

RESEARCH

Open Access



Polyploidization-driven transcriptomic dynamics in *Medicago sativa* neotetraploids: mRNA, smRNA and allele-specific gene expression

D. F. Santoro¹, G. Marconi^{1,2}, S. Capomaccio^{2,3}, M. Bocchini¹, A. W. Anderson¹, A. Finotti^{2,4}, M. Confalonieri⁵, E. Albertini^{1,2} and D. Rosellini^{1,2*}

Abstract

Whole genome duplication (WGD) is a powerful evolutionary mechanism in plants. Autopolyploids have been comparatively less studied than allopolyploids, with sexual autopolyploidization receiving even less attention. In this work, we studied the transcriptomes of neotetraploids ($2n = 4x = 32$) obtained by crossing two diploid ($2n = 2x = 16$) plants of *Medicago sativa* that produce a significant percentage of either $2n$ eggs or pollen. Diploid progeny from the same cross allowed us to separate the transcriptional outcomes of hybridization from those of WGD. This material can help to elucidate events at the base of the domestication of cultivated $4x$ alfalfa, the world's most important leguminous forage. Three $2x$ and three $4x$ progeny plants and $2x$ parental plants were used for this study. The RNA-seq data revealed that WGD did not dramatically affect the transcription of leaf protein-coding genes. The two parental genotypes did not contribute equally to the progeny transcriptomes, and genome-wide expression level dominance of the male parent was observed. A large majority of the genes whose expression level changed due to WGD presented increased expression, indicating that the $4x$ state requires the upregulation of approximately 2.66% of the protein-coding genes. Overall, we estimated that 3.63% of the protein-coding genes were transcriptionally affected by WGD and may contribute to the phenotypic novelty of the neotetraploid plants. Pathway analysis suggested that WGD could affect secondary metabolite biosynthesis, which in turn may influence forage quality. We found four times as many transcription factor genes among the polyploidization-affected genes than among those affected only by hybridization. Several of these belong to classes involved in stress response. Small RNA-seq revealed that very few miRNAs were significantly associated with WGD, but they target several hundred genes, and their role in the WGD response may be relevant. Integrated network analysis led to the identification of putative miRNA: mRNA interactions potentially involved in transcriptome reprogramming. Allele-specific expression analysis indicated that parent-of-origin bias was not a significant outcome of WGD, but we found that parentally biased RNA editing may be a significant source of variation in neopolyploids.

Keywords Alfalfa, Expression level dominance, Gene coexpression network, RNA-Seq, Sexual polyploidization, Small RNA, Whole genome duplication

*Correspondence:

D. Rosellini
daniele.rosellini@unipg.it

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Polyploidy is the presence of more than two complete sets of chromosomes within a nucleus. Polyploids are generally classified into two main categories: autopolyploids, arising from intraspecific whole genome duplication (WGD), and allopolyploids, deriving from the merging of the genomes of two (or more) species. More realistically, merging genomes creates a continuum based on how closely related the original genotypes are, from closely related genotypes creating polyploids containing sets of completely homologous chromosomes to polyploids containing sets of highly divergent homoeologous chromosomes.

Alterations in meiosis are considered the main path to polyploidy in natural populations through a process known as sexual polyploidization, which occurs when one or both gametes involved in fertilization are not reduced [1]. In particular, bilateral sexual polyploidization results from the fusion of a $2n$ egg cell and a $2n$ sperm nucleus, resulting in $4x$ offspring, which may initiate the establishment of a neopolyploid lineage [2]. In plants, polyploidy has long been recognized as a major force for evolution and speciation as well as a driver of genomic novelty and plasticity [3–7]. Barker et al. [8] estimated that among 4003 plant species of 47 genera, 11% are allopolyploid, and 13% are autopolyploid; moreover, most, if not all, extant diploid plant species are paleopolyploid, that is, they derive from the diploidization of ancient polyploids [9]. WGD probably favoured the domestication of agricultural plants [10]. About 50% of economically important crop species are polyploid, including alfalfa, banana, cotton, coffee, oilseed rape, potato, sweet potato, sugarcane, tobacco, and wheat [11, 12] emphasizing the importance of polyploid research for agriculture. In crop plants, the improvement of economically important traits by artificial polyploidy has long been investigated. There are examples of agricultural success of induced autopolyploids (banana, clover, ryegrass, sugar beet) and allopolyploids (festulolium, triticale).

In allopolyploids, combining different homoeoalleles from parental species can lead to new allelic interactions contributing to intergenomic heterozygosity and heterosis, which in turn can improve crop plasticity under environmental fluctuations and stressful conditions [13–15]. In autopolyploids, the possible presence of more than two alleles per locus can allow for higher-order gene interactions [16]. In $4x$ alfalfa, the positive effects of multiple alleles on biomass yield have been clearly demonstrated [17]. These allelic interactions are considered the basis for “progressive heterosis”, that is, higher vigour of double crosses $(AxB) \times (CxD)$ with respect to single crosses AxB or CxD , as demonstrated in potato, alfalfa and maize [18–20]. Consistent with this hypothesis,

autopolyploids from somatic WGD (sometimes termed colchiploids because they are most frequently obtained by colchicine treatments) may not be more vigorous in terms of biomass and are sometimes less vigorous than the original diploids [21–23]. For example, in *Solanum commersonii* and *S. bulbocastanum*, synthetic autotetraploids showed significant gene expression alterations and morpho-anatomical changes, but these changes did not consistently result in phenotypic superiority over diploids, suggesting variable effects of polyploidy depending on the species and genotype [24]. Recently, a thorough phenotypic characterization of the effects of somatic WGD on *Arabidopsis* plants at $2x$, $4x$, $6x$, and $8x$ ploidies showed that, at the $4x$ level, the growth rate decreased, cell size increased, stem biomass at flowering increased and the cell wall composition was modified [25]. Autotetraploids from sexual polyploidization likely contain more than two alleles per locus if they originate from unrelated diploid parents; thus, they can functionally be considered an intermediate condition between somatic autopolyploids and allopolyploids. Compared with their diploid parents, these tetraploids can display greater biomass and fertility [26, 27].

Polyploidization can cause rapid genomic structural changes, such as chromosomal rearrangements, gene loss, and genome fractionation, sometimes resulting in greater loss of genes from one of the subgenomes (subgenome dominance) [28–30]. WGD can also alter the associations among chromosomes in the nucleus [31, 32]. WGD has been demonstrated to affect transcription, resulting in deviations from the parental mean (nonadditive expression), expression level dominance, transgressive expression, and allele-specific expression (ASE) bias [30, 33–35]. ASE is defined as the differential expression of two alleles of a gene and is commonly observed in both plants and animals [35, 36]. It can arise through several different mechanisms: gene loss and silencing, activation of transposable elements, epigenetic modifications, and interactions between the divergent regulatory networks of the parents [35–38]. Among the transcriptional outcomes of WGD, the impact of WGD on the expression of small RNAs (smRNAs), including nonadditive expression, has been reported [39–43].

When studying the effects of polyploidization, an important issue is the need to differentiate the impact of hybridization from those of WGD [30, 44, 45]. Hybridization, both intraspecific and interspecific, has generally been demonstrated to influence gene expression more than WGD alone [22, 23, 44, 46–48]. In fact, interspecific hybridization combines two genomes with suites of partially diverged *cis*- and *trans*-regulatory elements [35, 49–52], that can interact and alter the transcriptional balance between parental genomes. In the case of autopolyploids

from sexual polyploidization, separating the impacts of hybridization and WGD is extremely relevant.

Compared with allopolyploidization, autopolyploidization is expected to cause fewer alterations in gene expression [53]. Limited gene expression imbalance was observed in some autopolyploids relative to their diploid parents [27, 46, 54–56].

Alfalfa (*Medicago sativa* L., $2n=4x=32$) is an important leguminous forage crop worldwide. The origin of cultivated $4x$ alfalfa can be found in sexual polyploidization events within and between the two wild diploid subspecies, *M. sativa* ssp. *caerulea* and ssp. *falcata*, which gives rise to cultivated *M. sativa* ssp. *x varia* [26, 57–59]. In previous work by our group, a *M. sativa* ssp. *falcata* plant (PG-F9, $2n=2x=16$) and a ssp. *falcata* *x caerulea* hybrid (12P, $2n=2x=16$) producing significant percentages of $2n$ eggs and pollen, respectively, were selected and characterized [27, 60, 61]. Three $2x$ and three $4x$ plants ($2n=4x=32$) were randomly selected from among the progeny of the PG-F9 \times 12P cross. These six plants are full sibs at different ploidy levels and are well suited for separating the effects of hybridization from those of polyploidization at the phenotypic and molecular levels. Compared with their diploid full siblings, the three neotetraploids presented increased biomass, earlier flowering, higher seed set in crosses with unrelated plants, higher seed weight, and larger leaves with larger cells. Microarray analyses revealed that the expression levels of several hundred genes related to diverse metabolic functions changed in response to polyploidization alone [27]. In this work, we further investigated the transcriptomes of these plant materials via RNA-seq and smRNA-seq. We conducted allele-specific expression analyses to gain further insight into the short-term transcriptional consequences of sexual tetraploidization in *M. sativa*. We show quantitatively limited but qualitatively relevant nonadditive mRNA expression and almost no significant allele-specific expression bias that can be ascribed specifically to WGD. Small RNA expression changes were also limited but potentially relevant. We conclude that the gene expression shifts induced by sexual polyploidization involving divergent alfalfa genotypes are not large but can impact several pathways involved in adaptation and phenotypic traits.

Materials and methods

Plant material and experimental conditions

Eight genotypes were used in this work [27]: the two $2x$ parental plants PG-F9 and 12P described above, three $2x$ progeny plants (S8, S16, S24) and three $4x$ progeny plants (S29, S48, S60). They were maintained by yearly cloning through stem cuttings. Three plants per genotype (biological replicates) were grown in pots filled with field soil,

sand and peat moss 2:1:1 and randomized on benches within a bee-proof cage under natural conditions at the departmental experimental station (S. Andrea d'Agliano, Perugia, Italy). All the plants were clipped at the beginning of flowering. Three-week-old shoots (vegetative growth stage) of the regrowth were sampled by taking the first three fully expanded leaves from several shoots of each of the three cloned plants per genotype. The leaves were then immediately flash-frozen in liquid nitrogen and stored at -80 °C until DNA or RNA extraction. Genomic DNA was extracted using the GenElute Plant Kit (Sigma-Aldrich, St. Louis, MO, USA). RNA isolation was performed using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Acid nucleic purity and concentration were first assayed using the NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). For sequencing, RNA quantification was performed using Bioanalyzer (Agilent) and equal amounts of total RNA per sample were used.

RNA-seq

RNA-seq (Novaseq6000 2×150 bp, ~ 18 million fragments per sample) was performed by Sequentia Biotech (Barcelona, Spain). The raw fastq files were cleaned and trimmed with Cutadapt 4.6 [62], using default parameters for Illumina sequencing. FastQC 0.11.9 was applied to check for quality of both the raw and the clean sequences. To reduce bias for multiple copies of the same exact gene in polyploid genomes, cleaned reads were aligned with STAR 2.7.10b [63] to a non-redundant (collapsed) version of the alfalfa genome sequence of Chen et al. [64], produced by Sequentia as described below. Bam files from the alignment were managed with samtools 1.19 [65] and a count matrix, accounting only for uniquely mapped reads, produced with FeatureCounts, from the subread package [66] with the help of the annotation previously produced. This matrix was then imported into R (R Code Team 2024) [67] and processed with DESeq2 [68] with an in-house built script. The raw matrix was first purged from low-expressed genes, and DESeq2 best practices were applied to retrieve differentially expressed genes. The hierarchical clustering map for differential expression genes is shown in Supplementary Fig. S1A. Genes were considered differentially expressed between genotypes or ploidy levels when their \log_2 fold change (FC) was ± 1.0 with *P*_{adj} < 0.05; *P* values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate [69]. The iTAK tool (version 1.6) was used for transcription factor prediction with default parameters (<http://itak.feilab.net/cgi-bin/itak/index.cgi>) [70]. Pathway enrichment analysis of nonadditively expressed or differentially expressed

genes was conducted using the ShinyGO V0.80 online tool [71], with *M. truncatula* used as the reference gene set. To achieve comprehensive gene functional annotation, *M. sativa* NCBI non-redundant protein IDs were converted into *M. truncatula* orthologs using gProfiler (<https://biit.cs.ut.ee/gprofiler/convert>). KOBAS (v 3.0 version, adjusted $P < 0.05$) was used to test the statistical enrichment of differentially expressed genes (DEGs) in KEGG pathways [72].

Small RNA (smRNA) sequencing

Small RNA libraries were generated and sequenced by Novogene Co., Ltd. Twenty-four sequencing libraries were constructed (eight samples \times three biological replicates) using the smRNA Library Prep Kit (Illumina, San Diego, CA, USA) following the manufacturer's recommendations. Briefly, 3' and 5' adaptors were ligated to the 3' and 5' ends of smRNA. The first strand of cDNA was subsequently synthesized after hybridization with a reverse transcription primer. The double-stranded cDNA libraries were generated through PCR enrichment and checked with a Qubit fluorometer (Thermo Fischer Scientific, MA, USA) and real-time PCR for quantification and a sensitive Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) for size distribution detection. The libraries with inserts between 18 and 40 bp were pooled and sequenced on the Illumina NovaSeq 6000 platform to generate 50 bp single-end reads. Raw fastq format reads were first processed through custom Perl and Python scripts to remove low-quality reads, adapter sequences or the insert tag, which contained poly A, T, G, or C. At the same time, the Q20, Q30, and GC contents of the raw data were calculated. All downstream analyses were based on clean, high-quality data. The smRNA tags were mapped to the *M. sativa* reference genome [64] by Bowtie (bowtie-0.12.9 version) [73] without mismatches. Mapped smRNA tags were used to search for known or conserved miRNAs against miRbase (version 20.0). miRDeep2 (mirdeep2_0_0_5 version) [74] and srna-tools-cli (https://github.com/sRNAworkbench/UEA_sRNA_Workbench) were used to identify candidate miRNAs and draw their secondary structures. Then, smRNA tags were mapped to the RepeatMasker (version 4.0.3) [75] and Rfam [76] databases to exclude reads originating from protein-coding genes, repetitive sequences, rRNAs, tRNAs, snRNAs and snoRNAs. The resulting reads were used for novel miRNAs prediction based on the hairpin structure of miRNA precursors using miREvo (miREvo_v1.1 version) [77] and miRDeep2 (mirdeep2_0_0_5 version) [74]. The miRNA target genes were predicted using psRobot (psRobot_v1.2 version) [78].

Small RNA differential expression, gene ontology and KEGG enrichment

The expression levels of known and unique miRNAs estimated via transcripts per million were statistically analysed and normalized [79]. The read counts of each gene were used as input data for DESeq2 (1.12.0) to obtain differentially expressed miRNAs (DEMs) [68]. The resulting P values were adjusted using the Benjamini and Hochberg's approach to control the false discovery rate [69]. The significant genes ($P_{adj} \leq 0.05$) with corresponding \log_2 fold change values of ± 1.0 were considered differentially expressed. A correlation analysis was performed to demonstrate experimental repeatability and reveal differences in gene expression among samples (Supplementary Fig. S1B). Hierarchical clustering was performed using the read counts of each sample as input data. The GO enrichment analysis of the DEMs' target genes was implemented using the ShinyGO online tool (8.08 version, $P_{adj} \leq 0.05$) [71] with *M. truncatula* as the reference species. Gene functions were determined using the KEGG pathway database. KOBAS (3.0 version, $P_{adj} \leq 0.05$) was used to test the statistical enrichment of the target genes in KEGG pathways [72].

Gene coexpression network analysis

The integrated network analysis between miRNA–target pairs with negative regulatory relationships was implemented using the SWIM tool [80, 81]. The read counts from both mRNA- and miRNA-related genes were used as input data to identify statistically significant master regulators between diploids and tetraploids. An adjusted p -value cutoff of 0.05 and a \log_2 fold change (\log_2FC) threshold of ± 1 were adopted to construct a gene coexpression network based on Pearson correlation.

Allele-specific expression analysis (ASE)

A custom reference genome was constructed for allele-specific expression analysis on the basis of a recently published allele-aware chromosome-level genome assembly for cultivated alfalfa [64]. This genomic assembly consists of 32 allelic chromosomes (8 chromosomes with four copies each) generated by the combination of high-fidelity single-molecule sequencing and Hi-C data. Since variant calling and RNA-Seq tools require a linear reference genome, we constructed a unified reference genome that includes all unique genes without redundancy, to avoid losing information due to possible gene content differences between the four haplotypes. To this end, two tools were combined: MashMap (<https://github.com/marbl/MashMap>) [82], a fast mapper specifically designed to work with long DNA sequences such as genome assemblies, and OrthoFinder (<https://github.com/davidemms/>

OrthoFinde) [83], which identifies orthogroups and orthologous genes. MashMap alignments allowed to identify shared regions between haplotypes and cross-references gene annotations to determine whether genes overlap these regions and should be considered homologous. Homology was assessed by checking if genes from different haplotypes occupy similar genomic positions and share alignment features. Additionally, OrthoFinder results were integrated to refine homology assessment, based on the principle that if two or more loci belong to the same orthogroup defined by OrthoFinder and are also located in the same homologous region identified by MashMap, they can be considered allelic. This combined approach ensured that each gene present in multiple haplotypes is represented only once in the final reference, providing a comprehensive yet non-redundant gene catalog for downstream analyses. Finally, a virtual linear chromosome was built by concatenating the chromosomal region of one allele for each protein-coding gene and separated by 50 Ns.

Genome sequencing of parental plants at low coverage (≥ 12) was performed by Sequentia. Illumina library preparation was carried out according to the manufacturer's instructions following the TruSeq DNA PCR-free Library Construction (insert size 350 bp) protocol. Libraries were then sequenced on a NovaSeq 6000 platform, which produced paired-end reads of 150 bp. Variants (SNPs) were identified between parents, and those that were homozygous in each parent were selected. A quality check was performed on the raw RNA-seq data using the software BBDuk, which removes low-quality portions while preserving the longest high-quality part of the NGS reads. The minimum length was set to 35 bp, and the quality score was 35. The sequences were aligned with the parental genomes to identify the parent-of-origin of the variants. To assess RNA editing, the variants found in the variant calling of the DNA-seq analysis for the parental individuals were compared with those found in the variant calling of the RNA-seq analysis of the same individuals.

High-quality RNA-seq reads were mapped against the custom linear reference using STAR (2.7.3a version) (<https://github.com/alexdobin/STAR>) [63]. Sambamba (0.6.7 version) (<https://lomereiter.github.io/sambamba/>) was used to identify and mark duplicate reads in the resulting BAM files [84], whereas Picard (2.21.1 version) (<https://broadinstitute.github.io/picard/command-line-overview.html#AddOrReplaceReadGroups>) was used to assign read group names. BAM files with RNA-seq reads were further processed with Opossum (<https://github.com/BSGOxford/Opossum>) to pre-process the RNA-seq reads before variant calling [85]. Freebayes (1.3 version) (<https://github.com/ekg/freebayes>) [86] was employed to

detect SNPs between parents using either DNA-seq or RNA-seq data with the following settings: ploidy 2, minimum quality filters for both mapping and base quality of 30, and a minimum coverage threshold of 12. The resulting VCF files were further filtered to keep only the SNPs located within exons. Freebayes was used to count reads mapped to each allele of the variants previously identified in parental samples using DNA-seq. VCF files used as references were further filtered to keep only the variants with a genotype that allowed us to determine the parental origin in the progeny analysis, that is, variants between parents that were homozygous in each parent and necessarily heterozygous in the progenies. These analyses allowed us to identify parent-of-origin bias in progeny gene expression. In the case of unbiased expression, 50% of reference and variant alleles are expected in the progenies. The reference allele frequency was calculated as reference reads (reads supporting the allele observed in the reference genome) divided by total reads (reference reads + alternative reads).

Results

Identification of differentially expressed genes (DEGs)

RNA-seq resulted in an average of 10,825,743 unique reads per sample (Supplementary Table S1A). We filtered out genes with low expression levels by setting the threshold for the minimum number of counts to 10, resulting in the retention of 53,053 transcripts out of the initial 75,092 across the entire dataset. Hierarchical clustering of the correlation matrix revealed that, as expected, both 2x and 4x full-sib progenies were more similar to each other than the parental genotypes (Supplementary Fig. S1A). The two parental genotypes showed significantly different transcription levels for as many as 9,700 genes, approximately 18.3% of all genes (Fig. 1A; Supplementary Table S2); this is to be expected for genetically unrelated plants. Most of these DEGs (59.6%) were expressed at lower levels in PG-F9 than in 12P.

The percentage of DEGs between individual progeny plants and their parents was considerably lower, ranging from 7.9 to 12.3% of all genes. Clear trends were observed between parents compared with the progenies: both 2x and 4x progenies presented a greater number of DEGs with PGF-9 than with 12P (Fig. 1A). Another consistent difference between parental genotypes in these comparisons was that the ratio between the number of upregulated and downregulated genes was reversed: the high transcriptional divergence between progenies and the PG-F9 female parent was due mainly to increased transcription levels in the hybrids, whereas the relatively lower divergence between the progenies and 12P was due mainly to decreased transcription levels in the hybrids. This finding is consistent with the ratio between

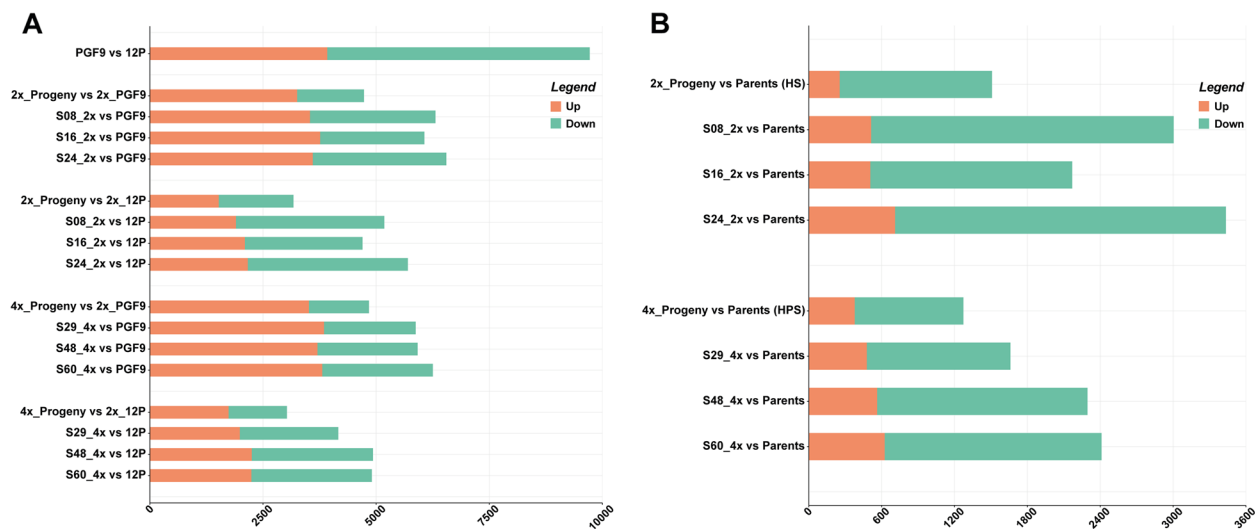


Fig. 1 Differentially expressed gene analysis in *M. sativa*. **A** Number of differentially expressed genes between progenies and single parents. **B** Number of nonadditively-expressed genes (NAGs) between progenies and the parental mean. Nonadditivity was assumed when the mean expression of the progeny was significantly different from the parental average. A log₂ fold change threshold of ± 1 with $P_{adj} \leq 0.05$ was adopted

up- and downregulated genes between the parents and is evidence of genome-wide expression level dominance of the male parent 12P in this cross, which occurs at both ploidy levels of hybrids.

Identification of nonadditively expressed genes (NAGs)

Nonadditive gene expression was assessed by comparing the expression levels of single progeny plants with the parental mean. Deviations from the parental mean were detected for a low percentage of genes (3.13–6.47%), corresponding to 1660–3435 genes per progeny (Fig. 1B). Individual progenies within ploidy level showed stochastic differences, but across all individual progenies at both ploidy levels, a strong prevalence of downregulated genes was observed. Notably, a greater number of NAGs were found on average for 2 \times (1509) progenies than for 4 \times (1273) progenies (significantly different according to the contingency two-tailed Fisher’s exact test with $P_{adj} \leq 0.05$; Fig. 1B; Supplementary Table S2).

Identification of genes affected by polyploidization

In our previous work [27], we defined the NAGs emerging from the 2x progeny–parent comparison as “hybridization sensitive” (HS), the NAGs found in the 4x progeny–parent comparison as “hybridization + polyploidization sensitive” (HPS), and the DEGs found in the 4x–2x progeny comparison as “polyploidization sensitive” (PS). The genes identified in this work belonging to these three groups are listed in Supplementary Table S3.

Within the two groups of NAGs, the HS group (pink circle in Fig. 2) contained 254 upregulated genes and

1,255 downregulated genes, for a total of 1,509 genes, which were nonadditive in 2x progenies (Supplementary Table S3). The HPS group (green circle in Fig. 2) contained 378 up- and 895 downregulated genes, for a total of 1,273 genes (Supplementary Table S3). These genes represent a low proportion of progeny genes deviating from the parental mean, regardless of ploidy (2.4–2.8). The genes of these two groups differed markedly in that the ratio of up- to downregulated genes was twice as high in the 4x progeny (0.42 for HPS genes vs. 0.20 for HS genes). This suggests a possible impact of WGD. The 515 genes unique to the HPS group (HPS_515, Fig. 2) may be associated with ploidy-induced transcriptional novelty.

The DEGs between the 4x and 2x progenies (1518 PS genes, blue circle in Fig. 2) were similar in number to those of the HS and HPS genes, but the upregulated (4x > 2x) genes were largely predominant, with an up/down ratio of 2.12 (Supplementary Table S3). This observation suggests that WGD mostly triggers increased transcription levels. Among the PS genes, 714 (PS_714) were exclusive to this group, that is, their transcription levels differed between the 2x and 4x progenies but not with respect to the parents. This subset of genes, as well as the HPS_515 genes mentioned above, may be the basis of polyploidization-induced novelty. Finally, 231 genes were common between the PS and HPS groups; thus, they were both nonadditively expressed with respect to the parents and differentially expressed between the 2x and 4x progenies. These HPS&PS_231 genes are candidates for being particularly ploidy affected.

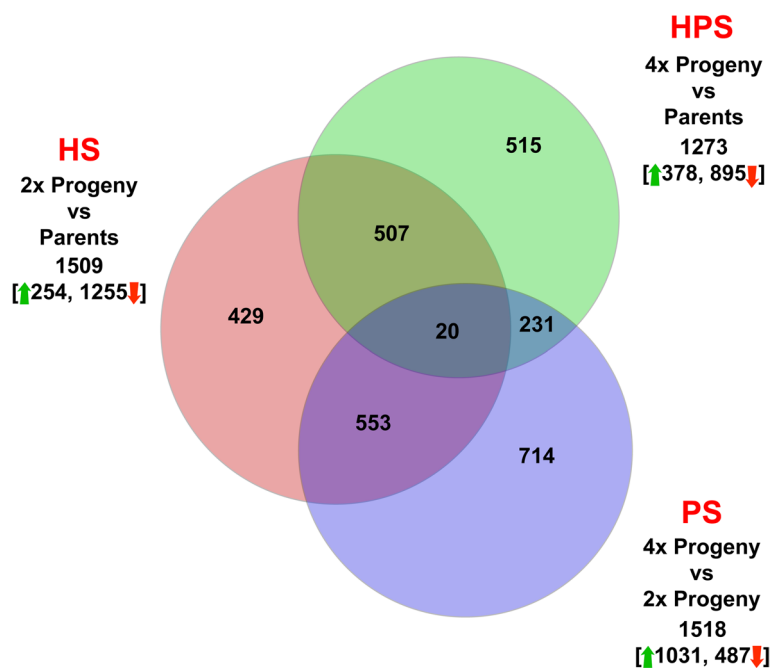


Fig. 2 Venn diagram indicating the numbers of genes that were either nonadditively or differentially expressed. The pink and green circles contain the numbers of nonadditively expressed genes (NAGs) with respect to the mean expression in parents in 2x (hybridization-sensitive, HS) or 4x progenies (hybridization + polyploidization sensitive, HPS); the blue circle contains the number of genes whose expression differed between 2x and 4x progenies (polyploidization-sensitive DEGs, PS). The upwards green arrows indicate upregulated genes, and the downwards red arrows indicate downregulated genes

Functional classification and enrichment analysis of NAGs and DEGs

We performed an enrichment analysis of Gene Ontology (GO) terms for the three groups of genes defined above (HPS_515, PS_714, and HPS&PS_231) in comparison with the genes solely affected by hybridization (HS_429) (Supplementary Table S4). The GO terms enriched exclusively in the PS_714 group, each containing more than 25 genes, were “Plasma membrane” (GO:0005886), “Transport” (GO:0006810), “Endomembrane system” (GO:0012505), “Protein-containing complex” (GO:0032991), followed by “Plastid” (GO:0009536), “Chloroplast” (GO:0009507), highlighting the strong association between plasma membrane categories with the photosynthetic apparatus as a result of WGD. Focusing on the GO terms enriched across all three groups of genes related to WGD (HPS_515, PS_714, and HPS&PS_231), we identified “Cell periphery” (GO:0071944), “Response to chemical” (GO:0042221), “Regulation of molecular function” (GO:0065009), “Cell communication” (GO:0007154) and “Molecular function regulator” (GO:0098772) as the key terms associated with more than 60 genes each. Among the most enriched GO terms, we also identified “DNA-binding transcription factor activity” (GO:0003700), “Transcription regulator activity”

(GO:0140110) and “Sequence-specific DNA binding” (GO:0043565).

We observed a clear trend regarding the ratio between the number of up- and downregulated genes related to each enriched GO term (Supplementary Table S4). For the HS_429 group, the average up/down ratio was 0.51, demonstrating that the transcription level in 2x progeny was lower than the parental average for the vast majority of these genes. Similar results were observed for the HPS_515 group, although with a more balanced up/down ratio of 0.82. For the PS_714 group, the up/down ratio was 1.69, and for the PS+HPS_245 group, it was 2.50, showing that the majority of ploidy-sensitive genes contributing to GO term enrichment had a higher transcription level in the 4x progenies than in the 2x progenies.

These four groups of genes were mapped against the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database to detect any WGD-responsive metabolic pathways. The distribution of KEGG pathways for DEGs in the four datasets is shown in Supplementary Table S5A. Very few genes were found within pathways identified to be affected by hybridization (HS_429 group). Some pathways, such as “metabolic pathways”, were specifically associated with the HPS_515 (59 genes), PS_714 (65 genes) and HPS&PS_231 (27 genes) subsets. These findings suggest that WGD triggered

metabolic changes. The “Starch and sucrose metabolism”, “Protein processing in endoplasmic reticulum” and “Ribosome biogenesis in eukaryotes” pathways were specifically enriched in the HPS_515 subset, while “Plant hormone signal transduction”, “Amino sugar and nucleotide sugar metabolism” and “Fatty acid metabolism” were specifically found in the PS_714 group (Supplementary Table S5A). In the HPS&PS_231 group, 15 genes were included in the KEGG pathway “Biosynthesis of secondary metabolites”, and 7 genes were involved in the pathway “alpha-Linolenic acid metabolism”. Finally, a few genes associated with the “Vitamin B6 metabolism”, and “Glucosinolate biosynthesis” pathways, which are known to play a role in the defence response, were identified [87–89].

To gain further insight into the pathways influenced by WGD, the MapMan approach (version 3.6.0RC1) was applied (Supplementary Table S5 B). We noticed upregulation of genes of the “Protein homeostasis” functional class in the HPS_515 and PS_714 group compared to the HS_429 group. Consistent with the results of the KEGG mapping, upregulated genes related to “secondary metabolism” were identified among the PS_714 and HPS&PS_231 groups. The “Phytohormone action” and “Redox homeostasis” pathways were also identified as potentially associated with WGD. Notably, five genes associated with “chromatin organization” category were found in the PS_714 group, thus corroborating the impact of WGC on chromatin structure and arrangement [32].

Identification of transcription factors responsive to polyploidization

We assessed the effect of WGD on transcription factor (TF) expression by examining the numbers of TFs present within the WGD-related groups (HPS_515, PS_714, and HPS&PS_231) in comparison with those in the HS_429 group, which were affected only by hybridization (Supplementary Table S6; Fig. 3). TFs were significantly more represented in the WGD-related groups (110 vs. 27 genes). This phenomenon was evident across several TF families, including MYB (13 vs. 3), AP2/ERF (12 vs. 3), bHLH (10 vs. 2) and WRKY (9 vs. 1) families. Tify TFs were represented only within the WGD-related groups with 7 genes, 5 of which were in the HPS&PS group. Finally, MYB, AP2/ERF, bHLH and WRKY TFs were overrepresented in the PS group (Supplementary Table S6) [90, 91]. Focusing on the HPS&PS_231 subset, we observed the upregulation of a putative WRKY TF 40 (LOC11435998), which is reported to be a transcriptional repressor of abscisic acid (ABA) signalling [92]. Additionally, a gene encoding the *Medicago truncatula* transcription factor bHLH35 (LOC11441152), which regulates stomatal development [93], was found to be specifically upregulated (Supplementary Table S6). The general picture of TF modulation by WGD may imply an effect of WGD on stress tolerance.

Small RNA-seq

Sequence analyses were performed on approximately 253 million filtered reads (Supplementary Table S1B;

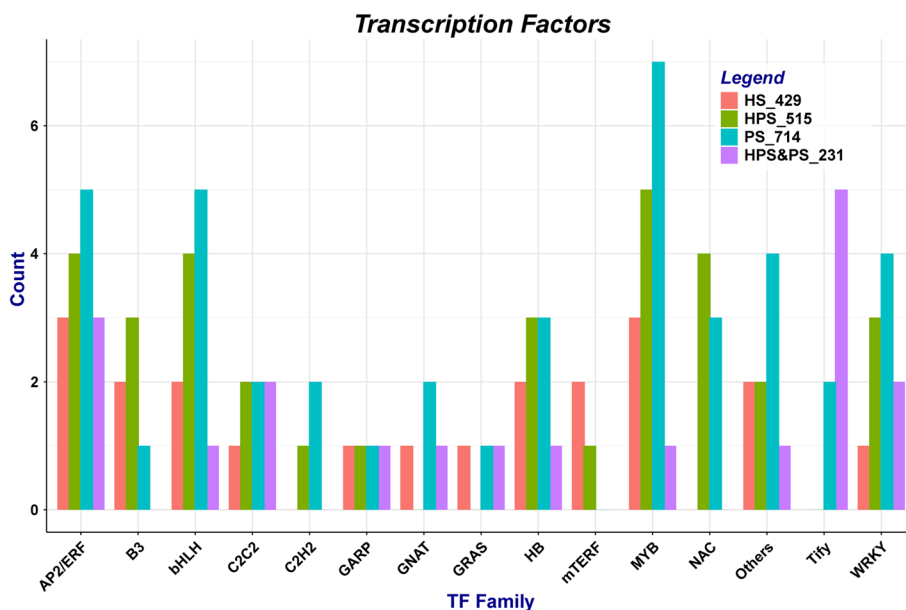


Fig. 3 Distribution of TF families among genes responsive to polyploidization. Each bar represents the number of genes belonging to each transcription factor family

Supplementary Table S7A, B). To assess assembly consistency, filtered unique reads were mapped to the reference *M. sativa* genome [64] using the alignment software Bowtie2, and an average mapping rate of 68.44% was obtained. The overall length distribution of smRNAs revealed that 21- and 24-nt smRNAs were the most represented classes (Supplementary Fig. S2). The different smRNA species mapped to distinct functional regions (Supplementary Fig. S3).

Identification of conserved and novel miRNAs

The smRNA sequences were compared with known miRNAs in miRBase to identify conserved miRNAs. Overall, a total of 257 unique mature miRNAs belonging to 154 different families were identified. The details of conserved miRNAs and miRNA precursors for each genotype are reported in Supplementary Table S7B. No quantitative differences were found between ploidy levels for the number of miRNAs, although marked genotype-specific expression patterns were observed (Supplementary Fig. S4). Among the highly conserved miRNA families, the miR2592 family was the most numerous, with 13 miRNAs. Among the nonannotated smRNAs, a total of 336 (286–314 per genotype) were

predicted as putative novel miRNAs (Supplementary Table S8); 28 of the 336 novel miRNAs were sequenced more than 100 times, whereas only 8 were sequenced more than 500 times. These 8 most abundant novel miRNAs were common to all genotypes (Supplementary Table S9), suggesting that they are miRNA candidates with a high level of confidence.

Differential expression analysis of miRNAs

In the PG-F9 vs. 12P comparison, 199 (83 upregulated and 116 downregulated) differentially expressed miRNAs (DEMs) were detected (Supplementary Table S10). As expected, due to genetic relatedness and stochastic variation between progeny genotypes within ploidy levels, fewer miRNAs were detected when comparing progeny with parents: 21 HS miRNAs (7 up- and 14 downregulated) and 18 HPS miRNAs (8 up- and 10 downregulated) were found to be nonadditively expressed when comparing the 2x progeny and 4x progeny with parents, respectively (Fig. 4A). In the 4x vs. 2x progeny comparison, 26 PS miRNAs (14 up- and 12 downregulated) were identified (Supplementary Table S10).

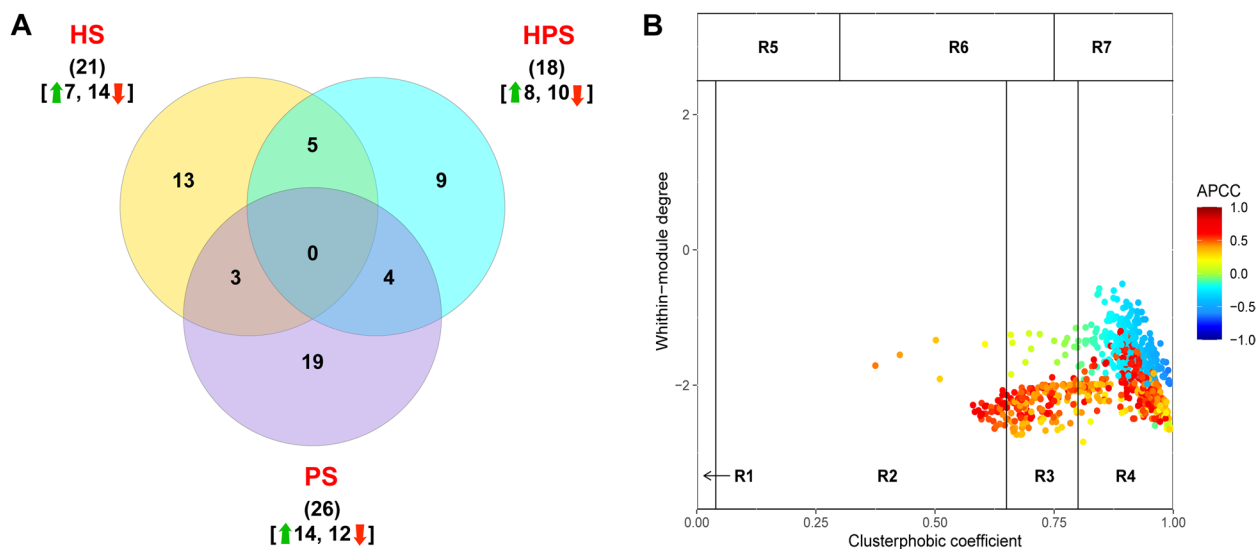


Fig. 4 Distribution and integrated network analysis of differentially expressed miRNAs for the three comparisons ($P_{adj} \leq 0.05$). **A** Venn diagram of miRNAs identified in the three groups of genes. Upwards red arrows indicate significantly upregulated miRNAs, whereas green downwards arrows indicate significantly downregulated miRNAs. The yellow and blue circles contain the numbers of nonadditively expressed miRNAs with respect to the mean expression in parents in 2x (hybridization-sensitive, HS) or 4x progenies (hybridization + polyploidization sensitive, HPS); the blue circle contains the number of miRNAs whose expression differed between 2x and 4x progenies (polyploidization-sensitive, PS). The upwards green arrows indicate upregulated genes, and the downwards red arrows indicate downregulated genes. **B** Heat cartography map generated by SWIM for the PS dataset. APCC: average Pearson correlation coefficient. The R 1–7 regions are defined by two topological properties: the claustrophobic coefficient K_{π} and the within-module degree Z_g . R1 and R2 are populated by 'peripheral nodes', that is, genes that mainly interact within their own community (low K_{π}) and have varying degrees of connection within their community (low to moderate Z_g). R3 contains 'nonhub connectors', which are characterized by nodes interacting with different communities (moderate K_{π}) but are not local hubs (moderate Z_g). R4 is characterized by high K_{π} and low Z_g , indicating that nodes primarily interact outside their own community [81]

Prediction and functional classification of WGD-associated miRNA target genes

We attempted to identify miRNA target genes specifically affected by polyploidization (Supplementary Table S11). A total of 94 target genes were identified for the HPS_9 miRNAs, 482 for the PS_19 miRNAs, and as many as 526 target genes for the HPS&PS_4 group miRNAs. These findings suggest that a substantial impact of WGD on gene expression can be mediated by a small number of differentially expressed miRNAs (Supplementary Table S11). The maximum number of targets was found for mtr-miR7696a-3p (237 target genes) and mtr-miR7696c-3p (232 target genes), which were exclusive to the HPS&PS_4 group. It is worth noting that both mtr-miR7696a-3p and mtr-miR7696c-3p miRNAs have been previously reported for their role in response to drought stress in *M. sativa* seedlings [94, 95]. Seven instances of negative correlations between these miRNAs and some target genes were found (Table 1). Among these miRNA–target pairs, a gene encoding the E3 ubiquitin-protein ligase MBR2, a key player in the final step of the ubiquitination process, was identified as a putative target of mtr-miR7696a-3p; ubiquitination has been shown to affect plant development [96]. Supporting this finding, a gene encoding for *Medicago truncatula* FBD-associated F-box protein At4g10400-like (LOC11446934), which contributes to the formation of E3 ubiquitin ligase complexes [97], was downregulated and targeted by three different miRNAs (Table 1). This suggests a combined regulation of ubiquitination by multiple miRNAs triggered by WGD, possibly affecting protein stability and function. We also observed the targeting of the *M. truncatula* putative P-loop containing nucleoside triphosphate hydrolase, leucine-rich repeat domain, L gene, a disease resistance protein known to be involved in recognizing pathogens and activating defence responses in plants [98].

The search for key regulatory miRNA genes involved in complex biological networks was performed using the SWIM tool [81] (Fig. 4B). A topological role was assigned to each node by intramodule (z_g) and intermodule (K_{π}) connections and the average Pearson correlation coefficient (APCC) between the expression patterns of each node in the coexpression network [80]. The heat cartography maps show the topological properties of the coexpression network for the miRNAs and their targets (Fig. 4B; Supplementary Table S12). Integrated network analysis identified 264 miRNA–target pairs with negative correlations, specifically in the PS and HPS&PS groups (Supplementary Table S12).

The functional annotation of these 264 downregulated mRNAs revealed GO terms associated with the regulation of biosynthetic and metabolic processes, enzymatic and metabolic activities, and protein catabolism (Fig. 5A). Four novel upregulated miRNAs were specific to the PS_19 group: novel_198, novel_217, novel_252 and novel_710. In addition, 3 upregulated miRNAs were specific to the HPS&PS_4 group and were negatively correlated with 26 genes: mtr-miR7696a-3p, mtr-miR7696a-5p and mtr-miR7696c-3p.

To predict any affected metabolic pathways, all target genes of the HS_13, HPS_9, PS_19 and HPS&PS_4 miRNA groups were mapped to the KOBAS database to assign their corresponding KEGG enrichment pathways (Fig. 5B). The pathways enriched in the HPS_9 group included several pathways associated with secondary metabolism, such as “isoquinoline alkaloid biosynthesis” and “biosynthesis of secondary metabolites”, which was consistent with the RNA-seq results (Supplementary Table S5A). In the PS_19 group, “ABC transporters” and “glycerophospholipid metabolism” emerged as the only two pathways, suggesting a focus on lipid transport and membrane transport (Fig. 5B). Additionally, the pathways ‘ether lipid metabolism’ and ‘inositol phosphate

Table 1 Negatively correlated miRNA–target pairs are putatively affected by polyploidization

miRNA id	Target genes	miRNA log ₂ fold change	mRNA log ₂ fold change	Target gene annotation
mtr-miR7696a-3p	MS.gene053533	1.0792	-1.5523	<i>Medicago truncatula</i> E3 ubiquitin-protein ligase MBR2 (LOC11446940)
	MS.gene66626		-1.0136	<i>Medicago truncatula</i> FBD-associated F-box protein At4g10400-like (LOC11446934)
mtr-miR7696a-5p	MS.gene84547	1.0817	-1.4728	<i>Medicago truncatula</i> putative P-loop containing nucleoside triphosphate hydrolase, leucine-rich repeat domain, L
mtr-miR7696c-3p	MS.gene66626	1.0792	-1.0136	<i>Medicago truncatula</i> FBD-associated F-box protein At4g10400-like (LOC11446934)
mtr-miR7696c-5p	MS.gene84547	1.5252	-1.4728	<i>Medicago truncatula</i> putative P-loop containing nucleoside triphosphate hydrolase, leucine-rich repeat domain, L
mtr-miR7696d-5p	MS.gene84547	1.5221	-1.4728	<i>Medicago truncatula</i> putative P-loop containing nucleoside triphosphate hydrolase, leucine-rich repeat domain, L
novel_335	MS.gene66626	1.4816	-1.0136	<i>Medicago truncatula</i> FBD-associated F-box protein At4g10400-like (LOC11446934)

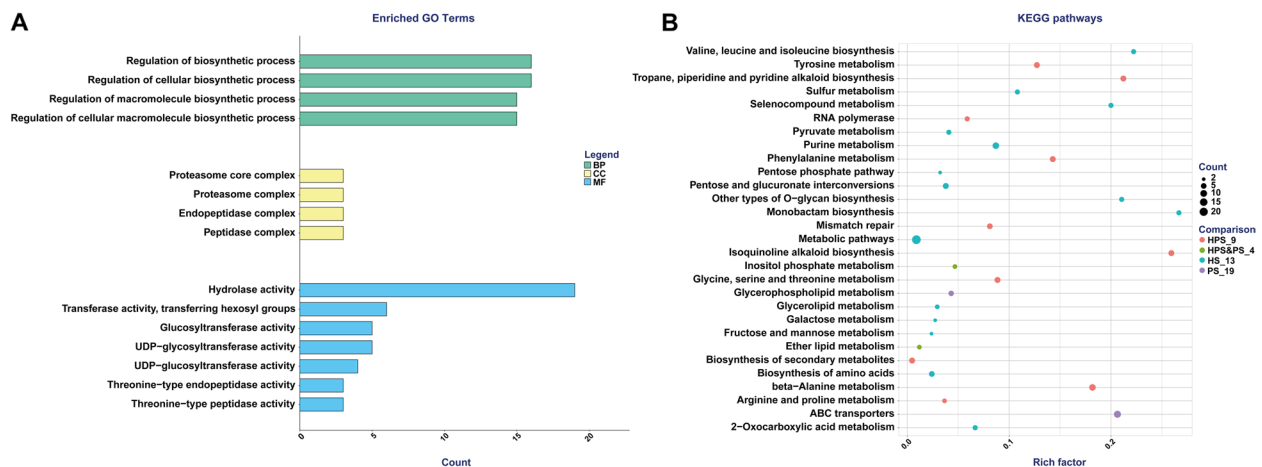


Fig. 5 Enrichment analysis of target genes. **A** GO enrichment analysis of the miRNAs in common between the PS_19 and HPS-PS_4 groups. The Y-axis indicates the subcategories, and the X-axis shows the numbers related to the total number of GO terms. BP, biological process; MF, molecular function; CC, cellular component. **B** Distribution of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways for target genes of the HS_13, HPS_9; PS_14 and HPS&PS_4 groups

metabolism' identified for the HPS&PS_4 group suggest that WGD may have affected lipid metabolism and membrane functions.

Allele-specific expression

To determine whether parental allele-specific expression was affected by WGD, a custom linear reference genome (available at: <https://doi.org/10.5281/zenodo.14354046>) was assembled as described above and used to identify parental alleles. This reference genome contained 75,092 protein-coding genes, approximately half the number of total genes in the source genome (158,894) [64], instead of the expected quarter. In fact, many genes were not present in all four homologues but were present in only three, two or even one chromosome (Supplementary Table S13). The number of single-copy genes, genes present in only one of the four homologues in the source genome, was 12,657 (432–2922 per chromosome).

A total of 480,991 variants (SNPs) between parental genomes were identified via DNA sequencing, and only homozygous variants in parental genomes were used in ASE analysis. The RNA-seq data yielded 239,474 variants, and the filtered clean reads were mapped back to each parental allele. The reference allele frequencies in progeny mRNAs were calculated as reference reads (reads corresponding to the allele present in the reference genome) divided by total reads (reference reads + alternative reads). The genes with at least one variant having a reference allele frequency of 0 or 100 (genes exhibiting complete parent-specific expression) were 377 in 2x progenies, 272 in 4x progenies, and 1,177 in both 2x and 4x progenies (Supplementary Table S14). The reference allele frequencies for all SNPs of each individual sample

were calculated. Histograms for each sample type were plotted (Supplementary Table S15A, Fig. 6A, B, C). As expected, the parents essentially expressed their respective alleles (Fig. 6A). However, genes with reference allele expression frequencies different from 0 to 100 were observed, although at very low frequencies; we hypothesize that RNA editing is the likely explanation for this observation (see below). As expected, most of the progeny, 2x or 4x, expressed both alleles. Parental expression bias was widespread but showed the same pattern between the 2x and 4x progenies (Fig. 6B, C), indicating that WGD was not a factor driving ASE.

To compare the allele expression frequencies of the 2x vs. 4x progeny, the nonparametric Wilcoxon test was used because of the lack of normality of the frequency values. Overall, 925 variants were significantly different ($P < 0.05$) between the 2x and 4x progenies (Supplementary Table S15B). No variant showed significant differences when the P values were adjusted using the Benjamini and Hochberg's correction ($FDR < 0.05$). We conducted a Gene Ontology (GO) enrichment analysis of the genes with at least one variant with a significant nonadjusted $P \leq 0.05$ (Supplementary Table S15B). Considering the 2x dataset, the most significantly enriched GO terms included "small molecule binding" (GO:0036094), "anion binding" (GO:0043168), "nucleotide binding" (GO:0000166), and "nucleoside phosphate binding" (GO:1901265), followed by "purine ribonucleotide binding" (GO:0032555), "purine nucleotide binding" (GO:0017076), "ribonucleotide binding" (GO:0032553), "carbohydrate derivative binding" (GO:0097367) and "purine ribonucleoside triphosphate binding" (GO:0035639) (Supplementary Table S16). The

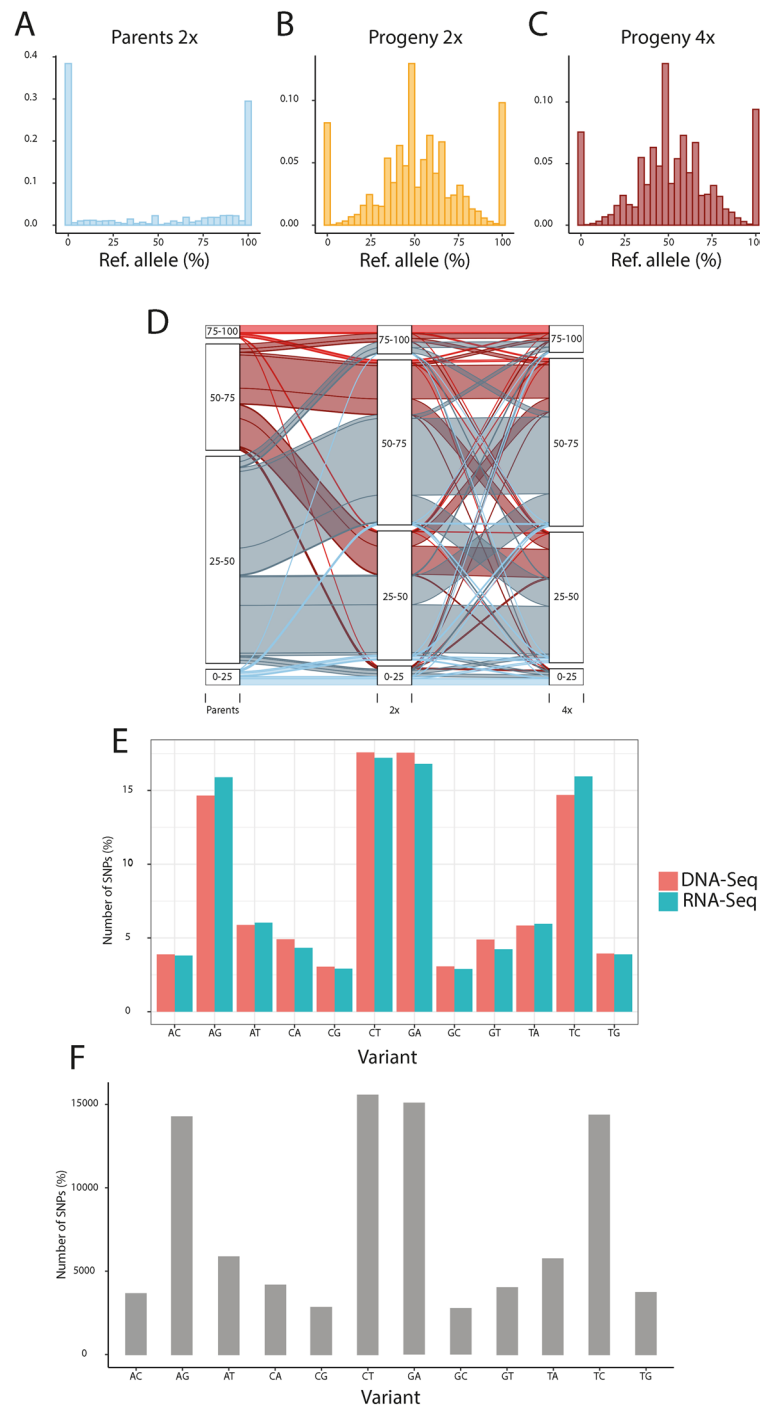


Fig. 6 Allele-specific expression analysis. **A-C** Allele-specific expression in parental genotypes (**A**), diploid (**B**) and tetraploid (**C**) progenies. In the absence of ASE, the two alleles are expected to be equally expressed (50% reference allele). **D** Reference allele frequency alluvial plot of each variant across all experimental groups and parental, diploid, and tetraploid progeny. **E** Total count for each variant type found in the DNA-seq or RNA-seq data of the parental individuals. The X-axis shows the variants as reference/alternative pairs; for example, AC means that the reference is A and the alternative is C. **F** Total count for each variant type found in the RNA-seq but not in the DNA-seq of the parental genomes. The X-axis shows the variants as reference/alternative pairs; for example, AC means that the reference is A and the alternative is C

4x group presented a distinct set of enriched GO terms: “biological_process” (GO:0008150), “catalytic activity” (GO:0003824), “metabolic process” (GO:0008152), “cellular process” (GO:0009987), “organic substance metabolic process” (GO:0071704), “cellular metabolic process” (GO:0044237) and “transferase activity” (GO:0016740) (Supplementary Table S16). An alluvial plot was used to visualize the average reference allele frequency of each variant across the 3 experimental groups: 2x parents, 2x progenies and 4x progenies (Fig. 6D). The white boxes representing allele frequency intervals are approximately the same size when 2x and 4x progenies were compared. The sizes change dramatically if we focus on the parental “25–50” box, which is almost double the size of the same 2x progeny box. The parental “25–50” box was almost split in half to become the “25–50” and “50–75” 2x boxes (Fig. 6D). Quantitatively, the number of variants with different expression intervals between parents and 2x progenies (6,398) or parents and 4x progenies (6,538) was almost the same but was much lower (4,054) between 2x and 4x progenies (Supplementary Table S15A). This further indicates a preponderant effect of hybridization over WGD on allele-specific expression.

RNA editing

Differences were observed when the variants found between parental genomes were compared with those found in the RNA-seq analysis of the same individuals (Fig. 6E). The variants found in the RNA-seq data but not in the DNA-seq data of the parental genotypes are likely to be *bona fide* RNA editing events. We counted 91,867 variants in mRNAs that were not present in the genomes (Fig. 6F). Of these, 55,142 were identified only in 12P, suggesting parent-biased RNA editing. In fact, the reference allele frequencies of the 12P parent presented high levels of variant heterozygosity at the mRNA level (Supplementary Fig. S5). A high number of C-to-T (U), A-to-G, and reciprocal transitions were observed, as described in the literature for plant organelles [99–101].

Discussion

In this work, we investigated the transcriptional effects of WGD by using 4x progenies obtained via bilateral sexual polyploidization from two 2x alfalfa parental plants. Since 2n gametes are the main route for polyploid formation in nature, we believe that our data can contribute to the knowledge of the putative advantages of WGD, possibly favouring domestication. Using control 2x plants obtained by crossing the same parental plants allowed us to separate the effects of WGD from those of hybridization. Overall, our data indicate that sexual autotetraploidization does not dramatically affect the transcription of leaf protein-coding genes in *M. sativa*,

despite its clear effect on the phenotype [27]. Allario et al. [102] came to the same conclusions in a study on *Citrus limonia*. Nonadditive gene expression was present for a limited percentage of genes in the progeny plants studied, and for these genes, lower expression with respect to the parental mean prevailed. Microarray data from the same genotypes [27] revealed much lower percentages than we observed using RNA-seq: 0.28% HS, 1.45% HPS, and 1.36% PS genes. We assume that the sensitivity of RNA-seq allowed us to detect differential expression more reliably. Palumbo et al. [103] recently published a transcriptomics study using three 2x and three 4x progeny plants from the PG-F9×12P cross without including the parental plants. Although they concentrated on reproductive tissue, they also used leaf tissue as a control. In leaves, 949 “ploidy-dependent” DEGs (509 upregulated, 440 downregulated), defined as “ploidy-sensitive” (PS) genes in this work, were detected; in our study (using three 2x and three 4x progeny plants but with different genotypes), more DEGs were detected (1518, 1031 upregulated, 487 downregulated). We searched for the core gene set deregulated by polyploidization in both studies and found only 33 genes. Among these genes, 20 were up- or downregulated in both studies, whereas 13 exhibited the opposite behaviour (Supplementary Table S17). Such little overlap between the genes identified in these two studies points to large transcriptional differences among full-sib plants and to a strong effect of the environmental conditions. Further studies with a larger number of genotypes tested in different environments may help to detect a larger set of ploidy-affected genes. Among the 20 genes showing consistent transcriptional behaviour in the two studies, five putative transcription factors (XP_013443539.1, XP_024637947.1, XP_003611347.1, XP_003623638.1, XP_003608660.2) were strongly upregulated; one of them, a putative ethylene responsive AP2-EREBP TF (XP_003608660.2), had a log₂_fold change > 2 in both studies. This finding suggests a possible effect of WGD on the stress response, which is consistent with the results of the above-described mRNA functional enrichment analysis.

We found that the two parental genotypes did not contribute equally to the progeny transcriptomes, with the male parent 12P being more similar to the progenies than PG-F9 at both ploidy levels (Fig. 1A). This can be explained by the fact that 12P is a *ssp. falcata* × *caerulea* hybrid, as are the progenies, which contain ¼ of the genes from *caerulea*. Liu and Wang [30] re-examined and questioned the evidence for gene expression dominance and subgenome dominance in hybrids and polyploids, emphasizing the influence of parental legacy. In this work, a direct comparison between 2x parents and their 2x and 4x full sibling progenies provided solid evidence

of the presence of genome-wide expression level dominance of the male parent 12P, irrespective of ploidy.

Stochastic differences in the numbers of nonadditively expressed genes were present between progeny genotypes of both ploidy levels and were more evident in 2x progenies, confirming previous microarray data [27]. Interestingly, 4x progenies deviated less than 2x progenies from the parental mean. Even though 2n gametes can transmit different amounts of heterozygosity to progeny depending on the mechanism of their formation (first or second division restitution) [104], it can be assumed that the 4x progenies carry all four parental alleles for a significant percentage of their genes. Therefore, it can be expected that 4x progenies are more genetically similar and deviate less from the parental mean than 2x progenies, which randomly receive only one allele per gene from each parental plant.

We observed that the ratio between up- and down-regulated genes among the nonadditive or differentially expressed genes increased when ploidy came into play, with a sharp increase for PS genes. This trend was reinforced when only the genes involved in GO term enrichment were considered. In other words, a large majority of the genes whose expression level changed due to WGD presented relatively high expression levels. Since the transcription levels were measured on a per-genome basis (equal total RNA amounts were used per sample),

upregulation in 4x plants indicates that genes respond more than proportionally to WGD. Considering that the epidermal leaf cell surface area increased by approximately 70% and leaf size 83% in 4x vs. 2x hybrids [27], we suggest that in neopolyploids, the 4x state requires some extra metabolic effort through the upregulation of some genes, which we estimated to constitute approximately 2.66% of the protein-coding genes (from Fig. 2: $378 + 1031 = 1409$; $1409/53053 = 0.0266$).

We propose an interpretation of the patterns of differential transcription (Table 2). While the large majority of the protein-coding genes were unaffected by either hybridization or WGD (line A), small percentages of genes were affected, which can contribute to the performance of the hybrids via different mechanisms, on the basis of the accepted hypothesis that quantitative changes in gene transcription can play a role in heterosis. Overall, we estimated that 3.63% of the protein-coding genes could contribute to heterosis and polyploidy-induced novelty.

Among the most enriched KEGG pathways, secondary metabolite biosynthesis was pinpointed by both KEGG and MapMan analyses. There are examples of WGD affecting the concentration of specific secondary metabolites [105–107]. In alfalfa, saponins are secondary metabolites that directly affect forage quality [108] and have diverse biological effects [109]. Metabolomic analyses

Table 2 Interpretation of different transcriptional responses to hybridization and to sexual polyploidization

Categories of protein-coding genes based on transcriptional response to WGD	Comparison			Interpretation	Effects in Neopolyploids
	2x progeny vs. parental mean	4x progeny vs. parental mean	4x progeny vs. 2x progeny		
A. Additive 95.7%	=	=	=	Transcription level of parents does not vary in hybrids irrespective of ploidy level	Do not contribute to heterosis and polyploid novelty
B. Hybridization-Sensitive (HS) 0.81%	<>	=	=	The presence of alleles from genetically distant parents in 2x hybrids perturbs transcription, whereas the copresence of both (coadapted) alleles from each parent in the 4x hybrids stabilizes transcription	Can contribute to heterosis at the 2x, not the 4x level, do not contribute to polyploid novelty
C. Hybridization + Polyploidization-sensitive (HPS) 1.41%	=	<>	=	The 4x state of hybrids amplifies transcription level differences from the parental average	Can contribute to heterosis at the 4x, not the 2x level, and to polyploid novelty
D. Polyploidization-Sensitive (PS) 1.78%	=	=	<>	Higher allelic diversity in 4x hybrids causes transcription differences from the parental average in the opposite direction with respect to 2x hybrids	Can contribute to heterosis at the 4x, not the 2x level, and to polyploid novelty
E. Shared between C and D (HPS&PS) 0.44%	=	<>	<>	Higher allelic diversity in the 4x hybrids perturbs transcription levels	Can contribute to heterosis at the 4x, not the 2x level, and to polyploid novelty

could be conducted on these materials to investigate the novelty of the secondary metabolite composition resulting from WGD.

We examined TF expression as affected by WGD and found four times as many TF genes as those affected by hybridization alone. Several of these belong to the TF classes involved in the stress response. Seven TFs of the Tify family were represented only among WGD-related genes. This family is involved in ABA signalling and stress responses [110–112]. The substantial increase in the expression of ABA-related TFs in tetraploids suggests that WGD may increase stress tolerance by pre-activating the ABA signalling cascade [90]. Recently, Janko et al. [52] introduced a thermodynamic model that emphasizes the role of regulatory divergence between parental species in shaping gene expression patterns in hybrids and polyploids. The model suggests that non-additive gene expression, including expression-level dominance, arises primarily from constrained interactions between inherited regulatory networks rather than from adaptive changes post-polyploidization. These regulatory constraints, where transcription factors from each parental genome exhibit differential affinities for homologous and heterologous promoters, can lead to transcriptome downsizing and parental genome dominance.

There were no quantitative differences in the numbers of miRNAs between 2x and 4x plants. This finding is consistent with evidence from maize, where the effect of ploidy per se did not trigger differential miRNA expression [113]. This contrasts with the observation that miRNA abundance is negatively correlated with ploidy in *A. thaliana* [114]. Large, significant differences were observed, instead, between genotypes within ploidies, and consequently, very few miRNAs were significantly associated with WGD. The potential impact of these few miRNAs should not be disregarded since they target more than 1000 genes. In particular, the 4 HPS&PS-specific miRNAs were collectively associated with more than 500 target genes, including 469 targeted by two drought stress-related miRNAs previously isolated from alfalfa seedlings [94, 95]. Overall, smRNA-seq and co-expression network analyses reinforce the hypothesis that WGD had an impact on stress tolerance stemming from RNA-seq. The potential of polyploidy to improve stress tolerance has been demonstrated [7, 14]. Recently, an increased ABA content and drought tolerance in 4x over 2x plants were demonstrated in *Lycium ruthenicum* [115]. Together, our results encouraged us to investigate the ability of neotetraploids to better tolerate abiotic stresses.

At least in the limited sample of neopolyploids examined here, parent-of-origin bias in allele-specific gene expression was not a significant outcome of alfalfa

sexual tetraploidization. Hybridization appears to be a much more important source of expression bias. Interestingly, we found that parentally biased RNA editing may be a significant source of variation. In plants, RNA editing is well known in plant organelles [99, 101, 116]. Our hypothesis of parent-of-origin-biased RNA editing demands confirmation through further study. Our ASE analysis required to assemble a custom reference genome including one copy of all protein-coding genes identified by Chen et al. [64]. In doing so, we observed extensive presence/absence variation between the four homologues, and we identified 12,657 single-copy genes in 158,894, or 7.97%. This finding is consistent with many studies in diploid and polyploid plants [117], but the scale appears to be larger than that reported in other studies. Zhou et al. [118] reported that in diploid potato (a highly heterozygous, clonally reproducing species), 1,878 out of a total of 76,394 protein-coding genes (2.46%) were present in only one of the two homologous chromosome sets. In a pairwise comparison between three maize inbred lines, only 22–75 protein-coding genes were line-specific [119]. A possible explanation for the comparatively high presence/absence variation between homologous chromosomes in tetraploid alfalfa is the high genome redundancy of autopolyploids, making them especially prone to post-WGD gene loss with respect to diploids. This hypothesis prompted us to sequence the genomes of our neopolyploids and their advanced generations to investigate early gene loss with respect to the diploid parents.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-025-06090-z>.

- Supplementary Material 1.
- Supplementary Material 2.
- Supplementary Material 3.
- Supplementary Material 4.
- Supplementary Material 5.
- Supplementary Material 6.
- Supplementary Material 7.
- Supplementary Material 8.
- Supplementary Material 9.
- Supplementary Material 10.
- Supplementary Material 11.
- Supplementary Material 12.
- Supplementary Material 13.
- Supplementary Material 14.
- Supplementary Material 15.
- Supplementary Material 16.
- Supplementary Material 17.

Supplementary Material 18.
 Supplementary Material 19.
 Supplementary Material 20.
 Supplementary Material 21.

Acknowledgements

We acknowledge funding from the Consorzio Interuniversitario Biotecnologie CIB-MUR Project CMPT219220 (PI: DR) and from the project MIUR-PRIN Project 2020HB9PR9 (PI: DR). The EU-RISE project "Polyploid" (PI: EA) supported AWA. We thank Gianni Barcaccia and Fabio Palumbo, Università di Padova, Italy, for helpful discussions. We are grateful to an anonymous referee for a very insightful review.

Authors' contributions

D.R. and E.A. designed the study. D.R., D.F.S., A.W.A. and M.C. managed, propagated, and maintained the plant materials. M.B., A.F., and D.R. sampled the plants and extracted the nucleic acids. G.P., S.C., and D.F.S. conducted the bioinformatic analyses. D.F.S., G.P., D.F.S. and D.R. interpreted the data. D.F.S. and D.R. drafted the manuscript. All the authors contributed to the manuscript and approved the final version.

Data availability

The plant materials used are available upon request. The datasets supporting the conclusions of this article are available in the NCBI SRA repository with the BioProject accession number PRJNA1134386. The custom linear alfalfa reference genome prepared for this work is available at: <https://doi.org/10.5281/zenodo.14354046>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Agricultural, Food and Environmental Sciences, University of Perugia, via Borgo XX giugno 74, Perugia 06121, Italy. ²Interuniversity Consortium for Biotechnology (CIB), Area Science Park, Padriciano 99, Trieste 34149, Italy. ³Department of Veterinary Medicine, University of Perugia, via S. Costanzo 4, Perugia 06126, Italy. ⁴Department of Life Sciences and Biotechnology, Section of Biochemistry and Molecular Biology, University of Ferrara, via Fossato di Mortara 74, Ferrara 44121, Italy. ⁵CREA Research Centre for Animal Production and Aquaculture (CREA-ZA), Viale Piacenza 29, Lodi 26900, Italy.

Received: 5 November 2024 Accepted: 9 January 2025

Published online: 25 January 2025

References

- Brownfield L, Köhler C. Unreduced gamete formation in plants: mechanisms and prospects. *J Exp Bot*. 2011;62:1659–68.
- De Storme N, Mason A. Plant speciation through chromosome instability and ploidy change: Cellular mechanisms, molecular factors and evolutionary relevance. *Curr Plant Biol*. 2014;1:10–33.
- Fawcett JA, Van de Peer Y. Angiosperm polyploids and their road to evolutionary success. *Trends Evol Biol*. 2010;2:16–21.
- Eric Schranz M, Mohammadin S, Edger PP. Ancient whole genome duplications, novelty and diversification: the WGD Radiation lag-Time Model. *Curr Opin Plant Biol*. 2012;15:147–53.
- Alix K, Gérard PR, Schwarzacher T, Heslop-Harrison JSP. Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. *Ann Bot*. 2017;120:183–94.
- Landis JB, Soltis DE, Li Z, Marx HE, Barker MS, Tank DC, et al. Impact of whole-genome duplication events on diversification rates in angiosperms. *Am J Bot*. 2018;105:348–63.
- Heslop-Harrison JS, Schwarzacher T, Liu Q. Polyploidy: its consequences and enabling role in plant diversification and evolution. *Ann Bot*. 2023;131:1–9.
- Barker MS, Arrigo N, Baniaga AE, Li Z, Levin DA. On the relative abundance of autopolyploids and allopolyploids. *New Phytol*. 2016;210:391–8.
- Jiao Y, Wickett NJ, Ayyampalayam S, Chanderali AS, Landherr L, Ralph PE, et al. Ancestral polyploidy in seed plants and angiosperms. *Nature*. 2011;473:97–100.
- Salman-Minkov A, Sabath N, Mayrose I. Whole-genome duplication as a key factor in crop domestication. *Nat Plants*. 2016;2:1–4.
- Baduel P, Bray S, Vallejo-Marin M, Kolář F, Yant L. The Polyploid hop: shifting challenges and opportunities over the evolutionary lifespan of genome duplications. *Front Ecol Evol*. 2018;6:AUG:1–19.
- Wang J, Qin J, Sun P, Ma X, Yu J, Li Y, et al. Polyploidy Index and its implications for the evolution of polyploids. *Front Genet*. 2019;10:1–11.
- Chen ZJ, Yu HH. Genetic and Epigenetic Mechanisms for Polyploidy and Hybridity. *Polyploid Hybrid Genom*. 2013;58:335–54.
- Van de Peer Y, Ashman TL, Soltis PS, Soltis DE. Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell*. 2021;33:11–26.
- Tossi VE, Martínez Tosar LJ, Laino LE, Iannicelli J, Regalado JJ, Escandón AS, et al. Impact of polyploidy on plant tolerance to abiotic and biotic stresses. *Front Plant Sci*. 2022;13:1–19.
- Bingham ET, Groose RW, Woodfield DR, Kidwell KK. Complementary gene interactions in alfalfa are greater in autotetraploids than diploids. *Crop Sci*. 1994;34:823–9.
- McCoy TJ, Rowe DE. Single cross alfalfa (*Medicago sativa* L.) hybrids produced via 2n gametes and somatic chromosome doubling: experimental and theoretical comparisons. *Theor Appl Genet*. 1986;72:80–3.
- Mok DWS, Peloquin SJ. Breeding value of 2n pollen (diplandroids) in tetraploid x diploid crosses in potatoes. *Theor Appl Genet*. 1975;46:307–14.
- Groose RW, Talbert LE, Kojis WP, Bingham ET. Progressive heterosis in autotetraploid alfalfa: studies using two types of inbreds. *Crop Sci*. 1989;29:1173–7.
- Washburn JD, McElfresh MJ, Birchler JA. Progressive heterosis in genetically defined tetraploid maize. *J Genet Genomics*. 2019;46:389–96.
- Riddle NC, Kato A, Birchler JA. Genetic variation for the response to ploidy change in *Zea mays* L. *Theor Appl Genet*. 2006;114:101–11.
- Chen ZJ. Molecular mechanisms of polyploidy and hybrid vigor. *Trends Plant Sci*. 2010;15:57–71.
- Miller M, Zhang C, Chen ZJ. Ploidy and hybridity effects on growth vigor and gene expression in *Arabidopsis thaliana* hybrids and their parents. G3 genes. *Genomes Genet*. 2012;2:505–13.
- Aversano R, Scarano MT, Aronne G, Caruso I, D'Amelia V, De Micco V, et al. Genotype-specific changes associated to early synthesis of autotetraploids in wild potato species. *Euphytica*. 2015;202:307–16.
- Corneillie S, De Storme N, Van Acker R, Fangel JU, De Bruyne M, De Rycke R, et al. Polyploidy affects plant growth and alters cell wall composition. *Plant Physiol*. 2019;179:74–87.
- Barcaccia G, Tavoletti S, Mariani A, Veronesi F. Occurrence, inheritance and use of reproductive mutants in alfalfa improvement. *Euphytica*. 2003;133:37–56.
- Rosellini D, Ferradini N, Allegrucci S, Capomaccio S, Zago ED, Leonetti P, et al. Sexual polyploidization in *Medicago sativa* L.: impact on the phenotype, gene transcription, and genome methylation. G3 genes, genomes. *Genet*. 2016;6:925–38.
- Cheng F, Wu J, Cai X, Liang J, Freeling M, Wang X. Gene retention, fractionation and subgenome differences in polyploid plants. *Nat Plants*. 2018;4:258–68.
- Wendel JF, Lisch D, Hu G, Mason AS. The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Curr Opin Genet Dev*. 2018;49:1–7.
- Liu C, Wang YG. Does one subgenome become dominant in the formation and evolution of a polyploid? *Ann Bot*. 2023;131:11–6.

31. Wan T, Liu ZM, Li LF, Leitch AR, Leitch IJ, Lohaus R, et al. A genome for gnetophytes and early evolution of seed plants. *Nat Plants*. 2018;4:82–9.
32. Zhang H, Zheng R, Wang Y, Zhang Y, Hong P, Fang Y, et al. The effects of Arabidopsis genome duplication on the chromatin organization and transcriptional regulation. *Nucleic Acids Res*. 2019;47:7857–69.
33. Grover CE, Gallagher JP, Szadkowski EP, Yoo MJ, Fligel LE, Wendel JF. Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol*. 2012;196:966–71.
34. Yoo MJ, Szadkowski E, Wendel JF. Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity (Edinb)*. 2013;110:171–80.
35. Hu G, Wendel JF. Cis–trans controls and regulatory novelty accompanying allopolyploidization. *New Phytol*. 2019;221:1691–700.
36. Gaur U, Li K, Mei S, Liu G. Research progress in allele-specific expression and its regulatory mechanisms. *J Appl Genet*. 2013;54:271–83.
37. Ma X, Xing F, Jia Q, Zhang Q, Hu T, Wu B, et al. Parental variation in CHG methylation is associated with allelic-specific expression in elite hybrid rice. *Plant Physiol*. 2021;186:1025–41.
38. Hamid R, Jacob F, Ghorbanzadeh Z, Jafari L, Alishah O. Dynamic roles of small RNAs and DNA methylation associated with heterosis in allotetraploid cotton (*Gossypium hirsutum* L). *BMC Plant Biol*. 2023;23:1–24.
39. Ng DWK, Lu J, Chen ZJ. Big roles for small RNAs in polyploidy, hybrid vigor, and hybrid incompatibility. *Curr Opin Plant Biol*. 2012;15:154–61.
40. Wendel JF, Jackson SA, Meyers BC, Wing RA. Evolution of plant genome architecture. *Genome Biol*. 2016;17:1–14.
41. Liu B, Sun G. Transcriptome and miRNAs analyses enhance our understanding of the evolutionary advantages of polyploidy. *Crit Rev Biotechnol*. 2019;39:173–80.
42. Palacios PM, Jacquemot MP, Tapie M, Rousselet A, Diop M, Remoué C, et al. Assessing the response of small RNA populations to allopolyploidy using resynthesized brassica napus allotetraploids. *Mol Biol Evol*. 2019;36:709–26.
43. Cavé-Radet A, Giraud D, Lima O, El Amrani A, Ainouche M, Salmon A. Evolution of small RNA expression following hybridization and allopolyploidization: insights from *Spartina* species (Poaceae, Chloridoideae). *Plant Mol Biol*. 2020;102:55–72.
44. Wang J, Tian L, Lee HS, Wei NE, Jiang H, Watson B, et al. Genomewide nonadditive gene regulation in arabidopsis allotetraploids. *Genetics*. 2006;172:507–17.
45. Xu P, Zhang X, Wang X, Li J, Liu G, Kuang Y, et al. Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nat Genet*. 2014;46:1212–9.
46. Hegarty MJ, Barker GL, Wilson ID, Abbott RJ, Edwards KJ, Hiscock SJ. Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. *Curr Biol*. 2006;16:1652–9.
47. Albertin W, Alix K, Balliau T, Brabant P, Davanture M, Malosse C, et al. Differential regulation of gene products in newly synthesized *Brassica napus* allotetraploids is not related to protein function nor subcellular localization. *BMC Genomics*. 2007;8:1–15.
48. Chen ZJ. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu Rev Plant Biol*. 2007;58:377–406.
49. Osborn TC, Chris Pires J, Birchler JA, Auger DL, Chen ZJ, Lee HS, et al. Understanding mechanisms of novel gene expression in polyploids. *Trends Genet*. 2003;19:141–7.
50. Bottani S, Zabet NR, Wendel JF, Veitia RA. Gene expression dominance in Allopolyploids: hypotheses and models. *Trends Plant Sci*. 2018;23:393–402.
51. Nieto Feliner G, Casacuberta J, Wendel JF. Genomics of Evolutionary Novelty in hybrids and polyploids. *Front Genet*. 2020;11:1–21.
52. Janko K, Eisner J, Cigler P, Tichopád Z. Unifying framework explaining how parental regulatory divergence can drive gene expression in hybrids and allopolyploids. *Nat Commun*. 2024;15:15.
53. Gregory TR. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biol Rev*. 2001;76:65–101.
54. Stupar RM, Bhaskar PB, Yandell BS, Rensink WA, Hart AL, Ouyang S, et al. Phenotypic and transcriptomic changes associated with potato autopolyploidization. *Genetics*. 2007;176:2055–67.
55. Riddle NC, Jiang H, An L, Doerge RW, Birchler JA. Gene expression analysis at the intersection of ploidy and hybridity in maize. *Theor Appl Genet*. 2010;120:341–53.
56. Pignatta D, Dilkes BP, Yoo SY, Henry IM, Madlung A, Doerge RW, et al. Differential sensitivity of the Arabidopsis thaliana transcriptome and enhancers to the effects of genome doubling. *New Phytol*. 2010;186:194–206.
57. Veronesi F, Mariani A, Bingham ET. Unreduced gametes in diploid *Medicago* and their importance in alfalfa breeding. *Theor Appl Genet*. 1986;72:37–41.
58. Small E. *Alfalfa and relatives: evolution and classification of Medicago*. Ottawa: NRC Research; 2011.
59. Palumbo F, Pasquali E, Albertini E, Barcaccia G. A review of unreduced gametes and neopolyploids in alfalfa: how to fill the gap between well-established meiotic mutants and next-generation genomic resources. *Plants*. 2021;10:999.
60. Tavoletti S. Cytological mechanisms of 2n egg formation in a diploid genotype of *Medicago sativa* subsp. *falcata*. *Euphytica*. 1994;75:1–8.
61. Barcaccia G, Tavoletti S, Falcinelli M, Veronesi F. Environmental influences on the frequency and viability of meiotic and apomeiotic cells of a diploid mutant of alfalfa. *Crop Sci*. 1997;37:70–6.
62. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;17:10–2.
63. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.
64. Chen H, Zeng Y, Yang Y, Huang L, Tang B, Zhang H, et al. Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the autotetraploid cultivated alfalfa. *Nat Commun*. 2020;11:1–11.
65. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10:1–4.
66. Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30:923–30.
67. R Core Team. R: A language and environment for statistical computing (version 4.4.0). 2024.
68. Love MI, Huber W, Anders S. Moderated estimation of Fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:1–21.
69. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing Author (s): Yoav Benjamini and Yosef Hochberg Source : Journal of the Royal Statistical Society. Series B (Methodological), Vol. 57, No. 1 (1995). *Publi J R Stat Soc*. 1995;57:289–300.
70. Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, et al. iTAK: a program for genome-wide prediction and Classification of Plant Transcription Factors, transcriptional regulators, and Protein Kinases. *Mol Plant*. 2016;9:1667–70.
71. Ge SX, Jung D, Jung D, Yao R. ShinyGO: a graphical enrichment tool for animals and plants. *bioRxiv*. 2018;36 December 2019:315150.
72. Bu D, Luo H, Huo P, Wang Z, Zhang S, He Z, et al. KOBAS-i: Intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res*. 2021;49:W317–25.
73. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10:1–10.
74. Friedländer MR, MacKowiak SD, Li N, Chen W, Rajewsky N. MiRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res*. 2012;40:37–52.
75. Chen N. Using repeat Masker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinforma*. 2004;5:4–40.
76. Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, et al. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res*. 2018;46:D335–42.
77. Wen M, Shen Y, Shi S, Tang T. MiREvo: an integrative microRNA evolutionary analysis platform for next-generation sequencing experiments. *BMC Bioinformatics*. 2012;13:1–10.
78. Wu HJ, Ma YK, Chen T, Wang M, Wang XJ, PsRobot. A web-based plant small RNA meta-analysis toolbox. *Nucleic Acids Res*. 2012;40:22–8.
79. Zhou L, Chen J, Li Z, Li X, Hu X, Huang Y, et al. Integrated profiling of MicroRNAs and mRNAs: MicroRNAs located on Xq27.3 associate with clear cell renal cell carcinoma. *PLoS One*. 2010;5:e15224.
80. Palumbo MC, Zenoni S, Fasoli M, Massonnet M, Farina L, Castiglione F, et al. Integrated network analysis identifies fight-club nodes as a class of hubs encompassing key putative switch genes that induce major

- transcriptome reprogramming during grapevine development. *Plant Cell*. 2014;26:4617–35.
81. Paci P, Colombo T, Ficon G, Gurtner A, Pavesi G, Farina L. SWIM: A computational tool to unveiling crucial nodes in complex biological networks. *Sci Rep*. 2017;2017:7.
 82. Kille B, Garrison E, Treangen TJ, Phillippy AM. Minmers are a generalization of minimizers that enable unbiased local Jaccard estimation. *Bioinformatics*. 2023;39:btad512.
 83. Emms DM, Kelly S, OrthoFinder. Phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20:1–14.
 84. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics*. 2015;31:2032–4.
 85. Oikonen L, Lise S. Making the most of RNA-seq: pre-processing sequencing data with opossum for reliable SNP variant detection. *Wellcome Open Res*. 2017;2:1–13.
 86. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. 2012;1–9. arXiv:1207.3907v2.
 87. Wang P, Yang C, Chen H, Song C, Zhang X, Wang D. Transcriptomic basis for drought-resistance in *Brassica napus* L. *Sci Rep*. 2017;7:1–20.
 88. Zhou Y, Yu H, Tang Y, Chen R, Luo J, Shi C, et al. Critical roles of mitochondrial fatty acid synthesis in tomato development and environmental response. *Plant Physiol*. 2022;190:576–91.
 89. Zhao Y, Zhang F, Mickan B, Wang D. Inoculation of wheat with *Bacillus* sp. wp-6 altered amino acid and flavonoid metabolism and promoted plant growth. *Plant Cell Rep*. 2023;42:165–79.
 90. Del Pozo JC, Ramirez-Parra E. Deciphering the molecular bases for drought tolerance in *Arabidopsis* autotetraploids. *Plant Cell Environ*. 2014;37:2722–37.
 91. Ndayambaza B, Jin X, Min X, Lin X, Yin X, Liu W. Genome-wide identification and expression analysis of the Barrel Medic (*Medicago truncatula*) and alfalfa (*Medicago sativa* L.) Basic Helix-Loop-Helix transcription factor family under Salt and Drought stresses. *J Plant Growth Regul*. 2021;40:2058–78.
 92. Chen H, Lai Z, Shi J, Xiao Y, Chen Z, Xu X. Roles of *Arabidopsis* WRKY18, WRKY40 and WRKY60 transcription factors in plant responses to abscisic acid and abiotic stress. *BMC Plant Biol*. 2010;10:1–15.
 93. Dong Y, Wang C, Han X, Tang S, Liu S, Xia X, et al. A novel bHLH transcription factor PebHLH35 from *Populus euphratica* confers drought tolerance through regulating stomatal development, photosynthesis and growth in *Arabidopsis*. *Biochem Biophys Res Commun*. 2014;450:453–8.
 94. Raza A, Charagh S, Karikari B, Sharif R, Yadav V, Mubarik MS, et al. miRNAs for crop improvement. *Plant Physiol Biochem*. 2023;201: 107857.
 95. Ruan Q, Bai X, Wang Y, Zhang X, Wang B, Zhao Y, et al. Regulation of endogenous hormone and miRNA in leaves of alfalfa (*Medicago sativa* L.) seedlings under drought stress by endogenous nitric oxide. *BMC Genomics*. 2024;25:1–19.
 96. Moon J, Parry G, Estelle M. The ubiquitin-proteasome pathway and plant development. *Plant Cell*. 2004;16:3181–95.
 97. Vierstra RD. The ubiquitin-26S proteasome system at the nexus of plant biology. *Nat Rev Mol Cell Biol*. 2009;10:385–97.
 98. Arya P, Acharya V. Plant STAND P-loop NTPases: a current perspective of genome distribution, evolution, and function: plant STAND P-loop NTPases: genomic organization, evolution, and molecular mechanism models contribute broadly to plant pathogen defense. *Mol Genet Genomics*. 2018;293:17–31.
 99. Small ID, Schallenberg-Rüdingler M, Takenaka M, Mireau H, Ostersetzer-Biran O. Plant organellar RNA editing: what 30 years of research has revealed. *Plant J*. 2020;101:1040–56.
 100. Shikanai T. RNA editing in plants: Machinery and flexibility of site recognition. *Biochim Biophys Acta - Bioenerg*. 2015;1847:779–85.
 101. Ichinose M, Sugita M. RNA editing and its molecular mechanism in plant organelles. *Genes (Basel)*. 2017;8:5.
 102. Allario T, Brumos J, Colmenero-Flores JM, Tadeo F, Froelicher Y, Talon M, et al. 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. *J Exp Bot*. 2011;62:2507–19.
 103. Palumbo F, Gabelli G, Pasquali E, Vannozzi A, Farinati S, Draga S, et al. RNA-seq analyses on gametogenic tissues of alfalfa (*Medicago sativa*) revealed plant reproduction- and ploidy-related genes. *BMC Plant Biol*. 2024;24:826.
 104. Mendiburu AO, Peloquin SJ. The significance of 2 N gametes in potato breeding. *Theor Appl Genet*. 1977;49:53–61.
 105. Lavania UC, Srivastava S, Lavania S, Basu S, Misra NK, Mukai Y. Autopolyploidy differentially influences body size in plants, but facilitates enhanced accumulation of secondary metabolites, causing increased cytosine methylation. *Plant J*. 2012;71:539–49.
 106. Tang Q, Xu Y, Gao F, Xu Y, Cheng C, Deng C, et al. Transcriptomic and metabolomic analyses reveal the differential accumulation of phenylpropanoids and terpenoids in hemp autotetraploid and its diploid progenitor. *BMC Plant Biol*. 2023;23:1–13.
 107. Dang Z, Xu Y, Zhang X, Mi W, Chi Y, Tian Y, et al. Chromosome-level genome assembly provides insights into the genome evolution and functional importance of the phenylpropanoid-flavonoid pathway in *Thymus mongolicus*. *BMC Genomics*. 2024;25:1–19.
 108. Pecetti L, Berardo N, Odoardi M, Piano E. Forage quality components in grazing-type lucerne (*Medicago sativa* L. complex). *J Agron Crop Sci*. 2001;187:145–52.
 109. Avato P, Migoni D, Argentieri M, Fanizzi FP, Tava A. Activity of saponins from *Medicago* species against HeLa and MCF-7 cell lines and their capacity to Potentiate Cisplatin Effect. *Anti-cancer Agents Med Chem (formerly Curr Med Chem. Anti-Cancer Agents)*. 2017;17:1508–18.
 110. Liu C, Zhang T. Expansion and stress responses of the AP2/EREBP superfamily in cotton. *BMC Genomics*. 2017;18:1–16.
 111. Sun Q, Wang G, Zhang X, Zhang X, Qiao P, Long L, et al. Genome-wide identification of the TIFY gene family in three cultivated *Gossypium* species and the expression of JAZ genes. *Sci Rep*. 2017;7:1–9.
 112. Xie Z, Nolan TM, Jiang H, Yin Y. AP2/ERF transcription factor regulatory networks in hormone and abiotic stress responses in *Arabidopsis*. *Front Plant Sci*. 2019;10:1–17.
 113. Shi X, Yang H, Chen C, Hou J, Ji T, Cheng J, et al. Dosage-sensitive miRNAs trigger modulation of gene expression during genomic imbalance in maize. *Nat Commun*. 2022;13:1–15.
 114. Ha M, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ, et al. Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. *Proc Natl Acad Sci U S A*. 2009;106:17835–40.
 115. Rao S, Tian Y, Xia X, Li Y, Chen J. Chromosome doubling mediates superior drought tolerance in *Lycium ruthenicum* via abscisic acid signaling. *Hortic Res*. 2020;7:1–18.
 116. Lo Giudice C, Pesole G, Picardi E. RED1db 3.0: a comprehensive collection of RNA editing events in plant organellar genomes. *Front Plant Sci*. 2018;9:1–9.
 117. Marroni F, Pinosio S, Morgante M. Structural variation and genome complexity: is dispensable really dispensable? *Curr Opin Plant Biol*. 2014;18:31–6.
 118. Zhou Q, Tang D, Huang W, Yang Z, Zhang Y, Hamilton JP, et al. Haplotype-resolved genome analyses of a heterozygous diploid potato. *Nat Genet*. 2020;52:1018–23.
 119. Sun S, Zhou Y, Chen J, Shi J, Zhao H, Zhao H, et al. Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nat Genet*. 2018;50:1289–95.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.