



OPEN ACCESS

EDITED BY

Anamika Pandey,
Selçuk University, Türkiye

REVIEWED BY

Ha Chu,
Vietnam National University, Hanoi, Vietnam
Muhammad Sajid,
Zhejiang University, China

*CORRESPONDENCE

Shailesh Kumar
✉ shailesh@nipgr.ac.in

RECEIVED 31 July 2025

ACCEPTED 01 October 2025

PUBLISHED 31 October 2025

CITATION

Hamid F, Zahra S and Kumar S (2025)
Molecular and expression analyses indicate
the role of fusion transcripts in mediating
abiotic stress responses in chickpea.
Front. Plant Sci. 16:1677098.
doi: 10.3389/fpls.2025.1677098

COPYRIGHT

© 2025 Hamid, Zahra and Kumar. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License](#)
(CC BY). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Molecular and expression analyses indicate the role of fusion transcripts in mediating abiotic stress responses in chickpea

Fiza Hamid¹, Shafaque Zahra² and Shailesh Kumar ^{1*}

¹Bioinformatics Lab, BRIC-National Institute of Plant Genome Research, New Delhi, India,

²Department of Pathology, School of Medicine, University of Virginia, Charlottesville, VA, United States

Understanding the transcriptome diversity is essential for deciphering the transcriptional level regulation. High-throughput sequencing technologies have facilitated the detection of fusion transcripts (FTs), which are chimeric mRNA molecules derived from gene fusions due to chromosomal rearrangements or via the splicing machinery at the RNA level. In this study, we investigated the transcriptome complexity in *Cicer arietinum* resulting from fusion events using high-throughput RNA-Seq datasets from five tissues, i.e., stem, leaves, buds, flowers, and pods, and two abiotic stress conditions, i.e., drought and salinity. Of the 328 unique FTs identified, 69% exhibited the presence of canonical splice sites at their junction, indicating their generation via trans-splicing. Functional annotation and enrichment analyses of fusion partners suggested that these transcripts may expand functional diversity. A total of 10 FTs were validated via RT-PCR followed by Sanger sequencing, which are the first FTs described in the important legume chickpea. Expression analysis of fusion transcripts across various tissues and under abiotic stress conditions revealed evidence of context-dependent regulation. Furthermore, 120 fusion gene pairs were found to be conserved across 17 chickpea genotypes, highlighting their potential biological significance and stability within the species. Overall, these findings suggest that fusion transcripts may contribute to regulatory mechanisms underlying abiotic stress responses in chickpea.

KEYWORDS

abiotic stress, *Cicer arietinum*, fusion transcripts, trans-splicing, transcriptome complexity

1 Introduction

Mature mRNA molecules are conventionally formed through transcription and post-transcriptional modifications, where introns are excised and stability is enhanced. Traditionally, mRNA was thought to originate solely from the alternative splicing of a single gene. However, advancements in sequencing technologies have uncovered the existence of novel transcripts, such as fusion transcripts (FTs), which arise by the joining of mRNA molecules derived from different genes (Gingeras, 2009; Gupta et al., 2018). FTs can be generated at the DNA level through genomic rearrangements such as translocation, deletion, duplication, or inversion and give rise to a fused gene (Li et al., 2009a; Annala et al., 2013); or at the RNA level through trans-splicing (Sutton and Boothroyd, 1986) or read-through transcription (Varley et al., 2014), resulting in an increase in transcriptome complexity without a corresponding increase in the gene number. While fusion transcripts are found in both unicellular and multicellular organisms, they were previously considered rare in nature and occasionally dismissed as transcriptional artifacts. Up to now, extensive research has characterized the cellular functions of these transcripts in cancer (Latysheva and Babu, 2016; Dupain et al., 2019). Recent studies indicate that in addition to their role in oncogenesis, FTs have also been reported under normal physiological conditions (Babiceanu et al., 2016; Chwalenia et al., 2017), and may act as potential regulators of their parental mRNA (Mukherjee et al., 2021). However, their presence and functional significance in plants remain largely unexplored.

In past decades, efficient big data analysis has facilitated the discovery of fusion transcripts in many plants, such as *Trifolium pratense* (Chao et al., 2018), *Camellia sinensis* (Qiao et al., 2019), *Oryza sativa* (Zhang et al., 2010), *Arabidopsis thaliana* (Singh et al., 2019), *Brassica rapa* (Tan et al., 2019), and *Zea mays* (Zhou et al., 2022). These plant-specific fusion transcripts add transcriptome complexity and introduce novel functions, which may be related to stress responses (Zhou et al., 2022), metabolic pathways (Hagel and Facchini, 2017), or phenotypic traits (Chen et al., 2017). For instance, maize exhibits various fusion events in response to viral infection, involving proteins like nodulin, flavone synthase, and cation-transporting ATPase (Zhou et al., 2022). In tomatoes, the fusion transcript *PFP-LeT6* is involved in leaf patterning (Kim et al., 2001), while in rice, the *GN2* chimeric gene controls plant height, heading date, and grain number (Chen et al., 2017). In *Arabidopsis thaliana*, a protein derived from the fusion of glutamine synthase and nodulin domains regulates root morphogenesis and flagellin-triggered signaling (Doskočilová et al., 2011). Studies across species also suggest that fusion events like domain shuffling and DNA fusion enhance the catalytic efficiency of an enzyme (Farrow et al., 2015; Li et al., 2016; Hagel and Facchini, 2017). These reports suggest that fusion events result in novel sequences with novel functions either as non-coding RNA or fusion proteins (Winzer et al., 2015; Qin et al., 2017).

Chickpea (*Cicer arietinum* L.) ranks as the third most cultivated legume crop globally, following common bean and pea (Nathawat et al., 2024). Cultivated in over 50 countries, the Indian

subcontinent leads in production, contributing approximately 70% of the global output (Koul et al., 2022). In addition to its economic significance, chickpeas are highly esteemed for their exceptional nutritional content, particularly their rich protein and carbohydrate content. Currently, chickpeas are grown on 15 million hectares worldwide, producing 15.9 million tons (FAO, 2023). However, this yield remains significantly below the crop's potential under optimal conditions. This gap is largely attributed to various biotic and abiotic stresses that hinder productivity. Abiotic stresses, such as salinity and drought, are key factors contributing to yield losses. Research has shown that chickpeas are particularly sensitive to salinity compared to other crops, making it a critical factor limiting yields (Flowers et al., 2010; Turner et al., 2013). The growing availability of high-throughput transcriptomic data offers valuable insights for tackling these stress-related challenges.

Despite advancements in sequencing technologies, the exploration of transcriptome diversity in chickpea, particularly due to fusion events, remains limited. A meta-analysis in *Arabidopsis thaliana*, *Cicer arietinum*, and *Oryza sativa* showed that the majority of fusion transcripts (~74%) were uniquely detected in only one transcriptomic sample (Chitkara et al., 2024). This lack of recurrence raises concerns about the biological reliability and reproducibility of these events. One possible explanation could be the inherent variability introduced by such a large and heterogeneous dataset. Alternatively, many of these uniquely detected fusion transcripts might represent random artifacts rather than true biological events. Moreover, the validation rate was reported to be generally low in *Arabidopsis thaliana* and *Oryza sativa*, which could be attributed to high false-positive prediction rates or the inherently low expression level of fusion transcripts. Notably, no experimental validation was performed for the fusion transcripts identified in *Cicer arietinum*.

Therefore, there is a compelling need for more targeted investigations under well-defined experimental conditions that allow biological replication. In the current study, we identified, annotated, and experimentally validated the fusion transcripts by utilizing in-house generated RNA-Seq datasets derived from specific abiotic stress conditions and tissue types, enabling the identification of high-confidence fusion transcripts by applying stringent filtering criteria. This enhances the reproducibility, which in turn strengthens the confidence in these fusion events as biologically relevant and potentially regulatory molecules, rather than technical noise. Further, stress-responsive and conserved fusion transcripts of chickpea were confirmed using a combination of second- and third-generation RNA-Seq datasets, along with experimental validation. To explore the features of fusion transcripts, we investigated their coding potential and the potential mechanism of formation. We predicted that 58% of fusion transcripts are non-coding in nature and 69% might arise via trans-splicing. While these findings suggest that non-coding fusion transcripts could play regulatory roles, their functional implications remain speculative and require further experimental validation. Additionally, a benchmarking analysis was performed to evaluate the performance of different fusion detection tools,

enabling the identification of the most effective methods for accurate fusion detection in Chickpea. This work provides deeper biological interpretation and functional insights into fusion transcript dynamics in Chickpea.

2 Materials and methods

2.1 Plant material, growth conditions, and stress treatment

Chickpea (Genotype ICC4958) seeds were cultivated by using the established protocol (Garg et al., 2010). Seedlings were grown in plastic pots filled with a sterilized mix of agro-peat and vermiculite in a 1:1 ratio, maintained at 22 °C with a 14-hour light cycle in a controlled growth chamber. Samples were collected at various developmental stages, including stems and leaves from the seedling stage and buds, flowers, and pods during the flowering stage. To induce drought stress, 21-day-old seedlings were removed from the pots and placed on tissue paper for five hours. For salinity stress, seedlings were immersed in a beaker filled with a 150 mM NaCl solution at 22 °C. Whole seedlings were collected after 5 hours of treatment and from three independent biological replicates. These samples were immediately flash-frozen in liquid nitrogen and stored at -80 °C until further analysis.

2.2 Total RNA isolation, cDNA library preparation, and sequencing

Total RNA was extracted from 100 mg of tissue using the RNeasy Plant Mini Kit (Qiagen). The quantity and quality of RNA were assessed using a NanoDrop Spectrophotometer (NanoDrop Technologies) and an Agilent Bioanalyzer. Only samples with a 260/280 ratio between 1.9 and 2.1, a 260/230 ratio between 2.0 and 2.4, and RNA Integrity Number (RIN) values greater than 7 were selected for Illumina sequencing. A total of 26 paired-end RNA-Seq libraries were prepared and sequenced. The quality of the reads was assessed with FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), and adapter trimming and filtering of low-quality reads were performed using fastp (Chen et al., 2018). High-quality reads were then aligned to the reference genome (ASM33114v1) using HISAT2-2.2.1 (Kim et al., 2019).

2.3 Computational prediction of fusion transcripts and their coding potential

Fusion transcripts were predicted using three different fusion detection tools, viz., FusionMap (Ge et al., 2011), STAR-Fusion (Haas et al., 2019), and MapSplice (Wang et al., 2010). These tools have different fusion detection strategies and filtering criteria, which help in identifying a broad range of fusion events. These tools were selected on the basis of available benchmarking publications, where

STAR-Fusion was ranked as the best tool in terms of its high sensitivity, accuracy, and execution time (Haas et al., 2019); and MapSplice and FusionMap show good sensitivity for fusion detection (Kumar et al., 2016). To assess the coding potential of fusion transcripts, we used CNIT (Guo et al., 2019), CPAT (Wang et al., 2013), and PlncPRO (Singh et al., 2017). CNIT uses a support vector machine (SVM) model to predict coding potential based on intrinsic sequence properties. CPAT uses a linear regression model to distinguish between coding and non-coding transcripts. PlncPRO employs a random forest classifier that integrates protein homology, sequence-based features, and 3-mer frequency patterns to differentiate coding transcripts from long non-coding RNAs (lncRNAs). Fusion transcripts were predicted as protein-coding or lncRNA depending on the consensus prediction of at least two of the three tools.

2.4 Gene ontology and KEGG pathway analysis

For all identified fusion transcripts, gene ontology (GO) and KEGG pathway analyses were performed using the DAVID web server (Sherman et al., 2022) with the DAVID knowledgebase v2024q2 (released on July 5, 2024), applying a significance cutoff of $p < 0.05$. These analyses help to identify biological processes, cellular components, and pathways potentially impacted by fusion events, shedding light on their functional relevance.

2.5 Expression analysis of fusion transcripts

The expression level of genes involved in fusion formation was analyzed based on Transcripts Per Million (TPM) using StringTie (Pertea et al., 2016) with the default parameters. To study the impact of fusion formation on the expression of their parental genes, we categorized the samples into two groups for each fusion transcript: (i) fusion-present (F/P) samples in which the corresponding fusion transcript was detected, and (ii) fusion-absent (F/A) samples lacking the respective fusion. The expression levels of each parental gene were compared between each sample of the two groups. Genes exhibiting ≥ 2 -fold change in expression in F/P samples compared to F/A samples were considered upregulated or downregulated.

2.6 Validation using long-read RNA-Seq data

For further validation of fusion transcripts, PacBio long-read RNA-Seq data (PRJNA613159, Jain et al., 2022) for *Cicer arietinum* (ICC4958) were downloaded from the NCBI Sequence Read Archive (SRA). The 200 bp junction sequences of the identified fusion transcripts, comprising 100 bp upstream of the 5' parental gene breakpoint and 100 bp downstream of the 3' parental gene breakpoint, were searched against the long-read RNA-Seq data

using BLASTn. Hits with greater than 80% sequence identity and alignment length exceeding 150 bp were considered significant. This approach confirmed the presence of predicted fusion transcripts within the long-read data, enhancing confidence in the identified fusion events. Additionally, it offered insights into the putative length of the fusion transcripts.

2.7 Identification of conserved fusion events

To identify intra-specific conserved fusions, 103 raw transcriptome data files from 17 *Cicer arietinum* genotypes were downloaded from the NCBI-SRA (<https://www.ncbi.nlm.nih.gov/sra>). A list of RNA-Seq samples from different genotypes used in this study is listed in [Supplementary Table S8](#). Fusion events were then identified across all genotypes using three fusion detection tools. A binary analysis was conducted to determine the presence or absence of each fusion gene pair in the respective genotypes. For inter-specific conserved fusion detection, fusion genes previously reported in *Arabidopsis thaliana* (via AtFusionDB: <http://www.nipgr.ac.in/AtFusionDB>) were utilized to identify their homologous fusion gene pairs in *Cicer arietinum* using the OrthoFinder tool.

2.8 Experimental validation of fusion transcripts

To validate fusion transcripts, primers were designed for the fusion transcript sequences taken from the genome, 200 bp upstream and 200 bp downstream of the breakpoint, using the OligoCalc (Kibbe, 2007) and primer BLAST. All primers used for validation are listed in [Supplementary Table S9](#). The cDNA synthesis was carried out with 2 µg of RNA using the Verso cDNA synthesis kit (ThermoScientific™). Following RT-PCR and gel electrophoresis, DNA bands were extracted and purified using the GenElute™ Gel Extraction kit (Sigma-Aldrich) and sent for Sanger sequencing at the NIPGR DNA sequencing facility. To quantify the expression of fusion transcripts under drought and salt stress, and across different tissues, quantitative Real-Time PCR was performed. EF1-α was used as an endogenous control gene in this experiment. The real-time PCR reaction mix (10 µl) consisted of 5 µl of 2X SYBR Green Master Mix (Applied Biosystems™), 10 µM of each primer, and 100 ng of cDNA template. PCR amplification was performed using an Applied Biosystems™ qPCR system with thermal cycling conditions of initial denaturation at 95 °C for 2 minutes, followed by 40 cycles of 95 °C for 15 seconds and 60 °C for 1 minute. Each PCR reaction included three biological replicates and three technical replicates. To ensure primer specificity, primer amplification efficiency and melt-curve analyses were performed ([Supplementary Figure S6](#)). The relative expression levels of fusion transcripts were calculated by using the 2^{-ΔΔCt} method (Livak and Schmittgen, 2001). The experimental data were presented as the mean with standard deviation (mean ± SD) derived from three independent biological

replicates and three technical replicates. Statistical analysis was conducted by comparing the means of control and stressed plants using one-way analysis of variance (ANOVA) followed by Student's t-test, with a significance level set at $P < 0.05$. P -values less than 0.05 were considered statistically significant.

2.9 Benchmark analysis of fusion detection tools

To benchmark the performance of various fusion detection tools for *Cicer*, we evaluated STAR-Fusion (Haas et al., 2019), SQUID (Ma et al., 2018), MapSplice (Wang et al., 2010), Tophat-Fusion (Kim and Salzberg, 2011), and FusionMap (Ge et al., 2011). The performance of each tool was assessed using both a public dataset (PRJNA288321, Garg et al., 2016) and our in-house dataset. For both datasets, we then searched the validated fusion transcripts (true fusions) in the results generated by each tool. The accuracy of fusion prediction was calculated following the method described by Kumar et al. (2016) (Kumar et al., 2016), as given below.

$$\text{Sensitivity (\%)} = \left(\frac{TP}{TF} \right) * 100$$

$$\text{Precision (\%)} = \frac{TP}{TP + FP} * 100$$

$$F \text{ measure} = 2 * (S * P) / (S + P)$$

S: Sensitivity, P: Precision, TP: True positive, TF: Total fusions, FP: False positive

3 Results

3.1 Identification of fusion transcripts in *Cicer arietinum*

To comprehensively investigate fusion events within the chickpea transcriptome, high-throughput paired-end RNA-Seq data of 150 base pairs (bp) read length were generated using the Illumina sequencing platform. The samples were sequenced from poly(A)-enriched RNAs extracted from five distinct organs, viz., buds, leaves, pods, flowers, and stem ([Figure 1A](#)), along with two abiotic stress conditions (i.e., drought and salinity). After adapter trimming and quality-check, high-quality reads were mapped onto the CDC Frontier Genome (ASM33114v1) (Varshney et al., 2013) using HISAT2 (v2.2.1) (Kim et al., 2019), and it was found that >95% of the reads mapped onto the genome from each paired-end RNA-Seq sample ([Supplementary Table S1](#)). In total, 122.47 GB of data was obtained from all samples, representing about 230-fold of the chickpea genome size, and around 852.8 million high-quality reads were generated.

FusionMap (Ge et al., 2011), STAR-Fusion (Haas et al., 2019), and MapSplice (Wang et al., 2010) tools were employed for genome-wide identification of fusion transcripts in 26 RNA-Seq

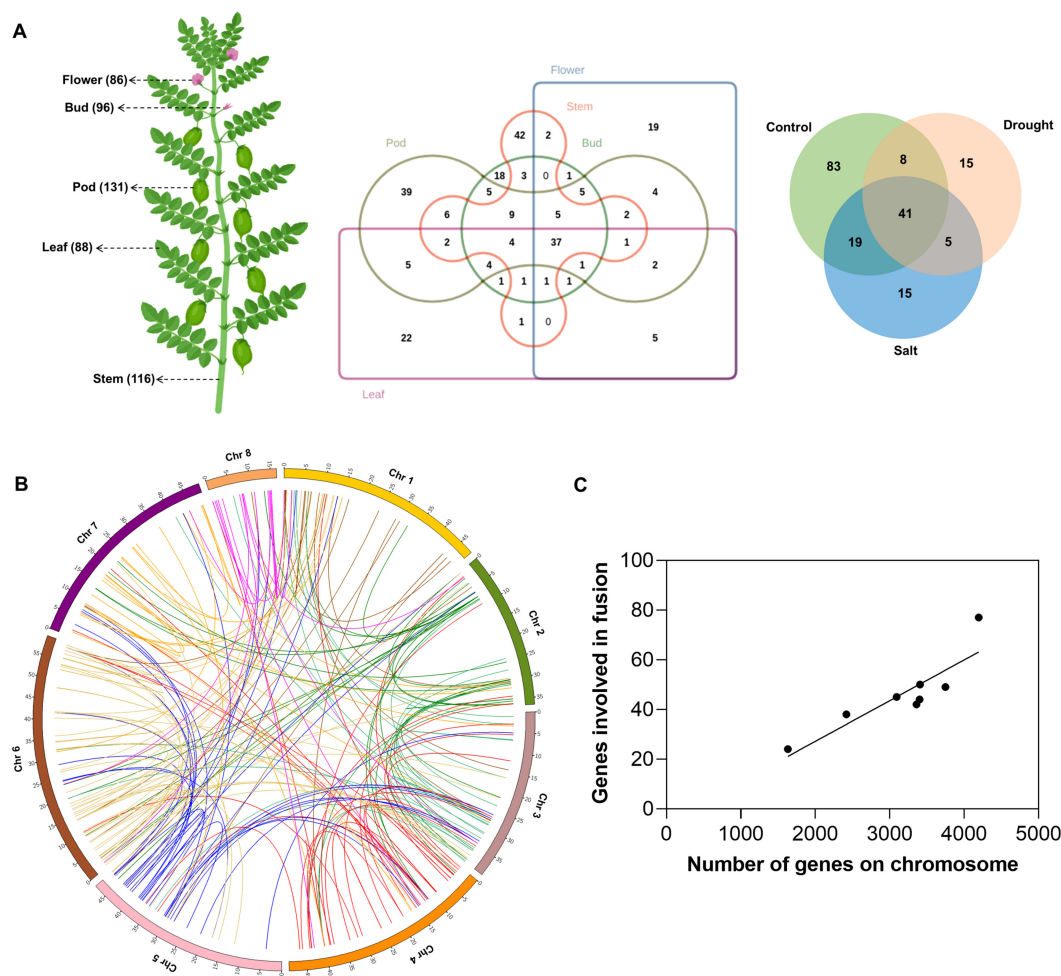


FIGURE 1
Distribution of fusion transcripts in *Cicer arietinum*, **(A)** The tissues included in this study and the number of fusions detected are indicated within parentheses following the tissue name. Venn diagram showing the overlap in fusion transcripts across different tissues and stress conditions, **(B)** Circos plot depicting the chromosomal distribution of parental genes involved in fusion events, with each connecting link representing a fusion event, **(C)** Correlation between the total number of genes per chromosome and the number of genes involved in fusion formation on that chromosome.

datasets of different tissues, and two abiotic stress conditions (Supplementary Table S1). The number of overlapping fusion transcripts detected between different tissues and stress samples is shown in Figure 1A. A total of 496, 721, and 533 fusion transcripts were identified by FusionMap (Ge et al., 2011), STAR-Fusion (Haas et al., 2019), and MapSplice (Wang et al., 2010), respectively (Supplementary Table S2). However, the number of unique fusion transcripts identified is 95, 109, and 140, respectively, resulting in 328 unique fusions (Supplementary Table S3) derived from 269 unique parental gene pairs. In total, 423 genes were involved in fusion formation, accounting for 1.4% (423/30,344) of the total annotated genes in *Cicer arietinum*. This proportion is similar to the percentage of genes involved in fusion events in other plant species, such as *Arabidopsis* (1%), soybean (1.7%), and rice (2.7%) (Cong et al., 2024). To verify the reliability of the predicted fusion transcripts, we randomly selected and validated 10 fusion transcripts in multiple biological replicates by PCR and Sanger sequencing (Supplementary Figure S1). The experimentally

validated junction sequences precisely matched the in-silico predicted breakpoint sites, confirming the accuracy of the fusion detection approach.

3.2 Features of fusion transcripts in *Cicer arietinum*

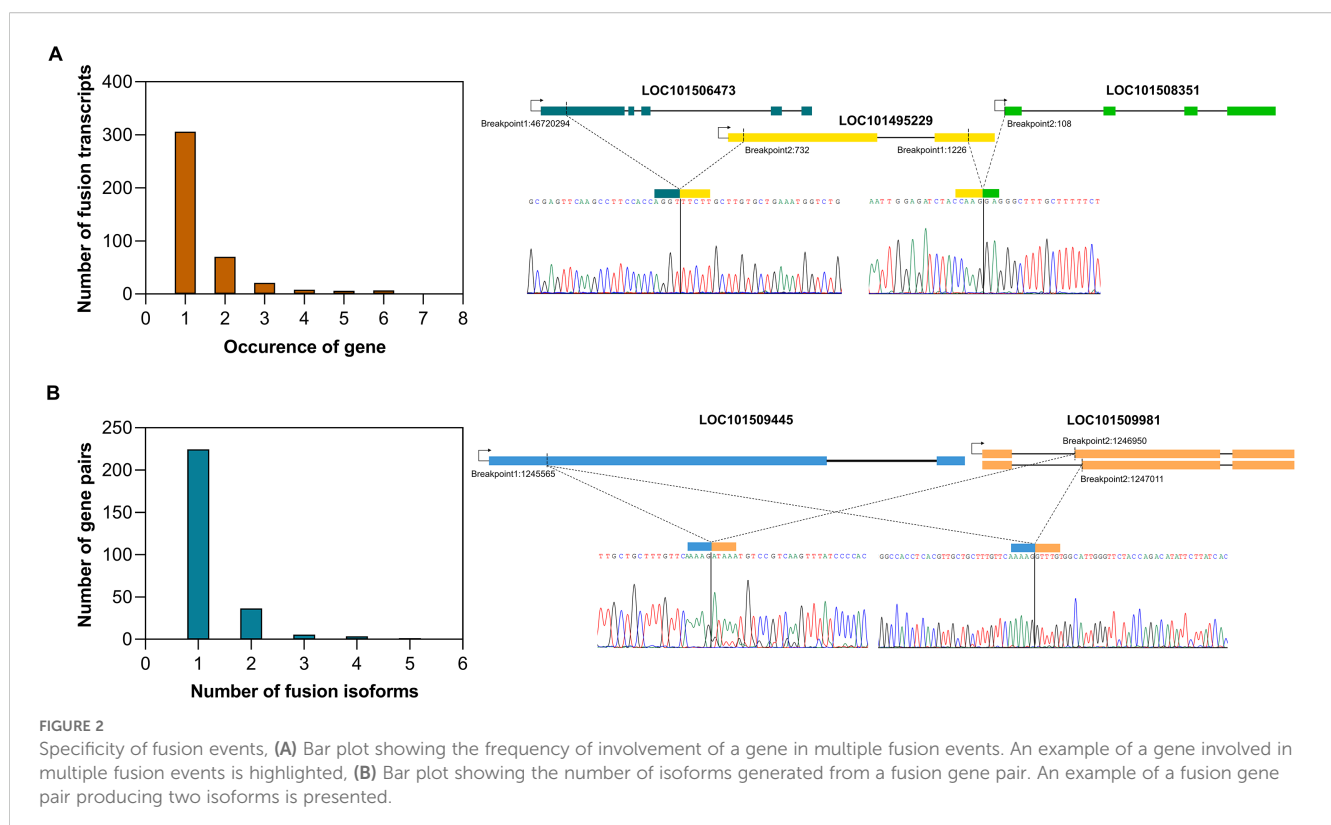
The prevalence of interchromosomal fusions (76%) as compared to intrachromosomal fusions (24%) in *Cicer arietinum* (Figure 1B) indicates that the formation of fusion transcripts is not strictly governed by the linear proximity of parental genes. It suggests that, beyond the linear arrangement of genes, spatial organization may play a role in facilitating fusion events in chickpea and thus underscores the value of further investigation into the spatial organization of the genome. A positive correlation between the number of genes participating in fusion events located at a particular chromosome and the total number of genes on that

chromosome suggests the lack of chromosome-level preference in fusion formation (Figure 1C). Out of the 423 genes involved in fusion formation, 306 participate in only one unique fusion event, while 70 genes form fusions with two different partner genes. This indicates that fusion events are highly specific, with individual genes preferentially associating with particular partners rather than engaging in multiple fusion events. However, the mechanisms behind partner selection in fusion events remain unknown.

Among all the validated fusion gene partners, each gene was found to fuse with only one partner, except for LOC101495229, which formed fusions with both LOC101508351 and LOC101506473 (Figure 2A). The number of fusion isoforms generated by a fusion gene pair was inversely related to the number of such gene pairs detected, highlighting the specificity of junction sites within fusion genes. Among the validated fusions, all gene pairs had a single junction site, except the LOC101509445_LOC101509981 fusion, which exhibited two isoforms. In both isoforms, the breakpoint in the LOC101509445 gene was identical, whereas LOC101509981 contributed two distinct breakpoints, each located at the exon boundaries of two different alternatively spliced transcripts of the LOC101509981 gene. It's a read-through fusion derived from two adjacent genes, where LOC101509445 is present on the reverse strand and LOC101509981 is present on the forward strand of chromosome 7. Here, the 5' gene belongs to ABC transporter I family member 1, and the 3' gene encodes for tRNA pseudouridine synthase. The two fusion events occur at the existing exon boundaries, hence producing in-frame fusions (Figure 2B).

To determine the fusion site within the parental genes, we analyzed the location of the breakpoint, whether it exists on the exon boundaries of both parental genes or one or none. It was observed that fusion transcripts where the breakpoint lies within the exon or in the UTR region were most abundant (57%), whereas fusion transcripts having a breakpoint at the exon border of either one or both parental genes were 25.3% and 17.6% respectively (Supplementary Table S3).

Trans-splicing is a known mechanism for fusion transcript formation (Li et al., 2009a). Junction pattern analysis showed that a significant portion of FTs exhibited canonical splice patterns GT-AG (69%); however, non-canonical splicing patterns were also found, such as GT-AT (4%), AT-AC (3%), CT-GC (3%), etc., implying apart from splicing, other unknown mechanisms may also contribute to fusion formation (Figure 3A). Another reported mechanism of fusion formation is transcriptional slippage mediated by the presence of short homologous sequences (SHSs) (Li et al., 2009b) at the junction, but less than 5% of total fusions exhibit SHSs at the breakpoints in *Arabidopsis*, soybean, rice, and maize (Cong et al., 2024). Similar results were observed in chickpea; 4% of total fusions showed SHSs at the junction, and 5.7% had both canonical splice sites and SHSs at their junction (Supplementary Table S4). All the experimentally validated fusions exhibited canonical splice sites at their junctions (e.g., LOC101515613_LOC113787786; Figure 3A), except for LOC101494819_LOC101493433, which showed a non-canonical GA-AG splice site. Additionally, three of the fusions also displayed short homologous sequences (SHSs) at the junctions, e.g., LOC101500131_LOC101505021 (Figure 3B).



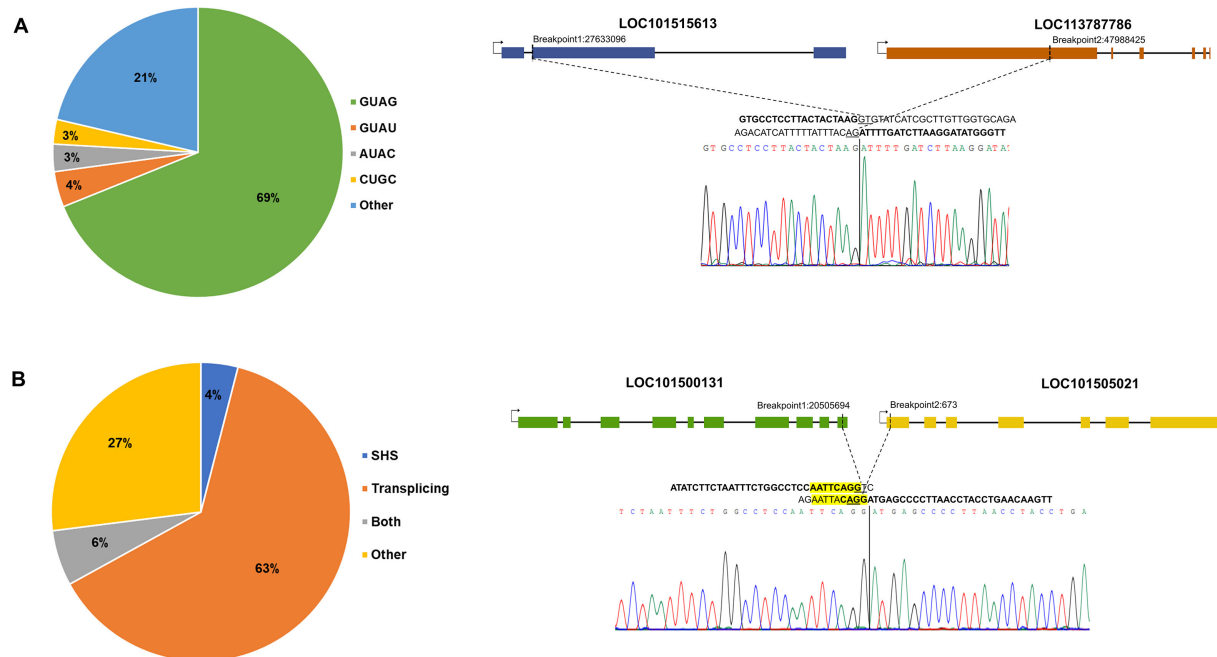


FIGURE 3

Junction site sequence analysis of fusion transcript, (A) The proportion of splice site patterns observed at fusion breakpoints is displayed, with an example fusion transcript demonstrating the canonical splice pattern (GT-AG) at the junction, (B) The contribution of different mechanisms to fusion generation is depicted, including trans-splicing and short homologous sequence (SHS) mediated fusion formation. An example fusion transcript displaying both canonical splice sites and SHS at the fusion junction is highlighted in yellow.

Predicting the coding potential of transcripts is essential for understanding their functional roles. Fusion transcripts were thus classified as protein-coding or lncRNA based on the computational predictions. In our analysis, only 12.5% of the fusion transcripts were predicted to possess protein-coding ability, while the remaining 87.5% of the transcripts lacked coding potential and are unlikely to be translated into proteins (Supplementary Table S5). However, it is important to note that computational prediction of non-coding potential does not necessarily imply a lack of function. Recent studies have shown that some transcripts annotated as lncRNAs can be translated into biologically active micro-peptides (Ruiz-Orera and Albà, 2019; Choi et al., 2019; Li et al., 2022; Pan et al., 2022; Patraquim et al., 2022). These findings underscore that while the majority of fusion transcripts may not encode proteins, they could still play important regulatory or functional roles as lncRNAs.

3.3 Properties of genes involved in fusion formation

Analysis of the transcriptome data showed a significant positive correlation ($R = 0.53$) between the expression of gene pairs involved in fusion formation, suggesting that these genes are linked in terms of expression (Figure 4A). Further to evaluate the impact of fusion formation on their parental genes, the expression levels of parental genes were compared between samples in which a particular fusion

was present and those in which it was absent. In most samples, no significant change in the expression of fusion parental genes (~75%) was observed, suggesting that fusion formation may be independently regulated (Figure 4B). However, a subset of parental genes displayed differential expression in fusion-present samples, with 24 genes exhibiting ≥ 2 -fold upregulation and only 3 showing ≥ 2 -fold downregulation (Supplementary Table S6).

Fusion transcripts may retain or be associated with the biological functions of their parental genes. To gain insights into the potential functions of the fusion transcripts, we employed gene ontology (GO) and KEGG pathway analyses of parental genes involved in fusion formation. The most enriched biological GO terms were translation, photosynthesis, response to light stimulus, and rRNA processing (Figure 4C). The most enriched terms in the molecular process analysis showed that fusion parental genes are involved in the structural constituent of the ribosome and in binding, such as RNA binding and chlorophyll-binding, and in enzymatic activities such as ATP hydrolysis, carboxylase, and dehydrogenase. The most enriched cellular component GO terms were cytosol, ribosome, and chloroplast. Pathway enrichment analysis showed that these genes are related to the Ribosome, carbon metabolism, and photosynthesis. Conclusively, fusion transcripts originate from genes with diverse functions that are distributed across various cellular compartments, and also show important enzymatic and binding activities. Notably, photosynthesis and ribosome-associated genes are majorly enriched among fusion precursor genes (Figure 4C).

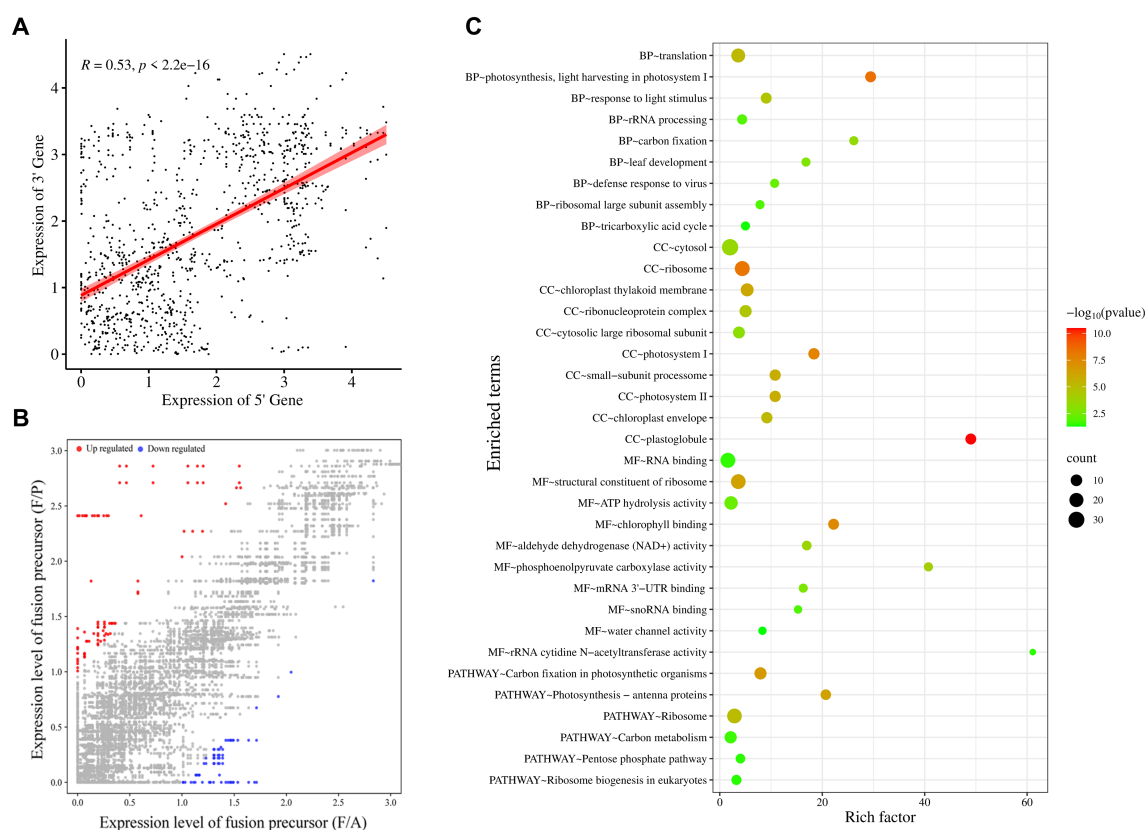


FIGURE 4

Functional categorization of parental genes involved in fusion. **(A)** Correlation between expression levels of 5' and 3' parental genes for each fusion event. Expression levels are presented as $\log_2(TPM + 1)$, with each point representing a fusion event. The R-value represents the Pearson correlation coefficient. **(B)** Scatterplot showing expression levels [$\log_2(TPM + 1)$] of parental genes in samples with (F/P) and without the corresponding fusions (F/A). Colored dots indicate parental gene expression (fold change ≥ 2): grey for no change, blue for downregulation, and red for upregulation. **(C)** Gene Ontology analysis representing biological processes (BP), cellular component (CC), molecular function (MF), and KEGG pathway analysis of genes involved in fusion transcript formation. These plots were generated using SRplot (Tang et al., 2023).

3.4 Validation of fusion transcripts from long-read RNA-Seq data

Transcript assembly of Illumina short reads to reconstruct full-length transcripts may introduce errors, particularly in regions with complex splicing patterns or novel splice sites. To overcome these challenges, long-read RNA sequencing technologies, such as PacBio, offer a significant advantage by capturing full-length transcripts in a single read, eliminating the need for assembly. For the identification of full-length fusion transcripts, publicly available long-read transcriptome sequencing datasets of *Cicer arietinum* were utilized. A BLASTn search of 200 bp fusion junction sequences, extracted from the results of fusion detection tools on Illumina datasets, against the PacBio long-read data identified a total of 95 fusion transcripts in the PacBio dataset, and their full length was predicted (Supplementary Table S7). The mean length of identified fusions was around 0.8 kb, and 70% of them had a length less than the mean length. We further compared the length of fusion transcripts and their parental genes and observed that the mean length of fusion transcripts was shorter than that of their parental

genes (Supplementary Figure S2). This indicates that only specific portions of the parental genes are involved in fusion formation.

3.5 Intra- and interspecific conserved fusion events

To explore intra-specific conserved fusion events, RNA-Seq data of 17 different chickpea genotypes were analyzed. The number of fusion transcripts identified across different genotypes varies significantly, ranging from 41 to 1031 (Figure 5A). This variation is likely attributed to differences in the number of samples analyzed for each genotype and the sequencing depth of the data. A fusion event was considered in a particular genotype if it was detected even in a single sample. After removing redundancy, interchromosomal fusion transcripts were found to be more prevalent than intrachromosomal fusion transcripts. Fusion genes identified in the ICC4958 genotype (269 fusion gene pairs) were searched across different genotypes (Figure 5B). Among these, 120 fusion gene pairs were conserved across multiple genotypes, while

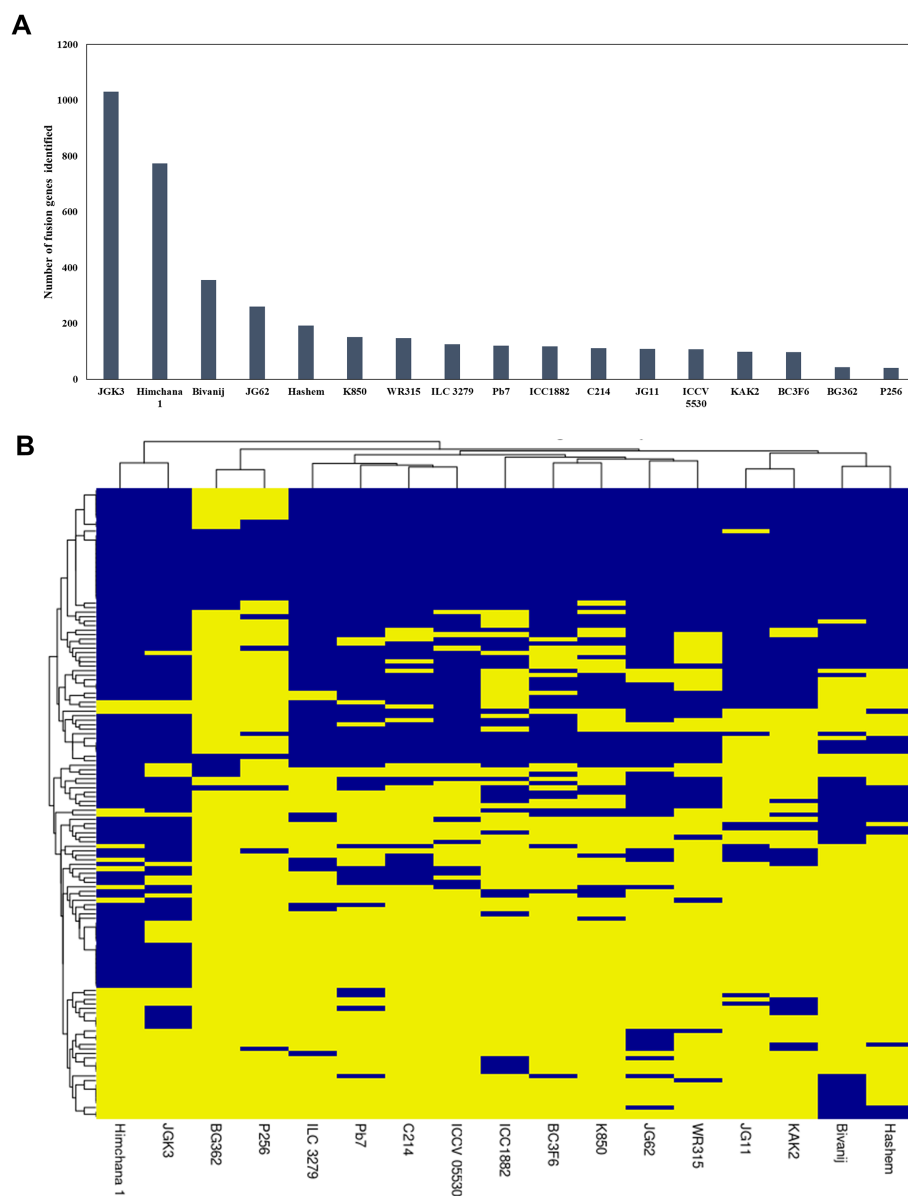


FIGURE 5

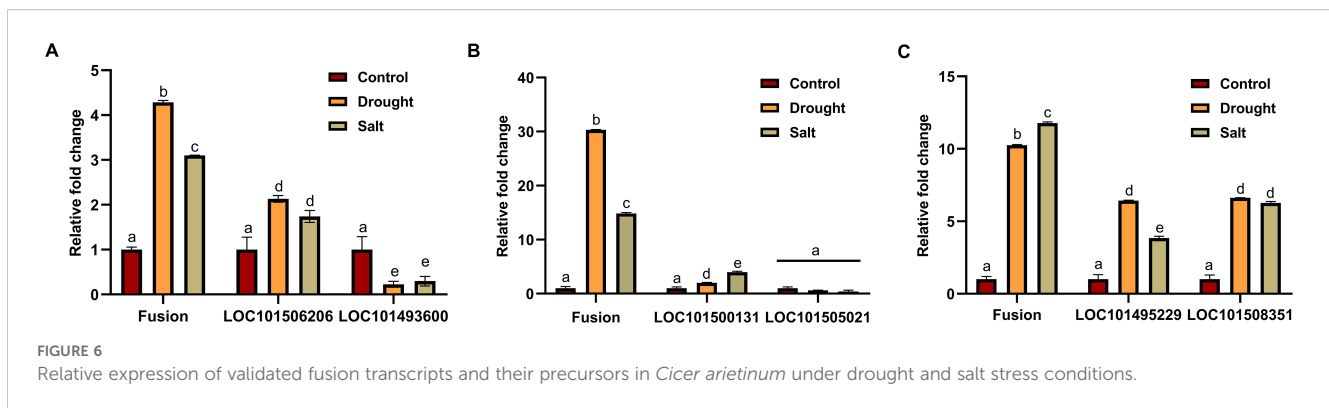
Fusion event profiles across different chickpea genotypes, (A) The number of fusion events identified in various chickpea genotypes, (B) Hierarchical clustering of 17 chickpea genotypes based on the presence or absence of fusion events identified in the ICC4958 genotype, with presence marked in blue and absence in yellow.

the rest were genotype-specific. Notably, 53 fusion gene pairs were detected in 10 or more genotypes, suggesting their potential biological significance and stability within the chickpea population.

To identify conserved FTs between Chickpea and *Arabidopsis*, FTs reported in AtFusionDB (Singh et al., 2019) were used. Among the 269 fusion genes identified in chickpea, 19 showed homology with 58 fusion events reported in *Arabidopsis* within AtFusionDB (Supplementary Figure S3A). Gene ontology analysis of these homologous genes revealed their involvement in essential biological processes, including metabolism and environmental responses (Supplementary Figures S3B–D). The presence of these fusion transcripts may play a critical role in enhancing biological functions linked to these processes.

3.6 Validation of fusion transcripts by qRT-PCR and Sanger sequencing

Expression of validated fusion transcripts under different abiotic stress and tissues by quantitative real-time PCR reveals that fusion transcripts are expressed at a very low level. Under different abiotic stress, three of the validated fusions (LOC101506206_LOC101493600, LOC101500131_LOC101505021, and LOC101495229_LOC101508351) showed differential expression, which implies that these fusions might play a crucial role during stress response (Figure 6). It was also observed that the relative fold change in expression of fusion transcripts was higher than that of their parental genes under



stress conditions, suggesting that fusion precursor formation is not solely governed by the expression of precursor genes but may involve additional regulatory mechanisms (Figure 6). Fusion transcripts can originate from parental genes that exhibit diverse responses to stress, including downregulation, upregulation, or no significant change in expression. Expression analysis of fusion transcripts across different tissues revealed their tissue-specific nature. For example, the LOC101500131_LOC101505021 fusion transcript exhibited upregulation in bud and downregulation in stem and flower, relative to its expression in leaves (Supplementary Figure S4). The LOC101506206_LOC101493600 is an interchromosomal fusion derived from a gene encoding a folylpolyglutamate synthase-like enzyme located on chromosome 5 and an uncharacterized gene from chromosome 8. This fusion produces an in-frame transcript that arises from an existing splice site of the 5' gene and a new splice site from the 3' gene. Since folylpolyglutamate synthase is a single-subunit enzyme, this fusion might add new domains, which may enhance its enzymatic activity (Muralla et al., 2008; Li et al., 2016). Interestingly, under stress conditions, the fusion transcript exhibited a higher relative fold change in expression compared to either of its precursor genes: LOC101506206, which was upregulated, and LOC101493600, which was downregulated, compared to control conditions.

The LOC101500131_LOC101505021 fusion was identified across all RNA-Seq samples with varying levels of expression. This fusion is formed by the joining of exon 10 of the LOC101500131 gene to exon 1 of the LOC101505021 gene, both genes are present on the forward strand of DNA and are involved in similar functions that are anaphase-promoting complex. This fusion uses a canonical splice site, however, not at the existing exon boundaries of the parental genes, and generates a frameshift fusion. Interestingly, this fusion showed significant upregulation in both drought and salt stresses as compared to the control condition. In contrast, the expression of the fusion precursor genes remains relatively unchanged in response to stress. Therefore, the regulation of the LOC101500131_LOC101505021 fusion appears to be independent of its parental genes.

The LOC101495229_LOC101508351 fusion is formed by joining exon 2 from LOC101495229 and exon 1 from LOC101508351. One of the parental genes acts as a ubiquitin ligase, while the other remains uncharacterized. This fusion occurs at the exon boundary of the LOC101508351 gene and uses

a new canonical splice site from the LOC101495229 gene. Both fusion transcript and its parental genes are stress-responsive; however, the relative change in fusion transcript expression under stress conditions is higher compared to parental genes.

3.7 Comprehensive assessment of fusion detection tools for *Cicer arietinum*

While many tools are available for fusion detection in humans, none of them is exclusively designed for plants, except EricScript-Plants (Benelli et al., 2012), which is limited to plant genomes available at Ensembl, and hence cannot be used for chickpea. In this study, we initially employed FusionMap, STAR-Fusion, and MapSplice for the comprehensive detection of fusion transcripts in chickpea. Further, a comparative analysis of five fusion detection algorithms was conducted to identify the best-performing tool for fusion detection for this plant species. This revealed the following order of tools based on sensitivity and F-measure: FusionMap > STAR-Fusion > MapSplice > SQUID > Tophat-Fusion. A similar ranking was observed with public data, except that STAR-Fusion performed better than FusionMap with public data (Supplementary Figures S5A, B). Of note, there is a small overlap in the fusions detected by these tools in both our dataset and public data (Supplementary Figure S5C, D). This could be due to false discoveries associated with individual software packages, or the fact that none of the tools is inclusive. Fusion gene pairs LOC101506206_LOC101493600 and LOC101503481_LOC101494793 were commonly detected by 3 out of the 5 fusion detection tools.

This analysis suggests that FusionMap detected the maximum number of true fusions, but it also identified a large number of false fusions, whereas STAR-Fusion identified fewer false fusions compared to FusionMap. Hence, FusionMap is the best performer in terms of sensitivity, whereas STAR-Fusion is the best performer in terms of precision. Hence, a combination of STAR-Fusion and FusionMap is a better choice for fusion transcript detection from RNA-Seq data for *Cicer* until more efficient tools become available.

Another factor that affects the detection of FTs from RNA-Seq data is the sequencing depth. It was observed that samples with higher sequencing depth showed a larger number of fusions than those with lower sequencing depth. For instance, in the P3 sample, the number of FTs detected is more than that in CS2 because the

sequencing depth of P3 is higher than CS2 (Supplementary Table S2).

4 Discussion

In recent years, it's already established that the fusion transcripts are not exclusive to tumors but also occur in normal human tissues and a wide variety of species (Babiceanu et al., 2016; Cong et al., 2024). Although significant progress has been made in mammalian fusion transcript research, studies on fusion transcripts in plants remain limited. While some studies have been conducted on model species such as *Arabidopsis* and rice, a comprehensive investigation of plant fusion transcripts is still lacking. A few plant-specific fusion transcripts databases are also available (Singh et al., 2019; Arya et al., 2024). Recently, a comprehensive profiling of fusion transcripts was conducted using public RNA-Seq datasets of *Arabidopsis*, Rice, and Chickpea (Chitkara et al., 2024). Most fusion transcripts were detected in only a single sample, raising concerns about their biological reproducibility and potential as artifacts. The current study addresses this gap through the use of in-house generated RNA-Seq datasets under defined abiotic stress conditions and across tissue types.

Here, we present the genome-wide identification of fusion transcripts in *Cicer arietinum* and explore their potential functional roles. Our study enhances the current understanding of fusion transcripts in legume plants. Notably, we found that interchromosomal fusion transcripts are more prevalent than those of intrachromosomal fusions in Chickpea, a pattern consistent with maize and soybean. In contrast, rice and *Arabidopsis* exhibit a higher frequency of intrachromosomal fusions (Cong et al., 2024). This indicates that the proportion of interchromosomal and intrachromosomal fusion transcripts is variable among different plants and may be influenced by several factors, including the gene density and compactness of the genome. The majority of gene pairs involved in fusion formation have no other partner genes, with only a few showing fusion isoforms, highlighting the specificity of these events. Additionally, we identified fusion transcripts that are tissue-specific or induced under specific stress conditions. Expression correlation analysis of genes involved in fusions revealed a correlation coefficient (r) of 0.5, aligning with the recent report that suggests genes with similar transcriptional activity are more likely to participate in conserved fusion events (Cong et al., 2024). Most parental genes involved in these fusions are multi-exonic and protein-coding, but only 12.5% of fusion transcripts showed protein-coding ability, while 87.5% were non-coding. Despite lacking coding capacity, these may still have regulatory or functional roles as lncRNAs.

Fusion transcript identification in different genotypes of chickpea revealed that some fusions are conserved across all genotypes, while others are specific to individual genotypes, indicating a link between genomic variation and fusion transcript occurrence. Homologous fusion events between *Cicer arietinum* and *Arabidopsis* suggest that the parental genes involved play key

roles in essential biological processes, underscoring the conservation of fusion events across different plant species.

Most fusion breakpoints exhibited canonical splice sites at the fusion junctions, implying their generation via the splicing mechanism. However, a few fusions showed overlapping sequences at the breakpoints, suggesting that SHSs also contribute to the formation of these fusions in *Cicer arietinum*. Further investigation revealed that SHSs differ among fusion transcripts and are mostly <10 bp in length. Due to the short reads generated by Illumina sequencing, full-length fusion transcripts could not be reliably detected. To address this limitation, we employed long-read RNA sequencing to identify the full-length structure of fusion transcripts. Of the 328 fusion events initially identified, 95 were confirmed in the long-read RNA-Seq data, likely due to the lower sequencing depth of this dataset.

Out of the predicted fusion transcripts, we confirmed the existence of 10 FTs, marking the first experimental validation of fusion transcripts in *Cicer arietinum*. Notably, three of these fusions were stress-responsive and exhibited upregulation in drought and salt stress, suggesting a potential function in the plant's stress response mechanisms. However, despite thorough validation efforts, we were unable to identify any fusion transcripts that were exclusively expressed under a specific stress and absent in the control condition. This indicates that while fusion transcripts might be involved in the stress response, their expression may not be strictly limited to stress conditions. These fusions may represent a broader regulatory mechanism that operates under both normal and stress conditions but becomes more pronounced in response to stress. The confirmation of stress-responsive fusion transcripts paves the way for exploring the functional implications of these fusions in stress adaptation and resilience. Future research could focus on dissecting the biological roles of these fusions, including their impact on gene expression and protein function under stress conditions. Additionally, investigating the potential regulatory networks and pathways associated with these fusions could provide a profound understanding of how *Cicer arietinum* adapts to environmental challenges.

5 Conclusion

This study systematically reveals the transcriptomic complexity arising due to fusion events in chickpea by combining transcriptomic datasets and experimental validation across diverse tissues (leaf, stem, bud, flower, and pod) and conditions (drought and salinity). Transcriptome analysis revealed the involvement of both coding and non-coding genes in the fusion events located either in close proximity or distant from each other. The parental genes involved in these fusions are associated with diverse biological pathways, suggesting the potential diverse roles of the fusion events that need to be further experimentally evaluated. Comparative analysis revealed both inter- and intraspecific conservation of fusion events, while also uncovering intra-species diversity, thereby underscoring their evolutionary significance. Expression

analysis revealed stress-responsive fusion transcripts, indicating their potential regulatory roles under abiotic stress. Overall, this work provides a comprehensive report of fusion transcript diversity in chickpea and offers a valuable foundation for future functional studies aimed at elucidating their biological and adaptive relevance.

Data availability statement

The datasets presented in this study are publicly available. This data can be found here: <https://www.ncbi.nlm.nih.gov/sra>, accession numbers PRJNA613159, PRJNA288321 and PRJNA1296865.

Author contributions

FH: Methodology, Data curation, Formal Analysis, Validation, Visualization, Writing – original draft. SZ: Formal Analysis, Writing – review & editing. SK: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This research is supported by the BT/PR40146/BTIS/137/4/2020, BT/PR40169/BTIS/137/71/2023 research grants by the Department of Biotechnology (DBT), New Delhi and core research grant of NIPGR, New Delhi.

Acknowledgments

FH gratefully acknowledges the Department of Biotechnology (DBT) for providing a research fellowship. The authors extend their gratitude to the DBT e-Library Consortium (DeLCON) for providing access to e-material and the Computational Biology & Bioinformatics Facility (CBBF) of NIPGR for their support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial

intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2025.1677098/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Chromatogram of 10 validated fusion transcripts from chickpea (A–J), where black line marks the junction site between the two fusion genes.

SUPPLEMENTARY FIGURE 2

Box plot depicting the relation between the predicted length of fusion transcripts and their parental genes.

SUPPLEMENTARY FIGURE 3

Inter-specific conserved fusion events, (A) Venn diagram showing orthologous fusion events shared between *Cicer arietinum* and *Arabidopsis thaliana*. (B–D) Gene Ontology (GO) enrichment analysis of orthologous fusion gene pairs, highlighting enriched categories in, (B) Molecular Function, (C) Biological Process, and (D) Cellular Component.

SUPPLEMENTARY FIGURE 4

Relative expression of three validated fusion transcripts (A–C) in *Cicer arietinum* across different tissues.

SUPPLEMENTARY FIGURE 5

Benchmark analysis of fusion detection tools using in-house generated RNA-seq data and publicly available RNA-seq data from chickpea, (A) Sensitivity of fusion detection tools evaluated for both datasets, calculated as Sensitivity (%) = (TP/TF) * 100, where TP represents true positives and TF is the total number of fusions, (B) F-measure for each tool, calculated as F-measure = 2 * (Sensitivity * Precision)/(Sensitivity + Precision), where Precision (%) = TP/(TP + FP) * 100. (C, D) Venn diagrams showing the overlapping fusion transcripts identified by different fusion detection tools in the in-house generated RNA-seq data (C) and the publicly available RNA-seq data, (D), highlighting common fusions across the tools. The performance metrics (sensitivity, specificity, and F-measure) were calculated based on the identification of validated fusion transcripts (true fusions) across all tools.

SUPPLEMENTARY FIGURE 6

Melt curve plot for the endogenous control gene and stress-responsive fusion transcripts, along with the primer amplification efficiency (%).

SUPPLEMENTARY TABLE 1

Summary of RNA-Seq reads and mapping statistics.

SUPPLEMENTARY TABLE 2

Summary of the number of fusion transcripts detected in each sample by three fusion detection methods and the sequencing depth of each sample.

SUPPLEMENTARY TABLE 3

List of fusion transcripts detected in RNA-Seq samples.

SUPPLEMENTARY TABLE 4

Information regarding Short homologous sequences (SHSs) detected at the junction of the fusion transcript.

SUPPLEMENTARY TABLE 5

Predicted coding potential of fusion transcripts.

SUPPLEMENTARY TABLE 6

List of differentially expressed genes in samples with and without fusion transcripts.

SUPPLEMENTARY TABLE 7

List of fusion transcripts detected in Long-read RNA-Seq data.

SUPPLEMENTARY TABLE 8

List of RNA-Seq samples used to study intra-specific diversity of fusion transcripts.

SUPPLEMENTARY TABLE 9

List of primers used for qRT-PCR of fusion transcripts and their parental genes.

References

- Annala, M. J., Parker, B. C., Zhang, W., and Nykter, M. (2013). Fusion genes and their discovery using high throughput sequencing. *Cancer Lett.* 340, 192–200. doi: 10.1016/J.CANLET.2013.01.011
- Arya, A., Arora, S., Hamid, F., and Kumar, S. (2024). PFusionDB: a comprehensive database of plant-specific fusion transcripts. *3 Biotech.* 14, 1–8. doi: 10.1007/S13205-024-04132-1
- Babiceanu, M., Qin, F., Xie, Z., Jia, Y., Lopez, K., Janus, N., et al. (2016). Recurrent chimeric fusion RNAs in non-cancer tissues and cells. *Nucleic Acids Res.* 44, 2859–2872. doi: 10.1093/NAR/GKW032
- Benelli, M., Pescucci, C., Marseglia, G., Severgnini, M., Torricelli, F., Magi, A., et al. (2012). Discovering chimeric transcripts in paired-end RNA-seq data by using EricScript. *Bioinformatics* 28, 3232–3239. doi: 10.1093/BIOINFORMATICS/BTS617
- Chao, Y., Yuan, J., Li, S., Jia, S., Han, L., Xu, L., et al. (2018). Analysis of transcripts and splice isoforms in red clover (*Trifolium pratense* L.) by single-molecule long-read sequencing. *BMC Plant Biol.* 18, 1–12. doi: 10.1186/S12870-018-1534-8
- Chen, X., Zhang, J., Liu, Y., Wang, Y., and Xu, Y. (2017). Emergence of a novel chimeric gene underlying grain number in rice. *Genetics* 205, 993–1002. doi: 10.1534/GENETICS.116.188201/-/DC1
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/BIOINFORMATICS/BTY560
- Chitkara, P., Ahmed, S., Kumar, S., and Singh, D. (2024). The landscape of fusion transcripts in plants: a new insight into genome complexity. *BMC Plant Biol.* 24, 1–18. doi: 10.1186/S12870-024-05900-0
- Choi, S. W., Kim, H. W., and Nam, J. W. (2019). The small peptide world in long noncoding RNAs. *Brief Bioinform.* 20, 1853–1864. doi: 10.1093/BIB/BBY055
- Chwalenia, K., Facemire, L., and Li, H. (2017). *Chimeric RNAs in cancer and normal physiology* Vol. 8 (Wiley Interdiscip Rev RNA). doi: 10.1002/WRNA.1427
- Cong, L., Zhang, S., Wang, Q., and Zhao, X. (2024). Conserved features and diversity attributes of chimeric RNAs across accessions in four plants. *Plant Biotechnol. J.* 22, 3151–3163. doi: 10.1111/PBI.14437
- Doskočilová, A., Kučera, R., and Barta, A. (2011). A nodulin/glutamine synthetase-like fusion protein is implicated in the regulation of root morphogenesis and in signaling triggered by flagellin. *Planta* 234, 459–476. doi: 10.1007/S00425-011-1419-7
- Dupain, C., Bocquet, M., and Renaudin, J. (2019). Discovery of new fusion transcripts in a cohort of pediatric solid cancers at relapse and relevance for personalized medicine. *Mol. Ther.* 27, 200. doi: 10.1016/J.YMTHE.2018.10.022
- FAO. (2023). Crops and livestock products. Available online at: <https://www.fao.org/faostat/en/#data/QCL/metadata>.
- Farrow, S. C., Hagel, J. M., and Facchini, P. J. (2015). Stereochemical inversion of (S)-reticuline by a cytochrome P450 fusion in opium poppy. *Nat. Chem. Biol.* 11, 728–732. doi: 10.1038/NCHEMBIO.1879
- Flowers, T. J., Galal, H. K., and Bromham, L. (2010). Salt sensitivity in chickpea. *Plant Cell Environ.* 33, 490–509. doi: 10.1111/J.1365-3040.2009.02051.X
- Garg, R., Sahoo, A., Tyagi, A. K., and Jain, M. (2010). Validation of internal control genes for quantitative gene expression studies in chickpea (*Cicer arietinum* L.). *Biochem. Biophys. Res. Commun.* 396, 283–288. doi: 10.1016/J.BBRC.2010.04.079
- Garg, R., Bhattacherjee, A., Jain, M., and Singh, A. (2016). Transcriptome analyses reveal genotype- and developmental stage-specific molecular responses to drought and salinity stresses in chickpea. *Sci. Rep.* 6, 1–15. doi: 10.1038/srep19228
- Ge, H., Liu, K., Juan, T., Fang, F., Newman, M., Hoeck, W., et al. (2011). FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution. *Bioinformatics* 27, 1922–1928. doi: 10.1093/BIOINFORMATICS/BTR310
- Gingras, T. R. (2009). Implications of chimeric non-co-linear transcripts. *Nature* 461, 206–211. doi: 10.1038/nature08452
- Guo, W., Qiu, D., and Zheng, T. (2019). CNIT: a fast and accurate web tool for identifying protein-coding and long non-coding transcripts based on intrinsic sequence composition. *Nucleic Acids Res.* 47, W516–W522. doi: 10.1093/NAR/GKZ400
- Gupta, S. K., Luo, L., and Yen, L. (2018). RNA-mediated gene fusion in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* 115, E12295–E12304. doi: 10.1073/PNAS.1814704115/SUPPL_FILE/PNAS.1814704115.SM02.MOV
- Haas, B. J., Dobin, A., Li, B., Stransky, N., Pochet, N., Regev, A., et al. (2019). Accuracy assessment of fusion transcript detection via read-mapping and *de novo* transcript assembly-based methods. *Genome Biol.* 20. doi: 10.1186/S13059-019-1842-9
- Hagel, J. M., and Facchini, P. J. (2017). Tying the knot: occurrence and possible significance of gene fusions in plant metabolism and beyond. *J. Exp. Bot.* 68, 4029–4043. doi: 10.1093/JXB/ERX152
- Jain, M., Bansal, J., Rajkumar, M. S., and Garg, R. (2022). An integrated transcriptome mapping the regulatory network of coding and long non-coding RNAs provides a genomics resource in chickpea. *Commun. Biol.* 5, 1–18. doi: 10.1038/s42003-022-04083-4
- Kibbe, W. A. (2007). OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.* 35. doi: 10.1093/NAR/GKM234
- Kim, M., Canio, W., Kessler, S., and Sinha, N. (2001). Developmental changes due to long-distance movement of a homeobox fusion transcript in tomato. *Science* 293, 287–289. doi: 10.1126/SCIENCE.1059805
- Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/s41587-019-0201-4
- Kim, D., and Salzberg, S. L. (2011). TopHat-Fusion: An algorithm for discovery of novel fusion transcripts. *Genome Biol.* 12, 1–15. doi: 10.1186/GB-2011-12-8-R72
- Koul, B., Panesar, P. S., and Chawla, M. (2022). Chickpea (*Cicer arietinum* L.) biology and biotechnology: from domestication to biofortification and biopharming. *Plants* 11. doi: 10.3390/PLANTS11212926
- Kumar, S., Vo, A. D., Qin, F., and Li, H. (2016). Comparative assessment of methods for the fusion transcripts detection from RNA-Seq data. *Sci. Rep.* 6, 21597. doi: 10.1038/srep21597
- Latysheva, N. S., and Babu, M. M. (2016). Discovering and understanding oncogenic gene fusions through data intensive computational approaches. *Nucleic Acids Res.* 44, 4487–4503. doi: 10.1093/NAR/GKW282
- Li, J., Lee, E. J., Chang, L., and Facchini, P. J. (2016). Genes encoding norcoclaurine synthase occur as tandem fusions in the Papaveraceae. *Sci. Rep.* 6, 1–12. doi: 10.1038/srep39256
- Li, S., Gao, J., and Zhao, Y. (2022). Micropeptides translated from putative long non-coding RNAs: Micropeptides translated from lncRNAs. *Acta Biochim. Biophys. Sin. (Shanghai)* 54, 292. doi: 10.3724/ABBS.2022010
- Li, H., Wang, J., Ma, X., and Sklar, J. (2009a). Gene fusions and RNA trans-splicing in normal and neoplastic human cells. *Cell Cycle* 8, 218–222. doi: 10.4161/CC.8.2.7358
- Li, X., Zhao, L., Jiang, H., and Wang, W. (2009b). Short homologous sequences are strongly associated with the generation of chimeric RNAs in eukaryotes. *J. Mol. Evol.* 68, 56–65. doi: 10.1007/S00239-008-9187-0
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402–408. doi: 10.1006/METH.2001.1262
- Ma, C., Shao, M., and Kingsford, C. (2018). SQUID: Transcriptomic structural variation detection from RNA-seq. *Genome Biol.* 19, 1–16. doi: 10.1186/S13059-018-1421-5
- Mukherjee, A., Singh, P., and Sinha, S. (2021). Computational analysis of sense-antisense chimeric transcripts reveals their potential regulatory features and the landscape of expression in human cells. *NAR Genom. Bioinform.* 3. doi: 10.1093/NARGAB/LQAB074
- Muralla, R., Lloyd, J., and Meinke, D. (2008). A bifunctional locus (BIO3-BIO1) required for biotin biosynthesis in arabidopsis. *Plant Physiol.* 146, 60. doi: 10.1104/PP.107.107409
- Nathawat, B. D. S., Sharma, O. P., Kumari, M., and Shivran, H. (2024). Effect of nutrients on wilt in chickpea. *Legume Res.* 47, 152–155. doi: 10.18805/LR-4490

- Pan, X., Zhou, T., and Zhang, Y. (2022). Functional micropeptides encoded by long non-coding RNAs: A comprehensive review. *Front. Mol. Biosci.* 9. doi: 10.3389/fmolb.2022.817517/XML/NLM
- Patraquim, P., Albuquerque, I., and Monteiro, F. (2022). Translation and natural selection of micropeptides from long non-canonical RNAs. *Nat. Commun.* 13, 1–15. doi: 10.1038/s41467-022-34094-y
- Perteu, M., Kim, D., Perteu, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* 11, 1650–1667. doi: 10.1038/NPROT.2016.095
- Qiao, Y., Li, C., and Huang, Y. (2019). Comprehensive identification of the full-length transcripts and alternative splicing related to the secondary metabolism pathways in the tea plant (*Camellia sinensis*). *Sci. Rep.* 9, 1–13. doi: 10.1038/s41598-019-39286-z
- Qin, F., Zhang, Y., Liu, J., and Li, H. (2017). SLC45A3-ELK4 functions as a long non-coding chimeric RNA. *Cancer Lett.* 404, 53–61. doi: 10.1016/J.CANLET.2017.07.007
- Ruiz-Orera, J., and Albà, M. M. (2019). Conserved regions in long non-coding RNAs contain abundant translation and protein–RNA interaction signatures. *NAR Genom. Bioinform.* 1, e2. doi: 10.1093/NARGAB/LQZ002
- Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., et al. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists, (2021 update). *Nucleic Acids Res.* 50, W216–W221. doi: 10.1093/NAR/GKAC194
- Singh, A., Roy, S., and Chaturvedi, A. (2017). PLncPRO for prediction of long non-coding RNAs (lncRNAs) in plants and its application for discovery of abiotic stress-responsive lncRNAs in rice and chickpea. *Nucleic Acids Res.* 45, e183–e183. doi: 10.1093/NAR/GKX866
- Singh, A., Zahra, S., Das, D., and Kumar, S. (2019). AtFusionDB: a database of fusion transcripts in *Arabidopsis thaliana*. *Database (Oxford)*. doi: 10.1093/DATABASE/BAY135
- Sutton, R. E., and Boothroyd, J. C. (1986). Evidence for trans splicing in trypanosomes. *Cell* 47, 527–535. doi: 10.1016/0092-8674(86)90617-3
- Tan, Y., Deng, J., and Zhao, Y. (2019). Single-molecule real-time sequencing facilitates the analysis of transcripts and splice isoforms of anthers in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *BMC Plant Biol.* 19. doi: 10.1186/S12870-019-2133-Z
- Tang, L., Hu, X., and Liu, S. (2023). SRplot: A free online platform for data visualization and graphing. *PLoS One* 18, e0294236. doi: 10.1371/JOURNAL.PONE.0294236
- Turner, M., Adhikari, S., and Subramanian, S. (2013). Salinity tolerance and ion accumulation in chickpea (*Cicer arietinum* L.) subjected to salt stress. *Plant Soil* 365, 347–361. doi: 10.1007/S11104-012-1387-0/METRICS
- Varley, K. E., Gertz, J., Bowling, K. M., Parker, S. L., Reddy, T. E., Pauli-Behn, F., et al. (2014). Recurrent read-through fusion transcripts in breast cancer. *Breast Cancer Res. Treat* 146, 287. doi: 10.1007/S10549-014-3019-2
- Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., et al. (2013). Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* 31, 240–246. doi: 10.1038/nbt.2491
- Wang, L., Park, H. J., Dasari, S., et al. (2013). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* 41, e74. doi: 10.1093/NAR/GKT006
- Wang, X., Haberer, G., and Mayer, K. F. (2010). MapSplice: Accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res.* 38, e178. doi: 10.1093/NAR/GKQ622
- Winzer, T., Gazda, V., He, Z., Kaminski, F., Kern, M., Larson, T. R., et al. (2015). Plant science. Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein. *Science* 349, 309–312. doi: 10.1126/SCIENCE.AAB1852
- Zhang, H., Jin, J., Tang, L., Zhao, Y., and Gu, X. (2010). Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res.* 20, 646. doi: 10.1101/GR.100677.109
- Zhou, L., Yang, L., and Ma, X. (2022). Genome-Wide Profiling of Alternative Splicing and Gene Fusion during Rice Black-Streaked Dwarf Virus Stress in Maize (*Zea mays* L.). *Genes (Basel)* 13. doi: 10.3390/GENES13030456