Pelita Perkebunan 40(2) 2024, 105–124
DOI: 10.22302/iccri.jur.pelitaperkebunan.v40i2.600
ISSN: 0215-0212 / e-ISSN: 2406-9574

Silico Study of *RKD4* Gene Function in *Coffea arabica* L. and Various Cultivated Plants Related to Embryo Formation Initiation

Rina Arimarsetiowati[1*)], Endang Semiarti[2)], Budi Setiadi Daryono[2)], Yohana Theresia Maria Astuti[3)] and Erwin Prastowo[4)]

[1)]Indonesian Coffee and Cocoa Research Institute, JL. PB. Sudirman No. 90, Jember 68118, Indonesia
[2)]Department of Tropical Biology, Faculty of Biology, Universitas Gadjah Mada, Jl. Teknika Selatan Sekip Utara, Yogyakarta, Indonesia
[3)]Faculty of Agriculture, Institut Pertanian Stiper (Instiper), Jl. Nangka II, Yogyakarta 55282, Indonesia
[4)]Indonesian Research Institute for Estate Crops, Jl. Salak No.1A, Bogor 16128, Indonesia
[*)]Corresponding author email: arimarsetiowati@gmail.com
Received: 12 March 2024 / Accepted: 3 July 2024

Abstract

Arabica coffee supplies 60% of world coffee production because it has a unique taste as a superior quality beverage. Arabica coffee micropropagation can be conducted by somatic embryogenesis technique which produces clonal, fast and uniform plants. The somatic embryogenesis (SE) process describes the integration of endogenous signals and gene reprogramming, which releases signals to initiate embryogenic processes. The use of endogenous auxin, either single or in combination with other plant growth regulators (PGRs) or stress, induces differential gene expression, which modifies the genetic program of somatic cells and regulates the transition to each stage during SE development. The *RKD4* gene (*RWP-RK DOMAIN-CONTAINING 4*) is a gene that plays a role in the early initiation of embryo formation and development. The characterization of *RKD4* genes in *C. arabica* is still limited and under-explored. The objective of this research is to explore the characteristics of *RKD4* gene by comparing the difference and similarities of *RKD4* gene in *C. arabica* and other cultivated plants. The method was initiated by identifying nucleotide sequences from the National Center for Biotechnology Information (NCBI) database. Furthermore, it consists of analysis of nucleotide alignment, alignment of amino acid sequences, protein analysis, protein motif functions discovery, analysis of phylogenetic tree, protein 2D and 3D-modelling and physicochemical properties. According to the analysis, there were 100 polymorphism points with a total number of mutations of 211 points. The phylogenetic tree shows *C. arabica* L. has a very close relationship with grapes (*Vitis vinifera*) based on the RKD4 protein, gene structures and protein motifs. There are nine highly conserved motifs found in the protein alignment. *C. arabica* L. had more methyl jasmonate element responses than *A. thaliana*. The findings are useful to understand the initiation of embryo formation mechanisms of *C. arabica* L and other cultivated plants during propagation through somatic embryogenesis in the long run.

Keywords: *Coffea arabica* L., dicotyl, initiation embryo, phylogenetic, RKD4

INTRODUCTION

Coffee is an important commodity that is cultivated throughout the world and has high economic value. Indonesia is on the fourth position as the largest coffee producer in the world after Brazil, Vietnam and Colombia. Of the 130 different coffee species, 2 commercial coffee species are cultivated, namely Arabica coffee (*C. arabica* L.) and Robusta coffee (*C. canephora*). Arabica coffee has a unique taste and supplies 60% of world coffee production. Arabica coffee has superior quality as a drink compared to Robusta coffee (van der Vossen *et al*., 2015).

Coffee propagation can be done generatively (seeds) and vegetatively (cuttings and grafting). *In vitro*-vegetative propagation can be done using a tissue culture technique through a somatic embryogenesis (SE) method. SE is a modern biotechnology method for fast and efficient plant propagation, which is clonal and can be mass-produced without depending on the season. This technique can produce superior coffee seeds with the same genetic characteristics as the parent (uniform) (Debnath, 2018). SE is developed from haploid or diploid somatic cells without going through gamete fusion and then develops into superior clones (Salaün, *et al*., 2021). Somatic cells undergo several morphological and biochemical changes to form somatic embryos starting from the globular, heart, torpedo and cotyledone phases.

The SE process involves three phases: induction of SE, formation of central meristematic, and development of somatic embryos (Elhiti *et al*., 2013). Each stage of the SE process describes the integration of endogenous signals and gene reprogramming, which releases signals to initiate embryogenic processes. The use of endogenous auxin, either single or in combination with other PGRs or stress, induces different gene expressions, which modify the genetic program of somatic cells and regulate the transition to each stage during SE development (Loyola-Vargas & Ochoa-Alejo, 2016). Most of these genes fall into one of four categories as follows: transcription factors (TFs), proteins acting in the cell cycle, PGR biosynthesis (especially auxin), and proteins involved in signaling pathways (Leljak-Levanic *et al*., 2015).

Changes in the cell genetic program that lead to the induction of SE have required the regulation of several genes (Riechmann *et al.,* 2000). In both angiosperms and gymnosperms, little is known about gene expression, an early stage of embryogenesis, which is critical for subsequent embryonic development (Trontin *et al*., 2016). In *Arabidopsis thaliana* as a plant model, SE can be induced from various explants throughout its life cycle and occurs either directly or indirectly through the callus phase. SE can be induced using the synthetic auxin 2,4-D or by over-expression of certain transcription factors, including *BABY BOOM* (*BBM*), *LEAFY COTYLEDON* (LEC), *WUSCHEL* (*WUS*), *RWP-RK DOMAIN-CONTAINING 4* (*RKD4*) and *AND WOUND INDUCTION OF DEDIFFERENTIATION 1* (*WIND1*) (Horstman *et al*., 2017).

RWP-RK DOMAIN-CONTAINING 4 (*RKD4*)/*GROUNDED* (GRD) is a transcription factor expressed by embryos that can induce SE (Waki *et al*., 2011). *RKD4* is expressed throughout the early embryo and in the suspensor. In line with its unique role during early embryogenesis, other *RKDs* only affect embryo sac development (Horstman *et al.*, 2017). *RKD4* is the only *RKD* factor that induces SE (Koszegi *et al*., 2011; Tedeschi *et al*., 2017; Waki *et al*., 2011). The *RKD4* gene (*RWP-RK DOMAIN-CONTAINING 4*) is a gene that plays a role in early zygotic embryo development (Smertenko & Bozhkov, 2014). Waki *et al*. (2011) succeeded in over-expressing the *indRKD4ox* gene with the synthetic glucocorticoid dexamethasone (Dex) induction to induce somatic embryogenesis in *A. thaliana* plant seeds. In orchid

plants, the insertion of the *AtRKD4* gene successfully induced the formation of somatic embryos ectopically in the leaves and protocorms of the hybrid orchid *Phalaenopsis* 'Sogo Vivien' (Mursyanti *et al*., 2015) and *Dendrobium phalaenopsis* (Setiari *et al*., 2018).

According to Kõszegi *et al*. (2011), the most abundant levels of *RKD4* transcripts are found in the reproductive organs of the *A. thaliana* genome. *RKD* genes have been implicated in germ cell differentiation and regulation of gametophyte-sporophyte transition in the liverwort *Marchantia polymorpha* (Koi *et al*., 2016; Rövekamp *et al*., 2016). In *Arabidopsis* zygotic embryo development, *AtRKD4* plays a role at zygote elongation stage and the first asymmetric division (Weindrich & Weijers, 2013). Loss of *RKD4* function results in reduced zygote elongation and causes certain abnormalities during previous cell divisions, suggesting its participation in the control of embryogenesis (Waki *et al*., 2011). *RKD4* mutations cause short suspensors and inhibition of embryo formation, induced overexpression of *RKD4* in *Arabidopsis* plant seeds, causes overproliferation of root cells, where somatic embryos develop (Horstman *et al*., 2017).

The aim of this research was to find out more about the characteristics of the *RKD4* gene by comparing the nucleotide structure of the *RKD4* gene in *C. arabica* L. and other cultivated plants of monocot and dicot plants. We also analyzed the similarities and differences between the amino acid motifs that make up the RKD4 protein and the structure of RKD4 protein. Moreover, we obtained information about the relationship of the *RKD4* gene between *C. arabica* L. and other cultivated plants. Thus, it will facilitate further studies related to biological functions for the *RKD4* family genes in *C. arabica* L. and other cultivated plants.

MATERIALS AND METHODS

The initial stage of bioinformatics analysis was preceded by searching the nucleotide sequence of the *C. arabica* L. *RKD4* gene which was downloaded from National Center for Biotechnology Information (NCBI). Apart from *C. arabica* L., there are 41 other cultivated plants (both monocots and dicots) that are used as comparisons. Species names, accession numbers, genome lengths, CoDing Sequence (CDS) lengths and chromosome locations for plant species were obtained from the NCBI database (Table 1).

Polymorphism Analysis of *RKD4* Gene

The nucleotide sequence is copied into a notepad for alignment. Naming each sequence name is done in the same format. Files are saved using the FASTA format. After the sequence alignment file has been successfully saved, the MultAlin software is opened to process the alignment data into one alignment. The alignment results are saved in GIF image format. The part of the nucleotide that is conserved is marked in red, while the part of the nucleotide that is not conserved is marked in black. Several nucleotides are marked in blue if several nucleotides have the same sequence.

After using MultAlin, polymorphism analysis was carried out using MEGA11 and DNASP software. This software can find out the similarities and differences in each sequence in the nucleotide alignment. The first step taken was to open the sequence data that had been compiled into one file with the FASTA extension using MEGA11 software.

Once the file is open in MEGA11, the next step is to align the sequence using the CLUSTAL-W plug-in. The "Align all" is selected. After the alignment results are successfully formed, the file is saved in MEGA (MEG) format. Next, the alignment file from MEGA11 is opened using DNASP software,

Table 1. List of nucleotide sequences of the *RKD4* gene in *C. arabica* L. and other cultivated plants

No.	Species	Class	Family	Accession number	Length of genome (bp)	CDS (bp)	Location of chromosom	GC Content (%)
1.	*Coffea arabica*	Dicotyledoneae	Rubiaceae	XM_027252817.1	1218	753	2C	43.10
2.	*Arabidopsis thaliana*	Dicotyledoneae	Brassicaceae	NM_124683.1	982	771	5	37.37
3.	*Prunus dulcis*	Dicotyledoneae	Rosaceae	XM_034349665.1	1780	1018	2	37.42
4.	*Prunus avium*	Dicotyledoneae	Rosaceae	XM_021970572.1	1994	1239	*Unknown*	37.11
5.	*Capsicum annuum*	Dicotyledoneae	Solanaceae	XM_016711407.2	1463	1007	3	33.42
6.	*Malus domestica*	Dicotyledoneae	Rosaceae	XM_029107382.1	1404	868	1	40.60
7.	*Camellia sinensis*	Dicotyledoneae	Theaceae	XM_028230451.1	1830	660	*Unknown*	30.60
8.	*Vitis vinifera*	Dicotyledoneae	Vitaceae	XM_019226223.1	1198	985	16	41.57
9.	*Sesamum indicum*	Dicotyledoneae	Pedaliaceae	XM_011086441.1	691	582	LG7	38.06
10.	*Juglans regia*	Dicotyledoneae	Juglandaceae	XM_018987129.1	12294	843	16	36.60
11.	*Macadamia integrifolia*	Dicotyledoneae	Proteaceae	XM_042624636.1	1415	1239	14	38.02
12.	*Carica papaya*	Dicotyledoneae	Caricaceae	XM_022053636.1	1239	741	*Unknown*	39.06
13.	*Hevea brasiliensis*	Dicotyledoneae	Euphorbiaceae	XM_021802314.1	1349	1129	*Unknown*	38.32
14.	*Manihot esculenta*	Dicotyledoneae	Euphorbiaceae	XM_021750489.2	776	483	1	37.37
15.	*Durio zibethinus*	Dicotyledoneae	Malvaceae	XM_022881331.1	1593	839	*Unknown*	36.78
16.	*Gossypium hirsutum*	Dicotyledoneae	Malvaceae	XM_016884988.2	1049	769	D03	37.84
17.	*Theobroma cacao*	Dicotyledoneae	Malvaceae	XM_007039017.2	1643	857	3	37.91
18.	*Brassica oleracea*	Dicotyledoneae	Brassicaceae	XM_013764601.1	727	543	C2	40.30
19.	*Raphanus sativus*	Dicotyledoneae	Brassicaceae	XM_018600378.1	913	699	*Unknown*	38.55
20.	*Brassica napus*	Dicotyledoneae	Brassicaceae	XM_013813182.3	915	702	A10	37.92
21.	*Jatropha curcas*	Dicotyledoneae	Euphorbiaceae	XM_012231078.1	1808	654	*Unknown*	31.41
22.	*Cucumis sativus*	Dicotyledoneae	Cucurbitaceae	XM_031882442.1	15,06	814	3	29.94
23.	*Cucumis melo*	Dicotyledoneae	Cucurbitaceae	XM_008468620.3	1086	699	6	33.88
24.	*Cucurbita maxima*	Dicotyledoneae	Cucurbitaceae	XM_023112393.1	1124	893	*Unknown*	42.34
25.	*Momordica charantia*	Dicotyledoneae	Cucurbitaceae	XM_022285451.1	1053	756	*Unknown*	41.21
26.	*Glycine soja*	Dicotyledoneae	Fabaceae	XM_028380324.1	1797	1333	6	38.11
27.	*Vigna radiata*	Dicotyledoneae	Fabaceae	XM_022786971.1	1735	1218	10	38.84
28.	*Glycine max*	Dicotyledoneae	Fabaceae	XM_006578834.4	1585	1094	4	38.61
29.	*Pisum sativum*	Dicotyledoneae	Fabaceae	XM_051022087.1	1793	1380	5	35.08
30.	*Arachis hypogaea*	Dicotyledoneae	Leguminosae	XM_025811904.2	1233	1064	Arahy.17	38.52
31.	*Spinacia oleracea*	Dicotyledoneae	Amaranthaceae	XM_021987361.1	2116	801	*Unknown*	31.04
32.	*Chenopodium quinoa*	Dicotyledoneae	Amaranthaceae	XM_021905859.1	1867	807	*Unknown*	30.63
33.	*Triticum estivum*	Monocotyledonae	Poaceae	XM_044581535.1	801	652	7D	51.06
34.	*Elaeis guineensis*	Monocotyledonae	Arecaceae	XM_010913942.3	1430	1040	1	43.15
35.	*Asparagus officinalis*	Monocotyledonae	Asparagaceae	XM_020411045.1	1640	858	5	35.18
36.	*Hordeum vulgare*	Monocotyledonae	Poaceae	XM_045099286.1	2992	893	6H	45.92
37.	*Eucalyptus grandis*	Dicotyledoneae	Myrtaceae	XM_018864935.1	1664	957	7	45.37
38.	*Phoenix dactylifera*	Monocotyledonae	Arecaceae	XM_039129640.1	1027	534	8	44.10
39.	*Mangifera indica*	Dicotyledoneae	Anacardiaceae	XM_044615000.1	668	494	17	39.67
40.	*Nicotiana tabacum*	Dicotyledoneae	Solanaceae	XM_016651587.1	697	503	*Unknown*	35.43
41.	*Dendrobium catenatum*	Monocotyledonae	Orchidaceae	XM_028696900.1	17678	478	*Unknown*	34.11
42.	*Zea mays*	Monocotyledonae	Poaceae	XM_008664535.3	954	789	10	55.66

this software can only read data from MEGA format. In the DNASP menu, the MEGA format file is opened. The "Analysis" menu, "DNA Polymorphism", the "File" tab, and "Save/Export Data As" are selected. Then NEXUS (NEX) is selected to save the format. The file in NEXUS format is opened using notepad, all parts are copied into an MS Word file. The polymorphism data that has been obtained is given a nucleotide number notation for each sequence that is different from the reference sequence. The reference sequence used is *C. arabica* L.

Analysis of Phylogenetic Relationship

The amino acid sequences of the RKD4 protein of *C. arabica* L. and 41 other cultivated plant species were aligned using MEGA11 software with the MUSCLE algorithm. The results of the alignment are used to build a phylogenetic kinship tree with parameters used based on Maximum-Likelihood analysis (Tamura-Nei model) with a level 5 category gamma distribution (LG+G). Analysis of the RKD4 protein sequence into a phylogenetic tree using Bootstrap replication 1000 times.

Analysis of *RKD4* Gene Structure and Protein Motifs

The nucleotide sequence and CDS of the *RKD4* gene of *C. arabica* L. and 41 other cultivated plants were used to construct an exon-intron organization map using the Gene Structure Display Server (GSDS) program (Hu *et al*., 2015). Analysis of the RKD4 protein motif begins by downloading the RKD4 protein sequence from NCBI in FASTA format and copying it into a notepad. The FASTA file was read using MultAlin, and protein sequence alignment data was obtained. In the protein data, it will get the pattern arrangement of the amino acids that make up the protein. The results from MultAlin are analyzed and read to see the similarities and differences in the protein motifs formed, notated in rectangular form, with colors according to the amino acid sequence motifs formed.

The next step is to use the web-based software MEME-Suites to identify and classify conserved protein motifs. The protein sequence was uploaded to the MEME-Suites server, the parameters were set to default with the number of motifs searched being 20 motifs. Known protein motifs will be colored automatically by this software. The results obtained are in the form of information that informs the motifs contained in each sequence and their differences from each other. The motifs formed were marked and compared with literature studies, then a table was created containing each protein motif along with its description and function. The description and function of the RKD4 protein motifs were analyzed using GenomNet software. GenomeNet is a Japanese network of database and computing services for genome research and related research fields in the biomedical sciences, operated by the Bioinformatics Center of Kyoto University.

Analysis of Physicochemical

The characteristics of the RKD4 protein of *C. arabica* L. and 41 other cultivated plants were carried out through physicochemical analysis which included peptide length, protein molecular weight, isoelectric point, instability index, aliphatic index, average hydropathy value, predicted using ProtParam software.

Two-Dimensional and Three-Dimensional Modeling of the RKD4 Protein

Two-dimensional modeling of the RKD4 protein was carried out using Chou & Fasman Secondary Structure Prediction Server (CFSSP) software. Protein sequences are entered into the CFSSP software processing column one by one and processed by pressing "PREDICT". Three-dimensional modeling of the RKD4

protein was carried out using Swiss Model-expasy software. Protein sequences were entered into the processing fields of the Swiss Model-expasy web server. Parameters are set by default. The process begins by pressing "Build Model" in the software.

Analysis of Cis-Acting Elements in Promoter Regions

The promoter regions of each *RKD4* gene from each plant species were downloaded from the NCBI database. Cis-Acting elements in the promoter region were analyzed using PlantCARE and categorized based on their respective functions

RESULTS AND DISCUSSION

Polymorphism Analysis of *RKD4* Gene in *C. arabica* L. and Other Cultivated Plants

Analysis of the *RKD4* gene sequence shows the presence of polymorphism at the nucleotide level. It was found that many mutations consisting of gaps appeared in this alignment. At the nucleotide level, variations between different gene families occur due to high genetic variation among their members. Figure 1 shows nucleotide alignment with parts of the nucleotides that are conserved among the *RKD4* genes marked in red, some parts that are conserved in this alignment are shown in blue, while non-conserved regions are shown in black.

A putative conserved nucleotide sequence pattern was observed in sequences 1110 to 1181 with some minor variations (Figure 1). The difference in the form of a gap seen in this alignment is due to the absence of a nucleotide sequence. Variations at the nucleotide level have the potential to cause changes in the sequence of amino acids that make up proteins. A more in-depth analysis was carried out using DNASP (Rozas *et al*., 2017) to determine the polymorphisms contained in the aligned sequences. Figure 2 shows the nucleotide sequence processed in DNASP software to determine the position and location of the polymorphism.

The nucleotide sample used as a reference for polymorphism analysis is *C. arabica* L. This polymorphism data shows many differences between the nucleotide sequences in *C. arabica* L. and other cultivated plants. This is because in this analysis different plant DNA sequences were used starting from the taxon class level so that many polymorphisms were found. Similarities in sequence are indicated by dots, while differences in nucleotide base sequence are indicated by different colored letters. There were 100 polymorphism points in the data, with a total number of mutations of 211 points.

Analysis of RKD4 Protein Motif

Sequence analysis of the RKD4 protein showed conservative amino acids. Figure 3 shows a motif analysis of the RKD4 protein indicating conserved regions in *C. arabica* L. and other cultivated plants. Highly conserved amino acids are shown in red, other amino acids are shown in blue, and non-conserved amino acids are shown in black.

At the protein level, characterization is carried out by observing the amino acid motif of each sequence analyzed. Figure 3 shows that the RWP-RK motif contains a transcription factor that is essential for the initial initiation of somatic embryo formation (Waki *et al*., 2011). This RWP-RK motif region has an important function in the initial induction of plant somatic embryos. *RKD4* mutant plants that lose their function are defective in zygotic cell elongation and subsequent cell division patterns. *RKD4* is transcribed in the early embryo. *RKD4* has the functional characteris-

1041 1050 1060 1070 1080 1090 1100 1110 1120 1130 1140 1150 1160 1170
Coffea_arabica
Carica_papaya
Cucumis_sativus
Cucumis_melo
Camellia_sinensis
Jatropha_curcas
Macadamia_integrifol
Glycine_soja
Vigna_radiata
Glycine_max
Pisum_sativum
Hevea_brasiliensis
Prunus_dulcis
Prunus_avium
Capsicum_annuum
Vitis_vinifera
Elaeis_guineensis
Malus_domestica
Eucalyptus_grandis
Hordeum_vulgare
Arachis_hypogeae
Cucurbita_maxima
Momordica_charantia
Durio_zibethinus
Theobroma_cacao
Gossypium_hirsutum
Juglans_regia
Spinacia_oleracea
Chenopodium_quinoa
Asparagus_officinali
Arabidopsis_thaliana
Raphanus_sativus
Brassica_napus
Brassica_oleracea
Sesamum_indicum
Manihot_esculenta
Mangifera_indica
Nicotiana_tabacum
Phoenix_dactylifera
Dendrobium_catenatium
Triticum_aestivum
Zea_mays
Consensus

Figure 1. Sequence alignment of *RKD4* gene in *C. arabica* L. and other cultivated plants using MultAlin Software

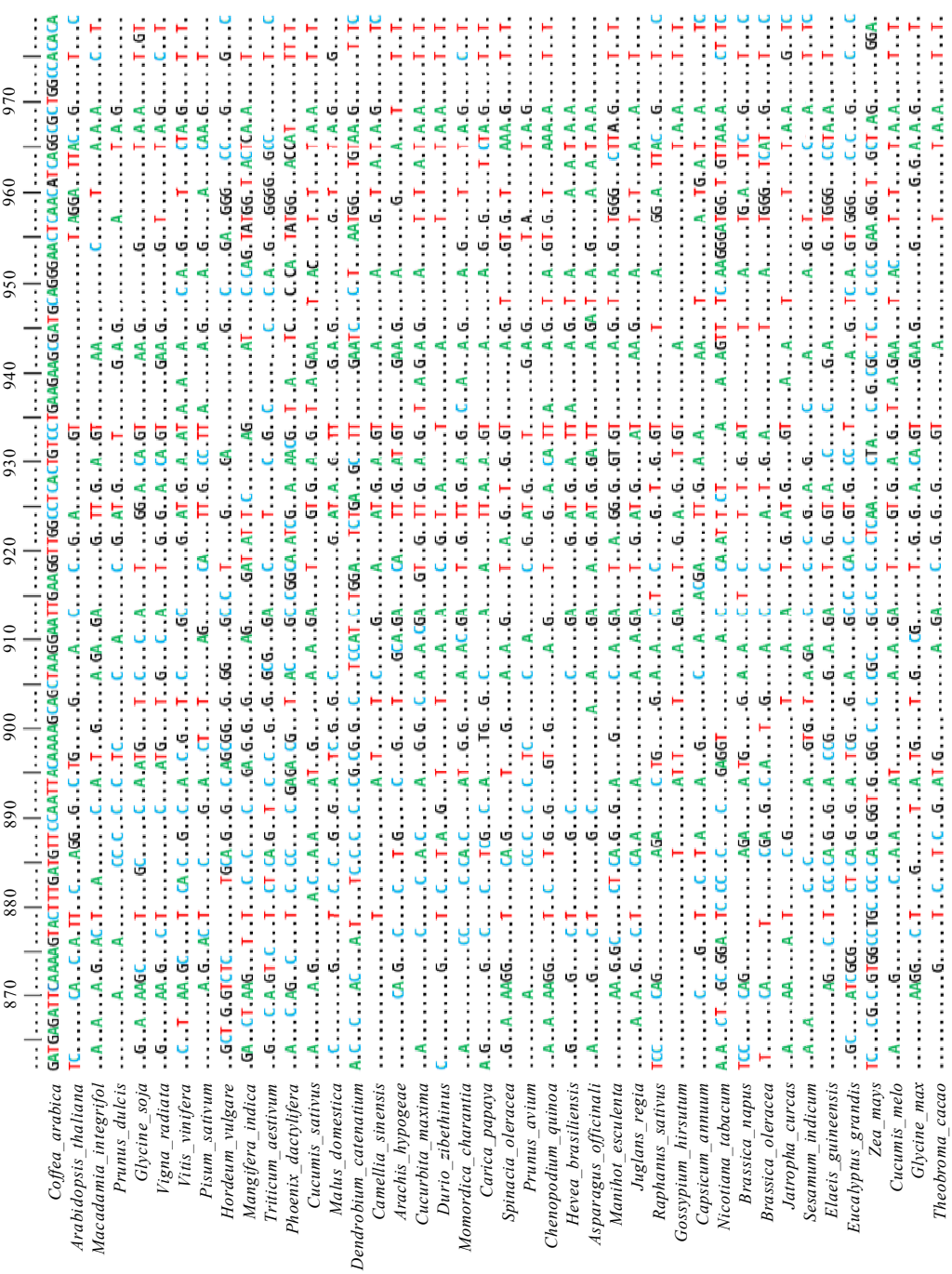

Figure 2. Polymorphism analysis of nucleotide sequences of *RKD4* gene in *C. arabica* L. and other cultivated plants using DNASP Software

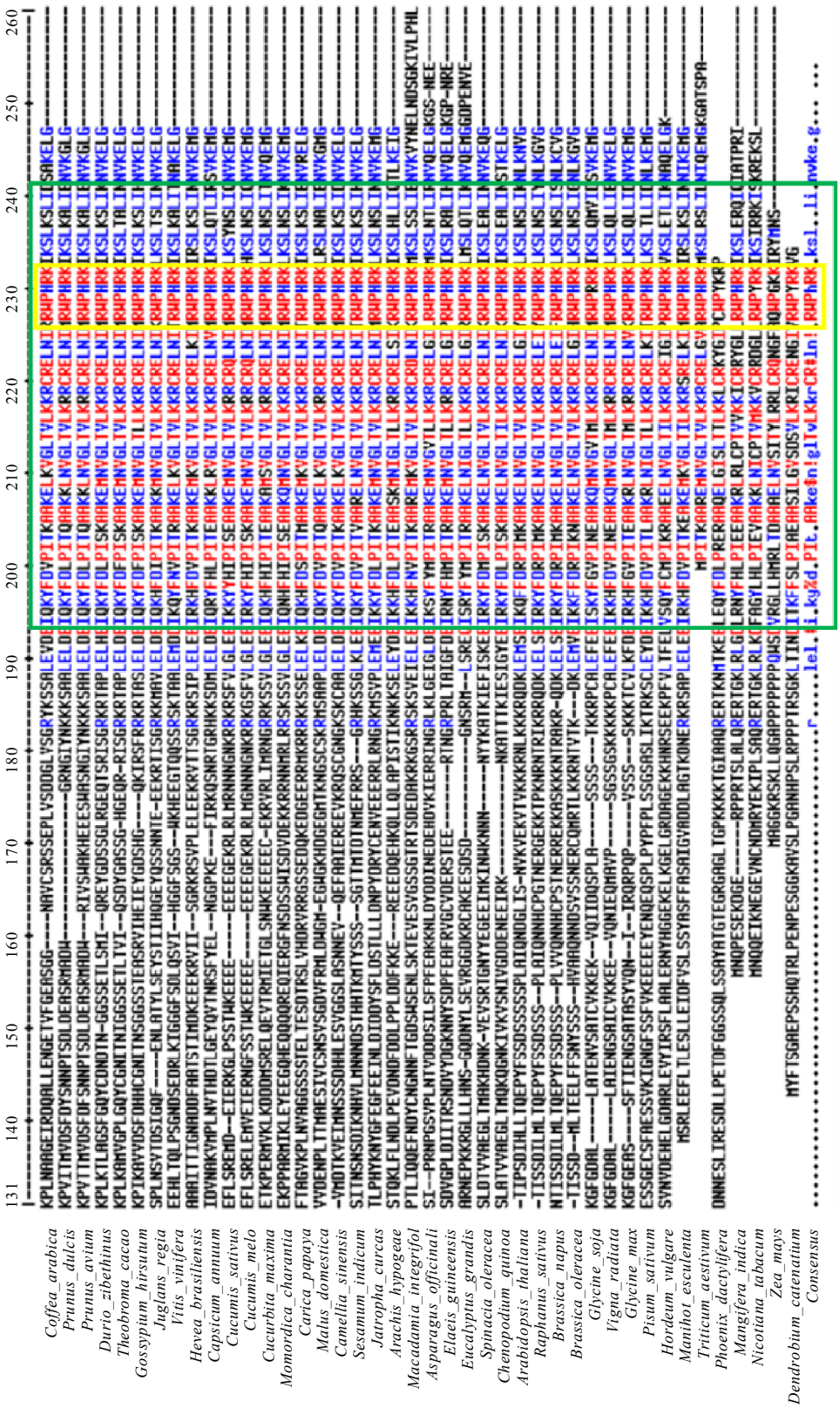

Figure 3. Amino acid sequences alignment of RKD4 protein in *C. arabica* L. and other cultivated plants

tics of a transcription factor and is capable of inducing ectopic embryo-specific genes when over-expressed in germination. Overexpression of RKD4 from primary somatic embryos is induced independently of external growth regulators. RKD4 is a novel key regulator of the early stages of plant development (Waki *et al*., 2011). The preserved parts are marked in blue and red (highly preserved). There are several motifs other than RWP-RK (yellow box) that have been successfully characterized using MEME software, marked with green boxes. Figure 3 Amino acid sequences alignment of RKD4 protein in *C. arabica* L. and other cultivated plants

Figure 4 Phylogenetic analysis, exon-intron of the *RKD4* gene and RKD4 protein motifs in *C. arabica* L. and other cultivated plants Figure 4 shows the relationship between the RKD4 protein in *C. arabica* L. and other cultivated plants, equipped with gene structures and protein motifs for each species. *C. arabica* L. has a very close relationship with grapes (*Vitis vinifera*). Meanwhile, *Arabidopsis thaliana* is closely related to *Brassica oleracea*. The structure of the *RKD4* gene shows the arrangement of exons and introns in each plant species. The RKD4 protein motif shows that there are several protein arrangement patterns in each protein sequence with the RKD4 protein. The results of RKD4 protein motif analysis show that 20 protein motifs make up the RKD4 protein sequence (Figure 5). It was found that there are similar protein regions between the *C. arabica* L. species and other cultivated plants from the twenty motifs in the RKD4 protein sequence. There are nine highly conserved motifs found in the protein alignment and their functions are shown in Table 2. The conserved amino acid sequence motifs found in the sequence alignment consist of several functions namely involved in nitrogen-controlled plant development, polysaccharide synthesis, regulating phosphorylation, expression genes in Archaea, and playing a role in signal transduction, catalyzing the formation of lysidine using lysine and ATP as substrates, as a DNA/RNA binding domain and found in various nuclei and cytoplasm, as an antitoxin and as a negative regulatory component that detects the presence of foreign objects by acting as factor binders or substrates for enzymes encoded by foreign bodies such as viral peptidases in certain Dnd systems as well as binding domains of the GAGA sequence, the main structural motif capable of binding DNA, regulating gene expression. It is important to characterize the function of a particular protein motif in a gene (Table 2). In this research, it was found that several amino acid motifs code for the same function, including being involved in plant development which is controlled by nitrogen and as a DNA/RNA binding domain and is found in various nuclei and cytoplasm.

Analysis of Physicochemical Characterization of RKD4 Protein in *C. arabica* L. and Other Cultivated Plants

Table 3 shows *C. arabica* L. has an RKD4 protein with 250 amino acids with a molecular weight of 27842.77 kDa. The highest physicochemical data for *Eucalyptus grandis* (318) and *Macadamia integrifolia* (305). All sequences are located in the nucleus.

Analysis of RKD4 Protein Structure

To understand the similarities and differences in the structure of the RKD4 protein structure in *C. arabica* L. and other cultivated plants, sequential two-dimensional protein analysis was carried out using the Chou & Fasman Secondary Structure Prediction (CFSSP) server. This server predicts the secondary structure of proteins from the amino acid sequence. On this server, the Chou & Fasman algorithm has been implemented. From these results, it can be seen the arrangement and folding of the proteins in each RKD4

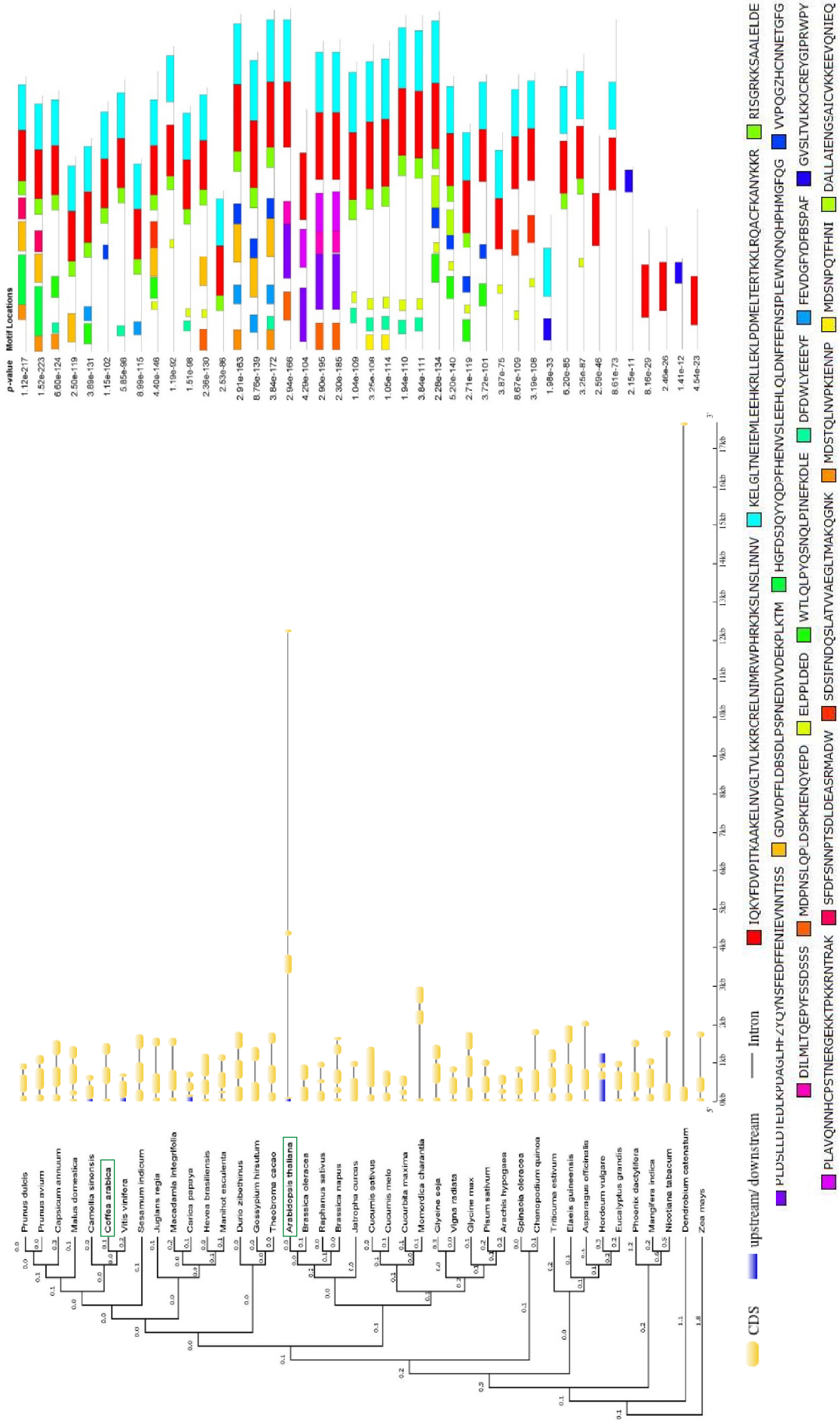

Figure 4. Phylogenetic analysis, exon-intron of the *RKD4* gene and RKD4 protein motifs in *C. arabica* L. and other cultivated plants

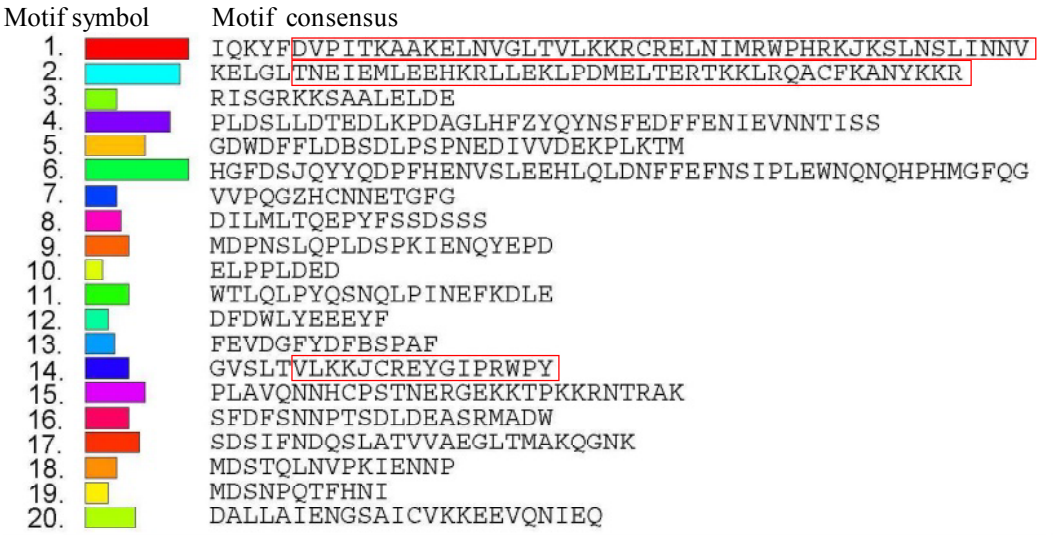

Figure 5. RKD4 protein motifs in *C. arabica* L. and other cultivated plants

Table 2. Description and function of RKD4 protein motifs in *C. arabica* L. and other cultivated plants

No.	Motifs of protein	Description	Function
1.	IQKYFDVPITKAAKELNVGLTVLKKRC RELNIMRWPHRKJKSL	PF02042, RWP-RK domain	Involved in nitrogen-controlled plant development.
2.	KELNVGLTVLKKRCRELNIMRWPHRKJK SLNSLI	PF11316, candidate of rhamnosyl transferase	Polysaccharide synthesis.
3.	TNEIEMLEEHKRLLEKLPD-MELTERTK KLRQ	PF18677, Archaellum regulatory network domain B, C-terminal	Regulates phosphorylation, gene expression in Archaea and plays a role in signal transduction.
4.	GVSLTVLKKJCREYGIPRWPY	PF02042, RWP-RK domain	Involved in nitrogen-controlled plant development.
5.	LKKJCREYGIPRW	PF11734, TilS substrate C-terminal domain	Catalyzes the formation of lysidine by using lysine and ATP as substrates.
6.	LTVLKKJCREYGIP	PF02037, SAP domain	As a DNA/RNA binding domain and is found in various nuclei and cytoplasm.
7.	TVLKKJCREYGIPR	PF18953, SAP domain containing new25	As a DNA/RNA binding domain and is found in various nuclei and cytoplasm.
8.	LTVLKKJCREYGIP	PF20306, Small protein found in certain Dnd DNA modification systems	As an antitoxin, as a negative regulatory component that detects the presence of foreign bodies by acting as a binding factor or substrate for enzymes encoded by foreign bodies such as viral peptidases in certain Dnd systems.
9.	KKJCREYGIPR	PF05225, *helix-turn-helix*, Psq domain	As a GAGA sequence binding domain, the main structural motif capable of binding DNA, regulates gene expression.

Table 3. Physicochemical characterization of the RKD4 protein in *C. arabica* L. and other cultivated plants

No.	Species	Number of AA	pI	MW	II	AI	GRAVY
1.	*Coffea arabica* L.	250	5.62	27842.77	50.82	82.76	-0.402
2.	*Arabidopsis thaliana*	256	6.72	29896.05	55.96	81.48	-0.729
3.	*Prunus dulcis*	275	5.93	32212.84	40.14	83.31	-0.522
4.	*Prunus avium*	256	5.58	29944.05	39.76	78.87	-0.636
5.	*Capsicum annuum*	261	7.07	30705.11	52.30	73.91	-0.636
6.	*Malus domestica*	195	6.00	22352.77	54.22	79.03	-0.511
7.	*Camellia sinensis*	213	7.64	24862.50	46.75	75.07	-0.635
8.	*Vitis vinifera*	272	5.67	31809.80	55.03	70.62	-0.783
9.	*Sesamum indicum*	193	9.32	22447.69	53.47	74.72	-0.780
10.	*Juglans regia*	280	5.40	32443.98	53.68	85.32	-0.591
11.	*Macadamia integrifolia*	305	5.45	35176.87	60.68	78.92	-0.683
12.	*Carica papaya*	246	8.38	28401.50	63.12	77.28	-0.783
13.	*Hevea brasiliensis*	264	8.34	30396.06	38.48	84.28	-0.579
14.	*Manihot esculenta*	160	9.59	18744.03	49.89	93.31	-0.488
15.	*Durio zibethinus*	252	5.72	28957.96	45.89	72.02	-0.616
16.	*Gossypium hirsutum*	224	7.68	26101.93	50.28	79.24	-0.497
17.	*Theobroma cacao*	255	6.03	29142.28	43.80	80.35	-0.460
18.	*Brassica oleracea*	180	6.92	20628.47	42.98	79.61	-0.708
19.	*Raphanus sativus*	232	6.47	27222.80	50.40	75.65	-0.944
20.	*Brassica napus*	233	8.20	27449.33	49.70	75.32	-0.877
21.	*Jatropha curcas*	217	5.36	25848.50	58.05	79.49	-0.811
22.	*Cucumis sativus*	230	6.78	27508.19	74.68	75.83	-0.971
23.	*Cucumis melo*	232	6.03	27590.31	75.25	82.33	-0.887
24.	*Cucurbita maxima*	251	6.24	29657.07	64.24	83.07	-0.690
25.	*Momordica charantia*	251	6.85	29690.78	66.13	75.38	-0.872
26.	*Glycine soja*	254	6.20	29106.58	56.94	82.91	-0.498
27.	*Vigna radiata*	264	6.41	30075.95	55.84	83.52	-0.414
28.	*Glycine max*	218	8.83	25089.95	60.09	83.62	-0.585
29.	*Pisum sativum*	266	5.90	31010.15	49.49	74.06	-0.647
30.	*Arachis hypogaea*	210	5.28	24612.04	56.90	85.48	-0.750
31.	*Spinacia oleracea*	266	6.18	31203.70	57.52	84.32	-0.671
32.	*Chenopodium quinoa*	268	5.75	31427.07	63.33	86.98	-0.646
33.	*Triticum estivum*	131	9.47	14949.47	69.18	78.32	-0.465
34.	*Elaeis guineensis*	271	6.22	31228.58	52.05	81.00	-0.623
35.	*Asparagus officinalis*	285	4.88	33350.64	58.33	80.04	-0.670
36.	*Hordeum vulgare*	207	9.18	23968.67	67.43	83.86	-0.616
37.	*Eucalyptus grandis*	318	5.48	35872.48	59.98	69.65	-0.677
38.	*Phoenix dactylifera*	177	5.15	19682.91	42.19	62.37	-0.803
39.	*Mangifera indica*	117	10.32	13661.84	59.61	85.98	-0.867
40.	*Nicotiana tabacum*	123	9.71	14277.67	37.60	85.69	-0.709
41.	*Dendrobium catenatum*	92	9.89	9959.38	48.11	7.39	-0.393
42.	*Zea mays*	142	7.87	15305.23	66.23	72,25	-0.633

Notes: AA: Amino Acids, MW: Molecular Weight in kDa, pI: point Isoelectric, II: Instability Index, AI: Aliphatic Index, GRAVY: Grand Average of Hydropathy.

protein sequence in *C. arabica* L. and other plants (Figure 6). There are two main structures in the RKD4 protein, namely Alpha helix, Beta sheet, and Beta turn (Figure 6). The Alpha helix protein structure is colored red, the Beta sheet protein structure is green, the Beta turn protein structure is blue and the random coil protein structure is yellow. *C. arabica* L. tends to have a greater number of Alpha helix protein structures than Beta turn and Beta sheet. Analysis of the three-dimensional structure of the RKD4 protein using Swiss Model-expasy software shows that *C. arabica* L. and other cultivated plants have protein structures that vary from those of various other plants (Figure 7).

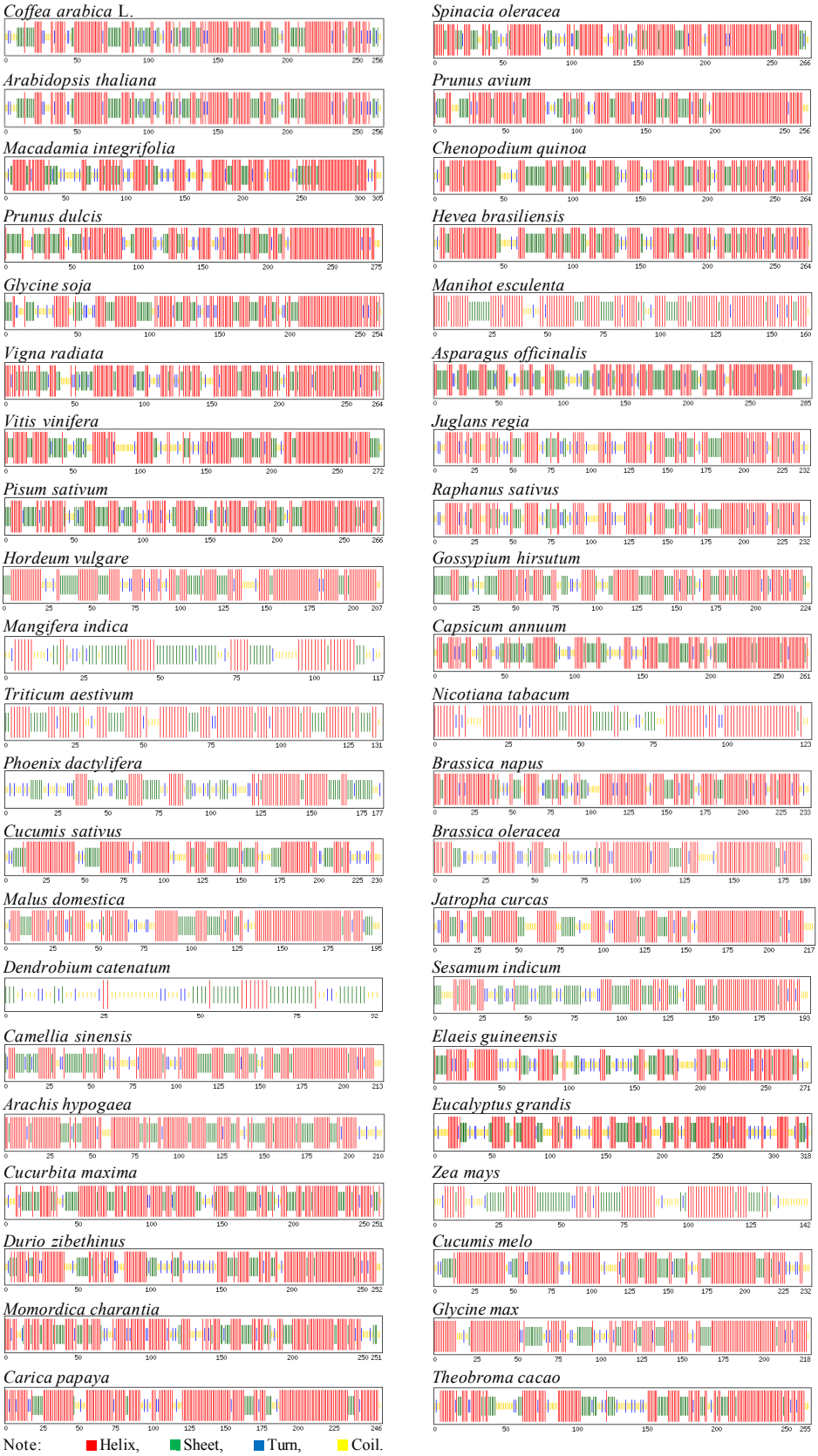

Figure 6. Two-dimensional RKD4 protein structure of *C. arabica* L. and other plants

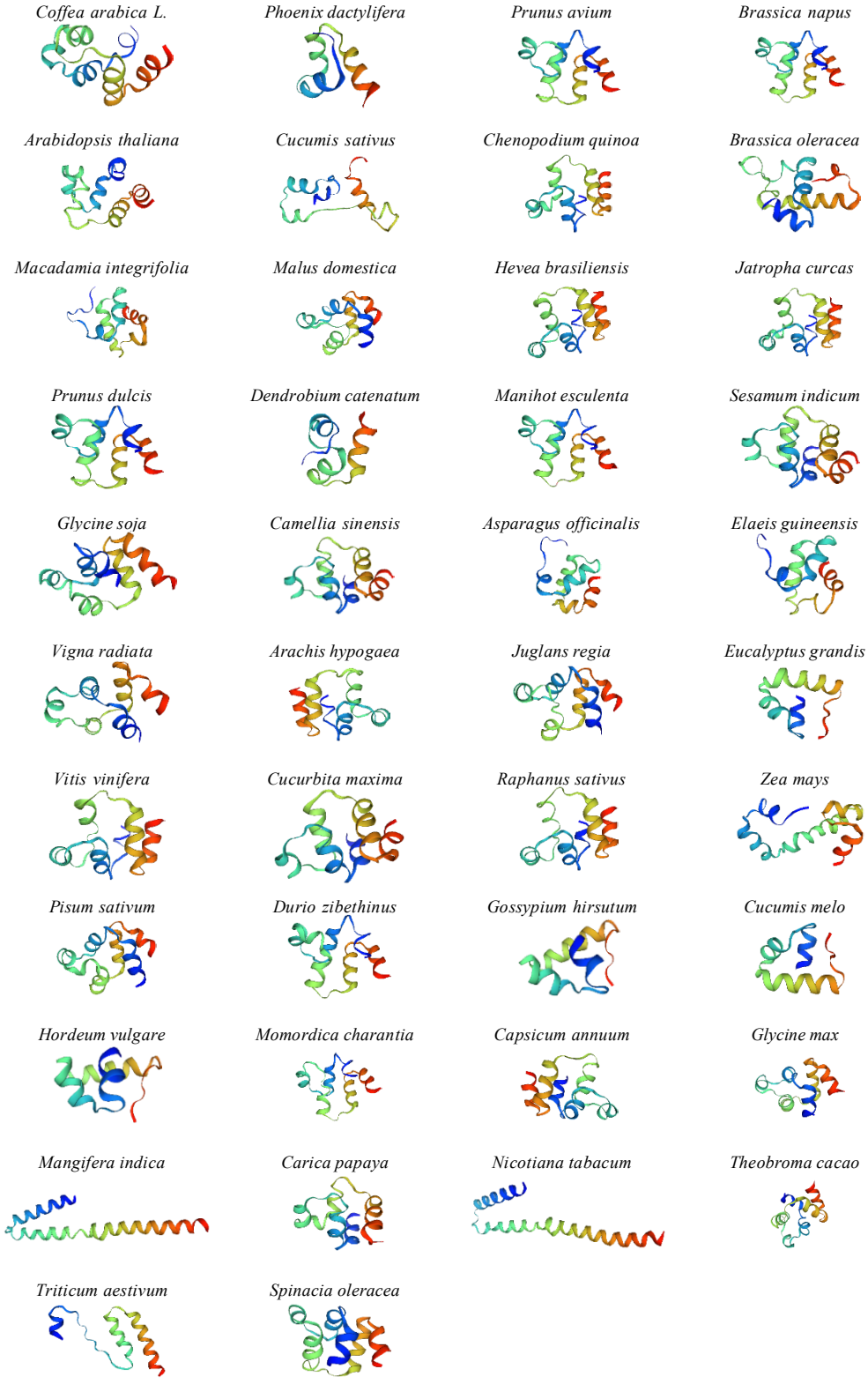

Figure 7. Three-dimensional structure of the RKD4 protein of *C. arabica* L. and other cultivated plants

Analysis of Cis-Acting Elements in Promoter Regions

The identification and characterization of the RKD4 protein containing the RWP-RK motif have improved the understanding of nitrogen response and gametophyte development in plant species (Chardin *et al*., 2014; Tedeschi *et al*., 2017). *C. arabica* L. is an important plant that globally has high economic value. Characterization and comparative analysis of the RKD4 protein from *C. arabica* L. and other cultivated plants will increase the understanding of nitrate response and the development of gametophyte regulation in these plant species. Several RWP-RK proteins have key roles in regulating the response to nitrogen availability. The nodulation-specific NIN protein is involved in nodule organogenesis and rhizobial infection under nitrogen starvation conditions. *Arabidopsis* NLP7 is a key player in the primary nitrate response. Some RKDs act as transcription factors involved in egg cell specification and differentiation or gametogenesis in algae modulated by nitrogen availability (Chardin *et al*., 2014). In this research, 42 RKD4 proteins of the *C. arabica* L. species and other cultivated plants have been identified and characterized.

Cis-Acting elements in the promoter region are responsible for modulating gene expression. To determine the potential expression response of the *RKD4* gene, cis-Acting elements have been identified in its promoter region. A total of 6 types of cis-Acting elements were found in all *RKD4* gene promoters of *C. arabica* L. and 41 other cultivated plants, including site binding-related elements, promoter-related elements, hormone-responsive elements, environmental stress-related elements, development-related elements, responsive to light (Figure 8). Promoter-related elements are the most abundant elements in all plants. This suggests that this gene functions in controlling when and where the *RKD4* gene is expressed in plants. All *RKD4* promoters contain TATA-box and CAAT-box promoter-associated elements, which are responsible for promoter function (Liu *et al*., 2020). In addition, the number and type of cis-Acting elements differed between the promoters of the *RKD4* gene in *C. arabica* L. and *Arabidopsis thaliana*, indicating the functional diversity of this gene (Figure 9).

C. arabica L. had more methyl jasmonate element responses than *A. thaliana*. Methyl jasmonate is a plant hormone that acts as an

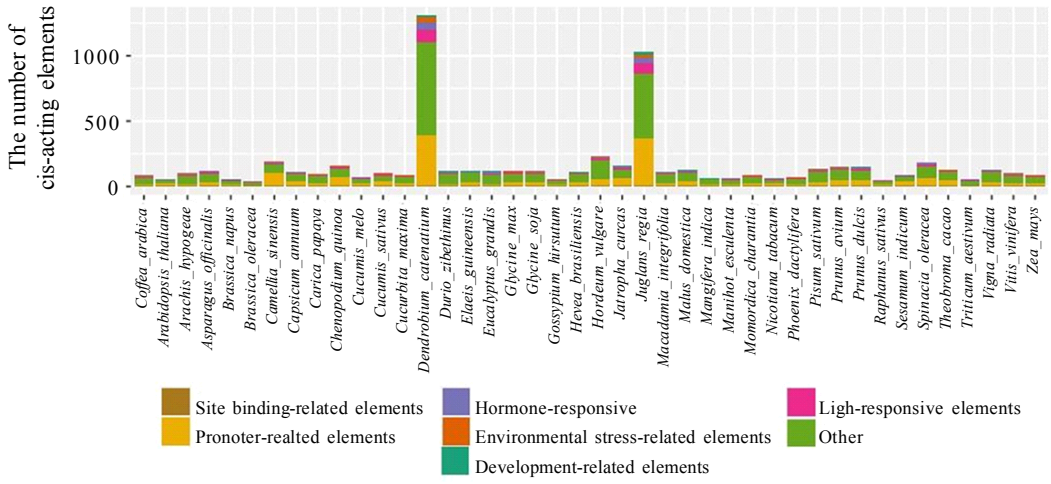

Figure 8. Number and type of *cis-Acting* elements in each *RKD4* promoter region of *C. arabica* L. and other cultivated plants

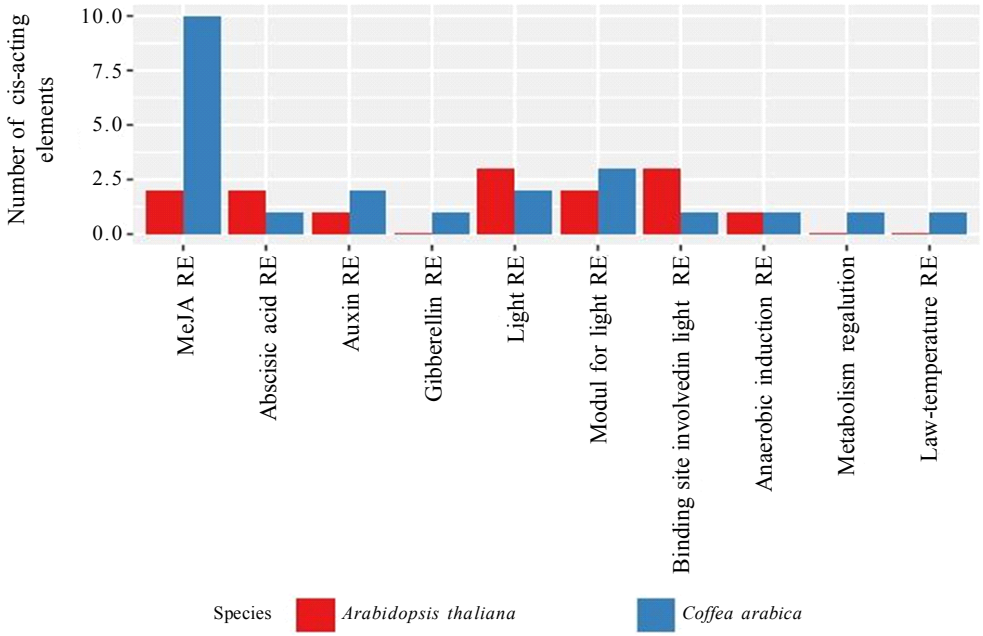

Figure 9. Analysis of *cis-Acting* elements related to development (metabolic regulation), environmental stress (anaerobic induced RE, low temperature RE), hormones (methyl jasmonate RE, abscisic acid RE, auxin RE, gibberellin RE), and light response (light RE, module light RE, sites binding involved in light RE) in the *RKD4* promoter region of *C. arabica* L. and *A. thaliana*. (RE: Element Response)

important cellular regulator mediating diverse developmental processes and defense responses (Seo *et al*., 2001). Cis-Acting elements are responsible for different expression levels of the *RKD4* gene in different tissues. Gene expression is influenced by many factors other than the presence of specific cis-Acting elements. For example, epigenetic modifications, such as DNA methylation, have substantial effects on gene expression (Zhang *et al*., 2018).

CONCLUSIONS

The characteristics of the *RKD4* gene will facilitate further studies related to biological functions for the *RKD4* family genes in *C. arabica* L. and other cultivated plants regarding the initiation key of embryo formation mechanisms during propagation through somatic embryogenesis as well as zygotic embryogenesis. There were 100 polymorphism points with a total number of mutations of 211 points. The phylogenetic tree shows *C. arabica* L. has a very close relationship with grapes (*Vitis vinifera*) based on the RKD4 protein, gene structures, and protein motifs. There are nine highly conserved motifs found in the protein alignment. *C. arabica* L. had more methyl jasmonate element responses than *A. thaliana*.

REFERENCES

Chardin, C.; T. Girin; F. Roudier; C. Meyer & A. Krapp (2014). The plant RWP-RK transcription factors: Key regulators of nitrogen responses of gametophyte development. *Journal of Experimental Botany*, 65, 5577–5587.

Debnath, S.C. (2018). *In thidiazuron: from urea derivative to plant growth regulator*

(eds. Naseem, A. & F. Mohammad). Springer, 139–158.

Elhiti, M.; C. Stasolla & A. Wang (2013). Molecular regulation of plant somatic embryogenesis. *In Vitro Cellular & Developmental Biology. Plant*, 49, 631–642.

Horstman, A.; M. Bemer & K. Boutilier (2017). A transcriptional view on somatic embryogenesis. *Regeneration,* 4, 201–216.

Hu, B.; J. Jin; A.Y. Guo; H. Zhang; J. Luo & G. Gao (2015). GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics*, 31, 1296–1297.

Koi, S.; T. Hisanaga; K. Sato; M. Shimamura; K.T. Yamato; K. Ishizaki; T. Kohchi & K. Nakajima (2016). An evolutionarily conserved plant RKD factor controls germ cell differentiation. *Current Biology,* 26, 1–7.

Kõszegi, D.; A.J. Johnston; T. Rutten; L. Altschmied; J. Kumlehn; S.E.J. Wüst; O. Kirioukhova; J. Gheyselinck; U. Grossniklaus & H. Bäumlein (2011). Members of the RKD transcription factor family induce an egg cell like gene expression program. *Plant Journal,* 67, 280–291.

Leljak-Levanic, D.; S. Mihaljevic & N. Bauer (2015). Somatic and zygotic embryos share common developmental features at the onset of plant embryogenesis. *Acta Physiologiae Plantarum,* 37, 1–14.

Loyola-Vargas, V.M. & N. Ochoa-Alejo (2016). *Somatic embryogenesis. An overview, in somatic embryogenesis. Fundamental Aspects and Applications* (Eds. V.M. Loyola-Vargas & N. Ochoa-Alejo). Cham: Springer, 1–10.

Liu, C.; D. Yuan; T. Liu; M. Xing; W. Xu; H. Zhang; H. Jin; C. Cai & S. Li (2020). Characterization and Comparative Analysis of RWP-RK Proteins from *Arachis duranensis*, *Arachis ipaensis*, and *Arachis hypogaea. International Journal of Genomics,* 2020, 119.

Mursyanti, E.; A. Purwantoro; A. Moeljopawiro & E. Semiarti (2015). Induction of somatic embryogenesis through overexpression of *AtRKD4* genes in *Phalaenopsis* "Sogo Vivien". *Indonesian Journal of Biotechnology*, 20, 42–53.

Riechmann, J.L.; J. Heard; G. Martin; L. Reuber; C. Jiang; J. Keddie; L. Adam; O. Pineda; O.J. Ratcliffe; R.R. Samaha; R. Creelman; M. Pilgrim; P. Broun; J. Z. Zhang; D. Ghandehari; B. K. Sherman; G. Yu (2000). Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. *Science*, 290, 2105–2110.

Rövekamp, M.; J.L. Bowman & U. Grossniklaus (2016). Marchantia MpRKD regulates the gametophyte–sporophyte transition by keeping egg cells quiescent in the absence of fertilization. *Current Biology,* 26, 1–8.

Rozas, J.; A. Ferrer-Mata; J.C. Sánchez-DelBarrio; S. Guirao-Rico; P. Librado; S.E. Ramos-Onsins & A. Sánchez-Gracia (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, 34, 3299–3302.

Salaün, C.; L. Lepiniec & B. Dubreucq (2021). Genetic and molecular control of somatic embryogenesis. *Plants,* 10, 1467.

Seo, H.S.; J.T. Song; J.J. Cheong; Y.H. Lee; Y.W. Lee; I. Hwang; J.S. Lee & Y.D. Choi (2001). Jasmonic acid carboxyl methyltransferase: A key enzyme for jasmonate-regulated plant responses. *Proceedings of the National Academy of Sciences,* 98, 4788–4793.

Setiari, N.; A. Purwantoro; S. Moeljopawiro & E. Semiarti (2018). Micropropagation of *Dendrobium phalaenopsis* orchid through overexpression of embryo gene AtRKD4. *AGRIVITA Journal of Agricultural Science*, 40, 284–294.

Smertenko, A. & P.V. Bozhkov (2014). Somatic embryogenesis: Life and death processes during apical-basal patterning. *Journal of Experimental Botany*, 65, 1343–1360.

Tedeschi, F.; P. Rizzo; T. Rutten ; L. Altschmied & H. Bäumlein (2017). RWP-RK domain-containing transcription factors control

cell differentiation during female gametophyte development in *Arabidopsis*. *New Phytologist,* 213, 1909–1924.

Trontin, J.F.; K. Klimaszewska; A. Morel; C. Hargreaves & M.A. Lelu-Walter (2016). *Molecular aspects of conifer zygotic and somatic embryo development: A review of genome-wide approaches and recent insights, in In Vitro Embryogenesis in Higher Plants,* (eds Germanà, M.A. & M. Lambardi). New York, NY: Springer, 167–207.

van der Vossen, H.; B. Bertrand & A. Charrier (2015). Next generation variety development for sustainable production of Arabica coffee (*Coffea arabica* L.): A review. *Euphytica,* 204, 243–256.

Waki, T.; T. Hiki; R. Watanabe; T. Hashimoto & K. Nakajima (2011). The *Arabidopsis* RWP RK protein RKD4 triggers gene expression and pattern formation in early embryogenesis. *Current Biology,* 21, 1277–1281.

Wendrich, J.R. & D. Weijers (2013). The *Arabidopsis* embryo as a miniature morphogenesis model. *New Phytologist,* 199, 14–25.

Zhang, Q.; Z. Liang ;X. Cui; C. Ji; Y. Li; P. Zhang; J. Liu; A. Riaz; P. Yao; M. Liu; Y. Wang; T. Lu; H. Yu; D. Yang; H. Zheng & X. Gu (2018). N6-Methyladenine DNA methylation in Japonica and Indica rice genomes and its association with gene expression, plant development, and stress responses. *Molecular Plant*, 11, 1492–1508.

Zimmerman, J.L. (1993). Somatic embryogenesis: a model for early development in higher plants. *Plant Cell,* 5, 1411–1423.

-o0o-