



# A Systematic Phylogenomic Classification of the Multidrug and Toxic Compound Extrusion Transporter Gene Family in Plants

Manduparambil Subramanian Nimmy<sup>1†</sup>, Vinod Kumar<sup>2†</sup>, Backiyarani Suthanthiram<sup>3</sup>, Uma Subbaraya<sup>3</sup>, Ramawatar Nagar<sup>1</sup>, Chellapilla Bharadwaj<sup>4</sup>, Pradeep Kumar Jain<sup>1</sup> and Panneerselvam Krishnamurthy<sup>3\*</sup>

## OPEN ACCESS

### Edited by:

Ingo Ebersberger,  
Goethe University Frankfurt, Germany

### Reviewed by:

Jurandir Magalhaes,  
Brazilian Agricultural Research  
Corporation (EMBRAPA), Brazil  
Gabriel V. Markov,  
UMR 8227 Laboratoire de Biologie  
Intégrative des Modèles Marins,  
France

### \*Correspondence:

Panneerselvam Krishnamurthy  
pselva7@gmail.com

<sup>†</sup>These authors have contributed  
equally to this work and share first  
authorship

### Specialty section:

This article was submitted to  
Technical Advances in Plant Science,  
a section of the journal  
Frontiers in Plant Science

**Received:** 22 September 2021

**Accepted:** 24 January 2022

**Published:** 15 March 2022

### Citation:

Nimmy MS, Kumar V,  
Suthanthiram B, Subbaraya U,  
Nagar R, Bharadwaj C, Jain PK and  
Krishnamurthy P (2022) A Systematic  
Phylogenomic Classification of the  
Multidrug and Toxic Compound  
Extrusion Transporter Gene Family in  
Plants. *Front. Plant Sci.* 13:774885.  
doi: 10.3389/fpls.2022.774885

<sup>1</sup> ICAR–National Institute for Plant Biotechnology, New Delhi, India, <sup>2</sup> Department of Molecular Biology and Genetic Engineering, Bihar Agricultural University, Bhagalpur, India, <sup>3</sup> Crop Improvement Division, ICAR–National Research Centre for Banana, Tiruchirappalli, India, <sup>4</sup> Division of Genetics, ICAR–Indian Agricultural Research Institute, New Delhi, India

Multidrug and toxic compound extrusion (MATE) transporters comprise a multigene family that mediates multiple functions in plants through the efflux of diverse substrates including organic molecules, specialized metabolites, hormones, and xenobiotics. MATE classification based on genome-wide studies remains ambiguous, likely due to a lack of large-scale phylogenomic studies and/or reference sequence datasets. To resolve this, we established a phylogeny of the plant MATE gene family using a comprehensive kingdom-wide phylogenomic analysis of 74 diverse plant species. We identified more than 4,000 MATEs, which were classified into 14 subgroups based on a systematic bioinformatics pipeline using USEARCH, blast+ and synteny network tools. Our classification was performed using a four-step process, whereby MATEs sharing  $\geq 60\%$  protein sequence identity with a  $\leq 1E-05$  threshold at different sequence lengths (either full-length,  $\geq 60\%$  length, or  $\geq 150$  amino acids) or retaining in the similar synteny blocks were assigned to the same subgroup. In this way, we assigned subgroups to 95.8% of the identified MATEs, which we substantiated using synteny network clustering analysis. The subgroups were clustered under four major phylogenetic groups and named according to their clockwise appearance within each group. We then generated a reference sequence dataset, the usefulness of which was demonstrated in the classification of MATEs in additional species not included in the original analysis. Approximately 74% of the plant MATEs exhibited synteny relationships with angiosperm-wide or lineage-, order/family-, and species-specific conservation. Most subgroups evolved independently, and their distinct evolutionary trends were likely associated with the development of functional novelties or the maintenance of conserved functions. Together with the systematic classification and synteny network profiling analyses, we identified all the major evolutionary events experienced by the

MATE gene family in plants. We believe that our findings and the reference dataset provide a valuable resource to guide future functional studies aiming to explore the key roles of MATEs in different aspects of plant physiology. Our classification framework can also be readily extendable to other (super) families.

**Keywords:** gene family classification, gene family evolution, MATEs, phylogenomics, synteny network, kingdom wide, USEARCH

## INTRODUCTION

A large class of proteins embedded in cell membranes facilitates both the inward and outward passage of cellular solutes based on diffusion or electrochemical gradients. These proteins serve to maintain cell integrity and can be broadly categorized into channels, pumps, transporters, and other carriers (Tang et al., 2020). Multidrug and toxic compound extrusion (MATE) proteins form one of the major multidrug transporter families that intercept the secondary transport of heterogenic cytotoxic compounds using either proton ( $H^+$ ) or sodium ion ( $Na^+$ ) gradients. These proteins are ubiquitous across all three domains of life and are presumed to have evolved from a common ancestral gene *via* duplication (Omote et al., 2006). The number of MATEs per species across the tree of life is small except for plants, in which the MATE transporter gene family flourished, resulting in an astonishing diversity of functional novelties. Plant MATEs have evolved for the efflux of an array of substrates ranging from xenobiotics and primary (e.g., citrate) as well as specialized metabolites (e.g., alkaloids and phenylpropanoids) to phytohormones (e.g., abscisic acid and salicylic acid), which play significant roles in detoxification (Diener et al., 2001; Li et al., 2002), disease resistance (Nawrath et al., 2002; Ishihara et al., 2008), iron homeostasis (Rogers and Guerinot, 2002), and aluminum tolerance (Magalhaes et al., 2007). MATEs have also been implicated in anther dehiscence and pollen development (Thompson et al., 2010), plant architecture regulation (Li et al., 2014; Suzuki et al., 2015), root development (Upadhyay et al., 2020), organ initiation (Burko et al., 2011), and senescence (Jia et al., 2019).

Classification of genes and proteins based on sequence homology and/or phylogeny into families often provide important indications of the evolutionary past of genes and common ancestry (Frech and Chen, 2010). In a broader sense, MATEs have been classified as one of the four families of multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) superfamilies, members of which may descend from primordial proteins most closely related to prokaryotic polysaccharide transporters (Hvorup et al., 2003). With a limited number of related sequences available, the MATEs of all phyla have been classified into 15 subfamilies ( $n = 203$ ; Hvorup et al., 2003) or three large subfamilies (prokaryotic NorM and DinF, and eukaryotic MATEs) encompassing 14 smaller subgroups ( $n = 861$ ; Omote et al., 2006). Within these classifications, plant MATEs are further clustered into two subfamilies (Hvorup et al., 2003) or one subgroup (Omote et al., 2006).

Advances in sequencing techniques and genome assembly pipelines have led to the availability of hundreds of diverse

plant genomes. Consequently, given the significance of MATEs in different aspects of plant growth, development, and defense (Takanashi et al., 2014; Upadhyay et al., 2019), the MATE gene family has been studied extensively in several plant species (Table 1). Based on its phylogenetic topology, the MATE gene family in plants can be classified into several groups and subgroups; however, as such classifications are not systematic, contradictions exist within the literature (Table 1). For instance, Arabidopsis MATEs were assigned to four groups by Wang et al. (2016) compared to five groups by Santos et al. (2017). In other plant species, three to seven phylogenetic groups have been described (Table 1). This confusion is further exaggerated in the case of subgroup assignment. For example, eight (Ia–IIIc) subgroups are assigned to poplar (*Populus trichocarpa*) MATEs (Li et al., 2017) while 10 subgroups (C1–1–C4–3) are assigned to soybean (*Glycine max*) MATEs (Liu et al., 2016). Intriguingly, the naming of groups/subgroups is also inconsistent between published studies (Table 1) and, albeit not without value, existing studies cannot be used to compare the functional and evolutionary relevance of MATEs across plant species. In this context, a systematic classification of the plant MATE gene family is urgently needed.

Many tools have been developed for the automated annotation of protein (super) families (Wu et al., 2003; Frech and Chen, 2010). However, (super) families are routinely further classified into families, subfamilies, and subgroups by manual intervention, and several classifications and nomenclature committees have been established for this purpose. Examples include the committees for aldo-keto reductase (AKR<sup>1</sup>; Hyndman et al., 2003), UDP-glycosyltransferase (UGT<sup>2</sup>; Mackenzie et al., 1997), and cytochrome P450 (CYP450<sup>3</sup>; Nelson, 2009). However, the establishment of such committees for all gene (super) families is not possible, and a systematic user-friendly classification pipeline would be highly beneficial.

Examining the conservation of gene collinearity and synteny provides critical information about the evolutionary history of genomes, gene families, and genes, as well as yielding important insights into the occurrence of ancient polyploidizations, chromosomal rearrangements, gene orthology, and functional novelties (Zhao et al., 2017). Recently, a phylogenomic synteny network and clustering approach was established to uncover and visualize the synteny relationships of target genes across kingdoms of interest (Zhao et al., 2017; Zhao and Schranz, 2017), which outperforms conventional gene family studies in which

<sup>1</sup><https://hosting.med.upenn.edu/akr/>

<sup>2</sup><https://prime.vetmed.wsu.edu/resources/udp-glucuronosyltransferase-homepage>

<sup>3</sup><https://drnelson.uthsc.edu/>

**TABLE 1** | Discrepancy in the classification of plant MATE gene family.

Plant species	CGS*	Subgroup	References†
<i>Arabidopsis thaliana</i>	I–V	/	Wang et al., 2016†
<i>Arabidopsis thaliana</i>	I–V	/	Santos et al., 2017
<i>Brassica rapa</i>	1–4	/	Qiao et al., 2020†
<i>Brassica juncea</i>	1–4	/	Qiao et al., 2020†
<i>Brassica napus</i>	1–4	/	Qiao et al., 2020†
<i>Brassica nigra</i>	1–4	/	Qiao et al., 2020†
<i>Brassica oleracea</i>	1–4	/	Qiao et al., 2020†
<i>Capsicum annuum</i>	I–V	/	Chen et al., 2020
<i>Cicer arietinum</i>	G1–G4	/	Zhang et al., 2020
<i>Citrus sinensis</i>	I–V	/	Juliao et al., 2020
<i>Glycine max</i>	C1–C4	C1-1, C1-2, C1-3, C2-1, C2-2, C3-1, C3-2, C4-1, C4-2, C4-3	Liu et al., 2016
<i>Gossypium arboreum</i>	MI–MIII	/	Lu et al., 2018
<i>Gossypium hirsutum</i>	C1–C4	C1-1, C1-2, C1-3, C1-4, C1-5, C2, C3, C4	Xu et al., 2019
<i>Gossypium raimondii</i>	MI–MIII	/	Lu et al., 2018
<i>Linum usitatissimum</i>	I–V	/	Ali et al., 2021†
<i>Malus × domestica</i>	I–V	/	Zhang et al., 2021†
<i>Medicago sativa</i>	C I–C IV	C I-1, C I-2, C I-3, C II-1, C II-2, C III-1, C III-2, C III-3, C IV-1, C IV-2, C IV-3	Min et al., 2019
<i>Medicago truncatula</i>	I–V	/	Wang et al., 2017
<i>Nicotiana tabacum</i>	I–V	Ia, Ib, Ic, Id, Ie, IIa, IIb, IIc, III, IVa, IVb	Gani et al., 2021
<i>Oryza sativa</i>	1–4	/	Du Z. et al., 2021
<i>Oryza sativa</i>	C1–C4	C1-1, C1-2, C2, C3-1, C3-2, C4-1, C4-2, C4-3, C4-4	Huang et al., 2019
<i>Oryza sativa</i>	I–V	/	Wang et al., 2016†
<i>Populus trichocarpa</i>	I–III	Ia, Ib, IIa, IIb, IIc, IIIa, IIIb, IIIc	Li et al., 2017
<i>Solanum lycopersicum</i>	I–V	/	Santos et al., 2017
<i>Solanum tuberosum</i>	I–V	/	Huang et al., 2021†
<i>Solanum tuberosum</i>	I–VI	/	Li et al., 2019
<i>Solanum tuberosum</i>	I–V	/	Chen et al., 2020
<i>Vaccinium corymbosum</i>	I–V	/	Chen et al., 2015
<i>Zea mays</i>	I–VII	/	Zhu et al., 2016

\*CGS denotes *clade, clusters, class, group, and/or subfamily*.

†Classification of MATEs in these studies is similar. In each other studies, classification of MATEs is performed by independent phylogenetic observation.

one or a limited number of species of interest are typically the focus. Synteny network and clustering analyses on a broader scale have also revealed the significance of deep and/or lineage-specific conservation, ancestral translocations, and ancient duplications in the development of functional and phenotypic novelties of

certain gene superfamilies, including APETALA2 (Kerstens et al., 2020), auxin response factors (Gao et al., 2020), and type III polyketide synthases (Naake et al., 2021). Such analyses are necessary to unravel the evolution of plant MATE gene families in greater detail.

Here, we report a kingdom-wide analysis of the MATE gene family and synteny network analyses using 74 species from different plant orders. We identify a total of 4,211 MATEs using a hidden Markov model (HMM) profile-based search and classify these into 14 subgroups based on sequence homology and synteny relationships using a four-step process. Following confirmation based on synteny network analyses, we establish a reference sequence dataset for the classification of the MATE gene family in other species. Together with the synteny network clustering analyses, our systematic classification framework reveals several ancient gene duplication and translocation events in the evolutionary history of plant MATE subgroups. We believe that our classification strategy could be effectively applied to other genes (super) families, and our findings provide a valuable resource for ongoing plant MATE research.

## MATERIALS AND METHODS

### Identification and Characterization of Multidrug and Toxic Compound Extrusion Transporters in 74 Plant Species

We used the genomes of 74 diverse plant species including 39 eudicots, 20 monocots, eight green algae, three basal angiosperms, and a single representative of the following groups: red algae, liverworts, mosses, and lycophytes (**Supplementary Table 1**). The proteomes of three cucurbit and two banana species were, respectively, downloaded from the Cucurbit Genomics (Zheng et al., 2019)<sup>4</sup> and Banana Genome Hub (Droc et al., 2013)<sup>5</sup> databases, and the proteomes of the 69 other species were downloaded from the Phytozome version 13<sup>6</sup> database (Goodstein et al., 2012). All downloaded proteomes comprised only the primary transcripts. The MatE.hmm file (Pfam No. PF01554.19) corresponding to the HMM of the MATE gene family was downloaded from the Pfam database (version 33; El-Gebali et al., 2019).<sup>7</sup> The MatE-domain-containing proteins of the 74 plant species were retrieved through “hmmsearch” within the HMMER3 software platform (Mistry et al., 2013), in which the MatE.hmm profile was searched with the gathering threshold (--cut\_ga) option against each downloaded proteome. The resulting output files were concatenated, and the amino acid sequences of all hits were interrogated using stand-alone InterProScan (version 5.48-83) software (Jones et al., 2014) to determine their Pfam (Pfam version 33.1) domain composition. The domains were visualized and extracted using the QKdomain

<sup>4</sup><http://cucurbitgenomics.org/>

<sup>5</sup><http://banana-genome-hub.southgreen.fr/>

<sup>6</sup><https://phytozome-next.jgi.doe.gov/>

<sup>7</sup><http://pfam.xfam.org/>

pipeline (Bailey et al., 2018).<sup>8</sup> Intron abundance was calculated using the Genestats pipeline by exploiting the GFF3 files of the corresponding species (Card, 2021).<sup>9</sup> Protein and domain lengths were inferred using the Seqtk tool (Shen et al., 2016). The presence of signal peptides and transmembrane helices was verified using the SignalP-5 (Armenteros et al., 2019)<sup>10</sup> and TMHMM version 2 (Krogh et al., 2001)<sup>11</sup> servers.

## Multiple Sequence Alignment and Phylogenetic Analyses

The full-length amino acid sequences of selected plant MATEs were aligned using MUSCLE version 3.8.1551 with default options (Edgar, 2004). The gappiest positions in the resulting alignment were removed with trimAl version 1.4 using the -gappypout option (Capella-Gutierrez et al., 2009). The trimmed alignment was then used for phylogenetic reconstruction analysis in MEGA X (Kumar et al., 2018), whereby an unrooted neighbor-joining (NJ) tree was generated with 1,000 bootstrap replications by applying the Poisson model, uniform rates, and pairwise deletion options. Concurrently, a maximum-likelihood (ML) tree was generated using the same trimmed alignment in RAxML version 8.2.12 (Stamatakis, 2014) with -m PROTGAMMAAUTO (this option determined JTTDCMUT likelihood as the best protein substitution model for the input alignment and used it for the tree search), -f a (this option would start rapid bootstrap analysis and search for best-scoring ML tree, simultaneously), and 100 bootstrap replications. The resulting phylogenies were visualized and annotated using iTOL v6 (Letunic and Bork, 2021).<sup>12</sup>

## Classification of Plant Multidrug and Toxic Compound Extrusion Transporters

Plant MATEs were classified based on homology and synteny relationships using a four-step process. First, a distance matrix was generated for 3,523 MATEs (which showed comparable protein length and domain architecture to the characterized plant MATEs) based on their protein sequence identity (PSI) using the -calc\_distmx command. The resulting distance matrix was utilized for agglomerative clustering using the -cluster\_aggd command. Protein clusters sharing  $\geq 60\%$  PSI were then retrieved using the minimum linkage (-linkage min) option. These three commands were all executed in the USEARCH version 11 program with default parameters<sup>13</sup> (Edgar, 2010). Second, a protein database was created with 3,369 MATEs grouped into 14 USEARCH clusters, each with  $\geq 50$  genes, using ncbi-blast+ version 2.6.0-1 (Camacho et al., 2009), against which all unclassified plant MATEs ( $n = 847$ ) were blasted using the -max\_target\_seqs 5 and -outfmt 6 options. Subgroups were then assigned to each query based on two conditions: (i)  $\geq 60\%$  identity with an e-value threshold  $\leq 1E-05$  relative to the subject; and (ii) an alignment

length covering  $\geq 60\%$  of the query's protein length or  $\geq 150$  amino acids. Third, if an unclassified query showed a synteny relationship with more than three plant MATEs in the same subgroup, the subgroup of those classified MATEs was assigned to the query. Finally, a new protein database was created based on the MATEs classified during the first three steps against which all remaining, unclassified MATEs were blasted, and subgroups were assigned to the queried MATEs that fulfilled the conditions outlined under step two.

## Genomic Synteny and Network Construction Analyses

To determine the synteny relationships between MATEs across the plant kingdom, genomic synteny, and network construction analyses were conducted for the proteomes of the selected 74 plant species as described in the synteny network (SynNet) pipeline (Zhao et al., 2017).<sup>14</sup> This was performed using Diamond version 2.0.6 (Buchfink et al., 2015) and MScanX (Wang et al., 2012). Briefly, Diamond was employed with the "number of hits" option (i.e., -k) set to five for all inter- and intra-pairwise all-vs.-all whole-genome comparisons, in which the proteomes of 74 plant species were blasted with themselves and one another reciprocally (in total 5,476 blasts were made). The resulting blast outputs were then utilized by MScanX to compute synteny blocks with default parameters (i.e., number of minimum match size for a collinear block = 5 genes; the number of maximum gaps allowed = 25 genes). Followed by, synteny blocks consisting of plant MATEs were extracted from the master synteny file and visualized in Cytoscape version 3.7 (Shannon et al., 2003). Prior to visualization, the MATE synteny file was curated, whereby certain syntenic pairs, having non-MATEs (i.e., a gene in the syntenic pair was not detected during the HMM search), were excluded. Clustering analyses of the synteny networks were performed in the R statistical environment (version 4.0.2), using the "igraph" (Csardi and Nepusz, 2006), "pheatmap" (Kolde, 2015), and "vegan" (Dixon, 2003) packages.

## RESULTS

### Systematic Identification, Classification, and Phylogeny of Plant Multidrug and Toxic Compound Extrusion Transporters

Phylogenomic profiling was performed based on the HMM profile of the MATE gene family (PF01554.19) to explore the complete repertoire of MATEs in the plant kingdom (Supplementary Table 1). An Hmmssearch was conducted separately for each of the 74 plant species to avoid the influence of proteome size/e-value threshold in the MATE detection. A total of 4,217 hits were detected (Supplementary Table 2) with 54 Pfam domains (Supplementary Table 3) and 55 Pfam domain compositions (Supplementary Table 4). While most hmmssearch hits showed two (81.2%) or one (15.9%) MATE domains, few showed non-specific domains in addition to 0–4 MATE domains (1.6%) or multiple MATE domains (i.e., more

<sup>8</sup><https://github.com/matthewmoscou/QKdomain>

<sup>9</sup><https://github.com/darencard/GenomeAnnotation/blob/master/genestats>

<sup>10</sup><http://www.cbs.dtu.dk/services/SignalP/>

<sup>11</sup><https://services.healthtech.dtu.dk/service.php?TMHMM-2.0>

<sup>12</sup><https://itol.embl.de/>

<sup>13</sup><https://drive5.com/usearch/>

<sup>14</sup><https://github.com/zaotao1987/SynNet-Pipeline/wiki>

than two; 0.7%; **Supplementary Table 4**). We presumed that the existence of non-specific or multiple MatE domains in a single hmmsearch hit resulted from an annotation and/or genome assembly pipeline error. InterProScan did not detect the MatE domain in 0.5% of the hmmsearch hits (**Supplementary Table 4**), implying that the MatE profile signal was too low in those cases. Gathering thresholds are family-specific bit score thresholds assigned manually by Pfam curators at the time a family is built to exclude any false positive matches (Punta et al., 2012). They are generally considered the reliable curated thresholds defining family membership. Because the hmmsearch was conducted using a gathering threshold option in this study, all of the identified hits were selected to represent the MATE family of plants and were designated as plant MATEs accordingly (**Figure 1**).

Plant MATEs were systematically classified in four subsequent steps using a set of bioinformatics tools by applying a  $\geq 60\%$  PSI threshold (**Figure 2**). The length of all characterized plant MATEs ranged from 370 to 615 amino acids and consisted of one or two MatE domains (**Supplementary Table 5**). Concurrently, the length of a single MatE domain corresponds to 161 amino acids (PF01554.19). Based on this, and to achieve a better sequence alignment as well as reliable phylogenies, only 3,524 MATEs were selected for the USEARCH cluster analyses, with full-length and MatE domain lengths ranging between 350 and 650, and between 140 and 340 amino acids, respectively (**Supplementary Table 6**). As the subfamilies of several gene superfamilies are assigned based on the  $\geq 60\%$  PSI (for example, the AKR gene superfamily; Hyndman et al., 2003), USEARCH was employed to cluster MATEs with  $\geq 60\%$  PSI based on the PSI distance matrix. Two linkage options are available in USEARCH for clustering. First, in the maximum linkage option, members in each cluster share  $\geq 60\%$  PSI with every member of the same cluster, while second, in the minimum linkage option, members in each cluster share  $\geq 60\%$  PSI with any of one member in the same cluster. Because the maximum linkage option generates many small clusters (data not shown), we applied the minimum linkage option, which identified 121 clusters with 1–718 members (**Supplementary Table 7**). Fourteen clusters (each comprising  $\geq 50$  members) were selected to represent the subgroups of the plant MATE gene family that covered 3,369 (out of 3,524) MATEs (i.e., USEARCH-classified MATEs).

To unravel the evolutionary relationships of the MATE subgroups and to assign logical subgroup names, an unrooted NJ-derived phylogenetic tree was reconstructed using the USEARCH-classified MATEs. Four major phylogenetic groups were identified and named according to Wang et al. (2016), and subgroups were designated based on their clockwise appearance within each phylogenetic group using capital letters (**Figure 3**). Groups I, II, III, and IV comprised two, five, four, and three subgroups, respectively. The tree topology of the NJ-derived phylogeny was similar to that of the ML-derived phylogeny (**Supplementary Figure 1**), suggesting that the clustering of subgroups under different methods is reliable.

Next, a BLASTP search was performed to determine the subgroup of unclassified (UnC) MATEs (i.e., the MATEs not used in the phylogeny reconstruction), in which the UnC MATEs were queried against the database of USEARCH-classified

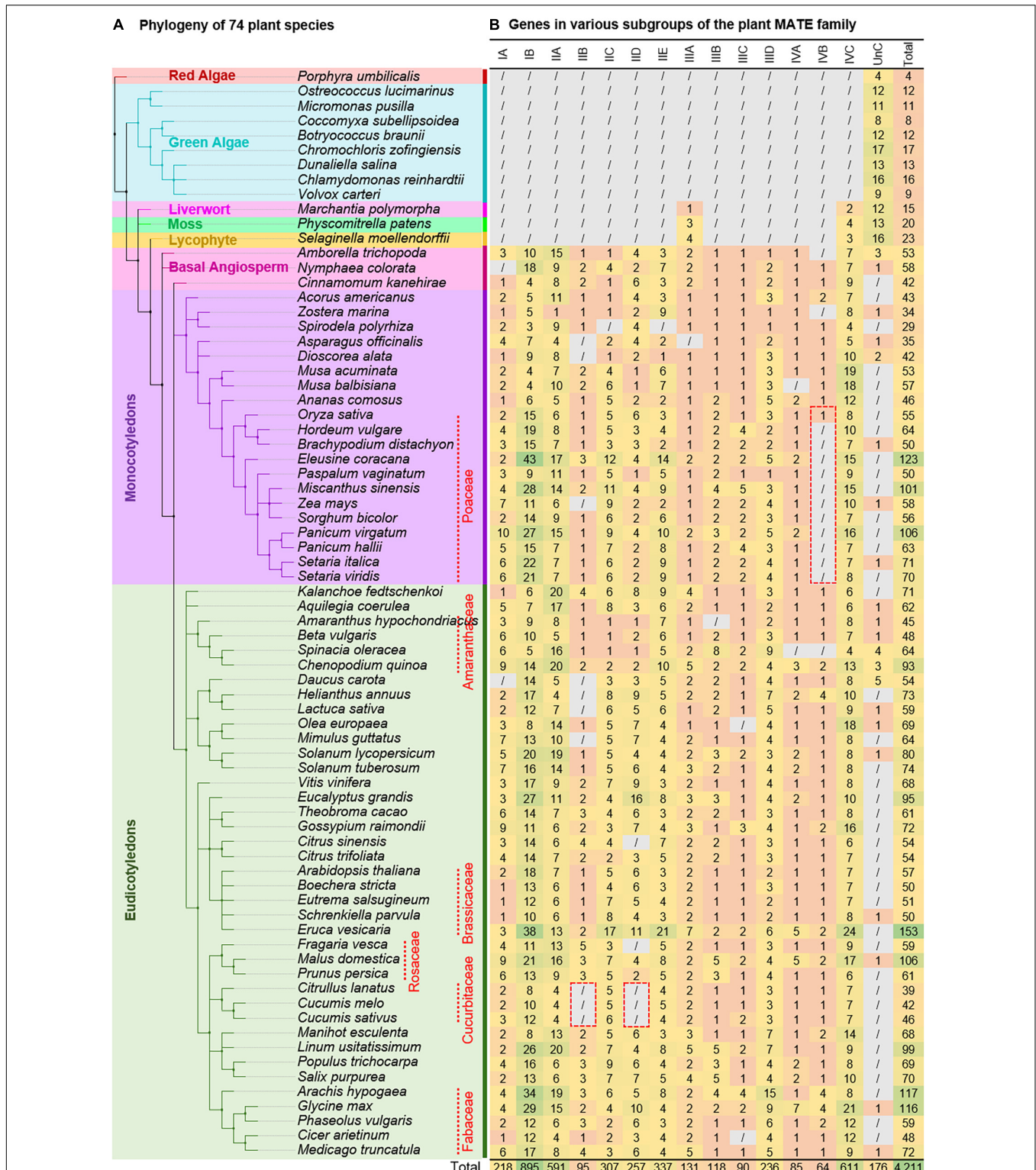
MATEs. A subgroup was assigned to 615 MATEs (termed as BLASTP-classified MATEs), which showed  $\geq 60\%$  PSI in  $\geq 60\%$  of its protein length or  $\geq 150$  amino acids with an e-value threshold  $\leq 1E-05$  to the classified MATEs. Subgroup information of USEARCH- and BLASTP-classified MATEs was incorporated into the synteny file. A subgroup of 40 MATEs (termed SYNTENY-classified MATEs) was determined that locate in similar synteny blocks to  $\geq 3$  MATEs of the same subgroup. A second blastp search was conducted using the classified and UnC MATEs, which identified a subgroup of 11 MATEs based on the stated blastp criteria. The same blastp output was utilized then with a relaxed condition of  $\geq 40\%$  PSI in  $\geq 40\%$  of its protein length or  $\geq 100$  amino acids with an e-value threshold  $\leq 1E-05$  to explore the group of still UnC MATEs. This identified the group of 73 UnC MATEs. Ultimately, six hmmsearch hits were identified as non-MATEs, which did not contain the MatE domain and to which a group/subgroup was not assigned based on any of the criteria described. Excluding these six genes, the number of plant MATEs identified was revised to 4,211 (**Figure 1**). Overall, the group and subgroup were determined to be 97.5% (4,108 in 4,211) and 95.8% (4,035 in 4,211) of the plant MATEs identified in this study. Intriguingly, a subgroup of 176 hmmsearch hits having MatE domain(s)—the majority of which belong to the algae lineage—was not determined based on our criteria and were tentatively designated as UnC MATEs (**Figure 1**). The complete attributes of all the identified MATEs are listed in **Supplementary Table 8**.

To assist the classification of the MATE gene family in other plant species, a reference sequence dataset (termed “RefPlantMATEs”) was prepared with the 3,444 classified MATEs (**Supplementary Text 1**). Members in the dataset comprised 1–2 MATE domains, and their protein and domain lengths ranged between 351 and 650, and between 79 and 342 amino acids, respectively (**Supplementary Table 8**). Using this dataset, the MATE gene family of wheat (*Triticum aestivum*; designated as TaMATEs) was systematically classified. Briefly, the 226 TaMATEs identified through hmmsearch, as previously described, were blasted against the RefPlantMATEs protein database, and a subgroup of 224 was successfully determined based on the blastp criteria (**Supplementary Table 9**). TaMATEs were clustered together with respect to their subgroups in the NJ-derived phylogenetic tree (**Supplementary Figure 2**), showing the direct benefits of RefPlantMATEs in MATE gene family classification in other plants species. A repository was created in GitHub,<sup>15</sup> where the step-wise classification of MATEs for additional species is described.

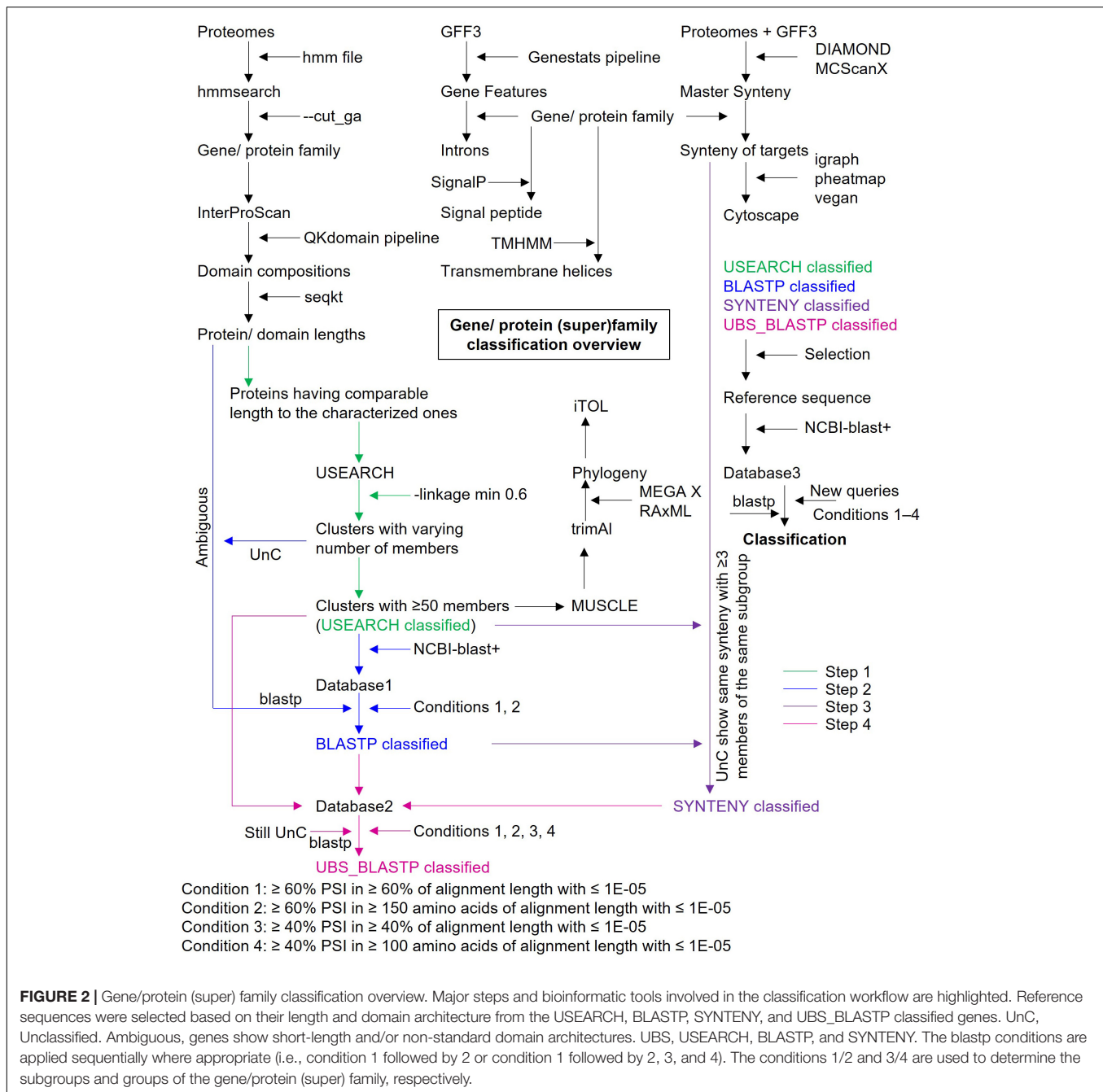
## Genomic Diversification of Plant Multidrug and Toxic Compound Extrusion Transporters

To better understand the evolutionary relationship of plant MATEs, we constructed a synteny network (**Figure 4**) using the established SynNet pipeline (Zhao et al., 2017; Zhao and Schranz, 2017). Synteny analysis revealed that 74% of

<sup>15</sup><https://github.com/pselva7/PlantMATEsClassification>



**FIGURE 1 |** Distribution of multidrug and toxic compound extrusion (MATE) transporters across the plant kingdom. The phylogeny of 74 plant species (A) was inferred using the NCBI Taxonomy Browser (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>). The number of MATEs in different subgroups of each plant genome is shown next to the species names (B). Numbers are colored based on abundance. “/” denotes not detected. Subgroups are assigned based on ≥ 60% protein sequence identity threshold using a set of bioinformatic tools (see section “Systematic Identification, Classification, and Phylogeny of Plant Multidrug and Toxic Compound Extrusion Transporters” for details).

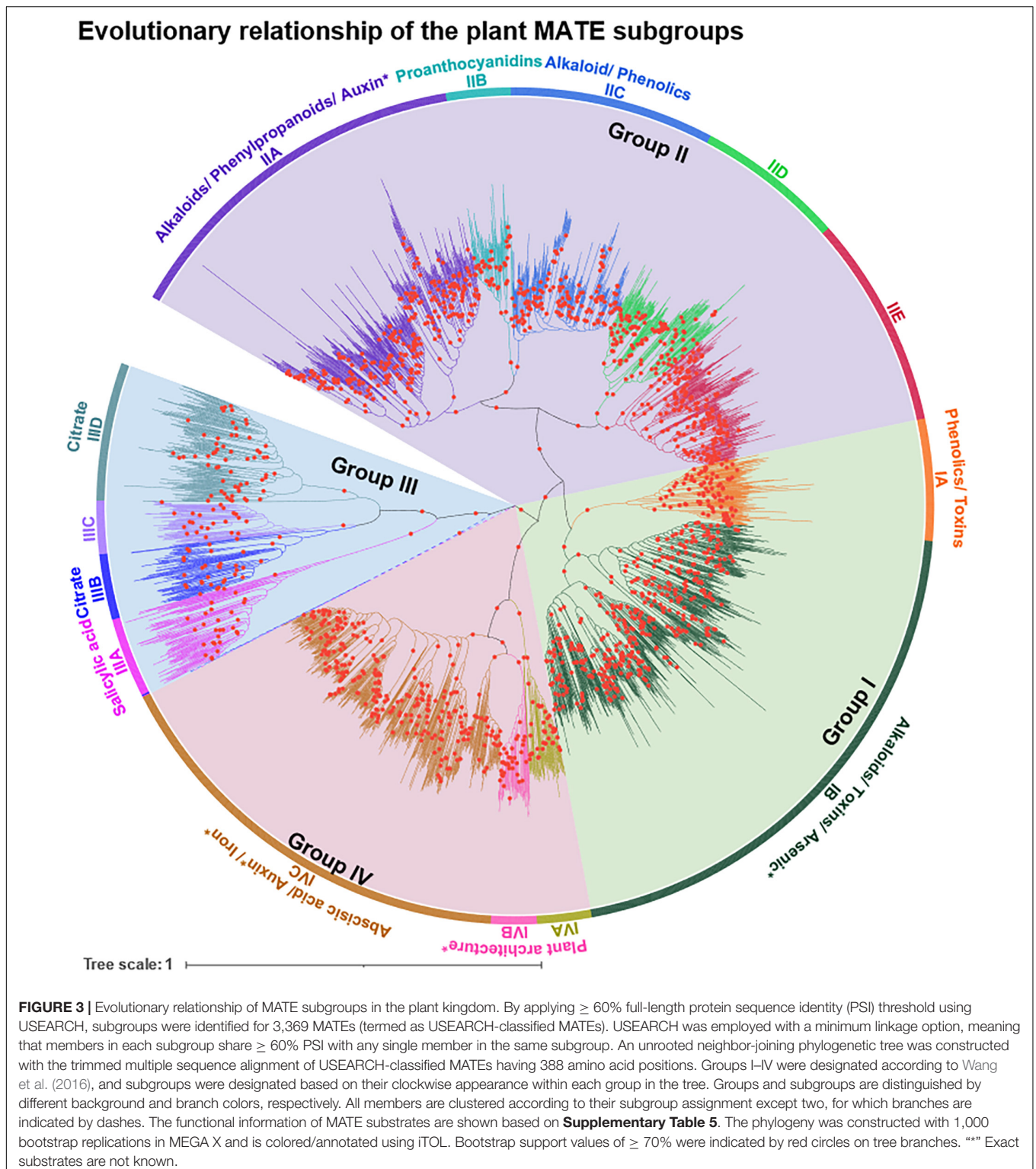


**FIGURE 2 |** Gene/protein (super) family classification overview. Major steps and bioinformatic tools involved in the classification workflow are highlighted. Reference sequences were selected based on their length and domain architecture from the USEARCH, BLASTP, SYNTENY, and UBS\_BLASTP classified genes. UnC, Unclassified. Ambiguous, genes show short-length and/or non-standard domain architectures. UBS, USEARCH, BLASTP, and SYNTENY. The blastp conditions are applied sequentially where appropriate (i.e., condition 1 followed by 2 or condition 1 followed by 2, 3, and 4). The conditions 1/2 and 3/4 are used to determine the subgroups and groups of the gene/protein (super) family, respectively.

the MATEs (3,124 out of 4,211) were located in synteny blocks across and/or within the genomes of the selected 74 plant species (Supplementary Table 10). Most species ( $n = 55$ ) retained  $\geq 60\%$  of the MATEs in the synteny blocks (Supplementary Table 11). Synteny conservation was usually greater within the members of the same plant orders/families, likely due to their close evolutionary relationships. For example, 70–90% of the MATEs identified in Amaranthaceae, Brassicaceae, Cucurbitaceae, Fabaceae, and Poaceae species were found in synteny blocks. Intriguingly, certain species, such as barley (*Hordeum vulgare*) and spinach (*Spinacia oleracea*), showed a

lower percentage of synteny MATEs compared to other closely related species (Supplementary Table 11), implying that the genome assembly of these species is relatively fragmented.

The synteny MATEs of the 74 plant species (Supplementary Table 10) formed  $> 50,000$  edges (i.e., the number of pairwise connections between MATEs; Supplementary Table 12). Almost all of these connections were intra-group or intra-subgroup types, while only 0.007% and 0.629% were inter-group and inter-subgroup types, respectively (Figure 4 and Supplementary Table 12). On account of connection preponderance, the synteny network was subdivided into 202 synteny communities

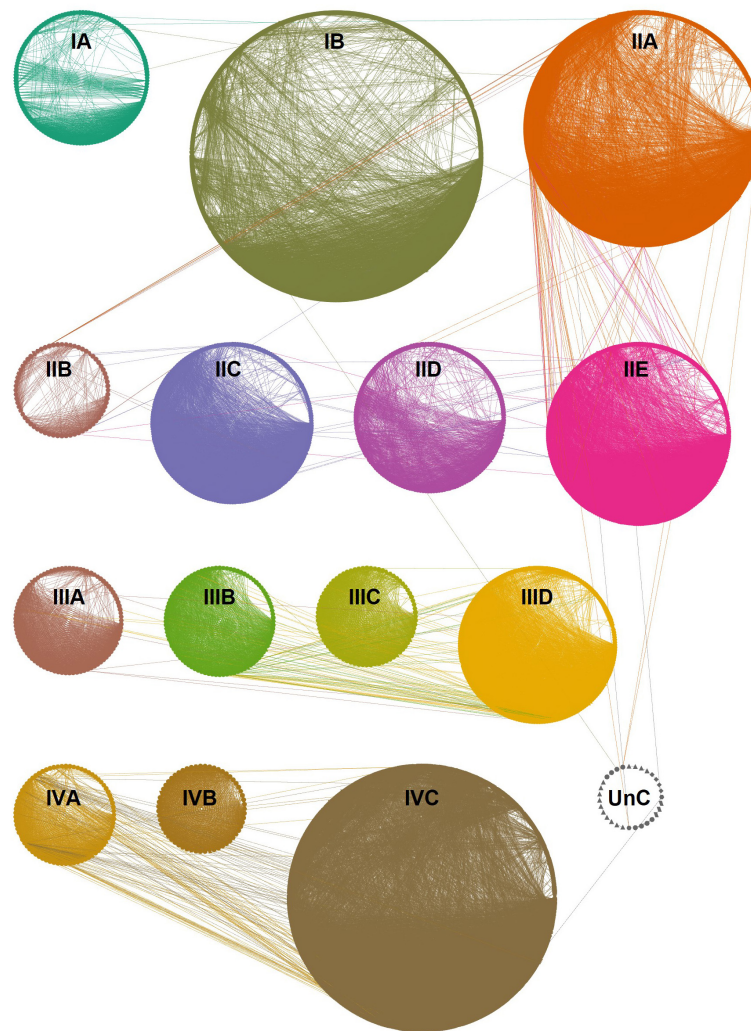


by igraph, the majority of which retained members of the same subgroups, while 21 communities retained members from different subgroups (**Supplementary Table 13**). The number of MATEs in each community varied widely between 2 to 213. Three subgroups comprised most communities (i.e., IB–62,

IIA–32, and IVC–21), while the other subgroups comprised 3–14 communities (**Supplementary Table 14**). The percentage of synteny MATEs was relatively low in subgroups IA, IIB, IIC, IID, and IIIA (60–69%) compared to other subgroups, in which 70–93% of MATEs were retained in synteny blocks.



### Synteny relationship between/ within subgroups of plant MATE gene family



**FIGURE 4 |** Global synteny network (SynNet) of the plant multidrug and toxic compound extrusion (MATE) transporter gene family. The network was established with all the SynNet relations of plant MATEs within and between genomes that included 3,124 nodes having 55,276 edges and 202 communities. The network was visualized in Cytoscape and the nodes are grouped based on their subgroup. Lines connecting the nodes denote pairwise synteny relationships. Subgroups are highlighted in different colors.

A phylogenetic profile was established using the synteny communities of the MATE gene family (Figure 5), in which 13 types of synteny communities were detected based on their species composition (Supplementary Table 15). Four types, namely AFO-specific (members representing two different species of the same order/family of the algae lineage), MoS-specific (members representing a single species of moss), LS-specific (members representing a single species of lycophytes), and basal-specific (members representing two different species of basal angiosperms), corresponded a smaller number of communities and genes (Supplementary Table 16), mainly because of the underrepresentation of those species in our dataset. While three communities of 29 genes were identified as BM-specific (members representing

at least one monocot and one basal angiosperm species), nine communities of 95 genes were identified as BE-specific (members representing at least one eudicot and one basal angiosperm species). These BE- and BM-specific communities suggest the occurrence of lineage-specific ancient translocation or genomic rearrangement events during the divergence of basal-eudicot-monocot lineages. Forty-four communities were categorized as angiosperm-wide, with members representing at least one eudicot and one monocot species, and covered 74% of synteny MATEs, showing that most MATEs are located in conserved synteny blocks across angiosperms. More than 51% of the communities (104 out of 202) covering 19% of the synteny MATEs were detected as lineage-specific, with members representing either different orders/families of eudicots

(eudicot-wide) and monocots (monocot-wide) or two different species of the same order/family of eudicots (EFO-specific) and monocots (MFO-specific; **Supplementary Table 16**). Lineage-specific communities suggest that ancient translocation or genomic rearrangement events may have occurred during the divergence of the eudicot-monocot lineages. Species-specific communities (i.e., ES- and MS-specific) corresponded to 15% of the total communities and 2% of the total genes that are usually paralogs generated from local duplication events (e.g., segmental duplication) within a single species.

## DISCUSSION

### Classification Pipeline

Classification of a gene family provides insights into the genetic structure, biological function, and evolutionary trends that are often essential in comparative genomics when tracing the origins and divergence of structural/functional features. Given the key roles of MATEs in different aspects of plant physiology (Takanashi et al., 2014; Upadhyay et al., 2019), the MATE gene family has been studied extensively in several plant species and has been classified into various subgroups (**Table 1**). As all existing studies have examined MATEs in only a few species and applied classifications based on phylogeny, ambiguity still exists (**Table 1**), largely due to a lack of reference sequences. To address this issue, we conducted a large-scale phylogenomic study to explore the complete repertoire of plant MATEs, which were then systematically classified using a bioinformatics pipeline involving USEARCH, blast+, and SynNet tools (**Figure 2**).

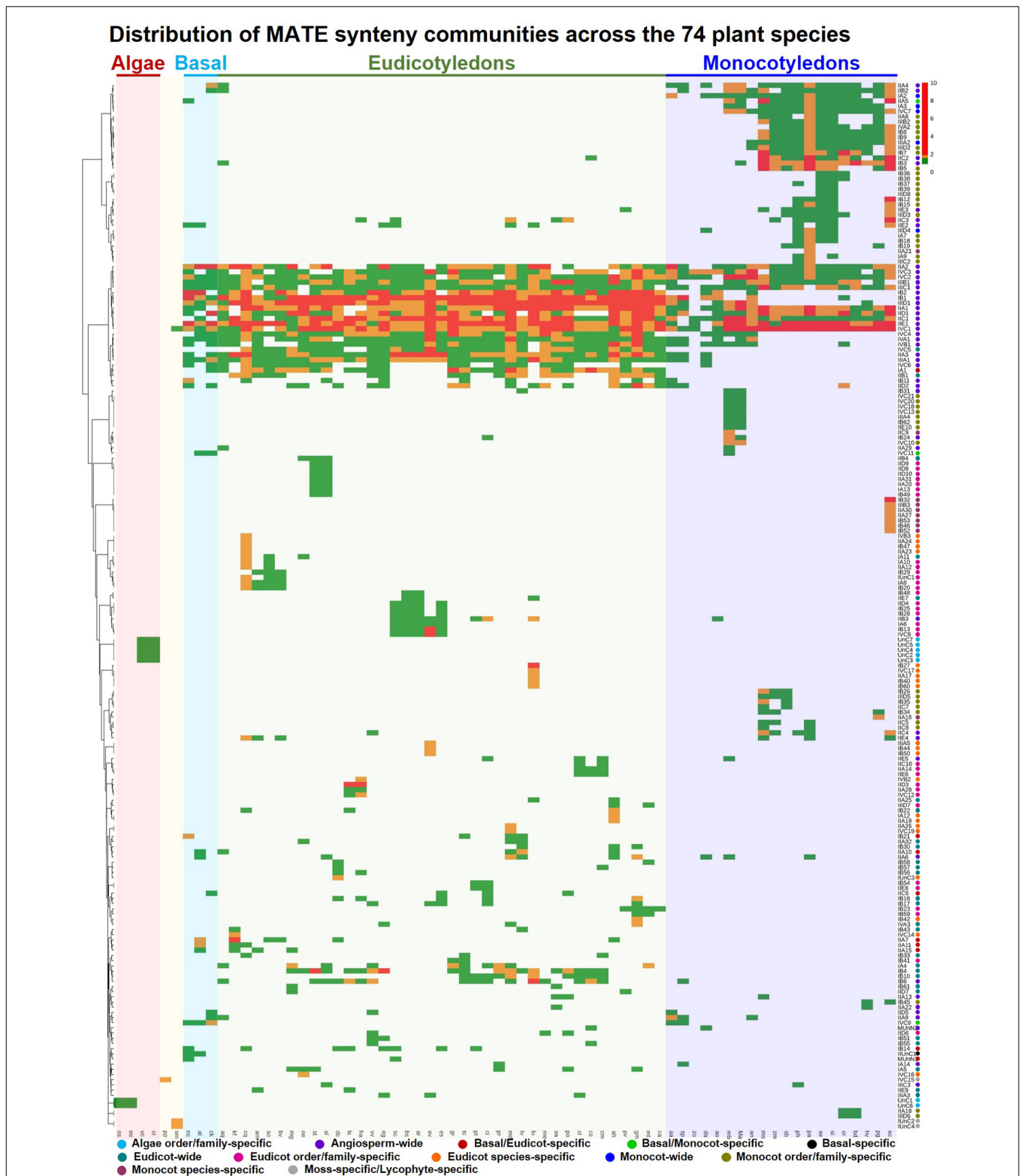
Generally, homologous proteins are grouped into families and subfamilies based on their sequence similarity. A threshold of  $\geq 40\%$  similarity for families and  $\geq 60\%$  similarity for subfamilies are widely implemented for gene superfamilies, such as the AKR (Hyndman et al., 2003), UGT (Paquette et al., 2003), and organic anion transporting polypeptide (Hagenbuch and Meier, 2003) gene superfamilies. Once the plant MATE repertoire was established by hmm profile searches, USEARCH was employed to determine the subgroups of plant MATEs based on a  $\geq 60\%$  PSI threshold with the minimum linkage option, and caution was taken during both of these steps. Hmsearch should be performed for individual species because the size of the proteome influences hit detection. For example, certain hits with low profile thresholds were not detected when the hmsearch was performed against the master proteome file (in which proteomes of all species were concatenated), but were otherwise detected against the individual proteomes. Likewise, input sequences for clustering in USEARCH are highly valued, and sequences of short-length or non-standard proteins should be avoided (Step 1). For instance, if the input dataset comprises fragment proteins (assuming 50–100 amino acids, which would be very low compared to most members of a target gene family) having considerable similarity to different subgroup members, two or more subgroups would be grouped into one cluster. Another important criterion was the selection of clusters to represent the subgroup/subfamily of a target gene family. Only clusters with a considerable number of members and

diverse species representations should be selected. For example, a small group of 13 Brassicaceae MATEs sharing  $\geq 60\%$  PSI with one another but not with any of the other members led to their designation as a subgroup (cluster no. 1 in **Supplementary Table 7**). These, however, exhibited synteny relationships with numerous members of the same existing subgroup (**Supplementary Table 12**), implying that these 13 MATEs belong to an existing subgroup. Based on this, we selected clusters with  $\geq 50$  members as MATE subgroups.

To assign logical names to the selected subgroups based on their evolutionary relationships, a family phylogeny was established using NJ and ML methods. The MATE subgroups were clustered under four major phylogenetic clades, which were designated as groups and named according to Wang et al. (2016) (**Figure 3**). The NJ tree was favored over the ML tree for naming purposes because the former was generated more quickly (1,000 bootstrap replications of 3,369 sequences took  $\sim 10$  h with 12 threads) compared to the latter (100 bootstrap replications of 3,369 sequences took  $\sim 10$  d with 24 threads). The resulting NJ and ML trees were compared for consistency based on phylogeny topology, and were confirmed to be the same (**Supplementary Figure 1**). After the USEARCH cluster analysis, a large fraction of the hmsearch hits were not classified and showed short and/or non-standard domain architectures (collectively termed “ambiguous proteins”). Ambiguous proteins may represent the incompleteness of the genome assembly and/or annotation pipeline, and may obtain their correct sequence structure in forthcoming genome assemblies of the respective species. However, these still require classification to uncover the emergence, expansion, and contraction of the gene family without bias.

Ambiguous proteins were, therefore, blasted against the USEARCH-classified MATE database with an option to output the top five hits (Step 2). Two criteria were implemented in each blastp search to classify ambiguous proteins. The first (i.e.,  $\geq 60\%$  PSI in  $\geq 60\%$  of protein lengths with a  $1E-05$  threshold) was mainly defined to classify short-length proteins. Because several hits showed multiple MATE domains or non-standard domains, and their amino acid lengths hindered them to satisfy the first criteria, the second criteria were defined (i.e.,  $\geq 60\%$  PSI in  $\geq 150$  amino acids of proteins with a  $1E-05$  threshold). Blastp searches determined subgroups for a large fraction of ambiguous proteins; however, several hits remained unclassified. As gene synteny is evolutionarily more stable than gene sequences (as is the case for the 13 Brassicaceae MATEs previously described), we applied the MATE synteny relationship to the unclassified hits (Step 3). Finally, we generated a further database against which the remaining unclassified MATEs were blasted to determine their subgroup and/or group (Step 4).

All of these steps need to be performed only once for each gene family and, thereafter, a blastp search is sufficient to classify the same gene family in additional species. The fasta header in the reference sequence dataset of the MATE gene family was prepared with five attributes (gene id, species, subgroup, protein length, and classification criteria) separated by pipe (`|`), which facilitates gene family classification in other species.



**FIGURE 5 |** Phylogenetic profile of MATE SynNet communities across 74 plant species. Columns and rows represent plant species and SynNet communities, respectively. The abundance of MATE genes in each community of each species is color-scaled using heatmap in the R environment. Synteny communities were categorized into 13 types based on their distribution among the 74 plant species. The differently colored circles on the right denote the category assignment (Supplementary Table 15).

The usefulness of the reference sequence dataset has already been demonstrated in classifying the MATE gene family of wheat (**Supplementary Table 9**), and an automated plant MATE classification pipeline is described in the GitHub repository (see text footnote 15). Thus, we believe our approach can be successfully applied to other poorly classified and/or unclassified gene (super) families in the plant kingdom.

## Synteny Network Substantiates the Classification Pipeline

Using the proposed classification pipeline (**Figure 2**), subgroups were determined for 95.8% (4,035 out of 4,211) of identified plant MATEs (**Supplementary Table 8**). More than 81% of the unclassified (UnC) MATEs (143 out of 176) belonged to algae and early land plants, highlighting their ancient origins. A similar phenomenon was observed during the classification of CYP450s, in which algae are not amalgamated with higher-order plants (Nelson and Werck-Reichhart, 2011). Thirty-three angiosperm MATEs with MatE signatures did not fulfill any of the classification criteria and, therefore, were designated as UnC. We assume that these proteins correspond to ambiguous sequences resulting from genome assembly/annotation pipeline errors, although may yield classifiable sequences in forthcoming assemblies.

Our subgroup assignment was verified by a synteny relationship given that genes sharing a common ancestry tend to be retained in a similar synteny block across and/or within genomes. Approximately 74% of the plant MATEs (3,124 out of 4,211) were identified as syntelogs (i.e., homologs retained in synteny), meaning that the remaining genes may have lost synteny owing to polyploidization and diploidization, gene translocation, genomic rearrangement, or redundancy in genome assembly/annotation. Alternatively, this may reflect the inability of the SynNet pipeline to detect certain tandem duplicates. This is highlighted using the chickpea (*Cicer arietinum*) MATE gene family (designated as CarMATEs; **Supplementary Table 17**), in which certain segmental/tandem duplicates with higher PSI were identified as non-syntelogs (i.e., homologs not retained in synteny; **Supplementary Table 18**).

Plant synteny MATEs formed 55,276 pairwise connections, most of which (99.37%) existed within each subgroup (**Supplementary Table 12**). This supports the MATE gene family phylogeny established through our proposed classification pipeline. For example, if the MATE phylogeny is considered alone (**Figure 3**) and sequence homology, as well as synteny, are disregarded, IIB and IIC may be assigned into one subgroup as both show only a marginal distance in the tree. However, both of these subgroups share < 60% PSI, and only 2% of their connections (out of ~470) were identified as inter-subgroup types (**Supplementary Table 12**). This implies that our designation of IIB and IIC as separate subgroups is valid. A similar phenomenon is observed in the case of other subgroups, such as IIIB↔IIIC (**Figure 3**). Overall, only 348 connections were identified as inter-subgroups (**Supplementary Table 12**), of which few (1–5%) occurred between different subgroups and the majority occurred between IIA↔IIE (20%), IIIB↔IIID (19%), and IVA↔IVC subgroups (35%; **Supplementary Table 19**). As inter-group and inter-subgroup connections provide evidence

for ancient gene duplications followed by lineage-specific gene losses (Gao et al., 2020), the IIA and IIE subgroups may descend from a recent common ancestor. The same logic can be extended to the IIIB/IIID and IVA/IVC subgroups.

## Evolution of the Multidrug and Toxic Compound Extrusion Transporter Gene Family in Plants

All the analyzed plant species contained MATEs, and the total MATE counts per species were twice as high in angiosperms compared to early-diverging land plants (**Figure 1**), possibly due to polyploidization. Of note, species MATE counts should be considered alongside the genome assembly version (**Supplementary Table 1**). This is because the number of MATEs we identified for species differs from existing data in some instances, which we believe reflects likely differences in the use of genome assemblies/annotations, selection criteria, and data download sources. For example, our CarMATEs count ( $n = 48$ ) is relatively low compared to that reported by Zhang et al. (2020), who identified 56 CarMATEs using the reference genome of the CDC Frontier available in the NCBI database (BioProject: PRJNA190909). At the same time, compared with previous studies, we identified comparable numbers of MATEs in the dicot model *Arabidopsis* (*Arabidopsis thaliana*;  $n = 57$ ) and the monocot model rice (*Oryza sativa*;  $n = 55$ ; Huang et al., 2019; Qiao et al., 2020), which further validates our identification strategy based on the MATE-domain based hmm profile search.

Depending on the evolutionary history of the analyzed species, the MATE count in each subgroup varied considerably, and a definitive expansion pattern was observed among them. The expansion was substantial for subgroups IB, IIA, and IVC (each containing > 500 MATEs), moderate for the IA, IIC, IID, IIE, and IIID subgroups (each containing ~200–300 MATEs), and weak for IIB, IIIA, IIIB, IIIC, IVA, and IVB subgroups (each containing ~60–130 MATEs). Consistent with the expansion trend, subgroups IB, IIA, and IVC retained a maximum number of synteny communities, whereas those of other subgroups retained only a few (**Supplementary Table 14**). This suggests that the subgroups IB, IIA, and IVC flourished during plant evolution, with a higher copy-number variation and synteny diversification, while other subgroups evolved more conservatively. For example, AtEDS5 is an essential component of salicylic acid-dependent signaling for disease resistance (Nawrath et al., 2002), while ADS1 (Sun et al., 2011), ZRIZI (Burko et al., 2011), and AtBIGE1A/ZmBIGE1 (Suzuki et al., 2015) are known to be involved in different aspects of plant architecture regulation. AtEDS5 was identified as a member of subgroup IIIA, whereas ADS1, ZRIZI, AtBIGE1A, and ZmBIGE1 were identified as members of subgroup IVC (**Supplementary Table 5**). Both of these subgroups are evolutionarily most conserved, containing members detected in early land plants through to angiosperms (**Figure 1**), but followed a contrasting evolutionary path. This implies that the evolutionary trends of MATE subgroups may be associated with the occurrence of functional novelties and/or conserved functions.

Even though we considered all hits of the hmm profile search of the MATE gene family and classified them as fully as possible,

several species-specific subgroup losses were observed (**Figure 1**). We presume that the loss of subgroups in certain species may not be true but may have occurred due to incomplete genome assemblies/annotations in these species. Examples include the loss of IIIA in *Asparagus officinalis*; because the IIIA subgroup was detected in all analyzed species except algae, its loss in *Asparagus* is highly unlikely. Nevertheless, the following logical observations can be stated based on our results: (i) subgroup IID may have been lost during Cucurbitaceae evolution because its members were not detected in any of the three analyzed genomes of Cucurbitaceae; (ii) subgroup IVB was not detected in any of the 12 Poaceae species (except rice), suggesting its loss during Poaceae evolution; and (iii) the absence of subgroup IIB in three Cucurbitaceae and two Asteraceae species implicates an order/family-specific loss. Notably, members of the subgroup IIB are involved in proanthocyanidin metabolism (**Supplementary Table 5**), and proanthocyanidins have been detected in cucumber fruit spines (Liu et al., 2019). Therefore, to further verify whether the loss of IIB is true in Asteraceae/Cucurbitaceae, we examined two additional species from Asteraceae and seven additional species from Cucurbitaceae (**Supplementary Table 20**). Results suggest that the loss of IIB in Asteraceae and the loss of IIB as well as IID in Cucurbitaceae may be true because all the additional species of Asteraceae/Cucurbitaceae lacked the homologs of those subgroups (**Supplementary Table 21**). This suggests the following two possibilities, of which we favor the former: (i) IIB members may be present in Asteraceae/Cucurbitaceae species, but were not detected in the current assembly versions of the corresponding analyzed genomes; and (ii) biosynthesis/accumulation of proanthocyanidins in Asteraceae/Cucurbitaceae species may have been mediated by members other than those of subgroup IIB.

To unravel structural feature evolution among the subgroups of the MATE gene family, we further examined the RefPlantMATEs dataset given that the inclusion of all MATEs (some of which are possibly redundant or the products of assembly/annotation error) could potentially result in bias. Based on this, the following observations were made (**Supplementary Figure 3**): (i) the subgroups of groups I/II, III, and IV mostly comprised of 6–10, > 10, and 0–5 introns, respectively; (ii) only a small proportion of all the subgroups (except IIIA) constitute of a single MatE domain, and (iii) the proportions of MATEs with 6–10 TMHMMs was substantial in subgroups IIE, IIIA, IIIB, IIIC, IIID, IVA, and IVB, compared to 11–15 TMHMMs in the other subgroups. These observations suggest that the subgroups of the MATE gene family evolved distinct structural features during the evolutionary history of plants.

## Insights on the Synteny Conservation of Characterized Plant Multidrug and Toxic Compound Extrusion Transporters

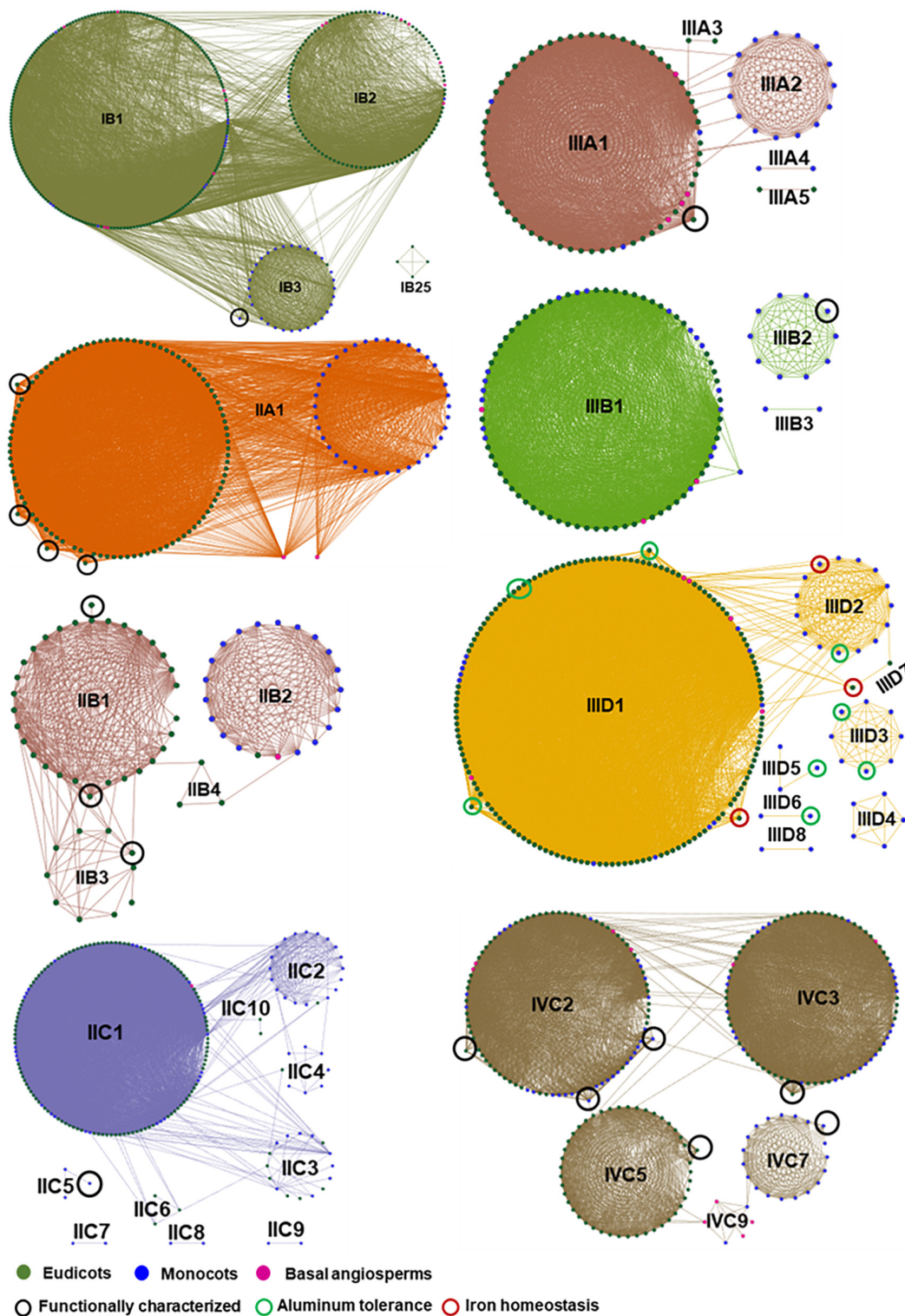
Phylogenetic clustering of synteny communities revealed distinct conservation and diversification patterns of genomic contexts in the MATE gene family of plants (**Figure 5**). Several communities showed deep conservation across the angiosperms, while several other communities showed lineage-, order/family-,

and species-specific diversification. Examples were observed in all subgroups at this level. At least 50 plant MATEs have been functionally characterized to date, representing 10 subgroups and 21 synteny communities (**Supplementary Table 5**). Notably, no MATEs from the subgroups IID, IIE, IIIC, and IVA have been previously characterized, which we believe expands the functional versatility of plant MATEs.

Subgroup IA was one of the moderately expanded subgroups with the lowest synteny gene percentage (**Supplementary Table 14**). AtALF5, conferring toxin resistance (Diener et al., 2001), and AtDTX18, involved in hydroxycinnamic acid amide transport (Dobritzsch et al., 2016), were identified in the Brassicaceae-specific synteny community IA6. IA1 contained members from every eudicot species, except Brassicaceae, and represented the largest IA community (**Supplementary Figure 4** and **Supplementary Table 15**). We presume that a gene translocation event in the Brassicaceae ancestry separated the IA6 community from IA1. Moreover, almost all of the IA communities (except IA14) retained exclusively eudicot or monocot genes (**Supplementary Figure 4**), indicating an ancient translocation event towards the divergence of the eudicot and monocot lineages. IB was the most diversified subgroup, accounting for the highest gene count and synteny communities (**Supplementary Table 14**), yet only three of its members were functionally characterized (**Supplementary Table 5**). Because tobacco (*Nicotiana tabacum*) was not included in our study, we are uncertain whether Nt-JAT1, which is involved in nicotine translocation (Morita et al., 2009), is retained in conserved synteny blocks. OsMATE2, which modulates arsenic accumulation (Das et al., 2018), was located in the IB3 community, which comprised only monocot genes and showed considerable connections to IB1 and IB2 communities (**Figure 6**). AtDTX1, which mediates the efflux of endogenous and exogenous toxic compounds from the cytoplasm (Li et al., 2002), was present in the Brassicaceae-specific community IB25, which showed no connection to any other IB communities. This suggests a gene transposition event in the ancestral Brassicaceae species.

Subgroup IIA was the third most diversified subgroup, with 73.6% of genes located in 32 synteny communities (**Supplementary Table 14**), and its members accounted for alkaloid as well as phenylpropanoid transport and auxin homeostasis (**Supplementary Table 5**). Synteny of Nt-JAT2, involved in nicotine sequestration (Shitan et al., 2014), was not identified, as tobacco was not included in our study. AtDTX30, involved in auxin homeostasis (Upadhyay et al., 2020) was not found in conserved synteny blocks, indicating a species-specific synteny loss. Flavonoid and anthocyanin transporters, namely AtFFT (Thompson et al., 2010), MtMATE2 (Zhao et al., 2011), MTP77 (Mathews et al., 2003), and anthoMATE1/anthoMATE3 (Gomez et al., 2009), were all located in the highly conserved angiosperm-wide community IIA1 (**Figure 6**). Of note is the presence of large communities (i.e., IIA1–IIA3, each consisting of > 70 genes and > 1,000 connections) with very few inter-community connections and diverse species (eudicot, monocot, and basal angiosperms) compositions (**Supplementary Figure 4**). This implies that

### Characterized plant MATEs in synteny communities of different subgroups



**FIGURE 6 |** Characterized plant MATE transporters in synteny communities of different identified subgroups. Green, blue, and pink nodes represent different species of eudicot, monocot, and basal angiosperm lineages, respectively. Characterized MATEs were inferred from **Supplementary Table 5** and are highlighted with a black circle in all subgroups except IIID. Green and brown circles in subgroup IIID indicate that the MATEs are characterized for aluminum tolerance and iron homeostasis, respectively. Lines connecting the nodes denote pairwise synteny relationships. Subgroups are highlighted in different colors.

ancient duplication events during the emergence of angiosperms contributed to the expansion/diversification of the subgroup IIA. IIB was one of the most conserved subgroups, with only four synteny communities (**Figure 6**). MtMATE1 (Zhao and Dixon, 2009), VvMATE2/VvMATE2 (Pérez-Díaz et al., 2014), and AtTT12 (Debeaujon et al., 2001) involved in proanthocyanidin metabolism are all located in the eudicot-wide synteny communities of IIB1 and IIB3. Of note, the IIB2 community comprised only monocot genes, one early eudicot gene, and one basal angiosperm gene that share almost no connections with IIB1/IIB3 (**Figure 6**). This suggests an ancient gene translocation event during lineage differentiation. Members of the IIC subgroup accounted for nicotine sequestration (NtMATE1 and NtMATE2; Shoji et al., 2009), phenolics efflux (OsPEZ1; Ishimaru et al., 2011), and novel aluminum tolerance mechanisms (ZmMATE2; Maron et al., 2010). While the synteny of NtMATE1 and NtMATE2 is not known and OsPEZ1 lost synteny, ZmMATE2 was identified in a Poaceae-specific IIC5 synteny community. Although IIC showed moderate expansion during plant evolution with 10 synteny communities, inter-community connections were commonly found (**Figure 6**).

Subgroup IIIA constituted of two major (IIIA1 and IIIA2) and three minor (IIIA3–IIIA5) synteny communities (**Figure 6**). AtEDS5 is involved in salicylic acid-dependent signaling (Nawrath et al., 2002) and was located in the IIIA1 community, members of which show inter-community connections with IIIA2 and IIIA3, suggesting that genes of IIIA4 and IIIA5 are the products of synteny loss. The MATEs of citrate transporters governing both iron homeostasis and aluminum tolerance (Durrett et al., 2007) represented six of the eight synteny communities in the IIID subgroup (**Supplementary Table 5**), in which eudicot MATEs show substantial synteny conservation while those of monocots are diversified to a lesser extent (**Figure 6**). ZmMATE1 (Maron et al., 2010) and OsFRDL2 (Yokosho et al., 2016) were located in two different Poaceae-specific communities, implying loss of synteny events. Since *Lotus japonicus* was not included in our study, we are uncertain whether the nodule-specific LjMATE1 (a member of IIID), which assists the translocation of iron from the root to nodules (Takanashi et al., 2013), is retained in conserved synteny blocks. Recently, ZmMATE6 was identified as contributing to aluminum tolerance through citrate transport (Du H. et al., 2021). ZmMATE6 was a member of the IIIB subgroup located in the Poaceae-specific synteny community (i.e., IIIB2), which had no connection with other communities (**Figure 6**).

In plant architecture maintenance, AtDET50, ZRIZI, OsDG1, ZmDG1, ADS1, AtBIGE1A, and ZmBIGE1 have diverse functions, such as organ initiation (Burko et al., 2011), grain filling (Qin et al., 2021), senescence promotion (Jia et al., 2019), and lateral organ-size regulation (Suzuki et al., 2015). These all represent IVC, the second-most diversified subgroup with the second-highest synteny gene percentage (**Supplementary Table 14**) and five synteny communities (IVC1, IVC2, IVC3, IVC5, and IVC7; **Supplementary Table 5**). These five communities were found to be partly connected based on one or more genes (**Supplementary Figure 4**). AtDTX50 involved in abscisic acid transport (Zhang et al., 2014) was

located in IVC1, which was conserved very deeply across the angiosperms (**Supplementary Figure 4** and **Supplementary Table 15**). While ZRIZI/OsDG1/ZmDG1 and ADS1 were located in the deeply conserved angiosperm-wide communities IVC2 and IVC3, respectively, the organ initiation and organ size regulators AtBIGE1A and ZmBIGE1 (Suzuki et al., 2015) were located in eudicot-wide IVC5 and monocot-wide IVC7 communities, respectively. Interestingly, IVC5 and IVC7 were interconnected by IVC9 with basal angiosperm genes (**Figure 6**). This indicates an ancient translocation event during the differentiation of eudicot and monocot lineages. A member of IVB (i.e., AtBIGE1B) has also been reported to mediate the regulation of organ initiation and organ size; however, this gene showed comparably less activity than AtBIGE1A (Suzuki et al., 2015). Inter-community connections involving at least one gene were observed among all the synteny communities of IVC except for IVC13–IVC17, and most of the large communities (i.e., IVC1–IVC4) retained genes from basal angiosperms, eudicots, and monocot lineages. This suggests that expansion/diversification of IVC started before or during the emergence of vascular/flowering plants *via* duplication events.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

MN, VK, and PK conceived, designed, and performed all the research, and wrote the first draft. BS, US, RN, CB, and PJ contributed to the drafting, revising, and approving of the draft contents. All authors contributed to the article and approved the submitted version.

## FUNDING

This research was supported in part by an in-house project of ICAR–NIPB, India to MN and an INSPIRE Faculty Fellowship Scheme of the Department of Science and Technology (DST), India to PK (Award No. IFA17-LSPA94).

## ACKNOWLEDGMENTS

We would like to thank Editage (www.editage.com) for English language editing.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.774885/full#supplementary-material>

## REFERENCES

- Ali, E., Saand, M. A., Khan, A. R., Shah, J. M., Feng, S., Ming, C., et al. (2021). Genome-wide identification and expression analysis of detoxification efflux carriers (DTX) genes family under abiotic stresses in flax. *Physiol. Plant.* 171, 483–501. doi: 10.1111/ppl.13105
- Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi: 10.1038/s41587-019-0036-z
- Bailey, P. C., Schudoma, C., Jackson, W., Baggs, E., Dagdas, G., Haerty, W., et al. (2018). Dominant integration locus drives continuous diversification of plant immune receptors with exogenous domain fusions. *Genome Biol.* 19:23. doi: 10.1186/s13059-018-1392-6
- Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176
- Burko, Y., Geva, Y., Refael-Cohen, A., Shleizer-Burko, S., Shani, E., Berger, Y., et al. (2011). From organelle to organ: zRIZI MATE-type transporter is an organelle transporter that enhances organ initiation. *Plant Cell Physiol.* 52, 518–527. doi: 10.1093/pcp/pcr007
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinform.* 10:421. doi: 10.1186/1471-2105-10-421
- Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Card, D. (2021). *Genestats*. Available online at: <https://github.com/darencard/GenomeAnnotation/blob/master/genestats>. (accessed on June 29, 2021).
- Chen, L., Liu, Y., Liu, H., Kang, L., Geng, J., Gai, Y., et al. (2015). Identification and expression analysis of mate genes involved in flavonoid transport in blueberry plants. *PLoS One* 10:e0118578. doi: 10.1371/journal.pone.0118578
- Chen, Q., Wang, L., Liu, D., Ma, S., Dai, Y., Zhang, Z., et al. (2020). Identification and expression of the multidrug and toxic compound extrusion (MATE) gene family in *Capsicum annuum* and *Solanum tuberosum*. *Plants* 9:1448. doi: 10.3390/plants9111448
- Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *Inter J. Complex Syst.* 1695, 1–9. doi: 10.1186/1471-2105-12-455
- Cytochrome (2021). *P450 Homepage*. Available online at: <https://drnelson.uthsc.edu/> [accessed on Sep 1, 2021].
- Das, N., Bhattacharya, S., Bhattacharyya, S., and Maiti, M. K. (2018). Expression of rice MATE family transporter OsMATE2 modulates arsenic accumulation in tobacco and rice. *Plant Mol. Biol.* 98, 101–120. doi: 10.1007/s11103-018-0766-1
- Debeaujon, I., Peeters, A. J. M., Leon-Kloosterziel, K. M., and Koornneef, M. (2001). The TRANSPARENT TESTA12 gene of *Arabidopsis* encodes a multidrug secondary transporter-like protein required for flavonoid sequestration in vacuoles of the seed coat endothelium. *Plant Cell* 19, 2023–2038. doi: 10.1105/tpc.13.4.853
- Diener, A. C., and Fink, G. R., Gaxiola, R. A. (2001). *Arabidopsis* ALF5, a multidrug efflux transporter gene family member, confers resistance to toxins. *Plant Cell* 13, 1625–1637. doi: 10.1105/tpc.010035
- Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* 14, 927–930. doi: 10.1111/j.1654-1103.2003.tb02228.x
- Dobritzsch, M., Lubken, T., Eschen-Lippold, L., Gorzalka, K., Blum, E., Matern, A., et al. (2016). MATE transporter-dependent export of hydroxycinnamic acid amides. *Plant Cell* 28, 583–596. doi: 10.1105/tpc.15.00706
- Droc, G., Larivière, D., Guignon, V., Yahiaoui, N., This, D., Garsmeur, O., et al. (2013). The banana genome hub. *Database* 2013:bat035. doi: 10.1093/database/bat035
- Du, H., Ryan, P. R., Liu, C., Li, H., Hu, W., and Yan, W. (2021). ZmMATE6 from maize encodes a citrate transporter that enhances aluminum tolerance in transgenic *Arabidopsis thaliana*. *Plant Sci.* 311:111016. doi: 10.1016/j.plantsci.2021.111016
- Du, Z., Su, Q., Wu, Z., Huang, Z., Bao, J., Li, J., et al. (2021). Genome-wide characterization of MATE gene family and expression profiles in response to abiotic stresses in rice (*Oryza sativa*). *BMC Ecol. Evo.* 21:141. doi: 10.1186/s12862-021-01873-y
- Durrett, T. P., Gassmann, W., and Rogers, E. E. (2007). The FRD3-mediated efflux of citrate into the root vasculature is necessary for efficient iron translocation. *Plant Physiol.* 144, 197–205. doi: 10.1104/pp.107.097162
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi: 10.1093/nar/gky995
- Frech, C., and Chen, N. (2010). Genome-wide comparative gene family classification. *PLoS One* 5:e13409. doi: 10.1371/journal.pone.0013409
- Gani, U., Sharma, P., Tiwari, H., Nautiyal, A. K., Kundan, M., Wajid, M. A., et al. (2021). Comprehensive genome-wide identification, characterization, and expression profiling of MATE gene family in *Nicotiana tabacum*. *Gene* 783:145554. doi: 10.1016/j.gene.2021.145554
- Gao, B., Wang, L., Oliver, M., Chen, M., and Zhang, J. (2020). Phylogenomic synteny network analyses reveal ancestral transpositions of auxin response factor genes in plants. *Plant Methods* 16:70. doi: 10.1186/s13007-020-00609-1
- Gomez, C., Terrier, N., Torregrasa, L., Violet, S., Fournier-Level, A., Verriès, C., et al. (2009). Grapevine MATE-type proteins act as vacuolar H<sup>+</sup>-dependent acylated anthocyanin transporters. *Plant Physiol.* 150, 402–415. doi: 10.1104/pp.109.135624
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944
- Hagenbuch, B., and Meier, P. J. (2003). Organic anion transporting polypeptides of the OATP/SLC21 family: phylogenetic classification as OATP/SLCO superfamily, new nomenclature and molecular/functional properties. *Pflugers Arch.* 447, 653–665. doi: 10.1007/s00424-003-1168-y
- Huang, J. J., An, W. J., Wang, K. J., Jiang, T. H., Ren, Q., Liang, W. H., et al. (2019). Expression profile analysis of MATE gene family in rice. *Biol. Plant.* 63, 556–564. doi: 10.32615/bp.2019.099
- Huang, Y., He, G., Tian, W., Li, D., Meng, L., Wu, D., et al. (2021). Genome-wide identification of mate gene family in potato (*Solanum tuberosum* L.) and expression analysis in heavy metal stress. *Front. Genet.* 12:650500. doi: 10.3389/fgene.2021.650500
- Hvorup, R. N., Winnen, B., Chang, A. B., Jiang, Y., Zhou, X. F., and Saier, M. H. (2003). The multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) exporter superfamily. *Eur. J. Biochem.* 270, 799–813. doi: 10.1046/j.1432-1033.2003.03418.x
- Hyndman, D., Bauman, D. R., Heredia, V. V., and Penning, T. M. (2003). The aldo-keto reductase superfamily homepage. *Chem. Biol. Inter.* 143–144, 621–631. doi: 10.1016/s0009-2797(02)00193-x
- Ishihara, T., Sekine, K. T., Hase, S., Kanayama, Y., Seo, S., Ohashi, Y., et al. (2008). Overexpression of the *Arabidopsis thaliana* EDS5 gene enhances resistance to viruses. *Plant Biol.* 10, 451–461. doi: 10.1111/j.1438-8677.2008.00050.x
- Ishimaru, Y., Kakei, Y., Shimo, H., Bashir, K., Sato, Y., Sato, Y., et al. (2011). A rice phenolic efflux transporter is essential for solubilizing precipitated apoplasmic iron in the plant stele. *J. Biol. Chem.* 286, 24649–24655. doi: 10.1074/jbc.M111.221168
- Jia, M., Liu, X., Xue, H., Wu, Y., Shi, L., Wang, R., et al. (2019). Noncanonical ATG8-ABS3 interaction controls senescence in plants. *Nat. Plant.* 5, 212–224. doi: 10.1038/s41477-018-0348-x
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031
- Juliao, M. H. M., Silva, S. R., Ferro, J. A., and Varani, A. M. (2020). A genomic and transcriptomic overview of MATE, ABC, and MFS transporters in *Citrus sinensis* interaction with *Xanthomonas citri* subsp. *citri*. *Plants* 9:794. doi: 10.3390/plants9060794
- Kerstens, M. H. L., Schranz, M. E., and Bouwmeester, K. (2020). Phylogenomic analysis of the APETALA2 transcription factor subfamily across angiosperms reveals both deep conservation and lineage-specific patterns. *Plant J.* 103, 1516–1524. doi: 10.1111/tpj.14843
- Kolde, R. (2015). *heatmap: Pretty Heatmaps*. Available online at: <https://CRAN.R-project.org/package=heatmap> (accessed on Mar 1, 2021).



- Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315
- Kumar, S., Stecher, G., Li, M., Niyaz, C., and Tamura, K. (2018). MEGA X: molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096
- Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301
- Li, L., He, Z., Pandey, G. K., Tsuchiya, T., and Luan, S. (2002). Functional cloning and characterization of a plant efflux carrier for multidrug and heavy metal detoxification. *J. Biol. Chem.* 277, 5360–5368. doi: 10.1074/jbc.M108777200
- Li, N., Meng, H., Xing, H., Liang, L., Zhao, X., and Luo, K. (2017). Genome-wide analysis of MATE transporters and molecular characterization of aluminum resistance in *Populus*. *J. Exp. Bot.* 68, 5669–5683. doi: 10.1093/jxb/erx370
- Li, R., Li, J., Li, S., Qin, G., Novak, O., Pencik, A., et al. (2014). ADP1 affects plant architecture by regulating local auxin biosynthesis. *PLoS Genet.* 10:e1003954. doi: 10.1371/journal.pgen.1003954
- Li, Y., He, H., and He, L. F. (2019). Genome-wide analysis of the MATE gene family in potato. *Mol. Biol. Rep.* 46, 403–414. doi: 10.1007/s11033-018-4487-y
- Liu, J., Li, Y., Wang, W., Gai, J., and Li, Y. (2016). Genome-wide analysis of MATE transporters and expression patterns of a subgroup of MATE genes in response to aluminum toxicity in soybean. *BMC Genom.* 17:223. doi: 10.1186/s12864-016-2559-8
- Liu, M., Zhang, C., Duan, L., Luan, Q., Li, J., Yang, A., et al. (2019). CsMYB60 is a key regulator of flavonols and proanthocyanidins that determine the colour of fruit spines in cucumber. *J. Exp. Bot.* 70, 69–84. doi: 10.1093/jxb/ery336
- Lu, P., Magwanga, R. O., Guo, X., Kirungu, J. N., Lu, H., Cai, X., et al. (2018). Genome-Wide analysis of multidrug and toxic compound extrusion (MATE) family in *Gossypium raimondii* and *Gossypium arboreum* and its expression analysis under salt, cadmium, and drought stress. *G3* 8, 2483–2500. doi: 10.1534/g3.118.200232
- Mackenzie, P. I., Owens, I. S., Burchell, B., Bock, K. W., Bairoch, A., Bélanger, A., et al. (1997). The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmacogenetics* 7, 255–269. doi: 10.1097/00008571-199708000-00001
- Magalhaes, J. V., Liu, J., Guimaraes, C. T., Lana, U. G., Alves, V. M., Wang, Y. H., et al. (2007). A gene in the multidrug and toxic compound extrusion (MATE) family confers aluminum tolerance in sorghum. *Nat. Genet.* 39, 1156–1161. doi: 10.1038/ng2074
- Maron, L. G., Pineros, M. A., Guimaraes, C. T., Magalhaes, J. V., Pleiman, J. K., Mao, C., et al. (2010). Two functionally distinct members of the MATE (multidrug and toxic compound extrusion) family of transporters potentially underlie two major aluminum tolerance QTLs in maize. *Plant J.* 61, 728–740. doi: 10.1111/j.1365-313X.2009.04103.x
- Mathews, H., Clendennen, S. K., Caldwell, C. G., Liu, X. L., Connors, K., Matheis, N., et al. (2003). Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. *Plant Cell* 15, 1689–1703. doi: 10.1105/tpc.012963
- Min, X., Jin, X., Liu, W., Wei, X., Zhang, Z., Ndayambaza, B., et al. (2019). Transcriptome-wide characterization and functional analysis of MATE transporters in response to aluminum toxicity in *Medicago sativa* L. *Peer J.* 7:e6302. doi: 10.7717/peerj.6302
- Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., and Punta, M. (2013). Challenges in homology search: hMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 41:e121. doi: 10.1093/nar/gkt263
- Morita, M., Rischer, H., Inze, D., Sawada, K., Oksman-Caldentey, K. M., Shitan, N., et al. (2009). Vacuolar transport of nicotine is mediated by a multidrug and toxic compound extrusion (MATE) transporter in *Nicotiana tabacum*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2447–2452. doi: 10.1073/pnas.0812512106
- Naake, T., Maeda, H. A., Proost, S., Tohge, T., and Fernie, A. R. (2021). Kingdom-wide analysis of the evolution of the plant type III polyketide synthase superfamily. *Plant Physiol.* 185, 857–875. doi: 10.1093/plphys/kiab086
- Nawrath, C., Heck, S., Parinithawong, N., and Métraux, J. P. (2002). EDS5, an essential component of salicylic acid-dependent signaling for disease resistance in *Arabidopsis*, is a member of the MATE transporter family. *Plant Cell* 14, 275–286. doi: 10.1105/tpc.010376
- NCBI Taxonomy Browser. (2021). <https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>. (Last accessed on May 10, 2021).
- Nelson, D., and Werck-Reichhart, D. (2011). A P450-centric view of plant evolution. *Plant J.* 66, 194–211. doi: 10.1111/j.1365-313X.2011.04529.x
- Nelson, D. R. (2009). The cytochrome p450 homepage. *Hum. Genomics* 4, 59–65. doi: 10.1186/1479-7364-4-1-59
- Omote, H., Hiasa, M., Matsumoto, T., Otsuka, M., and Moriyama, Y. (2006). The MATE proteins as fundamental transporters of metabolic and xenobiotic organic cations. *Trends Pharmacol. Sci.* 27, 587–593. doi: 10.1016/j.tips.2006.09.001
- Paquette, S., Möller, B. L., and Bak, S. (2003). On the origin of family 1 plant glycosyltransferases. *Phytochemistry* 62, 399–413. doi: 10.1016/s0031-9422(02)00558-7
- Pérez-Díaz, R., Ryngajllo, M., Pérez-Díaz, J., Peña-Cortés, H., Casaretto, J. A., González-Villanueva, E., et al. (2014). VvMATE1 and VvMATE2 encode putative proanthocyanidin transporters expressed during berry development in *Vitis vinifera* L. *Plant Cell Rep.* 33, 1147–1159. doi: 10.1007/s00299-014-1604-9
- Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., et al. (2012). The Pfam protein families database. *Nucleic Acids Res.* 40, D290–D301. doi: 10.1093/nar/gkr1065
- Qiao, C., Yang, J., Wan, Y., Xiang, S., Guan, M., Du, H., et al. (2020). A genome-wide survey of mate transporters in Brassicaceae and unveiling their expression profiles under abiotic stress in rapeseed. *Plants* 9:1072. doi: 10.3390/plants9091072
- Qin, P., Zhang, G., Hu, B., Wu, J., Chen, W., Ren, Z., et al. (2021). Leaf-derived ABA regulates rice seed development via a transporter-mediated and temperature-sensitive mechanism. *Sci. Adv.* 7:eabc8873. doi: 10.1126/sciadv.abc8873
- Rogers, E. E., and Guerinot, M. L. (2002). FRD3, a member of the multidrug and toxin efflux family, controls iron deficiency responses in *Arabidopsis*. *Plant Cell* 14, 1787–1799. doi: 10.1105/tpc.001495
- Santos, A., Chaves-Silva, S., Yang, L., Maia, L., Chalfun-Junior, A., Sinharoy, S., et al. (2017). Global analysis of the MATE gene family of metabolite transporters in tomato. *BMC Plant Biol.* 17:185. doi: 10.1186/s12870-017-1115-2
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* 11:e0163962. doi: 10.1371/journal.pone.0163962
- Shitan, N., Minami, S., Morita, M., Hayashida, M., Ito, S., Takanashi, K., et al. (2014). Involvement of the leaf-specific multidrug and toxic compound extrusion (MATE) transporter Nt-JAT2 in vacuolar sequestration of nicotine in *Nicotiana tabacum*. *PLoS One* 9:e108789. doi: 10.1371/journal.pone.0108789
- Shoji, T., Inai, K., Yazaki, Y., Sato, Y., Takase, H., Goto, Y., et al. (2009). Multidrug and toxic compound extrusion-type transporters implicated in vacuolar sequestration of nicotine in tobacco roots. *Plant Physiol.* 149, 708–718. doi: 10.1104/pp.108.132811
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Sun, X., Gilroy, E. M., Chini, A., Nurmberg, P. L., Hein, I., Lacomme, H., et al. (2011). ADS1 encodes a MATE-transporter that negatively regulates plant disease resistance. *New Phytol.* 192, 471–482. doi: 10.1111/j.1469-8137.2011.03820.x
- Suzuki, M., Sato, Y., Wu, S., Kang, B. H., and McCarty, D. R. (2015). Conserved functions of the MATE transporter BIG EMBRYO1 in regulation of lateral organ size and initiation rate. *Plant Cell* 27, 2288–2300. doi: 10.1105/tpc.15.00290
- Takanashi, K., Shitan, N., and Yazaki, K. (2014). The multidrug and toxic compound extrusion (MATE) family in plants. *Plant Biotechnol.* 31, 417–430. doi: 10.5511/plantbiotechnology.14.0904a
- Takanashi, K., Yokosho, K., Saeki, K., Sugiyama, A., Sato, S., Tabata, S., et al. (2013). LjMATE1: a citrate transporter responsible for iron supply to the nodule infection zone of *Lotus japonicus*. *Plant Cell Physiol.* 54, 585–594. doi: 10.1093/pcp/pct019

- Tang, R. J., Luan, M., Wang, C., Lhamo, D., Yang, Y., Zhao, F. G., et al. (2020). Plant membrane transport research in the post-genomic era. *Plant Commun.* 1:100013. doi: 10.1016/j.xplc.2019.100013
- Thompson, E. P., Wilkins, C., Demidchik, V., Davies, J. M., and Glover, B. J. (2010). An Arabidopsis flavonoid transporter is required for anther dehiscence and pollen development. *J. Exp. Bot.* 61, 439–451. doi: 10.1093/jxb/erp312
- Udp-glycosyltransferase [UGT] (2021). *Nomenclature Committee*. Available online at: <https://prime.vetmed.wsu.edu/resources/udp-glucuronosyltransferase-homepage> (accessed on Sep 1, 2021).
- Upadhyay, N., Kar, D., and Datta, S. (2020). A multidrug and toxic compound extrusion (MATE) transporter modulates auxin levels in root to regulate root development and promotes aluminium tolerance. *Plant Cell Environ.* 43, 745–759. doi: 10.1111/pce.13658
- Upadhyay, N., Kar, D., Deepak, M. B., Nanda, S., Rahiman, R., Panchakshari, N., et al. (2019). The multitasking abilities of MATE transporters in plants. *J. Exp. Bot.* 70, 4643–4656. doi: 10.1093/jxb/erz246
- Wang, J., Hou, Q., Li, P., Yang, L., Sun, X., Bhagavatula, L., et al. (2017). Diverse functions of multidrug and toxin extrusion (MATE) transporters in citric acid efflux and metal homeostasis in *Medicago truncatula*. *Plant J.* 90, 79–95. doi: 10.1111/tbj.13471
- Wang, L., Bei, X., Gao, J., Li, Y., Yan, Y., and Hu, Y. (2016). The similar and different evolutionary trends of MATE family occurred between rice and *Arabidopsis thaliana*. *BMC Plant Biol.* 16:207. doi: 10.1186/s12870-016-0895-0
- Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40:e49. doi: 10.1093/nar/gkr1293
- Wu, C. H., Huang, H., Yeh, L. S., and Barker, W. C. (2003). Protein family classification and functional annotation. *Comput. Biol. Chem.* 27, 37–47. doi: 10.1016/s1476-9271(02)00098-1
- Xu, L., Shen, Z. L., Chen, W., Si, G. Y., Meng, Y., Guo, N., et al. (2019). Phylogenetic analysis of upland cotton MATE gene family reveals a conserved subfamily involved in transport of proanthocyanidins. *Mol. Biol. Rep.* 46, 161–175. doi: 10.1007/s11033-018-4457-4
- Yokosho, K., Yamaji, N., Fuji-Kashino, M., and Ma, J. F. (2016). Functional analysis of a MATE gene OsFRDL2 revealed its involvement in Al-induced secretion of citrate, but a lower contribution to Al tolerance in rice. *Plant Cell Physiol.* 57, 976–985. doi: 10.1093/pcp/pcw026
- Zhang, H., Zhu, H., Pan, Y., Yu, Y., Luan, S., and Li, L. (2014). A DTX/MATE-type transporter facilitates abscisic acid efflux and modulates ABA sensitivity and drought tolerance in Arabidopsis. *Mol. Plant.* 7, 1522–1532. doi: 10.1093/mp/ssu063
- Zhang, W., Liao, L., Xu, J., Han, Y., and Li, L. (2021). Genome-wide identification, characterization and expression analysis of MATE family genes in apple (*Malus × domestica* Borkh). *BMC Genom.* 22:632. doi: 10.1186/s12864-021-07943-1
- Zhang, X., Weir, B., Wei, H., Deng, Z., Zhang, X., Zhang, Y., et al. (2020). Genome-wide identification and transcriptional analyses of MATE 2 transporter genes in root tips of wild *Cicer* spp. under aluminium stress. *bioRxiv* [Preprint] doi: 10.1101/2020.04.27.063065
- Zhao, J., and Dixon, R. A. (2009). MATE transporters facilitate vacuolar uptake of epicatechin 3'-O-glucoside for proanthocyanidin biosynthesis in *Medicago truncatula* and Arabidopsis. *Plant Cell* 21, 2323–2340. doi: 10.1105/tpc.109.067819
- Zhao, J., Huhman, D., Shadle, G., He, X. Z., Sumner, L. W., Tang, Y., et al. (2011). MATE2 mediates vacuolar sequestration of flavonoid glycosides and glycoside malonates in *Medicago truncatula*. *Plant Cell* 23, 1536–1555. doi: 10.1105/tpc.110.080804
- Zhao, T., Holmer, R., de Bruijn, S., Angenent, G. C., van den Burg, H. A., and Schranz, M. E. (2017). Phylogenomic synteny network analysis of MADS-box transcription factor genes reveals lineage-specific transpositions, ancient tandem duplications, and deep positional conservation. *Plant Cell* 29, 1278–1292. doi: 10.1105/tpc.17.00312
- Zhao, T., and Schranz, M. E. (2017). Network approaches for plant phylogenomic synteny analysis. *Curr. Opin. Plant Biol.* 36, 129–134. doi: 10.1016/j.cpb.2017.03.001
- Zheng, Y., Wu, S., Bai, Y., Sun, H., Jiao, C., Guo, S., et al. (2019). Cucurbit Genomics Database (CuGenDB): a central portal for comparative and functional genomics of cucurbit crops. *Nucleic Acids Res.* 47, D1128–D1136. doi: 10.1093/nar/gky944
- Zhu, H., Wu, J., Jiang, Y., Jin, J., Zhou, W., Wang, Y., et al. (2016). Genomewide analysis of MATE-type gene family in maize reveals microsynteny and their expression patterns under aluminum treatment. *J. Genet.* 95, 691–704. doi: 10.1007/s12041-016-0686-2

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Nimmy, Kumar, Suthanthiram, Subbaraya, Nagar, Bharadwaj, Jain and Krishnamurthy. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.