PLOS ONE

RESEARCH ARTICLE

Aster spathulifolius Maxim. a leaf transcriptome provides an overall functional characterization, discovery of SSR marker and phylogeny analysis

Sivagami Jean Claude, SeonJoo Park[ID]*

Department of Life Sciences, Yeungnam University, Gyeongsan, Gyeongbuk, South Korea

* sjpark01@ynu.ac.kr

Abstract

Aster spathulifolius Maxim. is belongs to the Asteraceae family, which is distributed only in Korea and Japan. The species is traditionally a medicinal plant and is economically valuable in the ornamental field. On the other hand, the *Aster* genus, among the Asteraceae family, lacks genomic resources and its molecular functions. Therefore, in our study the high-throughput RNA-sequencing transcriptome data of *A. spathulifolius* were obtained to identify the molecular functions and its characterization. The *de novo* assembly produced 98660 uniqueness with an N50 value of 1126bp. Total unigenes were procure to analyze the functional annotation against databases like non-redundant protein, Pfam, Uniprot, KEGG and Gene ontology. The overall percentage of functional annotation to the nr database (43.71%), uniprotein database (49.97%), Pfam (39.94%), KEGG (42.3%) and to GO (30.34%) were observed. Besides, 377 unigenes were found to be involved in the terpenoids pathway and 666 unigenes were actively engaged in other secondary metabolites synthesis, given that 261 unigenes were within phenylpropanoid pathway and 81 unigenes to flavonoid pathway. A further prediction of stress resistance (9,513) unigenes and transcriptional factor (3,027) unigenes in 53 types were vastly regulated in abiotic stress respectively in salt, heat, MAPK and hormone signal transduction pathway. This study discovered 29,692 SSR markers that assist the genotyping approaches and the genetic diversity perspectives. In addition, eight Asteraceae species as in-group together with one out-group were used to construct the phylogenetic relationship by employing their plastid genome and single-copy orthologs genes. Among 50 plastid protein-coding regions, *A. spathulifolius* is been closely related to *A. annua* and by 118 single copy orthologs genes, *O. taihangensis* is more neighboring species to *A. spathulifolius*. Apart from this, *A. spathulifolius* and *O. taihangensis*, genera have recently diverged from other species. Overall, this research gains new insights into transcriptome data by revealing and exposing the secondary metabolite compounds for drug development, the stress-related genes for producing resilient crops and an ortholog gene of *A. spathulifolius* for the robustness of phylogeny reconstruction among Asteraceae genera.

Introduction

The *Aster* genus is a perennial flowering plant in the family of Asteraceae. There are 32,020 species contains 190 genera of which many of the asters are in the Astereae tribe, the second largest tribe [1]. The *A. spathulifolius* Maxim (seashore spatulate aster) is a diploid species belonging to the genus *Astersa*. It is an endemic plant occurring only in the coastal region of Korea and Japan. *A. spathulifolius* is widely distributed in Dokdo Island, which composed of volcanic rocks, pyroclastic sedimentary rocks and highly salinity environments, with smog and strong wind [2], and also it has been reported that the plants are difficult to grow under such challenging environmental conditions [3, 4]. The *A. spathulifolius* has high economic value, both in medicine and for ornamental purposes [5]. In medicinal perspectives, *A. spathulifolius* used to treat asthma and diuresis, and in human cancer cell lines with labdane diterpene compounds (secondary metabolites) [6]. Besides, *A. spathulifolius* was studied for antibacterial activity and carcinoma cell lines against oral pathogens [7] and its simple leaf extraction were used to treat hyperglycemia that inhibits the gluconeogenesis by accelerating glucose and lipid metabolism and helps against diabetes and diabetic complications [8, 9]. *In vivo* study on mice (diet-induced obesity) demonstrated, an anti-adipogenic, anti-lipogenic, anti-obesity and anti-hyperlipidemia effects [10, 11] in the manner of weight loss and fat mass reduction [12]. To date, the complete plastid genome with 13.6 kb inversion has been documented as genomics data for *A. spathulifolius* in NCBI [13]. In this current study, we used RNA-sequencing data to identify the functional genes involved in the leaf of *A. spathulifolius* for the first time. Moreover, the total unigenes were highly used to evaluate their multiple functional annotations on *A. spathulifolius*. The species was further elaborated on the production of different antibiotic and secondary metabolites for drug development. Furthermore, this study will be useful to identify stress-specified unigene, which support crops for better improvement. Phylogeny based on plastid protein-coding region and single-copy nuclear gene (orthologs gene) among Asteraceae genera along with *A. spathulifolius* were documented on this study.

Materials and methods

Plant material, RNA isolation and illumina sequencing

The whole plant of *A. spathulifolius* was collected from Dokdo Island (South Korea). The herbarium was prepared and deposited at Yeungnam University Herbarium (YNUH2018asp001) for future purposes. For the isolation of RNA, 100 mg of leaf tissue of the plant collected from the habitat condition (Dokdo island, SK) was used, and further preserved in liquid nitrogen and stored at—80˚C. A total RNA isolation was performed following the protocol of Breitler *et al.* [14]. Using DNase 1 (Invitrogen DNase 1 Amplification kit, USA) treatment to obtain purified RNA from DNA contamination. The total RNA (47.50 μg) were used to sequence via Illumina HiSeq 2000 platform at Lab Genomics (Seongnam, South Korea), to generate total 21 GB reads of 151 bp of paired-end library preparation, the TruSeq standard mRNA kit was used for library preparation.

De novo assembly, alignment and transcript quantification

The paired-end raw reads were quality assessed by FastQC v0.11.218. The raw paired-end reads were quality assessed by FastQC v0.11.218. Phred score > = 30 reads were retained using Trimmomatic tool to trim reads below 30% quality from the 148035505 raw reads. The filtered reads was further used for transcriptome assembly using Trinity v2.8.2 tool [15]. Trinity, a *de novo* assembler tool based on sequence data into many individual de Bruijn graphs with

default k-mer size (k = 25) and which followed by Inchworm is to generate full-length isoforms from raw reads, Chrysalis cluster the contig which was generated by Butterfly modules to generate full-length transcripts or unigenes. Later Transdecoder v5.5.0 tool was used to retain peptide sequences for further analysis of blast. The Tophat v2.1.0 –Bowtie v2.3.5.1 alignment was used to align the raw reads to the retained transcripts [16, 17]. Estimation of abundance unigene was determined by RSEM v1.3.1(RNA-Seq by Expectation-Maximization) tool [18], which first generates and processes of total transcripts and then aligns to raw reads of *A. spathulifolius*. RSEM is used calculate the fragments per kilobase per million (FPKM) and transcripts per million (TPM) values of total unigenes to analysis the maximum number of gene expressed or abundant in *A. spathulifolius*. The both FPKM and TPM values are calculated to estimate the expression levels of unigenes involved in phenylpropanoid pathway (PPP) and caffeoyl quinic acid production of secondary metabolites.

Functional annotation and analysis

The *de novo* assembled sequences of *A. spathulifolius* were used to identify the most significant match against the National Centre for Biotechnology Information (NCBI) database like the non-redundant protein (nr) and nucleotide using BLASTX tool with an E-value (1E-05). The BLASTP tool used against the Uniprot database (https://www.uniprot.org/) was downloaded for uniprotein identification with E-vaue of 1E-05. Pfam related unigene was retained using HMMER. The nr, Pfam and Uniprot database results are further used to retrieve Gene Ontology (GO) terms. The retrieved GO terms were classified into three categories: Cellular Component, Molecular Function and Biological process. Further, KEGG Automated Annotation Server (KAAS) was used for pathway mapping of species specified data with 1E-05 parameter.

Transcription factor identification and stress gene finder

The transcriptional factor unigenes were identified using http://planttfdb.cbi.pku.edu.cn/ tool. The identified TFs were uploaded to obtain the Gene ontology function. The stress-related unigenes were estimated using the TRAPID program web tool (http://bioinformatics.psb.ugent.be/webtools/trapid/) and BlastX (TAIR database). The parameter used was 1E-05 to both the transcriptional factor and TRAPID analysis. Further confirmation of stress resistance genes by BlastP against the non-redundant proteins and the Uniprotein database has been performed.

Identification of SSRs marker

The MicroSAtellite identification tool (MISA) was used to identify the simple sequence repeats marker from the total assembled unigenes of *A. sphathulifolius* [19]. The parameters were set to identify perfect repeat nucleotide motifs with a minimum threshold of (1/10) (2/6) (3/5) (4/5) (5/5), and (6/5).

Phylogenetic analysis of plastid and nuclear gene

To determine the Aster group phylogenetic relationship with other known Asteraceae species (S1 Table), 50-plastid protein-coding region was chosen from the entire plastid genome. The plastid genomes of these species were retrieved from NCBI database. MAFFT was used to align the 50-plastid protein-coding region and then aligned sequences were loaded to the fast tree to construct the phylogeny tree among nine species. For the plastid-based phylogeny and visualization, the GeneiousR11 version was used. Transcriptome data for orthologs genes were used to draw the phylogenetic relationship between the eight species from Asteraceae as in-

groups and one out-group from Goodeniaceae family. The SRA raw database from NCBI was downloaded using the SRA tool kit and fastq-dumb tool to extract the leaf and right raw reads, *de novo* assembly done by Trinity with Trimmomatic tool and Transdecoder to find ORF (S2 Table). The longest ORF coding region was extracted from all contig to minimize the layoff. Then, the cdhit tool [20] was used to reduce the redundancy of amino acids. Orthofinder [21] tool were finally ran to analyze the single-copy orthologs genes (auto selection of out-group) to construct phylogeny tree.

Results and discussions

In total, 148035505 raw reads were generated from *A. spathulifolius* leaf transcriptome that accounts for approximately 21 GB of paired-end sequencing data and in that, 40.49% of the GC content were retained. The raw data were deposited at National Centre for Biotechnology Information (NCBI) Short Read Archive (SRA) database under the accession number SRR10724565. Overall, 163,022 assembled transcripts were generated with an average size of 908 bp; 98,660 unigenes with an average length of 722.83 bp and an N50 contig length of 1,126 bp. Read mapping of the raw reads showed that 89.1% were aligned by Tophat. The *De novo* based transcriptome-assembled details are given in Table 1.

Functional annotation of unigenes

The total unigenes of *A. sphathulifolius* annotated using blastP analysis against the non-redundant protein database showed 43,221 unigenes (43.71%) with an E-value cut of 1E-05. The unigenes showed top-hit species similarity with *Cynara cardunculus* var *scolymus* (24.41%), *Artemisia annua* (19.20%) and *Helianthus* (1.3%) and others (Fig 1). The results indicated that *Cynara* is more associate to *A. spathulifolius*. The annotations against proteins in the Pfam database showed 39,890 (39.94%) significant hits. Moreover, 46,869(49.97%) unigenes to the Uniprot, 29,880 (30.1%) to the GO classification, and 41,776 (42.34%) to the KEGG-KASS were annotated (Table 2). Indeed, there are unidentified unigenes are not known in this study due to no full discovery of differential function to the Asteraceae species. There are numerous unigenes are not identified through NCBI database though, this might number of Non-coding genes also included, which play great role in controlling the expression of transcriptome products [22].
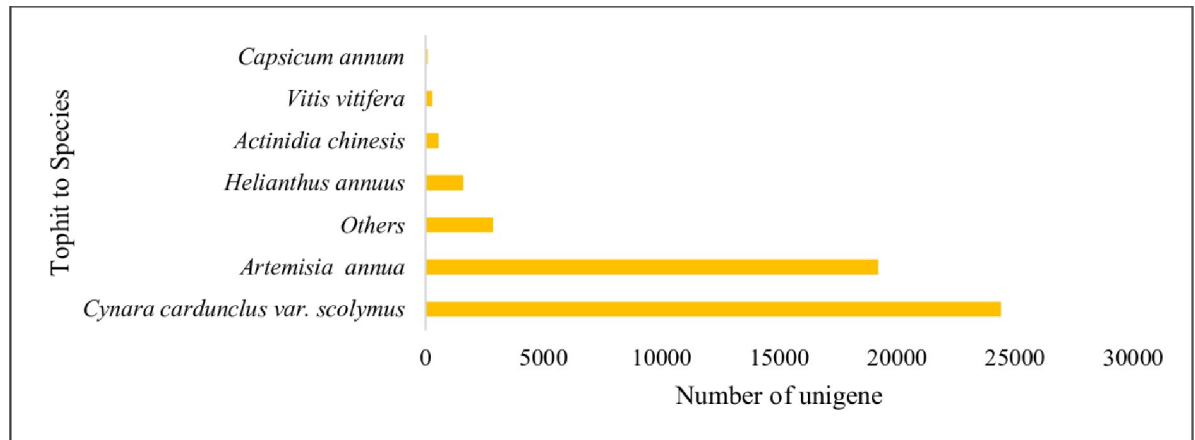
Functional classification of unigenes

The Gene Ontology classification was retrieved based on the annotated unigene to the nr, Pfam, InterPro, and Uniprot databases. In total, 28,621 unigenes were assigned to 64 classes. In terms of the biological functions, there were 9,206 unigene involved, in which 4,447

Table 1. Statistical information of *de novo* assembly of *A. spathulifolius* leaf.

Data Information	Stats
Number of raw reads	148035505
GC content	40.49%
Total transcripts	163,022
Total Unigene	98,860
Median contig length	425
smallest contig	207
largest contig	11,859
contig N50	1,412
Mapping to the raw data	89.91%

https://doi.org/10.1371/journal.pone.0244132.t001

Fig 1. Tophit to the species of nr database by blastP.

https://doi.org/10.1371/journal.pone.0244132.g001

responded to the abiotic stimulus, 9,823 to the protein modification process and 9024 responded to the stress. To molecular functional category, 12,124 unigenes were assigned to 26 classes involving, reproduction (4,288), post-embryonic development (3,755) and flower development (1,229). In the cellular components, there were 8,894 unigenes categorized and grouped into 27classes, where 5,085 unigenes were classified into the cellular component organization. (Fig 2).

KEGG annotation

Total unigenes of *A. spathulifolius* was performed using KAAS and KEGG map orientated database to reveal the molecular interaction networks and metabolic pathways. KEGG-KAAS analysis revealed 41,776 (42.34%) unigenes mapped to the different types of metabolic functions (Fig 3). 666 unigenes involved in the biosynthesis of secondary metabolisms, and 201 in the biosynthesis of antibiotics, respectively, which indicates that *A. spathulifolius* has an abundant of therapeutic benefit compounds, as previously stated elsewhere (Fig 4) [6–9]. In the terpenoid pathways, 377 unigenes are mapped and are classified as organic compounds that include terpenes, diterpene, and sesquiterpene. *A. spathulifolius* highly regulates the terpenoids production, such as mono, di, carotenoid, and sesqui and triterpenoids (Fig 5). In particular, 81 unigenes were involved to the flavonoid synthesis, and 261 unigenes were strongly expressed in phenylpropanoid pathway (PPP) and caffeoylquinic acid production (CQA; is a major chlorogenic acid presents in herbs, fruits, and vegetables). Previously many reports were highly focused on phenylpropanoid synthesis (Fig 6), to help in the reduction of lipid accumulation and adipogenesis [9–12, 23]. Therefore, we used RSEM to estimate the significance

Table 2. Functional annotation of different database and percentage of unigene annotated.

Database	Unigene	Percentage
Nr	47,282	47.80%
InterPro	54,781	55.55%
Pfam	39,890	39.94%
GO	29,921	30.34%
Uniprot	46,869	49.97%
KEGG	41,776	42.3%

https://doi.org/10.1371/journal.pone.0244132.t002
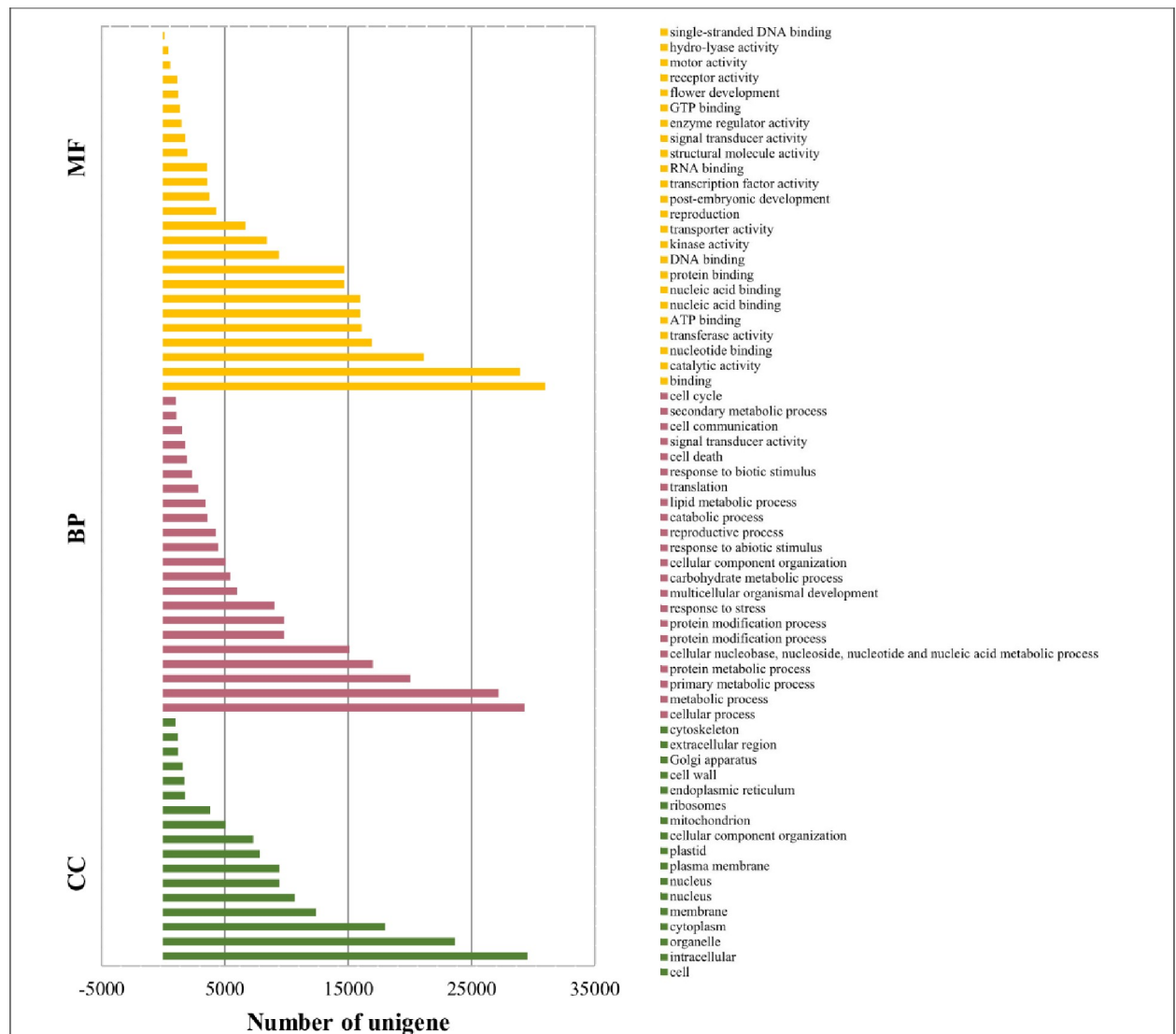
Fig 2. Gene ontology against *A. spathulifolius*: Top 60 enriched GO classification.

amounts of abundant unigenes among some of the PPP involved unigenes. RSEM are estimate to calculate the TPM and FPKM values are determines the expression that result in highly regulated and biosynthesized of PPP by-products in *A. spathulifolius*. In phenylpropanoid pathway, the high TPM values for Phenylalanine Ammonia Lyase (PAL) was 840648.07, which is main key product of PPP pathway that leads to different by-product throughout the PPP metabolites. For caffeic acid 3-O-methyltransferase that leads to synthesis of CQA was 44571.09, which is first product of PPP pathway that leads to different by-product throughout the secondary metabolites. In addition, cinnamoyl-CoA reductase was 39204.2 respectively and so on (S1 Table) this indicates the transcription frequency of a specific unigenes expression in *A. spathulifolius*. The most abundant unigenes for *de novo* assembled *A. spathulifolius* transcriptome expected to have gene abundantly in many pathway like the one flavonoids also observed. Furthermore, 137 unigenes were involved in the alpha-linolenic acid metabolism and 60 unigenes for in linoleic acid—lipid metabolism, which could not be produced by
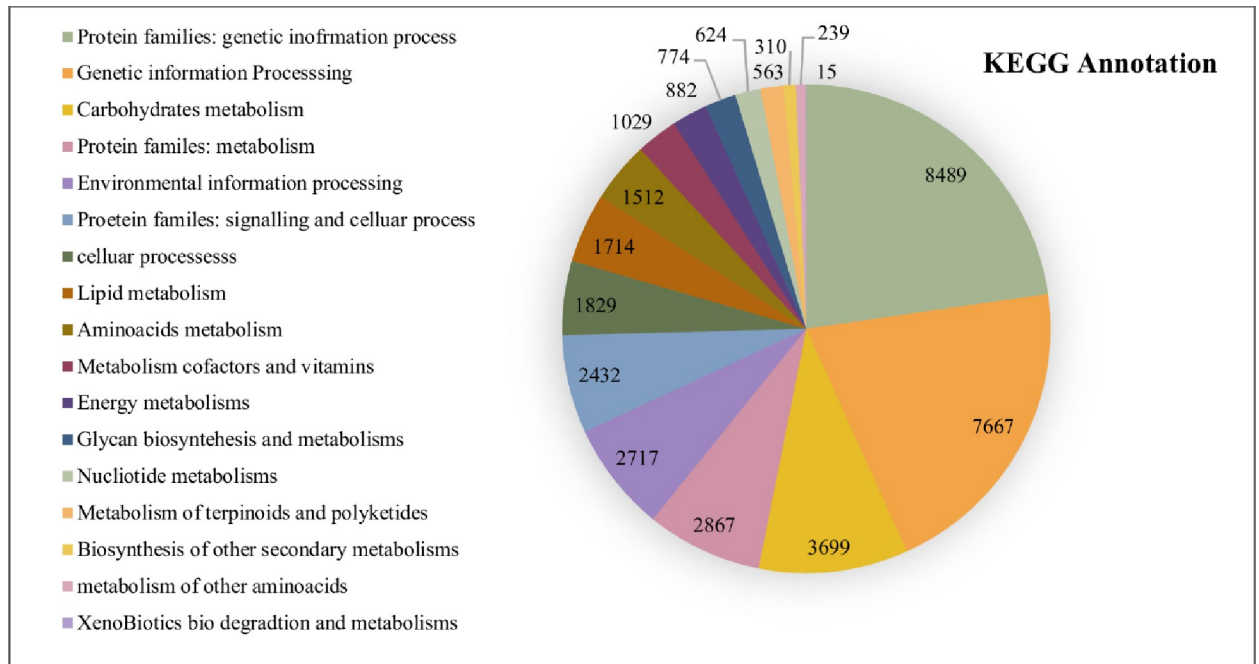
**Fig 3. KEGG-KASS annotation to the *A. spathulifolius*.**

mammals (S1 Fig). In addition, 54 unigenes were mapped in the folate biosynthesis (Vitamin B9), which plays an important role in DNA replication and cell division when the organism is exposed to stress conditions. As a result, it can be noted that *A. spathulifolius* has a stress-tolerant function that helps plants grow in harsh environments. 330 unigenes were involved in the
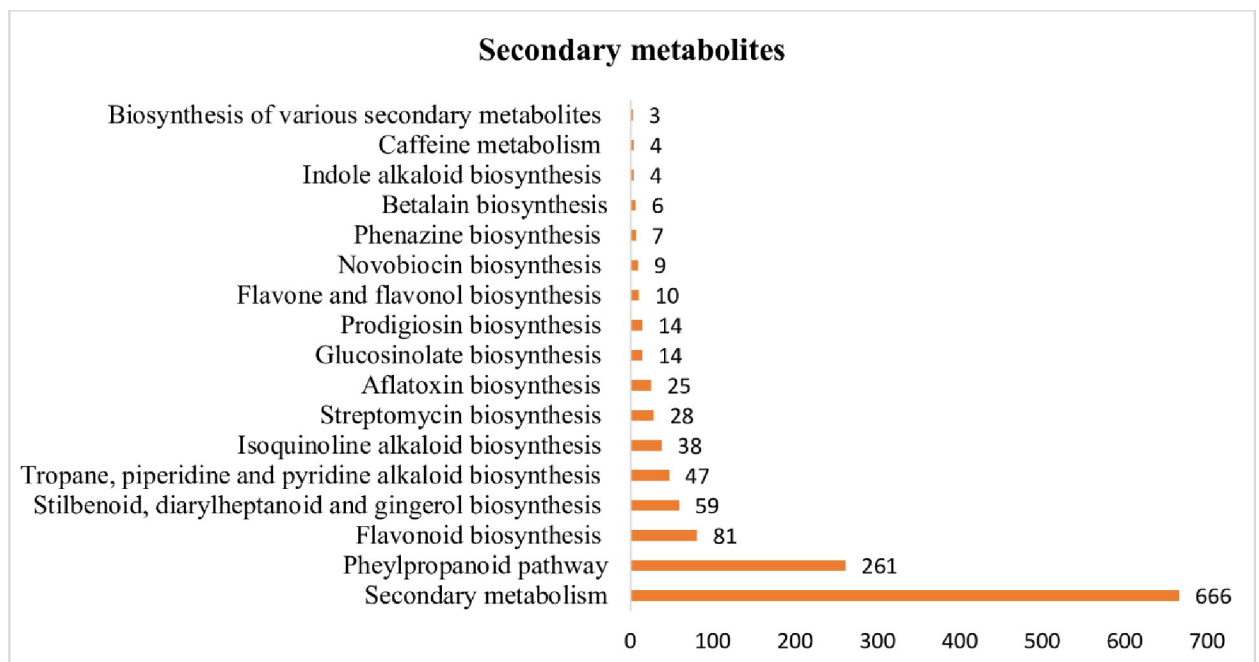
Fig 4. Biosynthesis of secondary metabolites pathway by KEGG-KAAS.
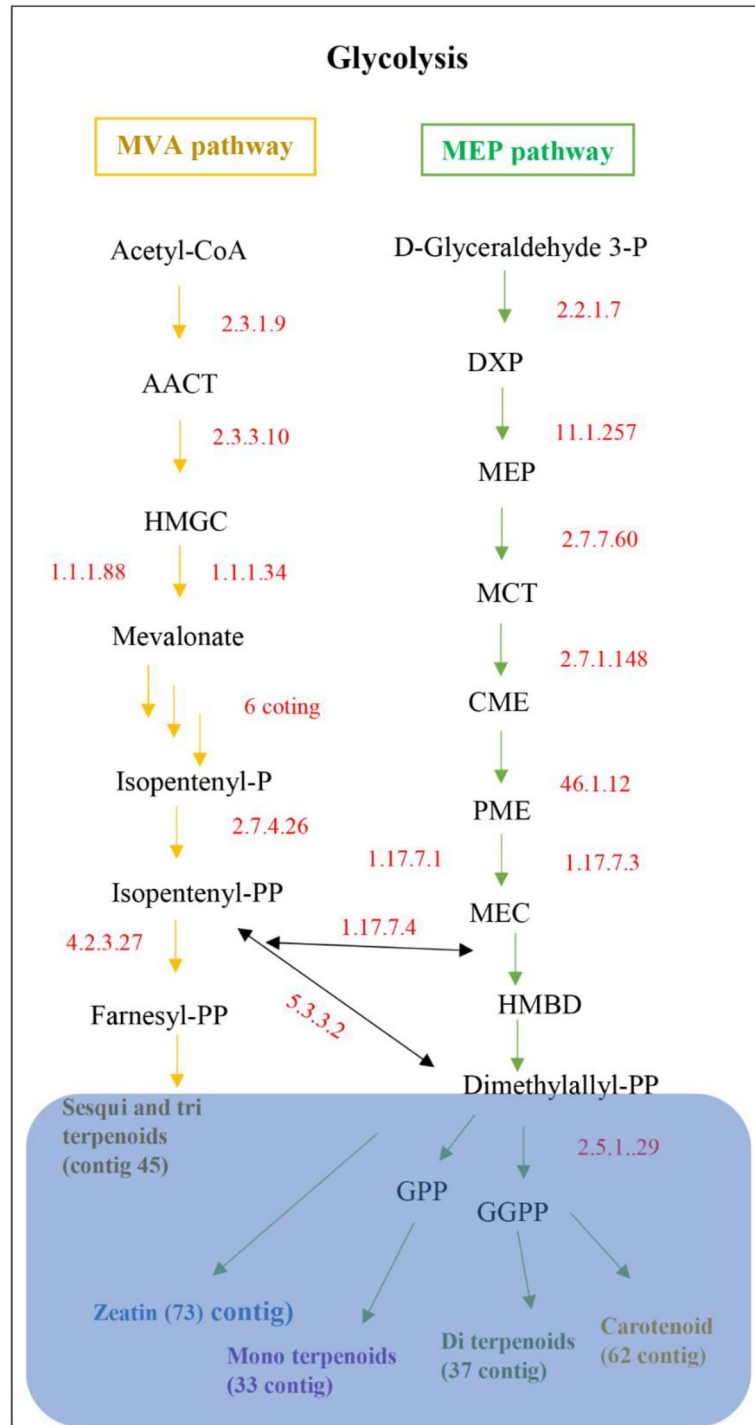
Fig 5. KEGG-KAAS analysis depicting 377 unigenes involved in MVA and MEP pathway and their secondary metabolism products from *A. spathulifolius* maxim. Down arrow marks indicates steps of enzymatic reaction involved.

thermogenesis process (oxidative phosphorylation) in the mitochondria cell and have assisted in a high-fat diet-induced process [10, 23, 24]. This study is useful for the pharmaceutical industry in the development of drugs in future.
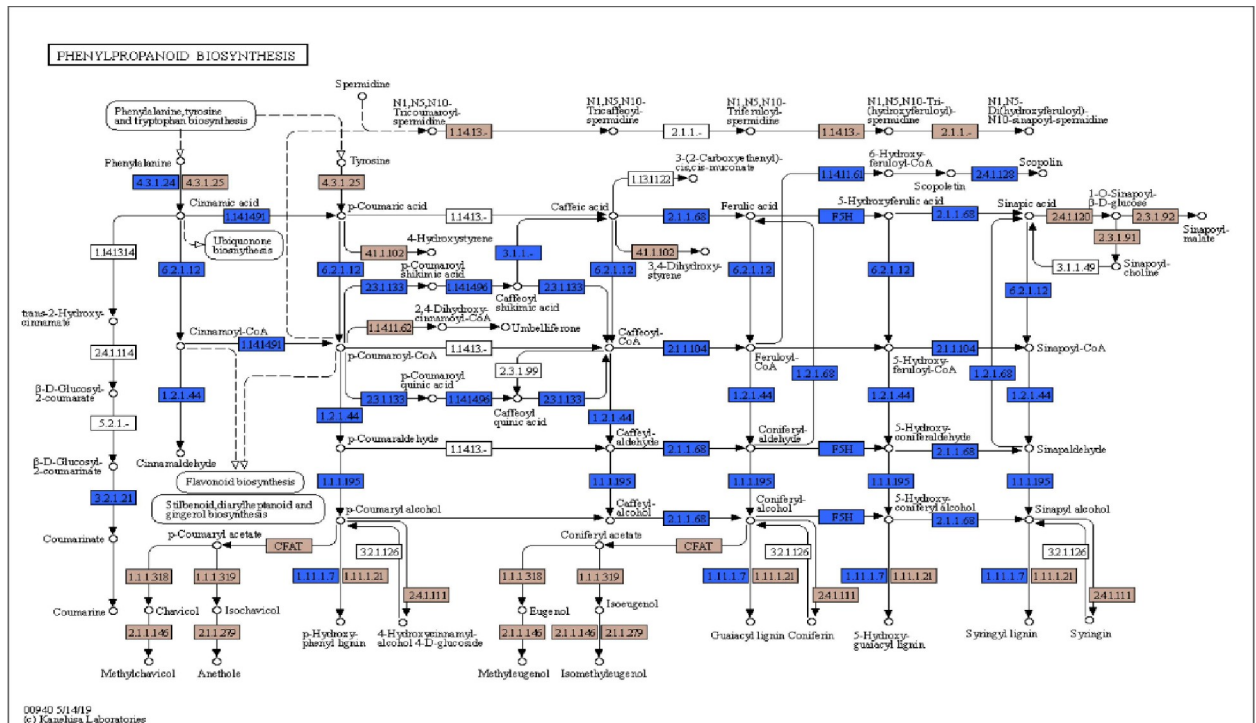
Fig 6. KEGG annotation showing unigenes involved in phenylpropanoid biosynthesis of *A. spathulifolius*.

https://doi.org/10.1371/journal.pone.0244132.g006

Stress resistance gene and transcriptional factor

In nature, the plants are ultimately exposed to different stress factors, such as salt, cold, drought and heat that limit plant growth and productivity. As a result, we found 9,513 unigenes related to stress. 9,024 unigenes were retrieved from *Arabidopsis thaliana* databases, in which 1335 responded to the salt, 932 to the cold, and 547 to the heat and in addition others stress response unigenes were observed (Fig 7). Recently, Baillo et al. [25] stated that Transcriptional factor genes are highly involved in biotic and abiotic stress and are excellent candidates for crop improvement [26]. The present study revealed 53 types of transcriptional factors by using TF-plant tool (Fig 8). 3,167 unigenes were hit to the reference species of *Arabidopsis thaliana*. The highest number was found to be bHLH (basic helix-loop-helix) type, which is involved in drought stress [27]. The B3 ARF (Auxin response factors), ABI3 (Abscisic acid Insensitive3) and RAV (Related to AB13/VP1)) type were the second most involved in dehydration [28] and heat stress [29]. Followed with the ethylene response factor (ERF), which plays a vital role in extreme heat, cold, salt, and drought, and ultimately helps plant to grow under any stress environment [30]. In addition, the bzip (Basic Leucine Zipper) type responds to the binding onto the promoter region and controlling the expression of the gene [31]. WRKY (WRKYGQK motif along with zinc-finger) responds to both biotic and abiotic responses [32] and C3H, which play a role in salt responses [33]. These transcription factors strongly regulates the stress-responsive genes in *A. spathulifolius* and thus normalize the plant growth in its regular life cycle.

Stress response unigenes in KEGG annotation

In total, 489 unigenes were observed from the KEGG annotated pathway of plant hormone signal transduction (Fig 8). As a result, an abundance of stress genes indicates the
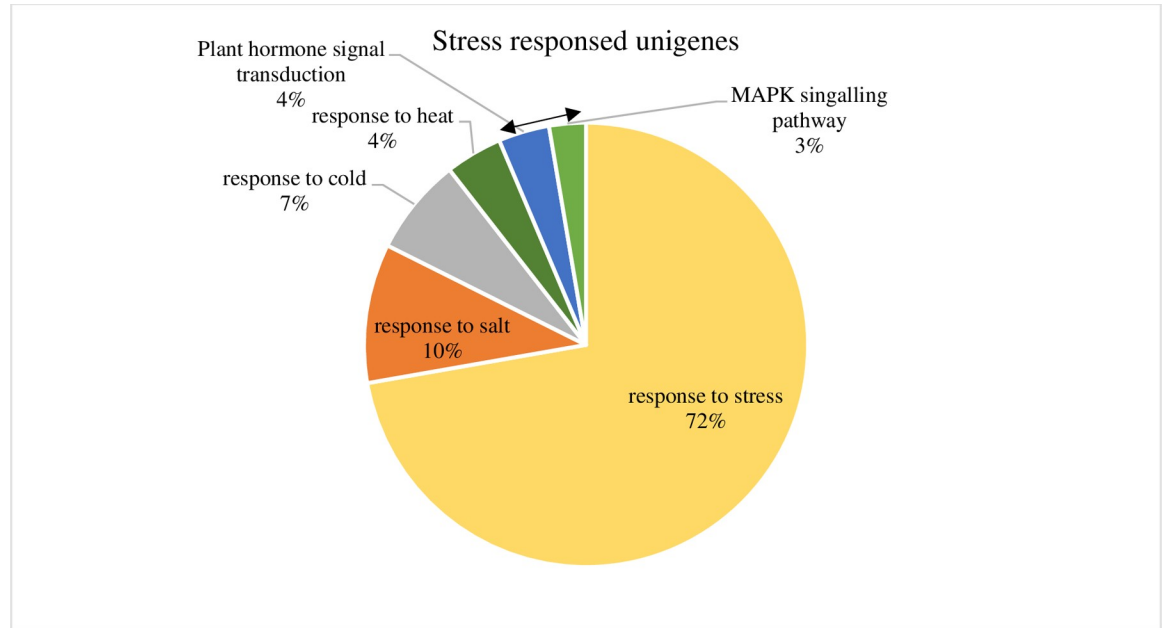
Fig 7. Unigenes expressed to stress related gene in Blastx (1E-05), KEGG and blast tool. Double side arrow mark indicates KEGG analyzed unigenes to the stress response.
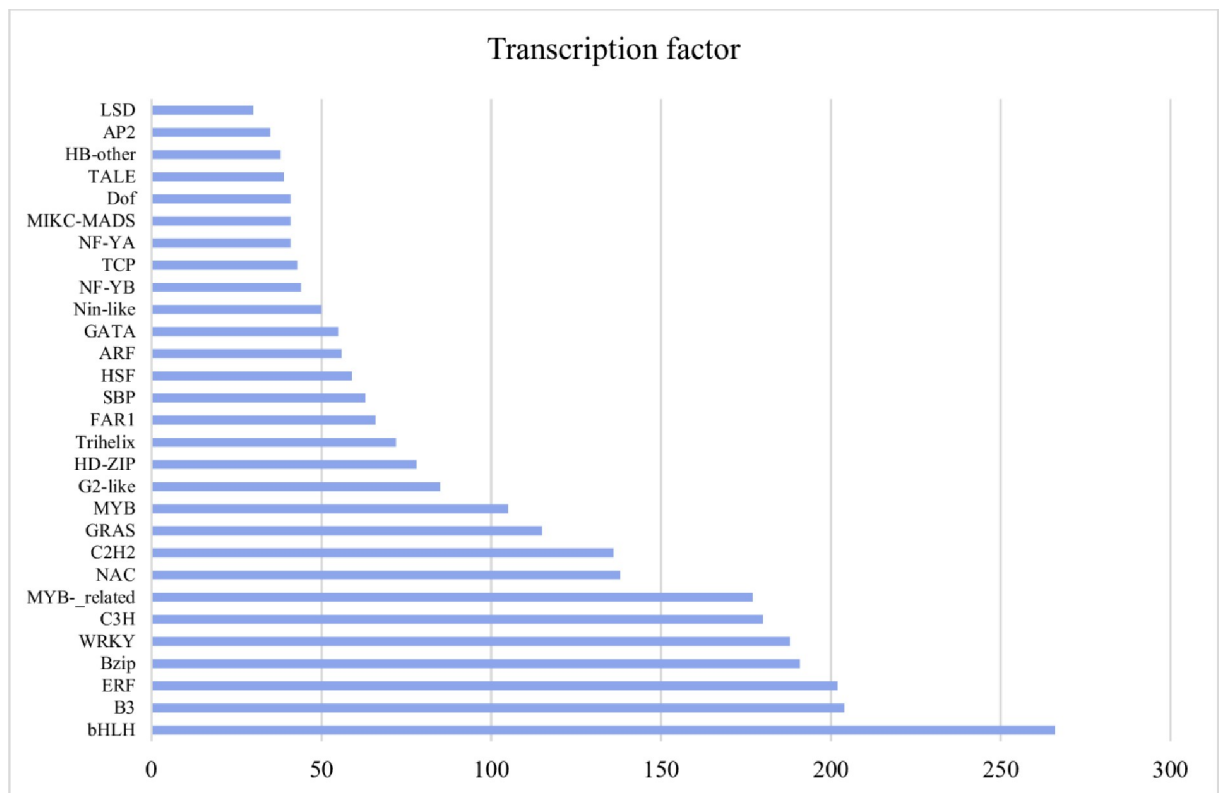
Fig 8. Top 30 enriched transcription factors unigenes in *A. spathulifolius* at 1E-05 parameter.

harsh environments that promote plants to produce stress-related unigenes. 352 unigenes were involved in the MAPK-signaling pathway (S2 Fig), where mitogen-activated protein kinase (MAPK) cascade plays a powerful role in biotic and abiotic stress regulating in plants [34]. 62 unigenes were mapped to the salt and cold tolerance pathway, corresponding to pathogen attacks, heavy metals, salinity, and drought conditions (S3 Fig). In addition, 402 unigenes were involved in plant-pathogen interaction [35].

SSR identification

Microsatellites marker are profusely present in the DNA genome. SSR markers are beneficial for research on genetic diversity, population genetics, genotype, plant breeding improvement, and genetic linkage analysis. Presently, the SSR marker is very cost-effective and well determines the genetic diversity among population species studies based on transcriptome data and studied previously in *Chrysanthemum indicum* based on intraspecific genetic divergence [36, 37]. Out of 163,022 assembled transcripts from *A. sphathulifolius*, 38,545-SSRs markers were identified, which included 3,907 SSRs in the compound formation and 29,698 containing SSRs with an additional 6,896 sequences, showing more than one SSR (Fig 9). Mostly tri-nucleotide repeats with 12,875 (52.05%) repeat motifs were found and followed by di-nucleotide repeats 10,272 (41.52%), tetra-nucleotide 1,027 (4.15%), penta-nucleotide 262 (1.05%) and hexa-nucleotide 299 (1.20%). SSRs with five tandem repeats (6,772) were most common in *A. spathulifolius* followed by six repeats (6,285), and seven repeats (3,628), eight repeats (2,454), nine repeats (1,269) and ten repeats (6,923). Among di-nucleotide repeats, AC/GT (18.68%) followed by AG/CT (10.49%) and AT/AT (14%). AAC/GTT (3.63%) showed the highest frequency of Tri-nucleotide repeats motifs followed by AAG/CTT (9.98%), ATC/ATG (15.04%), AGC/CTG (4.09%) and other motifs, which were uniformly distributed (Fig 10).

Phylogeny analysis of *A. spathulifolius*

The maternal gene-based phylogeny was constructed to examine the relationship between *A. spathulifolius* with eight in-group Asteraceae species such as *Artemisia*, *Chrysanthemum*, *Helianthus*, *Lactuca*, *Opisthopappus*, *Cynara*, *Mikrania* and one out-group namely *Scaevola* from the Goodeniaceae family. Based on the nucleotide alignment of 50 plastid protein-coding
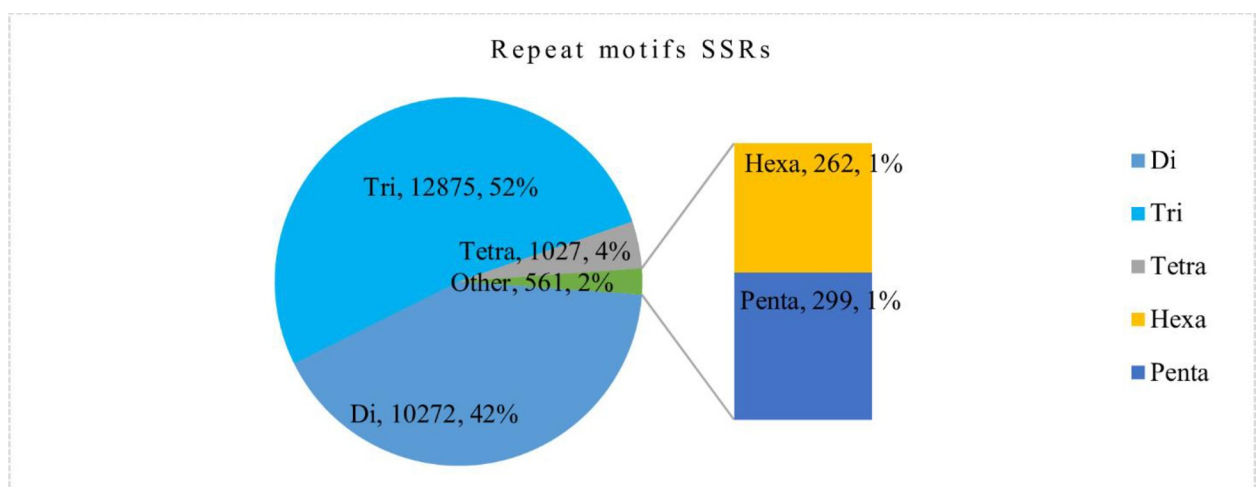
Fig 9. Distribution of repeats motifs of SSR in *A. spathulifolius*.

https://doi.org/10.1371/journal.pone.0244132.g009
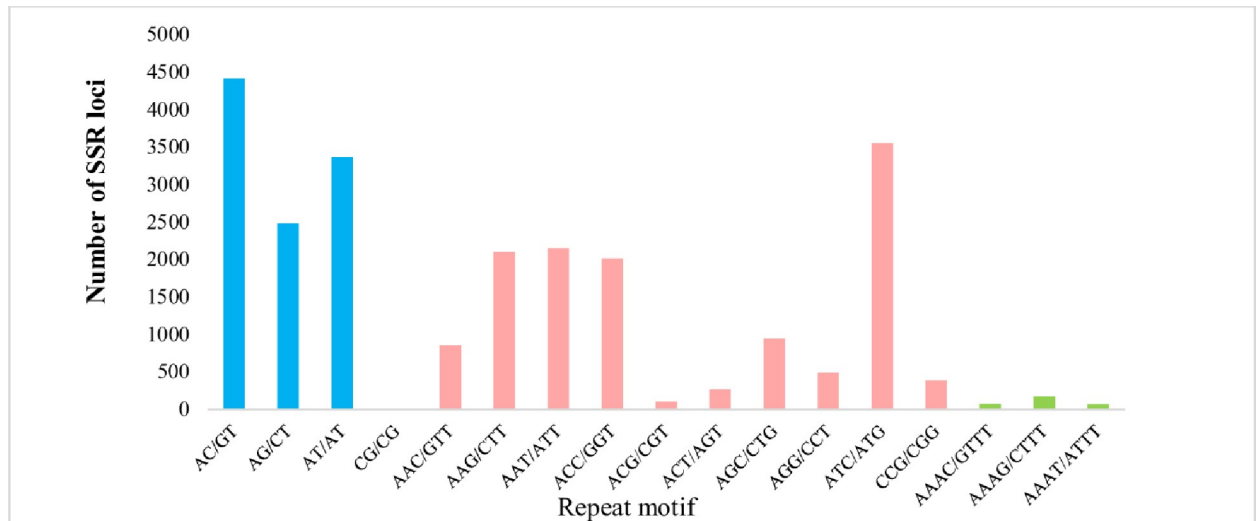
Fig 10. Distribution of repeats motifs of base in *A. spathulifolius*.

genes, the phylogenetic tree had a likelihood score >98% bootstrap value among the Asteraceae species. Plastid protein-coding phylogeny tree showed that *A. spathulifolius* forms one clade to *A. annua*, *O. taihangensis* and *C. x morifolium* lineage. To construct, single-copy orthologs genes based phylogeny complete ORF unigenes were retrieved from nine species (S2 Table). *A. spathulifolius* resulting to close affiliate with *O. taihangensis*, *A. annua*, and *C. x morifolium* in one clade with 89.80% bootstrap value than in the plastid protein-coding gene (Fig 11A). *C. cardunculus.var*. Scolymus and *L. sativa* were chosen as outgroup among other genera. Especially, in single-copy orthologs gene based phylogeny, *A. spathulifolius* found to be in one clade with *O. taihangensis* (Fig 11B), whereas, plastid protein-coding genes to *A. annua*. This result indicates that their phylogeny tree relationship varies due to past and high divergence in the nuclear gene than the plastid protein-coding gene and ultimately similar to the previous reports where *Cynara* and *Lactuca* were ancestors to the *Aster*, *Chrysanthemum*, *Helianthus* and to other genera [38]. Based on gene trees, first duplication event among Asteraceae genera (S4 Fig) shows that without out-group (*S. taccada*) *Cynara* gives rise to *Lactuca* and then the second duplication takes place to form the two new clades, of which, one is *Mikrania* and *Helianthus* (N4) branch. In other hand, *Artemisia* and *Chrysanthemum* (N5) branch followed by *Aster* and *Opisthopappus* (N6) genera. This pattern argues *Aster* and *Opisthopappus*, a recent divergence from other genera. Nevertheless, the more transcriptome species data will be needed to interpret a wide evolutionary relationship among the complete Asteraceae genera.

Conclusion

The main aim of our study was to analyze the species of the genus *Aster* (*A. spathulifolius*) species that has not yet been extensively studied. The transcriptome of *A. spathulifolius* leaf using Illumina high throughput RNA sequencing platform were used. The identification of the unitranscripts involved in molecular function, classification, and phylogenetic analysis was characterized. In this study, KEGG pathway analysis revealed the number of unigenes involved in flavonoid biosynthesis and highly expressed in *A. spathulifolius* leaf transcriptome. Medicinally significant aromatic amino acids (AAAs), especially high in phenylalanine compounds were found rich in *A. spathulifolius* regulating the Chlorogenic acid (CGA) biosynthesis. The
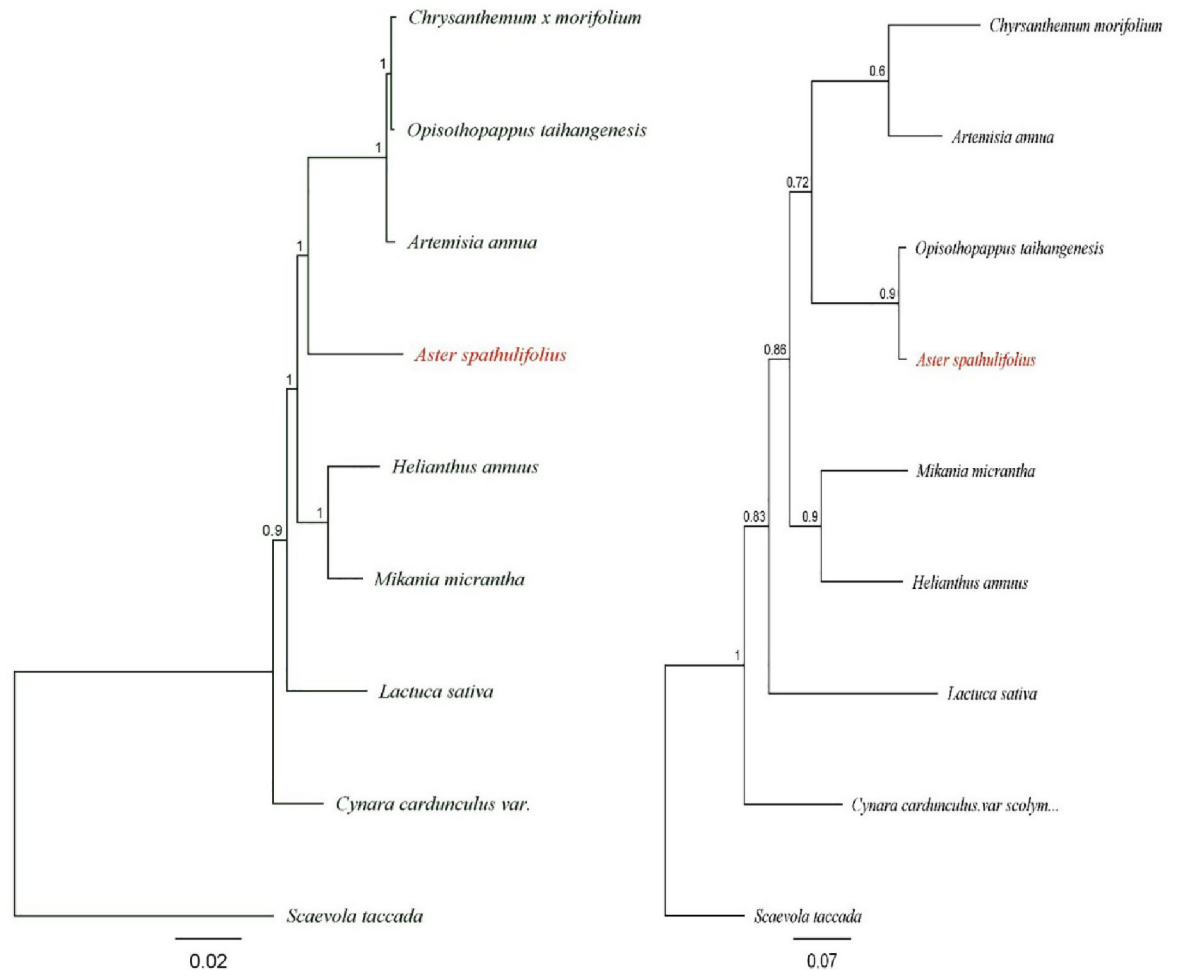
Fig 11. Phylogenetic relationships of major clades of Asteraceae phylogeny of plastid and single-copy orthologous genes. a)
Maximum likelihood divergence, estimation tree shows the *A. spathulifolius* relationship of plastid genome protein-coding region. b)
Divergence estimation of orthologous genes of nine species using Orthofinder and scale bar indicates the substitution rate.

https://doi.org/10.1371/journal.pone.0244132.g011

thermogenesis process may increase by the CGA production that alters glucose and lipid
metabolism. Furthermore, the *A. spathulifolius* plant, which grows in hostile environments, is
prone to produce generous amounts of stress response unigenes to combat abiotic stress. That
concludes the *A. spathulifolius* is more resistant to abiotic stress (heat, salt, drought, and oth-
ers). Likewise, to estimate the divergence of *A. spathulifolius*, phylogeny of Asteraceae species
using a single-copy orthologs gene from transcriptome data was constructed and studied. As a
result, researching medically essential and stress-resistance plants like *A. spathulifolius* is a
dynamic resource for the future.

Supporting information

S1 Fig. KEGG annotation to the alpha- linolenic acid metabolism in *A. spathulifolius*.
(DOCX)

**S2 Fig. MAPK- signaling pathway mapped to the KEGG annotation against *A. spathulifo-
lius*.**
(DOCX)

S3 Fig. Plant hormone signaling pathway in *A. spathulifolius* against KEGG annotation.
(DOCX)

S4 Fig. Orthofinder: Gene duplication prediction among the Asteraceae family, N1-N6 indicates the duplication of gene in order, AAN: *A. annua*, OTA: *O. taihangensis*, ASP: *A. spathulifolius*, HAN: *H. annuus*, MMI: *M. micranta*, LSA: *L. sativa*, CCA: *C. cardunculus*.
(DOCX)

S1 Table. RSEM calculate expression matrix show main gene involved in phenylpropanoid pathway at FPKM rate.
(DOCX)

S2 Table. NCBI plastid and SRA database accession number along with information of assembled unigene in total.
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: Sivagami Jean Claude, SeonJoo Park.

Data curation: Sivagami Jean Claude.

Formal analysis: Sivagami Jean Claude.

Funding acquisition: SeonJoo Park.

Investigation: Sivagami Jean Claude, SeonJoo Park.

Methodology: Sivagami Jean Claude.

Project administration: Sivagami Jean Claude, SeonJoo Park.

Resources: SeonJoo Park.

Software: SeonJoo Park.

Supervision: SeonJoo Park.

Validation: Sivagami Jean Claude, SeonJoo Park.

Writing – original draft: Sivagami Jean Claude.

Writing – review & editing: Sivagami Jean Claude, SeonJoo Park.

References

1. Nesom Guy L. et al. 2000. Generic conspectus of the tribe Astereae (Asteraceae) in North America and Central America, the Antilles, and Hawaii. Sida Botanical Miscellany 20: 1–100.

2. Jang YD, Park BJ. 2008. Geology of Dokdo volcanic: Rocks, minerals, age, and cause of formation. In: Research Institute for Ulleungdo& Dokdo Islands, Kyungpook National University, editors. Nature of Dokdo. Daegu: Kyeongbuk University Press. p. 10e51.

3. Park JH, Lee DH. 2008. Plants of Dokdo. In: Research Institute for Ulleungdo& Dokdo Islands, Kyungpook National University, editor. Nature of Dokdo. Daegu: Kyeongbuk University Press p. 166e221.

4. Jung Su-Young, Byun Jun-Gi, Park Soo-Hyun, Oh Seung-Hwan, Yang Jong-Cheol, Jang Jeong-Won, et al. The study of distribution characteristics of vascular and naturalized plants in Dokdo, South Korea. Korea. Journal of Asia-Pacific Biodiversity 7 (2014) e197ee205

5. Lee, C. B., Illustrated Flora of Korea. Hyangmoonsa, Seoul, pp.740 (1979).

6. Sung Ok Lee Sang Zin Choi, Sang Un Choi Kang Choon Lee, Young Won Chin Jinwoong Kim, et al. Lab-dane Diterpenes from *Aster spathulifolius* and Their Cytotoxic Effects on Human Cancer Cell Lines. Journal of Natural Products 2005 68 (10), 1471–1474. https://doi.org/10.1021/np058044e PMID: 16252909

7. Je-HyukLee, 2013. Antibacterial Activity and Cytotoxicity of Aster sphathulifolius Maxim. Against Oral-bacteria and Oral-cancer Cell Lines. Research Journal of Medicinal Plants, 8: 41–49.

8. Yin X, Huang Y, Jung DW, Chung HC, Choung SY, Shim JH, et al. Epub 2015 May 11. Anti-Diabetic Effect of Aster sphathulifolius in C57BL/KsJ-db/db Mice. J Med Food. 2015 Sep; 18(9):987–98. https://doi.org/10.1089/jmf.2014.3416 PMID: 25961463

9. Yin X.F., Jeon Y.E., Chung H.C. et al. In vitro efficacy evaluation for prevention of diabetes and diabetic complications using *Aster sphathulifolius.* Food SciBiotechnol (2015) 24: 301. https://doi.org/10.1007/s10068-015-0040-0

10. Kim Sa-Jic, Bang Chae-Young, Guo Yuan-Ri, and Se-Young Choung Anti-Obesity Effects of *Aster spathulifolius* Extract in High-Fat Diet-Induced Obese Rats. Journal of Medicinal FoodVol. 19, No. 4.2016 https://doi.org/10.1089/jmf.2015.3566 PMID: 26908215

11. Kim Sa-Jic and Choung Se-Young, Inhibitory effects of *Aster spathulifolius* extract on adipogenesis and lipid accumulation in 3T3-L1 preadipocytes, Royal Pharmaceutical Society, Journal of Pharmacy and Pharmacology, 68 (2015), pp. 107–118. https://doi.org/10.1111/jphp.12485 PMID: 26471469

12. Choa In-Jin, Choung Se Young, Hwang You-Cheol, KyuJeungAhn, Chung Ho Yeon, Jeong In-Kyung, Aster spathulifolius Maxim extract reduces body weight and fat mass in obese humans. Nutrition research 36 (2016) 671–678. https://doi.org/10.1016/j.nutres.2016.03.001 PMID: 27333958

13. Choi KS, Park S. The complete chloroplast genome sequence of Aster spathulifolius (Asteraceae); genomic features and relationship with Asteraceae. Gene. 2015; 572(2):214–221. https://doi.org/10.1016/j.gene.2015.07.020 PMID: 26164759

14. Breitler J.C., Campa C., Georget F., Bertrand B. & Etienne H. A single-step method for RNA isolation from tropical crops in the field Scientific Reports volume 6, Article number: 38368 (2016). https://doi.org/10.1038/srep38368 PMID: 27922073

15. Haas Brian J., Papanicolaou Alexie et al., De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. Nat Protoc. 2013 August; 8(8): https://doi.org/10.1038/nprot.2013.084 PMID: 23845962

16. Kim D., Pertea G., Trapnell C. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14, R36 (2013). https://doi.org/10.1186/gb-2013-14-4-r36 PMID: 23618408

17. Langmead B., Salzberg S. Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359 (2012). https://doi.org/10.1038/nmeth.1923 PMID: 22388286

18. Li B., Dewey C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323 (2011). https://doi.org/10.1186/1471-2105-12-323 PMID: 21816040

19. Beier Sebastian, Thiel Thomas, Thomas Münch Uwe Scholz, Mascher Martin, MISA-web: a web server for microsatellite prediction, Bioinformatics, Volume 33, Issue 16, 15 August 2017, Pages 2583–2585, https://doi.org/10.1093/bioinformatics/btx198 PMID: 28398459

20. Li Weizhong, Godzik Adam, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences, Bioinformatics, Volume 22, Issue 13, 1 July 2006, Pages 1658–1659, https://doi.org/10.1093/bioinformatics/btl158 PMID: 16731699

21. Emms D.M., Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20, 238 (2019) https://doi.org/10.1186/s13059-019-1832-y PMID: 31727128

22. Diamantopoulos MA, Tsiakanikas P, Scorilas A. Non-coding RNAs: the riddle of the transcriptome and their perspectives in cancer. Ann Transl Med. 2018; 6(12):241. https://doi.org/10.21037/atm.2018.06.10 PMID: 30069443

23. Yasuko Ito-Inaba Yamato Hida et al., Characterization of the plant uncoupling protein, SrUCPA, expressed in spadix mitochondria of the thermogenic skunk cabbage. Journal of Experimental Botany, Volume 59, Issue 4, March 2008, Pages 995–1005,https://doi.org/10.1093/jxb/ern024 PMID: 18308738

24. Ledesma Amalia, Lacobaa Mario García de et al., The mitochondrial uncoupling proteins. Genome Biol. 2002; 3(12): reviews3015.1–reviews3015.9. https://doi.org/10.1186/gb-2002-3-12-reviews3015 PMID: 12537581

25. Baillo EH, Kimotho RN, Zhang Z, Xu P. Transcription Factors Associated with Abiotic and Biotic Stress Tolerance and Their Potential for Crops Improvement. Genes (Basel). 2019; 10(10):771. https://doi.org/10.3390/genes10100771 PMID: 31575043

26. Pengfei Wang, Haili Wang, Yongmei Wang, Fengshan Ren, Wei Liu Analysis of bHLH genes from foxtail millet (Setariaitalica) and their potential relevance to drought stress (2018). https://doi.org/10.1371/journal.pone.0207344

27. Pereira, Andy, 2016. Plant Abiotic Stress Challenges from the Changing Environment, Frontiers in Plant Scienc.2016 Volume:7:1123. https://www.frontiersin.org/article/10.3389/fpls.2016.01123

28. Xu Y, Gao S, Yang Y, et al. Transcriptome sequencing and whole genome expression profiling of *chrysanthemum* under dehydration stress. BMC Genomics. 2013; 14:662. Published 2013 Sep 28. https://doi.org/10.1186/1471-2164-14-662 PMID: 24074255

29. Mizoi J 1, Shinozaki K, Yamaguchi-Shinozaki K AP2/ERF family transcription factors in plant abiotic stress responses. BiochimBiophysActa. 2012 Feb; 1819(2):86–96. https://doi.org/10.1016/j.bbagrm.2011.08.004 PMID: 21867785

30. Debbarma J, Sarki YN, Saikia B, Boruah HPD, Singha DL, Chikkaputtaiah C. Ethylene Response Factor (ERF) Family Proteins in Abiotic Stresses and CRISPR-Cas9 Genome Editing of ERFs for Multiple Abiotic Stress Tolerance in Crop Plants: A Review. MolBiotechnol. 2019 Feb; 61(2):153–172. https://doi.org/10.1007/s12033-018-0144-x PMID: 30600447

31. Wang Y, Zhang Y, Zhou R, et al. Identification and characterization of the bZIP transcription factor family and its expression in response to abiotic stresses in sesame. PLoS One. 2018; 13(7):e0200850. https://doi.org/10.1371/journal.pone.0200850 PMID: 30011333

32. Boeckler G.A., Gershenzon J., Unsicker S.B. (2011) Phenolic glycosides of the Salicaceae and their role as anti-herbivore defenses, Phytochemistry, 72, 1497–1509. https://doi.org/10.1016/j.phytochem.2011.01.038 PMID: 21376356

33. Phukan UJ, Jeena GS, Shukla RK. WRKY Transcription Factors: Molecular Regulation and Stress Responses in Plants. Front Plant Sci. 2016; 7:760. Published 2016 Jun 3. https://doi.org/10.3389/fpls.2016.00760 PMID: 27375634

34. Jiang A., Xu Z., Zhao G. et al. Genome-Wide Analysis of the C3H Zinc Finger Transcription Factor Family and Drought Responses of Members in Aegilopstauschii. Plant MolBiol Rep 32, 1241–1256 (2014) https://doi.org/10.1007/s11105-014-0719-z

35. Taj G, Agarwal P, Grant M, Kumar A. MAPK machinery in plants: recognition and response to different stresses through multiple signal transduction pathways. Plant Signal Behav. 2010; 5(11):1370–1378. https://doi.org/10.4161/psb.5.11.13020 PMID: 20980831

36. Wang H, Jiang J, Chen S, Qi X, Peng H, Li P, et al. (2013) Next-Generation Sequencing of the Chrysanthemum nankingense (Asteraceae) Transcriptome Permits Large-Scale Unigene Assembly and SSR Marker Discovery. PLoSONE 8(4): e62293. https://doi.org/10.1371/journal.pone.0062293 PMID: 23626799

37. Han Z, Ma X, Wei M, Zhao T, Zhan R, Chen W. SSR marker development and intraspecific genetic divergence exploration of *Chrysanthemum indicum* based on transcriptome analysis. BMC Genomics. 2018; 19(1):291. Published 2018 Apr 25. https://doi.org/10.1186/s12864-018-4702-1 PMID: 29695227

38. Liu Ping-Li & Wan Jun-Nan & Guo Yanping& Ge Song & Rao Guang-Yuan. (2012). Adaptive evolution of the chrysanthemyl diphosphate synthase gene involved in irregular monoterpene metabolism. BMC evolutionary biology. 12. 214. https://doi.org/10.1186/1471-2148-12-214 PMID: 23137178