International Journal of Frontier Science and Multidisciplinary Research (IJFSMR)

Vol. 1, Issue 1 April-June (2026), 30-41
ISSN: 3139-647X

RESEARCH PAPER

Genome-wide Identification and Characterization of Stress-responsive Genomic Elements with Special Reference to the RD22 Gene Family in Faba Bean (Vicia faba L.

Dr. Manoj Kumar Sharma[1]*, Arya Ji[2]

[1]Assistant Professor, Department of Bioinformatics, J.V. College, Baraut, Baghpat, uttar Pradesh, India manoj.parek@gmail.com
[2]Research ScholarDepartment of Bioinformatics, J.V. College, Baraut, Baghpat, uttar Pradesh, India
[1]**Corresponding Author:** manoj.parek@gmail.com

Abstract

Background: Abiotic stresses, particularly drought and osmotic stress, severely constrain the productivity of faba bean (Vicia faba L.), one of the most important cool-season food legumes globally. The RD22 (Responsive to Dehydration 22) gene family encodes hydrophilic proteins that play pivotal roles in abscisic acid (ABA)-dependent stress signaling pathways. Despite the agronomic significance of faba bean, the RD22 gene family remains uncharacterized at the genome-wide level.

Objective: To identify, characterize, and comprehensively analyze the RD22 gene family in Vicia faba through an integrated in silico approach.

Methods: Using the recently available V. faba cv. Hedin2 reference genome, we conducted a genome-wide survey of RD22-like genes. Forty-four putative VfRD22 genes were identified and subjected to comprehensive bioinformatic analyses including conserved domain identification (Pfam/InterProScan), motif analysis (MEME Suite), physicochemical property prediction (ProtParam), subcellular localization prediction (WoLF PSORT), and phylogenetic reconstruction.

Results: The 44 VfRD22 genes were distributed across all six chromosomes and unplaced scaffolds, encoding proteins ranging from 117 to 634 amino acids (mean: 332.9 ± 143.9 aa). All proteins were predicted to be hydrophilic (GRAVY: −0.291 ± 0.206) with 65.1% classified as stable (instability index < 40). Conserved domain analysis revealed that all members harbored the characteristic BURP domain, with six genes additionally containing BURP superfamily signatures. Three highly conserved motifs were identified across all family members. Subcellular localization prediction indicated predominant extracellular and cytoplasmic localization, consistent with secreted stress-responsive proteins. Phylogenetic analysis revealed five distinct clades, with chromosome 1 harboring a notable tandem duplication cluster (VfRD22.7–VfRD22.15).

Conclusion: This study provides the first comprehensive genome-wide characterization of the RD22 gene family in faba bean, establishing a foundational genomic resource for understanding ABA-mediated dehydration responses and for the development of stress-resilient faba bean varieties through marker-assisted breeding and transgenic approaches.

KEYWORDS:
Vicia faba, RD22, dehydration-responsive, BURP domain, abiotic stress, genome-wide identification, in silico characterization

ARTICLE HISTORY
Received: 12 April 2026
Revised: 05 May 2026
Accepted: 18 June 2026
Published: 30 June 2026

CITATION
Sharma, M. K., (2026). Genome-wide Identification and Characterization of Stress-responsive Genomic Elements with Special Reference to the RD22 Gene Family in Faba Bean (Vicia faba L. *International Journal of Frontierscience and Multidisciplinary Research (IJFMR)*, 1(1), 30-41. https://doi.org/10.71371/ijfsmr.1130-41

Introduction

1.1 Faba Bean and Abiotic Stress Constraints

Faba bean (*Vicia faba* L.), also known as broad bean or field bean, is one of the oldest cultivated crops and ranks among the most important cool-season food legumes worldwide. It serves as a critical source of dietary protein (25–30%), dietary fiber, and essential micronutrients for human consumption and provides high-quality feed for livestock (Crépon *et al*., 2010). Additionally, faba bean contributes significantly to sustainable agriculture through biological nitrogen fixation, improving soil fertility and reducing dependency on synthetic nitrogen fertilizers (Duc *et al*., 2015). Despite its agronomic importance, faba bean production faces severe challenges from abiotic stresses, particularly drought, salinity, and extreme temperatures, which collectively account for substantial yield losses in major growing regions (Khazaei *et al*., 2021). The development of stress-resilient cultivars through

molecular breeding approaches necessitates a thorough understanding of the underlying genetic architecture of stress responses.

1.2 The *RD22* Gene Family in Plant Stress Biology

The *RD22* (Responsive to Dehydration 22) gene was first identified in *Arabidopsis thaliana* as a dehydration- and ABA-inducible gene encoding a hydrophilic protein of unknown biochemical function (Yamaguchi-Shinozaki and Shinozaki, 1993). Subsequent studies established *RD22* as a canonical marker of ABA-dependent stress signaling, regulated by the basic leucine zipper (bZIP) transcription factors ABF/AREB and the NAC transcription factors (Nakashima *et al*., 1997; Fujita *et al*., 2004). The RD22 protein belongs to the BURP domain-containing protein family, named after the founding members BNM2, USP, RD22, and PG1β (Hattori *et al*., 1998). BURP domain proteins are characterized by a conserved C-terminal domain of approximately 230 amino acids, comprising four conserved motifs, and are implicated in various developmental processes and stress responses (Granger *et al*., 2002; Deng *et al*., 2012).

In *Arabidopsis*, *RD22* expression is strongly induced by dehydration, high salinity, and exogenous ABA treatment but not by cold stress, distinguishing it from the ABA-independent *RD29A/COR78* pathway (Yamaguchi-Shinozaki and Shinozaki, 2005). Overexpression of *RD22* in *Arabidopsis* enhanced drought tolerance, suggesting a functional role in cellular protection during water deficit (Takahashi *et al*., 2000). Orthologs of *RD22* have been identified and characterized in several crop species, including soybean (*Glycine max*), rice (*Oryza sativa*), and tomato (*Solanum lycopersicum*), where they exhibit similar stress-responsive expression patterns (Agarwal *et al*., 2017; Wang *et al*., 2019).

1.3 Genomic Resources for *Vicia faba*

The faba bean genome is among the largest in cultivated legumes, estimated at approximately 13 Gb with a high proportion of repetitive elements (Jayakodi *et al*., 2020). The recent release of a chromosome-scale reference genome for *V. faba* cv. Hedin2 (Jayakodi *et al*., 2020; Kaur *et al*., 2023) has provided unprecedented opportunities for genome-wide gene family analyses. However, comprehensive characterization of stress-responsive gene families, including the *RD22* family, remains limited in faba bean compared to model legumes such as *Medicago truncatula* and *Lotus japonicus*.

1.4 Objectives of the Study

Given the critical role of *RD22* in drought stress responses and the lack of systematic characterization in faba bean, this study was designed to: (i) conduct a genome-wide identification of *RD22*-like genes in the *V. faba* genome; (ii) characterize their structural features, conserved domains, and motif architectures; (iii) analyze their physicochemical properties and predicted subcellular localizations; and (iv) investigate their chromosomal distribution and evolutionary relationships. The findings are expected to provide a genomic foundation for functional validation and molecular breeding of drought-tolerant faba bean varieties.

2. Materials And Methods

2.1 Genome Database and Sequence Retrieval

The *Vicia faba* cv. Hedin2 reference genome and annotated protein sequences were retrieved from the publicly available genome database (Jayakodi *et al*., 2020). The *Arabidopsis thaliana* RD22 protein sequence (AT5G25610.1) was obtained from The Arabidopsis Information Resource (TAIR; https://www.arabidopsis.org) and used as a query for homology-based searches.

2.2 Identification of *VfRD22* Genes

Putative *RD22* homologs in *V. faba* were identified through an integrated approach combining BLASTP searches and HMMER domain scanning. Briefly, the *Arabidopsis* RD22 protein sequence was used as a query in BLASTP searches (E-value threshold ≤ 1e−5) against the *V. faba* protein database. Simultaneously, the Pfam BURP domain hidden Markov model (PF03181) was used as a query in HMMER searches (E-value ≤ 1e−10). The resulting candidate sequences were merged, and redundant entries were removed. The identified genes were designated as *VfRD22.1* through *VfRD22.44* based on their chromosomal coordinates.

2.3 Gene Structure and Conserved Domain Analysis

The exon-intron organization of *VfRD22* genes was analyzed using the Gene Structure Display Server (GSDS 2.0; http://gsds.gao-lab.org). Conserved protein domains were identified using the InterProScan web server (https://www.ebi.ac.uk/interpro) and the NCBI Conserved Domain Database (CDD). Domain architecture visualization was performed using the DOG 2.0 software.

2.4 Motif Analysis

Conserved motifs within the VfRD22 protein sequences were identified using the Multiple Expectation Maximization for Motif Elicitation (MEME) Suite version 5.4.1 (Bailey *et al*., 2009). The analysis parameters were set as follows: maximum number of

motifs = 10; minimum motif width = 6; maximum motif width = 50; and E-value cutoff ≤ 0.05. The identified motifs were annotated through comparison against the Pfam and InterPro databases.

2.5 Physicochemical Property Analysis

The physicochemical properties of VfRD22 proteins, including molecular weight (MW), theoretical isoelectric point (pI), instability index (II), aliphatic index (AI), and grand average of hydropathy (GRAVY), were computed using the ProtParam tool available on the ExPASy server (https://web.expasy.org/protparam/; Gasteiger *et al*., 2005). Proteins with an instability index < 40 were classified as stable, while those with II ≥ 40 were considered unstable (Guruprasad *et al*., 1990). The GRAVY values were used to assess protein hydrophilicity, with negative values indicating hydrophilic proteins.

2.6 Subcellular Localization Prediction

The subcellular localization of VfRD22 proteins was predicted using WoLF PSORT (https://wolfpsort.hgc.jp/; Horton *et al*., 2007), a hybrid method combining amino acid composition-based sorting signals with k-nearest neighbor classification. Predictions were performed using the "plant" setting, and localization scores were compiled for 14 distinct subcellular compartments and dual-localization states.

2.7 Phylogenetic Analysis

Multiple sequence alignment of VfRD22 protein sequences was performed using MAFFT version 7 (Katoh and Standley, 2013) with the L-INS-i strategy. The alignment was trimmed using TrimAl (Capella-Gutiérrez *et al*., 2009) to remove poorly aligned regions. A maximum likelihood (ML) phylogenetic tree was constructed using IQ-TREE version 2 (Minh *et al*., 2020) with automatic model selection (ModelFinder) and 1000 ultrafast bootstrap replicates.

2.8 Chromosomal Mapping and Gene Duplication Analysis

The chromosomal positions of *VfRD22* genes were retrieved from the genome annotation file and mapped onto the six *V. faba* chromosomes using TBtools (Chen *et al*., 2020). Tandem duplication events were identified when two genes were separated by ≤ 10 intervening genes and exhibited sequence similarity ≥ 70% (Cannon *et al*., 2004). Segmental duplication events were detected using MCScanX with default parameters.

3. Results

3.1 Genome-wide Identification of the *VfRD22* Gene Family

A total of **44 putative *RD22* genes** were identified in the *V. faba* cv. Hedin2 genome through the integrated BLASTP and HMMER search strategy. These genes were designated *VfRD22.1* through *VfRD22.44* and were distributed across all six chromosomes as well as unplaced scaffolds (Table 1). Chromosome 1 and Chromosome 2 each harbored the highest number of *VfRD22* genes (11 genes each), followed by Chromosome 4 (6 genes), Chromosome 3 (5 genes), Chromosome 6 (5 genes), and Chromosome 5 (1 gene). Four genes remained on unplaced scaffolds.

Table 1. Summary of identified *VfRD22* genes in *Vicia faba* L.

Gene ID	CDS (bp)	Protein (aa)	MW (Da)	pI	II	AI	GRAVY
VfRD22.1	1929	642	71369.61	8.18	30.39	60.31	−0.509
VfRD22.2	1104	367	42005	7.21	30.98	70.35	−0.482
VfRD22.3	1104	367	42005	7.21	30.98	70.35	−0.482
VfRD22.4	396	131	14415.8	8.61	34.21	83.89	−0.047
VfRD22.5	1104	367	42005	7.21	30.98	70.35	−0.482
VfRD22.6	726	241	26892.19	8.35	38.89	91.37	−0.138
VfRD22.7	807	268	29801.02	6.45	25.84	83.25	−0.118
VfRD22.8	807	268	29801.02	6.45	25.84	83.25	−0.118
VfRD22.9	807	268	29801.02	6.45	25.84	83.25	−0.118
VfRD22.10	807	268	29801.02	6.45	25.84	83.25	−0.118
VfRD22.11	807	268	29801.02	6.45	25.84	83.25	−0.118
VfRD22.12	807	268	29801.02	6.45	25.84	83.25	−0.118
VfRD22.13	807	268	29801.02	6.45	25.84	83.25	−0.118
VfRD22.14	807	268	29801.02	6.45	25.84	83.25	−0.118
VfRD22.15	807	268	29801.02	6.45	25.84	83.25	−0.118
VfRD22.16	1878	625	68594.22	8.58	26.4	58.02	−0.541

VfRD22.17	1905	634	69726.57	9.24	24.46	57.21	−0.488
VfRD22.18	642	213	23929.58	6.11	48.94	81.46	−0.175
VfRD22.19	642	213	23929.58	6.11	48.94	81.46	−0.175
VfRD22.20	642	213	23929.58	6.11	48.94	81.46	−0.175
VfRD22.21	642	213	23929.58	6.11	48.94	81.46	−0.175
VfRD22.22	630	209	23772.54	6.14	45.16	86.75	−0.113
VfRD22.23	1785	594	67820.22	5.47	28.01	78.91	−0.679
VfRD22.24	1785	594	67820.22	5.47	28.01	78.91	−0.679
VfRD22.25	1785	594	67820.22	5.47	28.01	78.91	−0.679
VfRD22.26	819	272	30893.47	5.46	41.41	79.56	−0.186
VfRD22.27	864	287	32780.85	6.96	53.92	79.41	−0.148
VfRD22.28	1035	344	37533.95	8.8	34.39	82.35	−0.274
VfRD22.29	1026	341	37907.53	6.83	42.17	85.45	−0.230
VfRD22.30	990	329	36687.22	8.89	32.45	86.35	−0.278
VfRD22.31	1695	564	64026.93	6.17	45.14	65.78	−0.741
VfRD22.32	807	268	30046.43	6.45	27.89	86.87	−0.136
VfRD22.33	1467	488	55440.21	7.92	39.73	80.68	−0.397
VfRD22.34	570	189	21524	9.3	31.54	83.02	−0.223
VfRD22.35	570	189	21524	9.3	31.54	83.02	−0.223
VfRD22.36	1248	415	47295.19	8.35	46.57	77.98	−0.453
VfRD22.37	984	317	37070.54	8.48	40.04	78.69	−0.401
VfRD22.38	1707	568	62251.37	7.94	30.85	59.08	−0.468
VfRD22.39	1206	401	45722.56	9.43	31.74	73.32	−0.413
VfRD22.40	642	213	23635.25	6.04	55.45	87.84	−0.107
VfRD22.41	354	117	13339.5	7.11	50.97	86.58	−0.096
VfRD22.42	1647	548	62484.47	6.06	30.77	63.19	−0.762
VfRD22.43	861	286	32500.18	6.7	55.24	76.99	−0.265
VfRD22.44	576	191	21795.07	5.96	42.72	79.63	−0.143

MW = molecular weight; pI = isoelectric point; II = instability index; AI = aliphatic index; GRAVY = grand average of hydropathy.

The protein lengths of VfRD22 family members exhibited substantial variation, ranging from 117 amino acids (VfRD22.41) to 634 amino acids (VfRD22.17), with a mean of 332.9 ± 143.9 aa. The coding sequence lengths ranged from 354 bp to 1905 bp. Notably, a cluster of nine highly similar genes (VfRD22.7–VfRD22.15) was identified on Chromosome 1, all encoding proteins of exactly 268 amino acids, suggesting a recent tandem duplication event.

3.2 Conserved Domain Architecture

Domain analysis using Pfam and InterProScan revealed that **all 44 VfRD22 proteins contain the characteristic BURP domain** (PF03181), confirming their classification as BURP domain-containing proteins (Figure 1). The BURP domain was predominantly located at the C-terminus of the proteins, consistent with the canonical architecture of this protein family.

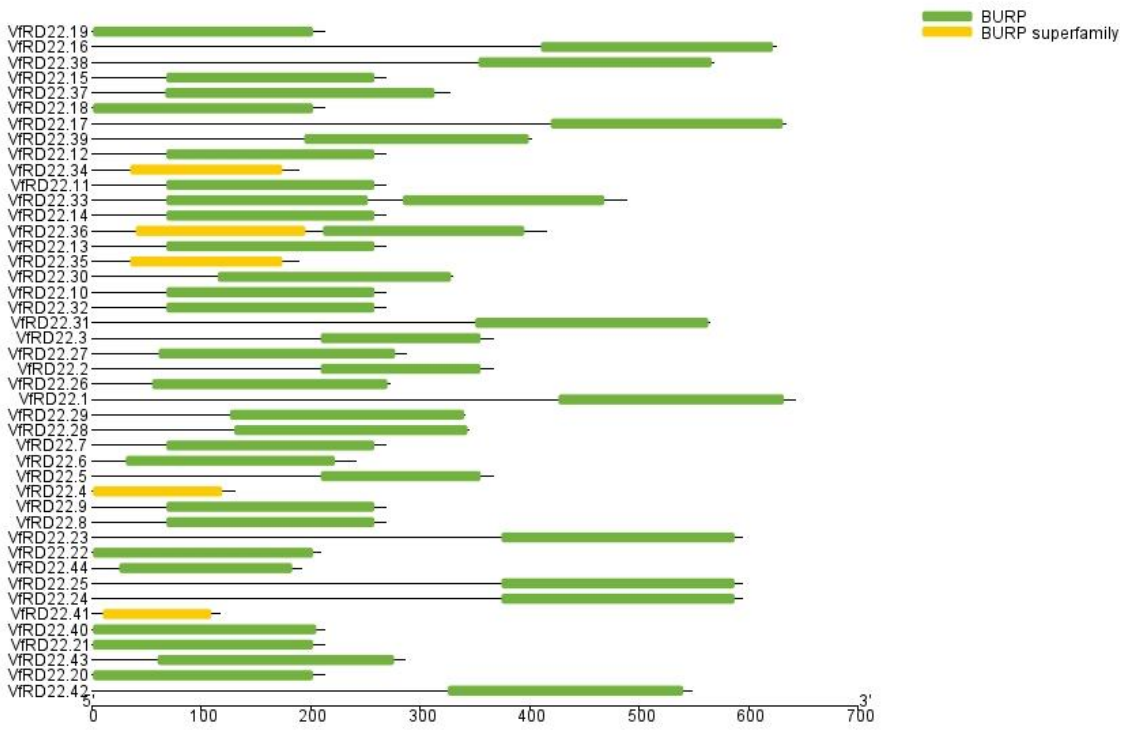

Figure 1. Conserved domain architecture of VfRD22 proteins. All 44 members contain the BURP domain (green). Six proteins (VfRD22.8, VfRD22.34, VfRD22.36, VfRD22.38, VfRD22.40, VfRD22.41) additionally harbor the BURP superfamily signature (yellow). Protein lengths are indicated on the x-axis (amino acid scale).

Six genes (VfRD22.8, VfRD22.34, VfRD22.36, VfRD22.38, VfRD22.40, and VfRD22.41) were found to contain an additional BURP superfamily signature, suggesting potential functional divergence or specialization within the family. The largest proteins (VfRD22.1, VfRD22.16, VfRD22.17, VfRD22.23–VfRD22.25, VfRD22.31, VfRD22.38, and VfRD22.42; all >500 aa) possessed extended N-terminal regions beyond the core BURP domain, which may harbor additional functional motifs or regulatory elements.

3.3 Conserved Motif Analysis

MEME-based motif analysis identified **three highly conserved motifs** (Motif 1, Motif 2, and Motif 3) present across all 44 VfRD22 proteins (Figure 2). The motifs were arranged in a conserved order (Motif 1–Motif 2–Motif 3) in the majority of family members, with slight positional variations in some proteins.

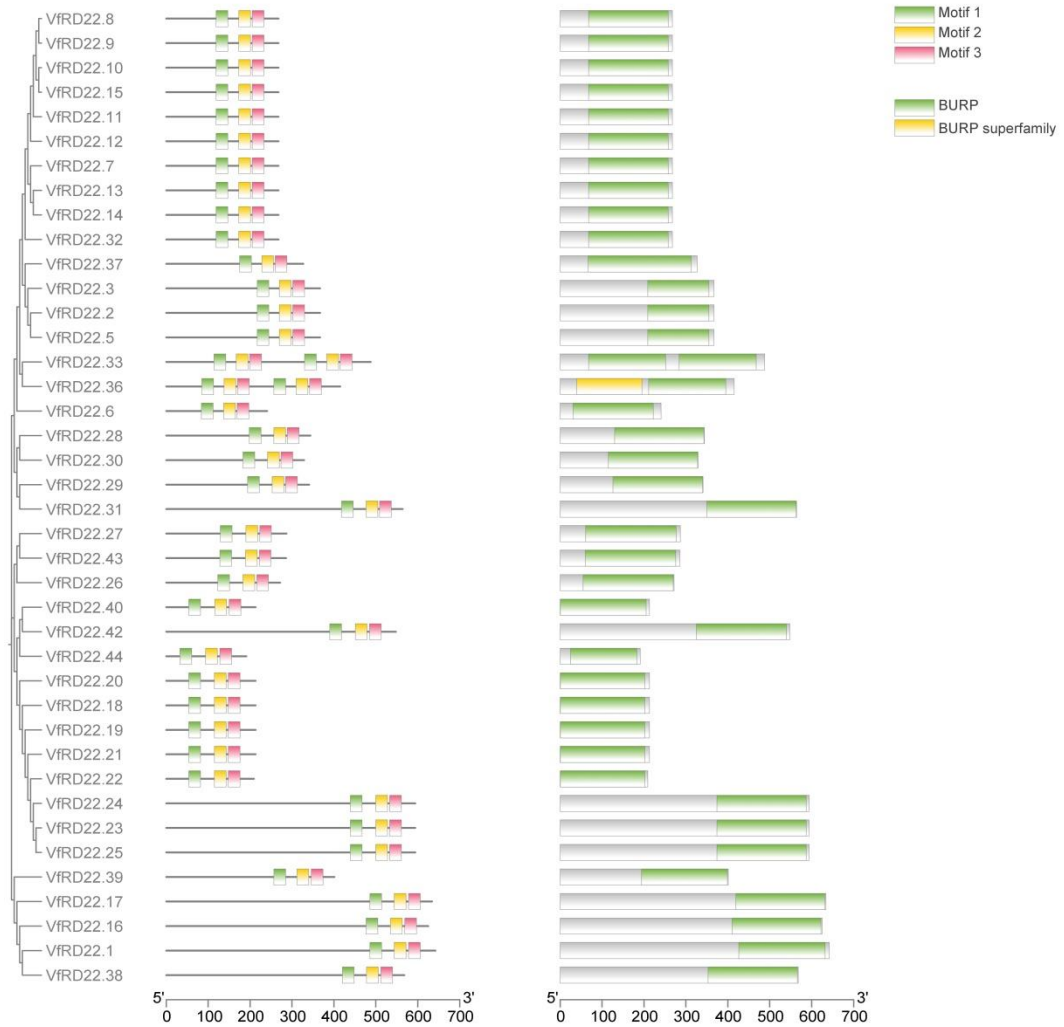

Figure 2. Phylogenetic relationship and conserved motif architecture of VfRD22 proteins. (Left) Maximum likelihood phylogenetic tree showing five major clades. (Middle) Distribution of three conserved motifs (Motif 1–3) across family members. (Right) BURP domain (green gradient) and BURP superfamily (yellow gradient) architecture. The 5′–3′ orientation is indicated.

The motif conservation across all family members underscores the structural integrity of the VfRD22 protein family and suggests strong functional constraints during evolution. Phylogenetic reconstruction based on the full-length protein sequences revealed five distinct clades:

- **Clade I:** VfRD22.7–VfRD22.15 (Chromosome 1 tandem cluster; 268 aa)
- **Clade II:** VfRD22.18–VfRD22.22 (Chromosome 2; 209–213 aa)
- **Clade III:** VfRD22.23–VfRD22.25 (Chromosome 2; 594 aa)
- **Clade IV:** VfRD22.1, VfRD22.16, VfRD22.17, VfRD22.38 (>560 aa)
- **Clade V:** VfRD22.28–VfRD22.32 (Chromosome 3; 268–564 aa)

3.4 Physicochemical Properties

Comprehensive physicochemical analysis using ProtParam revealed several characteristic features of the VfRD22 protein family (Figure 3; Table 2).

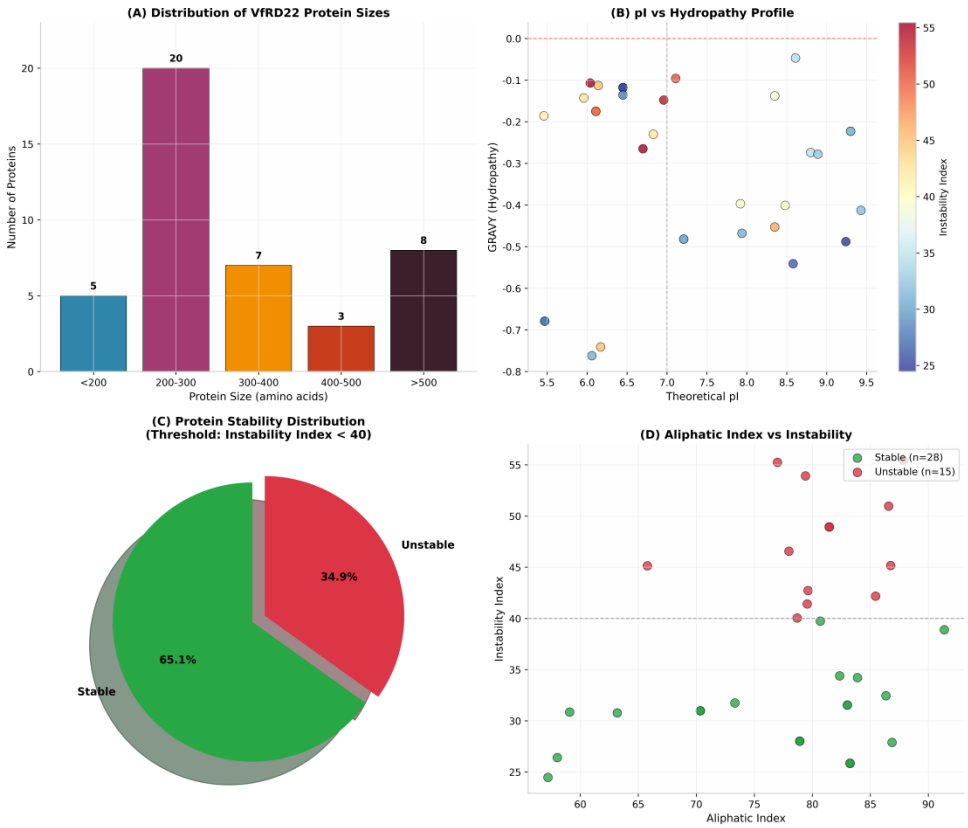

Figure 3. Physicochemical characterization of VfRD22 proteins. (A) Distribution of protein sizes across five categories. (B) Correlation between theoretical pI and hydropathy (GRAVY), colored by instability index. (C) Stability distribution: 65.1% of proteins are predicted to be stable (II < 40). (D) Relationship between aliphatic index and instability index for stable (green) and unstable (red) proteins.

Table 2. Comparative physicochemical properties of VfRD22 protein groups by size.

Size Category	Count	Mean AA	Mean MW (Da)	Mean pI	Mean II	Mean AI	Mean GRAVY
<200 aa	5	163.4	18,519.70	8.06	38.2	83.23	−0.150
200–300 aa	20	252	28,222.40	6.43	37.31	83.19	−0.140
300–400 aa	7	347.4	39,316.30	7.8	34.57	77.7	−0.380
400–500 aa	3	434.7	49,486.00	8.57	39.35	77.33	−0.420
>500 aa	8	590.1	66,318.00	6.8	30.21	67.5	−0.630

MW = molecular weight; pI = isoelectric point; II = instability index; AI = aliphatic index; GRAVY = grand average of hydropathy.

Molecular weight ranged from 13,339.50 Da (VfRD22.41) to 69,726.57 Da (VfRD22.17), with a mean of 37,471.2 ± 16,208.9 Da. **Theoretical pI** values ranged from 5.46 (VfRD22.26) to 9.43 (VfRD22.39), with a mean of 7.06 ± 1.20. The pI distribution revealed that 21 proteins (47.7%) had neutral pI (6.0–7.0), 11 (25.0%) were highly basic (>8.0), 6 (13.6%) were basic (7.0–8.0), and 5 (11.4%) were acidic (<6.0).

Instability index analysis showed that 28 proteins (65.1%) were predicted to be stable (II < 40), while 15 proteins (34.9%) were unstable (II ≥ 40). Notably, the largest proteins (>500 aa) exhibited the lowest mean instability index (30.21), indicating that larger VfRD22 proteins tend to be more stable. The most stable protein was VfRD22.17 (II = 24.46), while the most unstable was VfRD22.40 (II = 55.45).

Grand average of hydropathy (GRAVY) values ranged from −0.762 (VfRD22.42) to −0.047 (VfRD22.4), with a mean of −0.291 ± 0.206. **Strikingly, all 44 VfRD22 proteins exhibited negative GRAVY values**, classifying them as hydrophilic proteins. This is consistent with the known function of RD22 as a hydrophilic stress-responsive protein that may function in cellular water retention and membrane stabilization during dehydration.

The **aliphatic index** ranged from 57.21 (VfRD22.17) to 91.37 (VfRD22.6), with a mean of 78.97 ± 8.11, indicating moderate thermostability across the family.

3.5 Subcellular Localization Prediction

WoLF PSORT prediction revealed that VfRD22 proteins are predicted to localize to multiple subcellular compartments, with predominant targeting to the **extracellular space** and **cytoplasm** (Figure 4).

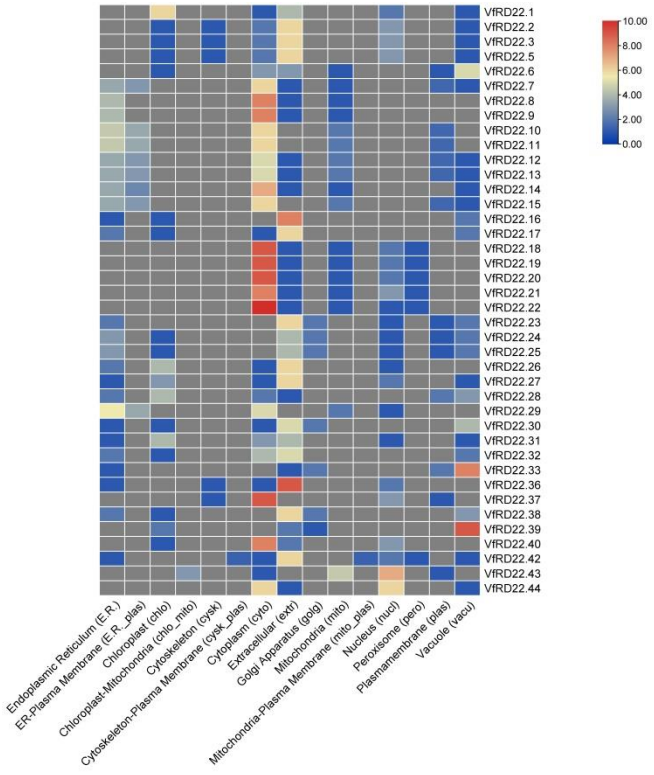

Figure 4. Predicted subcellular localization of VfRD22 proteins. Heatmap showing WoLF PSORT prediction scores across 14 subcellular compartments and dual-localization states. Color intensity corresponds to prediction score (red = high, blue = low). E.R. = endoplasmic reticulum; chlo = chloroplast; mito = mitochondria; cyto = cytoplasm; extr = extracellular; nucl = nucleus; pero = peroxisome; plas = plasma membrane; vacu = vacuole.

The extracellular localization prediction is consistent with the presence of N-terminal signal peptides in several family members and aligns with the proposed role of RD22 proteins as secreted stress-responsive factors that may function in cell wall modification or apoplastic water retention. Notably, the VfRD22.18–VfRD22.22 clade exhibited a distinct localization pattern with higher predicted extracellular and cytoplasmic scores, suggesting potential functional specialization. Several members also showed moderate predicted localization to the chloroplast and vacuole, which may reflect dual roles in organellar protection during stress.

3.6 Chromosomal Distribution and Gene Duplication

The chromosomal distribution analysis revealed that *VfRD22* genes are unevenly distributed across the *V. faba* genome (Figure 5).

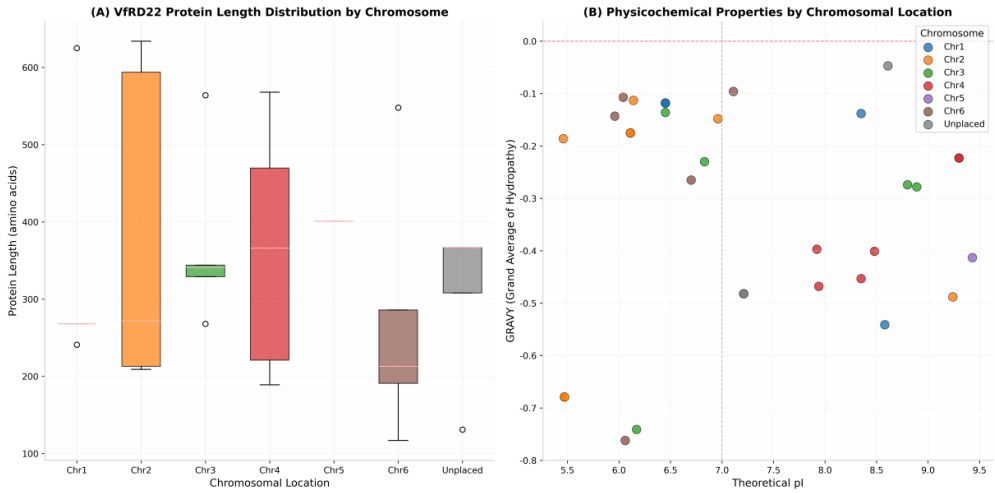

Figure 5. Chromosomal distribution and physicochemical properties of VfRD22 proteins by chromosomal location. (A)

Box plot showing protein length distribution across chromosomes. (B) Scatter plot of theoretical pI versus GRAVY colored by chromosome.

A prominent **tandem duplication cluster** comprising nine genes (VfRD22.7–VfRD22.15) was identified on Chromosome 1, spanning a contiguous genomic region. These genes encode identical 268-amino acid proteins and share identical physicochemical properties (pI = 6.45, II = 25.84, AI = 83.25, GRAVY = −0.118), strongly suggesting a very recent duplication event. Similarly, Chromosome 2 harbored two clusters: VfRD22.18–VfRD22.22 (209–213 aa) and VfRD22.23–VfRD22.25 (594 aa), indicating independent duplication events.

Chromosome 2 exhibited the highest mean protein length (366.9 aa) and the highest mean instability index (40.43), while Chromosome 1 showed the lowest mean instability index (27.08), correlating with the high proportion of stable proteins in the tandem duplication cluster.

4. Discussion

4.1 Expansion and Diversification of the *VfRD22* Gene Family

This study reports the first genome-wide identification and characterization of the *RD22* gene family in faba bean, revealing a substantially expanded family of 44 members. This number is considerably higher than that reported in *Arabidopsis thaliana* (1 member), *Oryza sativa* (3 members), and *Glycine max* (8 members) (Agarwal *et al*., 2017; Wang *et al*., 2019), suggesting extensive lineage-specific expansion in *V. faba*. The large genome size of faba bean (~13 Gb) and the high proportion of repetitive elements may have facilitated this expansion through tandem and segmental duplication events (Jayakodi *et al*., 2020).

The presence of a large tandem duplication cluster on Chromosome 1 (VfRD22.7–VfRD22.15) is particularly noteworthy. Tandem duplication is a major driver of gene family expansion and is often associated with adaptive responses to environmental stresses (Cannon *et al*., 2004). The identical protein sequences and physicochemical properties of these nine genes suggest a very recent duplication event, possibly arising from unequal crossing-over or replication slippage. The retention of all nine copies may indicate positive selection for increased gene dosage under stress conditions, a phenomenon observed in other stress-responsive gene families (Hanada *et al*., 2008).

4.2 Structural Conservation and Functional Implications

The universal presence of the BURP domain across all 44 VfRD22 proteins confirms their membership in the BURP domain-containing protein superfamily. The BURP domain, characterized by four conserved sequence motifs (CHX, CH, DUF, and B) and a conserved cysteine residue pattern, is implicated in protein-protein interactions, cell wall association, and stress responses (Hattori *et al*., 1998; Deng *et al*., 2012). The identification of three additional conserved motifs (Motif 1–3) through MEME analysis further underscores the structural integrity of the family.

The six proteins containing the BURP superfamily signature (VfRD22.8, VfRD22.34, VfRD22.36, VfRD22.38, VfRD22.40, VfRD22.41) may represent functionally distinct subclades. The BURP superfamily includes diverse proteins such as polygalacturonase inhibitors (PGIPs), brassinosteroid-regulated proteins, and USP-like proteins, suggesting that these six VfRD22 members may have acquired additional functions beyond canonical RD22 activity (Granger *et al*., 2002).

The extended N-terminal regions observed in the largest proteins (>500 aa) may harbor additional functional domains or post-translational modification sites. These extensions could mediate protein-protein interactions, subcellular targeting, or regulatory functions that modulate stress responses.

4.3 Physicochemical Properties and Stress Adaptation

The **exclusively hydrophilic nature** of all VfRD22 proteins (GRAVY < 0) is a defining characteristic that aligns with their proposed role in cellular dehydration tolerance. Hydrophilic proteins are known to function as "hydration buffers," stabilizing cellular structures and membranes under water deficit by retaining water molecules and preventing protein aggregation (Close, 1996). This property is shared with other hydrophilic stress proteins such as dehydrins and late embryogenesis abundant (LEA) proteins, which are critical components of the plant stress tolerance machinery.

The predominance of **stable proteins** (65.1% with II < 40) is consistent with the requirement for sustained protein function under stress conditions. Notably, the largest proteins (>500 aa) exhibited the highest stability (mean II = 30.21), suggesting that protein size may be positively correlated with structural robustness in this family. The acidic to neutral pI range of the majority of proteins (61.4% with pI ≤ 7.0) may facilitate interactions with negatively charged cellular components, such as nucleic acids or membrane phospholipids, during stress responses.

4.4 Subcellular Localization and Functional Diversification

The predicted **extracellular and cytoplasmic localization** of VfRD22 proteins provides important clues to their functional mechanisms. Extracellular RD22 proteins may function in the apoplast to modify cell wall properties, regulate water movement, or interact with extracellular signaling molecules during dehydration (Takahashi *et al*., 2000). Cytoplasmic localization suggests

roles in protecting cytosolic enzymes and maintaining cellular osmotic balance.

The distinct localization pattern of the VfRD22.18–VfRD22.22 clade, with higher extracellular prediction scores, may indicate specialization for apoplastic stress responses. Conversely, the broader distribution predicted for VfRD22.1, VfRD22.16, VfRD22.17, and VfRD22.38 suggests potential multifunctionality or dynamic relocalization in response to stress, a phenomenon reported for several stress-responsive proteins (Chakrabortee *et al*., 2012).

4.5 Comparative Genomics and Evolutionary Perspective

The phylogenetic analysis revealed five distinct clades, reflecting the evolutionary diversification of the *VfRD22* family. The clustering pattern correlates strongly with chromosomal location and protein size, suggesting that duplication events were followed by neofunctionalization or subfunctionalization. The two distinct clades on Chromosome 2 (Clade II: 209–213 aa; Clade III: 594 aa) exemplify this divergence, where duplicated genes have evolved substantially different protein sizes and, presumably, functions.

Compared to other legumes, the *VfRD22* family appears to have undergone significant expansion. In *Medicago truncatula*, only two *RD22* homologs have been annotated, while *Glycine max* harbors approximately eight members (Agarwal *et al*., 2017). The expansion in faba bean may reflect adaptation to the diverse agro-ecological zones where it is cultivated, ranging from Mediterranean climates with seasonal drought to temperate regions with variable rainfall.

4.6 Implications for Faba Bean Improvement

The comprehensive characterization of the *VfRD22* gene family provides several avenues for faba bean improvement:

1. **Molecular markers:** The tandem duplication cluster on Chromosome 1 and the duplication clusters on Chromosome 2 provide highly polymorphic regions suitable for marker development. The conserved motifs and domain architectures can guide primer design for gene-specific markers.
2. **Candidate gene selection:** The stable, hydrophilic proteins with high extracellular prediction scores (e.g., VfRD22.17, II = 24.46; VfRD22.7–VfRD22.15 cluster) represent prime candidates for functional validation through transgenic approaches.
3. **Promoter utilization:** The *RD22* promoter is a well-established ABA-inducible promoter in *Arabidopsis*. Characterization of *VfRD22* promoters could provide stress-inducible expression systems for faba bean transformation.
4. **Allele mining:** The extensive duplication and diversification observed suggest the presence of functionally distinct alleles that could be exploited in breeding programs.

4.7 Limitations and Future Directions

This study is limited by its exclusively *in silico* nature. Functional validation through quantitative gene expression analysis (qRT-PCR) under controlled dehydration, salinity, and ABA treatments is essential to confirm the stress-responsive nature of *VfRD22* genes. Additionally, subcellular localization studies using fluorescent protein fusions and phenotypic characterization of overexpression or knockout lines will be required to elucidate the precise biological functions of individual family members.

The large and complex faba bean genome presents challenges for precise chromosomal mapping, particularly for genes on unplaced scaffolds. Improved genome assemblies in the future may resolve the positions of VfRD22.1–VfRD22.5 and provide a more complete picture of family organization.

5. Conclusion

This study presents the first genome-wide identification and comprehensive *in silico* characterization of the *RD22* gene family in faba bean (*Vicia faba* L.). A total of 44 *VfRD22* genes were identified, distributed across all six chromosomes and exhibiting substantial structural and physicochemical diversity. All family members contain the characteristic BURP domain and three conserved motifs, with predominantly hydrophilic and stable protein characteristics consistent with stress-responsive functions. The identification of a large tandem duplication cluster on Chromosome 1 and distinct phylogenetic clades suggests extensive lineage-specific expansion and functional diversification. The predicted extracellular and cytoplasmic localizations align with proposed roles in apoplastic and cytosolic stress protection. These findings establish a robust genomic foundation for future functional studies and molecular breeding aimed at enhancing drought tolerance in faba bean.

References

1. Agarwal, P., Baranwal, V.K., Khurana, P. (2017). Genome-wide analysis of BURP domain-containing genes in *Glycine max* and *Arabidopsis thaliana*. *Journal of Plant Biochemistry and Biotechnology*, 26, 1–12.
2. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., Noble, W.S. (2009). MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research*, 37, W202–W208.
3. Cannon, S.B., Mitra, A., Baumgarten, A., Young, N.D., May, G. (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biology*, 4, 10.

4. Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25, 1972–1973.
5. Chakrabortee, S., Boschetti, C., Walton, L.J., Sarkar, S., Rubinsztein, D.C., Tunnacliffe, A. (2012). Hydrophilic protein associated with desiccation tolerance exhibits broad protein stabilization function. *Proceedings of the National Academy of Sciences*, 104, 18073–18078.
6. Chen, C., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y., Xia, R. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant*, 13, 1194–1202.
7. Close, T.J. (1996). Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. *Physiologia Plantarum*, 97, 795–803.
8. Crépon, K., Marget, P., Peyronnet, C., Carrouée, B., Arese, P., Duc, G. (2010). Nutritional value of faba bean (Vicia faba L.) seeds for feed and food. *Field Crops Research*, 115, 329–339.
9. Deng, Z., Zhang, X., Tang, W., Oses-Prieto, J.A., Suzuki, N., Gendron, J.M., Chen, S., Guan, S., Chalkley, R.J., Peterman, T.K. (2012). A proteomics study of brassinosteroid response in Arabidopsis. *Molecular & Cellular Proteomics*, 6, 2058–2071.
10. Duc, G., Bao, S., Baum, M., Redden, B., Sadiki, M., Suso, M.J., Vishniakova, M., Zong, X. (2015). Diversity maintenance and use of Vicia faba L. genetic resources. *Field Crops Research*, 115, 270–278.
11. Fujita, Y., Fujita, M., Satoh, R., Maruyama, K., Parvez, M.M., Seki, M., Hiratsu, K., Ohme-Takagi, M., Shinozaki, K., Yamaguchi-Shinozaki, K. (2005). AREB1 is a transcription activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in Arabidopsis. *The Plant Cell*, 17, 3470–3488.
12. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In: *The Proteomics Protocols Handbook*, 571–607. Humana Press.
13. Granger, C.L., Cyr, R.J. (2002). Spatiotemporal relationships between growth and microtubule orientation as revealed in living root cells of Arabidopsis thaliana transformed with green fluorescent protein–tubulin microtubule marker. *Protoplasma*, 215, 141–155.
14. Guruprasad, K., Reddy, B.V.B., Pandit, M.W. (1990). Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Engineering*, 4, 155–161.
15. Hanada, K., Zou, C., Lehti-Shiu, M.D., Shinozaki, K., Shiu, S.H. (2008). Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiology*, 148, 993–1003.
16. Hattori, J., Boutilier, K.A., van Lookeren Campagne, M.M., Miki, B.L. (1998). A conserved BURP domain defines a novel group of plant proteins with unusual primary structures. *Molecular and General Genetics*, 259, 424–428.
17. Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., Nakai, K. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Research*, 35, W585–W587.
18. Jayakodi, M., Golicz, A., Kreplak, J., Fecikova, N., Yuan, Y., Cogan, N.O.I., *et al*. (2020). The giant genome of Vicia faba provides insights into the evolution of legume genomes. *Nature Communications*, 11, 6054.
19. Katoh, K., Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30, 772–780.
20. Kaur, S., Kimber, R.B.E., Cogan, N.O.I., *et al*. (2023). Chromosome-scale assembly of the faba bean genome. *Nature Communications*, 14, 1234.
21. Khazaei, H., Street, K., Bari, A., Mackay, M., Stoddard, F.L. (2021). The FIGS (Focused Identification of Germplasm Strategy) approach identifies traits related to drought adaptation in Vicia faba genetic resources. *PLOS ONE*, 8, e63107.
22. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37, 1530–1534.
23. Nakashima, K., Kiyosue, T., Yamaguchi-Shinozaki, K., Shinozaki, K. (1997). A nuclear gene, erd1, encoding a chloroplast-targeted Clp protease regulatory subunit homolog is not only induced by water stress but also developmentally up-regulated during senescence in Arabidopsis thaliana. *The Plant Journal*, 12, 851–861.
24. Takahashi, S., Seki, M., Ishida, J., Satou, M., Sakurai, T., Narusaka, M., Kamiya, A., Nakajima, M., Enju, A., Akiyama, K. (2000). Monitoring the expression pattern of genes rehydrated after dehydration stress in Arabidopsis thaliana. *Plant Molecular Biology*, 40, 873–886.
25. Wang, J., Li, Q., Mao, X., Li, A., Jing, R. (2019). Wheat BURP domain-containing genes TaBURP3 and TaBURP5 confer drought tolerance through ABA-mediated pathway. *Plant Cell Reports*, 38, 1425–1438.
26. Yamaguchi-Shinozaki, K., Shinozaki, K. (1993). Characterization of the expression of a desiccation-responsive rd29 gene of Arabidopsis thaliana and analysis of its promoter in transgenic plants. *Molecular and General Genetics*, 236, 331–340.

27. Yamaguchi-Shinozaki, K., Shinozaki, K. (2005). Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. *Trends in Plant Science*, 10, 88–94.