Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

First Insights into Genetic Diversity of Indonesian Typhonium flagelliforme Using Transcriptome-Based EST- SSR Markers

Devit Purwoko
National Research and Innovation Agency
Siti Zulaeha
National Research and Innovation Agency
Gemilang Rahmadara
National Research and Innovation Agency
Suparjo Suparjo
National Research and Innovation Agency
Teuku Tajuddin
National Research and Innovation Agency
Ani Kurniawati
IPB University
Willy Bayuardi Suwarno
IPB University
Sobir Sobir
sobir@apps.ipb.ac.id

IPB University

Research Article

Keywords: genetic diversity, marker development, microsatellite, roddent tubber, transcriptome assembly

Posted Date: October 29th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7790550/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Typhonium flagelliforme, an Araceae species recognized for its anticancer metabolites distributed throughout its roots, tubers, stems, and leaves, remains underexplored at the molecular level. In this study, we developed expressed sequence tag–simple sequence repeat (EST-SSR) markers from the *T. flagelliforme* transcriptome and evaluated their utility in assessing the genetic diversity of Indonesian germplasm as well as their transferability to other Araceae species. Transcriptome assembly yielded 41,392 unigenes, of which 9,868 were successfully annotated. Mining of 18,225 unigenes identified 24,404 potential EST-SSRs, with an overall frequency of one SSR per 4.36 kb and di-nucleotide repeats (45.83%) being the most abundant motif. From these, 38 primer pairs were randomly selected, and 32 produced clear amplicons, of which 23 (60.5%) were polymorphic across 14 accessions. A total of 15 alleles were detected, with an average of 2.87 alleles per locus. The mean observed and expected heterozygosity were 0.238 and 0.620, respectively, while the polymorphism information content (PIC) ranged from 0.416 to 0.580, with an average of 0.515, indicating a moderate level of polymorphism. Genetic structure analysis and principal coordinate analysis (PCoA) consistently grouped the accessions into three distinct genetic clusters, underscoring the effectiveness of the developed markers in resolving population structure. To our knowledge, this represents the first genetic diversity analysis of Indonesian *T. flagelliforme* germplasm using EST-SSR markers. The newly developed resources provide a valuable foundation for molecular characterization, germplasm conservation, and breeding applications in *T. flagelliforme* and potentially related Araceae species.

Introduction

Typhonium flagelliforme is recognized for its remarkable pharmacological potential, primarily attributed to its anticancer metabolites. Notably, bioactive compounds with anticancer properties are distributed throughout the plant, including the roots, tubers, stems, and leaves (Choo et al. 2001). Phytochemical investigations have revealed the presence of diverse classes of compounds such as alkaloids, steroids, flavonoids, glycosides (Sianipar et al. 2023), and glycolipids (Huang et al. 2004). Among its major metabolites, β-sitosterol and phytol have been highlighted as key contributors to the inhibition of cancer cell growth (Setiawati et al. 2016). In addition to its anticancer potential, *T. flagelliforme* exhibits a broad spectrum of pharmacological activities, including antioxidant, antimicrobial, and anti-inflammatory properties (Mohan et al. 2011; Farida et al. 2014; Mirgane et al. 2020; Septaningsih et al. 2021). However, despite its therapeutic promise, the broader cultivation and application of this species remain constrained. A key limiting factor is the scarcity of genomic and genetic resources, which impedes breeding programs and the development of value-added products. Although previous studies have predominantly focused on phytochemical profiling and assessments of biological activity, comprehensive molecular characterization, especially the development of genetic markers, remains limited and underexplored.

Various molecular markers have been developed and widely used for genetic diversity analysis, DNA fingerprinting, cultivar identification, marker-assisted breeding, genetic linkage mapping, and quantitative

trait locus mapping (Bidyananda et al. 2024; Abdul Aziz and Masmoudi 2025). Among these markers, simple sequence repeats (SSRs) are known to have a high level of polymorphism, are informative, and are widely distributed throughout the plant genome (Senan et al. 2014; Vieira et al. 2016). The advantages of SSRs include high information content, good reproducibility, locus specificity, ease of analysis, high intraspecific polymorphism, multiallelic nature, and wide genome coverage (Umang et al. 2022; Srivastava et al. 2023; Alves et al. 2024). SSR markers have been widely applied in the evaluation of genetic structure and diversity (Purwoko et al. 2019) and are classified into two categories based on their source, namely expressed sequence tag–SSR (EST–SSR) derived from transcribed RNA sequences, and genomic SSR (g–SSR) derived from genomic sequences (Purwoko et al. 2025). Compared with g-SSRs, EST-SSRs generally have a higher degree of interspecies transferability, are easier to identify, are more evolutionarily conservative, and are often associated with specific phenotypes (Purwoko et al. 2021; Topu et al. 2023). Previous studies on the diversity and characterization of *T. flagelliforme* have primarily relied on morphological traits (Erdayani et al. 2023), molecular approaches using SRAP (Erdayani et al. 2023) and RAPD markers (Nurhasanah et al. 2024). The discovery, characterization, and validation of simple sequence repeats (SSRs) from the *de novo* transcriptome assembly of *T. flagelliforme* were recently reported by Purwoko et al. (2025). However, these efforts have not yet been extended to explore their application in assessing the genetic diversity of *T. flagelliforme* or their transferability to related species.

EST-SSR development has been carried out in various plant species (Liu et al. 2021; Gao et al. 2022; Kapoor et al. 2023; Mortazavian et al. 2024; Ren et al. 2025). The use of Next Generation Sequencing (NGS) technology enables the rapid, efficient, accurate, and relatively low-cost acquisition of large-scale genomic data (Satam et al. 2023; Panahi et al. 2024). NGS-based de novo transcriptome sequencing has become a leading method for functional gene identification, differential gene expression analysis, transcriptome profiling, and molecular marker discovery in non-model species (Kapoor et al. 2023; Scholpp et al. 2024). Before the advent of NGS, SSR marker development was a time-consuming, expensive, and labor-intensive process, including the identification of SSR motifs, which requires significant economic investment (Squirrell et al. 2003; Merritt et al. 2015).

Currently, the availability of molecular markers for *T. flagelliforme* is still very limited (Purwoko et al. 2025), and reports on the development of polymorphic SSRs in this species are lacking. These limitations have become obstacles in the genetic study of *T. flagelliforme*. With the advancement of NGS technology, especially RNA-Seq which is starting to replace the microarray platform (Gao et al. 2025), various studies have successfully developed RNA-Seq-based SSR markers in *Cephalotaxus oliveri* (Liu et al. 2021), *Paris polyphylla* (Gao et al. 2022), *Phyllanthus emblica* (Kapoor et al. 2023), *Pseudostellaria heterophylla* (Xu et al. 2023), *Cuminum Cyminum* (Mortazavian et al. 2024), and *Radix Ardisia* (Ren et al. 2025). Therefore, the development of RNA-Seq-based SSR markers to support *T. flagelliforme* breeding strategies is urgently needed. In this study, SSRs were mined, developed and validated a genic SSR marker asset for *T. flagelliforme* from previous transcriptomic data, followed by the characterization and development of 15 EST-SSR markers. These newly developed markers were applied to assess genetic variation and elucidate the phylogenetic relationships among 14 *T. flagelliforme* germplasm accessions.

In addition, their cross-species transferability was evaluated in other members of the Araceae family. This report expected to provide important genetic resources for genetic diversity analysis and strengthen breeding efforts for this species.

Materials and Methods

Plant Material

All plant materials used in this study were domesticated wild accessions maintained under controlled conditions in the experimental screenhouse of the Biotechnology Laboratory Facility, Habibie-BRIN Science and Technology Area (KST), Setu, South Tangerang, Indonesia (6°21'33"S, 106°39'50"E). The cultivated plants were no endangered or protected species were involved. To assess microsatellite variation of the EST–SSR markers developed in this study, 14 accessions (Supplementary Table 1) were analyzed, comprising 11 individuals from Java (5 from West Java, 4 from Central Java, and 1 from East Java), one accession from Bali, one from South Sumatra, and one from East Borneo (Fig. 1).

Sequence Assembly, EST SSRs Identification and Primer Design

The raw dataset from leaf transcriptome sequencing, used for SSR development in this study, was previously generated in our laboratory by Purwoko et al. (2025). The raw Illumina transcriptome sequence reads of *T. flagelliforme* are available in the Sequence Read Archive (NCBI/SRA) under BioProject number PRJNA1142076, accessible at the following link: https://www.ncbi.nlm.nih.gov/sra/PRJNA1142076. The sequencing data were downloaded in the form of raw reads, then examined using FastQC. Trimmomatic was then used to filter the sequencing data by removing low-quality data, adapter sequences, and contaminants, resulting in high-quality reads (v0.39). The high-quality reads were then assembled into transcripts using rnaSPAdes v3.15.3. The assembled FASTA sequences were subsequently utilized for SSR identification, primer design, and functional gene analysis. SSR primers were designed from transcriptomic sequences using PrimerPro version 1.0 (http://webdocs.cs.ualberta.ca/~yifeng/primerpro/), with parameters adapted from Worth et al. (2019).

Functional Annotations and Classification of EST-SSRs

Sequences containing SSRs with a minimum motif length of 20 bp were then subjected to synteny analysis against *Oryza sativa* chromosome data. To predict putative gene functions, SSR-containing sequences were compared against the non-redundant Arabidopsis protein database available in The Arabidopsis Information Resource (TAIR; http://www.arabidopsis.org/index.jsp) using BLASTx with an e-value cut-off of 10 − 3. GO analysis based on biological process, molecular function and cellular component categorization was done using ShinyGO 0.80 (http://bioinformatics.sdstate.edu/go/) (Ge et al. 2020). Furthermore, pathway enrichment was conducted using the KEGG database (Kanehisa et al.

2025), and the distribution of enriched pathways, including those associated with SSR-containing sequences, was illustrated through bar plots generated in ShinyGO (Ge et al. 2020).

DNA extraction, Validation and EST-SSR Marker Amplification

Genomic DNA was extracted from fresh leaf tissues using the Genomic DNA Mini Kit for plants (Geneaid™, GP100). Primers were designed based on the following parameters: annealing temperature (55–60°C), GC content (40–60%), primer length (18–20 bp), and target PCR product size (100–500 bp). EST-SSR primer validation was carried out on Serpong, Tanjung Bungkak and Karang Anyar accessions using PCR amplification with a total composition of 25 µL/reaction consisting of GoTaq® Green Master Mix, 2X (12.5 µL), 10 µM Forward primer (1 µL), Reverse primer (1 µL), 50 µM DNA template (5 µL) and sterile H_2O to a volume of 25 µL. Amplification was carried out using a Tommy PCR machine. The program used was as follows: one cycle of 95°C for 15 minutes, 94°C for 30 seconds, 30 cycles, 52°C for 90 seconds and 72°C for 60 seconds and a final extension of 60°C for 60 minutes and ended with a temperature of 4°C. Successfully validated primers will be used to identify the genetic diversity of the mouse ear plant. Amplification separation using 1.5% agarose in 0.5X TAE buffer at 50 V for 45 minutes electrophoresed with eXU MUPID. DNA standards (100bp ladder) were used so that the size of the sample DNA could be estimated. The agarose gel was stained with sybr safe® and then observed with a Vilber Lourmat UV transilluminator and documented using a digital camera.

Phylogenetic, Population Genetic Structure, and Transferability in Cross-Species

We used the 15 EST-SSR markers to analyze genetic diversity among 14 individual accessions and three different species: *Alocasia cucullata*, *Xanthosoma undipes* K.Kock, and *Metroxylon sagu*, were used to analyze the transferability of the 15 EST-SSR markers. The number of observed alleles (Na), expected heterozygosity (He), effective alleles (Ne), Shannon's information index (I) and Polymorphism information content (PIC) values were computed using Cervus version 3.0.7 (Kalinowski et al. 2007). The dissimilarity matrix calculation in this study was based on allele data, assuming a diploid ploidy level, using a simple dissimilarity index approach. To ensure the accuracy of the results, the analysis was supplemented with 1,000 bootstrap iterations. A phylogenetic tree was constructed using the Neighbor Joining method based on the previously calculated dissimilarity matrix. The entire analysis process, from dissimilarity matrix creation, bootstrap analysis, factorial analysis, to phylogenetic tree construction, was performed using DARwin software version 6.05 (Perrier & Jacquemoud-Collet, 2006), available at http://darwin.cirad.fr/darwin. Population genetic structure among the 14 *T. flagelliforme* individuals was inferred using STRUCTURE software version 2.3.4. To determine the most likely number of genetic clusters (K), ten independent simulations were performed for K values ranging from 1 to 7. Each run consisted of a burn-in period of 10,000 iterations followed by 10,000 Markov chain Monte Carlo (MCMC) repetitions to ensure convergence and accuracy. The optimal K value was identified using the ΔK method implemented in the online program Structure Selector (Li and Liu 2018). The resulting population

stratification patterns were visualized based on the optimal K value, as described previously (Evanno et al. 2005; Purwoko et al. 2019; Zulaeha et al. 2023).

Results

Statistic De novo Assembly and EST SSRs Mining

The assembled transcriptome comprised a total of 116,877 sequences with an overall length of 117,819,065 bp. The sequence size ranged from 40 bp to 15,253 bp, with an average length of 1,008 bp. The assembly quality was further supported by an N50 of 1,947 bp and an N90 of 341 bp, indicating a relatively robust representation of transcript sizes. From the 115,035 sequences examined for EST-SSRs, covering a total length of 55,951,160 bp, a total of 24,404 SSRs were identified. Among these, 18,225 sequences contained at least one SSR, while 4,483 sequences carried more than one motif. In addition, 3,444 SSRs were found in compound formation, suggesting notable complexity in the SSR distribution across the transcriptome (Table 1).

Table 1
De novo assembled EST-SSRs for T. flagelliforme

Category	Items	Numbers
Assembly statistic	Total sequence number	116,877
	Total sequence base	117,819,065
	Largest	15,253
	Smallest	40
	Average	1,008
	N50 (bp)	1,947
	N90 (bp)	341
EST-SSRs	Total number of sequences examined	115,035
	Total size of examined sequences (bp)	55,951,160
	Total number of identified SSRs	24,404
	Number of SSR containing sequences	18,225
	Number of sequences containing more than 1 SSR	4,483
	Number of SSRs present in compound formation	3,444

Based on the sequence length distribution illustrated in the graph, the assembled unigenes exhibited a declining trend in number with increasing sequence length. The category of 200–299 bp contained the highest number of unigenes, followed by the 300–399 bp and 400–499 bp categories. The number of

unigenes then gradually decreased across the medium-length range (500–999 bp) and further into sequences exceeding 2000 bp. Overall, the distribution indicates a predominance of short- to medium-length unigenes, with a relatively small proportion exceeding 2 kb. This pattern is consistent with the typical characteristics of de novo transcriptome assemblies, in which shorter transcripts are more frequently detected compared to longer ones. The number of assembled unigenes decreased progressively with increasing sequence length (Fig. 2).

Frequency and Distribution of EST-SSRs

The distribution of identified SSR motifs revealed a predominance of dinucleotide repeats, which accounted for 11,184 loci (45.83%) of the total SSRs (Fig. 3). Mononucleotide motifs represented the second most abundant category, followed by trinucleotide repeats, together comprising the majority of SSRs. In contrast, tetranucleotide, pentanucleotide, and hexanucleotide repeats were detected at much lower frequencies, indicating their relatively minor contribution to the overall SSR composition. This pattern suggests that short repeat motifs, particularly di- and mononucleotides, dominate the SSR landscape of the transcriptome.

Analysis of SSR motif types demonstrated that dinucleotide repeats were the most abundant, with AG/CT motifs showing the highest frequency (7,149), followed by AT/AT (2,710) and AC/GT (1,248). Among trinucleotide motifs, AAG/CTT was the most common (993), while other motifs such as AAT/ATT, ACC/GGT, ACG/CTG, and AGC/CGT occurred at relatively lower frequencies (ranging from 240 to 474). Notably, the trinucleotide motif CCG/CGG was also highly represented (2,034), ranking among the top abundant classes. In contrast, motifs such as CG/CG and AAC/GTT were rarely detected, with only 77 and 72 instances, respectively. Overall, the motif distribution highlights the dominance of AG/CT dinucleotide repeats and the presence of a few highly abundant trinucleotide motifs, suggesting a non-random pattern of SSR occurrence within the transcriptome (Fig. 4).

The distribution of tetranucleotide SSR motifs showed a clear predominance of AGCT/ATCG repeats, which were the most abundant with 45 occurrences (Fig. 5). Other frequently detected motifs included AATT/ATTT (37), ACGT/ATGC (36), and AGAT/ATCT (28). Several motifs such as AAAC/GTTT, AAAG/CTTT, ACAT/ATGT, and AGGG/CCTT were represented at intermediate frequencies (ranging from 20 to 31). In contrast, pentanucleotide and hexanucleotide occurred at low frequencies, each with fewer than 10 instances. Overall, the results indicate that tetranucleotide, pentanucleotide and hexanucleotide SSRs were less abundant than di- and trinucleotide repeats, but specific motifs such as AG/CT and CCG/CGG exhibit strong enrichment, suggesting possible functional or structural preferences in the transcriptome.

EST SSRs Functional Annotation and Classification

From the 18,225 detected SSRs, a total of 88 primer pairs were successfully designed, excluding mononucleotide repeats due to their high PCR error rate. Of these, 38 primer pairs were randomly

selected and synthesized by Macrogen Inc. (Seoul, Korea) for the development of polymorphic EST-SSR markers. Functional annotation of 38 sequences was performed using similarity searches against the Nr, Swiss-Prot, GO, eggNOG, KEGG, and Pfam databases via the BLASTx algorithm, applying a stringent E-value cut-off of 1×10^{-3} (Table 2). Among them, 36 unigenes (94.74%) were successfully annotated in the Nr database, while 30 (78.95%) showed matches in Swiss-Prot. GO annotation assigned terms to 25 unigenes (65.79%), and 24 (63.16%) were associated with Pfam protein families. Notably, the highest coverage was obtained from eggNOG, with 35 unigenes (92.11%) assigned, whereas KEGG pathway mapping yielded the lowest representation, with only 12 unigenes (31.58%) annotated. Collectively, these results highlight the broad functional coverage achieved, with Nr and eggNOG contributing the most comprehensive annotations.

Table 2
Summary of functional annotation results of 38 unigenes in T. flagelliforme transcriptome.

Annotated Database	Number of Unigene	Percentage (%)
Nr	36	94.74
Swiss-Prot	30	78.95
GO	25	65.79
eggNOG	35	92.11
KEGG	12	31.58
Pfam	24	63.16
Total Unigene	38	100

Functional annotation of SSR-containing unigenes showed that out of 38 sequences analyzed, 32 produced BLAST hits, 28 were successfully mapped to GO terms, and 25 received GO annotations. Additionally, all 38 sequences were annotated through InterProScan, indicating complementary functional insights from domain-based analysis. These results suggest that a substantial proportion of SSR-associated unigenes can be functionally characterized, reinforcing the reliability of the transcriptome dataset for downstream functional and comparative genomics studies.

Gene Ontology (GO) classification of SSR-containing unigenes revealed a wide functional distribution across the three major categories: biological process, cellular component, and molecular function (Fig. 6). Within the biological process category, the majority of unigenes were enriched in metabolic and cellular processes, particularly organic substance metabolism, nitrogen compound metabolism, and regulation of biological processes. For the cellular component category, the most highly represented terms included intracellular membrane-bounded organelles, cytoplasm, and plasma membrane, indicating that many unigenes are associated with fundamental cellular structures. In the molecular function category, unigenes were predominantly involved in catalytic activity and binding functions, with strong representation in transferase, hydrolase, and nucleotide-binding activities. Collectively, these

results highlight that SSR-associated unigenes are functionally diverse and play roles in essential biological, structural, and molecular activities.

The annotated unigenes were categorized into three main GO domains: cellular component, molecular function, and biological process, covering 578 functional groups (Fig. 7). Within the cellular component category, the most abundant terms were i*ntracellular membrane-bounded organelle* (1,917 unigenes, 14.69%) and *cytoplasm* (1,756, 13.46%), followed by *membrane* (1,448, 11.10%). For molecular function, the largest groups were *anion binding* (1,788, 5.56%) and *small molecule binding* (1,782, 5.54%), while specialized functions such as *squalene monooxygenase activity*, *auxin influx transmembrane transporter activity*, *oxidoreductase activity*, and l*igand-gated cation channel activity* were represented by a smaller proportion (6, 0.02%).

In the biological process domain, most unigenes were associated with o*rganic substance metabolic process* (1,906, 3.20%) and *cellular metabolic process* (1,862, 3.13%), reflecting the prominence of fundamental metabolic and cellular activities in *T. flagelliforme*. Other highly represented processes included *primary metabolic process* (1,860, 3.13%), n*itrogen compound metabolic process* (1,788, 3.00%), and *macromolecule metabolic process* (1,689, 2.84%). A smaller subset of unigenes was linked to *mitotic DNA replication* (6, 0.01%), *regulation of gene silencing by miRNA* (6, 0.01%), and *peptidyl-threonine phosphorylation* (6, 0.01%), potentially relating to fundamental cell cycle regulation, with phosphorylation being a key mechanism for controlling DNA replication and cell division, and miRNAs a mechanism to fine-tune gene expression in processes like the cell cycle and DNA damage response (Song et al. 2023; Quévillon Huberdeau and Simard 2019; Garcia-Garcia et al. 2016). These processes are interconnected, as phosphorylation regulates critical proteins involved in DNA replication, and miRNAs can respond to DNA damage, influencing cell cycle progression.

KEGG pathway enrichment analysis of SSR-containing unigenes showed that they were distributed across several essential biological pathways (Fig. 8). The most significantly enriched pathways included endocytosis and the plant MAPK signaling pathway, each containing a large number of genes. Other important pathways represented were ABC transporters, phagosome formation, homologous recombination, phosphatidylinositol signaling, and plant–pathogen interaction. Pathways with moderate unigene representation included DNA replication, ribosome biogenesis, and RNA degradation. These findings suggest that SSR-associated unigenes are functionally diverse and may play key roles in cellular transport, signaling, genetic stability, and defense responses.

The mitogen-activated protein kinase (MAPK) signalling cascade plays a pivotal role in regulating diverse aspects of plant biology, including growth, development, yield formation, and responses to both abiotic and biotic stresses (Jagodzik et al. 2018; Manna et al. 2023). Owing to its central regulatory function, this pathway represents a critical target for genetic and molecular interventions aimed at crop improvement. KEGG pathway enrichment analysis revealed the mitogen-activated protein kinase (MAPK) signalling cascade as a key regulatory network in *T, flagelliforme* (Fig. 9). Several genes associated with MAPKKK, MAPKK, and MAPK modules were significantly mapped (highlighted in red), indicating their involvement in multiple biological processes, including pathogen infection and defense responses, hormone-mediated

signalling (ethylene, jasmonate, and abscisic acid), abiotic stress tolerance (cold, salt, and drought), and developmental regulation such as root growth and stomatal development. The enrichment of these signalling components highlights the central role of MAPK cascades in integrating environmental cues with plant growth and defense mechanisms.

Development and validation of novel EST-SSRs

Primer design was conducted using Primer3 software, ensuring that all pairs met the established selection criteria. From the transcriptome of *Typhonium flagelliforme*, a total of 88 primer pairs for expressed sequence tag–derived SSRs (EST-SSRs) were successfully identified., providing a useful genomic resource for future functional and applied studies. A total of 38 primer pairs were randomly chosen for synthesis and subjected to preliminary polymorphism screening. In this study, a total of 38 EST-SSR primer pairs were randomly selected and tested across three *Typhonium flagelliforme* accessions (Serpong, Tanjung Bungkak, and Karang Anyar). Of these, 32 primers successfully produced amplicons, with 23 being polymorphic and 9 monomorphic. To further evaluate their utility, 15 primers were applied to 14 additional *T. flagelliforme* accessions.

A total of 15 EST-SSR primer pairs were successfully designed and characterized from *Typhonium flagelliforme* (Table 3). The repeat motifs identified included di-, tri-, and tetra-nucleotide types, with product sizes ranging from 100 to 300 bp. Several primers were associated with genes encoding proteins of known biological functions, such as *solute carrier family 13* (SSR1), *choline transporter-like* (SSR2), *deoxyhypusine synthase superfamily* (SSR4), *decaprenyl diphosphate synthase-like* (SSR5), *ribonuclease H domain* (SSR6), and *peptidase S8/S53 domain superfamily* (SSR11). Other primers were linked to regulatory or stress-responsive genes, including *ethylene insensitive 3* (SSR14) and P-type ATPases (SSR12 and SSR15), suggesting their potential relevance in developmental and adaptive processes. In contrast, several primers (e.g., SSR3, SSR7, SSR8, SSR10) were associated with hypothetical or uncharacterized proteins, indicating unexplored functional diversity within the transcriptome. The gene ontology (GO) annotations revealed varying levels of functional assignment, with the highest GO term association observed for SSR5 (15 annotations), followed by SSR14 (10 annotations) and SSR11 (8 annotations). Collectively, the characterized primers demonstrate not only structural variation in repeat motifs but also functional diversity, thereby providing valuable resources for genetic diversity analysis, functional marker development, and future breeding applications in *T. flagelliforme*.

Table 3
List and functional annotation results in *T. flagelliforme* primers.

No	Primer	Repeat Motif	Primer sequence (50 – 30)		Expected product size (bp)	Description of putative function	GO
1	SSR 1	(AT)15	F:	CTTCCTCTACCCCCTCTTCG	250–273	Solute carrier family 13	4
			R:	ATCGCATCGAATCCACAAA			
2	SSR 2	(AT)11	F:	GCCCACTTTGATCCTTGAGT	199 – 150	Choline transporter-like	2
			R:	TGGCAGACCTTGTGTTATGC			
3	SSR 3	(GCT)8	F:	TTGGCACATTTGAAGGTGAA	226–250	Nn	0
			R:	GGTTGTGCGTGTTCCTACAA			
4	SSR 4	(CT)14	F:	TGCTCTCTTTCTCTCGTCCC	273–300	Deoxyhypusine synthase superfamily	3
			R:	CGAAGACAACTCGTCATGGA			
5	SSR 5	(GA)23	F:	ATCGTCCTCTAGGCCCTTGT	237–250	Decaprenyl diphosphate synthase-like	15
			R:	CTCTTGCAAGGGGAAGACTG			
6	SSR 6	(AT)11	F:	TTCTTGTACTGCCTCCCCTG	106–150	Ribonuclease H domain	2
			R:	TGCATGTGCAAAGAGTCTCC			
7	SSR 7	(TC)10	F:	GTTTTCCCAAGCCTCCCC	263–300	JmjN domain	0
			R:	CCAGGAAGGTATTCCAAGCC			
8	SSR 8	(TG)12	F:	GAAAGGCAAGTTATTGGCGA	179–200	Agenet domain, plant type	0
			R:	AATTTTCTTCCAGGAACCCG			
9	SSR 9	(GAA)10	F:	GCCAACACGGGTAGAAGAAC	150–205	L-aspartate oxidase	7
			R:	CGAAAAAGGAATCAAACCCA			
10	SSR 10	(TG)12	F:	GAAAGGCAAGTTATTGGCGA	150–179	Agenet domain, plant type	0
			R:	AATTTTCTTCCAGGAACCCG			
11	SSR 11	(AG)16	F:	TATTCGCGTAGATCGCATGA	100–125	Peptidase S8/S53 domain superfamily	8
			R:	ACTGGAGTTAAAAGACGCCG			
12	SSR 12	(AGA)8	F:	TCAAAGCCCTTCCTTCACAC	150–222	P-type ATPase, subfamily IV	4
			R:	GTGCATATCACCACGCAGTC			

No	Primer	Repeat Motif	Primer sequence (50 – 30)		Expected product size (bp)	Description of putative function	GO
13	SSR 13	(AT)20	F:	GGCAGTGCACTAACATTACCA	150–170	Plus3-like superfamily	3
			R:	TTGGCCATAATCAGCATTGA			
14	SSR 14	(TAA)8	F:	CCCTATACACCCTCACCCTG	200–262	Ethylene insensitive 3	10
			R:	CCTCCCTTGCTTTTGTTGAC			
15	SSR 15	(CT)13	F:	CACACCCCCTCTTTAAACCC	150–187	P-type ATPase, cytoplasmic domain N	4
			R:	ACCGTCCAATTACAAGAGCG			

Transferability of T. flagelliforme EST-SSR markers in different Araceae plants

To assess the broader applicability of the developed markers, we tested whether primer pairs designed from EST sequences of *T. flagelliforme* could reliably amplify the corresponding SSR motifs in other Araceae species. The cross-transferability of EST-SSR primers was successfully evaluated across three species of Araceae (Supplementary Fig. 1). As summarized in Table 4, 13 of the 15 EST-SSR markers successfully amplified fragments of the expected size in *Xanthosoma undipes*, whereas only 7 and 8 markers were amplifiable in *Metroxylon sagu* and *Alocasia cucullata*, corresponding to amplification rates of 47–87%. These findings indicate clear differences in amplification efficiency across species, with notably high success in *X. undipes* and lower rates observed in *M. sagu* and *A. cucullata*. While most markers were consistently amplified, a few, such as SSR2, SSR5 and SSR6, produced amplification only in *Xanthosoma undipes*, indicating species-specific variability in marker performance.

Table 4
Cross-species amplification of the 15 microsatellite loci in Araceae species

Markers	Amplification		
	Alocasia cucullata	*Xanthosoma undipes*	*Metroxylon sagu*
SSR1	+	+	+
SSR2	-	+	-
SSR3	+	+	+
SSR4	+	+	-
SSR5	-	+	-
SSR6	-	+	-
SSR7	-	+	+
SSR8	+	+	+
SSR9	+	+	-
SSR10	-	+	+
SSR11	-	-	-
SSR12	+	+	-
SSR13	-	-	-
SSR14	+	+	+
SSR15	+	+	+
Total	8	13	7

Genetic diversity analysis of EST-SSR markers

The genetic diversity parameters generated from 15 EST-SSR loci across 14 accessions of *Typhonium flagelliforme* are presented in Table 5. The number of alleles per locus (Na) ranged from 2 to 3, with an average of 2.87 alleles, reflecting moderate allelic richness within the population. Observed heterozygosity (Ho) varied considerably among loci, from 0.071 (SSR2, SSR8, SSR11, SSR12) to 0.643 (SSR6), with a mean of 0.238. In contrast, the expected heterozygosity (He) was consistently higher, ranging from 0.537 to 0.677, and averaged 0.620, indicating a substantial proportion of genetic variation within the accessions. The effective number of alleles (Ne-I) ranged from 0.194 to 0.343 (mean 0.247), while the Shannon's information index (Ne-SI) ranged between 0.472 and 0.575 (mean 0.513). The polymorphic information content (PIC), which reflects marker informativeness, showed values between 0.480 and 0.580, with an average of 0.515, suggesting that the SSR markers employed were highly informative for genetic diversity assessment. Taken together, these results indicate that the selected EST-SSR loci are effective for detecting genetic variation in *T. flagelliforme*, with relatively high expected

diversity but lower observed heterozygosity, possibly due to limited gene flow or a degree of inbreeding within accessions.

Table 5
Summary of observed allele number (N), polymorphism information content (PIC), observed and expected heterozygosity (Ho and He) for 14 *T. flagelliforme* transcriptome accession.

No	SSR Loci ID	N	Ho	He	NE-I	NE-SI	PIC
1	SSR1	3	0.571	0.677	0.194	0.472	0.58
2	SSR2	3	0.071	0.627	0.224	0.504	0.537
3	SSR3	3	0.214	0.611	0.257	0.52	0.501
4	SSR4	3	0.071	0.548	0.329	0.568	0.421
5	SSR5	3	0.357	0.648	0.217	0.492	0.548
6	SSR6	3	0.643	0.659	0.207	0.484	0.561
7	SSR7	3	0.214	0.648	0.217	0.492	0.548
8	SSR8	3	0.071	0.653	0.215	0.489	0.552
9	SSR9	3	0.286	0.677	0.194	0.472	0.58
10	SSR10	3	0.143	0.624	0.238	0.508	0.523
11	SSR11	2	0.071	0.537	0.334	0.575	0.416
12	SSR12	2	0.071	0.553	0.327	0.565	0.424
13	SSR13	3	0.286	0.593	0.254	0.528	0.501
14	SSR14	3	0.214	0.606	0.259	0.523	0.498
15	SSR15	3	0.286	0.635	0.233	0.502	0.53
	Average	2.867	0.238	0.620	0.247	0.513	0.515

Genetic diversity was assessed using cluster analysis, which groups individuals based on allele patterns derived from SSR markers, thereby reflecting their genetic similarity or divergence. The clustering results were visualized in the form of a phylogenetic tree, which revealed three distinct genetic groups among the 14 accessions (Fig. 10). Cluster I grouped together accessions from West Java (S3, S4, S6, and S7), Banten (S2), and East Borneo (S14), indicating a high level of genetic similarity across these geographically separated regions. Cluster II consisted of accessions from Central Java (S10, S11), Sumatera (S1), East Java (S12), and West Java (S5), highlighting inter-regional genetic relatedness. In contrast, Cluster III formed a distinct group comprising accessions from Bali (S13) and Central Java (S8, S9), reflecting unique genetic divergence. Interestingly, the clustering pattern did not always correspond

strictly to geographic origin, suggesting possible gene flow or shared ancestral variation among populations from different islands.

Factorial analysis (Axes 1/2) revealed a clear separation of the 14 *Typhonium flagelliforme* accessions into three major groups (Fig. 11). The first cluster comprised accessions from Sumatera (S1), West Java (S5), East Java (S12), and Central Java (S10, S11), which grouped closely along the negative axis. The second cluster included accessions from Central Java (S8, S9), Bali (S13), and West Java (S4), positioned on the positive axis. The third cluster was formed by accessions from West Java (S3, S6, S7), Banten (S2), and East Borneo (S14), separated towards the lower right quadrant. These results corroborate the UPGMA dendrogram, highlighting both geographic associations and genetic divergence across regions, while also suggesting that accessions from Java exhibit broader genetic distribution compared to those from Sumatera and Borneo.

Vertical bar plots with distinct color patterns represent the estimated membership coefficients of each sample, as calculated by the Q statistic (Fig. 12a**)**. The ΔK estimator, derived from the likelihood probability [LnP(D)] across K values ranging from 1 to 7 (Fig. 12b**)**, was used to determine the optimal number of genetic clusters. For *T. flagelliforme* populations cultivated in Indonesia, K values between 2 and 7 were tested, with each calculation repeated ten times to ensure consistency. The ΔK trend analysis revealed a distinct peak at K = 3 (Fig. 12c), indicating three major genetic groups. Accordingly, all samples were assigned to three clusters: Group I (red) comprising seven samples, Group II (green) comprising two samples, and Group III (blue) comprising five samples. The NJ dendrogram (Fig. 10) and factorial analysis (Fig. 11) also separated the germplasms into three distinct groups, consistent with the STRUCTURE-based population stratification results, thereby validating the genetic clustering patterns observed among the 14 wild T. flagelliforme accessions.

Discussion

Characterization of the Typhonium flagelliforme transcriptome assembly

For species lacking a reference genome, transcriptome sequencing provides one of the most effective strategies for identifying functional genes and developing novel molecular markers [Unamba et al. 2015; Zalapa et al. 2012]. To our knowledge, this study represents the first report of EST-SSR marker development in *Typhonium flagelliforme*. Publicly available transcriptome data derived from *T. flagelliforme* leaves (accession number SRR8752356) were retrieved from the NCBI SRA database for paired-end sequencing and subsequent assembly. The resulting transcriptome comprised 116,877 sequences, spanning a total of 117,819,065 bp. The assembly quality was reflected in an N50 of 1,947 bp, which is substantially higher than those reported for other members of the Araceae family, including *Amorphophallus konjac* (N50 = 381 bp) (Zheng et al. 2013), A. *bulbifer* (N50 = 372 bp) (Zheng et al. 2013), *Zantedeschia rehmannii* (N50 = 1,476 bp) (Wei et al. 2016), *Colocasia esculenta* (N50 = 1,357 bp) (Wang et al. 2017), *Arum maculatum* (N50 = 1,699 bp) (Szenteczki et al. 2022), and *Aglaonema commutatum* (N50 = 1,340 bp) (Li et al. 2023).

Frequency and distribution of EST-SSRs

The rapid development of next-generation sequencing (NGS) technologies has revolutionized the identification and application of molecular markers, particularly simple sequence repeats (SSRs), in a wide range of plant species. By providing large-scale sequence information, NGS has facilitated the efficient discovery of SSRs in both model and non-model plants, including medicinal species. These advances have expanded the scope of SSR marker applications, from plant breeding and disease resistance studies to genetic diversity and evolutionary analyzes across diverse crops, vegetables, and fruits. The codominant nature, reproducibility, and high transferability of SSRs make them particularly advantageous compared with other molecular markers, thereby strengthening their role as reliable tools for assessing genetic variation (Vieira et al. 2016; Satam et al. 2023).

Typhonium flagelliforme, a distinctive medicinal herb of the Araceae family, was recognized for its potent anticancer properties and long-standing ethnomedicinal use in Indonesia. Despite its pharmacological importance, genomic resources for this species remain scarce, particularly with respect to molecular markers. This lack of information has limited comprehensive studies on its genetic diversity, population structure, and evolutionary relationships, thus constraining both conservation and breeding efforts. In the present study, we addressed this gap by mining expressed sequence tag–simple sequence repeats (EST-SSRs) from transcriptome data and developing novel genic SSR markers. These markers were validated and applied to assess the genetic variation among *T. flagelliforme* accessions, as well as to evaluate their cross-species transferability within the Araceae family. The outcomes provide new molecular insights and valuable tools for future genetic, evolutionary, and breeding studies of this underexplored medicinal plant.

In the present study, a total of 18,225 SSRs were identified from 115,035 asembled sequences of *T. flagelliforme*, corresponding to a density of one SSR per 0.326 kb of EST sequence. This frequency is considerably lower than that reported others in the EST-SSR marker development of medicinal plants, where averaging one SSR locus per 2.1 KB (Ren et al. 2025). Among the identified motifs, dinucleotide repeats were the most prevalent (45.83%), a trend consistent with observations in *Radix Ardisia* (Ren et al. 2025), *Cuminum cyminum* (Mortazavian et al. 2024), and *Paris polyphylla* ESTs (Gao et al. 2022), where dinucleotide motifs predominated over other repeat classes. In contrast, studies on *Phyllanthus emblica* have reported a predominance of trinucleotide motifs (Kapoor et al. 2023). Dissimilar things were reported in *Pseudostellaria heterophylla* that trinucleotides predominated compared to others (Xu et al. 2023). However, the type of dinucleotide motif found to be the most common SSR was AG/CT (7,149) followed by CCG/CGG (2,034) then AGCT/ATCG (45). This is different to motifs in *Pseudostellaria heterophylla*, in which the AT/AT and AAT/ATT motifs are predominant (Xu et al. 2023).

EST SSRs Functional Annotation and Classification

Not all SSRs are suitable for primer design, as some occur at positions that make primer construction unfeasible, such as near the beginning or end of a sequence. As noted by Feeney et al. (2014), primer design failure often arises from the absence of suitable clamping regions or constraints related to

melting temperature. Validation of designed primers is therefore essential to assess their amplification efficiency. The success of primer amplification depends on multiple factors, including internal stability, melting temperature, secondary structure formation, and possible competition between primers (Schoenbrunner et al. 2017). A total of 88 primer pairs were successfully designed, or approximately 0.48% of the 18,225 SSRs detected. This is lower than in its related plants: *Colocasia esculenta* (150 primer pairs) (Wang et al. 2017) and *Zantedeschia rehmannii* (200 primer pairs) (Wei et al. 2016).

For annotation analysis, EST sequences with SSR and having a primer (38 sequences randomly selected) were performed similarity searches against the Nr, Swiss-Prot, GO, eggNOG, KEGG and Pfam databases. Between 31.58% and 94.74% of the unigenes were successfully annotated in at least one public database, a proportion consistent with previous reports for leaf unigenes (Liu et al. 2021). The relatively high proportion of unigenes exceeding 500 bp in length within the assembled transcriptome likely contributed to the higher annotation success observed. Longer unigenes generally have a greater probability of producing significant BLAST hits in protein databases (Parchman et al. 2010). This trend has also been emphasized in earlier studies, which noted that the reliability of sequence similarity searches is partly influenced by query length, and that shorter reads generated through next-generation sequencing (NGS) often fail to align with known genes (Novaes et al. 2008). Gene Ontology (GO) analysis revealed that most SSR-containing unigenes were assigned to categories such as i*ntracellular membrane-bounded organelle*, *cytoplasm*, *anion binding*, o*rganic substance metabolic process, cellular metabolic process* and *nitrogen compound metabolic process* functions that underpin essential metabolic pathways in plant cells. Complementary KEGG and eggNOG classifications further demonstrated that SSR-containing unigenes contribute to a wide range of biological processes associated with the growth and development of *T. flagelliforme*, underscoring their potential role in key physiological functions.

Evaluation of genetic diversity and relationships among T. flagelliforme accessions

To assess the utility of EST-SSR markers in detecting polymorphism, three accessions (Serpong, Tanjung Bungkak, and Karang Anyar) were used for initial validation. Among the 38 primer pairs synthesized, 32 successfully generated amplicons, of which 23 were polymorphic and 9 monomorphic. Of these, 23 primer pairs (60.52%) yielded PCR products of the expected fragment size, and 15 polymorphic EST-SSR markers were ultimately selected for further screening across 14 accessions. EST-SSRs are advantageous compared with genomic SSRs located in non-coding regions, as they typically show higher amplification efficiency and better cross-species transferability (Varshney, Graner, & Sorrells, 2005). The successful amplification and polymorphism rates obtained in this study (59.3%) were substantially higher than those previously reported for *Zantedeschia aethiopica* using Sanger sequencing (Wei et al. 2012), supporting the accuracy and reliability of de novo transcriptome sequencing via Illumina RNA-seq. For comparison, in *Z. aethiopica*, 68 (40.9%) primer pairs amplified in 24 accessions, with 43 (63.2%) being polymorphic, while considerably higher amplification and polymorphism rates were reported in *Amorphophallus spp.* (Zheng et al. 2013), where 270 (84.4%) primer pairs amplified successfully, and 205 (89.1%) were polymorphic across 25 accessions. Lower amplification and polymorphism efficiencies

observed in *T. flagelliforme* may be attributed to several factors, including the presence of introns in the cDNA, primer–template mismatches caused by SNPs or InDels, potential errors during de novo assembly, and the inherently high heterozygosity of the species' genome (Varshney et al. 2007). Optimization of PCR conditions, such as lowering annealing temperatures or employing gradient PCR, may help to overcome amplification failures.

Earlier studies on *T. flagelliforme* have used molecular markers such as RAPD and SRAP (Erdayani et al. 2023; Nurhasanah et al. 2024) for cultivar identification and diversity analysis, but the information generated has been limited and inconsistent. By contrast, EST-SSR markers offer greater resolution for genetic analysis, as they target coding sequences that are typically more conserved across related taxa (Filho et al. 2010; Chen et al. 2024). In this study, 43 alleles were detected across 14 accessions using 15 polymorphic EST-SSR markers, with an average of 2.87 alleles per locus. The polymorphism information content (PIC) ranged from 0.416 to 0.580, with a mean of 0.515, which is higher than the mean PIC (0.420) reported in *Zantedeschia rehmannii* (Wei et al. 2016). These results underscore the relatively high discriminatory power of the markers developed here.

Cluster analysis based on the 15 polymorphic markers grouped the 14 accessions into three distinct clusters, further supporting the effectiveness of these EST-SSR markers in revealing genetic relationships and cultivar differentiation within *T. flagelliforme*. With the growing accessibility of genomic data, numerous EST datasets have been generated for a wide range of medicinal plants (Liu et al. 2021; Gao et al. 2022; Kapoor et al. 2023; Xu et al. 2023; Mortazavian et al. 2024; Ren et al. 2025). These resources have proven effective for developing gene-associated molecular markers that are directly linked to agronomically important traits (Kapoor et al. 2023; Li et al. 2024). However, only limited efforts have been made to design and evaluate EST-SSR or other gene-linked markers in *T. flagelliforme* (Purwoko et al. 2025). This scarcity highlights both the novelty and importance of our findings, underscoring the opportunity to advance research in marker development for this pharmacologically valuable species. The development of such markers not only enhances germplasm characterization but also provides valuable resources for genetic diversity studies, comparative mapping, and marker-assisted selection. Importantly, as EST-SSRs originate from coding regions, they can be directly associated with functional genes (Purwoko et al. 2021; Gao et al. 2022), making them particularly valuable for genetic association and linkage studies.

Conclusion

In this study, we identified 18,225 EST-SSRs from the transcriptome of *T. flagelliforme* and characterized their distribution, frequency, genomic locations, and functional annotations. From these, 15 highly informative EST-SSR markers were developed, displaying abundant polymorphism and revealing a moderate level of genetic diversity. Both genetic structure analysis and principal coordinate analysis (PCoA) consistently grouped the accessions into three distinct genetic clusters, underscoring the effectiveness of the developed markers in resolving population structure. Notably, more than 70% of the EST-SSRs demonstrated cross-species transferability, highlighting their potential for broader application

across related taxa. Collectively, these markers represent valuable genomic resources for *T. flagelliforme*, providing a foundation for future studies on genetic diversity, population structure, and evolutionary dynamics, as well as supporting conservation strategies and breeding programs for this important medicinal plant.

Declarations

Conflict of interest

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Funding

This research was fully supported by RIIM Collaborative Platform Biology Biomolecular Structure Biodiversity Batch I number B-2657/III.5/FR.06.00/6/2024 and Research Collaboration Program Batch 2 of the BRIN Biological and Environmental Research Organization with number B-1779/III.5/PR.03.06/7/2025.

Author Contribution

Conceptualization, DP, So and TT; methodology, DP and So; software, DP; validation, DP, TT; formalanalysis, DP and SZ; investigation, DP; resources, Su and GR; data curation, DP, and SZ; writing—originaldraft preparation, DP, TT, and So; writing—review and editing, DP, So, TT and SJS; visualization, DP, andSJS; supervision, So, TT, AK and WBS; project administration, DP; funding acquisition, DP. All authorshave read and agreed to the published version of the manuscript.

Acknowledgement

Thanks to Applied Botany Research Center and Talent Management BRIN for funding DBR research, laboratories and scholarships. The authors would like to thank the two anonymous reviewers for their insightful feedback on an earlier draft of this paper.

Data Availability

The data sets generated during the current study was deposited in the Sequence Read Archive (NCBI/SRA) under BioProject number PRJNA1142076, accessible at the following link: https://www.ncbi.nlm.nih.gov/sra/PRJNA1142076.

References

1. Abdul Aziz M, Masmoudi K (2025) Molecular breakthroughs in modern plant breeding techniques. Horticultural Plant Journal 11(1):15–41. https://doi.org/10.1016/j.hpj.2024.01.004.
2. Alves SIA, Dantas CWD, Macedo DB, Ramos RTJ (2024) What are microsatellites and how to choose the best tool: a user-friendly review of SSR and 74 SSR mining tools. Frontiers in genetics 15:1474611. https://doi.org/10.3389/fgene.2024.1474611
3. Bidyananda N, Jamir I, Nowakowska K, Varte V, Vendrame WA, Devi RS, Nongdam P (2024) Plant Genetic Diversity Studies: Insights from DNA Marker Analyses. International Journal of Plant Biology 15(3): 607–640. https://doi.org/10.3390/ijpb15030046
4. Chen, JA., Yu, PJ., Jheng, SW. et al (2024) Mining expressed sequence tag (EST) microsatellite markers to assess the genetic differentiation of five *Hynobius* species endemic to Taiwan. Sci Rep 14: 20898. https://doi.org/10.1038/s41598-024-71887-1
5. Choo CY, Chan KL, Takeya K, Itokawa H (2001) Cytotoxic activity of *Typhonium flagelliforme* (Araceae). Phytother Res 15(3):260–2. https://doi.org/10.1002/ptr.717
6. Erdayani E, Erlinawati I, Nurhasanah AN, Hastilestari BR, Widyajayantie D, Rahmawati S, Wahyuni S (2023) Identification of *Typhonium flagelliforme* and its closely related species in Java Island using morphology and SRAP markers. AIP Conf Proc 2606 (1): 040012. https://doi.org/10.1063/5.0118839
7. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14(8):2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x
8. Farida Y, Irpan K, Fithriani L (2014) Antibacterial and antioxidant activity of keladi tikus leaves extract (*Typhonium flagelliforme*) (Lodd) Blume. Procedia Chemistry 13: 209 213. https://doi.org/10.1016/j.proche.2014.12.029
9. Feeney M, Murphy K, Lopilato J (2014) Designing PCR primers painlessly. Journal of microbiology & biology education 15(1): 28–29. https://doi.org/10.1128/jmbe.v15i1.634
10. Filho LSDG, Silva PP, Santos JM, Barbosa GVS, Ramalho-Neto CE, Soares L, Andrade JCF, Almeida C (2010) Genetic similarity among genotypes of sugarcane estimated by SSR and coefficient of parentage. Sugar Tech 12(2):145–149. https://doi.org/10.1007/s12355-010-0028-2.
11. Gao X, Su Q, Yao B, Yang W, Ma W, Yang B, Liu C (2022) Development of EST-SSR Markers Related to Polyphyllin Biosynthesis Reveals Genetic Diversity and Population Structure in *Paris polyphylla*. Diversity 14(8):589. https://doi.org/10.3390/d14080589
12. Gao X, Yourick MR, Campasino K, Zhao Y, Sepehr E, Vaught C, Sprando RL, Yourick JJ (2025) An updated comparison of microarray and RNA-seq for concentration response transcriptomic study: case studies with two cannabinoids, cannabichromene and cannabinol. BMC genomics 26(1): 392. https://doi.org/10.1186/s12864-025-11548-3
13. Garcia-Garcia T, Poncet S, Derouiche A, Shi L, Mijakovic I, Noirot-Gros MF (2016) Role of Protein Phosphorylation in the Regulation of Cell Cycle and DNA-Related Processes in Bacteria. Frontiers in

microbiology 7: 184. https://doi.org/10.3389/fmicb.2016.00184

14. Ge SX, Jung D, Yao R (2020) ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics (Oxford, England) 36(8):2628–2629. https://doi.org/10.1093/bioinformatics/btz931
15. Huang P, Karagianis G, Waterman PG (2004) Chemical constituents from *Typhonium flagelliforme*. Zhong yao cai = Zhongyaocai = Journal of Chinese Medicinal Materials 27(3):173–175. PMID: 15272778.
16. Jagodzik P, Tajdel-Zielinska M, Ciesla A, Marczak M and Ludwikow A (2018) Mitogen-Activated Protein Kinase Cascades in Plant Hormone Signaling. Front Plant Sci 9:1387. https://doi.org/10.3389/fpls.2018.01387
17. Kapoor B, Sharma M, Sharma R, Zadokar A, Thakur A, Sharma P, Kumar S, Rozar KP, Kumar KS, Hegde N, Pandey D (2023) De novo transcriptome profiling and development of novel secondary metabolites based genic SSRs in medicinal plant *Phyllanthus emblica* L. (Aonla). Scientific reports 13(1):17319. https://doi.org/10.1038/s41598-023-44317-x
18. Li YL, Liu JX (2018) StructureSelector: A web based software to select and visualize the optimal number of clusters using multiple methods.Molecular Ecology Resources 18:176–177. https://doi.org/10.1111/1755-0998.12719
19. Li J, Wu K, Li L et al (2023) Transcriptomic analysis reveals biosynthesis genes and transcription factors related to leaf anthocyanin biosynthesis in *Aglaonema commutatum*. BMC Genomics 24: 28. https://doi.org/10.1186/s12864-022-09107-1
20. Li W, Liu S, Wang S, Li Y, Kong D, Wang A (2024) A single origin and high genetic diversity of cultivated medicinal herb *Glehnia littoralis* subsp. littoralis (Apiaceae) deciphered by SSR marker and phenotypic analysis. PloS one 19(8):e0308369. https://doi.org/10.1371/journal.pone.0308369
21. Liu H, Zhang Y, Wang Z, Su Y, Wang T (2021) Development and application of EST-SSR markers in *Cephalotaxus oliveri* from transcriptome sequences. Frontiers in genetics 12:759557. https://doi.org/10.3389/fgene.2021.759557
22. Manna M, Rengasamy B, Sinha AK (2023) Revisiting the role of MAPK signalling pathway in plants and its manipulation for crop improvement. Plant, cell & environment 46(8): 2277–2295. https://doi.org/10.1111/pce.14606
23. Mark A Szenteczki, Adrienne L Godschalx, Jérémy Gauthier, Marc Gibernau, Sergio Rasmann, Nadir Alvarez (2022) Transcriptomic analysis of deceptively pollinated *Arum maculatum* (Araceae) reveals association between terpene synthase expression in floral trap chamber and species-specific pollinator attraction. G3 Genes|Genomes|Genetics 12(9): jkac175. https://doi.org/10.1093/g3journal/jkac175
24. Merritt BJ, Culley TM, Avanesyan A, Stokes R, Brzyski J (2015) An empirical review: Characteristics of plant microsatellite markers that confer higher levels of genetic variation. Applications in plant sciences 3(8): apps.1500025. https://doi.org/10.3732/apps.1500025
25. Mirgane NA, Chandore A, Shivankar V, Gaikwad Y, Wadhawa GC (2021) Phytochemical study and screening of antioxidant, anti-inflammatory *Typhonium flagelliforme*. Research Journal of Pharmacy

and Technology 14: 2686–2690. https://doi.org/10.52711/0974-360X.2021.00474

26. Mohan S, Bustamam A, Ibrahim S, Al-Zubairi AS, Aspollah M, Abdullah R, Elhassan MM (2011) In Vitro Ultramorphological Assessment of Apoptosis on CEMss Induced by Linoleic Acid-Rich Fraction from *Typhonium flagelliforme* Tuber. Evidence-based complementary and alternative medicine: eCAM 2011: 421894. https://doi.org/10.1093/ecam/neq010
27. Mortazavian SSM, Arshadi-Bidgoli M, Sadeghi D, Bakhtiarizadeh MR (2024) Identified and validation of EST-SSR in the transcriptome sequences by RNA-Seq in cumin (*Cuminum Cyminum* L.). Plant Gene 40:100477. https://doi.org/10.1016/j.plgene.2024.100477.
28. Novaes E, Drost DR, Farmerie WG, Pappas GJ Jr, Grattapaglia D, Sederoff RR, Kirst M (2008) High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. BMC genomics 9: 312. https://doi.org/10.1186/1471-2164-9-312
29. Parchman TL, Geist KS, Grahnen JA, Benkman CW, Buerkle CA (2010) Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. BMC genomics 11: 180. https://doi.org/10.1186/1471-2164-11-180
30. Panahi B, Jalaly HM, Hamid R (2024) Using next-generation sequencing approach for discovery and characterization of plant molecular markers. Current Plant Biology 40:100412. https://doi.org/10.1016/j.cpb.2024.100412.
31. Perrier X, Jacquemoud-Collet JP (2006) DARwin software http://darwin.cirad.fr/darwin
32. Purwoko D, Cartealy IC, Tajuddin T, Dinarti D, Sudarsono S (2019) SSR identification and marker development for sago palm based on NGS genome data. Breeding Science 69(1):1–10. https://doi.org/10.1270/jsbbs.18061
33. Purwoko D, Zulaeha S, Tajuddin T, Khairiyah H, Fauzi RZ, Priyanti (2021) SSR Markers Characterization for Temu Ireng (*Curcuma aeruginosa* Roxb.) Generated From EST of *Curcuma longa*. Jurnal Bioteknologi Dan Biosains Indonesia. 8(2):160–173.
34. Purwoko D, Zulaeha S, Rahmadara G, Suparjo, Tajuddin T, Kurniawati A, Suwarno WB, Sobir (2025) Simple sequence repeats (SSRs) discovery, characterization and validation from de novo transcriptome assembly of *Typhonium flagelliforme.* BIO Web Conf 186(2025):03007. https://doi.org/10.1051/bioconf/202518603007
35. Quévillon Huberdeau M, Simard MJ (2019) A Guide to microRNA-Mediated Gene Silencing. FEBS Journal 286(4): 642–652. https://doi.org/10.1111/febs.14666.
36. Ren D, Wu W, Wen Q, Li Y (2025) Development of EST-SSR markers based on transcriptome for genetic analysis in *Radix Ardisia*. PeerJ 13:e19560. https://doi.org/10.7717/peerj.19560
37. Satam H, Joshi K, Mangrolia U, Waghoo S, Zaidi G, Rawool S, Thakare RP, Banday S, Mishra AK, Das G, Malonia SK (2023) Next-Generation Sequencing Technology: Current Trends and Advancements. Biology 12(7):997. https://doi.org/10.3390/biology12070997
38. Schoenbrunner NJ, Gupta AP, Young KKY, Will SG (2017) Covalent modification of primers improves PCR amplification specificity and yield. Biology methods & protocols 2(1): bpx011. https://doi.org/10.1093/biomethods/bpx011

39. Scholpp A-C, Schilbert HM, Viehöver P, Weisshaar B, Beckstette M, Neumann JM, Bednarz H and Niehaus K (2024) Differential gene expression in leaves and roots of *Hydrangea serrata* treated with aluminium chloride. Front Plant Sci 15:1412189. https://doi.org/10.3389/fpls.2024.1412189
40. Song HY, Shen R, Mahasin H, Guo YN, Wang DG (2023) DNA replication: Mechanisms and therapeutic interventions for diseases. MedComm 4(1): e210. https://doi.org/10.1002/mco2.210
41. Squirrell J, Hollingsworth M, Woodhead M, Russell J, Lowe AJ, Gibby M, Powell W (2003) How much effort is required to isolate nuclear microsatellites from plants? Molecular Ecology 12: 1339–1348. https://doi.org/10.1046/j.1365-294x.2003.01825.x
42. Senan S, Kizhakayil D, Sasikumar B, Sheeja TE (2014) Methods for Development of Microsatellite Markers: An Overview. Not Sci Biol 6(1):1–13. https://doi.org/10.15835/nsb619199.
43. Septaningsih DA, Yunita A, Putra CA, Herawati I, Achmadi SS, Heryanto R, Rafi M (2021) Phenolics profiling and free radical scavenging activity of *Annona muricata*, *Gynura procumbens*, and *Typhonium flagelliforme* leaves extract. Indonesian Journal of Chemistry 21: 1140–1147. https://doi.org/10.22146/ijc.62124
44. Setiawati A, Immanuel H, Utami MT (2016) The inhibition of *Typhonium flagelliforme* Lodd. Blume leaf extract on COX-2 expression of WiDr colon cancer cells. Asian Pacific Journal of Tropical Biomedicine 6(3): 251–255. https://doi.org/10.1016/j.apjtb.2015.12.012
45. Sianipar NF, Khoirunnisa A, Hadisaputri YE, Salam S, Purnamaningsih R, Gautama So I (2023) Mutant Plant Tipobio Variety of Rodent Tuber (*Typhonium Flagelliforme*): Fatty Acids Compounds and in Vitro Anticancer Activity. E3S Web of Conf 388(2023):01032. https://doi.org/10.1051/e3sconf/202338801032
46. Srivastava S, Aspak, Mukherjee S, Datta S, Burman S, Rana M, Kumar R, Katyayani KKS, Karnwal A, Kumar S, Shukla M, Kumar R (2023) Microsatellite markers for crop improvement: A review: Role of microsatellite markers. Journal of Applied and Natural Science 15(3):1018–1035. https://doi.org/10.31018/jans.v15i3.4615
47. Topu M, Sesiz U, Bektaş H, Toklu F, Özkan H (2023) Next-Generation-Sequencing-Based Simple Sequence Repeat (SSR) Marker Development and Linkage Mapping in Lentil (*Lens culinaris* L.). Life (Basel, Switzerland) 13(7):1579. https://doi.org/10.3390/life13071579
48. Umang U, Bharti PK, Husain A (2022) Bibliometric analysis of research aspects of simple sequence repeats in cultivars. International Journal of Health Sciences 6(S4):7922–7933. https://doi.org/10.53730/ijhs.v6nS4.11756
49. Unamba CIN, Nag A, Sharma RK (2015) Next generation sequencing technologies: the doorway to the unexplored genomics of non-model plants. Front Plant Sci 6:1074. https://doi.org/10.3389/fpls.2015.01074
50. Varshney RK, Chabane K, Hendre PS, Aggarwal RK, Graner A (2007) Comparative assessment of EST-SSR, EST-SNP and AFLP markers for evaluation of genetic diversity and conservation of genetic resources using wild, cultivated and elite barleys. Plant Science 173: 638–649. https://doi.org/10.1016/j.plantsci.2007.08.010

51. Vieira MLC, Santini L, Diniz AL, Munhoz C de F (2016) Microsatellite markers: What they mean and why they are so useful. Genet Mol Biol 39(3):312–328. https://doi.org/10.1590/1678-4685-GMB-2016-0027.
52. Wang L, Yin J, Zhang P, Han X, Guo W, Li C (2017) De novo assembly and characterization of transcriptome and microsatellite marker development for Taro (*Colocasia esculenta* (L.) Schott.). International Journal of Genetics and Molecular Biology 9(5): 26–36. https://doi.org/10.5897/IJGMB2017.0155
53. Wei Z, Sun Z, Cui B, Zhang Q, Xiong M, Wang X, Zhou D (2016) Transcriptome analysis of colored calla lily (*Zantedeschia rehmannii* Engl.) by Illumina sequencing: de novo assembly, annotation and EST-SSR marker development. PeerJ 4: e2378. https://doi.org/10.7717/peerj.2378
54. Worth JRP, Chang KS, Ha YH, Qin A (2019) Development of microsatellite markers for the Japanese endemic conifer *Thuja standishii* and transfer to other East Asian species. BMC Res Notes 12: 694. https://doi.org/10.1186/s13104-019-4716-z.
55. Xu L, Li P, Su J, Wang D, Kuang Y, Ye Z, Chen M (2023) EST-SSR development and genetic diversity in the medicinal plant *Pseudostellaria heterophylla* (Miq.) Pax. Journal of Applied Research on Medicinal and Aromatic Plants 33:100450. https://doi.org/10.1016/j.jarmap.2022.100450
56. Zalapa JE, Cuevas H, Zhu HY, Steffan S, Senalik D, Zeldin E, et al (2012) Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. Am J Bot 99(2):193–208. https://doi.org/10.3732/ajb.1100394
57. Zheng X, Pan C, Diao Y et al (2013) Development of microsatellite markers by transcriptome sequencing in two species of *Amorphophallus* (Araceae). BMC Genomics 14: 490. https://doi.org/10.1186/1471-2164-14-490
58. Zulaeha S, Fadhullah H, Cartealy IC et al (2023) Assessment of genetic diversity among sago palms (*Metroxylon sagu* Rottb.) in Bengkulu, Indonesia using simple sequence repeats. Genet Resour Crop Evol 71: 3083–3095. https://doi.org/10.1007/s10722-023-01819-x

Figures



Figure 1

Map depicting different locations from where experimental material procured in the study.

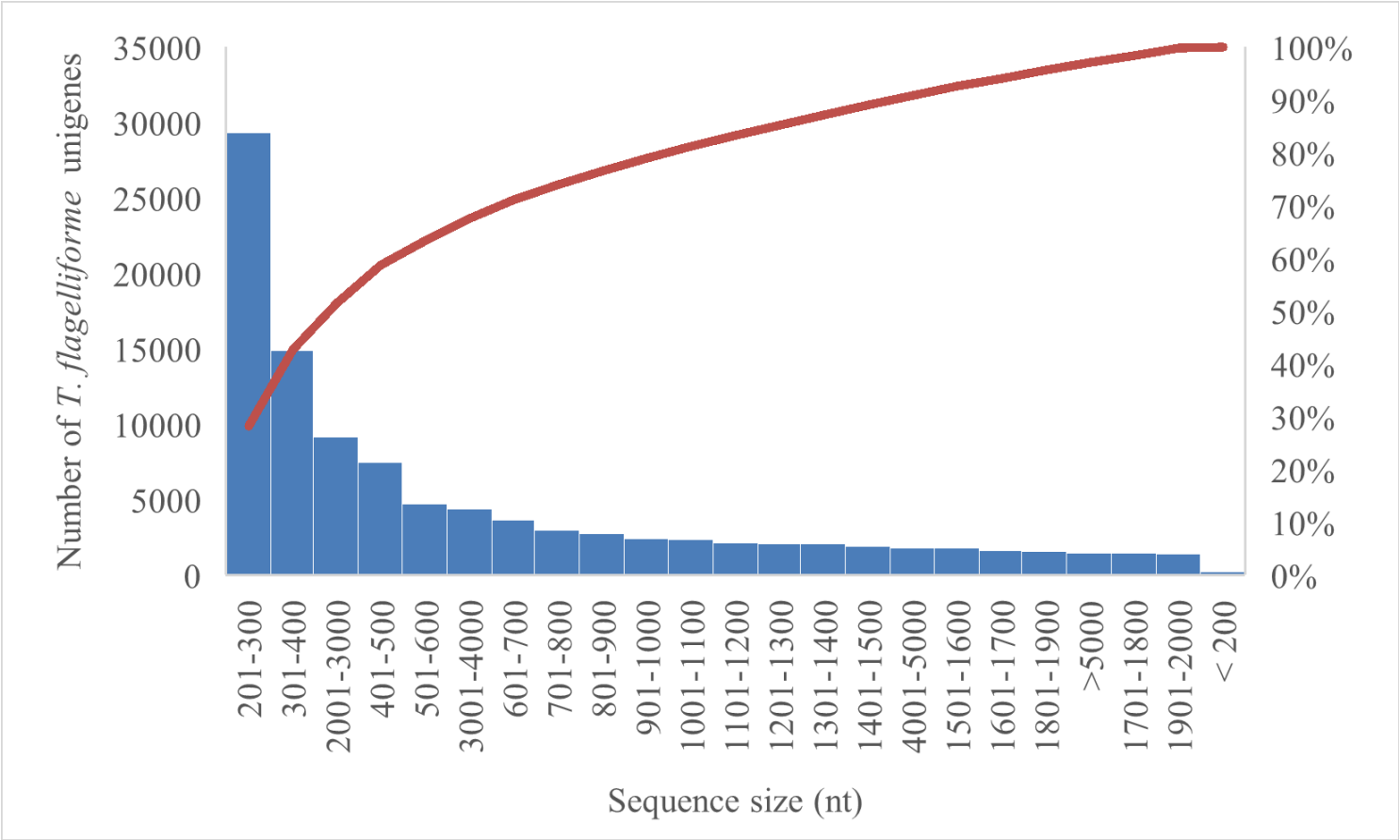


Figure 2

Sequence length distribution of the assembled unigenes.

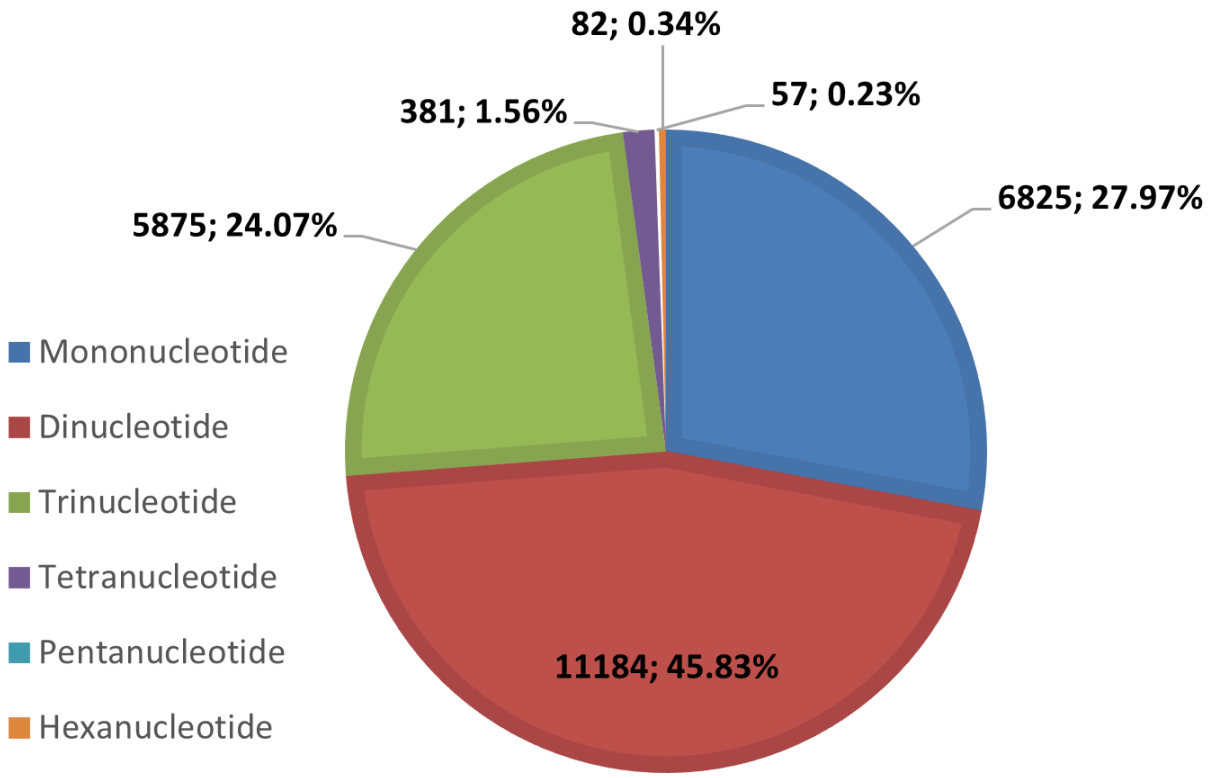


Figure 3

Frequency and Distribution of EST-SSRs in *T. flagelliforme*

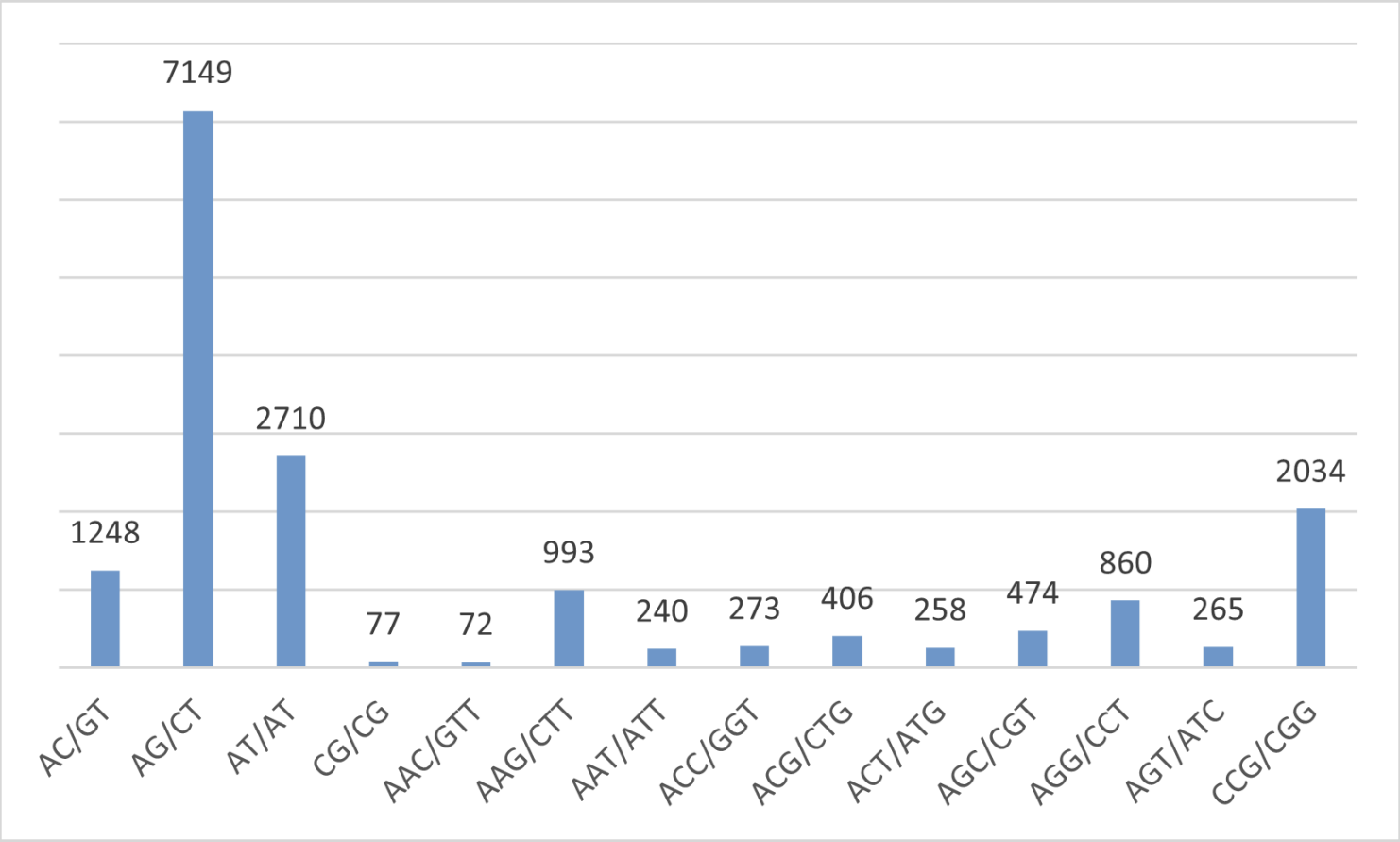


Figure 4

Frequency of identified EST-SSR motifs: dinucleotide and trinucleotide

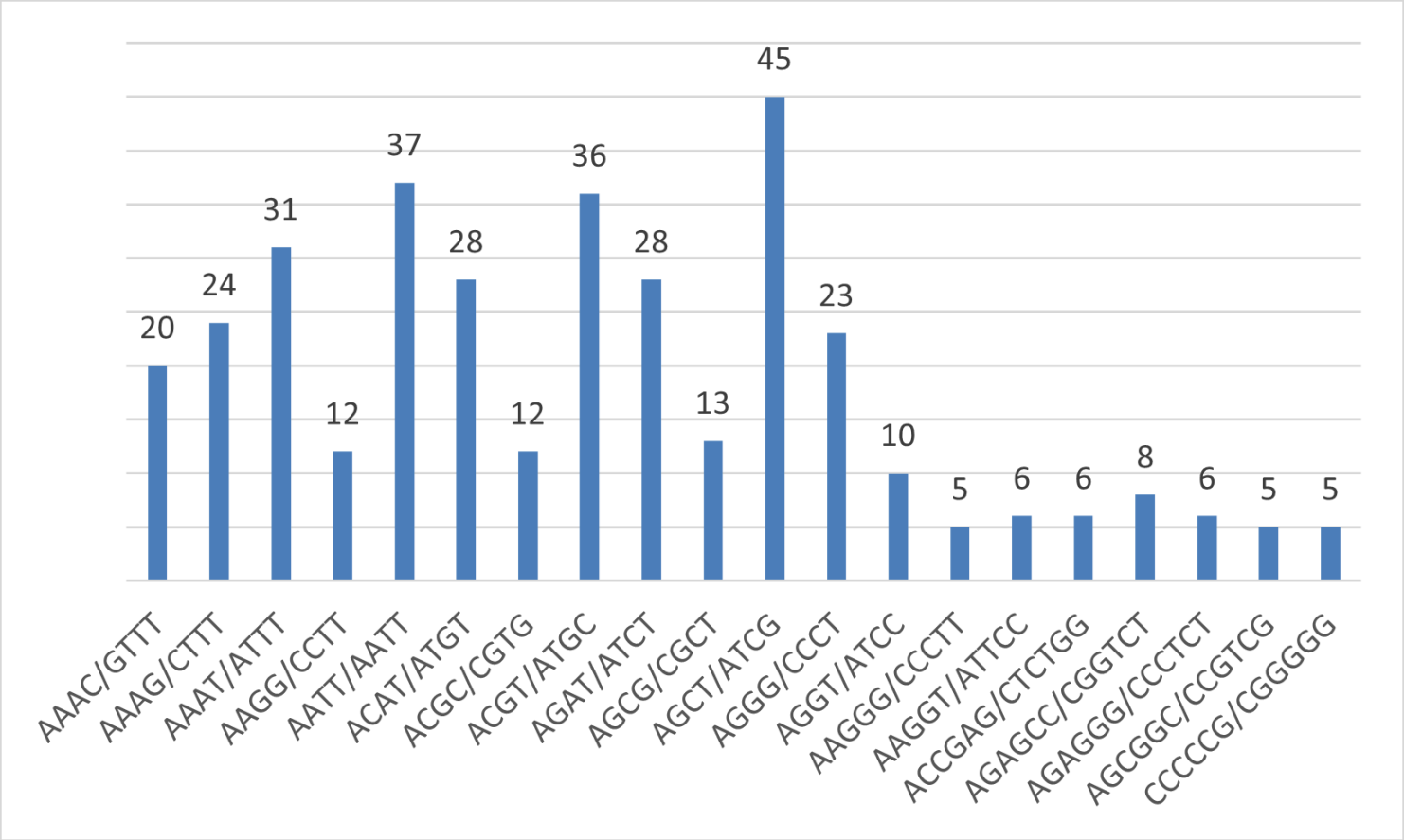


Figure 5

Frequency of identified EST-SSR motifs: tetranucleotide, pentanucleotide and hexanucleotide

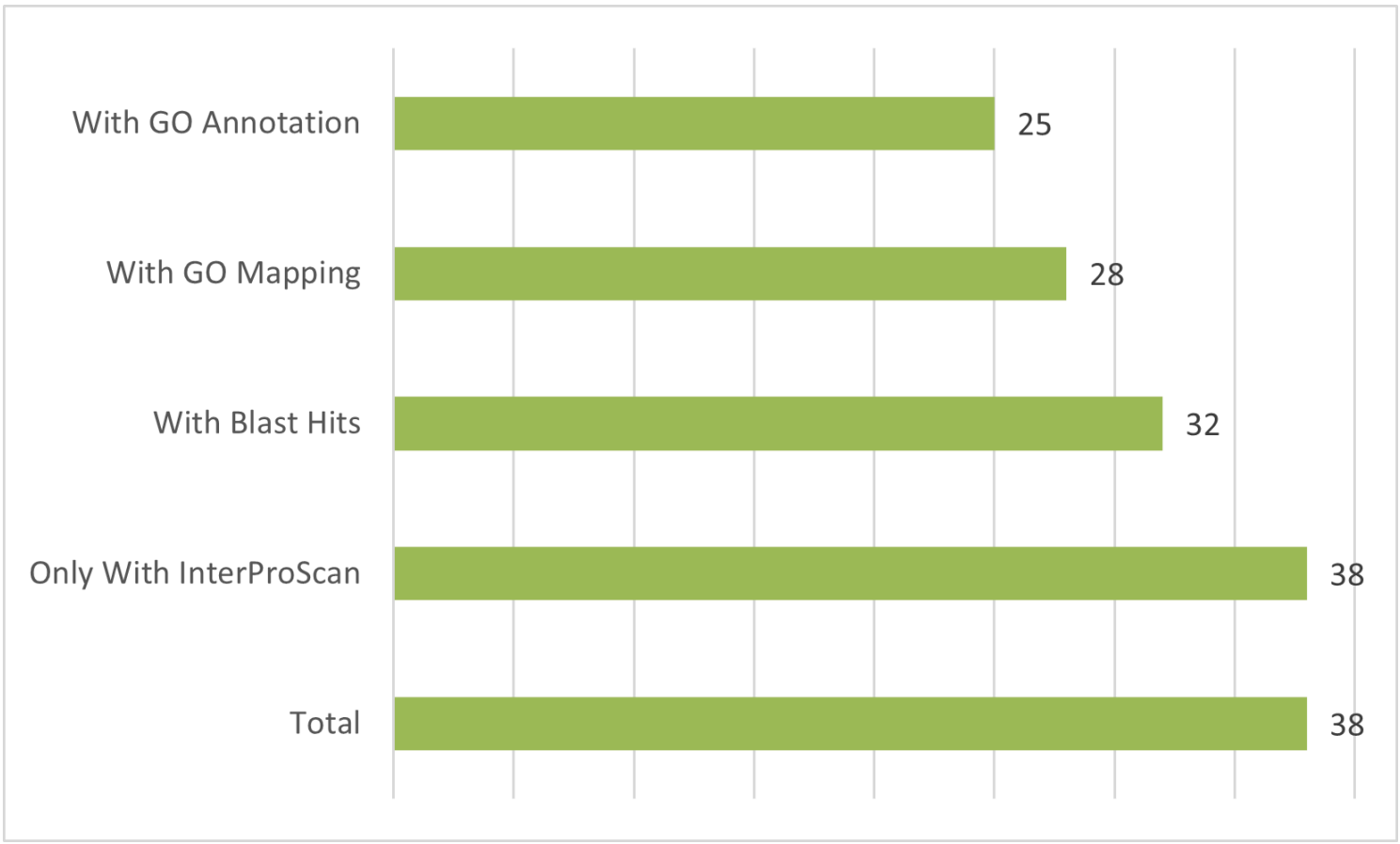


Figure 6

Distribution of SSR-containing unigenes from ShinyGO analysis

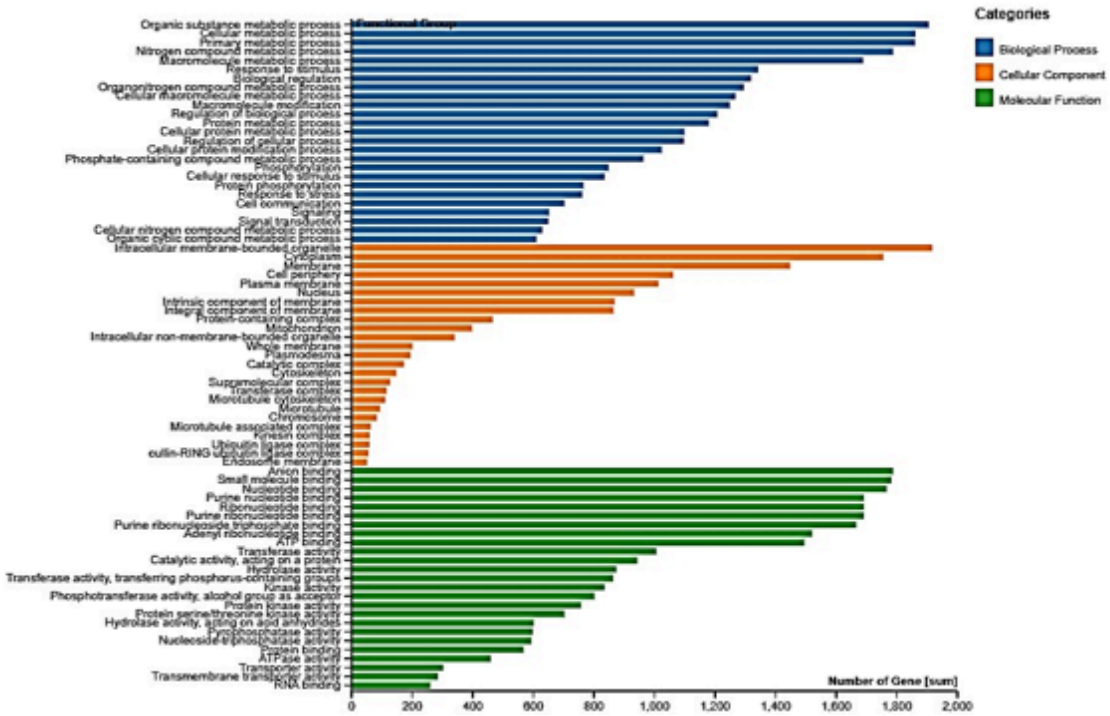


Figure 7

Gene ontology (GO) classification of SSR-containing unigenes.

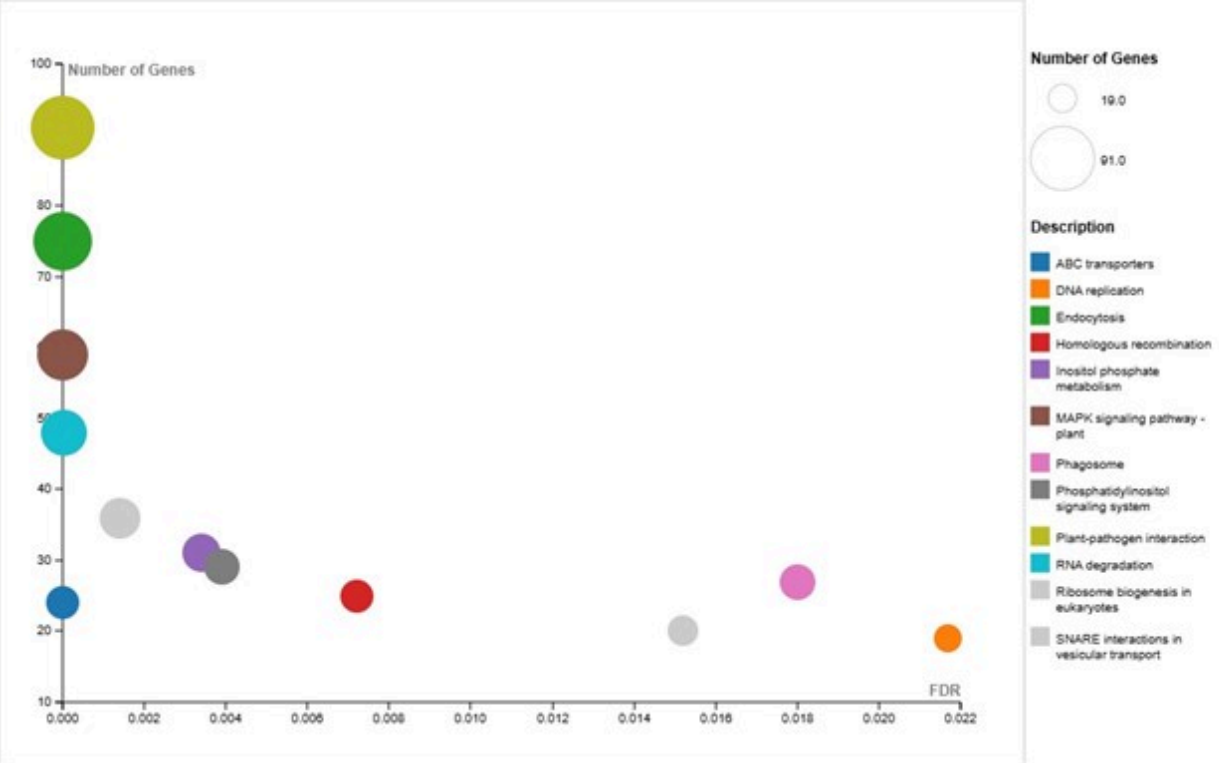


Figure 8

KEGG pathway enrichment analysis of SSR-containing unigenes.

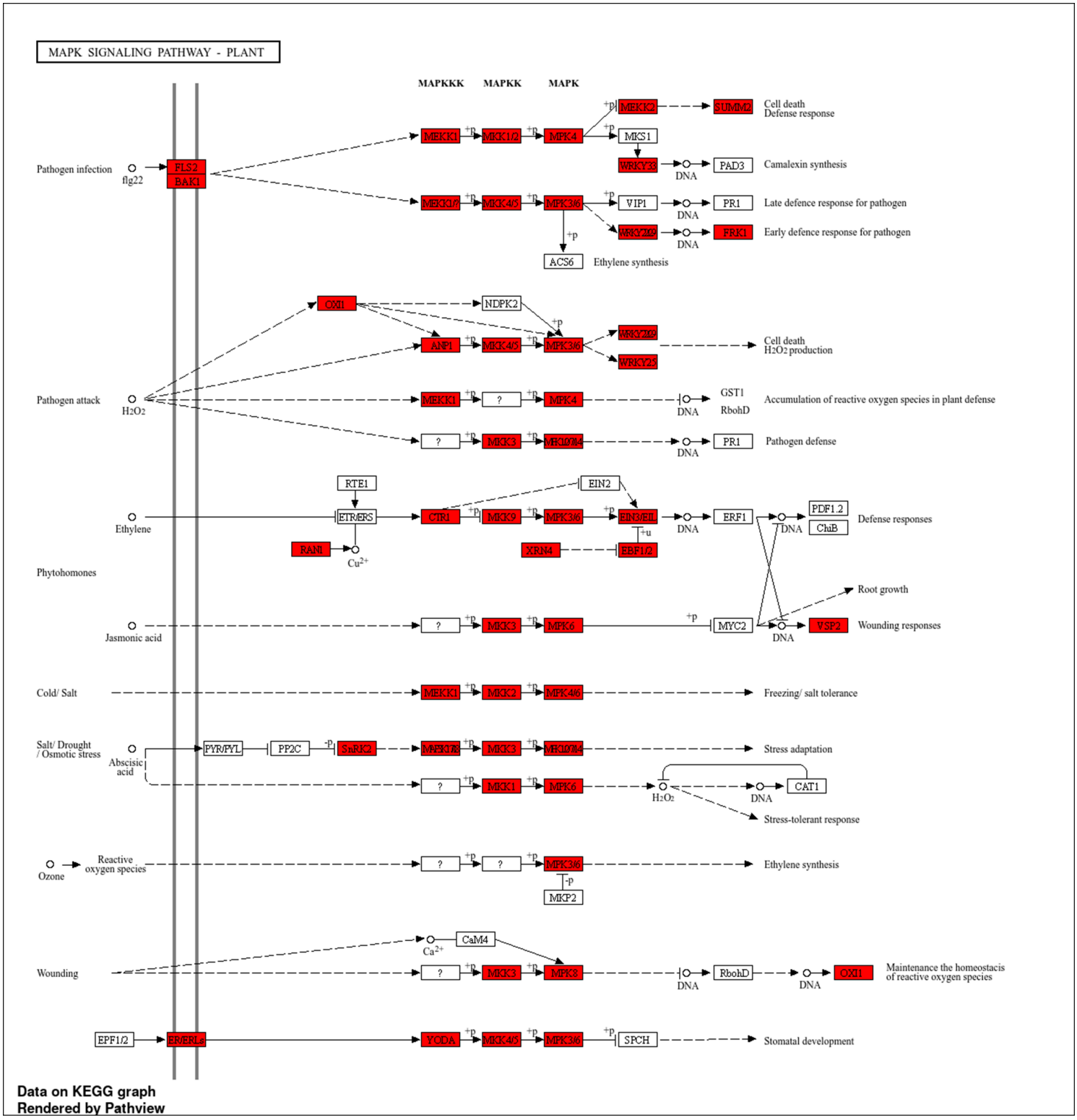


Figure 9

KEGG of mitogen-activated protein kinase (MAPK) pathway enrichment analysis of SSR-containing unigenes.

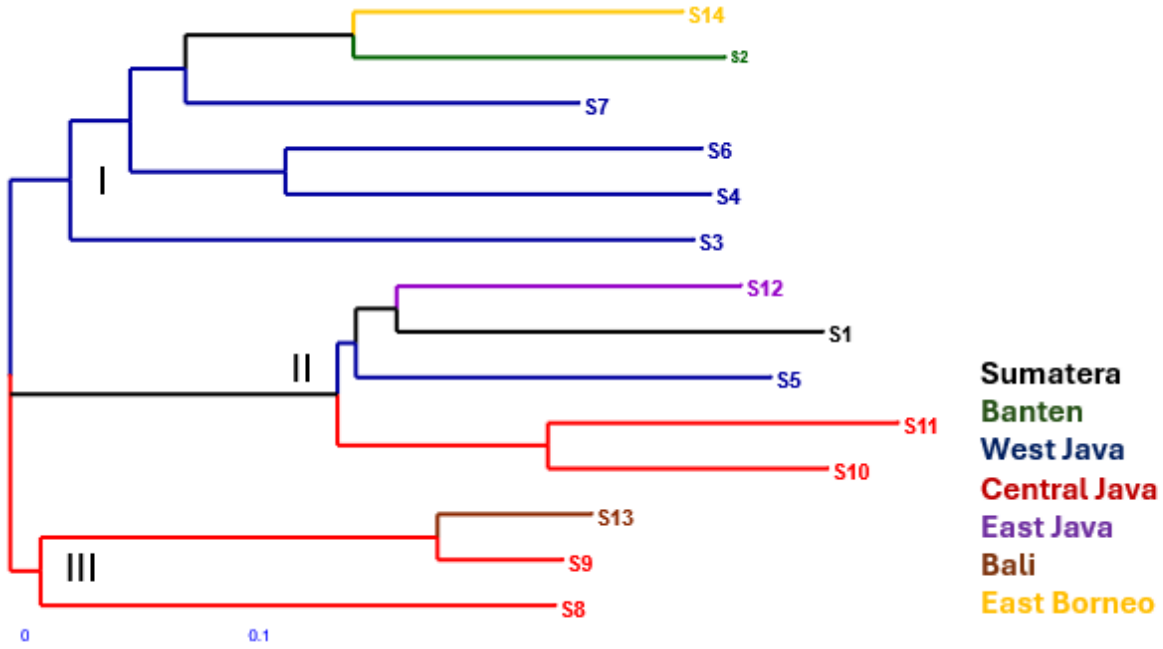


Figure 10

An NJ dendrogram of 14 *T. flagelliforme* accessions based on 15 polymorphic EST-SSR markers.

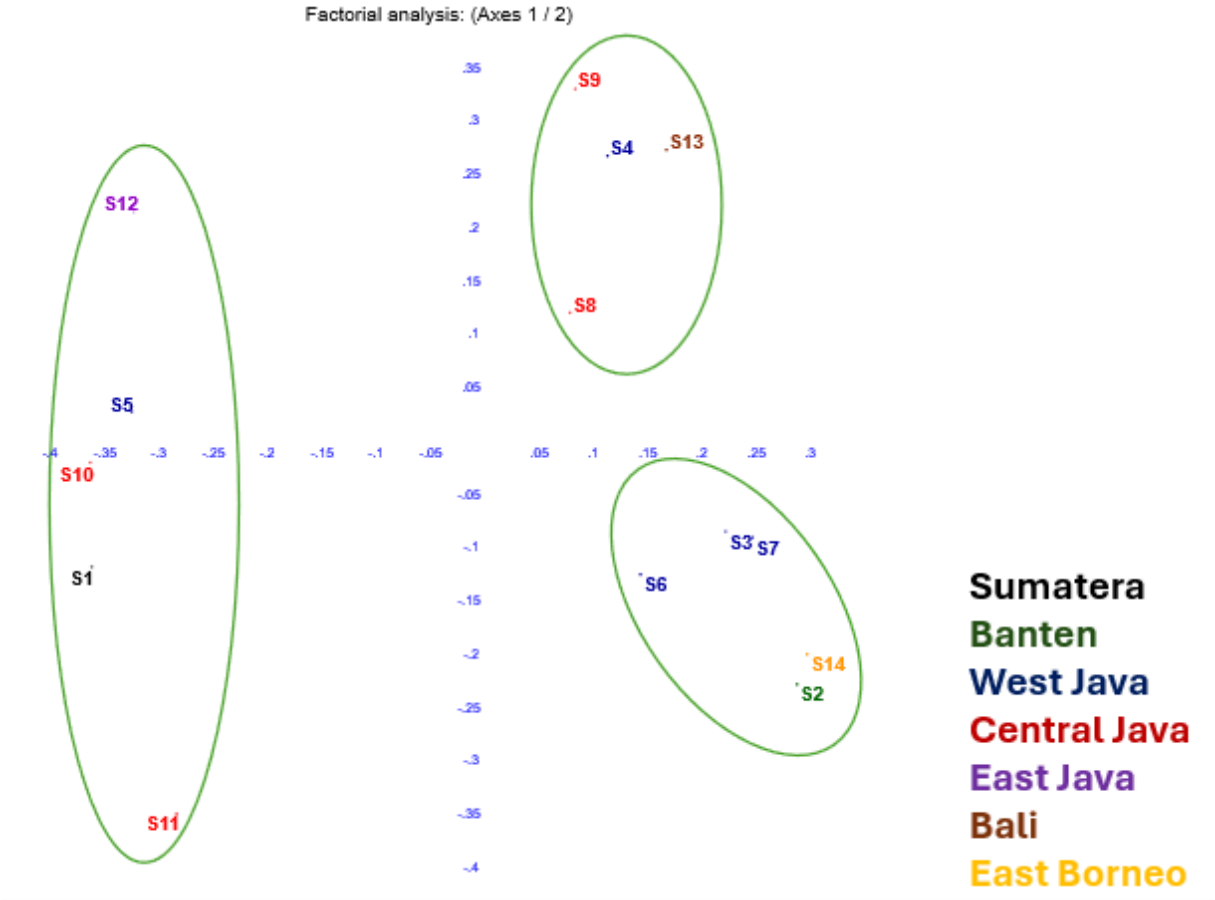


Figure 11

Factorial analysis based on Eigen values calculated from 15 SSR markers.

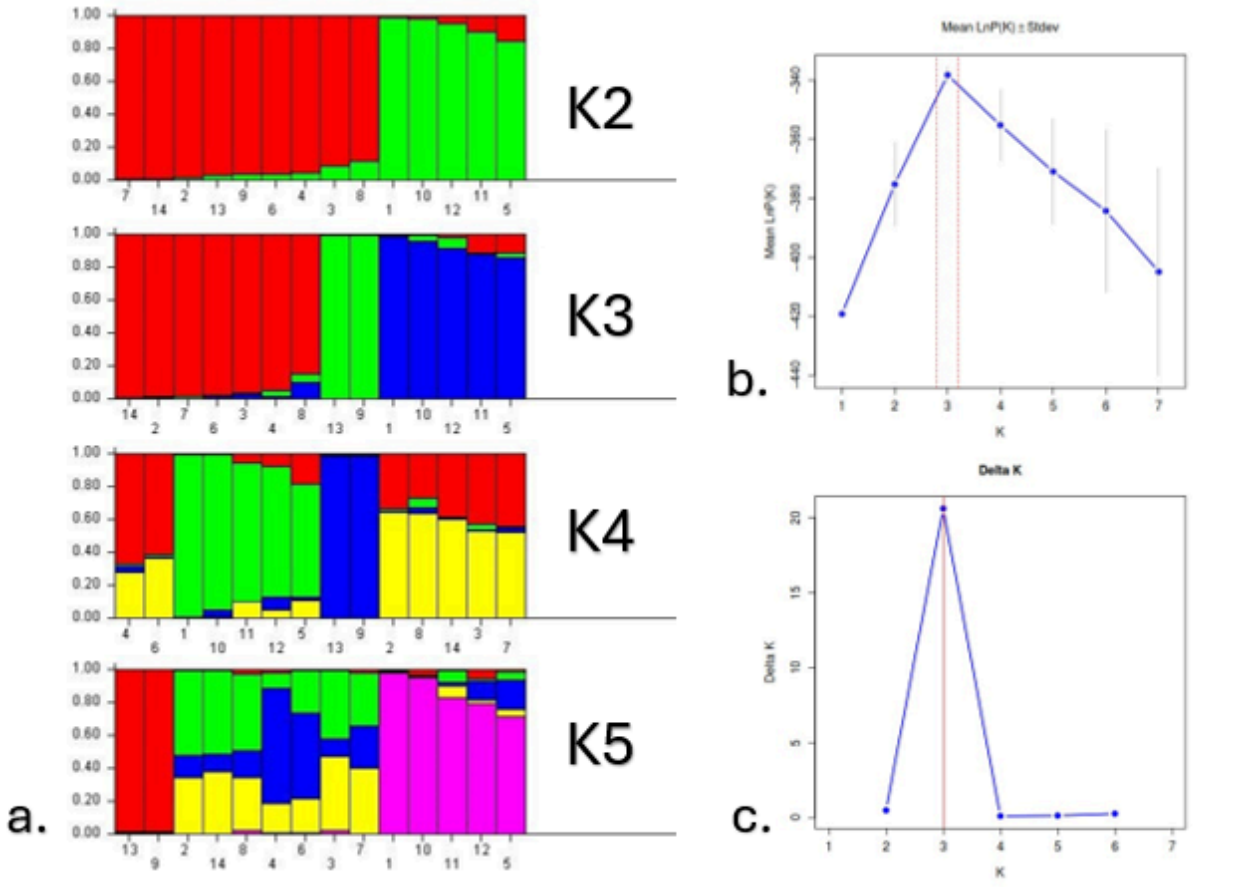


Figure 12

Results of STRUCTURE analysis for 14 populations *T. flagelliforme* using 15 EST-SSR markers: (a) Estimation of population structure based on STRUCTURE analysis. (b) Plot mean likelihood L(K) and variance per K value. (c) Estimation of population using Delta K value with cluster K ranging from 1 to 7.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.docx
- SupplementaryFig.1.docx