CROP BREEDING AND APPLIED BIOTECHNOLOGY

ARTICLE

De novo transcriptome assembly and development of EST-SSR derived functional domain markers for *Pandanus amaryllifolius*

Norhafizah Sidek[1], Boon Chin Tan[1], Katharina Mebus[1], Sima Taheri[1] and Chee How Teo[1*]

Crop Breeding and Applied Biotechnology
24(4): e501424414, 2024
Brazilian Society of Plant Breeding.
Printed in Brazil
http://dx.doi.org/10.1590/1984-70332024v24n4a51

Abstract: *Pandanus amaryllifolius, a tropical fragrant screw pine, is commonly used in cooking and traditional medicine. Although studies have analysed its metabolite, transcriptome, and proteome profiles, molecular markers for genetic improvement remain scarce. This study aimed to develop EST-SSR markers from the P. amaryllifolius transcriptome and utilise them to analyse their genetic diversity. We obtained 157,467 unigenes with an average length of 1,084 base pairs. From these, 66,820 EST-SSRs were identified, with dinucleotides being the most common (45.5%). Following functional domain marker (FDM) analysis, 7,208 markers were used for in silico PCR, which revealed 8,843 polymorphic bands (86.9%), with an average polymorphic information content (PIC) of 0.41, gene diversity of 0.54, and a Shannon index of 0.91. The dendrogram and PCA plot clearly separated samples from Malaysia and India. These findings provide a new alternative for EST-SSR marker validation and a foundation for genetic improvement and conservation strategies for P. amaryllifolius.*

Keywords: Pandanus amaryllifolius*, expressed sequence tag (EST)-simple sequence repeats (SSR), RNA-seq, functional domain marker (FDM), in silico PCR*

INTRODUCTION

Pandanus amaryllifolius Roxb. (common name: pandan), a tropical plant in the screw pine family Pandanaceae, is commonly found in Southeast Asia, such as Thailand, Malaysia, Indonesia, and India (Amnan et al. 2023). It holds significant economic value in the region due to its extensive culinary, medicinal, and cultural use. The plants have no flowers and rely entirely on vegetative propagation for reproduction (Wakte et al. 2009). *P. amaryllifolius* is well-known for its exotic fragrance from its leaves, similar to Basmati-India, Jasmine-Thai, and Kaorimai-Japan rice (Wongpornchai 2006). Due to its distinct and pleasant aroma, pandan leaves are often used in cooking to give flavour and natural colour to foods. This aroma is due to presence of the 2-acetyl-1-pyrroline (2AP), which is the major chemical compound in the leaves (Yahya et al. 2011).

Over time, research on *P. amaryllifolius* has increasingly focused on its phytochemical compounds and pharmacological potential. Although there are reports on the metabolite (Sa'adah et al. 2023), transcriptome (Bhatt et al. 2021) and proteome (Amnan et al. 2022) profiles, little attention has been paid to the genetic variations within *P. amaryllifolius,* hindering its use in genetic and breeding studies. In the past, Wakte et al. (2012) utilised inter-simple sequence

***Corresponding author:**
E-mail: cheehow.teo@um.edu.my
ORCID: 0000-0002-8118-2592

Received: 15 July 2024
Accepted: 15 September 2024
Published: 31 October 2024

[1] Centre for Research in Biotechnology for Agriculture, Universiti Malaya, 50603 Kuala Lumpur, Malaysia

repeat (ISSR) and amplified fragment length polymorphism (AFLP) markers to evaluate the genetic diversity between *P. amaryllifolius* from India, Sri Lanka and Thailand. Their findings indicated that the genetic diversity within the Indian population was notably low. However, the genetic diversity of *P. amaryllifolius* in Malaysia remains unstudied due to difficulties in accessing samples.

Simple sequence repeats (SSRs) marker system has been widely utilised to study the genetic diversity in plants (Wang et al. 2018). SSRs are repetitive DNA sequences found in both protein-coding and non-coding regions of DNA sequences. They are abundant, highly polymorphic, multiallelic, codominant, reproducible, and transferable (Boopathi 2020). Nevertheless, generating SSR markers from genomic DNA is costly and ineffective, resulting in limited transferability of these markers. Therefore, an expressed sequence tag (EST)-SSR markers system is preferable, particularly in organisms that lack a sequence database (Cai et al. 2019, Azevedo et al. 2022). The development of simple sequence repeat-functional domain markers (SSR-FDM), which integrates SSR markers with predicted protein domains, allows for identifying molecular markers associated with putative functional domains (Basu and Bhattacharya 2022).

This study aimed to investigate the evolutionary origin of *P. amaryllifolius*. We first sequenced the transcripts using the RNA-seq platform as well as assembled and annotated the unigenes of *P. amaryllifolius*. The generated data were used to evaluate its genetic diversity. To enhance the robustness and comprehensiveness of our genetic diversity study, we integrated an *in silico* PCR approach to validate the developed EST-SSR FDMs. These transcriptome and genetic variation data could provide new perspectives into the evolutionary origin of *P. amaryllifolius* and serve as a valuable genetic resource for molecular breeding to develop improved cultivars.

MATERIAL AND METHODS

Three-month-old disease-free *Pandanus amaryllifolius* plants purchased from a nursery in Selangor and Johor, Malaysia, were grown and maintained in a growth room at the Universiti Malaya, Malaysia. Total RNAs were extracted

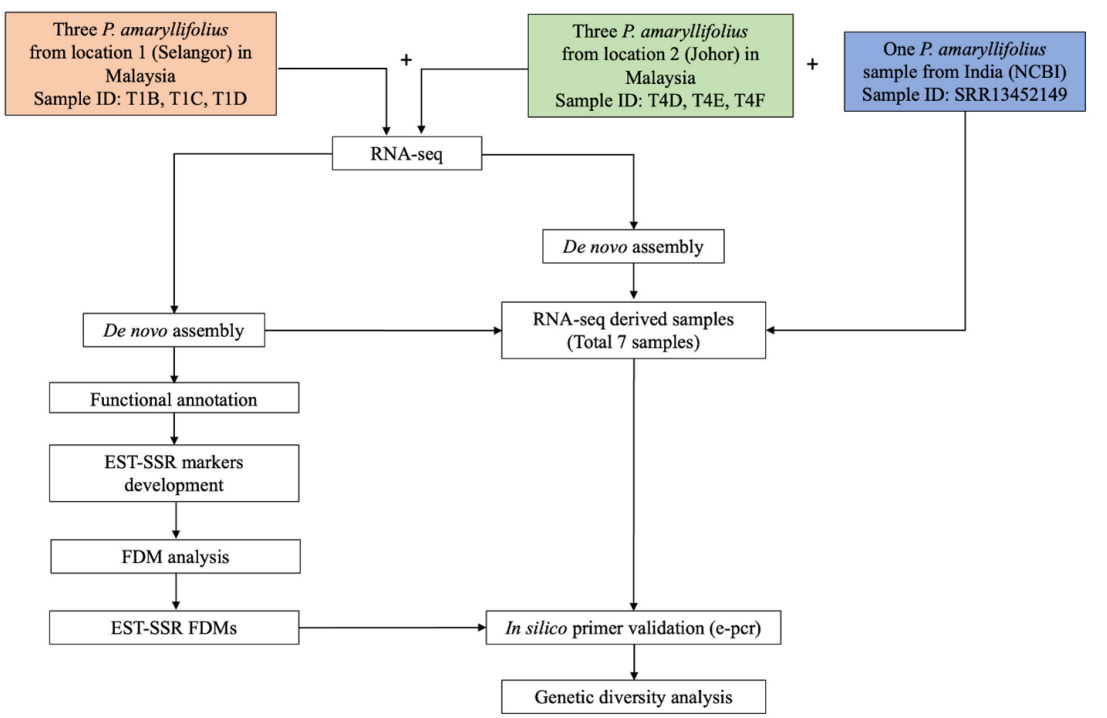


Figure 1. Flow chart showing steps involved in the genetic diversity analysis of *Pandanus amaryllifolius* using *in silico* PCR and RNA-seq-derived samples.

from the young leaves of six *P. amaryllifolius* plants using a modified CTAB method (Tattersall et al. 2005). The RNA samples were sent for sequencing at Beijing Genomics Institute (Shenzhen, China), utilising an Illumina HiSeq 4000 system. Figure 1 shows a flow chart outlining the steps in this study, from RNA-seq of the samples to the development of EST-SSR markers and genetic diversity analysis of *P. amaryllifolius*.

The Trinity program (v2.8.5) was used to assemble the clean reads to obtain the *P. amaryllifolius* transcriptome dataset (Grabherr et al. 2011). The TGICL pipeline (v2.0.6) was used to cluster and assemble the Trinity-assembled transcriptome to produce longer and unique contigs (Pertea et al. 2003). The functional annotation of the assembled unigenes was performed by searching transcripts against the NCBI non-redundant protein (Nr, https://www.ncbi.nlm.nih.gov/), Cluster of Orthologous Groups (COG, https://www.ncbi.nlm.nih.gov/research/cog-project/), Swiss-Prot (Boeckmann et al. 2003), and Gene Ontology (GO, http://geneontology.org/) databases. The search was conducted using BLASTx with an E-value cut-off of $1e^{-5}$ (Altschul et al. 1990). The KEGG metabolic pathway database (https://www.kegg.jp/blastkoala/) was used for assigning the molecular and higher-level functions to the *P. amaryllifolius* transcriptome dataset (Kanehisa et al. 2007). The functional classification of unigenes was performed using the WEGO2.0 (https://wego.genomics.cn/) (Zhang et al. 2018).

The detection of SSRs in unigenes was done using the Perl script MIcroSAtellite identification tool (MISA, http://pgrc.ipk-gatersleben.de/misa/misa.html) (Beier et al. 2017). The assembled transcriptome was examined for di-, tri-, tetra, penta-, and hexanucleotide repeat motifs. EST-SSR primers were designed using Primer 3 (http://bioinfo.ut.ee/primer3) (Untergasser et al. 2012). The sequences with SSRs were then examined for FDM using InterProScan (http://www.ebi.ac.uk/Tools/InterProScan/).

This study used six accessions of *P. amaryllifolius* from two different locations in Malaysia that were previously sent for RNA-seq and assembled using the Trinity pipeline (Grabherr et al. 2011). Additionally, we included one available *P. amaryllifolius* RNA-seq read from NCBI that was previously studied by Bhatt et al. (2021) (Figure 1). The mapping and transferability of the designed primers to six RNA-seq datasets from Malaysia and one published dataset from India were performed using an electronic polymerase chain reaction (ePCR) (Schuler 1997). A maximum 2 nucleotides mismatch with two gaps was set as a cut-off value for the ePCR result filter.

The NTSYSpc 2.11 V software package from Rohlf (2000) was used to analyse the EST-SSR marker binary matrices generated from *in silico* PCR. A dendrogram was generated using the unweighted pair group method with arithmetic mean (UPGMA) through the SAHN subprogram. Genetic differences between populations were measured using the genetic distance index (D), as described by Nei (1972).

Table 1. Summary of assembly results for *Pandanus amaryllifolius*

Features	
Total clean reads (Mb)	73.69
Total clean bases (Gb)	7.369
Q20 (%)	97.2
Total number of transcripts	236,058
Total length of transcripts (bp)	329,393,215
Mean length of transcripts (bp)	1,395.39
N50 value of transcripts	2,760
GC (%)	41.0
Total number of unigenes (bp)	157,467
Total length of unigenes (bp)	170,772,211
Mean length unigenes (bp)	1,084.50
N50 value of unigenes	2,514
GC (%)	41.0

RESULTS AND DISCUSSION

RNA-seq and *de novo* assembly are powerful tools for gene discovery and are effective methods for generating markers (Unamba et al. 2015). In this study, about 73,687,726 raw reads (7.369 Gb) were generated from sequencing, with the percentage of Q20 bases being 97.2%. The GC content was approximately 41.0%. After filtering, high-quality reads were assembled into 157,467 unigenes and 236,058 transcripts, reaching 170,772,211 bp and 329,393,215 bp in length, respectively (Table 1). Previously, Bhatt et al. (2021) reported 29.46 million reads, 34,672 transcripts

Table 2. Summary of functional annotation of unignenes of *Pandanus amaryllifolius*

Unigenes	Number	Percentage
Total unigenes	157,467	100.0
Nr-annotated	43,891	27.9
Swissprot-annotated	76,755	48.7
COG-annotated	40,807	25.9
TrEMBL-annotated	43,991	27.9
KEGG-annotated	17,209	10.9
GO-annotated	22,503	14.3
Overall	112,525	71.5

with an average transcript size of 802 bp in *P. amaryllifolius*, which is lower than our findings. Despite both assemblies being generated using the Trinity program (Grabherr et al. 2011), different sequencing depths might be the main factor of the assembly discrepancy.

Among the unigenes, 43,891 (27.9%), 76,755 (48.7%), 40,807 (25.9%), 43,991 (27.9%), 17,209 (10.9%) and 22,503

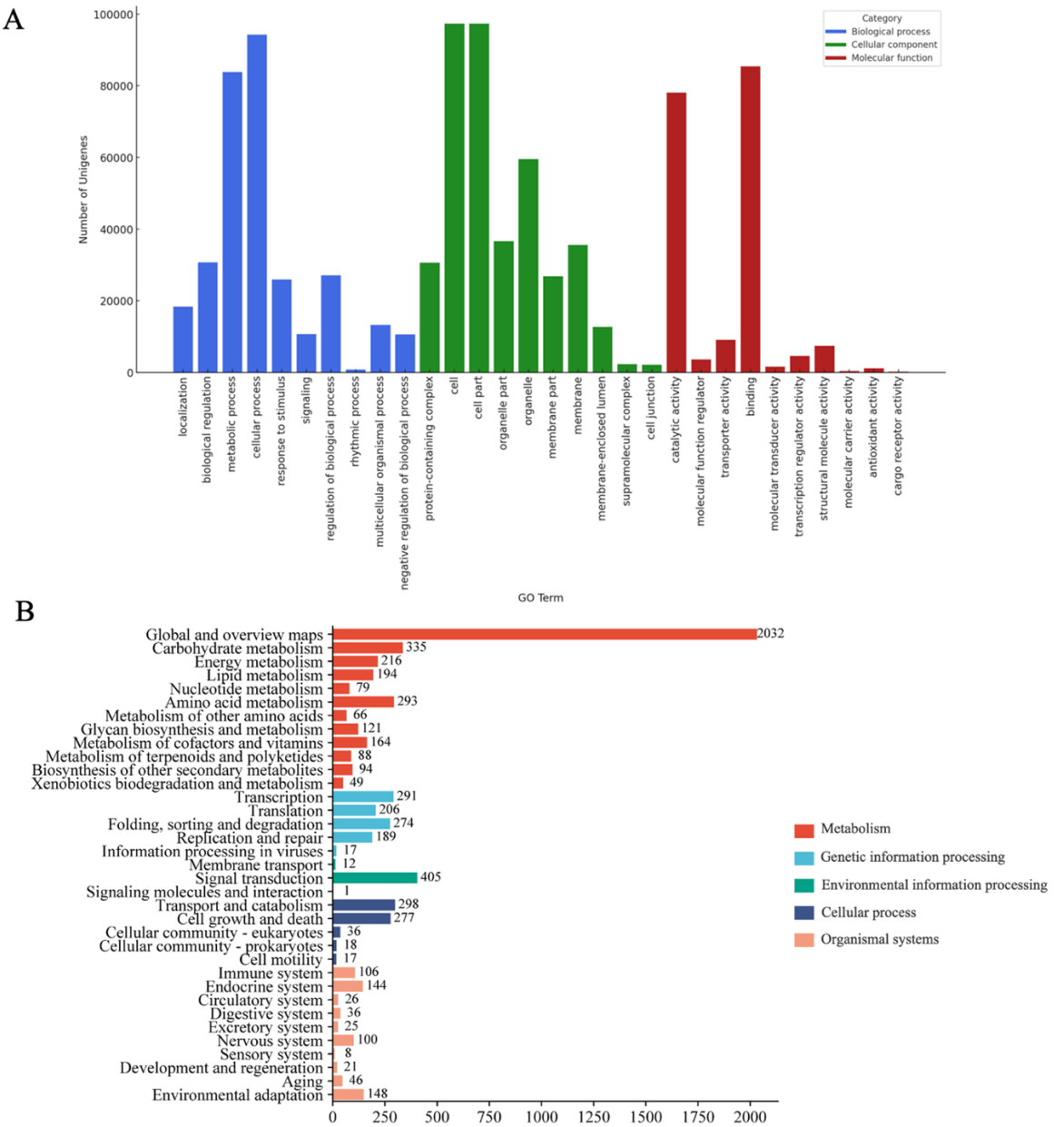


Figure 2. **A**) GO classification of the assembled unigenes of *Pandanus amaryllifolius*, classified into three categories: molecular function, cellular component, and biological process. **B**) Top 35 KEGG pathways associated with the assembled unigenes, categorised into five subgroups: metabolism, genetic information processing, environmental information processing, cellular processes, and organismal systems.

(14.3%) matched in the NR, SwissProt, COG, TrEMBL, KEGG, and GO databases, respectively. Overall, 112,525 (71.5%) unigenes were annotated in all six databases (Table 2), probably due to the substantial proportion of unigenes over 500 bp in length within the assembled transcripts. Generally, the longer unigenes were most likely to have BLAST matches in the protein databases (Parchman et al. 2010). However, there is still a large percentage of unannotated unigenes, which might be due to the lack of publicly available whole-genome sequencing data for *Pandanus* species.

Table 3. Summary of EST-SSRs identified from *Pandanus amaryllifolius*

Features	
Total number of sequences examined	157,467
Total size of examined sequences (bp)	170,772,211
Total number of identified SSRs	66,820
Number of SSR containing sequences	48,816
Number of sequences containing more than one SSR	13,054
Number of SSRs present in compound formation	6,596

The unigenes were grouped into five major categories and classified into 35 pathways (Figure 2A). The primary categories included metabolism, genetic information processing, environmental information processing, cellular processes, and organismal systems. The metabolism category was the most prominent, with "Global and overview maps" and "Carbohydrate metabolism" being the top pathways, reflecting the extensive involvement in energy production and biosynthetic pathways, similar to findings in other plants like *Rhodomyrtus tomentosa* and *Populus wulianensis* (Wu et al. 2020, Sun et al. 2023). The *P. amaryllifolius* unigenes can be annotated into 3 GO categories, which include 10 top subcategories for each GO main category (Figure 2B). A large number of unigenes were annotated in the cellular

Table 4. Summary of EST-SSRs identified from the unigenes of *Pandanus amaryllifolius*

Repeat motifs	Number of repeats								
	4	**5**	**6**	**7**	**8**	**9**	**>10**	**Total**	**Frequency** (%)
Mono-nucleotide									
A/T	-	-	-	-	-	-	20,516	20,516	30.7
C/G	-	-	-	-	-	-	1,304	1,304	2.0
								21,820	32.7
Di-nucleotide									
AG/CT	-	-	3,917	3,252	2,699	2,336	10,473	22,677	33.9
AT/AT	-	-	704	391	299	244	1,306	2,944	4.4
AC/GT	-	-	1,342	780	566	438	1,523	4,649	7.0
CG/CG	-	-	110	26	15	5	2	158	0.2
								30,428	45.5
Tri-nucleotide									
CCG/CGG	-	445	150	57	19	3	4	678	1.0
AGG/CCT	-	1,156	495	294	164	40	97	2,246	3.4
AAG/CTT	-	1,302	625	367	237	75	268	2,874	4.3
AGC/CTG	-	635	215	86	37	11	3	987	1.9
ACG/CGT	-	120	26	23	10	-	1	180	0.8
Others	-	2,560	1,096	569	370	70	179	4,844	7.3
								11,809	17.7
Quad-nucleotide									
AAAG/CTTT	-	71	46	3	5	-	-	128	2.0
AAAT/ATTT	-	167	60	3	1	1		233	0.4
AAGG/CCTT	-	5	8	2	-	-	-	15	0.02
AAAC/GTTT	-	14	17	-	-	-	-	31	0.05
Others		342	191	34	46	18	14	645	1.0
								1,052	1.6
Penta-nucleotide	684	130	19	6	3	2	3	847	1.8
Hexa-nucleotide	605	135	69	30	9	11	5	864	1.3
Total	1,289	7,080	9,082	5,923	4,480	3,254	35,712	66,820	100.0
Frequency (%)	2.0	11.0	14.0	9.0	7.0	5.0	53.0	100.0	

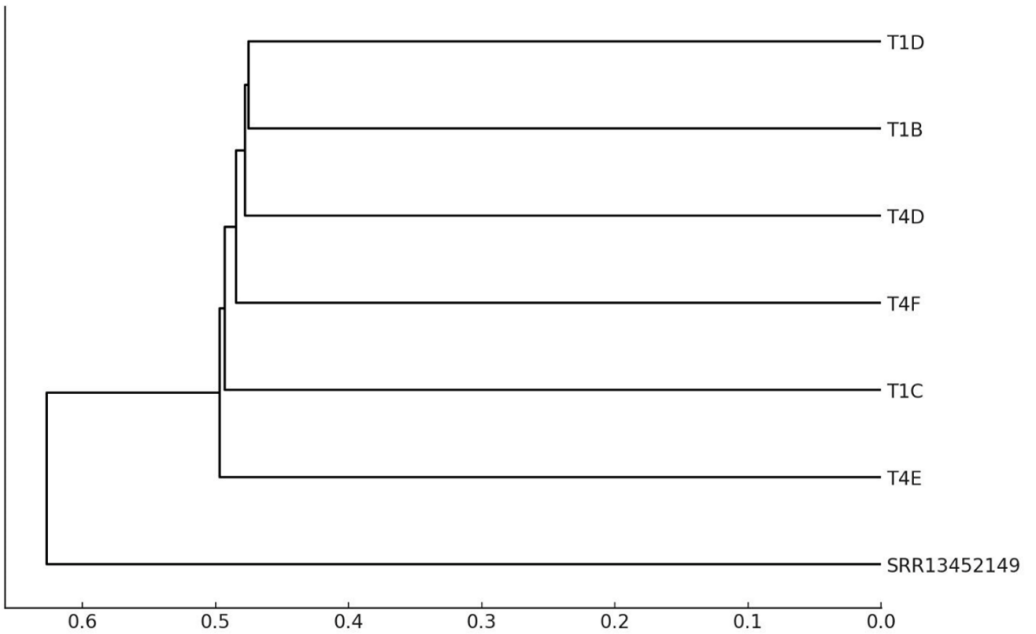


Figure 3. UPGMA dendrogram of six *Pandanus amaryllifolius* accessions and one accession from India based on EST-SSR markers.

component (CC) category. Most are represented by the terms cell, cell part, organelle, organelle part and membrane, which aligns with the structural and functional diversity required for cellular integrity and activity (Mathur 2021).

Approximately 66,820 unigenes encoded 48,816 potential EST-SSRs. Among these unigenes, 13,054 contain multiple SSRs, while 6,596 were identified to be present in compound formation (Table 3). A total of 170,772,211 kb of the sequenced data were analysed, revealing a frequency of at least one SSR every 2.6 kb within the expressed part of the *P. amaryllifolius* genome with 31% distribution frequency, which was remarkably higher than *Curcuma alismatifolia* (12.6%) (Taheri et al. 2019) and *Coriandrum sativum* (21.9%) (Tulsani et al. 2020), but lower than *Glycyrrhiza uralensis* (60.1%) (Liu et al. 2015).

The SSRs comprised 21,820 mononucleotide (32.7%), 30,428 dinucleotide (45.5%), 11,809 trinucleotide (17.7%), 1,052 quadnucleotide (1.6%), 847 pentanucleotide (1.3%), and 864 hexanucleotide (1.3%) motifs (Table 4). It was observed that dinucleotide motif was the most prevalent type. This was in agreement with other plant species, such as *Mentha piperita* (Kumar et al. 2015) and *Camellia sinensis* (Liu et al. 2018). The AG/CT dinucleotide repeat motif exhibited the highest prevalence among the other motif types, accounting for 33.9% of the total SSRs. The abundance of dinucleotide repeat sequences might be due to the UTRs' overexpression compared to open reading frames (Kumpatla and Mukhopadhyay 2005). Among the two types of mononucleotide repeat motifs, it was observed that the A/T motif exhibited the highest frequency, accounting for 30.7% of the total number of SSRs (Table 4). According to Gur-Arie et al. (2000), the presence of a high number of A/T bases implies a potential correlation between the phenomena and the ease of DNA melting in repeat sequences that are abundant in A/T bases.

A total of 7,208 developed EST-SSR FDM markers with 10,172 bands were identified to screen for polymorphisms among seven RNA-seq-derived samples (Supplementary File 1). Of these, 8,843 bands were polymorphic (86.9%). The average PIC was 0.41, which falls within the medium range according to the classification by Botstein et al. (1980). A total of 2,908 markers exhibited PIC >0.5, indicating high informativeness (Supplementary File 2). The PIC marker index has been widely used to assess the informativeness of EST-SSR in different cultivated species (Sharma et al. 2020, Lu et al. 2021).

The gene diversity was 0.54, and the Shannon index was 0.91 (Supplementary File 2). A dendrogram constructed from the polymorphic EST-SSR marker data distinctly separated the *P. amaryllifolius* samples from Malaysia from the one from India, with all six Malaysia samples clustering together (Figure 3). Based on the previous study (Wakte et al. 2012), a genetic diversity analysis of *P. amaryllifolius* using ISSR and AFLP markers on samples from India, Sri Lanka, and Thailand demonstrated diversity distinctions between geographical locations, further emphasising the impact of geographic isolation on genetic variation. However, the study by Wakte et al. (2012) concluded that *P. amaryllifolius* has low genetic diversity, as evidenced by their analysis showing only 1.6% polymorphism in AFLP markers and finding only one polymorphic marker using ISSR. This indicates a challenge when using conventional PCR to validate genetic markers. The low genetic diversity of *P. amaryllifolius* can be attributed to its vegetative propagation (Wakte et al. 2009). Due to this low genetic variation, it is difficult and laborious to detect the small differences in polymorphic bands using conventional PCR.

CONCLUSIONS

The *de novo* assembly of the *P. amaryllifolius* transcriptome using RNA-seq technology resulted in a substantial number of unigenes and EST-SSR primers. A total of 7,208 EST-SSR markers were developed after FDM analysis. Our analysis revealed significant genetic differentiation between samples from Malaysia and India, underscoring the impact of geographical isolation. However, the study was limited to a small sample size, highlighting the need for further research with more samples to fully understand the genetic diversity of this species. These findings provide a foundation for genetic improvement and conservation strategies for *P. amaryllifolius*, demonstrating the effectiveness of EST-SSR markers, RNA-seq-derived samples, and *in silico* PCR in uncovering genetic diversity.

ACKNOWLEDGEMENTS

This work was supported by the RU grant of Universiti Malaya (RU006-2019) and the MOHE-Top 100 grant (TU002B-2018). Supplementary files are available upon request to the corresponding author.

DATA AVAILABILITY

Transcriptome datasets supporting the conclusions of this article are available in NCBI (PRJNA908997).

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool**. Journal of Molecular Biology 215**: 403-410.

Amnan MAM, Aizat WM, Khaidizar FD and Tan BC (2022) Drought stress induces morpho-physiological and proteome changes of *Pandanus amaryllifolius*. **Plants 11**: 221.

Amnan MAM, Teo WFA, Aizat WM, Khaidizar FD and Tan BC (2023) Foliar application of oil palm wood vinegar enhances *Pandanus amaryllifolius* tolerance under drought stress. **Plants 12**: 785.

Azevedo ALS, Souza FRD, Campos RA, Reis DRDL, Machado JC, Machado MA and Resende M (2022) Development of microsatellite panels for molecular fingerprinting of Napier grass (*Cenchrus purpureus*) cultivars. **Crop Breeding and Applied Biotechnology 22**: e42522244.

Basu S and Bhattacharya S (2022) Development and characterization of EST-SSR derived functional domain marker (FDM) in *'Phaseolus vulgaris'* (common bean). **Plant Omics 15**: 17-24.

Beier S, Thiel T, Münch T, Scholz U and Mascher M (2017) MISA-web: a web server for microsatellite prediction. **Bioinformatics 33**: 2583-2585.

Bhatt V, Barvkar VT, Furtado A, Henry RJ and Nadaf A (2021) Fragrance in *Pandanus amaryllifolius* Roxb. despite the presence of a betaine aldehyde dehydrogenase 2. **International Journal of Molecular Sciences 22**: 6968.

Boeckmann B, Bairoch A, Apweiler R, Blatter M-C, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C and Phan I (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. **Nucleic Acids Research 31**: 365-370.

Boopathi NM (2020) **Genetic mapping and marker assisted selection: basics, practice and benefits**. 2nd edn, Springer, New York, 504p.

Botstein D, White RL, Skolnick M and Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **American Journal of Human Genetics 32**: 314.

Cai K, Zhu L, Zhang K, Li L, Zhao Z, Zeng W and Lin X (2019) Development and characterization of EST-SSR markers from RNA-Seq data in *Phyllostachys violascens*. **Frontiers in Plant Science 10**: 50.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R and Zeng Q (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. **Nature Biotechnology 29**: 644.

Gur-Arie R, Cohen CJ, Eitan Y, Shelef L, Hallerman EM and Kashi Y (2000)

Simple sequence repeats in *Escherichia coli*: abundance, distribution, composition, and polymorphism. **Genome Research 10**: 62-71.

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S and Tokimatsu T (2007) KEGG for linking genomes to life and the environment. **Nucleic Acids Research 36**: D480-D484.

Kumar B, Kumar U and Yadav HK (2015) Identification of EST–SSRs and molecular diversity analysis in *Mentha piperita*. **The Crop Journal 3**: 335-342.

Kumpatla SP and Mukhopadhyay S (2005) Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. **Genome 48**: 985-998.

Liu S, An Y, Li F, Li S, Liu L, Zhou Q, Zhao S and Wei C (2018) Genome-wide identification of simple sequence repeats and development of polymorphic SSR markers for genetic studies in tea plant (*Camellia sinensis*). **Molecular Breeding 38**: 1-13.

Liu Y, Zhang P, Song M, Hou J, Qing M, Wang W and Liu C (2015) Transcriptome analysis and development of SSR molecular markers in *Glycyrrhiza uralensis* Fisch. **PloS ONE 10**: e0143017.

Lu J, Zhang Y, Diao X, Yu K, Dai X, Qu P, Crabbe MJC, Zhang T and Qiao Q (2021) Evaluation of genetic diversity and population structure of *Fragaria nilgerrensis* using EST-SSR markers. **Gene 796**: 145791.

Mathur J (2021) Organelle extensions in plant cells. **Plant Physiology 185**: 593-607.

Nei M (1972) Genetic distance between populations. **The American Naturalist 106**: 283-292.

Parchman TL, Geist KS, Grahnen JA, Benkman CW and Buerkl CA (2010) Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. **BMC Genomics 11**: 1-16.

Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F and Parvizi B (2003) TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. **Bioinformatics 19**: 651-652.

Rohlf F (2000) NTSYS-pc: Numerical taxonomy and multivariate analysis system version 2.2. Exeter Publications, Setauket.

Sa'adah SM, Putri FR, Ibtisam AA, Arrohmah RS and Fitriyah F (2023) Phytochemical analysis of secondary metabolite compounds of Pandan wangi leaf extract (*Pandanus amaryllifolius*). **Journal of Natural Sciences and Mathematics Research 9**: 1-14.

Schuler GD (1997) Sequence mapping by electronic PCR. **Genome Research 7**: 541-550.

Sharma H, Kumar P, Singh A, Aggarwal K, Roy J, Sharma V and Rawat S (2020) Development of polymorphic EST-SSR markers and their applicability in genetic diversity evaluation in *Rhododendron arboreum*. **Molecular Biology Reports 47**: 2447-2457.

Sun L, Li J, Sun K, Wang H, Yang K, Chen Q and Lin M (2023) Development and characterization of EST-SSR markers in *Rhodomyrtus tomentosa* Hassk. based on transcriptome. **Genetic Resources and Crop Evolution 70**: 1691-1705.

Taheri S, Abdullah TL, Rafii M, Harikrishna JA, Werbrouck SP, Teo CH and Sahebi M (2019) De novo assembly of transcriptomes, mining, and development of novel EST-SSR markers in *Curcuma alismatifolia* (Zingiberaceae family) through Illumina sequencing. **Scientific Reports 9**: 1-14.

Tattersall EA, Ergul A, AlKayal F, DeLuc L, Cushman JC and Cramer GR (2005) Comparison of methods for isolating high-quality RNA from leaves of grapevine. **American Journal of Enology and Viticulture 56**: 400-406.

Tulsani NJ, Hamid R, Jacob F, Umretiya NG, Nandha AK, Tomar RS and Golakiya BA (2020) Transcriptome landscaping for gene mining and SSR marker development in coriander (*Coriandrum sativum* L.). **Genomics 112**: 1545-1553.

Unamba CI, Nag A and Sharma RK (2015) Next generation sequencing technologies: the doorway to the unexplored genomics of non-model plants. **Frontiers in Plant Science 6**: 1074.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M and Rozen SG (2012) Primer3—new capabilities and interfaces. **Nucleic Acids Research 40**: e115.

Wakte KV, Nadaf AB, Thengane RJ and Jawali N (2009) *Pandanus amaryllifolius* Roxb. cultivated as a spice in coastal regions of India. **Genetic Resources and Crop Evolution 56**: 735-740.

Wakte KV, Zanan RL, Saini A, Jawali N, Thengane RJ and Nadaf AB (2012) Genetic diversity assessment in *Pandanus amaryllifolius* Roxb. populations of India. **Genetic Resources and Crop Evolution 59**: 1583-1595.

Wang X, Yang S, Chen Y, Zhang S, Zhao Q, Li M, Gao Y, Yang L and Bennetzen JL (2018) Comparative genome-wide characterization leading to simple sequence repeat marker development for Nicotiana. **BMC Genomics 19**: 1-12.

Wongpornchai S (2006) **Pandan wangi.** In Peter KV (ed.) **Handbook of herbs and spices**. Elsevier, Amsterdam, p. 453-459.

Wu Q, Zang F, Xie X, Ma Y, Zheng Y and Zang D (2020) Full-length transcriptome sequencing analysis and development of EST-SSR markers for the endangered species *Populus wulianensis*. **Scientific Reports 10**: 16249.

Yahya F, Fryer PJ and Bakalis S (2011) The absorption of 2-acetyl-1-pyrroline during cooking of rice (*Oryza sativa* L.) with Pandan (*Pandanus amaryllifolius* Roxb.) leaves. **Procedia Food Science 1**: 722-728.

Zhang H, He L and Cai L (2018) **Transcriptome sequencing: RNA-Seq. Methods in Molecular Biology** 1754: 15-27.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.