



Transcriptome-Wide Identification of Nine Tandem Peptide Repeat Families in Roselle (*Hibiscus sabdariffa* L.)

Authors:

Christina Seok Yien Yong* and Nur Atheeqah-Hamzah

*Correspondence: chrisyong@upm.edu.my

Received: 22 August 2023; **Accepted:** 20 May 2024; **Early view:** 12 July 2024

To cite this article: Christina Seok Yien Yong and Nur Atheeqah-Hamzah. (in press).

Transcriptome-wide identification of nine tandem peptide repeat families in roselle (*Hibiscus sabdariffa* L.). *Tropical Life Sciences Research*.

Highlights

- Nine tandem repeat gene families were identified in the calyx transcriptome of roselle, namely the Ankyrin repeats (ANK), Armadillo repeats (ARM), Elongation Factor-hand domain repeats (EF-Hand), Huntingtin, Elongation factor 3, protein phosphatase 2A, yeast kinase TOR1 repeats (HEAT), Kelch repeats (Kelch), Leucine rich repeats (LLR), Pentatricopeptide repeats (PPR), Tetratricopeptide repeats (TPR) and WD40 repeats (WD40).
- Transcriptome dataset could be a valuable source to discover tandem repeat gene families in the functional fraction of genome. A total of 1,045 tandem repeat-containing proteins implicated in various biological processes including plant development, protein-protein interaction, innate immunity, and abiotic stress responses were discovered.
- Biotic and abiotic response related tandem repeats genes for potential application in breeding improved varieties.

Transcriptome-Wide Identification of Nine Tandem Peptide Repeat Families in Roselle (*Hibiscus sabdariffa* L.).

Christina Seok Yien Yong* and Nur Atheeqah-Hamzah

Department of Biology, Faculty of Science, Universiti Putra Malaysia, Jalan UPM, 43400 Serdang, Selangor, Malaysia

*Corresponding author: chrisyong@upm.edu.my

Running head: Tandem Peptide Repeat Family in Roselle

Received: 22 August 2023; **Accepted:** 20 May 2024; **Early view:** 12 July 2024

To cite this article: Christina Seok Yien Yong and Nur Atheeqah-Hamzah. (in press).

Transcriptome-wide identification of nine tandem peptide repeat families in roselle (*Hibiscus sabdariffa* L.). *Tropical Life Sciences Research*.

Abstract. Plants are rich in tandem repeats-containing proteins. It is postulated that the occurrence of tandem repeat gene families facilitates the adaptation and survival of plants in adverse environmental conditions. In this study, we intended to identify the tandem repeats in the transcriptome of a high potential tropical horticultural plant, roselle (*Hibiscus sabdariffa* L.). We analysed a total of 92,974 annotated *de novo* assembled transcripts using *in-silico* approach, and identified 6,541 transcripts that encoded proteins containing tandem repeats with length of 20-60 amino acid residues. Domain analysis revealed a total of nine tandem repeat gene families in the transcriptome of roselle, which are the Ankyrin repeats (ANK), Armadillo repeats (ARM), Elongation Factor-hand domain repeats (EF-Hand), Huntingtin, Elongation factor 3, protein phosphatase 2A, yeast kinase TOR1 repeats (HEAT), Kelch repeats (Kelch), Leucine rich repeats (LLR), Pentatricopeptide repeats (PPR), Tetratricopeptide repeats (TPR) and WD40 repeats (WD40). Functional annotation analysis further matched 6,236 transcripts to 1,045 known proteins that contained tandem repeats including proteins implicated in plant development, protein-protein interaction, immunity, and abiotic stress responses. Our findings provide new insights into the occurrence of tandem repeats in the transcriptome and lay the foundation to elucidate the functional associations between tandem peptide repeats and proteins in roselle, and facilitate the identification of novel biotic and abiotic response related tandem repeats genes that may be useful in breeding improved varieties.

Keywords: Abiotic Stress, Biotic Stress, Protein-Protein Interaction, Roselle Calyx, Tandem Peptide Repeat

Abstrak. Tumbuhan adalah kaya dengan protein yang mengandungi ulangan peptida tandem. Ianya dipostulasikan bahawa kejadian famili gen ulangan tandem ini membantu adaptasi dan kemandirian tumbuhan dalam keadaan persekitaran mudarat. Dalam kajian ini, kami ingin mengenalpasti ulangan tandem di dalam transkriptom satu tumbuhan hortikultur tropikal berpotensi tinggi, roselle (*Hibiscus sabdariffa* L.). Kami menganalisis sejumlah 92,974 transkrip beranotasi berhimpun secara *de novo* menggunakan kaedah *in-silico*, dan mengenalpasti 6,541 transkrip yang mengkodkan protein mengandungi ulangan tandem dengan kepanjangan 20-60 residu asid amino. Analisis domain mendedahkan sejumlah sembilan famili gen ulangan tandem di dalam roselle, iaitu Ankyrin repeats (ANK), Armadillo repeats (ARM), Elongation Factor-hand domain repeats (EF-Hand), Huntingtin, Elongation factor 3, protein phosphatase 2A, yeast kinase TOR1 repeats (HEAT), Kelch repeats (Kelch), Leucine rich repeats (LLR), Pentatricopeptide repeats (PPR), Tetratricopeptide repeats (TPR) and WD40 repeats (WD40). Analisis penganotasian fungsian seterusnya memadankan 6,236 transkrip kepada 1,045 protein diketahui yang mengandungi ulangan tandem termasuk protein terlibat dalam perkembangan, interaksi antara protein, imuniti, semulajadi, dan tindak balas tekanan abiotik tumbuhan. Penemuan kami ini memberikan pandangan baharu terhadap kejadian ulangan tandem di dalam transkriptom rosel dan menyediakan asas untuk menjelaskan asosiasi fungsian antara ulangan peptida tandem dan protein di dalam rosel, dan membantu pengenalpastian gen ulangan tandem novel yang mungkin berguna dalam pembiakan varieti lebih baik.

Kata Kunci: Tekanan Abiotik, Tekanan Biotik, Interaksi Protein-Protein, Kaliks Rosel, Ulangan Peptida Tandem

INTRODUCTION

Hibiscus sabdariffa L., commonly known as roselle, is a multi-purpose tropical and subtropical shrub from the family of Malvaceae that possesses great therapeutic potentials. It is a valuable medicinal crop that can be cultivated in arid and semi-arid regions due to its high drought resilience. It has been used traditionally as folk medicines for decades particularly in regions where modern medicines are not readily accessible (Riaz and Copra 2018). Nonetheless, many scientific researches have reported the numerous beneficial properties of the extracts from roselle, including antibacterial, nephro-protective, hepato-protective, anti-diabetic and anti-hypertensive among others in recent years (Da-Costa-Rocha *et al.* 2014; von Dentz *et al.*

2020). Thus, it holds nutraceutical and pharmaceutical significance for industrial uses (Jamini and Aminul-Islam 2021), and is also frequently used as a functional ingredient in the food and beverages industries (Cid-Ortega and Guerrero-Beltrán 2015). The global roselle market size is expected to reach \$252.6 million by 2030 with a compound annual growth rate of 7.5% from 2021 to 2030. The calyx is the most valued part of the plant attributable to its rich content of phytochemicals such as phenolic compounds, anthocyanins, and organic acids (Morales-Luna *et al.* 2019). While the phytochemicals and their bioactivities have warranted roselle considerable amount of attention from the industry players and scientific communities, the genetic aspects of this plant remain mostly unexplored.

Tandem peptide repeats (TRs) in proteins are generally made of conserved block of amino acids and found abundantly in the proteomes of different organisms across all domains of life. The tandem repeat gene families are found copiously in the genomes of eukaryotes in comparison to prokaryotes (Delucchi *et al.* 2020). Plants house the most TRs, and their occurrences are anticipated to facilitate the adaptation and survival of plants in unfavourable environmental conditions. The massive expansion of TR gene families in land plants over the course of evolution has supported possible roles in plant growth, development, and survival (Sharma and Pandey 2016). Thus, studying the TRs and their association with functional proteins is essential to understand plant's adaptability. These tandem repeat regions provide a rich source for genetic variability with a wide range of possible genotypes at a given locus, which confers variability in adaptive ability in plants (Nithianantharajah *et al.* 2007). Although a substantial number of proteins that contains these adjacently repeated amino acid sequences have been identified, the biological significance for this association is still not entirely understood in plants.

Tandem repeats in protein-coding genes are mainly derived from internal gene duplication events (Luo and Nijveen 2014). These tandem repeats can be described based on the length of the repeating motif, the number of repeated units and/or the similarity among the repeat units (Delucchi *et al.* 2020). Tandem repeats with motif length of 20–60 amino acid residues are of particular interest to researchers because these TRs can function as structurally independent units and are likely to be folded as part of a protein domain displaying complex patterns as functional domain repeats (Jernigan and Bordenstein 2015). Examples of TRs in this category including leucine rich repeats (20-30 amino acid), pentatricopeptide repeats (~35 amino acid) and Armadillo repeats (~40 amino acid). Studies have demonstrated that TR-containing proteins play many important biological roles in diverse organisms. For instances, leucine rich repeats-containing proteins are associated with innate immunity in plants and animals (Ng and Xavier 2011), tetratricopeptide repeats in proteins are known to mediate protein-protein interactions and formation of protein complexes; while pentatricopeptide repeats-containing proteins involve in organelle biogenesis and function, are

fundamentally important for biological processes such as photosynthesis and plant development (Barkan and Small 2014). Although progresses have been achieved in some important plants such as *Arabidopsis thaliana* and *Oryza sativa* (Sharma and Pandey 2016), TRs remain largely understudied in most plant species.

To our best knowledge, no work has reported the TRs in roselle including their occurrence and association with functional proteins. In this study, we had analysed an in-house calyx transcriptome dataset of roselle and identified the transcripts that encode proteins containing TRs with length of 20-60 amino acid residues. We described all the TR gene families identified and their associated proteins. Subsequently, we discussed the three most abundant TR gene families in roselle, and further deliberated on crucial roles of TR gene families in plant's responses to biotic and abiotic stresses. These findings enhance our understanding of TRs in roselle, their potential functions, and facilitate the identification of novel biotic and abiotic response related TR gene families for their use in breeding improved plant varieties.

MATERIALS AND METHODS

The preparation of plant materials, RNA extraction, RNA-sequencing library, NGS sequencing and preliminary data analysis were performed in our previous study (Hamzah *et al.* 2022), and are briefly described in the subsections below.

Samples Preparation, RNA Extraction, RNA-Sequencing Library Preparation and Sequencing

Hibiscus sabdariffa var UMKL (a red variety registered as 'HS2' in the National Crop List of Malaysia) was used in this study. The planting material was obtained from the Department of Agriculture Terengganu, Malaysia. Mature calyces were collected from six healthy individual roselle plants. The calyces were collected at maturation stages III and IV according to number of days post-anthesis (DPA) as described in the guidelines provided by the Federal Agricultural Marketing Authority (FAMA) under the Ministry of Agriculture and Food Industry, Malaysia. These two maturation stages are particularly important in roselle cultivation because these stages are often associated with the quality of the calyces, and determined calyx harvesting time (Hamzah *et al.* 2022). Three biological replicates each were harvested on 32nd DPA (stage III) and 59th DPA (stage IV) respectively. The calyx samples were submerged in liquid nitrogen immediately upon collection and stored at -80°C until use. RNA samples were extracted from the calyx tissues using a RNeasy Plant Mini Kit (Qiagen, Germany) according to the manufacturer's instructions. The quantity and quality of the RNA extracts were

subsequently determined using Qubit Fluorometer 2.0 (Life Technologies Corporation, Carlsbad, USA), Agilent 2100 Bioanalyzer (Agilent Technologies, Germany) via a Pico Chip and 1% agarose gel electrophoresis. Six sequencing libraries were prepared using ScriptSeq™ v2 RNA-Seq Library preparation kit (Epicentre, Madison, USA) according to the manufacturer's protocols. The sequencing was performed on an Illumina NextSeq 500 platform to generate 76bp paired-end reads.

Preliminary Data Analysis of Raw RNA Sequencing Reads

Raw reads were screened against the PhiX genome sequence to remove sequencing reads belonging to the sequencing control library using Bowtie2 version 2.2.3 (Langmead and Salzberg 2012). The sequencing reads were then subjected to adapters removal and base quality trimming ($Q > 30$). Bases with low quality scores and trimmed reads of less than 35bp were discarded with its pair. Adapter removal, base quality trimming and discard of short reads were performed using BBDuk (BBTools version 36, <http://jgi.doe.gov/data-and-tools/bbtools/>). The sequencing short reads were submitted to NCBI-SRA under the BioProject number PRJNA664826 with the accession numbers; SRX9171164, SRX9171165, SRX9171166 (Stage III), and SRX9171161, SRX9171162, SRX9171163 (Stage IV).

Assembly, Functional Annotation, and Identification of Tandem Peptide Repeats

The clean reads were assembled (Hamzah *et al.* 2022) and used in the current study for TRs identification. Good quality reads were *de novo* assembled into transcripts representing full or partial fragment of the transcriptome using Trinity Version 2.2.0 (Haas *et al.* 2013). The assembled transcripts were searched for protein coding region using TransDecoder ver 2.0.1 (<https://transdecoder.github.io>). Subsequently, the transcript sequences and the predicted peptide sequences were subjected to sequence similarity search using BLAST version 2.2.31+ with an *E*-value cut-off of 10^{-6} (Camacho *et al.* 2009) to protein sequences of known functions available in the SwissProt database (Bairoch and Apweiler 2000), and scanned against protein domain motifs in the Protein Family database (Finn *et al.* 2014) using HMMER version 3.1b2 with a GA bit scores and *E*-value cut-off of 10^{-5} (Finn *et al.* 2011), respectively. Tandem repeat families containing tandem repeat with lengths of 20-60 amino acid residues were identified.

RESULTS

Good Quality Sequencing Reads and Transcript Assembly Statistics

The sequencing of the six calyx tissue RNA libraries generated a total of 237,321,698 good quality reads (Table 1), after data quality assessment (Hamzah *et al.* 2022). *De novo* assembly of these good quality reads yielded 221,334 contigs with minimum contig length of 161bp, maximum contig length of 25,718 bp, and N50 score of 491bp (Table 2). Each of the samples have between 89-99% reads mapped to the *De novo* assembled transcriptome, signifying that most of the sequencing reads have been assembled and accounted for.

Table 1. Good quality reads generated from each of the roselle sample.

Sample	Good quality Reads (Pairs)
Calyx stage III (Biological replicate 1)	38,507,642
Calyx stage III (Biological replicate 2)	40,661,821
Calyx stage III (Biological replicate 3)	41,267,234
Calyx stage IV (Biological replicate 1)	40,981,601
Calyx stage IV (Biological replicate 2)	35,605,342
Calyx stage IV (Biological replicate 3)	40,298,058
Total	237,321,698

Table 2. Transcript assembly statistic of the roselle calyx's transcriptome.

Total number of contigs	221,334
Minimum contig length (bp)	161
Maximum contig length (bp)	25,718
Average contig length (bp)	432
N50 score (bp)	491
Total contig length (bp)	95,824,738

Tandem Repeat Gene Families Identified

In the absence of a genome sequence, we utilized the calyx transcriptome generated from our previous study (Hamzah *et al.* 2022) to identify novel tandem peptide repeat gene families present in roselle. Although the tandem peptide repeat gene families identified here do not represent the complete collection of these gene families in roselle, it provides a glimpse of the tandem peptide repeat gene families present in the functional genome fraction of this plant for the first time. In this study, we had analysed a total of 92,974 annotated *de novo* assembled

transcripts and identified 6,541 transcripts that encode protein sequences containing tandem repeat (TRs) with the lengths of 20-60 amino acid residues. These transcripts corresponded to nearly 7.1% of the total annotated transcripts analysed. Nine TR gene families were identified, namely Leucine rich repeats (LRR), Pentatricopeptide repeats (PPR), WD40 repeats (WD40), Ankyrin repeats (ANK), Tetratricopeptide repeats (TPR), Kelch-like repeats (Kelch), Elongation Factor hand domain repeats (EF-Hand), Armadillo repeats (ARM), and Huntingtin, elongation factor 3, protein phosphatase 2A, yeast kinase TOR1 (HEAT) repeats. Leucine rich repeat and Pentatricopeptide repeat gene families, represented by 2,389 and 2,091 transcripts respectively were the two more abundant TR families found in this study. In contrast, only 114 transcripts were related to proteins containing the HEAT repeats, making this TR the least common repeat gene family found in the roselle's transcriptome. Functional annotation further linked 6,236 of the transcripts to 1,045 TR-containing proteins (Fig. 1). The remaining 305 transcripts also encoded amino acid sequences containing TRs, but they were either matched to uncharacterised proteins or did not match to any known proteins searched. The TRs identified, their associated proteins and the number of transcripts associated with each of the protein are displayed in Supplementary Table S1 – Table S9. The TR-containing proteins identified in this study signified possible correlation between these TRs with specific functional proteins.

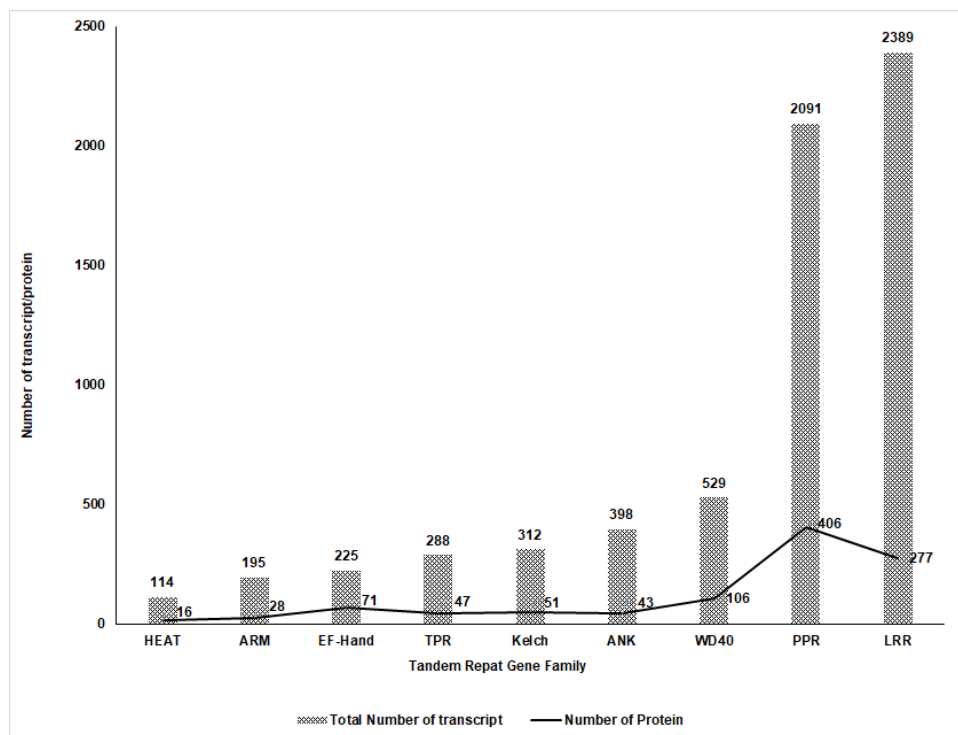


Figure 1: Tandem repeat Gene Families, the number of TR-containing proteins and the number of associated transcripts identified in the calyx transcriptome of roselle.

Tandem Repeat Gene Families and Associated Proteins

Leucine rich repeats

Leucine rich repeats (LRRs) are the most abundant tandem repeats identified from the roselle's calyx transcriptome. A total of 277 known proteins that encompassed LRRs were identified from 2,389 transcripts. Out of that, 108 proteins were identified as Leucine-rich repeat receptor-like protein kinases (LRR-RLK). Among the LRR-RLK, Leucine-rich repeat receptor-like Serine/Threonine protein kinases with 69 different proteins discovered (Supplementary Table S1) were the major type of LRR-RLK. Examples of LRR-RLK Serine/Threonine protein kinases identified were LRR receptor-like Serine/Threonine BAM1, LRR receptor-like Serine/Threonine EFR and LRR receptor-like Serine/Threonine ERECTA, LRR receptor-like Serine/Threonine RPK2 and LRR receptor-like Serine/Threonine FEI1. In addition, 39 non-Serine/Threonine types of LRR-RLK were also found. For instances, the LRR receptor-like protein kinases CORYNE, LRR receptor-like protein kinases TDR, LRR receptor-like protein kinases PXL2, LRR receptor-like protein kinases HAIKU2 and LRR receptor-like protein kinases HSL1. Furthermore, LRR receptor protein kinase (LRR-RK) with 23 proteins found contributed almost 21% to the overall LRR-kinases identified. Some of the LRR-RK proteins identified included Protein Brassinosteroid Insensitive 1, Brassinosteroid LRR receptor kinase, LRR receptor protein kinase EMS1, LRR receptor protein kinase CLAVATA1, Serine/threonine protein kinase BRI1-like and Somatic embryogenesis receptor kinase 1.

In addition to LRR-RLK and LRR-RK, a considerable number of disease resistance/susceptibility proteins (55 proteins) were also found to comprise LRR, which corresponded to approximately 19% of the total LRR-containing proteins identified. The Disease resistance RPP13-like proteins, Disease resistance RPP8-like proteins, Disease resistance protein RPS and Disease resistance protein RGA were among the more common ones found. The identification of these LRR-containing disease related proteins insinuated possible functional association between the LRR and these disease related proteins. Apart from that, we had also discovered many LRR-containing F-box proteins (54 proteins) in this study, such as the EIN3-binding F-box proteins, F-box protein FBW2 and F-box protein SKIP. Other LRR-containing proteins also included the plant intracellular RAS-related LRR proteins, LRR Extensin-like proteins, and transport inhibitor response proteins.

Pentatricopeptide repeats

Pentatricopeptide repeats (PPR), the second most abundant TR family found in the calyx transcriptome, were represented by 2,091 transcripts. Functional annotation assigned 96.3% of the transcripts to 395 PPR-containing protein homologs of *Arabidopsis thaliana* such as the

Pentatricopeptide repeat-containing proteins At1g62930, At1g71210, At1g73710 and At1g0990 (Supplementary Table S2). However, the specific functions of most of these proteins have not been corroborated. Thus, not much inference can be made pertaining to their functions. On the other hand, only 60 transcripts (less than 3% of the 2,091 transcripts) were annotated to 11 specific proteins with known functions, which were the Pentatricopeptide repeat-containing protein MRL1, Pentatricopeptide repeat protein for Germination on NaCl, Protein Nuclear Fusion Defective 5, Reticulon-like protein B22, Pentatricopeptide repeat-containing protein DOT4, Pentatricopeptide repeat-containing protein ELI1, two Transcription factor proteins, Proteinaceous RNase P1, Pentatricopeptide repeat-containing protein PNM and Pentatricopeptide repeat-containing protein OTP51. While, 16 transcripts showed no match to any proteins in the databases searched in this study. In total, 406 pentatricopeptide repeats-containing proteins (including those that matched to specific proteins and protein homologs of *Arabidopsis*) were discovered.

WD40 repeats

A total of 529 transcripts were found to encode amino acid sequences comprising the WD40 repeats. Functional annotation linked 452 transcripts (~85%) to 106 known proteins containing the WD40 repeat. However, the remaining 77 transcripts were either matched to uncharacterized proteins or showed no match to any proteins in the database searched. The Katanin p80 WD40 repeat-containing subunit B1 homolog and Topless-related proteins represented by 38 and 39 transcripts respectively, were the two more abundant WD40 repeat-containing proteins identified in this study (Supplementary Table S3). Other WD40 repeat-containing proteins identified were the WD-40 repeat-containing proteins MSI and Autophagy-related proteins. Transcriptional related proteins such as the Transcription initiation factor TFIID, Transcriptional corepressor LEUNIG, Zinc finger CCCH domain-containing proteins; RNA processing proteins such as the Pre-mRNA-processing factors; and ribosome biogenesis proteins such as U3 small nucleolar RNA-associated proteins and U4/U6 small nuclear ribonucleoprotein PRP4-like protein were also found to contain the WD40 repeats. In addition, several WD40 repeat containing proteins associated with cell cycle were also discovered.

Ankyrin repeats

Forty-three Ankyrin repeat (ANK) containing proteins were identified from 398 annotated transcripts. Most of the transcripts were annotated with known functional proteins except for four transcripts that did not match to any known proteins. The ADP-ribosylation factor GTPase-activating proteins consisting of AGD1 - AGD9 (96 transcripts) formed the biggest group of ANK repeat-containing proteins found in this study (Supplementary Table S4). The E3 ubiquitin-protein ligases represented by 80 transcripts were also among the more abundant

group of ANK repeat-containing proteins identified. In addition, several regulatory proteins that encompassed the ANK repeats were found, such as Regulatory Protein NPR 1, 2, 4 and 5. Other ANK repeat-containing proteins identified were Protein S-acyltransferase 24, Ankyrin repeat protein SKIP35, BTB/POZ domain-containing protein At2g04740 and Ankyrin repeat domain-containing protein EMB506.

Kelch-repeat

A total of 312 transcripts were found to encode amino acid sequences containing Kelch-repeat. Out of that, 301 transcripts matched to either known proteins (24 matches) or protein homologs from *Arabidopsis thaliana* (27 matches); while 11 transcripts did not match to any proteins in the database (Supplementary Table S5). F-box/kelch-repeat proteins (32 proteins represented by 195 transcripts) were the main type of proteins found. Among them, the F-box/kelch-repeat SKIP proteins (SKIP4, SKIP6, SKIP11, SKIP25, SKIP30) were the most abundant group of Kelch-containing proteins identified. The remaining 106 transcripts were annotated to various Kelch repeat-containing proteins such as Acyl-CoA-binding domain-containing protein 4 and 5, Adagio proteins 1 and 3, Serine/threonine-protein phosphatase BSL1, 2 and 3, RING finger protein B, and F-box protein AFR.

Tetratricopeptide repeats

We had identified 47 known proteins containing tetratricopeptide repeats (TPR) from 288 annotated transcripts. The Peptidyl-prolyl cis-trans isomerases consisted of six different isomerases (represented by 27 transcripts) were the most abundant type of TPR-proteins found (Supplementary Table S6). Three Hsp70-Hsp90 organizing proteins (represented by 24 transcripts), three Anaphase-promoting complex subunits (represented by 8 transcripts), and two UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferases (represented by 25 transcripts) were also found. We had also identified several envelope or membrane associated TPR-proteins such as the Outer envelope proteins, Alpha-soluble NSF attachment protein and Translocon complex. Other TPR-containing proteins included the Peroxisome biogenesis protein 5, Protein SGT1 homolog, Ethylene-overproduction protein 1, Clustered mitochondria protein and protein CTR9 homolog. Fifty-six transcripts out of the 288 transcripts were either matched to uncharacterised proteins or did not match to any known proteins in the searched databases in this study.

EF-hand domain repeats

Seventy EF-hand domain repeats-containing proteins were identified from 225 transcripts. Majority of these EF-hand-containing proteins were Calcium-dependent protein kinases (17 proteins represented by 46 transcripts) and Calcium-binding proteins (15 proteins represented

by 29 transcripts). Apart from that, five Calcineurin B-like proteins, three Calmodulin and seven Calmodulin-like proteins were also found. Other less abundant EF-hand-containing proteins found were the Calcineurin subunit B, Caltractin, Calumenin, Serine/threonine protein phosphatase 2A regulatory subunits, External alternative NAD(P)H-ubiquinone oxidoreductase B and Two pore potassium channel proteins (Supplementary Table S7). Nearly 73% of the EF-hand containing proteins identified in this study were either calcium-associated or -dependent proteins signifying their potential involvement in the regulation of calcium signalling.

Armadillo repeats

A total of 28 proteins containing Armadillo repeats were predicted from 195 transcripts. The U-box-domain proteins (13 proteins) represented by 55 transcripts were the most abundant group of ARM-containing proteins found in this study. This was followed by the ARM repeat-containing kinesin-like proteins (three proteins represented by 41 transcripts) and Importin subunit alpha proteins (six proteins represented by 39 transcripts). The remaining six ARM repeat-containing proteins were Proteins ARABIDILLO 1 and 2, ARM REPEAT PROTEIN INTERACTING WITH ABF2, Armadillo repeat-containing protein LFR, Vacuolar protein 8 and Phospholipase A I (Supplementary Table S8). While 14 transcripts which also encode amino acid sequences containing ARM repeats, they did not match to any known proteins in this study.

HEAT repeats

HEAT repeats were the rarest TRs found in the calyx transcriptome of roselle, with only 114 transcripts annotated to HEAT-containing proteins. Compared with other TRs, the number of HEAT repeat-containing proteins (16 proteins) found in this study were relatively lower. Among them, Protein SHOOT GRAVITROPISM 6 has the highest number of transcripts (52 transcripts) representing it (Supplementary Table S9). In addition, several regulatory or transcription related HEAT-containing proteins were found such as the Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A beta isoform, 26S proteasome non-ATPase regulatory subunit 1 homolog A, Phosphoinositide 3-kinase regulatory subunit 4 and Transcription initiation factor TFIID subunit 6. Two Microtubule-associated proteins were also found to contain HEAT-repeats, namely the protein TORTIFOLIA1 and protein SPIRAL2-like. Other proteins such as the CLIP-associated protein, Coatomer subunit gamma-2 protein, Importin Proteins and Protein MOR1 that encompassed the HEAT repeats were also identified.

DISCUSSION

Proteins encompass TRs are frequently associated with additional functional domains, which expands the functional diversity of proteins within the same gene family. Tandem repeats in proteins are frequently implicated in mediating protein-protein interactions. Thus, changes in these TR structural units will likely interfere with the protein functions (Parra *et al.* 2015). The occurrence of TRs was found to correlate positively with proteome size, which implies the probable association of TRs and protein's functions (Schaper and Anisimova 2015). To date, about one-half of all TRs recognized are common protein domains. Despite that, the tandem repeat gene families remain undiscovered in most plants, and the occurrence or association of the TRs with protein functions are still inadequately studied. The reason could be due to the high sequence divergence in repeating units that often poses technical challenges in studying repeat gene families (Parra *et al.* 2015). It is further complicated by the lack of genome sequences for most non-model plant species including roselle. In this study, we took advantage of our transcriptome dataset (Hamzah *et al.* 2022) to identify the TR-containing proteins in roselle. Although genome sequences give more comprehensive and accurate prediction of TRs occurrence, the current study has provided novel insights into the occurrence of TRs in the transcribed genome of roselle.

Comparison of Tandem Peptide Repeat Family in Arabidopsis, Rice and Roselle

Whole genome analyses on *Arabidopsis thaliana* and *Oryza sativa* had separately classified multigene repeat families in dicotyledon and monocotyledon, respectively. In *A. thaliana*, LRRs were identified as the most abundant repeat proteins, followed by PPRs and WD40 repeats proteins. While ANK, TPR, ARM and Kelch repeats proteins occurred at lower frequencies than the former three TR gene families, and HEAT repeats proteins were the least common in *Arabidopsis*. In *O. sativa*, PPRs were the most common type of repeat proteins found, followed by LRR and WD40. As for ANK, TTR, ARM, Kelch and HEAT repeats, similar levels of their relative abundances were observed in *O. sativa*, as in *A. thaliana* (Sharma and Pandey 2016). As for the EF-hand domain repeats, 250 and 243 EF-hand containing proteins were reported in the whole genomes of *Arabidopsis* (Day *et al.* 2002) and *Oryza* (Boonburapong and Buaboocha 2007), respectively.

In our work, we had also identified nine TRs gene families in the transcriptome of roselle, similar to those reported in the two model plants. Table 3 compares the number of TR-containing proteins identified from the genomes of *Arabidopsis thaliana* and *Oryza sativa* reported in previous studies, and the transcriptome of roselle in our study. In comparison to

the TR-containing proteins identified in the genomes of Arabidopsis and rice, the number of TR-containing proteins of the LRR-RLK, WD40, ANK, TPR, EF-Hand and ARM repeat gene families found in our study were relatively lower. The number of Kelch repeat-containing proteins discovered in our study fell between the counts observed in Arabidopsis and rice. While for the PPR and HEAT repeat families, the numbers of proteins found were comparable with those reported in Arabidopsis and rice.

Table 3. Comparison of Tandem Peptide Repeat-containing proteins predicted in the genomes of *Arabidopsis* and *Oryza*, and the number of TR-containing proteins identified in the transcriptome of roselle.

Tandem repeat	<i>A. thaliana</i>	<i>O. sativa</i>	<i>H. sabdariffa</i> (Current study)
LRR-RLK	223 (Wu et al. 2016)	292 (Hwang et al. 2011)	108
PPR	458 (Colcombet et al. 2013)	491 (Chen et al. 2018)	406
WD40	237 (Van Nocker & Ludwig 2003)	200 (Ouyang et al. 2012)	106
Kelch	97 (Sun et al. 2007)	28 (Sun et al. 2007)	51
ANK	105 (Becerra et al., 2004)	175 (Huang et al. 2009)	43
TPR	177 (Wei & Han 2017)	216 (Wei & Han 2017)	47
EF-Hand	250 (Day et al. 2002)	243 (Boonburapong & Buaboocha 2007)	70
ARM	108 (Mudgil et al. 2004)	158 (Sharma et al. 2014)	28
HEAT	17 (Sharma & Pandey 2016)	17 (Sharma & Pandey 2016)	16

The relatively lower number of TR-containing proteins found in our study was not unexpected, given the extensive studies and availability of the complete genome sequences of the two model plant species. The disparity in the data size and data type used (i.e. genome vs transcriptome) likely contributed to the lower number of TR-containing proteins found in our study. In addition, the calyx tissue-specific nature of our data also possibly influenced and constrained the number of TR-containing proteins identified in this study. Despite data limitation, the calyx's transcriptome has enabled the identification of nine TR gene families in roselle.

Although the number of TR-containing proteins found was lower in roselle compared to Arabidopsis and rice, it is worth noting the occurrence of tandem repeat proteins identified in roselle in term of abundance of each TR family was generally comparable to that of the Arabidopsis and rice, except for EF-Hand repeat-containing proteins. Notably, LRR, PPR and WD40 were found to be the more abundant tandem peptide repeats families in all three

species. The abundance of ANK, TPR, ARM, Kelch and HEAT repeat proteins are lower than the former three, with HEAT repeat proteins being the least common. However, there were only 70 EF-Hand repeat-containing proteins found in the transcriptome of roselle, in contrast to more than 200 EF-Hand repeat-containing proteins identified in the genomes of *Arabidopsis* and rice (Table 3). Apart from the differences in data sizes described above, the tissue- and maturation stage-specific nature of our transcriptome might have contributed to the discrepancy. Zeng *et al.* (2017) revealed that the number of EF-hand proteins in plant is directly influenced by tissue developmental stage and type. For instances, some EF-hand repeat-containing *CML* proteins are expressed exclusively in the flower; while others expressed specifically in the root tissue. Though the TRs predicted from the transcriptome in our study may not reflect the actual occurrence of TRs in the genome of roselle, the relative abundance of proteins in the tandem peptide repeat families identified, from the most common to the rarest, were fundamentally comparable to those reported in the whole genome analysis of *Arabidopsis thaliana* largely. This finding suggested that transcriptome data could be a valuable alternative to study TR gene family in non-model plants, providing insights to these important gene families in the absence of the genome sequences.

The Three Most Abundant Tandem Repeat Families in Roselle

Leucine rich repeats

Leucine rich repeats are the most prominent tandem repeat type discovered from the roselle's calyx transcriptome. Leucine-rich repeats are frequently associated with the mediation of protein-protein interaction of cell-cell communication and in plant innate immunity. Its slender, arc-shaped structure provides maximum surface area for protein-protein binding (Padmanabhan *et al.* 2009). Of the LRR-containing proteins identified in our study, more than half were leucine rich repeats receptor-like protein kinases (LRR-RLK). The LRR-RLKs are the largest subgroup of the RLK family and represent a complex gene family in plants (Liu *et al.* 2017; Li *et al.* 2018). All LRR-RLK proteins possess the N-terminal leucine rich repeats domain, a single transmembrane domain, and a C-terminal kinase domain, which are important for cellular signalling via interaction with various ligands (Dufayard *et al.* 2017; Chakraborty *et al.* 2019). In *Arabidopsis*, more than 600 RLKs were identified and out of those, 223 members are LRR-RLKs (Shiu *et al.* 2004; Wu *et al.* 2016). In rice, 1,100 RLKs had been identified with 292 members being the LRR-RLKs (Morillo and Tax 2006; Hwang *et al.* 2011). The expansion of the LRR-RLK gene family and their ubiquity in all land plants imply their importance in the evolution and survival of plants (Li *et al.* 2018). To date, only ~35% of the LRR-RLK members in *Arabidopsis* have been functionally assigned, and to a lesser extend in other species (Wu *et al.* 2016). The LRR receptor-like serine/threonine-protein kinases are the

most common group of LRR-RLK found in our study. This group of LRR-RLKs interacts with a diverse group of proteins to affect a wide array of biological processes related to signalling in plants (Afzal *et al.* 2008). Examples of LRR receptor-like serine/threonine-protein kinases found in our study include LRR receptor-like Serine/Threonine RPK2 which is a regulatory factor for anther development (Mizuno *et al.* 2007), LRR receptor-like serine/threonine-protein kinase At3g14840 (also known as LYSM RLK1-INTERACTING KINASE 1) required for Microbe Associated Molecular Pattern-triggered innate immunity (Le *et al.* 2014), and LRR receptor-like Serine/Threonine ERECTA regulates multiple development processes including flowering and stomatal cell differentiation (Shpak 2013; Pillitteri and Torii 2012).

We uncovered numerous transcripts that encode LRR-containing proteins that are related to disease resistance in this study. Plant disease resistance genes are an essential component of plant's genetic resistance mechanism. Majority of the immune receptors have the LRR domains, which display broad interaction surfaces and confer diverse classes of immune receptors with a platform to mediate protein-protein interactions (Padmanabhan *et al.* 2009; Ng and Xavier 2011; Jose *et al.* 2020). The highly mutable TRs enriched in LRRs commonly found in R genes of plants are believed able to facilitate resistance to newly emerging pathogen (Schaper and Anisimova 2015). Furthermore, majority of disease resistance genes (R genes) in plants encode the nucleotide-binding site LRR proteins (NBS-LRR) are actively involved in the detection of diverse plant pathogens (McHale 2006; Wan *et al.* 2012). NBS-LRR proteins are encoded by one of the most important gene families involved in plant immunity. Some of the NBS-LRR proteins discovered in our study were the disease resistance proteins RPP8, RPP13, RPS2, RPS4, RPS5 and RPM1. Based on past studies on *Arabidopsis*, RPP8 and RPP13 conferred resistance to downy mildew caused by *Peronospora parasitica* (Bittner-Eddy 2000; Mohr *et al.* 2010). RPS2, RPS4 and RPS5 and RPM1 were the key NBS-LRR proteins that provide resistance against *Pseudomonas syringae* (Tao *et al.* 2000; Shao *et al.* 2016).

LRR-RLKs involve in the control of plant growth and development were also found in our study, for example, Protein Brassinosteroid-insensitive-1 regulates plant growth and development via its interaction with Brassinosteroid, a common phytohormone in many plants (Friedrichsen *et al.* 2000; Navarro *et al.* 2015), LRR receptor protein kinase EMS1 and somatic embryogenesis receptor kinase 1 are involved in the determining anther cell fate in plants (Li *et al.* 2017), whereas CLAVATA1 and CLAVATA 2 are implicated in flower development (Jones *et al.* 2021). LRR-containing F-box proteins are one of the largest regulatory protein superfamilies in plants involved in protein degradation and ubiquitination controlling crucial processes such as floral organogenesis, senescence, embryogenesis, hormonal responses and seedling development (Lechner *et al.* 2006). The LRR at the C-terminal domain of F-box proteins is important for target protein recognition and binding (Hong *et al.* 2021). In our study,

22 LRR-containing F-box proteins were identified in the transcriptome of roselle. For example, ETHYLENE-INSENSITIVE binding F-box proteins (EIN3) is involved in ethylene signalling in plants by regulating EIN3 or EIN3-like turnover (Binder *et al.* 2007). F-box protein FBW2 regulates the expression of ARGONAUTE1 (AGO1) protein via proteolysis activities, and indirectly controls both microRNA- and small interfering RNA-directed silencing in plants (Hacquard *et al.* 2022).

Pentatricopeptide repeats

The abundance of PPR proteins found in our study agreed well with previous studies. PPR gene family is one of the largest and highly expanded nuclear-encoded protein families in terrestrial plants (Nakamura *et al.* 2012), with over 400 members in most sequenced plant species (Wang *et al.* 2021). Many PPR proteins play prominent roles in plant growth and development (Qu *et al.* 2019). These proteins are fundamentally sequence-specific RNA-binding proteins that serve as important site recognition factors for organellar RNA processing events such as RNA cleavage, degradation, stability, and splicing (Nakamura *et al.* 2012; Gutmann *et al.* 2020).

Breakthroughs in understanding PPR protein-RNA sequence recognition mechanisms have shed some lights on their involvement in regulating gene expressions of organellar transcripts (Barkan and Small 2014; Wang *et al.* 2021). Thus, they have profound effects on organelle biogenesis and functions including photosynthesis, respiration, and environmental responses (Barkan and Small 2014; Ren *et al.* 2020). We had identified PPR-containing proteins from the transcriptome of roselle that are involved in the transcriptional regulation of chloroplast and/or mitochondrial genes. For example, PPR-containing protein MRL1 acts on chloroplast gene and is essential for the production/stabilization of the processed transcript (Johnson *et al.* 2010), PPR-containing protein ELI1 and PPR-containing protein DOT4 are required for RNA editing in chloroplast (Hayes *et al.* 2013), Proteinaceous RNase P1 resides in mitochondria and chloroplast is necessary to support organellar function (Chen *et al.* 2019), Pentatricopeptide repeat-containing protein OTP51 is needed for intron splicing in the chloroplast (de Longevialle *et al.* 2008), and Pentatricopeptide repeat-containing protein PNM is involved in protein translation in mitochondria (Hammani *et al.* 2011). Mutational study on PPR protein-coding genes had also provided evidence that alteration to these genes caused dysfunction of mitochondria or/and chloroplasts, which resulted in growth retardation (Li *et al.* 2021).

WD40 repeats

The WD40 repeat is an important tandem repeat family found in multitude of eukaryotic proteins. The WD40 repeat-containing proteins act as scaffolding molecules that manoeuvre

activities of proteins involved in cellular, metabolic, and molecular pathways (Mishra *et al.* 2012). In our study, these repeats were found to associate with various proteins in which many of them act as regulatory factors or proteins involved in cell division, chromatin organization, pre-mRNA processing. Among the WD40 proteins identified in our study, Katanin p80 WD40 repeat-containing subunit B1 homolog is the most abundant WD40 repeat-containing protein. Katanin is a heterodimer microtubule severing ATPase essential for the organization and proper functions of the spindle, thus regulates the progression of cell division and plane orientation (Luptovčiak *et al.* 2017). It is also critical for fundamental biological processes such as cell elongation and morphogenesis (Takáč *et al.* 2017). Other WD40-proteins involved in cell division included Anaphase-promoting complex subunit essential in plant reproduction (Saleme *et al.* 2021) and Mitotic Checkpoint protein BUB3 required for gametophyte development (Lermontova *et al.* 2008).

TOPLESS proteins and topless-related proteins found in our study are among the most prominent transcriptional corepressors in plants (Martin-Arevalillo *et al.* 2017), which are involved in the regulations of hormone signalling (Dinesh *et al.* 2016) and interacts with WUSCHEL protein regulating plant stem cell homeostasis (Ikeda *et al.* 2009). The C-terminal WD40 repeats on the TOPLESS proteins serve as the protein-protein interaction surface (Collins *et al.* 2019). Other transcriptional factors such as Zinc finger CCCH domain-containing proteins are involved in mRNA processing (Peng *et al.* 2012), and Transcriptional co-repressor LEUNIG and its homologs regulate cell fate specification and hormone signalling (Sitaraman *et al.* 2008). The WD40 repeat-containing proteins MSI, which were also relatively abundant in the transcriptome of roselle, are a group of histone binding WD40-repeat proteins that form an essential part of a histone deacetylase complex that regulate gene expression (Mehdi *et al.* 2016). Several WD40 containing autophagy-related proteins (ATG16, ATG18a, ATG18c, ATG18d, ATG18f, ATG18g, ATG18h) were also found in our study. In general, ATGs are essential for maintaining cellular homeostasis and to prolong cell life under stress conditions. The evolutionary conserved ATG16 is one of the core autophagy proteins necessary for autophagosome formation and maturation (Xiong *et al.* 2018); while ATG18 was shown to respond to high salt and osmotic stress conditions (Wang *et al.* 2021).

Crucial Roles of Tandem Peptide Repeat Proteins Associated With Plant Resistance To Biotic And Abiotic Stresses for Breeding Resilient Plant Varieties

Tandem peptide repeat gene families are ubiquitously present in plant's genome and play pivotal roles in a myriad of biological processes essential for plant growth, development, and responses to biotic and abiotic stresses. For examples, LRR-RLPs are involved in both plant development (Kruijt *et al.* 2009) and plant-pathogen interaction (Wulff *et al.* 2009). Similarly,

ANK-repeat containing proteins are instrumental for plant growth and development, and responses to a range of biotic and abiotic stresses (Yuan *et al.* 2013). Acyl-CoA-binding domain-containing protein 4, the most abundant Kelch-peptide repeat-containing protein found in our study, is involved in transporting oleoyl-CoA from the chloroplast to endoplasmic reticulum (Leung *et al.* 2005) and regulating drought tolerance (Lai and Chye 2021) in *Arabidopsis*. Consequently, genes that encode TR-containing proteins may be expressed to serve diverse purposes in plants during normal or/and unfavourable growing conditions. The ubiquity of TRs in plants is possibly caused by their positive selection during evolution attributable to their correlation with plant growth and development, and adaptation. It is hypothesized that the ability of TR proteins to establish interaction with multiple ligands in response to biotic and abiotic stress contributed to their selection during evolution (Sharma and Pandey 2016). Through *in-silico* analysis, we found that other than the HEAT repeat families, all TR gene families discovered in this study have some members implicated in plant's responses to biotic and/or abiotic stresses. The identification of TR gene families and their respective proteins that are associated with plant's response to biotic and abiotic stresses is essential to promote the application of these genes in breeding more resilient varieties (Gottin *et al.* 2021).

For resistance to biotic stress, the identification of disease resistance LRR-containing proteins and LRR receptor-like serine/threonine protein kinase proteins strongly support the likely involvement of LRR domain in plant's immunity proteins. The Nucleotide Binding-LRR proteins such as Disease resistance protein TAO1 and Disease resistance protein ADR2 that were found in our study are some of the important proteins responsible for pathogen detection and host defence (DeYoung and Innes 2006). Other than LRR proteins, TPR containing Alpha-soluble NSF attachment protein also contributes to additive resistance in plant against pathogen (Lakhssassi *et al.* 2017); while TPR-protein SGT1 is required to induce host and non-host disease resistance in plants (Peart *et al.* 2002). The Importin subunit alpha that was found in our study is a group of ARM-repeat proteins, they function as nuclear transport receptors that implicated in nuclear translocation of immune regulatory proteins in animal and plant defence signalling (Wirthmueller *et al.* 2013). Phospholipase A 1 is also an ARM-containing protein, which confers resistance against biotic stress in plant (Canonne *et al.* 2011). Ankyrin repeat-containing proteins localized in the chloroplast are also involved in defence-related signalling in plants (Cao *et al.* 1997; Despres *et al.* 2001). Ankyrin-containing Protein ACCELERATED CELL DEATH 6 is a necessary defence response activator against virulent bacteria in plant (Lu *et al.* 2003); whereas Regulatory proteins NPR1, 3 and 4 are the key receptors interacting with plant defence hormone Salicylic acid, which play broad roles in plant immunity (Liu *et al.* 2020).

As for resistance against abiotic stress, we identified several Peptidyl-prolyl cis-trans isomerases and TPR repeat-containing thioredoxin TTL1 proteins in our study, which were reportedly involved in osmotic stress responses in plants (Sharma *et al.* 2003; Rosado *et al.* 2006). We also found other TPR proteins (Hsp70-Hsp90 organizing proteins 1, 2 and 3) that are particularly important for plant responses to heat stress. Hsp70-Hsp90 organizing proteins (HOP) function as co-chaperones that mediate protein-protein interactions in plant responses to heat stress (Kurek *et al.* 1999; Toribio *et al.* 2020). The transcriptional response in *Arabidopsis* was tremendously altered in *hop1 hop2 hop3* triple mutant during acclimatization to high temperature (Fernández-Bautista *et al.* 2018). Kelch-containing Acyl-CoA-binding domain-containing protein 4 (ACBP4) is involved in regulating drought tolerance (Lai and Chye 2021) in *Arabidopsis*. Ankyrin repeat-containing protein At3g12360 had been found to confer salt-stress tolerance in *A. thaliana* (Sakamoto *et al.* 2008). Several main EF-hand domain repeat calcium signal sensor proteins were found in our study, including Calcium-dependent protein kinases (CPDKs), Calmodulins (CaMs), Calmodulin-like proteins (CMLs) and Calcineurin B-like proteins (CBLs). These sensor proteins decipher the calcium signals triggered by environmental stimuli such as drought, heat and salt stresses, and thus are important for plant abiotic stress responses (Pandey *et al.* 2015; Shi *et al.* 2018). Leucine-rich repeat receptor-like kinase protein FLORAL ORGAN NUMBER1 (FON1), was demonstrated to increase drought tolerance in rice (Feng *et al.* 2014). Almost half of the ARM-repeat containing proteins found in our study are U-box/ARM proteins such as U-box/ARM protein 18 and 19. These proteins play crucial roles in plant to achieve proteome plasticity in response to environmental stress (Mudgil *et al.* 2004).

There are also TR-proteins found in our study that response to both abiotic and biotic stresses. Although LRR proteins are more often associated with plant's immunity, many LRR-proteins are also involved in regulating plant's responses to abiotic stress (Soltabayeva *et al.* 2022). Examples of LRR proteins that response to biotic and abiotic stresses found in our study are the LRR receptor-like kinases. Apart from LRR-proteins, other TR proteins such as the ANK repeats-containing E3 ubiquitin-protein ligases, play essential roles in regulating specific protein ubiquitination associated with plant immunity (Gao *et al.* 2021) and abiotic stress responses (Shu and Yang 2017). Pentatricopeptide repeat protein for germination on NaCl (PGN) is also implicated in biotic and abiotic stress tolerance in plants (Laluk *et al.* 2011). The WD40 repeat-containing Protein pleiotropic regulatory locus 1 is involved in regulating plant innate immunity and responses to environmental changes (Baruah *et al.* 2009).

The identification of TR-containing proteins responsible for plant's resistance to abiotic and biotic stresses may elucidate the molecular mechanisms behind plant stress adaptations. Studies on the natural variability, evolution and functions of TR family genes are critical to identify novel stress-tolerance genes and provide the foundation for their application in

breeding future climate-resilient varieties (Marone *et al.* 2013; Liao *et al.* 2017; Gottin *et al.* 2021). Furthermore, the discovery of highly mutable TRs enriched in LRRs that are commonly found in novel broad-spectrum R genes of plants and the NBS-LRR proteins, which provide effector triggered immunity in plants (Yu *et al.* 2022), could enhance the understanding of their function in disease resistance and further facilitate in producing disease-resistance varieties (Lee and Yeom, 2015; Zhang *et al.* 2021).

CONCLUSION

Tandem repeat often acts as a scaffold for substrate proteins, mediating protein-protein interactions that are fundamentally important from essential processes such as cell division to adaptive mechanisms such as responses to biotic and abiotic stresses in plants. In our study, we had identified nine TR gene families from the calyx tissue transcriptome of roselle and described their possible associated proteins. This study provides new insights to the occurrence and potential functions of the TR gene families in roselle, and thus established the fundamental knowledge for further studies to ascertain their biological functions in roselle. Given that the TR gene families prediction in our study was based on the calyx tissue transcriptome of roselle, the occurrence of the TRs might be influenced by the types of tissue studied and physiologically state of the plant. Nonetheless, the relative abundance of TR gene families predicted corresponded well with those found in the genomes of *Arabidopsis* and rice, where LRR, PPR and WD40 are more abundant compared to other TR gene families. The TR-proteins associated with plant's immunity and responses to abiotic stress found in this study, provide useful baseline data to explore novel candidate genes to facilitate breeding future tolerance varieties.

ACKNOWLEDGEMENTS

This work was supported by PUTRA IPS grant (GP-IPS/2017/9548800) provided by Universiti Putra Malaysia. We would also like to thank Dr. Benjamin Lau for proofreading the article.

AUTHORS' CONTRIBUTIONS

Christina Seok Yien Yong: Funding acquisition, Project administration, Visualization, Conceptualization, Resources, Methodology, Validation, Formal analysis, Writing – review & editing, Supervision.

Nur Atheeqah-Hamzah: Methodology, Investigation, Software, Validation, Data curation, Formal analysis.

REFERENCES

- Afzal A J, Wood A J and Lightfoot D A. (2008). Plant Receptor-Like Serine Threonine Kinases: Roles in Signaling and Plant Defense. *Molecular Plant Microbe Interactions* 21(5): 507–517. <https://doi.org/10.1094/MPMI-21-5-0507>.
- Bairoch A and Apweiler R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research* 28(1): 45-48. <https://doi.org/10.1093/nar/28.1.45>.
- Barkan A and Small I. (2014). Pentatricopeptide repeat proteins in plants. *Annual Review of Plant Biology* 65: 415-442. <https://doi.org/10.1146/annurev-arplant-050213-040159>.
- Baruah A, Šimková K, Hinch D K, Apel K and Laloi C. (2009). Modulation of ¹O₂-mediated retrograde signaling by the PLEIOTROPIC RESPONSE LOCUS 1 (PRL1) protein, a central integrator of stress and energy signaling. *The Plant Journal* 60: 22-32. <https://doi.org/10.1111/j.1365-313X.2009.03935.x>.
- Becerra C, Jahrmann T, Puigdomènech P and Vicient C M. (2004). Ankyrin repeat-containing proteins in Arabidopsis: characterization of a novel and abundant group of genes coding ankyrin-transmembrane proteins. *Gene* 340(1):111-21. <https://doi.org/10.1016/j.gene.2004.06.006>.
- Binder B M, Walker J M, Gagne J M, Emborg T J, Hemmann G, Bleecker A B and Vierstra R D. (2007). The Arabidopsis EIN3 binding F-Box proteins EBF1 and EBF2 have distinct but overlapping roles in ethylene signaling. *Plant Cell* 19(2):509-23. <https://doi.org/10.1105/tpc.106.048140>.
- Bittner-Eddy P D, Crute I R, Holub E B and Beynon J L. (2000). RPP13 is a simple locus in *Arabidopsis thaliana* for alleles that specify downy mildew resistance to different avirulence determinants in *Peronospora parasitica*. *The Plant Journal* 21(2):177-88. <https://doi.org/10.1046/j.1365-313x.2000.00664.x>.
- Boonburapong B and Buaboocha T. (2007) Genome-wide identification and analyses of the rice calmodulin and related potential calcium sensor proteins. *BMC Plant Biology* 7:4. <https://doi.org/10.1186/1471-2229-7-4>.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K and Madden T L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10: 421. <https://doi.org/10.1186/1471-2105-10-421>.
- Canonne J, Froidure-Nicolas S and Rivas S. (2011). Phospholipases in action during plant defense signaling. *Plant Signaling and Behavior* (1):13-8. <https://doi.org/10.4161/psb.6.1.14037>.

- Cao H, Glazebrook J, Clarke J D, Volko S and Dong X. (1997). The *Arabidopsis* NPR1 gene that controls systemic acquired resistance encodes a novel protein containing ankyrin repeats. *Cell* 88(1): 57-63. [https://doi.org/10.1016/s0092-8674\(00\)81858-9](https://doi.org/10.1016/s0092-8674(00)81858-9).
- Chakraborty S, Nguyen B, Wasti S D, Xu G. (2019). Plant leucine-rich repeat receptor kinase (LRR-RK): structure, ligand perception, and activation mechanism. *Molecules* 24(17): 1-37. <https://doi.org/10.3390/molecules24173081>.
- Chen G, Zou Y, Hu J and Ding Y. (2018). Genome-wide analysis of the rice PPR gene family and their expression profiles under different stress treatments. *BMC Genomics* 19: 720. <https://doi.org/10.1186/s12864-018-5088-9>.
- Chen T H, Sotomayor M and Gopalan V. (2019). Biochemical Studies Provide Insights into the Necessity for Multiple *Arabidopsis thaliana* Protein-Only RNase P Isoenzymes. *Journal of Molecular Biology* 431(3): 615-624. <https://doi.org/10.1016/j.jmb.2018.11.004>.
- Cid-Ortega S and Guerrero-Beltrán J A. (2015). Roselle calyces (*Hibiscus sabdariffa*), an alternative to the food and beverages industries: a review. *The Journal of Food Science and Technology* 52: 6859–6869. <https://doi.org/10.1007/s13197-015-1800-9>.
- Colcombet J, Lopez Obando M, Heurtevin L, Bernard C, Martin K, Berthomé R and Lurin C. (2013). Systematic study of subcellular localization of *Arabidopsis* PPR proteins confirms a massive targeting to organelles. *RNA Biology* 10(9): 1557-1575. <https://doi.org/10.4161/rna.26128>.
- Collins J, O'Grady K, Chen S and Gurley W. (2019). The C-terminal WD40 repeats on the TOPLESS co-repressor function as a protein–protein interaction surface. *Plant Molecular Biology* 100(1): 47-58. <https://doi.org/10.1007/s11103-019-00842-w>.
- Da-Costa-Rocha I, Bonnlaender B, Sievers H, Pischel I and Heinrich M. (2014). *Hibiscus sabdariffa* L.–A phytochemical and pharmacological review. *Food Chemistry* 165: 424-443. <https://doi.org/10.1016/j.foodchem.2014.05.002>.
- Day I S, Reddy V S, Shad Ali G and Reddy A S. (2002). Analysis of EF-hand-containing proteins in *Arabidopsis*. *Genome Biology* 3(10): RESEARCH0056. <https://doi.org/10.1186/gb-2002-3-10-research0056>.
- de Longevialle A F, Hendrickson L, Taylor N L, Delannoy E, Lurin C, Badger M, Millar A H and Small I. (2008). The pentatricopeptide repeat gene *OTP51* with two LAGLIDADG motifs is required for the cis-splicing of plastid *ycf3* intron 2 in *Arabidopsis thaliana*. *The Plant Journal* 56(1):157-68. <https://doi.org/10.1111/j.1365-313X.2008.03581.x>.
- Delucchi M, Schaper E, Sachenkova O, Elofsson A and Anisimova M. (2020). A new census of protein tandem repeats and their relationship with intrinsic disorder. *Genes* 11(4): 1-19. <https://doi.org/10.3390/genes11040407>.

- Despres B, Delseny M and Devic M (2001). Partial complementation of embryo defective mutations: a general strategy to elucidate gene function. *The Plant Journal* 27(2): 149-159. <https://doi.org/10.1046/j.1365-313x.2001.01078.x>.
- DeYoung, B J and Innes R W. (2006). Plant NBS-LRR proteins in pathogen sensing and host defense. *Nature Immunology* 7(12):1243-1249. <https://doi.org/10.1038/ni1410>.
- Dinesh D C, Villalobos L I A C and Abel S. (2016). Structural biology of nuclear auxin action. *Trends in Plant Science* 21(4): 302-316. <https://doi.org/10.1016/j.tplants.2015.10.019>.
- Dufayard J F, Bettembourg M, Fischer I, Droc G, Guiderdoni E, Périn C, Chantret N and Diévar A. (2017). New insights on leucine-rich repeats receptor-like kinase orthologous relationships in angiosperms. *Frontiers of Plant Science* 8: 1-18. <https://doi.org/10.3389/fpls.2017.00381>.
- Fernández-Bautista N, Fernández-Calvino L, Muñoz A, Toribio R, Mock H P and Castellano M M. (2018) HOP family plays a major role in long-term acquired thermotolerance in *Arabidopsis*. *Plant Cell and Environment* 41(8): 1852-1869. <https://doi.org/10.1111/pce.13326>.
- Finn R D, Bateman A, Clement J, Coggill P, Eberhardt R Y, Eddy S R, Heger A *et al* (2014). Pfam: the protein families database. *Nucleic Acids Research* D222-230. <https://doi.org/10.1093/nar/gkt1223>.
- Finn R D, Clements J and Eddy S R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* W29-37. <https://doi.org/10.1093/nar/gkr367>
- Friedrichsen D M, Joazeiro C A, Li J, Hunter T and Chory J. (2000). Brassinosteroid-insensitive-1 is a ubiquitously expressed leucine-rich repeat receptor serine/threonine kinase. *Plant Physiology* 123(4): 1247-1256. <https://doi.org/10.1104/pp.123.4.1247>.
- Gao C, Sun P, Wang, W and Tang D. (2021). *Arabidopsis* E3 ligase KEG associates with and ubiquitinates MKK4 and MKK5 to regulate plant immunity. *Journal of Integrative Plant Biology* 63(2): 327-339. <https://doi.org/10.1111/jipb.13007>.
- Gottin C, Dievar A, Summo M, Dro G., Périn C, Ranwez V and Chantret N. (2021). A new comprehensive annotation of leucine-rich repeat-containing receptors in rice. *The Plant Journal* 108(2):492-508. <https://doi.org/10.1111/tpj.15456>.
- Gutmann B, Royan S, Schallenberg-Rüdinger M, Lenz H, Castleden I R, McDowell R, Vacher M A, Tonti-Filippini J, Bond C S, Knoop V and Small I D. (2020). The Expansion and Diversification of Pentatricopeptide Repeat RNA-Editing Factors in Plants. *Molecular Plant* 13(2): 215–230. <https://doi.org/10.1016/j.molp.2019.11.002>.
- Haas B J, Papanicolaou A, Yassour M, Grabherr M, Blood P D, Bowden J, Couger M B *et al*. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity

- platform for reference generation and analysis. *Nature Protocol* 8: 1494–1512. <https://doi.org/10.1038/nprot.2013.084>.
- Hacquard T, Clavel M, Baldrich P, Lechner E, Pérez-Salamó I, Schepetilnikov M, Derrien B *et al.* (2022). The *Arabidopsis* F-box protein FBW2 targets AGO1 for degradation to prevent spurious loading of illegitimate small RNA. *Cell Reports* 39(2):110671. <https://doi.org/10.1016/j.celrep.2022.110671>.
- Hammani K, Gobert A, Hleibieh K, Choulier L, Small land Giegé P. (2011). An *Arabidopsis* dual-localized pentatricopeptide repeat protein interacts with nuclear proteins involved in gene expression regulation. *Plant Cell* 23(2):730-40. <https://doi.org/10.1105/tpc.110.081638>.
- Hamzah N A, Yong C S Y, Nallappan M. (2022) The first transcriptome dataset of roselle (*Hibiscus sabdariffa* L.) calyces during maturation. *Data in Brief* 45:108613. <https://doi.org/10.1016/j.dib.2022.108613>.
- Hayes M L, Giang K, Berhane B and Mulligan R M. (2013). Identification of two pentatricopeptide repeat genes required for RNA editing and zinc binding by C-terminal cytidine deaminase-like domains. *Journal of Biological Chemistry* 288(51):36519-29. <https://doi.org/10.1074/jbc.M113.485755>.
- Hong M, Kim J, Seo Y and Kim D. (2021). Regulation of Glycosylphosphatidylinositol-Anchored Protein (GPI-AP) Expression by F-Box/LRR-Repeat (FBXL) Protein in Wheat (*Triticum aestivum* L.). *Plants* 10(8):1606. <https://doi.org/10.3390/plants10081606>.
- Huang J, Zhao X, Yu H, Ouyang Y, Wang L and Zhang Q. (2009). The ankyrin repeat gene family in rice: genome-wide identification, classification and expression profiling. *Plant Molecular Biology* 71(3):207-26. <https://doi.org/10.1007/s11103-009-9518-6>.
- Hwang SG, Kim D S and Jang C S. (2011). Comparative analysis of evolutionary dynamics of genes encoding leucine-rich repeat receptor-like kinase between rice and *Arabidopsis*. *Genetica* 139:1023. <https://doi.org/10.1007/s10709-011-9604-y>.
- Ikeda M, Mitsuda N and Ohme-Takagi M. (2009). *Arabidopsis* WUSCHEL is a bifunctional transcription factor that acts as a repressor in stem cell regulation and as an activator in floral patterning. *Plant Cell* 21(11): 3493-3505. <https://doi.org/10.1105/tpc.109.069997>.
- Jamini T S and Aminul-Islam A K M. (2021). Roselle (*Hibiscus sabdariffa* L.): Nutraceutical and Pharmaceutical Significance, in: S M Sapuan, R Nadlene, A M Radzi and R A Ilyas (Eds). *Roselle*. Academic Press, Cambridge, 103-119.
- Jernigan K K and Bordenstein S R. (2015). Tandem-repeat protein domains across the tree of life. *PeerJ* 3: e732. <https://doi.org/10.7717/peerj.732>

- Johnson X, Wostrickoff K, Finazzi G, Kuras R., Schwarz C, Bujaldon S, Nickelsen J, Stern D B, Wollman F A and Vallon O. (2010). MRL1, a conserved Pentatricopeptide repeat protein, is required for stabilization of *rbcL* mRNA in *Chlamydomonas* and *Arabidopsis*. *Plant Cell* 22(1):234-48. [https://doi.org/ 10.1105/tpc.109.066266](https://doi.org/10.1105/tpc.109.066266).
- Jones D S, John A, VanDerMolen K R and Nimchuk Z L (2021). CLAVATA Signaling Ensures Reproductive Development in Plants across Thermal Environments. *Current Biology* 31(1):220-227. <https://doi.org/10.1016/j.cub.2020.10.008>.
- Jose J, Ghantasala S and Choudhury R S. (2020). *Arabidopsis* transmembrane receptor-like kinases (RLKs): A Bridge between extracellular signal and intracellular regulatory machinery. *International Journal of Molecular Sciences* 21(11): 1-29. <https://doi.org/10.3390/ijms21114000>.
- Kruijt M, de Kock M J D, de Wit P J G M. (2005). Receptor-like proteins involved in plant disease resistance. *Molecular Plant Pathology* 6:85–97. [https://doi: 10.1111/j.1364-3703.2004.00264.x](https://doi:10.1111/j.1364-3703.2004.00264.x).
- Kurek I, Aviezer K, Erel N, Herman E and Breiman A. (1999). The wheat peptidyl prolyl cis-trans-isomerase FKBP77 is heat induced and developmentally regulated. *Plant Physiology* 119(2): 693-704. [https://doi.org/ 10.1104/pp.119.2.693](https://doi.org/10.1104/pp.119.2.693).
- Lai S H and Chye M L. (2021). Plant Acyl-CoA-Binding Proteins—Their Lipid and Protein Interactors in Abiotic and Biotic Stresses. *Cells* 10(5): 1-21. <https://doi.org/10.3390/cells10051064>.
- Lakhssassi N, Liu S M, Bekal S, Zhou Z, Colantonio V, Lambert K, Berakat A and Meksem K. (2017). Characterization of the Soluble NSF Attachment Protein gene family identifies two members involved in additive resistance to a plant pathogen. *Scientific Reports* 7: 45226. <https://doi.org/10.1038/srep45226>.
- Laluk K, Abuqamar S and Mengiste T. (2011). The *Arabidopsis* mitochondria-localized pentatricopeptide repeat protein PGN functions in defense against necrotrophic fungi and abiotic stress tolerance. *Plant Physiology* 156(4): 2053-68. <https://doi.org/10.1104/pp.111.177501>.
- Langmead B and Salzberg S L (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9(4): 357–359. <https://doi.org/10.1038/nmeth.1923>.
- Le M H, Cao Y, Zhang X C and Stacey G. (2014). LIK1, a CERK1-interacting kinase, regulates plant immune responses in *Arabidopsis*. *PLOS One* 9(7): e102245. <https://doi.org/10.1371/journal.pone.0102245>.
- Lechner E, Achard P, Vansiri A, Potuschak T and Genschik P. (2006). F-box proteins everywhere. *Current Opinion in Plant Biology* 9(6):631-8. <https://doi.org/10.1016/j.pbi.2006.09.003>.

- Lee H and Yeom S. (2015). Plant NB-LRR proteins: tightly regulated sensors in a complex manner. *Briefings in Functional Genomics* 14(4): 233–242. <https://doi.org/10.1093/bfpg/elv012>.
- Lermontova I, Fuchs J and Schubert I. (2008). The Arabidopsis checkpoint protein Bub3.1 is essential for gametophyte development. *Frontiers in Bioscience* 13:5202-11. <https://doi.org/10.2741/3076>.
- Leung K C, Li H Y, Mishra G and Chye M L. (2005). ACBP4 and ACBP5, novel *Arabidopsis* acyl-CoA-binding proteins with kelch motifs that bind oleoyl-CoA. *Plant Molecular Biology* 55(2): 297-309. <https://doi:10.1007/s11103-004-0642-z>.
- Li X, Salman A, Guo C, Yu J, Cao S, Gao X, Li W, Li H and Guo Y. (2018). Identification and Characterization of *LRR-RLK* Family Genes in Potato Reveal Their Involvement in Peptide Signaling of Cell Fate Decisions and Biotic/Abiotic Stress Responses. *Cells* 7(9): 120. <https://doi.org/10.3390/cells7090120>.
- Li X, Sun M, Liu S, Teng Q, Li S and Jiang Y. (2021). Functions of PPR Proteins in Plant Growth and Development. *International Journal of Molecular Sciences* 22(20):11274. <https://doi.org/10.3390/ijms222011274>.
- Li Z, Wang Y, Huang J, Ahsan N, Biener G, Paprocki J, Thelen J J, Raicu V and Zhao D. (2017). Two SERK receptor-like kinases interact with EMS1 to control anther cell fate determination. *Plant Physiology* 173(1): 326-337. <https://doi.org/10.1104/pp.16.01219>.
- Liao Y, Hu C, Zhang X, Cao X, Xu Z, Gao X, Li L, Zhu J and Chen R. (2017). Isolation of a novel leucine-rich repeat receptor-like kinase (*OsLRR2*) gene from rice and analysis of its relation to abiotic stress responses. *Biotechnology, Biotechnological Equipment* 31(4): 717-724. <https://doi.org/10.1080/13102818.2016.1242377>
- Liu P L, Du L, Huang Y, Gao S M and Yu M. (2017). Origin and diversification of leucine-rich repeat receptor-like protein kinase (LRR-RLK) genes in plants. *BMC Evolutionary Biology* 17(1): 1-16. <https://doi.org/10.1186/s12862-017-0891-5>.
- Liu Y, Sun T, Sun Y, Zhang Y, Radojičić A, Ding Y, Tian H *et al.* (2020). Diverse Roles of the Salicylic Acid Receptors NPR1 and NPR3/NPR4 in Plant Immunity. *Plant Cell* 32(12): 4002-4016. <https://doi.org/10.1105/tpc.20.00499>.
- Lu H, Rate D N, Song J T and Greenberg J T. (2003). ACD6, a novel ankyrin protein, is a regulator and an effector of salicylic acid signaling in the Arabidopsis defense response. *Plant Cell* 15(10): 2408-20. <https://doi.org/10.1105/tpc.015412>.
- Luo H and Nijveen H. (2014). Understanding and identifying amino acid repeats. *Briefings in Bioinformatics* 15(4): 582–591. <https://doi.org/10.1093/bib/bbt003>.

- Luptovčiak I, Komis G, Takáč T, Ovečka M and Šamaj J. (2017). Katanin: A Sword Cutting Microtubules for Cellular, Developmental, and Physiological Purposes. *Frontiers of Plant Science*. 8:1982. <https://doi.org/10.3389/fpls.2017.01982>.
- Marone D, Russo M A, Laidò G, De Leonardis A M and Mastrangelo A M. (2013). Plant Nucleotide Binding Site-Leucine-Rich Repeat (NBS-LRR) genes: Active guardians in host defense responses. *International Journal of Molecular Sciences* 14. 7302-7326. <https://doi.org/10.3390/ijms14047302>.
- Martin-Arevalillo R, Nanao M H, Larrieu A, Vinos-Poyo T, Mast D, Galvan-Ampudia C, Brunoud G, Vernoux T, Dumas R and Parcy F. (2017). Structure of the *Arabidopsis* TOPLESS corepressor provides insight into the evolution of transcriptional repression. *The Proceedings of the National Academy of Sciences* 114(30): 8107–8112. <https://doi.org/10.1073/pnas.1703054114>.
- McHale L, Tan X, Koehl P and Michelmore R W. (2006). Plant NBS-LRR proteins: adaptable guards. *Genome Biology* 7: 212. <https://doi.org/10.1186/gb-2006-7-4-212>.
- Mehdi S, Derkacheva M, Ramström M, Kraleman L, Bergquist J, Hennig L. (2016). The WD40 domain protein MSI1 functions in a histone deacetylase complex to fine-tune abscisic acid signaling. *Plant Cell* 28(1): 42-54. <https://doi.org/10.1105/tpc.15.00763>.
- Mishra A K, Puranik S and Prasad M. (2012). Structure and regulatory networks of WD40 protein in plants. *Journal of Plant Biochemistry Biotechnology* 21(1): 32-39. <https://doi.org/10.1007/s13562-012-0134-1>.
- Mohr T J, Mammarella N D, Hoff T, Woffenden B J, Jelesko J G and McDowell J M. (2010). The *Arabidopsis* downy mildew resistance gene RPP8 is induced by pathogens and salicylic acid and is regulated by W box cis elements. *Molecular Plant Microbe Interactions* 10: 1303-15. <https://doi.org/10.1094/MPMI-01-10-0022>.
- Morales-Luna E, Pérez-Ramírez I F, Salgado L M, Castaño-Tostado E, Gómez-Aldapa C A and Reynoso-Camacho R. (2019). The main beneficial effect of roselle (*Hibiscus sabdariffa*) on obesity is not only related to its anthocyanin content. *Journal of the Science of Food and Agriculture* 99: 596-605. <https://doi.org/10.1002/jsfa.9220>.
- Morillo S A and Tax F E. (2006). Functional analysis of receptor-like kinases in monocots and dicots. *Current Opinion in Plant Biology* 9(5): 460-469. <https://doi.org/10.1016/j.pbi.2006.07.009>.
- Nakamura T, Yagi Y and Kobayashi K. (2012). Mechanistic insight into pentatricopeptide repeat proteins as sequence-specific RNA-binding proteins for organellar RNAs in plants. *Plant Cell Physiology* 53(7): 1171-1179. <https://doi.org/10.1093/pcp/pcs069>.
- Navarro C, Moore J, Ott A, Baumert E, Mohan A, Gill K S and Sandhu D. (2015). Evolutionary, comparative and functional analyses of the Brassinosteroid receptor gene, *BRI1*, in

- wheat and its relation to other plant genomes. *PLOS One* 10(5): 1-16. <https://doi.org/10.1371/journal.pone.0127544>.
- Mizuno S, Osakabe Y, Maruyama K, Ito T, Osakabe K, Sato T, Shinozaki K and Yamaguchi-Shinozaki K. (2007). Receptor-like protein kinase 2 (RPK 2) is a novel factor controlling anther development in *Arabidopsis thaliana*. *The Plant Journal* 50(5): 751-66. <https://doi.org/10.1111/j.1365-313X.2007.03083.x>.
- Mudgil Y, Shiu S H, Stone S L, Salt J N and Goring D R. (2004). A large complement of the predicted Arabidopsis ARM repeat proteins are members of the U-box E3 ubiquitin ligase family. *Plant Physiology* 134(1):59-66. <https://doi.org/10.1104/pp.103.029553>
- Ng A and Xavier R J. (2011). Leucine-rich repeat (LRR) proteins: integrators of pattern recognition and signaling in immunity. *Autophagy* 7(9): 1082-1084. <https://doi.org/10.4161/auto.7.9.16464>.
- Nithianantharajah J and Hannan A J. (2007). Dynamic mutations as digital genetic modulators of brain development, function and dysfunction. *BioEssays* 29:525–535. <https://doi.org/10.1002/bies.20589>.
- Ouyang Y, Huang X, Lu Z and Yao J. (2012). Genomic survey, expression profile and co-expression network analysis of *OsWD40* family in rice. *BMC Genomics* 13:100. <https://doi.org/10.1186/1471-2164-13-100>.
- Padmanabhan M, Cournoyer P and Dinesh-Kumar S P (2009). The leucine-rich repeat domain in plant innate immunity: a wealth of possibilities. *Cellular Microbiology* 11(2): 191-198. <https://doi.org/10.1111/j.1462-5822.2008.01260.x>.
- Pandey G K, Kanwar P, Singh A, Steinhorst L, Pandey A, Yadav A K, Tokas I *et al.* (2015). Calcineurin B-Like Protein-Interacting Protein Kinase CIPK21 Regulates Osmotic and Salt Stress Responses in Arabidopsis. *Plant Physiology* 169(1): 780-92. <https://doi.org/10.1104/pp.15.00623>.
- Parra R G, Espada R, Verstraete N and Ferreira D U. (2015). Structural and Energetic Characterization of the Ankyrin Repeat Protein Family. *PLoS Computational Biology* 11(12): e1004659. <https://doi.org/10.1371/journal.pcbi.1004659>.
- Peart J R, Lu R, Sadanandom A, Malcuit I, Moffett P, Brice D C, Schauser L *et al.* (2002). Ubiquitin ligase-associated protein SGT1 is required for host and nonhost disease resistance in plants. *The Proceedings of the National Academy of Sciences* 99(16):10865-9. <https://doi.org/10.1073/pnas.152330599>.
- Peng X, Zhao Y, Cao J, Zhang W, Jiang H, Li X, Ma Q, Zhu S and Cheng B. (2012). CCCH-type zinc finger family in maize: genome-wide identification, classification and expression profiling under abscisic acid and drought treatments. *PLOS One* 7(7): e40120. <https://doi.org/10.1371/journal.pone.0040120>.

- Pillitteri L J and Torii K U. (2012). Mechanisms of stomatal development. *Annual Review of Plant Biology* 63, 591–614. <https://doi.org/10.1146/annurev-arplant-042811-105451>
- Qu K, Wei L, Yu J and Wang. (2019). Identifying plant pentatricopeptide repeat coding gene/protein using mixed feature extraction methods. *Frontiers of Plant Science* 9: 1-10. <https://doi.org/10.3389/fpls.2018.01961>.
- Rosado A, Schapire A L, Bressan R A, Harfouche A L, Hasegawa P M, Valpuesta V and Botella M A. (2006). The Arabidopsis Tetratricopeptide Repeat-Containing Protein TTL1 Is Required for Osmotic Stress Responses and Abscisic Acid Sensitivity1. *Plant Physiology* 142(3): 1113–1126. <https://doi.org/10.1104/pp.106.085191>.
- Ren R C, Wang L L, Zhang L, Zhao Y J, Wu J W, Wei Y M, Zhang X S and Zhao X Y. (2020). DEK43 is a P-type pentatricopeptide repeat (PPR) protein responsible for the Cis - splicing of nad4 in maize mitochondria. *Journal of Integrative Plant Biology* 62(3): 299–313. <https://doi.org/10.1111/jipb.12843>.
- Riaz G and Chopra R. (2018). A review on phytochemistry and therapeutic uses of *Hibiscus sabdariffa* L. *Biomedicine & Pharmacotherapy* 102: 575-586. <https://doi.org/10.1016/j.biopha.2018.03.023>.
- Sakamoto H, Matsuda O and Iba K. (2008). *ITN1*, a novel gene encoding an ankyrin-repeat protein that affects the ABA-mediated production of reactive oxygen species and is involved in salt-stress tolerance in *Arabidopsis thaliana*. *The Plant Journal* 56: 411-422. <https://doi.org/10.1111/j.1365-313X.2008.03614.x>.
- Saleme M L S, Andrade I R and Eloy N B. (2021). The Role of Anaphase-Promoting Complex/Cyclosome (APC/C) in Plant Reproduction. *Frontiers of Plant Science* 12: 642934. <https://doi.org/10.3389/fpls.2021.642934>.
- Schaper E and Anisimova M. (2015). The evolution and function of protein tandem repeats in plants. *New Phytologist* 206(1): 397-410. <https://doi.org/10.1111/nph.13184>.
- Shi S, Li S, Asim M, Mao J, Xu D, Ullah Z, Liu G, Wang Q and Liu H. (2018). The *Arabidopsis* Calcium-Dependent Protein Kinases (CDPKs) and Their Roles in Plant Growth Regulation and Abiotic Stress Responses. *International Journal of Molecular Sciences* 9(7):1900. <https://doi.org/10.3390/ijms19071900>.
- Shao Z Q, Xue J Y, Wu P, Zhang Y M, Wu Y, Hang Y Y, Wang B and Chen J Q. (2016). Large-Scale Analyses of Angiosperm Nucleotide-Binding Site-Leucine-Rich Repeat Genes Reveal Three Anciently Diverged Classes with Distinct Evolutionary Patterns. *Journal of Plant Physiology* 170(4): 2095-109. <https://doi.org/10.1104/pp.15.01487>.
- Sharma A D, Wajapeyee N, Yadav V and Singh P. (2003). Stress-Induced Changes in Peptidyl-Prolyl *cis-trans* Isomerase Activity of *Sorghum bicolor* Seedlings. *Biologia Plantarum* 47: 367–371. <https://doi.org/10.1023/B:BIOP.0000023879.74558.48>.

- Sharma M and Pandey G K. (2016). Expansion and function of repeat domain proteins during stress and development in plants. *Frontiers of Plant Science* 6: 1-15. <https://doi.org/10.3389/fpls.2015.01218>.
- Sharma M, Singh A, Shankar A, Pandey A, Baranwal V, Kapoor S, Tyagi A K and Pandey G K. (2014). Comprehensive expression analysis of rice Armadillo gene family during abiotic stress and development. *DNA Research* 21(3): 267-83. <https://doi.org/10.1093/dnares/dst056>.
- Shiu S H, Karlowski W M, Pan R, Tzeng Y H, Mayer K F and Li W H. (2004). Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. *Plant Cell* 16(5): 1220-1234. <https://doi.org/10.1105/tpc.020834>.
- Shu K and Yang W. (2017). E3 ubiquitin ligases: ubiquitous actors in plant development and abiotic stress responses. *Plant and Cell Physiology* 58(9): 1461-1476. <https://doi.org/10.1093/pcp/pcx071>.
- Shpak E D. (2013). Diverse roles of ERECTA family genes in plant development. *Journal of Integrative Plant Biology* 55(12):1238-50. <https://doi.org/10.1111/jipb.12108>.
- Sitaraman J, Bui M and Liu Z. (2008). *LEUNIG_HOMOLOG* and *LEUNIG* Perform Partially Redundant Functions during Arabidopsis Embryo and Floral Development. *Plant Physiology* 147(2): 672–681. <https://doi.org/10.1104/pp.108.115923>.
- Soltabayeva A, Dauletova N, Serik S, Sandybek M, Omondi J O, Kurmanbayeva A and Srivastava S. (2022). Receptor-like Kinases (LRR-RLKs) in Response of Plants to Biotic and Abiotic Stresses. *Plants* 11(19):2660. <https://doi.org/10.3390/plants11192660>.
- Sun Y, Zhou X and Ma H. (2007). Genome-wide analysis of Kelch repeat- containing F-box family. *Journal of Integrative Plant Biology* 49(6): 940–952. <https://doi.org/10.1111/j.1744-7909.2007.00498.x>.
- Takáč T, Šamajová O, Pechan T, Luptovčíak I and Šamaj J. (2017). Feedback microtubule control and microtubule-actin cross-talk in *Arabidopsis* revealed by integrative proteomic and cell biology analysis of KATANIN 1 mutants. *Molecular & Cellular Proteomics* 16(9): 1591-1609. <https://doi.org/10.1074/mcp.M117.068015>.
- Tao Y, Yuan F, Leister R T, Ausubel F M and Katagiri F. (2000). Mutational analysis of the Arabidopsis nucleotide binding site-leucine-rich repeat resistance gene *RPS2*. *Plant Cell* 12(12): 2541-2554. <https://doi.org/10.1105/tpc.12.12.2541>.
- Toribio R, Mangano S, Fernández-Bautista N, Muñoz A and Castellano M. (2020). HOP, a Co-chaperone Involved in Response to Stress in Plants. *Frontiers of Plant Science* 11: 1-8. <https://doi.org/10.3389/fpls.2020.591940>.

- van Nocker S and Ludwig P. (2003). The WD-repeat protein superfamily in *Arabidopsis*: conservation and divergence in structure and function. *BMC Genomics* 4(1): 50. <https://doi.org/10.1186/1471-2164-4-50>.
- Von Dentz K E, Silva B S, Queiroz E A, Bomfim G F, Nascimento A F, Sugizaki M M, Luvizotto R A M. (2020). *Hibiscus sabdariffa* ethanolic extract modulates adipokine levels, decreases visceral fat and improves glycemic profile in high-fat/sugar diet-induced obese rats. *Nutrition and Food Science* 51(2): 222-233. <https://doi.org/10.1108/NFS-03-2020-0092>.
- Wang H, Ding Z, Gou M, Hu J, Wang Y, Wang L. Wang Y *et al.* (2021). Genome-wide identification, characterization, and expression analysis of tea plant autophagy-related genes (CsARGs) demonstrates that they play diverse roles during development and under abiotic stress. *BMC Genomics* 22(1): 121. <https://doi.org/10.1186/s12864-021-07419-2>.
- Wan H, Yuan W, Ye Q, Wang R, Ruan M, Li Z, Zhou G *et al.* (2012). Analysis of TIR- and non-TIR-NBS-LRR disease resistance gene analogous in pepper: characterization, genetic variation, functional divergence and expression patterns. *BMC Genomics* 13:502. <https://doi.org/10.1186/1471-2164-13-502>.
- Wang X, An Y, Xu P and Xiao J. (2021). Functioning of PPR Proteins in organelle RNA metabolism and chloroplast biogenesis. *Frontiers of Plant Science* 12: 627501. <https://doi.org/10.3389/fpls.2021.627501>.
- Wei K and Han P. (2017). Comparative functional genomics of the TPR gene family in *Arabidopsis*, rice and maize. *Molecular Breeding* 37: 152. <https://doi.org/10.1007/s11032-017-0751-4>.
- Wirthmueller L, Roth C, Banfield M J and Wiermer M. (2013). Hop-on hop-off: importin- α -guided tours to the nucleus in innate immune signaling. *Frontiers of Plant Science* 4, 1-8. <https://doi.org/10.3389/fpls.2013.00149>.
- Wu Y, Xun Q, Guo Y, Zhang J, Cheng K, Shi T, He K, Hou S, Gou X and Li J. (2016). Genome-Wide Expression Pattern Analyses of the Arabidopsis Leucine-Rich Repeat Receptor-Like Kinases. *Molecular Plant* 9(2):289-300. <https://doi.org/10.1016/j.molp.2015.12.011>.
- Wulff B B, Chakrabarti A, Jones D A. (2009). Recognition specificity and evolution in the tomato-*Cladosporium fulvum* pathosystem. *Molecular Plant Microbe Interaction* 22:1191–1202. <https://doi.org/10.1094/MPMI-22-10-1191>.
- Xing H, Fu X, Yang C, Tang X, Guo L, Li C, Xu C and Luo K. (2018). Genome-wide investigation of pentatricopeptide repeat gene family in poplar and their expression analysis in response to biotic and abiotic stresses. *Scientific Reports* 8: 2817. <https://doi.org/10.1038/s41598-018-21269-1>.

- Xiong Q, Li W, Li P, Yang M, Wu C and Eichinger L. (2018). The Role of ATG16 in Autophagy and The Ubiquitin Proteasome System. *Cells* 8(1):2. <https://doi.org/10.3390/cells8010002>.
- Yu S, Ali J, Zhou S, Ren G, Xie H, Xu J, Yu X *et al.* (2022). From Green Super Rice to green agriculture: Reaping the promise of functional genomics research. *Molecular Plant* 15(1): 9-26. <https://doi.org/10.1016/j.molp.2021.12.001>.
- Yuan X, Zhang S, Qing X, Sun M, Liu S, Su H, Shu H, Li X. (2013). Superfamily of ankyrin repeat proteins in tomato. *Gene* 523(2): 126-36. [https://doi:10.1016/j.gene.2013.03.122](https://doi.org/10.1016/j.gene.2013.03.122).
- Zhang Q, Wang Y, Wei H, Fan W, Xu C and Li T. (2021). CC_R-NB-LRR proteins MdRNL2 and MdRNL6 interact physically to confer broad-spectrum fungal resistance in apple (*Malus × domestica*). *The Plant Journal* 108: 1522-1538. <https://doi.org/10.1111/tpj.15526>.
- Zeng H, Zhang Y, Zhang X, Pi E, Zhu Y. (2017). Analysis of EF-Hand Proteins in Soybean Genome Suggests Their Potential Roles in Environmental and Nutritional Stress Signaling. *Frontiers of Plant Science* 8:877. doi: 10.3389/fpls.2017.00877.
- Zeytuni N, Zarivach R. (2012) Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure* 20(3):397-405. [https://doi:10.1016/j.str.2012.01.006](https://doi.org/10.1016/j.str.2012.01.006).