



OPEN

# Genome-wide in-silico analysis of ethylene biosynthesis gene family in *Musa acuminata* L. and their response under nutrient stress

Nosheen Tabassum<sup>1</sup>, Muhammad Shafiq<sup>1✉</sup>, Sameen Fatima<sup>1</sup>, Sana Tahir<sup>1</sup>, Bushra Tabassum<sup>2</sup>, Qurban Ali<sup>3✉</sup> & Muhammad Arshad Javed<sup>3</sup>

Ethylene is a gaseous phytohormone involved in plants' growth and developmental processes, including seed germination, root initiation, fruit ripening, flower and leaf senescence, abscission, and stress responses. Ethylene biosynthesis (EB) gene analysis in response to nitrogen (N) and potassium (K) stress has not yet been conducted in *Musa acuminata* (banana) roots. The genome mining of banana (*Musa acuminata* L.) revealed 14 putative 1-aminocyclopropane-1-carboxylate synthase (ACS), 10 1-aminocyclopropane-1-carboxylate oxidase (ACO), and 3 Ethylene overproducer 1 (ETO1) genes. ACS, ACO, and ETO1 proteins possessed amino acid residues ranging from 422–684, 636–2670, and 893–969, respectively, with molecular weight (Mw) ranging from 4.93–7.55 kD, 10.1–8.3 kD and 10.1–10.78 kD. The number of introns present in ACS, ACO, and ETO1 gene sequences ranges from 0–14, 1–6, and 0–6, respectively. The *cis*-regulatory element analysis revealed the presence of light-responsive, abscisic acid, seed regulation, auxin-responsive, gibberellin element, endosperm-specific, anoxic inducibility, low-temperature responsiveness, salicylic acid responsiveness, meristem-specific and stress-responsive elements. Comprehensive phylogenetic analyses ACS, ACO, and ETO1 genes of Banana with *Arabidopsis thaliana* revealed several orthologs and paralogs assisting in understanding the putative functions of these genes. The expression profile of *Musa acuminata* genes in root under normal and low levels of nitrogen and potassium shows that *MaACS14* and *MaACO6* expressed highly at normal nitrogen supply. *MaACS1* expression was significantly upregulated at low potassium levels, whereas, *MaACO6* gene expression was significantly downregulated. The functional divergence and site-specific selective pressures on specific gene sequences of banana have been investigated. The bioinformatics-based genome-wide assessment of the family of banana attempted in the present study could be a significant step for deciphering novel ACS, ACO, and ETO1 genes based on genome-wide expression profiling.

## Abbreviations

ACS	1-Aminocyclopropane-1-carboxylate synthase
ACO	1-Aminocyclopropane-1-carboxylate oxidase
ETO1	Ethylene overproducer 1
<i>M. acuminata</i>	<i>Musa acuminata</i>
pI	Isoelectric point
Mw	Molecular weight
AA	Amino acid
kD	Kilodaltons
NJ	Neighbor Joining

<sup>1</sup>Department of Horticulture, Faculty of Agricultural Sciences, University of the Punjab New Campus, Lahore, Pakistan. <sup>2</sup>School of Biological Sciences, University of the Punjab New Campus, Lahore, Pakistan. <sup>3</sup>Department of Plant Breeding and Genetics, Faculty of Agricultural Sciences, University of the Punjab New Campus, Lahore, Pakistan. ✉email: shafiq.iags@pu.edu.pk; saim1692@gmail.com

Ka/Ks	The ratio of nonsynonymous substitutions per nonsynonymous site (Ka) to the number of synonymous substitutions per synonymous site (Ks).
N	Nitrogen
K	Potassium

As a gaseous phytohormone, ethylene is produced in most plant tissues. It is important in regulating plant growth and developmental processes, including seed germination, root initiation, root gravitropism, fruit ripening, flower and leaf senescence, abscission, and stress responses. In plant tissues, ethylene production is typically low but increases at developmental stages such as ripening and senescence and in response to mechanical and environmental stresses<sup>1</sup>. Nutrient stress is a complex phenomenon that leads to low or high availability of nutrients to the plant<sup>2</sup>. Different genes are also involved in plant responses due to nutrient stresses<sup>3</sup>. Hence, to understand the nutrient stress response mechanism of *Musa acuminata* in roots, it is important to identify and analyze the genes involved in ethylene biosynthesis. Furthermore, the ethylene biosynthesis (EB) genes in response to nitrogen (N) and potassium (K) stress have not yet been analyzed in *Musa acuminata* (banana) roots.

Ethylene also acts as an important modulator and inducer of plant aging<sup>4</sup>. Ethylene is biosynthesized from Methionine pathways in higher plants in which ACS and ACC oxidase (ACO) catalyze the reactions from S-adenosylmethionine (SAM) to 1-aminocyclopropane-1-carboxylate (ACC) and ACC to ethylene, respectively<sup>5,6</sup>. This ethylene biosynthesis pathway has been well explained and mapped in the *Arabidopsis thaliana*<sup>7</sup>. Using advanced molecular biology techniques, various plant species have been used to isolate cDNA and genomic clones for both enzymes, encoded by multigene families<sup>8–13</sup>. These cDNA clones have been used to characterize the expression of individual members in various tissues and in response to recognized ethylene biosynthesis-inducing stimuli<sup>7</sup>. In *Arabidopsis*, ACS2, ACS4 to ACS9, and ACS11 form functional homodimers, whereas unfunctional homodimer forms in ACS1. ACS3 is a pseudogene. Whereas ACS10 and ACS12 encode aminotransferase. All genes play a specific role in the plant during growth and development. It was studied that *AtACS5* gene had the highest promoter activity in grown seedlings (2 weeks old). And this gene was localized at the reproductive stage from areoles where *AtACS4* and *AtACS7* were present in both veins and areoles<sup>14</sup>. Ethylene gas is a plant hormone and critical growth regulator. It synthesizes in response to biotic and abiotic stresses<sup>15–19</sup>. It also influences plant growth and developmental processes like germination, leaf and flower senescence and abscission, fruit ripening, and nodulation<sup>19,20</sup>. In most plant tissues, the level of ACS activity is almost directly proportional to the level of ethylene production<sup>21</sup>. The synthesis of ethylene involves simple yet highly regulated steps. ACS's conversion of S-adenosyl methionine to ACC is a rate-limiting step in ethylene biosynthesis. ACC is converted to ethylene by ACC oxidase (ACO). 1-Aminocyclopropane-1-carboxylate oxidase (ACO), is an O<sub>2</sub>-activating ascorbate-dependent nonheme iron enzyme, which involves in the catalyses of ACC in the ethylene biosynthetic pathway and converts 1-aminocyclopropane-1-carboxylic acid (ACC) to ethylene<sup>22</sup>. An ethylene overproducer mutation, *ETO1*, negatively regulates ACS and ethylene production. *ETO1* regulates the stability of 1-aminocyclopropane-1-carboxylate synthase (ACS) enzymes. It can also act as substrate-specific adapter connecting ACS enzymes like ACS5 to ubiquitin ligase complexes, leading to proteasomal degradation of ACS enzyme<sup>23,24</sup>. A constitutive triple-response phenotype in *Arabidopsis* was used to identify three mutants that were affected in the regulation of ethylene biosynthesis due to ethylene overproduction<sup>25,26</sup>. The post-transcriptional regulation of 1-aminocyclopropane-1-carboxylic acid synthase (ACS) has been affected by the *Arabidopsis* ethylene-overproducing mutants *ETO1*, *ETO2*, and *ETO3*. *ETO2* and *ETO3* are dominant mutations, but *ETO1* is inherited as a recessive mutation. The 12-amino-acid C-terminal region of ACS5 is predicted to be disrupted by the 1-bp insertion that causes the *ETO2* mutation. Since the steady-state level of ACS5 mRNA is not increased by the *ETO2* mutation, the mutation likely works post-transcriptionally<sup>27</sup>.

A comparative study of 1-aminocyclopropane-1-carboxylic acid synthase (ACS), 1-aminocyclopropane-1-carboxylate oxidase (ACO), and Ethylene overproducer 1 (*ETO1*) genes from ethylene biosynthesis pathway in *M. acuminata* was performed with *A. thaliana* would provide a starting point for understanding how the *M. acuminata* gene family response under different stresses. Genome-wide analysis identified ACS, ACO, and ETO genes in *M. acuminata*. *M. acuminata* gene's structures, functions, and conserved motifs were compared with *A. thaliana*'s ACS, ACO, and ETO gene to investigate their response under nitrogen (N) and potassium (k) stress. Furthermore, the expression of ACS, ACO, and ETO genes in *M. acuminata* in response to nitrogen (N) and potassium (k) stress was investigated. The present findings suggest the roles of ACS, ACO, and ETO genes in mediating abiotic stress and provide valuable information for further study on the function of ACS, ACO, and ETO genes in the growth, development, and stress responses of *M. acuminata*.

## Materials and methods

### Database search and retrieval of sequence

377 Amino acid (AA) sequence of PLP\_aminotran (CL0061)<sup>28</sup> present in *AtACS* (Accession no. NP\_191710.1), *ACO* (Accession no. NP\_179549.1) and full-length AA sequence of *ETO1* (Accession no. NP\_190745.6) retrieved from the *Arabidopsis thaliana* proteome database at NCBI Gene bank and was used for the identification of ACS, ACO and *ETO1* proteins-encoding genes in the banana proteome database at Phytozome <https://phytozome.jgi.doe.gov/pz/portal.html> using BLAST-P program<sup>29</sup>. The incorrect and redundant predicted sequences were manually removed, and then all putative *MaACS*, *MaACO*, and *MaETO* genes were further verified using Pfam database. The retrieved amino acid sequences were subjected to NCBI CDD (Conserved Domain Database) (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>)<sup>30</sup> with the default parameters.

## Determination of physical characteristics of banana proteins

Amino acid length, molecular weight, and theoretical pI of *MaACS*, *MaACO*, and *MaETO* proteins were predicted using ProtParam tool (<http://web.expasy.org/protparam/>)<sup>31</sup>. The information for gene IDs, chromosomal position, and gene and protein sequence was retrieved from Phytozome. All genes were renamed according to the order of their physical position. Subcellular localization of *MaACS*, *MaACO*, and *MaETO* was predicted using the online tool WoLF PSORT (<https://wolfsort.hgc.jp/>)<sup>32</sup>.

## Multiple sequence alignment and phylogenetic analysis

The amino acid sequences of ACS, proteins from banana, and arabidopsis were aligned using Clustal W version 2.1<sup>33,34</sup>. The phylogenetic tree was constructed with MEGAx.0 using a neighbor-joining (NJ) method<sup>35</sup>. The bootstrap values were calculated for 1000 iterations. Similar protein alignment and phylogenetics analysis methods were used for *ETO* and *ACO* proteins. 14 *MaACS* genes and 12 *Arabidopsis thaliana* protein sequences were used for phylogenetic analysis. Similarly, for *ACO*, 10 *MaACO* genes and 5 *Arabidopsis thaliana* protein sequences, and for *ETO*, 3 *MaETO* genes and 5 *Arabidopsis thaliana* protein sequences were used.

## Gene structure analysis and conserved motifs recognition

The genomic and coding sequences of identified genes were retrieved from the database to investigate the intron/exon arrangement of banana ACS, ACO, and *ETO* genes. These sequences were further used to draw the gene structure using Gene Structure Display Server (GSDS v2.0<sup>36</sup>) (available at <http://gsds.cbi.pku.edu.cn/>). Multiple EM for Motif Elicitation (MEME) programs (<http://meme.nbcr.net/meme/>)<sup>37</sup> was used to analyze the concluded protein sequences of the banana ACS, ACO, and *ETO*-1 with a maximum number of motif set as 20.

## Gene duplication and calculation of Ka and Ks substitution rates

Putative Gene pair was generated using the knowledge obtained from phylogenetic, motif, and domain analysis of *M. acuminata* ACS, ACO, and *ETO1* genes. These gene pairs were used to calculate the Ka and Ks values through tbtools<sup>38</sup>. The CDS, protein sequence, and gene pair of *MaACS*, *MaACO*, and *MaETO* were utilized. Using Ka/Ks ratio, molecular evolutionary rates of each gene pair were calculated. The Ka/Ks ratio less than 1 indicates the viability of purifying selection; however, Ka/Ks ratio greater than 1 indicates positive selection, and Ka/Ks = 1 indicates neutral selection<sup>39,40</sup>. Using the formula " $t = Ks/2\lambda$ ", with  $\lambda$  ( $6.05 \times 10^{-9}$ ), gene pair divergence was estimated to represent neutral substitution. The *MaACS*, *MaACO*, and *MaETO* like genes were mapped on scaffold using TBtool, and duplicated genes were connected on scaffold using red lines.

## Promoter analysis

To analyze the organ-specific expression profile of *MaACS*, *MaACO* and *MaETO* at various development stages, the *cis*-regulatory element evaluation of recognized banana ACS, ACO, and *ETO1* genes was performed by recovering 1000 base pairs sequence upstream from the starting site of banana genomic sequences (promoter region) using the Phytozome database. An ample number of *cis*-regulatory elements were analyzed in all the *Musa acuminata* 14 ACS, 10 ACO, and 3 *ETO* genes by employing the PlantCare database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>)<sup>41</sup>.

## Expression analysis

For expression profiling, Reads Per Kilo bases per Million mapped reads (RPKM) values from RNA-seq data were  $\log^2$  transformed, and the p-value was calculated using *t* test. Expression patterns with hierarchical clustering are displayed in Heatmap Illustrator in TBtools<sup>38</sup>.

## Putative microRNA target sites analysis

The micro-RNA (miRNA) datasets of *Musa acuminata* were retrieved from a knowledge-based database called Plant miRNA Encyclopedia (PmiREN, <http://www.pmiREN.com/>)<sup>42</sup>. There, to find out the miRNAs that target the *MaACS*, *MaACO*, and *MaETO* genes, CDS sequences of all *MaACS*, *MaACO*, and *MaETO* genes were searched for the complementary sequences of miRNAs with the help of psRNA Target (<https://plantgrn.noble.org/psRNA/Target/analysis?function=3>)<sup>43</sup> with default parameters.

## Ethical approval

It has been confirmed that the experimental samples of plants, including the collection of plant material, complied with relevant institutional, national, and international guidelines and legislation with appropriate permissions from Authorities of Department of Horticulture, Faculty of Agricultural Sciences, University of the Punjab New campus, Lahore, Pakistan.

## Results

### Identification of the ACS, ACO and *ETO1* genes in banana

ACS, ACO, and *ETO* genes, involved in the ethylene biosynthesis pathway, were detected in Banana Genome database. Accordingly, the initial analysis predicted 26, 18, and 8 protein sequences of ACS, ACO, and *ETO*, respectively. The proteins encoded by the same gene isoforms and proteins containing truncated ACS, ACO, and *ETO* DNA-binding domains were manually removed. Finally, 14 *MaACS*, 10 *MaACO*, and 3 *MaETO* non-redundant genes were identified and used for further analysis. The number of ACS genes was more than that of ACO and *ETO*.

The studied genes encoded proteins ranging from 443 to 2670 amino acids in length; *MaACS3* was the smallest protein, whereas *MaACO6* was the most significant protein. Molecular weight varied between 10.16 and 83.3. Isoelectric points ranged from 8.47 (*MaACS6*) to 4.79 (*MaACO7*) (Tables 1, 2, 3).

### Comparative phylogenetic relatedness of banana ACS, ACO and ETO1 gene family with Arabidopsis

To investigate the evolutionary relationships between *MaACS*, *MaACO* and *MaETO1* TFs and *Arabidopsis thaliana*, a neighbor-joining (NJ) phylogenetic tree was constructed by aligning their full-length protein sequences. The results showed that 14 *MaACS* proteins were distributed among 3 groups named I, II A, II B, and III (Table S1 and Fig. 1A).

Group I consisted of 3 *Arabidopsis* proteins while 3 belong to banana *MaACS1*, *MaACS3*. II-A group consisted of 5 *Arabidopsis* *AtACS4*, *AtACS5*, *AtACS8*, *AtACS9*, *AtACS11* and 6 genes of *MaACS4*, *MaACS5*, *MaACS6*, *MaACS7*, *MaACS8*, *MaACS9*. II-B Group contained 2 *Arabidopsis thaliana* proteins *AtACS12*, *AtACS10*, while *Musa acuminata* consisted of *MaACS12*, *MaACS13*, *MaACS14*. Group III contained only 1 *Arabidopsis* protein *AtACS7*, whereas banana included *MaACS10* and *MaACS11*.

The results for *ETO1* depicted that 3 *MaETO1* proteins were distributed among 3 groups named I, II, and III (Table S2 and Fig. 1B).

Group I consisted of 2 *Arabidopsis thaliana* *AtETO1\_951*, *AtETO1\_959* proteins while 2 belong to banana *MaETO1*, *MaETO2*. II groups consist of only 1 *Arabidopsis* *AtEOL2\_925*, and 0 banana genes. Group III contained only 1 *Arabidopsis* protein *AtEOL1\_888*, whereas banana included *MaETO3*. According to ACO results, the division of *MaACO* into groups represents that group I contains 3 *Arabidopsis thaliana* proteins *AtACO1*, *AtACO2*, *AtACO3*, and 3 *Musa acuminata* genes, i.e. *MaACO1*, *MaACO2*, *MaACO10*. Group II consisted of 1 *Arabidopsis thaliana* protein *AtACO5* and 2 *Musa acuminata* genes *MaACO5* and *MaACO6*. Similarly, group III contain 1 *Arabidopsis thaliana* protein *AtACO4* and 5 *Musa acuminata* genes *MaACO3*, *MaACO4*, *MaACO7*, *MaACO8*, *MaACO9* (Table S3 and Fig. 1C).

Proteins of common clade usually seem to show similarity in structure and functioning. So, all the ACS, ACO, and *ETO1* like proteins of similar Clades may have similar structures and functions.

### Gene structures and recognition of conserved motifs and domain

The organization of exon and intron helps verify the evolutionary relationship between genes or organisms<sup>48</sup>. Their distribution patterns and numbers are an evolutionary mark for a gene family. A comprehensive demonstration of Banana genes' exon–intron structures and phylogenetic revealed that the gene structure pattern was consistent with the phylogenetic analysis. The number of introns varied from 3 to 8 in *MaACS* gene, and no gene in *MaACS* is intron-less (Fig. 2A, Table S4).

In Group I, *MaACS1* possesses 3 introns and 4 exons. Group IIA comprises of *MaACS2* gene which has 3 introns and 4 exons. In group IIB, *MaACS3* contain 5 introns and 6 exons, while *MaACS4*, *MaACS5*, *MaACS6*, *MaACS7* genes contain 4 exons and 3 introns. In group III, *MaACS11* gene contains 5 exons and 4 intron and *MaACS13* gene has 9 exons and 8 introns, while, *MaACS8*, *MaACS9*, *MaACS10*, *MaACS12* and *MaACS14* consists of 4 exons and 3 introns.

According to Fig. 2B and Table S5, group I contain *MaETO1* and *MaETO2* genes consisting of 4 exons and 3 introns, while *MaETO3* in group III has 4 exons and 3 introns. Similarly, *MaACO1* and *MaACO2* of group I contain 4 exons and 3 introns, while *MaACO10* of group I have 2 exons and 1 intron. *MaACO5* of group II comprises 2 introns and 3 exons, while *MaACO6* contains 4 exons and 3 introns. *MaACO3*, *MaACO8*, and *MaACO9* of group III contain 4 exons and 3 introns, while *MaACO4* has 3 exons and 2 introns, and *MaACO7* consists of 7 exons and 6 introns (Fig. 2C, Table S6).

The conserved motif analysis also verified the classification of *MaACS* genes. All *MaACS* protein sequences were loaded into the MEME analysis tool to identify the conserved motifs. As a result, twenty conserved motifs were observed, which were statistically significant with *E*-values less than  $1 \times 10^{-40}$  (Fig. 3). The motifs of *MaACS* proteins identified by MEME were between 15–50 amino acids in length. Motif 1, Motif 2, Motif 5, Motif 6, Motif 7 and Motif 9 are common in all groups. Group I and Group IIA had similar motif patterns. Motif 3 is also common in all groups except in *AtACS-12\_Type 2* gene of group IIB. All groups (I, IIA, IIB and III) contain M-4 and M-8 other than Group III (*MaACS11*) and G-I (*MaACS3*), respectively. Meanwhile, IIB group members have relatively complex motif patterns compared with Group IIA (Fig. 3A).

In *MaETO* proteins, all the motifs are common in all groups. Motif pattern is also the same in all groups. Figure 3B shows that *MaETO* gene structures are similar to the corresponding *AtETO* gene structures. TPR-1 domain is only present in *AtETO* genes. BTB\_POZ domain is only present in *MaETO3* genes, similar to *atEOL-1* genes.

Motif analysis of *MaACO* genes shows that motif 3 is only present in *MaACO7*, *MaACO8* *MaACO9* of G-3. Motif 7 is only present in G-3 (*MaACO3*, *MaACO4*, *MaACO7*, *MaACO8*, *MaACO9*). *MaACO10* contains only 5 motifs i-e. 1, 2, 3, 4 and 5. *MaACO7* has repeats of motifs and domains. Both 2OG-Fell\_Oxy and DIOX\_N domains are present in all *MaACO* genes (Fig. 3C).

### Gene duplication of banana ACS, ACO and ETO1 genes

The date of duplication of the gene was also estimated through MEGA-X using pairwise alignment that provided Ks and Ka values and then Ka/Ks was calculated manually (Fig. 4). Ks depicts the number of synonymous substitutions per synonymous site, whereas Ka shows the number of nonsynonymous substitutions per nonsynonymous site and the ratio of nonsynonymous (Ka) versus synonymous (Ks) mutation was represented by Ka/Ks. The speculative date for gene duplication of the paralogous group *MaACS7\_MaACS4* was calculated to be 99.11Mya,

ACS gene	Source accession no		Chromosome no	Chromosome location	Direction (bp)	mRNA	Size (AA)			pI	Mw (D)	Gene expression of banana in roots (Fig. 6)	Orthologue from model <i>Arabidopsis thaliana</i> and or other plants			
	Phytozome	Banana Genome Hub					Peptide	Genome	Gene ID				Accession no	Annotation/ function in plants	References	
<i>MaACS1</i>	GSMUA_Achr4G29150_001	Ma04_g35640	4	27240988..27243490	Reverse	1461	486	2504	7.97	54,964.73	It showed high expression in roots during low potassium stress	AtACS2	NP_171655	These enzymes catalyze the conversion of S-adenosyl-L-methionine (SAM) into 1-aminocyclopropane-1-carboxylate (ACC), a direct ethylene precursor	44	
<i>MaACS2</i>	GSMUA_Achr4G24930_001	Ma04_g31490	4	24594460..24596343	Reverse	1275	481	1884	6.62	53,971.55	It did not show any expression in roots or leaves under N, K stress	AtACS6	NP_192867	It is a direct precursor of ethylene. Involved in bacterial flagellin-induced ethylene production	45	
<i>MaACS3</i>	GSMUA_AchrUn_randomG12290_001	Ma09_g19150	Un-characterized	58535286..58538567	Forward	1443	443	3282	7.61	49,795.41	It requires a high potassium level to express itself in roots	AtACS6	NP_192867	Involved in bacterial flagellin-induced ethylene production	45	
<i>MaACS4</i>	GSMUA_Achr2G08600_001	Ma02_g10500	2	12544533..12546250	Reverse	1341	463	1718	7.81	51,966.40	It expressed moderately in roots during low nitrogen level	AtACS9	NP_190539	ACS proteins' stability and regulation of such stability play a central role in ethylene biosynthesis	44	
<i>MaACS5</i>	GSMUA_Achr3G25830_001	Ma03_g27050	3	25860557..25862163	Forward	1299	449	1607	8.23	50,282.53	It did not express in roots or leaves either in any nitrogen or potassium level	AtACS8	NP_195491	Promotes early biosynthesis of ethylene that elicits plant immunity in Arabidopsis	42	
<i>MaACS6</i>	GSMUA_Achr5G08880_001	Ma05_g08580	5	6453130..6454804	Forward	1401	466	1675	8.47	52,017.26	It did not express in roots or leaves either in any nitrogen or potassium level	AtACS8	NP_195491	Promotes early biosynthesis of ethylene that elicits plant immunity in Arabidopsis	42	
<i>MaACS7</i>	GSMUA_Achr1G04420_001	Ma04_g24230	1	3704286..3705947	Forward	1305	462	1662	6.51	51,501.56	It did not express in roots or leaves either in any nitrogen or potassium level	AtACS8	NP_195491	Promotes early biosynthesis of ethylene that elicits plant immunity in Arabidopsis	42	
<i>MaACS8</i>	GSMUA_Achr4G01340_001	Ma04_g01260	4	1083786..1085438	Forward	948	462	1653	7.58	51,938.54	It did not express in roots or leaves either in any nitrogen or potassium level	AtACS8	NP_195491	Promotes early biosynthesis of ethylene that elicits plant immunity in Arabidopsis	42	
<i>MaACS9</i>	GSMUA_Achr1G14280_001	Ma01_g07800	1	10895407..10897053	Forward	1374	457	1647	7.15	51,130.24	It showed very slight expression in both low nitrogen and low potassium in roots	AtACS9	NP_190539	ACS proteins' stability and regulation of such stability play a central role in ethylene biosynthesis	44	
Continued																

ACS gene	Source accession no		Chromosome no	Chromosome location	Direction (bp)	mRNA	Size (AA)		pI	Mw (D)	Gene expression of banana in roots (Fig. 6)	Orthologue from model <i>Arabidopsis thaliana</i> and or other plants			
	Phytozome	Banana Genome Hub					Peptide	Genome				Gene ID	Accession no	Annotation/ function in plants	References
<i>Ma</i> ACS10	GSMUA_Achr10G27570_001	Ma10_g27510	10	30681506..30683062	Forward	477	440	1557	5.27	49,310.04	It did not express in roots or leaves either in any nitrogen or potassium level	AtACS7	NP_194350	It is involved in root gravitropism in <i>Arabidopsis thaliana</i>	46
<i>Ma</i> ACS11	GSMUA_Achr4G30870_001	Ma04_g37400	4	28274327..28275910	Forward		422	1584	5.35	47,224.60	It did not express in roots or leaves either in any nitrogen or potassium level	AtACS7	NP_194350	It is involved in root gravitropism in <i>Arabidopsis thaliana</i>	46
<i>Ma</i> ACS12	GSMUA_Achr1G18880_001	Ma01_g12130	1	14122481..14126525	Forward	1524	507	4045	6.99	55,839.15	It did not express in roots or leaves either in any nitrogen or potassium level	AtACS12	NP_001330155	This gene encodes aminotransferases	44
<i>Ma</i> ACS13	GSMUA_Achr3G12300_001	Ma03_g12030	3	9096959..9105359	Reverse	1.1494 2.486	684	8401	8.44	75,564.75	It did not express in roots or leaves either in any nitrogen or potassium level	AtACS12	NP_001330155	This gene encodes aminotransferases	44
<i>Ma</i> ACS14	GSMUA_Achr5G13730_001	Ma02_g09230	5	9876405..9880334	Forward	1464	503	6138	8.05	55,406.70	It expressed highly under low nitrogen or controlled nitrogen in roots and leaves both of plant	AtACS12	NP_001330155	This gene encodes aminotransferases	44

**Table 1.** Information about 14 *Ma*ACS genes discovered from the genome of *M. acuminata*, their orthologue in *Arabidopsis* and their functional annotation.



ACO gene	Source accession		Chromosome no	Chromosome location	Direction (bp)	Size (AA)		pI	M <sub>w</sub> (D) Gene ID	Gene expression of banana in roots (Fig. 6)	Orthologue from model <i>Arabidopsis thaliana</i> and or other plants			References
	Phytozome	Banana genome hub				Peptide	Gene ID				Accession no	Annotation/ function in plants		
MaACO1	GSMUA_Achr7T18420_001	Ma07_g9730	7	21215281..21216618	Forward	1338		4.99	110,033.70	It showed no expression in roots or leaves under N, K stress	AtACO3	NP_171655	Enzyme involved in the ethylene biosynthesis. May promote stem elongation by maximizing the extensibility of cells, possibly by activating ethylene biosynthesis, in response to very-long-chain fatty acids	47
MaACO2	GSMUA_Achr1T18250_001	Ma01_g11540	1	13635443..13636669	Forward	1227		4.97	101,618.97	It expressed moderately in roots during low potassium level	AtACO2	NP_192867	Required to mediate the 1-aminocyclopropane-1-carboxylic acid (ACC)-mediated reversion of the ABA-induced inhibition of seed germination via endosperm rupture	47
MaACO3	GSMUA_Achr10T16070_001	Ma10_g16100	10	23745515..23747172	Forward	1658		4.94	136,735.63	It expressed moderately high in roots during low nitrogen level	AtACO4	NP_192867	Activate ethylene biosynthesis in response to very-long-chain fatty acids	47
MaACO4	GSMUA_Achr6T12910_001	Ma06_g14390	6	8622468..8623805	Reverse	1338		5.00	109,682.80	It expressed moderately in roots during low potassium level	AtACO4	NP_190539	Enzyme involved in the ethylene biosynthesis. May promote stem elongation by maximizing the extensibility of cells, possibly by activating ethylene biosynthesis, in response to very-long-chain fatty acids	47
MaACO5	GSMUA_Achr6T00870_001	Ma06_g02600	6	598708..599734	Forward	1027		5.03	83,222.98	It expressed moderately in roots during low potassium and nitrogen level	AtACO5		Enzyme involved in ethylene biosynthesis	47
MaACO6	GSMUA_Achr5T09690_001	Ma05_g09360	5	6911660..6914329	Reverse	2670		4.86	221,359.93	It showed very slight expression in low nitrogen and low potassium in roots	AtACO5			
Continued														

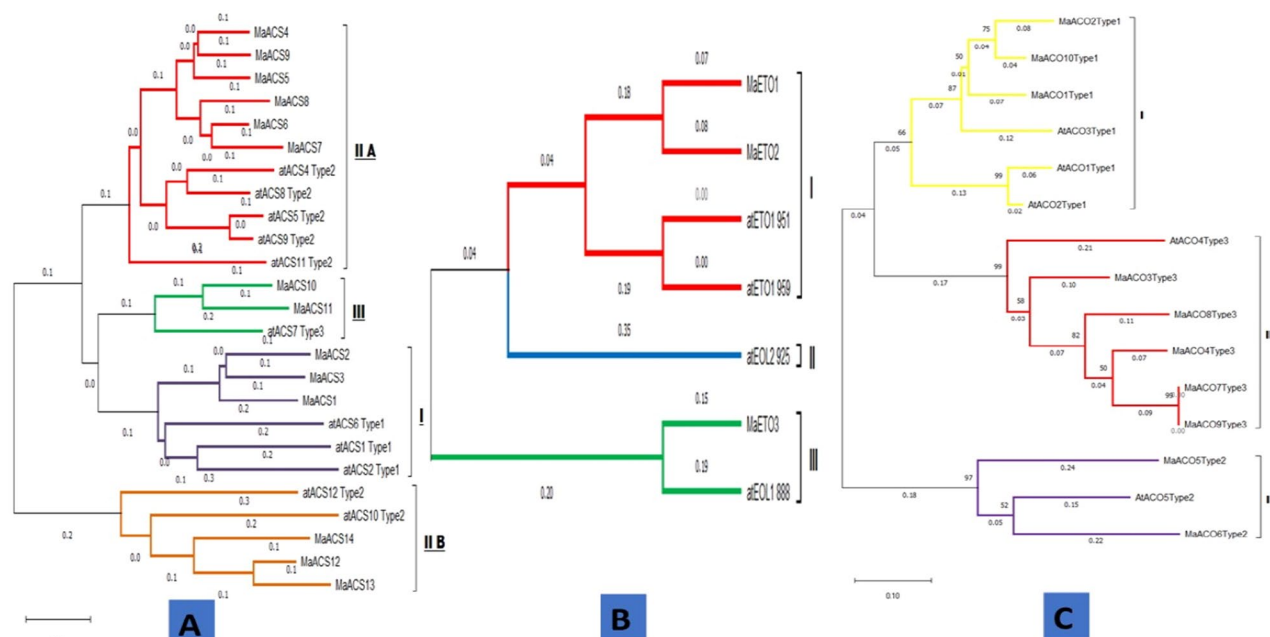
ACO gene	Source accession		Chromosome no	Chromosome location	Direction (bp)	Size (AA)		M <sub>w</sub> (D)	Gene ID	Gene expression of banana in roots (Fig. 6)	Orthologue from model <i>Arabidopsis thaliana</i> and or other plants			
	Phytozome	Banana genome hub				Peptide	pI				Gene ID	Accession no	Annotation/ function in plants	References
<i>MaACO7</i>	GSMUA_AchrUn_ran-domT20420_001	Ma06_g14370	Un	99012649..99015822	Forward	3074	4.79	252,277.03	AtACO4	It was not expressed in roots or leaves in any nitrogen or potassium level			Enzyme involved in the ethylene biosynthesis. May promote stem elongation by maximizing the extensibility of cells, possibly by activating ethylene biosynthesis, in response to very-long-chain fatty acids	47
<i>MaACO8</i>	GSMUA_AchrUn_ran-domT26960_001	Ma06_g14430	Un	132016004..132017996	Forward	1393	5.00	114,867.70	AtACO4	It showed moderate expression in both low nitrogen and low potassium in roots			Enzyme involved in the ethylene biosynthesis. May promote stem elongation by maximizing the extensibility cells	47
<i>MaACO9</i>	GSMUA_AchrUn_ran-domT20430_001	Ma06_g14420	Un	99027647..99028973	Forward	1327	4.92	108,789.51	AtACO4	It showed very slight expression in both low nitrogen and low potassium in roots			Activate ethylene biosynthesis in response to very-long-chain fatty acids	47
<i>MaACO10</i>	GSMUA_AchrUn_ran-domT28750_001	Ma0_g04770	Un	139442789..139443424	Forward	636	5.13	53,238.13	AtACO3	It did not express in roots or leaves either in any nitrogen or potassium level			Enzyme involved in the ethylene biosynthesis. May promote stem elongation by maximizing the extensibility of cells, possibly by activating ethylene biosynthesis, in response to very-long-chain fatty acids	47

**Table 2.** Information about 10 *MaACO* genes discovered from the genome of *M. acuminata*, their orthologue in *Arabidopsis* and their functional annotation.



ETO gene	Source accession		Chromosome no	Chromosome location	Direction (bp)	Size (AA)		pI	M <sub>w</sub> (D)
	Phytozome	Banana genome hub				Genomic	Peptide		
MaETO1	GSMUA_Achr6G32850_001	Ma06_g34510	6	32072728..32078672	Reverse	8518	969	6.02	107,866.66
MaETO2	GSMUA_Achr8G01920_001	Ma08_g01750	8	1432050..1436389	Reverse	5945	965	6.42	107,534.88
MaETO3	GSMUA_Achr4G21140_001	Ma04_g27750	4	21647264..21656370	Reverse	4340	893	5.78	101,212.79

**Table 3.** Information about 10 *MaACO* genes discovered from the genome of *M. acuminata*.



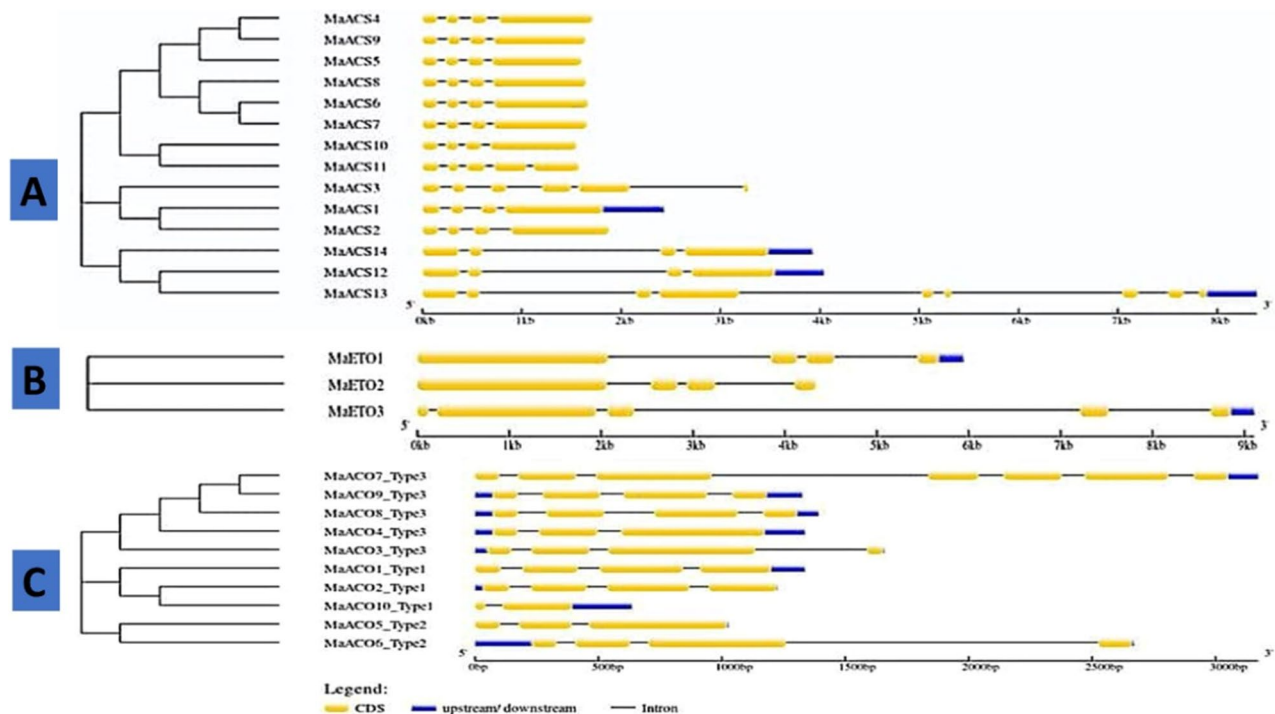
**Figure 1.** Phylogenetic analysis of *Musa acuminata* (*MaACS*) (A), *MaETO* (B) and *MaACO* (C) genes.

while for the other remaining 21 *ACS* paralogous pairs, the segmental duplication date of *M. acuminata* was estimated in the range from 46.08 to 88.78 Mya for paralogous pairs 15 and 11. All the paralogous groups in *M. acuminata* had a *Ka/Ks* ratio greater than 0.12, suggesting the possibility of considerable functional divergence after the duplication process (Fig. 4A, Table S7).

The date of gene duplication for *ETO* of *M. acuminata* was calculated 17.26 Mya for paralogous group *MaETO2\_MaETO1* (Table S8). Similarly, the gene duplication date for *MaACO* was calculated in 44 paralogous pairs. The segmental duplication date of banana was estimated from 0.45 to 24.13 Mya for paralogous pairs 38 and 14. In all paralogous groups of *MaACO* the *Ka/Ks* ratio is greater than 1.15 (Fig. 4B, Table S9).

### Analysis of *cis*-regulatory elements

The spatial-temporal transcriptomic expression of genes is affected by the presence and organization of various *cis*-regulatory elements at the binding site of transcription factors on the promoter region. In-silico analysis of *cis*-regulatory elements can be employed to evaluate the putative functions of genes. *Cis*-regulatory elements related to vital physiological processes such as response to light, seed-specific, endosperm-specific, hormone-specific, meristem-specific, and stress were observed (Fig. 5). Mainly, 9 out of 14 *MaACS*, all 3 *MaETO* and 5 out of 10 *MaACO* genes possess element involved in light responsiveness, 6 *MaACS* genes possess a fragment of a conserved DNA module that takes part in light responsiveness, and 9 *MaACS*, 2 *MaETO* and All 10 *MaACO* elements involved in the abscisic acid response. 1 *MaACS* and 1 *MaETO* gene possess elements involved in salicylic acid responsiveness, while 3 *MaACS* genes showed elements that show response in defense and stress, and 3 *MaACS* and 10 *MaACO* genes are related to meristem expression, 3 *MaACS* and 1 *MaETO* gene possess factors involved in low-temperature responsiveness, 4 *MaACS*, and 3 *MaETO* genes possess Auxin-responsive element, 2 *MaACS*, and 2 *MaETO* genes showed element specific to anoxic inducibility, 1 *MaACS* genes possess elements specific to seed regulation, 4 *MaACS* genes contain elements involved in endosperm expression, 6 *MaACS*, and 1 *MaETO* gene possess elements involved in element involved in gibberellin response, 6 *MaACO* genes essential for the anaerobic induction and all 10 *MaACO* genes involved in *MeJA*-responsiveness. The *cis*-regulatory



**Figure 2.** Phylogenetic relationship and gene structure of ACS, ETO and ACO genes from *M. acuminata*.

elements identified among 14 ACS, 3 ETO, and 10 MaACO genes of banana and their functional annotation are shown in (Fig. 5, Tables S10, S11 and S12).

### General miRNA expression dataset of *Musa acuminata*

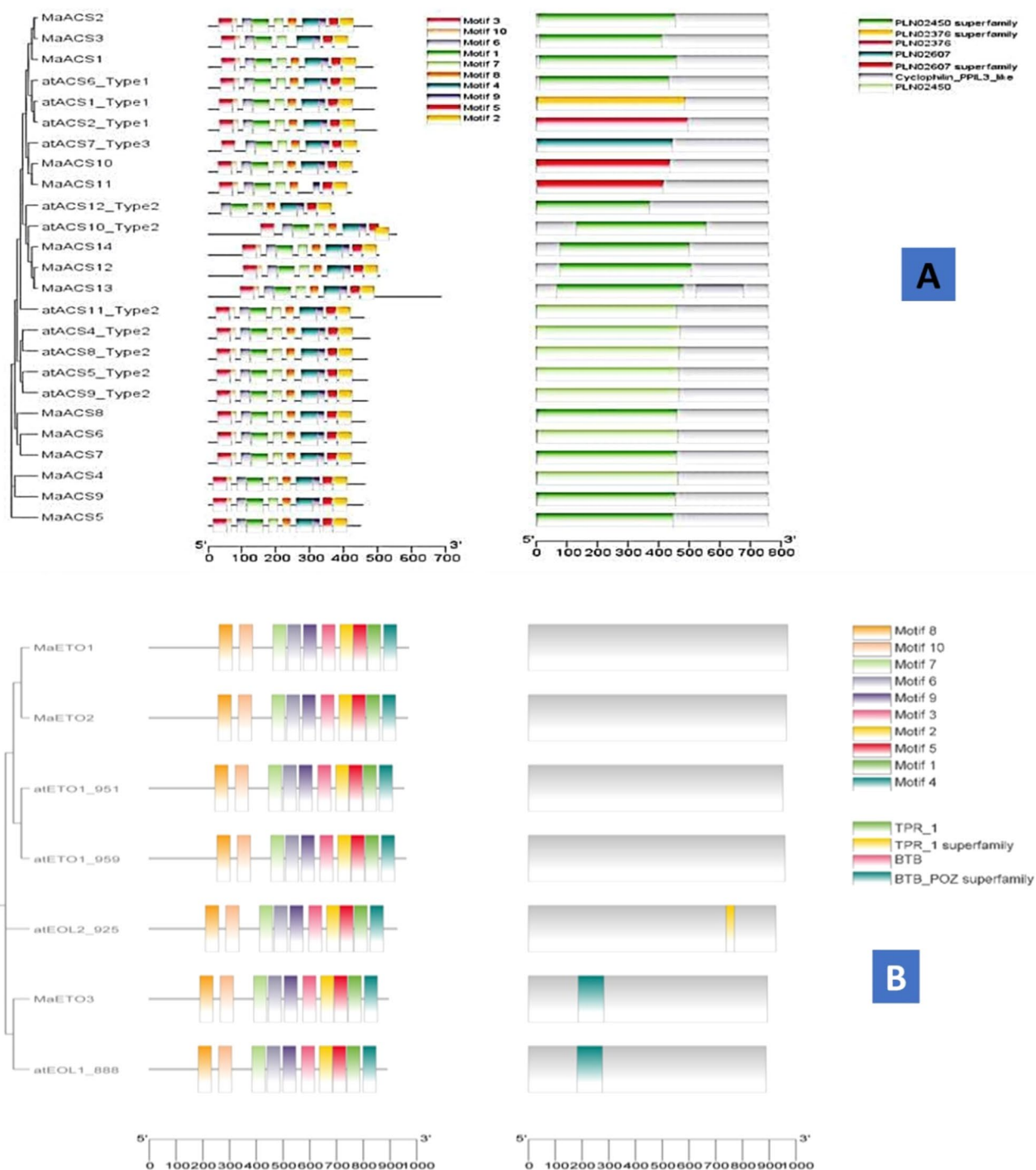
Heat map for the expression profile of *Musa acuminata* genes in root under normal and low nitrogen and potassium shows that *MaACS14* and *MaACO6* expressed highly at a normal nitrogen supply level. Still, they do not respond significantly at low nitrogen levels. *MaACO5*, *MaACO3*, *MaACO8* show very slight expression at normal nitrogen levels, while *MaACS14* express slightly at low nitrogen levels. On the other hand, *MaACS1* expresses well at low levels of potassium, whereas *MaACO6* gives a significant response at normal levels of potassium. *MaACS4*, *MaACO4* and *MaACO5* show slight behavior in response to low potassium. (Fig. 6). The expression of each banana gene in roots is explained in Table 1.

The mature miRNAs sequences were retrieved from Plant MicroRNA Encyclopedia database. Later, those miRNAs that could potentially target *MaACS* genes were identified with the help of the psRNA Target online tool (<https://plantgrn.noble.org/psRNA/Target/analysis>). Consequently, 49 miRNAs were found, targeting 8 out of 14 *MaACS* genes. The remaining 6 *MaACS* genes were not targeted by any of these miRNAs (Table 4). The number of miRNAs targeting these genes varies from 1 to 20 miRNAs per *MaACS* gene. *MaACS* 4, 9, 13, 14 are the genes targeted by only 1 miRNAs. On the other hand, *MaACS* 14 is targeted by 2 miRNAs. 3 miRNAs target *MaACS* 4. None of the gathered miRNAs targeted the remaining 6 *MaACS* genes. So, this indicates that *MaACS* 4 is the gene targeted by the maximum number of miRNAs. While discussing based on groups, Group II A was targeted the most by these miRNAs, which was targeted by 3 miRNAs. On the other hand, Group A was targeted by only 3 miRNAs, which is the least among all (Table 4, Table S13).

In *MaACO*, there are total 32 miRNAs that targeted 6 *MaACO* genes out of 10. The number of targeting miRNAs varies from 1 to 17. *MaACO5* targeted 2 miRNAs, and the remaining 4 genes were targeted by 1 miRNA. So, this depicts that *MaACO5* is the gene targeted by a maximum number of miRNAs (Table 4, Table S14). In *ETO* of banana, there are two 2 targeted miRNA for both *MaETO1* and *MaETO2*. The total number of targets is 11, the maximum number of miRNA targets on the *MaETO1* gene (Table 4, Table S15).

### Discussion

The coordination of genes, hormones, and environmental factors made the ripening process successful<sup>57</sup>. Genes like ACS, ACO and ETO involved in ethylene biosynthesis pathway<sup>5,6</sup> perform key role in ripening<sup>58</sup>. Banana genome database ([https://phytozome-next.jgi.doe.gov/info/Macuminata\\_v1](https://phytozome-next.jgi.doe.gov/info/Macuminata_v1)) implied to identify 14 ACS genes, 10 ACO genes and 3 ETO genes (Table 1) at the genome-wide level. The 14 banana ACS genes were classified into 4 subfamilies (Group I, IIA, IIB, III), whereas in ETO and ACO, there were divided into 3 subgroups (Group I, II, III) using the phylogenetic analysis (Fig. 1). The exon–intron structure and prediction can also be used as evidence for understanding the evolutionary relationships among genes or organisms<sup>48,59,60</sup>. The predicted exon–intron association revealed that all 14 genes have introns (Fig. 2A). The number of introns varied from 3 to 9 in *MaACS* gene (Fig. 2B), 3 to 4 in *MaETO* genes (Fig. 2B) and 1–6 in *MaACO* genes of banana (Fig. 2C).



**Figure 3.** The distribution of 10 motifs present in *MaACS* (A), *MaETO* (B) and *MaACO* (C) protein of banana.

In general, it shows that banana *MaACS* genes in the same group share similar exon–intron structures (Fig. 2). Exon–intron having similar structures have also been noticed in Arabidopsis, rice and soybean<sup>61,62</sup> which suggest that these structures are evolutionarily preserved.

Motif analysis shows that Motif 3 is common in all groups except in *AtACS-12\_Type 2* gene of group IIB. While, IIB group members have relatively complex motif patterns compared with Group IIA (Fig. 3A). Figure 3B shows that *MaETO* gene structures are similar to the corresponding *AtETO* gene structures. BTB\_POZ domain is only present in *MaETO* gene that was vital for *AtETO* genes (Fig. 3B). Furthermore, *MaACO7* has motifs and PLN02299 domain repeats, similar to *CoACO1* and *CoACO2* in *Camellia oleifera*<sup>63</sup>. PLN02403 domain is only present in *MaACO5* and *MaACO6* (Fig. 3C). The distribution of motifs among ACS proteins (Fig. 3) indicates evolutionary and structural relationships as deduced by the phylogenetic tree<sup>64–66</sup>. The motif data analysis by MEME (Fig. 3), and domain analysis using NCBI CDD distinct motifs were identified that were differentially

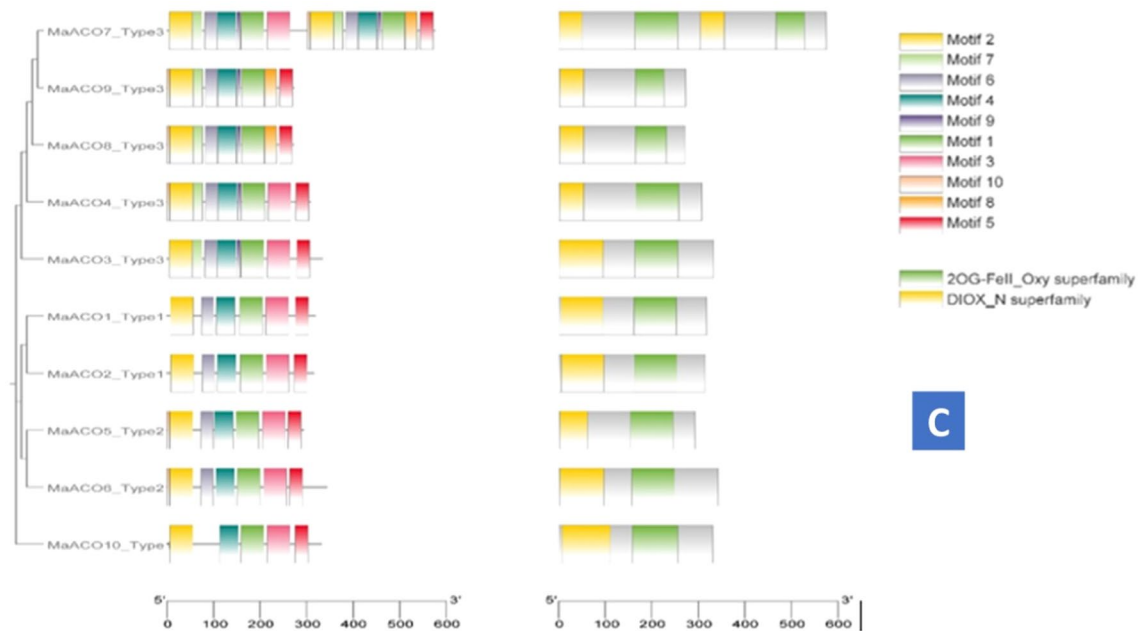
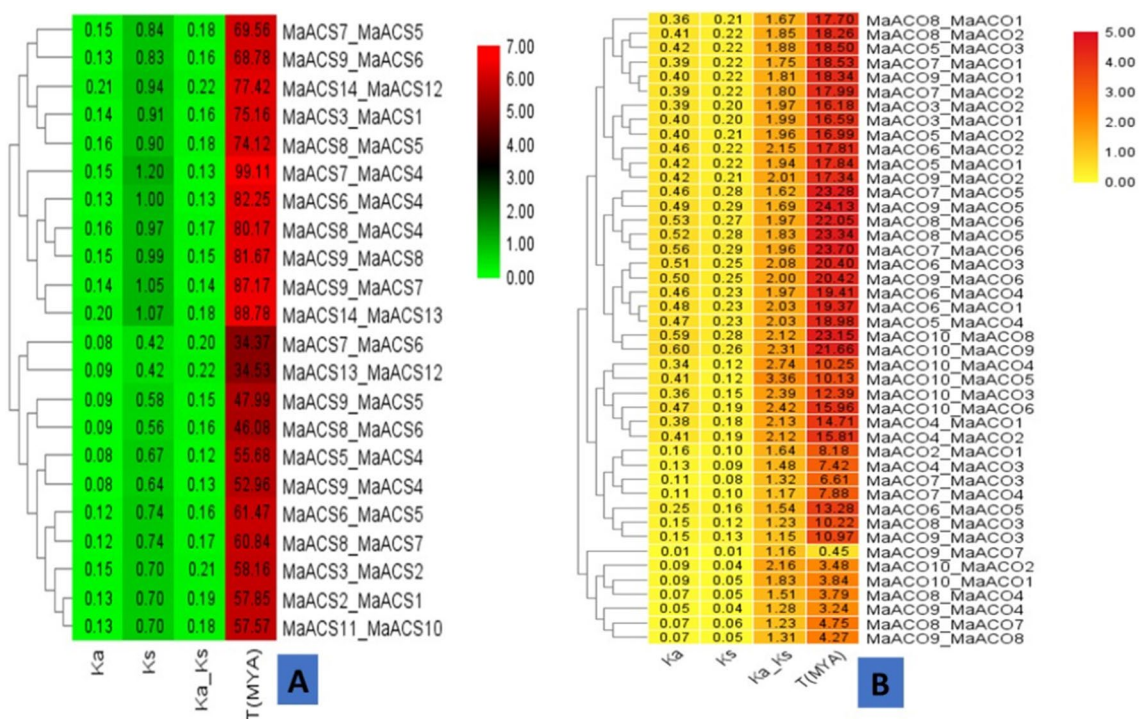


Figure 3. (continued)

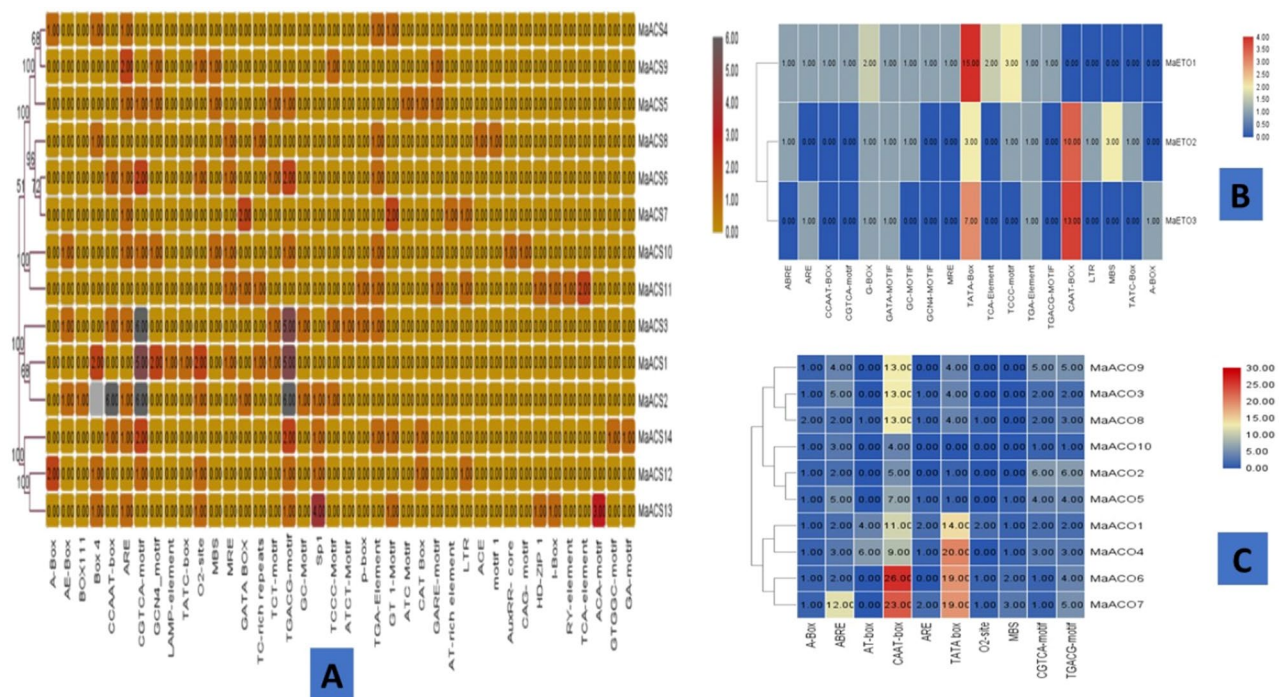


**Figure 4.** Time of gene duplication estimated for different paralogous pairs of *MaACS* (A) and *MaACO* (B) genes based on  $K_s$  and  $K_a$  values.

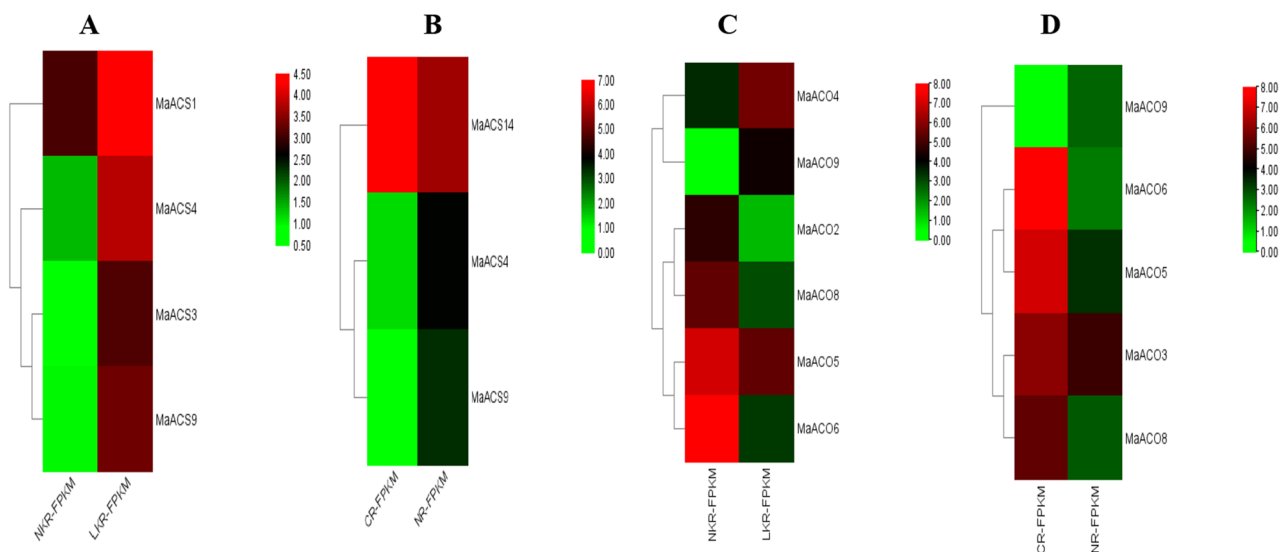
distributed among *MaACS* (Fig. 3). Meanwhile, at least one or two conservative motif types and spatial distributions in *MaACS* were present in the same subfamily while some differences were present, implying certain functional similarities of banana ACS members within the same subfamily. In addition, *MaACS* genes showed structural conservation in subfamilies and were consistent with other plants such as *Arabidopsis*, rice, cotton and chickpea<sup>61,67–70</sup>. In addition, as predicted by in silico analyses, 3 deduced *MaACS* harbored NLSs to help localize them to the nucleus, but subcellular localization analysis using online tool WoLF PSORT (<https://wolfpsort.hgc.jp/>), supposed nucleolus localization in almost all *MaACS* protein except *MaACS* 12 and *MaACS* 14.

The ratio of  $K_a/K_s$  provides an understanding of the selection pressure on substituting amino acids. Less than one ratio of  $K_a/K_s$  ( $K_a/K_s < 1$ ) suggests the possibility of purifying selection, whereas more than one ratio of  $K_a/K_s$  ( $K_a/K_s > 1$ ) suggests the likelihood of positive selection<sup>39,71,72</sup>. Generally, evaluation of selective pressure





**Figure 5.** Different *cis*-acting elements in putative *MaACS* (A), *MaETO* (B) and *MaACO* (C) promoters which are associated with abiotic stresses, hormone responses, growth and development.



**Figure 6.** The heat map shows the expression profile of the *M. acuminata* ACS (B) and ACO (D) genes in root under the normal level of nitrogen (CR) and low level of nitrogen (NR) and ACS (A) and ACO (C) genes in roots under the normal level of potassium and low level of potassium in different organs. The expression levels of *MaACS* and *MaACO* genes are revealed by different colors, which increase from green to red.

provides a particular lead for amino acid sequence altered in a protein and is also necessary for interpreting functional residues and protein shifts<sup>73</sup>. Ka/Ks ratios of the sequences from the different banana *MaACS* groups vary remarkably, while in *ETO*, only one gene showed Ka/Ks ratio. Despite the differences, all the estimated values of Ka/Ks were less than 1, suggesting that all the ACS sequences in each group undergo strong purifying selection pressure and positive selection might have acted on only a few sites during the process of evolution. In *MaACO*, all the estimated values of Ka/Ks which were more than one (> 1) shows the possibility of existence of significant positive selection after duplication (Fig. 4).

Heat map for the expression profile of *Musa acuminata* genes in root under normal and low nitrogen and potassium shows that *MaACS14* and *MaACO6* expressed highly at normal nitrogen supply. Whereas, *MaACS1* represents well at low potassium levels, whereas *MaACO6* gives a significant response at normal levels of

miRNA_Acc	Target_Acc	No. of target found	Function of miRNA
Mac-miR396	MaACS6	20	Induced by cold but greatly repressed by heat stress <sup>49</sup>
Mac-miR396	MaACS8	2	Induced by cold but greatly repressed by heat stress <sup>49</sup>
Mac-miR396	MaACS14	1	Induced by cold but greatly repressed by heat stress <sup>49</sup> Regulates transition of Arabidopsis root stem cells by transit-amplifying cells to form a regulatory circuit by repressing GRF <sup>50</sup>
Mac-miRN2003	MaACS2	3	
Mac-miRN2003	MaACS3	3	
Mac-miR159	MaACS4	5	Induced by cold but greatly repressed by heat stress <sup>49</sup>
Mac-miR319	MaACS4	11	Induced by cold but not by dehydration, salinity, or ABA <sup>51,52</sup>
Mac-miRN2002	MaACS13	1	
Mac-miRN2002	MaACS14	1	
Mac-miRN1993	MaACS9	1	
Mac-miRN2010	MaACS4	1	
Mac-miR390	MaACO7	6	
Mac-miR390	MaACO9	6	
Mac-miR391	MaACO5	1	
Mac-miR172	MaACO5	17	An important role during the banana fruit ripening process <sup>53</sup>
Mac-miR827	MaACO6	1	
Mac-miR1999	MaACO2	1	
Mac-miR171	MaETO1	7	Inhibit translation in plants <sup>54</sup>
Mac-miR428	MaETO1	1	The second most prevalent miRNA in NC, neural and ectoderm tissue but was not expressed in blastula <sup>55</sup>
Mac-miR 2009	MaETO2	2	
Mac-miR528	MaETO2	1	Significantly downregulated under cold stress and it was found to target genes encoding polyphenol oxidase (PPO) and copper-containing proteins <sup>56</sup>

**Table 4.** miRNA targets prediction of *MaACS*(A), *MaACO*(B) and *MaETO*(C).

potassium (Fig. 5). In contrast, ACS genes exhibited low expression under potassium and cadmium stress<sup>74</sup>. Similarly, ACS1, ACS4 and ACS7 expression were enhanced in response to UV treatment in tomatoes, while ACS3, ACS5 and ACS6 showed no variation<sup>75</sup>.

MicroRNAs are very important regulators of plants that regulate almost every biological process, ranging from growth and development to combating pathogens and maintaining proper internal conditions, as miRNA affects many genes of specified functions<sup>43,76–79</sup>. miRNAs are highly conserved among different species, as each microRNA performs a specific function, regardless of the type of species in which they were observed. In *Musa acuminata*, *MaACS4* is targeted by 3 miRNA i.e. Mac-miR159, Mac-miR319, and Mac-miRN2010, which are induced in cold stress and repressed in heat stress<sup>49</sup>, dehydration, salinity, or ABA<sup>52</sup> and involved in root formation and cell elongation<sup>51</sup>. Mac-miR396 is an important miRNA that targets *MaACS6*, *MaACS8* and *MaACS14* that activate in cold stress and repress in heat stress<sup>49</sup>. It also regulates the transition of Arabidopsis root stem cells by transit-amplifying cells to form a regulatory circuit by repressing GRF<sup>50</sup>. *MaACS14* is targeted by 2 miRNAs, Mac-miRN2002 and Mac-miR396, which are important for root stem cells in Arabidopsis<sup>50</sup>. *MaACO5* is targeted by 2 miRNA Mac-miR391 and Mac-miR172, which are involved in the ripening process of banana<sup>53</sup>. Mac-miR172 is a known MiRNA of banana<sup>53</sup>, has maximum targets on *MaACO5* and does not target any other protein in *MaACO*. In *MaETO*, both *MaETO1* and *MaETO2* are targeted by 2 miRNA. Mac-miR171 and Mac-miR428 both target on *MaETO1*, in which Mac-miR171 inhibits translation in plants<sup>54</sup>, and Mac-miR428 is expressed in neural and ectoderm tissue but was not expressed in blastula<sup>55</sup>. While Mac-miR N2009 and Mac-miR528 target *MaETO2*. Mac-miR528 targets many genes encoding copper-containing proteins and polyphenol oxidase (PPO), and is downregulated in cold stress<sup>56</sup>. It is reported that three specialized miRNAs (miR173, miR390 and miR828) have been identified and well characterized in *Arabidopsis*<sup>80–82</sup>. Only miR390 is present in *MaACO7* and *MaACO9*. This might suggest that Mac-miR396 in *MaACS* genes are important for root cell elongation and maintaining the regulatory mechanism between root and stem.

## Conclusion

A comprehensive analysis of genes (*ACS*, *ACO*, and *ETO1*) involved in ethylene biosynthesis in the *Musa acuminata* (Banana) genome was discussed in this study. The 15 genes of *ACS*, 10 genes of *ACO* and 3 genes of *ETO1* were categorised into subgroups. The structural and functional properties of each *MaACS*, *MaACO* and *MaETO* member were characterized under Nitrogen (N) and Potassium (K) stress in plant roots, where, *MaACO6* expressed highly at both normal and low level of N and P, and *MaACS14* expressed well at low nitrogen level, while, *MaACS4*, *MaACO4* and *MaACO5* responded to low potassium. Most genes were involved in root cell formation and maintaining the regulation mechanism between roots and stem, suggesting their role in plant root growth and development. The detailed computational inspection of Banana *ACS*, *ACO* and *ETO* proteins revealed in the current study might be selected for cloning purposes at the molecular level, portraying gene expression and studying their interaction with different transcription factors.



## Data availability

The datasets generated and/or analysed during the current study are available in the manuscript.

Received: 5 April 2023; Accepted: 30 December 2023

Published online: 04 January 2024

## References

- Yang, S. F. & Hoffman, N. E. Ethylene biosynthesis and its regulation in higher plants. *Annu. Rev. Plant Physiol.* **35**, 155–189 (1984).
- Reddy, K. J. *Physiology and Molecular Biology of Stress Tolerance in Plants* 187–217 (Springer, 2006).
- Zeng, H. *et al.* Role of microRNAs in plant responses to nutrient stress. *Plant Soil* **374**, 1005–1021 (2014).
- Kanojia, A., Xu, X. & Dijkwel, P. P. *The Plant Hormone Ethylene* 73–87 (Elsevier, 2023).
- Jiao, X. Z., Yip, W. K. & Yang, S. F. The effect of light and phytochrome on 1-aminocyclopropane-1-carboxylic acid metabolism in etiolated wheat seedling leaves. *Plant Physiol.* **85**, 643–647. <https://doi.org/10.1104/pp.85.3.643> (1987).
- Wang, T. T. & Yang, S. F. The physiological role of lipoxygenase in ethylene formation from 1-aminocyclopropane-1-carboxylic acid in oat leaves. *Planta* **170**, 190–196. <https://doi.org/10.1007/BF00397887> (1987).
- Binder, B. M. Ethylene signaling in plants. *J. Biol. Chem.* **295**, 7710–7725 (2020).
- Sadat-Hosseini, M., Bakhtiarizadeh, M. R., Boroomand, N., Tohidfar, M. & Vahdati, K. Combining independent de novo assemblies to optimize leaf transcriptome of Persian walnut. *PLoS One* **15**, e0232005. <https://doi.org/10.1371/journal.pone.0232005> (2020).
- Kanodia, P. *et al.* A rapid and simple quantitative method for specific detection of smaller coterminal RNA by PCR (DeSCO-PCR): Application to the detection of viral subgenomic RNAs. *RNA* **26**, 888–901. <https://doi.org/10.1261/rna.074963.120> (2020).
- Wu, L., Liu, S., Qi, H., Cai, H. & Xu, M. Research progress on plant long non-coding RNA. *Plants (Basel)* <https://doi.org/10.3390/plants9040408> (2020).
- Zhou, S. *et al.* Full-length transcriptome sequences of *Agropyron cristatum* facilitate the prediction of putative genes for thousand-grain weight in a wheat–*A. cristatum* translocation line. *BMC Genomics* **20**, 1025. <https://doi.org/10.1186/s12864-019-6416-4> (2019).
- Ziyuan, L. *et al.* Molecular cloning and functional analysis of lotus salt-induced NnDREB2C, NnPIP1-2 and NnPIP2-1 in *Arabidopsis thaliana*. *Mol. Biol. Rep.* **47**, 497–506. <https://doi.org/10.1007/s11033-019-05156-0> (2020).
- Sun, W. *et al.* Genome-wide identification of AGO18b-bound miRNAs and phasiRNAs in maize by cRIP-seq. *BMC Genomics* **20**, 656. <https://doi.org/10.1186/s12864-019-6028-z> (2019).
- Takeshi Yamagami, A. T., Yamada, K., Haddon, W. F. & Harden, L. A. Biochemical diversity among the 1-amino-cyclopropane-1-carboxylate synthase isozymes encoded by the Arabidopsis gene family. *J. Biol. Chem.* **278**, 49102–49112 (2003).
- Ali, S. Response of rice under salt stress. *Biol. Agric. Sci. Res. J.* **2022**, 6 (2022).
- Amjad, N., Rehman, M., Khalid, M. & Amjad, I. Construction of the infectious molecule of beta satellite associated with ageratum yellow vein disease of ageratum conyzoides. *Bull. Biol. Allied Sci. Res.* **2018**, 13–13 (2018).
- Rasheed, M. & Malik, A. Mechanism of drought stress tolerance in wheat. *Bull. Biol. Allied Sci. Res.* **2022**, 23–23 (2022).
- Yoshida, H. *et al.* Arabidopsis ETO1 specifically interacts with and negatively regulates type 2 1-aminocyclopropane-1-carboxylate synthases. *BMC Plant Biol.* **5**, 14 (2005).
- Haider, M. *et al.* Exploring morphological traits variation in *Gomphrena globosa*: A multivariate analysis. *Biol. Agric. Sci. Res. J.* **2023**, 21–21 (2023).
- Abeles, F. B., Morgan, P. W. & Saltveit, M. E. Jr. *Ethylene in Plant Biology* (Academic Press, 1992).
- Acaster, M. A. & Kende, H. Properties and partial purification of 1-aminocyclopropane-1-carboxylate synthase. *Plant Physiol.* **72**, 139–145 (1983).
- Kende, H. Ethylene biosynthesis. *Annu. Rev. Plant Biol.* **44**, 283–307 (1993).
- Fatima, S. *et al.* The genome-wide bioinformatics analysis of 1-aminocyclopropane-1-carboxylate synthase (acs), 1-aminocyclopropane-1-carboxylate oxidase (aco) and ethylene overproducer 1 (eto1) gene family of *Fragaria vesca* (woodland strawberry). *Bull. Biol. Allied Sci. Res.* **2023**, 38–38 (2023).
- Yoshida, H., Nagata, M., Saito, K., Wang, K. L. & Ecker, J. R. Arabidopsis ETO1 specifically interacts with and negatively regulates type 2 1-aminocyclopropane-1-carboxylate synthases. *BMC Plant Biol.* **5**, 1–13 (2005).
- Abbas, A., Rehman, A. & Javed, M. Exploring the potential of in vitro tissue culture in breeding programs of legume and pulse crops: Utilization and present condition. *Bull. Biol. Allied Sci. Res.* **2021**, 36–36 (2021).
- Guzman, P. & Ecker, J. R. Exploiting the triple response of Arabidopsis to identify ethylene-related mutants. *Plant Cell* **2**, 513–523 (1990).
- Vogel, J. P., Woeste, K. W., Theologis, A. & Kieber, J. J. Recessive and dominant mutations in the ethylene biosynthetic gene ACS5 of Arabidopsis confer cytokinin insensitivity and ethylene overproduction, respectively. *Proc. Natl. Acad. Sci. USA* **95**, 4766–4771 (1998).
- Finn, R. D. *et al.* Pfam: The protein families database. *Nucleic Acids Res.* **42**, D222–D230 (2014).
- Goodstein, D. *et al.* Phytozome comparative plant genomics portal (2014).
- Lu, S. *et al.* CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Res.* **48**, D265–D268 (2020).
- Gasteiger, E. *et al.* *The Proteomics Protocols Handbook* 571–607 (Springer, 2005).
- Horton, P., Park, K.-J., Obayashi, T. & Nakai, K. *Proceedings of the 4th Asia-Pacific Bioinformatics Conference* 39–48 (World Scientific, 2006).
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
- Thompson, J. D., Gibson, T. J. & Higgins, D. G. Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinform.* **1**, 2.3.1–2.3.22 (2003).
- Tamura, K. *et al.* MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).
- Hu, B. *et al.* GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics* **31**, 1296–1297 (2015).
- Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).
- Chen, C. *et al.* TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **13**, 1194–1202. <https://doi.org/10.1016/j.molp.2020.06.009> (2020).
- Yang, Z. & Bielawski, J. P. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**, 496–503 (2000).
- Hurst, L. D. The Ka/Ks ratio: Diagnosing the form of sequence evolution. *Trends Genet.* **18**, 486–487 (2002).
- Rombauts, S., Déhais, P., Van Montagu, M. & Rouzé, P. PlantCARE, a plant cis-acting regulatory element database. *Nucleic Acids Res.* **27**, 295–296 (1999).
- Guo, Z. *et al.* PmiREN: A comprehensive encyclopedia of plant miRNAs. *Nucleic Acids Res.* **48**, D1114–D1121 (2020).
- Samad, A. F. A. MicroRNA and transcription factor: Key players in plant regulatory network. *Front. Plant Sci.* **8**, 565 (2017).
- Yamagami, T. *et al.* Biochemical diversity among the 1-amino-cyclopropane-1-carboxylate synthase isozymes encoded by the Arabidopsis gene family. *J. Biol. Chem.* **278**, 49102–49112 (2003).

45. Liu, Y. & Zhang, S. Phosphorylation of 1-aminocyclopropane-1-carboxylic acid synthase by MPK6, a stress-responsive mitogen-activated protein kinase, induces ethylene biosynthesis in Arabidopsis. *Plant Cell* **16**, 3386–3399 (2004).
46. Huang, S.-J. *et al.* A type III ACC synthase, ACS7, is involved in root gravitropism in *Arabidopsis thaliana*. *J. Exp. Bot.* **64**, 4343–4360 (2013).
47. Qin, Y.-M. *et al.* Saturated very-long-chain fatty acids promote cotton fiber and Arabidopsis cell elongation by activating ethylene biosynthesis. *Plant Cell* **19**, 3692–3704 (2007).
48. Koralewski, T. E. & Krutovsky, K. V. Evolution of exon-intron structure and alternative splicing. *PLoS One* **6**, e18055. <https://doi.org/10.1371/journal.pone.0018055> (2011).
49. Fang, Z. *et al.* Genome-wide identification, structure characterization, and expression profiling of Dof transcription factor gene family in wheat (*Triticum aestivum* L.). *Agronomy* **10**, 294 (2020).
50. Rodriguez, R. E. *et al.* MicroRNA miR396 regulates the switch between stem cells and transit-amplifying cells in Arabidopsis roots. *Plant Cell* **27**, 3354–3366 (2015).
51. Gudimella, R. *et al.* Genome-wide regulatory network mapping of miRNA and transcription factors in banana roots. *Trop. Plant Biol.* **11**, 141–153 (2018).
52. Sunkar, R. & Zhu, J.-K. Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. *Plant Cell* **16**, 2001–2019 (2004).
53. Bi, F., Meng, X., Ma, C. & Yi, G. Identification of miRNAs involved in fruit ripening in Cavendish bananas by deep sequencing. *BMC Genomics* **16**, 1–15 (2015).
54. Brodersen, P. *et al.* Widespread translational inhibition by plant miRNAs and siRNAs. *Science* **320**, 1185–1190 (2008).
55. Ward, N. J. *et al.* microRNAs associated with early neural crest development in *Xenopus laevis*. *BMC Genomics* **19**, 1–13 (2018).
56. Zhu, H. *et al.* Micro RNA 528, a hub regulator modulating ROS homeostasis via targeting of a diverse set of genes encoding copper-containing proteins in monocots. *New Phytol.* **225**, 385–399 (2020).
57. Wee, C.-C., Subbiah, V. K., Arita, M. & Goh, H.-H. The applications of network analysis in fruit ripening. *Sci. Hortic.* **311**, 111785 (2023).
58. Ahmadzadeh, M., Chen, J.-T., Hasanzadeh, S., Ahmar, S. & Heidari, P. Insights into the genes involved in the ethylene biosynthesis pathway in *Arabidopsis thaliana* and *Oryza sativa*. *J. Genet. Eng. Biotechnol.* **18**, 1–20 (2020).
59. Poverennaya, I. & Roytberg, M. Spliceosomal introns: Features, functions, and evolution. *Biochemistry (Moscow)* **85**, 725–734 (2020).
60. Bondarenko, V. S. & Gelfand, M. S. Evolution of the exon-intron structure in ciliate genomes. *PLoS One* **11**, e0161476. <https://doi.org/10.1371/journal.pone.0161476> (2016).
61. Lijavetzky, D., Carbonero, P. & Vicente-Carbajosa, J. Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families. *BMC Evol. Biol.* **3**, 17 (2003).
62. Gu, X. *et al.* An update of DIVERGE software for functional divergence analysis of protein family. *Mol. Biol. Evol.* **30**, 1713–1719 (2013).
63. Hu, X. *et al.* Ethylene-regulated immature fruit abscission is associated with higher expression of CoACO genes in *Camellia oleifera*. *R. Soc. Open Sci.* **8**, 202340 (2021).
64. Bano, N., Fakhrah, S., Lone, R. A., Mohanty, C. S. & Bag, S. K. Genome-wide identification and expression analysis of the HD2 protein family and its response to drought and salt stress in *Gossypium* species. *Front. Plant Sci.* **14**, 1109031 (2023).
65. Gupta, S. *et al.* Insights into structural and functional diversity of Dof (DNA binding with one finger) transcription factor. *Planta* **241**, 549–562 (2015).
66. Malviya, N. *et al.* Genome wide in silico characterization of Dof gene families of pigeonpea (*Cajanus cajan* (L.) Millsp.). *Mol. Biol. Rep.* **42**, 535–552 (2015).
67. Xue, T. *et al.* Genome-wide investigation and co-expression network analysis of SBT family gene in *Gossypium*. *Int. J. Mol. Sci.* **24**, 5760 (2023).
68. Dong, C., Hu, H. & Xie, J. Genome-wide analysis of the DNA-binding with one zinc finger (Dof) transcription factor family in bananas. *Genome* **59**, 1085–1100. <https://doi.org/10.1139/gen-2016-0081> (2016).
69. Nasim, J., Malviya, N., Kumar, R. & Yadav, D. Genome-wide bioinformatics analysis of Dof transcription factor gene family of chickpea and its comparative phylogenetic assessment with Arabidopsis and rice. *Plant Syst. Evol.* **302**, 1009–1026 (2016).
70. Yang, X. & Tuskan, G. A. Divergence of the Dof gene families in poplar, Arabidopsis, and rice suggests multiple modes of gene evolution after duplication. *Plant Physiol.* **142**, 820–830 (2006).
71. Shen, T. *et al.* The chromosome-level genome sequence of the camphor tree provides insights into Lauraceae evolution and terpene biosynthesis. *Plant Biotechnol. J.* **20**, 244 (2022).
72. Hurst, L. D. The Ka/Ks ratio: Diagnosing the form of sequence evolution. *Trends Genet.* **9**, 486–487 (2002).
73. Morgan, C. C., Loughran, N. B., Walsh, T. A., Harrison, A. J. & O'Connell, M. J. Positive selection neighboring functionally essential sites and disease-implicated regions of mammalian reproductive proteins. *BMC Evol. Biol.* **10**, 39 (2010).
74. Huang, B. *et al.* Effect of potassium intake on cadmium transporters and root cell wall biosynthesis in sweet potato. *Ecotoxicol. Environ. Saf.* **250**, 114501 (2023).
75. Mannucci, A. *et al.* Differential expression of 1-aminocyclopropane-1-carboxylate synthase and oxidase gene family in micro-tom tomato leaves and roots under short daily UV radiation. *J. Plant Growth Regul.* **42**, 637–643 (2023).
76. Komatsu, S., Kitai, H. & Suzuki, H. I. Network regulation of microRNA biogenesis and target interaction. *Cells* **12**, 306 (2023).
77. Carbone, F. *et al.* Identification of miRNAs involved in fruit ripening by deep sequencing of *Olea europaea* L. transcriptome. *PLoS One* **14**, e0221460. <https://doi.org/10.1371/journal.pone.0221460> (2019).
78. Spanudakis, E. The role of microRNAs in the control of flowering time. *J. Exp. Bot.* **65**, 365–380 (2014).
79. Terzi, L. C. *Regulation of Flowering Time by RNA Processing* (Springer Link, 2008).
80. Allen, E., Xie, Z., Gustafson, A. M. & Carrington, J. C. MicroRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* **121**, 207–221 (2005).
81. Rajagopalan, R., Vaucheret, H., Trejo, J. & Bartel, D. P. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev.* **20**, 3407–3425 (2006).
82. Yoshikawa, M., Peragine, A., Park, M. Y. & Poethig, R. S. A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. *Genes Dev.* **19**, 2164–2175 (2005).

## Acknowledgements

This work was supported by the University of the Punjab, Lahore, Pakistan.

## Author contributions

Conceptualization, N.T. and M.S.; methodology, S.F.; software, S.T. and B.T.; validation, M.S. and Q.A.; formal analysis, M.A.J.; investigation, M.S.; resources, Q.A., and S.F.; data curation, M.S.; writing—original draft preparation, N.T. and Q.A.; writing—review and editing, S.T., Q.A., M.A.J., and M.S.; visualization, M.S. and M.A.J.;

supervision, M.S.; project administration, M.S. All authors have read and agreed to the published version of the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-51075-3>.

**Correspondence** and requests for materials should be addressed to M.S. or Q.A.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024