

Review

Not peer-reviewed version

Homoeologs in Allopolyploids: Navigating Redundancy as Both an Evolutionary Opportunity and a Technical Challenge - A Transcriptomics Perspective

[Gaetano Aufiero](#) , [Carmine Fruggiero](#) , [Davide D'Angelo](#) , [Nunzio D'Agostino](#) *

Posted Date: 4 July 2024

doi: 10.20944/preprints202407.0415.v1

Keywords: Polyploids; homeologs; allopolyploidization; RNA-sequencing; gene expression; expression level dominance; homoeolog expression bias; additivity; bioinformatics



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Review

Homoeologs in Allopolyploids: Navigating Redundancy as Both an Evolutionary Opportunity and a Technical Challenge - A Transcriptomics Perspective

Gaetano Aufiero ^{1,†}, Carmine Fruggiero ^{1,†}, Davide D'Angelo ¹ and Nunzio D'Agostino ^{1,*}

Department of Agricultural Sciences, University of Naples Federico II, 80055, Portici, Naples, Italy; gaetano.aufiero@unina.it, carmine.fruggiero@unina.it, davide.dangelo@unina.it, nunzio.dagostino@unina.it

* Correspondence: nunzio.dagostino@unina.it; Tel.: +39 0812539486

† These authors contributed equally to this work.

Abstract: Allopolyploidy in plants involves the merging of two or more distinct parental genomes into a single nucleus, a significant evolutionary process in the plant kingdom. Transcriptomic analysis provides valuable insights into the fate of duplicated genes, evolutionary novelties, and environmental adaptations exhibited by allopolyploid plants. However, transcriptome profiling is challenging due to genomic redundancy, which is further complicated by the presence of multiple chromosomes sets and the variations among homoeologs and allelic genes. Prior to transcriptome analysis, sub-genome phasing and homoeology inference are essential for obtaining a comprehensive view of gene expression. This review aims to clarify the terminology in this field, identify the most challenging aspects of transcriptome analysis, explain their inherent difficulties, and suggest reliable analytic strategies. Furthermore, bulk RNA-seq is highlighted as a primary method for studying allopolyploid expression, with an emphasis on the critical steps of read mapping and normalization in differential gene expression analysis.

Keywords: polyploids; homeologs; allopolyploidization; RNA-sequencing; gene expression; expression level dominance; homoeolog expression bias; additivity; bioinformatics

1. Introduction

Transcriptomics involves studying the complete set of RNA transcripts (i.e., the transcriptome) present in a specific cell or tissue type at a particular developmental stage and/or under specific physiological conditions.

In allopolyploid plants, the transcriptome reflects the expression of duplicated genomes resulting from hybridization, which can lead to complex gene expression patterns, including homoeolog expression bias, condition-specific differential homoeolog usage, expression level dominance, and transgressive regulation. Analyzing transcriptomes in allopolyploid plants helps researchers understand the evolution of duplicated genes [1–4] and how these plants adapt to new environments [5,6]. Additionally, it aids in identifying novel genes and isoforms [7–9] that may be important for breeding and biotechnological applications.

Due to the coexistence of highly similar gene pairs, which originated through speciation and were brought back together in the same genome by allopolyploidization (homoeologs genes) [10], transcriptome analysis in allopolyploid species poses numerous bioinformatics challenges. These include choosing optimal strategies and parameters for the assembly and mapping of RNA-seq reads, avoiding chimeric assemblies of homoeolog gene copies, distinguishing individual homoeolog expression levels, and detecting novel genes, alternative splicing and polyadenylation events [11–18]. Therefore, transcriptome analysis in allopolyploid is both a valuable and challenging research topic.

2. Terminology in Allopolyploid Gene Expression

Recent studies on gene expression in allopolyploids have revealed confusion regarding the terminology used. Clear definitions are essential to avoid misunderstandings and ensure effective scientific communication. This confusion often stems from variations in analytical methods and differences in the target organisms studied.

The term “genome dominance”, “genomic expression dominance” or “expression dominance” in allopolyploids was originally described by [19]. They used these terms to describe situations where the total expression level of a gene or group of homoeolog genes equals that of one of the parental genomes. Their study focused on a synthetic allopolyploid (A2D1), obtained from the inter-specific cross between the two diploid species *Gossypium arboreum* (genome A2) and *Gossypium thurberi* (genome D1) followed by chromosomal doubling with colchicine. In this synthetic allopolyploid, they observed that the expression levels of approximately 11,000 genes, both up- and down-regulated, were statistically equivalent to those of the D1 genome. Further studies on genome dominance in allopolyploid cotton were conducted by [20,21] applied the same concept to *Coffea* allopolyploid. Similarly, the term “parental dominance” has been used to describe a condition where the overall gene expression of the allopolyploid mirrors the overall expression level of one parent [22–24].

In other studies, the term “genome dominance” or “sub-genome dominance” is used to describe certain phenomena, e.g., lower DNA methylation, higher gene retention and preferential expression, attributable to one of the parental genomes of the allopolyploid [25–29]. Recently [30] defined “genomic dominance” in allopolyploids more generally as the situation in which one parental genome becomes dominant over the other. [31] suggested replacing the term “genome dominance” as defined by [19] with “expression level dominance”, to avoid the ambiguity of the word “genome”. Furthermore, they pointed out that “expression level dominance” cannot be detected without measuring the expression level in the two parental genomes. Therefore, it is difficult to calculate this value in paleopolyploids because the parents may have become extinct or have undergone substantial genomic and/or transcriptomic changes [32]. This concept was also highlighted in a review by [33]. The “expression level dominance” is independent of the relative contribution to expression level of the two homoeologs. In other words, an allopolyploid derived from two parental species (A and B), which differ in the expression level of gene X (10 and 20, respectively), might express gene X (as the overall expression of the two homoeologs) at the same level as one of the parents (i.e., 10 or 20). This expression level is independent of the relative contribution of the two homoeologs in the allopolyploid, as they could contribute equally or unequally to the total expression of gene X. Currently, several studies have used the term “expression level dominance” as a synonym for the term “genome expression dominance” sensu [19,29,30,34–38]. “Expression level dominance” could be balanced (the same number of duplicated gene pairs in the allopolyploid have the overall expression level of one of the parental species, e.g., considering 10 duplicated gene pairs in the allopolyploid, the total expression of 5 homoeolog gene pairs is equivalent to one of the parents while the total expression of the remaining 5 homoeolog gene pairs is equivalent to the other parent) or unbalanced (more homoeolog genes than half have the expression level of one of the parental species). According to [31], the term “homoeolog expression bias” can be used to describe the condition in which a duplicated gene in an allopolyploid is preferentially expressed by one homoeolog, generally resulting in an unequal expression level between the homoeologs (Figure 1). As observed for expression level dominance, homoeolog expression bias could be balanced (there is an equal number of genes preferentially expressed by each homoeolog, e.g., considering 10 duplicated genes, 5 are preferentially expressed by one homoeolog and the remaining 5 are preferentially expressed by the other) or unbalanced (the number of expressed genes favor one homoeolog over other). In other words, if the expression level of a gene is equivalent between the two diploid parents but becomes unequal between the homoeologs in an allopolyploid, then the gene exhibits an expression bias. Similarly, this bias is also observed if the diploid parents display non-equivalent expression levels while the homoeologs in the allopolyploid show equivalent expression levels [39]. This last case is more described as “no bias in progeny” [35,36,38], whereby, an expression bias in the parent is reverted to non-differential expression between homoeologs. It can be deduced

that to define an expression bias between homoeologs, as in the case of “expression level dominance”, it is necessary to also take into consideration the expression levels of the parental genomes. Therefore, three conditions generally occur (synthetically originated, naturally originated, missing parents).

In the “synthetically originated” condition, the allopolyploid has been synthetically originated and the expression levels of its parents are available [35]. In this case, there is great interest in studying the phenomena that are generally defined as ‘genome shock’ and ‘transcriptome shock’ [40–42]. That is, all those phenomena of rapid reorganization of genome and gene expression that occur after hybridization and duplication.

In the “naturally originated” condition, the allopolyploid originating from the natural hybridization of its two parent species and expression levels of its parents are available [38]. In this case, it is particularly interesting to observe which genetic, genomic, and epigenetic changes have occurred since the hybridization and polyploidization event over a long period. Nevertheless, the great challenge is represented by the identification of the right parentals (i.e., the expression profile of the parental species that generated the allopolyploid). Considering that not only the allopolyploid but also its parents, assuming they still exist, have undergone evolutionary mechanisms, we cannot be certain that the identified parents are truly representative of the original parents, especially due to polyploidization events that occurred several million years ago [32].

In the two conditions described above, if the expression levels of the parents are not equal, the expression bias of the homoeologs has been evaluated by comparing the expression value of each homoeolog with a parental mix expression value [20] or more frequently with the parental expression level estimated separately [35,38].

In “missing parents” condition, the parental expression levels are not available, as long as it is possible to identify the parental genomes in the allopolyploid, and an equal expression level of the two parental genomes is usually assumed [25,43]. If we assess homoeolog expression bias without knowing the parental expression levels, we cannot dismiss the possibility that the bias between homoeologs is merely a reflection of parental inheritance. Thus, the expression differences observed between homoeologs might have already been present in the parental species (Figure 1). In other words, for a given gene or group of genes, the relative contribution of expression level between homoeologs in the allopolyploid is equal to the relative contribution of expression level between parents. This situation could be described as an additive expression of homoeologs [33,44]. But generally, the term “additive” describes a situation in which the total expression of both homoeologs in an allopolyploid, for a gene or a group of genes, equals the arithmetic mean, also known as the mid-parent value (MPV) [20,24,34,35,38], or as discussed by [45], the sum of parental species expression (also defined as summed-parent value “SPV”). According to [45] the term “MPV” is used to calculate additivity by assuming that each parent contributes half of their genome to the offspring. However, in the case of neo-polyploidy, where the offspring receives each parent’s entire genome, parental values should be summed, not averaged, if they are truly additive. This is because the entire genome, rather than half, is passed on to offspring. However, MPV is the generally used method to calculate the additive expression value (Figure 1). As discussed by [33], there are at least two ways in which MPV is calculated. Either by measuring gene expression levels in a mixture of RNA from both parents, as demonstrated in studies using microarray approaches [46–48], or, more recently, through RNA-sequencing technology, where reads are counted relative to the expression levels of each parent separately and then averaged [34,35,38]. For example, suppose we have two parental species A and B, that differ in the expression level of the X gene (10 and 20, respectively). The MPV of X gene is $(10 + 20)/2 = 15$. If the allopolyploid derived from the two parental species A and B has an expression of X gene equal to 15, then we can say that the expression of X gene is additive in the allopolyploid. The additive condition of parental expression is generally used as reference to compare the overall allopolyploid expression of a gene or a group of genes, i.e., the zero hypothesis. If overall expression of allopolyploid is not additive, then it might be equally to one of the parents (expression level dominance) or it might be lower (transgressive down-regulation) or higher (transgressive up-regulation) than both parental expressions (Figure 1). In transgressive expression the gene expression is outside the range of parental expression level. For example, if the expression level of the X gene is

10 and 20 in parental species A and B, respectively, then the expression level of the same gene in allopolyploid is transgressive up-regulated if it is higher than 20, or transgressive down-regulated if it is lower than 10. In practice, differentially expressed genes (e.g., $|\log_2 \text{ fold change}| \geq 1$ and $\text{padj} \leq 0.05$) between the allopolyploid (considering the joint expression level of both homoeologs) and its parents are usually binned into 12 conditional states. The first binned could be done dividing genes whose expression is not distinguishable from additivity (e.g., $\text{padj} \geq 0.05$) and distinguishable from additivity (e.g., $\text{padj} \leq 0.05$). Genes in the latter group could be further classified into more specific nonadditive states (i.e., expression level dominance and transgressive down/up regulated) [19,20,34,35,38]. According to [19], the 12 differentially expressed states, distinguished by Roman numbers, are: additive (categories I, XII), expression level dominance (categories II, IV, IX, XI) and transgressive regulation (categories III, V, VI, VII, VIII, X) (Figure1). Genes not differentially expressed between the allopolyploid and the parents are assigned to the “no change” state (Figure 1). It is important to note that conditions mentioned above can be influenced by various factors, including the specific genes involved, the tissues being analyzed, and the environmental conditions [36].



Figure 1. According to [19], the 12 states, in which the differentially expressed genes between allopolyploid and parents are binned, are represented by Roman numbers from I to XII. The area of the circle represents the overall expression level of a single gene. AA, BB and AABB represent the two diploid parental species and allopolyploid species, respectively. Additivity (first and second rows) is calculated as mid-parent value (MPV). AA-ELD and BB-ELD (from third to sixth rows) represent the expression level dominance in favor of the parental genome AA and BB, respectively, i.e., the overall expression level of both homoeologs, relative to a gene, is equivalent to the overall expression level, for the same gene, of one of the parental species (AA or BB). TDR and TUR (seventh to twelfth) represent transgressive down-regulation and transgressive up-regulation, respectively, i.e., the overall expression level of both homoeologs, relative to a gene, is outside the range of parental expression level for the same gene. NO CHANGE (thirteenth row) represents the absence of differential expression genes between parents and the allopolyploid. The relative contribution of homoeolog expression can either proportionally reflect pre-existing differences between the parental species (parental legacy) or show bias towards one of the parental species (homoeolog expression bias, illustrated by two examples per state).

3. How Allopolyploidy Affects Bioinformatics Transcriptome Analyses?

Analysis to estimate gene expression profiles in allopolyploids is challenging, particularly in the steps of read mapping and normalization of gene expression levels. Intrinsic genomic redundancy complicates the analysis, the complexity of which is further amplified in polyploids due to the existence of multiple sets of chromosomes and homoeological variation coexisting with allelic variation. Allopolyploidization leads to genetic, epigenetic, transcriptomic and proteomic variations, compared to parental species, often associated with the activities of repetitive sequences and transposon elements [41,49–51].

As discussed above, in allopolyploids, the relative expression levels of homoeolog gene pairs can be unequal. While, compared to the parental species, the cumulative expression of each homoeolog pairs can be additive or not. Furthermore, homeologs could be mutually silenced on a tissue-specific basis [36,52]. Early research, based on the use of microarrays, often failed to distinguish expression between homoeolog genes [19,48]. This limitation meant that scientists could only assess the cumulative expression of a target gene, lacking information on the regulation of individual homoeolog copies, i.e., the relative contributions of each homoeolog copy to the overall expression were elusive. Consequently, when comparing gene expression between allopolyploid species and their parents, scientists could only infer whether the combined expression of homoeolog copies was comparable to MPV or parent's expression levels.

High-throughput RNA-sequencing and the availability of bioinformatics and genomic resources (such as the OneKP project, an international initiative that has generated large-scale gene sequencing data for over 1,000 phylodiverse plant species [53], and PloiDB, a database of ploidy estimates for seed plants ([54]), have significantly improved our ability to define gene expression profiles in allopolyploids and appropriately distinguish gene expression levels of homoeolog genes.

Gene expression levels in both allopolyploids and diploid organisms are typically measured using high-throughput RNA-sequencing (bulk RNA-seq). It targets mRNA and/or non-coding RNA libraries for the identification of differentially expressed genes. It aims to estimate the average of global expression values in each sample, elucidating the molecular mechanisms involved in a particular tissue at a given time point. In these experiments, mRNA is extracted, purified, and converted into cDNA. The cDNA is then fragmented and sequenced using next-generation technologies. The resulting sequenced reads are mapped to a reference genome or transcriptome, and the number of reads mapped to each gene, or other biological units indicates its expression level.

Furthermore, [55] conducted a study on the allopolyploid *Gossypium hirsutum* using single-cell RNA-seq (scRNA-seq) to investigate the mechanism of cotton fiber cell initiation from ovule epidermis. scRNA-seq enables profiling of the entire transcriptome of numerous individual cells isolated using microscopic droplets or wells. Molecular barcodes are attached to sequences before amplification to identify the cell origin of each transcript. The core of scRNA-seq analysis is the expression matrix, detailing the transcript counts for each gene and cell, which undergoes data

preprocessing and subsequent downstream analyses. Unlike bulk RNA-seq, scRNA-seq provides more precise insights into the heterogeneity of gene expression across different cell types and the molecular trajectory of cell differentiation during development. Nevertheless, bulk RNA-seq remains the primary method for studying expression in polyploids.

Better identification of homoeologs expression now allows for a deeper understanding of allopolyploidy effects on gene expression and regulation. This includes exploring homoeolog expression bias, dosage compensation, and epigenetic modifications, all of which play a critical role in phenotype variation, environmental responses, and progress in crop improvement. Choosing the best strategy to study the transcriptome of allopolyploids depends on several factors, including the nature of the biological question, the range of bioinformatics tools available, the quantity and divergence of sub-genomes within the allopolyploid, and the available information on the parental species.

Therefore, it is necessary to consider different analytical approaches in order to optimize the analysis.

The following paragraphs aim to review the intricate steps and prerequisites involved in transcriptome analysis of allopolyploids, focusing on RNA-seq. These include sub-genome phasing, homoeology inference, RNA-sequencing, read mapping and normalization in identification of differentially expressed genes.

4. Sub-Genome Phasing

In allopolyploid species multiple parental genomes, called sub-genomes, coexist in a single nucleus. A key step is to assign each homoeolog chromosome and gene copy to the corresponding sub-genome. Notoriously, sub-genome identification is simplified if the genomes of putative progenitors are characterized. Nonetheless, the assignment of homoeologs to the corresponding sub-genomes remains challenging due to divergence between the parental and allopolyploid genomes, the high similarity between sub-genomes, and the reshuffling events that characterize the allopolyploid genome. This reorganization includes phenomena such as the loss of chromosomes, homoeologs, and duplicated genes, known as “fractionation” [56–59]. Fractionation, if it affects the entire polyploid genome, contributes to “diploidization.” Homoeolog exchanges represent another aspect of this dynamic process, which involves the mispairing of chromosomes of different sub-genomes [60–63]. Because of these exchanges, DNA fragments of varying sizes are swapped between sub-genomes, potentially leading to deletions, duplications, and translocations. Although most recombinations between sub-genomes occur between homoeolog chromosomes, occasionally they also occur between non-homoeolog regions [62].

The identification of sub-genomes is often necessary as preparatory information for homoeolog annotation and correctly distinguishing them from paralogues and allelic variants. Additionally, this identification enables the analysis of differential expression between sub-genomes, offering insights into their distinct regulatory mechanisms and enhancing our understanding of the biology and evolution of the studied organism.

Sub-genome inference is achieved using supervised and unsupervised bioinformatic approaches. The supervised approach exploits data and sequences from diploid parents or closely related species. When such information is available, sub-genomes can be distinguished by mapping allopolyploid-derived DNA- or RNA-seq data on one or both reference genomes, taking advantage of sequence similarity or homeolog-specific polymorphisms [64–66]. To achieve this, specialized mapping and categorization algorithms designed specifically for allopolyploid organisms can be used, including PolyCat [67], SNIploid [64], and HANDS2 [68], particularly if a single reference genome is accessible. Alternatively, PolyDog [69] is instead used when both parental genomes are available (more details about mapping and categorization algorithms will be given in the subsequent paragraph). Further evidence to sub-genome assignment can be obtained from multi-locus synthetic phylogenetic trees [27] and synonymous nucleotide substitution approaches implemented in the WGDI pipeline [70].

When both the sub-genomes and the parental species genomes have been diverging for a long period of time, the sub-genomes are not easily distinguishable. To overcome this hurdle, the sub-genomes separation can be based on large phylogenetic data collection [71]. Specifically, by evaluating the phylogenies of all individual genes within the allopolyploid species and their homologs in the parental species, it becomes feasible to rank parental leaves based on distance. This process allows for the assignment of parental origins to each homoeolog gene pairs by selecting the leaf with the shortest phylogenetic distance.

The unsupervised strategy comes into play when diploid ancestors are unknown and the methods rely on differences in sub-genome properties, such as GC content or the presence of sub-genome-specific repetitive sequences (e.g., transposable elements). Various programs, including PolyCRACKER [72] and SubPhaser [73], implement this methodology and require chromosome-scale genome assembly. Recently, [74] provided a valuable benchmark and recommendations for assessing the accuracy of the WGDI and SubPhaser tools. Additionally, a comprehensive review on allopolyploid sub-genome identification was conducted by [75].

5. Homoeology Inference

The identification of homoeologs in transcriptome analysis is essential for obtaining a detailed and comprehensive view of gene expression. It helps determine whether varying expression levels of homoeologs across different sub-genomes influence the phenotype and reveals whether homoeologs have retained similar functions (sub-functionalization) or acquired new functions (neo-functionalization).

The process of identifying ortholog genes, i.e., genes derived from a single ancestral gene and now present in different species [76], overlaps in some ways with the identification of homoeologs in allopolyploids [10]. Indeed, homoeologs within the sub-genomes of an allopolyploid species could be considered as orthologs between distinct species, given their common origin from speciation events. Their identification does not imply that the genes have the same function but rather that they arise from a speciation event [76]. Nevertheless, in comparative genomics the “ortholog conjecture” states that ortholog genes, and therefore also homeologs, have similar functions [77]. In comparative genomics, homologs genes resulting from a gene duplication event are called paralogs, and those from recent duplications can maintain similar functions compared to distant paralogs separated by millions of years of evolution [78]. Paralogs that resulted from a duplication event that occurred before speciation are called outparalogs, while those that occurred from a post-speciation duplication are called inparalogs [79]. Since inparalogs can be grouped under a single gene and defined as co-orthologs (“orthologs group”), not all orthologs in the same group arise from speciation. This situation could be considered a “terminology muddle,” as defined by [80], since the term extends the relationship to include inparalogs relative to another gene in a different species [81]. In the context of a hybridization event, these genes would classify as homoeologs to each other [82]. According to this classification method, homoeolog relationships could vary from one-to-one, one-to-many, or many-to-many, depending on the number of duplications since the speciation of the progenitors [10]. It is important to note that in allopolyploids and polyploids in general, the presence of redundant genes as paralogs allows the latter to acquire, relatively quickly, mutations, leading to the emergence of sub- neo- and non-functionalizations [2]. Outparalogs, having had more time to evolve and potentially change function compared to in-paralogs, are typically not considered co-orthologs [83]. Similarly, they should not be classified as homoeologs for the same reason.

Programs developed to determine orthologs could be adapted to infer homoeologs in allopolyploid species, given the parallels between the two. This often requires a priori knowledge of sub-genomes in allopolyploids. Bioinformatics inference of homoeologs involves a combination of computational and analytical approaches. The choice of method depends on the specific requirements of the study and available resources. Various programs have been developed for this purpose; these can be mainly divided into graph-based and tree-based approaches. The graph-based method assumes that orthologs or homoeologs share sequence similarities. The graph construction phase identifies orthologs or homoeologs through similarity searches, considering pairs of genomes

or sub-genomes at a time. In these graphs, genes (or proteins) are depicted as nodes, and their evolutionary relationships are represented as edges [84].

The Bidirectional Best Hits (BBH) approach is the simplest and most widely used method based on sequence similarity [85]. This method assumes that two genes are homoeologs if each gene is the best match of the other when their respective sub-genomes are compared. In simpler terms, for a gene pair to be classified as homoeologs, each gene must show higher similarity (e.g., highest alignment score) to its pair than to all other genes in the other sub-genome [86]. Similarity is typically assessed using BLAST [87] or other sequence alignment algorithms. However, the BBH method has a limitation: it can only identify orthology in a 1-to-1 relationship [88]. To complement this method, synteny (the conservation of gene order between two chromosomal regions [89]) is often employed as additional evidence for identifying homoeologs [90]. Even when both methods are used, the ability to detect genes that have duplicated and relocated, both before and after allopolyploidization events, remains a challenge. For instance, in plants where gene duplication rates are high, the BBH method alone may fail to identify up to 60% of orthologous relationships in datasets involving 12 Viridiplantae species [88]. Even when BBH and synteny methods are combined, they may still fail to detect up to 26% of homoeologous relationships in cotton [86]. To address this limitation, other algorithms such as OMA [91] and OrthoFinder [92] have been developed to incorporate inparalog relationships into the graph construction process. The OMA pipeline has been adapted to infer homoeologs in recent studies [82]. In these studies, homoeologs were inferred in three polyploid species: upland cotton (*Gossypium hirsutum*), rapeseed (*Brassica napus*), and bread wheat (*Triticum aestivum*). Each sub-genome was treated as a separate genome, and orthologs were inferred using the standard OMA pipeline [91]. Similarly, [93] identified homoeologs in a synthetic allotetraploid wheat resulting from a cross between the diploid parent species *T. urartu* (AA) and *Ae. tauschii* (DD). They inferred orthologs between parental genomes using OrthoFinder.

Tree-based methods rely on the assumption that genes share an evolutionary history and employ a resolution technique known as “tree reconciliation.” In this approach, the structure of a gene tree is compared with that of a species tree, and they are reconciled based on the principle of parsimony, aiming to infer the least number of duplication and loss events during the gene’s evolution [83]. To begin this method, a multiple alignment of homologous sequences is first generated, which forms the basis for constructing a phylogenetic tree of the gene family. During the reconciliation phase, the nodes of this gene tree are classified as either duplication or speciation events by comparing them with the nodes of the species tree. This method is implemented in various software, including Ensembl Compara [94]. The Ensembl Compara pipeline utilizes the TreeBeST algorithm (<http://treesoft.sourceforge.net/treebest.shtml>), originally developed for TreeFam [95], to infer reconciled gene trees. This pipeline was utilized by [96] to infer orthologous genes for studying the evolutionary history of the polyploid family Cyprininae.

The effectiveness of tree-based methods hinges heavily on the accurate construction of multiple alignments and trees, which can be computationally intensive and thus limit their application to large datasets [78]. A hybrid approach that integrates graph- and tree-based methods has been implemented in HomoloGene. In contrast, other software employs a meta-approach to leverage predictions generated by multiple programs, such as MARIO [97] and MetaPhOrs 2.0 [98]. These methods have been extensively reviewed in several studies [10,78,83,86,99]. It is important to note that when inferring homoeologs based on RNA-seq-derived sequences, misidentification may occur if one or both homoeolog genes are not expressed.

6. RNA-Sequencing

In recent years, expression profiling in allopolyploid species has commonly involved the analysis of RNA-seq data obtained through next-generation sequencing (NGS) technology, with a notable preference for short-read sequencing. Despite the availability of several NGS technology providers, these short reads are often generated using Illumina instruments due to lower cost, high throughput and availability for a wide range of tools and pipelines also suitable for polyploid species [100,101]. For example, through the Illumina HiSeq 2000 platform, [37] studied the expression

patterns of homolog genes between the allotetraploid cotton *Gossypium hirsutum* and its diploid progenitors *Gossypium arboreum* and *Gossypium raimondii* at the early fiber development stage. While, through Illumina NextSeq 500, [102] performed a *de novo* transcriptome assembly of the tetraploid species *Tripsacum dactyloides*, a sister genus of *Zea mays*, in order to demonstrate that similar genes may be decaying into pseudogenes in both genera after a shared ancient polyploidy event. Considering that the cost of sequencing a genome is affected by its size, the cost for polyploid organisms could be double that for their diploid counterparts. Recently, [18] explored the expression of homoeologs in allohexaploid wheat and demonstrated cost-effective sequencing using 3' UTR RNA-seq with the Lasy-Seq protocol [103]. However, using short reads poses challenges in transcriptome assembly and isoform identification, especially in allopolyploids where homoeolog genes often produce highly similar isoforms [104]. These challenges are typically addressed by leveraging RNA-seq long reads. Long-read sequencing technologies such as PacBio Iso-Seq [105] and Nanopore-Based Direct RNA sequencing [106] offer improvements in *de novo* assembly, mapping accuracy, identification of transcript isoforms, and detection of structural variants. Moreover, direct RNA sequencing from native molecules eliminates amplification biases while preserving base modifications [107]. [104] used long-read sequencing to identify approximately 175,000 isoforms in nearly 45,000 genes in allotetraploid cotton, revealing that approximately 50% of homoeolog genes produce divergent isoforms in each sub-genome. Consistent with these findings, [108] subsequently identified numerous novel transcripts in response to salt stress. Additionally, [109] utilizing PacBio Iso-Seq, performed a comprehensive transcriptome analysis of the allopolyploid *Brassica napus*, leading to the construction of an isoform sequencing database.

7. Mapping Reads

Transcriptome analysis in allopolyploids can produce two types of valuable data. First, uniquely mapped reads can reveal the expression levels of homoeologs, provided there are sufficient polymorphisms between homoeolog sequences. Second, ambiguously mapped reads can indicate the absolute expression abundance for each pair of homoeologs. The mapping process can use the parental genomes as references, treating homoeologs as orthologs [12], or it can be based on the assembled genome of the allopolyploid itself.

Assembling an allopolyploid genome presents significant challenges, exacerbating the difficulties already encountered in diploid genome assembly. One major issue is the incorrect assembly of fragments from one sub-genome into another, leading to chimeric and fragmented contigs. This challenge is primarily due to repetitive sequences, such as satellite DNA, DNA transposons, retrotransposons and variation in gene copy-number. These repetitive elements are prevalent in allopolyploid genomes, and their accurate detection is essential for correct genome annotation [110–112].

It is important to note that using reference sequences discordantly annotated can distort the assessment of differential expression of homoeologs. This issue is particularly significant because annotations, often based on RNA-seq data, may incompletely or incorrectly annotate homoeolog exons, especially for genes expressed at low levels. Consequently, depending on the used tool, the mapping stage may be restricted to accurately annotated loci in both sub-genomes. In the absence of a reference genome, *de novo* assembly of the transcriptome involves reconstructing the transcript set from multiple samples, potentially under varying conditions, to achieve a transcriptome as comprehensive as possible. However, assembling the transcriptome of an allopolyploid presents nearly two additional challenges compared to its diploid parental species. The first challenge, known as “homoeologs collapse,” occurs when two homoeologs are erroneously assembled into a single hybrid transcript. The second challenge, referred to as “SNP shuffling” arises when a homoeolog is assembled incorporating polymorphisms from the other homoeolog, complicating accurate sequence reconstruction [113]. In cases where parental transcriptomes or genomes need to be assembled, accurately identifying the parents is crucial but not always straightforward. This complexity underscores the meticulousness required in transcriptome assembly of allopolyploids to avoid these pitfalls and ensure robust data interpretation.

Once reference sequences are identified, two distinct mapping approaches can be used. The first approach utilizes traditional algorithms originally developed for diploid organisms. Well-known tools such as STAR [114] and HISAT [115], are used to align RNA-seq reads directly to the reference genome. Additionally, pseudoalignment programs like Kallisto [116] provide alternative methods for efficient sequence alignment based on transcript abundance estimation. The second approach involves specialized pipelines specifically designed for polyploid organisms. These methods typically fall into two categories:

- Similarity-based mapping: These pipelines compare allopolyploid reads to loci in the genomes of the parental species to identify homoeolog sequences.
- Genotype-aware mapping: These pipelines aim to detect genotypic differences and classify allopolyploids reads to the corresponding sub-genomes.

Both approaches play crucial roles in overcoming the challenges unique to allopolyploid transcriptome analysis, ensuring accurate quantification of gene expression and facilitating comprehensive understanding of gene regulation in allopolyploid organisms. As a result, these pipelines generate sub-genome specific BAM files that serve as inputs for downstream bioinformatics analyses. Several tools leverage similarity search for such analyses, including HomeoRoq [117] and PolyDog. The HomeoRoq pipeline requires genomic sequences from both parental organisms and employs STAR to map RNA-seq reads to each reference sequence independently. Reads are then classified based on mismatch counts to determine their genomic origin. Similarly, PolyDog uses genomic sequences from both parental genomes as references. It employs GSNAP [118] as its default mapping tool to align RNA-seq reads. The final assignment of reads to specific sub-genomes is determined by considering quality scores, alignment length, and mismatch quantity.

Several tools utilize genotypic differences to classify RNA-seq reads to their respective sub-genomes in allopolyploid organisms. Among these tools are PolyCat and EAGLE-RC [119]. PolyCat uses a single reference sequence along with GSNAP as its default mapping tool. It utilizes SAMtools [120] for SNP calling to identify “homoeo-SNPs” between parental diploid reads and the reference genome. PolyCat then generates a SNP index, which is used to categorize allopolyploid reads into their respective sub-genomes based on matches with identified homoeo-SNPs. Similarly, EAGLE-RC utilizes an alignment in BAM format and a set of variants in VCF format, along with a single reference genome. It classifies reads based on genotype differences, EAGLE-RC also offers a mode (--ngi mode) that directly classifies reads using the likelihood derived from alignments different sub-genomes, without explicit genotype information. When aiming to determine homoeologs expression level, benchmark studies have demonstrated that traditional software originally designed for diploid species results in higher error rates, particularly exceeding 10% when using pseudoalignment algorithms. However, using sub-genome classification pipelines can significantly reduce mapping errors. For instance, tools like EAGLE-RC achieve error rates below 1%, while HomeoRoq typically achieves less than 2%, depending on variables such as species, expression level, and read length. It is crucial to note that discrepancies in expression level assessment primarily stem from genes with low expression levels. These genes can cause significant shifts in the calculation of expression ratios between homoeologs [15].

7. Normalization in Differential Expression Analysis

Transcriptome profiling is commonly applied to identify differentially expressed genes. In differential expression analysis, high-throughput RNA-seq data is used to determine if there are statistically significant differences in read abundance at each locus across different experimental conditions. This analysis can involve comparing homoeolog genes within the allopolyploid genome or assessing the global expression levels of each sub-genome relative to their respective parental genomes. These comparisons are crucial for addressing diverse biological questions, including identifying and understanding the mechanisms that enable polyploids exhibiting superior vegetative growth compared to their progenitor species.

Before identifying differentially expressed genes, it is crucial to address biases by normalizing read counts obtained during the quantification step. Errors in normalization can lead to significant consequences in downstream analysis, such as an increased rate of false positives in differential expression analysis [121]. The primary goal of normalization is to mitigate biases so that differences in normalized read counts accurately reflect true differences in gene expression between samples. Thus, a gene is deemed differentially expressed across biological conditions if it shows varying mRNA levels per sample under those conditions [122]. The biases that mainly impact the detection of differential expression include gene length, GC content (the proportion of guanine and cytosine nucleotides), and library size (total number of sequenced reads *per* sample, i.e., sequencing depth). Gene length and GC content primarily introduce biases in comparisons of gene counts within a single sample ("within-sample" bias). These factors are often disregarded in differential expression analysis between two or more biological conditions. However, they can affect the accuracy of expression level quantification if not appropriately normalized or accounted for. On the other hand, library size plays a crucial role in comparisons of the same gene's counts across different samples ("between-sample" bias). Differences in sequencing depth between samples can lead to misleading conclusions about differential expressions if not adjusted during normalization.

In RNA-seq libraries, read counts directly correspond to relative gene abundances, meaning highly expressed genes can dominate, thereby reducing the number of reads available for less expressed genes. Consequently, increasing library size enhances the detection sensitivity for poorly expressed genes. Normalizing samples with different library sizes is crucial because without normalization, a sample sequenced at a higher depth than another would show inflated read counts for all genes. This could potentially lead to erroneous conclusions regarding differential expressions between samples. The simplest method of between-sample normalization involves scaling raw read counts within each sample by a sample-specific factor that reflects its library size [122]. Several data normalization methods have been developed to mitigate this bias in RNA-seq analysis. FPKM (Fragments Per Kilobase Million) [123] and TPM (transcript per million) methods [124] primarily focus on "within-sample" normalization, adjusting for gene length within individual samples [125]. In contrast, TMM (edgeR's Trimmed Mean of M values) [126] and RLE (DESeq's relative log expression) [127], are applied for "between-sample" normalization where differences in sequencing depth between samples are addressed. No normalization method is without limitations, and each approach has scenarios where its assumptions may not hold. For instance, methods like TMM and RLE are effective in managing differences in mRNA levels *per* cell across conditions when there are only a few differentially expressed genes, particularly in cases with minimal asymmetry among highly expressed genes. However, these methods may struggle to handle situations involving a global shift in gene expression, such as when there is a proportional increase in mRNAs *per* cell due to changes in ploidy levels. This scenario can lead to substantial asymmetry in the number of highly expressed genes between conditions [122]. Studies have highlighted instances where varying ploidy levels can cause significant global shifts in gene expression, thereby challenging the applicability of traditional normalization methods [128–130].

Therefore, the choice of a normalization strategy depends on the specific biological question being addressed and the characteristics of the samples under study. For example, when comparing samples from the same cell type in diploid organisms, the number of mRNA molecules *per* cell (transcriptome size) typically remains relatively stable. However, this consistency does not hold when comparing samples with different ploidy levels, as observed in [129,131,132]. This is because polyploid organisms often exhibit larger cells, which are positively correlated with increased transcriptome size [133,134].

Traditional sample preparation methods that solely standardize by sample concentration, without accounting for biomass or cell count, may result in inaccurate interpretations [135]. For instance, when assessing whether changes in gene expression correlate with variations in photosynthesis *per* unit of surface area, it can be more insightful to compare expression levels among samples with equal biomass quantities. This approach provides a more robust basis for analyzing gene expression patterns than relying solely on sample concentration-standardized preparations

[132]. Therefore, particularly when comparing organisms with different ploidy levels, it is crucial to consider various scale of comparison: *per* transcriptome (RNA concentration-based), *per* cell, and *per* biomass (surface area or weights) [132,135]. Using these different approaches, studies have demonstrated that the polyploid plant *Tolmiea menziesii* (which has a lower cell density compared to the diploid *T. diplomenziesii*) tends to maintain diploid-like transcript abundance *per* biomass but increasing gene expression *per* cell. Notably, this enhanced expression is enriched for functions related to photosynthesis [132].

To address normalization issues related to biomass or cell count and mitigate transcription size bias, scientists have proposed the use of multiple different exogenous “RNA spike-in” [136]. These RNA spike-ins, known for their precise concentrations, are added in proportion to the biomass or number of cells prior to RNA-seq library preparation. Subsequently, these spike-ins serve as scaling factors for between-sample normalization after quantification. Implementing RNA spike-ins requires prior knowledge of the biomass quantity or the number of cells included in each sample for accurate normalization. When direct cell counting is impractical, researchers can estimate cell numbers based on DNA content using techniques such as flow cytometry [130,136].

In the context of the scenarios mentioned earlier, analyzing differentially expressed genes in allopolyploids can present more complexity compared to standard RNA-seq analyses. A widely used approach was introduced by [19], where they performed comprehensive comparisons including relative expression levels between the two parents, the hybrid allopolyploid versus its parents, and the allopolyploid versus the mid-parent value. Recently, [137] adopted a similar strategy, utilizing RNA-sequencing data to explore the transcriptional responses to saline stress in cotton and heterosis events in root traits in rice. In their studies, they introduced HYBRIDEXPRESS, a tool designed to facilitate comparative analyses between allopolyploid hybrids and parental lines. This tool encompasses transcriptomic data processing, graphic visualization, data normalization, identification of differentially expressed genes, classification of genes based on the stages proposed by [19], and functional annotation of these genes. Such tools, alongside advancements in sequencing technologies and methodologies for homoeology inference and sub-genome identification, hold significant promise in elucidating parental legacies, homoeolog expression biases in allopolyploids, understanding their evolutionary mechanisms, and providing crucial insights for crop breeding.

8. Conclusions

Transcriptome analysis in allopolyploid plants is crucial for unraveling the complexities of gene expression resulting from genome duplication through hybridization. This process not only sheds light on the evolution of duplicated genes and their adaptive potential in new environments but also reveals novel genes and regulatory mechanisms important for agricultural and biotechnological advancements. However, conducting such analysis poses significant bioinformatics challenges due to the presence of highly similar homoeolog genes and the need for precise methods in read mapping, assembly, and expression quantification. Overcoming these challenges requires meticulous experimental design and the application of advanced analytic strategies to ensure accurate interpretation of the transcriptomic data. Despite these hurdles, the insights gained from transcriptome studies in allopolyploid plants promise to deepen our understanding of plant evolution and facilitate targeted improvements in crop breeding and genetic engineering.

Author Contributions: Conceptualization, G.A., C.F. and N.D.A.; investigation, G.A., C.F., D.D.A; writing—original draft preparation, G.A and C.F.; writing—review and editing, D.D.A and N.D.A.; supervision, N.D.A. All authors have read and agreed to the published version of the manuscript.

Funding: This study was carried out within the AgriTech National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1032 17/06/2022, CN00000022). This manuscript reflects only the authors’ views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

Informed Consent Statement: Not applicable

Data Availability Statement: Not applicable

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Flagel, L.E.; Wendel, J.F.; Udall, J.A. Duplicate Gene Evolution, Homoeologous Recombination, and Transcriptome Characterization in Allopolyploid Cotton. *BMC Genomics* **2012**, *13*, 302, doi:10.1186/1471-2164-13-302.
2. Roulin, A.; Auer, P.L.; Libault, M.; Schlueter, J.; Farmer, A.; May, G.; Stacey, G.; Doerge, R.W.; Jackson, S.A. The Fate of Duplicated Genes in a Polyploid Plant Genome. *Plant J* **2013**, *73*, 143–153, doi:10.1111/tpj.12026.
3. Conant, G.C.; Birchler, J.A.; Pires, J.C. Dosage, Duplication, and Diploidization: Clarifying the Interplay of Multiple Models for Duplicate Gene Evolution over Time. *Curr Opin Plant Biol* **2014**, *19*, 91–98, doi:10.1016/j.pbi.2014.05.008.
4. Mutti, J.S.; Bhullar, R.K.; Gill, K.S. Evolution of Gene Expression Balance Among Homeologs of Natural Polyploids. *G3 (Bethesda)* **2017**, *7*, 1225–1237, doi:10.1534/g3.116.038711.
5. Griffiths, A.G.; Moraga, R.; Tausen, M.; Gupta, V.; Bilton, T.P.; Campbell, M.A.; Ashby, R.; Nagy, I.; Khan, A.; Larking, A.; et al. Breaking Free: The Genomics of Allopolyploidy-Facilitated Niche Expansion in White Clover. *The Plant Cell* **2019**, *31*, 1466–1487, doi:10.1105/tpc.18.00606.
6. Van de Peer, Y.; Ashman, T.-L.; Soltis, P.S.; Soltis, D.E. Polyploidy: An Evolutionary and Ecological Force in Stressful Times. *The Plant Cell* **2021**, *33*, 11–26, doi:10.1093/plcell/koaa015.
7. Qin, J.; Mo, R.; Li, H.; Ni, Z.; Sun, Q.; Liu, Z. The Transcriptional and Splicing Changes Caused by Hybridization Can Be Globally Recovered by Genome Doubling during Allopolyploidization. *Molecular Biology and Evolution* **2021**, *38*, 2513–2519, doi:10.1093/molbev/msab045.
8. Li, M.; Hu, M.; Xiao, Y.; Wu, X.; Wang, J. The Activation of Gene Expression and Alternative Splicing in the Formation and Evolution of Allopolyploid Brassica Napus. *Horticulture Research* **2022**, *9*, uhab075, doi:10.1093/hr/uhab075.
9. He, W.; Zhang, X.; Lv, P.; Wang, W.; Wang, J.; He, Y.; Song, Z.; Cai, D. Full-Length Transcriptome Reconstruction Reveals Genetic Differences in Hybrids of *Oryza Sativa* and *Oryza Punctata* with Different Ploidy and Genome Compositions. *BMC Plant Biology* **2022**, *22*, 131, doi:10.1186/s12870-022-03502-2.
10. Glover, N.M.; Redestig, H.; Dessimoz, C. Homoeologs: What Are They and How Do We Infer Them? *Trends Plant Sci* **2016**, *21*, 609–621, doi:10.1016/j.tplants.2016.02.005.
11. Gruenheit, N.; Deusch, O.; Esser, C.; Becker, M.; Voelckel, C.; Lockhart, P. Cutoffs and K-Mers: Implications from a Transcriptome Study in Allopolyploid Plants. *BMC Genomics* **2012**, *13*, 92, doi:10.1186/1471-2164-13-92.
12. Krasileva, K.V.; Buffalo, V.; Bailey, P.; Pearce, S.; Ayling, S.; Tabbita, F.; Soria, M.; Wang, S.; IWGS Consortium; Akhunov, E.; et al. Separating Homeologs by Phasing in the Tetraploid Wheat Transcriptome. *Genome Biol* **2013**, *14*, R66, doi:10.1186/gb-2013-14-6-r66.
13. Boatwright, J.L.; McIntyre, L.M.; Morse, A.M.; Chen, S.; Yoo, M.-J.; Koh, J.; Soltis, P.S.; Soltis, D.E.; Barbazuk, W.B. A Robust Methodology for Assessing Differential Homeolog Contributions to the Transcriptomes of Allopolyploids. *Genetics* **2018**, *210*, 883–894, doi:10.1534/genetics.118.301564.
14. Chen, L.-Y.; Morales-Briones, D.F.; Passow, C.N.; Yang, Y. Performance of Gene Expression Analyses Using de Novo Assembled Transcripts in Polyploid Species. *Bioinformatics* **2019**, *35*, 4314–4320, doi:10.1093/bioinformatics/btz620.
15. Kuo, T.C.Y.; Hatakeyama, M.; Tameshige, T.; Shimizu, K.K.; Sese, J. Homeolog Expression Quantification Methods for Allopolyploids. *Briefings in Bioinformatics* **2020**, *21*, 395–407, doi:10.1093/bib/bby121.
16. Voshall, A.; Moriyama, E.N. Next-Generation Transcriptome Assembly and Analysis: Impact of Ploidy. *Methods* **2020**, *176*, 14–24, doi:10.1016/j.ymeth.2019.06.001.
17. Hu, G.; Grover, C.E.; Arick, M.A., II; Liu, M.; Peterson, D.G.; Wendel, J.F. Homoeologous Gene Expression and Co-Expression Network Analyses and Evolutionary Inference in Allopolyploids. *Briefings in Bioinformatics* **2021**, *22*, 1819–1835, doi:10.1093/bib/bbaa035.
18. Sun, J.; Okada, M.; Tameshige, T.; Shimizu-Inatsugi, R.; Akiyama, R.; Nagano, A.J.; Sese, J.; Shimizu, K.K. A Low-Coverage 3' RNA-Seq to Detect Homeolog Expression in Polyploid Wheat. *NAR Genomics and Bioinformatics* **2023**, *5*, lqad067, doi:10.1093/nargab/lqad067.
19. Rapp, R.A.; Udall, J.A.; Wendel, J.F. Genomic Expression Dominance in Allopolyploids. *BMC Biology* **2009**, *7*, 18, doi:10.1186/1741-7007-7-18.
20. Flagel, L.E.; Wendel, J.F. Evolutionary Rate Variation, Genomic Dominance and Duplicate Gene Expression Evolution during Allotetraploid Cotton Speciation. *New Phytologist* **2010**, *186*, 184–193, doi:10.1111/j.1469-8137.2009.03107.x.
21. Bardil, A.; de Almeida, J.D.; Combes, M.C.; Lashermes, P.; Bertrand, B. Genomic Expression Dominance in the Natural Allopolyploid *Coffea Arabica* Is Massively Affected by Growth Temperature. *New Phytologist* **2011**, *192*, 760–774, doi:10.1111/j.1469-8137.2011.03833.x.

22. Chelaifa, H.; Monnier, A.; Ainouche, M. Transcriptomic Changes Following Recent Natural Hybridization and Allopolyploidy in the Salt Marsh Species *Spartina* × *Townsendii* and *Spartina Anglica* (Poaceae). *New Phytologist* **2010**, *186*, 161–174, doi:10.1111/j.1469-8137.2010.03179.x.
23. Ferreira de Carvalho, J.; Boutte, J.; Bourdaud, P.; Chelaifa, H.; Ainouche, K.; Salmon, A.; Ainouche, M. Gene Expression Variation in Natural Populations of Hexaploid and Allododecaploid *Spartina* Species (Poaceae). *Plant Syst Evol* **2017**, *303*, 1061–1079, doi:10.1007/s00606-017-1446-3.
24. Giraud, D.; Lima, O.; Rousseau-Gueutin, M.; Salmon, A.; Ainouche, M. Gene and Transposable Element Expression Evolution Following Recent and Past Polyploidy Events in *Spartina* (Poaceae). *Frontiers in Genetics* **2021**, *12*.
25. Schnable, J.C.; Springer, N.M.; Freeling, M. Differentiation of the Maize Subgenomes by Genome Dominance and Both Ancient and Ongoing Gene Loss. *Proceedings of the National Academy of Sciences* **2011**, *108*, 4069–4074, doi:10.1073/pnas.1101368108.
26. Schnable, J.; Wang, X.; Pires, J.; Freeling, M. Escape from Preferential Retention Following Repeated Whole Genome Duplications in Plants. *Frontiers in Plant Science* **2012**, *3*.
27. Edger, P.P.; Poorten, T.J.; VanBuren, R.; Hardigan, M.A.; Colle, M.; McKain, M.R.; Smith, R.D.; Teresi, S.J.; Nelson, A.D.L.; Wai, C.M.; et al. Origin and Evolution of the Octoploid Strawberry Genome. *Nat Genet* **2019**, *51*, 541–547, doi:10.1038/s41588-019-0356-4.
28. Bird, K.A.; Niederhuth, C.E.; Ou, S.; Gehan, M.; Pires, J.C.; Xiong, Z.; VanBuren, R.; Edger, P.P. Replaying the Evolutionary Tape to Investigate Subgenome Dominance in Allopolyploid Brassica Napus. *New Phytologist* **2021**, *230*, 354–371, doi:10.1111/nph.17137.
29. Glombik, M.; Copetti, D.; Bartos, J.; Stoces, S.; Zwierzykowski, Z.; Ruttink, T.; Wendel, J.F.; Duchoslav, M.; Dolezel, J.; Studer, B.; et al. Reciprocal Allopolyploid Grasses (*Festuca* × *Lolium*) Display Stable Patterns of Genome Dominance. *The Plant Journal* **2021**, *107*, 1166–1182, doi:10.1111/tpj.15375.
30. Kopecký, D.; Scholten, O.; Majka, J.; Burger-Meijer, K.; Duchoslav, M.; Bartoš, J. Genome Dominance in Allium Hybrids (*A. Ceba* × *A. Roylei*). *Frontiers in Plant Science* **2022**, *13*.
31. Grover, C.E.; Gallagher, J.P.; Szadkowski, E.P.; Yoo, M.J.; Flagel, L.E.; Wendel, J.F. Homoeolog Expression Bias and Expression Level Dominance in Allopolyploids. *New Phytologist* **2012**, *196*, 966–971, doi:10.1111/j.1469-8137.2012.04365.x.
32. Stebbins, G.L. Chromosomal Evolution in Higher Plants. *Chromosomal evolution in higher plants*. **1971**.
33. Buggs, R.J.A.; Wendel, J.F.; Doyle, J.J.; Soltis, D.E.; Soltis, P.S.; Coate, J.E. The Legacy of Diploid Progenitors in Allopolyploid Gene Expression Patterns. *Philos Trans R Soc Lond B Biol Sci* **2014**, *369*, 20130354, doi:10.1098/rstb.2013.0354.
34. Yoo, M.-J.; Szadkowski, E.; Wendel, J.F. Homoeolog Expression Bias and Expression Level Dominance in Allopolyploid Cotton. *Heredity* **2013**, *110*, 171–180, doi:10.1038/hdy.2012.94.
35. Wu, J.; Lin, L.; Xu, M.; Chen, P.; Liu, D.; Sun, Q.; Ran, L.; Wang, Y. Homoeolog Expression Bias and Expression Level Dominance in Resynthesized Allopolyploid Brassica Napus. *BMC Genomics* **2018**, *19*, 586, doi:10.1186/s12864-018-4966-5.
36. Li, M.; Wang, R.; Wu, X.; Wang, J. Homoeolog Expression Bias and Expression Level Dominance (ELD) in Four Tissues of Natural Allotetraploid Brassica Napus. *BMC Genomics* **2020**, *21*, 330, doi:10.1186/s12864-020-6747-1.
37. Peng, Z.; Cheng, H.; Sun, G.; Pan, Z.; Wang, X.; Geng, X.; He, S.; Du, X. Expression Patterns and Functional Divergence of Homologous Genes Accompanied by Polyploidization in Cotton (*Gossypium Hirsutum* L.). *Sci China Life Sci* **2020**, *63*, 1565–1579, doi:10.1007/s11427-019-1618-7.
38. Pootakham, W.; Sonthirod, C.; Naktang, C.; Yundaeng, C.; Yoocha, T.; Kongkachana, W.; Sangsrakru, D.; Somta, P.; Tangphatsornruang, S. Genome Assemblies of *Vigna Reflexo-Pilosa* (Créole Bean) and Its Progenitors, *Vigna Hirtella* and *Vigna Trinervia*, Revealed Homoeolog Expression Bias and Expression-Level Dominance in the Allotetraploid. *Gigascience* **2023**, *12*, giad050, doi:10.1093/gigascience/giad050.
39. Yoo, M.-J.; Liu, X.; Pires, J.C.; Soltis, P.S.; Soltis, D.E. Nonadditive Gene Expression in Polyploids. *Annual Review of Genetics* **2014**, *48*, 485–517, doi:10.1146/annurev-genet-120213-092159.
40. Bird, K.A.; VanBuren, R.; Puzey, J.R.; Edger, P.P. The Causes and Consequences of Subgenome Dominance in Hybrids and Recent Polyploids. *New Phytologist* **2018**, *220*, 87–93, doi:10.1111/nph.15256.
41. Qiu, T.; Liu, Z.; Liu, B. The Effects of Hybridization and Genome Doubling in Plant Evolution via Allopolyploidy. *Mol Biol Rep* **2020**, *47*, 5549–5558, doi:10.1007/s11033-020-05597-y.
42. Shimizu, K.K. Robustness and the Generalist Niche of Polyploid Species: Genome Shock or Gradual Evolution? *Current Opinion in Plant Biology* **2022**, *69*, 102292, doi:10.1016/j.pbi.2022.102292.
43. Nomaguchi, T.; Maeda, Y.; Yoshino, T.; Asahi, T.; Tirichine, L.; Bowler, C.; Tanaka, T. Homoeolog Expression Bias in Allopolyploid Oleaginous Marine Diatom *Fistulifera Solaris*. *BMC Genomics* **2018**, *19*, 330, doi:10.1186/s12864-018-4691-0.
44. Buggs, R.J.A. Unravelling Gene Expression of Complex Crop Genomes. *Heredity* **2013**, *110*, 97–98, doi:10.1038/hdy.2012.61.

45. Gianinetti, A. A Criticism of the Value of Midparent in Polyploidization. *Journal of Experimental Botany* **2013**, 64, 4119–4129, doi:10.1093/jxb/ert263.
46. Pumphrey, M.; Bai, J.; Laudencia-Chingcuanco, D.; Anderson, O.; Gill, B.S. Nonadditive Expression of Homoeologous Genes Is Established Upon Polyploidization in Hexaploid Wheat. *Genetics* **2009**, 181, 1147–1157, doi:10.1534/genetics.108.096941.
47. Chagué, V.; Just, J.; Mestiri, I.; Balzergue, S.; Tanguy, A.-M.; Huneau, C.; Huteau, V.; Belcram, H.; Coriton, O.; Jahier, J.; et al. Genome-Wide Gene Expression Changes in Genetically Stable Synthetic and Natural Wheat Allohexaploids. *New Phytologist* **2010**, 187, 1181–1194, doi:10.1111/j.1469-8137.2010.03339.x.
48. Chelaifa, H.; Chagué, V.; Chalabi, S.; Mestiri, I.; Arnaud, D.; Deffains, D.; Lu, Y.; Belcram, H.; Huteau, V.; Chiquet, J.; et al. Prevalence of Gene Expression Additivity in Genetically Stable Wheat Allohexaploids. *New Phytologist* **2013**, 197, 730–736, doi:10.1111/nph.12108.
49. Song, K.; Lu, P.; Tang, K.; Osborn, T.C. Rapid Genome Change in Synthetic Polyploids of Brassica and Its Implications for Polyploid Evolution. *Proceedings of the National Academy of Sciences* **1995**, 92, 7719–7723, doi:10.1073/pnas.92.17.7719.
50. Salmon, A.; Ainouche, M.L.; Wendel, J.F. Genetic and Epigenetic Consequences of Recent Hybridization and Polyploidy in *Spartina* (Poaceae). *Molecular Ecology* **2005**, 14, 1163–1175, doi:10.1111/j.1365-294X.2005.02488.x.
51. Hu, G.; Koh, J.; Yoo, M.-J.; Chen, S.; Wendel, J.F. Gene-Expression Novelty in Allopolyploid Cotton: A Proteomic Perspective. *Genetics* **2015**, 200, 91–104, doi:10.1534/genetics.115.174367.
52. Adams, K.L.; Cronn, R.; Percifield, R.; Wendel, J.F. Genes Duplicated by Polyploidy Show Unequal Contributions to the Transcriptome and Organ-Specific Reciprocal Silencing. *Proceedings of the National Academy of Sciences* **2003**, 100, 4649–4654, doi:10.1073/pnas.0630618100.
53. Leebens-Mack, J.H.; Barker, M.S.; Carpenter, E.J.; Deyholos, M.K.; Gitzendanner, M.A.; Graham, S.W.; Grosse, I.; Li, Z.; Melkonian, M.; Mirarab, S.; et al. One Thousand Plant Transcriptomes and the Phylogenomics of Green Plants. *Nature* **2019**, 574, 679–685, doi:10.1038/s41586-019-1693-2.
54. Halabi, K.; Shafir, A.; Mayrose, I. PloidyDB: The Plant Ploidy Database. *New Phytologist* **2023**, 240, 918–927, doi:10.1111/nph.19057.
55. Qin, Y.; Sun, M.; Li, W.; Xu, M.; Shao, L.; Liu, Y.; Zhao, G.; Liu, Z.; Xu, Z.; You, J.; et al. Single-Cell RNA-Seq Reveals Fate Determination Control of an Individual Fibre Cell Initiation in Cotton (*Gossypium hirsutum*). *Plant Biotechnology Journal* **2022**, 20, 2372–2388, doi:10.1111/pbi.13918.
56. Bennett, R.J.; Uhl, M.A.; Miller, M.G.; Johnson, A.D. Identification and Characterization of a *Candida albicans* Mating Pheromone. *Mol Cell Biol* **2003**, 23, 8189–8201, doi:10.1128/MCB.23.22.8189-8201.2003.
57. Langham, R.J.; Walsh, J.; Dunn, M.; Ko, C.; Goff, S.A.; Freeling, M. Genomic Duplication, Fractionation and the Origin of Regulatory Novelty. *Genetics* **2004**, 166, 935–945, doi:10.1093/genetics/166.2.935.
58. Tate, J.A.; Joshi, P.; Soltis, K.A.; Soltis, P.S.; Soltis, D.E. On the Road to Diploidization? Homoeolog Loss in Independently Formed Populations of the Allopolyploid *Tragopogon miscellus* (Asteraceae). *BMC Plant Biol* **2009**, 9, 80, doi:10.1186/1471-2229-9-80.
59. Cheng, F.; Wu, J.; Cai, X.; Liang, J.; Freeling, M.; Wang, X. Gene Retention, Fractionation and Subgenome Differences in Polyploid Plants. *Nature Plants* **2018**, 4, 258–268, doi:10.1038/s41477-018-0136-7.
60. Thomas, B.C.; Pedersen, B.; Freeling, M. Following Tetraploidy in an Arabidopsis Ancestor, Genes Were Removed Preferentially from One Homeolog Leaving Clusters Enriched in Dose-Sensitive Genes. *Genome Res.* **2006**, 16, 934–946, doi:10.1101/gr.4708406.
61. Gaeta, R.T.; Chris Pires, J. Homoeologous Recombination in Allopolyploids: The Polyploid Ratchet. *New Phytologist* **2010**, 186, 18–28, doi:10.1111/j.1469-8137.2009.03089.x.
62. Mason, A.S.; Wendel, J.F. Homoeologous Exchanges, Segmental Allopolyploidy, and Polyploid Genome Evolution. *Frontiers in Genetics* **2020**, 11.
63. Deb, S.K.; Edger, P.P.; Pires, J.C.; McKain, M.R. Patterns, Mechanisms, and Consequences of Homoeologous Exchange in Allopolyploid Angiosperms: A Genomic and Epigenomic Perspective. *New Phytologist* **2023**, 238, 2284–2304, doi:10.1111/nph.18927.
64. Peralta, M.; Combes, M.-C.; Cenci, A.; Lashermes, P.; Dereeper, A. SNIploid: A Utility to Exploit High-Throughput SNP Data Derived from RNA-Seq in Allopolyploid Species. *International Journal of Plant Genomics* **2013**, 2013, 1–6, doi:10.1155/2013/890123.
65. Freyman, W.A.; Johnson, M.G.; Rothfels, C.J. Homologizer: Phylogenetic Phasing of Gene Copies into Polyploid Subgenomes. *Methods in Ecology and Evolution* **2023**, 14, 1230–1244, doi:10.1111/2041-210X.14072.
66. Schiavinato, M.; Bodrug-Schepers, A.; Dohm, J.C.; Himmelbauer, H. Subgenome Evolution in Allotetraploid Plants. *The Plant Journal* **2021**, 106, 672–688, doi:10.1111/tpj.15190.
67. Page, J.T.; Gingle, A.R.; Udall, J.A. PolyCat: A Resource for Genome Categorization of Sequencing Reads From Allopolyploid Organisms. *G3 Genes | Genomes | Genetics* **2013**, 3, 517–525, doi:10.1534/g3.112.005298.
68. Khan, A.; Belfield, E.J.; Harberd, N.P.; Mithani, A. HANDS2: Accurate Assignment of Homoeallelic Base-Identity in Allopolyploids despite Missing Data. *Sci Rep* **2016**, 6, 29234, doi:10.1038/srep29234.

69. Page, J.T.; Udall, J.A. Methods for Mapping and Categorization of DNA Sequence Reads from Allopolyploid Organisms. *BMC Genetics* **2015**, *16*, S4, doi:10.1186/1471-2156-16-S2-S4.
70. Sun, P.; Jiao, B.; Yang, Y.; Shan, L.; Li, T.; Li, X.; Xi, Z.; Wang, X.; Liu, J. WGD: A User-Friendly Toolkit for Evolutionary Analyses of Whole-Genome Duplications and Ancestral Karyotypes. *Molecular Plant* **2022**, *15*, 1841–1851, doi:10.1016/j.molp.2022.10.018.
71. Schiavinato, M.; Marcet-Houben, M.; Dohm, J.C.; Gabaldón, T.; Himmelbauer, H. Parental Origin of the Allotetraploid Tobacco *Nicotiana Benthiana*. *The Plant Journal* **2020**, *102*, 541–554, doi:10.1111/tpj.14648.
72. Gordon, S.P.; Levy, J.J.; Vogel, J.P. PolyCRACKER, a Robust Method for the Unsupervised Partitioning of Polyploid Subgenomes by Signatures of Repetitive DNA Evolution. *BMC Genomics* **2019**, *20*, 580, doi:10.1186/s12864-019-5828-5.
73. Jia, K.-H.; Wang, Z.-X.; Wang, L.; Li, G.-Y.; Zhang, W.; Wang, X.-L.; Xu, F.-J.; Jiao, S.-Q.; Zhou, S.-S.; Liu, H.; et al. SubPhaser: A Robust Allopolyploid Subgenome Phasing Method Based on Subgenome-Specific k-Mers. *New Phytologist* **2022**, *235*, 801–809, doi:10.1111/nph.18173.
74. Zhang, R.-G.; Shang, H.-Y.; Jia, K.-H.; Ma, Y.-P. Subgenome Phasing for Complex Allopolyploidy: Case-Based Benchmarking and Recommendations. *Briefings in Bioinformatics* **2024**, *25*, bbad513, doi:10.1093/bib/bbad513.
75. Session, A.M. Allopolyploid Subgenome Identification and Implications for Evolutionary Analysis. *Trends in Genetics* **2024**, *0*, doi:10.1016/j.tig.2024.03.008.
76. Fitch, W.M. Distinguishing Homologous from Analogous Proteins. *Systematic Biology* **1970**, *19*, 99–113, doi:10.2307/2412448.
77. Altenhoff, A.M.; Studer, R.A.; Robinson-Rechavi, M.; Dessimoz, C. Resolving the Ortholog Conjecture: Orthologs Tend to Be Weakly, but Significantly, More Similar in Function than Paralogs. *PLOS Computational Biology* **2012**, *8*, e1002514, doi:10.1371/journal.pcbi.1002514.
78. Nevers, Y.; Defosset, A.; Lecompte, O. Orthology: Promises and Challenges. In *Evolutionary Biology—A Transdisciplinary Approach*; Pontarotti, P., Ed.; Springer International Publishing: Cham, 2020; pp. 203–228 ISBN 978-3-030-57246-4.
79. Sonnhammer, E.L.L.; Koonin, E.V. Orthology, Paralogy and Proposed Classification for Paralog Subtypes. *Trends in Genetics* **2002**, *18*, 619–620, doi:10.1016/S0168-9525(02)02793-2.
80. Ouzounis, C. Orthology: Another Terminology Muddle. *Trends in Genetics* **1999**, *15*, 445, doi:10.1016/S0168-9525(99)01845-4.
81. Kristensen, D.M.; Wolf, Y.I.; Mushegian, A.R.; Koonin, E.V. Computational Methods for Gene Orthology Inference. *Brief Bioinform* **2011**, *12*, 379–391, doi:10.1093/bib/bbr030.
82. Glover, N.M.; Altenhoff, A.; Dessimoz, C. Assigning Confidence Scores to Homoeologs Using Fuzzy Logic. *PeerJ* **2019**, *6*, e6231, doi:10.7717/peerj.6231.
83. Koonin, E.V. Orthologs, Paralogs, and Evolutionary Genomics. *Annual Review of Genetics* **2005**, *39*, 309–338, doi:10.1146/annurev.genet.39.073003.114725.
84. Altenhoff, A.M.; Glover, N.M.; Dessimoz, C. Inferring Orthology and Paralogy. *Methods Mol Biol* **2019**, *1910*, 149–175, doi:10.1007/978-1-4939-9074-0_5.
85. Overbeek, R.; Fonstein, M.; D'Souza, M.; Pusch, G.D.; Maltsev, N. The Use of Gene Clusters to Infer Functional Coupling. *Proceedings of the National Academy of Sciences* **1999**, *96*, 2896–2901, doi:10.1073/pnas.96.6.2896.
86. Glover, N.; Sheppard, S.; Dessimoz, C. Homoeolog Inference Methods Requiring Bidirectional Best Hits or Synteny Miss Many Pairs. *Genome Biology and Evolution* **2021**, *13*, evab077, doi:10.1093/gbe/evab077.
87. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic Local Alignment Search Tool. *Journal of Molecular Biology* **1990**, *215*, 403–410, doi:10.1016/S0022-2836(05)80360-2.
88. Dalquen, D.A.; Dessimoz, C. Bidirectional Best Hits Miss Many Orthologs in Duplication-Rich Clades Such as Plants and Animals. *Genome Biology and Evolution* **2013**, *5*, 1800–1806, doi:10.1093/gbe/evt132.
89. McCouch, S.R. Genomics and Synteny. *Plant Physiology* **2001**, *125*, 152–155, doi:10.1104/pp.125.1.152.
90. Dewey, C.N. Positional Orthology: Putting Genomic Evolutionary Relationships into Context. *Briefings in Bioinformatics* **2011**, *12*, 401–412, doi:10.1093/bib/bbr040.
91. Train, C.-M.; Glover, N.M.; Gonnet, G.H.; Altenhoff, A.M.; Dessimoz, C. Orthologous Matrix (OMA) Algorithm 2.0: More Robust to Asymmetric Evolutionary Rates and More Scalable Hierarchical Orthologous Group Inference. *Bioinformatics* **2017**, *33*, i75–i82, doi:10.1093/bioinformatics/btx229.
92. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics. *Genome Biology* **2019**, *20*, 238, doi:10.1186/s13059-019-1832-y.
93. Zhang, Z.; Lv, R.; Wang, B.; Xun, H.; Liu, B.; Xu, C. Effects of Allopolyploidization and Homoeologous Chromosomal Segment Exchange on Homoeolog Expression in a Synthetic Allotetraploid Wheat under Variable Environmental Conditions. *Plants (Basel)* **2023**, *12*, 3111, doi:10.3390/plants12173111.
94. Vilella, A.J.; Severin, J.; Ureta-Vidal, A.; Heng, L.; Durbin, R.; Birney, E. EnsemblCompara GeneTrees: Complete, Duplication-Aware Phylogenetic Trees in Vertebrates. *Genome Res.* **2009**, *19*, 327–335, doi:10.1101/gr.073585.107.

95. Li, H.; Coghlan, A.; Ruan, J.; Coin, L.J.; Hériché, J.-K.; Osmotherly, L.; Li, R.; Liu, T.; Zhang, Z.; Bolund, L.; et al. TreeFam: A Curated Database of Phylogenetic Trees of Animal Gene Families. *Nucleic Acids Res* **2006**, *34*, D572–580, doi:10.1093/nar/gkj118.
96. Xu, M.-R.-X.; Liao, Z.-Y.; Brock, J.R.; Du, K.; Li, G.-Y.; Chen, Z.-Q.; Wang, Y.-H.; Gao, Z.-N.; Agarwal, G.; Wei, K.H.-C.; et al. Maternal Dominance Contributes to Subgenome Differentiation in Allopolyploid Fishes. *Nat Commun* **2023**, *14*.
97. Pereira, C.; Denise, A.; Lespinet, O. A Meta-Approach for Improving the Prediction and the Functional Annotation of Ortholog Groups. *BMC Genomics* **2014**, *15*, S16, doi:10.1186/1471-2164-15-S6-S16.
98. Chorostecki, U.; Molina, M.; Prysycz, L.P.; Gabaldón, T. MetaPhOrs 2.0: Integrative, Phylogeny-Based Inference of Orthology and Paralogy across the Tree of Life. *Nucleic Acids Research* **2020**, *48*, W553–W557, doi:10.1093/nar/gkaa282.
99. Altenhoff, A.M.; Dessimoz, C. Phylogenetic and Functional Assessment of Orthologs Inference Projects and Methods. *PLOS Computational Biology* **2009**, *5*, e1000262, doi:10.1371/journal.pcbi.1000262.
100. Goodwin, S.; McPherson, J.D.; McCombie, W.R. Coming of Age: Ten Years of next-Generation Sequencing Technologies. *Nat Rev Genet* **2016**, *17*, 333–351, doi:10.1038/nrg.2016.49.
101. Heather, J.M.; Chain, B. The Sequence of Sequencers: The History of Sequencing DNA. *Genomics* **2016**, *107*, 1–8, doi:10.1016/j.ygeno.2015.11.003.
102. Gault, C.M.; Kremling, K.A.; Buckler, E.S. Tripsacum De Novo Transcriptome Assemblies Reveal Parallel Gene Evolution with Maize after Ancient Polyploidy. *The Plant Genome* **2018**, *11*, 180012, doi:10.3835/plantgenome2018.02.0012.
103. Kamitani, M.; Kashima, M.; Tezuka, A.; Nagano, A.J. Lasy-Seq: A High-Throughput Library Preparation Method for RNA-Seq and Its Application in the Analysis of Plant Responses to Fluctuating Temperatures. *Sci Rep* **2019**, *9*, 7091, doi:10.1038/s41598-019-43600-0.
104. Wang, M.; Wang, P.; Liang, F.; Ye, Z.; Li, J.; Shen, C.; Pei, L.; Wang, F.; Hu, J.; Tu, L.; et al. A Global Survey of Alternative Splicing in Allopolyploid Cotton: Landscape, Complexity and Regulation. *New Phytol* **2018**, *217*, 163–178, doi:10.1111/nph.14762.
105. Gonzalez-Garay, M.L. Introduction to Isoform Sequencing Using Pacific Biosciences Technology (Iso-Seq). In *Transcriptomics and Gene Regulation*; Wu, J., Ed.; Translational Bioinformatics; Springer Netherlands: Dordrecht, 2016; pp. 141–160 ISBN 978-94-017-7450-5.
106. Pyatnitskiy, M.A.; Arzumanyan, V.A.; Radko, S.P.; Ptitsyn, K.G.; Vakhrushev, I.V.; Poverennaya, E.V.; Ponomarenko, E.A. Oxford Nanopore MinION Direct RNA-Seq for Systems Biology. *Biology* **2021**, *10*, 1131, doi:10.3390/biology10111131.
107. Shi, H.; Zhou, Y.; Jia, E.; Pan, M.; Bai, Y.; Ge, Q. Bias in RNA-Seq Library Preparation: Current Challenges and Solutions. *Biomed Res Int* **2021**, *2021*, 6647597, doi:10.1155/2021/6647597.
108. Wang, D.; Lu, X.; Chen, X.; Wang, S.; Wang, J.; Guo, L.; Yin, Z.; Chen, Q.; Ye, W. Temporal Salt Stress-Induced Transcriptome Alterations and Regulatory Mechanisms Revealed by PacBio Long-Reads RNA Sequencing in *Gossypium Hirsutum*. *BMC Genomics* **2020**, *21*, 838, doi:10.1186/s12864-020-07260-z.
109. Yao, S.; Liang, F.; Gill, R.A.; Huang, J.; Cheng, X.; Liu, Y.; Tong, C.; Liu, S. A Global Survey of the Transcriptome of Allopolyploid Brassica Napus Based on Single-Molecule Long-Read Isoform Sequencing and Illumina-Based RNA Sequencing Data. *The Plant Journal* **2020**, *103*, 843–857, doi:10.1111/tpj.14754.
110. Navajas-Pérez, R.; Paterson, A.H. Patterns of Tandem Repetition in Plant Whole Genome Assemblies. *Mol Genet Genomics* **2009**, *281*, 579–590, doi:10.1007/s00438-009-0433-y.
111. Kyriakidou, M.; Tai, H.H.; Anglin, N.L.; Ellis, D.; Strömvik, M.V. Current Strategies of Polyploid Plant Genome Sequence Assembly. *Front. Plant Sci.* **2018**, *9*, doi:10.3389/fpls.2018.01660.
112. Kong, W.; Wang, Y.; Zhang, S.; Yu, J.; Zhang, X. Recent Advances in Assembly of Complex Plant Genomes. *Genomics Proteomics Bioinformatics* **2023**, *21*, 427–439, doi:10.1016/j.gpb.2023.04.004.
113. Gutierrez-Gonzalez, J.J.; Garvin, D.F. De Novo Transcriptome Assembly in Polyploid Species. *Methods Mol Biol* **2017**, *1536*, 209–221, doi:10.1007/978-1-4939-6682-0_15.
114. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* **2013**, *29*, 15–21, doi:10.1093/bioinformatics/bts635.
115. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A Fast Spliced Aligner with Low Memory Requirements. *Nat Methods* **2015**, *12*, 357–360, doi:10.1038/nmeth.3317.
116. Bray, N.L.; Pimentel, H.; Melsted, P.; Pachter, L. Near-Optimal Probabilistic RNA-Seq Quantification. *Nat Biotechnol* **2016**, *34*, 525–527, doi:10.1038/nbt.3519.
117. Akama, S.; Shimizu-Inatsugi, R.; Shimizu, K.K.; Sese, J. Genome-Wide Quantification of Homeolog Expression Ratio Revealed Nonstochastic Gene Regulation in Synthetic Allopolyploid Arabidopsis. *Nucleic Acids Research* **2014**, *42*, e46, doi:10.1093/nar/gkt1376.
118. Wu, T.D.; Reeder, J.; Lawrence, M.; Becker, G.; Brauer, M.J. GMAP and GSNAP for Genomic Sequence Alignment: Enhancements to Speed, Accuracy, and Functionality. In *Statistical Genomics: Methods and Protocols*; Mathé, E., Davis, S., Eds.; Springer: New York, NY, 2016; pp. 283–334 ISBN 978-1-4939-3578-9.

119. Kuo, T.; Frith, M.C.; Sese, J.; Horton, P. EAGLE: Explicit Alternative Genome Likelihood Evaluator. *BMC Medical Genomics* **2018**, *11*, 28, doi:10.1186/s12920-018-0342-1.
120. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079, doi:10.1093/bioinformatics/btp352.
121. Zyprych-Walczak, J.; Szabelska, A.; Handschuh, L.; Górczak, K.; Klamecka, K.; Figlerowicz, M.; Siatkowski, I. The Impact of Normalization Methods on RNA-Seq Data Analysis. *BioMed Research International* **2015**, *2015*, 621690, doi:10.1155/2015/621690.
122. Evans, C.; Hardin, J.; Stoebe, D.M. Selecting Between-Sample RNA-Seq Normalization Methods from the Perspective of Their Assumptions. *Brief Bioinform* **2017**, *19*, 776–792, doi:10.1093/bib/bbx008.
123. Trapnell, C.; Williams, B.A.; Pertea, G.; Mortazavi, A.; Kwan, G.; van Baren, M.J.; Salzberg, S.L.; Wold, B.J.; Pachter, L. Transcript Assembly and Quantification by RNA-Seq Reveals Unannotated Transcripts and Isoform Switching during Cell Differentiation. *Nat Biotechnol* **2010**, *28*, 511–515, doi:10.1038/nbt.1621.
124. Li, B.; Dewey, C.N. RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome. *BMC Bioinformatics* **2011**, *12*, 323, doi:10.1186/1471-2105-12-323.
125. Zhao, S.; Ye, Z.; Stanton, R. Misuse of RPKM or TPM Normalization When Comparing across Samples and Sequencing Protocols. *RNA* **2020**, *26*, 903–909, doi:10.1261/rna.074922.120.
126. Robinson, M.D.; Oshlack, A. A Scaling Normalization Method for Differential Expression Analysis of RNA-Seq Data. *Genome Biol* **2010**, *11*, R25, doi:10.1186/gb-2010-11-3-r25.
127. Anders, S.; Huber, W. Differential Expression Analysis for Sequence Count Data. *Nat Protoc* **2010**, *1*–1, doi:10.1038/npre.2010.4282.1.
128. Aanes, H.; Winata, C.; Moen, L.F.; Østrup, O.; Mathavan, S.; Collas, P.; Rognes, T.; Aleström, P. Normalization of RNA-Sequencing Data from Samples with Varying mRNA Levels. *PLOS ONE* **2014**, *9*, e89158, doi:10.1371/journal.pone.0089158.
129. Coate, J.E.; Doyle, J.J. Variation in Transcriptome Size: Are We Getting the Message? *Chromosoma* **2015**, *124*, 27–43, doi:10.1007/s00412-014-0496-3.
130. Pirrello, J.; Deluche, C.; Frangne, N.; Gévaudant, F.; Maza, E.; Djari, A.; Bourge, M.; Renaudin, J.-P.; Brown, S.; Bowler, C.; et al. Transcriptome Profiling of Sorted Endoreduplicated Nuclei from Tomato Fruits: How the Global Shift in Expression Ascribed to DNA Ploidy Influences RNA-Seq Data Normalization and Interpretation. *The Plant Journal* **2018**, *93*, 387–398, doi:10.1111/tpj.13783.
131. Coate, J.E.; Doyle, J.J. Quantifying Whole Transcriptome Size, a Prerequisite for Understanding Transcriptome Evolution Across Species: An Example from a Plant Allopolyploid. *Genome Biology and Evolution* **2010**, *2*, 534–546, doi:10.1093/gbe/evq038.
132. Visger, C.J.; Wong, G.K.-S.; Zhang, Y.; Soltis, P.S.; Soltis, D.E. Divergent Gene Expression Levels between Diploid and Autotetraploid *Tolmiea* Relative to the Total Transcriptome, the Cell, and Biomass. *American Journal of Botany* **2019**, *106*, 280–291, doi:10.1002/ajb2.1239.
133. Fomina-Yadlin, D.; Du, Z.; McGrew, J.T. Gene Expression Measurements Normalized to Cell Number Reveal Large Scale Differences Due to Cell Size Changes, Transcriptional Amplification and Transcriptional Repression in CHO Cells. *Journal of Biotechnology* **2014**, *189*, 58–69, doi:10.1016/j.jbiotec.2014.08.037.
134. Orr-Weaver, T.L. When Bigger Is Better: The Role of Polyploidy in Organogenesis. *Trends Genet* **2015**, *31*, 307–315, doi:10.1016/j.tig.2015.03.011.
135. Coate, J.E. Beyond Transcript Concentrations: Quantifying Polyploid Expression Responses per Biomass, per Genome, and per Cell with RNA-Seq. *Methods Mol Biol* **2023**, *2545*, 227–250, doi:10.1007/978-1-0716-2561-3_12.
136. Chen, K.; Hu, Z.; Xia, Z.; Zhao, D.; Li, W.; Tyler, J.K. The Overlooked Fact: Fundamental Need for Spike-In Control for Virtually All Genome-Wide Analyses. *Mol Cell Biol* **2016**, *36*, 662–667, doi:10.1128/MCB.00970-14.
137. Almeida-Silva, F.; Prost-Boxoen, L.; Van de Peer, Y. Hybridexpress: An R/Bioconductor Package for Comparative Transcriptomic Analyses of Hybrids and Their Progenitors. *New Phytologist* **2024**, *243*, 811–819, doi:10.1111/nph.19862.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.