Pertanika J. Trop. Agric. Sci. 44 (4): 697 - 711 (2021)

PERTANIKA JOURNALS

TROPICAL AGRICULTURAL SCIENCE

Journal homepage: http://www.pertanika.upm.edu.my/

Comparative Genomics of *Copia* and *Gypsy* Retroelements in Three Banana Genomes: A, B, and S Genomes

Sigit Nur Pratama[1], Fenny Martha Dwivany[1,2,3] and Husna Nugrahapraja[1,3]*

[1]School of Life Sciences and Technology, Institut Teknologi Bandung, Jalan Ganesha No. 10, Bandung, West Java, 40132, Indonesia
[2]Bali International Research Center for Banana, Badung, Bali, 80361, Indonesia
[3]University Center of Excellence for Nutraceuticals, Bioscience and Biotechnology Research Center, Institut Teknologi Bandung, Jalan Ganesha No. 10, Bandung, West Java, 40132, Indonesia

ABSTRACT

In plants, the proportion of transposable elements (TEs) is generally dominated by long terminal repeat (LTR) retroelements. Therefore, it significantly impacts on genome expansion and genetic and phenotypic variation, namely *Copia* and *Gypsy*. Despite such contribution, TEs characterisation in an important crop such as banana [*Musa balbisiana* (B genome), *Musa acuminata* (A genome), and *Musa schizocarpa* (S genome)] remains poorly understood. This study aimed to compare B, A, and S genomes based on repetitive element proportions and copy numbers and determine the evolutionary relationship of LTR using phylogenetic analysis of the reverse transcriptase (RT) domain. Genome assemblies were acquired from the Banana Genome Hub (banana-genome-hub.southgreen.fr). Repetitive elements were masked by RepeatMasker 4.0.9 before Perl parsing. Phylograms were constructed according to domain analysis using DANTE (Domain-based ANnotation of Transposable Elements), alignments were made using MAFFT 7 (multiple alignments using fast Fourier transform), and trees were inferred using FastTree 2. The trees were inspected using SeaView 4 and visualised with FigTree 1.4.4. We reported that B, A, and S genomes are composed of repetitive elements with 19.38%, 20.78%, and 25.96%, respectively. The elements were identified with dominant proportions in the genome are LTR, in which *Copia* is more abundant than *Gypsy*. Based on RT phylogenetic analysis, LTR elements are

ARTICLE INFO

Article history:
Received: 3 April 2021
Accepted: 11 June 2021
Published: 4 October 2021

DOI: https://doi.org/10.47836/pjtas.44.4.01

E-mail addresses:
sigitnurpratamashift@gmail.com (Sigit Nur Pratama)
fenny@sith.itb.ac.id (Fenny Martha Dwivany)
nugrahapraja@sith.itb.ac.id (Husna Nugrahapraja)
*Corresponding author

ISSN: 1511-3701
e-ISSN: 2231-8542

clustered into 13 ancient lineages in which *Sire* (*Copia*) and *Reina* (*Gypsy*) are shown to be the most abundant LTR lineages in bananas.

Keywords: Banana, B, A, and S genomes, reverse transcriptase, transposable elements

INTRODUCTION

Banana (*Musa* spp.) is one of the most consumed fruits and staple food in many countries across Asia and Africa (Food and Agriculture Organization [FAO], 2019). Its diversity is represented by the number of cultivars and genome diversity (A, B, S, and T genome) (D'Hont et al., 2000). A, B, and S genomes are publicly available in Banana Genome Hub and GenBank. Hence it is possible to characterise their genome organisation by observing the repetitive elements and transposable elements (TEs). Defined as stretches of DNA that are competent to integrate into new positions in the genome, TEs are competent to increase their copy number over time, and that rely on one or more enzymatic function provided by an autonomous element (Lisch, 2013).

To date, comprehensive genome analysis remains limited to two genome assemblies (*M. acuminata* and *M. balbisiana*) (D'Hont et al., 2012; Davey et al., 2013) despite two recent whole-genome sequencings (WGSs) of *Musa itinerans* (Wu et al., 2016) and *M. schizocarpa* (Belser et al., 2018) have been accomplished. However, these genome data can be conducted into a comparative study of *M. acuminata* (A genome) and *M. balbisiana* (B genome), a promising study to learn the structure and character of a gene or a gene family. For example, Nugrahapraja et al. (2021) successfully identified and characterised the pectin methylesterase (PME) gene family among A and B genome bananas from the comparative study of the genome. On the other hand, the characterisation of repetitive elements can also be studied by comparing these genome data. The characterisation of repetitive elements within the bananas' chromosomal genome is relatively easier than other plant species owing to the size of the paradoxically small genome (1 C ~ 600 Mbp) (Doležel et al., 1994). Plants repetitive sequences make up a genome proportion of 20% in *Arabidopsis* and more than 80% in *Zea mays* (Kaul et al., 2000; Vitte et al., 2014), which is highly dominated by long terminal repeat (LTR).

LTR elements are an extensive group, and their immense diversity is further divided into an enormous number of families. In eukaryote, the families are grouped into two superfamilies: *Copia/Ty1* and *Gypsy/Ty3*, characterised by their terminal repeat on both ends, flanking the sequence (Wicker et al., 2007). Once thought of as 'junk DNA', TEs have been known to create a variety of alterations of genes expression and function. It leads to numerous studies to inquire how TEs have played a crucial part in plant genome dynamics. LTR elements (*Copia* and *Gypsy*) have a significant impact in contributing to flowering plants diversity, evolutionary adaptation, and genome expansion (Ragupathy et al., 2013). Through a process called exaptation

(Hoen & Bureau, 2015), TEs could be evolutionarily adapted as functional genes, such as *Fhy33/Far* (light-responsive genes), *Sleeper* (transcriptional regulator in plant development), and *Mustang* (transcriptional regulator) families (Joly-Lopez et al., 2016; Knip et al., 2013; Lin et al., 2007).

Considering banana's potential development and challenges as important crop species, notably the characterisation of chromosomal genome organisation, *in silico* analysis was performed to characterise the structure of repetitive elements. This study also aimed to dissect the LTR (*Copia* and *Gypsy*) phylogeny of three banana genomes: A, B, and S genomes. In the future, such research can be used in genome mapping, evolutionary studies, omics studies, and further depict the dynamic of transposable elements.

MATERIALS AND METHODS

Data Retrieval

Whole genome sequence (WGS) of A (*Musa acuminata* 'DH-Pahang'), B (*Musa balbisiana* 'Pisang Klutuk Wulung'), and S genomes (*Musa schizocarpa*) (Belser et al., 2018; D'Hont et al., 2012; Davey et al., 2013; Martin et al., 2016) were used in the study. Complete WGS of three genomes were downloaded directly from Banana Genome Hub through the download menu of genome_sequences (https://banana-genome-hub.southgreen.fr/species-list) (Droc et al., 2013).

Repeat Masking and Parsing

Fasta format of WGS then masked using RepeatMasker 4.0.9 (Smit et al., 2015) implemented in Perl 5.8.0. The data were aligned using RMBlast 2.9.0 while tandem repeats were analysed using TRF 4.9.0 (Benson, 1999). Dfam 3.0 protein database (Hubley et al., 2016) and RepBase v20181026 (Jurka et al., 2005) were used as a library to identify the repeats. As for RepBase library, one should contact the Genetic Information Research Institute (GIRI) to attain the non-commercial license. The dictionary of parsing was built using build_dictionary.pl against RM.out and genome. The results of RepeatMasker were parsed with one_code_to_find_them_all.pl using fuzzy matching (Bailly-Bechet et al., 2014). CSV (comma-separated values) files, which comprised LTR, transposons, elem_sorted, and copy number created from parsing, were visualised using Office 365.

Phylogeny Analysis

Phylogram was constructed by harnessing the reverse transcriptase (RT) conserved domain of transposable elements class I using DANTE (Novák et al., 2010). The domain opted for its conserved bases. Thus, it could be easily used to dissect the diversity of superfamilies or lineages. The RT detection used the algorithm of robust alignment from the LAST program. The alignment was performed against REXdb Viridiplantae 3.0, a database for plant repetitive elements (Neumann et al., 2019). Extracted domains were aligned by MAFFT

7 (fft-NS-i), a progressive fast Fourier transform alignment program (Katoh & Standley, 2013). Aligned operational taxonomical units (OTUs) were transformed into a tree using FastTree 2 (Price et al., 2010). The tree was constructed with GTR+CAT substitution model, tree search normal+NNI+SPR, JC joins evaluation, and Shimodaira-Hasegawa (SH) test (1000) as a bootstrap alternative (Shimodaira, 2002). The tree produced was manually inspected using SeaView 4 with a bootstrap threshold of 0.5 (Gouy et al., 2010). Finally, the edited tree was visualised and annotated using FigTree 1.4.4 (Rambaut, 2018).

RESULTS AND DISCUSSION

Results

Repeats Genome Proportion and Copy Number. The structure of the banana genome represents its total repeats, repetitive elements proportion, and copy number (Figure 1). Overall, repeats covered from three bananas were observed proportional to their genome size. Repetitive elements composed 19.38%, 20.78%, and 25.96% of B, A, and S genomes, respectively. The proportions of *Copia* and *Gypsy* account for 8.79% and 7.51% in B genome, 9.32% and 8.12% in A genome, and 12.66% and 9.70% in S genome, a significant fraction of the repetitive elements. On the other hand, the proportion of elements such as non-LTR retroelements [short interspersed retrotransposable element (SINE), long interspersed retrotransposable element (LINE)], DNA elements were relatively low with less than 2% of the genome. Copy number visualisation shows that tandem repeats are abundant, followed by LTRs (*Copia* and *Gypsy*), while other elements are less abundant.

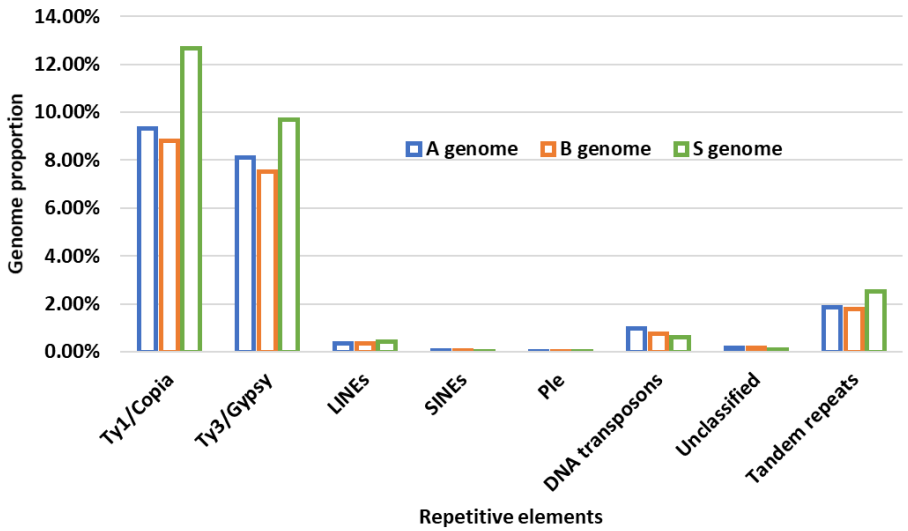


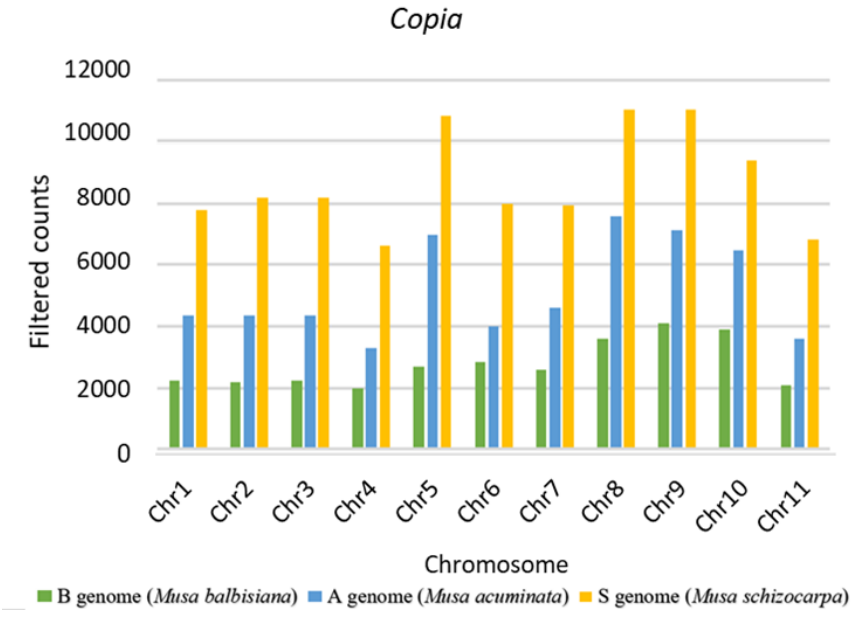
Figure 1. Proportion of repetitive elements in A, B, and S genome

Parsed Transposable Elements. As shown in Table 1, masked sequences analysis, was then parsed to produce a more explicit non-bias depiction of LTRs (*Copia* and *Gypsy*) copy number. Figure 2 shows a virtually similar trend in B, A, and S genomes that *Copia* is far more abundant than *Gypsy* in genomic and chromosomal levels as well. A glimpse at the copy number of *Copia* and *Gypsy* in the S genome illustrates an incredible abundance compared to two other genomes.

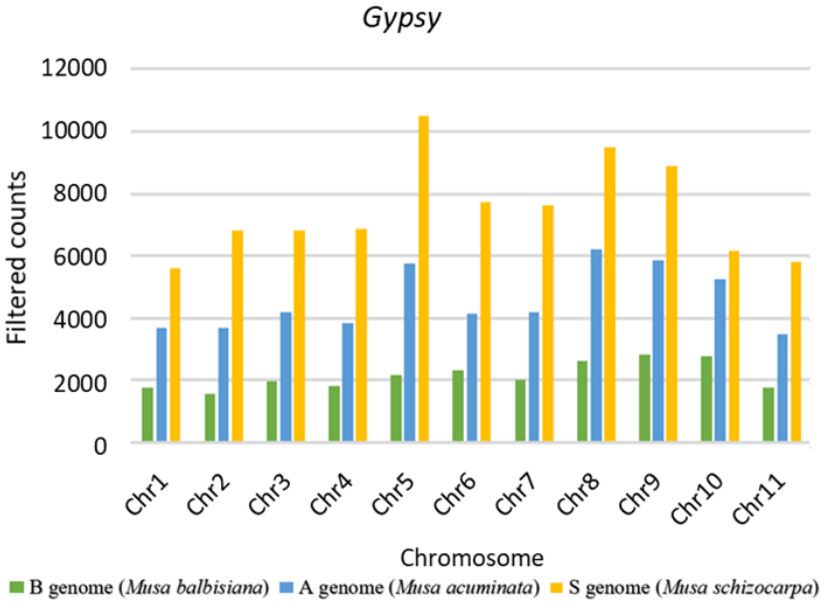
Table 1

*Copy number of LTRs (*Copia *and* Gypsy*)*

Species	LTRs copy number	
	Copia	*Gypsy*
A genome (*Musa acuminata*)	68857	59495
B genome (*Musa balbisiana*)	61070	46075
S genome (*Musa schizocarpa*)	95849	82317



(a)



(b)

Figure 2. Copy number of (a) *Copia* and (b) *Gypsy* in A, B, and S genomes described in total copies and chromosomal level

Ratio and divergence of parsed *Copia* and *Gypsy* elements were plotted as shown in Figure 3. Copies of the element are represented by blue dots, providing a general illustration of potentially full-length and active elements and those which have degraded over time. Provided that the blue dots are abundant, the ratio of elements is close to 1, possessing a low divergence. Thus, the ubiquity of *Copia* and *Gypsy* was inferred due to potentially active elements in which their bases were relatively less degraded. Figure 3 also shows an accordance trend as depicted in Figure 2 as more active and abundant *Copia* elements.

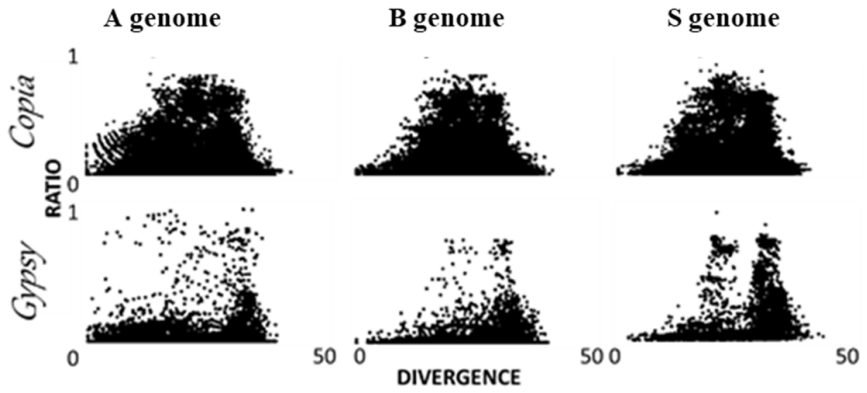


Figure 3. Parsed retroelements: divergence and ratio values of *Copia* and *Gypsy* in B, A, and S genomes

Figure 4 illustrates the ratio vs divergence of parsed elements compared from Chr1 of *M. balbisiana* to gain further perspective on how abundant and less abundant elements differ from each other regarding elements' activity. As seen from Figure 4, LINE and SINE are less abundant than LTR, such as *Copia*. On the other hand, DNA elements (*MuDR*, *Helitron*, and *hAT*) are relatively scarce. Therefore, they tend to be degraded (a ratio close to zero).

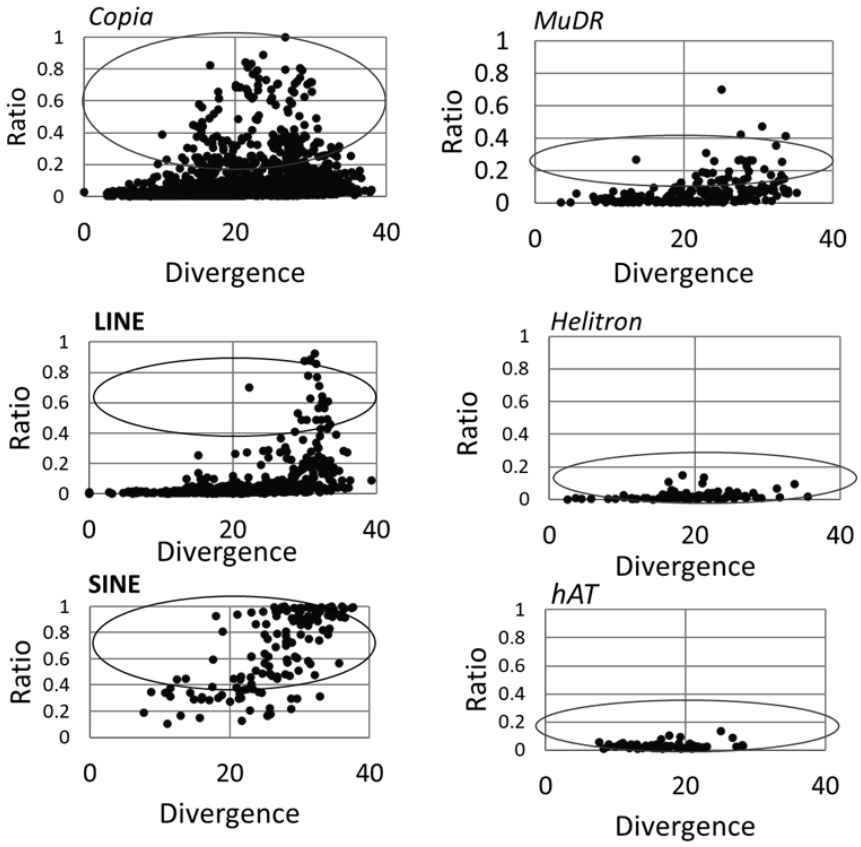


Figure 4. Ratio vs divergence comparison of *Copia*, class II elements (*MuDR*, *Helitron*, *hAT*), and non-LTRs (class LINE and SINE)

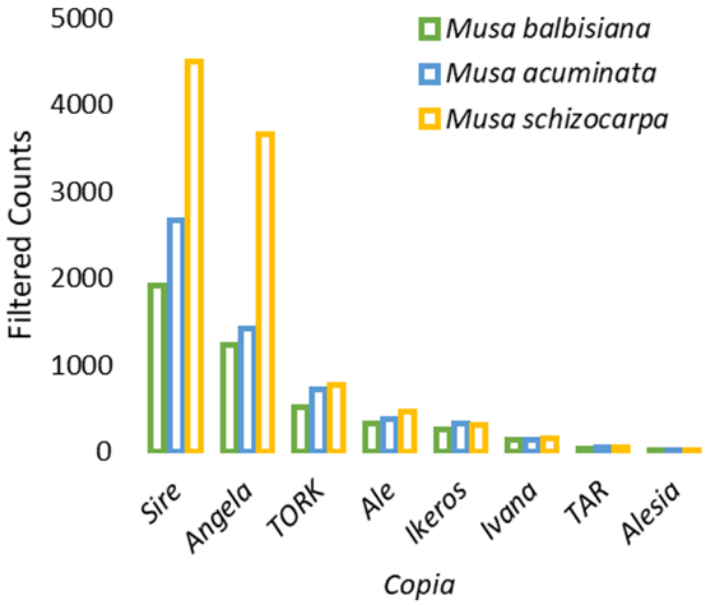
Reverse Transcriptase (RT) Domain Detection and Phylogenetic Analysis. *Copia* and *Gypsy* elements detected based on the RT domain are shown in Table 2. *Copia* lineages detected are *Ale*, *Alesia*, *Angela*, *Ikeros*, *Ivana*, *Sire*, *TAR*, and *Tork*. At the same time, *Gypsy* comprises *Galadriel*, *Tekay*, *Reina*, *CRM*, and *Retand*. The abundance of *Copia* and *Gypsy* lineages were further visualised as shown in Figure 5; the former is dominated by *Sire* lineages while the latter is shown mainly composed by *Reina* lineages.

Table 2

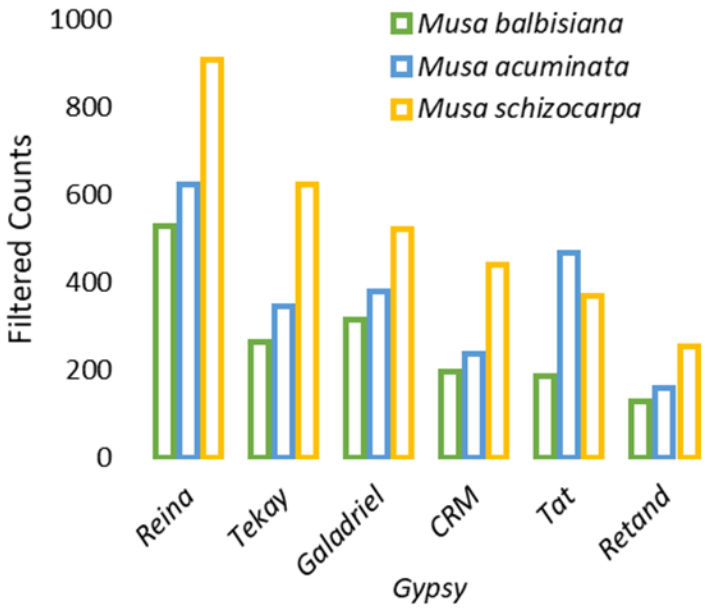
Hierarchical classification of Copia *and* Gypsy *in plants according to Neumann et al. (2019)*

Copia		Gypsy
*Ale**	Chromovirus	*Chlamyvir*
Alesia		*Tcn1*
*Angela**		*Galadriel**
Bianca		*Tekay**
Bryco		*Reina**
Lyco		*CRM**
Gymco-I, II, III, IV		*Chromo-unclass*
*Ikeros**	Non-chromovirus	*Phygy*
*Ivana**		*Selgy*
Osser		*OTA/Athila*
*Sire**		*OTA/TAT/Tat-I, II, III*
*TAR**		*Ogre*
*Tork**		*Retand**

Note. Asterisk (*): Identified elements based on RT domain: 13 elements in total comprising 8 *Copia* lineages and 5 *Gypsy* lineages



(a)



(b)

Figure 5. Filtered counts of (a) *Copia* and (b) *Gypsy* lineages in B genome (*Musa balbisiana*, green), A genome (*Musa acuminata*, blue), and S genome (*Musa schizocarpa*, yellow) showing numbers of non-bias

RT domains of *Copia* and *Gypsy* were constructed into inferred trees, as shown in Figure 6. Regardless of the type of genomes, both elements could be clustered into evolutionary lineages including two significant *Copia* and *Gypsy* superfamily clusters. *Copia* cluster encompassed two major lineages, designated *Sirevirus* clade and *Tork* clade. At the same time, *Gypsy* was divided into chromovirus and non-chromovirus. Major clades/lineages of *Copia* and *Gypsy* could be subdivided into several lineages as mentioned in filtered count results. Topologies acquired from individual genomes could be consistently inferred through a joint phylogram, as represented in Figure 6.

DISCUSSION

Stood at around 20%, the numbers of bananas' repeats proportions, including transposable elements (TEs) are less than the sister group in a subclass of Commelinids (the core of monocots) such as corn (*Zea mays*), wheat (*Triticum* sp.), and barley (*Hordeum vulgare*) with a proportion of more than 80% (Vitte & Panaud, 2005). Based on the size of the genome, with the S genome being the largest and the B genome being the smallest of all three, the proportions of TEs correlated with the size of the genome. The results are supported by the collinearity between regression analysis of various plants genomes against the proportion of TEs (Kidwell, 2002). This trend is similar to previous repeats calculations, although the results were

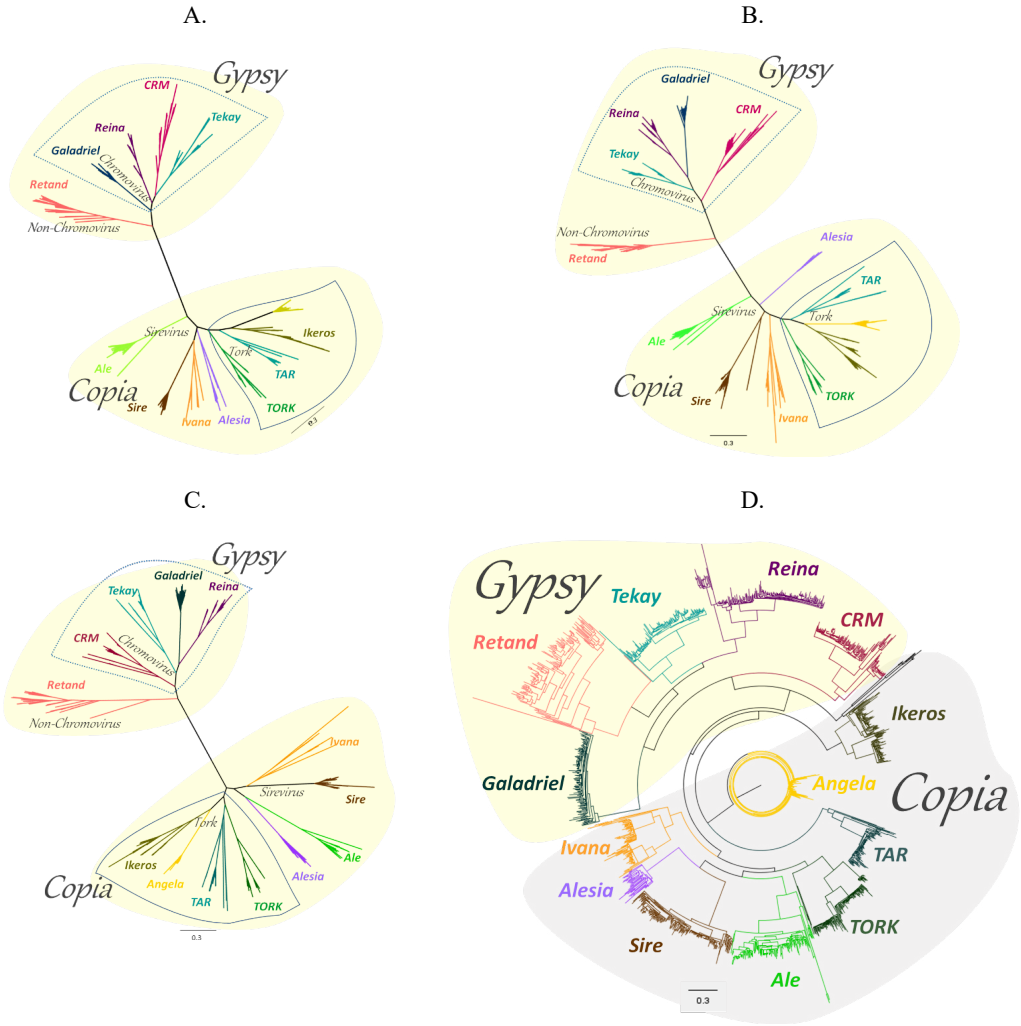


Figure 6. RT domain of *Copia* and *Gypsy* in banana phylogenetics with (A) *Musa acuminata* (AA), (B) *Musa balbisiana* (BB), (C) *Musa schizocarpa* (SS), and (D) joint phylogram of the three bananas

higher, making up about 30% of the genome (Hřibová et al., 2010). That said, TEs, particularly LTR retroelements (*Copia* and *Gypsy*), could transpose and contribute to genome expansion. At the same time, LTRs' underlying mechanism and contributions in affecting epigenetic and phenotype in bananas need further inquiry.

The dominance and prominent copy number of *Copia* and *Gypsy* in plants are not excluded in bananas. As elaborated before, the ubiquity of *Copia* and *Gypsy* resulted from the copy and paste transposition action, actively multiplying through intermediary RNA (Wicker et al., 2007). In other types of plants, *Gypsy* could outnumber *Copia*

in terms of proportion or copy number, for example, in Poales, such as *Oryza* spp. (Zhang & Gao, 2017). In the case of bananas, the abundance of *Gypsy* that somewhat lower compared to *Copia* could be understood due to its element association with spatial distributions: in the banana's genome, *Gypsy* is scattered broadly in heterochromatin, thus hampering the transcription cues from accessing the sequences in transposition (Domingues et al., 2012).

The activity of retroelements transposition could be described by their ratios close to 1. At the same time, divergences are close to zero (Bailly-Bechet et al., 2014). By contrast, DNA elements' cut and paste mechanism were shown less abundant and prone to degrade. In the meantime, non-LTR retroelements were less active to multiply. Although regarded as non-autonomous retroelements, SINEs maintain the transposition through which LINE transposition machinery enzymes are involved. The mechanism of LINE-dependent SINE transposition is also facilitated by SINE ability in recruiting RNA Pol III while LINE depends on RNA Pol II; the ratios and divergences of LINE were encountered similar to SINE patterns (Dewannieux et al., 2003).

Protein coding structures identified from domain searching were classified to chromodomain (CHD), endonuclease (ENDO), GAG, integrase (INT), protease (PROT), ribonuclease H (RH), reverse transcriptase (RT), transposase (TPase), and archeal ribonuclease H (aRH) according to REXdb (Neumann et al., 2019). Not only did we found the RT domain belonged to *Copia* and *Gypsy*, but *Parsaretrovirus/ Caulimovirus* RT within the genome described as endogenous banana streak virus (eBSV) was identified (Chabannes et al., 2013). In terms of *Copia* and *Gypsy* lineages abundance, *Sire* and *Reina* were consistently found the most ubiquitous LTR element in three genomes of the B, A, and S genomes, while other plants might differ. For instance, *Angela/Tork* and *TAT/Athila* are found prominent in paraphyletic monocots group of rice (*Oryza* spp.), *Sorghum* spp., foxtail millet (*Setaria italica*), and sugarcane (*Saccharum* spp.) (Du et al., 2010) while particular lineage such as *CRM* is absent in *Saccharum officinarum* (Domingues et al., 2012). As an addition to REXdb, a variety of lineages were grouped based on some classifications comprised GyDB (*Gypsy* Database) (Llorens et al., 2009) and unified classification (Wicker & Keller, 2007).

CONCLUSION

This study unravelled the bananas genome's overall composition, focusing on repetitive elements proportion among three genomes, *Copia* and *Gypsy* potential characteristics and their phylogeny. *Copia* and *Gypsy* were inferred as potentially active and full-length elements. Their phylogeny was illustrated in several lineages in which *Sire* and *Reina* account for a significant percentage of LTR ubiquity.

ACKNOWLEDGEMENTS

We want to acknowledge the Ministry of Research and Technology (RISTEK) Indonesia for funding this project (Competitive Research Grant from RISTEK/BRIN, Grant Number: 002/AMD/E1/KP.PTNBH/2020 for F. M. Dwivany and H. Nugrahapraja), the Banana Group - Institut Teknologi Bandung for support during this study, and Ms. Cindy Novianti for proofreading support.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

Bailly-Bechet, M., Haudry, A., & Lerat, E. (2014). "One code to find them all": A perl tool to conveniently parse RepeatMasker output files. *Mobile DNA*, *5*(1), 13. https://doi.org/10.1186/1759-8753-5-13

Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F. C., Falentin, C., Genete, M., Berrabah, W., Chèvre, A. M., Delourme, R., & Deniot, G. (2018). Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nature Plants*, *4*(11), 879–887. https://doi.org/10.1038/s41477-018-0289-4

Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research*, *27*(2), 573-580. https://doi.org/10.1093/nar/27.2.573

Chabannes, M., Baurens, F.-C., Duroy, P.-O., Bocs, S., Vernerey, M.-S., Rodier-Goud, M., Barbe, V., Gayral, P., & Iskra-Caruana, M.-L. (2013). Three infectious viral species lying in wait in the banana genome. *Journal of Virology*, *87*(15), 8624–8637. https://doi.org/10.1128/jvi.00899-13

D'Hont, A., Paget-Goy, A., Escoute, J., & Garreel, F. (2000). The interspecific genome structure of cultivated banana, *Musa* spp. revealed by genome DNA *in situ* hybridization. *Theoretical and Applied Genetics*, *100*(2), 177–183. https://doi.org/10.1007/s001220050024

D'Hont, A., Denoeud, F., Aury, J. M., Baurens, F. C., Carreel, F., Garsmeur, O., Noel, B., Bocs, S., Droc, G., Rouard, M., Da Silva, C., Jabbari, K., Cardi, C., Poulain, J., Souquet, M., Labadie, K., Jourda, C., Lengellé, J., Rodier-Goud, M., … Wincker, P. (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature*, *488*(7410), 213–217. https://doi.org/10.1038/nature11241

Davey, M. W., Gudimella, R., Harikrishna, J. A., Sin, L. W., Khalid, N., & Keulemans, J. (2013). A draft *Musa balbisiana* genome sequence for molecular genetics in polyploid, inter- and intra-specific *Musa* hybrids. *BMC Genomics*, *14*(1), 683. https://doi.org/10.1186/1471-2164-14-683

Dewannieux, M., Esnault, C., & Heidmann, T. (2003). LINE-mediated retrotransposition of marked *Alu* sequences. *Nature Genetics*, *35*(1), 41-48. https://doi.org/10.1038/ng1223

Doležel, J., Doleželová, M., & Novák, F. J. (1994). Flow cytometric estimation of nuclear DNA amount in diploid bananas (*Musa acuminata* and *M. balbisiana*). *Biologia Plantarum*, *36*(3), 351. https://doi.org/10.1007/BF02920930

Domingues, D. S., Cruz, G. M. Q., Metcalfe, C. J., Nogueira, F. T. S., Vicentini, R., de S Alves, C., & van Sluys, M. A. (2012). Analysis of plant LTR-retrotransposons at the fine-scale family level reveals individual molecular patterns. *BMC Genomics*, *13*(1), 137. https://doi.org/10.1186/1471-2164-13-137

Droc, G., Larivière, D., Guignon, V., Yahiaoui, N., This, D., Garsmeur, O., Dereeper, A., Hamelin, C., Argout, X., Dufayard, J.-F., Lengelle, J., Baurens F.-C., Cenci, A., Pitollat, B., D'Hont, A.,

Ruiz, M., Rouard, M., & Bocs, S. (2013). The Banana Genome Hub. *Database*, *2013*, bat035. https://doi.org/10.1093/database/bat035

Du, J., Tian, Z., Bowen, N. J., Schmutz, J., Shoemaker, R. C., & Ma, J. (2010). Bifurcation and enhancement of autonomous-nonautonomous retrotransposon partnership through LTR swapping in soybean. *Plant Cell*, *22*(1), 48–61. https://doi.org/10.1105/tpc.109.068775

Food and Agriculture Organization. (2019). *FAOSTAT: Crops*. http://www.fao.org/faostat/en/#data

Gouy, M., Guindon, S., & Gascuel, O. (2010). Sea view version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, *27*(2), 221–224. https://doi.org/10.1093/molbev/msp259

Hoen, D. R., & Bureau, T. E. (2015). Discovery of novel genes derived from transposable elements using integrative genomic analysis. *Molecular Biology and Evolution*, *32*(6), 1487–1506. https://doi.org/10.1093/molbev/msv042

Hřibová, E., Neumann, P., Matsumoto, T., Roux, N., Macas, J., & Doležel, J. (2010). Repetitive part of the banana (*Musa acuminata*) genome investigated by low-depth 454 sequencing. *BMC Plant Biology*, *10*(1), 204. https://doi.org/10.1186/1471-2229-10-204

Hubley, R., Finn, R. D., Clements, J., Eddy, S. R., Jones, T. A., Bao, W., Smit, A. F. A., & Wheeler, T. J. (2016). The Dfam database of repetitive DNA families. *Nucleic Acids Research*, *44*(D1), D81-D89. https://doi.org/10.1093/nar/gkv1272

Joly-Lopez, Z., Hoen, D. R., Blanchette, M., & Bureau, T. E. (2016). Phylogenetic and genomic analyses resolve the origin of important plant genes derived from transposable elements. *Molecular Biology and Evolution*, *33*(8), 1937–1956. https://doi.org/10.1093/molbev/msw067

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., & Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research*, *110*(1-4), 462-467. https://doi.org/10.1159/000084979

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kaul, S., Koo, H. L., Jenkins, J., Rizzo, M., Rooney, T., Tallon, L. J., Feldblyum, T., Nierman, W., Benito, M. I., Lin, X., Town, C. D., Venter, J. C., Fraser, C. M., Tabata, S., Nakamura, Y., Kaneko, T., Sato, S., Asamizu, E., Kato, T., … Somerville, C. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, *408*(6814), 796-815. https://doi.org/10.1038/35048692

Kidwell, M. G. (2002). Transposable elements and the evolution of genome size in eukaryotes. *Genetica*, *115*(1), 49-63. https://doi.org/10.1023/A:1016072014259

Knip, M., Hiemstra, S., Sietsma, A., Castelein, M., de Pater, S., & Hooykaas, P. (2013). DAYSLEEPER: A nuclear and vesicular-localized protein that is expressed in proliferating tissues. *BMC Plant Biology*, *13*(1), 211. https://doi.org/10.1186/1471-2229-13-211

Lin, R., Ding, L., Casola, C., Ripoll, D. R., Feschotte, C., & Wang, H. (2007). Transposase-derived transcription factors regulate light signaling in *Arabidopsis*. *Science*, *318*(5854), 1302-1305. https://doi.org/10.1126/science.1146281

Lisch, D. (2013). How important are transposons for plant evolution?. *Nature Reviews Genetics*, *14*(1), 49–61. https://doi.org/10.1038/nrg3374

Llorens, C., Muñoz-Pomer, A., Bernad, L., Botella, H., & Moya, A. (2009). Network dynamics

of eukaryotic LTR retroelements beyond phylogenetic trees. *Biology Direct*, *4*(1), 41. https://doi.org/10.1186/1745-6150-4-41

Martin, G., Baurens, F. C., Droc, G., Rouard, M., Cenci, A., Kilian, A., Hastie, A., Doležel, J., Aury, J.-M., Alberti, A., Carreel, F., & D'Hont, A. (2016). Improvement of the banana "*Musa acuminata*" reference sequence using NGS data and semi-automated bioinformatics methods. *BMC Genomics*, *17*(1), 243. https://doi.org/10.1186/s12864-016-2579-4

Neumann, P., Novák, P., Hoštáková, N., & MacAs, J. (2019). Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. *Mobile DNA*, *10*(1), 1. https://doi.org/10.1186/s13100-018-0144-1

Novák, P., Neumann, P., & Macas, J. (2010). Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics*, *11*(1), 378. https://doi.org/10.1186/1471-2105-11-378

Nugrahapraja, H., Putri, A. E., & Martha, D. F. (2021). Genome-wide identification and characterization of the pectin methylesterase (PME) and pectin methylesterase inhibitor (PMEI) gene family in the banana A-genome (*Musa acuminata*) and B-genome (*Musa balbisiana*). *Research Journal of Biotechnology*, *16*(2), 179–191.

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLOS One*, *5*(3), e9490. https://doi.org/10.1371/journal.pone.0009490

Ragupathy, R., You, F. M., & Cloutier, S. (2013). Arguments for standardizing transposable element annotation in plant genomes. *Trends in Plant Science*, *18*(7), 367-376. https://doi.org/10.1016/j.tplants.2013.03.005

Rambaut, A. (2018). *FigTree v. 1.4.4*. http://Tree.Bio.Ed.Ac.Uk/Software/Figtree/

Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. *Systematic Biology*, *51*(3), 492-508. https://doi.org/10.1080/10635150290069913

Smit, A., Hubley, R., & Grenn, P. (2015). *RepeatMasker Open-4.0.7*. http://www.repeatmasker.org/

Vitte, C., Fustier, M. A., Alix, K., & Tenaillon, M. I. (2014). The bright side of transposons in crop evolution. *Briefings in Functional Genomics and Proteomics*, *13*(4), 276–295. https://doi.org/10.1093/bfgp/elu002

Vitte, C., & Panaud, O. (2005). LTR retrotransposons and flowering plant genome size: Emergence of the increase/decrease model. *Cytogenetic and Genome Research*, *110*(1–4), 91–107. https://doi.org/10.1159/000084941

Wicker, T., & Keller, B. (2007). Genome-wide comparative analysis of *copia* retrotransposons in Triticeae, rice, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual *copia* families. *Genome Research*, *17*(7), 1072-1081. https://doi.org/10.1101/gr.6214107

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, *8*(12), 973-982. https://doi.org/10.1038/nrg2165

Wu, W., Yang, Y.-L., He, W.-M., Rouard, M., Li, W.-M., Xu, M., Roux, N., & Ge, X.-J. (2016). Whole genome sequencing of a banana wild relative *Musa itinerans* provides insights into lineage-specific diversification of the *Musa* genus. *Scientific Reports*, *6*(1), 31586. https://doi.org/10.1038/srep31586

Zhang, Q. J., & Gao, L. Z. (2017). Rapid and recent evolution of LTR retrotransposons drives rice genome evolution during the speciation of AA-genome *Oryza* species. *G3: Genes, Genomes, Genetics*, *7*(6), 1875–1885. https://doi.org/10.1534/g3.116.037572