered 25,544 bp of the 25,669 bp potential target regions, achieving 99.4% coverage. The custom panel included a total of 107 primer pairs divided into two pools of 53 and 54 primer pairs each, respectively (Supplementary Table S5). The 25,669 bp target sequence was uploaded as a BED file and used as a reference for the initial screening of variants from the EMS-treated and non-treated potato plants. Next, similar to pre-screening sequencing, a cross-specific consensus sequence was generated



for each target gene by aligning three non-treated controls for each cross and used as a reference to further screen and call true EMS-induced SNPs.

#### *Ampliseq library preparation and sequencing*

The quantified genomic DNA was normalized to 10 ng/μL for all the 102 potato lines. A sample pooling strategy was employed to minimize the sequencing efforts and Ampliseq sequencing was performed in two steps: In a first step, a total of 13 DNA pools containing each an equal amount of seven to eight target DNA samples were prepared from the 87 EMS-treated and 15 non-treated potato lines. Ion AmpliSeq™ Library Kit 2.0 (Thermo Fisher Scientific, Carlsbad, CA, USA) was used for the targeted amplification of the 13 DNA pools. For each of the 13 DNA pools, two separate multiplex PCR reactions were setup, one with 53 and the other with 54 AmpliSeq™ primer sets for target amplification. The resulting amplicons from the two PCRs were combined together for each DNA pool, partially digested with FuPa reagent (Thermo Fisher Scientific, Carlsbad, CA, USA) and ligated using IonXpress™ barcode adapters. The barcoded libraries were purified using Agencourt® AMPure® XP Kit (Beckman Coulter, Mississauga, ON, Canada) according to the manufacturer's specification and further quantified by qPCR using Ion Universal Library Quantitation Kit (Thermo Fisher Scientific). The quantified libraries were equalized to 30 pM and used in template preparation for sequencing. Template preparation was performed using Ion PGM™ Hi-Q™ Chef Kit (Thermo Fisher Scientific) on an Ion Chef Instrument (Thermo Fisher Scientific) as per the manufacturer's instructions. Each sequencing template was prepared by pooling 6 to 8 barcoded libraries and loaded on to a 316 v2 chip, achieving a minimum of 250–300× coverage of the 107 targeted and generated amplicons. The sequencing was performed on an Ion Torrent Personal Genome Machine (PGM) (Thermo Fisher Scientific, Carlsbad, CA, USA) using Ion PGM HiQ Sequencing kit and 850 flows (Thermo Fisher Scientific).

In a second step, pooled libraries containing EMS-induced mutations were identified and their individual DNA samples were used as templates to construct individual AmpliSeq libraries for re-sequencing using the same gene panel as described above, but with a higher coverage (10,000–18,500×).

#### *Read mapping and variants call for SNP detection*

The sequencing data analysis was performed on the local IonTorrent Server using pre-installed Ion Torrent Suite Software v 5.6.0 (ThermoFisher Scientific) as in Fofana et al. [48]. The 25,669 bp target genomic region uploaded as a BED file was used as a reference for mapping the sequencing reads. A Torrent Variant Caller 5.8 plugin was selected for SNP and INDEL detection in target genes. The default variant caller plugin setup was selected with the parameters for a minimum value of allele frequency, variant quality score and coverage of the reads for

INDEL identification set to 0.1, 10 and 10, respectively for the listed parameters whereas the same parameters were set to 0.1, 10 and 5 for SNP identification. Variant calls with allele frequency  $\geq 50\%$  were considered after the filtration criteria. The position of the variants, nucleotide changes, allele type and type of mutations in EMS-treated lines *versus* non-treated lines were summarized and reported. The mutation rates were determined as previously described and the mutational effects of the EMS-induced variations were predicted using Phyre 2 investigator software as reported previously by Ghose et al. [52].

### **Transcriptome Sequencing of High and Low Glycoalkaloid Potato Lines**

#### *Plant materials and RNA extraction*

Following the validation of the EMS-induced mutations and SGA quantification, low and high SGA potato lines were identified among the EMS-treated and non-treated lines, respectively. It was therefore of interest to determine if low and high SGA-producing EMS-treated and control lines differ at the transcriptional levels. A mutant line (EMS4\_901) with low SGA content (50.4 mg/100 g DW) and a control line (CTL4\_492) with high SGA content (165.1 mg/100 g DW) were used as plant materials for the RNA transcriptomic study. Three tubers from each low and high glycoalkaloid lines were used for skin tissue sampling, with each tuber considered as a replicate. The tubers were thoroughly washed under running water and cleaned with 70% ethanol prior to sampling. Tuber tissue was collected as per the method described by Grunenfelder et al. [15]. Upon collection, the tissue was immediately frozen in liquid nitrogen, and RNA extraction and purification were performed using Spectrum™ Plant Total RNA kit (Sigma-Aldrich, Toronto, ON, Canada) following supplier's guidelines. The DNase-treated total RNA was diluted in 100  $\mu$ L nuclease-free water and supplemented with 1  $\mu$ L of RNase OUT. RNA quality was visualized on agarose gel and quantified using NanoDrop (Thermo Scientific, Madison, WI, USA) before shipping the ethanol-precipitated total RNA samples to Macrogen Inc. (Macrogen Inc, Seoul, South Korea) for transcriptome sequencing, following Macrogen's Sample Submission Guide [53]. Once received, the RNA samples were re-pelleted, and the quality control was performed using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). RNA QC was ensured before moving to the library preparation as all samples had an RNA integrity number (RIN)  $\geq 7.0$  and for absence of DNA contamination.

#### *Library construction, transcriptome sequencing, and data analysis*

To perform the RNA transcriptome analysis, cDNA libraries were constructed following the Macrogen's in-house workflow (Macrogen Inc, Seoul, South Korea). Briefly, using 1  $\mu$ g total RNA as starting material, the cDNA libraries were generated with the TruSeq® Stranded Total RNA

library preparation kit (Illumina). The libraries were quantified by qPCR and paired-ends ( $2 \times 100$  bp) sequencing was performed on a HiSeq2500 platform (Illumina, Inc). The raw sequencing read data were submitted to the GenBank Short Read Archive (SRA) database under the Bioproject ID# PRJNA533334.

The quality control check of the raw sequencing data at each cycle was done using FastQC V0.10.0 [54]. Trimming the data for the removal of adapters, low quality bases and shorter reads ( $<36$ ) was carried out using Trimmomatic V0.32 program [55]. The reference genome sequence of *Solanum tuberosum* (GCF\_000226075.1, SolTub\_3.0) and the annotation data were downloaded from the NCBI. The reference genome was indexed using HISAT [56] and the trimmed paired end reads were indexed, aligned and mapped to the reference genome using HISAT [56]. After alignment, the processed reads were assembled into transcripts and their abundance was estimated using StringTie v1.3.3b [57,58]. StringTie infers the relative transcript abundance as values of Fragments per Kilobase of transcript per Million reads mapped (FPKM), which were further normalized to the library size. StringTie, as the successor of Cufflinks, was used to determine the known genes and transcript expression levels and to investigate novel transcripts and novel alternatively spliced transcripts, when they existed. Genes with FPKM = 0 across all samples were excluded. The FPKM +1 values were log2 transformed and quantile-normalized to produce more even data distribution and to reduce systematic bias from the pre-processed core libraries and reported as processed normalized FPKM. Differentially expressed genes (DEGs) analysis was performed based on the normalized processed FPKM data from the 12 paired-comparison including the three replicates and 2 time points in each of the two lines using scripts of the comprehensive R archive network (CRAN), version R 3.4.3 (<http://cran.r-project.org>). Statistical significance of the differential expression data was determined by performing independent *t*-test and fold change considering that there was no difference between groups null hypothesis. False discovery rate (FDR) was controlled by adjusting *P*-value using the Benjamini-Hochberg algorithm [59]. Genes that satisfied the absolute fold change ( $|fc|$ ) of  $\geq 2$  and  $p < 0.05$  in the independent *T*-test from at least one of the paired-comparison were reported for each sample and transcript/gene. For DEG sets, correlation analysis, hierarchical clustering (Euclidian method, complete linkage), and multidimensional scaling analyses were performed using the heatmap.2 function provided by the R3.4.3 package gplots option for data visualization. Functional classification and gene-set enrichment of the DEGs was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID, <http://david.abcc.ncifcrf.gov/>) tool v6.8 [60,61] based on Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG, <http://www.genome.jp/kegg/pathway.html>) and other functional annotation databases including NCBI (<http://ncbi.nlm.nih.gov>) and Uniprot (<http://www.uniprot.org>).

### RNAseq Data Validation Using Quantitative RT-PCR

To validate the RNAseq data, quantitative RT-PCR (qPCR) was performed using gene specific primers. The primers were designed from 16 differentially expressed genes selected in 12 gene families through RNAseq data (Supplementary Table S6). The One-step qPCR approach [62] was used. The qScript™ 1-Step SYBR Green qRT-PCR kit (Quanta Biosciences, Beverly, MA, USA) and CFX96 Real Time system (BioRad, Laboratories, Mississauga, ON, Canada) were used to perform qPCR. Each 25 µL PCR reaction contained a total of 100 ng RNA and 300 nM for each primer. PCR cycling conditions included, cDNA synthesis at 50 °C for 10 min, pre-denaturation at 95 °C for 5 min, denaturation using 40 cycles at 95 °C for 10 s, annealing at 60 °C for 30 s, and holding at 95 °C for 10 s. A melting curve was generated between 60 °C and 95 °C to ensure the specificity of amplification for each gene. For each sample group, three biological replicates with three technical replicates were analyzed. Relative fold change in expression was calculated using  $2^{-\Delta\Delta Cq}$  method [63] and the data was normalized to *18S rRNA* (X67238), used as a housekeeping gene.

### Relationship between DEGs and EMS-Induced Mutations in SGA Biosynthesis

To establish a potential link between EMS-induced mutations in the low SGA line and its RNA transcriptomic profile and identify putative DEGs associated with SGA biosynthesis, a phylogenetic analysis was performed. A total of 68 protein sequences including 39 KEGG-enriched DEGs putatively involved in SGA biosynthetic genes were selected from the RNAseq DEGs dataset and used as input sequences along with 28 known SGA biosynthetic genes belonging to gene families such as glycosyltransferases, cytochromes, amino acyl transferases, transcription factors, sterol biosynthesis gene families downloaded from NCBI [37,38,64–66] (Supplementary Table S7). The 68 protein sequences were aligned using ClustalW and a phylogenetic tree was constructed using Neighbor-Join and BioNJ algorithm using 5,000 bootstrap replicates [67]. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. Evolutionary analyses were conducted using MEGA6 [68].

## RESULTS

### Detection of EMS-Induced Mutations through Pre-Screening Amplicon Sequencing of First 4 Target Genes

After pre-screening the 1,750 EMS-treated and 140 non-treated lines by amplicon resequencing of *SGT1*, *SGT2*, *AS1* and *AS2*, a subpopulation of 839 EMS-treated lines was selected from 57 positive pools and used as the

core germplasm collection in multiple screenings. A selection was further carried out within this core EMS population and it was reduced to 364 EMS-treated lines for the detection of individual mutant lines. A total of 71 EMS-induced mutations were detected in 166 individual EMS-treated lines. Of these 71 mutations, 10 were located in introns, whereas 61 appeared to be located in coding regions (exons) of the 4 target genes, occurring in 133 EMS-treated lines, with 42 functional mutations found in 98 of the 133 lines (Table 1). A total of 8, 10, 15 and 9 functional mutations were observed in *SGT1*, *SGT2*, *AS1* and *AS2* genes, respectively. These 98 lines included 7 lines carrying mutations in both *SGTs* and *ASs*, with 5 plants having mutations in *AS2* and *SGT2*, 1 plant carrying mutations in *AS2* and *SGT1*, and 1 plant with mutation in *AS1* and *SGT1*. These 98 lines were further considered in downstream analysis. The mutation frequency in the exonic regions of the 4 targets genes ranged between 1/43 kb and 1/436 kb with individual mutation frequencies of 1/43, 1/45, 1/228, and 1/436 kb, for *SGT2* (16 SNPs), *AS2* (16 SNPs), *SGT1* (11 SNPs), and *AS1* (18 SNPs), respectively. Of the 22 SNPs positions assayed by KASP, twelve of the 22 targeted mutations were thus validated in 36 of the 49 lines genotyped. Interestingly, 7 of the 10 targeted mutations in *SGT1* and *SGT2* were validated in 19 lines tested but the other 10 from 13 M2 families were not (Supplementary Table S8). Altogether, the calculated average mutation frequency was 1/75 kb across the 4 genes in the 1750 EMS lines screened.

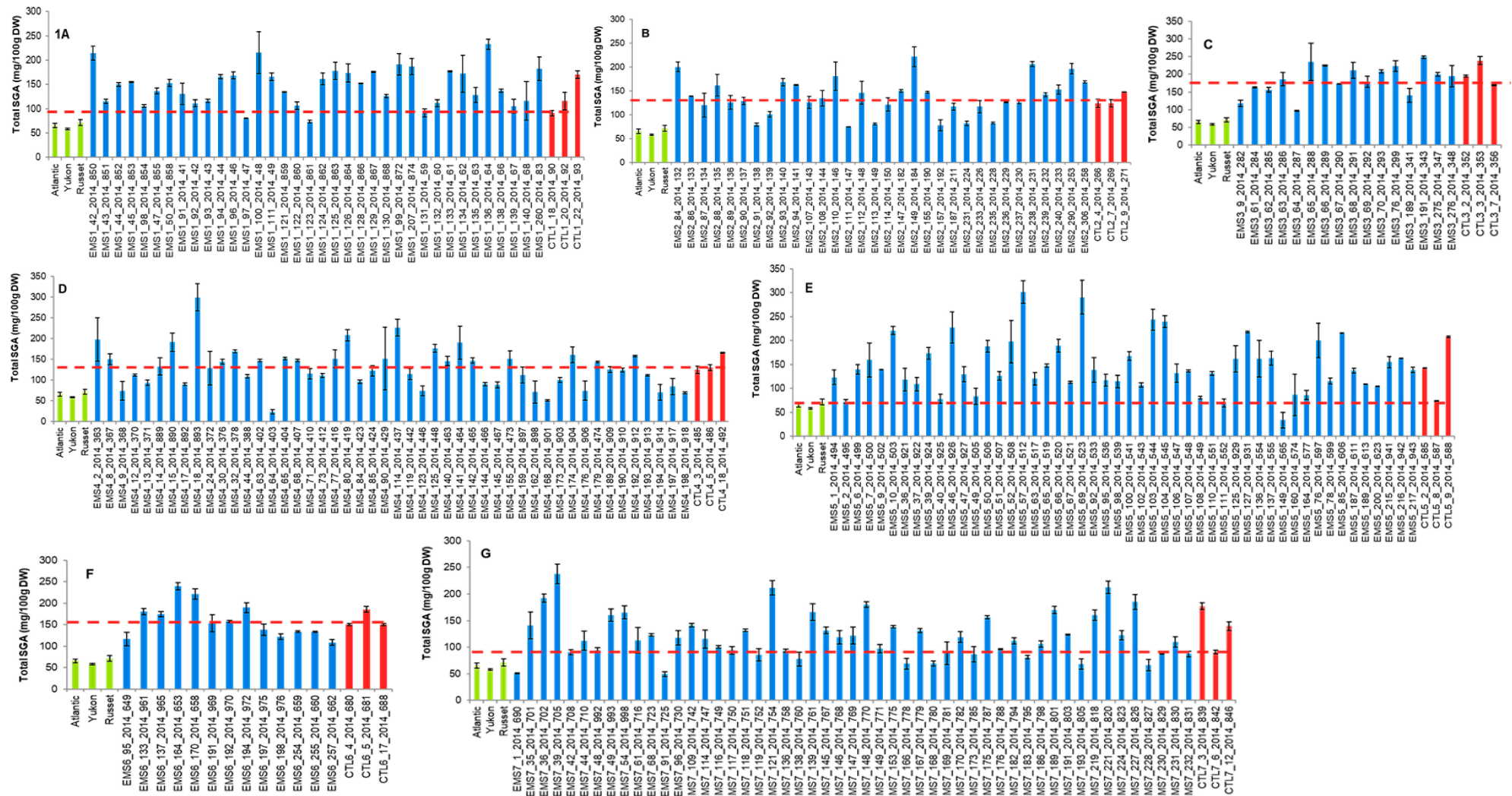
**Table 1.** Number of EMS-induced mutations identified in the exonic regions of *SGT1*, *SGT2*, *AS1* and *AS2* genes from 133 EMS-treated lines following the targeted amplicon sequencing.

Mutation type	SGT1	SGT2	AS1	AS2
Sense	3	6	3	7
Missense	7	10	12	6
Nonsense	1	0	3	3
Total	11	16	18	16

### Effect of EMS Mutations on Tuber SGA Production

To determine the genotype to phenotype relationship and assess whether the induced mutations led to altered SGA production in the detected EMS mutant lines, SGA content was assessed in 246 EMS-treated and 21 non-treated lines along with 3 commercial potato varieties (Figure 1A–G). SGA content in the tuber of commercial varieties ranged from 58 mg/100 g DW (Yukon Gold) to 71 mg/100 g DW (Russet Burbank) (Figure 1A–G; Supplementary Figure S3), whereas it ranged in the non-treated controls from 73 mg/100 g DW (CTL5\_587) (Figure 1E) to 239 mg/100 g DW (CTL3\_353) (Figure 1C). The highest tuber SGA variation was observed in EMS-treated lines, ranging from 23 mg/100 g DW (EMS4\_403) (Figure 1D)





**Figure 1.** Variations of total SGA content in EMS-treated and non-treated diploid potato lines in comparison with three commercial potato varieties. Green bars, commercial potato varieties; Blue bars, EMS-treated diploid potato lines; Red bars, non-treated diploid potato controls derived from 7 crosses (A–G, respectively). The baseline was set at the lowest non-treated control for comparison with EMS-treated and commercial lines..

to 301 mg/100 g DW (EMS5\_512) (Figure 1E). Forty five (45) of the 246 EMS-treated lines (18%) showed a lower total SGA content compared with their respective non-treated controls, whereas 34 (14%) had lower or comparable SGA level with the commercial varieties tested (Figure 1A–G). Taken together, 34 lines (14%) of the EMS-treated lines revealed a lower total SGA content than their respective non-treated controls and had lower or comparable levels with the commercial varieties evaluated. The mutant lines with low SGA profiles than commercial varieties included 2 lines in EMS1, 6 lines in EMS2, 1 line in EMS3, 13 lines in EMS4, 2 lines in EMS5, and 6 in EMS7 (Figure 1A–G). Although a mutant line (EMS6\_662) showed a lower SGA level compared to its non-treated controls, the SGA level was higher than those found in the commercial varieties (Figure 1F). Chaconine to solanine ratio ranged from 2.8 to 3.4 in the commercial varieties, 1.5 to 2.6 in the non-treated controls, and 1.0 to 3.4 in the EMS-treated lines. The chaconine to solanine ratio was the highest (3.4x) in line EMS7\_749, whereas it was the lowest (1.0x) in line EMS6\_653.

#### **Validation of Induced Mutations in SGTs and SGA Biosynthetic Pathway Extended Coverage by Ampliseq Gene Panel Sequencing**

To further confirm the EMS-induced mutations observed in SGTs and to increase the coverage of the SGA biosynthesis pathway, an Ampliseq panel sequencing of 9 target genes reported to be involved in the core SGA biosynthetic pathway was performed in 87 EMS-treated and 15 non-treated lines. For each of the 9 target genes sequenced, a 98% base alignment was achieved, with an average coverage depth of 18,500x. A total of 657 EMS-induced SNPs with an allelic frequency of  $\geq 50\%$  were observed in 80 of the 87 EMS-treated lines. Of these mutations, 131 mutations were exonic and detected in 67 of the 87 lines subjected to Ampliseq panel sequencing. Among the 131 exonic mutations, 55 were missense and 6 were nonsense mutations, for a total of 61 mutations with predictable functional effects, and were found in 56 EMS-treated lines (Tables 2 and 3). Of the 18 functional mutations detected in *SGT1* and *SGT2* through the Miseq amplicon sequencing procedure, 13 were confirmed by the Ampliseq panel screening. A high proportion (78%) of heterozygous mutations were observed and 28% of the mutations were predicted to have medium to high mutational effects (Table 3). The results also showed that 3 of the 6 observed nonsense mutations were from EMS-treated lines in cross 4. Nonsense (stop codon) mutations were detected in 4 of the 9 target genes namely *SGT2*, *GAME4CH6*, *GAME4CH12* and *GAME12* (Tables 2 and 3, Figure 2). The mutation frequency in the exonic regions of the targeted genes was the highest in *SGT2* (1/16 Kb) and the lowest in *GAME7* (1/341 Kb) (Table 2). The average mutation frequency was determined to be 1/47 Kb.

**Table 2.** Summary of EMS-induced mutations identified in the exons of 9 targeted SGA biosynthetic genes from 87 mutagenized lines.

Gene	Mutations in exon	Sense mutations	Functional mutations		Zygosity of functional mutations		Mutation Frequency (Kb <sup>-1</sup> )
			Missense	Nonsense	Hetero *	Homo *	
SGT1	11	7	4	--	4	--	1/135
SGT2	53	32	19	2	12	9	1/16
SGT3	13	7	6	--	6	--	1/49
GAME4CH6	3	2	--	1	1	--	1/204
GAME4CH12	20	7	12	1	11	2	1/28
GAME6	7	6	1	--	1	--	1/272
GAME7	2	2	--	--	--	--	1/341
GAME11	7	3	4	--	4	--	1/138
GAME12	15	5	8	2	8	2	1/22
Total	131	71	54	6	47	13	

\* Hetero, heterozygous; Homo, homozygous. Variant calls with frequency  $\geq 50\%$  were considered for SNP variants.

**Table 3.** Details of 50 unique SNP positions leading to 61 EMS-induced functional mutations detected across 9 target SGA genes, observed in exonic regions of 56 mutagenized lines.

Gene	Position	Ref	Variant *	Allele call	SNP	Amino acid		Mutation type	Mutation Sensitivity	No of line **	Cross ID
						Position	Change				
AME11	211	C/C	C/G	Hetero	SNP	3	D > E	Missense	High	2	5,6
GAME11	290	C/C	C/A	Hetero	SNP	30	P > T	Missense	Medium	2	5,6
GAME11	732	A/A	A/G	Hetero	SNP	146	K > R	Missense	Low	1	5
GAME11	940	C/C	C/A	Hetero	SNP	191	T > K	Missense	Low	1	7
GAME12	979	T/T	T/G	Hetero/Homo	SNP	10	S > A	Missense	Low	9	2,4,7
GAME12	1168	G/G	G/A	Hetero	SNP	37	G > D	Missense	Low	6	1,3,4,7
GAME12	2073	-	-/T	Homo	INS	88	STOP	Nonsense	High	1	7

**Table 3.** *Cont.*

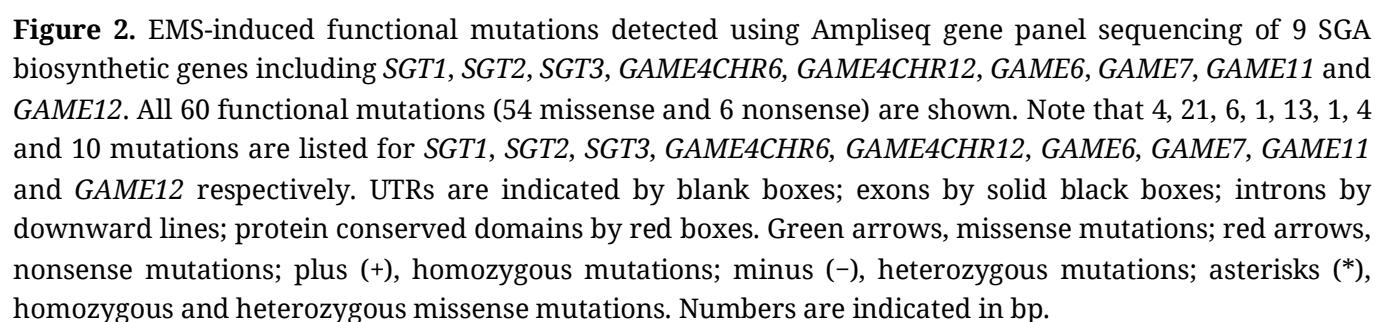
Gene	Position	Ref	Variant *	Allele call	SNP	Amino acid		Mutation type	Mutation Sensitivity	No of line **	Cross ID
						Position	Change				
GAME12	2527	C/C	C/T	Hetero	SNP	155	A > V	Missense	Low	2	4
GAME12	3373	G/G	G/A	Hetero	SNP	232	S > N	Missense	Low	3	3,7
GAME12	3381	T/T	(T/C)/(T/G)	Hetero	SNP	234	S > A	Missense	Low	3	2
GAME12	3392	T/T	T/C	Hetero	SNP	238	P > S	Missense	Medium	3	3,7
GAME12	3972	-	-/A	Hetero	INS	306	STOP	Nonsense	Medium	2	1
GAME4CHR12	219	A/T	A/T	Hetero	SNP	63	T > S	Missense	Low	1	6
GAME4CHR12	224	T/T	T/G	Hetero	SNP	64	I > M	Missense	Low	1	6
GAME4CHR12	246	C/C	C/A	Hetero	SNP	72	L > I	Missense	Low	1	6
GAME4CHR12	642	T/T	T/G	Hetero	SNP	140	L > V	Missense	Medium	3	4
GAME4CHR12	648	T/T	T/G	Hetero	SNP	142	S > A	Missense	Low	6	4
GAME4CHR12	695	C/C	C/T	Hetero	SNP	157	STOP	Nonsense	Medium	5	4,6
GAME4CHR12	710	C/C	C/A	Hetero	SNP	162	S > R	Missense	Low	10	4,6,7
GAME4CHR12	770	C/C	C/T	Hetero	SNP	185	A > V	Missense	Medium	4	4,6
GAME4CHR12	2960	A/A	A/C	Hetero/Homo	SNP	279	D > A	Missense	High	7	7
GAME4CHR12	2986	G/G	G/A	Hetero	SNP	288	V > I	Missense	Low	5	4,6
GAME4CHR12	5266	A/A	A/C	Hetero/Homo	SNP	482	I > L	Missense	Low	18	4,7
GAME4CHR6	1637	-	-/GTTTG	Hetero	INS	134	STOP	Nonsense	Medium	1	4
GAME6	841	G/G	G/T	Hetero	SNP	194	V > L	Missense	Medium	6	5,7
GAME7	1127	A/A	A/G	Hetero	SNP	114	K > E	Missense	Low	1	4
SGT1	33	GT/GT	GT/AC	Hetero	MNP ***	2	V > T	Missense	Low	4	7
SGT1	1016	(GGC/GGC)/ (G/G)	(GGC/AAT)/ (G/A)	Hetero	MNP ***	329	A > M	Missense	Medium	1	7

**Table 3.** *Cont.*

Gene	Position	Ref	Variant *	Allele call	SNP	Amino acid		Mutation type	Mutation Sensitivity	No of line **	Cross ID
						Position	Change				
SGT1	1053	A/A	A/T	Hetero	SNP	342	T > S	Missense	Low	5	4,5,6
SGT2	396	G/G	G/A	Homo	SNP	5	S > N	Missense	Low	4	4,6
SGT2	659	A/A	(A/T)/(A/G)	Hetero/Homo	SNP	93	I > L	Missense	Low	5	4
SGT2	662	C/C	C/T	Hetero/Homo	SNP	94	L > F	Missense	High	6	4,5
SGT2	665	CA/CA	CA/GG	Hetero	MNP ***	95	H > G	Missense	Medium	3	4
SGT2	710	G/G	G/C	Hetero	SNP	110	E > Q	Missense	Low	1	7
SGT2	962	G/G	G/A	Hetero	SNP	194	V > I	Missense	Medium	1	4
SGT2	1269	C/C	C/T	Hetero	SNP	296	T > I	Missense	Low	10	2,4
SGT2	1271	G/G	G/A	Hetero/ Homo	SNP	297	V > I	Missense	Low	7	4
SGT2	1362	-	-/TAGCA	Hetero	INS	327	STOP	Nonsense	Medium	1	4
SGT2	1430	C/C	C/G	Hetero	SNP	350	L > V	Missense	Medium	1	4
SGT2	1433	A/A	A/T	Hetero/Homo	SNP	351	T > S	Missense	Low	7	4,7
SGT2	1579	-	-/G	Homo	INS	399	STOP	Nonsense	Medium	4	1,2
SGT2	1619	G/G	G/A	Hetero/Homo	SNP	413	D > N	Missense	Low	13	4,7
SGT2	1734	T/T	T/C	Hetero	SNP	451	M > T	Missense	Low	4	2,4
SGT2	1833	A/A	A/T	Hetero	SNP	484	D > V	Missense	Low	5	4,7
SGT3	46	G/G	G/C	Hetero	SNP	8	E > Q	Missense	Medium	6	4,7
SGT3	445	A/A	A/C	Hetero	SNP	141	M > L	Missense	Low	5	5,7
SGT3	481	T/T	T/G	Hetero	SNP	153	S > A	Missense	Low	5	4,7
SGT3	640	A/A	A/G	Hetero	SNP	206	K > E	Missense	Low	7	4,7
SGT3	1045	GAGAA	GAGAA/-	Hetero	DEL	341	DEK	Frameshift	Medium	3	7
SGT3	1058	A/A	(A/T)/(A/G)	Hetero	SNP	345	H > L	Missense	Low	2	7
SGT3	1072	C/C	C/T	Hetero	SNP	350	R > C	Missense	Low	3	7

\* Position leading to heterozygous and homozygous variants are shown; \*\* Number of lines in which the given mutation was identified. A total of 56 non-redundant EMS lines were found to carry mutations in at least one of the 8 target genes;\*\*\* MNP, Multi-nucleotide polymorphism.





Among the EMS-treated lines showing low SGA content, EMS4\_403 (23 mg/100 g DW) carried missense mutations in *SGT3* and *GAME7*. In EMS4\_898 (71 mg/100 g DW), three positions were found to be mutated in *GAME7* only, whereas frameshift mutations leading to a stop codon were identified in *SGT2* and *GAME4CH6* in line EMS4\_901 (50.4 mg/100 g DW) (Supplementary Tables S9 and S10). *GAME4CH12* showed missense mutations at multiple positions, including positions 642, 648, 710, 2960, 5266, and as many as 12 lines (EMS4\_892, EMS4\_446, EMS4\_467, EMS4\_906, EMS4\_917, EMS7\_690, EMS7\_752, EMS7\_760, EMS7\_778, EMS7\_795, EMS7\_805, EMS7\_831) carried mutations at one or more of these positions and displayed low SGA phenotypes (Supplementary Table S9; Figure 1D,G). In particular, lines EMS4\_892 showed missense mutations predicted with medium to low mutational effect at positions 642, 648, 710, and 5266 (Supplementary Table S9, Table 1) and displayed a low SGA content in its tuber (Figure 1D). In total, 9 EMS-treated lines including 5 lines in cross 4 (EMS4\_892, EMS4\_446, EMS4\_467, EMS4\_906, EMS4\_917) and 4 lines cross 7 (EMS7\_690, EMS7\_778, EMS7\_795, EMS7\_805) carried mutations in *GAME4CH12* and produced low SGA content compared to the non-treated controls (Figure 1D,G). Moreover, nonsense and missense mutations were observed in *GAME12* of 8 low SGA lines including EMS1\_861 (73 mg/100 g DW), EMS1\_47 (80 mg/100 g DW), EMS7\_690 (51 mg/100 g DW) and EMS7\_805 (68 mg/100 g DW) (Supplementary Table S9). Interestingly, the low SGA lines EMS7\_725 (49 mg/100 g DW) carried mutations in *SGT3* and *GAME6*. Similar to *GAME6* with the low SGA line EMS7\_725, *GAME11* was found to be altered alongside with *SGT2* in the low SGA line EMS7\_827 (66 mg/100 g DW) (Supplementary Table S10).

**Table 4.** Summary statistics of transcriptomic sequencing in EMS4\_901 and CTL4\_492 lines.

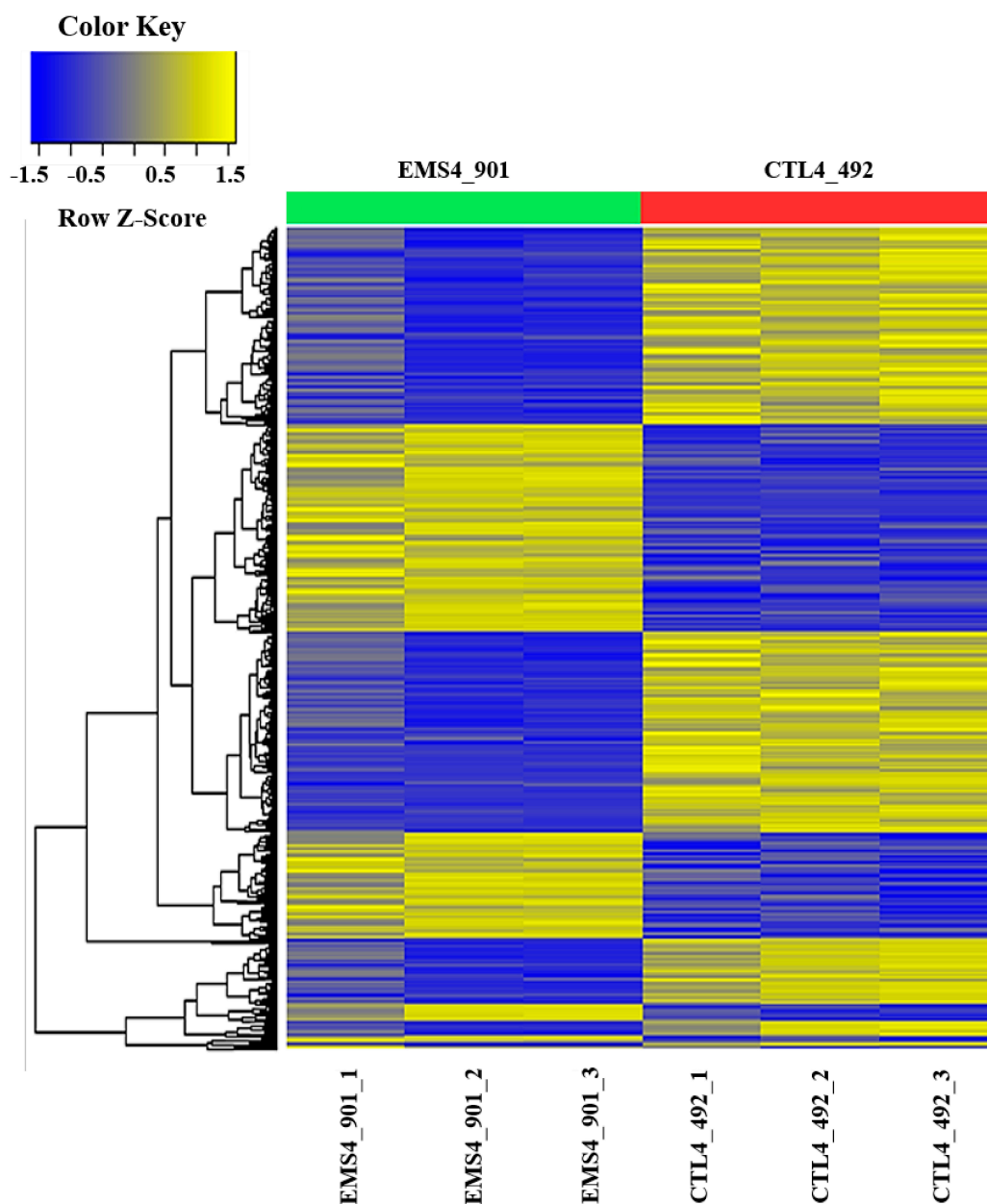
Sample ID	Rep	Raw reads <sup>a</sup>	Trimmed reads <sup>a</sup>	Processed reads <sup>b</sup>	Mapped reads <sup>c</sup>	Multiple mapped reads <sup>d</sup>	Overall read mapping (%)
CTL4_492	1	35,908,538	35,634,122	17,817,061	16,549,295	6,607,553	92
	2	36,993,012	36,719,068	18,359,534	17,305,401	8,060,440	94
	3	38,982,380	38,797,704	19,398,852	15,978,574	1,880,628	82
EMS4_901	1	39,881,154	39,644,754	19,822,377	17,577,195	5,547,320	88
	2	40,244,172	39,988,602	19,994,301	18,213,429	6,880,254	91
	3	36,821,422	36,609,540	18,304,770	15,809,529	3,602,122	86
Average		38,138,446	37,898,965	18,949,483	16,905,571	5,429,720	89

<sup>a</sup> Statistics on raw and trimmed total pair-end reads; <sup>b</sup> Statistics on total mate-paired clean reads; <sup>c</sup> Number of reads mapped to reference; <sup>d</sup> Number of reads multiple mapped to reference.

### EMS-Induced Mutation Leads to Contrasted Transcriptome in the Wild Type and Mutant Lines

Transcriptome sequencing was successfully conducted on the three biological replicates in each of the low SGA EMS mutant (EMS4\_901) and the high SGA control (CTL4\_492) lines. Mapped reads in the triplicated samples ranged from 16 to 17 million in the control line and from 16 to 18 millions in the EMS lines, with an average of 89% and 88% overall mapped reads, respectively (Table 4).

Following the mapping, transcript assembly, data transformation and quantile normalization, a total of 21,855 processed expressed genes were observed in at least one replicate of the two lines (Supplementary Table S11). Further statistical analyses of this dataset revealed a total of 1096 significant and differentially expressed genes between the two lines across all replicates, and of which 456 and 640 genes were upregulated and downregulated, respectively in EMS4\_901 when compared to CTL4\_492 (Supplementary Table S12). Fold change of the upregulated genes in EMS4\_901 ranged between 2.0 to 36.4 whereas that of the downregulated genes ranged from 2.0 to 3194 in the same line, in comparison with CTL4\_492. A consistent and contrasting gene expression profile was observed between the low SGA mutant (EMS4\_901) and the high SGA control (CTL4\_492) lines and the altered genes included *UDP-glycosyltransferase 74F2-like* (XP\_006351615.1), *UDP-glucose flavonoid 3-O-glucosyltransferase 6-like* (XP\_006353724.1), *UDP-glucuronate 4-epimerase 6-like* (XP\_006349916.1), *cytochrome P450 76A1* (XP\_006341183.1), *cytochrome P450 83B1-like* (XP\_006358770.1), *flavonoid 3',5'-methyltransferase-like* (XP\_006363473.1), *methyltransferase-like protein 2* (XP\_006367297.1) and *aspartate aminotransferase* (XP\_006343100.1) (Figure 3). Interestingly, genes sharing similar conserved domains with SGT or GAME genes were found significantly downregulated in EMS4\_901 line compared to CTL4\_492 (Table 5). In particular, two isoforms of *cytochrome p450 82C4-like* (XP\_006352674.1; XP\_015166509.1) and a *cytochrome P450 83B1-like* (XP\_006358770.1) gene sharing 40–45% sequence similarity and same conserved domain as *GAME4CH6*, *GAME4CH12*, *GAME6*, and *GAME7* were found to be significantly downregulated in EMS4\_901 compared to CTL4\_482 (Table 5). The *UDP-glycosyltransferase 74F2-like* (XP\_006351615.1) was 45% similar to *SGT2*, whereas *glutamine-fructose-6-phosphate aminotransferase 2* (XP\_006345399.1) was weakly (15%) similar to *GAME12* gene. All these isoforms were also found to be downregulated in EMS4\_901 (Table 5).



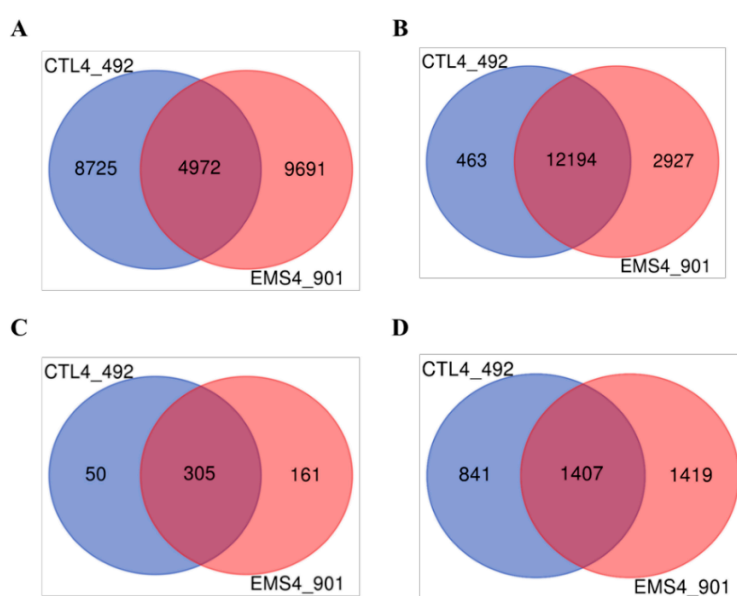
**Figure 3.** Heatmap showing the hierarchical clustering of 1096 differentially expressed genes in the mutant (EMS4\_901) low and the control (CTL4\_492) high SGA producing potato lines. Complete linkage and Euclidian distance were used as a measure of similarity to display expression patterns of DEGs with absolute fold ( $|fc|$ )  $\geq 2$  and  $T$ -test  $p$ -value  $< 0.05$ . Each row corresponds to a gene and each column corresponds to an individual replicate for each cultivar. The mutant line EMS4\_901 and the control line CTL4\_492 are color-coded in green and red, respectively. The normalized expression z-score scale is indicated by the color key in the top left corner. Blue, down-regulated genes; Yellow, up-regulated genes.

**Table 5.** Selected DEGs showing similar conserved domain with SGTs and GAME genes and found downregulated in EMS4\_901 compared to CTL4\_492.

Protein ID	Description	Conserved domain similar to *	EMS4_901/CTL4_492 (Fold change)
XP_006366795.2	probable glycosyltransferase At5g03795	<i>SGT</i>	-7.568989
XP_006351615.1	UDP-glycosyltransferase 74F2-like	<i>SGT</i>	-6.491434
XP_006367297.1	methyltransferase-like protein 2	<i>SMT</i>	-2.390494
XP_006357028.1	glycosyltransferase family protein 64 C3-like	<i>SGT</i>	-2.380108
XP_006366572.1	beta-1,3-galactosyltransferase 7-like	<i>SGT</i>	-2.208003
XP_006355647.1	probable methyltransferase PMT23	<i>SMT</i>	-2.017099
XP_006358770.1	cytochrome P450 83B1-like	<i>GAME</i>	-7.106240
XP_006352674.1	cytochrome P450 82C4-like	<i>GAME</i>	-2.027230
XP_006355680.1	7-deoxyloganic acid glucosyltransferase-like	<i>SGT</i>	-2.347019
XP_006366409.1	ethylene-responsive factor RAP2-7-like	<i>GAME</i>	-2.863531
NP_001274974.1	flavonoid 3',5'-methyltransferase-like	<i>SGT</i>	-6.488554
XP_006367260.2	UDP-glycosyltransferase 73C5-like	<i>SGT</i>	-2.181309

\* *SGT*, steroidal alkaloid glycosyltransferase gene family; *SMT*, steroidal methyl transferases; *GAME*, glycoalkaloid metabolism gene.

Using FPKM values  $\geq 1$  from the transcriptomes data, the proportion of known transcripts, known genes, novel transcripts and alternatively spliced variants were found to be higher in EMS4\_901 than in CTL4\_492 by 7%, 19%, 31%, and 26%, respectively. Furthermore, a total of 4,972 known transcripts, 12,194 known genes, 305 novel transcripts and 1,407 novel alternative splicing transcripts were found to be shared between CTL4\_492 and EMS4\_90 (Figure 4).

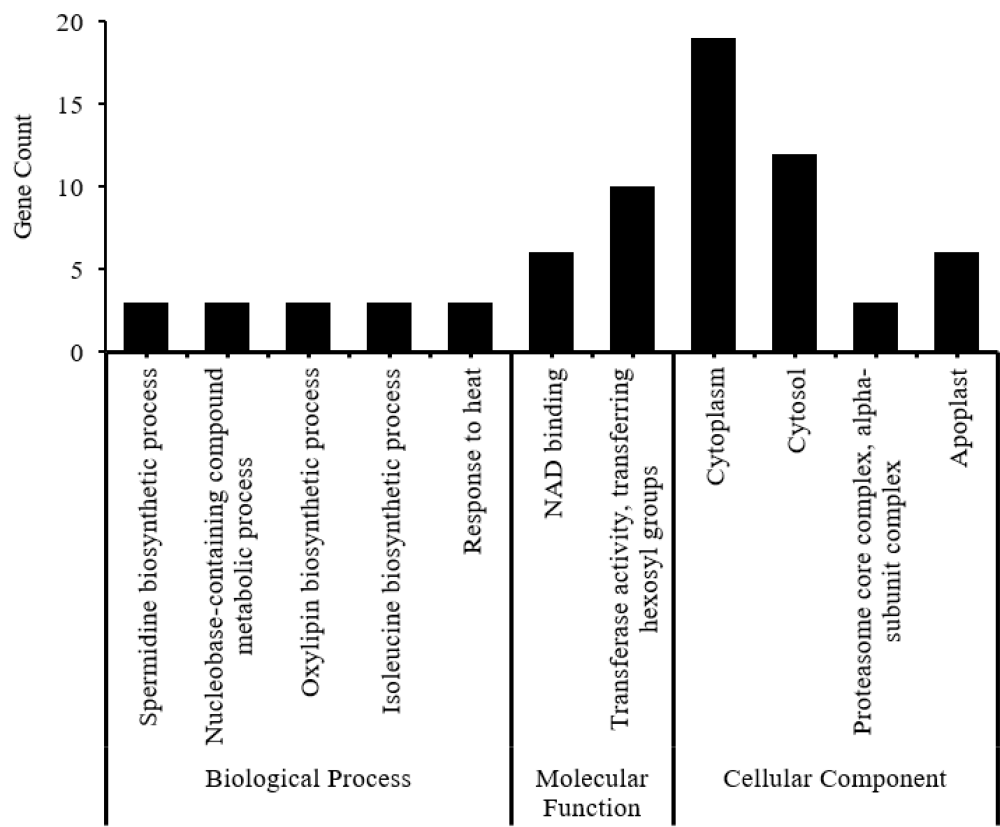
**Figure 4.** Venn diagram showing the differential number of transcripts observed between the high (CTL4\_492) and the low (EMS4\_901) SGA producing diploid potato lines. **A**, known transcripts; **B**, known genes; **C**, novel transcripts; **D**, novel alternative splicing transcripts. The high SGA line (CTL4\_492) and le mutant low SGA (EMS4\_901) are represented by the blue and the red circles, respectively.



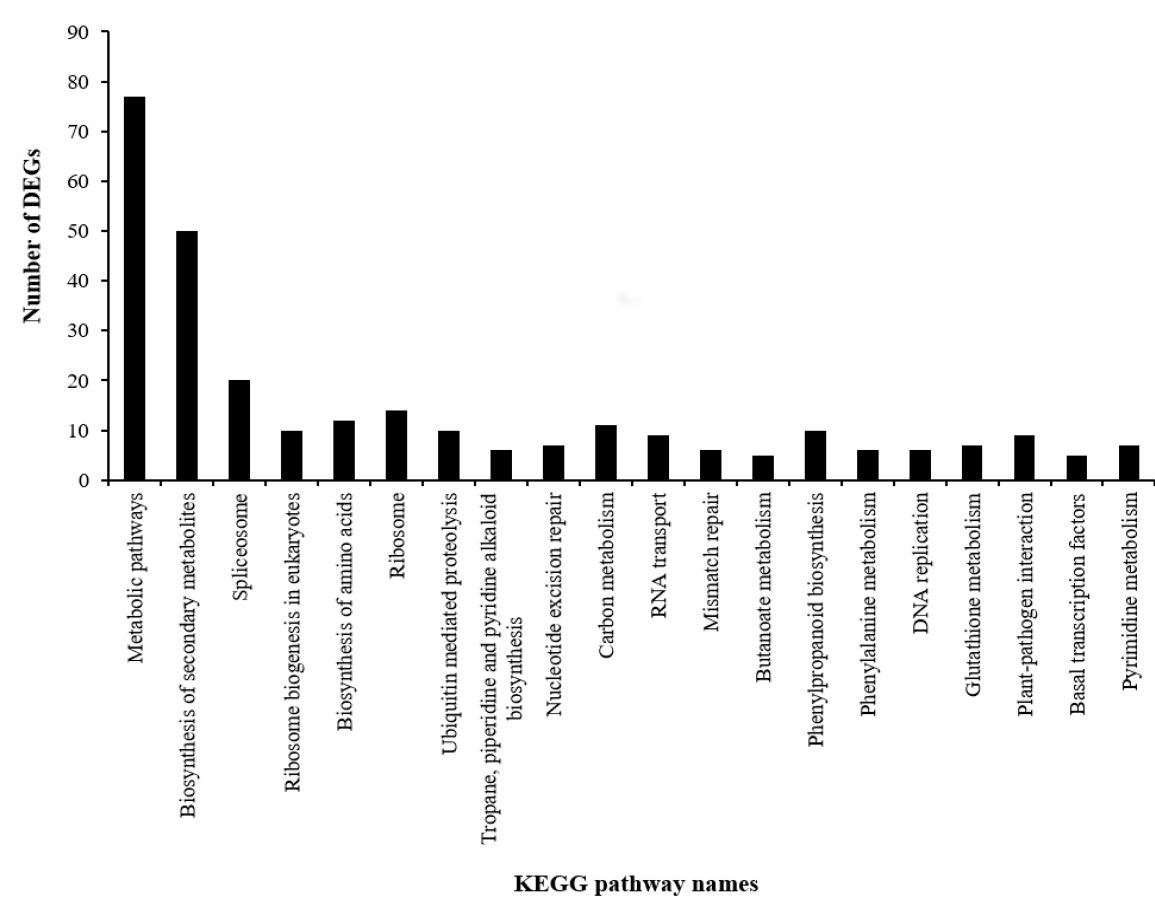
Gene Ontology Classification and KEGG Pathway Analysis of DEGs

To identify GO categories, classify pathways, and assign biological functions associated with the 1096 sets of differentially expressed genes across all six replicates, GO and KEGG pathway enrichment analysis were performed using DAVID tools. The GO enrichment analysis showed 71 DEGs classified into 11 GO terms, with 5, 2, and 4 GO terms associated with biological processes (15 DEGs), molecular functions (16 DEGs), and cellular components (40 DEGs), respectively (Figure 5). KEGG pathway analysis assigned 287 DEGs with high confidence (Bonferroni corrected  $P < 0.05$  and  $FDR < 0.05$ ) to 20 KEGG pathways, and genes involved in metabolic pathways and biosynthesis of secondary metabolites accounted for 44% of the 287 DEGs (Figure 6). Of interest was the finding of significant enrichment of DEGs in various secondary metabolite biosynthesis pathways including phenylpropanoid biosynthesis ( $p < 0.001$ ), isoquinoline alkaloid biosynthesis ( $p < 0.01$ ) and tropane, piperidine and pyridine alkaloid biosynthesis ( $p < 0.05$ ) (Figure 7). DEGs corresponding to UDP-glycosyltransferases (XP\_006367260.2 and XP\_006346605.1) were found to be enriched under the biosynthesis of secondary metabolites (sot01110) pathways (Supplementary Table S13).

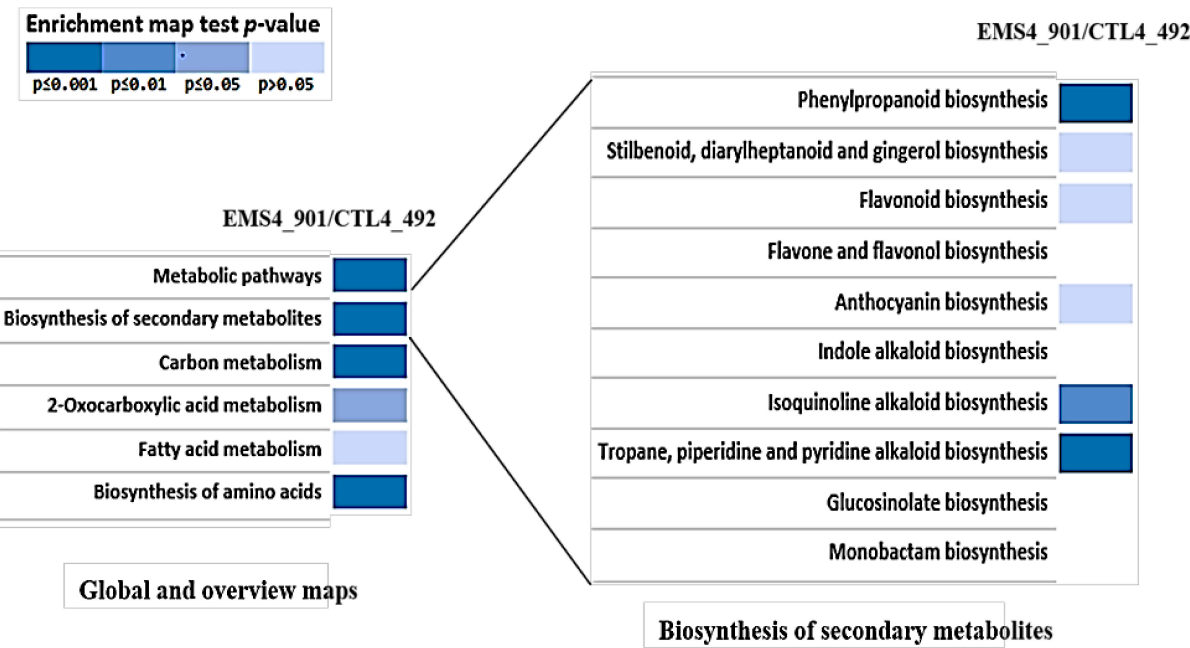
Gene Function Classification (GO)



**Figure 5.** GO term classification of DEGs between the between the high (CTL4\_492) and the low (EMS4\_901) SGA producing diploid potato lines using David Tool. Bar plots represent the number of DEGs enrichment in each GO term gene in each biological process, molecular function and cellular component.



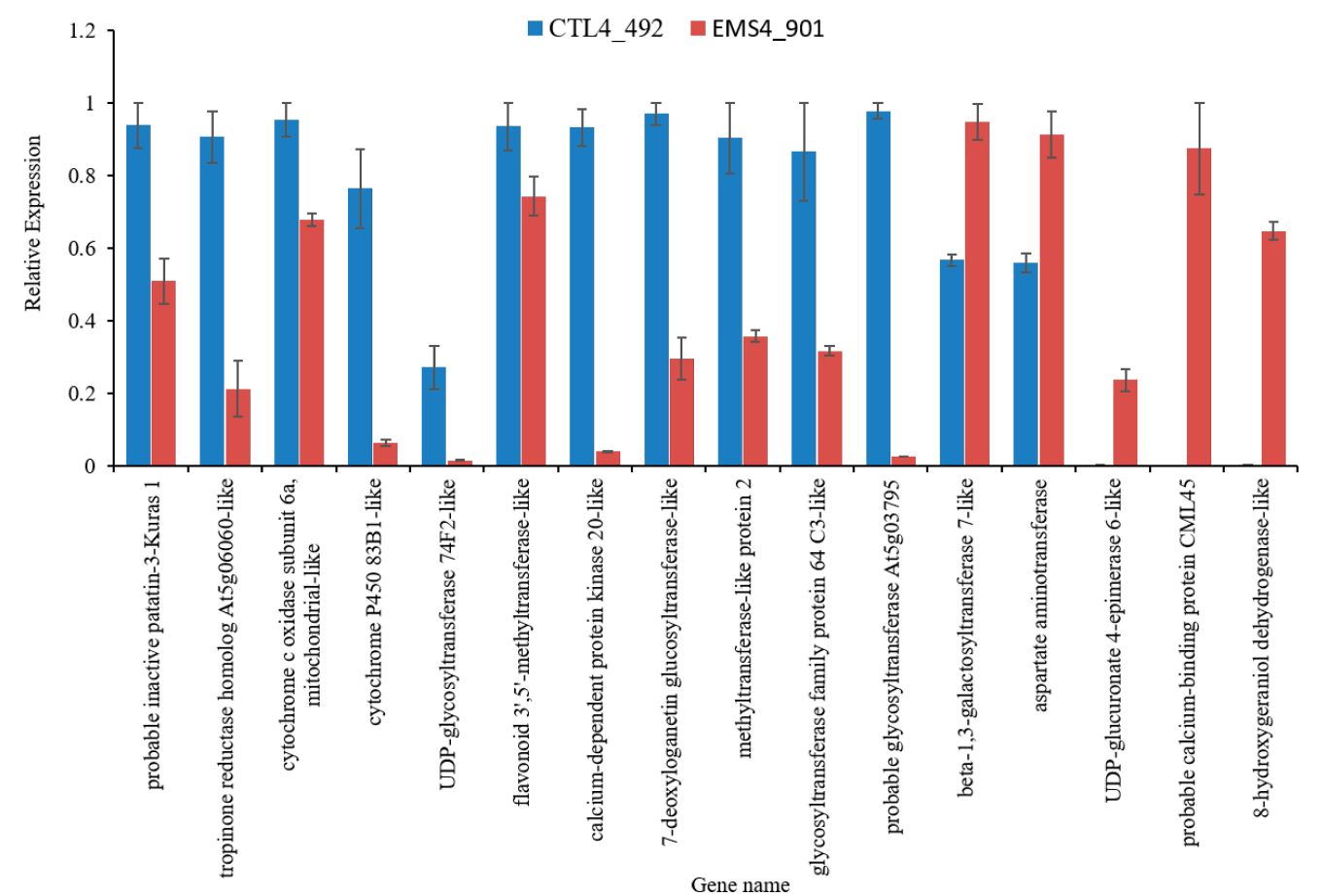
**Figure 6.** KEGG pathway enrichment of DEGs between the low (EMS4\_901) and the high (CTL4\_492) SGA producing diploid potato lines. Only the top 20 annotated pathways (FDR < 0.05) are shown.



**Figure 7.** Gene-set enrichment analysis showing significantly enriched KEGG pathways in the low (EMS4\_901) and the high (CTL4\_492) SGA producing diploid potato lines. An emphasis was put on secondary metabolite biosynthesis pathway. Color-codes indicate levels of enrichment significance as shown in the top left corner as expressed by the expression ratio of EMS4\_901/CTL4\_492.

RNAseq Data Validation Using Quantitative RT-PCR

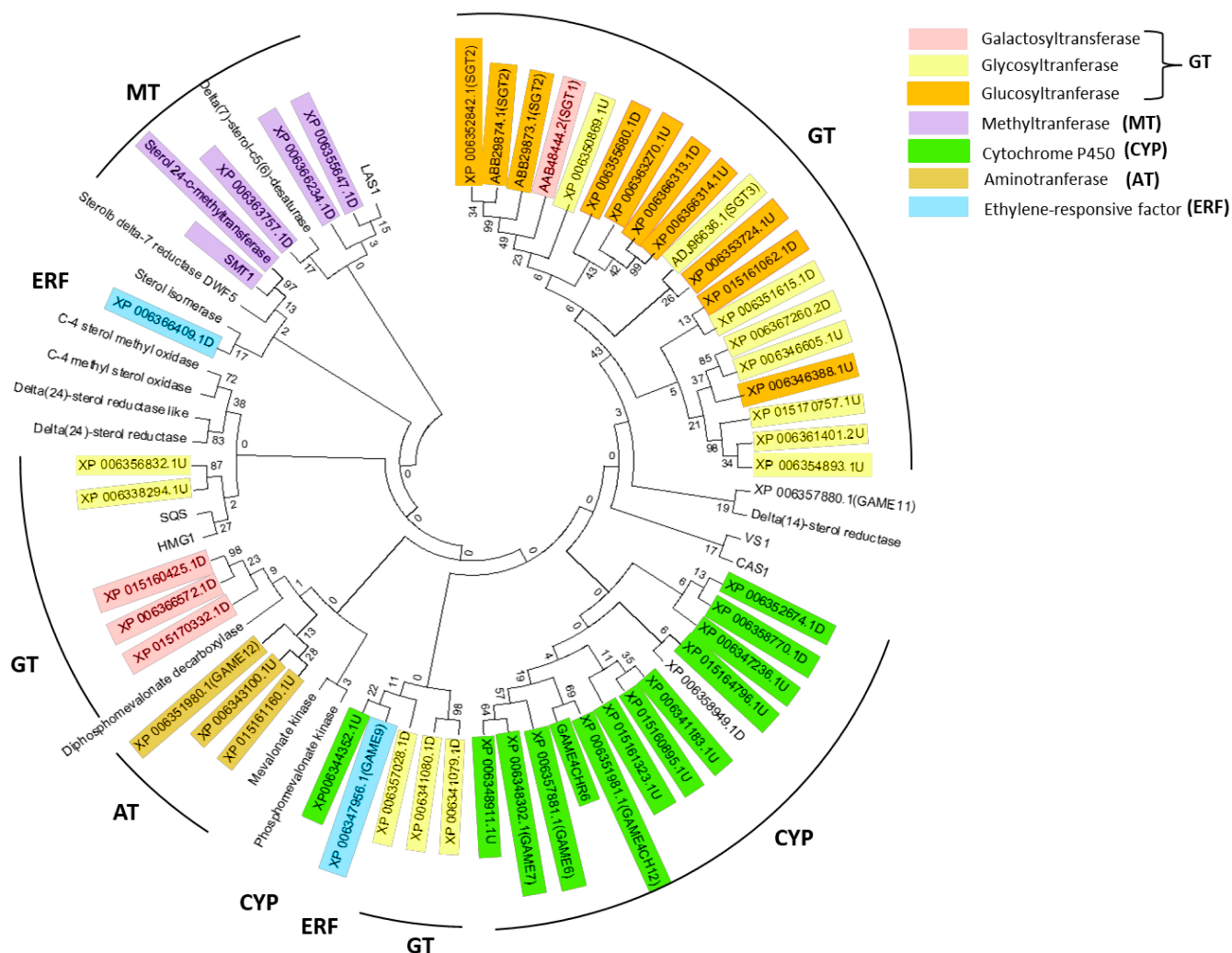
To validate the RNAseq data, qPCR targeting 16 genes showed consistent gene expression pattern between the RNAseq profiling and qPCR analyses for 15 of the 16 genes tested (Figure 8; Table 6). The expression of *beta-1,3- galactosyltransferase 7-like* was the only gene found to be upregulated in EMS4\_901 in the qPCR, whereas it was downregulated from the RNseq profiling. Transcription levels were found to be very low in CTL4\_492 for *UDP-glucuronate 4-epimerase 6-like*, *probable calcium-binding protein CML-45* and *8-hydroxygeraniol dehydrogenase –like* when compared to EMS4\_901 (Figure 8).



**Figure 8.** Quantitative RT-PCR expression patterns for sixteen genes randomly selected from the RNAseq data for validation. The gene expression data are expressed as relative expression from three independent replicates and normalized to the *18S RNA* expression. Error bars indicate the standard deviation of means ( $n = 3$ ).

**Table 6.** RNAseq gene expression validation using qPCR. Fold change expression obtained by qPCR is shown along with the expression pattern observed in the RNAseq profiling.

S. No	Protein ID	Description	Fold change (EMS4_901/CTL4_492) by RNAseq	Fold change (EMS4_901/CTL4_492) by qPCR
1	XP_015161448.1	probable inactive patatin-3-Kuras 1	-170.0	-1.9
2	XP_006352698.1	tropinone reductase homolog At5g06060-like	-15.0	-4.2
3	XP_006367050.1	cytochrome c oxidase subunit 6a, mitochondrial-like	-9.3	-1.4
4	XP_006358770.1	cytochrome P450 83B1-like	-7.1	-12
5	XP_006351615.1	UDP-glycosyltransferase 74F2-like	-6.5	-19.6
6	XP_006363473.1	flavonoid 3',5'-methyltransferase-like	-6.5	-1.3
7	XP_006364742.1	calcium-dependent protein kinase 20-like	-5.3	-25
8	XP_006366313.1	7-deoxyloganetin glucosyltransferase-like	-3.7	-3.3
9	XP_006367297.1	methyltransferase-like protein 2	-2.4	-2.5
10	XP_006357028.1	glycosyltransferase family protein 64 C3-like	-2.4	-2.7
11	XM_006366733.2	probable glycosyltransferase At5g03795	-8	-40
12	XP_006366572.1	beta-1,3-galactosyltransferase 7-like	-2.2	1.7
13	XP_006343100.1	aspartate aminotransferase	2.6	1.6
14	XP_006349916.1	UDP-glucuronate 4-epimerase 6-like	8.2	1180
15	XP_006338443.1	probable calcium-binding protein CML45	15	8740
16	XP_006364647.1	8-hydroxygeraniol dehydrogenase-like	37	539



**Figure 9.** Phylogenetic relationship between KEGG-enriched DEGs and known SGA biosynthesis genes. Maximum Likelihood method based on the JTT matrix-based model [1] was used to build the tree. Bootstrap values are indicated on the tree. GT, glycosyltransferase, CYP, cytochrome P450, MT, methyltransferase, AT, Amino transferase, ERF, ethylene responsive factor. Gene IDs followed by number 1 or 2 and letter D or U indicate that 1 or 2 transcripts of the genes were down-regulated (D) or up-regulated (U).

### Linking Transcriptome Data to Loss-of-function in SGA Biosynthetic Genes

To link EMS-induced mutations in the low SGA line with its RNA transcriptomic profile and identify putative DEGs associated with SGA biosynthesis, the phylogenetic tree constructed using 39 KEGG-enriched DEGs putatively involved in SGA metabolism and 28 genes known for their roles in SGA biosynthesis showed five main groups representing five gene families, namely glycosyltransferases (GT), methyltransferases (MT), cytochrome P450 (CYP), aminotransferases/transaminases (AT), and ethylene -responsive transcription factor (ERF) (Figure 9). Of the 22 DEGs with glycosyltransferase activity, 14 strongly clustered with the known *SGT1*, *SGT2* and *SGT3* under the GT group. The rest of the 8 DEGs with GTs



activity were spread into 3 small clusters. Similarly to the GT group, 8 of the 9 DEGs with cytochromes activity clustered with the GAME genes known as cytochromes under the CYP group. All DEGs with amino acyl transferase activity formed a group with the known amino acyl transferase (AT group). The DEGs for methyltransferases and transcription factors for sterol biosynthesis were grouped with known sterol biosynthesis genes under MT and ERF groups, respectively (Figure 9). Interestingly, it was observed that 11 of the 22 DEGs in the GT group, 2 DEGs in the CYP group, 3 DEGs in the MT group, and the single ERF were found to be down-regulated in the RNAseq transcriptomic analysis.

## DISCUSSION

Anti-nutritional factors including SGAs have always been of great concerns for potato breeders, retailers, processors, and consumers [7,9,69]. To address this issue, different approaches including traditional breeding [70,71], RNAi [24,72], TALENs [31], and transgenic [39] approaches have been attempted. Whereas the traditional breeding methods are more sustainable, they are also time-consuming and less efficient [73]. TALENs and RNAi are effective but costly, demand technical expertise, and generate limited number of germplasm with potential off-target effects [74–77], and the transgenic approaches have always been of concern to public due to their inherent nature [78,79]. In the current study, a diploid potato germplasm panel generated by a conventional mutagenesis approach using EMS mutagenesis [44] was characterized at the genomic and phenotypic levels for the identification of mutant lines altered in SGA production. Using next generation sequencing (NGS) of SGA biosynthesis genes and SGA profiling, we show that mutation of *GAME4CH6* or *SGT3* is sufficient to alter SGA profiles. To our knowledge, the study is unique in its approach by screening a large mutant diploid potato population and identifying mutant lines with altered SGA profiles.

The generation of mutant alleles in potato using EMS mutagenesis has been scarce and is relatively new [44,80]. The use of the chemical mutagenesis approach in plant reverse genetics combined with next generation sequencing allow a high throughput screening capacity of germplasm [81,82]. In the current report, an F<sub>1</sub> diploid mutant potato population was screened by Reverse genetics using 9 SGA biosynthesis genes and, as expected, mostly G to A or C to T transition mutations were observed in all targeted genes as previously reported [81]. The calculated mutation frequency in the targeted genes ranged from 1/16 Kb to 1/341 Kb, ranges that fall within those recently reported in flax [48,83], rapeseed [84,85] and tomato [86] treated with different EMS doses. A total of 20% of the EMS-induced SNP variations were located in exons, of which 46% were functional and 22% of these functional exonic mutations were homozygous, thus having potential to cause loss-of- [87] or gain-of-function effects [88]. In fact, although DNA mutations can alter all amino acids, not all substitutions have the same mutational effects on gene

functionalities [89]. Hence, depending on the amino acids, homozygous missense mutations with low or medium mutational effects can be either neutral with no effect or hypermorphic, in which case the mutation may cause an increased contribution of the mutated gene to the original function [48,90]. The high heterozygosity (78%) of the functional mutations observed here is in agreement with the F<sub>1</sub> nature and the clonally propagated option used for the screened plants. The wide variety of mutations detected in this study, including nonsense mutations via deletions and frameshift, missense mutations via transitions, insertions and deletions demonstrates the effectiveness of EMS-mutagenesis in developing genetically diverse germplasm with novel allelic variants. In the current study, functional mutations were found in almost all key SGA biosynthetic genes as also likely occurred in other metabolic pathways. This finding highlights the high potential of this population as a premium genetic resource for potato breeding programs and for the gene functional characterization in potato by plant biologists.

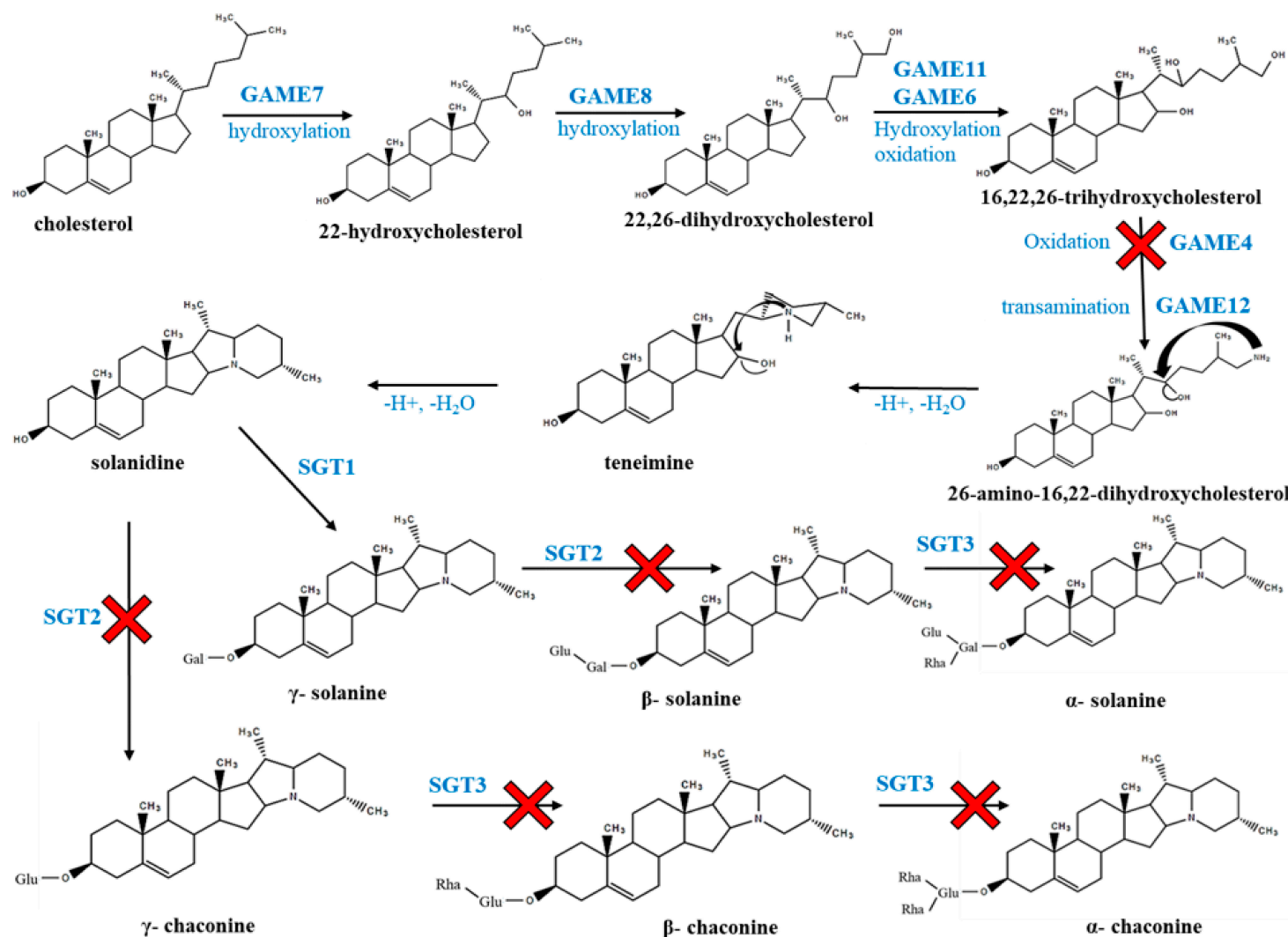
By phenotyping a set of EMS-treated clones, a wide range of SGA patterns were observed in tuber periderm, and 14% of the EMS-treated lines showed SGA contents lower or equal to commercial varieties whereas 21% of the lines showed lower SGA content than the safety limit established by USDA [11]. SGA content in the commercial potato cultivars was shown to vary from 58 to 71 mg/100 g DW, which is lower than the threshold (100 mg/100 g DW) set for a safe cultivar release, in agreement with the previously reported tuber periderm SGA content [15,22]. Since all plant materials used in the current study were grown, processed and analyzed under the same conditions throughout the study, the observed SGA values most likely reflect the actual SGA phenotype of each line, especially when a highly sensitive method such as UPLC-MS was used. Given the higher toxicity of chaconine compared to solanine, the ratio of chaconine to solanine is very important in potato tubers [91]. The chaconine to solanine ratio was found to be 1:1 in some EMS-treated lines as opposed to the previously reported ratio ranging from 1.2 to 2.6 [22], albeit their total SGA content still remained higher, suggesting a metabolic flow rechanneling in one or the other route after the disruption of one the two SGTs (*SGT1*, *SGT2*). Interestingly, 4 mutant lines showed SGA levels as low as 23 to 50 mg/100 g DW, some of which carried heterozygous missense and frame shift mutations in *SGT3*, *GAME4CH6*, *GAME4CH12*, and/or *GAME6*. Similar observations were previously reported after EMS mutagenesis for metabolic pathways in flax [48], egg plants [92], peppers [93] and tomato [94] for metabolites ranging from SDG lignan to anthocyanin and carotenoid biosynthesis.

By associating genotype to SGA phenotype, potato lines with altered SGA content revealed the presence of mutations in key SGA biosynthetic genes. As such, mutations in *GAME6* and *GAME11* were found alongside with *SGT3* and *SGT2*, respectively, and led to low SGA levels in EMS7\_725 (49 mg/100 g DW) and EMS7\_827 (66 mg/100 g DW). *GAME6* oxidizes the

22,26-dihydroxycholesterol at C22 whereas *GAME11* hydroxylates at C16 leading to the 16, 22, 26-trihydroxycholestreol [37]. The impairment of *GAME6* and *GAME11*, even at a heterozygous level, seems to have a critical impact on SGA production in these two lines despite a potential roles for *SGT2* and *SGT3* mutations. Likewise, some of the EMS-treated lines (EMS7\_690, EMS7\_805) carried heterozygous missense mutations in *GAME12* and showed low SGA content (51 mg/100 g DW, 68 mg/100 g DW, respectively), whereas some other EMS-treated lines (EMS3\_288) showed high SGA content (235 mg/100 g DW) despite carrying heterozygous missense mutations in *GAME12*, a gene encoding for the enzyme that incorporates nitrogen into the steroidal alkaloid aglycone intermediate backbone [34]. It is known that the transamination activity is critical for glycoalkaloid production [32,34,95]. Here, the low SGA lines carrying mutations in *GAME12* also showed mutations in *GAME4CH12* and *SGT2* (for EMS7\_690) and in *GAME6*, *GAME4CH12*, and *SGT1* (for EMS7\_805). We therefore speculate that the heterozygous mutation present in *GAME12* may have not altered its function enough to significantly block the pathway and that the alternate allelic variant within the heterozygous background seems to be functional enough for proceeding with the production of 26-amino-furostanol prior to its glycosylation, as it could be seen for line EMS3-288 carrying mutation only in *GAME12*. Alternatively, a yet unknown transaminase may have also contributed to the transamination process along with the heterozygous allele of *GAME12* (Figure 10). Hence, we conclude that the low SGA phenotype observed in EMS7\_690 (51 mg/100 g DW), and in EMS7\_805 (68 mg/100 g DW) may be associated only with the effects of heterozygous mutations in more than one of *GAME6* and *GAME4CH12* genes. Contrary to *GAME12*, the missense heterozygous mutation in *SGT3* and *GAME7* on one hand (EMS4\_403; 23 mg/100 g DW) and the frameshift heterozygous mutations leading to stop codon in *GAME4CH6* and *SGT2* on the other hand (EMS4\_901; 50.4 mg/100 g DW), led both to low SGA phenotypes. *SGT3* adds rhamnose units to both solanine and chaconine [96] and is therefore critical at this glycosylation step, whereas *GAME7* hydroxylates cholesterol into 22-hydroxycholesterol in the early step of SGA biosynthesis. Taken together, heterozygous mutations of both genes may have led to the low SGA phenotype observed in EMS4\_403, a plant producing very small tubers with poor yield (not shown). *GAME4a* (CYP88B) [66], oxidases furostanol-type saponin aglycone into furostanol-26-aldehyde that serves as a substrate for the transamination by *GAME12* [34]. The frameshift mutations identified in *SGT2* and *GAME4CH6* and observed in EMS4\_901 (50.4 mg/100 g DW) were located upstream to the PSPG and CYP450 conserved active domains of these two genes, respectively. This observation supports and increases the likelihood for an alteration in the gene function by these frameshift mutations. But, uncertainties around the mutation of which of the two genes (*SGT2* or *GAME4CH6*) that may have led to the low SGA level in this EMS4\_901 line can arise. It is well known that mutation of either *SGT1* or

*SGT2* does not alter the total SGA level [36] and that RNAi-mediated silencing of *GAME4* gene resulted in SGA reduction in both potato leaves and tubers [34,66]. It is thus reasonable to assume that the low SGAs observed in the line EMS4\_901 is mostly attributable to the frameshift mutation in *GAME4CH6*. However, the roles for mutations in other genes including uncharacterized transcription factors regulating SGA biosynthetic genes can not be ruled out, and only a formal complementation test in a wild type would provide a definite proof to this assumption. Nonetheless, we conclude that mutations of *GAME4CH6* or *SGT3* are critical for SGA production (Figure 10) and that the EMS4\_901 mutant line thus becomes an ideal candidate for molecular and phenotypic comparative analysis between low and high SGA producing lines. As such, by comparing EMS4\_901 with its non-treated control CTL4\_492, a high SGA producing line, altered and contrasting transcriptomic patterns were observed between the two lines and downregulated metabolic pathways were observed in the mutant line compared to the wild type. These data were validated by qPCR. Importantly, DEGs pertinent to SGA biosynthesis were found to form tight clusters with the known *SGT1*, *SGT2* and *SGT3*, cytochromes, *GAME* genes or transcription factors playing roles in SGA biosynthesis, linking transcriptomic DEGs and SGA biosynthetic genes. Moreover many of these gene transcripts were found to be downregulated in the mutant line, providing further proof of the mutant status and altered metabolism of this EMS line.

Taken together, this study is the first to demonstrate that EMS-mutagenesis can be successfully used to alter the SGA biosynthetic pathway and the content of SGA in the tubers of diploid potatoes. It also paves the way for further functional studies such as complementation tests or gene editing. Given the high level of heterozygous mutant alleles observed and the clonal propagation process, most of these mutations can be maintained in the somaclones. Nonetheless, self-pollination, whenever possible, will be required to fix the heterozygous mutations as homozygous for the true potato seed production and for developing inbred cultivars as performed in corn. However, this will require solving first the self-incompatibility issues frequently encountered in wild diploid potatoes [41] and a full genome sequencing will help uncovering all potential mutations in the genome, including the promoter regions toward a gene regulatory network analysis. The current study focused only on the SGA biosynthetic pathway. But, many other biological processes may have also been altered within the population as shown by the wide morphological variations already reported [44], making this mutant population a valuable genetic resource for potato biologists and breeders for the functional characterization of biological processes such as light-induced greening or self-compatibility toward mutant diploid potato cultivar development.



**Figure 10.** EMS-induced alteration of SGA biosynthesis genes leading to  $\alpha$ -solanine and  $\alpha$ -chaconine from cholesterol in potato [34,97]. Red "X" marks indicate the steps/genes showing EMS-induced frameshift mutations in *GAME4* and *SGT2* genes and missense mutation in *SGT3* gene of the low SGA producing EMS4\_901 and EMS4\_403 lines.

## SUPPLEMENTARY MATERIALS

The following supplementary materials are available online at <https://doi.org/10.20900/cbgg20200017>,

Supplementary Table S1. List of genes targeted for screening the potato lines during first and second round of sequencing

Supplementary Table S2. List of primers used in Multiplex PCR I for the amplification of four target genes

Supplementary Table S3. List of primers used in Multiplex PCR II for the amplification four target genes

Supplementary Table S4. List of 22 KASP assays designed to validate SNPs in *AS1*, *AS2*, *SGT1* and *SGT2* genes

Supplementary Table S5. Ion AmpliSeq Custom panel with 107 primer pairs

Supplementary Table S6. List of gene specific primers designed for qRT-PCR validation of 16 differentially expressed genes identified between low SGA (EMS4\_901) and its non-treated high SGA line (CTL4\_492)

Supplementary Table S7. List of 28 known SGA biosynthetic genes used for phylogenetic tree construction

Supplementary Table S8. Details of SNPs in *AS1*, *AS2*, *SGT1* and *SGT2* genes validated using KASP assays in 49 EMS-treated lines.

Supplementary Table S9. List of 61 functional mutations (55 missense, 6 nonsense) observed in 56 individual EMS-treated lines, shown with line and cross IDs

Supplementary Table S10. Summary of SGA content and mutations identified in individual EMS mutant lines

Supplementary Table S11. List of 21, 855 genes expressed in at least one replicate of the CTL4\_492 and EMS4\_901 during RNA sequencing

Supplementary Table S12. List of 1, 096 significantly expressed DEGs identified between CTL4\_492 and EMS4\_901 lines

Supplementary Table S13. Summary of DEGs identified in 20 key pathways

Supplementary Figure S1. UPLC chromatograms of  $\alpha$ -solanine (A),  $\alpha$ -chaconine (B) and solanidine (C) along with calibration curves for  $\alpha$ -solanine (D),  $\alpha$ -chaconine (E) and solanidine (F).

Supplementary Figure S2. The details of the 9 target gene structures used in Ampliseq sequencing for a total of 27,040 bp target. *GAME11*, 1619 bp; *GAME12*, 4226 bp; *GAME4CH12*, 5404 bp; *GAME4CH6*, 4052 bp; *GAME6*, 2341 bp; *GAME7*, 3961 bp; *SGT1*, 1694 bp; *SGT2*, 2018 bp; *SGT3*, 1725 bp. A gap of 1007 bp and 364 bp were present in *GAME4CH12* and *GAME4CH6* sequences, respectively making the total amplifiable target regions of 25,669 bp.

Supplementary Figure S3. Individual pictures composing Figure 1A–G. Pictures are shown enlarged for consultation only.



## DATA AVAILABILITY

All relevant data are within the manuscript and the supporting Supplementary files provided. The authors confirm that all raw variant call data underlying the findings will be made fully available without restriction upon request to any qualified researcher. The raw RNAseq dataset can be found in the GenBank Short Read Archive (SRA) database under the Bioproject ID# PRJNA533334.

## AUTHOR CONTRIBUTIONS

BF: Conception, coordination, design, experiments, data analysis, interpretation and writing of the manuscript; AS and KG: Performed experiments, data analysis, interpretation, drafting and revision of the manuscript; JM: UPLC and mass spectrometry data acquisition and analysis of SGA extracts, revision of manuscript; DM and AS: EMS population nursery and tuber processing; BB: provided original plant materials, coordination, and revision of the manuscript;. All authors read, commented and approved the manuscript.

## CONFLICT OF INTEREST

All authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## FUNDING

This study was conducted as part of the AAFC base project 1320 funded to BF. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## ACKNOWLEDGEMENTS

The authors thank Rick Doucet, Leanne Carter and all summer students at the Charlottetown Research and Development Centre, Charlottetown, PEI, for their help during the course of this project.

## REFERENCES

1. FAgriculture and Agri-Food Canada (AAFC). 2018 Horticulture sector reports: Potato Market Information Review 2016–2017. Available from: [https://www.agr.gc.ca/resources/prod/doc/pdf/potato\\_market\\_review\\_revue\\_marche\\_pomme\\_terre\\_2017-eng.pdf](https://www.agr.gc.ca/resources/prod/doc/pdf/potato_market_review_revue_marche_pomme_terre_2017-eng.pdf). Accessed 2019 Mar 19.
2. Camire ME, Kubow S, Donnelly DJ. Potatoes and human health. *Crit Rev Food Sci Nutr*. 2009 Dec 10;49(10):823-40.
3. Nie X, Zhang G, Lv S, Guo H. Steroidal glycoalkaloids in potato foods as affected by cooking methods. *Int J Food Prop*. 2018 Jan 1;21(1):1875-87.
4. Mottram DS, Wedzicha BL, Dodson AT. Food chemistry: acrylamide is formed in the Maillard reaction. *Nature*. 2002 Oct;419(6906):448-9.

5. Rommens CM, Yan H, Swords K, Richael C, Ye J. Low-acrylamide French fries and potato chips. *Plant Biotechnol J*. 2008 Oct;6(8):843-53. doi: 10.1111/j.1467-7652.2008.00363.x
6. Halford NG, Curtis TY, Muttucumaru N, Postles J, Elmore JS, Mottram DS. The acrylamide problem: a plant and agronomic science issue. *J Exp Bot*. 2012 Feb 16;63(8):2841-51. doi: 10.1093/jxb/ers011
7. Omayio DG, Abong GO, Okoth MW. A review of occurrence of glycoalkaloids in potato and potato products. *Curr Res Nutr Food Sci J*. 2016 Nov 15;4(3):195-202. doi: 10.12944/CRNFSJ.4.3.05
8. Percivala GC, Karimb MS, Dixon GR. Influence of light-enhanced glycoalkaloids on resistance of potato tubers to *Fusarium sulphureum* and *Fusarium solani* var. *coeruleum*. *Plant Pathol*. 1998;47:665-70. doi: 10.1046/j.1365-3059.1998.00284.x
9. Friedman M. Potato glycoalkaloids and metabolites: roles in the plant and in the diet. *J Agric Food Chem*. 2006 Nov 15;54(23):8655-81. doi: 10.1021/jf061471t
10. Valkonen JP, Keskitalo M, Vasara T, Pietilä L, Raman KV. Potato glycoalkaloids: a burden or a blessing? *CRC Crit Rev Plant Sci*. 1996 Jan 1;15(1):1-20. doi: 10.1080/07352689609701934
11. Aziz A, Randhawa MA, Butt MS, Asghar A, Yasin M, Shibamoto T. Glycoalkaloids ( $\alpha$ -chaconine and  $\alpha$ -solanine) contents of selected Pakistani potato cultivars and their dietary intake assessment. *J Food Sci*. 2012 Mar;77(3):T58-61. doi: 10.1111/j.1750-3841.2011.02582.x
12. Smith DB, Roddick JG, Jones JL. Potato glycoalkaloids: some unanswered questions. *Trends Food Sci Technol*. 1996 Apr 1;7(4):126-31. doi: 10.1016/0924-2244(96)10013-3
13. Bejarano L, Mignolet E, Devaux A, Espinola N, Carrasco E, Larondelle Y. 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*. 2000 Nov;80(14):2096-100. doi: 10.1002/1097-0010(200011)80:14<2096::AID-JSFA757>3.0.CO;2-6
14. Dimenstein L, Lisker N, Kedar N, Levy D. Changes in the content of steroidal glycoalkaloids in potato tubers grown in the field and in the greenhouse under different conditions of light, temperature and daylength. *Physiol Mol Plant Pathol*. 1997 Jun 1;50(6):391-402. doi: 10.1006/pmpp.1997.0098
15. Grunenfelder LA, Knowles LO, Hiller LK, Knowles NR. Glycoalkaloid development during greening of fresh market potatoes (*Solanum tuberosum* L.). *J Agric Food Chem*. 2006 Aug 9;54(16):5847-54. doi: 10.1021/jf0607359
16. Petersson EV, Nahar N, Dahlin P, Broberg A, Tröger R, Dutta PC, et al. Conversion of exogenous cholesterol into glycoalkaloids in potato shoots, using two methods for sterol solubilisation. *PLoS One*. 2013 Dec 9;8(12):e82955. doi: 10.1371/journal.pone.0082955
17. Friedman M, McDonald GM, Filadelfi-Keszi M. Potato glycoalkaloids: chemistry, analysis, safety, and plant physiology. *CRC Crit Rev Plant Sci*. 1997 Jan 1;16(1):55-132. doi: 10.1080/07352689709701946

18. Krits P, Fogelman E, Ginzberg I. Potato steroidal glycoalkaloid levels and the expression of key isoprenoid metabolic genes. *Planta*. 2007 Dec 1;227(1):143-50. doi: 10.1007/s00425-007-0602-3
19. Ginzberg I, Tokuhisa JG, Veilleux RE. Potato steroidal glycoalkaloids: biosynthesis and genetic manipulation. *Potato Res*. 2009 Feb 1;52(1):1-5. doi: 10.1007/s11540-008-9103-4
20. Maga JA, Fitzpatrick TJ. Potato glycoalkaloids. *Crit Rev Food Sci Nutr*. 1980 Jul 1;12(4):371-405. doi: 10.1080/10408398009527281
21. Friedman M. Analysis of biologically active compounds in potatoes (*Solanum tuberosum*), tomatoes (*Lycopersicon esculentum*), and jimson weed (*Datura stramonium*) seeds. *J Chromatogr A*. 2004 Oct 29;1054(1-2):143-55. doi: 10.1016/j.chroma.2004.04.049
22. Friedman M, Roitman JN, Kozukue N. Glycoalkaloid and calystegine contents of eight potato cultivars. *J Agric Food Chem*. 2003 May 7;51(10):2964-73. doi: 10.1021/jf021146f
23. Chowański S, Adamski Z, Marciniak P, Rosiński G, Büyükgüzel E, Büyükgüzel K, et al. A review of bioinsecticidal activity of Solanaceae alkaloids. *Toxins*. 2016 Mar;8(3):60. doi: 10.3390/toxins8030060
24. Nakayasu M, Umemoto N, Ohyama K, Fujimoto Y, Lee HJ, Watanabe B, et al. A dioxygenase catalyzes steroid 16 $\alpha$ -hydroxylation in steroidal glycoalkaloid biosynthesis. *Plant Physiol*. 2017 Sep 1;175(1):120-33. doi: 10.1104/pp.17.00501
25. Nakayasu M, Akiyama R, Lee HJ, Osakabe K, Osakabe Y, Watanabe B, et al. Generation of  $\alpha$ -solanine-free hairy roots of potato by CRISPR/Cas9 mediated genome editing of the St16DOX gene. *Plant Physiol Biochem*. 2018 Oct 1;131:70-7. doi: 10.1016/j.plaphy.2018.04.026
26. Wang CC, Sulli M, Fu DQ. The role of phytochromes in regulating biosynthesis of sterol glycoalkaloid in eggplant leaves. *PLoS One*. 2017 Dec 13;12(12):e0189481. doi: 10.1371/journal.pone.0189481
27. Lichtenthaler HK, Rohmer M, Schwender J. Two independent biochemical pathways for isopentenyl diphosphate and isoprenoid biosynthesis in higher plants. *Physiol Plant*. 1997 Nov;101(3):643-52. doi: 10.1111/j.1399-3054.1997.tb01049.x
28. Ginzberg I, Thippeswamy M, Fogelman E, Demirel U, Mweetwa AM, Tokuhisa J, et al. Induction of potato steroidal glycoalkaloid biosynthetic pathway by overexpression of cDNA encoding primary metabolism HMG-CoA reductase and squalene synthase. *Planta*. 2012 Jun 1;235(6):1341-53. doi: 10.1007/s00425-011-1578-6
29. Manrique-Carpintero NC, Tokuhisa JG, Ginzberg I, Holliday JA, Veilleux RE. Sequence diversity in coding regions of candidate genes in the glycoalkaloid biosynthetic pathway of wild potato species. *G3 (Bethesda)*. 2013 Sep 1;3(9):1467-79. doi: 10.1534/g3.113.007146
30. Kumar A, Fogelman E, Weissberg M, Tanami Z, Veilleux RE, Ginzberg I. Lanosterol synthase-like is involved with differential accumulation of steroidal glycoalkaloids in potato. *Planta*. 2017 Dec 1;246(6):1189-202. doi: 10.1007/s00425-017-2763-z

31. Sawai S, Ohyama K, Yasumoto S, Seki H, Sakuma T, Yamamoto T, et al. Sterol side chain reductase 2 is a key enzyme in the biosynthesis of cholesterol, the common precursor of toxic steroidal glycoalkaloids in potato. *Plant Cell*. 2014 Sep 1;26(9):3763-74. doi: 10.1105/tpc.114.130096
32. Cárdenas PD, Sonawane PD, Pollier J, Bossche RV, Dewangan V, Weithorn E, et al. GAME9 regulates the biosynthesis of steroidal alkaloids and upstream isoprenoids in the plant mevalonate pathway. *Nat Commun*. 2016 Feb 15;7(1):1-6. doi: 10.1038/ncomms10654
33. HAN SR, Campbell WF, Salunkhe DK. Ultrastructural localization of solanidine in potato tubers. *J Food Biochem*. 1989 Oct;13(5):377-89. doi: 10.1111/j.1745-4514.1989.tb00406.x
34. Itkin M, Heinig U, Tzfadia O, Bhide AJ, Shinde B, Cardenas PD, et al. Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science*. 2013 Jul 12;341(6142):175-9. doi: 10.1126/science.1240230
35. Umemoto N, Nakayasu M, Ohyama K, Yotsu-Yamashita M, Mizutani M, Seki H, et al. Two cytochrome P450 monooxygenases catalyze early hydroxylation steps in the potato steroid glycoalkaloid biosynthetic pathway. *Plant Physiol*. 2016 Aug 1;171(4):2458-67. doi: 10.1104/pp.16.00137
36. McCue KF, Allen PV, Shepherd LV, Blake A, Maccree MM, Rockhold DR, et al. Potato glycoesterol rhamnosyltransferase, the terminal step in triose side-chain biosynthesis. *Phytochemistry*. 2007 Feb 1;68(3):327-34. doi: 10.1016/j.phytochem.2006.10.025
37. Mariot RF, De Oliveira LA, Voorhuijzen MM, Staats M, Hutten RC, van Dijk JP, et al. Characterization and Transcriptional Profile of Genes Involved in Glycoalkaloid Biosynthesis in New Varieties of *Solanum tuberosum* L. *J Agric Food Chem*. 2016 Feb 3;64(4):988-96. doi: 10.1021/acs.jafc.5b05519
38. Thagun C, Imanishi S, Kudo T, Nakabayashi R, Ohyama K, Mori T, et al. Jasmonate-responsive ERF transcription factors regulate steroidal glycoalkaloid biosynthesis in tomato. *Plant Cell Physiol*. 2016 May 1;57(5):961-75. doi: 10.1093/pcp/pcw067
39. Shepherd LV, Hackett CA, Alexander CJ, McNicol JW, Sungurtas JA, Stewart D, et al. Modifying glycoalkaloid content in transgenic potato–Metabolome impacts. *Food Chem*. 2015 Nov 15;187:437-43. doi: 10.1016/j.foodchem.2015.04.111
40. Srivastava A, Bhardwaj V, Singh BP, Khurana SP. Potato diversity and its genetic enhancement. In: *Gene Pool Diversity and Crop Improvement*. Cham (Switzerland): Springer; 2016. p. 187-226. doi: 10.1007/978-3-319-27096-8\_6
41. Lindhout P, Meijer D, Schotte T, Hutten RC, Visser RG, van Eck HJ. Towards F<sub>1</sub> hybrid seed potato breeding. *Potato Res*. 2011 Dec 1;54(4):301-12. doi: 10.1007/s11540-011-9196-z
42. Eijlander R, Ramanna MS, Jacobsen E. Selection of vigorous and fertile S-homo-and heterozygous tester clones from self-incompatible diploid potato, *Solanum tuberosum* L. *Euphytica*. 1997 Sep 1;97(1):97-111. doi: 10.1023/A:1003041006157

43. Phumichai C, Mori M, Kobayashi A, Kamijima O, Hosaka K. Toward the development of highly homozygous diploid potato lines using the self-compatibility controlling *Sli* gene. *Genome*. 2005 Dec 1;48(6):977-84. doi: 10.1139/g05-066
44. Somalraju A, Ghose K, Main D, Bizimungu B, Fofana B. Development of pre-breeding diploid potato germplasm displaying wide phenotypic variations as induced by ethyl methane sulfonate mutagenesis. *Can J Plant Sci*. 2018 Oct 23;99(2):138-51. doi: 10.1139/cjps-2018-0189
45. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012 Apr;9(4):357. doi: 10.1038/nmeth.1923
46. SnpEff & SnpSift. SnpEff & SnpSift manual. Available from: [http://snpeff.sourceforge.net/SnpEff\\_manual.html#output](http://snpeff.sourceforge.net/SnpEff_manual.html#output). Accessed 2020 Oct 27.
47. Fu YX, Huai H. Estimating mutation rate: how to count mutations? *Genetics*. 2003 Jun 1;164(2):797-805.
48. Fofana B, Ghose K, McCallum J, You FM, Cloutier S. UGT74S1 is the key player in controlling secoisolariciresinol diglucoside (SDG) formation in flax. *BMC Plant Biol*. 2017 Dec;17(1):35. doi: 10.1186/s12870-017-0982-x
49. Liu W, Zhang N, Li B, Fan S, Zhao R, Li LP, et al. Determination of  $\alpha$ -chaconine and  $\alpha$ -solanine in commercial potato crisps by QuEChERS extraction and UPLC-MS/MS. *Chem Pap*. 2014 Nov 1;68(11):1498-504. doi: 10.2478/s11696-014-0617-8
50. Nikolić D, Gödecke T, Chen SN, White J, Lankin DC, Pauli GF, et al. Mass spectrometric dereplication of nitrogen-containing constituents of black cohosh (*Cimicifuga racemosa* L.). *Fitoterapia*. 2012 Apr 1;83(3):441-60. doi: 10.1016/j.fitote.2011.12.006
51. Nishio SY, Hayashi Y, Watanabe M, Usami SI. Clinical application of a custom AmpliSeq library and ion torrent PGM sequencing to comprehensive mutation screening for deafness genes. *Genet Test Mol Biomarkers*. 2015 Apr 1;19(4):209-17. doi: 10.1089/gtmb.2014.0252
52. Ghose K, McCallum J, Sweeney-Nixon M, Fofana B. Histidine 352 (His352) and tryptophan 355 (Trp355) are essential for flax UGT74S1 glucosylation activity toward secoisolariciresinol. *PLoS One*. 2015;10(2). doi: 10.1371/journal.pone.0116248
53. MacroGen. Sample submission guide. Available from: <https://dna.macrogen.com/support/retrieveFaq.do?menuCd=SUP200>. Accessed 2020 Oct 27.
54. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinform*. 2018;43:11.10.1-33. doi: 10.1002/0471250953.bi1110s43
55. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014 Aug 1;30(15):2114-20. doi: 10.1093/bioinformatics/btu170

56. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015 Apr;12(4):357-60. doi: 10.1038/nmeth.3317
57. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015 Mar;33(3):290. doi: 10.1038/nbt.3122
58. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc*. 2016 Sep;11(9):1650. doi: 10.1038/nprot.2016.095
59. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995 Jan;57(1):289-300. doi: 10.1111/j.2517-6161.1995.tb02031.x
60. Huang DW, Sherman BT, Zheng X, Yang J, Imamichi T, Stephens R, et al. Extracting biological meaning from large gene lists with DAVID. *Curr Protoc Bioinform*. 2009 Sep;27(1):13-1. doi: 10.1002/0471250953.bi1311s27
61. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009 Jan 1;37(1):1-3. doi: 10.1093/nar/gkn923
62. Hu J, Jiang J, Wang N. Control of citrus Huanglongbing via trunk injection of plant defense activators and antibiotics. *Phytopathology*. 2018 Feb 19;108(2):186-95. doi: 10.1094/PHYTO-05-17-0175-R
63. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods*. 2001 Dec 1;25(4):402-8. doi: 10.1006/meth.2001.1262
64. Arnqvist L, Dutta PC, Jonsson L, Sitbon F. Reduction of cholesterol and glycoalkaloid levels in transgenic potato plants by overexpression of a type 1 sterol methyltransferase cDNA. *Plant Physiol*. 2003 Apr 1;131(4):1792-9. doi: 10.1104/pp.102.018788
65. Hardigan MA, Laimbeer FP, Newton L, Crisovan E, Hamilton JP, Vaillancourt B, et al. Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc Natl Acad Sci U S A*. 2017 Nov 14;114(46):E9999-10008. doi: 10.1073/pnas.1714380114
66. Nahar N, Westerberg E, Arif U, Huchelmann A, Guasca AO, Beste L, et al. Transcript profiling of two potato cultivars during glycoalkaloid-inducing treatments shows differential expression of genes in sterol and glycoalkaloid metabolism. *Sci Rep*. 2017 Mar 3;7:43268. doi: 10.1038/srep43268
67. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 1985 Jul;39(4):783-91. doi: 10.1111/j.1558-5646.1985.tb00420.x
68. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013 Dec 1;30(12):2725-9. doi: 10.1093/molbev/mst197
69. McCue KF, Breksa A, Vilches A, Belknap WR. Modification of potato steroidal glycoalkaloids with silencing RNA constructs. *Am J Potato Res*. 2018 Feb 1;95(1):9-14. doi: 10.1007/s12230-018-9658-9



70. Sanford LL, Sinden SL. Inheritance of potato glycoalkaloids. *Am Potato J.* 1972 Jun 1;49(6):209-17. doi: 10.1007/BF02864830
71. Sanford LL, Deahl KL, Sinden SL. Glycoalkaloid content in foliage of hybrid and backcross populations from a *Solanum tuberosum* × *S. chacoense* cross. *Am Potato J.* 1994 Apr 1;71(4):225-35. doi: 10.1007/BF02849288
72. Paudel JR, Davidson C, Song J, Maxim I, Aharoni A, Tai HH. Pathogen and pest responses are altered due to RNAi-mediated knockdown of GLYCOALKALOID METABOLISM 4 in *Solanum tuberosum*. *Mol Plant Microbe Interact.* 2017 Nov 12;30(11):876-85. doi: 10.1094/MPMI-02-17-0033-R
73. Bhat JA, Ali S, Salgotra RK, Mir ZA, Dutta S, Jadon V, et al. Genomic selection in the era of next generation sequencing for complex traits in plant breeding. *Front Genet.* 2016 Dec 27;7:221. doi: 10.3389/fgene.2016.00221
74. Jiang M, Instrell R, Saunders B, Berven H, Howell M. Tales from an academic RNAi screening facility; FAQs. *Brief Funct Genomics.* 2011 Jul 1;10(4):227-37. doi: 10.1093/bfpg/elr016
75. Senthil-Kumar M, Mysore KS. Caveat of RNAi in plants: the off-target effect. In: *InRNAi and Plant Gene Function Analysis*. Totowa (US): Humana Press; 2011. p. 13-25. doi: 10.1007/978-1-61779-123-9\_2
76. Gupta RM, Musunuru K. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *J Clin Invest.* 2014 Oct 1;124(10):4154-61. doi: 10.1172/JCI72992.
77. Boettcher M, McManus MT. Choosing the right tool for the job: RNAi, TALEN, or CRISPR. *Mol Cell.* 2015 May 21;58(4):575-85. doi: 10.1016/j.molcel.2015.04.028
78. Rani SJ, Usha R. Transgenic plants: Types, benefits, public concerns and future. *J Pharm Res.* 2013 Aug 1;6(8):879-83. doi: 10.1016/j.jopr.2013.08.008
79. Wunderlich S, Gatto KA. Consumer perception of genetically modified organisms and sources of information. *Adv Nutr.* 2015 Nov;6(6):842-51. doi: 10.3945/an.115.008870
80. Muth J, Hartje S, Twyman RM, Hofferbert HR, Tacke E, Prüfer D. Precision breeding for novel starch variants in potato. *Plant Biotechnol J.* 2008 Aug;6(6):576-84. doi: 10.1111/j.1467-7652.2008.00340.x
81. Sikora P, Chawade A, Larsson M, Olsson J, Olsson O. Mutagenesis as a tool in plant genetics, functional genomics, and breeding. *Int J Plant Genomics.* 2011;2011. doi: 10.1155/2011/314829
82. Vicente-Dolera N, Troadec C, Moya M, del Rio-Celestino M, Pomares-Viciano T, Bendahmane A, et al. First TILLING platform in Cucurbita pepo: a new mutant resource for gene function and crop improvement. *PLoS One.* 2014;9(11):e112743. doi: 10.1371/journal.pone.0112743
83. Chantreau M, Grec S, Gutierrez L, Dalmais M, Pineau C, Demailly H, et al. PT-Flax (phenotyping and TILLinG of flax): development of a flax (*Linum usitatissimum* L.) mutant population and TILLinG platform for forward and reverse genetics. *BMC Plant Biol.* 2013 Dec 1;13(1):159. doi: 10.1186/1471-2229-13-159
84. Harloff HJ, Lemcke S, Mittasch J, Frolov A, Wu JG, Dreyer F, et al. A mutation screening platform for rapeseed (*Brassica napus* L.) and the detection of

- sinapine biosynthesis mutants. *Theor Appl Genet*. 2012 Mar 1;124(5):957-69. doi: 10.1007/s00122-011-1760-z
85. Wang N, Wang Y, Tian F, King GJ, Zhang C, Long Y, et al. A functional genomics resource for *Brassica napus*: development of an EMS mutagenized population and discovery of FAE1 point mutations by TILLING. *New Phytol*. 2008 Dec;180(4):751-65. doi: 10.1111/j.1469-8137.2008.02619.x
86. Okabe Y, Asamizu E, Saito T, Matsukura C, Ariizumi T, Brès C, et al. Tomato TILLING technology: development of a reverse genetics tool for the efficient isolation of mutants from Micro-Tom mutant libraries. *Plant Cell Physiol*. 2011 Nov 1;52(11):1994-2005. doi: 10.1093/pcp/pcr134
87. Uchida N, Sakamoto T, Kurata T, Tasaka M. Identification of EMS-induced causal mutations in a non-reference *Arabidopsis thaliana* accession by whole genome sequencing. *Plant Cell Physiol*. 2011 Apr 1;52(4):716-22. doi: 10.1093/pcp/pcr029
88. Bailly A, Wang B, Zwiewka M, Pollmann S, Schenck D, Lüthen H, et al. Expression of TWISTED DWARF 1 lacking its in-plane membrane anchor leads to increased cell elongation and hypermorphic growth. *Plant J*. 2014 Jan;77(1):108-18. doi: 10.1111/tpj.12369
89. Rogers EE, Eide DJ, Guerinot ML. Altered selectivity in an *Arabidopsis* metal transporter. *Proc Natl Acad Sci U S A*. 2000 Oct 24;97(22):12356-60. doi: 10.1073/pnas.210214197
90. Lu C, Ward PS, Kapoor GS, Rohle D, Turcan S, Abdel-Wahab O, et al. IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature*. 2012 Mar;483(7390):474-8. doi: 10.1038/nature10860
91. Haase NU. Glycoalkaloid concentration in potato tubers related to storage and consumer offering. *Potato Res*. 2010 Dec 1;53(4):297-307. doi: 10.1007/s11540-010-9162-1
92. Xiao X, Lin W, Li W, Gao X, Lv L, Ma F, et al. The analysis of physiological variations in M<sub>2</sub> generation of *Solanum melongena* L. Mutagenized by ethyl methane sulfonate. *Front Plant Sci*. 2017 Jan 19;8:17. doi: 10.3389/fpls.2017.00017
93. Arisha MH, Shah SN, Gong ZH, Jing H, Li C, Zhang HX. Ethyl methane sulfonate induced mutations in M<sub>2</sub> generation and physiological variations in M<sub>1</sub> generation of peppers (*Capsicum annuum* L.). *Front Plant Sci*. 2015 Jun 4;6:399. doi: 10.3389/fpls.2015.00399
94. Shikata M, Hoshikawa K, Ariizumi T, Fukuda N, Yamazaki Y, Ezura H. TOMATOMA update: phenotypic and metabolite information in the micro-tom mutant resource. *Plant Cell Physiol*. 2016 Jan 1;57(1):e11. doi: 10.1093/pcp/pcv194
95. Wang CC, Meng LH, Gao Y, Grierson D, Fu DQ. Manipulation of light signal transduction factors as a means of modifying steroidal glycoalkaloids accumulation in tomato leaves. *Front Plant Sci*. 2018 Apr 12;9:437. doi: 10.3389/fpls.2018.00437
96. Cárdenas PD, Sonawane PD, Heinig U, Bocobza SE, Burdman S, Aharoni A. The bitter side of the nightshades: Genomics drives discovery in Solanaceae

- steroidal alkaloid metabolism. *Phytochemistry*. 2015 May 1;113:24-32. doi: 10.1016/j.phytochem.2014.12.010
97. Cárdenas PD, Sonawane PD, Pollier J, Bossche RV, Dewangan V, Weithorn E, et al. GAME9 regulates the biosynthesis of steroidal alkaloids and upstream isoprenoids in the plant mevalonate pathway. *Nat Commun*. 2016 Feb 15;7:10654. doi: 10.1038/ncomms10654

How to cite this article:

Somalraju A, Ghose K, Main D, McCallum J, Bizimungu B, Fofana B. Altered Steroidal Glycoalkaloid (SGA) Biosynthesis in Diploid Potatoes as Induced by EMS-Mutagenesis. *Crop Breed Genet Genom*. 2020;2(4):e200017. <https://doi.org/10.20900/cbgg20200017>