Saudi Journal of Biological Sciences 30 (2023) 103676

HOSTED BY

Contents lists available at ScienceDirect

Saudi Journal of Biological Sciences

journal homepage: www.sciencedirect.com

Original article

Phoenix dactylifera (date palm; Arecaceae) putative lectin homologs: Genome-wide search, architecture analysis, and evolutionary relationship

Makarim Elfadil M. Osman [a,1], Rieham Sallah H. Osman [a,1], Sara A.A Elmubarak [a,1], Amina I. Dirar [b], Emadeldin Hassan E. Konozy [a,c,*]

[a] Department of Biotechnology, Africa City of Technology (ACT), Khartoum, Sudan
[b] Medicinal, Aromatic Plants and Traditional Medicine Research Institute (MAPTRI), National Center for Research, Mek Nimr Street, Khartoum, Sudan
[c] Pharmaceutical Research and Development, Centre Faculty of Pharmacy, Karary University, Omdurman, Khartoum State, Sudan

ARTICLE INFO

Article history:
Received 25 February 2023
Revised 10 April 2023
Accepted 27 April 2023
Available online 4 May 2023

Keywords:
Phoenix dactylifera
Ture lectins
Domain rearrangements
Localization
ACP

ABSTRACT

The date palm, *Phoenix dactylifera*, is a vital crop in nations in the Middle East and North Africa. The date palm was thought to have outstanding traditional medicinal value because it was abundant in phytochemicals with diverse chemical structures. The date palm's ability to withstand harsh environments could be partly attributed to a class of proteins known as lectins, which are carbohydrate-binding proteins that can bind sugar moieties reversibly and without changing their chemical structures. After scanning the genome *of P. dactylifera* (GCF 009389715.1), this *in silico* study discovered 196 possible lectin homologs from 11 different families, some specific to plants. At the same time, others could also be found in other kingdoms of life. Their domain architectures and functional amino acid residues were investigated, and they yielded a 40% true-lectin with known conserved carbohydrate-binding residues. Further, their probable subcellular localization, physiochemical and phylogenetic analyses were also performed. Scanning all putative lectin homologs against the anticancer peptide (ACP) dataset found in the AntiCP2.0 webpage identified 26 genes with protein kinase receptors (Lec-KRs) belonging to 5 lectin families, which are reported to have at least one ACP motif. Our study offers the first account of *Phoenix*-lectins and their organization that can be used for further structural and functional analysis and investigating their potential as anticancer proteins.

1. Introduction

Phoenix dactylifera is the date or date palm plant; it belongs to the Arecaceae family, which comprises about 200 genera and over 2,500 species (Al-Alawi et al., 2017). The tree is one of the ancient and essential staple crops native to North Africa and the Middle East, where it is religiously emblematic and has been cited several times in the Holy Quran (Lemlem et al., 2018). The plant fruit is rich in carbohydrates, antioxidants, and potassium but low in lipids and protein (Assirey 2015, Al-Shwyeh 2019). Different parts of the Palm tree are commonly used in folk and traditional medicine and offer various medicinal values such as anti-inflammatory, anti-oxidant, anti-tumor, antimicrobial, antiulcer, etc. (Gangwar et al., 2014, Rahmani et al., 2014, Lemlem et al., 2018). Such medicinal properties could be attributed to several metabolites, including carbohydrate-binding proteins; lectins, which are studied extensively for their biotechnological and medical applications (Konozy and Osman 2022, Konozy et al., 2022, Konozy et al., 2022).

Lectins are diverse group of proteins with heterogenous families endowed with carbohydrate recognition capacities; they have been isolated, purified, and studied to structural minutes from different parts of the plants and other organisms (Mishra et al., 2019). These proteins have diverse levels of specificity toward various

* Corresponding author at: Biotechnology Department, Africa City of Technology (ACT), Blue Nile Bridge Ave., 4A/2 Alamlak, Bahri Central, Khartoum 11111, Sudan.
E-mail addresses: makarim84@act.gov.sd (M.E.M. Osman), rsalah@act.gov.sd (R.S.H. Osman), saraezo@act.gov.sd (S.A.A Elmubarak), amina.dirar@ncr.gov.sd (A.I. Dirar), ehkonozy@act.gov.sd (E.H.E. Konozy).
[1] These authors contributed equally to this work.
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.sjbs.2023.103676
1319-562X/

kinds of carbohydrates, which allows for a general classification into five major classes based on specificity i.e., galactose/*N*-acetylgalactosamine, mannose, fucose, *N*-acetylglucosamine, and *N*-acetylneuraminic acid (Sharon and Lis 2007). Variation in primary structures over time is trivial. Hence, from a structural point of view, the variation of lectin folds is an interesting phenomenon. Plant lectins have been divided into seven categories based on their evolution and structural characteristics: the amaranth family, chitin-binding lectin, Cucurbitaceaee phloem lectin, and jacalin-related lectin, legume lectin, monocot-mannose-binding lectin, and type-2 ribosome inactivating lectin. Depending on the number of carbohydrate-binding sites, plant lectins have been also classified into merolectins, hololectins, chimerolectins, and superlectins (Van Damme 2022). On the other hand, based on their sugar-binding specificity, cellular localization, and carbohydrate-binding domains, animal lectins have been, as well, classified into several groups such as C-type lectins, I-type lectins, and S-type lectins (galectins), P-type lectins, pentraxins, etc. (Matejuk and Duś 1998, Peumans et al., 1998). Of the main perquisites of lectin is the existence of at least one glycan-binding domain (CBD); according to the structure and folding pattern of CBD in plants, 12 groups could be identified: *Agaricus bisporus* Imbach agglutinin family (ABA), the amaranthine family, the *Galanthus nivalis* L. agglutinin (GNA) family, class V chitinases (CRA), the cyanovirin family, the *Euonymus europaeus* L. lectin (EUL) family, the Hevein family, the jacalin-related lectin (JRL) family, the Legume lectin family, the Lysin Motif (LysM) domain family, the *Nicotiana tabacum* L. agglutinin (Nictaba) family, and the Ricin-B lectin family (Van Damme 2014) (Table 1).

Plant lectins can be targeted extracellularly, to the cell wall or the cell/organ membrane, deposited in the vacuole, or found free in the cytoplasm or nucleus. Plant lectins can also vary in how they are localized within the cell. Identifying their biological roles and pathways is essential for ensuring that plants respond optimally to diverse environmental challenges and stimuli as well as their involvement in plant growth and development (Van Damme et al., 2004). For instance, lectin receptor-like kinases (Lec-RLK) are membrane-bound sensory proteins, they detect external environmental changes then induces cascade of downstream signals which triggers the expression of other stress-related genes and transcriptional factors. Contrarily, the ER-associated chaperons F-box-Nictaba and Calnexin (CNX)/Calreticulin (CRT) are involved in the processes of glycoproteins turnover and degradation. When a plant is stressed, a nucleoplasmic Nictaba modifies chromatin by attaching to the histone proteins' *O*-NAcGlc, which causes the production of defense-related genes (for more information, see (Osman et al., 2023)).

Although several lectins have been purified and characterized from the Arecaceae family, and the genome of *Phoenix dactylifera* was made available, the lectins from the date palm plants remain uncharacterized. Therefore, in this *in silico* investigation, we have retrieved all protein-containing lectin domains from the NCBI database and characterized them with respect to their carbohydrate-binding motifs, domain architectures, subcellular localization, physiochemical properties, and evolutionary relationships.

2. Methodology

2.1. Screening for lectin homologs from Phoenix dactylifera genome

The date palm (Taxon ID: 42345) *Phoenix dactylifera* genome assembly (scaffold level) (GCF_009389715.1) (Gross-Balthazard et al., 2019) deposed in the National Center for Biotechnology Information NCBI (https://www.ncbi.nlm.nih.gov/) was screened for the presence of lectin homologs from different families both plant-specific and universal families. A sequence model from each family was aligned against the *P. dactylifera* genome using the NCBI-BLASTp tool (Table 1). This was followed by a second BLAST using the retrieved sequence with the highest identity. All sequences were then collected and tested for the presence of at least one lectin domain using the InterProScan 5.0 online server (https://www.ebi.ac.uk/interpro/) (Jones et al., 2014). The Pfam ID for each family was used to identify potential lectin-like domains, and the start and end points were collected. The difference in lectin domain architecture in each family was illustrated using visualization of protein domain structures (DOG 2.0) (Ren et al., 2009).

2.2. Identification of true lectins

The analysis of conserved amino acids responsible for carbohydrate binding was performed using trimmed sequences corresponding to lectin-like domains. Sequences with multiple lectin domains were trimmed, and each domain was used as a separate entry. Multiple sequence alignments using the Clustal W tool built in BioEdit 7.2 (Hall 1999) were done, and each sequence's sugar-binding amino acid residues were identified.

2.3. Phylogenetics analysis

The phylogenetic trees were constructed using aligned trimmed lectin-like domain sequences. The maximum likelihood method and the JTT matrix-based substitution model which rely on large scale of aligned protein sequences to measure the amino acids frequencies and the substitution rate were selected (rate of heterogenicity: Gamma, and #rate category 5) in IQ-TREE web server (https://iqtree.cibiv.univie.ac.at/) to analyze the evolutionary relationship (Trifinopoulos et al., 2016). The final bootstrap for consensus trees was inferred from 1000 replicates. The tree topologies were illustrated by iTOL v6 (https://itol.embl.de/) (Letunic and Bork 2021).

2.4. Characterization of lectin homolog sequences

Obtained lectin homologs were checked for the presence of signal peptide and transmembrane domain sequences using the SignalP 5.0 server (Almagro Armenteros et al., 2019) (https://services.healthtech.dtu.dk/service.php?SignalP-5.0) and TMHMM 2.0 webtool (Krogh et al., 2001) (https://services.healthtech.dtu.dk/service.php?TMHMM-2.0), respectively. Furthermore, each sequence was analyzed for the presence of potential glycosylation sites for *N*- glycan using NetNGlyc 1.0 server (Gupta and Brunak 2001) (https://services.healthtech.dtu.dk/service.php?NetNGlyc-1.0) and *O*-glycan using NetOGlyc 4.0 server (https://services.healthtech.dtu.dk/service.php?NetOGlyc-4.0) (Steentoft et al., 2013).

2.5. Physiochemical characterization of primary and secondary lectin structures

The molecular weight, isoelectric points (*pI*), and the total amino acid composition for each lectin homologs sequence were predicted using the Expasy ProtParam webtool (Gasteiger et al., 2005) (https://web.expasy.org/protparam/). The percentage of the alpha chains and beta sheets were determined by the PHD secondary structure prediction method tool (Rost 1996) (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_phd.html).

Table 1
Plant lectin families and sequence models ID used for screening against *Phoenix dactylifera* genome:

Lectin family	Model species	ID model seq.	Database	Pfam ID	Pfam domain	Biological distribution
ABA	*Agaricus bisporus*	Q00022.3	UniProtKB	PF07367	FB_lectin	Plant-specific
Amaranthin	*Amaranthus caudatus*	AAL05954.1	Genbank	PF07468	Agglutinin	Plant-specific
CRA	*Robinia pseudoacacia*	ABL98074.1	Genbank	PF00704	GH-18	Plants, animals
Cyanovirin	*Nostoc ellipsosporum*	P81180.2	UniProtKB	PF08881	CVNH	Plant-specific
EUL	*Euonymus europaeus*	ABW73993.1	Genbank	–	–	Plant-specific
GNA	*Galanthus nivalis*	P30617	UniProtKB	PF01453	Bulb lectin	Prokaryotes, fungi, plants, animals
Hevein	*Hevea brasiliensis*	ABW34946.1	Genbank	PF00187	Chitin Bind-1	Plant-specific
JRL	*Artocarpus integer*	AAA32680.1	Genbank	PF01419	Jacalin	Plant-specific
Legume	*Glycine* max	P05046	UniProtKB	PF00139	Leg_B	Plants, animals
LysM	*Brassica juncea*	BAN83772.1	Genbank	PF01476	LysM	Plant-specific
Nictaba	*Nicotiana tabacum*	AAK84134.1	Genbank	PF14299	PP2	Plant-specific
Ricin-B	*Ricinus communis*	PDB:2AA1_B	Genbank	PF00652	Ricin_B	Prokaryotes, plants, animals
Malectin	*Xenopus laevis*	NP_001085212.1	Genbank	PF11721	Malectin	Prokaryotes, plants, animals
Calnexin	*Arabidopsis thaliana*	P29402	Genbank	PF00262	Calreticulin	Yeast/fungi, plants, animals
M−type	*Homo sapiens*	Q9UKM7.2	UniProtKB	PF01532	EDEMs	Yeast/fungi, plants, animals

2.6. Subcellular localization and targeting of lectin homologs

The potential subcellular compartment localization was identified for each sequence using web server CELLO2GO web server for protein subcellular localization prediction with functional gene ontology annotation (https://cello.life.nctu.edu.tw/cello2go/) (Yu et al., 2014). This was followed by the prediction of signal classes (signal peptide (SP), transmembrane (TM), intracellular (IC), and unconventional secretion (USP)) using the OutCyte 1.0 server (https://www.outcyte.com/) (Zhao et al., 2019). Lectins hydropathic characteristics were plotted using Kyte-Doolittle scale (Kyte and Doolittle 1982) available at (https://www.novoprolabs.com/tools/protein-hydropathy).

2.7. Anticancer motifs in lectin homologs genes

To identify the anti-cancerous properties of *P. dactylifera,* each putative lectin protein sequence was screened for anticancer motifs against ACP/AMP primary dataset using Motif Scan Page AntiCP2.0 (Agrawal et al., 2021) (https://webs.iiitd.edu.in/raghava/anticp2/motif_scan.php).

3. Results

3.1. Identification of putative lectin gene homologs from Phoenix dactylifera genome

Phoenix dactylifera (Arecaceae) is a monocot plant with 36 pairs of chromosomes (n = 18) (Shabana et al., 2010); its genomic scaffold assembly was used to screen for the presence of lectin gene homologs. Fifteen lectin families were used as search models; 8 of these families are found only on plants, while the seven other families exist in plants and other kingdoms. (Table 1). A total of 196 genes were retrieved from 11 tested families out of 15; ABA, Amaranthin, Cyanovirin, and Ricin-B were not found. Legume, Nictaba, and JRL lectin homologs comprised the most significant number of lectins, with a total of 133 (57%) genes. At the same time, the *Phoenix* EUL family contained only a single putative gene (Table 2). The high abundance of legume genes is consistent with other plants, such as *Glycine* max, Fabaceae (94(26%)) (Van Holle and Van Damme 2015), *Sorghum bicolor,* Poaceae (55 (45%)) (Osman et al., 2022), and *Arabidopsis thaliana*, Brassicaceae (54 (25%)) (Eggermont et al., 2017).

3.2. Identification of true lectins

The similarities in the lectin domain between members of the same family do not guarantee carbohydrate-binding ability. Amino acid residues responsible for binding sugar moieties should be conserved, and the lectin domain should only bind to them reversibly and without altering their chemical structures (Peumans and Van Damme 1995). Table 2 indicates the prevalence of true-lectins (40%) and lectin-likes (60%), identified upon investigating and predicting the presence of sugar-binding residues. The tree topology for the lectin families with true and lectin-like members that shows their domain evolutionary relationship is available in Fig. 1.

Proteins with Hevein domains are ubiquitous in plants (Damme et al., 1998). Such proteins are considered defense proteins against fungi due to their ability to mediate the *N*-acetyl-D-glucosamine or chitin molecules expressed on their surface coat. The chitin-binding domain from Hevein can be classified as C6, C8, and C10 based on the number of conserved Cystine residues. *Phoenix*-Hevein-like domains are classified as C8. However, three genes lack the N-terminal Cystine residue. Binding the Chitin molecules requires three conserved aromatic amino acids, Trp21, Trp23, and Tyr30, as well as a Serine residue Ser19 (Asensio et al., 2000). While both residues corresponding to Ser19 and Tyr30 were highly conserved in the *Phoenix*-Hevein-like domain, Both Tryptophan residues were substituted by Tyr residues except for one sequence (XP_008801201.1) which retains Typ21 and Tyr23. Though Trp21 and Trp23 are commonly found in most plant Hevein lectins, the substitutes Tyr21 and Tyr23 were reported in the binding site of the wheat germ agglutinin (WAG1) and contributed to the stabilization of the Neuraminayl-N acetyl galactosamine (NeuNAc-Gal) through non-polar interaction with the aromatic side chain of the Tyr23 (Wright 1990). Only 3 sequences of *Phoenix*-Hevein domains satisfy the presence of all the structural bases for sugar binding as well as the 8 cystine residues (Additional file S1: Figure S1).

EUL is a large heterogenous family, with the EUL-S3 domain to be found in lower plants, monocots, and dicots (Fouquaert et al., 2009). EUL has a promiscuous sugar specificity (Fouquaert and Van Damme 2012). Nonetheless, this was attributed to the involvement of a single binding site, mediated by conserved Asp130, Trp143, Asp145, Tyr147, Asn148, and Gln149 reported in the model protein *Euonymus europaeus* (EEA) (Agostino et al., 2015). *Phoenix*-EUL protein precursor is built from three tandemly arrayed EUL-S3 domains, but none of the three domains have all six of the required-binding residues. Instead, Asp145 and Tyr147 were replaced by positive amino acids Lys145 and Lys147 like the *A. thaliana* (ArathEULS3) and *Physcomitrella patens* (PhypaEULS3) (Fouquaert and Van Damme 2012). Such changes add

Table 2
Identified lectin families from *Phoenix dactylifera*:

Lectin families	# Predicted genes			Predicted signal classes†	M.Wt.	Subcellular localization
	Total # (%)	Lectin-like	True lectins		KDa	
CRA	10 (5.1)	6	4	SP(8), USP(2)	33 –51	Extracellular, Cell membrane, Chloroplast, Vacuoles, Lysosome
EUL	1 (0.5)	0	1	IC(1)	94	Nuclear, Extracellular, Cytoplasm, Cell membrane, Mitochondria, Peroxisome
GNA	14 (7.1)	7	7	SP(11), USP(3)	14 – 92	Extracellular, Cell membrane, Lysosome, Mitochondria, Nucleus, Chloroplast
Hevein	6 (3.1)	3	3	SP(6)	29 – 34	Extracellular, Nucleus, Cytoplasm, Lysosome, Vacuoles
JRL	21 (10.7)	14	6	SP(10), USP(11)	12 – 66	Cell membrane, Extracellular, Cytoplasm, Chloroplast, Mitochondria, Nucleus
Legume	65 (33.2)	49	16	SP(58), TM(4), IC(3)	52 – 86	Cell membrane, Chloroplast, Cytoplasm, Extracellular, Vacuoles, Lysosome
LysM	7 (3.6)	6	1	SP(7)	66 – 72	Cell membrane, Chloroplast, Extracellular, Cytoplasm, Mitochondria
Nictaba	27 (13.7)	19	10	USP(17), TM(1), IC(9)	17 – 39	Cell membrane, Cytoplasm, Nucleus, Extracellular, Mitochondria
Malectin	18 (9.2)	12	6	SP(11), TM(2), IC(5)	75 – 130	Cell membrane, Cytoplasm, Nucleus, Extracellular
Calnexin	9 (4.6)	3	6	SP(1), USP(5), IC(3)	38 – 60	Cytoplasm, ER, Nucleus, Mitochondria
M−type	17 (8.7)	0	17	USP(16), TM(1),	11 – 60	Cytoplasm, Cell membrane, Lysosome, Mitochondria, Extracellular
Total	**196**	**118**	**78**			
%	**100**	**60.2**	**39.8**			

†Predicted signal classes: SP: signal peptide, TM: transmembrane, IC: intracellular, and USP: unconventional peptide secretion.

to the promiscuity of the protein's specificity (Additional file S1: Figure S2).

Glycosyl hydrolase 18 (GH-18) is a widely expressed domain in prokaryotes and eukaryotes (i.e., fungi, plants, arthropods, and nematodes) responsible for remodeling chitin (Grover 2012). The class V chitinase-related agglutinin (CRA) is closely related to the GH-18 superfamily; however, the agglutinin is devoid of the enzymatic activity of the GH-18 (Van Damme et al., 2007). To annotate *Phoenix*-CRA homologs as true-lectins, they should have changes in the highly conserved catalytic motif of their GH-18 domain (DxDxE). Four homologs have alternation in the second Asp and/or third Glu residues, while the first Asp remains highly conserved. Phylogenetic analysis indicated that these true lectin orthologs are closely related and clustered in two clades (Fig. 1B; Additional file S1: Figure S3).

The universal LysM domain binds *N*-acetyl-D-glucosamine and Chitin-containing structures. Subtle differences between eukaryotic and prokaryotic exist in the domain's functional residues, including the absence of stabilizing Cystine residues from the prokaryotic LysM domain and plant LysM PF01476 domain (Ohnuma et al., 2008). Structural and functional analysis of the green algae *Volvox carteri* revealed that titration of $(GlcNAc)_n$ into the protein solution resulted in a gradual shift in resonance for the following residues Gly92, Asp93, Thr94, Phe95, Ile98, Leu121, and Gln122 (Kitaoku et al., 2017). Searching for these residues in LysM peptides from *A. thaliana* showed that only Asp93, Thr94, Ile98, and Ala99 were conserved (Eggermont et al., 2017). Such residues were also conserved in a single LysM domain from *P. dactylifera*. Due to the lack of conclusive information, empirical studies in plant LysM lectin homologs should be performed for definitive results and conformation of the carbohydrate-binding amino acids (Additional file S1: Figure S4).

Jacalin-related lectins are classified into mannose-JRL and galactose-JRL based on their specificity. However, the variation in sugar specificity between the groups does not occur due to the differences in the sugar-binding residues but rather due to co– or posttranslational proteolytic modifications that cleave the protomer resulting in the formation of an extra loop that extends the binding site allowing for galactose accommodation (Bourne et al., 2002). Highly conserved N-terminal Gly and C-terminal Asp, as well as the semi-conserved Asn96, Tyr141, and Tyr142, are reported in man-JRL from *Calystegia sepium* (Jeyaprakash et al., 2005). In contrast, the N-terminal Gly28 and the C-terminal Gly149 and Asp153 were reported to bind mannose at the binding loop located at the top of the β-prism fold in *Pteria Penguin*'s JRL (PPL3 and PPL4) (Ogawa et al., 2019). Searching for these residues in *Phoenix*-JRLs revealed that the N-terminal Gly, C-terminal Gly, Asp, and the second Tyr/Trp were present. Sequences (7) with such a combination were cautiously considered lectin and needed to be confirmed experimentally. (Additional file S1: Figure S5).

The ubiquitous plant Nictaba and its closely related Cucurbitaceae Phloem Lectins PP2 are GlcNAc-specific lectins (Delporte et al., 2015). The binding is facilitated by four amino acid residues, Trp15, Trp22, Glu138, and Glu145 (Schouppe et al., 2010). 53% of the identified Nictaba homologs were true lectins retaining the four critical residues (Additional file S1: Figure S6).

The mannose-Bulb lectin superfamily can be found in all living kingdoms. The binding of mannose-related structures is achieved by the consensus sequence motif QxDxNxVxY. *Galanthus nivalis* GNA lectin domain contains three motifs array (Ramachandraiah and Chandra 2000). Investigating *Phoenix*-GNA domains resulted in identifying seven genes with domains containing at least a single consensus motif near the N-terminal. In all sequences, the second motif is highly variable compared to the third motif which has 4 out of 5 conserved residues that vary between sequences (Additional file S1: Figure S7).

Legume lectins have different specificities toward carbohydrate moieties, which include mannose/glucose, *N*-acetylglucosamine, galactose/*N*-acetylgalactosamine, and complex glycans ((Ramos et al., 2000). This is endowed with the variation in the fourth carbohydrate-binding residue located at the D-loop of the protein monomer. However, the EcorL from *Erythrina coralderone*'s binding residues, i.e., Ala88, Asp89 (A-loop), Gly107 (B-loop), and Asn133 (C-loop) (Loris et al., 1998, Osman and Konozy 2017), are highly conserved and found in all plant legume lectin homologs including 33% of *P. dactylifera* legume lectins which are all clustered in a single clade (Fig. 1D; Additional file S1: Figure S8).

Calnexin/calreticulin mediates the sugar-binding by six key residues Tyr165, Lys167, Tyr186, Met189, Asp/Glu217, and Asp/Glu426 (Leach and Williams 2004). Only *Phoenix*-Calnexin/calreticulin peptides constructed by two domains are considered true lectins. Where only the first domain contains the complete

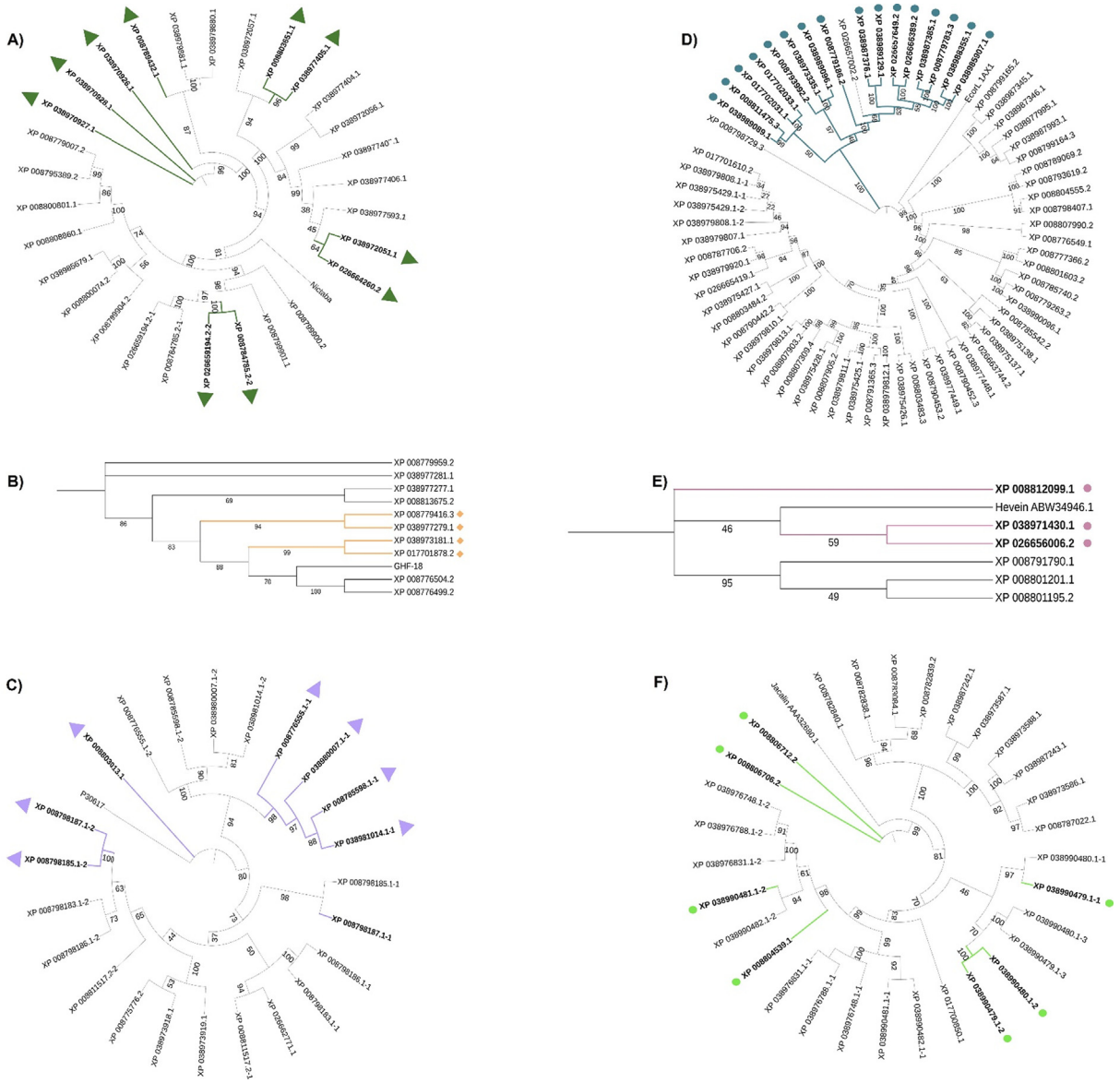


Fig. 1. Maximum likelihood analysis of lectin homologs from *Phoenix dactylifera*. Bootstraps were inferred from 1000 replicates. A) Nictaba family, the dark green triangle symbols represent true lectins; the Nictaba sequence from *Nicotiana tabacum* Nictaba lectin (AAK84134.1) was used as a reference ortholog. B) CRA family, the orange diamond symbols represent true lectins; the human GHF18 chitinase (Q9BZP6.1) was used as a reference ortholog. C) GNA family, the purple triangle symbols represent true lectins; the GNA from *Galanthus nivalis* Bulb lectin (P30617) was used as a reference ortholog. D) Legume family, cyan-green circle symbols represent true lectins; the. *Erythrina corallodendron* EcorL (1AX1) was used as a reference ortholog. E) Hevein family, the pink circle symbols represent true lectins; the *Hevea brasiliensis* Hevein lectin (ABW34946.1) was used as a reference ortholog. F) JRL family, lime green circle symbols represent true lectin; the *Artocarpus integer* Jacalin lectin (AAA32680.1) was used as a reference ortholog.

set of reported amino acid residues. These domains are all branched from the same clade (Fig. 2).

M−type is closely related to alpha-mannosidase. However, it lacks the catalytic enzymatic activity due to the alternation in critical amino acid residues responsible for the mannosidase activity. These residues are Glu123, Arg187, Cys253, and Cys280, plus two Cystine residues at 24 and 366 that stabilize the enzyme by forming disulfide bridges (Hosokawa et al., 2001). Alignment of *Phoenix*-M−type homologs showed that the entire set is devoid of alpha-mannosidase critical residues, and the Cystine resides (Additional file S1: Figure S9).

Of the twelve Malectin putative genes identified in this study, six are true lectin sequences with the reported sugar-binding significant residues Tyr67, Leu89, Tyr116, Phe117, and Gly186 (Schallus et al., 2008) (Fig. 3).

3.3. General overview

All *P. dactylifera* putative genes were characterized by their general assembly, localization, and physiochemical properties. The domain architectures are generally very simple compared to other plants like *G.* max and *A. thaliana* (Van Holle and Van Damme 2015, Eggermont et al., 2017). Members of four families are strictly chimeric peptides, with either a protein kinase domain (PF00069 and PF07714), as in the Legume, LysM, and Malectins, or a glycosyl hydrolase-19 (GH-19) domain (PF00182), as in the Hevein-like lectins, joined to the lectin domain by the C-terminal. Other chimeric domains such as F-box (PF00646) found in Nictaba, and S-locus glycoprotein (PF00956) and PAN2 (PF08278) were reported in GNA three homologs. Merolectins (a single domain lectin) and hololectins with two or three tandemly arrayed lectin domains

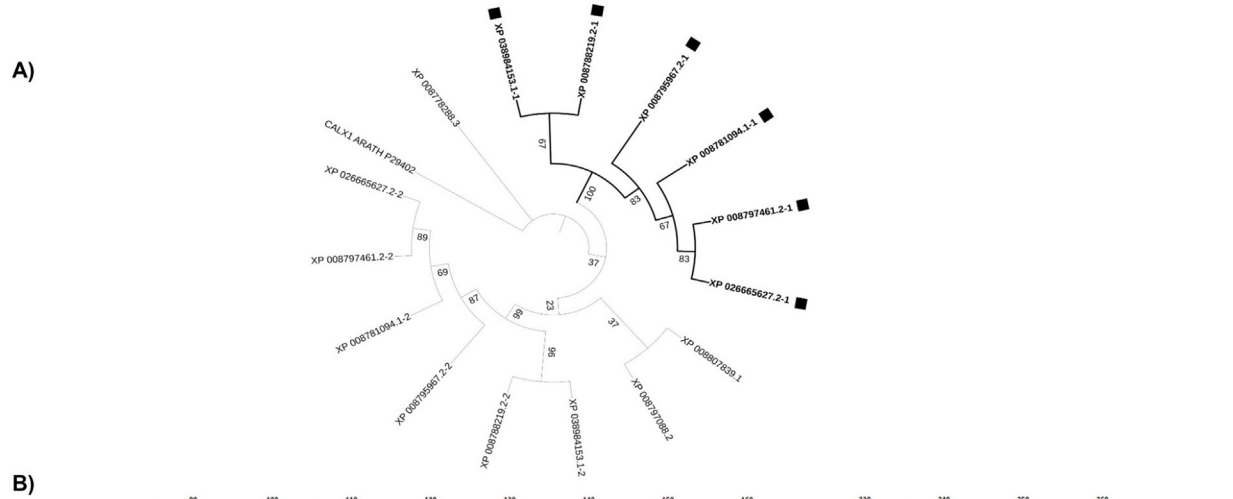


Fig. 2. Identification of true-calnexin/calreticulin lectins from *Phoenix dactylifera*. A) The maximum likelihood phylogenetic tree of calnexin/calreticulin. The black square blocks represent the clade of true lectins. B) Partial alignment of calnexin/calreticulin homologs. The different color one-letter codes are used to represent amino acids. The red-shaded boxes represent the conserved critical amino acids required for carbohydrate binding. The black dot represents lectins that retain all conserved amino acids needed for the binding. Lectins ID and accession numbers are shown on the left, and the number of domains that resides in the same protein is indicated after the dash. CA1X1_ARATH from *Arabidopsis thaliana* (P29402) was used as a reference for the alignment and an ortholog in the phylogenetic analysis.

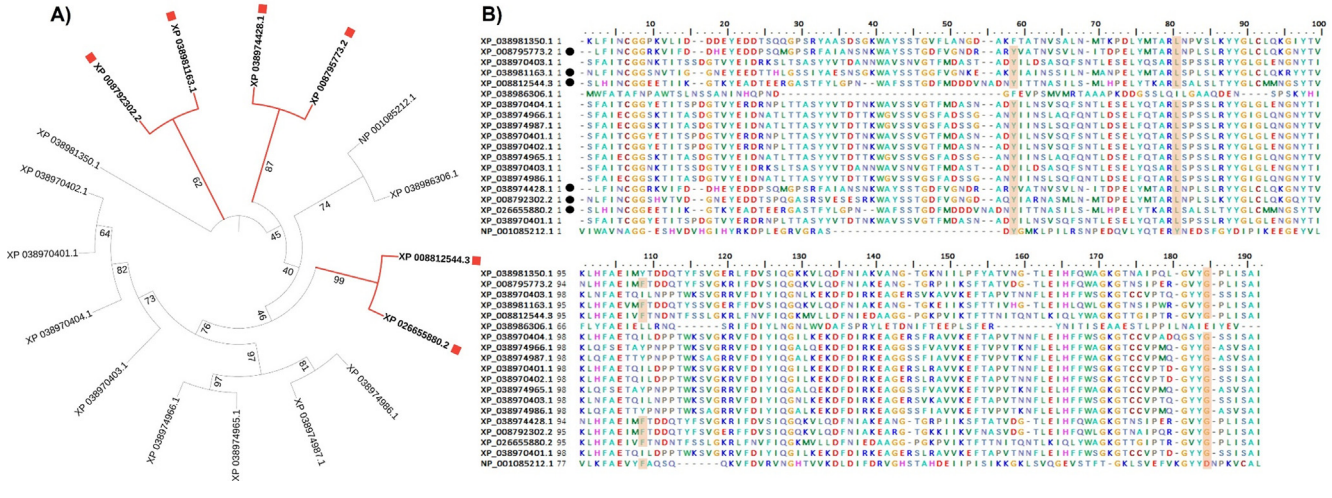


Fig. 3. Identification of true-Malectins from *Phoenix dactylifera*. A) The maximum likelihood phylogenetic tree of Malectins. The red square blocks represent the clade of true-lectins. B) Partial alignment of Malectin homologs. The different color one-letter codes are used to represent amino acids. The red-shaded boxes represent the conserved critical amino acids required for carbohydrate binding. The black dot represents lectins that retain all conserved amino acids needed for the binding. Lectins ID and accession numbers are shown on the left, and the number of domains that resides in the same protein is indicated after the dash. *Xenopus laevis* (NP_00085212.1) was used as a reference model for the alignment and an ortholog in the phylogenetic analysis.

comprised around 44% of the total identified putative genes (Fig. 4). EUL, JRL, and Nictaba homologs are inducible lectins that respond to environmental triggers; they lack signal peptides and transmembrane domains, and they are hydrophilic proteins. Various members of these families are predicted to be glycosylated. JRL homolog's secondary structure is mainly constructed from β-sheets, and random coils and they are rich in glycine > 10% and serine > 9% residues, unlike Nictaba, where the α-helix and β-sheets equally cover around 40% of the peptides' structure and leucine residues are more predominant compared to other residues. Chitinase-related families CRA (33 – 51 kDa) and Hevein (28 – 33 kDa) putative gene sequences are proceeded by N-terminal signal peptide; this is true for all sequences except two CRA true-lectins. Hevein-like homologs were predicted to be secreted extracellularly, while CRAs are mainly targeted to the cell membrane and the chloroplast. M−type (10 – 60 kDa) and Calnexin/calretic-

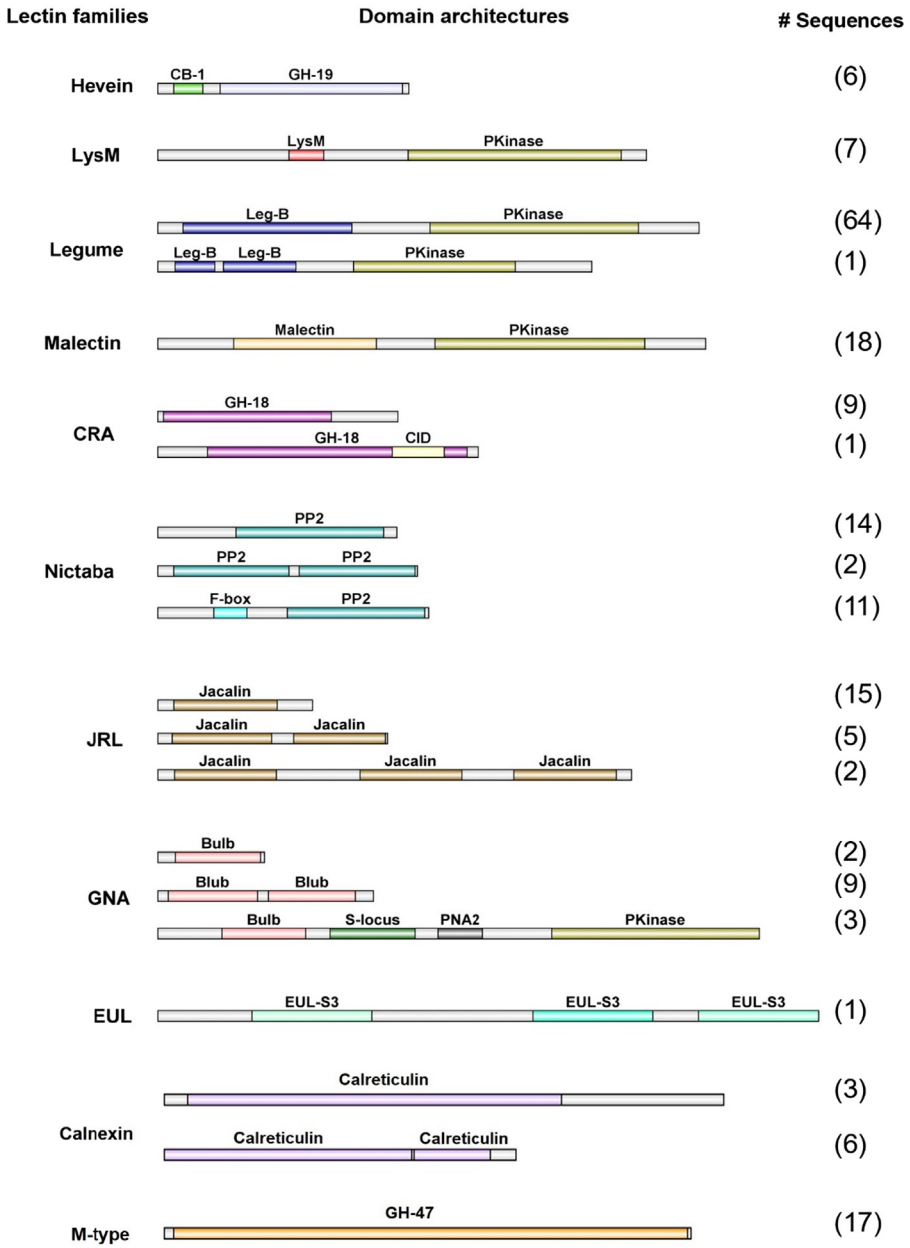


Fig. 4. Lectin families and true-lectin homologs representation and prevalence in *Phoenix dactylifera* genome. LysM: Lysin motif, GH-47: glycosyl hydrolase-47, GH–19: glycosyl hydrolase–19 (PF00182), CB–1: carbohydrate binding domain–1, PP2: Phloem protein 2, F–box (PF00646), PKinase: Protein kinase (PF00069, PF07714), S-locus glycoprotein (PF00956), and PAN2 (PF08278).

ulin (37 – 60 kDa) are devoid of the signal peptide, and the plasma membrane, cytoplasm, and endoplasmic reticulum (ER) are the most probable subcellular locations these peptides are likely to be targeted to. Lectin-receptor Kinases (Lec-RKs) are the most predominant lectin assemblies and are characteristically rich in Leucine and Serine amino acid residues (∼9 – 13%). Such lectins are mainly targeted extracellularly or to the cell membrane. (Fig. 5 and Additional files S2 and S3). According to the hydropathic analysis of lectin sequences, EUL, JRL, Nictaba, and M−type members have average negative hydrophilic index readings along the length of the sequences with many hydrophobic regions and energy peaks that are insufficient to prevent the proteins from crossing the membrane bilayer. Legume, LysM, GNA, and members of the calnexin/calreticulin family, on the other hand, have a positive hydrophobic peak with a hydropathic index (> +1.8), indicating that these lectins are transmembrane proteins. With sufficient energy, CRA have an average positive hydrophobic reading across the whole sequence, which enables the protein to traverse the lipid bilayer (Fig. 6).

3.4. ACP/AMP properties of Phoenix-lectin homologs

Plant lectins play defense and protective functions against various plant illnesses and infectious diseases. Such property made this class of proteins a target for extensive medical and biotechnological studies to find plant-based alternative replacements to many pharmaceutical drugs. We were motivated to search for molecular bases that can contribute to some of these lectins' anti-cancer (ACP)/anti-microbial (AMP) properties. Characteristically, residues like A, F, K, L, and W have a higher composition in ACPs, with A, F, and K preferably positioned N-terminally. In contrast, W and L are placed at the C-terminal. Motifs like LAKL, FAKL,

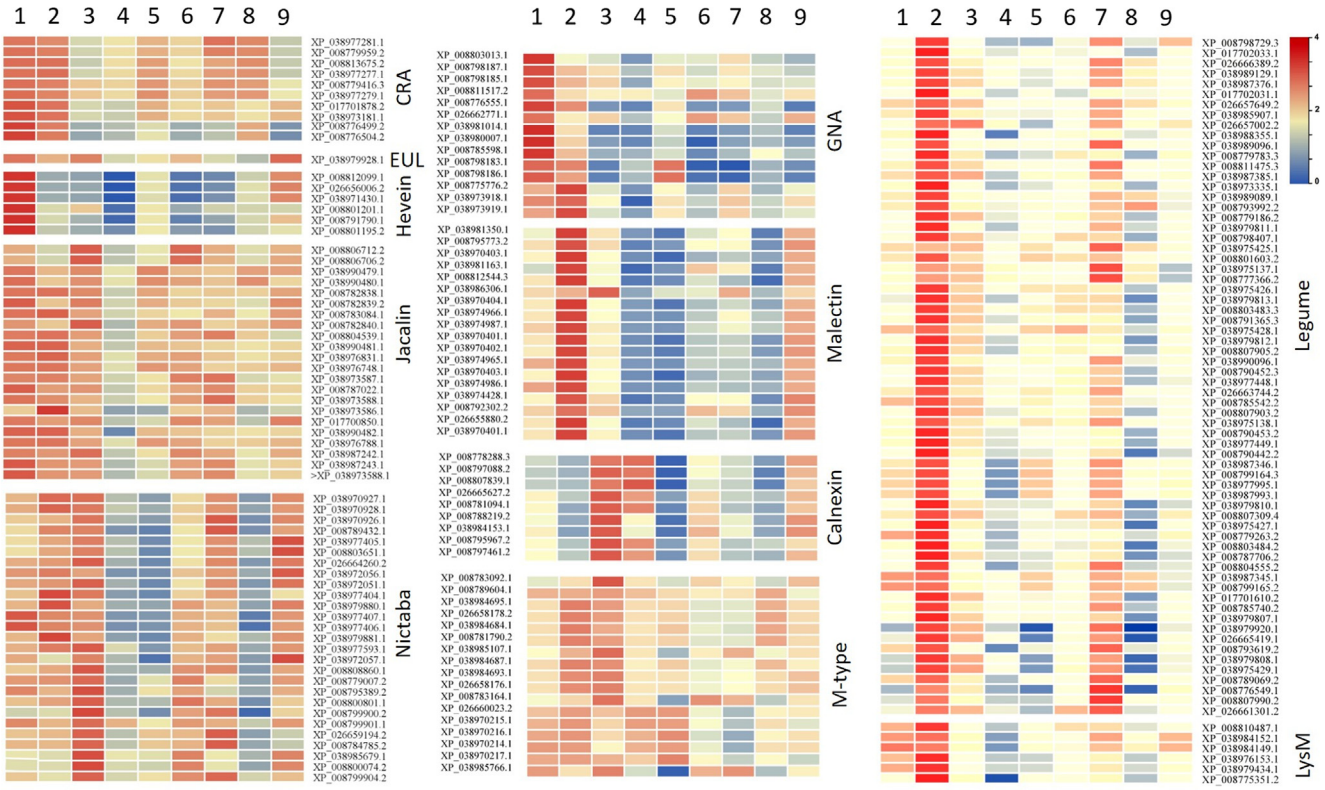


Fig. 5. Heatmap of *Phoenix dactylifera* lectin homologs from different families. The numbers above the heatmap blocks represent the various locations. 1: Extracellular, 2: Plasma membrane, 3: Cytoplasm, 4: ER, 5: Lysosome, 6: Mitochondria, 7: Chloroplast, 8: Vacuoles, and 9: Nucleus.

KLAK, and LAKLA were reported (Agrawal et al., 2021). Several lectin homolog members belonging to the LysM, Legume, Malectin, GNA, and M−type were reported to have these features. LAKL motif was present in almost all of these proteins. Except for the M−type sequence, where the ACP motif is located inside the lectin domain (GH-47), all other sequences are considered Lec-RKs, and the ACP motif is located inside the protein Kinase domain. (Table 3, Additional file S3).

4. Discussion

The economic and agriculturally important date palm tree *Phoenix dactylifera* genome was expanded by the events of genome-wide and a massive segmental duplication; this provided the date palms with the opportunity to flourish in North Africa and the Middle East, where they become equipped to withstand heat, salinity, and drought as well as biotic stresses (Al-Mssallem et al., 2013). Such events allow for the fixation of beneficial mutations that either maintains the original function or permit the acquisition of new functions by sub-functionalization or neofunctionalization (Álvarez-Lugo and Becerra 2021). Although the *P. dactylifera* genome retains a relatively high number of lectin gene homologs (196 genes), only 40% of these genes contain true lectin domains capable of actual carbohydrate binding based on alignments with previously reported members of each lectin family. Evidently, the high number of genes occurs due to previous duplication events. These processes might also generate pseudogenes with similar architectural and structural annotations to true genes that could lose their presumed function due to disabling mutations (Panchy et al., 2016). Theoretically, though, chimeric lectins-like might still be expressed and function through the other fused domain, while the pseudogenes of mero- and hololectins architecture are rendered nonfunctional or have altered affinity or specificity to the specific carbohydrate due to such mutations. Further studies focusing on the other 60% of the *P. dactylifera* lectins-like structural bases for carbohydrate binding will provide more conclusive evidence on the effect of the changed amino acids binding residues.

The complexity degree of the domain rearrangement in *Phoenix*-lectins, as determined by *in silico* study, is lower than that seen in *G.* max (Van Holle and Van Damme 2015) and *A. thaliana* (Eggermont et al., 2017) lectins. Fusion, fission, and N- and C-terminal losses are the mechanism by which different domain arrangements might occur, with fusion being the most prevalent. These arrangements are governed by genetic means, including exon-shuffling, nonallelic homologous recombination, transposition, or nonhomologous end joining. Hence, allowing for an evolutionary jump and the creation of novel genetic phenotypes (Kersting et al., 2012). G-type-RK lectin (GNA) containing S-locus domain that controls the inability of flowering plants to achieve self-fertilization, and PAN (apple-like) domain that assist in protein–protein or protein-carbohydrate interaction, are believed to occur as a result of fusion event (Xing et al., 2013). The loss of N-terminal signal peptide is observed in several members of lectin families from *Phoenix* (true and lectin-like), i.e., CRA, Legume, GNA, Malectin, Calnexin/calreticulin, and M−type. This indicates a different role for these lectins based on their other possible subcellular localization they might end up targeted to.

Several factors dictate the broad functional spectrum of endogenous plant lectin. These aspects include 1) the presence of plant-specific lectin families versus other more common families across the kingdoms, as well as 2) the high number of lectins and lectin-likes, 3) the variation in domain rearrangements between members of the same family, 4) their potential subcellular localization, 5) the wide range of sugar specificity, and 6) the final 3D structures.

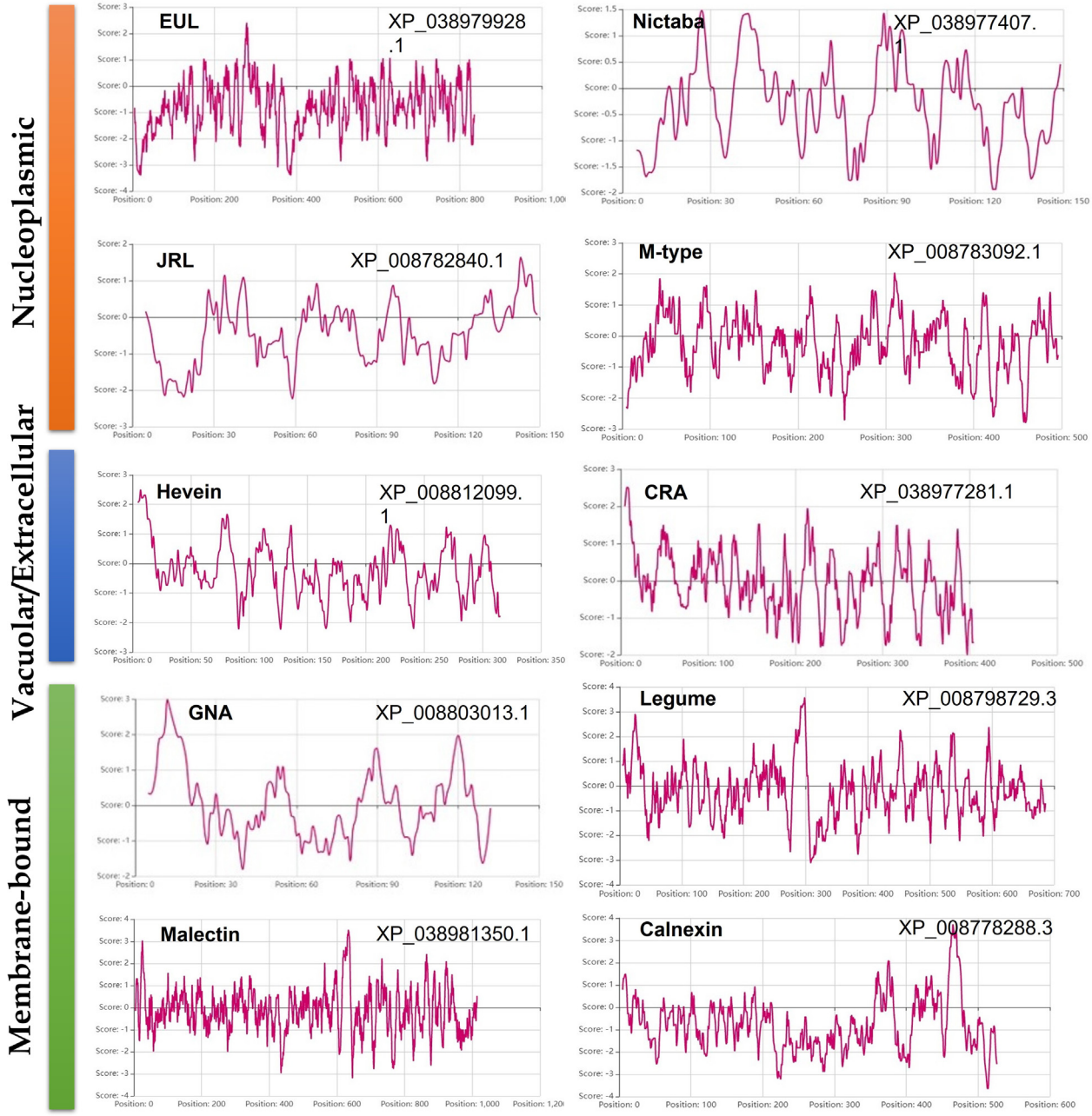
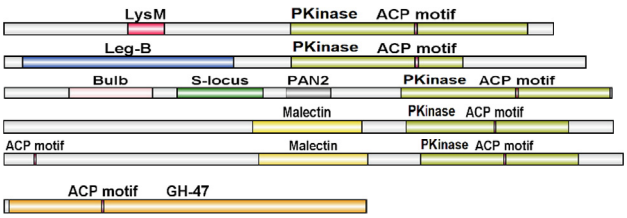


Fig. 6. Hydropathic plots for lectin models from each family. X-axis: the position of amino acids, Y-axis: the hydropathy index.

Table 3
Identification of Phoenix dactylifera ACP-motifs, their abundance in lectin families, and location within specific peptides:

Lectin family	# Seq	ACP motif	Motif location
LysM	1	LAKA (1)	
Legume	7	LAKL (4), KALK (2), KLLA (2)	
GNA	3	LAKL (3)	
Malectin	10	LAKL (8), ILLDK (1), LRPT (1)	
	4	LAKL- KALK (1), LAKL- LAKL (1), LYDD-LAKL (1), LTDK- LAKL (1)	
M−type	1	KLLA (1)	

Plants frequently encounter numerous pathogenic microbes that threaten their continued existence. As a result, they employ a variety of immunological defense mechanisms to evade microbial invaders and predators. Two interconnected mechanisms are

employed, the microbe-associated molecular pattern-triggered immunity (MAMPs) and the microbial effector-triggered immunity (RLK). Both mechanisms would trigger many signaling pathways that finally lead to the termination of the pathogen life cycle and may develop an immune memory to protect against similar future attacks (Jones and Dangl 2006, Liu et al., 2020). The MAMPs triggered immunity is associated with several kinase receptors that lead to defense gene activations, radical releases, and/or programmed cell death (Sussholz et al., 2020). These receptor kinases often possess a versatile number of extracellularly protruding leucine-rich repeats that assist in adapting to a wide range of interactions (Dufayard et al., 2017) or lectin domain involved in external sensation and subsequent internal triggering of downstream response (Sun et al., 2020). Malectin is a novel ER-anchored carbohydrate-binding protein that recognizes and binds Glc2-high-mannose *N*-glycan (Schallus et al., 2008); it is not found in plant cells; however, its analogues Malectin-like domains (MLDs) are component of many membrane-bound plant receptors and play a vital role in plant innate immunity. In *Arabidopsis*, Malectin could also be found in receptor-like serine/threonine-protein kinases like receptor kinases, and their roles in *A. thaliana* defense are well studied (Hok et al., 2011, Martín-Dacal et al., 2022). Lectin receptor kinases (Lec-RKs) constitute two domains, the N-terminal extracellular lectin domain, the C-terminal intracellular cytosolic Ser/Thr kinase domain, and in addition to transmembrane domains. Based on the lectin domain architecture, Lec-RKs have been grouped into types; G-, C-, L-, and LysM types (Petrova et al., 2021). Lec-RKs are found at different localities, such as vacuole, cytoplasm, nucleus, plasma membrane, and cell wall. Thus, their physiological significance may vary accordingly. For example, the lectins traced in vacuoles are assigned to play a defense role against herbivorous animals and are often abundantly accumulated in seeds and storage organs. However, in other tissues with no storage function, the lectin is usually found in lower concentrations (Damme et al., 1998). These lectins can also play different roles in sensing the environment and passing signals. *Medicago truncatula*, a legume plant, expresses a lectin-like kinase (LLK) during symbiosis. The lectin expression upsurges during nitrogen depletion and decreases upon the symbiotic bacterium infection (Babosha 2008). The inducible plant lectins, localized in the nucleus and/or cytoplasm, are expressed in response to the pathogenic attack and have been discussed in many articles (Vandenborre et al., 2011, Macedo et al., 2015). These lectins are also believed to regulate the plant responses toward abiotic and biotic stresses (Van Damme et al., 2004, Van Damme et al., 2004, Babosha 2008, Shumayla et al., 2016). Functionally, JRLs are believed to act against many disease stresses. Interestingly, JRL from *A. thaliana* interacts with β-glucosidase PYK10/BGLU23; both proteins are localized in the different cellular compartments, and they interact when a pathogen destroys the cellular structures; hence regulate the size of the β- glucosidase complex upon infection (Nagano et al., 2008). The nucleoplasmic JRLs can also respond to various other stress environmental stimuli. For instance, OsJAC1, a Jacalin lectin from *Oryza sativa,* is upregulated by ionizing radiation in a time and dose-dependent manner and subsequently modulates the DNA repair by activating the DNA damage checkpoints (Jung et al., 2019). None-chimeric ArathNictabas (AN3 cytoplasmic, AN4, and AN5 nuclear proteins) are synthazied in trace amounts during plant growth and development; however, upon infection with *Pseudomonas syringae,* the levels of the nuclear AN4 and AN5 increases significantly and facilitate higher tolerance against *P. syringae* in transgenic plants. AN4 interacts with two other defense enzymes, TGG1 and BGLU23 (Eggermont et al., 2018). F-box-Nictaba is considered functional homolog of the mammalian F-box proteins; in *A. thaliana,* F-box-Nictaba can interact with Skp1 to form the SCF complex, which is involved in ubiquitin tagging of proteins fated for degradation (Stefanowicz et al., 2012). CRA and Hevein-Like lectins are chitin-elicitor defense proteins. They play many physiological roles; CRA was reported to locate in quantitative trait loci (QTL) regions associated with leaf, panicle, and roots morphology in *S. bicolor*; this lectin is mainly targeted to the cell wall and the chloroplast where it function in cell wall loosening by controlling its assembly and degradation (Rai et al., 2016, Osman et al., 2022). Plant lectins can also play other physiological functions. Calnexin/calreticulin lectins are endoplasmic reticulum chaperons that assist in *N*-glycosylated protein's general folding, mediating the formation of disulfide bonds and the isomerization of Proline residues (Kozlov and Gehring 2020).

Also, it was shown that plant lectins are essential for nutrient uptake and enhancing crop output, particularly when the plant is experiencing nutrient stress. In roots of the wheat, wheat germ agglutinin (WGA) interacts with rhizobacterium *Azospirilla brasilense* to form nitrogen-fixing symbiosis, such interaction allows the plant to surpass the sever nitrogen starvation and undergoes a whole vegetation period through the upregulation of the WGA (Antonyuk and Evseeva 2006). A *Galanthus nivalis* agglutinin-related and apple domain lectin-1 in *Arabidopsis* (AtGAL1) binds the high-mannose glycan of the purple acid phosphatase (AtPAP26) and enhances its function to facilitate the scavenging of phosphate (Pi) in phosphate-starved *Arabidopsis* plant (Ghahremani et al., 2019). Another study also highlighted the role of plant lectin in the modulation and update of iron in plants under iron-deficiency stress. A mannose-binding lectin-1 (MNB1) is involved in the regulation of the response of *Arabidopsis* growing under iron-deficiency. It's expression is inhibited by iron-deficiency stress, which allows for enhancing the accumulation of iron and increase the plant tolerance to the stress (Song et al., 2022).

Such studies feature the application of plant lectins in increasing the plant tolerance and resistance against various types of environmental and biological stressors, and their potential use as a tool for improving crop productivity and sustainability. However, with no current studies on *Phoenix dactylifera*'s lectins; these proteins' physiological functions and potential roles in response to various environmental stimuli and stresses remain elusive. Some of the reported *P. dactylifera*'s lectins contain anticancer (ACP/AMP) motifs. Combined with the plant's traditional medicinal values can pave the way for future research in medical, pharmaceutical, and agricultural applications.

5. Conclusion

Plant lectins' sugar specificity, structural makeup, and biological roles are diverse. This study is the first to describe the lectins of Phoenix dactylifera; the number of their putative genes, domain configurations and architecture, some aspects of their physiochemical properties, and subcellular locations were highlighted. However, further experimental research is needed to address the biophysiochemical and functional features of mature proteins to comprehend the overall dynamics that enable the plant to endure the hostile environment.

Authors contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Makarim El-Fadil M. Osman, Rieham Sallah H. Osman, Sara A. A. Elmubarak, and Amina I Dirar. Emadeldin Hassan E. Konozy and Makarim El-Fadil M. Osman wrote the first draft of the manuscript. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Ethical approval

No specific permits were required, and all repository data servers were freely accessible.

8. Availability of data and materials

All data generated and analyzed during this study are included in the main article and its supplementary data provided (additional file documents 1–5). Genomic data of *Phoenix dactyliefera* is freely accessible, and all related UTR links were supplied within the article under relevant mentions.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Authors reporting data in additional files 1, 2, and 3 related to the alignment for each family inclusive the phylogenetic tree analysis, the physiochemical characterization of *Phoenix dactylifera* putative lectins, and their amino acid compositions (respectively). Supplementary data to this article can be found online at https://doi.org/10.1016/j.sjbs.2023.103676.

References

Agostino, M., Velkov, T., Dingjan, T., Williams, S.J., Yuriev, E., Ramsland, P.A., 2015. The carbohydrate-binding promiscuity of *euonymus europaeus* lectin is predicted to involve a single binding site. Glycobiology 25 (1), 101–114.

Agrawal, P., Bhagat, D., Mahalwal, M., Sharma, N., Raghava, G.P.S., 2021. Anticp 2.0: an updated model for predicting anticancer peptides. Brief Bioinform. 22 (3). https://doi.org/10.1093/bib/bbaa153.

Al-Alawi, R.A., Al-Mashiqri, J.H., Al-Nadabi, J.S.M., Al-Shihi, B.I., Baqi, Y., 2017. Date palm tree (*phoenix dactylifera* l.): Natural products and therapeutic options. Front Plant Sci. 8:845 https://doi.org/10.3389/fpls.2017.00845.

Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., Nielsen, H., 2019. Signalp 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 37 (4), 420–423.

Al-Mssallem, I.S., Hu, S., Zhang, X., Lin, Q., Liu, W., Tan, J., Yu, X., Liu, J., Pan, L., Zhang, T., 2013. Genome sequence of the date palm *phoenix dactylifera* l. Nat. Commun. 4 (1), 1–9.

Al-Shwyeh, H.A., 2019. Date palm (*phoenix dactylifera* l.) fruit as potential antioxidant and antimicrobial agents. J. Pharm. Bioallied Sci. 11 (1), 1–11. https://doi.org/10.4103/jpbs.JPBS_168_18.

Álvarez-Lugo, A., Becerra, A., 2021. The role of gene duplication in the divergence of enzyme function: A comparative approach. Front. Genet., 12

Antonyuk, L., Evseeva, N., 2006. Wheat lectin as a factor in plant-microbial communication and a stress response protein. Microbiology 75, 470–475.

Asensio, J.L., Cañada, F.J., Siebert, H.-C., Laynez, J., Poveda, A., Nieto, P.M., Soedjanaamadja, U., Gabius, H.-J., Jiménez-Barbero, J., 2000. Structural basis for chitin recognition by defense proteins: Glcnac residues are bound in a multivalent fashion by extended binding sites in hevein domains. Biol. Chem. 7 (7), 529–543.

Assirey, E.A.R., 2015. Nutritional composition of fruit of 10 date palm (*phoenix dactylifera* l.) cultivars grown in saudi arabia. JTUMED. 9(1):75-79 https://doi.org/https://doi.org/10.1016/j.jtusci.2014.07.002.

Babosha, A.V., 2008. Inducible lectins and plant resistance to pathogens and abiotic stress. Biochem. (Mosc.) 73 (7), 812–825. https://doi.org/10.1134/S0006297908070109.

Bourne, Y., Astoul, C.H., Zamboni, V., Peumans, W.J., Menu-Bouaouiche, L., Van Damme, E.J., Barre, A., Rougé, P., 2002. Structural basis for the unusual carbohydrate-binding specificity of jacalin towards galactose and mannose. Biochem. J. 364 (1), 173–180.

Damme, E.J.V., Peumans, W.J., Barre, A., Rougé, P., 1998. Plant lectins: a composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. Crit. Rev. Plant Sci. 17 (6), 575–692.

Delporte, A., Van Holle, S., Lannoo, N., JM Van Damme, E., 2015. The tobacco lectin, prototype of the family of nictaba-related proteins. Curr. Protein Pept. Sci., 16 (1):5-16

Dufayard, J.-F., Bettembourg, M., Fischer, I., Droc, G., Guiderdoni, E., Périn, C., Chantret, N., Diévart, A.J.F.i.P.S., 2017. Corrigendum: New insights on leucine-rich repeats receptor-like kinase orthologous relationships in angiosperms. Front. Plant Sci.(8):381

Eggermont, L., Verstraeten, B., Van Damme, E.J., 2017. Genome-wide screening for lectin motifs in *arabidopsis thaliana*. Plant genome. 10 (2):plantgenome2017.2002.0010

Eggermont, L., Stefanowicz, K., Van Damme, E.J., 2018. Nictaba homologs from *arabidopsis thaliana* are involved in plant stress responses. Front. Plant Sci. 8, 2218.

Fouquaert, E., Peumans, W.J., Vandekerckhove, T., Ongenaert, M., Van Damme, E.J., 2009. Proteins with an *euonymus* lectin-like domain are ubiquitous in embryophyta. BMC Plant Biol. 9 (1), 1–17.

Fouquaert, E., Van Damme, E.J., 2012. Promiscuity of the *euonymus* carbohydrate-binding domain. Biomolecules 2 (4), 415–434.

Gangwar, A.K., Ghosh, A.K., Saxena, V., 2014. Antiulcer activity of chloroform extract of *phoenix dactylifera* linn. leaves. World J. Pharm. Res. 3 (2), 3193–3200.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein identification and analysis tools on the expasy server. The proteomics protocols handbook.571-607

Ghahremani, M., Tran, H., Biglou, S.G., O'Gallagher, B., She, Y.M., Plaxton, W.C., 2019. A glycoform of the secreted purple acid phosphatase atpap26 co-purifies with a mannose-binding lectin (atgal1) upregulated by phosphate-starved arabidopsis. Plant, Cell & Environ. 42 (4), 1139–1157.

Gross-Balthazard, M., Dario, C., Alain, L., Marc, L., Khaled, M., Sylvie, F., Ulise, R., Jianwei, Z., Jayson, T., Seunghee, L., 2019. Genome-wide association mapping of date palm fruit traits. Nat, Commun.

Grover, A., 2012. Plant chitinases: genetic diversity and physiological roles. Crit. Rev. Plant Sci. 31 (1), 57–73.

Gupta, R., Brunak, S., 2001. Prediction of glycosylation across the human proteome and the correlation to protein function. Pac. Symp. Biocomput.

Hall, T.A., 1999. Bioedit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/nt. Nucleic acids symposium series. 41 (41):95-98

Hok, S., Danchin, E.G., Allasia, V., Panabières, F., Attard, A., Keller, H.J.P., cell, environment, 2011. An *arabidopsis* (malectin-like) leucine-rich repeat receptor-like kinase contributes to downy mildew disease. Plant Cell Environ., 34 (11):1944-1957

Hosokawa, N., Wada, I., Hasegawa, K., Yorihuzi, T., Tremblay, L.O., Herscovics, A., Nagata, K., 2001. A novel er α-mannosidase-like protein accelerates er-associated degradation. EMBO Rep. 2 (5), 415–422.

Jeyaprakash, A.A., Jayashree, G., Mahanta, S., Swaminathan, C., Sekar, K., Surolia, A., Vijayan, M., 2005. Structural basis for the energetics of jacalin–sugar interactions: promiscuity versus specificity. J. Mol. Biol. 347 (1), 181–188.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., 2014. Interproscan 5: genome-scale protein function classification. Bioinformatics 30 (9), 1236–1240.

Jones, J.D.G., Dangl, J.L., 2006. The plant immune system. Nature 444 (7117), 323–329. https://doi.org/10.1038/nature05286.

Jung, I.J., Ahn, J.-W., Jung, S., Hwang, J.E., Hong, M.J., Choi, H.-I., Kim, J.-B., 2019. Overexpression of rice jacalin-related mannose-binding lectin (osjac1) enhances resistance to ionizing radiation in *arabidopsis*. BMC Plant Biol. 19 (1), 1–16.

Kersting, A.R., Bornberg-Bauer, E., Moore, A.D., Grath, S., 2012. Dynamics and adaptive benefits of protein domain emergence and arrangements during plant genome evolution. Genome Biol. Evol. 4 (3), 316–329.

Kitaoku, Y., Fukamizo, T., Numata, T., Ohnuma, T., 2017. Chitin oligosaccharide binding to the lysin motif of a novel type of chitinase from the multicellular green alga, *volvox carteri*. Plant Mol. Biol. 93 (1), 97–108.

Konozy, E.H.E., Osman, M.E.-f.M., 2022. Plant lectin: A promising future anti-tumor drug. Biochimie

Konozy, E.H.E., Osman, M.E.-f.M., Dirar, A.I., 2022. Plant lectins as potent anti-coronaviruses, anti-inflammatory, antinociceptive and antiulcer agents. Saudi J. Biol. Sci.:103301

Konozy, E.H.E., Osman, M.E.-f.M., Dirar, A.I., Ghartey-Kwansah, G., 2022. Plant lectins: A new antimicrobial frontier. Biomedicine & Pharmacotherapy. 155:113735

Kozlov, G., Gehring, K., 2020. Calnexin cycle–structural features of the er chaperone system. FEBS J. 287 (20), 4322–4340.

Krogh, A., Larsson, B., Von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden markov model: Application to complete genomes. J. Mol. Biol. 305 (3), 567–580.

Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157 (1), 105–132.

Leach, M.R., Williams, D.B., 2004. Lectin-deficient calnexin is capable of binding class i histocompatibility molecules in vivo and preventing their degradation. J. Biol. Chem. 279 (10), 9072–9079.

Lemlem, A., Alemayehu, M., Endris, M., 2018. Date palm production practices and constraints in the value chain in afar regional state, ethiopia. Adv. Agric. 2018, 6469104. https://doi.org/10.1155/2018/6469104.

Letunic, I., Bork, P., 2021. Interactive tree of life (itol) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 49 (W1), W293–W296.

Liu, J., Huang, Y., Kong, L., Yu, X., Feng, B., Liu, D., Zhao, B., Mendes, G.C., Yuan, P., Ge, D., Wang, W.-M., Fontes, E.P.B., Li, P., Shan, L., He, P., 2020. The malectin-like receptor-like kinase letum1 modulates nlr protein summ2 activation via mekk2 scaffolding. Nat. Plants 6 (9), 1106–1115. https://doi.org/10.1038/s41477-020-0748-6.

Loris, R., Hamelryck, T., Bouckaert, J., Wyns, L., 1998. Legume lectin structure. (BBA)-Protein Struct. Mol. Enzymol. 1383 (1), 9–36.
Macedo, M.L.R., Oliveira, C.F.R., Oliveira, C.T., 2015. Insecticidal activity of plant lectins and potential application in crop protection. Molecules 20 (2), 2014–2033.
Martín-Dacal, M., Fernández-Calvo, P., Jiménez-Sandoval, P., López, G., Garrido-Arandía, M., Rebaque, D., del Hierro, I., Torres, M.Á., Kumar, V., Berlanga, D.J.J.b., 2022. A group of *arabidopsis thaliana* leucine rich repeat-malectin receptor kinases are required for the activation of immune responses triggered by cellulose and mixed-linked glucan oligosaccharides. bioRxiv.2022.2008. 2009.503277 https://doi.org/https://doi.org/10.1101/2022.08.09.503277.
Matejuk, A., Duś, D., 1998. Animal lectins–structure and function. Postepy Hig. Med. Dosw. 52 (5), 445–470.
Mishra, A., Behura, A., Mawatwal, S., Kumar, A., Naik, L., Mohanty, S.S., Manna, D., Dokania, P., Mishra, A., Patra, S.K., Dhiman, R., 2019. Structure-function and application of plant lectins in disease biology and immunity. Food Chem. Toxicol. 134,. https://doi.org/10.1016/j.fct.2019.110827 110827.
Nagano, A.J., Fukao, Y., Fujiwara, M., Nishimura, M., Hara-Nishimura, I., 2008. Antagonistic jacalin-related lectins regulate the size of er body-type β-glucosidase complexes in *arabidopsis thaliana*. Plant Cell Physiol. 49 (6), 969–980.
Ogawa, T., Sato, R., Naganuma, T., Liu, K., Lakudzala, A.E., Muramoto, K., Osada, M., Yoshimi, K., Hiemori, K., Hirabayashi, J., 2019. Glycan binding profiling of jacalin-related lectins from the *pteria penguin* pearl shell. Int. J. Mol. Sci. 20 (18), 4629.
Ohnuma, T., Onaga, S., Murata, K., Taira, T., Katoh, E., 2008. Lysm domains from *pteris ryukyuensis* chitinase-a: a stability study and characterization of the chitin-binding site. J. Biol. Chem. 283 (8), 5178–5187.
Osman, M.E.M., Dirar, A.I., Konozy, E.H.E., 2022. Genome-wide screening of lectin putative genes from *sorghum bicolor* l., distribution in qtls and a probable implications of lectins in abiotic stress tolerance. BMC Plant Biol. 22 (1), 1–15.
Osman, M.E.M., Konozy, E.H.E., 2017. Insight into erythrina lectins: properties, structure and proposed physiological significance. Open Bioact. Compd. J. 5 (1).
Osman, M.E.M., Osman, R.S.H., Ghartey-Kwansah, G., Konozy, E.H.E., 2023. Plant lectins: Implications in tolerance and resistance. Annu. Plant Rev. Online 6, 1–25. https://doi.org/10.1002/9781119312994.apr0804.
Panchy, N., Lehti-Shiu, M., Shiu, S.-H., 2016. Evolution of gene duplication in plants. Plant Physiol. 171 (4), 2294–2316.
Petrova, N., Nazipova, A., Gorshkov, O., Mokshina, N., Patova, O., Gorshkova, T., 2021. Gene expression patterns for proteins with lectin domains in flax stem tissues are related to deposition of distinct cell wall types. Front. Plant Sci. 12,. https://doi.org/10.3389/fpls.2021.634594 634594.
Peumans, W.J., Damme, E.J.V.J.B., Reviews, G.E., 1998. Plant lectins: versatile proteins with important perspectives in biotechnology. Biotechnol. Genet. Eng. Rev. 15 (1), 199–228.
Peumans, W.J., Van Damme, E., 1995. Lectins as plant defense proteins. Plant Physiol. 109 (2), 347.
Rahmani, A.H., Aly, S.M., Ali, H., Babiker, A.Y., Srikar, S.J.I.j.o.c., medicine, e., 2014. Therapeutic effects of date fruits (*phoenix dactylifera*) in the prevention of diseases via modulation of anti-inflammatory, anti-oxidant and anti-tumour activity. Int J Clin Exp Med. 7(3):483
Rai, K.M., Thu, S.W., Balasubramanian, V.K., Cobos, C.J., Disasa, T., Mendu, V., 2016. Identification, characterization, and expression analysis of cell wall related genes in *sorghum bicolor* (l.) moench, a food, fodder, and biofuel crop. Front. Plant Sci. 7, 1287.
Ramachandraiah, G., Chandra, N.R., 2000. Sequence and structural determinants of mannose recognition. Proteins 39 (4), 358–364.
Ramos, M.V., Grangeiro, T.B., Cavada, B.S., Shepherd, I., Lopes, R.O.d.M., Sampaio, A. H., 2000. Carbohydrate/glycan-binding specificity of legume lectins in respect to their proposed biological functions. Braz Arch Biol Technol., 43:349-359
Ren, J., Wen, L., Gao, X., Jin, C., Xue, Y., Yao, X., 2009. Dog 1.0: Illustrator of protein domain structures. Cell Res. 19 (2), 271–273.
Rost, B., 1996. [31] phd: Predicting one-dimensional protein structure by profile-based neural networks. Methods Enzymol., Elsevier. 266, 525–539.
Schallus, T., Jaeckh, C., Fehér, K., Palma, A.S., Liu, Y., Simpson, J.C., Mackeen, M., Stier, G., Gibson, T.J., Feizi, T., 2008. Malectin: a novel carbohydrate-binding protein of the endoplasmic reticulum and a candidate player in the early steps of protein n-glycosylation. J. Mol. Cell Biol. 19 (8), 3404–3414.
Schouppe, D., Rougé, P., Lasanajak, Y., Barre, A., F Smith, D., Proost, P., Van Damme, E. J., 2010. Mutational analysis of the carbohydrate binding activity of the tobacco lectin. Glycoconj. J., 27(6):613-623
Shabana, H., Al-Ani, B., Zaid, A., 2010. On the status of chromosomes of the date palm *(phoenix dactylifera* l.). IV International Date Palm Conference 882.253-268.
Sharon, N., Lis, H., 2007. Specificity and affinity. In: Lectins. Springer, dordrecht. https://doi.org/10.1007/978-1-4020-6953-6_4: 63-103.
Shumayla, Sharma, S., Pandey, A.K., Singh, K., Upadhyay, S.K., 2016. Molecular characterization and global expression analysis of lectin receptor kinases in bread wheat (*triticum aestivum*). PloS one. 11(4):e0153925 https://doi.org/10.1371/journal.pone.0153925
Song, H., Chen, F., Wu, X., Hu, M., Geng, Q., Ye, M., Zhang, C., Jiang, L., Cao, S., 2022. Mnb1 gene is involved in regulating the iron-deficiency stress response in arabidopsis thaliana. BMC Plant Biol. 22 (1), 1–13.
Steentoft, C., Vakhrushev, S.Y., Joshi, H.J., Kong, Y., Vester-Christensen, M.B., Schjoldager, K.T.B., Lavrsen, K., Dabelsteen, S., Pedersen, N.B., Marcos-Silva, L., 2013. Precision mapping of the human o-galnac glycoproteome through simplecell technology. EMBO J. 32 (10), 1478–1488.
Stefanowicz, K., Lannoo, N., Proost, P., Van Damme, E.J., 2012. *Arabidopsis* f-box protein containing a nictaba-related lectin domain interacts with n-acetyllactosamine structures. FEBS Open Bio. 2, 151–158.
Sun, Y., Qiao, Z., Muchero, W., Chen, J.G., 2020. Lectin receptor-like kinases: The sensor and mediator at the plant cell surface. Front. Plant Sci. 11,. https://doi.org/10.3389/fpls.2020.596301 596301.
Sussholz, O., Pizarro, L., Schuster, S., Avni, A., 2020. Slrlk-like is a malectin-like domain protein affecting localization and abundance of leeix2 receptor resulting in suppression of eix-induced immune responses. Plant J. 104 (5):1369-1381 https://doi.org/https://doi.org/10.1111/tpj.15006.
Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., Minh, B.Q., 2016. W-iq-tree: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. 44 (W1), W232–W235.
Van Damme, E.J.M., 2022. 35 years in plant lectin research: a journey from basic science to applications in agriculture and medicine. Glycoconj. J. 39 (1), 83–97. https://doi.org/10.1007/s10719-021-10015-x.
Van Damme, E.J., Barre, A., Rougé, P., Peumans, W.J., 2004. Cytoplasmic/nuclear plant lectins: a new story. Trends Plant Sci. 9 (10), 484–489.
Van Damme, E.J., Culerrier, R., Barre, A., Alvarez, R., Rougé, P., Peumans, W.J., 2007. A novel family of lectins evolutionarily related to class v chitinases: an example of neofunctionalization in legumes. Plant Physiol. 144 (2), 662–672.
Van Damme, E.J.M., 2014. History of plant lectin research. Lectins: Methods and Protocols.3-13 https://doi.org/10.1007/978-1-4939-1292-6_1.
Van Holle, S., Van Damme, E.J., 2015. Distribution and evolution of the lectin family in soybean (*glycine max*). Molecules 20 (2), 2868–2891.
Vandenborre, G., Smagghe, G., Van Damme, E.J.J.P., 2011. Plant lectins as defense proteins against phytophagous insects. Phytochemistry 72 (13), 1538–1550.
Wright, C.S., 1990. 2.2 a resolution structure analysis of two refined n-acetylneuraminyl-lactose–wheat germ agglutinin isolectin complexes. J. Mol. Biol. 215 (4), 635–651. https://doi.org/10.1016/s0022-2836(05)80174-3.
Xing, S., Li, M., Liu, P., 2013. Evolution of s-domain receptor-like kinases in land plants and origination of s-locus receptor kinases in brassicaceae. BMC Evol. Biol. 13 (1), 1–11.
Yu, C.-S., Cheng, C.-W., Su, W.-C., Chang, K.-C., Huang, S.-W., Hwang, J.-K., Lu, C.-H., 2014. Cello2go: a web server for protein subcellular localization prediction with functional gene ontology annotation. PloS One 9 (6), e99368.
Zhao, L., Poschmann, G., Waldera-Lupa, D., Rafiee, N., Kollmann, M., Stühler, K., 2019. Outcyte: a novel tool for predicting unconventional protein secretion. Sci. Rep. 9 (1), 19448. https://doi.org/10.1038/s41598-019-55351-z.