University of Naples Federico II
Department of Agriculture

Joint international Ph.D.
Environmental Resources Sciences
XXV Cycle (2010-2013)

EFFECT OF POLYPLOIDIZATION ON THE TRASCIPTOME AND METABOLOME IN SINTETHIC POLYPLOIDS OF *SOLANUM spp*

Ph.D. Dissertation
by Fasano Carlo

TUTOR:
Dott. Riccardo Aversano

COORDINATOR:
Prof. Maria A. Rao

INDEX

1. Introduction

1.1. Polyploidy: definition, diffusion and importance

Polyploidy is the heritable condition of possessing more than two complete sets of chromosomes. Although most eukaryotic species are diploid, meaning they have two sets of chromosomes (one set inherited from each parent), polyploidy is very common in nature. The traditional view is that genomic repetition, or better genomic duplication through polyploidy, has played a central role in generating the biodiversity through the evolutionary history of living organisms. In the animal kingdom, polyploidy is rarely found in invertebrates (e.g. insects) and vertebrates (e.g. bony fishes, amphibians and reptiles) (1)probably because of their delicate schemes of sex determination and development(2). By contrast, in the plant kingdom polyploidy is very much spread, especially among flowering plants.

The prevalence of polyploidy in plants has classically been estimated by comparative analyses of chromosome numbers and other biosystematic approaches. In the early 1900 Müntzing (1936) and Darlington (1937)speculated that about one half of all angiosperm species were polyploid. Later, Grant (1963)and Lewis (1980) reported that at least 70% (perhaps 80%) of monocots and dicots were of polyploid origin. Similarly, Masterson (1994) estimated that 70% of all angiosperms experienced at least one round of genome duplication in their ancestry based on the comparison of leaf guard cell size in fossil and extant taxa from a few angiosperm families (*Platanaceae*, *Lauraceae*, *Magnoliaceae*). More recently, Otto and Whitton (2000)suggested that roughly 2 – 4% of speciation events in angiosperms may have involved polyploidy and therefore "polyploidization may be the single most common mechanism of sympatric speciation in plants".

Although these cytogenetically-based approaches accomplished to unravel the prevalence of polyploidy among angiosperms, they definitely underestimated the role that polyploidy has played in shaping modern plant genomes. In the last decade, a renewed and growing interest from the scientific community on polyploidy has spread over, especially among the so-called "omic" sciences (see the special issue of New Phytologist, 2010; 186 (1) dedicated to plant polyploidy). In particular, one of the most spectacular advances of the genomics era has been a great appreciation of the pervasiveness and importance of genome doubling in plant evolution. The availability of extensive EST data sets and the saturated linkage maps, revealed that all plant genomes harbor evidence of cyclical, recurrent episodes of genome doubling. Recently, plant genome sequencing data provided evidence that species once considered classical diploids, such as *Arabidopsis thaliana*

(L.) Heynh., *Oryza sativa* L., *Populustrichocarpa* (Torr. & Gray) and *Vitisvinifera* L., contain signatures of ancient polyploidization events in their genomes and are therefore paleopolyploids (9, 10).

In addition, molecular and computational tools from the post-sequencing era have provided new ways to probe the history of genomes, leading to the discovery that even the smallest plant genomes considered to be diploid incurred at least one round of whole-genome duplication, possibly predating the origins of the angiosperms (11). Taken the evidences as a whole, the emerging picture is that polyploidy is a dynamic process that has contributed to evolutionary diversification and speciation in all eukaryotes through successive rounds of polyploidization and rediploidization (12, 13)(Fig. 1).

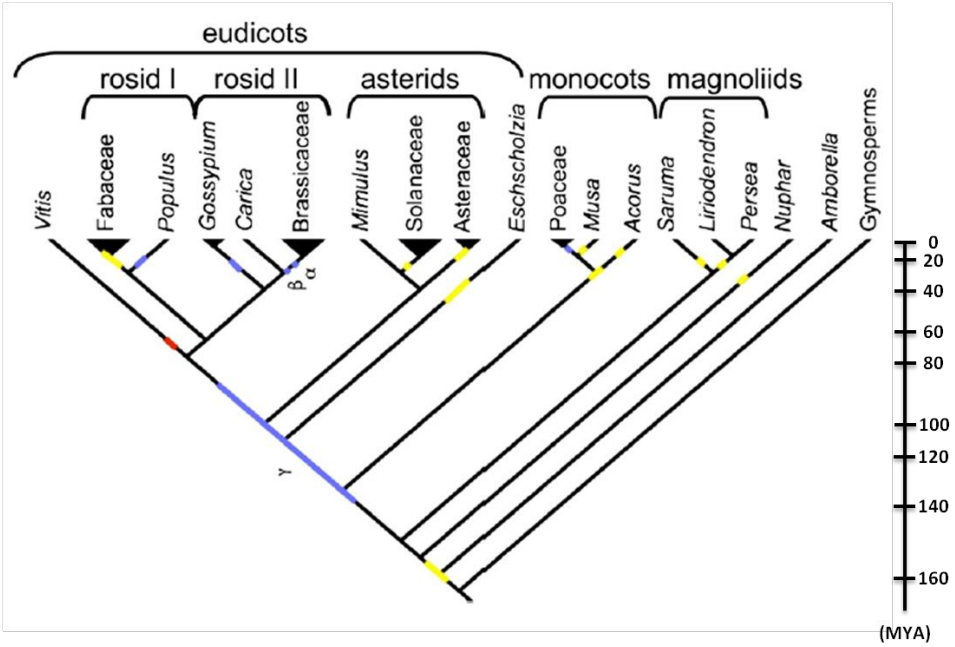

Figure 1. Philogenetically tree for angiosperms (following Soltis *et al*., 2000), with modifications reflecting more recent analyses, including Jansen *et al*., (2007); Moore *et al*., (2007), depicting putative locations of genome duplication events. The α, β, and γ duplication events of Bowers *et al*., (2003)have been added to this topology. The α event likely occurred within *Brassicaceae*. The position of β is less certain; it may also have occurred within *Brassicaceae*, or perhaps earlier in *Brassicales*. Blue bars represent genome duplications inferred from comparative analyses of physical and/or genetic maps. The red bar on the *Vitis* lineage represents the Velasco *et al*., (2007) inference of a second distinct tetraploidy evident in the *Vitis* genome. Yellow bars represent genome-wide duplication events.The timing and nature of the γ event are controversial.Modified by Soltis & Soltis (1999).

The large diffusion of polyploid plant species obviously contradicted the view of polyploids as evolutionary dead end. Polyploids often show novel phenotypes that are not present in their diploid progenitors or exceed the range of contributing species (18-20), possibly extending the types

of environments over which normal development can take place. In fact, it has been often noted that there is a higher frequency of polyploid plants at higher latitudes and altitudes, particularly in arctic populations (17, 21-24). A possible and common explanation is that polyploids are capable of tolerating and invading harsher environments than their diploid counterparts due to changes in physiology and/or increased genetic buffering provided by extra genome copies (14, 25). Indeed, one of the most common features of polyploids is the *Gigas* plant type, characterized by an enhanced vigour and larger plant organs, tissue and cells, higher tolerance to mineral and nutrient stress, and enhanced resistance to environmental stresses (24, 26). However, nowadays the mechanisms that best explain current polyploidy distributions are still unclear.

Many domesticated crops have long been recognized as polyploids. Wheat, canola, tobacco, peanut, plum, apple, pear, sugarcane and cotton, for example, possess allopolyploid genomes. Other domesticated crops have a history of autopolyploidy, including coffee, banana, watermelon, strawberries, potato and alfalfa (Tab. 1). Given that, one could say not only that the life on earth is predominantly a polyploid plant phenomenon, but also that civilization has been mainly powered by polyploid food, fodder and fiber (27).

Table 1. Examples of polyploid crops classified by use. The somatic chromosome number is reported in brackets. From Aversano *et al*, 2012.

Cereals	*Triticumaestivum*(6x= 42); *T. durum*(4x= 28); *Avena sativa* (6x=42); *A. nuda* (6x=42)
Forage grasses	*Dactylisglomerata* (4x=28); *Festuca arundinacea* (4x=28); *Agropyron repens* (4x=28); *Paspalum dilatatum* (4x=40)
Legumes	*Medicago sativa* (4x=32); *Lupinus alba* (4x=40); *Trifolium repens* (4x=32); *Arachis hypogaea* (4x=40); *Lotus corniculatus* (4x=32)
Industrial plants	*Nicotiana tabacum* (4x=48) *Coffea spp*. (4x=44 fino a 8x); *Brassica napus* (4x=38); *Saccharum officinalis* (8x=80) *Gossypium hyrsutum* (4x=52)
Tuber plants	*Solanum tuberosum* (4x=48); *Ipomea batatas* (6x=96); *Dioscorea sativa* (6x=60)
Fruit trees	*Prunus domestica* (6x=48); *Musa* spp (3x=33; 4x=44); *Citrus aurantifolia* (3x=27); *Actinidia deliciosa* (4x=116); *Prunus cerasus* (4x=32)

1.2. Origin and classification of polyploids

Polyploidy in angiosperms has been studied for a century, dating to the work of De Vries (29, 30) and to early interest in putative chromosome duplication in maize (Zea mays) (31). It was Winkler (1916) who introduced for the first time the term polyploidy and observed that different chromosome numbers within some genera followed an arithmetic progression of multiplication (within the *Poaceae*, for example, some species are 2n=14, others 2n=28, 42 and so on).

There is a basic distinction between autopolyploids and allopolyploids. Both have multiple sets of chromosomes, but in the former these are of the same type and have the same origin, whereas in the latter both the type and the origin are different. Indeed in autopolyploids all genomes are identical or very similar and arise via genome duplication within the same species. Allopolyploids, by contrast, contain two or more distinct genomes, and can arise via hybridization of two different species together with genome doubling. The total number of chromosome sets is indicated by the prefix; for example, tri- (3), tetra- (4), penta- (5), hexa- (6) and octa- (8). Alternatively, polyploids could be defined as polysomic polyploids and disomic polyploids. These terms are not based on origin (often unknown) or genome differentiation, but on genetic behavior at meiosis. Polysomic polyploids are made up of homologous chromosomes that segregate from

multivalents or random bivalents. By contrast, disomic polyploids are made up of two or more sets of homeologous chromosomes that segregate from bivalents formed only between homologous chromosomes (33).

How are polyploids formed in nature? Two prevailing models may explain their origins. According to the "one-step" model, both auto- and allopolyploids can arise after unreduced gamete formation by the union of two unreduced gametes from the same plant or different plants, as proposed, for example, for autotetraploid *Dactylis glomerata* (34) and allotetraploid *Tragopogon* species (35). Alternatively, the "two-step" model proposes that the production of either an auto- or allotetraploid may involve a 'triploid bridge', in which triploids are formed within a diploid population, and backcrossing to diploids (3x gamete X haploid gamete) or selfing of the triploid (3x gamete X haploid gamete produced by a single individual) produces a tetraploid. The "two-step" method has been considered a significant pathway to polyploid formation (36, 37), but some (38, 39) have suggested that the "one-step" process involving the union of two unreduced gametes may be more common than often considered because 2n gametes have been reported in a number of species, most notably those that also produce polyploids (Fig. 2).

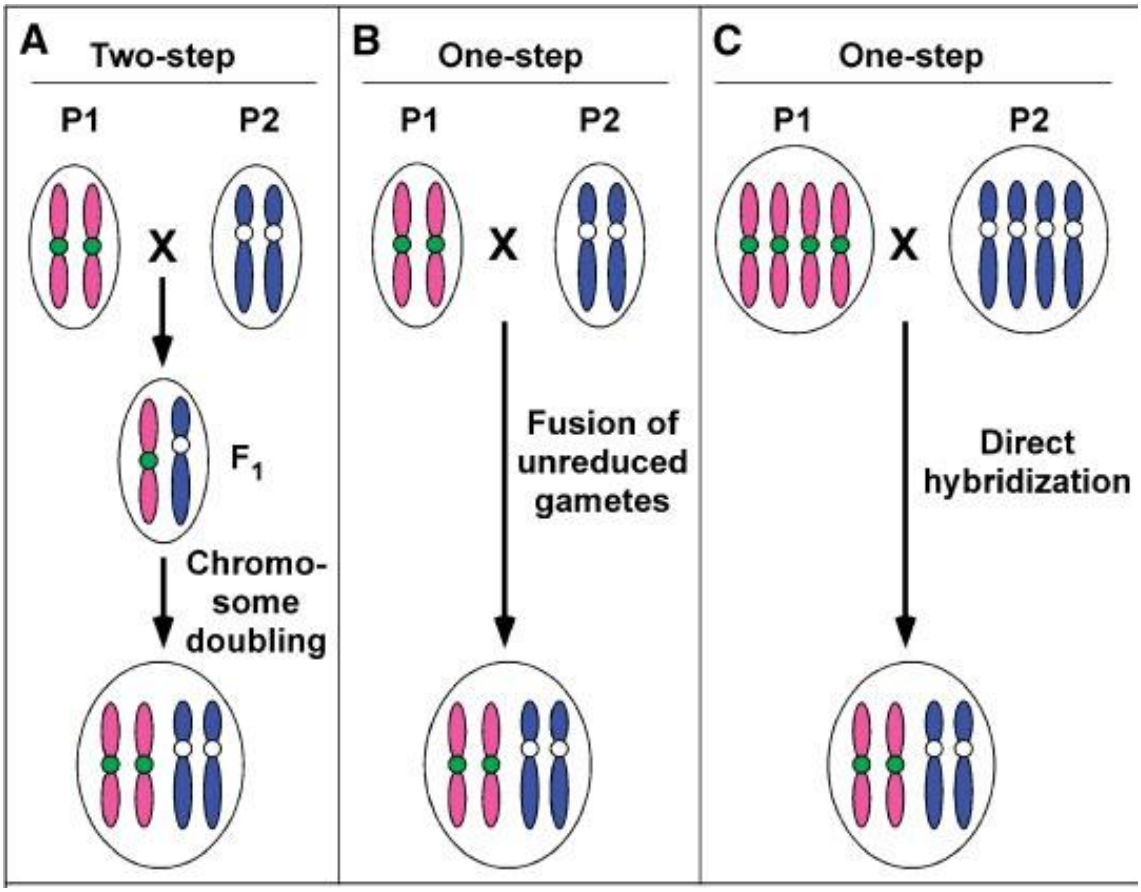

Figure 2. Two models (‘‘one-step’’ and ‘‘two-step’’) for the formation of polyploids. A: An allotetraploid, combination is formed by hybridization between two diploid progenitors followed by chromosome doubling (two-step model). For simplicity, each diploid species has one pair of chromosomes. P1 and P2 represent two progenitors. B: Fusion of unreduced male and female gemetes of two diploid progenitors leads to the production of an allotetraploid (one-step model). C: An allotetraploid is immediately formed by direct hybridization between two autotetraploid species (one-step model). Modified from Chen & Ni (2006).

Anyway, in either model the unreduced gametes are believed to be a major mechanism of polyploid formation. Meiotic aberrations related to spindle formation, spindle function, and cytokinesis have been implicated as the cause of 2n gamete production in non hybrid crop cultivars (38). For example, a parallel spindle (“ps”) orientation at anaphase Il results in the reconstitution of diploid nuclei in microsporogenesis, and pre-mature cytokinesis that immediately follows the first meiotic division creates diploid nuclei that never undergo a second meiotic division. Therefore, parallel spindle is genetically equivalent to a first-division restitution (FDR) mechanism and leads at the end of meiosis to the formation of a dyad of two 2n microspores. This mechanism is controlled by a single recessive gene, as proved in potato by Mok & Peloquin (1975) (Fig. 3). Another well documented cause of unreduced gamete formation is the omission (“os”) of the second meiotic division. When “os” is present, after the first division there is no second division,

and the chromatids just fall apart. Of the resulting two 2n megaspores, one gives rise to the 2n female gametophyte and the other degenerates. Omission of the second division is also a simply inherited Mendelian recessive (41) that has been detected in a large number of *Solanum* species (41-43). Omission of the second division is genetically equivalent to a second-division restitution (SDR) mechanism (Fig. 3)

In nature both allopolyploidy and autopolyploidy are common, but probably allopolyploidy predominates. As for autopolyploids, for a long time they were believed to suffer from several evolutionary disadvantages when compared with allopolyploids, and this led Clausen *et al*., (1945) and Stebbins (1971) to propose that autopolyploids are rare and represent evolutionary dead-ends. In particular, multivalent formation after genome multiplication may cause meiotic irregularities and result in reduced fertility compared with the diploid progenitors. Views on the importance of autopolyploidy have changed considerably since 1981. Genetic studies conducted during the 80s and early 90s indicated unambiguously that autopolyploidy was much more common than traditionally maintained. Importantly, autopolyploids should exhibit tetrasomic or higher level inheritance, rather than disomic inheritance. For instance, at given locus (es. A) with two alleles (es. A and a), five genotypes are possible in an auto/polysomic polyploid: aaaa, Aaaa, AAaa, AAAa and AAAA. A specific terminology has to be applied according to the number of dominant alleles per locus: nulliplex, simplex, duplex, triplex and quadruplex, respectively. By contrast, only three genotypes characterize allo/disomic polyploidy: AA, Aa and aa.T he higher number of alleles per locus provides the opportunity for a larger number of intra-locus interactions: a total of 11 (6 of first order –between two alleles-; 4 of second order –between three alleles- and 1 of the third order –between four alleles) vs. one of the first order, as well as inter-locus interactions. Thus, maximum heterozygosity can be achieved. For traits whose genetic variance is almost entirely non-additive, this is very important. Indeed, breeder must look for maximum heterozygosity to optimize heterotic combinations (46, 47).

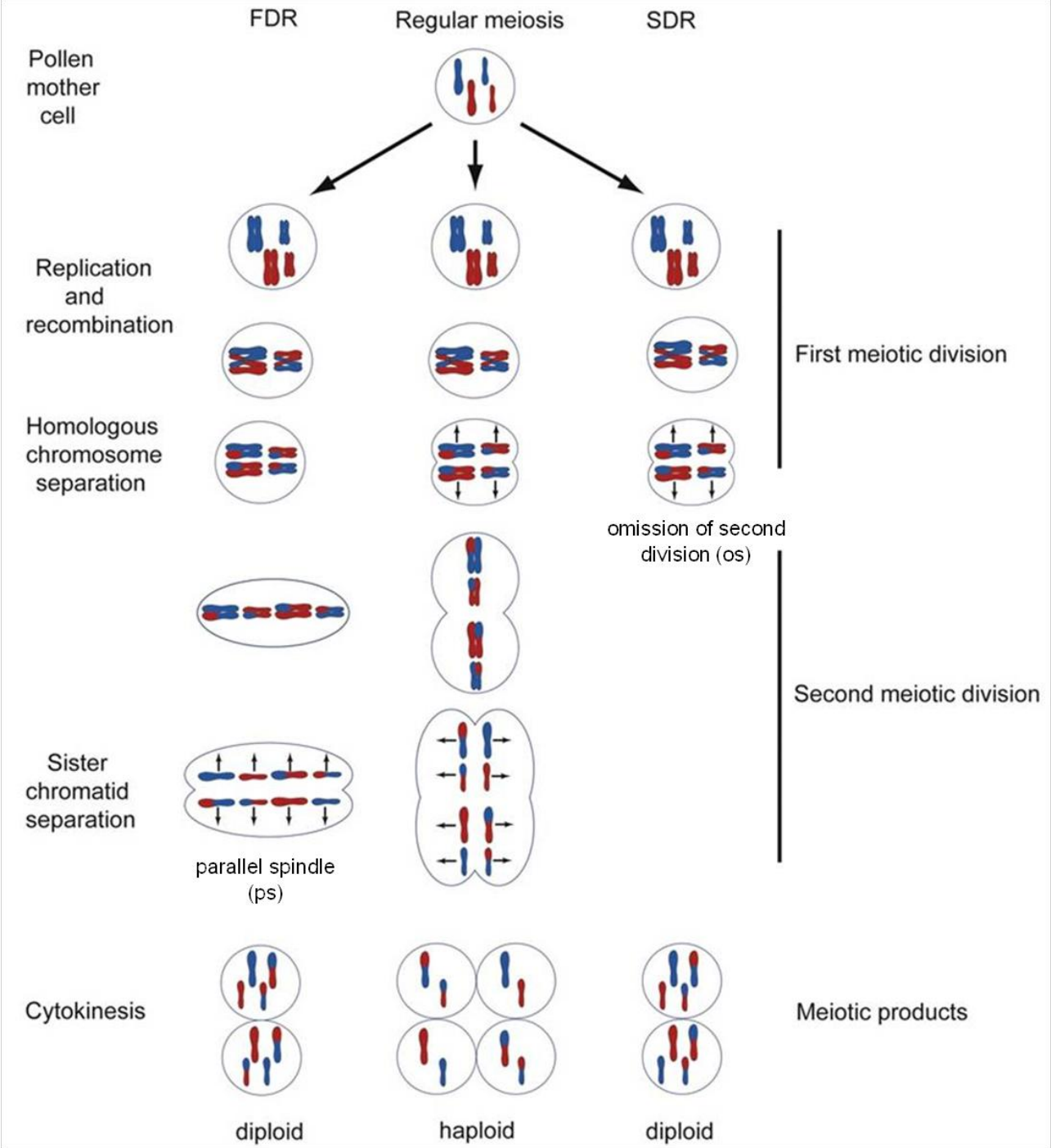

Figure 3. A model for the origin of nuclear restitution by first division restitution (FDR) or second division restitution (SDR). In wild-type meiosis, after replication and recombination, homologous chromosomes are separated in meiosis I and sister chromatids are separated in meiosis II. The spores are haploid and carry recombined genetic information. FDR causes a failure in homologous chromosome separation during meiosis I, while separation of sister chromatids occurs, resulting in the formation of 2n spores containing recombined non-sister chromatids. SDRleads to an omission of the second meiotic division, giving rise to diploid spores with recombined sister chromatids. Small arrows indicate the direction of chromosome or chromatid movement. Modified from Köhler et *al*., (2010).

Although it is now clear that autopolyploids are much more common than traditionally maintained, they are probably not as common as allopolyploids. We stress the word 'probably',

however, because we still have no firm estimate of the relative abundance of autopolyploids. There are numerous examples of single taxonomic species with multiple cytotypes. For instance, potato species form a polyploid series with a basic chromosome number (x) of 12 and a minimal chromosome differentiation among taxa. Most of them (~70%) occur as diploids (2n = 24), and the others are triploids (2n = 36), tetraploids (2n = 48), pentaploids (2n = 60), and hexaploids (2n = 72). The main cultivated potato is a tetrasomic tetraploid originating from South America. *S. tuberosum* Group Andigena is cultivated mainly under short days. *S. tuberosum* Group Tuberosum is the potato cultivated worldwide under long-day conditions. Also, some diploids, triploids, and pentaploids are still cultivated in South America, as well as a large number of wild tuber-bearing species (~160). They grow in a wide range of habitats from the southern part of the United States through Mexico and Central America down to southern Chile (49).

1.3. Genetic, epigenetic and metabolic consequences of polyploidization

Polyploidization entails a reorganization of the full genome to enable harmonious cohabitation in the same nucleus of genes of different origin (in the case of allopolyploids) or of duplicated genes from identical or very similar genomes (in the case of autopolyploids). The structural and functional modifications induced by polyploidization are probably the key to the success of polyploids in nature. Indeed, polyploidization may greatly affect the chromosome structure, gene expression regulation, epigenetic landscape, and secondary metabolism of newly formed offspring (Fig. 4).

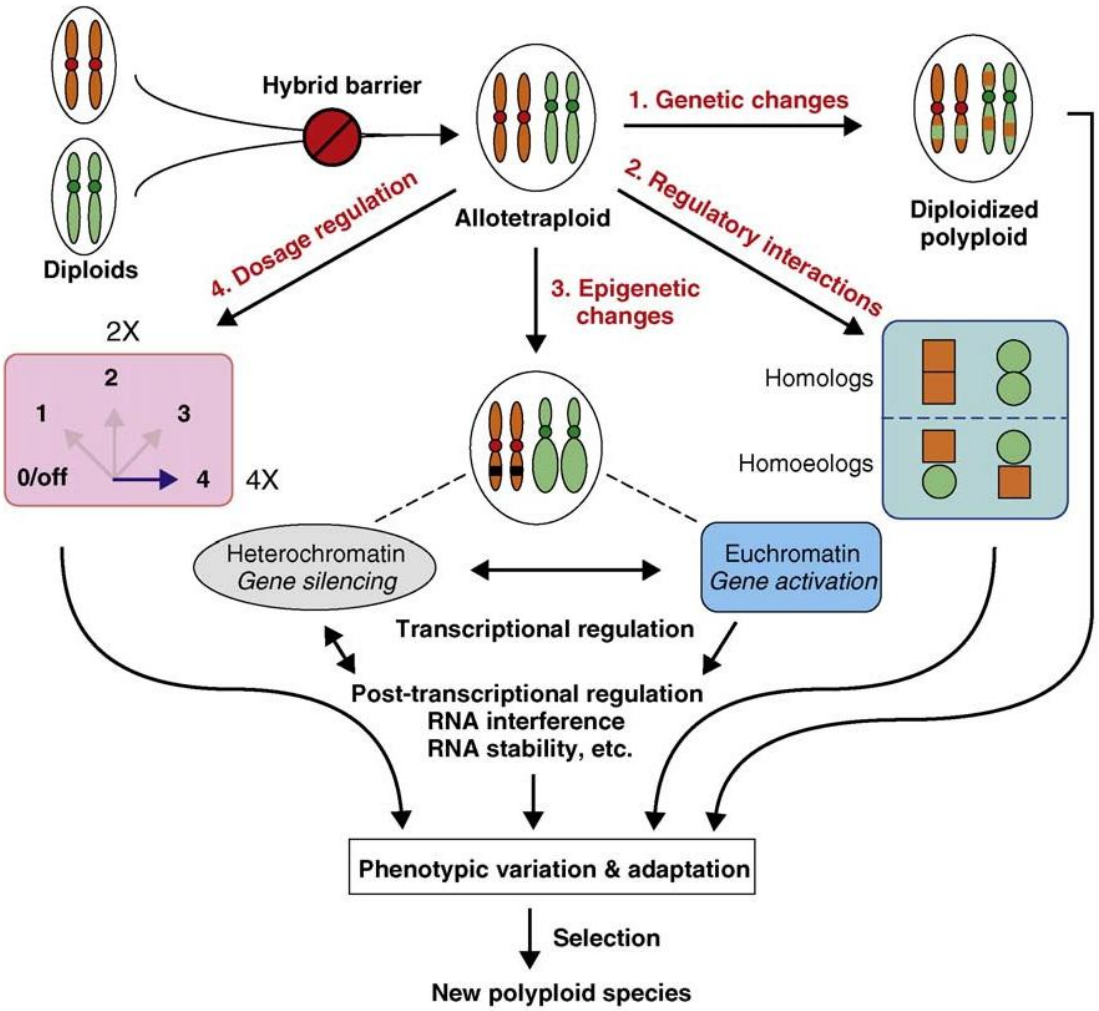

Figure 4. Mechanisms of the full genome reorganization in polyploids. Polyploid formation is usually impaired by the hybridization barrier (red stop sign) between species. A tetraploid may undergo rapid changes via genetic and epigenetic mechanisms, dosage regulation, and regulatory interactions. First, genetic changes (e.g. chromosomal rearrangements and gene loss) may cause gene expression and phenotypic variation. Some duplicate genes may diverge expression patterns. Second, regulatory proteins produced from orthologous genes may generate incompatible products (heterodimers between red squares and blue circles), leading to silencing of one homoeologous gene in allopolyploids. Alternatively, homoeologous proteins may perform better than the homozygous forms,which may explain overdominant effects on hybrid vigor in allopolyploids. Third, gene expression patterns may be reprogrammed by epigenetic changes in new polyploids. Hybridization or allopolyploidy induces formation of heterochromatin and euchromatin, resulting in gene silencing or activation via transcriptional and post-transcriptional mechanisms. Finally, dosage regulation suggests additive effects of gene expression in autopolyploids, leading to increased levels of expression and phenotypic variation. The dosage may augment the effects of hybridity (or genome merger) on gene expression and phenotypes. Gene expression changes may be selected by natural and/or artificial forces that facilitate adaptation and development. Modified from Jackson & Chen (2010).

1.3.1. Genetic consequences of polyploidization

Whole-genome duplication events have been associated with genome rearrangement, including aberrant recombination, transposable element activation, meiotic/mitotic defects, and intron expansions and contractions (51). The first study demonstrating whole-genome changes after chromosome doubling was carried out in *Brassica* spp. by Song *et al*., (1995), who reported nuclear modifications in newly synthesized polyploids on the basis of restriction fragment length polymorphism (RFLP) patterns. The authors found rapid and extensive genomic changes in early generations from F2 to F5. Similarly, Lukens *et al*., (2006), using different independently resynthesized *Brassica* lines, observed that changes in S_0 were uncommon, whereas in early subsequent generations several fragment losses occurred due to homeologous recombination resulting in non-reciprocal translocations. In wheat allotetraploids, two studies investigated the changes associated with allopolyploidy by extensive genome-wide analysis with AFLP marker, followed by validation by DNA gel-blotting (54, 55). Perhaps a most astonishing finding emerged from these studies is the spectrum of changes observed, as rapid physical sequence elimination may affect up to 15% of the genomic loci in some species combinations (55). Indeed, this large-scale genome reduction at largely low-copy genomic regions was corroborated by a later demonstration that tandem DNA repeats specific to one of the parental species also underwent rapid elimination that is visible at the cytological level (56). Remarkably, this pervasive retraction force of allopolyploidy was independently documented in wheat by flow cytometric (57, 58), and were also found in *Triticale*, a man-made allopolyploid species by crossing wheat and rye (59) followed by cochicine-mediated genome doubling (60-63). In these series of studies by using multiple molecular markers including AFLP, RFLP, inter-simple sequence repeat (ISSR), inter-retrotransposon amplified polymorphism (IRAP) and retrotransposon-microsatellite amplified polymorphism (REMAP), it was documented that allopolyploidy in *Triticale* was associated with genome-wide rearrangements with a prominent type of change being elimination of the rye-specific fragments that often represent retrotransposons or their derivatives. Similar observations were also made in other plant systems (64-66). By contrast, results coming from the molecular analyses of *Gossypium* (67), *Spartina anglica* (68, 69), and *Tragopogon* natural allopolyploids (70), showed that these species presented a very low frequency of genomic changes, probably because they displayed a high level of tolerance to genome doubling and interspecific hybridization. Taken as whole, these findings suggested that although it is believed that genome rearrangements may be necessary for restoring nuclear-cytoplasmic compatibility (17), it is clear that certain species may be able to

respond to allopolyploidy and adapt to a novel genomic environment without a corresponding reorganization of the genome.

As autopolyploids arise by doubling of similar homologous genomes, it is believed that substantial alterations might not be required after their formation (71, 72). According to most findings reported in the literature, two different structural changes may occur following autopolyploidization: short- and long-term changes. The former occur immediately after polyploidization and may be important for the successful establishment of polyploids. The latter, occur during the whole lifespan of polyploids and potentially explain their long-term persistence. Immediate loss of 17% of total DNA has been observed in synthetic autopolyploids of *Phlox drummondii*, with a further reduction of up to 25% upon the third generation (73). In the exclusively bivalent pairing *Elymuse longatus*, both newly synthesized autopolyploids as well as natural accessions showed a loss of 10% DNA compared with their diploid progenitors (74). Again, a synthetic autotetraploid of *Paspalum notatum* exhibited 9.5% sequence elimination relative to its diploid progenitor (75).

Long-term structural changes were detected in established autopolyploids. For example, the young autotetraploids *Biscutella laevigata* showed successive down-sizing of its genome along its post-glacial recolonization pathway(76). Genome size variation was also detected in established autopolyploids of *Triticeae*(74). Whereas the multivalent forming *Hordeum bulbosum* showed no significant change in genome size compared with its diploid progenitor, the strictly bivalent forming *Hordeummurinum* subsp. *murinum* showed a 23% reduction in the amount of DNA. However, the bivalent forming *Hordeummarinum* subsp. *gussoneanum* unexpectedly showed the exact additive genome size of its progenitor, indicating that at least in some cases there is no association between genome down-sizing and pairing behavior of chromosomes. Genome arrangements might also involve the relocation of chromosomal segments. Asymmetrical relocation and loss of rDNA loci has been reported for the autopolyploid *Hepatica nobilis* var. *pubescens* (77). In a synthetic autotetraploid lineage of *Arabidopsis thaliana* grown for 20-30 generations, translocation of the 45S rDNA from chromosome 4 to chromosome 3 has been identified (78). This accession forms > 85% bivalents, which contrasts with multivalent pairing observed in newly synthesized autopolyploids (79), suggesting that diploidization proceeded through structural rearrangements within 30 generations.

Along with the described cases of short- and long-term changes, no major chromosomal rearrangements were found after genome doubling in *A. thaliana* (79, 80), *S. phureja* (81), *S. bulbocastanum*, *S. commersonii* (82), *and Citrus limonia* (83).

1.3.2. Epigenetic consequences of polyploidization

Ploidy-induced structural genomic changes are strongly associated with epigenetic alterations. The latter refer to heritable changes in gene expression that do not entail a change in nucleotide sequence. The classical epigenetic modification include DNA (cytosine) methylation, and histone modifications, such as methylation, acetylation, phosphorylation, ubiquitylation, and sumoylation (84). In addition, microRNA and small interfering RNAs are increasingly being recognized as important players in the various epigenetic mechanisms. This important feature, along with the knowledge of naturally occurring DNA methylation polymorphisms or epigenetic alleles (epialleles) accumulated within and between natural plant populations (85-88) has led to the belief that epigenetic variations may have played an important role in genome evolution (89, 90). So, while genetic mutations can explain the cause of variability mainly over evolutionary time, epigenetic changes, which are potentially reversible, offer a flexible expedient for a polyploid cell to respond to genomic shock, especially in the early stages of polyploid formation, leading to various genetic variations.

Some studies found that DNA methylation changes occurred in newly formed allopolyploids of *Arabidopsis*(91) and wheat (55). In natural allopolyploids, a subset of *A. thaliana* loci, including transposons and centromeric repeats, were heavily methylated and subjected to homologous genome-specific RNA-mediated DNA methylation; however, reduced methylation at some loci was also observed (92). In a synthetic allotetraploid (*Aegilopssharonensis* × *Ae.umbellulata)*, some genetic alterations were associated with comparative hypomethylation of the promoter region within the *Ae. umbellulata*-derived rDNA units (93). In another example, the endoreduplication of cells during tomato fruit ripening was not accompanied by changes in the level of methylation, but the patterns of methylation changed (94). This suggests that methylation patterns may be rapidly established on dramatically multiplied chromosomes (95). Moreover, in allopolyploids, some studies have shown polyploidy-induced demethylation and activation of dormant mobile elements as a consequence of genomic instability or "genomic shock". For example, activation of transposons appears to have occurred in some allopolyploid lines of *Arabidopsis*(96). In newly synthesized wheat allopolyploids, Kashkush *et al.,* (2003) showed that activation of certain transposable elements could result in silencing of the adjacent downstream genes. Active transposable elements have the potential for insertional mutagenesis and changes in phenotype while altering local patterns of gene expression, but they may also facilitate rapid genomic reorganisation in new polyploids. Because polyploids contain duplicate copies of all genes,

transposon insertions into "single" copy genes are less deleterious; thus, transposable elements can multiply and persist for far longer in polyploids than in diploids (98).

In autotetraploid, methylation changes after genome doubling have not been detected at random marker loci in *Paspalum* (75) and watermelon (99), but not in other systems. In autotetraploid *A. thaliana* (100, 101) it was reported ploidy-dependent methylation changes at a hygromycin-resistance transgene with subsequent silencing (and modified resistance phenotype). Investigating the mechanisms beyond such polyploidy-associated transcriptional gene silencing, Baubec *et al*., (2010) found that two epigenetic marks, namely DNA methylation and histone methylation, cooperate to give rise to a 'double lock' on transcriptional silencing, thus generating an extremely stable epigenetic state. More recently, Aversano *et al*., (2013) found alterations in cytosine methylation level in synthetic autopolyploids of two wild *Solanum* species, *S. bulbocastanum* and *S. commersonii* compared their diploid parents. The authors, identified various types of changes, including both hypo- and hyper methylation at CG and CHG sites, known to be the most affected targets of DNA methyltransferase (103). For both species, no change occurred concurrently in all tetraploids analyzed with respect to their diploid parent, revealing a stochastic trend in the changes observed. The occurrence of stochastic methylation alterations in autopolyploids was also reported in maize (104). Finally, enhanced cytosine methylation as a result of genome duplication were reported by Ochogavía *et al*.,(2009) in *Eragrostis curvula* and Lavania *et al*., (2012) in *Cymbopogon* spp.

1.3.3. Polyploidization effect on metabolism

Polyploidization can greatly influence plant metabolism both in a quantitative and qualitative manner. The evidence derived from chemical analysis undertaken on induced allopolyploids indicates that such polyploids are generally richer in phenolic compounds and also express greater enzymatic diversity than either of the parents. The newer chemical types so developed offer the distinct advantage amenable for natural selection. The production of hybrid compounds in allopolyploids has been explained as the modification of the products of one parental genome by the genome of other parent (107-110). Probably, intergenomic complementation of the blockages/limiting factors attendant in the biosynthetic pathway of one parent by the other brings about such an effect. The production of an esterase isozyme in an allooctoploid *Triticale* not present in either parent has been considered as an example of the chemical basis of heterosis (111, 112)t hus supporting the argument that induced allopolyploids can be significantly different from their diploid ancestors. In *Coffea* an allotetraploid with improved liquor quality has been developed by

interspecific hybridization between *C. arabica* and an induced tetraploid *C. canephora*(113). The amphidiploids produced from a cross between *O. cimumgratissimum* (rich in eugenol) and *O. viride* (rich in thymol) provides a successful example of chemical heterosis. This allotetraploid exhibits increases for both vegetative and floral characters as well as for the concentration of secondary metabolites i.e. eugenol and thymol rich essential oil (114). More recently, it was been shown how *Nicotiana* allopolyploids changed profile of secondary metabolites manufactured in defense of herbivory between individuals (115) and between individuals and their parents (116). In nature, such physiological differences may affect allopolyploid evolution and survival. Moreover many genes involved in plant metabolism was found upregulated in the allotetraploids (117).

In autopolyploids the basic genetic material remains the same, but gene dosage is multiplied. Chromosome doubling is often accompanied by conspicuous changes in secondary metabolism. A common model to explain these changes relies on the assumption that the lower ratio of nuclear membrane to chromatin results in more chromatin coming into contact with the nuclear membrane, thereby enhancing gene activity, improving water relationships, hormonal status and photosynthetic rate on a per cell basis (118), coupled with enlarged cell size. All of these factors could have a positive impact on secondary metabolism and the production of active phytopharmaceuticals (119). Often polyploidy brings about an increase in the concentration of secondary metabolites, although a reduction is observed in some cases. In addition, an altered chemical profile of metabolic products is frequently encountered, which is itself of potential value for qualitative improvement through ploidy manipulation. Levy (1976) found qualitative differences in the glycoflavone profiles of 14 of the 15 autotetraploids induced from cultivars of *Phlox drummondi*. This included 14 instances of additional flavonoids in the autotetraploid that were lacking in the progenitor diploid, and eight where flavonoids were present in the diploid but absent in the autotetraploid. Lavania (1988) was able to attain qualitative improvement in the perfume value of the essential oil of vetiver autotetraploids, which had an enhanced proportion of khusinol at the cost of a reduction in the concentration of β-vetivone compared to the diploid progenitor. Such 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. The multiplicity of allelic combinations that is possible in polyploids offers the advantage of producing hybrid enzymes and enzymatic diversity, and thus provides a greater level of versatility and homeostasis than is possible by breeding and selection at the diploid level (112, 121). Dhawan and Lavania (1996) reported, as polyploidization consequences, a lower concentration per unit dry weight of essential oils in autotretraploids of *Menthaspicata* than in the diploid genotypes. In

synthetic autotetraploids of *S. commersonii,* Caruso *et al*., (2011), found no qualitative differences in the metabolite profiles between the diploid and its tetraploids. By contrast, the results showed that the phenylpropanoid content was generally significantly higher in the tetraploids than in the diploid *S. commersonii*. Furthermore, in the same study, evidences of a significant increase of minor glycoalkaloids (solanidenediol triose, solanidadienollycotetraose, and solanidenollycotetraose) as well as a reduction of major glycoalkaloids (dehydrodemissine and dehydrocommersonine) were concomitantly observed.

1.4. Gene expression regulation in polyploids

Gene expression that accompany polyploidization has been the subject of numerous investigations and the general picture of the phenomenon is still far to be well understood. One generalization that has emerged from the literature is that gene expression in polyploids often is non-additive. In particular, repression or silencing of gene expression has frequently been found in synthetic and natural allopolyploids, and this is observed far more frequently that up-regulation or novel gene expression.

Using cDNA-AFLP techniques, approximately 5% of loci were repressed in *Tragopogon* (70) and cotton (123) polyploids, whereas about 10% of genes were reported as repressed in *Arabidopsis* polyploids (124). In synthetic *Triticum aestivum*allohexaploids (125), about 8% of transcripts displayed altered expression, and >95% of them were reduced or absent. Large-scale microarray studies in a range of polyploid plant species have confirmed that allopolyploids exhibit considerable transcriptome alterations as compared with their diploid progenitors. Such changes are mostly due to the reunion in a common nucleus of previously diverged regulatory hierarchies (each from the diverged parental genomes), which likely entails non-additive gene expression. In particular, Wang *et al*., (2006) (117) in a detailed microarray study that examined the regulation of 26000 genes in *Arabidopsis neopolyploids* (obtained crossing a self-compatible *A. thaliana* autotetraploid with the *A. arenosa* tetraploid) detected a strong expression dominance of the *A. arenosa* parent, coupled with suppression of the *A. thaliana* genome. The extent of this suppression was impressive; approximately 94% of the genes up-regulated in the *A. thaliana* parent but not in *A. arenosa* were subsequently down-regulated (suppressed) to the level of the *A. arenosa* after allotetraploidization. The authors inferred a genome-wide bias against *A. thaliana* gene expression, consistent with overall plant phenotype (the allotetraploids looked more like *A. arenosa* than *A. thaliana*) and the direction of previously reported nucleolar dominance (126). Similarly, in cotton, comparison of two synthetic allopolyploids with their parents (*G. arboreum* x *G. thurberi* and *G. arboreum* × *G. bickii*)

revealed substantial dominance of the maternal, *G. arboreum* expression phenotype. Both studies highlighted an important and emergent property of allopolyploidy: genes do not behave as simple additive combinations of the parental genomes. The genomic dominance appears to be quite common, insomuch as it has been validated by genome-wide expression analyses also in synthetic polyploids of wheat, *Senecio*, *Brassica*, and *Spartina* (127-131).

This lack of additivity in gene expression levels raises the fundamental question of what is responsible for nonadditivity in gene expression. With the aim to separate the effects of hybridization and genome duplication Hegarty *et al*., (2005, 2006) compared the transcriptomic alteration between the synthetic triploid hybrid *Senecio* x *Baxteri* with its progentiors, the diploid *S. squalidus* and the tetraploid *S. vulgaris*. The results suggested that both hybridization and genome doubling affected the transcriptome but that these effects were distinct from each other. In particular, it was suggested that hybridization induced transcriptome shock, whereas genome doubling has a "calming effect". On the same conclusions got Flagel *et al*., (2008)who compared, in cotton (*Gossypium* spp.) the gene expression changes between a synthetically produced F1 hybrid and a natural allopolyploid of the same original parentage. The major expression changes were observed in the natural allopolyploid (compared with its progenitor species) likely due to the mutations and the sub- or neo-functionalized duplicate genes that had assumed altered expression patterns over evolutionary time in the polyploid. Similar findings were reported for the proteome of synthesized Brassica allotetraploids and diploid hybrids, where Albertin *et al*., (2006) found that ~89% of protein expression differences could be attributed to hybridization rather than to polyploidization.

Such results suggested that autopolyploidy might produce less dramatic expression changes than allopolyploidy. However, few data are available concerning the regulation of gene expression in autopolyploid plants so far. A study examining the expression of 18 maize (*Zea mays*) genes in haploid, diploid, triploid and tetraploid plants showed that for most genes the transcript level per cell was directly proportional to the structural gene dosage. A few genes, however, displayed negative or positive (greater than expected) correlation with ploidy (135). Similarly, in synthetic autotetraploids of *Isatis indigotica*, Lu *et al*., (2006) have demonstrated highlighted differences in gene expression relative to the diploid progenitors (up- and down-regulation) as well as gain/loss of expression. Interestingly, inbred maize autotetraploids displayed stronger inbreeding depression than diploids (137). A transcriptome follow-up study in an inbred maize ploidy series found frequent but low-level changes (138). Comparison of diploid to triploid and tetraploid maize revealed that the most common trend was a decrease in expression, mirroring the decrease in size

observed in maize autopolyploids. In *Citrus* transcriptome expression changes between autotetraploid and diploid were studied using a microarray containing 21 081 genes. The number ofgenes differentially expressed was less than 1% and the maximum rate of gene expressionchange within a 2-fold range, by contrast the authors found large alterations in anatomy and physiology in response to gene doubling (83).

In *Solanum phureja* autotetraploids, 10% of the ~ 9000 genes analyzed using microarray displayed changes in expression level within a twofold range, with no "ploidy-regulated" genes identified (81). Finally, Pignatta *et al*., (2010) carried out a microarray analysis in three equivalent lineages of matched diploids and autotetraploids of *Arabidopsis thaliana*. Additionally, they compared the expression levels of GFP transgenes driven by endogenous enhancer elements (enhancer traps) in diploids and autotetraploid of 16 transgenic lines. Unlike the expectation that true ploidy-dependent changes should occur in independently derived autopolyploid lineages, the microarray analysis detected few changes associated with polyploidization. By contrast, the enhancer-trap analysis revealed that multiple loci in the genome are potentially subject to ploidy regulation.

Overall what are the mechanisms causing gene expression changes following polyploidization are still unclear. Some author speculated that altered gene expression patterns might arise from regulatory perturbations linked to altered cell volumes and stoichiometric of transcriptional factors in complex ways to generate widespread changes in gene expression (123).

Knowledge about gene expression regulation in wild potato species and in particular in their induced polyploids is of fundamental importance for breeding. It is well known how *S. commersonii* is resistant to abiotic stresses as cold treatment (140) and *S. bulbocastanum* is resistant to biotic stress as well as late blight (141). The information generated on plant genome and gene structure has drastically changed strategies in plant research and is starting to impact plant breeding (142). We are in a situation where we know the genes sequence but not their functions. Functional genomics is a way to decipher gene's role in complex trait like abiotic stress response (143). The regulation of expression of MYB transcriptional factor familygenes (TFs) lead to the accumulation of secondary metabolities as anthocyanins. They play a crucial role in the resistance to abiotic stresses (144, 145).

The MYB superfamily constitutes one of the most abundant groups of transcription factors described in plants. Among the others, MYB TF have been implicated in a diversity of plant-specific processes, including the determination of cell shape, regulation of trichome length, density,

developmental processes and defence responses(144, 145). The protein contains three domains, an N-terminal DNA-binding domain, a central transcriptional activation domain and a C-terminal domain involved in transcriptional repression. However, while the functions of some plant MYB genes are relatively well understood they are, at present, quite distinct from their animal counterparts and their functions appear to be highly diverse and remain rather unclear. Therefore the perturbation of MYB TF expression after polyploidization, may have significant consequences. MYB proteins are involved in the regulation of numerous physiological process and are notoriously know to regulate the phenylpropanoid pathway. Among them, a member of R2R3 MYB TFs gene family, named *an1* in potato, seems to be a key role in the anthocyanin accumulation (146).

In conclusion, if we assume that the differential expression of the genome in all polyploids compared with their parents has a role in the evolutionary trajectory of the neopolyploids, then it is tempting to speculate that the magnitude of transcriptomic changes in the neopolyploid can have a deciding role over which polyploids will persevere in evolution. Moreover it appears that different species, and even different genotypes of the same species, can vary significantly in their responses to hybridization, genome doubling or the combination of the two.

1.5. Aims of the thesis

So far many studies carried out on allopolyploids have contributed very much to gain new insights into polyploidy-related processes. However, allopolyploidy inherently changes two parameters at once and does not allow separating the effects of chromosome number change versus hybrid formation. Given that, autopolyploids are the ideal plant system in which the changes are associated only with ploidy. In particular, potato and its wild relatives represent an excellent model system since 1) the cultivated potato, *Solanum tuberosum*, has been defined as a classic autopolyploid on the basis of its tetrasomic inheritance; 2) the *Solanum* species are particularly amenable to ploidy manipulations; 3) potato breeding often fall back upon polyploidization of incongruent wild relatives to access their largely untapped diversity and to overcome the sexual barriers hampering the introgression of useful traits from wild germplasm to cultivated genepool.

Since many years at the University of Naples a wide multidisciplinary program is being carried out to understand events underlying early polyploid formation in potato. In particular, two diploid wild *Solanum* species, *S. commersonii* and *S. bulbocastanum*, were chosen to accomplish this goal. They have several useful traits for the genetic improvement of the cultivated potato. In particular, *S. commersonii* is a noteworthy species owing to its high resistance to low temperatures

as well as its cold acclimation capacity(140)together with a high specific gravity of tubers (147). *S. bulbocastanum* is highly resistant to all known races of *Phytophthora infestans*, even under intense disease pressure (141). We previously produced synthetic tetraploids from each wild species by oryzalin treatment, an antimitotic agent. The differences between *S. commersonii* and *S. bulbocastanum* synthetic tetraploids and the diploid parents they derived from were studied at genetic, epigenetic, morpho-anatomical (82), physiological and biochemical (122) levels. Information gained has been used for both basic and applied research.

To understand the complex dynamics of polyploidization and to expand our knowledge of such phenomenon in *Solanum* species, and more in general in crops, a genome-wide analyses at transcriptomic and metabolomic levels was carried out in the search presented here. In particular, the following objectives were pursued:

- To understand the gene expression regulation in synthetic autopolyploids of *S. commersonii* and *S. bulbocastanum* their diploid parents. Towards this goal, a high-throughput transcriptome analysis was carried out using a custom microarray able to detect all the expressed genes. We studied complex transcriptome changes associated to genome doubling and then we confirmed the whole transcriptome analyses by independent qPCR experiments.
- To study the effects of chromosome doubling on the metabolome of *S. commersonii* and *S. bulbocastanum* autotetraploids. A "global" profiling approach has been carried out by a liquid chromatography-electrospray ionization/multi-stage mass spectrometry (LC-ESI(+)-MS) analysis on 170 targeted molecules involved in primary (aminoacids, sugars, organic acids etc) and secondary (flavonoids, alkaloids etc) metabolism.
- To enhance own knowledge about the changes in anthocyanin pathway in tetraploids versus ofdiploids of *S. commersonii* and *S. bulbocastanum*. Total flavonoids resulted drastically accumulated in tetraploids of both species in our metabolomic analyses. The role of the R2R3 MYB transcription factor *an1*in the regulation of the anthocyanin synthesis pathwaty is well known. We identified the gene structure of a R2R3 MYB transcription factor *an1* in both diploid parental genotypes. The gene expression study of *an1* and two structural enzymes of the anthocyanin biosynthetic pathways (*ans* and *dfr*) was performed on all the 4x and 2x genotypes using qPCR assays.

2. Materials and Methods

2.1. Plant material

A clone (named CMM1T) belonging to accession PI 243503 of diploid (2n=2x=24) *S. commersonii* and a clone (named BLB1C) belonging to accession PI 275190 of diploid (2n=2x=24) *S bulbocastanum* species were used in this study. Both PIs were kindly provided by Dr. J. Bamberg of the University of Wisconsin (Madison, USA). They were maintained and propagated in vitro on MS medium (150), including salts, vitamins, sucrose (30 g/l) and microagar (9 g/l) and adjusted to pH 5.8. The cultures were maintained in a growth chamber at 25 °C ± 2 under a 16/18 (light/dark) photoperiod at 125 μEm-2 s-1 irradiance provided by a cool white fluorescent tube (Philips). Two subcultures on the growing medium at 3-week intervals were necessary to obtain a sufficient number of fully developed plantlets for somatic chromosome doubling. This study also included four tetraploids clones of *S. commersonii* (CMM15, CMM24, CMM27 and CMM30) and four tetraploids of S. *bulbocastanum* (BLB10, BLB22, BLB25 and BLB26), obtained as previously reported (151). All plant materials were maintained and propagated in vitro on MS medium (150). For the analysis, five plants of each genotype were grown together in the same greenhouse conditions at 20-25 °C under natural light. Plants were irrigated three times a week depending on the evaporative demand of environment.

2.2. Nucleic acids extraction

Fully developed young leaves were collected from individual plants on the same date. To perform DNA extraction, a leaf pool was created by mixing equal amounts of leaf powered from five different plants. DNA was isolated using the DNeasy Plant mini kit (Qiagen) following the manufacturer's instructions. DNA quality and integrity was checked by gel electrophoresis and spectrophotometric assay.

To perform RNA extraction, a leaf pool was created by mixing equal amounts of leaf from five different plants. One ml of TRIZOL reagent (Life Technologies) were added per 100 mg of homogenize samples obtained by homogenizer (TissueLyser by Qiagen). Sequential RNA extraction was performed following Life Technologies instructions manual. The quality and the integrity of RNA were checked after electrophoresis of 1 μg of RNA samples on 1% agarose gel stained with SYBR® Safe (Life Technologies). The amounts and the purity degree of RNA extracted were estimated using NanoDrop spectrophotometer (Thermo Fisher Scientific).

2.3. Microarray analyses

2.3.1. Chip design

The 90k microarray chip has been synthesized on Combimatrix platform in the frame of a collaboration with prof. Massimo Delledonne (Department of Biotechnology, University of Verona (Italy). The chip contains 90k siliceous electrodes supporting up to 30000 in situ synthesized DNA probes with 3 replications. Each probe consists of 35-mer oligonucleotides designed to be specific for a different unique transcripts. As negative control nine bacterial oligonucleotide sequences provided by Combimatrix were utilized. The Custom Array™ 90K was designed with three replicates of each probe randomly distributed across the array to allow measurement of the variability within the array.

For the chip design the following workflow was applied:

1. Download of all *S. tuberosum* TC sequences (23453) from the SolEST database (http://biosrv.cab.unina.it/solestdb/)
2. Analyses of all *S. tuberosum* TC sequences orientation
3. Download of all *S. tuberosum* mRNA complete sequences (46) from the SolEST database
4. Probe identification in the 3'-UTR region of all these sequences
5. Download of all *Medicago sativa* probe sequences from Combimatrix microarray design by Prof. Massimo Delledonne
6. Homology analyses between potato and *Medicago* probes
7. Chip production with potato specific probes plus *Medicago* probes with a high potato homology

We choose to design oligo probes for the identification of the expressed sequence in the 3'-UTR region of genes because these region is the more specific region of the sequence conserved through the evolution. To reach this goal we must know the orientation of all TC potato sequence present in the SolEST database. This was a problem because the TC sequences came from the assembly of EST sequence and normally ESTs have no orientation. So we analyzed orientation of all TC potato sequences. A blast alignment of TC sequences against the protein sequence database was made using a software of the BLAST suite (BLASTx). The results is that only 15234 sequences out of about 24000 TC sequences have a positive blast alignment thanks to which it was possible to assigne an orientation to these sequences. So 13207 sequences have a forward orientation instead 2027 had a reverse orientation. About 9000 TC sequence had a negative blast result so it's

impossible for us to say where was the 3'-UTR region of these sequences. To overcome this problem we chose 6000 TC sequences among the 9000 as the more expressed TC sequences (assembled with the highest number of EST sequences). Finally OligoArray 2.1 was the software used to design the oligo probes. The sequences used to produce oligo probes were:

1. 15234 TC sequences with a positive BLAST
a. 13207 TC FORWARD orientation
b. 2027 TC REVERSE orientation
2. 9000 TC with a negative BLAST
a. 6000 TC FORWARD orientation
b. 6000 TC REVERSE orientation

Each probe identified a sequence of about 45 oligonucleotides in the 3'-UTR region of these sequences. The total number of potato probes present on the chip was 27234. The last step for the design of the microarray was homology analyses between the probe set of *Medicago sativa* present in the specific Combimatrix microarray and all potato TC sequences. Six hundred seven *M. sativa* probes positively matched with potato sequences, making them usefully to identified potato expressed sequences. So we also added these 607 probes to the probe set of potato microarray, totaling 27841 probes.

2.3.2. aRNA synthesis and marking

The synthesized and labeled antisense-RNA (aRNA) were performed using the Kreatech's kit RNA ampULSe: Amplification and Labeling Kit for CombiMatrix arrays with Cy5 dye. In this kit two successful technologies are fused. First of all aRNA amplification guarantees a relative abundance of aRNA fragments. This first step is followed by fluorescent labeling with KREATECH's proprietary Universal Linkage System (ULS) technology that use unmodified nucleotides to label aRNA resulting in higher aRNA yields and a better representation of the transcriptome by avoiding the strain on the enzyme to incorporate modified nucleotides. Steps are here reported.

1. Starting material is eukaryotic Total RNA
2. Reverse transcription to synthesize first strand cDNA using T7 Oligo(dT) primers
3. Second strand synthesis
4. Clean up cDNA and use as template for aRNA synthesis
5. In vitro transcription (IVT) to synthesize aRNA

6. Clean up aRNA
7. ULS labeling of aRNA
8. aRNA fragmentation
9. Hybridization

Starting from 1 µg of total RNA reverse transcription were performed to synthesize first strand cDNA using T7 Oligo(dT) primers following manufacturer's instructions. Yield and quality of amplified aRNA were verified after IVT and purification steps using NanoDrop spectrophotometer (Thermo Fisher Scientific). For all RNAs OD260/280 were >1,9 and OD260/230 were >2,1. Five µg of aRNA were labeled per reaction following manufacturer's instructions and after ULS labeling steps labeling density were checked. To measure degree of labeling (DoL) NanoDrop spectrophotometer (Thermo Fisher Scientific) was used to read absorbance at A260 and A650 for determining the DOL of Cy5-ULS labeled aRNA and then these values were put in the following formulas:

$$\text{ng / µL} = \frac{A_{260} * \text{dilution factor} * 40}{\text{cuvet length (in cm)}}$$

$$\text{pmol / µL} = \frac{A_{\text{dye at max}} * \text{dilution factor}}{\text{cuvet length} * \varepsilon_{\text{dye}} * 10^{-6}}$$

εdye Cy3 Reagent = 150,000
εdye Cy5 Reagent = 250,000

Degree of labeling (amount of dyes per 100 nucleotides)

$$\text{Labeling \%} = \frac{340 * \text{pmol}_{\text{dye}}}{\text{ng}_{\text{nucleic acid}} * 1000} \times 100\%$$

Optimal DoL value for labeled aRNA is between 1,0 and 3,6. DoL values lower than 1.0 may not produce enough signal, whereas DoL values higher than 3.6 might cause either high background levels or quenching of signal.

2.3.3. Hybridization and imaging

Labelled aRNA was hybridized on the PotatoArray. The microarray slides were pre-hybridized, onto the rotisserie in the hybridization oven, for 30 min at 45°C in the Pre-hybridization Solution (SOL. A). The microarray was then hybridized filling the hybridization chamber with the Hybridization Solution (SOL. B), and incubated for 14 h at 45°C in the hybridization oven. After hybridization the slides were iteratively washed with different Wash Solution (SOL. C, D, E) until the final wash with PBST Wash Solution (SOL. F), and incubated at room temperature for 1 min.

Three hybridization replica per each genotype (CMM1T, 15, 24, 27, 30; BLB1C, 10, 22, 25, 26) per three biological replica were performed for a total of 30 hybridization on 4 different arrays. Slides were then scanned using a GenePix® 4400A Microarray Scanners (Molecular Devices) with a minimum resolution of 5µm and an adjustable focus and an acquisition software (ScanArray Express Microarray Analysis System Version 4.0) according to the manufacturer's instructions. For more details see CustomArray® 90K Microarray - Hybridization and Imaging Protocol (PTL020) available at website http://www.combimatrix.com/support_docs.htm. Five µg of labelled target sample were used per microarray to hybridize samples to the CombiMatrix CustomArray™ 90K microarray. Hybridization was performed following manufacturer's instructions. The hybridized CustomArray™ 90K were stripped of targets using the CombiMatrix Stripping Kit following manufacturer's instructions, and then re-used according to the hybridization protocol described in manufacturer's instructions. The composition of the solution used is reported:

PRE-HYBRIDIZATION SOLUTION (SOL. A)	VOLUME FOR 120 µL
2X Hyb Solution stock[a]	60 µl
Nuclease-free water	41 µl
50X Denhart's solution	12 µl
Salmon sperm DNA (10 mg/ml)	1 µl
1% SDS	6 µl

[A]2X HYB SOLUTION STOCK	VOLUME FOR 10 ML
20X SSPE	6 ml
10% Tween-20	100 µl
0.5 M EDTA	560 µl
Nuclease-free water to	3,34 ml

HYBRIDIZATION SOLUTION (SOL. B)	VOLUME FOR 120 µL
2X Hyb Solution stock[a]	60 µl
DI Formamide	30 µl
Labeled RNA (5µg)	10 µl
Salmon sperm DNA (10 mg/ml)	1 µl
1% SDS	5 µl
Nuclease-free water to	120 µl

6X SSPET WASH SOLUTION (SOL. C)	VOLUME FOR 10 ML
20X SSPE	3 ml
10% Tween-20	50 µl
Nuclease-free water to	6,95 ml

3X SSPET WASH SOLUTION (SOL. D)	VOLUME FOR 10 ML
20X SSPE	1,5 ml
10% Tween-20	50 µl
Nuclease-free water to	8,45 ml

0,5X SSPET WASH SOLUTION (SOL. E)	VOLUME FOR 10 ML
20X SSPE	50 µl
10% Tween-20	9,7 ml
Nuclease-free water to	250 ml

PBST WASH SOLUTION (SOL. F)	VOLUME FOR 10 ML
10X PBS	2 ml
10% Tween-20	100 µl
Nuclease-free water to	7,9 ml

2.3.4. Data analysis

After scanning microarray TIF images were processed to generate numerical data using the Molecular Devices GENEPIX PRO V.7 copy@right 2012. Signal probe medians and standard deviations were further imported in SPSS software. 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. Following data normalization and quality control all values were log transformed (log base 2). Finally, probe signals having a variability coefficient higher than 0.5, as well as spikes and factory probes, were filtered out. Also, probe showing the 10% uppermost and the 10% lowest signal intensity were deleted. To visually check normalization Box plot were provided. The median of the data is represented by the line in the center of the rectangular box; the two ends of the rectangles represent the upper quartile and the lower quartile. The other two values always shown are the maximum and minimum value of the data set. Expression data were normalized using the SPSS software package and probes with significant change (P-value <0.01) in their hybridization signals were identified by running a two-factor ANOVA module included in the TIGR Multiple Experiment Viewer Software v4.0 (http://www.tm4.org/). Co-regulations were investigated by Pearson metric. In particular, a Mack-Skillings (two factor designs, NxK,) nonparametric test was used for 2X vs 4X and for CMM Vs BLB analyses. A Willcoxon, Mann-Whitny test (one factor, two experimental groups) nonparametric test was used for CMM1T Vs CMM 4X and for BLB1C Vs BLB 4X. Finally a t-test with 100 permutations and a critical p-value <0.01 was used for the single analyses:

1. CMM1T Vs CMM15
2. CMM1T Vs CMM24
3. CMM1T Vs CMM27
4. CMM1T Vs CMM30
5. BLB1C Vs BLB10
6. BLB1C Vs BLB22
7. BLB1C Vs BLB25
8. BLB1C Vs BLB26

Blast2GO (http://blast2go.bioinfo.cipf.es/ - Conesa et al 2005) was used to provide automatic and high-throughput annotation, gene ontology mapping, and categorization of TCs showing differential transcription signal. An expectation value < E 10-10 was used. InterPro database was

used to assign protein domain and KEGG database to map the genes on metabolic pathways. Weighted gene co-expression network analysis was performed as described previously (152-153) using an R-script modified for this analysis and a network threshold cut off of 0.01 Cytoscape Ver. 2.6.3 (154) and NetworkAnalyzer ver. 1.0 (155) were used to visualize the global gene network and calculate topology parameters. Candidate network drivers were identified by ranking nodes based on their degree index and also by comparing to index values determined for closeness, radiality and eccentricity.

2.4. Real-time PCR analysis

The cDNA was synthesized using SuperScript™ III Kit (Life Technologies). In a 0,2 ml RNase-free tube, was added:

COMPONENT	VOLUME
Total mRNA	1 µg
T7-Oligo(dT) Primer	1 µl
DEPC-treated water to	10 µl

The mix was incubated at 70°C for 10 min, and then placed on ice for 1 min, and centrifuged briefly to collect the contents. After that was added:

COMPONENT	VOLUME
5X First-Strand buffer	4 µl
0.1 M DTT	2 µl
10 mM dNTP Mix	1 µl
RNaseOUT (40 U/µl)	1 µl
SuperSript III RT (200 U/µl)	2 µl
Total Reaction Volume	20 µl

The tube was centrifuged briefly and then incubated at 46°C for 2 min, following an incubation at 70°C for 10 min.

Real Time PCR assays were performed using the Qiagen 2x QuantiFast SYBR Green PCR Master Mix (Qiagen). Each gene expression assay contained, together with the forward and reverse primers, master mix (HotStarTaq Plus DNA Polymerase, QuantiFast SYBR Green PCR Buffer, SYBR Green, and ROX passive reference dye) and sscDNA target as follow:

COMPONENT	VOLUME/FINAL CONCENTRATION
2x Master Mix	6 µl/1X
Primer FW [100µM]	1,8 µl/300 nM
Primer FW [100µM]	1,8 µl/300 nM
Template cDNA [1:10]	1 µl
Nuclease-free water to	1,4 µl
Total Reaction Volume	12 µl

For quantitative real-time evaluation of mRNA expression levels of potato genes, sequence-specific primers were designed used Primer3Plus web site (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/). The real-time PCR were run in triplicate on 96-well fast PCR plates (SARSTEDT) with an equal loading of 1:10 dilution of cDNA/well. The following table outlined the program used on a ABI PRISM 7900HT Instrument (Life Science) run in fast mode.

STEP	TIME	TEMPERATURE
PCR initial activation step	5 min	
TWO-STEP CYCLING		
Denaturation		
Combined	10 s	95 °C
annealing/extension	30 s	60 °C
Number of cycles	40	

Data acquisition were 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 analyzed using the ABI PRISM 7900HT Sequence Detection System Version 2.1 (SDS 2.1). Relative quantification was measured using the Comparative Ct (Threshold Cycle) Method. The endogenous control gene (Aprt) were used to normalize for the cDNA of each sample. The ΔCt for each of the triplicate (Ct of the target gene minus Ct of the endogenous control) and then the average ΔCt of the triplicate were calculated. When a single value within each duplicate differed substantially from the other value the skewed value was excluded from the analysis. To calculate the ΔΔCt, the ΔCt of each sample was subtracted to the chosen reference sample. The relative quantity was then calculated following the equation $RQ=2^{-\Delta\Delta CT}$ where 2 represents doubling of the amount of the product of amplification after each PCR cycle.

2.5. Chemical analyses

The metabolome analyses were done in collaboration with Dr. Diretto G. and Giuliano G. (156) in the laboratory of Italian National Agency for New Technologies, Energy, and Sustainable Development, Casaccia Research Center, Rome 00123, Italy.

2.5.1. Spectrophotometric analyses of total pigments

The protocol of Porra *et al*., (1989) was used for spectrophotometric determination of leaf chlorophylls and carotenoids. Briefly, 2-3 mg have been extracted with 80% acetone, by mixing the samples for 10' at 30hz using a Mixer Mill 300 (Qiagen) and OD446.6, OD446.8, OD663.6 and OD470 have been, subsequently, measured. For anthocyanin measurement, freeze-dried leaf tissue powder (5 mg) were extracted overnight in 0.5 mL of methanol containing 50 mM HCl. After addition of 0.375 mL of water and 0.5 mL of chloroform, the mixture was vortexed, centrifuged, and the OD535 of the upper phase determined.

2.5.2. LC-ESI(+)-MS analysis of leaf polar metabolome

Five mg of ground lyophilized leaf powder were extracted with 0.75 mL 75% methanol - 0,1 % Formic Acid spiked Cold (with 10 μg/ml Formononetin); after vortexing for 30 seconds, samples were shaked for 15 min at 15 hz using a Mixer Mill 300 (Qiagen) and kept at RT for 5 min (twice). After centrifugation for 15 min at 20,000 g at 4°C, 0,6 mL of supernatants were removed and transfer to HPLC tubes. For each genotype, at least five independent extractions were performed. LC-MS analyses were carried out using a Discovery LTQ-Orbitrap mass spectrometry system (Thermo Fischer Scientific) operating in positive electrospray ionization (ESI), coupled to an Accela U-HPLC system (Thermo Fischer Scientific, Waltham, MA). LC separations were performed using a C18 Luna column (150 x 2.0 mm, 3 μm) purchased from Phenomenex. The mobile phases used were water – 0.1% Formic Acid (A), acetonitrile – 0.1% Formic Acid (B). The gradient was: 95% A : 5% B for one minute, followed by a linear gradient to 25% A : 75% B over 40 minutes; LC conditions were kept for 2 more minutes, before going back the initial LC conditions in 18 minutes. Ten μl of each sample were injected and a flow of 0.2 mL were carried out during the whole LC runs. Detection was performed continuously from 230 to 800 nm with an online Accela Surveyor photodiode array detector (PDA, Thermo Fischer Scientific, Waltham, MA). All solvents used were LC-MS grade quality (CHROMASOLV® from Sigma-Aldrich). Metabolites were relatively quantified on the basis of the internal standard amounts. ESI-MS ionization were as described previously (159). Identification was performed by 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 on Pubchem database (http://pubchem.ncbi.nlm.nih.gov/) for monoisotopic masses identification, or on Metabolomics Fiehn Lab Mass Spectrometry Adduct Calculator (http://fiehnlab.ucdavis.edu/staff/kind/Metabolomics/MS-Adduct-Calculator/) in case of adduct ion detection.

The ANOVA analyses was applied to the absolute metabolite quantifications and pairwise comparisons was used in Tukey's HSD test running on XLSTAT 2013 software (http://www.xlstat.com).

3. Results

3.1. Whole gene expression analysis

3.1.1 Raw microarray data statistical analyses

Comparative transcriptome analysis of the two wild potato species *S. commersonii* and *S. bulbocastanum* were performed in three independent series of experiments. We used the diploid (2n=2x=24) parental genotypes CMM1T and BLB1C of *S. commersonii* and *S. bulbocastanum* respectively, and four different tetraploid (2n=4x=48) genotypes for each species, namely CMM15, CMM24, CMM27, CMM30 for *S. commersonii* and BLB10, BLB22, BLB25, BLB26 for *S. bulbocastanum*. After hybridization, scanning and data extraction, microarray data were statistically analysed. For each experiments median and standard deviation were calculated and all the data were normalized with a common central mean as shown in Figure 5. The box plots indicated that there were no anomalies or outlines samples in the dataset.

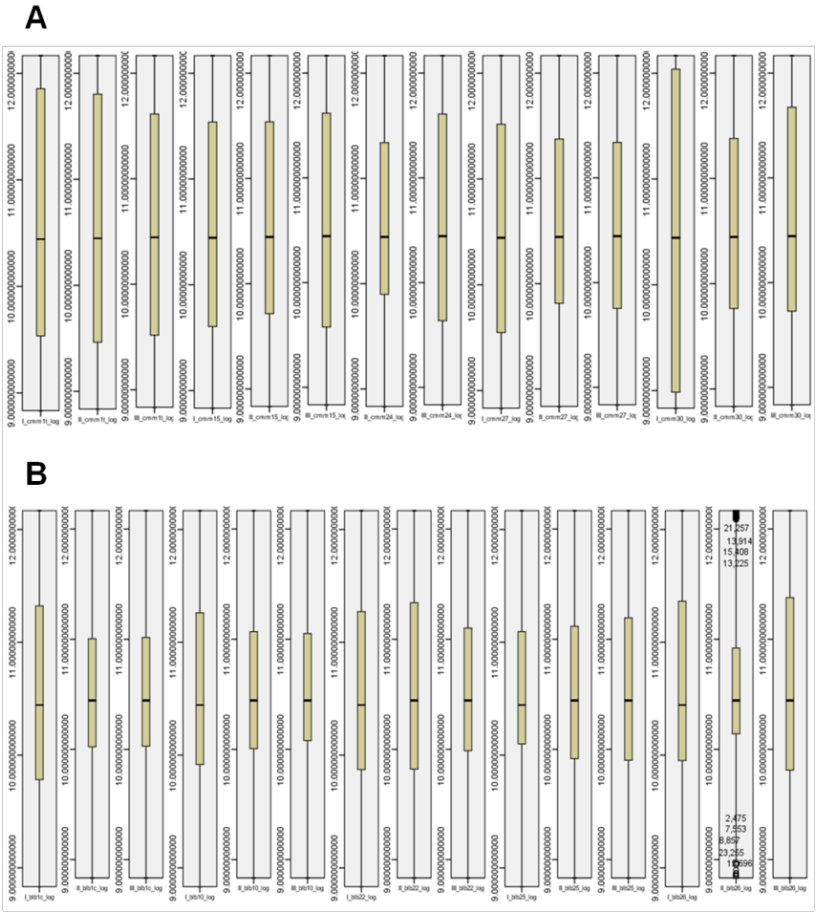

Figure 5. Normalization box plot: Box plot of all 29 samples using a common central mean. Section A: *S. commersonii* samples (triplicates for CMM1T, CMM15, CMM27 and CMM30, duplicates for CMM24). Section B: *S. bulbocastanum* samples (triplicates for BLB1C, BLB10, BLB22, BLB25 and BLB26).

To study microarray data several statistical test were chosen to analyse different relationships existing among samples or groups. For each analysis the up- and down-regulated genes were shown (Fig. 6). The Mack-Skillings nonparametric test is conceived to find differential expressed genes between two factor designs (NxK), corresponding, in our case, to the ploidy level (2x and 4x) and the species (*S. commersonii* and *S. bulbocastanum*). Analysing the fold change of differentially expressed genes, the Mack-Skillings' test found 1.602 and 7.853 genes, between tetraploids and diploids (4X Vs. 2X) and between *S. commersonii* and *S. bulbocastanum* (CMM Vs. BLB), respectively. Among the 1.602 genes, 1.238 were up-regulated in the tetraploid genotypes and 364 were instead down-regulated. Examining the 7.853 genes differentially expressed between *S. commersonii* and *S. bulbocastanum*, 5.649 were up-regulated in *S. commersonii* and 2.204 in *S. bulbocastanum*.

The Willcoxon, Mann-Whitny nonparametric test analyses the statistical differences between two groups dissimilar for one factor. In our experiment, the two groups were represented, within each species, by the diploid parent and the corresponding tetraploids, while the divergent factor was the ploidy level. The total number of genes whose expression changed in CMM4X Vs. CMM1T comparison was 1.742, whereas 969 differentially expressed genes were found in BLB4X Vs. BLB1C. In *S. commersonii*, 783 (45%) out of 1.742 genes were up-regulated in the tetraploids, while 959 (55%) were down-regulated. In *S. bulbocastanum*, 93.7% (908) of the total differentially expressed genes (969) were up-regulated and only 61 genes were down-regulated in the diploid parent (BLB1C) in comparison to the its tetraploids.

A t-test with 100 permutations and a critical p-value of 0.01 were used to relate the transcriptome profile of each tetraploid with the corresponding diploid parent. As for *S. commersonii*, the tetraploids CMM15 and CMM30 displayed a higher number of up-regulated genes (1.626 and 1.429, respectively) compared with the CMM1T, then down-regulated genes (1.194 and 1.282, respectively). By contrast, CMM24 and CMM27 behaved oppositely (1.575 and 882 up-regulated, and 2.150 and 2.062 down-regulated, respectively). As for *S. bulbocastanum*, analyses showed general trends of up-regulation in synthetic tetraploids in respect to BLB1C diploid parent. In particular, the frequency of up-regulated genes varied from 91.3% (2.385 out of 2.612) in BLB10 to 24.3% (1.348 out of 1.778).

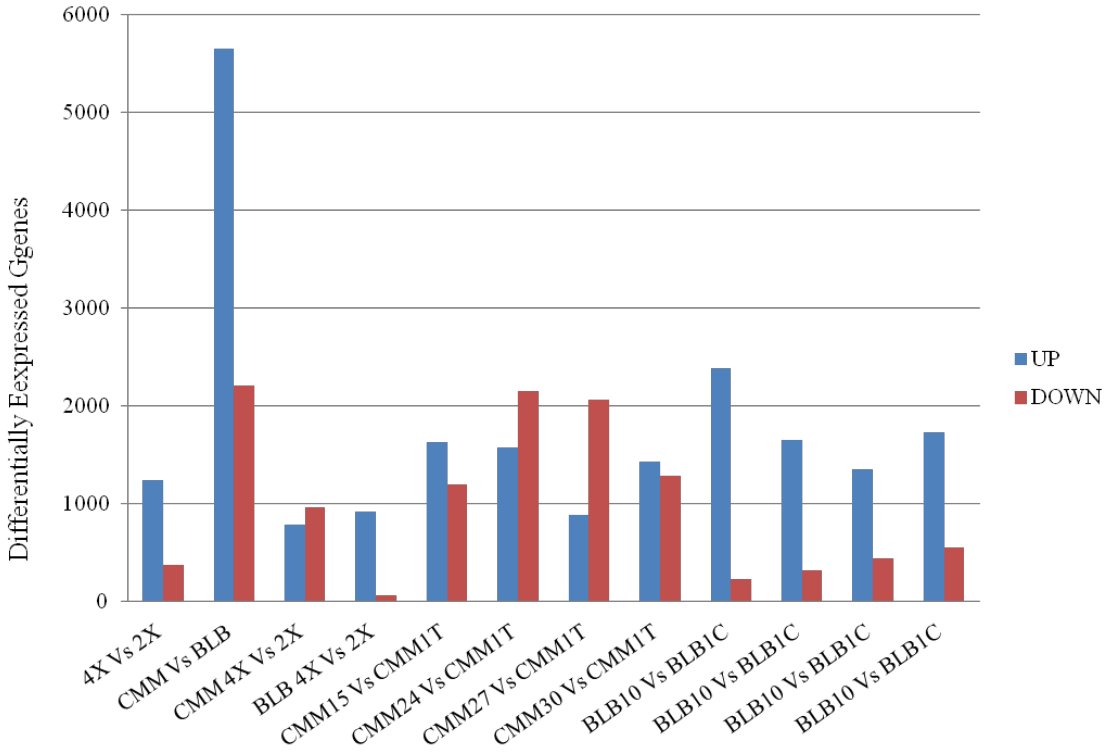

Figure 6. Bar chart describing the number of differentially expressed genes in all comparisons studied. The up- and down-regulated genes were colored in blue and red, respectively.

In Table 2 and 3were listed the differentially expressed genes between the diploid parent and the tetraploids they derived from of *S. commersonii* and *S. bulbocastanum*, respectively. A fold change cut-off of 1.5 was applied.

As for *S. commersonii*, the fold change varied from + 1.83 (probe id_5872) to -2.24 (probe id_24276), with only 3 genes up-regulated and 48 down-regulated (Tab. 2). As for *S. bulbocastanum*, fold change ranged from +2.39 (probe id_7971) to -2.84 (probe id_19706) with 16 genes up-regulated in tetraploids and 4 down-regulated (Tab. 3).

Table 2. List of differentially expressed genes (cutoff threshold 1.5fold change) in *S. commersonii* tetraploids versus diploids CMM1T (CMM4x Vs. CMM2x).

Probe	Cluster ID	Annotation	FC
id_5872	SOLTU007288	surfeit locus protein 6	1.83
id_203	SOLTU000291	gpi-anchoredproteinlorelei	1.73
id_15126	SOLTU019444	beta-amylase	1.60
id_12782	SOLTU015936	---NA---	-1.51
id_11287	SOLTU013701	Wrkytranscriptionfactor	-1.52
id_13206	SOLTU016617	#N/D	-1.53
id_21688	SOLTU032446	hypotheticalprotein SDM1_3t00007 [*Solanumdemissum*]	-1.53
id_23695	SOLTU037741	flowering locus d	-1.54
id_14111	SOLTU018029	Coppertransporter	-1.54
id_7135	SOLTU008354	#N/D	-1.55
id_6443	SOLTU007764	#N/D	-1.57
id_22667	SOLTU034436	glucose-methanol-choline oxidoreductase family protein	-1.58
id_20010	TC118593	#N/D	-1.58
id_24319	SOLTU004773	flavonolsulfotransferase-like	-1.58
id_5124	SOLTU006621	Threoninealdolase	-1.59
id_18485	SOLTU027721	---NA---	-1.60
id_16195	SOLTU021222	Protein	-1.60
id_16114	SOLTU020977	zincknuckle family protein	-1.60
id_11454	SOLTU013931	#N/D	-1.61
id_8743	SOLTU009681	polygalacturonase-1 non-catalyticsubunit beta	-1.61
id_2593	SOLTU004130	s-acyltransferase tip1	-1.61
id_7242	SOLTU008447	40s ribosomalprotein s28	-1.63
id_11609	SOLTU014150	2-cys peroxiredoxin-likeprotein	-1.65
id_5861	SOLTU007280	Protein	-1.65
id_19174	SOLTU028537	Phytoenesynthase	-1.67
id_991	SOLTU001603	oas-tl4 cysteinesynthase	-1.68
id_7558	SOLTU008682	dual specificity protein phosphatase diacylglycerol catalytic region	-1.69
id_5945	SOLTU007358	ethanolamine-phosphatecytidylyltransferase	-1.69
id_19609	SOLTU028980	snf2 family dna-dependent atpase	-1.71
id_22482	SOLTU034011	rna-dependentrnapolymerase 2	-1.72
id_2978	SOLTU004647	Histonedeacetylase	-1.74
id_3766	SOLTU005531	dual specificity protein phosphatase family protein	-1.74
id_16215	SOLTU021260	set1 ash2 histone methyltransferase complex subunit ash2	-1.75
id_19086	SOLTU028456	senescence-associatedprotein 101	-1.75
id_6874	SOLTU008126	#N/D	-1.77
id_10977	SOLTU013265	#N/D	-1.78
id_14432	SOLTU018440	microfibrillar-associatedprotein	-1.79
id_7870	SOLTU008932	#N/D	-1.80
id_23131	SOLTU035814	Samt	-1.80
id_17187	SOLTU025291	#N/D	-1.83
id_3332	SOLTU004993	caseinkinase i isoform delta-like	-1.85
id_14364	SOLTU018345	speckle-typepoz	-1.88
id_6487	SOLTU007806	---NA---	-1.88
id_2225	SOLTU003661	gonadotropin beta chain	-1.92
id_11064	SOLTU013393	#N/D	-2.05

Probe	Cluster ID	Annotation	FC
id_24047	NP7275573	hypothetical protein MTR_8g075960 [Medicagotruncatula]	-2.05
id_6438	SOLTU007761	heat stress transcription factor a-6b-like	-2.11
id_9412	SOLTU010588	---NA---	-2.13
id_6843	SOLTU008103	---NA---	-2.18
id_20659	SOLTU030345	ubiquitin carboxyl-terminal hydrolase 12-like	-2.19
id_24276	SOLTU004179	histone h3	-2.24

Table 3. List of differentially expressed genes (cutoff threshold 1.5 fold change) in *S. bulbocastanum* tetraploids versus diploids BLB1C (BLB4x Vs. BLB2x).

Probe	Cluster ID	Annotation	FC
id_7971	SOLTU009015	50s ribosomalprotein l18	2.39
id_8575	SOLTU009528	---NA---	2.18
id_9540	SOLTU010796	trypsin domain-containing	2.06
id_9931	SOLTU011410	cytochrome p450	2.00
id_2125	SOLTU003586	probable wrky transcription factor 40-like	1.97
id_7525	SOLTU008656	lrr receptor-like serine threonine-protein kinase rpk2-like	1.88
id_5088	SOLTU006591	Protein	1.76
id_942	SOLTU001600	Protein	1.74
id_7981	SOLTU009022	f-box kelch-repeat protein at1g51550-like	1.66
id_7882	SOLTU008944	cytochrome c biogenesisfc	1.66
id_17022	SOLTU024812	2 calciumsensor	1.64
id_7240	SOLTU008447	---NA---	1.61
id_68	SOLTU000027	replication factor a related protein	1.60
id_8864	SOLTU009813	Protein	1.58
id_7646	SOLTU008754	signal recognition particle 54 kdachloroplastic-like	1.57
id_14381	SOLTU018374	zeatin o-glucosyltransferase-like	1.50
id_12733	SOLTU015866	Pullulanase	-1.54
id_5938	SOLTU007349	sucrose-phosphatesynthase	-1.88
id_24151	SOLTU010421	Protein	-2.48
id_19706	SOLTU029113	Diphenoloxidase	-2.84

Among the data, there were no ploidy up-regulated genes changing simultaneously both in *S. bulbocastanum* and *S. commersonii*. In other words, when we looked for differentially expressed genes in 4X Vs. 2X comparisons, no common genes between the two species were found. However, we found three genes always down-regulated in *S. commersonii* tetraploids in respect to CMM1T, and, conversely, always up-regulated in *S. bulbocastanum* tetraploids in respect to BLB1C (Tab. 4 and Fig. 7). In particular, the three common differentially expressed genes were a transcription factor with a protein domain named far1 (cluster ID: SOLTU027534), a membrane protein with the function of sodium bile acid co-transporter (cluster ID: SOLTU033001), and an uncharacterized protein (cluster ID: SOLTU010790).

Table4. Common differentially expressed genes in *S. commersonii* and *S. bulbocastanum* autotetraploids versus their respective diploids.

Probe	Cluster ID	Annotation	GOs
id_18364	SOLTU027534	protein far1-related sequence 3-like	F:binding; P:response to abiotic stimulus
id_21965	SOLTU033001	sodium bile acid cotransporter 7-like	C:plastid; F:transporter activity; C:membrane; P:transport
id_9535	SOLTU010790	PREDICTED: uncharacterized protein LOC100793388 [Glycine max]	C:plastid; C:membrane

In Figure 7 was shown, for both species, the differential expression values (as fold change) of the three genes, in each tetraploid compared to the respective diploid parent. The transcription factor *far1* (cluster IDs: SOLTU027534), showed a -0.67, -1.07, -0.8 and -0.84 fold changes compared to CMM1T, respectively in CMM15, CMM24, CMM27, CMM30. In *S. bulbocastanum* tetraploids,the *far1* gene was *vice-versa* up-regulated respect to BLB1C and the fold changes were of 0.59 in BLB10, 0.53 in BLB22, 0.7 in BLB25 and 0.33 in BLB26. The sodium bile acid co-transporter 7 and (cluster IDs: SOLTU033001) was the gene with a lowest average fold change value (-0.6) compared to CMM1T and with -0.7, -0.46, -0.79 and -0.59 fold changes in CMM15, CMM24, CMM27 and CMM30, respectively. The analyses of the expression level differences of sodium bile acid co-transporter 7 gene showed a fold change of 0.6 in BLB10, 0.9 in BLB22, 0.73 in BLB25 and 0.54 in BLB26 when compared with the expression of the diploid parent BLB1C. Finally, the uncharacterized protein SOLTU010790 (cluster IDs: SOLTU010790) was down-regulated in *S. commersonii*, and displayed a -0.85 fold change value in CMM15, -0.55 in CMM24, -0.7 in CMM27 and -0.8 in CMM30. In *S. bulbocastanum* it was differentially expressed in BLB10 with a fold change of 0.61, 0.52, 1.38 and 1.58 in BLB22, BLB25 and BLB26 respectively.

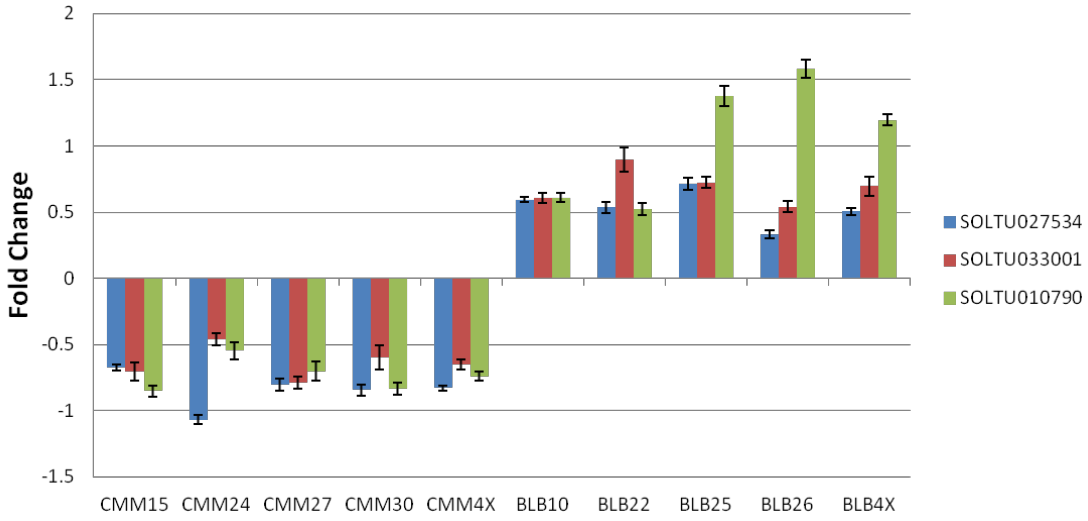

Figure 7. Common differentially expressed genes. Bar chart describing relative expression level of each tetraploid in comparison to the respective diploid parent, both for *S. commersonii* and *S. bulbocastanum*. For each species the mean values of relative expression level of all tetraploids were reported (CMM4X e BLB4X). Error bars represented the standard errors.

3.1.2 Bioinformatics analyses

Results of bioinformatics analyses of differentially expressed genes between the diploid parents and their synthetic tetraploidswere displayed in Figure 8 (for *S. commersonii*) and Figure 9 (for *S. bulbocastanum*). Blast2GO software assigned the following Gene Ontology (GO) categories to each differentially expressed gene: biological process (Fig .8A), molecular function (Fig. 8B), and cellular components (Fig. 8C). Within each functional category, specific terms representing the gene product properties were assigned according to the controlled GO vocabulary (ontology).

In *S. commersonii*, the GO annotation for biological process category resulted in identification of a total number of 28 functional terms. The average number of sequences per GO class ranged from 151 ("response to stress") to 5 ("tropism"). Four terms, within biological process category, counted more than 100 sequences, namely "response to stress" (151), "catabolic process" (128), "protein modification process" (122), and "transport" (118). As for the second GO annotation category (molecular function), a total number of 14 functional terms were found, whereof the most represented class were "nucleotide binding" (205 genes) and "protein binding" (163), and the least were "lipid binding" and "nuclease activity" (10 sequences each). It must be highlighted that, along with "nucleotide binding" terms, only the "protein binding" (163 sequences) class accounted more than 100 genes. Sixteen GO classes were found within cellular components category. "Plastid" class was the most crowed (261 sequences) one and "nuclear envelope" (7 genes) the less.

In *S. bulbocastanum*, GO annotation assigned 25 classes to the biological process category with the higher number of genes attributed to "response to stress" (73) and only 5 differentially expressed genes in the "pollination" class (Fig. 9A). There was no class with more than 100 differentially expressed genes, but 6 ("response to stress", 73; "catabolic process", 72; "transport", 66; "protein modification process", 68; "carbohydrate metabolic process", 55; "signal transduction", 52) displayed a number of sequences higher than 50. The GO annotation for the molecular function category identified 13 classes among the *S. bulbocastanum* 4X Vs. 2X differentially expressed genes (Fig. 9B). "Nucleotide binding" was the biggest group with 129 genes, followed by "protein binding" (109). By contrast the smallest class was "nuclease activity" (7). Finally, as for "cellular component" category, the differentially expressed genes were grouped in 15 classes, where "plastid" was the the biggest (144) and cytoskeleton the smallest (5) (Fig. 9C). There were no other categories with an higher number of differentially expressed sequences than 100.

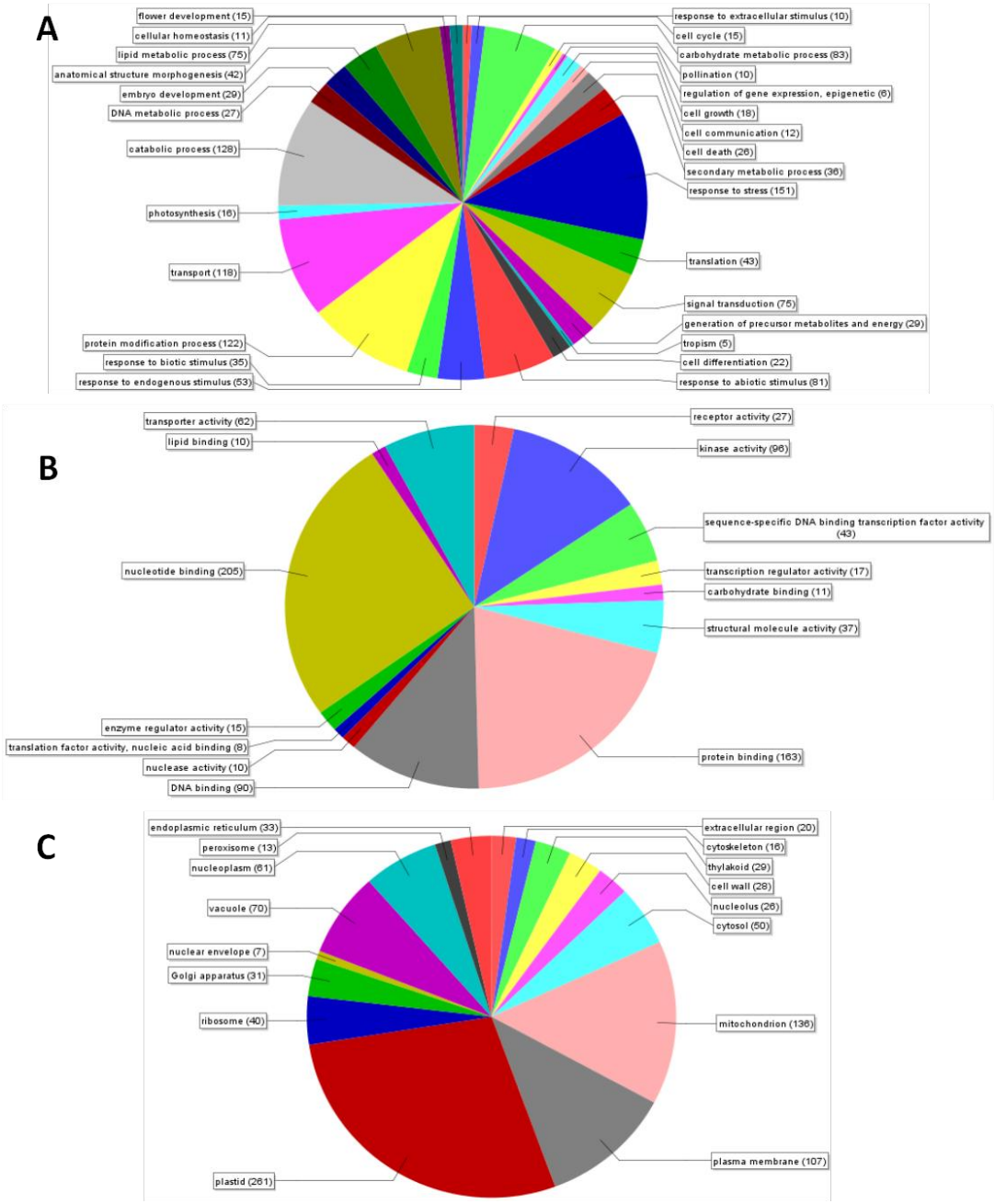

Figure 8. Blast2Go graphs for *S. commersonii* 4x Vs. 2x. A: GO "biological process" terms; B: GO "molecular function" terms; C: GO "cellular comparison compartment" terms. The number of differentially expressed genes for the correspondent GO term was reported in brackets.

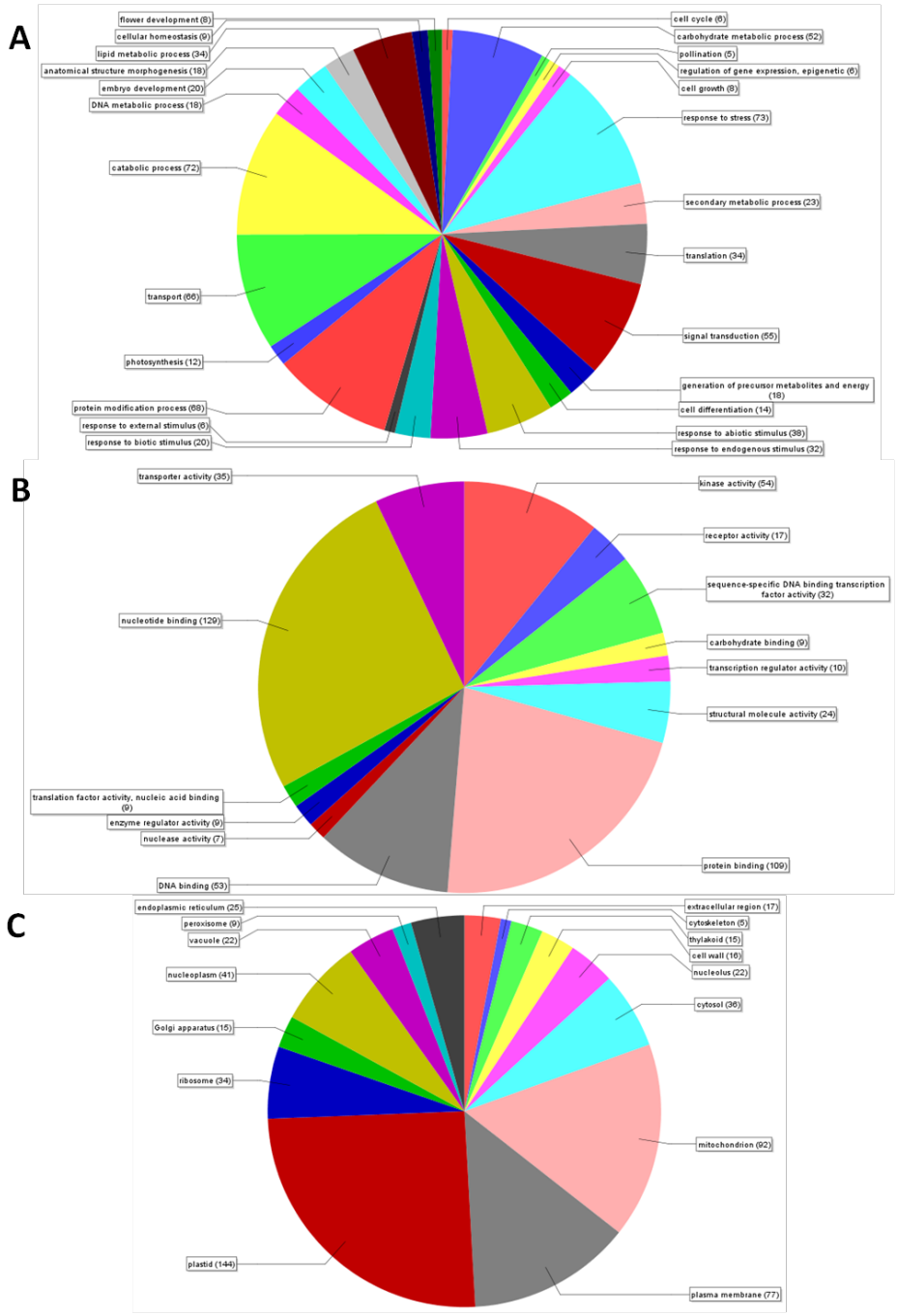

Figure 9. Blast2Go graphs for *S. bulbocastanum* 4x Vs. 2x. A: GO "biological process" terms; B: GO "molecular function" terms; C: GO "cellular comparison compartment" terms. The number of differentially expressed genes for the correspondent GO term was reported in brackets.

In order to visually map clusters of differentially expressed genes involved in common pathways and processes, the *Kyoto Encyclopedia of Genes and Genomes* (KEGG) tool was used. Such approach helped us to focus our attention on specific pathway, which might have been involved inpolyploidization. As for *S. commersonii*, 105 KEGG pathwaysshowed changes in the structural enzymes or transcription factor expression levels. Visual analysis found the phenylalanine, tyrosine and tryptophan biosynthesis pathway was the most affected pathway in

terms of number of differentially expressed genes. Indeed, seven genes (10 E.C. codes) were up- (two genes) and down-regulated (five genes) between 4X and 2X *S. commersonii* genotypes. (Fig. 10A, Tab. 5). In *S. bulbocastanum* 97 KEGG pathways highlighted alterated expression levels. The starch and sucrose metabolism pathway was the most perturbed pathway, where 12 enzymes (13 E.C. codes) were found differentially expressed between BLB1C and the tetraploids they derived from (Fig. 10B, Tab. 6).

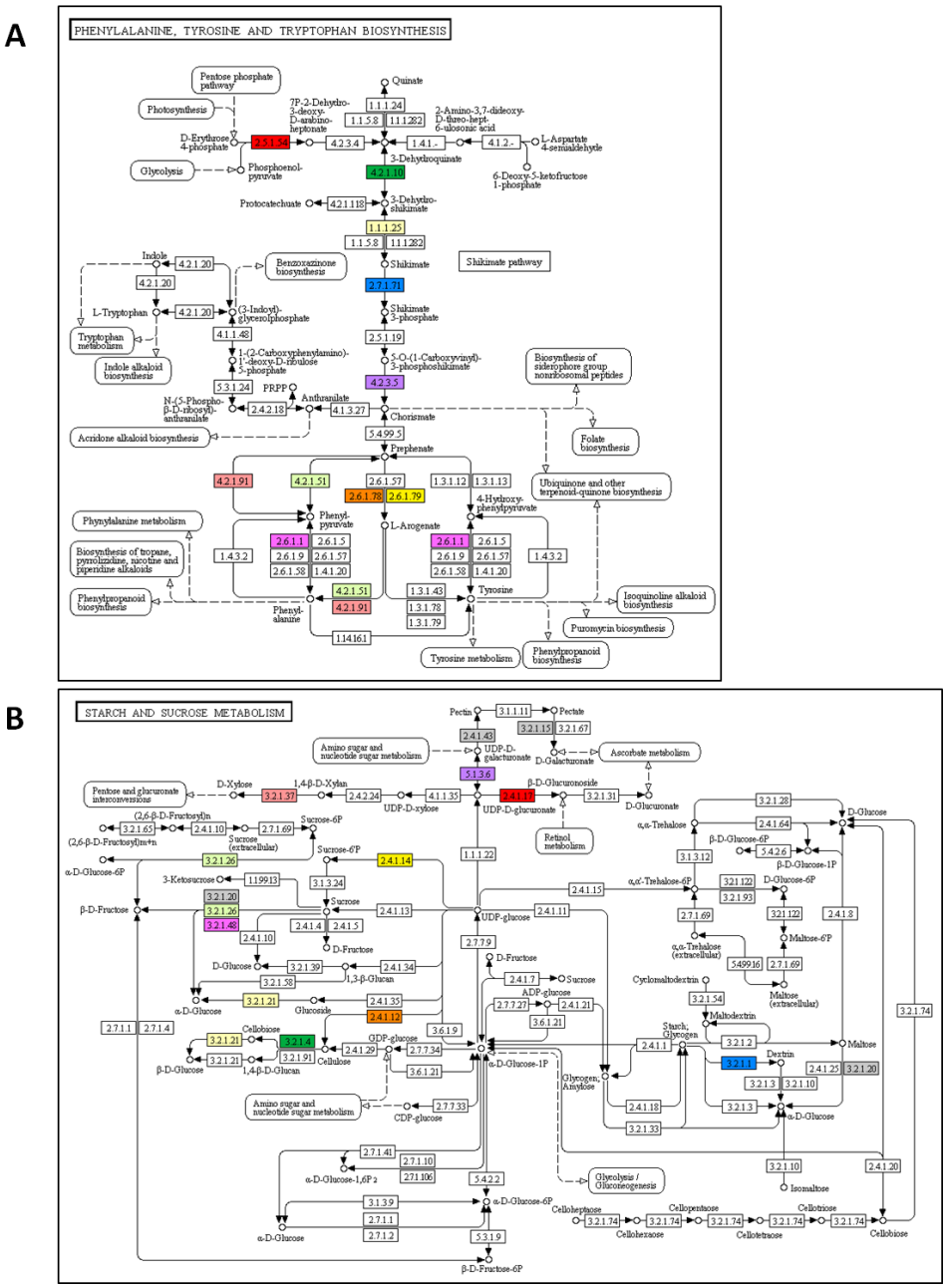

Figure 10. KEGG Metabolic pathways. A: the phenylalanine, tyrosine and tryptophan biosynthesis pathway with the differentially expressed enzymes in coloured boxes for *S. commersonii* 4x Vs. 2x. B: the starch and sucrose metabolism pathway with the differentially expressed enzymes in coloured boxes for *S. bulbocastanum* 4x Vs. 2x.

Table 5. Differentially expressed enzymes found in *S. commersonii* 4x Vs. 2x in the phenylalanine tyrosine and tryptophan biosynthesis pathway.

Enzyme	Cluster ID	Annotation	FC
EC:4.2.3.5	SOLTU027198	Chorismatesynthase	0.78
EC:2.7.1.71	SOLTU005562	Shikimatekinase	0.61
EC:2.5.1.54	SOLTU009774, SOLTU004953	3-deoxy-d-arabino-heptulosonate 7-phosphate synthase	-0.61
EC:4.2.1.91, EC:4.2.1.51	SOLTU017372	Arogenatedehydratase 2	-0.75
EC:4.2.1.10, EC:1.1.1.25	SOLTU027061	Dehydroquinatedehydrataseshikimatedehydrogenase	-0.98
EC:2.6.1.78, EC:2.6.1.79, EC:2.6.1.1	SOLTU005035	Aspartateaminotransferase	-1.16

Table 6. Differentially expressed enzymes found in *S. bulbocastanum* 4x Vs. 2x in the starch and sucrose metabolism pathway.

Enzyme	Cluster ID	Annotation	FC
EC:3.2.1.4	SOLTU006586	endo-beta-glucanase precursor	0.81
EC:3.2.1.15	SOLTU016606	Histonedeacetylase 2	0.78
EC:2.4.1.12	SOLTU026207	glucomannan 4-beta-mannosyltransferase	0.74
EC:3.2.1.21	SOLTU038394	lysosomal beta glucosidase-like	0.59
EC:3.2.1.20	SOLTU006410	alphaglucosidase-likeprotein	0.51
EC:2.4.1.17	SOLTU019830	probableglucuronoxylanglucuronosyltransferase irx7-like	0.48
EC:3.2.1.37	SOLTU035845	xylan 1 4-beta-xylosidase	0.46
EC:2.4.1.43	SOLTU025463	alpha-galacturonosyltransferase 1-like	0.41
EC:3.2.1.26, EC:3.2.1.48	SOLTU026727	Neutralalkalineinvertase	0.38
EC:3.2.1.1	SOLTU015866	pullulanase	-1.54
EC:2.4.1.14	SOLTU007349	sucrose-phosphatesynthase	-1.88
EC:5.1.3.6	SOLTU010421	protein	-2.48

The *Weighted gene co-expression network* (WGCN) analysis of all differentially expressed genes (4X Vs. 2X) resulted in 42 and 61 networks (r=0.98) in *S. commersonii* and *S. bulbocastanum*, respectively (Fig. 11). Each network was annotated based on the GO classesmostly represented in it. The networks annotated as "cellular process", "plastid" (20 networks each) and "binding" (12 networks) were the most present ones in *S. commersonii*WGCN analysis (4X Vs. 2X) (Fig. 12A). Out of 86 networks in *S. commersonii* 17 were not classified (N.C.) since the genes belonging to them were not annotated. Among the 61 *S. bulbocastanum* networks, the "cellular amino acid and derivative metabolic process", "nucleobase, nucleoside, nucleotide and nucleic acid

metabolic process" and "response to stress" were the most represented with 12, 10 ad 7 networks, respectively (Fig. 12B).

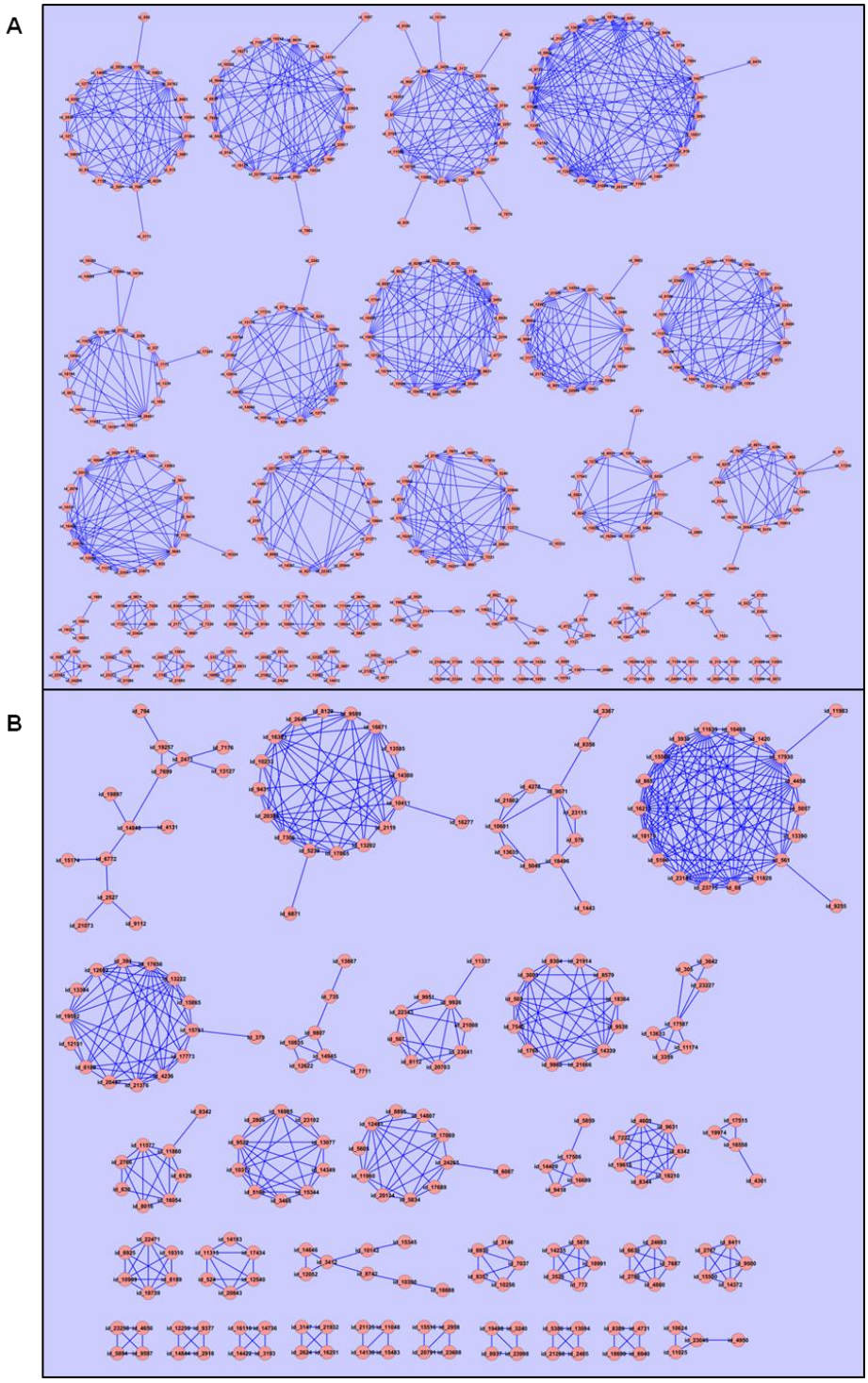

Figure 11. Weighted gene co-expression network (WGCN) of differentially transcribed genes. A percent score was calculated to determine the number of genes within the network modules that overlap their altered fold change value. Each round box representsa differentially expressed gene and the lines linked co-expressed genes. A: co-expression networksfor *S. commersoni* 4x Vs. 2x. B: co-expression networks for *S. bulbocastanum* 4x Vs. 2x.

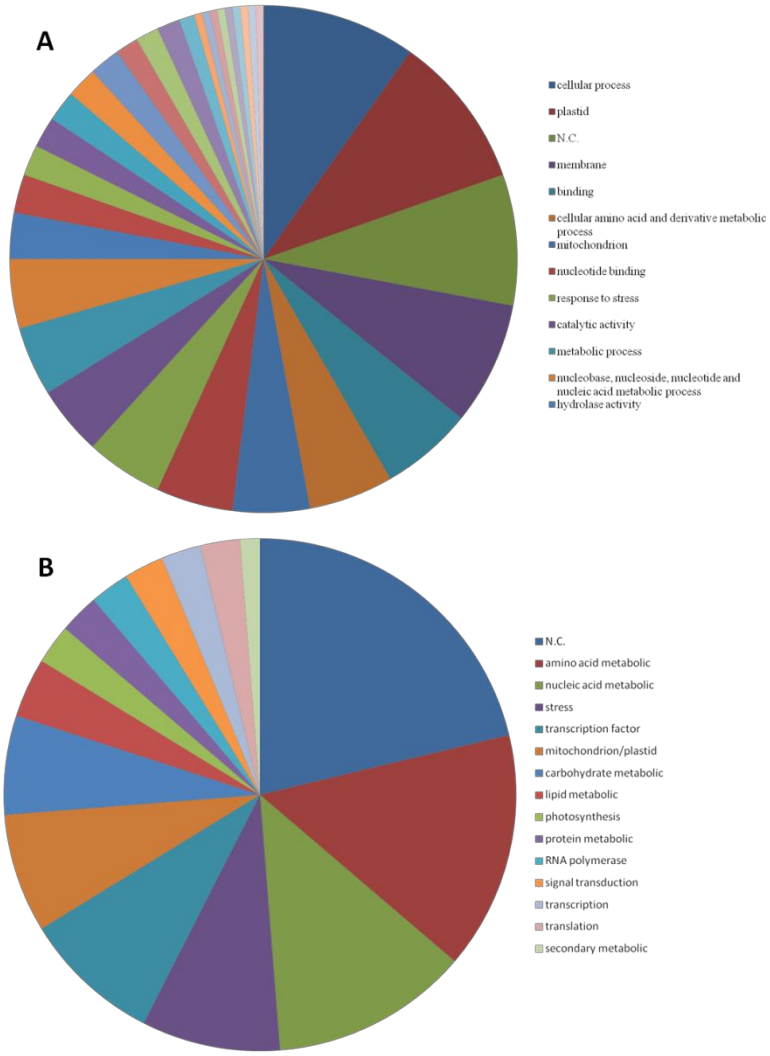

Figure 12. Pie-graph for GO term annotation for genetic co-expression networks. A: results for *S. commersonii* 4x Vs. 2x. B: results for *S. bulbocastanum* 4x Vs. 2x.

3.1.3 qPCR validation of microarray data

Quantitative real-time PCR (qPCR) was used for confirming gene expression results obtained from microarray analysis. Towards this goal, we focus on two genes (cluster IDs: SOLTU027534 and SOLTU033001), whose expression changes were ploidy-dependent (Fig. 13). On average, *far1* showed a 0.5-fold change (down-regulated) in the *S. commersonii* tetraploids and 3.5-fold change (up-regulated) in *S. bulbocastanum* ones. SOLTU033001 was down-regulated in *S. commersonii* tetraploids (0.5-fold change) and up-regulated in the 4x *S. bulbocastanum* (13-fold change).

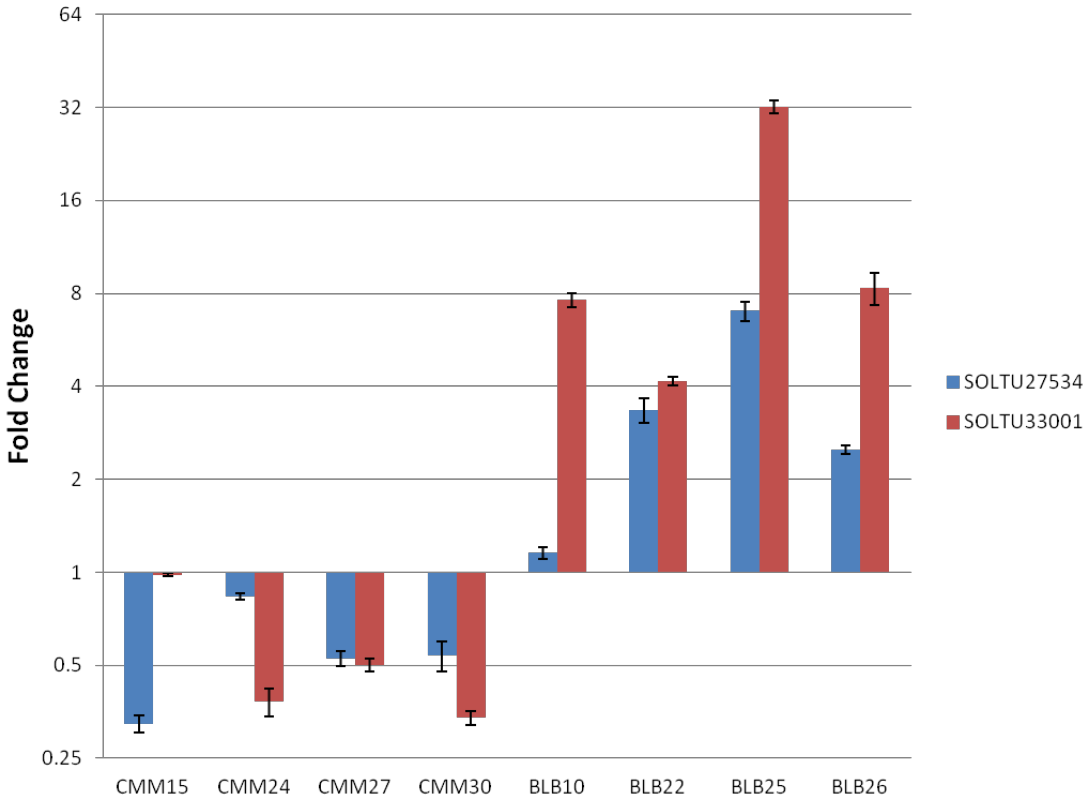

Figure 13. Bar-graph of common differentially expressed genes qPCR Fold Change (FC). qPCR FC of the two common differentially expressed genes (SOLTU27534 and SOLTU33001) was reported for each tetraploid Vs. matched diploid parental comparison in *S. commersonii* (CMM15, CMM24, CMM27 e CMM30) and *S. bulbocastanum* (BLB10, BLB22, BLB25 e BLB26). Error bars represented the standard errors.

3.2 Analysis of metabolites

The qualitative profiles of leaf carotenoids and anthocyanins of diploid CMM1T, BLB1C and their respective synthetic tetraploids were determined using a spectrophotometric assay to evaluate the effect of polyploidization. As shown in Figure 14, the total carotenoids extracts increased in 4x *S. commersonii* (Fig. 14A) as well as the total amount of anthocyanins (Fig. 14B). The analysis of variance showed that the differences in metabolite contents among genotypes were all significant at $P \leq 0.01$, for both species (data not shown). The increased accumulation of anthocyanins in the *S. commersonii*tetraploids was clearly visible during the extraction protocol (Fig. 14C).

In order to produce an accurate and sensitive profiling of CMM1T, BLB1C and their respective synthetic tetraploids, a liquid chromatography-electrospray ionization/multi-stage mass spectrometry (LC-ESI(+)-MS) analysis was performed on 170 targeted molecules. The one-way ANOVA showed statistically significant difference among genotypes at $P \leq 0.01$ and Tukey HSD post hoc test revealed ploidy-dependent changes in metabolite accumulation for 70 and 23 metabolites, in *S. commersonii* and *S. bulbocastanum* respectively. Compounds belonging to both

primary (aminoacids, sugars, organic acids etc) and secondary (flavonoids, alkaloids etc) metabolism were affected. In order to better highlight the differences between diploid parents and their synthetic tetraploids, metabolome data were expressed as a ratio between the metabolite amount in the tetraploid and the diploid parents. This will resultin a Fold Change of 1 in the diploid parent values, more than 1 in case of an accumulation in tetraploids and *vice-versa* between 0 and 1 ifthere was an accumulation in diploid. When a metabolite was not detected in the diploid or in all tetraploidsit was not possible to obtained FC value. In both cases the value reported in the Table 7 and 8 was 0 for the genotype in which the metabolite was not detected and the absolute value for those genotypes in which it was found.

As for *S. commersonii*, comparing the quantity of alkaloids and saponins of tetraploids to those of cmm 1T (Tab. 7), we found that the amount of Calystegine B1, Calystegine B2 and Pennogenin-Tetraglycoside was higher in tetraploids than the diploid CMM1T, resulted in an average FC of more than 2 for each metabolites. By contrast the γ-Tomatine decreased in concentration after polyploidization insomuch as was not detected at all in tetraploids. Similarly, among the nitrogen-containing organic compounds (amines, amides and aminoacids), N-CaffeoylPutrescine, FerulolylPutrescine, Tyramine, Aspartate, Glutamic Acid, and Threonine/Homoserine sensibly increased their content in tetraploids. Impressively N-CaffeoylPutrescine was accumulated by average of 261 times more between tetraploids, the Glutamic Acid by 6 times and the Aspartate was the less accumulated in 4x genotypes with an average FC of 1.5. By contrast, the N,N'-DiCaffeoylSpermidine, Arginine, Asparagine, Glutamine, Methionine and three isomeric leucineshowed the opposite trend showing a common reduction in the tetraploids of about 50%. The phenylpropanoid pathway revealed interesting ploidy dependent changes. For instance, the Malvidin-3-O-Rutinoside-5-O-Glucoside, Myricetin, Petunidin-3-O-Rutinoside-5-O-Glucoside, and Quercetin-Hexose-Deoxyhexose-Hexose-$C_{10}H_8O_3$ were not found in the diploid CMM1T, but were detected in high amount in all synthetic tetraploids. By contrast, Malvidin-3-O-Acetyl-Glucoside and Kaempferol-Hexose-Deoxyhexose-Hexose-$C_9H_6O_2$were consistently lost in tetraploids after polyploidization even if also in the diploid they were presence at very low but detectable levels.

Adenine, Guanosine and their respective nucleosides altogether experienced a drastic decreased in tetraploids also by 100 times (Adenine). Most of phenolic acids tested were up accumulated in the synthetic tetraploids of CMM1T, but not Benzyl Alcohol-Dihexose, KetoGlutaric Acid, Mevalonic Acid and Mevalonic Acid-5-P (average 4x FC 0,47, 0,24, 0,31 and 0,64 respectvely). Interestingly, a ploidy-dependent fall down response was observed for the

isoprenoid 3-DemethylUbiquinol-10 (50 times less accumulated in tetraploids), the polyol D-Pinitol and the vitamins Niacin and Nicotinamide, in fact none of the last three was detected in the tetraploids. Conversely, for the same metabolic classes, Fructose-6-P, Xylitol, Ascorbic Acid and the N5,N10-Formyl-TetrahydroFolic Acid accumulate in tetraploids with the Ascorbic Acid that was not detected at all in the diploid parental samples (Tab. 6).

As for *S. bulbocastanum,* three different alkaloids and saponins metabolites, Malonyl-Saikosaponin A, Pennogenin-Tetraglycoside, a-Solanine, showed a reduction in the tetraploids of a mean FC values of 0.5. Among the nitrogen-containing organic compounds (amines, amides and aminoacids), 8 out 8 metabolites (Spermine, Phospho-L-Serine, Delphinidin-3-O-p-Coumaroyl-Rutinoside-5-O-Glucoside, Myricetin, Peonidin-3-Coumaroyl-Rutinoside-5-O-Glucoside, Quercetin-Hexose-Deoxy-Hexose-Hexose-C9H6O2, Adenosine, Guanosine) were accumulated more in the diploid BLB1C than in 4x genotypes. The polyamine Spermine was detected at low levels in BLB1C but in no one of the tetraploids. The flavonoid Quercetin-Hexose-Deoxy-Hexose-Hexose-C9H6O2 was reduced by ¼ in tetraploids, whereas the other nitrogen-containing organic compounds were subject to a less reduction. All the organics acids (Benzyl Alcohol-Hexose-Pentose, Chlorogenic Acid, p-CoumaroylQuinic Acid, Cryptochlorogenic Acid, PhosphoEnolPyruvic Acid, Quinic Acid), isoprenoids (Mg-Protoporphyrin-MonoMethyl-Ester, Protoporphyrin IX) as well the sugars (1,2-AnydroMyoinositol, Erythritol) metabolites were accumulated in 4x genotypes. All doubling their amounts in tetraploids with the remarkable exception of PhosphoEnolPyruvic Acid that increased of 57 times. The only vitamin differentially accumulated in *S. bulbocastanum* tetraploids was the Riboflavin that was reduced to 0.6 FC (Tab. 8).

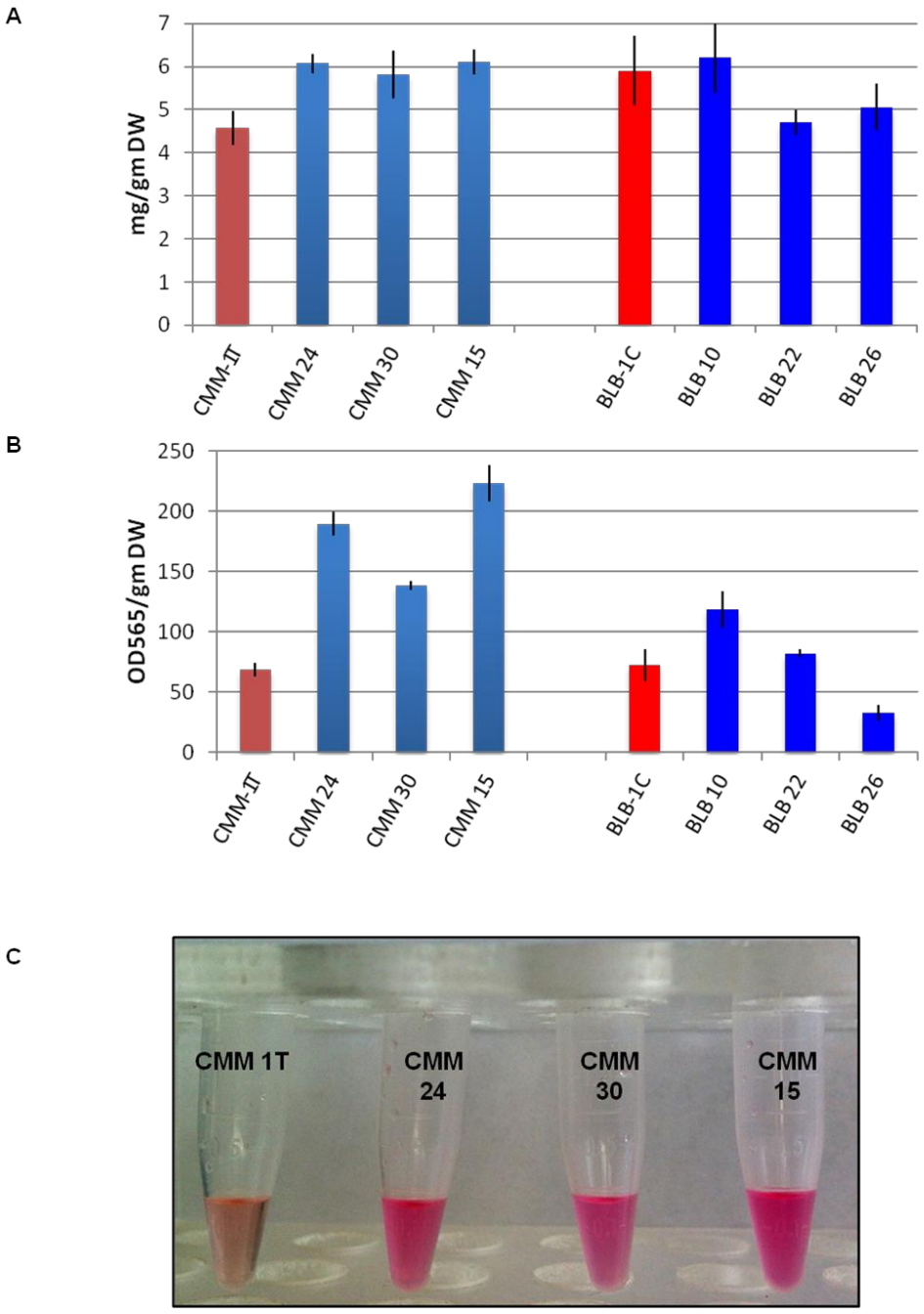

Figure 14. Bar graph of spectrophotometric assays. A: quantification of total carotenoid. B: quantification of total anthocyanin. The diploids (CMM1T and BLB1C) were identified by red coloured bars whereas the blue bars correspond to tetraploids genotypes (CMM15, CMM24, CMM27 e CMM30) in *S. commersonii* and (BLB10, BLB22, BLB25 e BLB26) in *S. bulbocastanum*. C: visual accumulation of total anthocyanin during extraction and purification protocol in tetraploids and parental diploid of *S. commersonii*.

Table 7. List of differentially accumulated metabolites in *S. commersonii* synthetic tetraploids and diploid parent they derived from. The amount of each metabolite was expressed as a ratio between the amounts in 2x parental genotype (CMM1T) and the amounts in 4x tetraploid (CMM15, CMM24, CMM30). Metabolites colored in red were accumulated in tetraploids. Metabolites colored in green were accumulated in diploids.

Metabolic Class	CMM1T[a]	CMM15	CMM24	CMM30
Alkaloids and saponins				
Calystegine B1	1.00 ± 0.13 *b*	3.57 ± 0.72 *a*	1.57 ± 0.21 *b*	1.56 ± 0.22 *b*
Calystegine B2/CE-GABA[b]	1.00 ± 0.14 *b*	3.58 ± 0.70 *a*	1.60 ± 0.21 *b*	1.69 ± 0.36 *ab*
Pennogenin-Tetraglycoside	1.00 ± 0.11 *b*	1.83 ± 0.48 *ab*	1.70 ± 0.39 *b*	2.63 ± 0.63 *a*
γ-Tomatine	1.28 ± 0.18 *a*	0.00 ± 0.00 *b*	0.00 ± 0.00 *b*	0.00 ± 0.00 *b*
Amides				
N-CaffeoylPutrescine	1.00 ± 0.18	283.17 ± 35.12	226.59 ± 20.65	274.26 ± 21.28
N.N'-DiCaffeoylSpermidine	1.00 ± 0.32 *a*	0.00 ± 0.00 *c*	0.37 ± 0.24 *bc*	0.56 ± 0.08 *ab*
FerulolylPutrescine	1.00 ± 0.11 *c*	1.84 ± 0.27 *a*	1.85 ± 0.26 *a*	1.45 ± 0.18 *b*
Amines				
Tyramine	1.00 ± 0.17 *c*	2.17 ± 0.36 *ab*	1.43 ± 0.25 *b*	2.78 ± 0.25 *a*
Aminoacids				
Arginine	1.00 ± 0.09 *a*	0.37 ± 0.05 *b*	0.36 ± 0.05 *c*	0.75 ± 0.07 *a*
Asparagine	1.00 ± 0.08 *a*	0.16 ± 0.01 *b*	0.14 ± 0.02 *c*	0.70 ± 0.05 *a*
Aspartate	1.00 ± 0.10 *b*	1.63 ± 0.15 *b*	1.58 ± 0.11 *ab*	1.20 ± 0.07 *ab*
Glutamic Acid	1.00 ± 0.06 *c*	6.73 ± 0.73 *ab*	7.40 ± 0.57 *a*	4.83 ± 0.39 *b*
Glutamine	1.00 ± 0.03 *a*	0.22 ± 0.03 *b*	0.19 ± 0.02 *c*	0.79 ± 0.06 *a*
Leucine/IsoLeu./NorLeu.	1.00 ± 0.15 *a*	0.35 ± 0.06 *c*	0.29 ± 0.05 *c*	0.68 ± 0.07 *b*
Methionine	1.00 ± 0.13 *a*	0.59 ± 0.12 *b*	0.55 ± 0.11 *b*	0.70 ± 0.04 *ab*
Threonine or Homoserine	1.00 ± 0.08 *a*	2.84 ± 0.37 *b*	2.46 ± 0.26 *a*	3.11 ± 0.29 *a*
Flavonoids.Anthocyanins and other Phenylpropanoids				
Caffeic Acid Glucoside	1.00 ± 0.13 *a*	1.89 ± 0.26 *b*	1.39 ± 0.15 *a*	1.68 ± 0.30 *a*
Trans-Chalcone	1.00 ± 0.20 *a*	0.65 ± 0.09 *ab*	0.54 ± 0.10 *b*	0.54 ± 0.06 *b*
Cinnamic Aldehyde	1.00 ± 0.17 *a*	0.10 ± 0.07 *b*	0.09 ± 0.04 *b*	0.20 ± 0.06 *b*
CAG	1.00 ± 0.19 *a*	0.15 ± 0.02 *b*	0.42 ± 0.11 *c*	0.17 ± 0.14 *b*
Coumarin	1.00 ± 0.28 *a*	0.00 ± 0.00 *c*	0.00 ± 0.00 *c*	0.49 ± 0.38 *b*
GCA[b]	1.00 ± 0.12 *a*	0.79 ± 0.07 *ab*	0.73 ± 0.11 *ab*	0.65 ± 0.07 *b*
KHDH-$C_{10}H_8O_3$[b]	1.00 ± 0.15 *d*	7.18 ± 0.99 *c*	10.92 ± 1.14 *b*	20.79 ± 3.11 *a*
KHDH-$C_9H_6O_2$[b]	0.01 ± 0.44 *a*	0.00 ± 0.00 *b*	0.00 ± 0.00 *b*	0.00 ± 0.00 *b*
Malvidin-3-O-Acetyl-Glucoside	0.02 ± 0.05 *a*	0.00 ± 0.00 *b*	0.00 ± 0.00 *b*	0.00 ± 0.00 *b*
Malvidin-3-O-Glucoside	1.00 ± 0.26 *a*	0.00 ± 0.00 *b*	0.00 ± 0.00 *b*	0.67 ± 0.26 *ab*
MCRG[b]	1.00 ± 0.38 *d*	27.53 ± 4.66 *a*	13.64 ± 2.16 *b*	5.80 ± 1.38 *c*
MCpRG[b]	1.00 ± 0.11 *c*	3.06 ± 0.50 *a*	2.02 ± 0.26 *b*	1.78 ± 0.38 *b*
MRG[b]	0.00 ± 0.00 *c*	659.60 ± 101.58 *a*	528.79 ± 117.0 *a*	181.49 ± 65.40 *b*
Myricetin	0.00 ± 0.00 *d*	100.02 ± 18.86 *a*	86.44 ± 9.00 *b*	16.01 ± 4.55 *c*
Petunidin-3-O-Glucoside	1.00 ± 0.24 *a*	0.00 ± 0.00 *b*	0.00 ± 0.00 *b*	0.59 ± 0.15 *c*
PRG[b]	0.00 ± 0.00 *c*	1356.29 ± 215.69 *a*	1320.1 ± 297.6 *a*	865.70 ± 191.9 *b*
Quercitin or Delphinidin	1.00 ± 0.09 *c*	2.71 ± 0.47 *b*	2.37 ± 0.34 *b*	4.01 ± 0.58 *a*
QHDH-$C_{10}H_8O_3$[b]	0.00 ± 0.00 *c*	184.61 ± 24.79 *b*	167.42 ± 33.59 *b*	1026.8 ± 165.0 *a*
QRG[b]	1.00 ± 0.06 *c*	1.88 ± 0.24 *b*	1.93 ± 0.19 *b*	4.10 ± 0.53 *a*

Metabolic Class	CMM1T[a]	CMM15	CMM24	CMM30
Quercetin-3-O-Trisaccharide	1.00 ± 0.05 *c*	2.22 ± 0.32 *b*	1.94 ± 0.23 *b*	6.99 ± 0.92 *a*
Quercetin-3-Di-Glucoside	1.00 ± 0.09 *c*	2.37 ± 0.35 *b*	2.16 ± 0.27 *b*	3.72 ± 0.52 *a*
Nucleotides and Nucleosides				
Adenine	1.00 ± 0.07 *a*	0.01 ± 0.00 *b*	0.01 ± 0.01 *b*	0.00 ± 0.00 *b*
Adenosine	1.00 ± 0.07 *a*	0.38 ± 0.07 *b*	0.38 ± 0.04 *b*	0.33 ± 0.04 *b*
Guanine	1.00 ± 0.24 *a*	0.00 ± 0.00 *b*	0.10 ± 0.03 *b*	0.00 ± 0.00 *b*
Guanosine	1.00 ± 0.16 *a*	0.18 ± 0.03 *b*	0.24 ± 0.03 *b*	0.11 ± 0.03 *b*
Organic/Phenolic Acids and Esters				
Anthranilic Acid	1.00 ± 0.08 *b*	1.54 ± 0.15 *a*	1.55 ± 0.16 *a*	1.40 ± 0.15 *a*
Benzyl Alcohol-Dihexose	1.00 ± 0.35 *a*	0.67 ± 0.23 *ab*	0.40 ± 0.10 *b*	0.33 ± 0.09 *b*
Caffeic Acid	0.00 ± 0.00 *d*	3315.95 ± 502.82 *b*	2322.2 ± 305.1 *c*	5286.8 ± 600.0 *a*
Chlorogenic Acid	1.00 ± 0.08 *b*	46.68 ± 6.40 *a*	46.64 ± 3.93 *a*	40.94 ± 4.09 *a*
Cinnamic Acid	0.00 ± 0.00 *c*	72.77 ± 18.59 *b*	99.95 ± 36.29 *a*	78.81 ± 34.74 *b*
Cryptochlorogenic Acid	1.00 ± 0.09 *b*	44.23 ± 3.50 *a*	48.63 ± 4.39 *a*	42.38 ± 4.30 *a*
Glyceric Acid-2-P or -3-P	0.00 ± 0.00 *d*	1304.98 ± 161.92 *b*	1008.1 ± 137.0 *c*	1824.9 ± 222.6 *a*
KetoGlutaric Acid	1.00 ± 0.21 *a*	0.11 ± 0.02 *b*	0.14 ± 0.06 *b*	0.49 ± 0.16 *b*
2-MethylCitric Acid	0.00 ± 0.00 *d*	469.13 ± 84.97 *c*	1001.0 ± 125.3 *a*	659.87 ± 81.44 *b*
Mevalonic Acid	1.00 ± 0.08 *a*	0.15 ± 0.02 *b*	0.08 ± 0.06 *b*	0.69 ± 0.06 *ab*
Mevalonic Acid-5-P	1.00 ± 0.09 *a*	0.71 ± 0.11 *ab*	0.62 ± 0.12 *b*	0.58 ± 0.10 *b*
Quinic Acid	1.00 ± 0.18 *bc*	2.00 ± 0.45 *a*	1.37 ± 0.23 *b*	0.99 ± 0.19 *bc*
Sinapic Acid or DHA[b]	0.00 ± 0.00 *c*	6450.10 ± 1019.38 *a*	6720.6 ± 819.3 *a*	22543 ± 3286 *b*
Peptides				
Cysteinyl-Glycine	0.00 ± 0.00 *d*	218.45 ± 47.67 *a*	39.27 ± 18.90 *c*	141.92 ± 35.57 *b*
Glutathione	1.00 ± 0.25 *c*	354.79 ± 41.53 *ab*	207.81 ± 14.90 *b*	318.38 ± 25.31 *a*
Polar Isoprenoids				
3-DemethylUbiquinol-10	1.00 ± 0.12 *a*	0.00 ± 0.00 *b*	0.00 ± 0.00 *b*	0.00 ± 0.00 *b*
Polar Lipids				
GlyceroPhosphoCholine	1.00 ± 0.09 *c*	20.78 ± 1.89 *b*	31.19 ± 2.88 *b*	74.14 ± 7.41 *a*
PhosphoCholine	0.00 ± 0.00 *c*	1136.09 ± 115.26 *b*	811.23 ± 97.11 *bc*	3000.5 ± 258.8 *a*
Sugar.Polyols and Phosphates				
Erythrose-4-P	1.00 ± 0.24 *a*	0.47 ± 0.09 *ab*	0.31 ± 0.07 *b*	0.30 ± 0.03 *b*
F6-P or G6-P or I6-P[b]	1.00 ± 0.06 *b*	2.13 ± 0.10 *ab*	2.37 ± 0.27 *ab*	2.70 ± 0.23 *b*
Maltitol	1.00 ± 0.09 *a*	0.73 ± 0.12 *b*	0.71 ± 0.07 *b*	0.69 ± 0.11 *b*
D-Pinitol	0.01 ± 0.40 *a*	0.00 ± 0.00 *b*	0.00 ± 0.00 *b*	0.00 ± 0.00 *b*
Xylitol or Arabitol or Ribitol	1.00 ± 0.42 *c*	3.61 ± 0.64 *a*	3.31 ± 0.74 *ab*	1.94 ± 0.71 *b*
Vitamins				
Ascorbic Acid	0.00 ± 0.00 *c*	846.04 ± 299.50 *a*	183.35 ± 74.31 *b*	194.39 ± 77.66 *b*
FTFA[b]	1.00 ± 0.41 *b*	2.37 ± 0.94 *ab*	2.96 ± 0.64 *a*	1.63 ± 0.33 *ab*
Niacin	0.10 ± 0.12 *a*	0.00 ± 0.00 *b*	0.00 ± 0.00 *b*	0.00 ± 0.00 *b*
Nicotinamide	0.09 ± 0.17 *a*	0.00 ± 0.00 *b*	0.00 ± 0.00 *b*	0.00 ± 0.00 *b*
Riboflavin	1.00 ± 0.11 *a*	0.73 ± 0.17 *b*	0.64 ± 0.10 *b*	0.64 ± 0.12 *b*
Thiamine	1.00 ± 0.05 *b*	0.00 ± 0.00 *a*	0.29 ± 0.11 *b*	0.34 ± 0.06 *c*

[a]) Means with some letters are not significantly different in according to Tukey Test (P<0).

[b]) CE-GABA: *N-Carboxy-Ethyl-GABA*; CAG: *Coniferyl Alcohol-β-D-Glucoside*; GCA: *4'-O-β-D-Glucosyl-cis-p-Coumaric Acid*; KHDH: *Kaempferol-Hexose-Deoxyhexose-Hexose*; MCRG: *Malvidin-3-Caffeoyl-Rutinoside-5-O-Glucoside*; MCpRG: *Malvidin-3-O-p-Coumaroyl-Rutinoside-5-O-Glucoside*; MRG: *Malvidin-3-O-Rutinoside-5-O-*

Glucoside; PRG: *Petunidin-3-O-Rutinoside-5-O-Glucoside*; QHDH: *Quercetin-Hexose-Deoxyhexose-Hexose*; QRG: *Quercetin-3-O-Rutinoside-7-O-Glucoside*; DHA: *3.5-Dimethoxy-4-Hydroxycinnamic Acid*; F6-P: *Fructose-6-P*; G6-P: *Glucose-6-P*; I6-P: *Inositol-6-P*; FTFA: *N5.N10-Formyl-TetrahydroFolic Acid*.

Table 8. List of differentially accumulated metabolites in *S. bulbocastanum* synthetic tetraploids and the diploid parent they derived from. The amount of each metabolite was expressed as a ratio between the amounts in 2x parental genotype (BLB1C) and the amounts in 4x tetraploid (BLB10, BLB22, BLB26). Metabolites colored in red were accumulated in tetraploids. Metabolites colored in green were accumulated in diploids.

Metabolic Class	BLB1C[a]	BLB10	BLB22	BLB26
Alkaloids and saponins				
Malonyl-Saikosaponin A	1.00 ± 0.10 *a*	0.54 ± 0.06*b*	0.45 ± 0.09*b*	0.44 ± 0.07*b*
Pennogenin-Tetraglycoside	1.00 ± 0.11*a*	0.68 ± 0.07*b*	0.56 ± 0.04 b	0.65 ± 0.11*b*
α-Solanine	1.00 ± 0.16*a*	0.51 ± 0.07*b*	0.33 ± 0.05*c*	0.42 ± 0.07*bc*
Amines				
Spermine	1.00 ± 0.25*a*	0.00 ± 0.00*b*	0.00 ± 0.00*b*	0.00 ± 0.00*b*
Aminoacids				
Phospho-L-Serine	1.00 ± 0.18*ab*	0.77 ± 0.08*ab*	0.50 ± 0.21*b*	0.78 ± 0.17*ab*
Flavonoids.Anthocyanins and other Phenylpropanoids				
DCRG[b]	1.00 ± 0.21*a*	0.65 ± 0.10*b*	0.24 ± 0.06*c*	0.03 ± 0.01*c*
Myricetin	1.00 ± 0.14*a*	0.39 ± 0.05*c*	0.28 ± 0.06*c*	0.66 ± 0.16*b*
PCRG[b]	1.00 ± 0.10*a*	0.00 ± 0.00*b*	0.00 ± 0.00*b*	0.76 ± 0.14*a*
QHDHH[b]	1.00 ± 0.19*a*	0.57 ± 0.10*b*	0.19 ± 0.04*c*	0.04 ± 0.01*c*
Nucleotides and Nucleosides				
Adenosine	1.00 ± 0.13*a*	0.19 ± ±0.02*d*	0.72 ± 0.06*b*	0.46 ± 0.06*c*
Guanosine	1.00 ± 0.07*a*	0.14 ± 0.03*c*	0.60 ± 0.10*b*	0.67 ± 0.13*b*
Organic/Phenolic Acids and Esters				
BenzylAlcohol-Hexose-Pentose	1.00 ± 0.18*b*	2.07 ± 0.23*a*	2.14 ± 0.48*a*	2.05 ± 0.41*a*
Chlorogenic Acid	1.00 ± 0.14*c*	2.64 ± 0.22*a*	1.31 ± 0.08*bc*	2.03 ± 0.35*ab*
p-CoumaroylQuinic Acid	1.00 ± 0.16*b*	2.00 ± 0.33*a*	1.38 ± 0.27*a*	2.47 ± 0.54*b*
Cryptochlorogenic Acid	1.00 ± 0.13*d*	2.70 ± 0.24*a*	1.32 ± 0.06*c*	2.04 ± 0.34*b*
PhosphoEnolPyruvic Acid (PEP)	0.00 ± 0.00*b*	68.05 ± 11.64 *a*	45.83 ± 13.11*a*	0.00 ± 0.00*b*
Quinic Acid	1.00 ± 0.32*b*	1.77 ± 0.13*a*	1.53 ± 0.93*ab*	2.30 ± 0.50*a*
PolarIsoprenoids				
PME[b]	1.00 ± 0.17*b*	1.56 ± 0.13*a*	1.41 ± 0.15*a*	1.55 ± 0.19*a*
Protoporphyrin IX	1.00 ± 0.24*d*	2.07 ± 0.63*b*	1.52 ± 1.24*c*	2.18 ± 0.42*a*
Sugar.Polyols and Phosphates				
1.2-AnydroMyoinositol	1.00 ± 0.17*c*	1.59 ± 0.35*b*	2.39 ± 0.30*a*	1.40 ± 0.20*bc*
Erythritol or Threitol	1.00 ± 0.39*c*	1.68 ± 0.57*b*	1.94 ± 0.95*ab*	2.12 ± 0.62*a*
Xylitol or Arabitol or Ribitol	1.00 ± 0.29 b	1.43 ± 0.27*a*	1.31 ± 0.22*a*	1.46 ± 0.27*a*
Vitamins				
Riboflavin	1.00 ± 0.15*a*	0.38 ± 0.09*c*	0.66 ± 0.07*b*	0.75 ± 0.12*b*

[a]) Means with some letters are not significantly different in according to Tukey Test (P<0).
[b]) PCRG: *Peonidin-3-Coumaroyl-Rutinoside-5-O-Glucoside*; QHDHH: *Quercetin-Hexose-Deoxy-Hexose-Hexose*-C9H6O2; DCRG: *Delphinidin-3-O-p-Coumaroyl-Rutinoside-5-O- Glucoside*; PME: *Mg-Protoporphyrin-MonoMethyl-Ester*.

3.3 R2R3-MYB transcription factor analysis

A genomic fragment corresponding to R2R3-MYB transcription factor was isolated from CMM1T and BLB1C, and the sequence compared to *S. tuberosum* anthocyanin 1 (*an1*) gene (alleles 777 and 816,corresponding to GenBankcodes: AY841130 and AY841128, respectively). In order to determine the anthocyanin 1 (*an1*) genestructure in CMM1T and BLB1C, four primer pairs were designed on *S. tuberosum* reference sequence to cover the entire genomic sequence (from the exon 1 to the 3 'UTR). The PCR experiments resulted in amplicons of the expected sizes for both species. All the PCR products were purified, sequenced, and then aligned to the afore-mentioned *S. tuberosum an1* references. The high homology found revealed the existence of *an1* gene in both wild species. In particular, it was possible to recognize the typical *an1* gene organization, composed of three exons and two introns. The "R2R3" binding domain was spanned in the three exons, whereas the "variable region" (encoding a non-DNA-binding domain) was located in the third exon downstream the R3 domain.

The complete nucleotide sequence of *an1* was obtained for *S. commersonii*; by contrast only the exon1 and a small upstream fragment of the exon3 were obtained for *S. bulbocastanum*. As for *S. commersonii* CMM1T a series of polymorphisms (SNPs, INDELs, substitutions) were detected in introns 1 and 2, as well as inthe coding sequence of exons 2 and 3 (Fig. 15). A list of all the polymorphisms found in CMM1T relative to both *S. tuberosum* alleles (777 and 816) was enlisted in Table 8. In order to identify the effects of such polymorphisms on protein sequence, the amino acid sequence of CMM1T *an1* was predicted and a protein alignment carried out (Tab. 9). Ahistidine (H. basic aminoacid) present in both *an1* alleles 816 and 777, was substituted into asparagine (N. hydrophilic aminoacid) in the highly conserved R2 domain. In the exon1,within the R2 domain we found an Alanine already present in *an1* 777 allele (data not shown). In the exon2 a substitution of Aspartic Acid and Tryptophan (DW) in Glutamic Acid and Proline (EP).was found within R3 domain. The remaining amminoacid substitutions within R3 domain were identified in the exon3. For three of them (I $\rightarrow$ M. A $\rightarrow$I and N $\rightarrow$C) no changes in the amminoacid groups were observed (neutral non polar for the first two and neutral polar for the third one). The first mutation on the exon 3 was detected at 856 nucleotide (nt) and resulted in a substitution of a Cysteine (C. neutral polar) in a Lysine (K. basic). More downstream at the 987 nt an Aspartic Acid (D. acid) was replaced by Histidine (H. basic), at 1051 nt a Threonine (T. neutral polar) by a Methionine (M. neutral non-polar) and at 1087 nt a Serine (S. neutral non-polar) was swapped by an Alanine (A. neutral polar). Finally, in the predicted R2R3 domain sequence of CMM1T was presented a conserved motif (A/S/G) NDV typical of dicot anthocyanin promoting MYB (158).

As for *S. bulbocastanum*, the sequence alignments revealed a high polymorphism (data not shown). In BLB1C the non-polar aminoacid Isoleucine (I. neutral non-polar) was replaced by the hydrophilic aminoacid Threonine (T. neutral polar), due to a nucleotide substitutions of a thymine (typical of *an1*-816/*an1*-777 alleles) with a cytosine in the R2 part of exon1 (Fig. 15). More downstream in the exon1 a different Isoleucine (I. neutral non-polar) was substitute by a Valine (V. neutral non-polar). The last R2 domain polymorphism found in BLB1C *an1* sequence was a specific allele 816 Threonine (T. neutral polar).

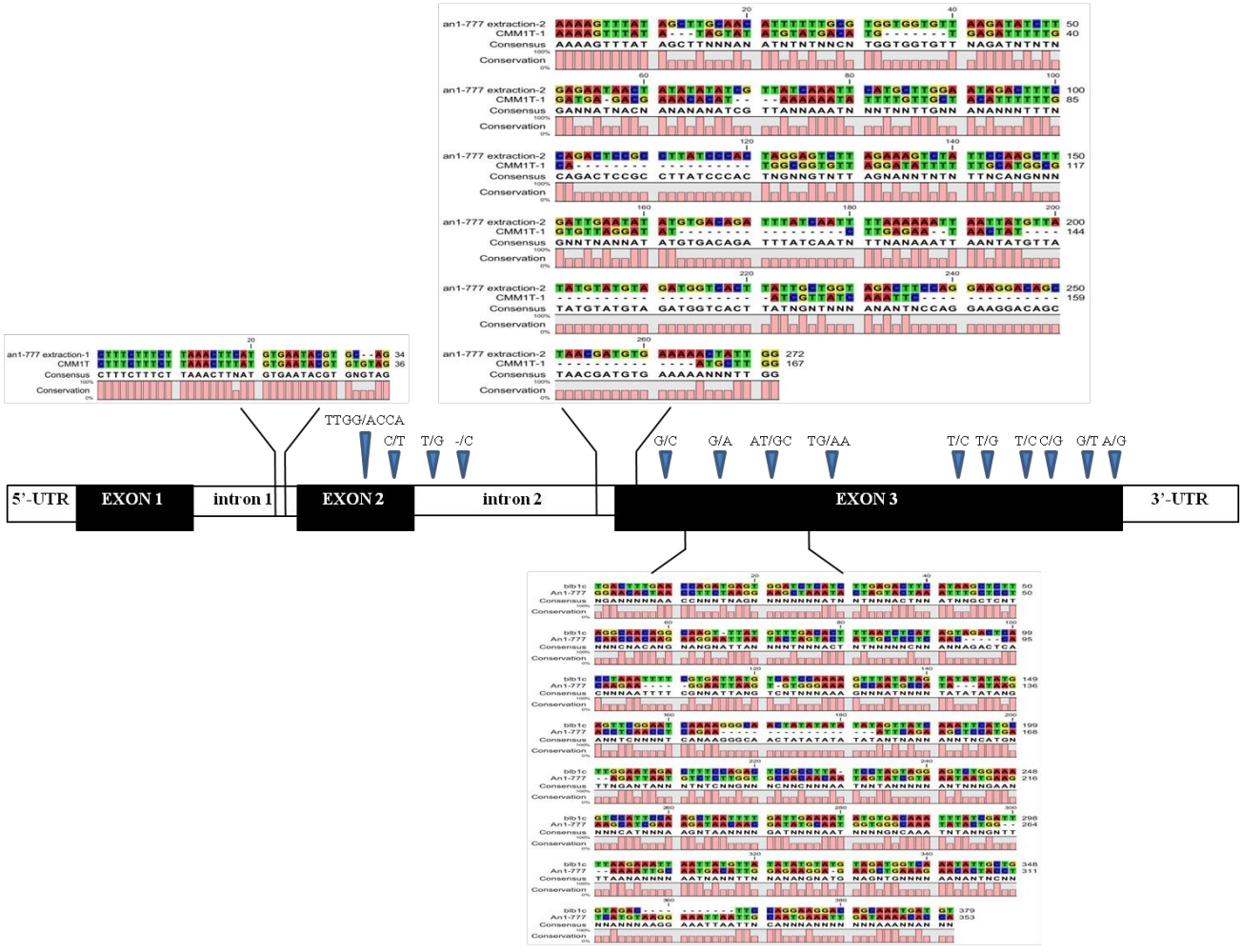

Figure 15. *an1* gene structure sequence and polymorphisms. In the upper part are reported polymorphisms or mutations for CMM1T genotype; in the bottom part polymorphisms or mutations for BLB1C genotype. Only polymorphisms between the wild potato *an1* sequence and cultivated *S. tuberosum* were showed.

Table 9. List of polymorphisms in the coding sequence region of *an1* in CMM1T genotype.

	nt	***an1*-816**	***an1*-777**	**CMM1T**	**aa[a] substitutions**
exon 1	115	c	a	c	
	133	a	g	g	H → N
exon 2	256	a	g	g	
	318-322	attgg	attgg	aacca	DW→ EP
	368	t	t	c	
exon 3	856-859	tgt	tgt	aaa	C → K
	948	c	c	g	I → M
	964-966	gc	gc	at	A→ I
	975	c	t	c	
	976	a	a	t	
	977	a	a	g	N→ C
	981	c	c	t	
	987	g	g	c	D → H
	1051	c	c	t	T → M
	1087	g	g	c	S → A
	1197	g	g	a	

[a]: H: Histidine; N: Asparagine; D: Aspartic acid; W: Tryptophane; E: Glutamic acid; P: Proline; C: Cysteine; K: Lysine; I: Isoleucine; M: Methionine; T: Threonine; S: Serine.

The expression analysis by qPCR of *an1* gene and two structural enzymes located downstream the anthocyanin biosynthetic pathway were shown in Figure 16. The transcription factor *an1* expression was down regulated in all the tetraploids especially in CMM27 in which it was present 10 times less than CMM1T, expect for CMM24 where it was slightly up-regulated of about 1.5 FC. Similarly, the anthocyanidin synthase (*ans*) genes were down regulated in all tetraploids of both species with an average value of fold change of 0.5. Finally, the dihydroflavonol 4-reductase (*dfr*) gene was donw regulated in all synthetic tetraploid of *S.bulbocastanum*, in CMM24,CMM27 and CMM30, but not in CMM15, where it was up regulated 1.2 FC.

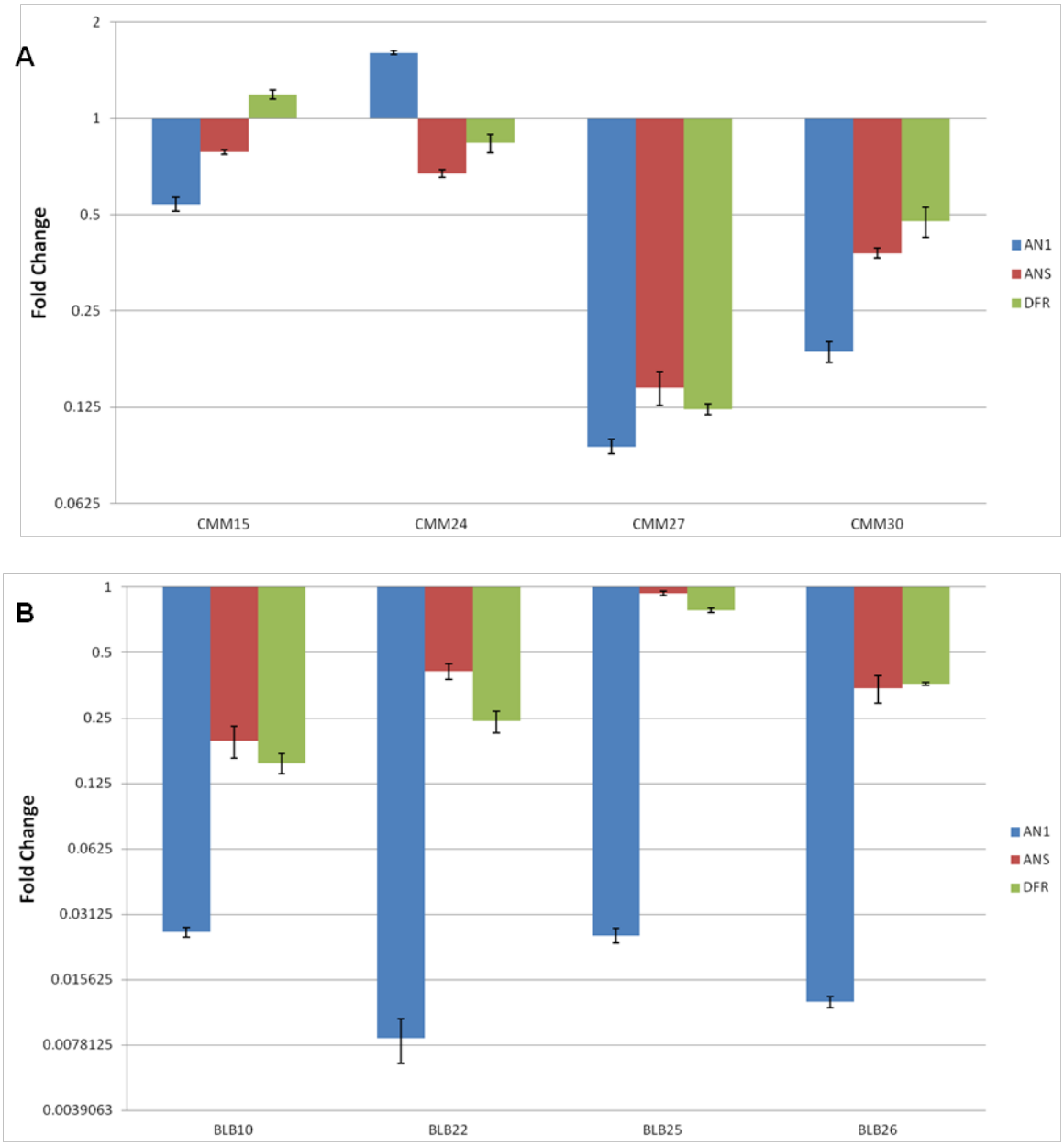

Figure 16. Bar graph for qPCR experiments. Quantitative gene expression of R2R3 MYB transcription factor *an1* (blue bar) and structural enzyme *ans* (red bar) and *dfr* (green bar) were showed as Fold Change between 4x VS. 2x. A: *S. commersonii* tetraploids (CMM15, CMM24, CMM27 and CMM30). B: *S. bulbocastanum* tetraploids (BLB10, BLB22, BLB25 and BLB26).

4. Discussion

The interest in polyploids rises high in the scientific community principally for two reasons. The first is the widespread occurrence of polyploidy in nature. Ohno and colleagues hypothesized that all eukaryotic organisms have experienced genome duplication and it is a fact that many crop species are polyploids. Therefore, it was stated that "life on earth is predominantly a polyploid phenomenon and civilization depends mainly on use of polyploid crops". The second is that one or more rounds of genome doubling occurred in the evolution history of about 70% of today angiosperm organisms (paleopolyploids). Indeed "omics" sciences investigations have found footprints of ancestral genome duplication (7, 159, 160), in above suspicious diploid species, such as *A. thaliana*. Thus, polyploidization is considered a major evolutionary force in plants. Accepted important effects of polyploidization are the instantaneous speciation due to immediate reproductive isolation between newly formed polyploids and their parents and the possibility of newly formed polyploids to overcome sterility and produce viable offspring. Following the pioneering work by McClintock (161), at molecular level many authors described as "genomic shock" the scenario in which cells go through after a polyploidization event, leading to several such as at genome, transcriptome, metabolome and epigenome levels. In particular, the effects of polyploidization on trascriptome, has been widely studied in the last few years. As for allopolyploids, recent microarray studies provided evidences that dramatic changes in gene expression profiles occur in allopolyploids of *Nicotiana*, *Senecio*, *Gossypium*, *Coffea*, *Spartina*, *Tragopogon*, *Triticum* and *Arabidopsis*. By contrast, little information exists for there trascriptome profiling (autopolyploids) were carried out in power species, such as lemon (83) and potato (81).

Results of all these studies reported gene expression alteration between diploids and tetraploids of the same species and showed that the amplitude of this "transcriptome shock" is less strong in autopolyploids than in allopolyploids. What is not clear is whether these are true responses to genome doubling *per se*, and therefore reproducible. However, independent replication of the polyploidization process was not always employed so far and investigations on allopolyploidization effects are confounded by hybridization. Given that, we applied this criterion by analyzing the transcriptome of two diploid wild potato species, *Solanum commersonii* and *S. bulbocastanum,* and their corresponding four independent synthetic tetraploids (CMM15, CMM24, CMM27, CMM30 and BLB10, BLB22, BLB25 and BLB26, respectively). Thanks to Combimatrix microarray platform, we were able to monitor the gene expression of more than 21000 known expressed genes.

We detected a high number of differentially expressed genes (7853, 37%) between the two species regardless of ploidy level (CMM Vs. BLB). Such results were not unexpected considering phylogenetic distance and phenotypic differences exiting between these two wild potato species, which have undergone a different evolutionary pathway.

Indeed, elaborating the data with a different type of statistical tool to find difference in gene expression in all the tetraploid in comparison with the diploid parents (CMM4X Vs. 2X and BLB4X Vs. 2X), we found that the percentage of differentially expressed genes in *S. commersonii* (8.3%) was double that and to *S. bulbocastanum* (4.6%). Therefore, while *S. commersonii* transcriptome alterations were consistent with previous studies, for *S. bulbocastanum*, this was not true. *S. bulbocastanum* tetraploids when the more conservative Willcoxon, Mann-Whitby non parametric test was used, might be explained by the fact that this species is phylogenetically more distant from *S. tuberosum* than *S. commersonii*. Since *S. tuberosum* was the reference genome chosen to design the microarray probes, we might hypothesize that the efficiency of the microarray probes to detect RNA molecules were lower in *S. bulbocastanum* than in *S. commersonii*. Consequently, increasing the statistically significance of the analysis (BLB4X VS. BLB2X) we found a drescrepances of differentially expressed genes in the two species. If we support this idea we should also state that our microarray analysis was able to analyse the expression of fewer genes in *S. bulbocastanum*. Moreover lower percentage of differentially expressed genes were found in previously transcriptome studies. Using *Arabidopsis thaliana* Affymetrix Genechip in a heterologous hybridization study to determinate the gene expression response to autopolyploidization in *Isatis indigotica* Lu *et al*., (2006), found ~4,3% of differentially expressed genes.

In conclusion, once again we found that ploidy induced alterations caused subtle expression changes, but with different modalities, in *S. commersonii* and *S. bulbocastanum* genomes.

When we looked deeper to find ploidy-sensitive genes (either up- or down-regulated), we found significant differences between *S. bulbocastanum* and *S. commersonii* modalities to respond to genome doubling. As for *S. bulbocastanum* polyploids, the expression of most genes increased with ploidy, insomuch as the up-regulated genes were higher than 93% (BLB4X Vs. 2X). This trend was expected following genome doubling, since polyploidy has the general effect of increasing gene expression level in a per cell basis in proportion to the gene dosage conferred by ploidy level. This phenomenon was observed by Guo *et al*., (1996), who found increased expression of most genes directly proportional to increased ploidy levels in an euploid maize series (monoploid, diploid, triploid, and tetraploid). They suggested a gene dosage-

dependent effect to explain such results. Gene regulation in dosage-dependent manner was also reported in *Saccharomyces cerevisiae* where microarrays were used to measured gene regulation in 1X, 2X, 3X and 4X cells (163). As for *S. commersonii* synthetic tetraploids, the up- and down-regulated genes were equally distributed among the differentially expressed genes (45% up-regulated and 55% down-regulated). In this case a gene-dosage deregulation of protein-dcoding gens probably occurred because of perturbations of regulatory networks able to buffer the expression of doubled loci. This opposite behaviour displayed by our wild species, mirrored the results reported by Yu *et al*., (2010) who studied the transcriptome alterations following autopolyploidization in two different *A. thaliana* ecotypes, Columbia and Landsberg. Thus, upon shift from di- to autotetraploidy, Columbia responded with the alteration of gene expression of several hundred genes, whereas Landsberg showed minimal altered gene expression. The authors classified these ecotypes as responder and nonresponder, respectively. Moreover they suggested that Columbia but not Landsberg possesses one or more genetic factors (buffer genes) that are capable of sensing the alteration of genome dosage and inducing gene expression alterations. In our study this concept might be applied not about the sensibility of a genotype to the autopolyploidization process but about how the regulatory transcription network response in a genotype specific way to the genome doubling.

Besides the genome-wide transcriptional response to ploidy change, the main objective of our research was also to look for genes whose ploidy-dependent regulation was common to both species. Our analysis did not show any gene always up- and down-regulated in synthetic tetraploids of both species. However we found three common differentially expressed genes changing in tetraploids but in a species-specific way. In fact they were down-regulated in all *S. commersonii* tetraploids and up-regulated in all *S. bulbocastanum* ones in comparison with the respective diploid parents. For two of them the differentially expression was confirmed by independent qPCR experiments and, as expected, relative expression levels were significantly higher than microarray data. There are several possible causes for changes in gene expression in polyploids. An increase in the copy number of all chromosomes may equally affect all genes and should result in a uniform increase in gene expression. However, it is possible that some genes deviate from this assumption because they respond to regulating factors that do not change proportionally with ploidy. Additionally, polyploidy can change the structural relationship between certain cellular components and might alter the progress of mitosis and meiosis. These effects might modify gene expression through reversible regulation or through persistent epigenetic resetting. In our work the mechanisms beyond the changes in gene

expression after genome doubling deserve further work, but we are aware that only a corresponding epigenetic study will lead us to a deeper knowledge about polyploidization.

Gene Ontology (GO) was constructed for genes with expression patterns affected by autopolyploidization. Ontologies covered almost all important terms of biological processes, molecular functions and cellular compartment. Interestingly, *S. commersonii* and *S. bulbocastanum* shared most of GO groups and showed a common way of response to autopolyploidization, at least for the most represented classes. Of relevance were the terms "response to stress" and "transport" within biological process category, "nucleotide binding", and "protein binding" within molecular function, and "plastid" within cellular compartment class.

Our results provided evidence that "carbohydrate metabolic process" was one of our most represented term in *S. bulbocastanum* tetraploids, as already reported in *A. thaliana* polyploids (164). In the same paper the authors found the enriched GO class "cell wall sugar" as one of the most differentially regulated in the 4x genotypes of *A. thaliana*. GO biological process term. Interestingly, we also found that several structural enzymes of the starch and sucrose metabolic pathways were differentially expressed between tetraploids and diploid of *S. bulbocastanum*. A transcriptome analysis carried out in *Citrus lemon* found the cinnamoyl-CoA reductases gene (*ccr*) up-regulated in response to autopolyploidization (83). Moreover *ccr* play a role in cell wall structure, participates in phenylpropanoid biosynthesis and is key enzyme of lignin biosynthesis (165). As a major polymer of cell walls, lignin provides the mechanical support for plants to prevent lodging (166). Lu *et al*., (2006) found a group of lignin biosynthesis related genes (LBRGs) with higher expression in autotetraploid of *Isatis indigotica*. Higher expression of LBRGs may be responsible for the better anti-lodging ability in autotetraploid *Isatis indigotica*. One of the more intriguing aspects of the differential expression of LBRGs is the question of whether or not it is random or subject to natural selection in autopolyploidy. Guo *et al*., (1996) found the sucrose synthase (EC 2.4.1.13) that showed expression response directly proportional to genome dosage in a maize series (up-regulated in the 4x>3x>2x>1x). We didn't found this enzyme participating in starch and sucrose metabolism, and it is probably due to a well-known limitation of microarrays, which lacks of statistical power to detect relatively small changes, because they were too subtle, masked by noise, or tissue specific. It will be interesting to confirm the ploidy related up-regulation of this gene in a more sensitive technique as qPCR.

The most recurring GO terms among molecular function category found was "nucleotide binding" in tetraploids of both wild potato species, and "signal transduction" only in *S. bulbocastanum* 4x genotypes. Both terms were found in synthetic autotetraploids of *A. thaliana*

(164) and *S. phureja* (81), and only the latter term in autopolyploids of *Isatis indigotica* (162). In the genome-wide comparison of gene responsive to autopolyploidy in *Isatis indigotica*, the authors founds several genes differentially expressed between tetraploid and diploid involved in signal transduction and transcriptional regulation. Most of these signal transduction genes (STGs) with high expression levels are known to encode plant receptorlike kinases (RLKs), which play a key role in plant cell communication with each other and with the environment (167). It is believed that signal transduction pathways converge on TFs, and almost all signal transduction genes are regulated by specific TFs (168). The difference response to genome doubling between *S. bulbocastanum* and *S. commersonii* might be caused by a different regulation of STGs and/or TFs expressions.

We found different histone genes (histone h3, histone methyltransferase and, histone deacetylase 2) differentially expressed between tetraploid and diploid of both species. Therefore, since histnes are involved in transcriptional regulation and chromative structure such as nucleosome assembly, it may be expected that subtle changes in gene expression of histone genes in tetraploids might also lead to slight alterations in genes found differentially expressed in our polyploids. H1 histone gene was found up-regulated in *Citrus* autitetraploids and, as speculated by authors, this probabily contributed to the altered transcriptional and phenotypic changes observed (83).

The involvement of transporter proteins in the polyploidization resulted by the "transporter" GO terms that we found in both species in consistent with proper "trasporter activity" and correlated "ions" categories found in *A. thaliana* transcriptome study (164). Nevertheless we found that a bile acid co-transporter gene expression was ploidy-dependent, since ti was down-regulated in all *S. commersonii* 4x and always up-regulated in *S. bulbocastanum* 4x.

GO term "response to stress" was highly represented in *S. commersonii* and *S. bulbocastanum* tetraploids. This finding was correlated with our metabolome analysis in which we found an accumulation of many metabolites known to have an important role in mediating response to abiotic and biotic stress, such flavonoids (anthocyanins, carotenoids and other phenylpropanoids). The same categories was found also in tetraploid of *A. thaliana* transcriptome study (164) and in tetraploid of *Citrus* (83). Interestingly, in *A. thaliana* the authors along with "stress" term, highlighted a correlated "auxin" category as well. Similarly, in *Citrus* tetraploids was found a higher score at "response to abscisic acid (ABA) stimulus" categories, suggesting a pre-adapted to abiotic constraints as a consequence to polyploidization.

The last GO categories that we would like to discuss were the cellular compartment "plastid" and "mitochondrion" found to be highly represented in the tetraploids of both species. The related categories "phosphorylation" and "light" were found in the study in *Citrus* and *A. thaliana*, respectively (83, 164). In particular Allario and colleagues (2011) found that two genes involved in "phosphorylation", plastocyanin chloroplast precursor (PC) and a Chlorophyll A-B, binding protein (Lhca2) were up-regulated in tetraploids in comparison to the matched parental diploid. Moreover some important changes reported in polyploids, compared with their related diploids, such as increased chloroplast number per cell, and an increase in chlorophyll content (169, 170) may be the result of genome doubling.

The three enriched GO terms ("auxin", "light", "ions") found by Yu *et al*., (2010), not only matched with GO terms "response to stress", "plastid" and "transporter" as abovementioned but also with the KEGG phenylalanine, tyrosine and tryptophan biosynthesis pathway drastically perturbed in *S. commersonni* tetraploids comparison. Our results showed a drastic alteration on aromatic aminoacids metabolic pathway, tryptophan, Tyrosine and Phenylalanine. As for tryptophane, in two recent papers it was revealed that a small family of tryptophan aminotransferases catalyze formation of indole-3-pyruvic acid (IPA) from L-tryptophan (L-Trp), the first step in a pathway for auxin biosynthesis (171, 172). In must be pointed out that tryptophan is also the precursor for alkaloids. Caruso *et al*., (2011) analysed the content of glycoalkaloids (GAs) in the same *S. commersonni* genotyes object of our research. The authors found that the minor glycoalkaloids were higher in the tetraploids than in the diploid progenitor CMM1T whereas the contents of the major GAs were significantly higher in the diploid CMM1T than in the tetraploids. They discussed about the lower leaf content of major GAs in the tetraploids. For these metabolites, there may well be a strong gene balance at the diploid level and gene duplication may have detrimental effects, as recently reviewed by Birchler & Veitia (1981). In the cultivated potato, the edible part is the tuber, whereas in this study, the GA content was investigated at the leaf level. Therefore, since GAs may display human toxicity and our tetraploid genotypes of *S. commersonii* are being used in breeding programs, it will be interesting to determine whether the GA level reduction detected in the leaves also occurs in the tubers of our synthetic *S. commersonii* tetraploids and in hybrid materials.

As for Tyrosine, it is been reported its important role in photosynthesis. Indeed, in chloroplasts (photosystem II), it acts as an electron donor in the reduction of oxidized chlorophyll. In this process, it undergoes deprotonation of its phenolic OH-group. This radical is subsequently reduced in the photosystem II by the four core manganese clusters. In the phD

thesis by Caruso (151), the author aimed to assess the occurrence and extent of genetic, epigenetic, and anatomical changes occurring in the same genotypes subject of our research. It is well known that leaf and cell areas may affect photosynthesis. In the literature, studies on the effects of polyploidy on photosynthetic rates gave contrasting results. In polyploids plants of *Triticum* and *Aegilops*, photosynthetic rates per leaf area decreased or did not change (174). Many authors (175-177) reported either a positive or a negative correlation between leaf thickness and net CO_2 assimilation (ACO_2). Since in both species analyzed in that PhD thesis for lamina and mesophyll thickness generally there were no significant differences between diploid and tetraploid derivatives, ACO_2 could be indirectly unaffected by change in ploidy level. This hypothesis is consistent with preliminary results obtained *in S. bulbocastanum* for the measure of photosynthetic rate (ACO_2) and efficiency, as showed by chlorophyll fluorescence evaluations. In *S. commersonii*, since generally there were not significant differences between diploid and tetraploids in leaf area and since experimental evidences proved that ACO_2 is inversely related to leaf area (178, 179), the author stated that probably photosynthetic rate could result unvaried between diploid and tetraploid genotypes.

As previously stated, in our synthetic tetraplioid of *S. bulbocastanum* and *S. commersonii* we observed a drastic alteration also of Phenylalanine amino acid pathway. It is the starting compound used in the flavonoid biosynthesis and the precursor of Lignin. Indeed, Phenylalanine is converted to cinnamic acid by the enzyme phenylalanine ammonia-lyase (PAL) which activity has been reported to lead to an increased accumulation of soluble phenols, flavonoids (quercetin and kaempferol), and lignin in diploid plants of Chamomile (*Matricaria chamomilla* L.) in comparison to tetraploids (180). PAL was also found up-regulated of 2,5 fold change in autotetraploids of *Isatis indigotica* (162).

Even if we did not found many alterations on phenylpropanoid, flavonoid and anthocyanin gene expression pathways, they were the class of secondary metabolite that showed more changes after polyploidization as evidenced by our metabolome experiments. A major qualitative and quantitative impact was found in *S. commersonii* tetraploids. Beside the fact that not many works studied the relationship between polyploidy and metabolites alterations, our results of the metabolome alterations in the tetraploids of both species were surprising and could change the previous vision emerging from literature. Some alterations on few metabolite classes or single metabolites were previously reported linked to ploidy levels and the general idea, as well as for transcriptome, was that autopolyploidization did not have a great impact on gene expression and metabolome (71, 119, 181). In commenting our data, first of all we want to underline the extension of metabolites changing their accumulation or depauperize in tetraploids

that covered all metabolic classes (with differences within each class). In *S. commersonii,* 39% of metabolites studied showed a different pattern between the two levels of ploidy. By contrast, in *S. bulbocastanum* this percentage decreased up to 13%. The difference in the number of altered metabolites reproduced the difference in gene expression changes between the species. Moving on the details, "Sugar, Polyols and Phosphates" were the most represented metabolite classes in *S. bulbocastanum* tetraploids among the primary metabolism classes. They showed a relationship with the GO data ("carbohydrate metabolic process" include 52 differentially expressed genes). Also Yu and colleagues (2011) found genes responsible for sugar biosynthesis altered in tetraploids of *A. thaliana*. In *Setaria tomentosa* the autotetraploids were found to be superior in starch, total soluble sugars and lipid content (182). A recent work by Caruso *et al*., (2011), on the same clone of *S. commersonii* used here showed that phenylpropanoids were accumulated in tetraploids. Our results obtained with a more sensible and specific technique not only confirm but magnified the fact that flavonoids, anthocyanin, phenylpropanoids and organic acids showed a drastic differentially metabolism between tetraploids and diploid parents of both species. In *S. bulbocastanum* four phenylpropanoids and six organic acids showed an altered metabolism. All the organic acids were accumulated in all tetraploids. By contrast, all the phenylpropanoids were accumulated more in the parental diploids. In *S. commersonii* we detected 21 secondary metabolites and 13 organic acids with changed accumulation in response to polyploidization. Besides the huge changing in secondary metabolite profile, we want to underline the range of this variations. We found metabolites as 3,5-Dimethoxy-4-Hydroxycinnamic Acid, Caffeic Acid or Petunidin-3-O-Rutinoside-5-O-Glucoside strongly accumulated in all the tetraploids and not detectable in the diploid. There were also others with an opposite trend, present in the diploid parents, but not detectable in polyploids. Indeed the accumulation of the total anthocyanin was so dramatic that it was evident even during the extraction protocol. In addition, an altered chemical profile of metabolic products is frequently encountered, which is itself of potential value for qualitative improvement through ploidy manipulation. Levy (1976) found qualitative differences in the glycoflavone profiles of 14 of the 15 autotetraploids induced from cultivars of *Phlox drummondi*. This included 14 instances of additional flavonoids in the autotetraploid that were lacking in the progenitor diploid, and eight where flavonoids were present in the diploid but absent in the autotetraploid. In view of the role of phenylpropanoids in suppressing pathogens and in response to abiotic stress (183, 184), our results are interesting from the breeding standpoint. Indeed, tetraploid genotypes of *S. commersonii* are produced and used in breeding schemes aimed at overcoming sexual barriers (185). Moreover anthocyanins are important also

for human health thanks to their anti-oxidant nature and evidence suggests they may possess analgesic properties in addition to neuroprotective and anti-inflammatory activities (186).

Our data on anthocyanin metabolism lead our research in the field of the genetic and gene regulation of these compounds. Previous studies on the anthocyanin synthesis pathway discovered all the structural enzymes involved and the transcription factors that regulate their activities in potato (148, 187). However, they did not explain how the anthocyanin synthesis is regulated. This is a very interesting double consecutive substitutions in CMM1T respects to *an1 S. tuberosum* reference sequence. In fact this is typical of *S. tuberosum* genotypes which presents high amount of anthocyanin pigments in leave tissues. Our gene expression results for *an1*, *ans* and *dfr* in *S. bulbocastanum* were consistent with the down-accumulation that we found at metabolic level. The same did not hold true for *S. commersonii*. Expect for a faint up-regulation of *an1* transcription factor in CMM24 and *dfr* in CMM15, the general trend was a strong down-regulation of all the genetic players of the anthocyanin biosynthesis pathway. The hypothesis to explain this inconsistency results should lie on two basis. First, the mutations that we found on an1 sequence should prevent its function as transcription factor decreasing the efficiency to bond the promoter DNA sequences of the structural enzymes or other partners of the complex transcription machinery also important in the regulation of the anthocyanin biosynthesis. Second, due to the complexity of the biosynthetic pathways of anthocyanins, including many branches and enzymes, we hypothesized that, after chromosome doubling, an increased number of genes may lead to an increased concentration and activity of some enzymes. More specific experiments may be necessary to explain the accumulation of these important secondary metabolites, to discover new molecules and new allelic variants of the known transcription factors and enzymes unidentified with our experiments.

5. Conclusion

We believe that this work contributed to extend the limited studies on gene expression and much more on metabolite changes associated to autopolyploidization. As far as we know, it is the first time that in more than one species transcriptome followed by metabolome analyses were carried out on independent autotetraploids. We think that four main evidences emerged from this research.

- After polyploidization a substantial variability has been observed for many differentially expressed genes in synthetic tetraploids, confirming the limited number of results availble in different species (81, 83, 139, 162, 164). This marks the fact that in newly formed autopolyploids about 10% of all the transcriptome is subjected to alteration. Most of responses to polyploidization were species- and genotype-dependent. Comparing 4x vs. 2x, it was clear that in *S. commersonii* there was a higher number of significant differences than in *S. bulbocastanum* in gene expression. Interestingly, we found few but consistent genes whose expression changed in a ploidy linked way. It should be pointed out that it is important to demonstrate the existence of these kind of genes not only because they may be functional linked to a molecular model of polyploidization events, but also because they could function as molecular markers for a successful polyploidization event.

- We showed how the modern bioinformatic analyses based on the gene ontology (GO) and on the metabolic pathways mapping (KEGG) could relate transcriptome data to previously study and to our metabolome analyses. Comparing our GO and KEGG data with the one obtained by the work of Yu and colleagues (2010) in a distant species as *A. thaliana*, we identified comparable changing in the sugar and starch metabolism for *S. bulbocastanum* polyploids and in the response to stress related secondary metabolites (e.g. anthocyanins) for *S. commersonii* polyploids. This may well be correlated to the possibility of identifying important metabolic classes that are linked to genome doubling and finally result in useful breeding activities.

- Our results were the first study on wide metabolome alteration in response to polyploidization. The metabolome alterations in the tetraploids of

both species were surprising and could change the previous vision emerging from literature in which only few alterations on a small pool of metabolite classes or single metabolites were previously reported. We provided evidence that a large amount of metabolites changed their accumulation or depauperized in tetraploids (with differences within each class). A major qualitative and quantitative impact was found in *S. commersonii* tetraploids. By contrast, in *S. bulbocastanum* lower but subtle changes occurred in important metabolite classes (glycoalkaloids, sugars). The knowledge that autopolyploidization might alter important metabolic classes, will make the genome doubling of wild diploid potato species not only a way to overcome sexual incompatibility barriers, but also a way to improve the usefulness of this species in breeding programs.

- The last consideration is based on our data on anthocyanin accumulation.

In *S. bulbocastanum* our results on gene expression analysis for *an1*, *ans* and *dfr* were convergent with the down-accumulation found. By contrast, in *S. commersonii* polyploids a drastic accumulation of these important secondary metabolites following chromosome doubling was found in all tetraploids. However, we were not able to explain this phenomena based on the regulation of R2R3-MYB TF known to be a major player in anthocyanin pathway regulation. We believe that answering this question may produce an important scientific finding with practical breeding applications. We think that the answer may be found in the epigenetic changes following autopolyploidization, as several studies are showing that these changes play an important role in polyploidization events.

6. References

1. Becak ML & Becak W (1998) Evolution by polyploidy in Amphibia: new insights. *Cytogenetics and cell genetics* 80(1-4):28-33.

2. Chen ZJ & Ni Z (2006) Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *Bioessays* 28(3):240-252.

3. Müntzing A (1936) The evolutionary significance of autopolyploidy. *Hereditas* 21:263–378.

4. Darlington CD (1937) Recent advances in cytology. *Recent advances in cytology*. (2nd Ed).

5. Grant V (1963) *The origin of adaptations* (Columbia University Press).

6. Lewis W (1980) b. Polyploidy in Angiosperms: Dicotyledons pp. 241-268. *Polyploidy-Biological Relevence. Plenum Press, New York.*

7. Masterson J (1994) Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* 264(5157):421-424.

8. Otto SP & Whitton J (2000) POLYPLOID INCIDENCE AND EVOLUTION. *Annual Review of Genetics* 34(1):401-437.

9. Jaillon O, *et al*. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449(7161):463-467.

10. Tang H, *et al*. (2008) Synteny and collinearity in plant genomes. *Science* 320(5875):486-488.

11. Soltis PS & Soltis DE (2009) The Role of Hybridization in Plant Speciation. *Annual Review of Plant Biology* 60(1):561-588.

12. Bowers JE, Chapman BA, Rong J, & Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422(6930):433-438.

13. Furlong RF & Holland PWH (2004) Polyploidy in vertebrate ancestry: Ohno and beyond. *Biological Journal of the Linnean Society* 82(4):425-430.

14. Soltis PS & Soltis DE (2000) The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences* 97(13):7051-7057.

15. Moore MJ, Bell CD, Soltis PS, & Soltis DE (2007) Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences* 104(49):19363-19368.

16. Velasco R, *et al*. (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PloS one* 2(12):e1326.

17. Soltis DE & Soltis PS (1999) Polyploidy: recurrent formation and genome evolution. *Trends in Ecology & Evolution* 14(9):348-352.

18. Ramsey J & Schemske DW (2002) Neopolyploidy in Flowering Plants. *Annual Review of Ecology and Systematics* 33(ArticleType: research-article / Full publication date: 2002 / Copyright © 2002 Annual Reviews):589-639.

19. Ainouche M, *et al*. (2009) Hybridization, polyploidy and invasion: lessons from *Spartina* (Poaceae). *Biological Invasions* 11(5):1159-1173.

20. Fawcett JA, Maere S, & Van de Peer Y (2009) Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. *Proceedings of the National Academy of Sciences* 106(14):5737-5742.

21. Stebbins Jr C (1950) Variation and evolution in plants. *Variation and evolution in plants*.

22. Muldal S (1952) The chromosomes of the earthworms: I. The evolution of polyploidy. *Heredity* 6(1):56-76.

23. Abbott RJ & Brochmann C (2003) History and evolution of the arctic flora: in the footsteps of Eric Hultén. *Molecular Ecology* 12(2):299-313.

24. Brockmann C, Brysting A.K., Alsos, I.G., Borgen, L., Grundt, H.H., Scheen, A.C., Elven. R., (2004) Polyploidy in arctic plants. *Biol. J. Linn. Soc*. 82:521-536.

25. Hegarty M & Hiscock S (2007) Polyploidy: Doubling up for evolutionary success. *Current Biology* 17(21):R927-R929.

26. Saleh B, Allario T, Dambier D, Ollitrault P, & Morillon R (2008) Tetraploid citrus rootstocks are more tolerant to salt stress than diploid. *Comptes rendus biologies* 331(9):703-710.

27. Bennett MD (2004) Perspectives on polyploidy in plants – ancient and neo. *BIOL J LINN SOC* 82(4):411-423.

28. Aversano R, *et al*. (2012) Molecular tools for exploring polyploid genomes in plants. *International journal of molecular sciences* 13(8):10316-10335.

29. Lutz AM (1907) A PRELIMINARY NOTE ON THE CHROMOSOMES OF OEligNOTHERA LAMARCKIANA AND ONE OF ITS MUTANTS, O. GIGAS. *Science* 26(657):151-152.

30. Gates RR (1909) The stature and chromosomes of Oenothera gigas. *De Vries. Arch. f. Zellforschung* 3:525–552.

31. Kuwada Y (1911) Meiosis in the pollen mother cells of Zea Mays L. *Bot Mag* 25:163-181.

32. Winkler H (1916) *Über die experimentelle Erzeugung von Pflanzen mit abweichenden Chromosomenzahlen* (Fischer).

33. Carputo D, *et al*. (2010) Glycoalkaloid Profile in Potato Haploids Derived from Solanum tuberosum–S. bulbocastanum Somatic Hybrids. *Chemistry & Biodiversity* 7(8):1885-1892.

34. Bretagnolle F & Lumaret R (1995) Bilateral polyploidization in Dactylis glomerata L. subsp. lusitanica: occurrence, morphological and genetic characteristics of first polyploids. *Euphytica* 84(3):197-207.

35. Ownbey M (1950) Natural hybridization and amphiploidy in the genus Tragopogon. *American Journal of Botany*:487-499.

36. Harlan JR (1975) On Ö. Winge and a prayer: the origins of polyploidy. *The Botanical Review* 41(4):361-390.

37. deWet JM (1979) Origins of polyploids. *Basic life sciences* 13:3-15.

38. Bretagnolle F & Thompson JD (1995) Tansley Review No. 78. Gametes with the Stomatic Chromosome Number: Mechanisms of Their Formation and Role in the Evolution of Autopolypoid Plants. *New Phytologist* 129(1):1-22.

39. Sato T, Maceira N, Lumaret R, & Jacquard P (1993) Flowering characteristics and fertility of interploidy progeny from normal and 2n gametes in Dactylis glomerata L. *New Phytologist* 124(2):309-319.

40. Mok DWSP, S.J. (1975) Three mechanisms of 2n pollen formation in diploid potatoes. *Can J Genet Cytol* 17:217-225.

41. Werner JE, Peloquin, S.J., (1990) Inheritance of two mechanism of 2n egg formation in 2x potatoes. *J. Hered*. 81:371-374.

42. NiJs TPMD, PELOQUIN, S. J., (1977) 2n Gametes in potato species and their function in sexual polyploidization. *Euphytica* 26:585-600.

43. Stelly D, Peloquin, SJ, Palmer, RG, Crane, CF (1984) Mayer's hemalum-methyl salicylate: a stain-clearing technique for observations within whole ovules. *Stain Techno* 159:155-161.

44. Clausen J, Keck DD, & Hiesey WM (1945) Experimental studies on the nature of species. 2. Plant evolution through amphiploidy and autoploidy, with examples from the Madiinae. *Experimental studies on the nature of species. 2. Plant evolution through amphiploidy and autoploidy, with examples from the Madiinae*.

45. Stebbins GL & Price HJ (1971) THE DEVELOPMENTAL GENETICS OF THE CALCAROIDES GENE IN BARLEY. I. DIVERGENT EXPRESSION AT THE MORPHOLOGICAL AND HISTOLOGICAL LEVEL. *Genetics* 68(4):527-538.

46. Mendiburu AO & Peloquin S (1977) The significance of 2n gametes in potato breeding. *Theoretical and Applied Genetics* 49(2):53-61.

47. Dunbier M & Bingham E (1975) Maximum Heterozygosity in Alfalfa: Results Using Haploid-derived Autoteraploids. *Crop Science* 15(4):527-531.

48. Köhler C, Mittelsten Scheid O, & Erilova A (2010) The impact of the triploid block on the origin and evolution of polyploid plants. *Trends in Genetics* 26(3):142-148.

49. Carputo D, Frusciante L, & Peloquin SJ (2003) The Role of 2n Gametes and Endosperm Balance Number in the Origin and Evolution of Polyploids in the Tuber-Bearing Solanums. *Genetics* 163(1):287-294.

50. Jackson S & Chen ZJ (2010) Genomic and expression plasticity of polyploidy. *Current opinion in plant biology* 13(2):153-159.

51. Hufton AL & Panopoulou G (2009) Polyploidy and genome restructuring: a variety of outcomes. *Current opinion in genetics & development* 19(6):600-606.

52. Song K, Lu P, Tang K, & Osborn TC (1995) Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. *Proceedings of the National Academy of Sciences* 92(17):7719-7723.

53. Lukens LN, *et al*. (2006) Patterns of Sequence Loss and Cytosine Methylation within a Population of Newly Resynthesized Brassica napus Allopolyploids. *Plant Physiology* 140(1):336-348.

54. Ozkan H, Levy AA, & Feldman M (2001) Allopolyploidy-Induced Rapid Genome Evolution in the Wheat (Aegilops–Triticum) Group. *The Plant Cell Online* 13(8):1735-1747.

55. Shaked H, Kashkush K, Ozkan H, Feldman M, & Levy AA (2001) Sequence Elimination and Cytosine Methylation Are Rapid and Reproducible Responses of the Genome to Wide Hybridization and Allopolyploidy in Wheat. *The Plant Cell Online* 13(8):1749-1759.

56. Han F, Fedak G, Guo W, & Liu B (2005) Rapid and Repeatable Elimination of a Parental Genome-Specific DNA Repeat (pGc1R-1a) in Newly Synthesized Wheat Allopolyploids. *Genetics* 170(3):1239-1245.

57. Eilam T, *et al*. (2007) Genome size and genome evolution in diploid Triticeae species. *Genome* 50(11):1029-1037.

58. Eilam T, Anikster Y, Millet E, Manisterski J, & Feldman M (2008) Nuclear DNA amount and genome downsizing in natural and synthetic allopolyploids of the genera Aegilops and Triticum. *Genome* 51(8):616-627.

59. Gustafson JP & Bennett MD (1976) Preferential Selection for Wheat-rye Substitutions in 42-Chromosome Triticale1. *Crop Sci*. 16(5):688-693.

60. Bento M, *et al*. (2008) Polyploidization as a Retraction Force in Plant Genome Evolution: Sequence Rearrangements in Triticale. *PLoS ONE* 3(1):e1402.

61. Ma XF & Gustafson JP (2005) Genome evolution of allopolyploids: a process of cytological and genetic diploidization. *Cytogenetic and genome research* 109(1-3):236-249.

62. Ma XF & Gustafson JP (2006) Timing and rate of genome variation in triticale following allopolyploidization. *Genome* 49(8):950-958.

63. Ma XF & Gustafson JP (2008) Allopolyploidization-accommodated genomic sequence changes in triticale. *Ann Bot* 101(6):825-832.

64. Levy AA & Feldman M (2004) Genetic and epigenetic reprogramming of the wheat genome upon allopolyploidization. *Biological Journal of the Linnean Society* 82(4):607-613.

65. Lim KY, *et al*. (2004) Dynamic Changes in the Distribution of a Satellite Homologous to Intergenic 26-18S rDNA Spacer in the Evolution of Nicotiana. *Genetics* 166(4):1935-1946.

66. Pontes O, *et al*. (2004) Chromosomal locus rearrangements are a rapid response to formation of the allotetraploid Arabidopsis suecica genome. *Proceedings of the National Academy of Sciences* 101(52):18240-18245.

67. Liu B, Brubaker CL, Mergeai G, Cronn RC, & Wendel JF (2001) Polyploid formation in cotton is not accompanied by rapid genomic changes. *Genome* 44(3):321-330.

68. Salmon A, Ainouche ML, & Wendel JF (2005) Genetic and epigenetic consequences of recent hybridization and polyploidy in Spartina (Poaceae). *Molecular Ecology* 14(4):1163-1175.

69. Ainouche ML, Baumel A, Salmon A, & Yannic G (2004) Hybridization, polyploidy and speciation in Spartina (Poaceae). *New Phytologist* 161(1):165-172.

70. Soltis DE, *et al*. (2004) Recent and recurrent polyploidy in Tragopogon (Asteraceae): cytogenetic, genomic and genetic comparisons. *Biological Journal of the Linnean Society* 82(4):485-501.

71. Comai L (2005) The advantages and disadvantages of being polyploid. *Nature reviews. Genetics* 6(11):836-846.

72. Parisod C, Holderegger R, & Brochmann C (2010) Evolutionary consequences of autopolyploidy. *The New phytologist* 186(1):5-17.

73. Raina S, *et al*. (1994) Associated chromosomal DNA changes in polyploids. *Genome* 37(4):560-564.

74. Gustafson P, *et al*. (2009) Genome size in natural and synthetic autopolyploids and in a natural segmental allopolyploid of several Triticeae species. *Genome* 52(3):275-285.

75. Martelotto LG, *et al*. (2007) Genome rearrangements derived from autopolyploidization in< i> Paspalum</i> sp. *Plant Science* 172(5):970-977.

76. König C, Műllner, A.N. (2005) Phylogeographic investigation in Biscutella L. (Brassicaceae) based on morphological, genome size and molecular (ITS) data. *Vienna, Austria: XVII International Botanical Congress*:400.

77. Weiss-Schneeweiss H, Schneeweiss, GM, Stuessy, TF, Mabuchi, T, Park, JM, Jang, CG, Sun, BY (2007) Chromosomal stasis in diploids contrasts with genome restructuring in auto- and allopolyploid taxa of Hepatica (Ranunculaceae). *New Phytologist* 174:669–682.

78. Weiss H, Maluszynska, J. (2000) Chromosomal rearrangement in autotetraploid plants of Arabidopsis thaliana. *Hereditas* 133:255–261.

79. Santos JL, *et al*. (2003) Partial Diploidization of Meiosis in Autotetraploid Arabidopsis thaliana. *Genetics* 165(3):1533-1540.

80. Ozkan H, Tuna M, & Galbraith DW (2006) No DNA loss in autotetraploids of Arabidopsis thaliana. *Plant Breeding* 125(3):288-291.

81. Stupar RM, *et al*. (2007) Phenotypic and Transcriptomic Changes Associated With Potato Autopolyploidization. *Genetics* 176(4):2055-2067.

82. Aversano R, *et al*. (2013) Stochastic changes affect Solanum wild species following autopolyploidization. *Journal of Experimental Botany* 64(2):625-635.

83. Allario T, *et al*. (2011) Large changes in anatomy and physiology between diploid Rangpur lime (Citrus limonia) and its autotetraploid are not associated with large changes in leaf gene expression. *Journal of Experimental Botany* 62(8):2507-2519.

84. Lauria M & Rossi V (2011) Epigenetic control of gene regulation in plants. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1809(8):369-378.

85. Vaughn MW, *et al*. (2007) Epigenetic natural variation in Arabidopsis thaliana. *PLoS biology* 5(7):e174.

86. Bossdorf O, Richards CL, & Pigliucci M (2008) Epigenetics for ecologists. *Ecology Letters* 11(2):106-115.

87. Richards EJ (2006) Opinion - Inherited epigenetic variation - revisiting soft inheritance. *Nat. Rev. Genet*. 7(5):395-U392.

88. Zhai J, *et al*. (2008) Small RNA-directed epigenetic natural variation in Arabidopsis thaliana. *PLoS genetics* 4(4):e1000056.

89. Rapp RA & Wendel JF (2005) Epigenetics and plant evolution. *New Phytologist* 168(1):81-91.

90. Liu B & Wendel JF (2003) Epigenetic phenomena and the evolution of plant allopolyploids. *Molecular phylogenetics and evolution* 29(3):365-379.

91. Madlung A, *et al*. (2002) Remodeling of DNA Methylation and Phenotypic and Transcriptional Changes in Synthetic Arabidopsis Allotetraploids. *Plant Physiology* 129(2):733-746.

92. Chen ZJ (2007) Genetic and Epigenetic Mechanisms for Gene Expression and Phenotypic Variation in Plant Polyploids. *Annual Review of Plant Biology* 58(1):377-406.

93. Shcherban A, Badaeva E, Amosova A, Adonina I, & Salina E (2008) Genetic and epigenetic changes of rDNA in a synthetic allotetraploid, Aegilops sharonensis× Ae. umbellulata. *Genome* 51(4):261-271.

94. Teyssier E, *et al*. (2008) Tissue dependent variations of DNA methylation and endoreduplication levels during tomato fruit development and ripening. *Planta* 228(3):391-399.

95. Fulneček J, Matyášek R, & Kovařík A (2009) Faithful inheritance of cytosine methylation patterns in repeated sequences of the allotetraploid tobacco correlates with the expression of DNA methyltransferase gene families from both parental genomes. *Molecular Genetics and Genomics* 281(4):407-420.

96. Madlung A, *et al*. (2005) Genomic changes in synthetic Arabidopsis polyploids. *The Plant Journal* 41(2):221-230.

97. Kashkush K, Feldman M, & Levy AA (2003) Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature genetics* 33(1):102-106.

98. Hegarty MJ, *et al*. (2005) Development of anonymous cDNA microarrays to study changes to the Senecio floral transcriptome during hybrid speciation. *Mol Ecol* 14(8):2493-2510.

99. Wang Z, Gerstein M, & Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics* 10(1):57-63.

100. Saze H, Scheid OM, & Paszkowski J (2003) Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nature genetics* 34(1):65-69.

101. Scheid OM, Afsar K, & Paszkowski J (1998) Release of epigenetic gene silencing by trans-acting mutations in Arabidopsis. *Proceedings of the National Academy of Sciences* 95(2):632-637.

102. Baubec T, *et al*. (2010) Cooperation of multiple chromatin modifications can generate unanticipated stability of epigenetic states in Arabidopsis. *The Plant Cell Online* 22(1):34-47.

103. Law JA & Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet*. 11(3):204-220.

104. Zhang M, *et al*. (2007) Endosperm-specific hypomethylation, and meiotic inheritance and variation of DNA methylation level and pattern in sorghum (Sorghum bicolor L.) inter-strain hybrids. *Theoretical and Applied Genetics* 115(2):195-207.

105. Ochogavía AC, Cervigni G, Selva JP, Echenique VC, & Pessino SC (2009) Variation in cytosine methylation patterns during ploidy level conversions in Eragrostis curvula. *Plant molecular biology* 70(1-2):17-29.

106. Lavania UC, *et al*. (2012) Autopolyploidy differentially influences body size in plants, but facilitates enhanced accumulation of secondary metabolites, causing increased cytosine methylation. *The Plant Journal* 71(4):539-549.

107. Alston R, Rösler H, Naifeh K, & Mabry T (1965) Hybrid compounds in natural interspecific hybrids. *Proceedings of the National Academy of Sciences of the United States of America* 54(5):1458.

108. Fahselt D & Ownbey M (1968) Chromatographic comparison of Dicentra species and hybrids. *American Journal of Botany*:334-345.

109. Levy M & Levin DA (1975) The novel flavonoid chemistry and phylogenetic origin of Phlox floridana. *Evolution*:487-499.

110. Mears JA (1980) Chemistry of polyploids: a summary with comments on Parthenium (Asteraceae-Ambrosiinae). *Polyploidy*, (Springer), pp 77-101.

111. Schwartz D & Laughner WJ (1969) A molecular basis for heterosis. *Science* 166(3905):626-627.

112. Barber H (1970) Hybridization and the evolution of plants. *Taxon*:154-160.

113. Owuor JBO (1985) lnterspecific hybridization between Coffea arabica L. and tetraploid C. canephora P. EXFr. I1. Meiosis in F1 hybrids and back crosses to C. arabica. *Euphytica* 34:355-360.

114. Khosla M, *et al*. (1990) Polyploidy breeding in Ocimum for evolving high yielding, better quality strains of essential oil importance. *11th International congress of essential oils, fragrances and flavours, New Delhi, India, 12-16 November 1989. Vol. 3. Biosciences*., (Aspect Publishing), pp 75-80.

115. Lou Y & Baldwin IT (2003) Manduca sexta recognition and resistance among allopolyploid Nicotiana host plants. *Proceedings of the National Academy of Sciences of the United States of America* 100(Suppl 2):14581-14586.

116. Pearse IS, Krügel T, & Baldwin IT (2006) Innovation in anti-herbivore defense systems during neopolypoloidy – the functional consequences of instantaneous speciation. *The Plant Journal* 47(2):196-210.

117. Wang J, *et al*. (2006) Genomewide Nonadditive Gene Regulation in Arabidopsis Allotetraploids. *Genetics* 172(1):507-517.

118. Levin DA (2002) *The role of chromosomal change in plant evolution* (Oxford University Press on Demand).

119. Dhawan OP & Lavania UC (1996) Enhancing the productivity of secondary metabolites via induced polyploidy: a review. *Euphytica* 87(2):81-89.

120. Lavania U (1988) Development of fertile autotetraploid strain in Hyoscyamus muticus L. *Trop. Agric* 65:277-278.

121. Manwell C & Baker CA (1970) Molecular biology and the origin of species: heterosis, protein polymorphism and animal breeding. *Molecular biology and the origin of species: heterosis, protein polymorphism and animal breeding*.

122. Caruso I, *et al*. (2011) Secondary Metabolite Profile in Induced Tetraploids of Wild Solanum commersoniiDun. *Chemistry & Biodiversity* 8(12):2226-2237.

123. Adams KL, Cronn R, Percifield R, & Wendel JF (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of Sciences* 100(8):4649-4654.

124. Wang J, *et al*. (2004) Stochastic and epigenetic changes of gene expression in Arabidopsis polyploids. *Genetics* 167(4):1961-1973.

125. He P, Friebe BR, Gill BS, & Zhou J-M (2003) Allopolyploidy alters gene expression in the highly stable hexaploid wheat. *Plant molecular biology* 52(2):401-414.

126. Chen ZJ & Pikaard CS (1997) Epigenetic silencing of RNA polymerase I transcription: a role for DNA methylation and histone modification in nucleolar dominance. *Genes & development* 11(16):2124-2136.

127. Chagué V, *et al*. (2010) Genome-wide gene expression changes in genetically stable synthetic and natural wheat allohexaploids. *New Phytologist* 187(4):1181-1194.

128. Pumphrey M, Bai J, Laudencia-Chingcuanco D, Anderson O, & Gill BS (2009) Nonadditive Expression of Homoeologous Genes Is Established Upon Polyploidization in Hexaploid Wheat. *Genetics* 181(3):1147-1157.

129. Chelaifa H, Monnier A, & Ainouche M (2010) Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species Spartina × townsendii and Spartina anglica (Poaceae). *New Phytologist* 186(1):161-174.

130. Marmagne A, Brabant P, Thiellement H, & Alix K (2010) Analysis of gene expression in resynthesized Brassica napus allotetraploids: transcriptional changes do not explain differential protein regulation. *New Phytologist* 186(1):216-227.

131. Buggs RJA, *et al*. (2009) Gene loss and silencing in Tragopogon miscellus (Asteraceae): comparison of natural and synthetic allotetraploids. *Heredity* 103(1):73-81.

132. Hegarty MJ, *et al*. (2006) Transcriptome Shock after Interspecific Hybridization in Senecio Is Ameliorated by Genome Duplication. *Current Biology* 16(16):1652-1659.

133. Flagel L, Udall J, Nettleton D, & Wendel J (2008) Duplicate gene expression in allopolyploid Gossypium reveals two temporally distinct phases of expression evolution. *BMC Biology* 6(1):16.

134. Albertin W, *et al*. (2006) Numerous and Rapid Nonstochastic Modifications of Gene Products in Newly Synthesized Brassica napus Allotetraploids. *Genetics* 173(2):1101-1113.

135. Guo M, Davis D, & Birchler JA (1996) Dosage Effects on Gene Expression in a Maize Ploidy Series. *Genetics* 142(4):1349-1355.

136. Lu BB, *et al*. (2006) Cloning and Characterization of a Differentially Expressed Phenylalanine Ammonialyase Gene (IiPAL) After Genome Duplication from Tetraploid Isatis indigotica Fort. *Journal of Integrative Plant Biology* 48(12):1439-1449.

137. Riddle NC, Kato A, & Birchler JA (2006) Genetic variation for the response to ploidy change in Zea mays L. *Theoretical and Applied Genetics* 114(1):101-111.

138. Riddle N, Jiang H, An L, Doerge R, & Birchler J (2010) Gene expression analysis at the intersection of ploidy and hybridity in maize. *TAG Theoretical and Applied Genetics* 120(2):341-353.

139. Pignatta D, *et al*. (2010) Differential sensitivity of the Arabidopsis thaliana transcriptome and enhancers to the effects of genome doubling. *New Phytologist* 186(1):194-206.

140. Palta J, ; Simon, G. (1992) Breeding potential for improvement of freezing stress resistance, genetic separation of freezing tolerance, freezing avoidance and capacity to cold acclimate. *Advances in Plant Cold Hardiness*, ed Li PC, L. (CRC Press, Boca Raton), pp 299-310.

141. Song J, *et al*. (2003) Gene RB cloned from Solanum bulbocastanum confers broad spectrum resistance to potato late blight. *Proceedings of the National Academy of Sciences* 100(16):9128-9133.

142. Tabata S (2002) Impact of genomics approaches on plant genetics and physiology. *J Plant Res* 115(4):271-275.

143. Vij S & Tyagi AK (2007) Emerging trends in the functional genomics of the abiotic stress response in crop plants. *Plant Biotechnology Journal* 5(3):361-380.

144. Butelli E, *et al*. (2012) Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *The Plant Cell Online* 24(3):1242-1255.

145. Zhang B, *et al*. (2012) A putative functional MYB transcription factor induced by low temperature regulates anthocyanin biosynthesis in purple kale (Brassica Oleracea var. acephala f. tricolor). *Plant Cell Rep* 31(2):281-289.

146. Cedroni M, Cronn R, Adams K, Wilkins T, & Wendel J (2003) Evolution and expression of MYB genes in diploid and polyploid cotton. *Plant molecular biology* 51(3):313-325.

147. Yanhui C, *et al*. (2006) The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. *Plant molecular biology* 60(1):107-124.

148. Jung C, *et al*. (2009) The potato developer (D) locus encodes an R2R3 MYB transcription factor that regulates expression of multiple anthocyanin structural genes in tuber skin. *Theoretical and Applied Genetics* 120(1):45-57.

149. Carputo D, Garreffa P, Mazzei M, Monti L, & Cardi T (1998) Fertility of somatic hybrids Solanum commersonii (2 x, 1EBN)(+) S. tuberosum haploid (2 x, 2EBN) in intra-and inter-EBN crosses. *Genome* 41(6):776-781.

150. Murashige T & Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol Plantarum* 15(3):473-497.

151. Caruso I (2010) Phenotypic, metabolic and genetic consequences of polyploidization in Solanum species. Environmental Resources Sciences Ph.D. Dissertation (University of Naples Federico II).

152. Langfelder P & Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9(1):559.

153. Zhang B & Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology* 4(1):1128.

154. Cline MS, *et al*. (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat. Protocols* 2(10):2366-2382.

155. Assenov Y, Ramírez F, Schelhorn S-E, Lengauer T, & Albrecht M (2008) Computing topological parameters of biological networks. *Bioinformatics* 24(2):282-284.

156. Diretto G, *et al*. (2007) Metabolic Engineering of Potato Carotenoid Content through Tuber-Specific Overexpression of a Bacterial Mini-Pathway. *PLoS ONE* 2(4):e350.

157. Porra R, Thompson, WA, Kriedermann, PE (1989) Determination of accurate extinction coefficients and simultaneous equation for assaying chlorophyll a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochim Biophys Acta* 975:384–394.

158. Lin-Wang K, *et al*. (2010) An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae. *BMC Plant Biology* 10(1):50.

159. Jiao Y, *et al*. (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature* 473(7345):97-100.

160. Wendel JF (2000) Genome evolution in polyploids. *Plant molecular biology* 42(1):225-249.

161. McClintock B (1984) The significance of responses of the genome to challenge. *Science* 226(4676):792-801.

162. Lu B, *et al*. (2006) A genome-wide comparison of genes responsive to autopolyploidy inIsatis indigotica usingArabidopsis thaliana Affymetrix genechips. *Plant Molecular Biology Reporter* 24(2):197-204.

163. Galitski T, Saldanha AJ, Styles CA, Lander ES, & Fink GR (1999) Ploidy regulation of gene expression. *Science* 285(5425):251-254.

164. Yu Z, *et al*. (2010) Impact of natural genetic variation on the transcriptome of autotetraploid Arabidopsis thaliana. *Proceedings of the National Academy of Sciences* 107(41):17809-17814.

165. Wadenbäck J, *et al*. (2008) Lignin biosynthesis in transgenic Norway spruce plants harboring an antisense construct for cinnamoyl CoA reductase (CCR). *Transgenic Research* 17(3):379-392.

166. Humphreys JM & Chapple C (2002) Rewriting the lignin roadmap. *Current Opinion in Plant Biology* 5(3):224-229.

167. Morris ER & Walker JC (2003) Receptor-like protein kinases: the keys to response. *Current Opinion in Plant Biology* 6(4):339-342.

168. Shao W, *et al*. (2005) Parallel Profiling of Active Transcription Factors Using an Oligonucleotide Array-Based Transcription Factor Assay (OATFA). *Journal of Proteome Research* 4(4):1451-1456.

169. Jellings AJ & Leech RM (1984) Anatomical Variation in First Leaves of Nine Triticum Genotypes, and its Relationship to Photosynthetic Capacity. *New Phytologist* 96(3):371-382.

170. Mathura S, Fossey A, & Beck SL (2006) Comparative study of chlorophyll content in diploid and tetraploid black wattle (Acacia mearnsii). *Forestry* 79(4):381-388.

171. Tao Y, *et al*. (2008) Rapid Synthesis of Auxin via a New Tryptophan-Dependent Pathway Is Required for Shade Avoidance in Plants. *Cell* 133(1):164-176.

172. Stepanova AN, *et al*. (2008) TAA1-Mediated Auxin Biosynthesis Is Essential for Hormone Crosstalk and Plant Development. *Cell* 133(1):177-191.

173. Birchler JA & Newton KJ (1981) MODULATION OF PROTEIN LEVELS IN CHROMOSOMAL DOSAGE SERIES OF MAIZE: THE BIOCHEMICAL BASIS OF ANEUPLOID SYNDROMES. *Genetics* 99(2):247-266.

174. Warner DA & Edwards GE (1989) Effects of Polyploidy on Photosynthetic Rates, Photosynthetic Enzymes, Contents of DNA, Chlorophyll, and Sizes and Numbers of Photosynthetic Cells in the C4 Dicot Atriplex confertifolia. *Plant Physiology* 91(3):1143-1151.

175. El-Sharkawy M & Hesketh J (1965) Photosynthesis among species in relation to characteristics of leaf anatomy and CO2 diffusion resistances. *Crop Science* 5(6):517-521.

176. Delaney RH & Dobrenz AK (1974) Morphological and Anatomical Features of Alfalfa Leaves as Related to CO2 Exchange1. *Crop Sci*. 14(3):444-447.

177. Dornhoff GM & Shibles R (1976) Leaf Morphology and Anatomy in Relation to CO2-exchange Rate of Soybean Leaves1. *Crop Sci*. 16(3):377-381.

178. AUSTIN RB, MORGAN CL, FORD MA, & BHAGWAT SG (1982) Flag Leaf Photosynthesis of Triticum aestivum and Related Diploid and Tetraploid Species. *Annals of Botany* 49(2):177-189.

179. Bhagsari AS & Brown RH (1986) Leaf Photosynthesis and its Correlation with Leaf Area1. *Crop Sci*. 26(1):127-132.

180. Kováčik J , Klejdus Bi , Hedbavny J , & Zoń J (2010) Copper Uptake Is Differentially Modulated by Phenylalanine Ammonia-lyase Inhibition in Diploid and Tetraploid Chamomile. *Journal of Agricultural and Food Chemistry* 58(18):10270-10276.

181. Madlung A (2013) Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity* 110(2):99-104.

182. Bala S & Sachdeva SK (1990) Biochemical variability in cytotypes of Setaria tomentosa (Roxb.) Kunth. *Acta Botanica Indica* 18(2):216-221.

183. Wegener C & Jansen G (2007) Soft-rot Resistance of Coloured Potato Cultivars (Solanum tuberosum L.): The Role of Anthocyanins. *Potato Res*. 50(1):31-44.

184. Lorenc-Kukula K, Jafra S, Oszmianski J, & Szopa J (2004) Ectopic Expression of Anthocyanin 5-O-Glucosyltransferase in Potato Tuber Causes Increased Resistance to Bacteria. *Journal of Agricultural and Food Chemistry* 53(2):272-281.

185. Carputo D, *et al*. (1997) Endosperm balance number manipulation for direct in vivo germplasm introgression to potato from a sexually isolated relative (Solanum commersonii Dun.). *Proceedings of the National Academy of Sciences* 94(22):12013-12017.

186. Korte G, *et al*. (2009) An examination of anthocyanins' and anthocyanidins' affinity for cannabinoid receptors. *Journal of medicinal food* 12(6):1407-1410.

187. Zhang Y, Jung C, & De Jong W (2009) Genetic analysis of pigmented tuber flesh in potato. *Theoretical and Applied Genetics* 119(1):143-150.