Pertanika J. Trop. Agri. Sci. 46 (3): 815 - 838 (2023)

PERTANIKA JOURNALS

TROPICAL AGRICULTURAL SCIENCE

Journal homepage: http://www.pertanika.upm.edu.my/

In silico Comparative Analysis of Gene and Protein of Plant Lectins

Fathiya Khairiya[1], Fenny Martha Dwivany[1,2,3]*, Sony Suhandono[1], Sofia Safitri Hessel[1], Ima Mulyama Zainuddin[4] and Trina Ekawati Tallei[5]

[1]School of Life Sciences and Technology, Institut Teknologi Bandung, Jalan Ganesha No. 10, Bandung 40132, West Java, Indonesia
[2]Bali International Research Center for Banana, Badung 80361, Bali, Indonesia
[3]Bioscience and Biotechnology Research Center, Institut Teknologi Bandung, Bandung 40132, West Java, Indonesia
[4]Department of Biosystems, KU Leuven, Leuven 3000, Belgium
[5]Department of Biology, Faculty of Mathematics and Natural Sciences, Sam Ratulangi University, Manado 95115, North Sulawesi, Indonesia

ABSTRACT

Lectins are a family of proteins that can recognize and bind specific carbohydrates. Plant lectins play various roles in plant defense and can be utilized as insecticidal, antibacterial, antifungal, and antiviral agents. This study compares genes, proteins, and carbohydrate-binding motifs between 15 plant lectins using *in silico* methods. The lectin genes of *Artocarpus hypargyreus* Hance, *Hordeum vulgare* var. Betzes, *Triticum aestivum* L. cv. Marshall, *Galanthus nivalis* L., *Allium sativum* L., *Phaseolus vulgaris*, *Lens culinaris* subsp. *tomentosus*, *Robinia pseudoacacia*, *Glycine max*, *Cicer arietinum*, *Pisum sativum*, *Canavalia ensiformis*, *Amaranthus caudatus*, *Amaranthus hypochondriacus*, and *Musa acuminata* subsp. *malaccensis* were obtained from National Center for Biotechnology Information and Banana Genome Hub. The gene comparison results revealed different characteristics of the 15 plant lectin genes, with *A. hypargyreus* having the shortest lectin gene and *G. max* having the longest. Overall, the 15 plant lectin genes have 1–3 exons. Domain predictions revealed the presence of five domains: jacalin, chitin_bind_1, B_lectin, legume lectin, and agglutinin. Furthermore,

ARTICLE INFO

Article history:
Received: 14 December 2022
Accepted: 06 March 2023
Published: 09 June 2023

DOI: https://doi.org/10.47836/pjtas.46.3.06

E-mail addresses:
fenny@sith.itb.ac.id (Fenny Martha Dwivany)
fathiyakhairiya@students.itb.ac.id (Fathiya Khairiya)
sony@sith.itb.ac.id (Sony Suhandono)
hesselsafitrisofia@gmail.com (Sofia Safitri Hessel)
ima.zainuddin@kuleuven.be (Ima Mulyama Zainuddin)
trina_tallei@unsrat.ac.id (Trina Ekawati Tallei)
*Corresponding author

ISSN: 1511-3701
e-ISSN: 2231-8542

there were 2 protein sequences from the jacalin domain, 2 protein sequences from the chitin_bind_I domain, 2 protein sequences from the B_lectin domain, and 4 protein sequences from the legume lectin domains that have complete carbohydrate-binding motifs compared to consensus motifs from literature. The data obtained from this study has not been previously reported and can be utilized for future lectin protein production with synthetic biology approaches. This method will allow scientists to obtain plant bioparts for lectin production using a heterologous system, even without plant samples.

Keywords: *In silico* comparative study, plant lectins, synthetic biology

INTRODUCTION

Lectins are a family of proteins with heterogeneous structures that can recognize and bind to specific carbohydrates. Lectins are found in various organisms, including bacteria, fungi, animals, and plants, and serve various functions. Lectins function in mediating biological processes such as cell-to-cell communication and cell migration. In addition, cell interactions with several types of lectins are known to cause cell agglutination or deposition of glycoconjugates and polysaccharides, which makes lectins in some organisms naturally able to act as agglutinins. Several studies discussed the results of conventional lectin isolation from various organisms and demonstrated that lectins in plants have higher protein yields than other lectin-producing organisms such as fungi and animals (Lam & Ng, 2011).

In plants, there are several lectin domains, including jacalin, chitin-binding, nictaba (*Nicotiana tabacum* agglutinin), *Galanthus nivalis* agglutinin (GNA), legume, hevein, amaranthin, and *Euonymus europaeus* lectin (EEL). The main function of plant lectins is to act as a defense system and to facilitate adaptation to abiotic and biotic stresses. Additional studies have also examined the function of plant lectins as insecticidal, antibacterial, antifungal, and antiviral agents (Tsaneva & Van Damme, 2020).

Several plant lectins, including legume, hevein, nictaba, and jacalin, have been shown in previous studies to have strong insecticidal activity against Lepidoptera, Coleoptera, and Diptera (Vandenborre et al., 2011). Plant lectins can act as antifungals by inhibiting pathogenic fungal growth (Coelho et al., 2017). Plant lectins with jacalin-related domains derived from bananas (*Musa acuminata*) are known to exhibit antiviral activity by recognizing glycans on the surface of the HIV-1 virus envelope, thereby inhibiting the entry of the virus into the host cell (Swanson et al., 2010). Another study showed that a *Lens culinaris*-derived plant lectin with a legume domain possesses antiviral activity against multiple SARS-CoV-2 variants (Wang et al., 2021). Another study demonstrated that lectins from *Archidendron jiringa* inhibited the growth of several bacteria, including *Bacillus subtilis* and *Staphylococcus* sp. (Charungchitrak et al., 2011).

A controversy in the utilization of lectins was reported in a publication by Singh et al. (2014), which is that wild-type banana lectin is a T-cell mitogen. A novel molecularly engineered lectin, H84T banana lectin (H84T), was reported to have real potential for clinical use against influenza by blocking the ability of the influenza virus to fuse with endosomes (Covés-Datson et al., 2020). The H84T was reported to be a potent inhibitor of the Ebola virus by blocking cellular entry and transcription/replication (Covés-Datson et al., 2019). Clinical use of lectins has been slowed due to concerns about their mitogenic activity toward immune cells. The H84T banana lectin was engineered by mutating a single amino acid, histidine, to threonine at position 84 to preserve the antiviral activity and lose the mitogenicity (Covés-Datson et al., 2021).

Importantly, *in silico* comparative analysis of plant lectin genes and proteins is required to characterize plant lectin sequences. In this case, *in silico* methods can be utilized in gene structure characterization and protein structure and function prediction. *In silico* methods have been widely used to characterize proteins and enzymes from various eukaryotic and prokaryotic species. Understanding the characteristics of lectin genes and proteins from various plant species is essential for developing more advanced lectin proteins. Parallel to this study, we also conducted an *in silico* analysis of the ability of wild-type and modified *M. acuminata* (H84T) and *M. balbisiana* lectins to inhibit the activity of the SARS-CoV-2 spike protein using molecular docking and molecular dynamics simulation. The results demonstrated that the modeled lectins and the SARS-CoV-2 RBD had good protein-protein interactions and strong binding (Hessel et al., 2022).

In addition to the study of plant lectins, methods for optimizing the production of lectins have also been investigated, and synthetic biology may offer a novel approach in this regard. The application of synthetic biology in agriculture, health, and pharmacy has been reviewed by Scott et al. (2015). However, this method has not yet been documented in plant lectin synthesis. Thus, the possibility remains open.

Therefore, this research employed an *in silico* method to compare plant lectin genes and proteins. A comparison of the carbohydrate-binding site motifs on plant lectins was also carried out to analyze the motifs related to their function in binding specific carbohydrates. This research's findings can serve as a reference for characterizing plant lectin sequences in the context of synthetic biology-based lectin protein production.

MATERIALS AND METHODS

Collection and Compilation of Plant Lectin Gene Sequences

Plant lectin gene sequences from 14 different plant species were obtained from the National Center for Biotechnology Information (NCBI) platform (https://www.ncbi.nlm.nih.gov/). The banana lectin gene was retrieved from the Banana Genome Hub platform (version 2) (https://banana-genomehub.southgreen.fr). The lectin gene sequences (accession/hit numbers)

acquired from 15 plants were *Musa acuminata* subsp. *malaccensis* (Ma09_t10460.1), *Artocarpus hypargyreus* Hance (KY924610.1), *Hordeum vulgare* var. Betzes (M29280.1), *Triticum aestivum* L. cv. Marshall (M25537.1), *Galanthus nivalis* L. (M55556.1), *Allium sativum* L. (U58948.1), *Phaseolus vulgaris* (AJ439715.1), *Lens culinaris* subsp. *tomentosus* (AJ421799.2), *Robinia pseudoacacia* (AB012633.1), *Glycine max* (NM_001250281.3), *Cicer arietinum* (XM_004509655.2), *Pisum sativum* (L11745.1), *Canavalia ensiformis* (AF308777.1), *Amaranthus caudatus* (AF401479.1), and *A. hypochondriacus* (AF143954.1).

Plant Lectin Protein Sequence Prediction

The prediction of 15 plant lectin protein sequences was made using the BLASTX program (https://blast.ncbi.nlm.nih.gov/Blast.cgi?LINK_LOC=blasthome&PAGE_TYPE=BlastSearch&PROGRAM=blastx) on the NCBI web page. The principle of BLASTX is to compare the input gene sequences (nucleotides) with the protein sequence database in NCBI.

Prediction and Visualization of Plant Lectin Gene Structure

Structural prediction of the 15 plant lectin gene sequences was performed using the FGENESH+ online tool on the Softberry web page (http://www.softberry.com/berry.phtml?topic=fgenes_plus&group=programs&subgroup=gfs) using the similar protein-based gene prediction approach and GeneWise (version 2022) (https://www.ebi.ac.uk/Tools/psa/genewise/) (Xiong, 2006). The homologous protein sequences were obtained from the previous step's BLASTX results of protein sequences. Visualization of the structures was carried out using Illustrator for Biological Sequence (IBS) (version 1.0.3) (W. Liu et al., 2015). Following this, the genomes, exons, and introns sizes of the 15 plant lectin gene sequences were compared.

Plant Lectin Protein Phylogenetic Tree Construction

The MEGA-X (version 10.1.7) was employed to construct the phylogenetic tree of the 15 plant lectin protein sequences using the MUSCLE alignment method and the Maximum Likelihood parameter with a 1,000x bootstrap (Kamalesha et al., 2022). The phylogenetic tree construction was done to determine the clustering and evolutionary relationships of the 15 plant lectin protein sequences (Kumar et al., 2018).

Analysis of Plant Lectin Protein Domains and Motifs

Protein domains and motif analysis of plant lectins were performed using the Conserved Domain (CD)-Search Tool (version 2020) (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) on the NCBI web page with an e-value threshold of 0.01 (Lu et al., 2020) and visualized using IBS. The Multiple Em for Multiple Elicitation (MEME)-Suite (version 5.4.1) (https://meme-suite.org/meme/tools/meme) (Bailey et al., 2009) was used for the prediction of plant lectin motifs. The parameter settings used were a minimum motif length

parameter of 10 amino acids (aa) and a maximum motif length of 50 aa, with other parameters remaining in the default settings (Dwivany et al., 2021; Kamalesha et al., 2022). Consensus motif detection and validation were performed using the InterPro online tool (version 88.0) (https://www.ebi.ac.uk/interpro/search/sequence/) (Mitchell et al., 2019). In addition, the motifs of carbohydrate-binding sites on plant lectin protein sequences were analyzed and compared to the consensus motifs from previous studies using multiple sequence alignment (MSA) with the MUSCLE method integrated into MEGA-X (version 10.1.7) (Kumar et al., 2018).

RESULTS AND DISCUSSION

In silico Characterization of Plant Lectin Genes

The lectin genes from 14 plant species and one banana species obtained from NCBI and Banana Genome Hub were used to identify the lectin type (Table 1).

Table 1

The identified lectin types from 15 plant species

No.	Plant species	Family	Accession number/Hits	Type of lectins
1	*Musa acuminata* subsp. *malaccensis*	Musaceae	Ma09_t10460.1	Lectin
2	*Artocarpus hypargyreus* Hance	Moraceae	KY924610.1	Lectin
3	*Hordeum vulgare* var. Betzes	Poaceae	M29280.1	Lectin
4	*Triticum aestivum* L. cv. Marshall		M25537.1	Wheat germ agglutinin isolectin D
5	*Galanthus nivalis* L.	Amaryllidaceae	M55556.1	Lectin GNA 2
6	*Allium sativum* L.		U58948.1	Mannose-specific lectin
7	*Phaseolus vulgaris*	Fabaceae	AJ439715.1	Lec4-B17 gene
8	*Lens culinaris* subsp. *tomentosus*		AJ421799.2	Lectin
9	*Robinia pseudoacacia*		AB012633.1	Lectin
10	*Glycine max*		NM_001250281.3	Lectin (LOC732576)
11	*Cicer arietinum*		XM_004509655.2	Seed lectin-like
12	*Pisum sativum*		L11745.1	Lectin (Blec4)
13	*Canavalia ensiformis*		AF308777.1	Con A
14	*Amaranthus caudatus*	Amaranthaceae	AF401479.1	Agglutinin
15	*Amaranthus hypochondriacus*		AF143954.1	Agglutinin

Several plant lectin types identified here were the amaranthins, the chitinase-related agglutinins, the GNA-related lectins, the hevein domain lectins, the jacalin-related lectins, and the legume lectins (Van Damme, 2022). The *in silico* comparative analysis of the lectin genes and proteins between these 15 plant species has not been reported yet.

The prediction and visualization of the 15 identified plant lectin genes by utilizing the FGENESH+ and IBS provided information on the length of the gene, exon, and poly-A signal (Figure 1).

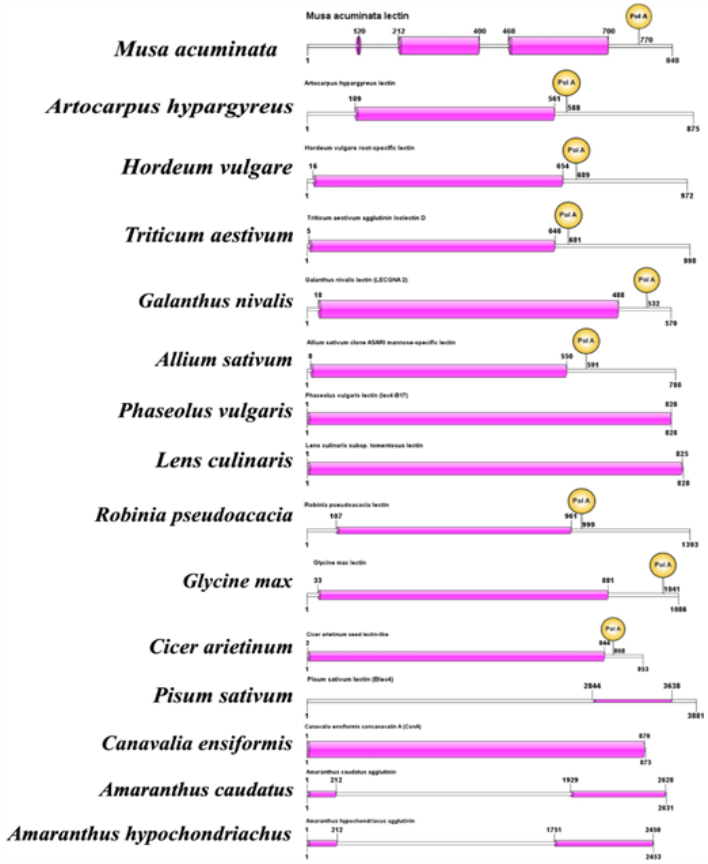


Figure 1. Gene structure of the 15 plant lectin genes

None of the 15 plant lectin gene sequences examined contained a transcription start site (TSS). As a result, the gene sizes were calculated starting from the first exon to the poly-A signal. A poly-A signal was found in nine plant lectin genes, but none were found in the six remaining genes. Therefore, the sizes of these six lectin genes were calculated from the first to the last exon. The lectin gene of *G. max*, measuring 1,009 bp, was the longest one. Furthermore, the shortest gene detected was

from *A. hypargyreus*, with a length of 480 bp. Plant lectin gene sequences ranged from 495 to 912 bp.

The number of exons in the lectin gene from 15 plants was also different; the highest number of exons detected was three in *M. acuminata*. Meanwhile, two exons were found in the lectin gene of *A. caudatus* and *A. hypochondriacus*. The remaining plant lectin genes had only one exon in each sequence. Table 2 displays the structural characteristics of the 15 plant lectin genes, including exon number and length, coding sequence (CDS) length, and mRNA length.

Table 2
Structural characteristics of the 15 plant lectin genes

No.	Plant species	Exon number	CDS length	mRNA length
1	*Musa acuminata* subsp. *malaccensis*	3	426	495
2	*Artocarpus hypargyreus* Hance	1	453	480
3	*Hordeum vulgare* var. Betzes	1	639	674
4	*Triticum aestivum* L. cv. Marshall	1	642	677
5	*Galanthus nivalis* L.	1	471	515
6	*Allium sativum* L.	1	543	584
7	*Phaseolus vulgaris*	1	828	828
8	*Lens culinaris* subsp. *tomentosus*	1	828	828
9	*Robinia pseudoacacia*	1	855	893
10	*Glycine max*	1	849	1,009
11	*Cicer arietinum*	1	843	867
12	*Pisum sativum*	1	755	755
13	*Canavalia ensiformis*	1	870	870
14	*Amaranthus caudatus*	2	912	912
15	*Amaranthus hypochondriacus*	2	912	912

The lectin gene sequences identified in *M. acuminata*, *A. caudatus*, and *A. hypochondriacus* occur at multiple locations in the genome. It indicates a duplication event in the lectin gene. One of the most important duplication events in the history of evolution and lectin expansion in the jacalin family occurred in the genome of *M. acuminata* (Van Holle & Van Damme, 2019).

Lectins are diverse in almost every aspect, including sequences, structures, binding site architectures, quaternary structures, carbohydrate affinities, and specificities. Not only that, their larger biological roles and potential applications

are wide as well (Van Damme, 2014). Another *in silico* comparative analysis study showed that the legume lectin sequences ranged from 700 to 900 bp. Hundreds of structures of plant lectins have been characterized, and legume lectins are the most extensively studied (Moraes Filho et al., 2017).

Until this day, lectins have been extensively studied in terms of their molecules and proteins due to various reports on their biomedical applications. However, proteomics and genomic analysis are still needed to explore various aspects of these proteins. More sequence information on lectins would be important for structural references and functional characterization, so more focus can be directed toward developing genetic engineering to produce recombinant lectins (Singh et al., 2014). Therefore, more information about different types of plant lectin genes and proteins from various species is required.

The 15 plant lectin proteins were characterized by identifying protein motifs and domains to predict protein classification, identify conserved sequences, and predict protein function (Table 3, Figure 2).

Table 3

Domain identification of 15 plant lectin proteins using the CD-search tool

No.	Lectin source	Identified domain
1	*Musa acuminata* subsp. *malaccensis*	Jacalin
2	*Artocarpus hypargyreus* Hance	
3	*Hordeum vulgare* var. Betzes	Chitin_bind_1
4	*Triticum aestivum* L. cv. Marshall	
5	*Galanthus nivalis* L.	B_lectin
6	*Allium sativum* L.	
7	*Phaseolus vulgaris*	Legume lectin
8	*Lens culinaris* subsp. *tomentosus*	
9	*Robinia pseudoacacia*	
10	*Glycine max*	
11	*Cicer arietinum*	
12	*Pisum sativum*	
13	*Canavalia ensiformis*	
14	*Amaranthus caudatus*	Agglutinin
15	*Amaranthus hypochondriacus*	

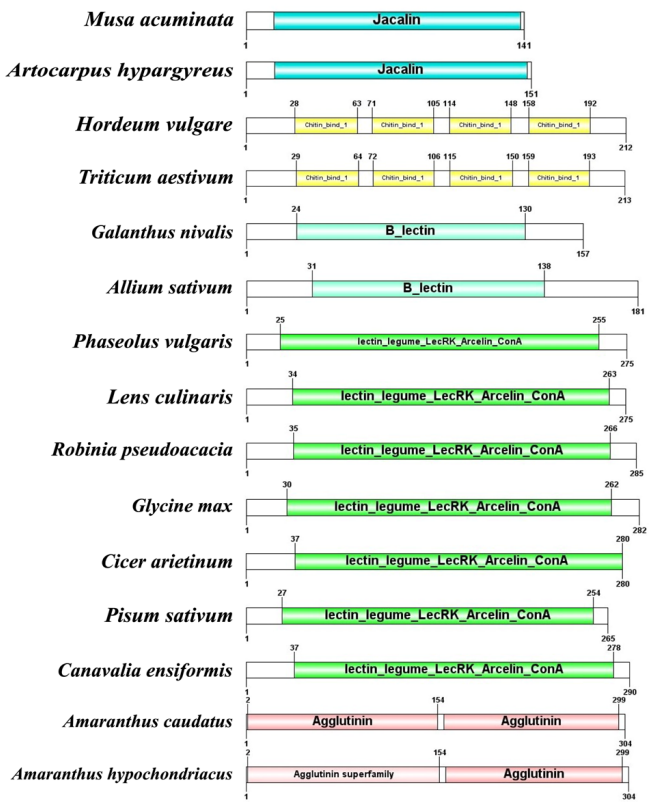


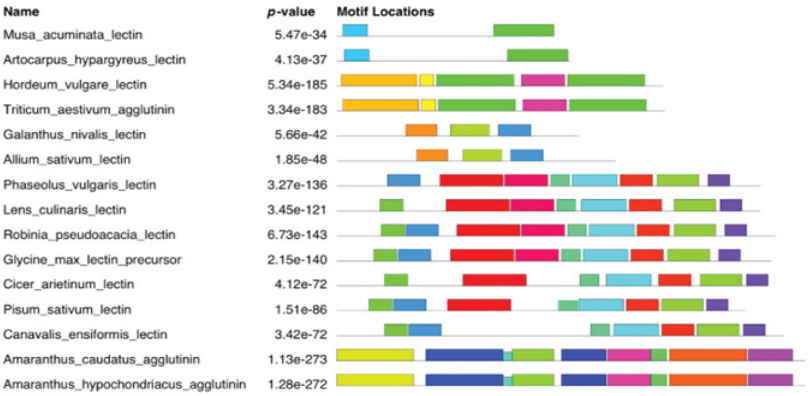
Figure 2. Visualization of domain positions on 15 plant lectin protein sequences

The results indicated that the location of the domains in the plant lectin protein sequences consisting of the same domain was quite uniform. In *M. acuminata* and *A. hypargyreus*, the jacalin domain was located between the 15 and 149 aa residues. In *H. vulgare* and *T. aestivum*, the four locations of the chitin_bind_1 domain were discovered in relatively consistent locations, specifically between the residues of 28–64 aa, 71–106 aa, 114–150 aa, and 158–193 aa. The B_lectin domain of *G. nivalis* and *A. sativum* was located between the 28 and 138 aa residues. The lectin_legume_LecRK_Arcelin_ConA domain, commonly called the "legume lectin domain", was located between 24 and 280 aa residues in seven species. Agglutinin domains were found in *A. caudatus* and *A. hypochondriacus* between residues 2–154 aa and 159–299 aa.

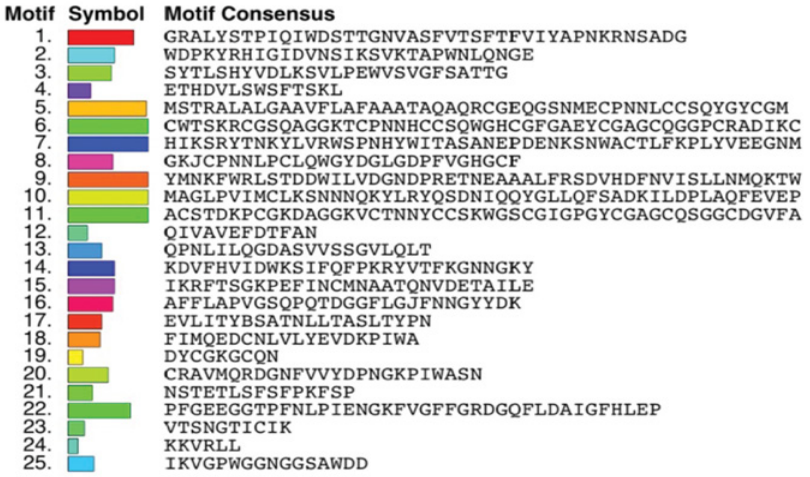
Identifying protein motifs from plant lectins yielded 25 consensus motifs (Figure 3). The results of the validation of these motifs are shown in Table 4.

The most important characteristic of lectins is a domain that can recognize and bind to carbohydrate structures specifically and reversibly (Peumans & Van Damme, 1995). Plant lectins can be classified based on the lectin domain, which consists of consensus motif sequences (Van Damme et al., 2008). A study showed that some lectin

domains are highly conserved in plants (Van Holle et al., 2017). Combinations of domains allow lectin domains to have new functions. The high diversity of domain architectures can make domains in lectins have more specialized roles (Dang & Van Damme, 2016).



(a)



(b)

Figure 3. A motif prediction of 15 plant lectin proteins: (a) location of predicted consensus motifs based on MEME-Suite, (b) 25 predicted consensus motif sequences

Table 4

Consensus motifs found in plant lectin protein sequences and InterPro detection results

No.	Consensus motif	InterPro detection
1	GRALYSTPIQIWDSTTGNVASFVTSFTFVIYAPNKRNSADG	Legume lectin

Table 4 *(Continue)*

No	Consensus motif	InterPro detection
2	WDPKYRHIGIDVNSIKSVKTAPWNLQNGE	Legume lectin
3	SYTLSHYVDLKSVLPEWVSVGFSATTG	Legume lectin
4	ETHDVLSWSFTSKL	-
5	MSTRALALGAAVFLAFAAATAQAQRCGEQGSNMECPNNLCCSQYGYCGM	Chitin-bd_1
6	CWTSKRCGSQAGGKTCPNNHCCSQWGHCGFGAEYCGAGCQGGPCRADIKC	Chitin-bd_1
7	HIKSRYTNKYLVRWSPNHYWITASANEPDENKSNWACTLFKPLYVEEGNM	Agglutinin
8	GKJCPNNLPCLQWGYDGLGDPFVGHGCF	-
9	YMNKFWRLSTDDWILVDGNDPRETNEAAALFRSDVHDFNVISLLNMQKTW	Agglutinin
10	MAGLPVIMCLKSNNNQKYLRYQSDNIQQYGLLQFSADKILDPLAQFEVEP	Agglutinin
11	ACSTDKPCGKDAGGKVCTNNYCCSKWGSCGIGPGYCGAGCQSGGCDGVFA	Chitin-bd_1
12	QIVAVEFDTFAN	-
13	QPNLILQGDASVVSSGVLQLT	-
14	KDVFHVIDWKSIFQFPKRYVTFKGNNGKY	-
15	IKRFTSGKPEFINCMNAATQNVDETAILE	Agglutinin
16	AFFLAPVGSQPQTDGGFLGJFNNGYYDK	-
17	EVLITYBSATNLLTASLTYPN	-
18	FIMQEDCNLVLYEVDKPIWA	Bulb-type _lectin
19	DYCGKGCQN	-
20	CRAVMQRDGNFVVYDPNGKPIWASN	Bulb-type_ lectin
21	NSTETLSFSFPKFSP	-
22	PFGEEGGTPFNLPIENGKFVGFFGRDGQFLDAIGFHLEP	Jacalin-like lectin
23	VTSNGTICIK	-
24	KKVRLL	-
25	IKVGPWGGNGGSAWDD	Jacalin-like lectin

Plant Lectin Phylogenetic Tree

According to the constructed phylogenetic tree shown in Figure 4, lectin proteins derived from *M. acuminata* and *A. hypargyreus* belong to the same clade, which exhibits a close relationship under their similar domains, namely, Jacalin. Then, *H. vulgare*, and *T. aestivum* species are also closely related in the tree, as indicated by the chitin_bind_1 domain that they possess. Furthermore, *G. nivalis* and *A. sativum* are positioned in the same clade, corresponding to the domain they both share, B_lectin. Two other species, *A. caudatus*, and *A. hypochondriacus*, form a clade according to their matching agglutinin domains.

Seven species, including *C. arietinum*, *P. sativum*, *C. ensiformis*, *L. culinaris*, *R. pseudoacacia*, *P. vulgaris*, and *G. max*, form a clade that demonstrates the close relationship of all seven lectin proteins. Consistently, each of the seven proteins shares a domain known as the legume lectin domain. In addition, it was observed that the legume lectin clade resulted from a different tree branching than the other species with the four other domains (jacalin, chitin_bind_1, B_lectin, and agglutinin). Legume lectin domains are mentioned in another study as being strongly conserved, which emphasizes their importance. The amaranthins, the chitinase-related agglutinins, the GNA-related lectins, the hevein domain lectins, the jacalin-related lectins, and the legume lectins are only found in vascular plants (Naithani et al., 2021).

The lectin genes in higher plants belong to the group of multi-copy genes. Duplication of genes plays an important role in the process of evolution and expansion of a gene (Van Holle et al., 2017). Tandem and segmental duplication are mechanisms for lectin gene expansion and play a role in the adaptation process of plants to various biotic and abiotic stresses (Jiang et al., 2010). Each plant lectin family shows a different evolutionary path as it is formed by full and partial gene duplication based on addition,

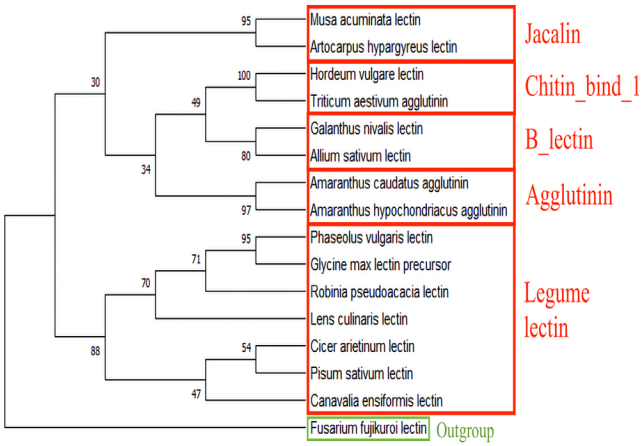


Figure 4. Phylogenetic tree of the 15 plant lectin protein sequences. The red boxes show the relationship between lectin sequences possessing the same domains

reduction, and/or recombination in each protein domain (Naithani et al., 2021).

In silico Analysis of Carbohydrate Binding Sites Motif on Plant Lectins

Carbohydrate binding motifs on plant lectin sequences serve as recognition sites for specific carbohydrates, which will later function as binding sites between lectin proteins and carbohydrates (monosaccharides to oligosaccharides). It is related to the function of lectin proteins in plants, such as their roles in pathogen defense mechanisms, abiotic stress responses, and the antiviral, antifungal, and antibiotic properties of plant lectins (Song et al., 2014).

Research on carbohydrate binding site motifs on legume lectins has previously been carried out by Cummings et al. (2017), which explained that there are 4 loops (consisting of loops A, B, C, and D) associated with carbohydrate binding sites. Below are the results obtained for CBS motif analysis of sequences with legume lectin domains.

As depicted in Figure 5, the key amino acid D132 (aspartic acid at position 132) in loop A contributes to establishing hydrogen bonds between the side chain and carbohydrate ligands (Katoch & Tripathi, 2021). It is denoted by a thick blue box containing a caret symbol (^). In addition, loop A consists of conserved amino acids, namely G133 (glycine in position 133) and F136 (phenylalanine in position 136), which are denoted by a red box and an asterisk symbol (*). Furthermore, loop B has G154 (glycine in position 154) as a key amino acid that forms hydrogen bonds. One of the key causes for preserving the structure of legume lectins is hydrogen bonding, which also happens to be one of the few factors that affect lectins' carbohydrate-binding specificity (Katoch & Tripathi, 2021). The MSA comparison reveals that in the same amino acid position, *P. sativum* and *C. ensiformis* contain key amino acids that are not consistent with previous research, specifically Q154 (glutamine in position 154) in *P. sativum* and R154 (arginine in position 154) in *C. ensiformis* (Cummings et al., 2017).

Loop C consists of several conserved amino acids, namely V185 (valine in the position of 185), E186 (glutamic acid in

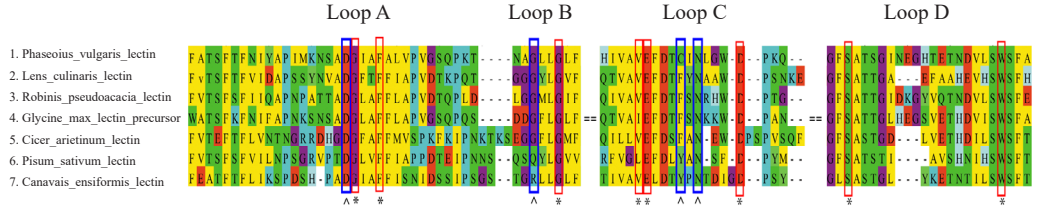


Figure 5. Comparison of carbohydrate-binding site motifs in plant lectin sequences with legume lectin domains based on the similarity of motifs in a previous study (Cummings et al., 2017)

Note. Red boxes and asterisk (*) symbols indicate conserved amino acids, while bold blue boxes and caret (^) symbols signify key amino acids that play a role in hydrogen bond formation

the position of 186), and D197 (aspartic acid in the position of 197), as indicated by a red box and an asterisk (*) in Figure 5. Nonetheless, according to the MSA results, there are two sequences with differences in the amino acid column/position where valine is the conserved amino acid. The *G. max* sequence has isoleucine in position 185, whereas the *P. sativum* sequence has a leucine in position 185. Loop C also has a conserved key amino acid, N192 (asparagine in position 192), which plays a role in hydrogen bond formation with the binding sugar's hydroxyl group through amide acid groups (Cummings et al., 2017). Furthermore, loop C also has key amino acids in the form of hydrophobic amino acids, including C190 (cysteine in position 190), F190 (phenylalanine in position 190), and V185 (valine in position 185). These hydrophobic amino acids may act as hydrophobic cavities that play a role in non-covalent interactions, although those interactions do not directly affect the carbohydrate-binding specificity of legume lectins (Katoch & Tripathi, 2021; Srinivas et al., 2000). Key amino acids are indicated by a bold blue box and a caret symbol (^). According to the MSA results, hydrophobic amino acids, part of the key amino acids, were not found in *P. sativum* and *C. ensiformis* lectin sequences. Instead, both sequences had a neutral amino acid, Y190 (tyrosine at position 190), instead of a hydrophobic one. The reason behind this is unknown, although several factors, such as random mutagenesis, may be the cause. Further structural analysis should be conducted to understand the effects of this substitution on the protein's structure. In loop D, the conserved amino acids were S285 (serine in position 285) and W303 (tryptophan in position 303), as indicated in Figure 5 by a red box and an asterisk (*).

The consensus motifs from this analysis were then compared to the reference literature (loops A-D), Cummings et al. (2017). Our findings revealed three sequences, namely *G. max*, *P. sativum*, and *C. ensiformis*, exhibiting a mismatch of key or conserved amino acids. However, the sequences from *P. vulgaris*, *L. culinaris*, *R. pseudoacacia*, and *C. arietinum* showed motif similarity with the referenced literature.

Similar to the previous study by Raval et al. (2004), the second domain, jacalin, was detected in two species with the same carbohydrate binding site motif. Jacalin carbohydrate binding site motifs are "GG", which forms a GG loop with said glycine amino acids positioned at 15 and 16, and "GXXXD" (with X representing any amino acid), which forms a binding loop in the position ranging through the amino acids 141 to 145 (Figure 6). Based on the MSA result, *M. acuminata* and *A. hypargyreus* consist of the jacalin domain consensus motif under the reference.

The GXXXD motif was considered a preferred characteristic in banana lectin (BanLec) sequences to be used and/or modified as a ligand-binding substance. The motif indicates ligand-binding sites (Covés-Datson et al., 2021; Meagher et al., 2005). The previously mentioned parallel study by Hessel et al. (2022) utilized this motif

Figure 6. Comparison of carbohydrate-binding site motifs in plant lectin sequences with jacalin domain based on the similarity of motifs found in a previous study (Raval et al., 2004). Jacalin's conserved motifs are "GG" and "GXXXD"

Note. The asterisk (*) symbol indicates conserved amino acids; "X" indicates any amino acid

as a selection parameter for banana lectin gene sequences. The sequences were then modeled and docked with the SARS-CoV-2 spike protein receptor-binding domain (RBD). The findings revealed an abundance of interactions between the two proteins and an adequate number of hydrogen bonds and salt bridges. These bonds accommodate the high binding affinities obtained from molecular docking analysis. It also showed that a single lectin excelled in the stability of interaction and binding with the SARS-CoV-2 spike protein RBD through molecular dynamics analysis.

The third domain, chitin_bind_1, is 43 residues long and was detected in two species that showed the same carbohydrate binding site motif as a previous study by Wright et al. (1991). This domain is frequently present in plant proteins that bind to N-acetylglucosamine and its polymers and is assumed to play a role in the recognition and binding of chitin subunits (Butler et al., 1991; Wright et al., 1991). It is known that chitin_bind_1 contains a motif of eight "C" or cysteine amino acids, which play a role in forming four disulfide bonds and are often found at the N-terminus (Wright et al., 1991). According to the MSA results, these eight "C" amino acid motifs were found in *H. vulgare* and *T. aestivum* species (Figure 7). The cysteine amino acids were positioned at the 30th, 39th, 44th, 45th, 51st, 58th, 62nd, and 67th amino acids, with the motif found being "CGXXXXXXXCXXXXCCSXXGXCGXXXXXCXXXCXXXXC", where X indicates any amino acid.

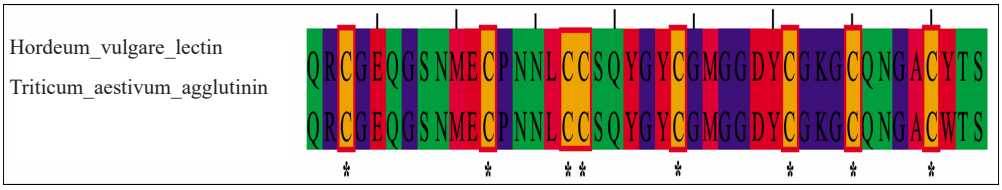


Figure 7. Comparison of carbohydrate-binding site motifs on plant lectin sequences with the chitin_bind_1 domain based on the similarity of motifs found in a previous study (Wright et al., 1991)

Note. The asterisk (*) symbol indicates conserved amino acids

These cysteine residues indicate that motifs in proteins containing the chitin_bind_1 domain are highly conserved. It is to be expected because the disulfide-rich domains of larger proteins, such as lectin and agglutinin, have a higher degree of sequence identity and are predicted to have descended from a common ancestor (Wright et al., 1991).

The B_lectin domain was found in two species that also possess the same carbohydrate binding site motif as the consensus of the B_lectin motif, according to a previous study by Zhao et al. (2017). B_lectin is known to have a conserved motif of QXDXNXVXY, where X represents any amino acid. Although the functionality of the motif is still in question, it confers the exclusive binding specificity of lectins containing the 5-amino acid motif (Pereira et al., 2015). This motif was detected in *G. nivalis* and *A. sativum* species, where it was found in the 56th through 64th and 88th through 96th amino acids of the aligned sequences with varying X amino acids (Figure 8).

The agglutinin domain is known to have no specific carbohydrate binding site. A study explained that the agglutinins in *Amaranthus* sp., more commonly known as amaranthin, assemble a binding site from a loop formed from two agglutinin subunits at the N- and C-terminals. This phenomenon is called a "head-to-tail arrangement", which will later form a binding site for carbohydrate binding. According to a comparison between the results and literature, *A. caudatus* and *A. hypochondriacus* are found to have the ability to assemble binding sites due to the presence of two agglutinin domain subunits in each lectin sequence of these two species, which allows the formation of a loop from a head-to-tail arrangement (Van Damme et al., 2008).

The analysis of the similarity or likeness of carbohydrate-binding site motifs in the 15 plant lectin protein sequences studied based on MSA revealed that 12 sequences matched the consensus of motifs found in the literature. The *G. max*, *P. sativum*, and *C. ensiformis* lectin sequences also exhibited a discrepancy between the consensus motif and the literature or incompleteness of carbohydrate-binding site motifs.

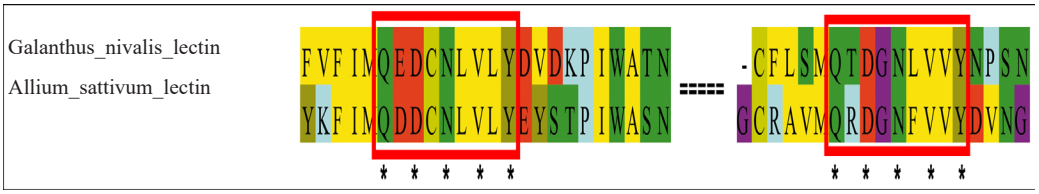


Figure 8. Comparison of carbohydrate-binding site motifs in plant lectin sequences with the B_lectin domain based on the similarity of motifs found in a previous study (Zhao et al., 2017). The conserved motif of B_lectin is "QXDXNXVXY"

Note. The asterisk (*) symbol indicates conserved amino acids; "X" indicates any amino acid

The Potential Roles of Plant Lectin

The five domains in the 15 plant lectin protein sequences have their respective functions or roles. The jacalin domain is a protein domain capable of binding monosaccharides and oligosaccharides with high specificity. At the beginning of its discovery, numerous jacalin domains were found in *Artocarpus* sp. (jackfruit), originating from the Moraceae family, but other studies have also found jacalin-related lectins in several plant families, such as the Asteraceae, Convolvulaceae, Musaceae, and Fagaceae families (Mann et al., 2001). Another study explained that the jacalin-related lectin domain plays roles in plant defense mechanisms, including disease resistance mechanisms, and is associated with abiotic stress signals (Esch & Schaffrath, 2017). In addition, previous studies have also shown that the jacalin-related lectin domain has a role in the transfer of proteins to the site of the pathogen attack by recognizing and binding to oligosaccharides that are typically present in pathogen infection processes (Esch & Schaffrath, 2017). Furthermore, other studies also explained that lectins derived from *M. acuminata* containing the jacalin domain could inhibit replication and block the entry of the HIV-1 virus (Swanson et al., 2010). Additionally, the modified lectin derived from *M. acuminata* (H84T) had inhibitory potential toward SARS-CoV-2 infection by blocking the binding of the SARS-CoV-2 spike protein to the ACE2 cellular receptor (Christodoulou et al., 2021). Lectins derived from *A. hypargyreus* are known to have the potential to act as immunomodulators by inducing T-lymphocyte activation (Zeng et al., 2019).

Chitin_bind_1, often found in the Poacea family, is a protein domain that recognizes and binds chitin subunits (type I). It also plays a role in plant defense mechanisms against pathogenic infections. A study showed broad-spectrum antifungal activity in plant chitinase proteins, which showed that the chitinase gene from *H. vulgare* could increase the fungal resistance of food crops (Kirubakaran & Sakthivel, 2007). In addition, results from other studies indicate the potential of lectins derived from *Triticum* sp. to be used as a therapeutic agent for leukemia (Ryva et al., 2019).

The next domain is B_lectin, which is a bulb-type mannose-specific lectin. The name "bulb-type lectin" corresponds to this type of lectin, which is often found in the bulb of a plant. This domain is found in several plant families, such as Amaryllidaceae and Aliaceae, and plays a role in the specific binding of mannose (Barre et al., 1997). B_lectin has the potential to be used as a natural insecticidal agent against specific receptors found on the insect midgut (Fitches et al., 2001).

The following domain is lectin_legume_LecRK_Arcelin_ConA, a combination of legume lectin, lectin-like receptor kinases, arcelin, concavalin A, and alpha-amylase. It is commonly known as legume lectin and is widely found in legumes in the Fabaceae family. This domain has a role in recognizing and binding specific glycoconjugates on cell surfaces and intracellular structures, where

they can serve as potential target molecules for development in agriculture, health, and pharmaceuticals (Lagarda-Diaz et al., 2017). In addition, legume lectins have shown potential to be developed as natural antifungal, antibacterial, and insecticidal agents for Coleoptera, Diptera, Lepidoptera, Hymenoptera, Isoptera, Neuroptera, and Homoptera, as well as possessing the ability to induce apoptosis in cancer cells (B. Liu et al., 2010; Gautam et al., 2018).

The last domain is agglutinin, a specific lectin that can agglutinate erythrocytes. Amaranthin has specific carbohydrate binding sites and can agglutinate A, B, and O red blood cells (Rinderle et al., 1989). In addition, agglutinins can inhibit the proliferation of tumor cells and are used as dyes for detecting human colon cancer through colonoscopy (Quiroga et al., 2015).

We then conducted a literature study concerning the production of lectin protein (yield) based on previous research. Table 5 lists the sources of plant lectin proteins in descending order of protein production and shows protein production and different organs that are the origin of lectins. Observation shows that the production of plant lectins through conventional methods varies widely from low (low yield) to high (high yield) categories. Plant lectin protein development can be carried out by synthetically utilizing its sequences as biological parts in producing plant lectin protein. Plant lectin protein sequences have the potential to be synthesized and further utilized in agriculture, health, and pharmacy by having information concerning the characteristics of each gene and the completeness of its carbohydrate binding site motifs.

The findings of this study show that several plant lectin sequences are potential candidates for utilization as biological parts by using the synthetic biology approach, especially in the coding sequence section. Some considerations for selecting the candidates include the completeness of the carbohydrate-binding site motif as well as the genetic characteristics such as CDS length (Table 2), which should be considered when producing lectin proteins through a synthetic biological approach because the longer the CDS, the more costly it will be compared to the shorter CDS. In addition, the completeness of the carbohydrate-binding site motif is also considered because it will affect the product proteins' ability to bind specific carbohydrates and their other roles as antifungal, antibacterial, antiviral, and insecticidal agents.

Candidate plant lectin sequences that can be utilized as biological parts for coding sequences include *M. acuminata*, *A. hypargyreus*, and *G. nivalis*. The first sequence is a lectin derived from *M. acuminata* with a CDS length of 426 bp, contains a jacalin domain, and has a complete carbohydrate binding site motif in the form of "GG" and "GXXXD" amino acids, according to a previous study (Raval et al., 2004). Banana lectin has shown potential as a candidate antiviral by inhibiting replication and blocking the entry of the HIV-1 virus (Swanson et al., 2010). In addition, previous studies

Table 5

Plant lectin sequences characteristics comparison

Species	Yield of lectin protein production (mg/g)	Protein domain	Completeness of carbohydrate-binding motifs	Organ of lectin origin
Amaranthus caudatus	1.600	Agglutinin	Two agglutinin subunits	Seed
Amaranthus hypochondriacus	1.280	Agglutinin	Two agglutinin subunits	Seed
Cicer arietinum	0.950	Legume lectin	Yes	Seed
Musa acuminata subsp. *malaccensis*	0.440	Jacalin	Yes	Fruit
Artocarpus hypargyreus Hance	0.400	Jacalin	Yes	Seed
Canavalia ensiformis	0.400	Legume lectin	No	Seed
Triticum aestivum L	0.300	Chitin_bind_1	Yes	Grain
Lens culinaris subsp. *tomentosus*	0.147	Legume lectin	Yes	Seed
Phaseolus vulgaris	0.130	Legume lectin	Yes	Seed
Hordeum vulgare var. Betzes	0.100	Chitin_bind_1	Yes	Grain
Galanthus nivalis L.	0.080	B_lectin	Yes	Tuber
Allium sativum L.	0.005	B_lectin	Yes	Tuber
Pisum sativum	0.005	Legume lectin	No	Seed
Robinia pseudoacacia	0.001	Legume lectin	Yes	Seed
Glycine max	0.001	Legume lectin	No	Seed

described the single amino acid-modified BanLec (H84T), which showed the potential for inhibiting SARS-CoV-2 infection by blocking the binding between the SARS-CoV-2 spike protein and the ACE2 cellular receptor (Christodoulou et al., 2021). Next is a lectin from *A. hypargyreus* with a CDS length of 453 bp and a jacalin domain with a complete carbohydrate binding site motif. Lectins from this species have potential as immunomodulators by inducing T-lymphocyte activation (Zeng et al., 2019).

The following candidate is a bulb-type lectin (B_lectin), a domain containing a lectin from *G. nivalis* with a CDS length of 471 bp. According to our reference

literature, this type of lectin has a complete carbohydrate binding site motif, which is "QXDXNXVXY", where X indicates any amino acid (Zhao et al., 2017). This lectin has shown potential as a natural insecticide with specific receptors in the insect midgut and exhibits antimetabolite activity in insects (Fitches et al., 2001).

CONCLUSION

This study compared the lectin gene lengths of 15 plant species. *Artocarpus hypargyreus* had the shortest lectin gene (480 bp), while *G. max* had the longest (1,009 bp). The length of the other plant lectin genes ranged from 495 to 912 bp. *Musa acuminata*, with three exons, had the most exons of any of the 15 lectin genes, while other plant lectin genes had 1-2 exons. The lectin proteins had five predicted domains: jacalin, chitin_bind_1, B_lectin, legume lectin, and agglutinin. Two protein sequences contained the jacalin domain; two contained the chitin_bind_1 domain; two contained the B lectin domain; and four contained the legume lectin domain. All domains' carbohydrate-binding site motifs were consistent with the consensus motifs found in the literature. The study results gave us gene, protein, and carbohydrate-binding motif comparison data from some lectin-producing plants, providing an understanding of future protein production, especially using a synthetic biology approach to obtain plant bioparts to develop more advanced lectin proteins.

ACKNOWLEDGEMENTS

The authors thank the Banana Research Group – Institut Teknologi Bandung (ITB), Indonesia, for engaging in discussions and providing technical support. This study was funded by the 2022 National Competitive Basic Research Grant from the Ministry of Education, Culture, Research, and Technology, Republic of Indonesia, for Fenny Dwivany (Grant Number: 083/E5/PG.02.00.PT/2022).

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., & Noble, W. S. (2009). MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Research*, *37*(Suppl_2), W202–W208. https://doi.org/10.1093/nar/gkp335

Barre, A., Van Damme, E. J. M., Peumans, W. J., & Rougé, P. (1997). Curculin, a sweet-tasting and taste-modifying protein, is a non-functional mannose-binding lectin. *Plant Molecular Biology*, *33*, 691–698. https://doi.org/10.1023/A:1005704616565

Butler, A. R., O'Donnell, R. W., Martin, V. J., Gooday, G. W., & Stark, M. J. (1991). Kluyveromyces lactis toxin has an essential chitinase activity. *European Journal of Biochemistry*, *199*(2), 483–488. https://doi.org/10.1111/j.1432-1033.1991.tb16147.x

Charungchitrak, S., Petsom, A., Sangvanich, P., & Karnchanatat, A. (2011). Antifungal and antibacterial activities of lectin from the seeds of *Archidendron jiringa* Nielsen. *Food Chemistry*, *126*(3), 1025–1032. https://doi.org/10.1016/j.foodchem.2010.11.114

Christodoulou, I., Rahnama, R., Ravich, J. W., Seo, J., Zolov, S. N., Marple, A. N., Markovitz, D. M., & Bonifant, C. L. (2021). Glycoprotein targeted CAR-NK cells for the treatment of SARS-CoV-2 infection. *Frontiers in Immunology*, *12*, 763460. https://doi.org/10.3389/fimmu.2021.763460

Coelho, L. C. B. B., dos Santos Silva, P. M., de Menezes Lima, V. L., Pontual, E. V., Paiva, P. M. G., Napoleão, T. H., & dos Santos Correia, M. T. (2017). Lectins, interconnecting proteins with biotechnological/pharmacological and therapeutic applications. *Evidence-Based Complementary and Alternative Medicine*, *2017*, 1594074. https://doi.org/10.1155/2017/1594074

Covés-Datson, E. M., Dyall, J., DeWald, L. E., King, S. R., Dube, D., Legendre, M., Nelson, E., Drews, K. C., Gross, R., Gerhardt, D. M., Torzewski, L., Postnikova, E., Liang, J. Y., Ban, B., Shetty, J., Hensley, L. E., Jahrling, P. B., Olinger, G. G., White, J. M., & Markovitz, D. M. (2019). Inhibition of Ebola virus by a molecularly engineered banana lectin. *PLOS Neglected Tropical Diseases*, *13*(7), e0007595. https://doi.org/10.1371/journal.pntd.0007595

Covés-Datson, E. M., King, S. R., Legendre, M., Gupta, A., Chan, S. M., Gitlin, E., Kulkarni, V. V., García, J. P., Smee, D. F., Lipka, E., Evans, S. E., Tarbet, E. B., Ono, A., & Markovitz, D. M. (2020). A molecularly engineered antiviral banana lectin inhibits fusion and is efficacious against influenza virus infection *in vivo*. *Proceedings of the National Academy of Sciences*, *117*(4), 2122–2132. https://doi.org/10.1073/pnas.1915152117

Covés-Datson, E. M., King, S. R., Legendre, M., Swanson, M. D., Gupta, A., Claes, S., Meagher, J. L., Boonen, A., Zhang, L., Kalveram, B., Raglow, Z., Freiberg, A. N., Prichard, M., Stuckey, J. A., Schols, D., & Markovitz, D. M. (2021). Targeted disruption of pi–pi stacking in Malaysian banana lectin reduces mitogenicity while preserving antiviral activity. *Scientific Reports*, *11*, 656. https://doi.org/10.1038/s41598-020-80577-7

Cummings, R. D., Marylinn, E. E., & Surolia, A. (2017). *Essentials of glycobiology* (3rd ed.). Cold Spring Harbor Laboratory Press.

Dang, L., & Van Damme, E. J. M. (2016). Genome-wide identification and domain organization of lectin domains in cucumber. *Plant Physiology and Biochemistry*, *108*, 165–176. https://doi.org/10.1016/j.plaphy.2016.07.009

Dwivany, F. M., Sukriandi, N., Meitha, K., & Brotosudarmo, T. H. P. (2021). *In silico* characterization of the structure of genes and proteins related to β-carotene degradation in *Musa acuminata* 'DH-Pahang' and *Musa balbisiana* 'Pisang Klutuk Wulung'. *Pertanika Journal of Tropical Agricultural Science*, *44*(2), 429–447. https://doi.org/10.47836/pjtas.44.2.10

Esch, L., & Schaffrath, U. (2017). An update on jacalin-like lectins and their role in plant defense. *International Journal of Molecular Sciences*, *18*(7), 1592. https://doi.org/10.3390/ijms18071592

Fitches, E., Woodhouse, S. D., Edwards, J. P., & Gatehouse, J. A. (2001). *In vitro* and *in vivo* binding of snowdrop (*Galanthus nivalis* agglutinin; GNA) and jackbean (*Canavalia ensiformis*; Con A) lectins within tomato moth (*Lacanobia oleracea*) larvae; mechanisms of insecticidal action. *Journal of Insect Physiology*, *47*(7), 777–787. https://doi.org/10.1016/S0022-1910(01)00068-3

Gautam, A. K., Gupta, N., Narvekar, D. T., Bhadkariya, R., & Bhagyawant, S. S. (2018). Characterization of chickpea (*Cicer arietinum* L.) lectin for biological activity. *Physiology and Molecular Biology of Plants*, *24*(3), 389–397. https://doi.org/10.1007/s12298-018-0508-5

Hessel, S. S., Dwivany, F. M., Zainuddin, I. M., Celik, I., Emran, T. bin, & Tallei, T. E. (2022). *Banana lectin as a potential candidate for anti-SARS CoV-2 by targeting the receptor-binding domain: A computational analysis* [Unpublished Master's thesis]. Bandung Institute of Technology.

Jiang, S. Y., Ma, Z., & Ramachandran, S. (2010). Evolutionary history and stress regulation of

the lectin superfamily in higher plants. *BMC Evolutionary Biology*, *10*, 79. https://doi.org/10.1186/1471-2148-10-79

Kamalesha, G., Dwivany, F. M., Nugrahapraja, H., & Rahma Putri, R. (2022). *In silico* comparisons of the Ethylene Response Factor 1 (ERF1) gene between Malaysian wild banana (*Musa acuminata ssp. malaccensis*) and Pisang Klutuk Wulung (*Musa balbisiana*). *Pertanika Journal of Tropical Agricultural Science*, *45*(2), 519–545. https://doi.org/10.47836/pjtas.45.2.12

Katoch, R., & Tripathi, A. (2021). Research advances and prospects of legume lectins. *Journal of Biosciences*, *46*, 104. https://doi.org/10.1007/s12038-021-00225-8

Kirubakaran, S. I., & Sakthivel, N. (2007). Cloning and overexpression of antifungal barley chitinase gene in *Escherichia coli*. *Protein Expression and Purification*, *52*(1), 159–166. https://doi.org/10.1016/j.pep.2006.08.012

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Lagarda-Diaz, I., Guzman-Partida, A., & Vazquez-Moreno, L. (2017). Legume lectins: Proteins with diverse applications. *International Journal of Molecular Sciences*, *18*(6), 1242. https://doi.org/10.3390/ijms18061242

Lam, S. K., & Ng, T. B. (2011). Lectins: Production and practical applications. *Applied Microbiology and Biotechnology*, *89*, 45–55. https://doi.org/10.1007/s00253-010-2892-9

Liu, B., Bian, H.-J., & Bao, J.-K. (2010). Plant lectins: Potential antineoplastic drugs from bench to clinic. *Cancer Letters*, *287*(1), 1–12. https://doi.org/10.1016/j.canlet.2009.05.013

Liu, W., Xie, Y., Ma, J., Luo, X., Nie, P., Zuo, Z., Lahrmann, U., Zhao, Q., Zheng, Y., Zhao, Y., Xue, Y., & Ren, J. (2015). IBS: An illustrator for the presentation and visualization of biological sequences. *Bioinformatics*, *31*(20), 3359–3361. https://doi.org/10.1093/bioinformatics/btv362

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., & Marchler-Bauer, A. (2020). CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Research*, *48*(D1), D265–D268. https://doi.org/10.1093/nar/gkz991

Mann, K., Farias, C. M. S. A., del Sol, F. G., Santos, C. F., Grangeiro, T. B., Nagano, C. S., Cavada, B. S., & Calvete, J. J. (2001). The amino-acid sequence of the glucose/mannose-specific lectin isolated from *Parkia platycephala* seeds reveals three tandemly arranged jacalin-related domains. *European Journal of Biochemistry*, *268*(16), 4414–4422. https://doi.org/10.1046/j.1432-1327.2001.02368.x

Meagher, J. L., Winter, H. C., Ezell, P., Goldstein, I. J., & Stuckey, J. A. (2005). Crystal structure of banana lectin reveals a novel second sugar binding site. *Glycobiology*, *15*(10), 1033–1042. https://doi.org/10.1093/glycob/cwi088

Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., Brown, S. D., Chang, H.-Y., El-Gebali, S., Fraser, M. I., Gough, J., Haft, D. R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., … Finn, R. D. (2019). InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research*, *47*(D1), D351–D360. https://doi.org/10.1093/nar/gky1100

Moraes Filho, R. M., Rossiter, J. G., Cavalcanti Junior, E. A., & Martins, L. S. S. (2017). *In silico* comparative analysis of legume lectins. *Journal of Genetics and Genomes*, *1*(1), 1000103.

Naithani, S., Komath, S. S., Nonomura, A., & Govindjee, G. (2021). Plant lectins and their many roles: Carbohydrate-binding and beyond. *Journal of Plant Physiology*, *266*, 153531. https://doi.org/10.1016/j.jplph.2021.153531

Pereira, P. R., Winter, H. C., Verícimo, M. A., Meagher, J. L., Stuckey, J. A., Goldstein, I. J., Paschoalin, V. M. F., & Silva, J. T. (2015).

Structural analysis and binding properties of isoforms of tarin, the GNA-related lectin from *Colocasia esculenta*. *Biochimica et Biophysica Acta - Proteins and Proteomics*, *1854*(1), 20–30. https://doi.org/10.1016/j.bbapap.2014.10.013

Peumans, W. J., & Van Damme, E. J. M. (1995). Lectins as plant defense proteins. *Plant Physiology*, *109*(2), 347–352. https://doi.org/10.1104/pp.109.2.347

Quiroga, A. V., Barrio, D. A., & Añón, M. C. (2015). Amaranth lectin presents potential antitumor properties. *LWT - Food Science and Technology*, *60*(1), 478–485. https://doi.org/10.1016/j.lwt.2014.07.035

Raval, S., Gowda, S. B., Singh, D. D., & Chandra, N. R. (2004). A database analysis of jacalin-like lectins: Sequence-structure-function relationships. *Glycobiology*, *14*(12), 1247–1263. https://doi.org/10.1093/glycob/cwh140

Rinderle, S. J., Goldstein, I. J., Matta, K. L., & Ratcliffe, R. M. (1989). Isolation and characterization of amaranthin, a lectin present in the seeds of *Amaranthus caudatus*, that recognizes the T- (or cryptic T)-antigen. *Journal of Biological Chemistry*, *264*(27), 16123–16131. https://doi.org/10.1016/S0021-9258(18)71595-0

Ryva, B., Zhang, K., Asthana, A., Wong, D., Vicioso, Y., & Parameswaran, R. (2019). Wheat germ agglutinin as a potential therapeutic agent for leukemia. *Frontiers in Oncology*, *9,* 100. https://doi.org/10.3389/fonc.2019.00100

Scott, D., Abdelhakim, D., Miranda, M., Höft, R., & Cooper, H. D. (2015). *Potential positive and negative impacts of components, organisms and products resulting from synthetic biology techniques on the conservation and sustainable use of biodiversity and associated social, economic and cultural considerations.* https://www.cbd.int/doc/meetings/cop/cop-12/information/cop-12-inf-11-en.pdf

Singh, S. S., Devi, S. K., & Ng, T. B. (2014). Banana lectin: A brief review. *Molecules*, *19*(11), 18817–18827. https://doi.org/10.3390/molecules191118817

Song, M., Xu, W., Xiang, Y., Jia, H., Zhang, L., & Ma, Z. (2014). Association of jacalin-related lectins with wheat responses to stresses revealed by transcriptional profiling. *Plant Molecular Biology*, *84*, 95–110. https://doi.org/10.1007/s11103-013-0121-5

Srinivas, V. R., Acharya, S., Rawat, S., Sharma, V., & Surolia, A. (2000). The primary structure of the acidic lectin from winged bean (*Psophocarpus tetragonolobus*): Insights in carbohydrate recognition, adenine binding and quaternary association. *FEBS Letters*, *474*(1), 76–82. https://doi.org/10.1016/S0014-5793(00)01580-5

Swanson, M. D., Winter, H. C., Goldstein, I. J., & Markovitz, D. M. (2010). A lectin isolated from bananas is a potent inhibitor of HIV replication. *Journal of Biological Chemistry*, *285*(12), 8646–8655. https://doi.org/10.1074/jbc.M109.034926

Tsaneva, M., & Van Damme, E. J. M. (2020). 130 years of plant lectin research. *Glycoconjugate Journal*, *37*, 533–551. https://doi.org/10.1007/s10719-020-09942-y

Van Damme, E. J. M. (2014). History of plant lectin research. In J. Hirabayashi (Ed.), *Lectins. Methods in molecular biology* (Vol. 1200, pp. 3–13). Humana Press. https://doi.org/10.1007/978-1-4939-1292-6_1

Van Damme, E. J. M. (2022). 35 years in plant lectin research: A journey from basic science to applications in agriculture and medicine. *Glycoconjugate Journal*, *39*, 83–97. https://doi.org/10.1007/s10719-021-10015-x

Van Damme, E. J. M., Lannoo, N., & Peumans, W. J. (2008). Plant lectins. *Advances in Botanical Research*, *48*, 107–209. https://doi.org/10.1016/S0065-2296(08)00403-5

Van Holle, S., de Schutter, K., Eggermont, L., Tsaneva, M., Dang, L., & Van Damme, E. (2017). Comparative study of lectin domains in model species: New insights into evolutionary dynamics. *International Journal of Molecular Sciences*, *18*(6), 1136. https://doi.org/10.3390/ijms18061136

Van Holle, S., & Van Damme, E. J. M. (2019). Messages from the past: New insights in plant lectin evolution. *Frontiers in Plant Science*, *10*, 36. https://doi.org/10.3389/fpls.2019.00036

Vandenborre, G., Smagghe, G., & Van Damme, E. J. M. (2011). Plant lectins as defense proteins against phytophagous insects. *Phytochemistry*, *72*(13), 1538–1550. https://doi.org/10.1016/j.phytochem.2011.02.024

Wang, W., Li, Q., Wu, J., Hu, Y., Wu, G., Yu, C., Xu, K., Liu, X., Wang, Q., Huang, W., Wang, L., & Wang, Y. (2021). Lentil lectin derived from *Lens culinaris* exhibit broad antiviral activities against SARS-CoV-2 variants. *Emerging Microbes and Infections*, *10*(1), 1519–1529. https://doi.org/10.1080/22221751.2021.1957720

Wright, H. T., Sandrasegaram, G., & Wright, C. S. (1991). Evolution of a family of N-acetylglucosamine binding proteins containing the disulfide-rich domain of wheat germ agglutinin. *Journal of Molecular Evolution*, *33*(3), 283–294. https://doi.org/10.1007/BF02100680

Xiong, J. (2006). *Essential bioinformatics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511806087

Zeng, Q., Lin, F., Zeng, L., Deng, Y., & Li, L. (2019). Purification and characterization of a novel immunomodulatory lectin from *Artocarpus hypargyreus* Hance. *International Immunopharmacology*, *71*, 285–294. https://doi.org/10.1016/j.intimp.2019.03.027

Zhao, Y., Jian, Y., Liu, Z., Liu, H., Liu, Q., Chen, C., Li, Z., Wang, L., Huang, H. H., & Zeng, C. (2017). Network analysis reveals the recognition mechanism for dimer formation of bulb-type lectins. *Scientific Reports*, *7*, 2876. https://doi.org/10.1038/s41598-017-03003-5