Food Chemistry: Molecular Sciences 10 (2025) 100238

Contents lists available at ScienceDirect

Food Chemistry: Molecular Sciences

journal homepage: www.sciencedirect.com/journal/food-chemistry-molecular-sciences

Deciphering the complex molecular architecture of the genetically modified soybean FG72 through paired-end whole genome sequencing

Fan Wang [a], Shengtao Lu [a,b], Wenting Xu [a], Litao Yang [a,b,c,*]

[a] Joint International Research Laboratory of Metabolic and Developmental Sciences, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, PR China
[b] Zhejiang Yuzhi Biotechnology Company Ltd, Ningbo 315032, PR China
[c] Yazhou Bay Institute of Deepsea Sci-Tech, Shanghai Jiao Tong University, Sanya 572024, PR China

ARTICLE INFO

Keywords:
GM soybean
FG72
Molecular characterization
Paired-end whole genome sequencing
T-DNA integration
Copy number
Structural variation

ABSTRACT

The clear molecular characterization of genetically modified (GM) plants and animals is a prerequisite for obtaining regulatory approval and safety certification for commercial cultivation. This characterization includes the identification of the transferred DNA (T-DNA) insertion site, its flanking sequences, the copy number of inserted genes, and the detection of any unintended genomic alterations accompanying the transformation process. In this study, we performed a comprehensive molecular characterization of the well-known GM soybean event FG72 using paired-end whole-genome sequencing (PE-WGS). We examined the T-DNA insertion site, flanking sequences, the entire structure and copy number of the T-DNA integration, the presence of plasmid backbone sequences, and genome-wide structural variations (SVs). Our analysis revealed that the T-DNA integrated into two distinct sites on chromosome 15 of the host genome, accompanied by a translocation of host genomic sequences. One site harbored a partial T-DNA integration, while the other site contained two tandem repeats of the full T-DNA. Importantly, no plasmid backbone sequences were detected in the host genome, indicating a clean T-DNA integration during the biolistic transformation process. Furthermore, we identified numerous genome-wide SVs, with chromosome 15 ranking second among all 20 chromosomes in terms of SV frequency, and most of these variations occurring within gene-coding regions. These results provide a refined and comprehensive molecular characterization of the FG72 soybean event, which could further support its commercial approval and cultivation. Our work highlights the utility of the PE-WGS approach as a sensitive and labor-efficient alternative to conventional molecular characterization techniques, generating comprehensive data to facilitate the safety assessment of GM crops during research and commercialization pipelines.

1. Introduction

The development and commercialization of genetically modified (GM) crops have accelerated in recent decades, driven by the promise of enhanced traits such as improved pest and disease resistance, abiotic stress tolerance, and improved nutritional profiles (James, 2021). The commercialization of GM crops has raised significant global concerns regarding their potential impact on human health and the environment. Comprehensive risk assessment is a critical step in ensuring the safety and responsible development of these novel organisms. At the core of this risk assessment process is the detailed molecular characterization of GM crops, which provides crucial insights into the genetic modifications and their potential consequences (EFSA, 2011). By thoroughly understanding the molecular characteristics of GM plants, risk assessors and regulatory authorities can make informed decisions about their safety and suitability for commercial release (Herman, 2013).

The molecular characterization of GM plants involves a comprehensive examination of the transferred DNA (T-DNA) integration within the host genome. This analysis includes determining the precise location, copy number, and structural integrity of the inserted genetic elements (Kawall, 2019). Assessing the stability of the T-DNA integration across multiple generations is also essential, as it ensures the consistent and predictable expression of the desired traits (Halpin, 2005). Furthermore, the molecular characterization process aims to identify

* Corresponding author at: Joint International Research Laboratory of Metabolic and Developmental Sciences, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, PR China
E-mail address: yylltt@sjtu.edu.cn (L. Yang).

https://doi.org/10.1016/j.fochms.2024.100238
Received 20 September 2024; Received in revised form 9 December 2024; Accepted 22 December 2024
Available online 24 December 2024

any unintended DNA insertions that may have occurred during the transformation process. The presence of such unintended genetic modifications could potentially lead to unanticipated effects, which must be carefully evaluated and mitigated (Goodwin et al., 2019).

Traditionally, GM plant molecular characterization has relied on various techniques, such as Southern blotting, polymerase chain reaction (PCR), and Sanger sequencing (Holst-Jensen et al., 2012). While these methods have been widely used, they often have limitations in terms of resolution, throughput, and the ability to provide a comprehensive understanding of the genetic modifications (Kovalic et al., 2012).

In recent years, the advent of next-generation sequencing (NGS) technologies has revolutionized the identification of T-DNA integration in GM plants, offering significant advantages over traditional methods (Chen et al., 2021; Kovalic et al., 2012; Yang et al., 2013; Zhang et al., 2022). NGS-based approaches provide a more comprehensive and higher-resolution analysis of the genetic modifications introduced into the host genome (Park et al., 2017). Kovalic et al. and Siddique et al. utilized PE-WGS to analyze the integration site and flanking regions of GM maize events, providing a comprehensive understanding of the transgenic insert and its genomic context (Kovalic et al., 2012; Siddique, Wei, Li, Zhang, & Shi, 2019). A combination of NGS techniques, including long-read sequencing and mate-pair whole-genome sequencing (MP-WGS), was established to characterize the molecular profile of GM events, revealing the precise structure of the transgenic insert and identifying potential unintended genomic alterations (Srivastava et al., 2014; Zastrow-Hayes et al., 2015). Edwards et al. used T-DNA-specific primers and high-throughput sequencing to reveal the transgene insertion in GM *Camelina sativa* (Edwards, Hornstein, Wilson, & Sederoff, 2022). Furthermore, our lab developed the pipelines of TranSeq and LIFE-Seq based on various NGS strategies (PE-WGS, MP-WGS, and PacBio SMRT sequencing) to decipher the full molecular characterization of various GM crops besides the T-DNA integration site and flanking sequences rice events (Xu, Zhu, Li, Hu, & Mao, 2022; Yang et al., 2013; Zhang et al., 2020; Zhang et al., 2022). Additionally, several bioinformatic tools for identifying the T-DNA insertion have been specially developed, such as TranSeq, TDNAscan, TC-hunter, and T-LOC (Börjesson et al., 2022; Li, Wang, You, Zhou, & Zhou, 2022; Sun et al., 2019; Yang et al., 2013).

The herbicide-tolerant soybean is the most widely commercialized GM crop worldwide. The GM soybean event FG72 has been approved for planting and use as raw material for food/feed in several countries for more than 15 years, having been developed with tolerance to the herbicides glyphosate and isoxaflutole by Bayer CropScience (EFSA, 2015). However, the molecular characterization of FG72 was initially reported through conventional molecular analyses, such as Southern blotting, PCR amplification, and Sanger sequencing (EFSA, 2015). A comprehensive molecular characterization at the whole-genome level is still lacking.

Therefore, the present study aims to decipher the full molecular characterization and SVs accompanying the transgene transformation of FG72 using a PE-WGS strategy. This in-depth analysis will further confirm and improve upon the previously presented data on the molecular profile of FG72. Furthermore, by simplifying the experimental and data analysis workflows of the PE-WGS approach, we hope to promote its wider application in the comprehensive molecular characterization and risk assessment of GM crops.

2. Materials and methods

2.1. Plant Material and Genomic DNA Extraction

Seed powder of the GM soybean event FG72 and its recipient line Jack were provided by the developer. The FG72 event was produced by biolistic transformation, incorporating two gene cassettes: HPPDPf W336 and 2mEPSPS, which express the novel proteins 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) and modified *p*-hydroxyphenylpyruvate dioxygenase (HPPD), respectively (EFSA, 2015). Soybean genomic DNA was isolated and purified using a Plant Genomic DNA Kit (Cat. no. DP3053; TIANGEN Company, China). The quality of the extracted genomic DNA was evaluated using a Nanodrop ND-8000 (ThermoFisher, USA) and 1 % agarose gel electrophoresis. The extracted DNA was used for Illumina paired-end sequencing, PCR, and droplet digital PCR (ddPCR) analyses.

2.2. Paired-End Whole-Genome Sequencing

For PE-WGS library construction, approximately 5 μg of genomic DNA was fragmented to a peak size of 500 bp by sonication with a Diagenode Bioruptor UCD-300TM-EX (Denville, NJ, USA). The Illumina TruSeq DNA Sample Preparation Kit.

(Illumina, San Diego, CA, USA) was used to construct the PE library, with a mean insert size of 500 bp. The library quality was assessed using the Agilent Bioanalyzer DNA 1000 kit (Agilent Technologies, USA). Library sequencing was conducted on an Illumina HiSeq 2000 DNA sequencer, generating 2 × 150 bp paired-end short sequence reads after removing index sequences, short reads, and poor-quality reads. Sufficient fragment sequences (reads) were obtained, and the effective genome coverage reached 40 X for comprehensive analysis (as described below).

2.3. PE-WGS data analysis

Compressed sequence files were unzipped and converted into the FastQ format, which served as input for the quality control tool FastQC (Chen, Zhou, Chen, & Gu, 2018). The BWA-mem algorithm (Jung & Han, 2022) was used to map the reads to the soybean reference genome (NCBI RefSeq assembly accession: *GCF_000004515.6*) and plasmid sequence. SAMtools (Li et al., 2009) was used to transform the file format, sort, and index the resulting BAM files. A custom Python script in a Linux shell was used to extract the junction reads. Reads containing the junction region sequence were identified through IGV analysis. The DNA fragment containing the junction sequence was obtained through the sequence assembly software MEGAHIT (Li, Liu, Luo, Sadakane, & Lam, 2015). The inserted junction location was further identified using BLAST against the reference genome. To determine the copy number of exogenous genes, BWA was used to map the reads to the reference genome and transgene sequences. The copy number was then estimated by comparing the coverage of the reference genome with the transgene coverage, and the read mapping pattern was visualized using IGV (Robinson, Thorvaldsdottir, Turner, Mesirov, & Alkan, 2023). To detect SVs in FG72, the tools DELLY and Manta were used with default parameters to analyze rearrangements and translocations caused by the exogenous DNA insertion (Chen et al., 2016; Rausch et al., 2012).

2.4. Conventional PCR and Sanger sequencing

Primers for the exogenous DNA integration site and flanking sequence confirmation were designed based on the assembled contigs and are listed in Table S1. Conventional PCR was performed in a final volume of 25 μL, containing 50 ng of genomic DNA, 1× PCR buffer, 0.2 mM of each dNTP, 0.4 μM of each primer, and 1 U of Taq DNA polymerase (Thermo Fisher Scientific, USA). The thermal cycling conditions were as follows: initial denaturation at 95 °C for 5 min, followed by 35 cycles of 95 °C for 30 s, 58 °C for 30 s, and 72 °C for 1 min, with a final extension step at 72 °C for 7 min. The PCR products were analyzed by agarose gel electrophoresis using a 1.5 % (*w/v*) agarose gel stained with GelRed (Biotium, USA). The gel was visualized under UV light, and the band sizes were compared to a DNA size marker to confirm the expected amplicon sizes. Following PCR amplification, the expected PCR amplicons were purified using the Qiagen MinElute PCR Purification Kit. The purified amplicons were then subjected to Sanger sequencing at BGI

Genomics Co., Ltd. in Shanghai, China, to obtain high-quality Sanger sequencing data.

2.5. Droplet Digital PCR Analysis

To verify the copy number of exogenous genes, primer/TaqMan probe sets were designed based on the sequences of hppdPfW336, 2mepsps, and Lectin (Table S2). The 20 μL ddPCR reactions included 10 μL of 2× ddPCR Supermix (Bio-Rad, Foster City, CA, USA), optimized concentrations of each primer/probe set, 1 μL of DNA template, and ddH2O to the final volume. The QX200 Droplet PCR system (Bio-Rad, Foster City, CA, USA) was used for droplet generation in eight-well cartridges and subsequent processing. The ddPCR assay conditions were as follows: 5 min at 95 °C, followed by 40 cycles of 30 s at 95 °C and 60 s at 58 °C, with a ramp rate of 2.0 °C/s. After cycling, each sample was incubated at 98 °C for 10 min and then cooled to 4 °C. After PCR amplification, the PCR plates were transferred to a QX200 droplet reader (Bio-Rad, Pleasanton, CA, USA), and data acquisition and analysis were performed using QuantaSoft (Bio-Rad, Pleasanton, CA, USA). All analyses were performed in triplicate, with three parallels in each instance.

3. Results

3.1. The principle and bioinformatics pipeline for molecular characterization analysis

The overall procedure includes the steps of plant genomic DNA extraction, DNA fragmentation, sequencing library construction, sequencing, and data analysis (Fig. 1). Data analysis is the most crucial and complex step in molecular characterization analysis. Since most previously reported tools (TDNAscan, T-LOC, and TC-hunter, etc.) besides TranSeq mainly focused on identifying the T-DNA insertion site and flanking region, they cannot satisfy the aim of comprehensive molecular characterization analysis. We built a fully functional bioinformatics pipeline that contained two parts: one for deciphering the T-DNA integration, and the other for revealing the unintended genome SVs accompanied by T-DNA integration (Fig. 1b).

In the T-DNA integration analysis, all quality-refined reads are mapped to the plasmid sequence and the soybean reference genome sequences using BWA-mem, respectively. Then, three categories of reads (Fig. 1a) were filtered and retained: i) paired reads mapped to the soybean reference genome (type A); ii) paired reads mapped to the plasmid sequence (type C); iii) reads with one end mapped or partially mapped to the soybean genome and the other end mapped to the plasmid sequence (discordant reads: type B; split reads: type D, E). The reads of types B, D, and E were further analyzed using a custom-developed Python script in a Linux environment (Additional file 1; Fig. 1b). The integrated site and its flanking sequence were obtained using the sequence assembly software Megahit based on the extracted reads of types B, D, and E. The paired reads were also aligned to the reference sequence using BLASTN and IGV to eliminate false-positive reads from the host soybean genome, which significantly reduced the number of candidate reads for further analysis. Here, we optimized the BWA parameter (bwa mem -t 8) to allow split alignments and identify junction reads quickly using a multi-threaded approach. Each step can be run independently, and it can analyze several samples automatically to optimize the use of available computational resources. According to the reads of types B, C, D, and E, the copy number and plasmid backbone residence were calculated through the new parameter of 'Samtools depth' and evaluated through IGV analysis, respectively.

For identifying the SVs that occurred in the FG72 event, two modules (DELLY and Manta) with default parameters were used. The extracted data from a reference sequence file and a sequence alignment file were used to filter and identify the predicted SVs. The resulting candidate SV sites were captured in a VCF (Variant Call Format) file. These candidate SVs were filtered by supporting evidence filters and then ranked by

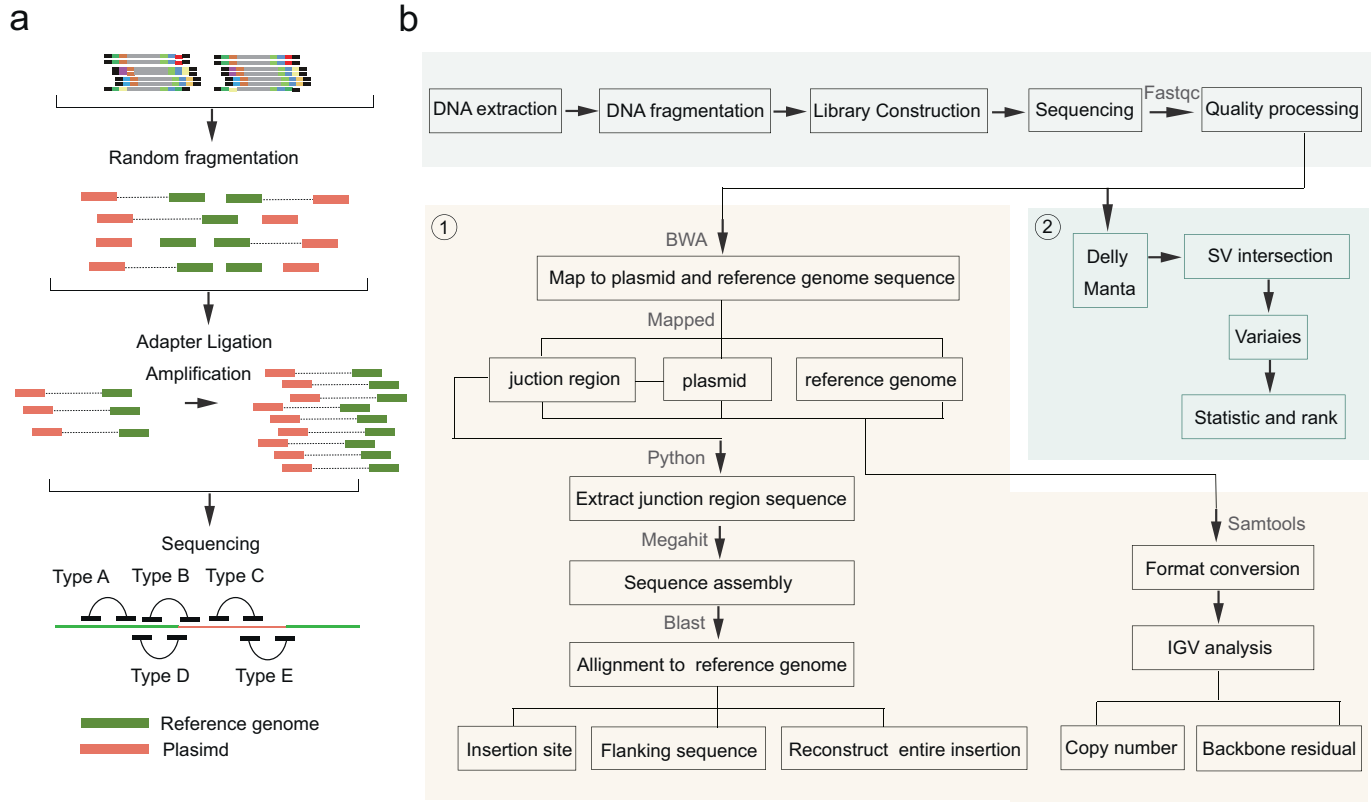

Fig. 1. Principle and bioinformatics pipeline for the analysis of T-DNA integrations and accompanying genomic SVs. (a) The types of sequenced paired-end reads. (b) The workflow and established bioinformatics pipelines. ① Pipeline for T-DNA integration analysis. ② Pipeline for structural variation (SV) evaluation.

normalized high-quality read count for each predicted variant.

In the PE-WGS of FG72, the reads with 150 bp in length were generated by paired-end sequencing. The sequencing data was evaluated with Fastqc and processed for removing the adapter sequence and filtering low-quality reads with Phred scores below 20. After quality control processing, a total of 314.66 M cleaned reads were obtained with a sequencing coverage of 45.06 for FG72. The sequencing data of the non-GM soybean cultivar JACK with 36.39 sequencing coverage was obtained from NCBI with the accession number SRX4968779 (Table S3). All the trimmed read pairs were used for further analysis according to the above bioinformatics pipeline.

3.2. Two T-DNA insertion sites and flanking sequences were identified and confirmed in FG72 soybean

To identify the T-DNA insertion sites and flanking sequences, reads of types B, C, D, and E were filtered from the developed pipeline and used to detect the T-DNA insertion sites and flanking sequences through a Python script in Linux. A total of 179 reads were collected, located on five chromosomes: 161 reads on chromosome 15, 7 reads on chromosome 6, 7 reads on chromosome 8, 2 reads on chromosome 14, and 2 reads on chromosome 18 (Table S4). To eliminate potential false-positive reads originating from the native genome, each discordant read and split read were aligned to the reference sequence using BLASTN and IGV. After detailed analysis, the reads on chromosomes 6, 8, 14, and 18 were recognized as false positives due to high sequence similarity between the T-DNA and the soybean genome, as well as potential cross-contamination during the Illumina sequencing library construction process.

From the 161 reads on chromosome 15, four breakpoints were identified at 13,715,991, 15,127,131, 15,127,156, and 16,897,687, representing the junctions between the genomic and plasmid sequences. IGV visualization results showed that there are two T-DNA insertions on chromosome 15 in FG72, one insertion (IS1) between 13,715,991 and 15,127,156 (Fig. 2a), and another (IS2) between 15,127,131 and 16,897,687 (Fig. 2b). Additionally, a large DNA fragment of the soybean genome, from 13,715,991 to 16,897,687, was observed to be translocated (Fig. 2c).

Five contigs were obtained from the assembly analysis of these reads (Additional file 2). BLASTN analysis of the contigs revealed the detailed insertion sites and flanking sequences. Contig 1 contains the left flanking sequence of IS1 and partial Ph4a748 ABBC, Contig 2 contains the left flanking sequence of IS2 and partial 3' HistonAt, Contig 3 contains partial T-NOS and 3' HistonAt, Contig 4 contains the right flanking sequence of IS2 and partial 3' HistonAt, and Contig 5 contains two soybean genome DNA fragments adjacent to chromosome 15: 13,715,991 and chromosome 15: 16,897,687. Contig 3 and Contig 4 overlapped with partial 3' HistonAt.

Combining the results from IGV and the assembled contigs, we deduced that there are two T-DNA insertions in FG72, accompanied by one large DNA fragment (from chromosome 15: 15,127,130 to 16,897,686) of host genome translocation (Fig. 2d). IS1 is inserted at chromosome 15: 13,715,991 and contains partial sequence of Ph4a748 ABBC. IS2 is inserted at 15,127,131 with a 25 bp host genomic DNA deletion, and the both ends of IS 2 were the element of 3' HistonAt.

We also confirmed the arrangement of the T-DNA insertions through conventional PCR and Sanger sequencing. A total of six primer sets (Fig. 2a, b, and c) were designed according to the revealed insertion sites and five contigs, and used for PCR amplification and sequencing analysis. The expected PCR amplicons were all observed in FG72, but not in the wild-type JACK (Fig. 2e). The Sanger sequencing analysis further confirmed that the IS 1 contains the 158 bp Ph4a748 ABBC promoter sequence.

3.3. The two copies of T-DNA were determined in FG72

Based on the filtered reads that mapped to the plasmid from FG72 and JACK (types B, C, D, and E), the copy number of the entire T-DNA and its individual elements (T-NOS, 3'histonAt, hppdPfW336, and 2mepsps, etc.) was estimated by comparing the average depth of the host genome and the transgene. This provided the quantitative relationship between the target region and the entire genome. The formula for calculating copy number is: $Copy\ number = N \times L/(D \times l)$, where N is the number of mapped reads, L is the read length, D represents the overall sequencing depth, and l represents the length of the target gene. Table 1 shows the calculated copy numbers of the entire T-DNA and its individual elements. For example, the copy number of the entire T-DNA was 2.28, and the copy number of the individual elements ranged from 4.14 to 2.04 in FG72, while the copy number of the T-DNA in JACK was 0.01. In FG72, the copy numbers of TPotp Y and TPotp C were around 4 due to their high sequence similarity. The copy numbers of Ph4a748 ABBC and Ph4a748 were 3.11 and 4.11, respectively, due to the partial integration of Ph4a748 ABBC and the high sequence similarity between Ph4a748 ABBC and Ph4a748. In Table 1, we also found that the non-transgenic soybean line Jack indeed shows 14 reads aligning to the 2mepsps element, specifically in the 5649–5713 bp region of the T-DNA. Upon closer examination, we discovered that these 14 reads are identical, with each read aligning to a 65 bp segment and containing 7 mismatches when compared to the 2mepsps element. This alignment is most likely the result of sequence homology between the 2mepsps region and a segment of the soybean reference genome, rather than indicating the presence of the 2mepsps element in the Jack line. Based on the calculated copy numbers, we deduced that there are two tandem repeats of the T-DNA inserted at the IS2 site.

We also used the ddPCR method to confirm the accuracy of the copy number calculation from the PE-WGS analysis, using hppdPfW336 and 2mepsps as examples. In the ddPCR analysis, the copy number is calculated using the formula: $Copy\ Number = C_{target} \times N_{ref}/C_{ref}.$, where C_{target} is the amount of the target gene, C_{ref} is the amount of the reference gene, and N_{ref} represents the copy number of the reference gene (lectin, which has a copy number of 1) in the haploid soybean genome. The ddPCR quantified the copy numbers of hppdPfW336 and 2mepsps as 1.68 and 1.69, respectively (Table 2). Compared to the PE-WGS results, the ddPCR values were slightly lower, which we speculate may be due to the tandem repeats of the T-DNA integration in FG72, as DNA molecules with tandem repeats might tend to be distributed into a single droplet. These results indicate that PE-WGS is slightly more accurate and effective for copy number analysis of inserted T-DNAs in GM crops. Nonetheless, we can conclude that two copies of the T-DNA were integrated into the soybean genome based on the combined results from PE-WGS and ddPCR.

3.4. The complete structure of insertion site 2 (IS2) was revealed through assembly analysis

To elucidate the structural organization of the T-DNA integrations (IS1 and IS2), with particular emphasis on the arrangement within IS2, we performed de novo assembly of the filtered type C reads using Megahit. This assembly yielded six contigs (designated contigs 6–11; Additional file 3) with lengths ranging from 379 bp to 3893 bp. We visualized these contigs using the Circoletto tool (Fig. 3a), where color bands represent local alignments generated by BLAST, with band width proportional to alignment length. Twisted color bands indicate reverse complementarity between sequences. The six contigs collectively encompassed nearly the entire T-DNA sequence. Notably, contig 11 contained both termini of the T-DNA (Fig. 3b), suggesting the presence of two tandem repeats in a head-to-tail orientation within IS2.

By integrating these contigs with those obtained from flanking sequence analysis (contigs 1–5), we constructed a schematic representation of the complete T-DNA integration arrangement (Fig. 3c). Our

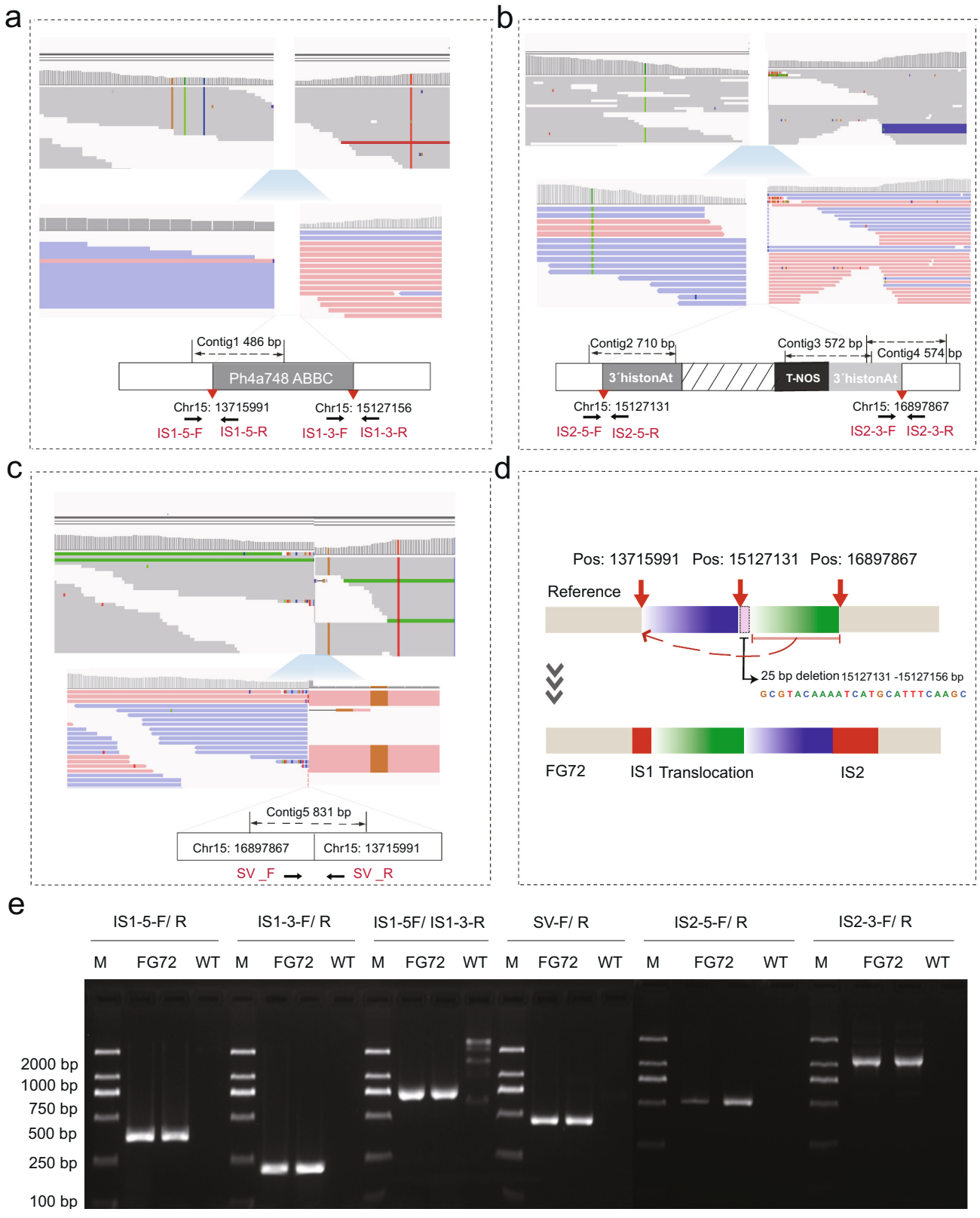

Fig. 2. Characterization of the T-DNA integration sites and flanking sequences using PE-WGS analysis and PCR **validation**. (a) Identification of the IS1 integration site through IGV visualization and de novo analysis of the filtered reads covering the IS1 region, followed by validation using designed PCR primer sets. (b) Identification of the IS2 integration site through IGV visualization and de novo analysis of the filtered reads covering the IS2 region, followed by validation using designed PCR primer sets. (c) Identification of the host genome DNA translocation through IGV visualization of the filtered reads and validation using designed PCR primer sets. (d) Schematic representation of the T-DNA insertions and the accomplished host genome DNA translocation on chromosome 15 in FG72. (e) PCR amplification results using the designed primers. Lane M: DL2000 DNA marker.

Table 1
The determined copy number of all the elements of T-DNA using PE-WGS analysis.

Element Name	Element Length (l)	FG72				JACK			
		Mapped reads Number (N)	Read length (L)	Depth	Copy number	Mapped reads Number	Read length (L)	Depth	Copy number
T-NOS	291	178	150	45.06	2.04	0	100	36.39	0
hppdPfW336	1076	650	150	45.06	2.01	0	100	36.39	0
TPotp Y	371	442	150	45.06	3.97	0	100	36.39	0
5'tev	141	120	150	45.06	2.83	0	100	36.39	0
Ph4a748 ABBC	1289	1203	150	45.06	3.11	0	100	36.39	0
Ph4a748	1014	1261	150	45.06	4.14	0	100	36.39	0
Intron1h3At	480	348	150	45.06	2.41	0	100	36.39	0
TPotp C	371	440	150	45.06	3.95	0	100	36.39	0
2mepsps	1337	827	150	45.06	2.06	14	100	36.39	0.03
T-HistonAt	686	648	150	45.06	3.14	0	100	36.39	0
T-DNA region	7296	5003	150	45.06	2.28	14	100	36.39	0.01

Table 2
The determined copy number of *hppdPfW336* and *2mepsps* gene using ddPCR.

Gene	Positive droplets			Amounts				SD	RSD (%)	Copy number
	1	2	3	1	2	3	Average			
2mepsps	12,228	11,420	11,838	2520	2540	2510	2523.33	15.28	0.61	1.69
hppdPfW336	15,814	14,033	11,706	2480	2520	2540	2513.33	30.55	1.22	1.68
Lectin	14,728	13,356	11,165	1472	1500	1520	1497.33	24.11	1.61	/

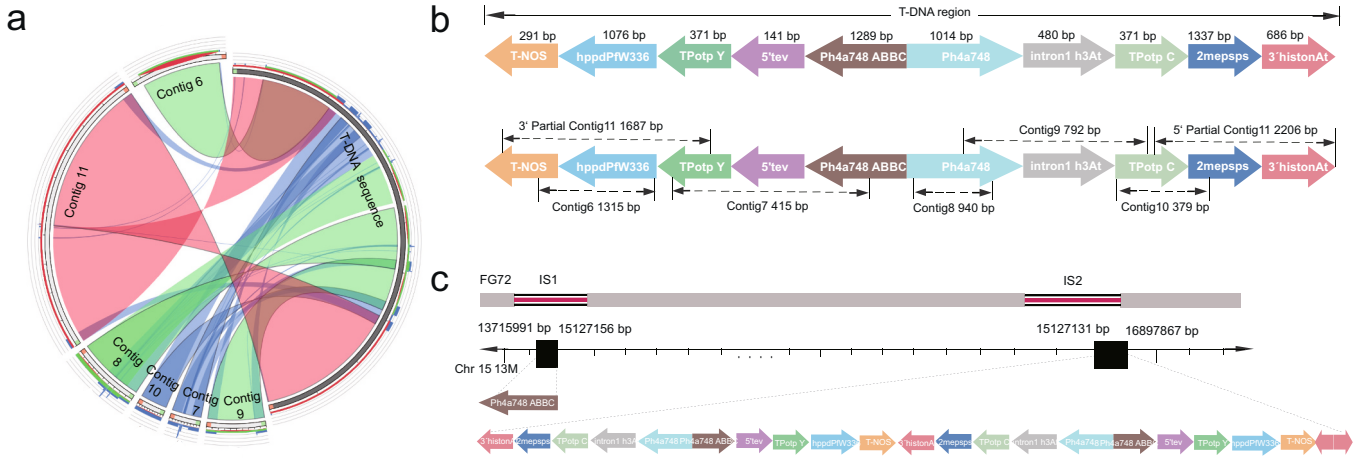

Fig. 3. Schematic of the entire transgene integration in FG72 event. (a) Visualization of assembled contigs sequence comparison results with T-DNA. (b) Schematic diagram of the linear alignment between the assembled contigs and the T-DNA sequence. (c) Schematic diagram of the two T-DNA insertions in FG72.

analysis revealed two T-DNA insertions (IS1 and IS2) on chromosome 15 of FG72. IS1 comprises a 158 bp partial Ph4a748ABBC promoter, while IS2 consists of two partial 3'histoneAt sequences in a head-to-head orientation, followed by two tandem repeats of the transformed T-DNA fragment in a head-to-tail orientation. Additionally, we observed a translocation of a genomic fragment (Chromosome 15: 15,127,156 to 16,897,866) to the 3′ end of IS1.

Despite elucidating the overall arrangement of the T-DNA integrations, obtaining a single contig containing the full DNA sequence of IS2 through de novo assembly remained challenging. This difficulty stems from limitations inherent to current short-read NGS platforms (typically generating 150–300 bp reads), which are ill-suited for assembling transgene insertions with repetitive sequences or complex structural variations. The presence of multiple similar or repetitive sequences can lead to misassemblies or collapsed assemblies, potentially obscuring the true copy number and arrangement.

To address these limitations, we designed primers based on the obtained contigs and amplified the inter-contig regions. This approach allowed us to confirm the complete sequence of IS2, including the two partial 3'histoneAt sequences in a head-to-head orientation. Our findings underscore the potential utility and necessity of long-read sequencing technologies for comprehensive analysis of complex T-DNA integrations.

3.5. *No plasmid backbone sequences were detected in FG72*

We mapped the cleaned raw data of FG72 and JACK to the sequence of the transformed plasmid pSF10 and visualized the results using IGV. The IGV results showed that the plasmid backbone, containing the *ORI ColE1* gene, the *bla* gene, and the *ORI f1* gene, was not observed in FG72 and JACK (Fig. 4), indicating that there is no plasmid backbone sequence inserted into FG72. Upon comparing the IGV view of the wild-type (WT) and FG72, we identified 14 sporadic sequences, each 100 bp in length, that partially align with the plasmid sequence. These alignments span 65–66 bp and contain seven base mismatches. These 14 reads correspond to the 5649–5713 bp region of the T-DNA. We observed that these reads map to the 2mepsps region (Fig. 4). This mapping is likely due to sequence homology between the plasmid and

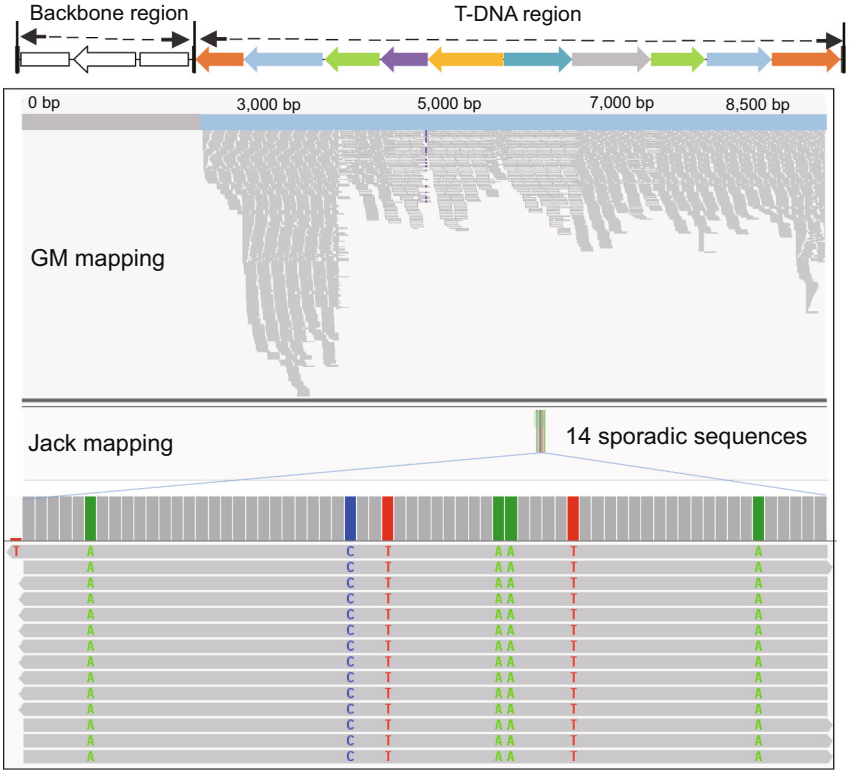

Fig. 4. IGV visualization of the sequencing data alignment to the plasmid pSF10 sequence for GM FG72 and non-GM JACK.

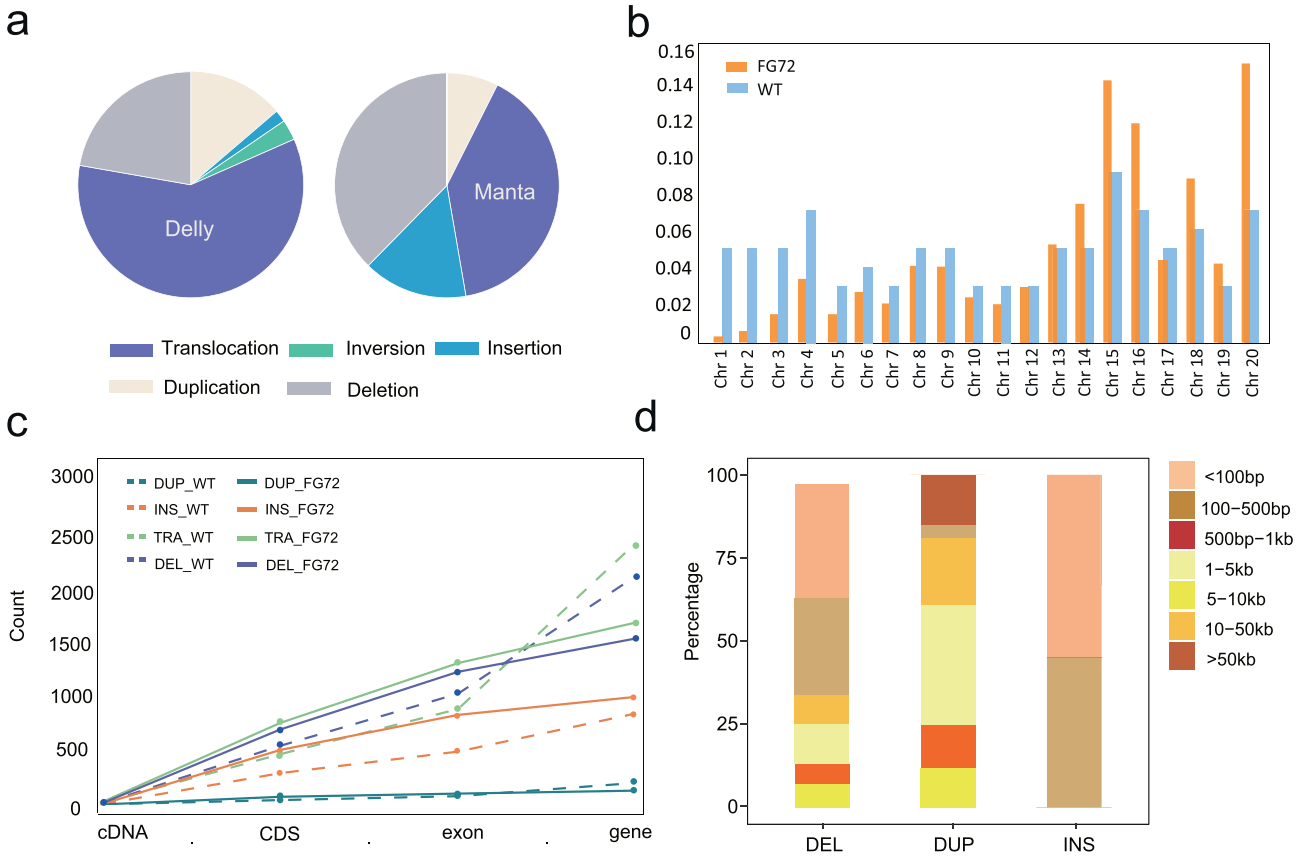

Fig. 5. SV detection. (a) The predicted SVs of FG72 using two algorithms. (b) Percentage of variations detected on each chromosome in WT and FG72. (c) Annotation of the genomic regions harboring the detected variants in WT and FG72. (d) Statistical summary of the variation sizes in FG72.

the soybean genome, rather than the presence of the transgenic element in the WT. Such sequence similarities can lead to ambiguous mapping results when using the BWA tool.

3.6. Structural variations (SVs) were observed in both FG72 and its recipient

The integration of T-DNA into the genome typically results in various structural changes, including base deletions at boundary sequences, duplications, deletions, and chromosomal translocations of DNA sequences in the recipient genomes. Structural variations (SVs) occur naturally in plant genomes and can be induced by various factors, including tissue culture processes, transgene transformation, and environmental conditions (Ruprecht, Carroll, & Persson, 2014; Yuan, Bayer, Batley, & Edwards, 2021). We investigated numerous SVs in both the transgenic line FG72 and its recipient, JACK, relative to the soybean reference genome using DELLY and Manta. As illustrated in Fig. 5, four types of SVs were widely observed in FG72 and JACK: translocation, deletion, insertion, and duplication. Among all SVs, two notable changes around the transgene insertion sites in FG72 were the translocation of a host genomic fragment (Chromosome 15: 15,127,156 to 16,897,866) at the first insertion site (IS1) and a 25 bp deletion of a host genome sequence at the second insertion site (IS2). Translocations and deletions were the two most prevalent types of SVs in FG72 (Fig. 5a). SVs were detected across all 20 chromosomes, with Chromosome 15 exhibiting the second-highest and highest number of SVs in FG72 and JACK, respectively (Fig. 5b). Annotation of the variation regions revealed that most variations occur within gene body regions in both FG72 and JACK (Fig. 5c). In FG72, we observed that the variations had approximate length ranges of less than 100 bp for deletions and insertions, and 1 kb–5 kb for duplications (Fig. 5d). These results indicate that SVs frequently occur not only in transgenic plants but also in common lines or cultivars.

4. Discussion

GM soybean FG72 was developed through biolistic transformation, which resulted in complex T-DNA integration according to data obtained from Southern blotting, conventional PCR techniques, and Sanger sequencing by its developer. In recent years, several studies have validated NGS as a powerful technique for identifying T-DNA insert sites and flanking sequences. However, few studies have provided comprehensive information on T-DNA integrations in GM crops developed by biolistic transformation, particularly for complex T-DNA integrations involving rearrangements of multiple T-DNAs and host genome translocations.

In this study, we successfully revealed the full, complex molecular characterization of the commercialized GM soybean event FG72 at the whole-genome level, deciphering its intricate T-DNA integrations through an established bioinformatics pipeline. Our analysis identified two T-DNA inserts (IS1 and IS2) in chromosome 15 of FG72. IS1 contains 158 base pairs of a partial Ph4a748ABBC promoter, while IS2 comprises two partial 3′ histone At sequences in a head-to-head orientation, followed by two tandem repeats of a transformed T-DNA fragment in a head-to-tail orientation (Fig. 2). Additionally, a genomic fragment (Chromosome 15: 15,127,156 to 16,897,866) has translocated to the 3′ end of IS1 (Fig. 3). These results further confirm the previously reported molecular characterizations by the developer.

PE-WGS coupled with our developed pipeline, demonstrated significant advantages over traditional techniques in analyzing complex T-DNA integrations, such as providing complete data, high accuracy, lower cost, and shorter experimental durations.

Our bioinformatics pipeline, modified from TranSeq, offers multiple functions for molecular characterization analysis, including identifying insert sites, flanking sequences, copy numbers of inserted DNA, plasmid backbone unintended integrations, and SVs. Currently, available bioinformatics tools (e.g., TDNAscan and T-LOC) mainly focus on identifying the T-DNA insert site and its flanking sequence and cannot extend to analyzing copy numbers, entire insertion structures, plasmid backbone unintended integrations, and SVs (Li et al., 2022; Sun et al., 2019). Compared to these tools, our pipeline shows excellent comprehensiveness in data analysis.

Using PE-WGS, we can easily obtain the two T-DNA insertion sites and their flanking sequences without additional experiments. This is challenging using conventional TAIL-PCR or genome walking methods, which often miss additional insert sites as reported previously (Wu et al., 2015; Xu et al., 2022). For copy number estimation, PE-WGS calculates the copy number of each T-DNA element with high accuracy and throughput, outperforming ddPCR and Southern blotting (Xu, Liang, Wang, & Yang, 2024). The accuracy in copy number determination can be further improved by increasing the sequencing depth. In PE-WGS, the copy number of each element can be obtained simultaneously. However, ddPCR or Southern blotting requires individually designing and testing primers and probes for each element, necessitating more experimental duration, labor, and cost (Xu et al., 2024). PE-WGS offers high accuracy not only in insert site and flanking sequence identification but also in copy number determination.

According to the developer's report on FG72, an orthogonal combination of eight probes hybridization and ten restriction enzymes digestion for Southern blotting analysis was used to sketch the basic framework and copy number of inserted T-DNA in FG72. Additionally, 12 primer sets were used to amplify, sequence, and confirm the T-DNA integrations (EFSA, 2015). The workload, experimental duration, and costs are substantial compared to PE-WGS analysis. PE-WGS is a well-established method that can be performed by one technician within 72 h. This streamlined workflow significantly reduces the overall workload and experimental duration compared to the step-by-step processes required for Southern blotting, Sanger sequencing, and PCR-based methods. Moreover, advancements in library preparation and automated sample handling have further improved efficiency. Many sample preparation steps, such as DNA extraction, fragmentation, and adapter ligation, can be automated, reducing hands-on time and minimizing potential errors introduced by manual handling. However, traditional methods such as Southern blotting, Sanger sequencing, and PCR-based techniques would require at least six months to reveal the insert site, flanking sequence, copy number, and entire sequence of T-DNA insertion. Southern blotting, in particular, is labor-intensive and time-consuming. The process of DNA extraction, restriction enzyme digestion, gel electrophoresis, and hybridization with specific probes can take several days. Additionally, analyzing and interpreting Southern blot results requires significant expertise, further adding to the workload (Glowacka et al., 2016). The high cost of specialized equipment, reagents, and labor-intensive protocols makes Southern blotting a relatively expensive approach for routine GMO characterization.

While PE-WGS shows clear advantages in GMO characterization, it remains challenging to reveal the entire arrangement and sequence of complex T-DNA integrations. In the case of IS2 in FG72, we could not obtain the entire sequence directly through de novo analysis (Fig. 3). This limitation is primarily due to its reliance on short-read sequencing, which typically generates reads ranging from 100 to 300 base pairs in length in currently used NGS platforms. In instances where T-DNA is present as multiple copies within the host genome, either as tandem or scattered insertions, PE-WGS may struggle to distinguish between these different T-DNA copies and accurately determine their organization and sequence. The assembly of short reads into a contiguous sequence can lead to gaps or misassemblies, particularly in regions with repetitive sequences, further hindering the comprehensive characterization of complex T-DNA integration sites. For example, a total of six contigs were obtained in identifying IS2 in FG72 (Fig. 3). These limitations underscore the need for alternative sequencing approaches, such as long-read technologies, to overcome the challenges in revealing the entire arrangement and sequence of complex T-DNA integrations in GMOs. Combining PE-WGS and long-read sequencing might offer a more

comprehensive, accurate, and efficient approach to uncover the detailed molecular features of T-DNA insertions, including the insert site, flanking sequences, copy number, and entire sequence, which is crucial for understanding the genetic stability and potential unintended effects of GMOs.

The observation of SVs across all 20 chromosomes in transgenic soybean is consistent with the current understanding of plant genomics. Notably, the comparable amount and type of SVs observed between the transgenic line FG72 and its non-transgenic counterpart JACK suggest that the majority of these variations are not a direct result of the transgene transformation process. This finding aligns with the work of Batista et al., who demonstrated that the extent of genomic variation in transgenic plants was often comparable to or less than that observed in plants derived from conventional breeding methods (Batista, Saibo, Lourenco, & Oliveira, 2008).

Anderson et al. further corroborated this perspective, showing that tissue culture-induced mutations were widespread in transgenic soybean lines, affecting various genomic regions (Anderson et al., 2016). While our study revealed a higher prevalence of SVs in gene bodies, potentially impacting gene function, it's crucial to interpret these results cautiously. Such variations could indeed lead to alterations in protein expression, metabolic pathways, or phenotypic traits. However, as Schnell et al. pointed out, not all SVs necessarily result in functional changes; some may be silent or have minimal impact on the plant's overall physiology (Schnell et al., 2015).

To better understand the implications of these SVs, it would be valuable to evaluate their stability and inheritance patterns across multiple generations. This approach, as emphasized by Ladics et al., is critical in assessing the genetic stability of transgenic crops as part of the safety assessment process (Ladics et al., 2015). It would provide insights into the long-term impact and potential accumulation of these variations over time.

Despite the presence of SVs in transgenic plants, it's important to note that the overall safety assessment should primarily focus on the phenotypic and compositional characteristics of the transgenic plant. Parrott et al. (2010) argued that if these characteristics fall within the range of natural variation observed in conventional varieties, the risk associated with the SVs may be considered low (Parrott et al., 2010). This perspective underscores the importance of contextualizing genetic changes within the broader spectrum of plant variability.

In conclusion, while the presence of SVs in transgenic soybean warrants careful examination, it should be evaluated within the context of natural genomic variation and compared to changes induced by conventional breeding methods. A comprehensive, science-based approach to risk assessment is crucial, considering both the genetic changes and their phenotypic manifestations. This approach ensures a balanced evaluation of transgenic crop safety, acknowledging the complexity of plant genomes and the various factors contributing to genetic diversity in both GM and non-GM crops.

CRediT authorship contribution statement

Fan Wang: Writing – original draft, Methodology, Formal analysis, Data curation. **Shengtao Lu:** Formal analysis, Data curation. **Wenting Xu:** Validation, Formal analysis. **Litao Yang:** Writing – review & editing, Investigation, Funding acquisition, Conceptualization.

Declaration of competing 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.

Acknowledgments

This work was funded by the Science and technology Innovation 2030 (2022ZD04020), National key research and development plan (2023YFF1001600) and Agricultural Research Fund of the Science and Technology Commission of Shanghai Municipality, China (21N31900100).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochms.2024.100238.

Data availability

Data will be made available on request.

References

Anderson, J. E., Michno, J. M., Kono, T. J. Y., Stec, A. O., Campbell, B. W., Curtin, S. J., & Stupar, R. M. (2016). Genomic variation and DNA repair associated with soybean transgenesis: A comparison to cultivars and mutagenized plants. *BMC Biotechnology, 16*. doi:ARTN 41.10.1186/s12896-016-0271-z.

Batista, R., Saibo, N., Lourenco, T., & Oliveira, M. M. (2008). Microarray analyses reveal that plant mutagenesis may induce more transcriptomic changes than transgene insertion. *Proceedings of the National Academy of Sciences of the United States of America, 105*(9), 3640–3645. https://doi.org/10.1073/pnas.0707881105

Börjesson, V., Martinez-Monleon, A., Fransson, S., Kogner, P., Johnsen, J. I., Milosevic, J., & López, M. D. (2022). Large-scale discovery of mouse transgenic integration sites reveals frequent structural variation and insertional mutagenesis. *BMC Genomics, 23*(1). https://doi.org/10.1186/s12864-022-08376-0

Chen, S. F., Zhou, Y. Q., Chen, Y. R., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics, 34*(17), 884–890. https://doi.org/10.1093/bioinformatics/bty560

Chen, X. H., Dong, Y., Huang, Y. L., Fan, J. M., Yang, M. S., & Zhang, J. (2021). Whole-genome resequencing using next-generation and Nanopore sequencing for molecular characterization of T-DNA integration in transgenic poplar 741. *BMC Genomics, 22*(1). https://doi.org/10.1186/s12864-021-07625-y

Chen, X. Y., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., & Saunders, C. T. (2016). Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics, 32*(8), 1220–1222. https://doi.org/10.1093/bioinformatics/btv710

Edwards, B., Hornstein, E. D., Wilson, N. J., & Sederoff, H. (2022). High-throughput detection of T-DNA insertion sites for multiple transgenes in complex genomes. *BMC Genomics, 23*(1). https://doi.org/10.1186/s12864-022-08918-6

EFSA. (2011). Guidance for risk assessment of food and feed from genetically modified plants. *EFSA Journal, 9*(5), 2150.

EFSA. (2015). Scientific opinion on an application (EFSA-GMO-BE-2011-98) for the placing on the market of herbicide-tolerant genetically modified soybean FG72 for food and feed uses, import and processing under regulation (EC) no 1829/2003 from Bayer CropScience. *EFSA Journal, 13*(17), Article 4167, 4129 pp.

Glowacka, K., Kromdijk, J., Leonelli, L., Niyogi, K. K., Clemente, T. E., & Long, S. P. (2016). An evaluation of new and established methods to determine T-DNA copy number and homozygosity in transgenic plants. *Plant, Cell & Environment, 39*(4), 908–917. https://doi.org/10.1111/pce.12693

Goodwin, L. O., Splinter, E., Davis, T. L., Urban, R., He, H., Braun, R. E., & Murray, S. A. (2019). Large-scale discovery of mouse transgenic integration sites reveals frequent structural variation and insertional mutagenesis. *Genome Research, 29*(3), 494–505. https://doi.org/10.1101/gr.233866.117

Halpin, C. (2005). Gene stacking in transgenic plants - the challenge for 21st century plant biotechnology. *Plant Biotechnology Journal, 3*(2), 141–155. https://doi.org/10.1111/j.1467-7652.2004.00113.x

Herman, R. A. (2013). Unintended compositional changes in genetically modified (GM) crops: 20 years of research. *Journal of Agricultural and Food Chemistry, 61*(48), 11695–11701. https://doi.org/10.1021/jf400135r

Holst-Jensen, A., Bertheau, Y., de Loose, M., Grohmann, L., Hamels, S., Hougs, L., & Wulff, D. (2012). Detecting un-authorized genetically modified organisms (GMOs) and derived materials. *Biotechnology Advances, 30*(6), 1318–1335. https://doi.org/10.1016/j.biotechadv.2012.01.024

James, C. (2021). *Global status of commercialized biotech/GM crops in 2021*. ISAAA Brief No. 57, International Service for the Acquisition of Agri-biotech Applications. Ithaca, NY.

Jung, Y., & Han, D. (2022). BWA-MEME: BWA-MEM emulated with a machine learning approach. *Bioinformatics, 38*(9), 2404–2413. https://doi.org/10.1093/bioinformatics/btac137

Kawall, K. (2019). New possibilities on the horizon: Genome editing makes the whole genome accessible for changes. *Frontiers in Plant Science, 10*. https://doi.org/10.3389/fpls.2019.00525

Kovalic, D., Garnaat, C., Guo, L., Yan, Y. P., Groat, J., Silvanovich, A., & Bannon, G. (2012). The use of next generation sequencing and junction sequence analysis bioinformatics to achieve molecular characterization of crops improved through modern biotechnology. *Plant Genome, 5*(3), 149–163. https://doi.org/10.3835/plantgenome2012.10.0026

Ladics, G. S., Bartholomaeus, A., Bregitzer, P., Doerrer, N. G., Gray, A., Holzhauser, T., & Glenn, K. (2015). Genetic basis and detection of unintended effects in genetically modified crop plants. *Transgenic Research, 24*(4), 587–603. https://doi.org/10.1007/s11248-015-9867-7

Li, D. H., Liu, C. M., Luo, R. B., Sadakane, K., & Lam, T. W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct graph. *Bioinformatics, 31*(10), 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., & Proc, G. P. D. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics, 25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, S. F., Wang, C. Y., You, C. J., Zhou, X. P., & Zhou, H. B. (2022). T-LOC: A comprehensive tool to localize and characterize T-DNA integration sites. *Plant Physiology, 190*(3), 1628–1639. https://doi.org/10.1093/plphys/kiac225

Park, D., Park, S. H., Ban, Y. W., Kim, Y. S., Park, K. C., Kim, N. S., & Choi, I. Y. (2017). A bioinformatics approach for identifying transgene insertion sites using whole genome sequencing data. *BMC Biotechnology, 17*. https://doi.org/10.1186/s12896-017-0386-x

Parrott, W., Chassy, B., Ligon, J., Meyer, L., Petrick, J., Zhou, J. G., & Levine, M. (2010). Application of food and feed safety assessment principles to evaluate transgenic approaches to gene modulation in crops. *Food and Chemical Toxicology, 48*(7), 1773–1790. https://doi.org/10.1016/j.fct.2010.04.017

Rausch, T., Zichner, T., Schlattl, A., Stütz, A. M., Benes, V., & Korbel, J. O. (2012). DELLY: Structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics, 28*(18), i333–i339. https://doi.org/10.1093/bioinformatics/bts378

Robinson, J. T., Thorvaldsdottir, H., Turner, D., Mesirov, J. P., & Alkan, C. (2023). Igv.Js: An embeddable JavaScript implementation of the integrative genomics viewer (IGV). *Bioinformatics, 39*(1). https://doi.org/10.1093/bioinformatics/btac830

Ruprecht, C., Carroll, A., & Persson, S. (2014). T-DNA-induced chromosomal translocations in feronia and anxur2 mutants reveal implications for the mechanism of collapsed pollen due to chromosomal rearrangements [J]. *Mol Plant, 7*(10), 1591–1594. https://doi.org/10.1093/mp/ssu062

Schnell, J., Steele, M., Bean, J., Neuspiel, M., Girard, C., Dormann, N., & Macdonald, P. (2015). A comparative analysis of insertional effects in genetically engineered plants: Considerations for pre-market assessments. *Transgenic Research, 24*(1), 1–17. https://doi.org/10.1007/s11248-014-9843-7

Siddique, K., Wei, J. J., Li, R., Zhang, D. B., & Shi, J. X. (2019). Identification of T-DNA insertion site and flanking sequence of a genetically modified maize event IE09S034 using next-generation sequencing technology. *Molecular Biotechnology, 61*(9), 694–702. https://doi.org/10.1007/s12033-019-00196-0

Srivastava, A., Philip, V. M., Greenstein, I., Rowe, L. B., Barter, M., Lutz, C., & Reinholdt, L. G. (2014). Discovery of transgene insertion sites by high throughput sequencing of mate pair libraries. *BMC Genomics, 15*(1). https://doi.org/10.1186/1471-2164-15-367

Sun, L., Ge, Y. B., Sparks, J. A., Robinson, Z. T., Cheng, X. F., Wen, J. Q., & Blancaflor, E. B. (2019). TDNAscan: A software to identify complete and truncated T-DNA insertions. *Frontiers in Genetics, 10*. https://doi.org/10.3389/fgene.2019.00685

Wu, L., Di, D.-W., Zhang, D., Song, B., Luo, P., & Guo, G.-Q. (2015). Frequent problems and their resolutions by using thermal asymmetric interlaced PCR (TAIL-PCR) to clone genes in Arabidopsis T-DNA tagged mutants. *Biotechnology & Biotechnological Equipment, 29*(2), 260–267. https://doi.org/10.1080/13102818.2014.998161

Xu, J. M., Zhu, J. S., Li, M. Z., Hu, H., & Mao, C. Z. (2022). Progress on methods for acquiring flanking genomic sequence. *Yi Chuan, 44*(4), 313–321. https://doi.org/10.16288/j.yczz.21-415

Xu, W., Liang, J., Wang, F., & Yang, L. (2024). Comparative evaluation of gene copy number estimation techniques in genetically modified crops: Insights from southern blotting, qPCR, dPCR and NGS. *Plant Biotechnology Journal*. https://doi.org/10.1111/pbi.14466

Xu, W. T., Zhang, H. W., Zhang, Y. C., Shen, P., Li, X., Li, R., & Yang, L. T. (2022). A paired-end whole-genome sequencing approach enables comprehensive characterization of transgene integration in rice. *Communications Biology, 5*(1). https://doi.org/10.1038/s42003-022-03608-1

Yang, L. T., Wang, C. M., Holst-Jensen, A., Morisset, D., Lin, Y. J., & Zhang, D. B. (2013). Characterization of GM events by insert knowledge adapted re-sequencing approaches. *Scientific Reports, 3*. https://doi.org/10.1038/srep02839

Yuan, Y. X., Bayer, P. E., Batley, J., & Edwards, D. (2021). Current status of structural variation studies in plants [J]. *Plant Biotechnol J, 19*(11), 2153–2163. https://doi.org/10.1111/pbi.13646

Zastrow-Hayes, G. M., Lin, H. N., Sigmund, A. L., Hoffman, J. L., Alarcon, C. M., Hayes, K. R., & Beatty, M. K. (2015). Southern-by-sequencing: A robust screening approach for molecular characterization of genetically modified crops. *Plant. Genome, 8*(1). https://doi.org/10.3835/plantgenome2014.08.0037

Zhang, H. W., Zhang, Y. C., Xu, W. T., Li, R., Zhang, D. B., & Yang, L. T. (2022). Development and performance evaluation of whole-genome sequencing with paired-end and mate-pair strategies in molecular characterization of GM crops: One GM rice 114-7-2 line as an example. *Food Chemistry: Molecular Sciences, 4*. https://doi.org/10.1016/j.fochms.2021.100061

Zhang, Y. C., Zhang, H. W., Qu, Z., Zhang, X. J., Cui, J. J., Wang, C. H., & Yang, L. T. (2020). Comprehensive analysis of the molecular characterization of GM rice G6H1 using a paired-end sequencing approach. *Food Chemistry, 309*. https://doi.org/10.1016/j.foodchem.2019.125760