Theor Appl Genet (2014) 127:391–405
DOI 10.1007/s00122-013-2226-2

ORIGINAL PAPER

Allelic variation in genes contributing to glycoalkaloid biosynthesis in a diploid interspecific population of potato

Norma Constanza Manrique-Carpintero · James G. Tokuhisa · Idit Ginzberg · Richard E. Veilleux

Received: 10 January 2013 / Accepted: 22 October 2013 / Published online: 5 November 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract

Key message **Variation for allelic state within genes of both primary and secondary metabolism influences the quantity and quality of steroidal glycoalkaloids produced in potato leaves.**

Abstract Genetic factors associated with the biosynthesis and accumulation of steroidal glycoalkaloids (SGAs) in potato were addressed by a candidate gene approach and whole genome single nucleotide polymorphism (SNP) genotyping. Allelic sequences spanning coding regions of four candidate genes [3-hydroxy-3-methylglutaryl coenzyme A reductase 2 (*HMG2*); 2,3-squalene epoxidase; solanidine galactosyltransferase; and solanidine glucosyltransferase (*SGT2*)] were obtained from two potato species differing in SGA composition: *Solanum chacoense* (*chc* 80-1) and *Solanum tuberosum* group Phureja (*phu* DH). An F_2 population was genotyped and foliar SGAs quantified. The concentrations of α-solanine, α-chaconine, leptine I, leptine II and total SGAs varied broadly among F_2 individuals. F_2 plants with *chc* 80-1 alleles for *HMG2* or *SGT2* accumulated significantly greater leptines and total SGAs compared to plants with *phu* DH alleles. Plants with *chc* 80-1 alleles at both loci expressed the greatest levels of total SGAs, α-solanine and α-chaconine. A significant positive correlation was found between α-solanine and α-chaconine accumulation as well as between leptine I and leptine II. A whole genome SNP genotyping analysis of an F_2 subsample verified the importance of *chc* 80-1 alleles at *HMG2* and *SGT2* for SGA synthesis and accumulation and suggested additional candidate genes including some previously associated with SGA production. Loci on five and seven potato pseudochromosomes were associated with synthesis and accumulation of SGAs, respectively. Two loci, on pseudochromosomes 1 and 6, explained phenotypic segregation of α-solanine and α-chaconine synthesis. Knowledge of the genetic factors influencing SGA production in potato may assist breeding for pest resistance.

Introduction

Biologically active secondary metabolites called steroidal glycoalkaloids (SGAs) evolved in potato species as part of their defense against pathogens and insects (Friedman 2006; Nema et al. 2008). Different SGAs, the most common of which are the triose glycosides of solanidine aglycone, α-chaconine and α-solanine, have been reported in *Solanum* (Distl and Wink 2009; Ginzberg et al. 2009). The unusual leptines have been found only in rare accessions of *Solanum chacoense* and are known for conferring resistance to Colorado potato beetle (*Leptinotarsa decemlineata* Say), one of the most serious pests of the potato crop (FAO 2008; Sinden et al. 1980, 1986). Besides their role in plant resistance, SGAs can have toxic effects in humans due to their anticholinesterase and membrane disruption activities (Morris and Lee 1984). For that reason, an upper safe limit of 200 mg of total SGAs per kg fresh weight of potato

Communicated by G. Bryan.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-013-2226-2) contains supplementary material, which is available to authorized users.

N. C. Manrique-Carpintero · J. G. Tokuhisa · R. E. Veilleux (✉)
Department of Horticulture, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA
e-mail: potato@vt.edu

I. Ginzberg
Department of Vegetable Research, The Volcani Center, Agricultural Research Organization, Bet Dagan, Israel

tubers is recognized for the release of commercial cultivars (Valkonen et al. 1996).

SGA accumulation varies in all parts of the potato plant, depending on developmental and environmental factors. (Distl and Wink 2009; Friedman and Dao 1992; Friedman and McDonald 1997; Kolbe and Stephan-Beckmann 1997; Mweetwa 2009; Nema et al. 2008). Interspecific crosses and somatic hybrids have given rise to progeny with unusual types and unexpected levels of SGAs compared to the parents (Laurila et al. 1996; Sanford et al. 1996; Yencho et al. 1998). Molecular markers associated with genomic regions hosting dominant genes with additive effects on SGA biosynthesis as well as recessive genes controlling high content of SGAs have been identified (Boluarte-Medina et al. 2002; Hutvágner et al. 2001; Ronning et al. 1999; Sagredo et al. 2006; Van Dam et al. 2003; Yencho et al. 1998). Even though molecular markers and quantitative trait loci (QTL) have been associated with synthesis and accumulation of SGAs, neither the identification of underlying genes nor marker assisted selection (MAS) has been possible in potato breeding. The continuous phenotypic variations among progeny due to the effects of multiple genes and their interactions, the high level of heterozygosity in cultivated and wild potato species, and polyploidy are the factors that have interfered with mapping genes responsible for SGA variance in composition and levels. The recently sequenced potato genome (The Potato Genome Sequencing Consortium 2011) and the Illumina single nucleotide polymorphism (SNP) genotyping platform developed using elite potato germplasm (Hamilton et al. 2011) enable the identification of candidate genes within targeted genomic regions. Genes involved in the glycoalkaloid biosynthetic pathway, genic regions surrounding QTL, and markers associated with SGAs provide important resources for the identification of candidate genes involved in the biosynthesis and accumulation of SGAs.

The primary metabolic activities of the mevalonic acid, isoprenoid condensation, and sterol pathways contribute to the biosynthesis of various terpenoids including SGAs (Fig. 1). Although the intermediate products of these pathways are precursors in common with a variety of terpenoid products, some of the steps are catalyzed by isoenzymes that are associated with specific terpenoid products. For instance, the enzyme 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR) catalyzes the formation of mevalonic acid. The specific isoenzymes HMG2 and HMG3 are associated with the production of sesquiterpenoid phytoalexin defense compounds, while HMG1 has been associated with SGA accumulation after wounding (Choi et al. 1992, 1994; Krits et al. 2007). Other steps of these pathways are represented by gene families including the oxidation of

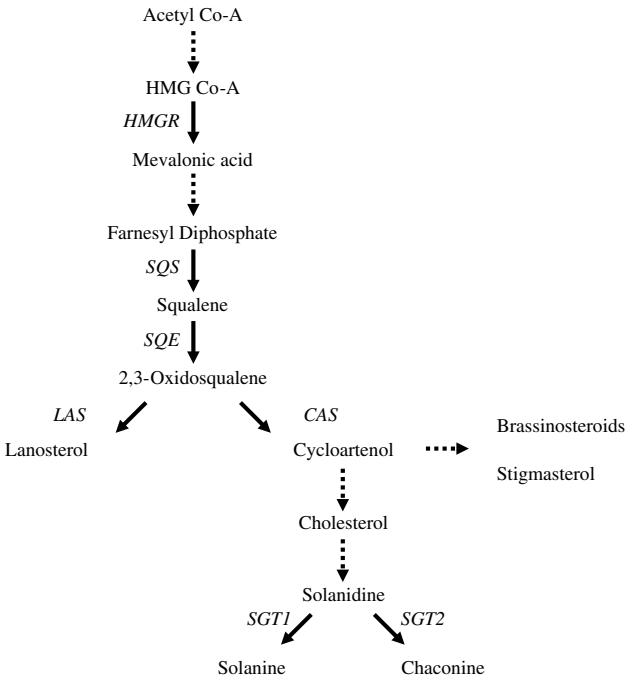

Fig. 1 Potato steroidal glycoalkaloid biosynthetic pathway. *Dashed line arrows* indicate step with multiple enzymatic reactions. *Continuous line arrows* have beside the abbreviation of gene performing the reaction. *HMGR* 3-hydroxy-3-methylglutaryl coenzyme A reductase, *SQS* squalene synthase, *SQE* squalene epoxidase, *LAS* lanosterol synthase and *CAS* cycloartenol synthase, *SGT1* solanidine galactosyltransferase and *SGT2* solanidine glucosyltransferase. Brassinosteroids, stigmasterol and additional phytosterols are major products of sterol biosynthesis in addition to cholesterol

squalene to 2,3-oxidosqualene, catalyzed by squalene epoxidase (SQE) in the isoprenoid condensation stage of triterpene–sterol formation. Inhibition of this enzyme activity decreased sterol biosynthesis (Wentzinger et al. 2002). Cholesterol is the precursor of SGAs (Arnqvist 2007); hence, SQE may have a regulatory role in SGA biosynthesis.

Enzymatic steps involved in the conversion of cholesterol to the SGA aglycone in Solanaceae have recently been suggested that include several steps of hydroxylation, oxidation and transamination (Itkin et al. 2013). Three enzymes that catalyze the final glycosylation reactions of the solanidine skeleton have been identified in potato (McCue et al. 2005, 2007a, b; Moehs et al. 1997). Solanidine galactosyltransferase (SGT1) and solanidine glucosyltransferase (SGT2) catalyze the glycosylation of solanidine to γ-solanine and γ-chaconine, respectively. The final reaction for formation of α-solanine and α-chaconine is mediated by rhamnosyltransferase (SGT3). Functional genomic analysis using antisense constructs of *SGT1* and *SGT2* resulted in alteration of the α-solanine/α-chaconine ratio, but with unchanged total SGA levels (McCue et al.

2007b). This suggests that at this step glycosyltransferase genes only define the specific structures of SGAs, while the total level of SGA may be controlled by genes linked to upstream metabolism in the pathway.

The main goal of this study was to identify allele-specific sequence polymorphisms of four candidate genes involved in the biosynthesis of SGAs (*HMG2, SQE, SGT1* and *SGT2*) for a diploid interspecific hybrid potato population derived from a cross between *S. chacoense* and *S. tuberosum* group Phureja. A second objective was to identify new candidate genes associated with the dissimilar levels of SGA accumulation observed in our F_2 population by whole genome SNP genotype analysis of selected individuals.

Materials and methods

Plant material

Two clones, *S. chacoense* USDA 8380-1 (*chc* 80-1) and *S. tuberosum* group Phureja clone DH OT-B × N–B P5/AR2 (*phu* DH), and an F_2 population from a cross between them were used in this study. *Chc* 80-1 produces many different types of SGAs, some at high levels, whereas *phu* DH does not accumulate SGAs. *Phu* DH is a dihaploid derived from the sexual progeny of two intermonoploid somatic hybrids from *S. tuberosum* group Phureja BARD clones, selected for tuberization under long photoperiod (Haynes 1972; Johnson et al. 2001; Lightbourn and Veilleux 2007; Veilleux 1990) (Figure S1). In addition two *chc* 80-1 and *phu* DH, 11 anther-derived monoploids were used to identify the allele-specific sequences of SGA candidate genes in the parental clones. With only 12 chromosomes the monoploids eliminated confusion between allelic differences and duplicate genes.

Due to self-incompatibility in the F_1 hybrids between *phu* DH × *chc* 80-1, we obtained the F_2 population from crosses between sibling F_1 plants. The F_1 plants were grown under controlled environment (Conviron, Winnipeg) set to 60 % relative humidity, 14 h photoperiod, 250 μmol $m^{-2}s^{-1}$ light intensity and day/night temperatures of 20/16 °C. At 77 days after germination leaves 4–6 from the shoot apex were harvested for SGA extraction and leaf tissue was collected for DNA extraction. Then the F_1 plants were transferred to the greenhouse and crosses between them were done to develop the F_2 population. The total level of SGA quantified in the F_1 plants was used to select specific crosses (Figure S1). The F_2 population was started from 200 seeds of five different families with 40 seed per family and grown under a 16 h photoperiod in a growth chamber. SGA and DNA samples were taken at 45 days after transplant to Deepots (Stuewe & Sons, Tangent, OR).

DNA extraction and PCR amplification

Homologous nucleotide sequences from solanaceous species for *HMG2*, *SQE*, *SGT1* and *SGT2* were obtained from the available data in the GenBank database between 2008 and 2010. These sequences were aligned to observe conserved coding regions for primer design (Table S1). Primers were designed spanning functional domains (Manrique-Carpintero et al. 2013) using DNASTAR® Lasergene 9 core suite software for sequence analysis and assembly. Standard PCR was performed using 100–200 ng of genomic DNA template and products were gel-purified (Qiagen Gel Extraction Kit, Hilden, Germany). PCR products from diploid heterozygous *chc* 80-1 and *phu* DH were cloned in the pGEM-T vector system (Promega, Madison, WI). At least five colonies from the cloned PCR products were sequenced in both directions (Tufts University Core Facility, Boston, MA). PCR products from monoploid lines were sent for direct sequencing in both strands for each sample. Due to the presence of PCR recombinant artifact sequences within *SGT1* reconditioning PCR was used (Judo et al. 1998; Lenz and Becker 2008; Thompson et al. 2002). In this case the final PCR product was derived from two sequential PCR reactions of 20 cycles each. In the second reaction 2 μl of product from the first reaction were used as template instead of genomic DNA. Three different reconditioned PCR reactions were done per DNA template from monoploid lines. PCR products from *phu* DH monoploids were used for direct sequencing, while PCR products from *chc* 80-1 monoploids were cloned and sequenced. In addition to allelic sequencing, flanking sequences neighboring *SGT1* in *chc* 80-1 monoploids were obtained by high-efficiency thermal asymmetric interlaced PCR (hiTAIL-PCR) according to (Liu and Chen 2007).

Identification of allelic sequences and genotyping

Forward and reverse nucleotide sequences from the same PCR sample were aligned to identify consensus sequences. Visual inspection of peak quality in chromatograms was used to resolve conflicting sites in the sequence analysis. The set of consensus sequences identified for each genotype and gene was aligned to identify haplotype alleles. A minimum alignment of at least three identical consensus sequences was used to determine allelic sequences. The sequences were confirmed by comparing with candidate entries in the Genebank. TaqMan® technology and electrophoresis separation of allele-specific PCR products were used to genotype *chc* 80-1 and *phu* DH alleles in the segregating population. Allele-specific primers were designed for tag SNP and polymorphic sites detected in the allelic sequences (Table S2). An allelic genotyping strategy was defined for each gene:

HMG2, Even though tag SNPs that differentiated *HMG2* alleles were identified, the intron size difference between *phu* DH alleles (P1 and P2) and conformational configuration due to a SNP in the *chc* 80-1 allele (C1) were used to screen the F_2 population. DNA fragment analysis of simple PCR products, using primers flanking the *HMG2* intron, was done in the high-resolution QIAxcel system® that discriminated the three alleles.

SQE, TaqMan® SNP genotyping assay and simple PCR-based DRR-GL marker system (Ramkumar et al. 2010) were used to screen *SQE* alleles (P1, C1 and C2) at two SNP positions. The TaqMan® assay for one *SQE* SNP was designed by Applied Biosystems® to differentiate the C2 allele from C1 and P1. PCR and genotyping, by fluorescent allelic discrimination, was performed in an Applied Biosystems real-time 7300 fast instrument. The PCR-based DRR-GL marker system was used to detect the P1 allele from C1 and C2. Two allele-specific primers were designed at the SNP, each with a polymorphism at the last base position from the 3′ end. The primers were used in forward and reverse orientation from the SNP position. The tag SNP was located at a position that divided the gene fragment into two different size portions. In a multiplex reaction, using the regular flanking primers of the gene fragment and the allele-specific primers, we could discriminate a co-dominant amplification pattern for the target SNP.

SGT1, Multiple polymorphic sites in the sequence and a promoter region were used to design allele-specific primers for five alleles at *SGT1* (C1, C2, C3, P1 and P2). Single PCR reactions with allele-specific primers at forward and reverse position were necessary to detect C3 and P2 alleles. Another reaction with flanking primers and one allelic primer was used to screen for C1. Finally, a multiplex reaction differentiated C2 and P1 alleles.

SGT2, To screen *SGT2* alleles (C1 and P1), a region containing a 9 bp deletion and a segment with two SNPs were used to design allele-specific primers. A multiplex reaction with flanking primers and the allele-specific primers discriminated C1 and P1.

Sequence alignment and editing were done using DNASTAR® Lasergene 9 core suite. SNPs were analyzed using DNA Sequence Polymorphism (DnaSP5) version 5.0 (Librado and Rozas 2009). The allelic sequence alignments per gene used for analysis in DnaSP5 were done by CLUSTAL *W* method (Thompson et al. 1994) using Molecular Evolutionary Genetics Analysis (MEGA5) version 5.0 (Tamura et al. 2011). Polymorphisms identified in the *chc* 80-1 and *phu* DH allelic sequences were used to screen the F_2 population. PCR was performed in 15 µl of 1X Immomix™ (Bioline, Taunton, Mass.), 2.5 mM $MgCl_2$, 0.2 µM of each primer, and 50–100 ng of genomic DNA template with standard cycling conditions.

SGA extraction and quantification

After 72 h of lyophilization (Labconco, Kansas City, MO), freeze-dried leaf tissue from two parental lines, nine F_1 plants and 148 F_2 plants was ground and SGAs were extracted following a modified procedure (Edwards and Cobb 1996). Leaf powder (30 mg) and 1 ml of extraction buffer (5 % v/v acetic acid, 0.02 M heptane sulfonic acid) were mixed by vortex and ultrasonication for 2 s (Digital sonifier® cell disruptor, Branson Ultrasonic Corporation, NY, USA with 20 % amplitude). The extract was maintained for 15 min at 1,200 rpm in a microtube thermal-mixer, and centrifuged 3–5 min at 16,000 rpm. The supernatant was filtered (50 µm, Macherey–Nagel Inc., Bethlehem, PA, USA) and the precipitate used for a second cycle of extraction. SGAs were concentrated and cleaned by solid phase extraction using Sep-Pak® classic C_{18} cartridge columns (250 mg) (Waters, Milford, MA, USA). The columns were activated with 5 ml of methanol (MeOH), and equilibrated with 5 ml of extraction buffer. Leaf extract was applied and a series of washes: 7.5 ml of water, 5 ml of 50 mM ammonium bicarbonate (NH_4HCO_3), 5 ml of 50 mM NH_4HCO_3:MeOH (1:1 v/v), and 7.5 ml of water. SGAs were eluted with 1.2 ml of elution buffer (80 % v/v MeOH, 0.5 % v/v formate).

SGAs were separated by high performance liquid chromatography (HPLC, Agilent HP 1200 Series, Santa Clara, CA) on a C-18 reverse-phase column (Agilent Eclipse XDB-C18, 5 µm pore size and 4.6 × 150 mm). SGAs were eluted using a binary gradient system consisting of Solvent A (30 % acetonitrile, 6 mM Tris-HCl, pH 8.0) and Solvent B (80 % acetonitrile, 6 mM Tris-HCl, pH 7.6) at a flow rate of 0.3 ml/min at 25 °C column temperature. The gradient elution was: 0–0.5 min, 0 % B 0.5–8.5 min, 0–30 % B; 8.5–12 min, 30–100 % B; 12–16 min 100 % B, 16–16.5 min, 100–0 % B; 16.5–21 min, 0 % B. Quantification of SGAs was based on peak absorbance area at A_{202nm}, which was converted to µg using a response factor [3.4 ng/mAU, calibration with purified α-solanine and α-chaconine (Sigma) for all four SGAs analyzed (α-solanine, α-chaconine, leptine I and leptine II)] as determined by Mweetwa et al. (2012). Concentrations were expressed in µg of SGAs per mg of dry weight leaf tissue (µg mg^{-1} DW).

SNP chip analysis

Genomic DNA was isolated from a subsample of 43 F_2 plants selected based on F_2 individuals with sufficient viable stored tubers. Different numbers of plants coming from families 3, 5, 7, 8 and 9 were available (2, 1, 5, 23 and 12, respectively). The SNP genotyping facility at Michigan State University processed the samples on an Illumina iScan Reader utilizing the Infinium® HD Assay

Ultra (Illumina, Inc., San Diego, CA) and the Infinium 8303 Potato Array. The 8,303 SNP data were filtered and used for contingency table analysis and ANOVA based on synthesis and accumulation of α-solanine, α-chaconine, leptine I and leptine II. SNPs that were monomorphic for all individuals, and SNPs with a no-call rate >10 % (>5 progeny with missing genotypes) were eliminated from the initial data set. From 3,196 analyzed SNPs per compound, significant SNPs in the contingency tables were selected based on *P* value <0.02, F_2 expected genotypic segregation ratios of 1:2:1 or 1:1 based on Chi-square tests, and strong association of homozygous alleles with synthesis of each compound (0 individuals). In addition, significant SNPs in ANOVAs were selected based on $R^2 > 0.2$, *P* value <0.02, F_2 expected genotypic segregation ratios of 1:2:1 or 1:1 based on Chi-square tests, and SNPs with superscaffold and pseudomolecule information on the potato genome browser (http://solanaceae.plantbiology.msu.edu/cgi-bin/gbrowse/potato/). Finally, superscaffolds with significant SNPs were mapped using the megabase (Mb) physical distances published by Felcher et al. (2012). These physical distances used as reference the recently sequenced potato genome (The Potato Genome Sequencing Consortium 2011). Superscaffolds with significant SNPs were screened on the potato genome browser to identify possible candidate genes related to biosynthesis and/or regulation of SGAs. The new putative candidate genes were listed using the potato genome gene model references per PGSC Assembly Version 3.

Statistical analysis

A completely randomized design was used to grow F_2 and parental plants. The amounts of α-solanine, α-chaconine, leptine I and leptine II were estimated in three subsamples of leaf tissue collected per plant. Three F_2 samples with highly variable subsamples were removed from the analyses. ANOVA of SGA response variables were conducted using allelic state as the source. Allelic sequences per gene, identified in the segregating F_2 population, were consolidated in a unique code from *phu* DH (PP), *chc* 80-1 (CC) and heterozygous (CP). Chi-square analyses were done to confirm phenotypic (presence/absence) and genotypic segregating ratios in the F_2 population. All statistical analyses were done using JMP® 9. SAS Institute Inc., Cary, NC, USA.

Results

Allelic sequences and sequence polymorphism

Allelic sequences from coding region fragments of four candidate genes within the SGA biosynthetic pathway were obtained. PCR amplification, cloning and sequencing from *chc* 80-1 and *phu* DH as well as direct sequencing of monoploid PCR products were used to identify the sequences of *SQE*, *HMG2*, *SGT1* and *SGT2* occurring in heterozygous diploid plants. However, a different strategy: PCR amplification, cloning and sequencing from *chc* 80-1 monoploids, was necessary to identify *SGT1* allelic sequences. Haplotypes defined the alleles for *phu* DH and *chc* 80-1. Thirteen allelic sequences were found for four candidate genes. In *chc* 80-1, a single allele for *HMG2* and *SGT2* and two alleles for *SQE* were found, while in *phu* DH a single allele for *SQE* and *SGT2* and two alleles for *HMG2* and *SGT1* were observed. Analysis of 20 different cloned sequences confirmed the presence of multiple alleles for *SGT1* in *chc* 80-1 monoploids due to possible gene duplication. In two different monoploids of *chc* 80-1 we found the same three allelic sequences, suggesting that there were three copies of this gene in the *chc* 80-1 parent. In fact allelic-specific primers designed for each of the three copies amplified fragments that always co-segregated in both *chc* 80-1 monoploids and the F_2 individuals. In addition, we used hiTAIL-PCR analysis to identify two different promoter regions and a specific *chc* 80-1 sequence for the 3′untranslated region (UTR) adjacent these sequences (data not shown).

Alignment of allelic sequence fragments revealed synonymous, non-synonymous, and indel features (Table 1). In 13 allelic sequences we found 106 SNPs generating 108 mutations. An average of 687 nucleotide sites was analyzed per gene. The sequenced fragments for *HMG2* and *SQE* were within regions with one and three introns, respectively. The proportion of SNPs was greater in the introns than in the exons for these genes (62 and 86 %, respectively). Regarding the frequency of SNPs only in the exons for the four candidate genes, *HMG2* (1/40 bp), *SQE* (1/134 bp), *SGT1* (1/16 bp) and *SGT2* (1/63 bp), the greatest frequency of SNPs was in *SGT1* and the lowest in *SQE*. Genes in primary metabolism, *HMG2* and *SQE*, showed more silent (synonymous and non-coding region) SNPs than non-synonymous SNPs. On the other hand, genes in secondary metabolism, *SGT1* and *SGT2*, had more non-synonymous than synonymous SNPs. Other sequence polymorphisms found were indels, three in intronic regions and three in exonic regions. Two of the indels were located in the exon of one *chc* 80-1 allele; these shifted the reading frame and produced stop codons.

F_2 population phenotypic analysis

Total SGA, α-solanine, α-chaconine, leptine I and leptine II content were determined for parental lines and their F_2 progeny (Table 2). The foliar extract of the *phu* DH female parent did not exhibit detectable SGAs, whereas the *chc*

Table 1 Summary statistics of analyzed allelic sequence sites (excluding sites with gaps on the total alignment), and sequence polymorphism found in four candidate genes of the glycoalkaloid biosynthetic pathway

Gene	*N*	Total sites (bp)	Total SNP sites (S)	Total number of mutations	SNPs in exonic regions	SNPs in intronic regions	Silent[a] SNPs	Non-synonymous SNPs	Indels (location)
HMG2	3	490	26	26	10	16	23	3	2 (Intron)
SQE	3	884	22	22	3	19	21	1	1 (Intron)
SGT1[b]	5	815	49	51	51	0	14	37	2 (Exon)
SGT2[b]	2	558	9	8	8	0	3	5	1 (Exon)
Average	3.3	686.8	26.5	27	18.0	8.8	15.3	11.5	1.5
Total	13	2,747	106	108	72	35	61	46	6

N number of allelic sequences

[a] Synonymous and noncoding SNPs

[b] SGT1 and SGT2 both consist of a single exon with no introns

Table 2 Mean, range, standard deviation (SD), quartiles, and covariance (CV) of steroidal glycoalkaloid estimations in a diploid F_2 potato population *phu* DH × *chc* 80-1

	Total SGA (μg/mg DW)	α-Solanine (μg/mg DW)	α-Chaconine (μg/mg DW)	Leptine I (μg/mg DW)	Leptine II (μg/mg DW)	Sol/chac ratio	Leptine II/I ratio
Parents							
phu DH	0	0	0	0	0		
chc 80-1	190.7	10.5	8.8	113.0	58.4	1.2	0.5
F_2 Progeny						$n = 138$	$n = 32$
Mean	25.6	7.5	14.4	2.8	0.9	0.6	0.3
SD	25.8	8.0	14.0	8.2	2.7	0.3	0.1
Range	0–126.9	0–46.4	0–84.9	0–44.7	0–18.5	0.2–3.8	0.1–0.5
CV	101	106	97	291	314	59	30

The amount of accumulation of SGAs in μg of compound per mg of dried weight leaf tissue (DW). F_2 progeny sample size = 148 for all calculations. Ratios were analyzed in subsamples (*n*) with presence of those compounds

80-1 male parent produced all four SGAs, with greater levels of leptines I and II. The accumulation of SGAs among the 148 F_2 individuals varied broadly. Total SGAs ranged from undetectable to 127 μg mg^{-1} DW, α-solanine from undetectable to 46 μg mg^{-1} DW, α-chaconine from undetectable to 85 μg mg^{-1} DW, leptine I from undetectable to 45 μg mg^{-1} DW, and leptine II from undetectable to 19 μg mg^{-1} DW. Regression analyses showed a positive correlation between α-solanine and α-chaconine accumulation ($n = 140$, $r^2 = 0.84$, $P < 0.001$*) as well as for leptine I and leptine II ($n = 38$, $r^2 = 0.86$, $P < 0.001$*). However, there was no correlation between the accumulation of leptines with α-solanine or α-chaconine. We found that 35 and 85 F_2 plants exceeded the α-solanine and α-chaconine accumulation, respectively, of the parental line *chc* 80-1. However, there was not a single plant with more leptine I or leptine II than *chc* 80-1.

SGA quantitative data were converted to a binary format (1/0 = presence/absence) to study the genetics of SGA biosynthesis. Contingency table analyses of presence/absence between either α-solanine, α-chaconine, leptine I or leptine II showed different patterns of association in the synthesis of these compounds. Strong associations between synthesis of α-solanine and α-chaconine ($n = 148$, Likelihood ratio $\chi_1^2 = 52.2$, $P = 0.0001$*) and leptine I and leptine II ($n = 148$, Likelihood ratio $\chi_1^2 = 121.4$, $P = 0.0001$*) were found.

F_2 population genotyping analysis

The above described sequence polymorphisms were used to screen the allelic structure in a segregating F_2 population with 148 individuals. The observed F_2 genotypic segregation ratios in the progeny did not significantly differ from the 1:2:1 expected based on Chi-square analysis for all genes (Table S4). The data of five families were pooled based on a test of homogeneity (Strickberger 1968). ANOVA using the allelic composition at each candidate gene as source of variation and concentrations of total SGAs, α-solanine, α-chaconine, leptine I and leptine II as response variables, showed significantly greater levels of accumulation of total SGAs in plants homozygous

Table 3 Mean SGAs for individuals grouped by genotype at each of four genes segregating for parental alleles (P = phu, C = chc) in an F_2 population derived from a cross between *phu* DH × *chc* 80-1

Gene	Total SGA ± SE (μg/mg DW)	α-Solanine ± SE (μg/mg DW)	α-Chaconine ± SE (μg/mg DW)	Leptine I ± SE (μg/mg DW)	Leptine II ± SE (μg/mg DW)
HMG2					
CC	28.5 ± 4.6	7.5 ± 1.2	14.6 ± 2.6	4.9 a ± 1.6	1.6 a ± 0.6
CP	26.1 ± 3.3	7.9 ± 1.1	14.0 ± 1.6	3.3 a ± 1.1	1.0 ab ± 0.3
PP	21.9 ± 3.1	7.0 ± 6.7	14.9 ± 2.0	0.0 b ± 0.0	0.0 b ± 0.0
P value	0.526	0.875	0.952	0.026*	0.031*
SQE					
CC	29.9 ± 4.4	8.8 ± 1.4	15.6 ± 2.0	4.2 ± 1.6	1.21 ± 0.5
CP	23.0 ± 2.8	6.9 ± 0.8	13.6 ± 1.7	1.9 ± 0.7	0.6 ± 0.2
PP	25.8 ± 4.5	7.3 ± 1.5	14.5 ± 2.6	3.0 ± 1.6	1.0 ± 0.6
P value	0.369	0.438	0.740	0.318	0.526
SGT1					
CC	27.5 ± 4.2	7.3 ± 1.2	15.5 ± 2.5	3.5 ± 1.3	1.1 ± 0.5
CP	22.8 ± 2.6	6.8 ± 0.8	12.8 ± 1.4	2.5 ± 0.9	0.7 ± 0.2
PP	29.3 ± 5.1	9.7 ± 1.8	16.3 ± 2.6	2.5 ± 1.6	0.9 ± 0.5
P value	0.439	0.273	0.408	0.772	0.750
SGT2					
CC	35.1 a ± 6.3	10.4 ± 1.8	18.5 ± 3.5	4.6 ± 2.0	1.6 ± 0.8
CP	24.3 b ± 2.6	7.1 ± 0.8	13.4 ± 1.4	3.0 ± 0.9	0.9 ± 0.3
PP	19.7 b ± 2.5	6.1 ± 0.9	13.1 ± 1.6	0.4 ± 0.4	0.2 ± 0.2
P value	0.049*	0.070	0.184	0.116	0.120

The amount of accumulation of SGAs in μg of compound per mg of dried weight leaf tissue (DW) followed by standard error of mean (±SE)

The *chc* 80-1 and *phu* DH alleles are defined as C or P in the diploid genetic structure for *HMG2*, *SQE*, *SGT1* and *SGT2* genes. Each F_2 plant was characterized as to whether it had C, P or both alleles at each of four loci based on genetic polymorphism between C and P alleles. The numbers of individuals per mean are in Table 6

* Means followed by the same latter are not significantly different at the 0.05 α level using student *t* mean separation analysis

for *chc* 80-1 alleles for *SGT2* ($P < 0.05$*), and for leptine I and leptine II when *HMG2 chc* 80-1 alleles were present ($P < 0.03$* for both) (Table 3). Analysis by two-way ANOVAs revealed that the greatest levels of total SGAs, α-solanine and α-chaconine were associated with a significant interaction between *HMG2* and *SGT2* when homozygous chc 80-1 alleles were at both loci (Table 4). There was no interaction or significance of *SGT1* and *SQE* alleles with allelic state at our other candidate genes for any of the response variables.

The association of allelic states of each gene with synthesis of α-solanine, α-chaconine, leptine I and leptine II was studied in data converted to presence/absence format (1/0). For *HMG2,* α-solanine and α-chaconine were produced in many different allelic combinations, whereas homozygosity for *chc* 80-1 alleles was not found among the ten plants that had no α-solanine and the eight plants that had no α-chaconine ($n = 147$, Likelihood ratio $\chi_2^2 = 10.0$, $P = 0.01$*, and $n = 147$, Likelihood ratio $\chi_2^2 = 7.3$, $P = 0.03$*). There was strong association between presence of *HMG2 phu* DH homozygous alleles and lack of leptines, whereas *chc* 80-1 *HMG2* alleles were associated with synthesis (Leptine I, $n = 147$, Likelihood ratio $\chi_2^2 = 30.2$, $P < 0.0001$*; and Leptine II, $n = 147$, Likelihood ratio $\chi_2^2 = 24.2$, $P < 0.0001$*). A similar trend was found for *SGT2* and leptines. Contingency table analyses revealed an association of *phu* DH homozygous alleles and the absence of leptines; only 3.3 % (1 of 30) of plants with this genotype synthesized leptines I and II ($n = 148$, Likelihood ratio $\chi_2^2 = 14.1$, $P = 0.001$*, and $n = 148$, Likelihood ratio $\chi_2^2 = 10.6$, $P = 0.005$*). In contrast, *chc* 80-1 alleles were associated with the presence of leptines. For α-solanine and α-chaconine, no significant differences were detected with any *SGT2* genotypes. For *SQE* and *SGT1,* there was no association with the synthesis of any of the SGAs.

The F_2 population was made up of five families from specific crosses between sibling F_1 plants with various levels of SGAs. Significant differences were found among families for accumulation of specific and total SGAs (Table 5). Family 7 showed significantly greater levels of leptines than the other families. All five families exhibited the expected F_2 genetic segregation ratio of 1:2:1

Table 4 Effect of *HMG2* and *SGT2* genetic structure interaction on the accumulation of total SGAs, α-solanine and α-chaconine in a segregating F_2 population

HMG2, SGT2	Total SGA ± SE (μg/mg DW)	*HMG2, SGT2*	α-Solanine ± SE (μg/mg DW)	*HMG2, SGT2*	α-Chaconine ± SE (μg/mg DW)
CC, CC	63.1 a ± 9.3	CC, CC	16.8 a ± 2.9	CC, CC	32.0 a ± 5.1
CP, CP	29.2 b ± 4.1	PP, PP	9.2 b ± 2.4	PP, PP	18.3 b ± 4.3
CP, CC	28.6 bc ± 5.5	CP, CC	9.1 b ± 1.7	CP, CP	15.1 b ± 2.2
PP, PP	27.6 bc ± 7.8	CP, CP	8.6 b ± 1.3	CP, CC	14.8 b ± 3.0
CC, CP	20.9 bc ± 5.0	PP, CP	6.4 b ± 1.5	PP, CP	13.8 b ± 2.7
PP, CP	20.2 bc ± 4.9	CC, PP	5.9 b ± 2.9	PP, CC	13.2 b ± 6.8
CC, PP	20.1 bc ± 9.3	PP, CC	5.5 b ± 3.9	CC, PP	12.0 b ± 5.1
PP, CC	18.6 bc ± 12.3	CC, CP	5.2 b ± 1.6	CC, CP	10.2 b ± 2.8
CP, PP	13.5 c ± 6.8	CP, PP	3.7 b ± 2.1	CP, PP	9.8 b ± 3.8
P value	0.005*	*P* value	0.015*	*P* value	0.015*

The amount of accumulation of SGAs in μg of compound per mg of dried weight leaf tissue (DW) followed by standard error of mean (±SE)

The *chc* 80-1 and *phu* DH alleles are defined as C or P in the diploid genetic structure for *HMG2* and *SGT2* genes in the F_2 population

* Means followed by same latter are not significantly different at the 0.05 α level using student *t* mean separation analysis

Table 5 Effect of family pedigree on the accumulation of total SGAs, α-solanine, α-chaconine, leptine I and leptine II in a segregating F_2 population

Family	Total SGA ± SE (μg/mg DW)	α-Solanine ± SE (μg/mg DW)	α-Chaconine ± SE (μg/mg DW)	Leptine I ± SE (μg/mg DW)	Leptine II ± SE (μg/mg DW)
Family 3 (6 × 5 = L × H)	19.0 b ± 2.9	5.9 b ± 0.8	10.5 bc ± 1.3	1.7 b ± 1.3	0.8 b ± 0.6
Family 5 (2 × 1 = L × L)	14.2 b ± 2.6	5.2 b ± 1.0	8.5 c ± 1.4	0.4 b ± 0.3	0.1 b ± 0.1
Family 7 (3 × 5 = I × H)	34.1 a ± 6.6	7.0 ab ± 1.3	14.1 abc ± 2.4	10.0 a ± 3.3	2.9 a ± 1.0
Family 8 (5 × 2 = H × L)	32.4 a ± 5.2	10.4 a ± 1.8	20.5 a ± 3.2	1.2 b ± 0.6	0.3 b ± 0.1
Family 9 (2 × 5 = L × H)	26.8 ab ± 4.4	7.9 ab ± 1.4	15.5 ab ± 2.4	2.5 b ± 0.9	0.7 b ± 0.2
P value	0.015*	0.072	0.005*	<0.0001*	0.002*

The amount of accumulation of SGAs in μg of compound per mg of dried weight leaf tissue (DW) followed by standard error of mean (±SE)

The F_2 families come from specific crosses between F_1 plants with contrasting or similar levels of SGAs high (*H*), intermediate (*I*) or low (*L*). The ID number of F_1 plants used by cross and their level of SGAs in parenthesis

* Means followed by same latter are not significantly different at the 0.05 α level using student *t* mean separation analysis

(PP:PC:CC) per gene based on Chi-square test, except for *SQE* in family 5 (Supplementary Table 4). Some specific associations of SGA accumulation with allelic state were observed within families: plants with *SQE chc* 80-1 alleles in family 9 had significantly more α–solanine ($P = 0.02$*), those with *HMG2 chc* 80-1 alleles in family 8 and 9 had significant more leptine I as well as leptine II for family 9 ($P = 0.03$*, 0.02* and 0.02*), and those with *SGT1 phu* DH alleles in family 8 had greater levels of total SGAs and α–chaconine ($P = 0.04$* and 0.04*).

SNP chip analysis

A whole genome SNP genotype analysis of 43 F_2 plants was done to identify genomic regions and likely candidate genes related to synthesis and accumulation of SGAs. Association of significant SNPs with presence/absence of SGAs was studied by a contingency table analysis (Table S5). Significant SNPs were positioned in the potato genome browser (http://solanaceae.plantbiology.msu.edu/cgi-bin/gbrowse/potato/) and putative candidate genes located in neighboring regions were selected.

For synthesis of α-solanine and α-chaconine, we found 34 highly significant ($P < 0.02$*) SNPs. Most of the SNPs ($n = 33$) occurred on five superscaffolds clustered at the bottom of pseudochromosome 6, with a single significant SNP located at bottom of pseudochromosome 1. Eleven of the SNPs on pseudochromosome 6 were clustered in four superscaffolds and had the lowest *P* value ($P = 0.005$* for all 11) relative to the others (Fig. 2). There was no recombination in our population for these 11 SNPs, so that they segregated as two haplotypes: *chc* 80-1 and *phu*

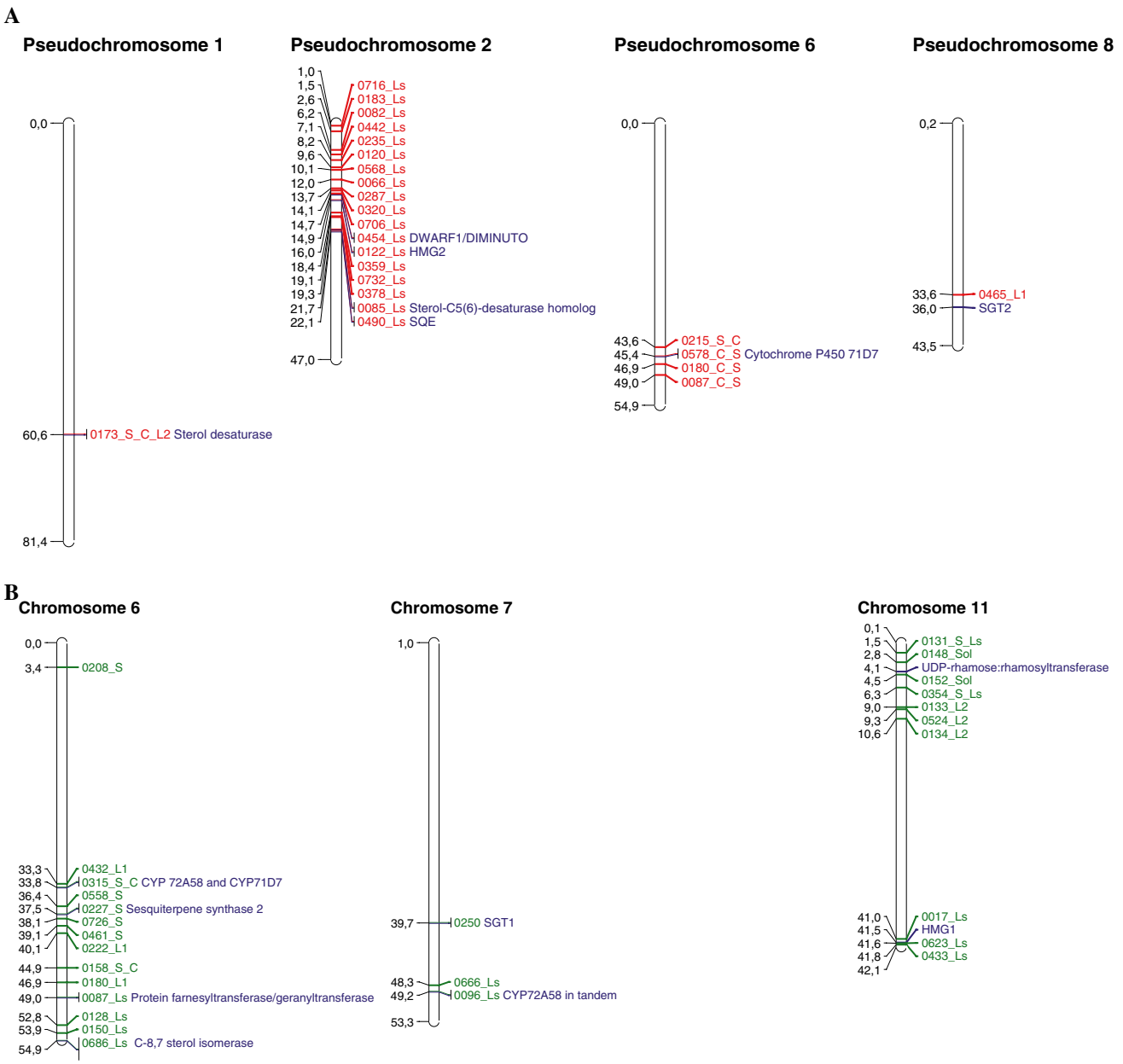

Fig. 2 Physical position of significant superscaffolds identified on pseudochromosomes of *Solanum tuberosum* group Phureja DM 1-3 516 R44. On the *left* of each pseudochromosome is the physical distances in Mb, and on the *right* superscaffold number (the last four numbers of each superscaffold PGSC0003DMB000000173 = 0173) and the compound associated (α-solanine *S*, α-chaconine *C*, leptine I *L1*, leptine II *L2* or Leptines *Ls*). Significant superscaffolds associated with synthesis (**a**) and accumulation (**b**). The location and names of putative candidate genes in the SGA biosynthetic pathway are shown in *blue* as is the location of three of our four candidate genes (*HMG2*, *SGT2*, and *SGT1*). An ortholog of our *SQE* candidate, *SQE* on pseudochromosome 2, is shown because of its location among a group of significant SNPs; however, the actual *SQE* that we amplified is not shown because there were no significant SNPs on pseudochromosome 4

DH. Effectively, the haplotype segregation yielded a 10 PP:22 PC:11 CC ratio ($\chi^2 = 0.97$ ns, for the expected 1:2:1 ratio). The four F_2 plants with *phu* DH alleles for the SNPs on both pseudochromosomes 1 and 6, either did not synthesize SGAs ($n = 3$) or express a low level of α-solanine and α-chaconine ($n = 1$). Since the locus on pseudochromosome 1 was homozygous in *phu* DH and heterozygous for *chc* 80-1 with one allele identical to *phu* DH, and the haplotypes for the 11 SNPs on pseudochromosome 6 were homozygous but different in both parents, the expected dihybrid segregation ratio would be 2:2:4:4:2:2 in the F_2 population for two independent loci, and the expected ratio did not differ from this ($\chi^2 = 0.56$ ns). Putative SGA-related candidate genes found in these regions

were sterol desaturase (PGSC0003DMG400001676) on pseudochromosome 1, and cytochrome P450 71D7 (PGSC0003DMG403020453) on pseudochromosome 6.

Leptine I and leptine II shared 58 of the same SNPs for synthesis of SGAs of the 59 and 60 that were found highly significant for leptine I and II, respectively (Table S5). In the contingency table analyses for presence/absence of leptines, 18 superscaffolds with 55 highly significant SNPs were situated between 1.5 and 22.1 Mb of pseudochromosome 2 (Fig. 2). There were four haplotypes in our F_2 population for this linkage block: ten individuals were homozygous for the *phu* DH haplotype, 13 homozygous for the *chc* haplotype, 18 heterozygous, and there were two individuals with recombinant haplotypes. None of the ten F_2 with only the *phu* DH haplotype synthesized leptines. One SNP (solcap_c1_15974) occurred within our *HMG2* candidate gene. Two significant SNPs, on pseudochromosomes 12 and 8 were also negatively associated with synthesis of leptines when F_2 individuals carried the *phu* DH allele. The SNP on pseudochromosome 8 was 1.6 Mb from our *SGT2* candidate gene, and demonstrated a similar segregation pattern with only two exceptions. The *chc* 80-1 alleles for one SNP on pseudochromosome 1 and two on pseudochromosome 7 were positively associated with synthesis of leptines. The SNP on pseudochromosome 1 occurred on the same superscaffold identified for α-solanine and α-chaconine and harbored a putative candidate gene, sterol desaturase.

Possible leptine biosynthetic and regulatory candidate genes occurring in regions nearby significant SNPs on pseudochromosomes 7, 2 and 8 included: cytochrome P450 with steroid hydroxylase activity CYP72A58 (PGSC0003DMG402012386) on pseudochromosome 7; *DWARF1/DIMINUTE* (*DWF1*) (PGSC0003DMG400011801), *HMG2* (PGSC0003DMG400003461), BRASSINAZOLE-RESISTANT 1 protein (*BZR1*) (PGSC0003DMG 400004501), sterol-C5(6)-desaturase homolog (PGSC0003 DMG400010415), *SQE* (PGSC0003DMG400003324) and N-acetyltransferase (PGSC0003DMG400013015) on pseudochromosome 2; *BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1* (*BAK1*) (PGSC0003 DMG400030075) and *SGT2* (PGSC0003DMG400017508) on pseudochromosome 8.

If we base our ANOVAs for allelic composition at each of 3,196 polymorphic SNPs on accumulation (amount of α-solanine, α-chaconine, leptine I and leptine II, as determined by HPLC) instead of synthesis, we identified 103 unique significant SNPs located on seven pseudochromosomes for the four different SGAs (Tables S6 and S7). For α-solanine and α-chaconine, 33 and four significant SNPs were found, respectively. Since three of the four significant SNPs for α-chaconine were also significant for α-solanine, we will consider them together. Of these 34 significant SNPs, 15 (45 %) and 11 (32 %) were located on pseudochromosomes 6 and 11 (Table S6); the other eight were scattered, one each on pseudochromosomes 1 and 7 and three each on pseudochromosomes 8 and 10. Clusters of significant superscaffolds surrounding a candidate gene would be expected due to linkage (Fig. 2). Within each cluster, we might expect the *P* value of the ANOVA to decrease as the SNPs approach a candidate gene. The 15 significant SNPs on pseudochromosome 6 occurred on seven superscaffolds spanning 10.7 Mb at the bottom arm and one superscaffold located at the top. Although all of the *P* values indicated a high level of significance, a nadir *P* value spanning three superscaffolds (PGSC0003DMB000000315, PGSC0003DMB000000227 and PGSC0003DMB 000000158) suggests the following candidate genes: CYP 72A58 (PGSC0003DMG400002351), CYP71D7 (PGSC 0003DMG400034479), sesquiterpene synthase 2 (PGSC 0003DMG400013034) and geraniol 10-hydroxylase (PGSC 0003DMG400020452). Likewise, the 11 significant SNPs on pseudochromosome 11 covered four superscaffolds, with the lowest *P* values occurring on superscaffold PGSC0003DMB000000148, implicating the candidate gene UDP-rhamnose:rhamnosyltransferase (*SGT3*) (PGSC 0003DMG400016194). The superscaffolds bearing significant SNPs on pseudochromosomes 1, 7, 8 and 10 did not have obvious candidate genes.

Leptine I and leptine II shared 39 significant SNPs of the 48 and 62 that were highly significant for accumulation of each compound (Tables S6 and S7). Since there is likely interconversion between the two leptines, we will consider only the 39 common significant SNPs. As with α-solanine and α-chaconine, pseudochromosomes 6 and 11 were implicated, accounting for ten (26 %) and 23 (59 %) of the significant SNPs. The other six were scattered, with one, two and three on pseudochromosomes 1, 7 and 8, respectively. Again, following the trend of selecting candidate genes from superscaffolds harboring SNPs with the lowest *P* values, we suggest three (PGSC0003DMB000000087, PGSC0003DMB000000150 and PGSC0003DMB000000686) on pseudochromosome 6 and three (PGSC0003DMB000000131, PGSC0003 DMB000000017 and PGSC0003DMB000000433) on pseudochromosome 11 as the most likely (Fig. 2). On pseudochromosome 6, the candidate genes included protein farnesyltransferase/geranyltransferase (PGSC0003 DMG400027009), 0-methyltransferase (PGSC0003DMG 400020095) and C-8, 7 sterol isomerase (PGSC0003 DMG400027684) and on pseudochromosome 11 UDP-glucose:glucosyltransferase (PGSC0003DMG400028670) and 3-hydroxy-3-metylglutaryl coenzyme A reductase 1 (PGSC0003DMG400046343). Pseudochromosome 1 and 7 had one and two highly significant SNPs. The same superscaffold that was positively associated with synthesis of α-solanine, α-chaconine and leptines on

pseudochromosome 1, as well as the same SNPs associated with synthesis of leptines on pseuchromosome 7 in the previous contingency table analysis were significant for accumulation of leptine. For both loci, homozygosity for the *chc* 80-1 homozygous alleles was associated with high accumulation and synthesis of leptines. SNPs on pseudochromosome 8 did not indicate obvious candidate genes.

Three of our four original candidate genes (*HMG2*, *SGT1* and *SGT2*) had SNPs that were included in the SolCAP 8303 Infinium array. We used available Solanaceae sequences to design and amplify the fourth candidate gene, *SQE*, since the potato genome sequence was not available at the time. Using the current potato genome browser, we can locate the *SQE* sequence amplified by our primers on pseudochromosome 4 (PGSC0003DMG400004923). We had nominated all of our candidate genes for inclusion in the SNP chip design; however, a homolog of *SQE* on pseudochromosome 2 was inadvertently selected. Our candidate *SQE* found on pseudochromosome 4 was 73 % similar at the nucleotide level with the *SQE* on pseudochromosome 2. Four (solcap_snp_c1_9695, 9696, c2_32413 and 32417) of six SNPs within this second *SQE* gene (PGSC0003DMG400003324) were significantly associated with leptine I and II, with heterozygotes intermediate in accumulation between the two homozygous states (Table 6). The SNPs occurred in exons 1, 5 and 6 of the *SQE* gene model; those in exons 5 and 6 were nonsynonymous. Homozygous states for the *chc* 80-1 allele for all SNPs were associated with greater leptine I and II accumulation. Of the five SNPs included on the SNP chip within the coding region of *HMG2* on pseudochromosome 2 (PGSC0003DMG400003461), three (solcap_snp_c1_15972, 15973 and 15974) were polymorphic in our population and their allelic states were significantly associated with leptine I and II accumulation. All three occurred within the first exon of *HMG2*, one (15973) was nonsynonymous, and another (15974) discriminated homozygotes with the *chc* allele expressing greater leptine accumulation. All five SolCAP SNPs within the coding region of *SGT2* on pseudochromosome 8 (PGSC0003DMG400017508) were not polymorphic in our population. For *SGT1* on pseudochromosome 7 (PGSC0003DMG400011749), three of 12 SNPs included on the Illumina chip were polymorphic in our population but none of the three was significantly associated with any SGA accumulation. Among the 22 SNPs in our candidate genes that were included on the potato Illumina array, only two within *SGT1* (solcap_snp_c2_23384 and _23385) were within our sequenced fragments.

Discussion

Allelic sequences and sequence polymorphism

Sequencing and haplotype analysis of fragments of coding regions were used to identify thirteen alleles of four candidate genes in two potato species. Fragments of genes in primary metabolism (*HMG2* and *SQE*) had more silent SNPs, whereas genes in SGA-specific secondary metabolism (*SGT1* and *SGT2*) had more nonsynonymous SNPs. Genes of primary metabolism are expected to have more selection pressure against mutations that may affect protein function than genes of secondary metabolism (Waxman and Peck 1998). Enzymes of the early steps of primary metabolism have greater constraints, being essential to a number of metabolic pathways during the plant life cycle, whereas secondary metabolism is characterized by the evolvement of related enzymes often with positive selection for diverse functions to produce a wide range of compounds for plant adaptation (Pichersky and Gang 2000). Gene duplication and subsequent divergence are part of the evaluative mechanisms to provide novel gene function (Todd et al. 1999). Variability in SGA compounds is defined by structural modifications of either the glycosidic residues or the aglycone. Variability in specialized genes and allelic interactions generate distinctive genetic combinations that relate to the synthesis of new SGAs, as suggested by interspecific crosses of potato with wild species (Sagredo et al. 2011;

Table 6 SNPs from candidate genes in the 8,303 Infinium array associated with leptine accumulation

	CC	CP	PP	*P* value
SQE				
Leptine I (μg/mg DW) ± SE	12.9 a ± 4.3	8.0 ab ± 3.2	0.0 b ± 0.0	0.030*
Leptine II (μg/mg DW) ± SE	4.4 a ± 1.8	2.6 ab ± 1.0	0.0 b ± 0.0	0.048*
HMG2				
Leptine I (μg/mg DW) ± SE	14.0 a ± 4.5	8.1 ab ± 3.2	0.0 b ± 0.0	0.016*
Leptine II (μg/mg DW) ± SE	4.8 a ± 1.9	2.6 ab ± 1.0	0.0 b ± 0.0	0.027*

The amount of accumulation of SGAs in μg of compound per mg of dried weight leaf tissue (DW) followed by standard error of mean (±SE)

ANOVA analyses for solcap_snp_c1_9695, 9696, c2_32413 and 32417 SNPs from *SQE* and solcap_snp_c1_15974 from *HMG2*

* Means followed by same latter are not significantly different at the 0.05 α level using student *t* mean separation analysis

Väänänen et al. 2005). For *SGT1,* we found three allelic sequences in the *chc* 80-1 monoploids; two had different promoter regions, whereas all three had the same 3′UTR. This, in addition to the observed co-segregation of the three alleles, suggested gene duplication in the high producing SGA clone, *chc* 80-1.

F_2 population phenotypic analysis

The accumulation of α-solanine and α-chaconine followed different distribution patterns in the segregating F_2 population than those for leptine I and leptine II. Although variable, most of the population accumulated α-solanine and α-chaconine (±95 %), whereas only ±25 % accumulated leptines I and II. There were positive correlations between the synthesis and accumulation of α-solanine and α-chaconine; and also between the accumulation of leptine I and leptine II. However, there was no correlation between the accumulation of either α-solanine or α-chaconine with leptines. This suggests that different genetic mechanisms could regulate synthesis and accumulation of these two groups of SGAs. Specialized genes controlling synthesis of SGAs based on aglycone structure were also suggested by Sanford et al. (1996). The synthesis of SGAs followed Mendelian segregation ratios, e.g., for the synthesis of α-solanine and α-chaconine, the F_2 population exhibited epistatic inheritance of two independent dominant genes with interaction of two or more genes for accumulation. On the other hand, synthesis of leptine I and leptine II followed a single gene Mendelian model in which a recessive gene was responsible for the synthesis of leptines. The variability in the level of these compounds in the progenies that accumulate all four of them, suggests the interaction of several genes. Similar conclusions of polygenetic control with a few dominant genes affecting biosynthesis of SGAs and recessive genes controlling leptines levels have been reached before (Hutvágner et al. 2001; Ronning et al. 1999; Sagredo et al. 2006; Sanford et al. 1996; Van Dam et al. 1999; Yencho et al. 1998).

F_2 population genotyping analysis

Genetic interaction between allelic states at two of our candidate genes was statistically significant. F_2 plants homozygous for *chc* 80-1 alleles for *HMG2* and *SGT2* exhibited the greatest levels of accumulation of total SGAs, α-solanine and α-chaconine. For leptines, the contingency analysis showed statistically significant association of *phu* alleles at both *SGT2* and *HMG2* and absence of leptines. Segregating populations have been previously used to elucidate genetic factors associated with solanine/chaconine and leptine accumulation. Our findings of *HMG2* on pseudochromosome 2 and *SGT2* on pseudochromosome 8 are in agreement with Sagredo et al. (2006) who suggested loci on the same two chromosomes with complementary epistasis associated with leptinine and leptine content, whereas the locus on chromosome 8 was directly responsible for leptine biosynthesis. *HMG2* could play a regulatory role since its expression has been induced by wounding and pathogen inoculation (Choi et al. 1992). Pseudochromosomes 1 and 8 could have genes that control synthesis of α-solanine-α-chaconine and leptines, respectively. Consistently, major loci on chromosome 1 have been associated with production of solanidine aglycone, total SGAs, α-solanine, α-chaconine and repression of leptines (Hutvágner et al. 2001; Ronning et al. 1999; Sørensen et al. 2008; Yencho et al. 1998).

SNP chip analysis

The SNP chip contingency table analyses identified loci that accurately explained synthesis of different SGAs. We found that the phenotypic segregation of the F_2 population could be explained by the epistatic interaction of two independent genes. Likewise, the SNP chip analysis detected two loci strongly associated with synthesis of α-solanine and α-chaconine. One significant SNP at 61 Mb on pseudochromosome 1 (between SSR markers STM5136 at 53.3 Mb and STM2030 at 65.7 Mb) corresponds to a significant QTL mapped previously by Sørensen et al. (2008), and suggested to be a biosynthetic gene acting early in the biosynthetic pathway of the solanidine aglycone. A candidate sterol desaturase mapped to this location could play a role in the post-squalene segment of the pathway leading to the cholesterol formation (Benveniste 2004). Accordingly, our candidate gene, cytochrome P450 CYP71D7 on pseudochromosome 6, may have similar catalytic properties as *GAME7* and *GAME8* in tomato, which were both proposed to add hydroxyl groups as steps in the transformation of cholesterol to tomatidenol in the synthesis of α-tomatine (Heftmann 1983; Itkin et al. 2013; Kaneko et al. 1976, 1977).

A few recessive genes have been previously associated with the synthesis of leptines (Boluarte-Medina et al. 2002; Hutvágner et al. 2001; Ronning et al. 1999; Sagredo et al. 2006). In our study, four loci were identified: two on pseudochromosome 1 and 7 that were positively associated with leptine I and leptine II synthesis, and two on pseudochromosome 2 and 8 that were negatively associated with their synthesis. However, these loci could not explain the 1:3 phenotypic ratio found for leptine synthesis in this population. The leptine significant SNP at 60.6 Mb of pseudochromosome 1 was nearby the abovementioned sterol desaturase at 60.1 Mb. The complementary leptine related candidate gene on pseudochromosome 7 is a family member of CYP72A58, a cytochrome P450 with steroid hydroxylase activity that could catalyze reactions in the conversion of cholesterol to

leptines. Recently, several cytochrome P450, three of them members of the CYP72A family, were mapped to chromosome 7 and were suggested to be involved in SGA biosynthesis (Itkin et al. 2013). The synthesis leading to SGAs, brassinosteroids and phytosterols from cycloartenol precursor is carried out in three parallel pathways C8 (cholesterol), C9 (campesterol) and C10 (stigmasterol) (Benveniste 2004; Schaller 2004); where enzymes catalyze the same reaction with intermediates related across the three pathways (Arnqvist 2007). The respective genes *DWF1*, *HMG2*, *BZR1*, sterol-C5(6)-desaturase homolog, *SQE* and *N*-acetyltransferase were presently selected on pseudochromosome 2 and *BAK1* and *SGT2* on pseudochromosome 8, both loci carried SNPs that were negatively associated with leptines. This is further supported by Nahar (2011) who showed that the level of cholesterol and SGAs was reduced in antisense *StDWF1* potato plants. *DWF1* and *HMG2* are tightly linked on pseudochromosome 2, and future sequence and expression analyses of *chc* 80-1 could elucidate the specific role of *DWF1* in leptine biosynthesis since some amino acid deletions were already identified in the cultivated potato by Nahar (2011). In addition, our data suggest clustering of genes that are related to a specific metabolism and that may consist of both primary and secondary metabolism functions (Osbourn 2010). The latter may also explain their coordination with respect to precursor availability (Ginzberg et al. 2012).

Regarding SGA accumulation, 103 unique significant SNPs were found. These mapped to seven pseudochromosomes on the potato genome. Pseudochromosomes 6 and 11 held the most significant SNPs for α-solanine and pseudochromosomes 1, 6, 7 and 11 for leptine I and leptine II. Some putative candidate genes affecting biosynthesis or regulating the expression of SGAs located nearby those loci. Association of SGA accumulation with multiple pseudochromosomes and loci within them agreed with the polygenic control that was proposed for this trait (Sagredo et al. 2006; Van Dam et al. 1999; Yencho et al. 1998). Overall, the present results demonstrate the genetic complexity of synthesis and accumulations of SGAs.

Acknowledgments This research was supported by Research Grant No. IS-4134-08 R from BARD, The United States—Israel Binational Agricultural Research and Development Fund. We thank Suzanne Piovano for technical assistance, Maichel Miguel Aguayo Bustos for programming help, Anne Ryan of the Laboratory for Interdisciplinary Statistical Analysis LISA for statistical consultation.

References

Arnqvist L (2007) Plant sterol metabolism with emphasis on glycoalkaloid biosynthesis in potato. Dept of plant biology and forest genetics. Swedish University of Agricultural Sciences, Uppsala, p 50

Benveniste P (2004) Biosynthesis and accumulation of sterols. Annu Rev Plant Biol 55:429–457

Boluarte-Medina T, Fogelman E, Chani E, Miller AR, Levin I, Levy D, Veilleux RE (2002) Identification of molecular markers associated with leptine in reciprocal backcross families of diploid potato. Theor Appl Genet 105:1010–1018

Choi D, Ward BL, Bostock RM (1992) Differential induction and suppression of potato 3-hydroxy-3-methylglutaryl coenzyme A reductase genes in response to *Phytophthora infestans* and to its elicitor arachidonic acid. Plant Cell 4:1333–1344

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

Distl M, Wink M (2009) Identification and quantification of steroidal alkaloids from wild tuber-bearing *Solanum* species by HPLC and LC-ESI-MS. Potato Res 52:79–104

Edwards EJ, Cobb AH (1996) Improved high performance liquid chromatographic method for the analysis of potato (*Solanum tuberosum*) glycoalkaloids. J Agr Food Chem 44:2705–2709

FAO (2008) Potato pest and disease management. International year of the Potato 2008. http://www.potato2008.org/en/potato/pests.html

Felcher KJ, Coombs JJ, Massa AN, Hansey CN, Hamilton JP, Veilleux RE, Buell CR, Douches DS (2012) Integration of two diploid potato linkage maps with the potato genome sequence. PLoS ONE 7:e36347

Friedman M (2006) Potato glycoalkaloids and metabolites: roles in the plant and in the diet. J Agr Food Chem 54:8655–8681

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

Friedman M, McDonald GM (1997) Potato glycoalkaloids: chemistry, analysis, safety, and plant physiology. Crit Rev Plant 16:55–132

Ginzberg I, Tokuhisa JG, Veilleux RE (2009) Potato steroidal glycoalkaloids: biosynthesis and genetic manipulation. Potato Res 52:1–15

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

Hamilton JP, Hansey CN, Whitty BR, Stoffel K, Massa AN, Van Deynze A, De Jong WS, Douches DS, Buell CR (2011) Single nucleotide polymorphism discovery in elite North American potato germplasm. BMC Genomics 12:302

Haynes FL (1972) The use of cultivated diploid *Solanum* species in potato breeding. In: Prospects for the potato in the developing world. International Potato Center (CIP), Lima, Peru

Heftmann E (1983) Biogenesis of steroids in *Solanaceae*. Phytochemistry 22:1843–1860

Hutvágner G, Bánfalvi Z, Milánkovics I, Silhavy D, Polgár Z, Horváth S, Wolters P, Nap JP (2001) Molecular markers associated with leptinine production are located on chromosome 1 in *Solanum chacoense*. Theor Appl Genet 102:1065–1071

Itkin M, Heinig U, Tzfadia O, Bhide AJ, Shinde B, Cardenas PD, Bocobza SE, Unger T, Malitsky S, Finkers R, Tikunov Y, Bovy A, Chikate Y, Singh P, Rogachev I, Beekwilder J, Giri AP, Aharoni A (2013) Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. Science 341:175–179

Johnson AAT, Piovano SM, Ravichandran V, Veilleux RE (2001) Selection of monoploids for protoplast fusion and generation of intermonoploid somatic hybrids of potato. Am J Potato Res 78:19–29

Judo MS, Wedel AB, Wilson C (1998) Stimulation and suppression of PCR-mediated recombination. Nucleic Acids Res 26:1819–1825

Kaneko K, Tanaka MW, Mitsuhashi H (1976) Origin of nitrogen in the biosynthesis of solanidine by *Veratrum grandiflorum*. Phytochemistry 15:1391–1393

Kaneko K, Tanaka MW, Mitsuhashi H (1977) Dormantinol, a possible precursor in solanidine biosynthesis, from budding *Veratrum grandiflorum*. Phytochemistry 16:1247–1251

Kolbe H, Stephan-Beckmann S (1997) Development, growth and chemical composition of the potato crop (*Solanum tuberosum* L.). II. Tuber and whole plant. Potato Res 40:135–153

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

Laurila J, Laakso I, Valkonen JPT, Hiltunen R, Pehu E (1996) Formation of parental-type and novel glycoalkaloids in somatic hybrids between *Solanum brevidens* and *S. tuberosum*. Plant Sci 118:145–155

Lenz TL, Becker S (2008) Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci–implications for evolutionary analysis. Gene 427:117–123

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

Lightbourn GJ, Veilleux RE (2007) Production and evaluation of somatic hybrids derived from monoploid potato. Am J Potato Res 84:425–435

Liu YG, Chen Y (2007) High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences. Biotechniques 43:649–656

Manrique-Carpintero NC, Tokuhisa JG, Ginzberg I, Holliday JA, Veilleux RE (2013) Sequence diversity in coding regions of candidate genes in the glycoalkaloid biosynthetic pathway of wild potato species. G3:g3.113.007146 Early Online

McCue KF, Shepherd LVT, Allen PV, Maccree MM, Rockhold DR, Corsini DL, Davies HV, Belknap WR (2005) Metabolic compensation of steroidal glycoalkaloid biosynthesis in transgenic potato tubers: using reverse genetics to confirm the in vivo enzyme function of a steroidal alkaloid galactosyltransferase. Plant Sci 168:267–273

McCue KF, Allen PV, Shepherd LV, Blake A, Maccree MM, Rockhold DR, Novy RG, Stewart D, Davies HV, Belknap WR (2007a) Potato glycosterol rhamnosyltransferase, the terminal step in triose side-chain biosynthesis. Phytochemistry 68:327–334

McCue KF, Allen PV, Shepherd LVT, Blake A, Rockhold DR, Novy RG, Stewart D, Davies HV, Belknap WR (2007b) Manipulation and compensation of steroidal glycoalkaloid biosynthesis in potatoes. Acta Hort 745:343–349

Moehs CP, Allen PV, Friedman M, Belknap WR (1997) Cloning and expression of solanidine UDP-glucose glucosyltransferase from potato. Plant J 11:227–236

Morris SC, Lee TH (1984) The toxicity and teratogenicity of Solanaceae glycoalkaloids, particularly those of the potato (*Solanum tuberosum*): a review. Food Technol Aust 36:118–124

Mweetwa MA (2009) Biosynthesis of steroidal glycoalkaloids in *Solanum chacoense* Bitter. Virginia Polytechnic Institute and State University, Blacksburg, p 118

Mweetwa AM, Hunter D, Poe R, Harich KC, Ginzberg I, Veilleux RE, Tokuhisa JG (2012) Steroidal glycoalkaloids in *Solanum chacoense*. Phytochemistry 75:32–40

Nahar N (2011) Regulation of sterol and glycoalkaloid biosynthesis in potato (*Solanum tuberosum* L.)—Identification of key genes and enzymatic steps. Department of plant biology and forest genetics. Swedish University of Agricultural Sciences, Uppsala, p 66

Nema PK, Ramayya N, Duncan E, Niranjan K (2008) Potato glycoalkaloids: formation and strategies for mitigation. J Sci Food Agr 88:1869–1881

Osbourn A (2010) Gene clusters for secondary metabolic pathways: an emerging theme in plant biology. Plant Physiol 154:531–535

Pichersky E, Gang DR (2000) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. Trends Plant Sci 5:439–445

Ramkumar G, Sivaranjani AKP, Pandey MK, Sakthivel K, Rani NS, Sudarshan I, Prasad GSV, Neeraja CN, Sundaram RM, Viraktamath BC, Madhav MS (2010) Development of a PCR-based SNP marker system for effective selection of kernel length and kernel elongation in rice. Mol Breeding 26:735–740

Ronning CM, Stommel JR, Kowalski SP, Sanford LL, Kobayashi RS, Pineada O (1999) Identification of molecular markers associated with leptine production in a population of *Solanum chacoense* bitter. Theor Appl Genet 98:39–46

Sagredo B, Lafta A, Casper H, Lorenzen J (2006) Mapping of genes associated with leptine content of tetraploid potato. Theor Appl Genet 114:131–142

Sagredo B, Lorenzen J, Casper H, Lafta A (2011) Linkage analysis of a rare alkaloid present in a tetraploid potato with *Solanum chacoense* background. Theor Appl Genet 122:471–478

Sanford L, Kobayashi R, Deahl K, Sinden S (1996) Segregation of leptines and other glycoalkaloids in *Solanum tuberosum* (4x) × *Solanum chacoense* (4x) crosses. Am J Potato Res 73:21–33

Schaller H (2004) New aspects of sterol biosynthesis in growth and development of higher plants. Plant Physiol Biochem 42:465–476

Sinden SL, Sanford LL, Osman SF (1980) Glycoalkaloids and resistance to the Colorado potato beetle in *Solanum chacoense* Bitter. Am Potato J 57:331–343

Sinden SL, Sanford LL, Cantelo WW, Deahl KL (1986) Leptine glycoalkaloids and resistance to the Colorado potato beetle (Coleoptera: Chrysomelidae) in *Solanum chacoense*. Environ Entomol 15:1057–1062

Sørensen K, Kirk H, Olsson K, Labouriau R, Christiansen J (2008) A major QTL and an SSR marker associated with glycoalkaloid content in potato tubers from *Solanum tuberosum* × *S. sparsipilum* located on chromosome I. Theor Appl Genet 117:1–9

Strickberger MW (1968) Genetics. Macmillan, New York

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

The Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–195

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Thompson JR, Marcelino LA, Polz MF (2002) Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. Nucleic Acids Res 30:2083–2088

Todd AE, Orengo CA, Thornton JM (1999) Evolution of protein function, from a structural perspective. Curr Opin Chem Biol 3:548–556

Väänänen T, Ikonen T, Rokka VM, Kuronen P, Serimaa R, Ollilainen V (2005) Influence of incorporated wild *Solanum* genomes on potato properties in terms of starch nanostructure and glycoalkaloid content. J Agr Food Chem 53:5313–5325

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

Van Dam J, Levin I, Struik PC, Levy D (1999) Genetic characterisation of tetraploid potato (*Solanum tuberosum* L.) emphasising genetic control of total glycoalkaloid content in the tubers. Euphytica 110:67–76

Van Dam J, Levin I, Struik PC, Levy D (2003) Identification of epistatic interaction affecting glycoalkaloid content in tubers of tetraploid potato (*Solanum tuberosum* L.). Euphytica 134: 353–360

Veilleux RE (1990) *Solanum phureja*: anther culture and the induction of haploids in a cultivated diploid potato species. Bajaj YPS edn. Springer-Verlag, New York

Waxman D, Peck JR (1998) Pleiotropy and the preservation of perfection. Science 279:1210–1213

Wentzinger LF, Bach TJ, Hartmann MA (2002) Inhibition of squalene synthase and squalene epoxidase in tobacco cells triggers an up-regulation of 3-hydroxy-3-methylglutaryl coenzyme a reductase. Plant Physiol 130:334–346

Yencho GC, Kowalski SP, Kobayashi RS, Sinden SL, Bonierbale MW, Deahl KL (1998) QTL mapping of foliar glycoalkaloid aglycones in *Solanum tuberosum* x *S. berthaultii* potato progenies: quantitative variation and plant secondary metabolism. Theor Appl Genet 97:563–574