Transcriptome and metabolome of synthetic Solanum autotetraploids reveal key genomic stress events following polyploidization

Carlo Fasano1*, Gianfranco Diretto2*, Riccardo Aversano1, Nunzio D’Agostino3, Antonio Di Matteo1, Luigi Frusciante1, Giovanni Giuliano2 and Domenico Carputo1

1Department of Agricultural Sciences, University of Naples Federico II, Portici 80055, Italy; 2Italian National Agency for New Technologies, Energy, and Sustainable Development, Casaccia Research Centre, Rome 00123, Italy; 3Consiglio per la ricerca in agricoltura e l’analisi dell’economia agraria – Centro di ricerca per l’orticoltura (CRA-ORT), via dei Cavalleggeri 25, Pontecagnano, Salerno 84098, Italy

Author for correspondence: Domenico Carputo
Tel: +39 081 2539225
Email: carputo@unina.it
Received: 13 July 2015
Accepted: 6 December 2015

New Phytologist (2016)
doi: 10.1111/nph.13878

Key words: autopolyploid, metabolome, nucleotide pool imbalance, potato, transcriptome, wild Solanum.

Summary

- Polyploids are generally classified as autopolyploids, derived from a single species, and allopolyploids, arising from interspecific hybridization. The former represent ideal materials with which to study the consequences of genome doubling and ascertain whether there are molecular and functional rules operating following polyploidization events.
- To investigate whether the effects of autopolyploidization are common to different species, or if species-specific or stochastic events are prevalent, we performed a comprehensive transcriptomic and metabolomic characterization of diploids and autotetraploids of Solanum commersonii and Solanum bulbocastanum.
- Autopolyploidization remodelled the transcriptome and the metabolome of both species. In S. commersonii, differentially expressed genes (DEGs) were highly enriched in pericentromeric regions. Most changes were stochastic, suggesting a strong genotypic response. However, a set of robustly regulated transcripts and metabolites was also detected, including purine bases and nucleosides, which are likely to underlie a common response to polyploidization.
- We hypothesize that autopolyploidization results in nucleotide pool imbalance, which in turn triggers a genomic shock responsible for the stochastic events observed. The more extensive genomic stress and the higher number of stochastic events observed in S. commersonii with respect to S. bulbocastanum could be the result of the higher nucleoside depletion observed in this species.

Introduction

Polyploidy, that is, the heritable condition of possessing more than two complete sets of chromosomes, is especially common among plants (Wood et al., 2009). Whole-genome polyploidization events predated the evolution of seed plants (Jaillon et al., 2007; Jiao et al., 2011), including Solanaceae (Tomato Genome Sequencing Consortium, 2012). Recently, plant genome sequencing data provided evidence that species once considered classical diploids, such as Arabidopsis thaliana, Oryza sativa, Populus trichocarpa and Vitis vinifera, contain signatures of ancient polyploidization in their genomes. Polyploids often show novel phenotypes that are not present in their diploid progenitors or exceed the range of contributing species (Ainouche et al., 2009; Fawcett et al., 2009). This provides them with a fitness advantage of great significance from the evolutionary standpoint.

Polyploids are generally classified into two main different types: autopolyploids, derived from an intraspecific increase in chromosome number, and allopolyploids, arising from interspecific hybridization concomitant with genome doubling. Polyploidization is a severe shock that may cause not only genome reshuffling but also changes in gene regulation, the epigenetic landscape and transposable element activity. A recent review by Tayale & Parisod (2013) reports several examples of genome changes, including both restructuring events and functional modifications. These may be very dramatic during the initial ‘revolutionary phase’ of polyploid formation, probably as a response to genomic shock (Levy & Feldman, 2004). The consequences of chromosome doubling have been widely studied in allopolyploids, whereas more limited data are available for autopolyploid plants (Parisod et al., 2010). This probably reflects the fact that autopolyploids used to be considered evolutionary dead-ends and that their identification in nature is more difficult (Soltis et al., 2010). As the plasticity of polyploids is to a large extent...
associated with changes at the transcriptional level (Jackson & Chen, 2010), one question of interest is how gene expression patterns shift in response to polyploidization. Advances in genomics and sequencing technology created unprecedented opportunities for discovering and monitoring the molecular effects of polyploidization (Aversano et al., 2012). For instance, large-scale microarray studies provided evidence that allopolyploids exhibit considerable transcriptome alterations as compared with their diploid progenitors. Such changes are mostly attributable to the reunion in a common nucleus of previously diverged regulatory hierarchies (from the parental species), which probably entails nonadditive gene expression (Wang et al., 2006; Yu et al., 2010; Bombarely et al., 2012; Higgins et al., 2012; Combes et al., 2013; Rambani et al., 2014; Jiang et al., 2015). Although nonadditive gene expression is a common feature of allopolyploidization, its magnitude largely depends on the specific allopolyploid. Data available in autopolyploids suggest that changes are less profound. Low numbers of genes were differentially regulated in polyploid Citrus × limonia, maize, and Arabidopsis thaliana (Pignatta et al., 2010; Riddle et al., 2010; Allario et al., 2011; Tan et al., 2015). In a Solanum phureja synthetic autopolyploid series (1 × 2 × 4 ×), c. 10% of analysed genes displayed variations in expression levels, but most changes were observed in the monoploid (Stupar et al., 2007). Pignatta et al. (2010) hypothesized that a large set of genes are potentially affected upon polyploidization, but only a small number become altered. Based on results in A. thaliana, Yu et al. (2010) suggested that autopolyploids may respond in a similar but more subtle way compared with allopolyploids.

Polyploidy impacts many plant traits, such as morphology, physiology and resistance to stress (Ramsey & Schemske, 2002). It can also greatly influence plant metabolism, in both a qualitative and quantitative manner, offering the opportunity for enhanced metabolic activity through altered gene expression, which leads to changes in the concentration of secondary metabolites. The large amount of possible allelic combinations may result in the production of hybrid enzymes and enzymatic diversity, providing additional layers of versatility and homeostasis. There are several reports on polyploidy induction to enhance the level of specific phytochemicals for commercial use (for a review, see Lavanía, 2005). For example, allotetraploids of Artemisia annua produced more terpenoids or triterpene-type compounds per gram of tissue than their diploid counterparts (Banyai et al., 2010). Nicotiana alloploids showed changes in the profile of secondary metabolites manufactured in defense against herbivory between individuals (Lou & Baldwin, 2003) and between individuals and their parents (Pearse et al., 2006). Evidence of the impact of genome doubling on metabolic profiles has also been reported in autopolyploids. Xing et al. (2011) reported that tetraploid Catharanthus roseus produced more vindoline, catharanthine and vinblastine than the diploid parent. Mishra et al. (2010) reported the use of polyploidy in Papaver somniferum as an effective tool for producing plants with high morphine content. Changes in the metabolic profile of the autopolyploids by a simple duplication of the basic genome were interpreted as being the result of an alteration in the mechanism(s) regulating the biosynthesis of individual compounds.

In spite of these findings and the possibilities offered now by liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS) analysis, in the literature, there are neither large-scale metabolomic studies nor integrative studies available aimed at linking transcriptomic and metabolomic data. The study of the metabolic and transcriptional remodelling of tuber-bearing Solanum species following autopolyploidization is of particular interest as modern potato (Solanum tuberosum), the third most important staple crop world-wide, originated from polyploidization events (Spooner et al., 2014). It would be of particular interest to ascertain whether the genomic and metabolic perturbations following autopolyploidization are common to different species, or if species-specific or even lineage-specific (stochastic) events are prevalent. In order to discriminate among the three event types, we produced a number of independent synthetic autotetraploids from two clones of two diploid wild potatoes (Solanum commersonii and Solanum bulbocastanum). These represent ideal materials in that they not only reveal the molecular machinery at work after genome doubling, but also allow the resolution of stochastic events. Solanum commersonii and S. bulbocastanum are widely used for potato improvement, being an important source of genes and allelic diversity. In addition, S. commersonii represents the first wild potato species whose genome has been sequenced (Aversano et al., 2015a). In autopolyploids of these species, DNA methylome and morpho-anatomic changes were stochastic and a systematic effect of autopolyploidization could not be identified (Aversano et al., 2013). In this report, we describe the comprehensive transcriptome and metabolome profiling of leaf tissues of S. commersonii and S. bulbocastanum, and present a model whereby autopolyploidization triggers, through nucleotide precursor pool imbalance, a genomic shock which in turn results in downstream stochastic events responsible for the ‘revolutionary’ phase following autopolyploidization.

Materials and Methods

Plant material

Synthetic autotetraploids of Solanum commersonii Dunal and S. bulbocastanum Dunal were generated as previously described (Caruso et al., 2011). In detail, a clone of diploid S. commersonii (PI 243503), coded CMM1T, and a clone of diploid S. bulbocastanum (PI 275190), coded BLB1C, were subjected to oryzaline treatment. Tetraploids deriving from CMM1T were coded CMM15, CMM24, CMM27 and CMM30. Those produced from BLB1C were coded BLB10, BLB22, BLB25 and BLB26. Clonal progenies of these experimental materials were stabilized for at least 3 yr (i.e. through five passages of clonal propagation through tubers) before being used in this study. All plants were also maintained and propagated in vitro on Murashige and Skoog (MS) medium (Murashige & Skoog, 1962), including salts, vitamins, sucrose (30 g l⁻¹) and microagar (9 g l⁻¹) and adjusted to pH 5.8. The cultures were maintained in a walk-in growth chamber at 25°C ± 2 under a 16 h:18 h (light: dark) photoperiod at 125 μmol m⁻² s⁻¹ irradiance.
provided by a cool white fluorescent tube (TL-D 58W/33-640 15L; Philips, Eindhoven, the Netherlands). To collect material for molecular and metabolomic analyses, 10 healthy and uniform plants per genotype were grown in 25-cm-diameter pots (spacing between plants was 30 × 30 cm), in a temperature-controlled greenhouse (24°C : 16°C, day : night, relative humidity 65–75%, and natural photoperiod). To minimize environmental effects, plants were grown according to a fully randomized design. Fully expanded third to fifth leaves were sampled at the appearance of the first flower buds, between 08:00 and 09:00 h, and frozen in liquid nitrogen.

**Microarray analysis**

The CombiMatrix CustomArray™ 90K platform (Mukilteo, WA, USA) was used. Chip design and annotation were derived from the SolEST database (D’Agostino et al., 2009). The CombiMatrix *S. tuberosum* chip was produced by the Plant Functional Genomics Center at the University of Verona. The chip contained 27,234 nonredundant probes in triplicate, composed of 35–40-mer oligos. Probes were designed based on tentative consensus (TC) sequences (23,453 probes) and singletons with a 3’ poly(A) tail (46 probes) using OligoArray v.2.1 (Rouillard et al., 2003). Oligo probes were designed to identify the 3’ untranslated region (UTR) of genes. BLASTx comparisons against the UniPortKB/Swiss-Prot database were exploited to determine the correct open reading frame and to define forward/reverse TC sequence orientation. There were 13,207 TC sequences which had a forward orientation, while 2027 had a reverse orientation. In the case of 9000 TC sequences, no BLAST hits were detected and it was not possible to assess where their 3’-UTR was located for these sequences. As a consequence, we filtered out 6000 TC sequences generated by assembling the largest number of expressed sequence tags and considered both the orientations for probe design. Nine bacterial oligonucleotide sequences provided by CombiMatrix, 40 probes designed on ArrayControl RNA spikes (ThermoFisher Scientific, Rockford, IL, USA) and 11 additional negative probes based on *Bacillus anthracis* strain 2002013094 (CP009902.1), *Haemophilus ducreyi* CLU5 (CP011227.1) and *Alteromonas* phage PM2 (AF155037.1) sequences were used as negative controls. Three to four replicates of each probe were distributed randomly across the array. A total of 10 microarrays were employed, one for each genotype. They were used three times, one for each biological replicate. Array hybridization, stripping and re-hybridization were performed following the CustomArray™ 90K Microarray: Hybridization and Imaging Protocol (PTL020; CombiMatrix CustomArray™) and the Stripping and Preparation of CombiMatrix 90K Microarrays for Re-hybridization protocol (PTL025), as recommended by CombiMatrix. From each genotype, RNA was isolated from a 100-μg pooled sample of leaves obtained by homogenizer (TissueLyzer; Qiagen, Valencia, CA, USA). Total RNA was extracted using the Trizol reagent (Life Technologies, Carlsbad, CA, USA) following the manufacturer’s protocol. Total RNA concentration and purity were determined using a NanoDrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA) (A260 : A280 > 1.8). The integrity of RNA was assessed with an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). The synthesis and labelling of antisense-RNA (aRNA) were performed using the RNA ampULSe kit (with Cy5 for CombiMatrix arrays) (Kreatech Biotechnologies, Amsterdam, the Netherlands). The purified, labelled aRNA was quantified using a spectrophotometer and 4 μg was hybridized to the CombiMatrix array according to the manufacturer’s directions. Imaging of array slides was performed using a GenePix® 4400A Microarray Scanner (Molecular Devices, Sunnyvale, CA, USA) at 5-μm resolution and controlled using the GenePix Pro v.7 software (Molecular Devices). The latter was also used for densitometry analysis and raw data extraction. Probe signals higher than negative control values plus twice the standard deviation were considered as ‘present’. Signal probe normalization and analyses were performed as previously reported (Di Matteo et al., 2010). In detail, signal probe medians and standard deviations were imported into the SPSS software package v.19 (SPSS Inc., Chicago, IL, USA). Normalization between arrays was achieved by correcting each probe median based on the ratio between the median of the array and the average median of arrays. Centred medians were log2 transformed and probe signals having a variability coefficient > 0.5, as well as spikes and factory probes, were filtered out. Also, probes showing the 10% highest and the 10% lowest average signal intensities were removed. To visually inspect normalization array, box plots were created (data not shown). Probes with significant changes in their hybridization signals were identified using TIGR MULTIPLE EXPERIMENT VIEWER v.4.0 (Saeed et al., 2003). Average gene expression values from microarray data are reported in Supporting Information Table S1. These and the other microarray data obtained in this work have been submitted to EMBL-EBI/ARRAYEXPRESS with provisory accession number A-MTAB-562. To find differentially expressed genes (DEGs) between each tetraploid and its diploid counterpart, a Student’s t-test with 100 permutations and a critical P-value < 0.01 was performed. To detect reproducible changes attributable to autopolyploidy, the combined tetraploids of each species were compared to the diploid parent using the same statistical procedure previously described. BLAST2GO (Conesa et al., 2005) was used to provide automatic and high-throughput annotation, gene ontology mapping, and categorization of TC sequences showing differential transcription signals. In particular, BLAST comparisons were performed using a threshold of the expectation value of 10^-10. DEGs were extracted for GO functional enrichment analysis. Enrichment analysis was performed using a false discovery rate (FDR) significance cut-off of 0.05. INTERProScan (Quevillon et al., 2005) was used to assign a protein domain, Enzyme Commission (EC) numbers and Kyoto Encyclopedia of Genes and Genomes annotations which were used to map genes on metabolic pathways. *Solanum tuberosum* TC sequences used for probe design were aligned along the potato reference sequence (v.4.03) by running GENOMETHREADER v.1.6.2 (Gremme et al., 2005) with default parameters. Circos v.0.64 (Krzywinski et al., 2009) was used to display differentially expressed transcripts as histogram plots. The ITAG v.1.0 annotation based on the potato genome sequence DM v3.4 (Potato...
Quantitative reverse transcription–PCR analysis

Quantitative reverse transcription (qRT–PCR) experiments were carried out in biological triplicate with the same RNA samples as used for microarray analysis, using the Qiagen 2× QuantiFast SYBR Green PCR Master Mix (Qiagen) and the ABI PRISM 7900HT Instrument (Life Technologies). Complementary DNA was synthesized from DNase-treated total RNA using the SuperScript™ III Kit (Life Technologies). Gene-specific primers were designed for the target genes as well as the adenine phosphoribosyltransferase (APRT) transcript (Table S2). Each 25-μl reaction included 300 nM of each primer and cDNA synthesized from 1 μg of total RNA (three replicates for each reaction) and began with a 50°C hold for 2 min and a 95°C hold for 10 min followed by 40 cycles at 95°C for 30 s, 55°C for 30 s, and 72°C for 20 s. Data acquisition was performed during the combined annealing/extension step. A melting curve analysis of the PCR products was performed to verify their specificity and identity. Results were then analysed using the ABI Prism 7900HT Sequence Detection System v.2.1 (Applied Biosystems, Foster City, CA, USA). Relative quantification was performed using the comparative cycle threshold (ΔΔCt) method. The endogenous control gene (APRT) was used to normalize the cDNA of each sample. The relative expression was estimated according to the ΔΔCt method (Livak & Schmittgen, 2001).

Liquid chromatography electrospray ionization mass spectrometry (LC-ESI-MS) analysis of leaf semi-polar metabolome

Five mg of ground lyophilized leaf powder from each biological replica was extracted with 0.75 ml of cold 75% (v/v) methanol and 0.1% (v/v) formic acid, spiked with 10 μl of 10 μM stock solution of acetonitrile. A biological replica consisted of a pool of two different plant genotypes. After vortexing for 30 s, samples were shaken for 15 min at 15 Hz using a Mixer Mill 300 (Qiagen) and kept at room temperature for 5 min (twice). After centrifugation for 15 min at 20 000 g at 4°C, 0.6 ml of supernatant was removed and transferred to high-performance liquid chromatography (HPLC) tubes. For each genotype, five independent extractions were performed. Liquid chromatography–mass spectrometry (LC-MS) analyses were carried out using the LTQ-Orbitrap Discovery mass spectrometry system (Thermo Fisher Scientific, Waltham, MA, USA) operating in positive electrospray ionization (ESI) mode, coupled to an Acela U-HPLC system (Thermo Fisher Scientific). LC separations were performed using a C18 Luna column (Phenomenex, Aschaffenburg, Germany) (150 × 2.0 mm; 3 μm). The mobile phases used were water – 0.1% formic acid (A) and acetonitrile – 0.1% formic acid (B). The gradient was: 95% A, 5% B for 1 min, followed by a linear gradient to 25% A: 75% B over 40 min. LC conditions were maintained for a further 2 min, before returning to the initial LC conditions for 18 min. Ten microlitres of extract were injected and an LC flow of 0.2 ml min⁻¹ was used during the whole LC runs. Detection was performed continuously from 230 to 800 nm with an online Accela Surveyor photodiode array detector (PDA; Thermo Fisher Scientific). All solvents used were LC-MS grade quality (Chromasolv®; Sigma-Aldrich). Metabolites were quantified relative to the internal standard values. ESI-MS ionization was carried out using the following parameters: the capillary temperature was set at 275°C; sheath, aux and sweep gas flow rates were 35, 15 and 5 l min⁻¹, respectively. Spray voltage was set at 3.20 kV, while capillary voltage and tube lens were set at, respectively, 30 and 85 V. Metabolite identification was performed through comparison of chromatographic and spectral properties of authentic standards and reference spectra, and on the basis of the m/z accurate masses, as reported in the Pubchem database for monoisotopic mass identification, or in the Metabolomics Fiehn Lab Mass Spectrometry Adduct Calculator for adduct ion detection. ANOVA was applied to the absolute metabolite quantifications, and pair-wise comparisons were performed using Tukey’s HSD (honest significance difference) test in PAST v.3.x (Hammer et al., 2001). Heatmaps, hierarchical clustering, correlation matrices and networks were generated as previously described (Dretto et al., 2010).

Results

Chromosome doubling remodels the transcriptome of independent tetraploids of CMM1T and BLB1C

There were 6947 transcripts which were differentially expressed in at least one tetraploid derived from clone CMM1T of S. commersonii, whereas 269 transcripts showed differential expression in all tetraploids (132 always up-regulated and 121 always down-regulated) (Fig. 1a; Table S3). In CMM24 and CMM27, most of the DEGs were down-regulated, whereas in CMM15 and CMM30 the numbers of up- and down-regulated genes were similar (Fig. 1b). There were 5509 differentially expressed transcripts which were detected in at least one tetraploid derived from clone BLB1C of S. bulboacastanum. One hundred and nineteen transcripts were differentially expressed in all tetraploids, of which 104 were always up-regulated and eight always down-regulated (Fig. 1a; Table S3). In all tetraploids of both clones, we checked whether there were common genes responding to polyploidization, 10 have a role in nucleotide-binding processes, including RNA binding, splicing, modification, processing and regulation of transcription and translation. In addition, two ion channels, two ribosomal proteins, one
photosystem II component, one gene involved in protein targeting, and seven genes of unknown function are also included in this set. The expression of a random sample of these genes was confirmed using (q)RT-PCR, the rates obtained being similar to those reported by others (Hegarty et al., 2006; Poole et al., 2007) (Fig. S1). Genes differentially expressed in all tetraploids of each species (Table S4) were mapped on the *S. tuberosum* reference genome (Potato Genome Sequencing Consortium, 2011), and their distribution on chromosomes is shown as a circular plot (Fig. 2). Chromosomes 9 and 8 had the lowest percentage of DEGs in tetraploids of CMM1T (3.5%) and BLB1C (1.7%), respectively, whereas chromosomes 10 and 9 had the highest percentage of DEGs in tetraploids of CMM1T (4.8%) and BLB1C (2.9%), respectively. Fig. 2 also shows the percentage of differentially expressed transcripts within nonoverlapping 2-Mb windows along the 12 chromosomes. Strikingly, in tetraploids of CMM1T, the pericentromeric regions of several chromosomes (6, 9 and 10 and, to a lesser extent, 3, 5 and 11) showed an unusually high percentage of DEGs. By contrast, only chromosome 1 showed a high percentage of DEGs in the pericentromeric region in BLB1C tetraploids. GO annotation of genes whose expression was affected by autopolyploidization in either species (Fig. S2) provided evidence that the most commonly recurring GO term within the molecular function category was ‘binding’. We did not find any particular enrichment in either of the species or any difference in distributions between the species using an FDR value of 0.05 in Fisher’s exact test.

Metabolite alterations following autopolyploidiation are predominantly stochastic, rather than species-specific or common

Semi-polar leaf extracts from the tetraploids and their diploid parents were subjected to metabolic profiling by liquid chromatography (LC) coupled to electrospray ionization high-resolution mass spectrometry (ESI-HRMS) operating in positive ion mode. Overall, 216 unique metabolites were identified and quantified in at least one tetraploid genotype of CMM1T (184 metabolites) and BLB1C (163 metabolites). Metabolite concentrations in the two species and in their synthetic tetraploids are shown in Fig. S3 and Table S6. Several amides, amino acids and phenylpropanoids were more abundant in CMM1T tetraploids, while some amines and sugars/polyols were more abundant in BLB1C tetraploids. Polyploidization affected both primary and secondary leaf metabolism of both species, but to different extents (Table 1). Compared with the diploid parent, 69 metabolites showed significant variations in all *S. commersonii* tetraploids, with alterations affecting mainly alkaloids, amino acids, flavonoids, organic acids and vitamins. Significant variations were observed for additional metabolites in only one or two tetraploids of CMM1T (Table 1a), suggesting that they were the result of stochastic, rather than systematic, responses to autopolyploidization. Consistent with transcriptome studies, BLB1C tetraploids displayed lower frequencies of significantly altered metabolites, with 21 being changed in all tetraploids with respect to the corresponding diploid and additional metabolites altered in only one or two
tetraploids (Table 1b). Again, amino acids, flavonoids, organic/phenolic acids and alkaloids showed the largest variation. Notably, most differentially accumulated metabolites found in tetraploids of the two species were different, or altered in opposite directions (Fig. S3). Out of 216 detected molecules, only 20 varied in all tetraploids and only five did so in the same direction in the two species: two phenolic acids (chlorogenic acid and its isomer, cryptochlorogenic acid) were increased in both CMM1T and BLB1C tetraploids, while three molecules involved in DNA metabolism (adenosine, guanine and guanosine) showed the opposite trend. Interestingly, the extent of variation for all five metabolites was much greater in CMM1T than in BLB1C tetraploids, while three molecules involved in DNA metabolism (adenosine, guanine and guanosine) showed the opposite trend. Interestingly, the extent of variation for all five metabolites was much greater in CMM1T than in BLB1C tetraploids (Table 2). The Pearson correlation coefficient was calculated for each metabolite pair and compounds were sorted genetically within different metabolic classes. In CMM1T tetraploids we found several significant (r ≥ 0.7; p < 0.05) correlations (Table S7a). Nucleotides and nucleosides were negatively correlated with most of the other metabolites, with the exception of a few amino acids (leucine, lysine, methionine, oxo-proline and phenylalanine), phenylpropanoids (chalcone, cinnamic aldehyde and coniferyl alcohol-aldehyde) and sugars (lycotetraose, maltilitol and pinitol), which displayed positive correlations. Alkaloids specifically accumulating in S. commersonii leaves, such as commersonine and dehydrocommersonine, were also highly correlated with most amino acids and phenylpropanoids, and with some organic/phenolic acids and sugars. The same pattern was observed for other alkaloids, such as acetoxy-hydroxy-dehydrotomatine, acetylleptinidine and calystegines A3, B1 and B2. By contrast, the metabolome of BLB1C tetraploids was characterized by few significant correlations, without consistent trends for metabolites belonging to the same class (Table S7b), with vitamins showing the highest number of significant correlations, positive with amines, amides and some amino acids (alanine, arginine, asparagine and aspartate), and negative with phenylpropanoids and peptides.

The folate pool is constitutively less abundant in S. commersonii than in S. bulbocastanum polyploids

Given the generalized decrease in nucleotide precursor pools in polyploids of both species, and the more extensive transcriptional and metabolic perturbations observed in tetraploids of S. commersonii, we searched for systematic differences between the two species in metabolites correlated with DNA metabolism. One such class of metabolites are folates, of which three (tetrahydrofolate (THF), 10-formyltetrahydrofolate (FTFH) and 5,10-methenyltetrahydrofolate (MTHF)) were detectable in our analyses (Fig. S3). THF is the product of the bifunctional thymidylate synthase/dihydrofolate reductase enzyme and, together with MTHF, is a cofactor in thymidylate de novo biosynthesis; FTFH acts as a donor of formyl groups in purine de novo biosynthesis, where it is a substrate for phosphoribosylaminomimidazole carboxamide formyltransferase (reviewed in Zrenner et al., 2006). The most abundant compound in the pool was FTFH, which was c. 30-fold more abundant in genotypes of S. bulbocastanum than in those of S. commersonii, followed by THF and MTHF, which were below the level of detection in S. commersonii.

Transcript-metabolite correlations reveal significant interactions between DEGs and the common differentially expressed metabolites

The DEGs listed in Table S3 were subjected to Pearson’s correlation analysis with the metabolites from genotypes of S. commersonii and S. bulbocastanum (Table S6). Whereas it was not possible to identify specific correlation trends according to the metabolic class in S. commersonii (Table S8a), in S. bulbocastanum most metabolites were characterized by a general negative correlation with the DEGs (Table S8b). The differentially expressed transcripts were mapped to the potato metabolic pathways using MapMan (Thimm et al., 2004). In agreement with metabolite data, we observed several transcripts involved in primary (glycolysis, gluconeogenesis, tricarboxylic acid and glyoxalate cycle, photorespiration and beta-oxidation) and secondary (phenylpropanoid and isoprenoid) plant metabolism (Table S3). Overall, with the exception of alkaloids and amides, we found alterations in gene expression in all the pathways taken into consideration in the metabolomic analysis. Finally, we sought to identify the DEGs that showed the highest co-regulation with the ‘common differentially accumulated metabolites’. The results of this correlation analysis are shown in Fig. 2 Chromosomal distribution and relative abundance of differentially expressed transcripts in synthetic polyploids using the Solanum tuberosum (TBR) chromosomes as reference. Distal and pericentromeric regions, as defined by markers described in Sharma et al. (2013), are shown in light and dark grey, respectively. Light grey bars above chromosomes represent the frequency of mapped transcripts within a window of 2 Mb. CMM and BLB, S. commersonii and S. bulbocastanum pseudo chromosomes. Red and green bars represent the fold change of up- and down-regulated genes in CMM and BLB, respectively. Blue histograms represent the percentage ratio of differentially expressed transcripts/total transcripts within a window of 2 Mb. * , percentage ratio ≥ 60.
Table S9 and summarized in the co-regulation network shown in Fig. 3. In the analysed clones of both species, a large number of genes (569 in tetraploids of CMM1T and 253 in those derived from BLB1C) showed significant ($|q| \geq 0.7$) co-regulation with the ‘common differentially accumulated metabolites’. However, the two sets of co-regulated genes in the two species were largely nonoverlapping, indicating that species-specific responses were prevalent. Further studies with a larger number of genotypes per species are necessary to confirm this species-specific response. Only 14 genes were significantly co-regulated with the ‘common differentially accumulated metabolites’ in both species (Table S10). They fall into two categories: genes that are regulated in a similar fashion in both species (Sotub05g006060, Unknown Protein; Sotub03g015260, Peptidoglycan-binding lysine domain protein; Sotub06g006820, Early tobacco anther 1; Sotub06g032910, 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 4; Sotub04g007480, nucleotide binding site-coding resistance gene analogue) and genes regulated in opposite fashions in the two species. The latter genes, which might provide a molecular basis for the differential responses to polyploidization of the two species, comprise Sotub11g024720 (Phosphoribide $\alpha$ oxidase); Sotub11g019460 (oligonucleotide/oligosaccharide binding (OB) -fold nucleic acid-binding); Sotub08g027910 (Pleckstrin homology); Sotub08g029480 (Phospholipase-like); Sotub12g022100 (Phox-like); Sotub01g012130 (Rhodopsin-like receptor); Sotub10g029710 (DNA/RNA helicase, DEAD/DEAH box type); Sotub02g005400 (Sacsin-like) and Sotub11g013430 (unknown protein).

**Discussion**

Perturbations of the transcriptome in *Solanum* autopolyploids largely depend on the diploid parent from which they are derived

A higher fraction of parental genes showed changes in expression in tetraploids of *S. bulbocastanum* versus *S. commersonii*. A second difference found was in the percentage of DEGs. Analysis of tetraploids of the former species confirmed that the gene transcription level may increase proportionally to gene dosage (Guo Table 1 Number of differentially accumulated metabolites ($P$-value $\leq 0.05$) in each (a) *Solanum commersonii* and (b) *Solanum bulbocastanum* tetraploid with respect to the diploid parent

<table>
<thead>
<tr>
<th></th>
<th>Common</th>
<th>CMM 15</th>
<th>CMM 24</th>
<th>CMM 30</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total (no.)</td>
<td>Up (no.)</td>
<td>Dw (no.)</td>
<td>% Tot</td>
</tr>
<tr>
<td>(a) Acids and esters</td>
<td>29</td>
<td>9</td>
<td>3</td>
<td>6.5</td>
</tr>
<tr>
<td>(a) Alkaloids and saponins</td>
<td>44</td>
<td>6</td>
<td>2</td>
<td>4.3</td>
</tr>
<tr>
<td>(a) Amides</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1.6</td>
</tr>
<tr>
<td>(a) Amines</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(a) Amino acids</td>
<td>23</td>
<td>2</td>
<td>5</td>
<td>3.8</td>
</tr>
<tr>
<td>(a) DNA metabolism</td>
<td>6</td>
<td>0</td>
<td>5</td>
<td>2.7</td>
</tr>
<tr>
<td>(a) Peptides</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1.1</td>
</tr>
<tr>
<td>(a) Phenylpropanoids</td>
<td>34</td>
<td>12</td>
<td>8</td>
<td>10.9</td>
</tr>
<tr>
<td>(a) Polar isoprenoids</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>(a) Polar lipids</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>1.1</td>
</tr>
<tr>
<td>(a) Sugars/polyols</td>
<td>19</td>
<td>1</td>
<td>4</td>
<td>2.7</td>
</tr>
<tr>
<td>(a) Vitamins</td>
<td>8</td>
<td>1</td>
<td>3</td>
<td>2.2</td>
</tr>
<tr>
<td>Total</td>
<td>184</td>
<td>37</td>
<td>32</td>
<td>37.5</td>
</tr>
<tr>
<td>(b) Acids and esters</td>
<td>31</td>
<td>3</td>
<td>0</td>
<td>1.8</td>
</tr>
<tr>
<td>(b) Alkaloids and saponins</td>
<td>26</td>
<td>2</td>
<td>6</td>
<td>4.9</td>
</tr>
<tr>
<td>(b) Amides</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(b) Amines</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>(b) Amino acids</td>
<td>23</td>
<td>0</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>(b) DNA metabolism</td>
<td>5</td>
<td>0</td>
<td>4</td>
<td>2.5</td>
</tr>
<tr>
<td>(b) Peptides</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(b) Phenylpropanoids</td>
<td>32</td>
<td>0</td>
<td>2</td>
<td>1.2</td>
</tr>
<tr>
<td>(b) Polar isoprenoids</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>1.2</td>
</tr>
<tr>
<td>(b) Polar lipids</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(b) Sugars/polyols</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(b) Vitamins</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>163</td>
<td>6</td>
<td>15</td>
<td>12.9</td>
</tr>
</tbody>
</table>

Common, metabolites differentially accumulated in all tetraploids of each species; Up and Dw, numbers of up-regulated and down-regulated metabolites, respectively; % Tot, fraction of differentially accumulated metabolites within all metabolites detected.
et al., 1996; Galitski et al., 1999; del Pozo & Ramirez-Parra, 2014). By contrast, results in *S. commersonii* imply that in tetraploids of this species a dosage-dependent regulatory mechanism is not present, as reported also in *S. phureja* (Stupar et al., 2007) and *Paulownia fortunei* (Zhang et al., 2014). The analysis shown in Fig. 2 revealed a third difference: in *S. commersonii* autopolyploids, DEGs were highly enriched in pericentromeric regions, while no preferential chromosomal region emerged as responsive to autopolyploidy-dependent gene regulation in *S. bulbocastanum* autopolyploids. The different response to polyploidization of the two species is consistent with their relatively high genetic distance, as demonstrated by marker analysis (Castillo & Spooner, 1997; Traini et al., 2013). It is also consistent with our transcriptome study, where c. 8000 genes (33%) were found to be differentially expressed between the two species.

Using more than one independent synthetic tetraploid of each species, we were able to address the question of whether there were genes that exhibited a common and predictable response to polyploidization. Overall, our data suggest that after autopolyploidization, for a large portion of genes, expression changes are genotype-specific. For a few others, rewiring seems to occur in a concerted and predictable manner. Previous studies on some of them (e.g. 6-phosphofructo-2-kinase, Glycosyltransferase, Phospholipase, Mitochondrial assembly regulatory factor, F-box protein 7) reported their differential expression related to abiotic and biotic stresses (Banzai et al., 2003; Rietz et al., 2004; Lorenc-Kukula et al., 2005). This limited number of ploidy-regulated genes was as expected (Pignatta et al., 2010). In

![Fig. 3 Co-regulation network of differentially accumulated common metabolites (from Table 2) and differentially expressed genes (DEGs) (from Table S3) in *Solanum commersonii* and *Solanum bulbocastanum* tetraploids. Only genes with correlations (\(r > 0.95\)) with at least one common metabolite are shown. Grey, green and red identify common, *S. commersonii*-specific and *S. bulbocastanum*-specific nodes, respectively.](image-url)
addition, it should be considered that microarray analysis may have failed to detect relatively small changes or may have been influenced by the cross-species nature of our study (Poole et al., 2007). For this reason we are currently profiling synthetic autotetraploids using RNA sequencing (RNA-Seq). This experiment is also including diploids derived from the same oryzalin treatment to exclude possible effects attributable to the antimitotic agent. GO annotation of DEGs of our tetraploids covered several important terms and also revealed that 'response to stress' was highly represented, as previously reported (Stupar et al., 2007; Yu et al., 2010; Allario et al., 2011; Tan et al., 2015).

Tetraploids of *S. commersonii* exhibit a larger number of differentially accumulated metabolites than those of *S. bulbocastanum*

The four metabolite classes most affected by polyploidization were organic acids, alkaloids, phenylpropanoids and amino acids, albeit with qualitative differences in terms of molecules affected. As organic/phenolic acids play a fundamental role both in cell metabolism and physiology and in the response to abiotic/biotic stress (Lopez-Bucio et al., 2000; Fernie et al., 2004), our data suggest that polyploidization dramatically remodels carbon metabolism, as already indicated by GO term analysis. In some cases, transcript and metabolite data could be directly linked: in tetraploids of *S. commersonii* we found higher expression of phosphoglycerate dehydrogenase, which uses hydroxypyruvate to produce phosphoglycerate (Toujani et al., 2013); similarly, caffeoyl CoA 3-0-methyltransferase, acting upstream of hydroxycinnamic acid biosynthesis (Maury et al., 1999), was overexpressed, in agreement with the accumulation of sinapic acid. Tetraploids of *S. commersonii* exhibited a larger number of differentially accumulated metabolites than those of *S. bulbocastanum*. Consistent with transcriptome data, most of the metabolite alterations were stochastic, rather than common. In *S. commersonii* tetraploids, glutamic acid, a key molecule in amino acid metabolism (Forde & Lea, 2007), was increased, while leucine and glutamine, which play essential roles in plant development, stress responses and nitrogen assimilation (Mentzen et al., 2008; Seabra et al., 2013; Yu et al., 2013), were reduced. Several transcripts were correlated with the changes in glutamate and glutamine concentrations, such as alanine/aspartate branched-chain amino acid transaminases (Lipman & Olsen, 2003). Tetraploids of both species exhibited a reduction in phenylalanine concentrations, which correlated well with the increased levels of phenylalanine ammonia-lyase (PAL) transcript, which is involved in the conversion of phenylalanine to cinnamic acid. PAL was also up-regulated in autotetraploids of *Isatis indigotica* (Lu et al., 2006). Further perturbations related to phenylalanine metabolism affected amides such as *N*-caffeylputrescine and feruloylputrescine. These molecules are precursors of a series of tropic acid-based alkaloids, such as hyoscyamine and calystegines (Medina-Bolivar & Flores, 1995; van der Rest et al., 2006), and protect plants against herbivore attacks (Kaur et al., 2010). Both amides and calystegines were highly induced in *S. commersonii* autotetraploids.

Common and species-specific responses to autopolyploidization: a model

In both species, the comparison of tetraploids and their diploid parents identified two organic acids (chlorogenic and cryptochlorogenic acid) and three molecules involved in DNA metabolism (adenosine, guanine and guanosine) that were, respectively, strongly increased and reduced in all autotetraploids. These metabolic changes are triggered by polyploidization in a nonstochastic, species-independent way. Chlorogenic and cryptochlorogenic acids are phenolic acids with marked antioxidant activity, and are involved in stress responses (Niggeweg et al., 2004; Cle et al., 2008). We hypothesize that cells of both species perceive genome doubling as a stress event, which leads to the accumulation of these metabolites. A similar effect was previously described in *Hylocereus* (Cohen et al., 2013). Reduction of adenosine, guanine and guanosine was much more extensive in *S. commersonii* tetraploids than in those of *S. bulbocastanum*. Their depletion might be diagnostic of the fact that intracellular metabolism cannot cope with the increased requirement for DNA building blocks caused by the increased DNA content. Intracellular dNTP pool sizes are known to be precisely regulated through both biochemical and genetic control mechanisms, and the malfunctioning of such mechanisms results in a phenomenon known as 'nucleotide pool imbalance’, leading to genomic instability events such as enhanced mutagenesis, stimulation of genetic recombination, chromosomal abnormalities, DNA breakdown, and cell death (Kaepplet et al., 1998; Mathews, 2006). Therefore, it is attractive to speculate that polyploidization results in a nucleotide pool imbalance, which in turn triggers genomic stress leading to a ‘revolutionary phase’. This can then cause the stochastic events responsible for the vast differences detected in all polyploids studied here. A prediction of this hypothesis is that the higher nucleotide precursor depletion observed in *S. commersonii* tetraploids would result in more extensive genomic stress and in a higher number of stochastic events with respect to those of *S. bulbocastanum*. This is exactly what we observed.

Do polyploids show any symptoms, other than nucleotide precursor depletion, of nucleotide pool imbalance? The answer to this question is yes: as already mentioned, nucleotide biosynthesis is a tightly regulated process, with several regulatory mechanisms acting at the gene expression level to maintain nucleotide pool homeostasis. Mapping of the transcriptome data on the general metabolic pathways by MAPMAN identified several differentially regulated transcripts involved in purine biosynthesis, which are all up-regulated in *S. commersonii* tetraploids, while in tetraploids of *S. bulbocastanum* both up- and down-regulation events are observed (Fig. S4). But why do tetraploids of *S. commersonii* show a more extensive nucleotide pool imbalance than tetraploids of *S. bulbocastanum?* The metabolomic data provide a plausible answer. *De novo* synthesis of purines is absolutely dependent on FTHF, a cofactor of GAR (glycinamide ribonucleotide) transformylase (Zrenner et al., 2006). We showed that FTHF is 30-fold less abundant in *S. commersonii* than in *S. bulbocastanum*. Therefore, a logical conclusion is that tetraploids of
S. commersonii are less capable of compensating purine nucleotide depletion caused by polyploidization by activating de novo purine biosynthesis, because of their lower endogenous FTHF levels.

A final very important question arises with respect to the model discussed here: what are the signal transduction mechanisms whereby nucleotide pool imbalance results in the genomic stress events triggering the ‘revolutionary phase’? A first answer to this question is provided by the functional classification of the genes showing significant correlations with the common differentially accumulated metabolites (Table S9). The most interesting genes are those that show the opposite behaviour in the polyploids belonging to the two species, that is, are induced in tetraploids of S. commersonii and repressed in those of S. bulbocastanum, and may underlie the different responses of the two species to nucleotide pool imbalance: these include an unusually high percentage of genes involved in signalling (including genomic stress signalling) and cellular homeostasis, such as ‘genome guardians’ maintaining genome stability (OB-fold proteins) (Flynn & Zou, 2010); post-transcriptional gene silencing (DNA–RNA helicases) (Dalmay et al., 2001); phosphoinositide signalling (phospholipase and pleckstrin homology domain) (Ferguson et al., 1995); cell death (pheophorbide α oxygenase) (Pruzinska et al., 2003); detection of a variety of (mostly extracellular) signals, including those from purine and pyrimidine ligands (rhodopsin-like receptors) (Ralevic & Burnstock, 1998); and plant–microbe interactions (LysM domain proteins (Gust et al., 2012), phox-like proteins (Sang et al., 2001) and DnaJ molecular chaperones (Anderson et al., 2011)).

In the light of the considerations made thus far, we would like to present a model (Fig. 4) whereby autopolyploidization affects the metabolome and transcriptome in S. commersonii and S. bulbocastanum, resulting in common, species-specific and stochastic remodelling. First, polyploidization induces an increased requirement for precursors of DNA synthesis (nucleosides). This is suggested by the observed reduction in purine base and nucleoside levels. It was not possible to measure those of pyrimidine bases and nucleosides, or of all nucleotides, probably because of their low levels and/or stability in our analytical conditions (Traut, 1994). Second, the reduction in DNA precursor levels results in nucleoside pool imbalance. This is suggested by the extensive transcriptional remodelling of nucleotide biosynthetic genes. Third, nucleoside pool imbalance is more extensive in S. commersonii than in S. bulbocastanum, because the former species is unable to compensate for the increased biosynthetic requirements, as a consequence of the decreased levels of FTHF, an essential cofactor for the biosynthesis of nucleosides. Fourth, nucleoside pool imbalance induces the accumulation of chlorogenic and cryptochlorogenic acids, two compounds with marked antioxidant activity which are involved in stress responses; as chlorogenic acid is known to bind some purine analogues such as caffeine with high affinity (Sondheimer et al., 1961), the increased levels of this acid may contribute further to purine nucleoside imbalance. Furthermore, nucleoside pool imbalance causes in both species the transcriptional remodelling of a series of key genes involved in stress signalling and in the control of cellular homeostasis (‘genome stress signalling genes’). Fifth, these genes induce a ‘genome stress syndrome’ that in turns triggers some of the species-specific and all of the stochastic events that we observed.

The model presented here may also explain the polyploid syndrome previously reported for several phenotypic traits in these and other polyploids (Stupar et al., 2007; Allario et al., 2011; Yao et al., 2011; Aversano et al., 2013, 2015b; Chen, 2013). Proving or disproving this model or parts thereof lies beyond the scope of this paper. However, its capacity to explain most of the observations resulting from the transcriptome and metabolome analyses and its predictive power make it, in our view, an interesting working hypothesis for future studies on polyploidization.

**Acknowledgements**

We thank M. Walters for language editing; C. Conicella, F. Consiglio and D. Rosellini for suggestions; R. Nocerino and R. Garramone for technical support; and the European Union for
financial support (From discovery to products: A next generation pipeline for the sustainable generation of high-value plant products; FP7 Contract 613153).

Author contributions

References


Supporting Information

Additional supporting information may be found in the online version of this article.