Pak. J. Bot., 59(2): February 2027.

ARTICLE | DOI: http://dx.doi.org/10.30848/PJB2027-2(1)

THE *wwe* DOMAIN OF RCD1 AS A NUCLEAR PHYLOGENETIC MARKER: A COMPARATIVE ANALYSIS WITH *rbc*L IN SELECTED LAND PLANTS

BEENA SIDDIQUA[1*], SYEDA QAMARUNNISA[2] AND RUBINA ABID[3]

[1]Department of Biotechnology. University of Karachi, Karachi 75270, Sind, Pakistan
[2]Dr A. Q. Khan Institute of Biotechnology and Genetic Engineering (KIBGE), University of Karachi, Main University Road, Karachi-75270, Sind, Pakistan
[3]Department of Botany. University of Karachi, Karachi 75270, Sind, Pakistan

*Corresponding author's email: beena.siddiqua@outlook.com

Abstract

Conserved genomic sequences are of evolutionary importance, as they give information regarding speciation and divergence of taxa. An important genomic sequence that has been used in the determination of plant taxonomic relationships is the gene for *rbc*L enzyme. With the increase in genomic data, the *rcd1* that belongs to the *sro* (Similar to RCD-One) family of genes has been found conserved within all land plants. It bears a functional *wwe* domain encoding sequence which is easily amplified and exists in low copy number per genome. The nucleotides of this domain vary between different species of a plant family while it clusters plants belonging to the same family into statistically supported clades during phylogenetic analyses. This observation prompted a need to identify the relationships it forms phylogenetically across larger variety of land plants. To achieve this objective, land plants belonging to diverse plant families were analysed using the *wwe* domain and the obtained phylogenetic patterns were compared to those obtained using *rbc*L. For these analyses, eighteen sequences were amplified using *rbc*L barcode primers and others were retrieved from online databases. The trees were constructed using *Physcomitrella patens* as outgroup as it belongs to order Funariales which is quite divergent to the angiosperms. To the best of our knowledge, a phylogenetic analysis comparing *rbc*L gene with *wwe* domain of *rcd*1 has not been reported previously. The results demonstrate a potential for using the *wwe* domain as a molecular marker in DNA taxonomic studies, as it has shown to resolve all analysed plants at familial level with high bootstrap support. However, further studies would confirm the claim.

Key words: RCD1; SRO; PARP; WWE; Poaceae; Phylogenetics; Brassicaceae; DNA barcode; Molecular Systematics

Introduction

Genes of importance for survival of living beings are found conserved within their genomes due to their evolutionary importance (Prakash, 2001). Such genes might be a part of either nuclear, mitochondrial or chloroplastic genomes. *rbc*L that encodes Large subunit of Ribulose Biphosphate Carboxylase is a widely studied gene that resolves phylogenies up to remarkable levels (Chen *et al.*, 2024; Tnah *et al.*, 2024). Similarly, other markers such as *Mat*K (Maturase-k) and *atp*B-*rbc*L intergenic spacer show no intra-species sequence variation and are reported within all plant genomes. These conserved genomic sequences are used for species identification and classification by taxonomists for DNA barcoding and molecular systematics (Erickson *et al*., 2008; Orozco-Sifuentes *et al*., 2023). Recent trend involves the exploration of multiple loci from the plant chloroplast, mitochondrial and nuclear genomes to aid in the resolution of phylogenies to a higher extent. It is believed that using a combination of markers from different subcellular genomes could aid in better understanding the link between the evolution of plants and deciphering dubious nodes in the tree of life (Young and Gillang, 2020). Another conserved gene-family across land plants is *sro* family (Similar-to-RCD1) which has 6 members (orthologues) in *Arabidopsis thaliana* (Jaspers *et al*., 2009; You *et al*., 2014). This gene-family encodes proteins that serve to regulate stress responses in plants, as well as normal developmental processes like growth, development, etc. RCD1 (Radical-induced Cell Death 1 protein) is the candidate protein which has been experimentally validated to be involved in several developmental and stress responsive pathways in various plants including the model plant for genomic-studies, *Arabidopsis thaliana* (Overmyer *et al*., 2000, Jaspers *et al*., 2009; Teotia & Lamb, 2009). RCD1 is classified as SRO type-I as it bears a WWE domain in addition to a central PARP and an N-terminal RST domain reported in other orthologues of SRO family. Along with its WWE and PARP-like domain, it also bears an intrinsically disordered region (IDR), all these components play an essential role in the overall function of this protein (Ahlfors *et al*., 2004, Jaspers *et al*., 2010a, Jaspers *et al*., 2010b, Vainonen *et al*., 2023).

Conserved across all land plants analysed to date including distant ancestral plants like *Amborella trichopoda* and *Physcomitrella patens*, *rcd1* could probably serve as a candidate gene for molecular phylogenetic and barcoding studies. However, its utility in the determination of phylogenetic relationships among plants is an area that requires investigation. Various studies have used domains of *rcd*1 and its conserved domains for evolutionary analysis (Jaspers *et al*., 2010a; Siddiqua *et al*., 2016; Christensen *et al*., 2019; Vainonen

Received: 25-08-2025 | Revised: 03-06-2026 | Accepted: 15-06-2026 | Online: 22-09-2026

et al., 2023). In a study, *wwe* was separately assessed for the analysis of domain conservation within land plants, and some other studies have used a widely conserved PARP signature domain in phylogenetic analyses (Jaspers *et al*., 2010a; Siddiqua *et al*., 2016). Using the *wwe* domain sequences for studying relationships among plants is particularly interesting owing to its less-copy number per genome, small size and easy amplification (Siddiqua *et al*., 2016). For the purpose of analysing its utility in phylogenetic studies, we amplified its sequence from some of the diverse plants and analysed the grouping patterns it forms in comparison to the popular barcode and phylogenetic marker *rbc*L gene amplified using DNA barcode primers (Kress *et al*., 2005; Kress & Erickson, 2007; Kress *et al*., 2009).

Recent advances in plant molecular research have highlighted the need for reliable nuclear markers that complement plastid genes in understanding evolutionary relationships across plant lineages (Li *et al*., 2026). Although plastid genes such as *rbc*L are widely used in plant phylogenetics, their maternal inheritance and highly conserved sequences can sometimes limit their ability to resolve evolutionary relationships. This has prompted interest in alternative nuclear regions that may capture evolutionary history more effectively and provide additional phylogenetic information (Young & Gillung, 2020). In this study, we evaluate the WWE domain of RCD1, a conserved nuclear region involved in stress responses, and compare its phylogenetic performance with the commonly used plastid marker *rbc*L across representative land plant species.

Materials and Methods

Plant growth and DNA extraction: NARC (National Agricultural Research Council, Islamabad, Pakistan) certified varieties of different plants namely: *Brassica napus* L.*, Brassica nigra* (L.) Koch*, Brassica juncea* (L.) Czern*, Brassica rapa* L. and *Arabidopsis thaliana* (L.) Heynh from Brassicaceae; *Oryza sativa* L.*, Sorghum bicolor* L. Moench*, Hordeum vulgare* L.*, Triticum aestivum* L. and *Zea mays* L. (Poaceae) *Medicago truncatula* (Fabaceae), Glycine max (Fabaceae) *Helianthus tuberosus* (Asteraceae) *Cucumis sativus* (Cucurbitaceae)*, Solanum lycopersicum* L*, Solanum pimpinellifolium* L. (Solanaceae) *Vitis vinifera* L. (Vitaceae) were grown till two-leaf stage (NARC accession numbers of all plant materials used in this study are provided in Table S1. The plants were grown in controlled environment of temperature which was set at 25℃ and 16 hour day light setting (Fig. 1). The genomic DNA was extracted from various species of plants using standard protocol employing CTAB buffer (Doyle and Doyle. 1987). The extracted DNA was suspended in 1X-Tris EDTA buffer and stored at 4°C till subsequent use. Its quality check was performed visually and using spectrophotometer by taking optical density.

Amplification of the amplicons: Amplification of the *rbc*L sequences were performed at 94°C for 4 minutes (initial denaturation), 35 cycles each consisting of 94°C for 30 seconds (denaturation), 52°C for 35 seconds (annealing) and 72°C for 1.5 minutes (extension), followed by 72°C for 10 minutes (final extension) using a PCR mix (Bioneer, inc) comprised of 1X PCR buffer, 1.5mM $MgCl_2$, 0.2mM dNTPs, 0.6µM forward and reverse primers, 1 unit of DNA Taq Polymerase, 100ng DNA template. The primers used for the amplification of *rbc*L region were the universal barcode primers (designated *rbc*La) 5'-ATGTCACC ACAAACAGAGAGACTAAAGC-3' and 5'-GTAAAAT CAAGTCCACCRCG-3' (Kress and Erickson 2007; Kress *et al*., 2009). Successfully amplified sequences were extracted using PCR-Purification Kit (Bioneer, inc), the purified DNA bands were sequenced from Bioneer, inc. The determined sequences were reassessed using triplicate replications for the sequence confirmation.

Supplementary Data:

Table S1. Germplasm Accessions of certified plant varieties from NARC, Islamabad.

No.	Plant species	Accession/Cultivar	GenBank ID
1.	*Brassica carinata*	16195-NARC	KP827653
2.	*Brassica napus*	1679-NARC	KP827651
3.	*Brassica juncea*	1664-NARC	KP827655
4.	*Glycine max*	4068-NARC	KP827665
5.	*Brassica nigra*	1508-NARC	KP827654
6.	*Brassica oleracea*	1739-NARC	KP827652
7.	*Sorghum bicolor*	23774-NARC	KP827646
8.	*Hordeum vulgare*	4064-NARC	KP827658
9.	*Triticum aestivum*	10727-NARC	KP827657
10.	*Helianthus tuberosus*	016374-NARC	GU817765.1
11.	*Cucumis sativus*	Long Green	KP827670
12.	*Zea mays*	Ageeti 2002	KP827659
13.	*Oryza sativa*	Khushbooo	KP827660
14.	*Medicago truncatula*	6475-NARC	
15.	*Arabidopsis thaliana*	Col-0	KP827656

Table S2. List of indigenous plant sequences submitted to GenBank.

Species	Sequence type	GenBank Id
Brassica napus	*rbc*L	KP827651
Brassica oleracea	*rbc*L	KP827652
Brassica carinata	*rbc*L	KP827653
Brassica nigra	*rbc*L	KP827654
Brassica juncea	*rbc*L	KP827655
Arabidopsis thaliana	*rbc*L	KP827656
Triticum aestivum	*rbc*L	KP827657
Hordeum vulgare	*rbc*L	KP827658
Zea mays	*rbc*L	KP827659
Oryza sativa	*rbc*L	KP827660
Solanum lycopersicum	*rbc*L	KP827661
Gossypium hirsutum	*rbc*L	KP827663
Vitis vinifera	*rbc*L	KP827664
Glycine max	*rbc*L	KP827665
Helianthus annus	*rbc*L	KP827666
Jatropha curcas	*rbc*L	KP827667
Nicotiana tabacum	*rbc*L	KP827669
Cucumis sativus	*rbc*L	KP827670

Table S3. Accessions of the online sequences (retrieved from NCBI GenBank).

Plant	Common name	Taxonomic family	Accessions/Data source for wwe/ RCD1 sequences	Accessions/Data source for *rbc*L sequences
Sorghum bicolor	Sorghum	Poaceae	XM_002441616.1/NCBI	AM849341.1/NCBI
Saccharum officinarum	Sugarcane	Poaceae	KP827639/NCBI	EF125120.1/NCBI
Zea mays	Maize	Poaceae	NM_001060759.1/NCBI	Z11973.1/NCBI
Setaria italica	Foxtail millet	Poaceae	XM_004951431.1/NCBI	EF125138.1/NCBI
Oryza sativa	Asian rice (Japonica)	Poaceae	NM_001063823.1/NCBI	KF731214.1/NCBI
Oryza sativa japonica	Asian rice	Poaceae	-	-
Oryza sativa indica	Asian rice	Poaceae	-	-
Triticum aestivum	Wheat	Poaceae	KP827645/NCBI	AB917066.1/NCBI
Hordeum vulgare	Barley	Poaceae	CCA61109.1/NCBI	-
Triticum urartu	Wheat ancestor	Poaceae	-	-
Oryza brachyantha	-	Poaceae	XP_006653007.1/NCBI	-
Oryza minuta	-	Poaceae	AEV41038.1/NCBI	-
Oryza officinalis	-	Poaceae	XP_002441661.1/NCBI	-
Capsella rubella	Red Shepherd's Purse	Brassicaceae	XM_006303880.1/NCBI	HM849845.1/NCBI
Arabidopsis thaliana	Thale cress	Brassicaceae	AT1G32230.1/TAIR	-
Arabidopsis lyrata	-	Brassicaceae	XM_002890929.1/NCBI	FN594842.1/NCBI
Brassica rapa	Turnip	Brassicaceae	Brara.I02610.1.p/Phytozome	JN847839.1/NCBI
Brassica napus	Canola	Brassicaceae	KG670500/NCBI	HM849821.1/NCBI
Brassica juncea	-	Brassicaceae	KP827635/NCBI	AY167979.1/NCBI
Brassica carinata	-	Brassicaceae	KP827634/NCBI	JN847835.1/NCBI
Brassica nigra	-	Brassicaceae	KP827636/NCBI	HM849822.1/NCBI
Gossypium hirsutum	Upland cotton	Malvaceae	TA26333_3635/TIGR	M77700.1/NCBI
Gossypium raimondii	Tetraploid cotton	Malvaceae	TA9990_29730/TIGR	JQ034250.1/NCBI
Theobroma cacao	cocoa tree	Malvaceae	Thecc1EG016592t4/GreenPhyl	AF022125.1/NCBI
Citrus sinensis	Sweet orange	Rutaceae	XM_006471451.1/NCBI	AB505951.1/NCBI
Selaginella moellendorfii	-	Selaginellaceae	-	AF093256.1/NCBI
Ricinus communis	Castor Bean	Euphorbiaceae	XP_002518985.1/NCBI	AB233871.1/NCBI
Jatropha curcas	Physic nut	Euphorbiaceae	Jcr4S06073.40/Kasuza Genome Resource	-
Manihot esculenta	Cassava	Euphorbiaceae	cassava4.1_006711m_MANES/Phytozome	AB233880.1/NCBI
Populus trichocarpa	Black cottonwood	Salicaceae	XM_002303403.2/NCBI	-
Populus tremula	-	Salicaceae	-	AJ418827.1/NCBI
Linum usitatissimum	Flax	Linaceae	Lus10030991/Phytozome	JX664057.1/NCBI
Prunus persica	Peach	Rosaceae	EMJ15994.1/NCBI	AF411493.1/NCBI
Malus domestica	Apple	Rosaceae	cMDP0000234325_MALDO/GreenPhyl	-
Fragaria vesca	Wild strawberry	Rosaceae	XM_004300155.1/NCBI	-
Medicago truncatula	Barrel medic	Fabaceae	XM_003621248.1/NCBI	-
Cicer arietinum	Chickpea	Fabaceae	XM_004489700.1/NCBI	AF308707.1/NCBI
Glycine max	Soybean	Fabaceae	XM_003516930.2/NCBI	Z95552.1/NCBI
Phaseolus vulgaris	Common Bean	Fabaceae	PHAVU_002G137600g/Phytozome	KF022497.1/NCBI
Nicotiana tabacum	Cultivated tobacco	Solanaceae	DV159437/TIGR	KC825342.1/NCBI
Nicotiana benthamiana	-	Solanaceae	-	-
Nicotiana tomentosiformis	Wild tobacco	Solanaceae	-	KM025251.1/NCBI
Solanum lycopersicum	Tomato	Solanaceae	XM_004245886.1/NCBI	KJ652188.1/NCBI
Solanum tuberosum	Potato	Solanaceae	XM_006365907.1/NCBI	M76402.1/NCBI
Helianthus annuus	Sunflower	Asteraceae	TA16125_4232	AY215124.1/NCBI
Helianthus tuberosus	Jerusalem artichoke	Asteraceae	TA1763_4233/TIGR	GU817765.1/NCBI
Helianthus ciliaris	-	Asteraceae	TA598_73280/TIGR	-
Helianthus paradoxus	Paradox sunflower	Asteraceae	TA3045_73304/TIGR	-
Helianthus argophyllus	Silverleaf sunflower	Asteraceae	TA1408_73275/TIGR	-
Helianthus petiolaris	Prairie Sunflower	Asteraceae	TA1528_4234/TIGR	-
Cichorium intybus	Chicory	Asteraceae	TA853_13427/TIGR	HM849895.1/NCBI
Lactuca sativa	Lettuce	Asteraceae	BQ858405/TIGR	L14073.1/NCBI
Taraxacum officinale	Dandelion	Asteraceae	-	AY395562.1/NCBI
Physcomitrella patens	-	Funariaceae	-	Phpat.016G023800.1/Phytozome

Sequence retrieval and phylogeny reconstruction: Additionally, some sequences were retrieved from the Phtozome and NCBI (National Center for Biotechnology Information) using (Basic Local Alignment Search Tool) search tool (Altschul *et al*., 1997) following the default settings and selecting single highest-scoring sequence from each unique plant species. The data-search was conducted in the NCBI dataset "RefSeq Genome Database" and the presence of Open Reading Frames was performed using "ORF prediction tool" (Goodstein *et al*., 2012; Rombel *et al*., 2002), the domains were searched in the sequences using CD-search from the "Conserved Domains Database" (Marchler-Bauer *et al*., 2005). The phylogenetic reconstructions were performed using BEAST v1.8.0 (Drummond *et al*., 2012) alongside BEAUTi v1.8.0. Uncor-related Log-normal relaxed clock model was used with Yule Model of tree priors. The MCMC operation was conducted with logP of 200 and chain length of 800000 for reconstructions. Model for such analyses was determined using the software name "JmodelTest" (Darriba *et al*., 2012). The consensus trees were saved and visualized using FigTree version 1.4 (Drummond *et al*., 2012). The sequence data used in the construction involving plants from diverse families due to wide availability of sequences from the NCBI database (Table S3).

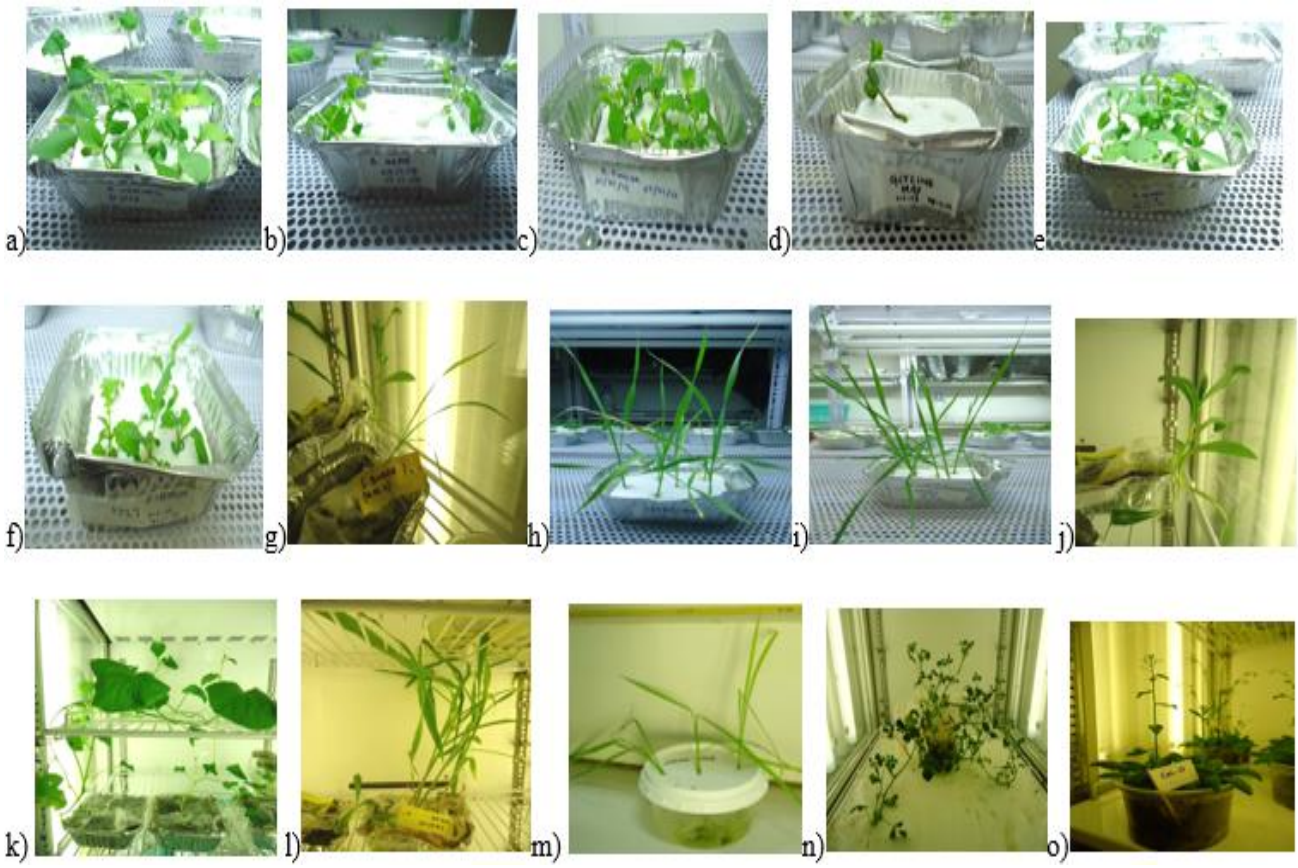

Fig. 1. Plants grown under ideal conditions for DNA extraction; (a) *Brassica carinata* (b) *Brassica napus* (c) *Brassica juncea* (d) *Glycine max* (e) *Brassica nigra* (f) *Brassica oleracea* (g) *Sorghum bicolor,* (h) *Hordeum vulgare* (i) *Triticum aestivum* (j) *Helianthus tubersosus* (k) *Cucumis sativus* (l) *Zea mays* (m) *Oryza sativa* (n) *Medicago truncatula and* (o) *Arabidopsis thaliana.*

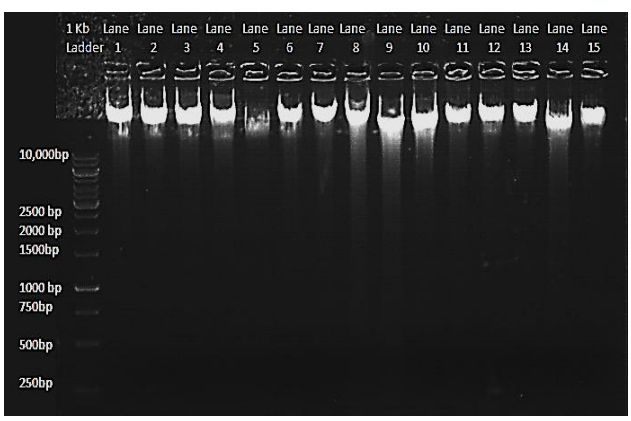

Fig. 2. DNA Extraction from *Brassica oleracea* (Lane 1), *Brassica carinata* (Lane 2), *Brassica juncea* (Lane 3), *Brassica nigra* (Lane 4*), Brassica napus* (Lane 5), *Oryza sativa* (Lane 6), *Sorghum bicolor* (Lane 7), *Zea mays* (Lane 8), *Helianthus tuberosus* (Lane 9), *Jatropha curcas* (Lane 10), *Arabidopsis thaliana* (Lane 11), *Cucumis sativus* (Lane 12), *Hordeum vulgare* (Lane 13), *Triticum aestivum* (Lane 14), *Medicago truncatula* (Lane 15).

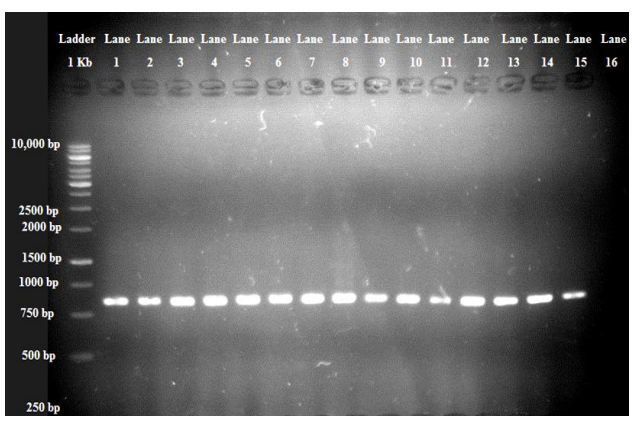

Fig. 3. Amplification of *rbc*L Sequences from Crops*: rbc*L sequences amplified as 800 bp. fragments. Lane 1: *Brassica oleracea*. Lane 2: *Brassica carinata*, Lane 3: *Brassica juncea*, Lane 4: *Brassica nigra*, Lane 5: *Brassica napus*, Lane 6: *Arabidopsis thaliana*, Lane 7: *Hordeum vulgare*, Lane 8: *Triticum aestivum*, Lane 9: *Saccharum officinarum*, Lane 10: *Hordeum vulgare*, Lane 11: *Oryza sativa*, Lane 12: *Sorghum bicolor*, Lane 13: *Zea mays*, Lane 14: *Brassica oleracea*, Lane 15: *Brassica nigra*, Lane 16: No template blank.

Results

Fresh leaves from various certified plant varieties (Table S1) grown under controlled conditions (Fig. 1) were successfully used for DNA extraction (Fig. 2). From this DNA; sequences of the *rbc*L were successfully amplified (Fig. 3). The sequences were submitted to Genbank and accessions issued are enlisted in Table S2. The confirmation of their identity *in-silico* was made by performing similarity searches in the NCBI-BLAST which was also used for retrieving closely matching orthologous sequences that belonged to the species of plants other than the amplified ones for use in tree construction.

However, a closer analysis of each constructed small family tree constructed using sequences from dicot family Brassicaceae and monocot family Poaceae depicts

differences in the arrangement of plants. This is important as the arrangement and cladding pattern must be supported morphologically to be of relevance. In reference to Poaceae, the sub-family Pooideae has two major clades differentiated based on the photosynthetic pathway employed. BOP clade was seen intact in *rbc*L plants, however, Andropogeneae (Sorghum and Maize) is seen intact only in the *wwe*-based tree while they occupy a closer position in the *rbc*L based tree. Triticeae (*Triticum aestivum* and *Hordeum vulgare*) is resolved with 1 bootstrap support in *rbc*L while *wwe* marker places both these species nearer to each other but separates *Hordeum vulgare* at a separate speciation step with *Oryza sativa*. Upon a closer examination of the Brassicaceae family, it can be seen that plants of Brassica genus (*Brassica napus, Brassica carinata, Brassica nigra* and *Brassica juncea*) have been resolved distinctly in both *rbc*L and *wwe*-based trees. Best resolution of the Brassicaceae has been observed with *rbc*L marker that has clustered all members of Brassica genus in a distinct clade with the highest BS of 1.

Phylogenetic trees reconstructed using WWE domain sequences and rbcL sequences are presented in Figs. 4 and 5, respectively. In the phylogenetic constructed using sequence data from diverse families, representing wide taxa had revealed that the complete sequence of SRO type I as well as of *wwe* groups plants into their respective families with a high bootstrap support (0.9-1) as seen with the plant families Asteraceae, Fabaceae, Brassicaceae, Solanaceaea, Malvaceae and Rutaceae. This trend was similar to the trend observed in case of *rbc*L sequences had also grouped the plants into their respective families with a high bootstrap support. Another important trend in all trees was that the plants of monocot family Poaceae were separated as a separate group from the other families that belong to the dicot group of plants with high bootstrap supports i.e. BS=1 in case of *wwe* tree and BS=0.95 in case of *rbc*L tree (Figs. 4 and 5, respectively). This fact, coupled with the small size and easy amplification makes *rbc*L and *wwe* better markers for phylogenetic reconstruction as compared to *rcd*1, which has a larger size and certain sites that may not be informative. When *Amborella trichopoda* was placed in the analysis, it was separated from the whole angiospermae clade as a distinct subgroup. This trend is supported morphologically as dicots have various features like fruit, flowers anatomically different from Poaceae. While, *Amborella trichopoda* belongs to Amborellaceae family which has morphological features particularly distinct from the monocots and dicots and is believed to the ancestral plant of all angiosperms. *Physcomitrella patens* a plant belonging to the Bryophitina division was separated and treated as outgroup to the remaining plants of the Angiospermae division in all the major trees constructed using various markers tested.

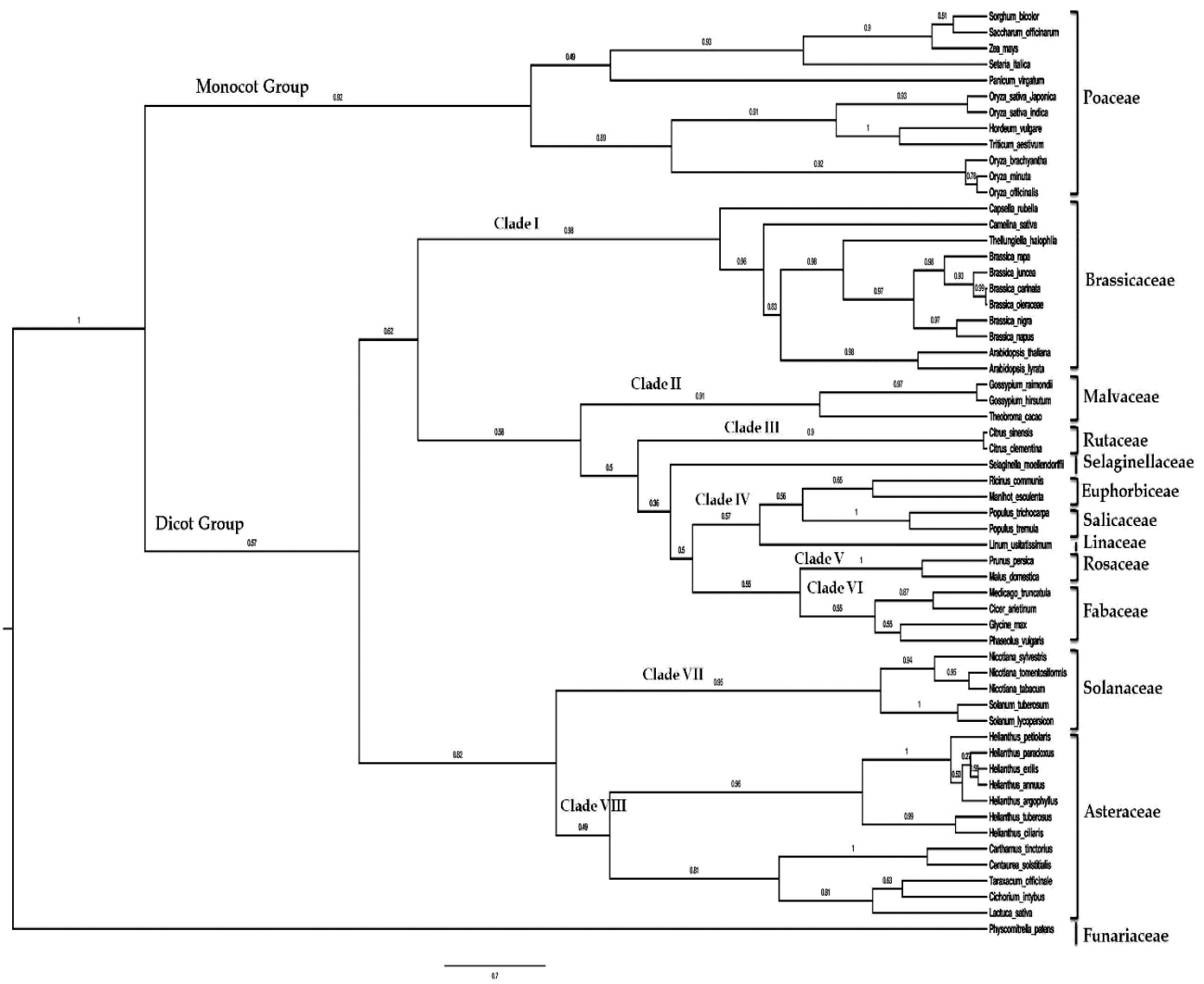

Fig. 4. Phylogenetic tree of the *wwe* sequences from multiple plant families.

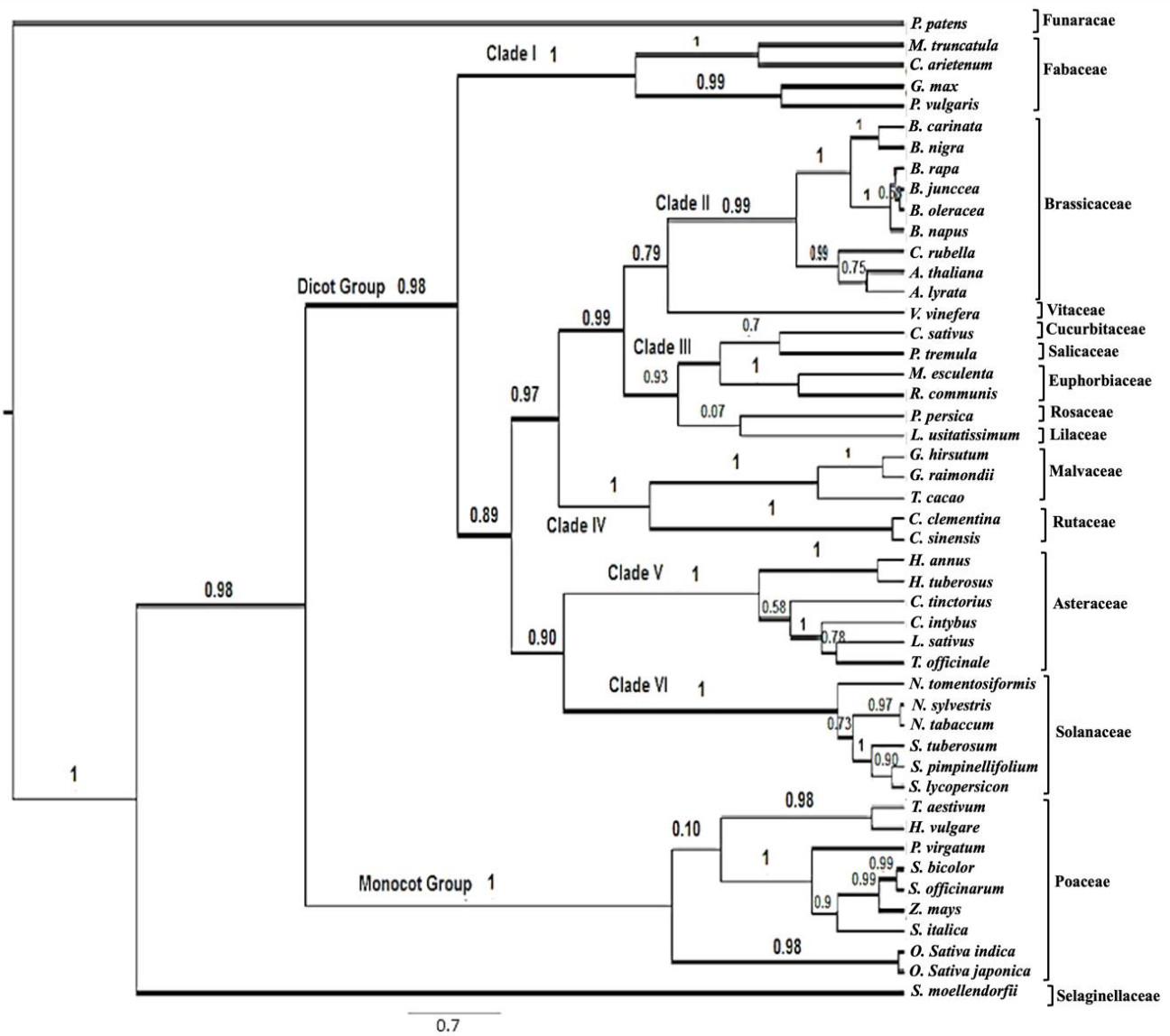

Fig. 5. Phylogenetic tree of the *rbc*L sequences from multiple plant families.

Discussion

Genes of house-keeping nature are highly conserved within the genomes of the members of a kingdom and evolve slowly as compared to tissue-specific genes (Joshi *et al.,* 2022). Plant genomes have also retained such genes that play key developmental roles and are important for the survival of plants despite domestication pressures (Meyer and Purugganan, 2013; Wendel *et al*., 2016). Such sequences can be used for the analysis of relationships among the different genera. Within the kingdom Plantae, sequences belonging to nuclear, mitochondrial and chloroplastic genomes have assisted in the studies of molecular systematics and DNA barcoding with the help of their conserved and variable sites. The inferred relationships are considered successful if the clades thus formed are supported with high values of probability and the morphological evidences of similarity (Orozco-Sifuentes *et al*., 2023). Over time, it has been found that a combination of such conserved regions from the nuclear (18S rDNA) and chloroplastic genomes (*rbc*L, *mat*K (megakaryocyte-associated tyrosine kinase) genes gave a substantially good approximation of their divergence of species over time. The importance of nuclear genes cannot be ruled out as it is bi-parentally inherited while the chloroplastic genome is maternally inherited. Due to the maternal inheritance of chloroplastic genome and hybridization at times, the inferred phylogenies using a chloroplast region alone may not be true. The support data provided through nuclear genes thus enhances the resolution and internal support for studying plant evolutionary pathways. (Morton, 2011).

The *wwe* sequences are from the nuclear genome that have important regulatory roles and have shown conservation of several important sites within the plant families (Jaspers *et al*., 2010a; Mur *et al*., 2006; Siddiqua *et al*., 2016). To depict this fact on the level of molecular phylogeny, the trees constructed using *wwe* sequences were compared to the trees obtained using sequences of *rbc*L gene sequences. The *rbc*L and *wwe* phylogenies were well-resolved with various plant families forming highly supported independent clades. Within Poaceae, the BOP clade (*Oryza*, *Triticum*, *Hordeum*) and Andropogoneae clade (*Sorghum, Zea)* are resolved in both wwe and *rbc*L trees. Within Poaceae, the BOP clade (Oryza, Triticum and

Hordeum) and the Andropogoneae clade (Sorghum and Zea) were resolved in both WWE- and rbcL-based phylogenies. The WWE tree separated Hordeum vulgare slightly from Triticum aestivum, whereas the rbcL tree grouped both species together with a bootstrap support of 0.98, reflecting their close taxonomic relationship. The whole gene *rcd*1 has several uninformative sites that could obscure clear phylogenetic relationship determination. However, this trend was not seen in the phylogenetic trees of *wwe* sequences. Thus showing support for the fact that the *wwe* also evolves uniformly over time, as does the popular *rbc*L gene that is widely used in the phylogenetic reconstruction studies. The *wwe* effectively resolving the plant families Brassicaceae, Fabaceae, Solanaceae, Asteraceae and Poaceae into strongly supported independent clades. Same clades were also resolved by the *rbc*L gene based trees, strengthening the point of view that the *wwe* sequence being of an appropriately small length could potentially be used for the assessment of relationships amongst plant genera at familial level. (Siddiqua *et al*., 2016). It was also noted that when a sequence from *Amborella trichopoda* was included in the analysis, it was separated from the angiospermae clade and was placed near *Physcomitrella patens* which was used as outgroup in the current analyses. Looking at the scale of time, this separation may be justified with the help of evolutionary data that suggests *Amborella trichopoda* as an ancestral plant to other angiosperms (Soltis *et al*., 2013). It might also be linked to its divergence from the modern plants that have evolved over time through the practices of selective breeding. The plant species *Physcomitrella patens* has been used as outgroup for training data in the phylogenetic reconstructions, as it belongs to the division Bryophyta, which is believed to be one of the first land plants. Hence, it is then sufficiently divergent from other plants used in these analyses. The evolution of monocots from their dicot ancestors, which is believed to date back to around 200 million years back, is however obscured by a lack of fossil record (Wolfe *et al.,* 1989). In both *wwe* and *rbc*L phylogenies, dicot group is placed closer to Funariales showing its evolution earlier than monocots.

Evolutionary patterns observed at the level of *rbc*L shows *Arabidopsis thaliana* at a separate lineage to Brassica genus is most appropriate when making assessment with regard to the morphological data (He *et al*., 2021). The Brassica genus in the *wwe* based trees is scattered on two branches, however, it forms an independent clade using the *rbc*L gene sequences. “Triangle of U” explains relationships between Brassica species in terms of evolution, where *Brassica nigra* (genome designated as BB) and *Brassica oleracea* (CC) are the ancestors of *Brassica carinata* (designated BBCC), while *Brassica nigra* (BB) and *Brassica rapa* (AA) are ancestors of *Brassica junccea* (AABB) and *Brassica rapa* (AA) and *Brassica oleracea* (CC) are ancestors of *Brassica napus* (AACC) (Nagaharu, 1935). The *wwe* based tree places AACC and AA in close proximity and shows them evolving at a separate incident to *Arabidopsis thaliana*. On this level, the best interpretation of the actual relationship comes from the *rbc*L-based tree, followed by the *wwe*-based tree. Based on the triangle of U, where the genomes of *Brassica nigra* (BB) and *Brassica oleracea* (CC) were combined to give rise to the hybrid variety *Brassica carinata* (BBCC) genome of 17 chromosomes (Murat *et al*., 2015; Nagaharu, 1935). This when represented on the level of a tree should ideally put BBCC close to BB and CC genomes. On the level of *wwe* based tree, BBCC is indeed placed between BB and CC genomes, while BB is at the upper position and CC is at lower position. On the other hand, *rbc*L based tree shows the genomes BB and BBCC evolved independently at a separate lineage while CC evolved separately and placed the genomes AA and AABB on the two sides of the CC genome. At this level, we find *wwe*-based tree better at resolving the relationship among the species of Brassica genus in comparison to the *rbc*L based tree. However, the evolutionary pattern of AACC being a hybrid of AA and CC is best represented by the *rbc*L based tree as compared to the the *wwe*-based tree. The *rbc*L based tree places these three genomes in close proximity, while the *wwe*-based tree shows the BB genome in a separate lineage to the genome of its hybrid AACC. On the level of the monocot family Poaceae, *H. vulgare* and *T. aestivum* (both belong to the sub-family Pooideae) form a separate clade with BS=1 (Fig. 4). This is in accordance to the data gathered and analysed using a combination of different marker genes by grass phylogeny working group (Barker *et al*., 2001). While the *rbc*L based tree also grouped *H. vulgare* and *T. aestivum* of the sub-family Pooideae in a single clade with BS of 0.98 (Fig. 5). This shows support for the fact that resolution of plants is also observed at the level of certain subfamilies.

Analysis performed using universal trees, where *wwe* sequences from all the plant taxa were combined, has revealed that the *wwe* domain have evolved uniformly over the course of time. It has grouped the individual plant taxa into the clades represented by their families. These plant families have historically been formed by the plant taxonomists after studying the details of each individual plant’s morphology. Among these sequences, the *wwe*-based tree gives a clear picture of a speciation event that separated the monocots and dicots into separate branches of life. This dichotomy is apparent in the angiosperms as dicots bear fruits with two cotyledons while monocots have a single cotyledon in their fruits. Clearly, the plants belonging to monocot and dicot groups are distinct from each other by the morphology of their seeds (with dicots bearing two cotyledons and monocots bearing single cotyledon). *wwe* has a smaller size (in the order of several bp), conserved status across the kingdom Plantae, low copy number and a prediction efficiency comparable to the *rbc*L sequences. Another disambiguate was with a member of the family Asteraceae, *Lactuca sativa*, which was present within the Asteraceae clade in the wwe-based tree. The *wwe* tree groups the members of the sub-families Nicotianoideae and Solanoideae within sub-clades, giving a better approximation of their morphological relationship. This suggests *wwe* to be more appropriate for the phylogenetic analysis than complete gene sequences, as the complete gene sequence might be long enough to harbour several informatively misleading sites. Hence, this research is suggestive of the fact that the *wwe* can be effectively used as a candidate with the combination of different marker genes for making high-precision assessment of phylogenetic relationships at both familial as well as sub-familial levels.

In addition to its methodological value, this study contributes to broader discussions in plant evolutionary biology, ecology, and conservation. The recovery of phylogenetic relationships that agree with established evolutionary frameworks using the WWE domain of *RCD1* highlights the usefulness of nuclear genes associated with stress and defense responses in studying plant diversification. Because *RCD1* plays a role in redox signaling and stress-induced cell death, variation within its conserved WWE domain may reflect long-term evolutionary responses to environmental pressures that have shaped plant adaptation to different habitats. From an ecological perspective, the ability of this domain to resolve relationships across diverse plant lineages supports its application in comparative studies linking evolutionary history with ecological traits such as stress tolerance and habitat preference. In terms of conservation, the use of additional nuclear markers can improve phylogenetic resolution, helping to identify evolutionarily distinct lineages and hidden diversity, particularly in non-model or poorly studied plant groups. Together, these findings suggest that the WWE domain of *RCD1* can serve as a useful complementary marker for integrating evolutionary, ecological, and conservation-based research in plants.

Competing Interests: The authors have no relevant financial or non-financial interests to disclose.

Author's Contributions: BS conducted lab experiments, analysed the results and wrote the manuscript; SQN supervised the research, prepared phylogenetic trees and helped in result analysis; RA rectified the work with useful insights.

References

Ahlfors, R., S. Lang, K. Overmyer, P. Jaspers, M. Brosche, A. Tauriainen, H. Kollist, H. Tuominen, E. Belles-Boix, M. Piippo, D. Inze, E. T. Palva and J. Kangasjärvi. 2004. Arabidopsis radical-induced cell death-1 belongs to the WWE protein–protein interaction domain protein family and modulates abscisic acid, ethylene, and methyl jasmonate responses. *Plant Cell*, 16: 1925-1937.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.*, 25: 3389-3402.

Barker, N.P., L.G. Clark and J.I. Davis. 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). *Ann. Missouri Bot. Gard.*, 88: 373-457.

Chen, Z., L. Gao, H. Wang and S. Feng. 2024. Molecular identification and phylogenetic analysis of *Cymbidium* species (Orchidaceae) based on the potential DNA barcodes *matK*, *rbcL*, *psbA–trnH* and internal transcribed spacer. *Agronomy*, 14: 933.

Christensen, L.F., L. Staby, K. Bugge, C. O'Shea, B.B. Kragelund and K. Skriver. 2019. Evolutionary conservation of the intrinsic disorder-based radical-induced cell death1 hub interactome. *Sci. Rep.*, 9: 18927.

Darriba, D., G. L. Taboada, R. Doallo and D. Posada. 2012. jModelTest 2: more models, new heuristics and high-performance computing. *Nat. Methods*, 9: 772.

Doyle, J.J. and J.L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.*, 19: 11-15.

Drummond, A.J., M.A. Suchard, D. Xie and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.*, 29: 1969-1973.

Erickson, D.L., J. Spouge, A. Resch, L.A. Weigt and J.W. Kress. 2008. DNA barcoding in land plants: developing standards to quantify and maximize success. *Taxon*, 57: 1304-1316.

Goodstein, D.M., S. Shu, R. Howson, R. Neupane, R.D. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten and D.S. Putnam. 2012. Phytozome: A comparative platform for green plant genomics. *Nucl. Acids Res.*, 40: 1178-1186.

He, Z., R. Ji, L. Havlickova, L. Wang, Y. Li, H.T. Lee, J. Song, C. Koh, J. Yang, M. Zhang, I.A. P. Parkin, X. Wang, D. Edwards, G.J. King, J. Zou, K. Liu, R.J. Snowdon, S.S. Banga, I. Machackova and I. Bancroft. 2021. Genome structural evolution in *Brassica* crops. *Nat. Plants*, 7: 757-765.

Jaspers, P., K. Overmyer, M. Wrzaczek, J.P. Vainonen, T. Blomster, J. Salojärvi, A.R. Ramesha and J. Kangasjärvi. 2010a. The RST and PARP-like domain containing SRO protein family: analysis of protein structure, function and conservation in land plants. *BMC Genom.*, 11: 170.

Jaspers, P., M. Brosché, K. Overmyer and J. Kangasjärvi. 2010b. The transcription factor interacting protein RCD1 contains a novel conserved domain. *Plant Signal. Behav.*, 5: 78-80.

Jaspers, P., T. Blomster, M. Brosché, J. Salojärvi, R. Ahlfors, J.P. Vainonen, A.R. Reddy, R. Immink, G. Angenent, F. Turck, K. Overmyer and J. Kangasjärvi. 2009. Unequally redundant RCD1 and SRO1 mediate stress and developmental responses and interact with transcription factors. *Plant J.*, 60: 268-279.

Joshi, C.J., W. Ke, A. Drangowska-Way, E.J. O'Rourke and N.E. Lewis. 2022. What are housekeeping genes? Re-evaluating their definition, characteristics and evolutionary conservation. *Trends Gen.*, 38: 624-637.

Kress, W.J. and D.L. Erickson. 2007. A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH–psbA* spacer region. *PLoS One*, 2: e508.

Kress, W.J., D.L. Erickson, F.A. Jones, N.G. Swenson, R. Perez, O. Sanjur and E. Bermingham. 2009. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proc. Natl. Acad. Sci. USA*, 106: 18621-18626.

Kress, W.J., K.J. Wurdack, E.A. Zimmer, L.A. Weigt and D.H. Janzen. 2005. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. USA*, 102: 8369-8374.

Li, X., Z. Hu, W. Yu, H. Xie, X. Wang, P. Huang, X. Zhang, J. Yang, Y. Li, W. Zhao, X. Yang, Z. Zhang, W. Pu and Z. Li. 2026. Advances and challenges in plant molecular marker technologies and their applications in the artificial intelligence empowered era. *Front. Plant Sci.*, 16: 1757949.

Marchler-Bauer, A., J.B. Anderson, P.F. Cherukuri, C. DeWeese-Scott, L.Y. Geer, M. Gwadz, S. He, D.I. Hurwitz, J.D. Jackson, Z. Ke, C.J. Lanczycki, C.A. Liebert, C. Liu, F. Lu, G.H. Marchler, M. Mullokandov, B.A. Shoemaker, V. Simonyan, J.S. Song, P.A. Thiessen, R.A. Yamashita, J.J. Yin, D. Zhang and S.H. Bryant. 2005. CDD: a conserved domain database for protein classification. *Nucl. Acids Res.*, 33: 192-196.

Meyer, R.S. and M.D. Purugganan. 2013. Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.*, 14: 840-852.

Morton, C.M. 2011. Newly sequenced nuclear gene (*Xdh*) for inferring angiosperm phylogeny. *Ann. Missouri Bot. Gard.*, 98: 63-89.

Mur, L.A.J., P. Kenton, R. Atzorn, O. Miersch and C. Wasternack. 2006. The outcomes of concentration-specific interactions between salicylate and jasmonate signalling include synergy, antagonism and oxidative stress leading to cell death. *Plant Physiol.*, 140: 249-262.

Murat, F., A. Louis, F. Maumus, A. Armero, R. Cooke, H. Quesneville and H.R. Crollius. 2015. Understanding Brassicaceae evolution through ancestral genome reconstruction. *Genom. Biol.*, 16: 262.

Nagaharu, U. 1935. Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Japan J. Bot.*, 7: 389-452.

Orozco-Sifuentes, M.M., R.G. Castillo-Godina and L.G. Campos-Múzquiz. 2023. DNA barcoding of crop plants. In: *Genomics, Transcriptomics, Proteomics and Metabolomics of Crop Plants*. Academic Press, pp. 199-228.

Overmyer, K., H. Tuominen, R. Kettunen, C. Betz, C. Langebartels, H. Sandermann and J. Kangasjärvi. 2000. Ozone-sensitive *Arabidopsis rcd1* mutant reveals opposite roles for ethylene and jasmonate signaling pathways in regulating superoxide-dependent cell death. *Plant Cell*, 12: 1849-1862.

Prakash, C.S. 2001. The genetically modified crop debate in the context of agricultural evolution. *Plant Physiol.*, 126: 8-15.

Rombel, I.T., K.F. Sykes, S. Rayner and S.A. Johnston. 2002. ORFFINDER: a vector for high-throughput gene identification. *Gene*, 282: 33-41.

Siddiqua, B., S. Qamarunnisa and A. Azhar. 2016. RCD1 homologues and their constituent WWE domain in plants: analysis of conservation through phylogenetic methods. *Biologia*, 71: 642-650.

Soltis, D.E., M.A. Gitzendanner, G. Stull, M. Chester, A. Chanderbali, S. Chamala, I. Jordon-Thaden, P.S. Soltis, P.S. Schnable and W.B. Barbazuk. 2013. "The Potential of Genomics in Plant Systematics." *Taxon* 62 (5): 886-898.

Teotia, S. and R.S. Lamb. 2009. The paralogous genes *RADICAL-INDUCED CELL DEATH1* and *SIMILAR TO RCD ONE1* have partially redundant functions during *Arabidopsis* development. *Plant Physiol.*, 151: 180-198.

Tnah, L.H., S.L. Lee, C.T. Lee, K. K.S. Ng, C.H. Ng and N. Zawiah. 2024. DNA barcode identification of cultivated and wild tropical fruit species. *3 Biotech*., 14: 7.

Vainonen, J.P., R. Goossens, J. Krasensky-Wrzaczek, R.D. Masi, I. Danciu, T. Puukko, N. Battchikova, C. Jonak, L. Wirthmueller, M. Wrzaczek, A. Shapiguzov and J. Kangasjärvi. 2023. Poly (ADP-ribose)-binding protein RCD1 is a plant PARylation reader regulated by photoregulatory protein kinases. *Commun. Biol.*, 6: 429.

Wendel, J.F., S.A. Jackson, B.C. Meyers and R.A. Wing. 2016. Evolution of plant genome architecture. *Genome Biol.*, 17: 37.

Wolfe, K.H., M.L. Gouy, Y.W. Yang, P.M. Sharp and W.H. Li. 1989. Date of the monocot dicot divergence estimated from chloroplast DNA-sequence data. *Proceedings of the National Academy of Sciences,* USA, 86: 6201-6205.

You, J., W. Zong, H. Du, H. Hu and L. Xiong. 2014. A special member of the rice SRO family, OsSRO1c, mediates responses to multiple abiotic stresses through interaction with various transcription factors. *Plant Mol. Biol.*, 84: 693-705.

Young, A.D. and J.P. Gillung. 2020. Phylogenomics—principles, opportunities and pitfalls of big-data phylogenetics. *Syst. Entomol.*, 45: 225-247.