

Article

Assessing loss of regulatory divergence, genome–transcriptome incongruence, and preferential expression switching in abaca/banana backcrosses

Nelzo C. Ereful^{1,2*}, Antonio G. Lalusin³, and Antonio C. Laurena²

*Corresponding author: ncereful@up.edu.ph; nelcaster7@yahoo.com

¹Plant Physiology Laboratory, Institute of Plant Breeding, College of Agriculture and Food Science, University of the Philippines Los Baños, Laguna, Philippines

²Philippine Genome Center for Agriculture, University of the Philippines Los Baños, Laguna, Philippines

³Institute of Crop Science, College of Agriculture and Food Science, University of the Philippines Los Baños, Laguna, Philippines

Abstract: The *Musa textilis* var. Abuab has high fiber quality (FQ) but has low resistance against abaca bunchy top virus (AbBTV); the *Musa balbisiana* var. Pacol, has low FQ but high resistance against AbBTV. Their backcrosses (BC₂ and BC₃) possess both desirable traits. Analysis using RNA-seq showed that the regulatory divergence of Abuab and Pacol is largely explained by cis differences with 27.4% and 22.3% if we are to assess it using BC₂ and BC₃, respectively. Cis differences between the two genotypes are significantly reduced from BC₂ to BC₃ due to changes in genomic constitution. Trans, on the other hand, is robust to changes in allelic composition. All these are attributed to the loss of heterozygosity in the BC₃ relative to BC₂. Further analysis showed that both backcrosses exhibited genome-wide preferential expression of Pacol- over Abuab-specific alleles, despite the wider genetic presence of the latter in the hybrids. The ratio of the two genotype-specific expressed transcripts and the ratio of their corresponding genetic make-up are significantly disproportionate, a phenomenon which we refer here as “genome–transcriptome incongruence”. We also observed preferential expression switching in which several genes prefer Abuab- (or Pacol-) specific allele in the BC₂ but switched to Pacol- (or Abuab-) specific allele in the BC₃ genome.

Keywords: abaca (*Musa textilis*), allelic imbalance, regulatory divergence, banana (*M. balbisiana*), allele-specific expression

1. Introduction

Abaca (*M. textilis* Née), also known as Manila hemp in the international community, is a close relative of banana. It is widely cultivated in the Philippines which supplies 85% of the world’s demand [1]. Ecuador, Costa Rica and other Southeast Asian countries such as Indonesia also supply abaca fiber, although, at lower proportions.

This crop is mainly cultivated for its fiber which is used to make ropes, currency notes, textiles, among others. Its uses have further broadened in the automotive and aerospace engineering industries due to its high tensile strength. In cars, it is used as an underfloor protection [2]. Recently, due to the COVID19 pandemic, its use has expanded in the medical industry to fabricate personal protective equipment (PPE) owing to its high medical grade quality [3].

Abaca (T genome) is placed under the Callimusa section, members of which have a ploidy of $2n = 20$. *M. acuminata* (A genome), along with *M. balbisiana* (B genome), are

diploid but are generally triploid. The double haploid Pahang belonging to *M. acuminata* subspecies *malaccensis* has a ploidy of $2n=22$ with 523-megabase genome [4]. Varieties with various ploidy (diploid, triploid, and tetraploid) resulted from crossing these A- and B-genomes. Majority of cultivated banana varieties are predominately triploid [5].

Despite the importance of this crop, studies on its molecular biology are relatively lagging compared to other fiber plants such as jute, cotton, and hemp. The genome sequence of abaca (var. Abuab) has been recently decoded using high-throughput sequencing [6]. This will immensely advance our understanding of its molecular sequence and functional divergence to other *Musa* spp. and will aid breeders into its improvement. A recent paper revealed the high genetic diversity of the abaca germplasm, with Shannon diversity index, $I = 0.68$, using 150 accessions across the Philippines [7].

In this paper, we assayed the parental genotypes abaca var. Abuab and wild banana (*M. balbisiana*) var. Pacol and their backcross hybrids (BC₂ and BC₃) for allele-specific expression (ASE) imbalance using RNA-seq. This happens when in F₁ hybrids, one of the two alleles is driven significantly higher expression levels compared to the other allele. F₁ hybrids are host to two co-residing genomes and asymmetric expression is attributed to cis-regulatory divergence between these two specific alleles.

ASE has been a subject of a number of papers in different organisms such as humans [8, 9], *Drosophila* [10,11,12], maize [13], stickleback [14], coffee [15], and rice [16,17]. There are several approaches in assessing ASE which include single-base extension [18], pyrosequencing [10], and most recently RNA-seq (e.g. [12]). By comparing ASE in the hybrids and the relative expression level of the same gene(s) in the parents, cis- and/or trans-regulatory divergence can be estimated. This approach has been used extensively to estimate the contributions of cis and/or trans regulatory factor on the evolutionary divergence of species of *Drosophila* [10,11,12], stickleback [14], and coffee [15]. A separate study by [19] showed the asymmetric expression of Stay-Green 1, a key chlorophyll degradation gene, in the banana AAB/ABB cultivars.

Allelic imbalance is mostly assayed in F₁ hybrids. However, there are several studies which used introgression lines such as monkeyflower [20], *Drosophila* [21], mouse [22], and *Solanum* [23] to assess ASE imbalance.

In this study, we sequenced the total transcript population of both backcrosses including their parents using RNA-seq to elucidate ASE imbalance in the backcrosses and the regulatory divergence in the parental genotypes. To our knowledge, no study has been performed to assess genome-wide allelic imbalance in *Musa* spp. hybrids, moreso, in backcrosses; thus, this investigation.

2. Materials and Methods

The abaca (*M. textilis*) var. Abuab, wild banana (*M. balbisiana*) var. Pacol and their backcrosses, BC₂ and BC₃, were grown and maintained at Feeds and Industrial Crops Section (FICS) collection site of the Institute of Plant Breeding (IPB), UPLB (14°09'09.7"N 121°15'39.2"E). A schematic diagram is shown on how backcrosses were generated (Fig. 1).

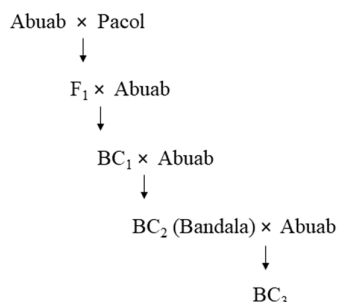


Figure 1. F₁BC₂ (Bandala) and F₁BC₃ were created by crossing Abuab and Pacol. Abuab serves as the maternal recipient; Pacol as the pollen donor at F₀. In the succeeding generations, Abuab serves as the pollen donor. Note that BC₂ is locally named as “Bandala.”

2.1. RNA extraction

The central whorl of the stalk samples of three-month-old suckers were collected on 22 June 2021 between 1000H and 1100H. Samples were snap-frozen in liquid N and were stored in –80 °C freezer until further extraction.

Three replicates of the same variety were pooled in equal weight (~0.33 g) to make up to 1.0 g. This was previously referred to as biological averaging [24] which was found to be more cost-efficient while maintaining statistical power [25]. Also, we preferred pooling of samples due to the challenge associated with the extraction of RNA samples from abaca and banana. Pooling of samples for RNA-seq has been performed in recent studies [26,27]. Pooled samples were then ground using mortar and pestles. All autoclavable equipment and reagents were sterilized to avoid unwanted contamination.

We followed the modified cetyl trimethylammonium bromide (CTAB) protocol with 2% β-mercaptoethanol method with LiCl precipitation for RNA extraction as previously described by Dr. John Carlson of Schatz Center for Tree Molecular Genetics at the Pennsylvania State University. RNA was reconstituted using 800μL SSTE with 4uL Monarch® DNase in 20 uL DNase buffer, incubated at 37 °C for 1 hour. DNase was removed by adding 24:1 Chloroform–Isoamyl Alcohol. After spinning for 20 min at 14,000 rpm at 4 °C, the aqueous phase was obtained and 0.1 volume of Na acetate and 2 volumes of absolute ethanol was added. Samples were incubated for 1 hour at –80° C then were spun for 30 min at 14,000 rpm at 4 °C. The pellet was obtained and was washed with 75% ethanol by spinning. Pellets were dried for 15 min and a 50-uL of nuclease-free water was added. Bands were resolved using 2% agarose gels, stained with GelRed®. RNA was quantified using Nanodrop spectrophotometry. Its quality is assessed using BioAnalyzer™ (Agilent Technologies Inc.).

The quality and quantity of total RNA was assessed using a NanoDrop 2000 spectrophotometer (ThermoFisher Scientific, Inc.) and an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

2.2. Bioinformatics analysis

Statistical analysis was performed using Rstudio v.1.4.1717 [28] while the computational pipeline was done using Linux environment (Debian 4.9.246-2). All R scripts and Linux commands are stored in our github repository site: <https://bit.ly/3tP4BtE> (accessed on 12 June 2022).

2.2.1. Reads pre-processing

Reads were quality-checked using FastQC [29]. No adapters and low-quality reads were found, therefore, no further pre-processing step was performed.

2.2.2. Mapping

Reads generated from the parental genotypes *M. textilis* var. Abuab and *M. balbisiana* var. Pacol were aligned against the genome references Abuab [6] and double haploid of Pisang Klutuk Wulung (DH-PKW) [5], respectively, using STAR v.2.7.7a [30] with the following arguments: --outFilterMatchNmin 0), with the following options: --outFilterScoreMinOverLread 0.3 --outFilterMatchNminOverLread 0.3. Because Abuab (*M. textilis*) and Pacol (*M. balbisiana*) are inter-specifically divergent, we searched for transcript orthologs using OrthoVenn2 [31] with protein fasta sequence of both reference assemblies as inputs, implementing E-value of 1e-5 and Inflation Index of 1.5.

To assess regulatory divergence between the parents and ASE in the backcrosses, we followed the pipeline as previously described [20,32,33] with modifications. Reads from

BC₂ and BC₃ genotypes were competitively aligned against the concatenated Abuab and DH-PKW genome reference assemblies, also using STAR, with the same parameters as described above.

2.2.3. Read count quantification and normalization

For ASE analysis in the backcrosses, uniquely mapping reads aligning to each reference assembly were quantified using subread featureCount function [34]. We counted the primary alignment alone (option: --primary) with the following additional options: -t exon, the feature type to count read against, and; -g transcript_id, the attribute type to summarize counts. These unique reads aligning to orthologous transcripts of merged pseudo-reference assemblies represent ASE [33]. A raw read count corresponding to the various genotypic counts of the parents and allele-specific counts from the backcrosses (columns) and transcript orthologs (rows) was created.

Reads with a total row sum of 0 were removed to increase computational speed and reduce data size. We normalized the reads using edgeR's Trimmed Means of M values with lowly expressed genes being filtered out [35]. The total normalized read counts between the parents were calculated. The total should be at least 20 (Abuab + Pacol \geq 20) to be considered for further analysis. We added 1 for read counts with 0 values to avoid 0 numerator or denominator, the sum, however, should be at least 20. The expression ratios and their Log(2) fold-change(FC) between the two parental read counts and the binomial exact test of equal proportion ($p = 0.5$, with FDR-adjusted P -values; in R) were calculated. Any significant difference between the parental orthologous transcripts gives estimates of the parental expression divergence (P). Likewise, in the BC₂ and BC₃ genotypes, the Log(2)FC between the two genotype-specific allele (H) in each backcross and their binomial exact test ($p = 0.5$; using R) were calculated (see github repository for custom R scripts we created).

Trans effects were estimated by subtracting H from P ($T = P - H$) [11,36,37]. We then tested the datasets using Fisher exact test followed by FDR analysis.

For cis and/or trans regulatory assignments (cis, trans, cis \times trans, cis – trans and cis + trans), we considered the most conservative analysis (i.e. binomial and Fisher exact tests, FDR < 0.5%) in the Results and Discussions as previously performed (e.g. [12,17]). For cis/trans regulatory assignments, see our github repository site: <https://bit.ly/3tP4BtE>.

2.2.4. Allelic imbalance polymorphism in the backcrosses

To assess expression imbalance between heterozygous sites (i.e., orthologous genes between the two genotype-specific alleles) in BC₂ and BC₃, we implemented binomial exact test, with a null hypothesis of equal proportion ($p = 0.50$). We used a custom R script to execute this command (see github site). Transcript orthologs which exhibit an FDR < 0.05 are said to be significantly asymmetrically expressed. Additionally, we calculated the log-transformed expression ratios of the read counts (Abuab-/Pacol-specific allele) to the base 2 (Log₂FC). A biological threshold of $|\text{Log}_2 \text{FC}| \geq 1$, binomial ($p = 0.5$, FDR < 0.05) was implemented. Only features with total normalized read count of 20 reads between the two specific alleles were considered for further analysis ($\text{Abuab}_{\text{BC}_2} + \text{Pacol}_{\text{BC}_2} \geq 20$; $\text{Abuab}_{\text{BC}_3} + \text{Pacol}_{\text{BC}_3} \geq 20$) to ensure genes are expressed and to avoid inclusion of artifacts, as previously performed (e.g. [12,16,37,38]). We incremented read counts with a value of 1 for genotype-specific alleles with 0 read counts to avoid a numerator or a denominator of 0.

We also tested the proportions between parental genotypes and genotype-specific alleles using z-score test for two population proportions, two-tailed with $N = 33,511$ (for BC₂) and $N = 33,394$ (BC₃). These are combined numbers of the predicted genes for *M. textilis* Abuab (33,277) and *M. balbisiana* DH-PKW (35,148) [6].

2.2.5. Relative transcript accumulation ratio in the backcrosses

We concatenated the normalized read counts of both BC₂ and BC₃ into a single matrix. The sum of the normalized read counts of the genotype-specific alleles in each

backcross was calculated. Only transcript orthologs with a total normalized read count of 20 in either or both backcrosses were considered for further analysis. As performed above, we added 1 for specific alleles with 0 values. Expression ratios between the two genotype-specific alleles (Abuab/Pacol) were \log_2 -transformed in both hybrids. Genes which are asymmetrically expressed at a biological threshold of $|\log_2FC| \geq 1$ and statistical threshold of $FDR < 0.5\%$ (binomial test of equal proportion, $p = 0.5$) in either or both genotype-specific alleles were considered for further classification. Genes which exhibit $FDR < 0.5\%$ are said to be asymmetrically expressed.

Transcript orthologs which are not significantly differentially expressed using binomial exact test ($FDR < 0.5\%$) in both backcrosses are said to exhibit no ASE imbalance.

2.2.6. GO enrichment analysis

Functional annotations of the genes were taken from the annotation provided for Abuab reference sequence [6]. GOfuncR (hypergeometric test, overrepresented at $FWER < 0.1$) [39] was used for GO functional enrichment analysis. Unless otherwise indicated, we used 0.1 as statistical threshold for most of the GO enrichment analyses since FWER provides a conservative measure of error rate. The participation of these genes to the predicted pathways was further confirmed in published literature.

3. Results and Discussion

The abaca (*M. textilis*) var. Abuab was found to have high-fiber quality but is susceptible to AbBTV [40]. On the other hand, Pacol, a *M. balbisiana* variety possesses phenotypic characters contrasting to that of Abuab, i.e., low fiber quality but resistant against AbBTV [41,42,43]. These two *Musa* spp. were crossed to create the F_1 which was successively crossed to Abuab to create the backcrosses, BC_2 and BC_3 (see Fig. 1) to identify allelic imbalance polymorphism and map cis and/or trans regulatory divergence.

Pacol is a diploid and was shown to have a chromosome number of $2n = 22$, while Abuab has $2n = 10$. Selection of F_1 and BC_1 individuals with abaca-like phenotypic traits and successive backcrossing of the selected progenies to Abuab showed a ploidy level of $2n = 20$ for both BC_2 and BC_3 based on morphological characterization and karyotyping analysis [42].

We chose backcrosses (not the F_1 hybrids) for allelic imbalance assay as these are economically important genotypes with the desirable traits of industrial significance – superior FQ and AbBTV resistance. Recent studies on several organisms showed that asymmetrically expressed genes were found associated with traits of interest including rice [44], yeast [45], and banana [19]. Therefore, it would be interesting to identify asymmetrically expressed genes as they could be potential candidates for the traits of interest.

BC_2 or Bandala, has been planted in multilocation trials sites in the different regions of the Philippines – Luzon, Visayas, and Mindanao [46,47]. BC_3 , on the other hand, was relatively recently developed. Based on initial field assessment, it has superior fiber quality and is resistant against bunchy top virus. Molecular study using RNA-seq revealed that it is the best among five abaca genotypes as it induces significantly higher number of genes associated with both desirable traits [16].

We generated multi-dimensional scaling (MDS) plot using the normalized read counts to visualize the relationship among the genotypes and each co-residing genomes in the backcrosses. Results (Fig. 2) clearly indicate that Dim.1 largely explains the divergence of the two genotypes with Abuab and Pacol and their respective backcrosses clearly separated. Dim 2, on the other hand, separates the parental genotypes from the backcrosses.

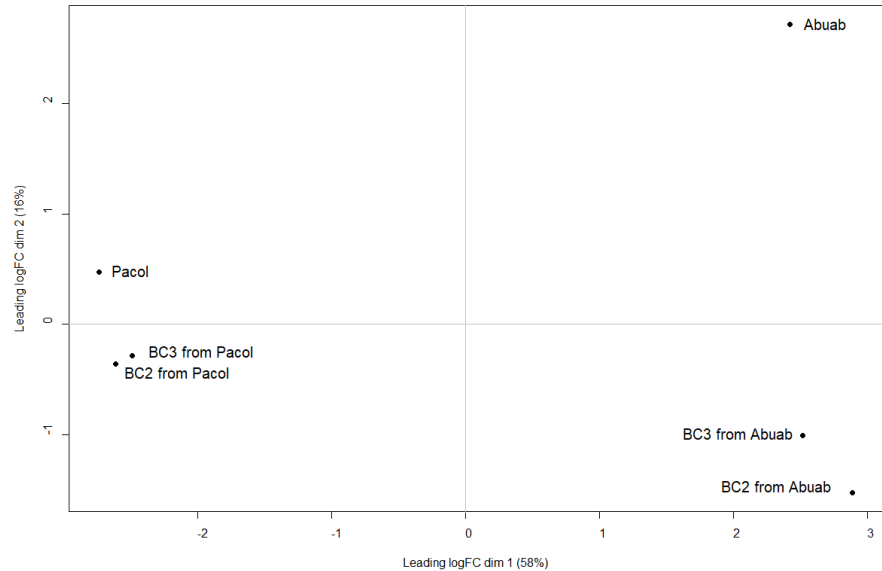


Figure 2. MDS plot of the parental genotypes and the genotype-specific alleles derived from each backcross. Legend: “BC₂ from Pacol” and “BC₃ from Pacol” indicate the total population of transcripts derived from Pacol in the BC₂ and BC₃ genomes, respectively. “BC₂ from Abuab” and “BC₃ from Abuab” indicate the total population of transcripts derived from Abuab in the BC₂ and BC₃ genomes, respectively.

3.1. Regulatory divergence between Abuab and Pacol

The genome-wide expression performance of the abaca-specific genes of BC₂ and BC₃ along with their parents were shown in a companion paper [16]. However, assessment of regulatory divergence between the parents and asymmetric expression in their backcrosses has not been performed.

Briefly, we aligned the parental reads against Abuab reference sequence [6] for Abuab and DH-PKW [5] for Pacol. The backcross reads, on the other hand, were mapped against the concatenated reference assemblies (see Materials and Method). Only the uniquely mapping reads were quantified. Transcripts with corresponding orthologs between the two genotype-specific alleles with a total normalized read counts of 20 were retained for further analysis (unmatched transcripts were dropped as previously performed [48]).

Results showed that there were 14,868 transcripts orthologs or transcripts with matching sequences between the two *Musa* spp. peptide reference assemblies. This accounts for nearly 50% of transcript population of both varieties suggestive of their degree of expression divergence.

Regulatory differences between intra- or interspecifically related organisms are often estimated using their F₁ progenies. In our case, we used backcross hybrids to estimate regulatory divergence between the two inter-specifically divergent *Musa* spp. We found 9295 genes with read counts greater than 20 between the parents (Abuab + Pacol ≥ 20). Results indicate that regulatory divergence between Abuab and Pacol is largely explained by cis differences with 27.4% and 22.3% as estimated and assayed in BC₂ and BC₃, respectively (excludes ambiguous and conserved; Table 1; red points in Fig. 3; see also Table S1 and Table S2 for lists of orthologous genes diverging in cis and/or trans using BC₂ and BC₃, respectively). If we are to extrapolate these values, the degree of expression divergence using F₁ will be mostly explained by cis differences with proportion beyond these

values. This, however, is a speculation and is an interesting area of further inquiry. The finding that cis regulatory factors explaining interspecific divergence has been widely reported in literature (e.g. [10,14,17,49,50]). On the other hand, trans regulatory factor (purple in Fig. 3) and its interaction with cis (synergistic, antagonistic, and compensating) may also explain their divergence, albeit, very modestly.

Further analysis showed that the number of cis-diverging genes is significantly reduced from BC₂ to BC₃, while the conserved transcript orthologs significantly increased ($P < 0.0001$; Table 1). This is attributed to the loss of heterozygosity (or increase of homozygosity) in the BC₃ genome relative to BC₂. These results suggest that with the changes in the genomic constitution, cis is significantly affected due to the loss of Pacol (or gain of Abuab) genomic segments. Its interactions with trans are, likewise, significantly affected (synergistic, antagonistic, and compensatory; $P < 0.0001$). Points representing these interactions, including trans, noticeably disappeared in BC₃ (Fig. 3B) relative to BC₂ (Fig. 3A).

Trans, on the other hand, is robust to changes in genomic composition ($P = 0.3817$). In recent papers, trans is significantly affected by changes in external factors [17,51] but not in changes in genetic make-up. On the contrary, cis is robust to environmental changes [17,52].

Table 1. Number of transcript orthologs exhibiting cis and/or trans regulatory divergence between Abuab ang Pacol assayed using their BC₂ and BC₃. χ^2 test (with P -values, two-tailed; test for 2x2 contingency table with Yates' continuity correction) indicates the significant difference of the number of genes classified in each regulatory category between the two backcrosses.

Regulatory factors/interactions	F ₁ BC ₂ (%)	F ₁ BC ₃ (%)	χ^2 stat with Yates correction	P -value (two tailed)
Cis	2647 (27.4)	2070 (22.3)	66.73	<0.0001
Cis + trans (synergistic)	361 (3.7)	121 (1.3)	112.48	<0.0001
Cis × trans (antagonistic)	478 (5.0)	14 (0.2)	429.55	<0.0001
Compensating	441 (4.6)	203 (2.2)	81.21	<0.0001
Trans	161 (1.7)	139 (1.5)	0.79	0.3736
Ambiguous	2955 (30.6)	3090 (33.2)	15.08	0.0001
Conserved	2613 (27.1)	3658 (39.4)	322.74	<0.0001
Total	9656	9295		

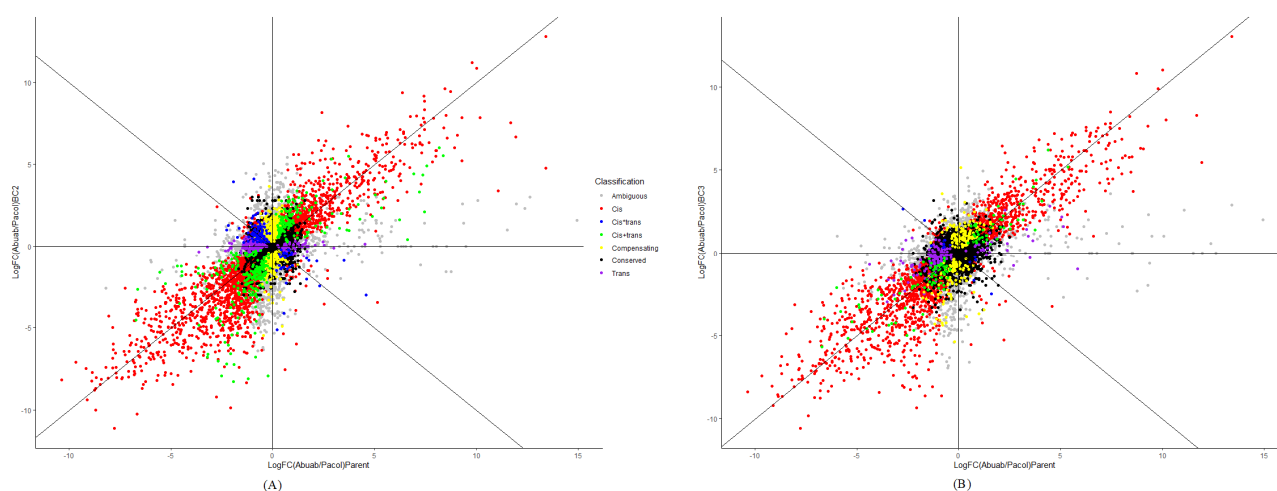


Figure 3. The cis and/or trans regulatory divergence architecture between Abuab and Pacol as estimated using their (A) BC₂ and (B) BC₃ backcross hybrids. Note the loss of cis and/or trans regulatory divergence from BC₂ to BC₃. Figure generated using ggplot2 [53] in Rstudio [28].

3.2. Asymmetric expression

We wanted to test whether asymmetric expression exists at sites heterozygous between two genotype-specific alleles in the backcross lines. This will shed hints on the biased directionality of expression of orthologous genes between the two co-residing genomes.

3.2.1. Allelic imbalance in the BC₂ genome

Briefly, we concatenated and normalized read counts of the two genotype-specific alleles in the backcrosses. MA plot showed symmetrical data cloud in both BC₂ and BC₃ (Figs. S1 and S2, respectively) which indicates that normalization was effective. We implemented binomial exact test on the normalized read counts (at $|\text{Log}_2\text{FC}| \geq 1$; $p = 0.5$, $\text{FDR} < 0.05$, see Materials and Methods) to identify genes which are asymmetrically expressed from the expected ratio of 1:1. We found 3971 transcript orthologs which satisfy these criteria, of which 1703 and 2268 preferentially expressed Abuab- and Pacol-specific alleles, respectively in BC₂ (Table S3). (Note that these values exclude symmetrically expressed genes. That is, those that do not satisfy both/either biological and/or statistical thresholds).

Apparently, these results revealed that BC₂, on a transcriptome-wide scale, preferentially expressed the *M. balbisiana* (banana) Pacol over the *M. textilis* (abaca) Abuab allele. This is in contrast against its expected genomic constitution in which BC₂ theoretically hosts 87.5% Abuab and 12.5% Pacol alleles. There is a significant discrepancy of these proportions to the ratio of the expressed genotype-specific transcripts in BC₂, 43% Abuab–57% Pacol ($z = 70.4$, $P < 0.0001$; z -score test for two population proportions). This shows that the genetic ratios of the two alleles in the backcross hybrids and its corresponding transcript expression ratios are significantly disproportionate, a condition which we refer here as “genome – transcriptome incongruence”.

In the BC₂ nucleus, both parent-specific alleles are exposed to the same trans regulatory factors. Therefore, allelic imbalance between two heterozygous sites has been ascribed to cis-divergence and varying allele-specific epigenetic landscape [10,11].

As polysaccharide and lignin were found associated with high-fiber quality [16], we searched for such genes allelically imbalanced in BC₂. We found 16 genes encoding for

either putative cellulose synthase or cellulose synthase, 14 of which preferred the Abuab-specific allele ($|\text{Log}_2\text{FC}| \geq 1$; $\text{FDR} < 0.05$). Abuab confers the high-fiber quality of the backcrosses. Six genes encoding for Cinnamoyl-CoA reductase, an enzyme involved in lignin biosynthesis [54] were found asymmetrically expressed, three genes preferred each specific allele.

GO analysis using GOfuncR [39] of the Abuab-specific transcript orthologs significantly differentially expressed using binomial exact test ($\text{FDR} < 0.05$) showed enrichment of genes associated with cellular component biogenesis, cellular component organization or biogenesis, and ribonucleoprotein complex biogenesis (Biological Process, $\text{FWER} < 0.1$) (see Table S4). On the other hand, no GO terms were found enriched in Pacol-specific transcript orthologs ($\text{FWER} < 0.1$) using GOfuncR.

3.2.2. Allelic imbalance in the BC₃ genome

In BC₃, 3561 were found significantly asymmetrically expressed between the two alleles ($|\text{Log}_2\text{FC}| \geq 1$; $p = 0.5$, $\text{FDR} < 0.05$) (Table S5). Of these, 1609 (45%) and 1952 (55%) preferentially expressed the Abuab- and Pacol-specific alleles, respectively. Similar to Bandala (BC₂), there is a broader expression proportion of the banana allele as compared to the abaca allele. This is contrary to its genetic background in which BC₃ is made up of 93.75% Abuab and 6.25% Pacol alleles. The expression proportion of Pacol allele is still significantly large considering its modest genetic presence in BC₃ ($z = 88.30$; $P < 0.00001$; z -score test for two population proportions, two-tailed). There appears to be consistent observation that the genetic ratio between the two alleles and their corresponding transcript expression ratio are incongruent in the backcrosses.

There are 3971 asymmetrically expressed in BC₂; and only 3561 in the BC₃ genotypic line ($\text{FC} \geq 2$; $p = 0.5$, $\text{FDR} < 0.05$). This shows that BC₃ exhibited lesser number of genes asymmetrically expressed as compared to BC₂ potentially due to diminishing or gradual loss of heterozygosity (or increasing homozygosity) as a consequence of backcrossing to the recurring parent. This has been partly observed in a previous paper by [55] which states that the loss of heterozygosity, a common form of allelic imbalance, happens when heterozygotic lines becomes homozygous because one of the two alleles gets lost.

Fifteen (15) genes encoding for either cellulose synthase or putative cellulose synthase were found to exhibit bias expression mostly preferring Abuab allele (14 genes) with one gene preferring Pacol-specific allele ($\text{FC} \geq 2$; $\text{FDR} < 0.05$). Notably, a gene encoding for Photosystem I assembly protein was the most allelically imbalanced gene exhibiting 13 \times fold and preferentially expressing Abuab-specific allele. This suggests an active photosynthetic activity during fiber synthesis at the vegetative stage with preferential expression favoring Abuab allele. No GO terms were found enriched in either of the specific alleles (at $\text{FWER} < 0.1$) using GOfuncR.

3.2.3. Extreme cases of allelic imbalance. Genes encoding for Photosystem I assembly protein (LOCUS_023583-RA; 13 \times $\log_2(\text{FC})$), Pre-mRNA-processing-splicing factor 8A (LOCUS_012301-RA; 11 \times), Serine carboxypeptidase-like 50 (LOCUS_015865-RA; 11 \times), and ABC transporter G family member 31 (LOCUS_002982-RA; 9.5 \times), were observed to exhibit the highest expression Log_2FC , all preferring Abuab-specific allele in BC₂ (Table S2). Interestingly, the same genes were found to exhibit extremely high fold expression in BC₃ (Table S3) with Abuab allele being favored: Photosystem I assembly protein (LOCUS_023583-RA; 13.6 \times), ABC transporter G family member 31 (9.4 \times), Pre-mRNA-processing-splicing factor 8A (LOCUS_012301-RA; 11.6 \times), and Serine carboxypeptidase-like 50 (LOCUS_015865-RA; 9.9 \times) (note: all fold expression values are in Log_2FC). These findings highlight the active engagement of photosynthesis, transcription, and transport during synthesis of fiber components (sugars and lignin) in both backcrosses with Abuab being preferred over Pacol-specific allele.

Pacol allele, on the other hand, confers the abaca bunchy top virus (AbBTV) resistance. We found four genes (Mba11_g21030, Mba05_g06300, Mba06_g30590, and Mba04_g19640) encoding for (Enhanced) Disease Resistance protein favoring Pacol allele

commonly asymmetrically expressed in both backcrosses (note: because Pacol is a *M. balbisiana* var., we used the DH-PKW naming system). Because the genotypes were not infected, genes associated with disease resistance are modestly detected. Challenging the genotypes with AbBTV is highly recommended in succeeding studies to identify allelically imbalanced genes under the biotic stress.

Furthermore, we found 957 genes commonly preferring Abuab-specific alleles in both backcrosses; 1409 genes, preferring Pacol-specific allele (Fig. 4). These results suggest the consistency of a specific allele being preferentially expressed.

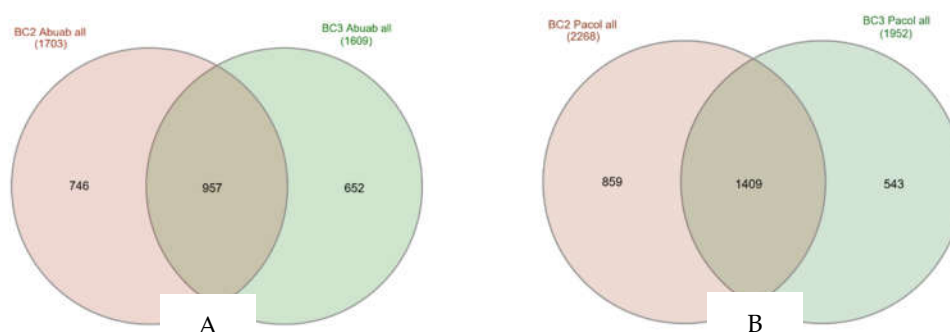


Figure 4. Number of genes commonly and uniquely asymmetrically expressed preferring (A) Abuab-specific allele, and (B) Pacol-specific alleles, between the two backcrosses. (Note: In both figures, Abuab all. and Pacol all. refer to Abuab and Pacol alleles, respectively).

3.3. Relative transcript accumulation ratio in the backcrosses

Above, we showed that cis-differences explained the divergence between the two parents and the prevalence of allelic imbalance polymorphism in each backcross. To assess changes in the preferential expression between alleles, the relative Abuab/Pacol log-transformed accumulation ratios in these backcrosses were analyzed. Fig. 2 (Table S6) depicts the over-all cis differences attributed to the variations in the genetic segments between the two backcrosses. Differences in the magnitude (level of expression) and direction (preference), are ascribed to cis differences with the addition of Abuab segments (loss of Pacol segments) acquired by BC₃ as a consequence of further backcrossing BC₂ to the recurring parent.

Results indicate that there are 10,112 transcript orthologs in either/both backcrosses which have more than 20 total normalized read counts. Of these, 2980 genes (29.5%) exhibit commonly asymmetrically expressed between the backcrosses signifying common cis-regulatory differences (red dots lying closely in the $y = x$ curve in Fig. 1; binomial exact test of equal proportion, $FDR < 0.5\%$). These are genes exhibiting conserved cis differences between the two hybrids with significant asymmetric expression (termed here as “conserved asymmetric”). GO enrichment analysis showed genes to be significantly associated with intracellular transport, cellular localization, and establishment of localization in cell (Biological Process, BP) and purine ribonucleoside triphosphate binding, and ATP binding (molecular function, MF, $FWER < 0.05$) (Table S7). This shows that these genes are involved during transport and binding of a range of molecules. On the other hand, 4,329 (42.8%) did not exhibit asymmetric expression (no ASE; black dots in Fig. 5). This proportion suggests that an enormous number of the total transcript population are symmetrically expressed in the backcrosses; hence, “conserved symmetric”.

There were 1576 (15.6%) genes allelically imbalanced in BC₂ which were symmetrically expressed in BC₃; hence, BC₂-specific ASE imbalance (yellow dots in Fig. 5). On the other hand, there were 1227 (12.1%) genes exhibiting asymmetric expression in BC₃ which were symmetrically expressed in BC₂; hence, BC₃-specific ASE (blue dots in Fig. 5). Apparently, there is a reduction of biased directionality of ASE from BC₂ to BC₃. These changes are ascribed to the augmentation or further “dilution” of Abuab segments (or loss of Pacol allele) in the BC₃ genome, thus, the reduction of heterozygous (or increase of homozygous) cis regulatory elements. Increasing homozygosity means decreasing cis differences.

We expect that all asymmetrically expressed genes in both BC₂ and BC₃ (red dots) should lie on the $y=x$ curve. However, there are also genes allelically imbalanced in both backcrosses lying on the $y=-x$ curve (red points). These genes exhibit preferential expression switching (or allelic imbalance switching) in which a particular gene prefers Abuab- (or Pacol-) specific allele in BC₂ but switches to Pacol- (or Abuab-) specific allele in BC₃, or vice-versa. There are 59 genes (FDR<0.5%) which exhibit this kind of swinging expression behavior and are significantly enriched in protein domain specific binding (MF, GO:0019904, FWER <0.05; Table S5). There were 34 genes preferentially expressing Pacol-over Abuab-specific allele in BC₂ but then preferred Abuab over Pacol allele in BC₃ (FDR<0.5%). On the contrary, there were 25 genes preferentially expressing Abuab over Pacol allele in BC₂ but switched vice-versa in BC₃. This phenomenon demonstrates the transcriptional versatility of an advanced backcross genotypes owing to heterozygosity.

Preferential expression switching has been reported previously, however, between contrasting water regimen; i.e., non- and water-stress conditions in rice [17]. In the current study, preferential expression switching happens between two backcrosses.

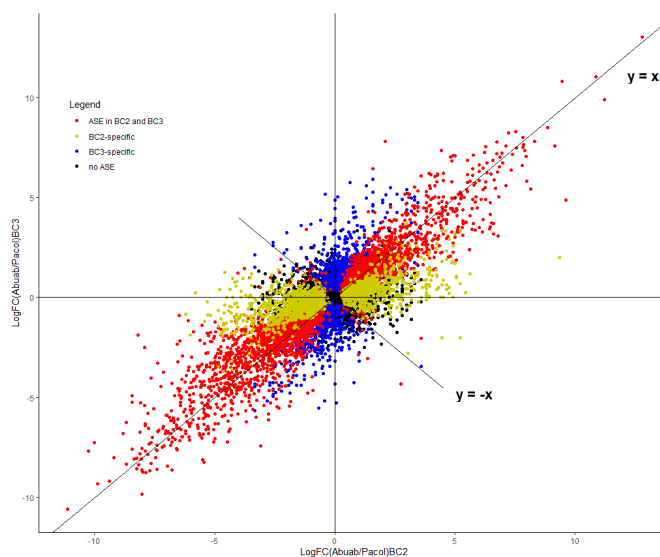


Figure 5. The overall cis differences between BC₂ and BC₃ (discussed in the text). The relative abundance of transcript isoforms (expressed as log-transformed Abuab/Pacol ratios) exhibiting ASE imbalance in BC₂ alone (blue), BC₃ alone (yellow), and both (red). Black points are genes exhibiting no ASE imbalance. Figure generated using ggplot2 (<https://ggplot2.tidyverse.org/>; [53]) in Rstudio [28].

4. Conclusions

Backcross hybrids (BC₂ and BC₃) from *M. textilis* (abaca var. Abuab) × *M. balbisiana* (wild banana var. Pacol) were generated to identify cis and/or trans regulatory divergence between the parents. We, likewise, interrogated whether there exists asymmetric expression at sites heterozygous to these backcross hybrid lines. Abuab has high FQ but has low resistance against AbBTV; Pacol, has low FQ but high resistance against AbBTV. Previous studies showed that their backcrosses possess both desirable traits. Using RNA-seq, results indicated that both backcrosses exhibited genome-wide preferential expression of Pacol- over Abuab-specific alleles, despite the wider genetic presence of the latter in the hybrids: 87.5% and 93.75% in BC₂ and BC₃, respectively. We call such observation as “genome-transcriptome incongruence” in which the ratio of the two genotype-specific expressed transcripts and the ratio of their corresponding genetic make-up are significantly disproportionate. Further analysis showed that regulatory divergence of Abuab and Pacol is largely explained by cis differences with 27.4% and 22.3% if we are to assess it using BC₂ and BC₃, respectively. The evolutionary divergence of the two interspecifically divergent genotypes is mostly explained by cis differences and we speculate that it would be higher than these proportions if we are to estimate it using F1. Because of the loss of heterozygosity (or increase in homozygosity) in the BC₃ relative to BC₂ genome, cis differences between the two backcrosses are significantly reduced from BC₂ to BC₃. This casts clues that cis variations are significantly affected by changes in allelic composition. Trans, on the other hand, is unaffected to changes in allelic composition. Taken all these together, we have provided preliminary findings consistent with literature in which hybrids are endowed with transcriptional versatility which confer them with the desirable traits, in this case, high FQ and resistance against AbBTV.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Figure S1: MA plot between Abuab- and Pacol-specific alleles in BC₂ or Bandala; Figure S2. MA plot between Abuab- and Pacol-specific alleles in BC₃; Table S1. Regulatory divergence between abaca and banana as estimated using BC₂; Table S2. Regulatory divergence between abaca and banana as estimated using BC₃; Table S3. Asymmetric expression in BC₂ using normalized read counts with binomial exact test of equal proportion; Table S4. GO enrichment analysis of allelically imbalanced genes in Abuab allele in BC₂; Table S5. Asymmetric expression in BC₃ using normalized read counts with binomial; Table S6. Relative transcript accumulation and preferential expression switching list of genes; Table S7. GO enrichment of genes allelically imbalanced in both BC₂ and BC₃.

Author Contributions: N.C.E. conceptualized the project, acquired partial funds from the Department of Science and Technology – Philippine Council for Agriculture, Aquatic and Natural Resources Research and Development (DOST-PCAARRD, Philippines), and performed laboratory activities, bioinformatics analysis, and writing of manuscript. A.G.L. (Antonio C. Laurena) undertook project management, sought additional funding for laboratory supplies, and hosting of the Balik Scientist. A.G.L. (Antonio G. Lalusin) undertook creation of the backcrosses, management, and maintenance of the abaca varieties. All authors have read and agreed to the published version of the manuscript.

Funding: This project has been supported by the Abaca Genomics project (N9-197-21) and Balik Scientist Program (N601600 and N602000), both funded by DOST-PCAARRD. Likewise, the APC was funded by DOST-PCAARRD.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: All datasets have been uploaded in EMBL-EBI ArrayExpress (www.ebi.ac.uk/arrayexpress) with an assigned accession number: E-MTAB-10990.

Acknowledgments: NCE would like to thank DOST-PCAARRD for providing partial funds for lab supplies and conferring a Balik Scientist award. We thank all the staff of the Philippine Genome Center for Agriculture, Biochem Lab, Plant Physio, and Genetics Lab of IPB for additional support on administrative and lab activities.

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

References

1. PhilFIDA. Philippine Fiber Industry Development Authority. Available online: <http://www.philfida.da.gov.ph/index.php/archived-articles/19-philippine-abaca-helps-in-global-environment-conservation> (accessed on 14 September 2021).
2. Sinha, A. K.; Narang, H. K.; Bhattacharya, S. Mechanical properties of natural fibre polymer composites, *J. Polym. Eng.* 2017, 37(9), 879–895. <https://doi.org/10.1515/polyeng-2016-0362>
3. Ocampo K. R. Demand rises for Philippine abaca as raw material for PPE. Inquirer. Available online: shorturl.at/xIO04 (accessed on 10 May 2020).
4. D'Hont, A.; Denoeud, F.; Aury, J.M.; Baurens, F.C.; Carreel, F.; Garsmeur, O.; Noel, B.; Bocs, S.; Droc, G.; Rouard, M.; et al. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 2012, 488, 213–216.
5. Wang, Z.; Miao, H.; Liu, J.; Xu, B.; Yao, X.; Xu, C.; Zhao, S.; Fang, X.; Jia, C.; Wang, J.; et al. *Musa balbisiana* genome reveals subgenome evolution and functional divergence. *Nat. Plants* 2019, 5, 810–821. <https://doi.org/10.1038/s41477-019-0452-6>
6. Galvez, L.C.; Koh, R.B.L.; Barbosa, C.F.C.; Asunto, J.C.; Catalla, J.L.; Atienza, R.G.; Costales, K.T.; Aquino, V.M.; Zhang, D. Sequencing and de Novo Assembly of Abaca (*Musa textilis* Née) var. Abuab Genome. *Genes* 2021, 12, 1202. <https://doi.org/10.3390/genes12081202>
7. Yllano, O.; Diaz, G.; Lalusin, A.; Laurena, A.; Mendoza, E. Genetic Analyses of Abaca (*Musa textilis* Née) Germplasm from its Primary Center of Origin, the Philippines, Using Simple Sequence Repeat (SSR) Markers. *Philipp. Agric. Sci.* 2020, 103, 311–321.
8. Tung J.; Fédrigo O.; Haygood R.; Mukherjee S.; Wray GA. Genomic features that predict allelic imbalance in humans suggest patterns of constraint on gene expression variation. *Mol. Biol. Evol.* 2009, 26(9):2047–59. doi: 10.1093/molbev/msp113.
9. Castel, S.E.; Aguet, F.; Mohammadi, P.; et al. A vast resource of allelic expression data spanning human tissues. *Genome Biol.* 2020, 21, 234. <https://doi.org/10.1186/s13059-020-02122-z>
10. Wittkopp, P.J.; Haerum, B.K.; Clark, A.G. Evolutionary changes in cis and trans gene regulation. *Nat.* 2004, 430, 85–88. <https://doi.org/10.1038/nature02698>
11. Wittkopp, P.J.; Haerum, B.K.; Clark A.G. Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat. Genet.* 2008, 40(3), 346–350 <https://doi.org/10.1038/ng.77>.
12. McManus, C.J.; et al. Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Res.* 2010, 20, 816–825. <https://doi.org/10.1101/gr.102491.109>.
13. Springer N.M.; Stupar R.M. Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize. *Plant Cell.* 2007, 19(8):2391–2402. <https://doi.org/10.1105/tpc.107.052258>.
14. Verta, J.P.; Jones, F.C. Predominance of cis-regulatory changes in parallel expression divergence of sticklebacks. *eLife.* 2019, 8, e43785. <https://doi.org/10.7554/eLife.43785> (2019).
15. Combes, M.C.; et al. Regulatory Divergence between Parental Alleles Determines Gene Expression Patterns in Hybrids. *GBE.* 2015, 7(4): 1110–1121. doi:10.1093/gbe/evv057 (2015).
16. Ereful, N.C.; Lalusin, A.G.; Laurena, A.C. RNA-Seq Reveals Differentially Expressed Genes Associated with High Fiber Quality in Abaca (*Musa textilis* Née). *Genes.* 2022, 13, 519. <https://doi.org/10.3390/genes13030519>.
17. Ereful, N.C.; Laurena, A.; Liu, L.Y.; et al. Unraveling regulatory divergence, heterotic malleability, and allelic imbalance switching in rice due to drought stress. *Sci. Rep.* 2021, 11, 13489. <https://doi.org/10.1038/s41598-021-92938-x>
18. Cowles, C.; Hirschhorn, J.; Altshuler, D.; et al. Detection of regulatory variation in mouse genes. *Nat. Genet.* 2002, 32, 432–437. <https://doi.org/10.1038/ng992>
19. Wu, Q.; Ma, Z.; Qin, Y.; et al. Imbalanced expression of stay-green 1 alleles in banana AAB/ABB cultivars prevents high-temperature-induced green ripening as in AAA Cavendish fruit. *Postharvest Biol. Technol.* 2019, 158, 110980.
20. Kerwin R.E.; Sweigart A.L. Rampant Misexpression in a Mimulus (Monkeyflower) Introgression Line Caused by Hybrid Sterility, Not Regulatory Divergence. *Mol Biol Evol.* 2020, 37(7):2084–2098.
21. Meiklejohn C.D.; Coolon J.D.; Hartl D.L.; Wittkopp P.J. The roles of cis- and trans-regulation in the evolution of regulatory incompatibilities and sexually dimorphic gene expression. *Genome Res.* 2014, 24(1):84–95.
22. Nagase, H.; Mao, J.H.; Balmain, A. Allele-specific Hras Mutations and Genetic Alterations at Tumor Susceptibility Loci in Skin Carcinomas from Interspecific Hybrid Mice. *Cancer Research.* 2003, 63, 4849–4853.
23. Guerrero R.F.; Posto A.L.; Moyle L.C.; Hahn M.W. Genome-wide patterns of regulatory divergence revealed by introgression lines. *Evolution.* 2016, 70(3):696–706.
24. Biswas, S.; Agrawal, Y.N.; Mucyn, T.S.; Dang, J.L.; Jones, C.D. Biological Averaging in RNA-Seq. *arXiv.* 2013, arXiv:1309.0670.
25. Assefa, A.T.; Vandesompele, J.; Thas, O. On the utility of RNA sample pooling to optimize cost and statistical power in RNA sequencing experiments. *BMC Genom.* 2020, 21, 312. <https://doi.org/10.1186/s12864-020-6721-y>.
26. Miao G.; Qin Y.; Guo J.; Zhang Q.; Bao Y. Transcriptome characterization and expression profile of Coix lacryma-jobi L. in response to drought. *PLoS ONE.* 2021, 16(9): e0256875. <https://doi.org/10.1371/journal.pone.0256875>
27. Kumar, S.; Seem, K.; Kumar, S.; et al. RNA-seq analysis reveals the genes/pathways responsible for genetic plasticity of rice to varying environmental conditions on direct-sowing and transplanting. *Sci. Rep.* 2022, 12, 2241. <https://doi.org/10.1038/s41598>
28. RStudio Team. RStudio: Integrated Development for R; RStudio, PBC: Boston, MA, USA, 2020. <https://www.rstudio.com/about>
29. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. Available online: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (accessed on 12 September 2021)

30. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 2013, 29, 15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
31. Xu, L.; Dong, Z.; Fang, L.; et al. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species, *Nucleic Acids Research*. 2019, 47 (W1): W52–W58, <https://doi.org/10.1093/nar/gkz333>.
32. He F.; Arce A.; Schmitz G.; Beyer A.; Koornneef M.; Novikova P.; de Meaux J. The footprint of polygenic adaptation on stress-responsive cis-regulatory divergence in the Arabidopsis genus. *Mol. Bio. Evol.* 2016, 33(8):2088–101.
33. Lovell, J.T.; Jenkins, J.; Lowry, D.B. et al. The genomic landscape of molecular responses to natural drought stress in *Panicum hallii*. *Nat. Commun.* 2018, 9, 5213. <https://doi.org/10.1038/s41467-018-07669-x>
34. Liao, Y.; Smyth, G.K.; Shi, W. Feature Counts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014, 30, 923–930. <https://doi.org/10.1093/bioinformatics/btt656>.
35. Robinson, M.D.; Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*. 2010, 11, R25.
36. Tirosh I.; Reikhav S.; Levy A. A.; Barkai N. A yeast hybrid provides insight into the evolution of gene expression regulation. *Science*. 2009. 324:659–662
37. Shi, X.; Ng, D. W. K.; Zhang, C.; Comai, L.; Ye, W.; Chen ZJ. Cis- and trans-regulatory divergence between progenitor species determines gene-expression novelty in Arabidopsis allopolyploids. *Nat. Commun.* 2012, 3, 950. <https://doi.org/10.1038/ncomms1954>
38. Bell, G. D.; Kane, N. C.; Rieseberg, L. H.; Adams, K. L. RNA-seq analysis of allele-specific expression, hybrid effects, and regulatory divergence in hybrids compared with their parents from natural populations. *GBE*. 2013, 5(7), 1309–1323.
39. Grote S. GOfuncR: Gene ontology enrichment using FUNC. R package version. 2021. 1.14.0.
40. Parac, E.P.; Lalusin, A.G.; Pangga, I.B.; Sta. Cruz, F.C. Characteristics of Selected Hybrids of Abaca (*Musa textilis* Nee) with Resistance to Bunchy Top. *Philipp. Agric. Scientist*. 2020, 103(1):1–12.
41. Lalusin, A.G.; Villavicencio, M.L.H. Abaca (*Musa textilis* Nee) Breeding in the Philippines. In *Industrial Crops; Handbook of Plant Breeding*; Springer: New York, NY, USA, 2015, 265–289.
42. Labrador, D.A.; Lalusin, A.G.; Mendoza, M.R.R.; dela Viña, C.B. Morphological Characterization and Karyotype Analysis of Abaca (*Musa textilis* Nee) and its Hybrids with *Musa balbisiana* Colla. *Philipp. Agric. Scientist*. 2020, 103(1): 13–28.
43. CFC; UNIDO; FAO; FIDA. Abaca Improvement of Fiber Extraction and Identification of Higher Yielding Varieties. Final Technical Report CFC/FIGHF/09. Activities in the Philippines. 2004. Available online: <https://www.yumpu.com/en/document/view/27575439/abaca-activities-in-the-philippines-unido> (accessed on 2 September 2021)
44. Ereful, N.C.; Liu, LY.; Tsai, E.; et al. Analysis of Allelic Imbalance in Rice Hybrids Under Water Stress and Association of Asymmetrically Expressed Genes with Drought-Response QTLs. *Rice* 2016, 9, 50. <https://doi.org/10.1186/s12284-016-0123-4>
45. Naranjo S.; Smith J.D.; Artieri C.G.; Zhang M.; Zhou Y.; Palmer M.E.; et al. Dissecting the Genetic Basis of a Complex cis-Regulatory Adaptation. *PLoS Genet.* 2015, 11(12): e1005751. doi:10.1371/journal.pgen.1005751
46. Boguero, A.P.B.; Pardocho, M.A.L.; Mendoza, M.R.; Abustan, M.A.; Lalusin, A.G. Molecular Screening of Abaca (*Musa textilis* Nee). *Philipp. J. Crop. Sci.* 2016. 41(2): 13 – 19.
47. Lalusin, A.G. Revitalizing the Abaca Industry through S & T Interventions for Higher Crop Productivity Using High-Yielding and Bunchy Top-Resistant Abaca Hybrids; Terminal Report 2016–2020; Institute of Plant Breeding, College of Agriculture and Food Science, University of the Philippines: Laguna, Philippines, 2020.
48. Zhou, Y.; Zhu, J.; Tong, T.; Wang, J.; Lin, B.; Zhang, J. A statistical normalization method and differential expression analysis for RNA-seq data between different species. *BMC Bioinform.* 2019, 20, 163. <https://doi.org/10.1186/s12859-019-2745-1>.
49. Campbell, M.T.; et al. The genetic basis of cis-regulatory divergence between the subspecies of cultivated rice (*Oryza sativa*). 2019. Preprint at <https://doi.org/10.1101/511550> (accessed on 01 October 2021)
50. Osada, N.; Miyagi, R.; Takahashi, A. Cis- and Trans-regulatory Effects on Gene Expression in a Natural Population of *Drosophila melanogaster*. *Genetics*. 2017, 206(4): 2139–2148. <https://doi.org/10.1534/genetics.117.201>
51. Tirosh, I.; Reikhav, S.; Levy, A. A.; Barkai, N. A yeast hybrid provides insight into the evolution of gene expression regulation. *Science*. 2009, 324, 659–662.
52. Cubillos, F. A.; et al. Extensive cis-regulatory variation robust to environmental perturbation in Arabidopsis. *Plant Cell* 2014, 26, 4298–4310.
53. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. Springer, New York, USA, 2016. <https://ggplot2.tidyverse.org>.
54. Kawasaki T.; Koita H.; Nakatsubo T.; et al. Cinnamoyl-CoA reductase, a key enzyme in lignin biosynthesis, is an effector of small GTPase Rac in defense signaling in rice. *Proc Natl Acad Sci.* 2006, 103(1):230–235. doi:10.1073/pnas.0509875103
55. Schwarzenbach, H. Loss of Heterozygosity, Editor(s): Stanley Maloy, Kelly Hughes, *Brenner's Encyclopedia of Genetics* (Second Edition), Academic Press, 2013, 271–273, <https://doi.org/10.1016/B978-0-12-374984-0.00882-2>.