Fruit Research

ARTICLE

Open Access

https://doi.org/10.48130/frures-0026-0030

Fruit Research **2026**, 6: e037

Genomic identification of banana hybrid progeny using PacBio HiFi sequencing

Caihong Jia[1#], Junting Feng[1#], Zhiwei Jia[2#], Jingyi Wang[1], Mengling Zhu[3], Jiawei Zhao[4], Zhuoyu Zhang[4], Huigang Hu[2*] and Zhuo Wang[1*]

[1] *National Key Laboratory of Tropical Crop Breeding, Institute of Tropical Bioscience and Biotechnology & Sanya Research Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China*

[2] *South Subtropical Crops Research Institute, Chinese Academy of Tropical Agricultural Sciences, Zhanjiang 524091, China*

[3] *National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China*

[4] *College of Tropical Crops, Yunnan Agricultural University, Puer 665099, China*

[#] Authors contributed equally: Caihong Jia, Junting Feng, Zhiwei Jia

[*] Correspondence: huhuigang@sina.com (Hu H); wangzhuocd@126.com (Wang Z)

Abstract

The accurate identification of hybrid progeny is essential for improving the efficiency of banana (*Musa* spp.) crossbreeding, yet traditional morphological methods are time-consuming, unstable, and ineffective for distinguishing hybrids at early developmental stages. In this study, the highly fertile and stress-tolerant landrace 'Dongguan Dajiao' (*Musa* × *paradisiaca*, DGDJ) was used as the maternal parent and crossed with two wild diploid species, *Musa acuminata* (XGYJ, AA) and *Musa balbisiana* (HNYJ, BB), as the paternal parents, producing the progeny lines X1701, X1702, X1931, and X1932. PacBio HiFi sequencing was performed using the Revio platform, followed by k-mer analysis, genome assembly, and parental read coverage analysis to systematically evaluate the genome size, heterozygosity, and genomic composition in both parents and the progeny. Flow cytometry confirmed that the maternal parent DGDJ is triploid (3*x*), the paternal parents XGYJ and HNYJ are diploid (2*x*), and all four progeny lines are tetraploid (4*x*). The results revealed that DGDJ possesses a complex genomic constitution comprising the A, B, and S genomes with a heterozygosity level of 5.77%. All progeny were identified as tetraploids, with paternal contributions ranging from approximately 14% to 17% and maternal inheritance exceeding 62%, confirming their status as true hybrids. Collectively, this study establishes a k-mer-based method for identifying banana hybrid progeny using HiFi sequencing data, offering critical technical support for germplasm innovation and crossbreeding in banana.

Keywords: Banana hybrid progeny, K-mer analysis, PacBio HiFi sequencing

Citation: Jia C, Feng J, Jia Z, Wang J, Zhu M, et al. 2026. Genomic identification of banana hybrid progeny using PacBio HiFi sequencing. *Fruit Research* 6: e037 https://doi.org/10.48130/frures-0026-0030

Introduction

Banana (*Musa* spp.) belongs to the Musaceae family and the *Musa* genus. It is a widely distributed monocotyledonous plant that originated in the Malay Archipelago or New Guinea in Southeast Asia and is mainly cultivated in tropical and subtropical regions[1,2]. As a large evergreen perennial herbaceous plant, banana is an important fruit widely consumed throughout the world that is a key food staple in many countries and regions[3,4]. Most contemporary banana cultivars primarily derive from two wild species: *M. acuminata* Colla (A genome) and *M. balbisiana* Colla (B genome) and hybridization between them. This interspecific crossbreeding has generated diverse genomic constitutions, including diploids (AA, BB, AB), triploids (AAA, AAB, ABB), and tetraploids (ABBB, AAAA)[5].

Banana production has long been threatened by diseases and further hindered by a lack of genetic diversity. Sexual hybridization is a key breeding method for improving banana, which allows the integration of favorable traits from both parents. Depending on the genetic distance between the parent lines, sexual hybridization is mainly classified as either interspecific and distant hybridization. However, breeding banana via hybridization is particularly challenging because of several biological constraints: The prevalence of polyploidy, parthenocarpy (seedless fruit development), low pollen fertility, asynchronous flowering times among different cultivars, and the clonal propagation of offspring.

Musa × *paradisiaca* L., an indigenous cultivar used for fresh consumption and widely distributed across South China, is characterized by its distinctive rhomboid-shaped fruit cross-section. This landrace exhibits a rustic growth habit (i.e., low input requirements), along with notable cold tolerance and resistance to *Fusarium* wilt caused by *Fusarium oxysporum* f. sp. *cubense*, reflecting strong overall stress resilience. The fruit pulp is soft in texture and contains relatively high levels of soluble sugars and titratable acids, contributing to a unique sweet–sour flavor. However, its fruit quality remains significantly inferior to that of major commercial cultivars such as 'Cavendish' (AAA genome) and 'Pisang Awak' (ABB genome). In general, *M.* × *paradisiaca* ('Dongguan Dajiao', DGDJ) is distinguished by its excellent stress tolerance but inferior flavor profile. DGDJ is a triploid (3*x*) landrace. More importantly, *M.* × *paradisiaca* is capable of natural pollination, a trait rarely observed in cultivated bananas[6]. Artificial pollination can greatly enhance its fruit set rate, making it a highly effective maternal parent in interspecific hybridization breeding programs aimed at developing stress-tolerant banana cultivars.

At present, the identification of banana hybrids and their progeny primarily relies on morphological observations[7]. However, most morphological markers are controlled by multiple genes and are highly susceptible to environmental influences, making them unstable and unreliable for accurate determination[8]. In addition, bananas are large and space-consuming plants, and morphological

https://doi.org/10.48130/frures-0026-0030

assessments are both time-consuming and labor-intensive[9]. Moreover, banana seedlings exhibit limited morphological differentiation at early stages, making it difficult to distinguish between progeny and parental lines, which severely constrains breeding efficiency[10]. DNA-based identification methods for banana hybrids remain limited, with recent reports only on *copia*-intra-retrotransposon amplified polymorphism (IRAP) markers[11], internal transcribed spacers (ITS) sequencing[12], fluorescence in situ hybridization (FISH)[13], and random amplified polymorphic DNA (RAPD) [14].

So far, the genomes of several wild banana genotypes, such as *M. acuminata*[1], *M. balbisiana*[15], and *M. troglodytarum*[16] have been successfully assembled. These datasets provide the necessary resources for a comprehensive genetic background analysis of *M.* × *paradisiaca*. In this study, we performed PacBio HiFi long-read sequencing on the parental lines and their hybrid progeny to systematically characterize their genome size, heterozygosity, and genomic composition. Our main objectives were to verify the hybrid status of the progeny, determine their ploidy levels, and elucidate the parental contributions to their genomes. This approach provides a precise and efficient method for identifying banana hybrids and supports germplasm innovation in banana breeding.

Materials and methods

Plant material collection, sequencing, and flow cytometry

The parental accessions used in this study were obtained from the banana germplasm repository of the South Subtropical Crops Research Institute, Chinese Academy of Tropical Agricultural Sciences (CATAS). XGYJ (*M. acuminata* subsp. *malaccensis*) was collected from Xishuangbanna, Yunnan Province, and HNYJ (*M. balbisiana*) was collected from Hainan Province. DGDJ is a local landrace widely cultivated in South China. To determine the genomic genotypes of the parental lines, we aligned the HiFi sequencing reads of XGYJ and HNYJ to a composite reference genome consisting of the A genome of *M. acuminata* subsp. *malaccensis*[1] and the B genome of *M. balbisiana*[15]. According to the coverage ratios, XGYJ and HNYJ were confirmed as AA and BB genotypes, respectively, which is consistent with their morphological classification, following the diagnostic traits of Simmonds and Shepherd[5].

Clean, uncontaminated leaf samples were collected from all parental lines and hybrid progeny. For flow cytometry analysis, approximately 0.5 cm^2 of fresh leaf tissue was chopped in a nucleic extraction buffer, stained with propidium iodide, and analyzed using a CytoFLEX flow cytometer (Beckman Coulter). The ploidy level of each sample was determined by comparing the fluorescence intensity with that of diploid control samples. Genomic DNA was extracted using the cetyltrimethylammonium bromide (CTAB) method[17], followed by library construction and PacBio HiFi sequencing on the Revio platform (PacBio).

Genome size and heterozygosity estimation

HiFi sequencing data were analyzed using Jellyfish (v2.3.0) to extract the 19-mer frequency spectra. GenomeScope2 (v2.0)[18] was subsequently applied to estimate genome size and heterozygosity. Briefly, GenomeScope2 models the k-mer frequency spectrum as a mixture of heterozygous and homozygous k-mers, from which it infers the genome size, heterozygosity rate, and repeat content. The ploidy parameter (-p) was set according to the ploidy levels determined by flow cytometry as described in the section 'Plant material collection, sequencing, and flow cytometry': p = 2 for diploid samples (XGYJ and HNYJ), p = 3 for the triploid maternal parent (DGDJ), and p = 4 for all hybrid progeny. Other parameters were kept as the default. GenomeScope2 provides an initial estimation of genome size and heterozygosity, whereas the actual ploidy levels were confirmed by flow cytometry, as described in the section 'Plant material collection, sequencing, and flow cytometry'.

Genomic assessment of the genotypes

Reference genomes for the banana A, B, S, and T genomes were obtained from the Banana Genome Hub (https://banana-genome-hub.southgreen.fr), corresponding to *M. acuminata* (DH-Pahang_v4, A genome), *M. balbisiana* (DH-PKW, B genome), *M. schizocarpa* (v1.0, S genome), and *M. troglodytarum* (Fe'i, T genome), respectively. The A, B, S, and T genome sequences were concatenated to construct a composite reference genome ($RG^{A+B+S+T}$). HiFi sequencing reads from the parental lines and progeny were aligned to this composite reference using minimap2 (v2.24)[19] with the parameter '--secondary=no'. The number of mapped reads and total bases aligned to each subgenome (A, B, S, and T) were calculated to determine the genomic contribution ratios. The proportion of each subgenome (A, B, S, T) in a sample was calculated as the number of bases mapped to that subgenome divided by the total number of mapped bases, expressed as a percentages, as reported in Table 1.

In addition, the A and B genomes were merged to generate another composite reference genome (RG^{A+B}). HiFi sequencing reads from the parental lines and hybrid progeny were then aligned to RG^{A+B} using minimap2 (v2.24) with the parameter '--secondary=no'. These alignments facilitated the calculation and assessment of genomic genotypes for both parental and hybrid progeny samples.

Validation of true hybrid progeny

The analysis was performed following the kmerGWAS method described by Voichek and Weigel[20]. The 31-mer counts were extracted from the HiFi sequencing data using KMC (v3.2.3)[21]. All k-mers from the individual samples were merged and filtered with the parameters '--mac 1 -p 0' to retain informative k-mers. For heritability analysis, 100 million k-mers were randomly subsampled from each hybrid combination. Principal component analysis (PCA) of the hybrid groups was performed using PLINK (v1.90b6.9)[22].

Genome assembly

Genome assemblies were generated using Hifiasm (v0.24.0)[23], with the -l0 parameter. The --n-hap parameter was set to 2, 3, and 4 according to the ploidy of each sample (diploid XGYJ and HNYJ, triploid DGDJ, and tetraploid progeny). Primary and alternative

Table 1. Genotypic proportions of each genome (%).

Sample	A (%)	B (%)	S (%)	T (%)
DGDJ	38.39	43.23	14.84	3.54
HNYJ	2.49	90.76	3.04	3.71
X1701	51.31	32.61	13.01	3.07
X1702	52.04	32.53	12.86	2.58
X1931	29.06	56.54	11.63	2.76
X1932	28.78	56.62	11.71	2.88
XGYJ	78.44	3.19	15.24	3.14

Values represent the percentage of bases mapped to each subgenome (A, B, S, T) relative to the total mapped bases for each sample. The T genome contributed less than 4% in all samples.

Table 2. Genome statistics of parent and hybrids.

Genome	num_seqs	sum_len	min_len	avg_len	max_len
DGDJ	14,553	1,947,668,911	7,441	133,832.80	43,112,517
HNYJ	3,863	851,980,049	9,267	220,548.80	27,489,044
XGYJ	8,421	1,117,300,856	8,451	132,680.30	23,313,597
X1701	2,536	1,964,185,925	11,558	774,521.30	41,161,095
X1702	2,428	1,964,474,777	9,408	809,091.80	36,305,017
X1931	5,951	2,005,365,046	12,654	336,979.50	47,283,826
X1932	4,711	1,982,262,728	11,611	420,773.20	25,176,664

contigs were merged to create genome assemblies for each sample. The genome size of each assembly is provided in Table 2.

Results

Genome sequencing data statistics and genome size estimation

This study focused on two hybrid combinations: XGYJ × DGDJ and HNYJ × DGDJ. Sequencing of the parental lines (XGYJ, HNYJ, and DGDJ) and their hybrid progeny produced a total of 231 Gb of HiFi data, with an average coverage depth of 66× and an average haplotype coverage of 20× (Supplementary Table S1).

Genome assessment of the parental lines indicated that XGYJ, HNYJ, and DGDJ had genome sizes of 412.1, 427.6, and 484.8 Mb, respectively, with heterozygosity levels of 0.597%, 0.594%, and 5.77% (Fig. 1, Supplementary Table S2). Among these parental lines, DGDJ exhibited the highest level of genomic heterozygosity, which may be attributed to its hybrid nature incorporating both A and B genomes.

X1701 and X1702, two progeny derived from the DGDJ cross, exhibited genome sizes of 422.8 and 416.6 Mb, with heterozygosity levels of 6.34% and 6.42%, respectively. Similarly, X1931 and X1932, offspring from the DGDJ cross, displayed genome sizes of 420.1 and 431.6 Mb and heterozygosity levels of 5.77% and 5.91% (Fig. 1, Supplementary Table S2), respectively. Genome size estimations performed using GenomeScope2 yielded haploid genome sizes for both parental and progeny samples consistently exceeding 400 Mb (Fig. 1). It should be noted that GenomeScope2 estimates the haploid (monoploid) genome size, which corresponds to one complete set of chromosomes. For tetraploid progeny, the total genome size is approximately four times the haploid value (~1.6–2.0 Gb), consistent with the assembly results presented in Table 2.

Flow cytometry analysis was performed to confirm the ploidy levels. The relative fluorescence intensity of DGDJ was approximately 10,000 (triploid, 3*x*), whereas XGYJ and HNYJ showed intensities of around 6,700 (diploid, 2*x*). All four hybrid progeny (X1701, X1702, X1931, X1932) exhibited fluorescence intensities of approximately 13,000, confirming their tetraploid (4*x*) status (Supplementary Table S3, Supplementary Fig. S1). These results are consistent with the 3*x* × 2*x* hybridization strategy used.

Genotype of parental lines and hybrid progeny

Hybridization leads to significant changes in the offspring's genome. Theoretically, when the A genome is used as the paternal parent, the offspring genotype would be AABB; when the B genome is used as the paternal parent, the offspring genotype would be

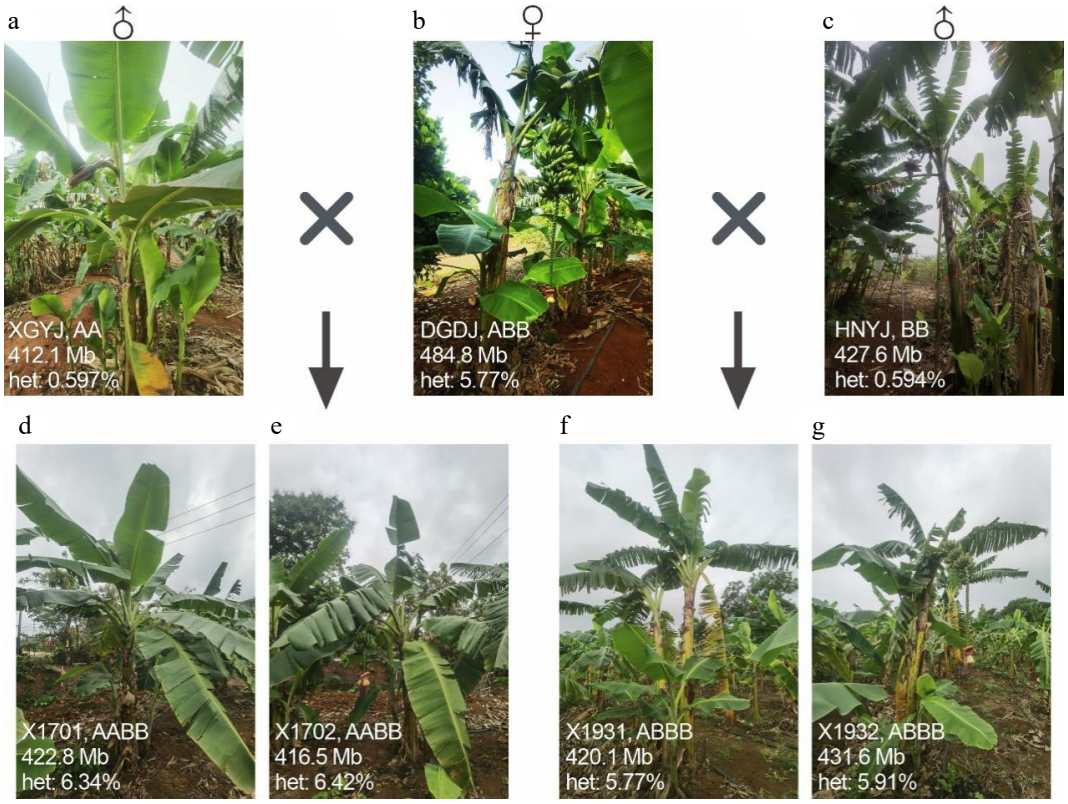

Fig. 1 Assessment genome size and heterozygosity of the parental lines and hybrid progeny. Panels (a)–(c) show the parental lines of the two hybrid combinations, whereas (d)–(g) show the hybrid progeny. X1701 and X1702 were derived from the cross DGDJ × XGYJ; X1931 and X1932 were derived from the cross DGDJ × HNYJ. The genome size and heterozygosity level shown in each panel were estimated using GenomeScope2. The ploidy parameter was set to p = 2 for XGYJ and HNYJ, p = 3 for DGDJ, and p = 4 for all hybrid progeny. The estimated genome size represents the haploid genome size.

https://doi.org/10.48130/frures-0026-0030

ABBB. To investigate the genotypic characteristics of the parental lines and hybrid progeny, this study aligned HiFi sequencing data to a combined reference genome that includes the published assemblies of the A, B, S, and T genomes. Genomic composition analysis revealed that XGYJ predominantly consists of A (78.44%) and S (15.24%) genomes (Table 1), suggesting that although XGYJ is morphologically classified as a diploid *M. acuminata* (AA), it contains a substantial proportion of S genome sequences, indicating historical introgression from *M. schizocarpa*. In contrast, HNYJ exhibited 90.76% B genome representation (Table 1), strongly supporting its characteristic diploid BB genomic configuration. Overall, the T genome did not participate in the genome of either hybrid combination (Table 1), indicating that it was not involved in the hybridization process of the hybrid banana. Genomic profiling of DGDJ revealed a tripartite composition with contributions from the A (38.39%), B (43.23%), and S (14.84%) genomes, suggesting the complex hybrid origin of DGDJ. Comparative analysis of hybrid progeny demonstrated paternal-specific dosage effects: When XGYJ was the paternal parent, the offspring X1701 and X1702 exhibited increased A-genome dosage (51.31% and 52.04%, respectively), whereas paternal HNYJ led to increased B-genome dosage in X1931 and X1932 (56.54% and 56.62%, respectively) (Table 1). These findings clearly indicate that the choice of hybrid combination significantly influences the distribution of genomic dosage in banana hybrids. The offspring X1701 and X1702 with the A-genome as the paternal parent exhibited a B-genome content of 32.61%, whereas the offspring X1931 and X1932 with B-genome as the paternal parent had a B-genome content of 56.54%. In contrast, the B genome content in DGDJ was 43.23%. The genomic content of the offspring is closely related to the genotype of the paternal parent.

To align with the Simmonds classification system[5], we independently aligned the parental lines and hybrid progeny to a merged reference genome of the A and B genomes. HNYJ exhibited an A:B ratio of 0.042, confirming its canonical BB genome; XGYJ showed an A:B ratio of 12.9 (Fig. 2), characteristic of a typical AA genome; and DGDJ displayed a balanced A:B ratio of 1.017 (Fig. 2), indicating an AB genome dosage close to equilibrium, with a $(AB)_n$ genotype pattern. The hybrid progeny X1701 and X1702 (XGYJ × DGDJ) exhibited $(A_{1.63–1.67}B)_n$ genotypes; X1931 and X1932 (HNYJ × DGDJ) displayed $(AB_{1.63–1.65})_n$ genotypes (Fig. 2), where '*n*' indicates the ploidy level (tetraploid, $n = 2$) in the context of the Simmonds classification. Genomic dosage analysis revealed that the paternal parents XGYJ and HNYJ contributed single sets of A and B genomes, respectively, although DGDJ contributed consistently across both hybrid combinations.

Genome assembly and verification of the hybrid progeny as true hybrids of parental combinations

Genome assembly using Hifiasm software revealed distinct genome sizes for the parental lines and hybrid progeny. The diploid genomes of XGYJ and HNYJ measured 1,117.3 and 852.0 Mb, respectively, but the triploid DGDJ's genome size was 1.95 Gb (Table 2). The hybrid progeny genomes ranged from 1.96 to 2.0 Gb. Except for the slightly larger genome of DGDJ (Table 2), all other genome sizes were consistent with the expected ploidy level. According to the HiFi sequencing data, k-mers ($k = 31$) were extracted and filtered to retain only those occurring at least twice. The merged k-mer datasets from both parental lines and hybrid progeny generated 151.0 million and 127.8 million k-mers for the XGYJ × DGDJ and HNYJ × DGDJ hybrid combinations, respectively. A random subsample of 10 million k-mers per cross was subjected to PCA. In the XGYJ × DGDJ cross, Principal Component (PC) 1 and PC2 explained 68.8% and 21.9% of the total variance, respectively, accounting for a cumulative 90.7% of the variation (Fig. 3). The progeny X1701 and X1702 were positioned near the perpendicular bisector between DGDJ and XGYJ, confirming their identity as true offspring of the XGYJ × DGDJ cross. Notably, PC1 alone explained 68.8% of the total variance, with X1701, X1702, and DGDJ clustering along the left boundary of PC1, indicating closer genetic affinity to the maternal parent DGDJ (Fig. 3). For the HNYJ × DGDJ cross, PC1 and PC2 explained 61.7% and 24.0% of the variance, respectively, with a cumulative explanatory power of 85.7% (Fig. 3). The progeny X1931 and X1932 were located near the perpendicular bisector of the HNYJ–DGDJ lineage, confirming their hybrid origin. Similarly, both the progeny and DGDJ clustered on the left side of PC1, further indicating a closer genetic

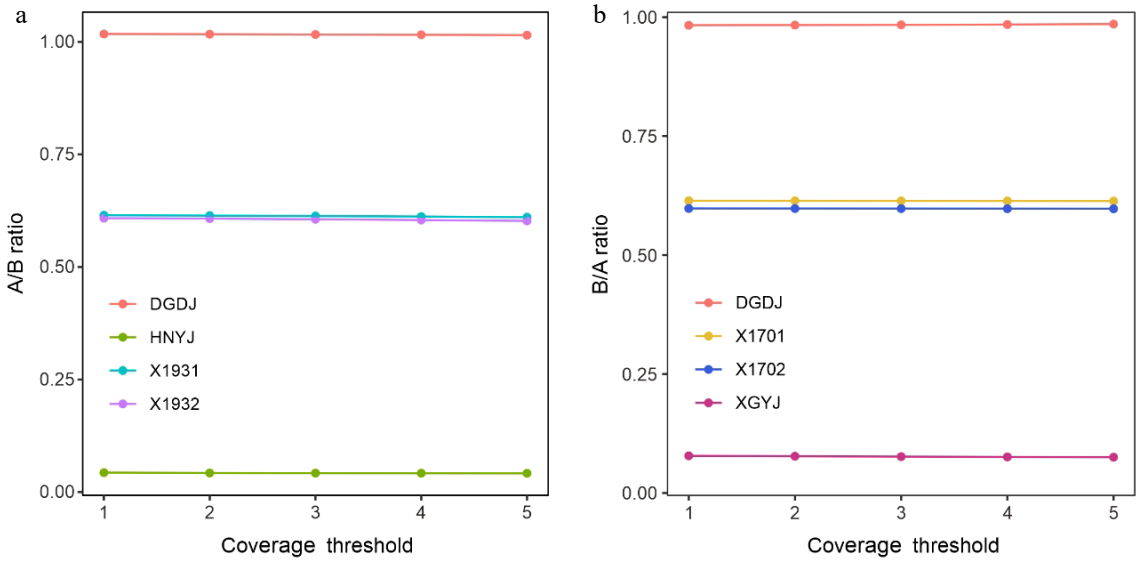

Fig. 2 Genotypic evaluation of parental lines and hybrid progeny. HiFi sequencing reads were aligned to the published reference (A + B) genomes, and the ratio of bases mapped to the A and B genomes was calculated for each parent and hybrid offspring. Panel (a) represents the hybrid combination between DGDJ and HNYJ, whereas (b) represents the hybrid combination between DGDJ and XGYJ.

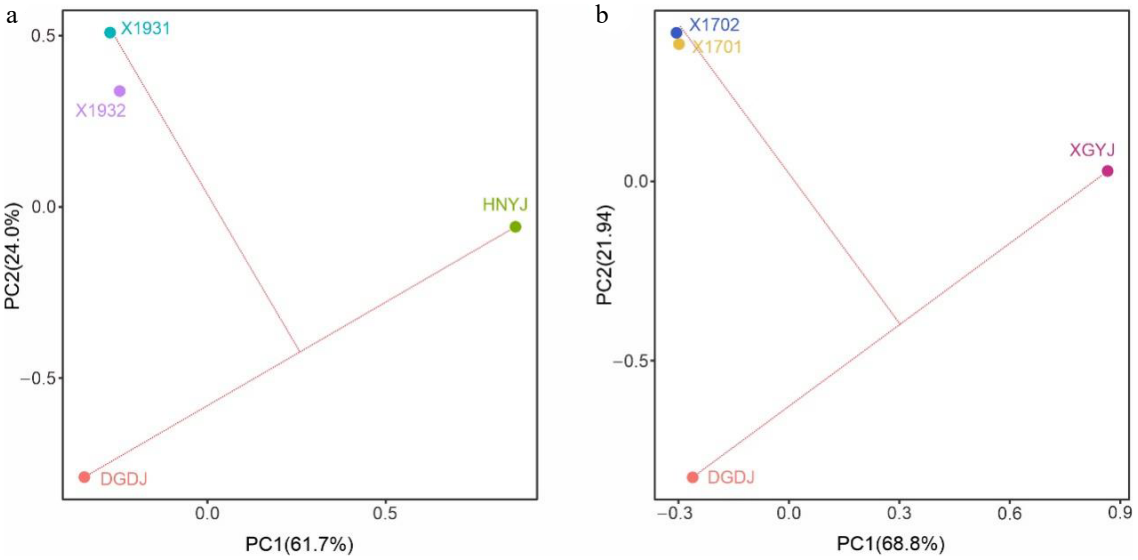

Fig. 3 Principal component analysis of k-mers from the parental lines and hybrid progeny. (a) PCA plot for the HNYJ × DGDJ cross. PC1 and PC2 explain 61.7% and 24.0% of the variance, respectively. The hybrid progeny X1931 and X1932 also cluster between the maternal and paternal parents. (b) PCA plot for the XGYJ × DGDJ cross. PC1 and PC2 explain 68.8% and 21.9% of the total variance, respectively. The hybrid progeny X1701 and X1702 are located near the perpendicular bisector between the two parents, confirming their status as true hybrids. In both crosses, the progeny and the maternal parent DGDJ cluster on the left side of PC1, indicating closer genetic affinity to the maternal parent.

relationship with DGDJ (Fig. 3). These results collectively demonstrate that X1701, X1702, X1931, and X1932 are true hybrid progeny derived from their respective crosses, excluding the possibility of self-pollination or admixture.

Genetic inheritance of parental genomes in progeny

According to the k-mer distribution patterns of the parental lines and hybrid progeny, this study quantified the proportion of the parental genomes' inheritance in different hybrid combinations. In the progeny X1701 and X1702, 14.3% and 14.4% of k-mers, respectively, were inherited from the paternal parent XGYJ, and 62.6% and 64.2% originated from the maternal parent DGDJ. Additionally, 22.1% and 21.5% of k-mers were shared between both parents (Fig. 4). Genomic profiling detected 36.5 million k-mers in XGYJ, of which 85% were inherited by X1701. In DGDJ, 79.3 million k-mers were identified, with 92.3% transmitted to X1701 (Supplementary Table S4). X1702 exhibited a similar pattern, with 83.3% of paternal k-mers and 91.6% of maternal k-mers being inherited (Supplementary Table S4). These findings indicate that the vast majority of parental k-mers were inherited by the progeny, further confirming that X1701 and X1702 are true hybrid offspring resulting from their respective crosses.

In progeny X1931 and X1932, 17.1% and 16.6% of the k-mers, respectively, were inherited from the paternal parent HNYJ, whereas 48.3% and 46.9% originated from the maternal parent DGDJ. Additionally, 34.6% and 36.5% of k-mers were shared between both parents (Fig. 4). Genomic analysis identified 54.5 million k-mers in HNYJ, of which 83.7% were inherited by X1931. In DGDJ, 82.4 million k-mers were detected, with 88.7% transmitted to X1931 (Supplementary Table S4). X1932 exhibited similar inheritance patterns, with 87.5% of paternal k-mers and 90.9% of maternal k-mers inherited (Supplementary Table S4). These findings indicate that the vast majority of parental k-mers were transmitted to the progeny, further confirming that X1931 and X1932 are authentic hybrid offspring of their respective crosses.

Discussion

Hybridization serves as a pivotal mechanism for generating genetic variation in crops, forming the foundation for cultivar development, population construction, and quantitative trait locus (QTL) mapping of key agronomic traits. In bananas, hybridization is a crucial breeding strategy. Currently, it remains the most promising conventional approach for improving banana. The history of breeding banana hybrids dates back to the 1920s, initially aimed at addressing the susceptibility of the 'Gros Michel' cultivar (AAA genome) to *Fusarium oxysporum* f. sp. *cubense* Race 1 (*Foc* TR1)[24]. Compared with vegetative selection and artificial mutagenesis, hybridization offers distinct advantages for breeding bananas[25].

Previous studies have demonstrated the introgression of the S genome (*M. schizocarpa*) into cultivated banana[26,27]. In this study, an S genome content exceeding 14% was detected in both XGYJ and DGDJ, indicating introgression of the S genome. Although XGYJ is morphologically classified as a diploid *M. acuminata* (AA), its substantial S genome proportion (15.24%) suggests historical introgression from *M. schizocarpa* (Table 1). Variations in S genome dosage among the progeny were attributable to differences in the paternal contributions (Table 1). Similarly, the S genome's introgression led to near 1:1 genomic dosage ratios between the A and B genomes in DGDJ (Fig. 2). Therefore, we propose that the genotype of DGDJ cannot be simplistically designated as 'ABB'. This genomic complexity likely underlies historical challenges in phenotypic classification of *M.* × *paradisiaca* using the conventional A/B genome framework.

Certain cultivated banana types, such as plantain, Pisang Awak, and *M.* × *paradisiaca*, exhibit partial fertility and are frequently used in distant hybridization with wild *Musa* resources (A, B, S, and T genomes). Natural hybridization occurs when *M.* × *paradisiaca* coexists with wild diploid bananas, resulting in hybrid progeny[6]. Our preliminary artificial pollination experiments using DGDJ as the maternal parent yielded two or three seeds per fruit finger (personal observation). Identifying hybrid progeny is a critical component of

https://doi.org/10.48130/frures-0026-0030

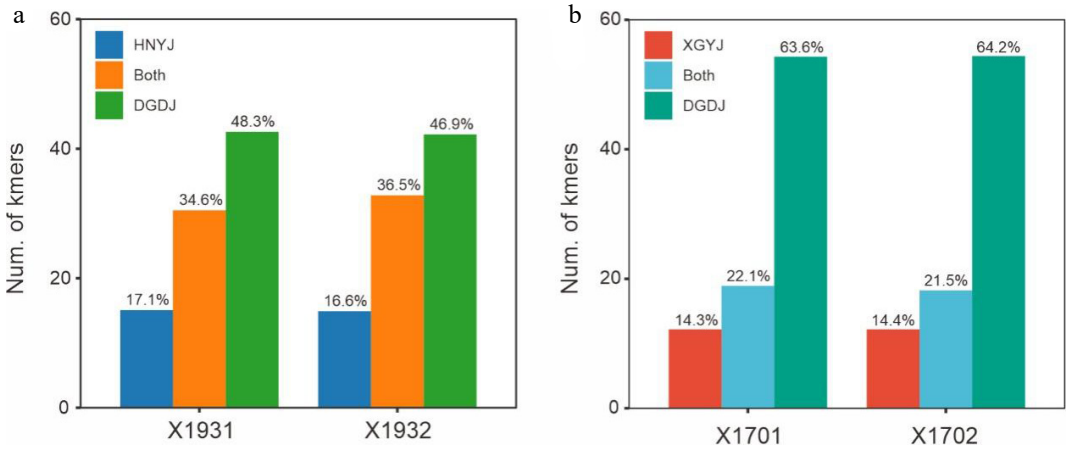

Fig. 4 Analysis of genetic inheritance in hybrid progeny. (a) Proportional composition of k-mers in X1931 and X1932, classified as paternal-specific (derived from HNYJ, steel blue), maternal-specific (derived from DGDJ, green), or shared (orange). (b) Proportional composition of k-mers in X1701 and X1702. The k-mers are classified as paternal-specific (derived from XGYJ, red), maternal-specific (derived from DGDJ, green), or shared (cyan). In all progeny, maternal-specific k-mers account for the largest proportion (> 48%), and the proportion of inherited paternal k-mers ranges from 14.3% to 17.1%, consistent with a 3*n* (maternal) × *n* (paternal) inheritance model.

banana breeding research. Morphological observation remains the primary method for authenticating banana hybrids; however, this approach is time-consuming and labor-intensive, particularly when distinguishing progeny from parents with minimal phenotypic differences. Polymerase chain reaction (PCR)-based genotyping techniques, such as those using simple sequence repeat (SSR) and ITS markers, have been applied in hybrid population analysis[25,28]. With advances in sequencing technology and decreasing costs, high-throughput whole-genome sequencing has emerged as the gold standard for precise genotyping[27]. In this study, we applied a k-mer-based approach using PacBio HiFi sequencing to accurately determine hybrids' authenticity and quantify the parental genetic contributions. This method offers a more efficient and precise solution for breeding banana hybrids.

The 3*x* × 2*x* hybridization strategy is a classical approach in banana breeding, producing numerous hybrid progeny[29,30]. The world's first hybrid cultivar, 'Goldfinger' AAAB (FHIA-01), was developed through distant hybridization between the improved diploid SH-3142 (AA, paternal) and the triploid 'Santa Catarina Prata' (AAB, Pome subgroup; maternal)[31]. Flow cytometry has confirmed tetraploidy in hybrid progeny derived from crosses between *M.* × *paradisiaca* and wild diploids, suggesting a 3*n* + *n* hybridization model[6]. In the present study, flow cytometry directly validated that DGDJ is triploid (3*x*), the paternal parents XGYJ and HNYJ are diploid (2*x*), and all four progeny are tetraploid (4*x*), providing unambiguous support for the 3*n* + *n* inheritance model. Similarly, studies on hybridization in *Populus* have demonstrated that triploids can serve as donors of unreduced gametes in interploidy crosses, acting as crucial intermediate vectors for ploidy variation in species[32]. GenomeScope analysis also validated tetraploidy (4*x*) in our progeny X170 and X193, indicating consistent tetraploid offspring regardless of the paternal genome type (A or B) (Fig. 3). Wild species typically produce *n* gametes but may occasionally generate 2*n* gametes. Our k-mer analysis revealed significantly lower paternal k-mer ratios in the progeny compared with the maternal contributions (Fig. 4, Supplementary Table S4), confirming the transmission of normal *n* gametes from the paternal parent. Notably, approximately 90% of the maternal k-mers were inherited, indicating predominant maternal genetic retention. These findings support the hypothesis that DGDJ participates in hybridization via female triploid gametes, contributing to its retained fertility.

Although the primary goal of this study was to identify true hybrids and characterize their genomic composition, the use of PacBio HiFi sequencing provides higher resolution for complex, polyploid genomes such as banana. For routine hybrid screening in large breeding populations, short-read sequencing (e.g., Illumina) or molecular markers remain more cost-effective alternatives. However, HiFi sequencing offers advantages in resolving subgenome-specific inheritance patterns and detecting structural variations that are difficult to capture with short reads. The method presented here is therefore most suitable for in-depth characterization of selected hybrids rather than for routine large-scale screening.

Conclusions

As a representative cultivar characterized by strong disease resistance but suboptimal fruit quality, plantain serves as an ideal maternal parent in hybridization because of its high compatibility. The 3*x* × 2*x* hybridization strategy, a classical approach in banana breeding, was utilized to produce hybrid progeny. Through flow cytometry and k-mer analysis, we elucidated the genetic backgrounds of these hybrids: Paternal parents contributed normal *n* gametes, whereas the maternal parent DGDJ produced 3*n* gametes, confirming a 3*n*+*n* inheritance model. All four progeny were confirmed as tetraploids (4*x*) by flow cytometry and genomic analysis. Additionally, we applied a k-mer-based method using PacBio HiFi sequencing data for identifying hybrid progeny, which provides a precise tool for characterizing complex genomic compositions in polyploid bananas. This approach may accelerate the screening of new hybrid germplasm and enhance the genetic resources available for the banana industry.

Author contributions

The authors confirm contributions to the paper as follows: study conception and design: Hu H, Wang Z; draft manuscript preparation: Jia C, Feng J; bioinformatics analysis: Feng J, Zhu M, Zhao J, Zhang Z; critical revision: Jia Z, Wang J. All authors reviewed and approved the final version of the manuscript.

Data availability

All data generated or analyzed during this study are included in this published article and supplementary information files.

Acknowledgments

The authors express their gratitude to the funding agencies and projects that supported this work, including the Key R&D Project of Hainan Province Yazhou Bay Joint Project (ZDYF2025GXJS138), the Central Public-interest Scientific Institution Basal Research Fund (CATASCXTD202308), the Central Public-interest Scientific Institution Basal Research Fund (No. 1630052022002), and the project of the National Key Laboratory for Tropical Crop Breeding (NKLTCBCXTD06).

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary information This accompanies this paper online at: https://doi.org/10.48130/frures-0026-0030.

Dates

Received 10 April 2026; Revised 3 July 2026; Accepted 23 July 2026; Published online 8 September 2026

References

[1] D'Hont A, Denoeud F, Aury JM, Baurens FC, Carreel F, et al. 2012. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488:213–217

[2] Davey MW, Gudimella R, Harikrishna JA, Sin LW, Khalid N, et al. 2013. "A draft *Musa balbisiana* genome sequence for molecular genetics in polyploid, inter- and intra-specific *Musa* hybrids". *BMC Genomics* 14:683

[3] De Langhe E, Vrydaghs L, De Maret P, Perrier X, Denham T. 2009. Why Bananas Matter: an introduction to the history of banana domestication. *Ethnobotany Research and Applications* 7:165–177

[4] Suman S. 2017. Plant tissue culture: a promising tool of quality material production with special reference to micropropagation of banana. *Biochemical and Cellular Archives* 17:1–26

[5] Simmonds NW, Shepherd K. 1955. The taxonomy and origins of the cultivated bananas. *Botanical Journal of the Linnean Society* 55:302–312

[6] Lin YE, Chiu HL, Wu CS, Chaw SM. 2024. Phylogenomics identifies parents of naturally occurring tetraploid bananas. *Botanical Studies* 65:19

[7] Anuradha C, Ramajayam D, Mayilvaganan M, Backiyarani S, Mol PP, et al. 2024. Molecular characterization of red banana and its somaclonal variant: a comprehensive study. *3 Biotech* 14:19

[8] Biswas MK, Yi G. 2016. Genes and markers: application in banana crop improvement. In *Banana: Genomics and Transgenic Approaches for Genetic Improvement*eds Mohandas S, Ravishankar K. Singapore: Springer. pp. 35–50 doi: 10.1007/978-981-10-1585-4_3

[9] Inta W, Traiperm P, Ruchisansakun S, Janssens SB, Viboonjun U, et al. 2023. Evolution and classification of Musaceae based on male floral morphology. *Plants* 12:1602

[10] Wong C, Kiew R, Loh JP, Gan LH, Set O, et al. 2001. Genetic diversity of the wild banana *Musa acuminata* Colla in Malaysia as evidenced by AFLP. *Annals of Botany* 88:1017–1025

[11] Nair AS, Teo CH, Schwarzacher T, Harrison PH. 2005. Genome classification of banana cultivars from South India using IRAP markers. *Euphytica* 144:285–290

[12] Nwakanma DC, Pillay M, Okoli BE, Tenkouano A. 2003. PCR-RFLP of the ribosomal DNA internal transcribed spacers (ITS) provides markers for the A and B genomes in *Musa* L. *Theoretical and Applied Genetics* 108:154–159

[13] Čížková J, Hřibová E, Christelová P, Van den Houwe I, Häkkinen M, et al. 2015. Molecular and cytogenetic characterization of wild Musa species. *PLoS One* 10:e0134096

[14] Martin KP, Pachathundikandi SK, Zhang CL, Slater A, Madassery J. 2006. RAPD analysis of a variant of banana(*Musa* sp.) cv. grande naine and its propagation via shoot tip culture. *In Vitro Cellular & Developmental Biology-Plant* 42:188–192

[15] Wang Z, Miao H, Liu J, Xu B, Yao X, et al. 2019. *Musa balbisiana* genome reveals subgenome evolution and functional divergence. *Nature Plants* 5:810–821

[16] Li Z, Wang J, Fu Y, Jing Y, Huang B, et al. 2022. The *Musa troglodytarum* L. genome provides insights into the mechanism of non-climacteric behaviour and enrichment of carotenoids. *BMC Biology* 20:186

[17] Murray MG, Thompson WF. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research* 8:4321–4325

[18] Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications* 11:1432

[19] Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100

[20] Voichek Y, Weigel D. 2020. Identifying genetic variants underlying phenotypic variation in plants without complete genomes. *Nature Genetics* 52:534–540

[21] Kokot M, Długosz M, Deorowicz S. 2017. KMC 3: counting and manipulating k-mer statistics. *Bioinformatics* 33:2759–2761

[22] Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, et al. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4:7

[23] Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nature Methods* 18:170–175

[24] Jones DR. 2000. *Diseases of banana, Abacá and Enset*. Wallingford: CAB International. 544 pp

[25] Batte M, Nyine M, Uwimana B, Swennen R, Akech V, et al. 2020. Significant progressive heterobeltiosis in banana crossbreeding. *BMC Plant Biology* 20:489

[26] Heslop-Harrison JS, Schwarzacher T. 2007. Domestication, genomics and the future for banana. *Annals of Botany* 100:1073–1084

[27] Martin G, Cottin A, Baurens FC, Labadie K, Hervouet C, et al. 2023. Interspecific introgression patterns reveal the origins of worldwide cultivated bananas in New Guinea. *The Plant Journal* 113:802–818

[28] Zeng H, Huang B, Xu L, Wu Y. 2024. Banana classification using Sanger sequencing of the ribosomal DNA internal transcribed spacer (ITS) region. *Plants* 13:2173

[29] Tomekpe K, Jenny C, Escalant JV. 2004. A review of conventional improvement strategies for *Musa*. *InfoMusa* 13:2–6

[30] Ortiz R, Swennen R. 2014. From crossbreeding to biotechnology-facilitated improvement of banana and plantain. *Biotechnology Advances* 32:158–169

[31] Rowe P, Rosales F. 1993. Diploid breeding at FHIA and the development of Gold finger (FHIA-01). *InfoMusa* 2:9–11

[32] Wang J, Huo B, Liu W, Li D, Liao L. 2017. Abnormal meiosis in an intersectional allotriploid of *Populus* L. and segregation of ploidy levels in 2x × 3x progeny. *PLoS One* 12:e0181767