GMR

The Original by FUNPEC-RP

In silico characterization and phylogenetic analysis of a mannose-specific lectin in *Allium* species

H.J. Jimenez[1,3], L.S.S. Martins[2,3], R. Carvalho[2,3], A.V.V. Montarroyos[1] and R.M. Moraes Filho[1,3]

[1] Departamento de Agronomia, Universidade Federal Rural de Pernambuco, Recife, PE, Brasil
[2] Departamento de Biologia, Universidade Federal Rural de Pernambuco, Recife, PE, Brasil
[3] Programa de Pós Graduação em Melhoramento Genético de Plantas, Universidade Federal Rural de Pernambuco, Recife, PE, Brasil

Corresponding author: R.M. Moraes Filho
E-mail: romulommfilho@yahoo.com.br

Genet. Mol. Res. 18 (2): gmr18187
Received October 26, 2019
Accepted June 12, 2019
Published June 18, 2019
DOI http://dx.doi.org/10.4238/gmr18187

ABSTRACT. The *Allium* genus stands out for its uses in human food and also for its medicinal properties. Many representatives of the Amaryllidaceae family are known for producing mannose binding lectins (MBL). In plants, lectins act as reserves of proteins that can be used for plant growth and development and also in defense against herbivores and pathogens, being toxic to some aphids and sucking insects. We examined physicochemical characteristics, such as isoelectric points and hydropathicity, of 22 sequences of MBL protein from *Allium* species and from other representatives of the Amaryllidaceae family present in public databases. Phylogenetic analysis, identification of functional domains and 3D homology modeling were also performed. We found two conserved functional motifs in the MBL sequences. It was observed that for all species the MBL had a hydrophilic character and great variation in isoelectric points. The phylogenetic analysis was not consistent with the taxonomic classification of the species evaluated at the infrageneric level. However, the methods proved efficient for the separation up to

the level of tribes within the Amaryllidaceae family. The generated 3D models also provide a better understanding of their tertiary structures and molecular functions.

Key words: Bioinformatics; Homology modeling; Conserved domains

INTRODUCTION

The Allium genus is the most numerous representative of the Allioideae subfamily; it contains more than 800 species, being thus one of the largest genera among the monocotyledons. Species of this genus are present throughout the world, ranging from the Arctic Circle to Europe, Asia, North America and Africa (Rahn, 1998; Fritsch et al., 2010). The main *Allium* genus center of origin extends from the Mediterranean basin to Central Asia and a less important one is located in western North America (Li et al., 2010). The basic chromosome number of the group is n = 8; however, variations in ploidy (2n = 14-68) and other basic numbers (n = 7, 9, 10 and 11) may occur (Fritsch et al., 1998; Zhou et al., 2007).

The species of this genus are widely used in human food and condiments, especially *A. cepa*, *A. sativum*, *A. fistulosum*, *A. schoenoprasum* and *A. ampeloprasum* (Souza, 2012). Some representatives of the genus have anti-inflammatory, vermifuge, antiseptic, diuretic, hypotensive and antipyretic properties, and produce allicin and garlicine, which confer antibacterial properties (Harris et al., 2001; Benkeblia, 2003). Many representatives of the Allioidae subfamily stand out because they contain mannose-binding lectins (Smeets et al., 1997). Lectins are proteins or glycoproteins that contain at least one carbohydrate binding site without demonstrating catalytic utility or immunological characteristics (Sharon et al., 2004), where they bind reversibly to eukaryotic glycoconjugate mono- and oligosaccharides (Vasconcelos et al., 2009). The first report about a lectin was from *Ricinus communis* in 1888, in a study about the toxic effects of the extracts of this plant (Franz, 1988; Sharon et al., 2004).

Regarding the three-dimensional structure, plant lectins have been subdivided into: only one carbohydrate binding domain (merolectin), two or more identical binding domains (hololectins), at least two different carbohydrate binding domains (superlectin), one or more carbohydrate binding domains and a domain that exerts biological activity independent of the carbohydrate binding domain (chimerolectin) (Van Damme et al., 1998; Vandenborre et al., 2011). There are also those that have two or more identical sugar binding sites (multilectins). Since their discovery, lectins have shown a great number of chemical and biological properties, which has allowed their use for structural and functional investigation of carbohydrates, especially glycoproteins, and to examine variations that occur on the cell surface (Sharon et al., 2004). They can be observed in microorganisms, plants and animals, being part of the plasma membrane or inside the cell.

Endogenous lectins mediate biological processes such as cell-cell recognition, extracellular matrix interactions, gametic fertilization, embryonic development, cell growth, cell differentiation, cell signaling, cell adhesion and migration, apoptosis, immunomodulation and inflammation, parasite-host interaction, folding and targeting of glycoprotein, mitogenic induction and homeostasis (Ghazarian et al., 2011), and has attracted great interest due to its variety of biological activities such as cellular agglutination

(Khan et al., 2007), antitumoral (Liu et al., 2012), immunomodulatory (Rubinstein et al., 2004), antifungal (Herre et al., 2004) and antiviral effects (Zuo et al., 2012). In recent years, it has become clear that lectins play two major roles in plants. First, they are deposits of proteins that can be boosted for the growth and development of plants, and secondly, in the defense of plants against herbivores and pathogens (Barre et al., 2002; Vandenborre et al., 2011; Dias et al., 2015; Troegeler et al., 2017).

Lectins usually occur more abundantly in seeds and storage tissues, where they are seen in subcellular organelles with storage function. Lectins accumulate during the growth and development or reproductive phase of the plant life cycle and are mobilized and used later (Peumans et al., 1995; Barre et al., 2002; Follmer et al., 2004; Vandenborre et al., 2011; Dias et al., 2015). Genes encoding proteins have some regions that are very well conserved, because of their structure or molecular functions, while other regions evolve faster in terms of nucleotide substitutions and insertions or deletions (Watson et al., 2005). To understand the role of proteins within groups of organisms, phylogenetic studies can help to clarify questions about how proteins are related in different species, and whether they may have evolved from a common ancestor (Kasap et al., 2010; Andrade et al., 2011). Phylogenetic results are currently continuously improving due to the increasing availability of a large amount of biological data and new approaches and methods of analysis (Kasap et al., 2010; Andrade et al., 2011).

The objectives of this study were: (1) Characterize, compare and identify conserved domains in amino acid sequences of a mannose binding lectin from different *Allium* species available in public databases. (2) Perform a phylogenetic analysis based on the sequences found, to understand the relationships between species of the genus, as well as at the Amaryllidaceae family level. (3) Develop three-dimensional models of representatives of the genus, based on the homology modeling methodology, in order to allow a better understanding of their structures and molecular functions. This comparative analysis provides useful theoretical information for studies involving these proteins, considering their effective potential for use in phylogenetic studies and studies on plant resistance to insect action.

MATERIAL AND METHODS

Sequences retrieval

A mannose binding lectin (MBL) from *A. sativum* was retrieved from the NCBI (National Center for Biotechnology Information) platform. Homologous sequences from other species were obtained with the BLAST (Basic Local Alignment Search Tool) search algorithm on the NCBI, where a total of 22 sequences were found (Table 1).

Sequence Analysis

Physicochemical parameters of MBLs present in species of the Amaryllidaceae family were analyzed by the ProtParam server (http://web.expasy.org/protparam) (Gasteiger et al., 2005). The presence of signal peptide cleavage sites was investigated using the TOPCONS server (http://topcons.cbr.su.se/pred) (Tsirigos et al., 2015). The identification of the functional domains of the protein, its classification and ontology were performed

using the Prodom server (http://prodom.prabi.fr/prodom/) (Servant et al., 2002), which is a database of families of protein domains of homologous segments. Estimates of the functional effects caused by mutations of amino acid sequences were predicted by the SNAP2 server (https://rostlab.org/services/snap2web/) (Hecht et al., 2015).

Alignment and Phylogenetic Analysis

Protein sequences were aligned using the ClustalW algorithm and phylogenetic trees were generated in MEGA software 7.0.21 (Kumar et al., 2016). Phylogenetic trees were constructed using the Maximum parsimony (MP) method with a bootstrap test with 1000 replicates.

Prediction, Evaluation and Validation of the Tertiary Structure Models

The tertiary structure prediction of the MBLs was performed by the Phyre2 server (http://www.sbg.bio.Ic.ac.uk/phyre2) (Kelley et al., 2015) in multi-template mode. This server uses advanced methods of homology detection to build a 3D model. The 3D structure of the protein was then visualized by the UCSF Chimera (Pettersen et al., 2004) package. The quality of the model was evaluated using the Molprobity server (http://molprobity.biochem.duke.edu/) (Chen et al., 2010) by Ramachandran analysis. The Z-score was calculated using the ProSA-web interactive server (https://prosa.services.come.sbg.ac.at/prosa.php) to recognize errors in three-dimensional structures (Wiederstein et al., 2007). Minimization of energy and correction of small errors in the three-dimensional model were performed by the Yasara force Field (Krieger et al., 2009) server and KiNG (Chen et al., 2009) software.

RESULTS

Since the initial studies with MBLs in monocotyledoneous species, these proteins were also identified in several species of the Amaryllidaceae family (Van Damme et al., 1991; Mo et al., 1993; Saito et al., 1997), which made it clear that MBLs were not exclusive of a single family of monocotyledons, but were rather widespread within the group (Van Damme et al., 1991; Smeets et al., 1997).

In this study, we analyzed sequences of MBLs in several species of the genus *Allium*, from protein sequences retrieved from the NCBI database. The physicochemical properties of the protein sequences were analyzed by the ProtParam server. Table 1 shows the molecular weight, the theoretical isoelectric point (pI) and the Grand mean of hydropathicity (GRAVY) of the protein sequences evaluated.

The isoelectric point (pI) is important for solubility, subcellular localization and interaction of a protein. At pHs different from the isoelectric point, the protein solubility increases due to the appearance of positive or negative charges on the protein chains, which favors the charge-moment interaction of the water dipole. At a time when a protein solution is in its isoelectric state, that is, when the protein manifests zero net charge in an aqueous system, protein-protein interactions increase, so less water interacts with protein molecules, which that their molecules approach, aggregate, and precipitate. The knowledge of the pI of a protein is of great importance, since, based on it, one can predict the net charge of the protein at a certain pH, which will be very important for the practice of experimental studies

as in the process of protein purification and maintenance of the same in a solution that favors its maintenance in the soluble form (Vojdani, 1996; Carneiro, 1997).

Table1. Analysis of the primary structure and subcellular localization of mannose binding lectins in *Allium* species.

Species	GB-ID	SPCS	No. aa	MW (kDa)	pI	GRAVY	CN
A. ampeloprasum	AAC37361.1	30-31	180/150	19.23/16.16	9.07/8.62	-0.056/-0.136	16
A. angulosum	AMZ79750.1	29-30	177/148	19.17/16.18	6.27/5.43	-0.005/-0.221	8
A. albidum	AMZ79752.1	30-31	176/147	18.75/15.76	7.61/6.44	-0.003/-0.2333	8
A. obliquum	AMZ79761.1	29-30	180/151	19.45/16.45	5.42/5.07	-0.055/-0.154	8
A. oreoprasum	AMZ79751.1	29-30	177/148	19.09/16.10	5.11/4.87	-0.014/-0.210	8
A. consanguineum	AMZ79762.1	29-30	180/151	19.46/16.48	6.25/5.41	-0.063/-0.147	8
A. senescens	AMZ79757.1	29-30	179/150	19.43/16.45	7.61/6.44	-0.041/-0.175	16
A. ramosum	AMZ79755.1	29-30	173/144	18.52/15.54	4.85/4.65	-0.092/-0.122	16
A. roylei	AMZ79763.1	30-31	178/149	19.18/16.20	8.73/8.64	-0.061/-0.153	8
A. sativum	AAB64237.1	30-31	181/151	19.29/16.22	8.84/6.40	-0.065/-0.090	8
A. scorodoprasum	AMZ79756.1	29-30	176/147	18.86/15.88	7.61/6.44	-0.008/-0.238	8
A. fistulosum	BAP75574.1	29-30	177/149	18.88/15.99	7.65/7.94	-0.092/-0.039	8
A. altaicum	ADN26577.1	29-30	177/149	18.94/16.12	9.05/8.96	-0.134/-0.028	8
A. x proliferum	AMZ79754.1	29-30	178/149	19.10/16.12	8.72/8.62	-0.025/-0.195	8
A. pskemense	AMZ79764.1	29-30	188/159	19.80/16.82	6.04/5.73	-0.070/-0.128	8
A. hookeri	AMZ79753.1	29-30	177/148	19.04/16.06	4.92/4.73	-0.012/-0.241	7
A. carolinianum	AMZ79759.1	29-30	179/150	19.33/16.29	6.24/6.44	-0.046/-0.149	16
A. triquetrum	ABA00714.1	29-28	173/145	18.36/15.51	7.56/6.41	-0.104/-0.189	9
A. tuberosum	AMZ79760.1	29-30	176/147	18.94/15.96	7.61/5.36	-0.055/-0.295	16
A. ursinum	AAC49858.1	28-29	166/138	17.67/14.83	5.01/4.79	-0.092/-0.199	7
*A. ascalonicum**	AAC37360.1	27-26	177/151	18.83/16.24	9.22/9.17	0.069/-0.108	8
*A. cepa**	AIS22686.1	3-2	146/144	15.71/15.51	5.29/5.3	-0.099/-0.126	8

GB-ID: Genbank Identification, SCSP: Signal Peptide Conserved Sites, No. aa: Number of amino acids CN: Chromosomic number (n). GRAVY: Grand Average of Hydropathicity: pI: Isoelectric point. MW: Molecular Weight. *: Partial sequence.

There is a correlation between the subcellular localization and the pI of a protein (Andrade et al., 1998; Nandi et al., 2006). Usually proteins in the cytoplasm have pI < 7.4, while those in the nucleus have a more alkaline pI (7.4 < pI < 8.1) (Andrade et al., 1998; Nandi et al., 2006). It has also been shown that pI can vary widely, depending on the insertions and deletions between orthologs, and the ecology of the organism (Kiraga et al., 2007).

The isoelectric point of the sequences was between 4.65 (*A. ramosum*) and 8.96 (*A. altaicum*), showing that this lectin exhibits great variation of chemical properties between species analyzed. The mean hydrophobicity indicates solubility, with positive values for hydrophobic and negative for hydrophilic proteins. Sequences ranged in size from 138 (*A. ursinum*) to 159 (*A. pskemense*) amino acids. The mean hydrophobicity ranged from -0.295 (*A. tuberosum*) to -0.028 (*A. altaicum*), indicating that this protein is hydrophilic in character for all species. The molecular weight varied from 14.83 kDa (*A. ursinum*) to 16.82kDa (*A. pskemense*). Molecular weight is employed in chromatography and electrophoresis, for example, to isolate proteins. However, it is not possible to generalize about it in relation to protein functions (Nelson et al., 2005).

The multiple sequence alignment is shown in Figure 1. The contoured region indicates location of the signal peptide. In the observed alignment, it was possible to verify

regions of high similarity, where most of the positions presented conservation among the sequences analyzed.

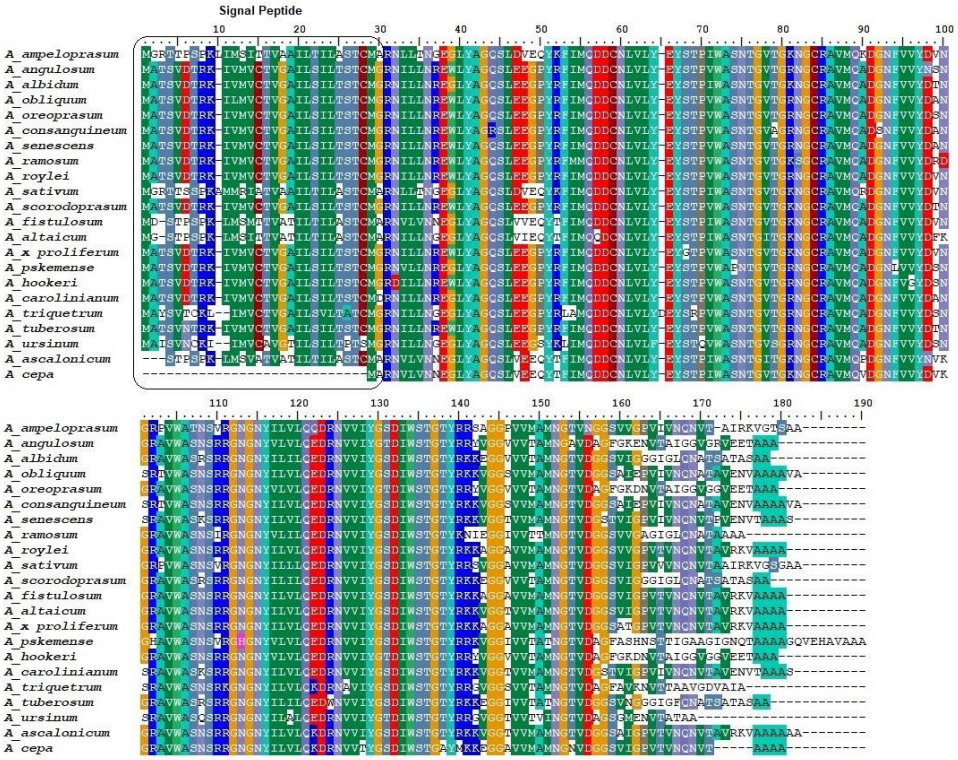

Figure 1. Alignment of mannose binding lectin sequences from species of the *Allium* genus. Sequences were aligned by ClustalW; identical and similar residues are displayed in the same color.

For the functional characterization of MBLs and prediction of the functional effects of mutations, five species were selected (Figure 2). Two functional domains were identified by the PRODOM server: PD330654 and PD585253 in the amino acid sequences of the mannose-specific lectin. The PD330654 functional domain was found in all observed sequences and is associated with mannose binding, actin monomer binding and pollen recognition functions, being an integral component of the membrane which indicates its relevance for protein activity. No information was found on the activities of domain PD585253 in the server. The SNAP2 server has shown that the functional domains PD330654 and PD585253 are the regions of the protein most sensitive to mutations. This server provided a map with the presence of different shades with possible substitution at each position of these proteins. The red color indicates strong signs for mutation, white coloration shows low effect indicator and blue color indicates neutrality. The occurrence of a greater amount of mutation-sensitive sites in the PD330654 domain explains its conservation and occurrence in all sequences analyzed.

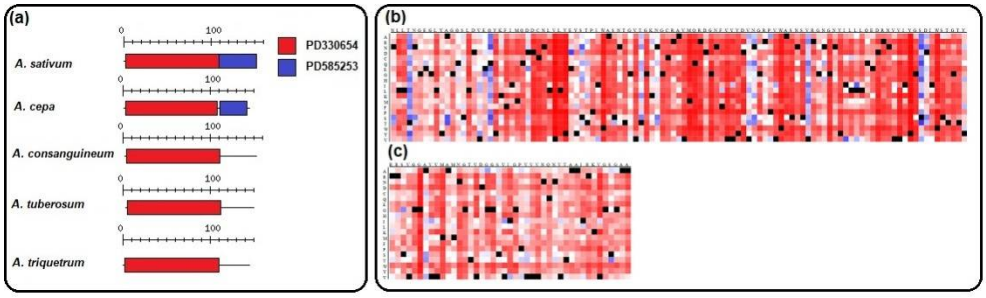

Figure 2. Prediction of the functional domains of the mannose binding lectins observed by the ProDom server (a) and prediction of functional effects of the mutations in amino acids by the SNAP2 server for domains PD330654 (b) and PD585253(c) of *Allium sativum*.

In the phylogenetic analyzes, the *Allium* genus was shown to be monophyletic and robustly separated from the outgroup (Figure 3).

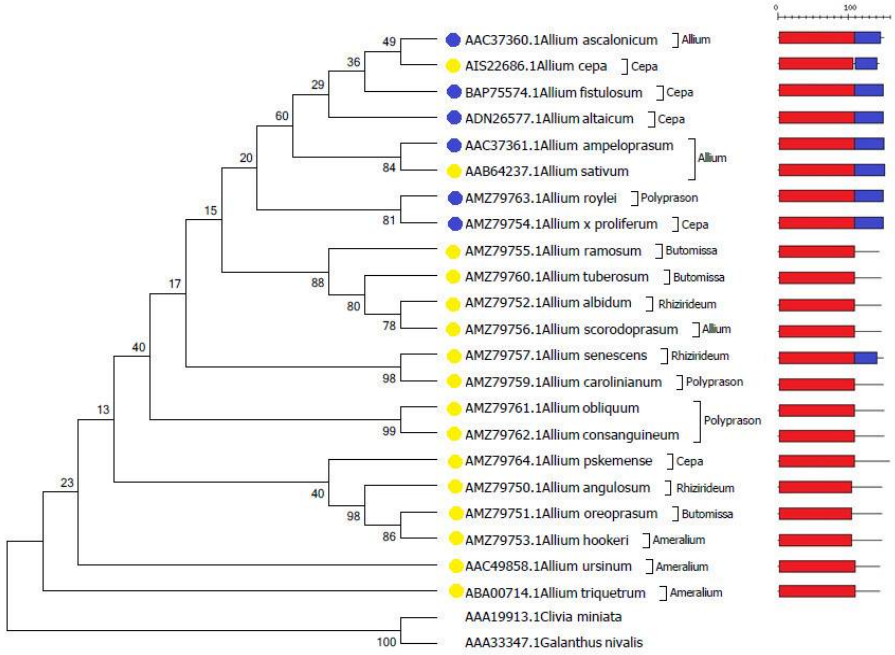

Figure 3. Phylogenetic tree of *Allium* species generated by the maximum parsimony (MP) method. Yellow circles indicate acidic mannose binding lectins (MBLs) and blue circles indicate alkaline MBLs. The species are represented by their respective subgenres. The red-colored rectangular figure indicates the PD330654 functional domain and the blue-colored rectangular figure indicates the PD585253 functional domain.

Several clades were found in *Allium*, comprising species of subgenus *Cepa*, *Allium*, *Polyprason*, *Butomissa*, *Rhizirideum* (representatives of the 3rd evolutionary line of the genus) and *Amerallium* (representative of the 1st evolutionary lineage of the genus). In the *Amerallium* clade, it was possible to observe that the species A. *triquetrum* (Section Briseis) and *A. ursinum* (Section Arctoprasum) presented an isolated group in the phylogenetic tree

(BS = 23%), being close also of the species *A. hookeri* (Section Bromatorrhiza). These results are similar to the results found by means of ITS markers for this genus (Li et al., 2010), where the authors find a close relationship between species of Sections Briseis (*A. triquetrum*), Arctoprasum (*A. ursinum*) and Bromatorrhiza (*A. hookeri*). The *Amerallium* clade species used in this study represent two relatively small groups in the Mediterranean and East Asia, thus belonging to the Old World clade. However, *Amerallium* is one of the major subgenres in the evolutionary line of *Allium*, being extremely morphologically and ecologically diversified (Li et al., 2010). *Amerallium* encompasses three distinct geographic groups: one consisting of all native *Allium* species of North America and the remainder encompassing two smaller groups in the Mediterranean region and East Asia (Li et al., 2010).

The *Butomissa* clade was comprised of the species *A. tuberosum*, *A. ramosum* and *A. oreoprasum*, and it was possible to observe that the species *A. tuberosum* and *A. ramosum* were grouped together with *A. scorodoprasum* and *A. albidum*, that belong to the subgenus *Allium* and *Rhizirideum* respectively. The grouping of species of the *Allium* and *Butomissa* subgenus was also observed in previous analyzes (Li et al., 2010), where the authors indicate that species of the subgenus Allium (*A. heldreichii*) may present close phylogenetic relationships with *A. tubero*sum (BS = 61%). The author suggests that additional sequences of multiple markers would be important to test this relationship. Our results indicated that species of the *Butomissa* clade showed proximity in the phylogenetic tree with representatives of the clade *Rhizirideum*. *Butomissa* occupies a position between nearby representatives of the subgenus *Rhizirideum* (Li et al., 2010) and it was observed that the growth form and the morphology of the chromosomes of species of the *Butomissa* clade are similar to those of the clade *Rhizirideum* (*A. angulosum*, *A. albidum* and *A. senescens*) (Kruse et al., 1992; Friesen et al., 1998).

The *Rhizirideum* clade covered the species *A. angulosum*, *A. albidum* and *A. senescens*, which formed grouping in the phylogenetic tree. Analyzes with ITS markers show that the section *Rhizirideum* forms polytomies and their relation is in addition to the resolution by these markers, thus presenting problems as to their phylogeny (Li et al., 2010). Small stages of time between occurrences of speciation would promptly clarify the polytomies visualized in the section *Rhizirideum*, and the occurrence of the polyploid complex in the *A. senescens* group could be related to the recent origin of these species (He et al., 1999). Our results showed that species of the subgenus *Allium*, *Cepa* and *Polyprason* formed monofiletism along the phylogenetic tree. However previous observations suggest that these subgenera are not monophyletic, thus suggesting reconsidering the systematic position of some species (Li et al., 2010).

Protein-based phylogenetic reconstructions have been widely used to elucidate the role of proteins within clusters of organisms. Phylogenetic studies may help to clarify questions about how proteins are related in different species, and whether they may have evolved from a common ancestor (Kasap et al., 2010; Andrade et al., 2011; Van Holle et al., 2017; Moraes Filho et al., 2016). Phylogenies based on some lectin families demonstrate similarity with the phylogeny of angiosperms (Van Holle et al., 2017). Figure 4 illustrates the phylogenetic reconstruction of Amaryllidaceae with the inclusion of species of the subfamily Amaryllidoideae, with representatives of the tribes Amaryllideae, Hippeastreae, Galantheae, Narcisseae, Haemantheae and Lycorideae. The phylogeny obtained is in accordance with the botanical classification APG-IV, indicating that the use of lectin

sequences in question is suitable for the separation of species up to the tribe level, but not at the infrageneric level. Our results showed that the use of MBLs as phylogenetic markers did not offer enough resolution to resolve the relationships within the *Allium* genus and that the sequences of this protein are not the best model for this kind of analysis.

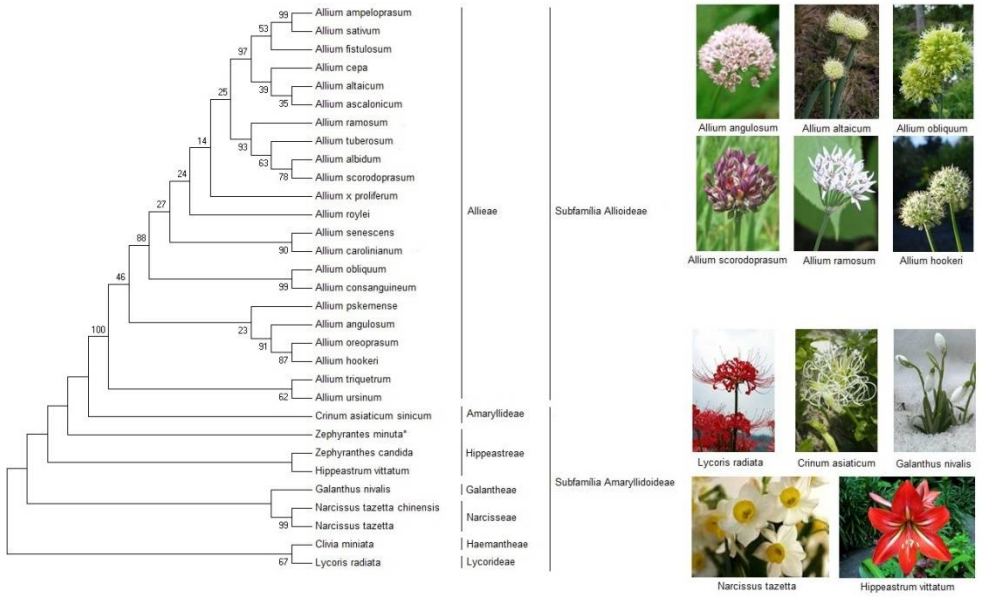

Figure 4. Phylogenetic tree of the species of the Amaryllidaceae family generated with the maximum parsimony (MP) method. The tribes of the subfamilies Allioideae and Amaryllioidae are indicated to the left of the representative bars of the subfamilies. The species included were: *Crinum asiaticum* (GenBank: AAA534507.1), *Zephyranthes minuta* (GenBank: AAN73327.1), *Zephyranthes candida* (GenBank: AAM27447.1), Hippeastrum vittatum (GenBank: AAP57409.1), *Galanthus nivalis* (GenBank: AAA33346.1), *Narcissus tazetta* (GenBank: ACR15122.1), *Narcissus tazetta var. chinensis* (GenBank: ADN05761.1), *Clivia miniata* (GenBank: AAA19913.1) and *Lycoris radiata* (GenBank: AAP20877.1). Figures obtained at https://pfaf.org/. * The species *Zephyrantes minuta* is identified in GenBank with the synonymy *Amaryllis minuta*.

For the prediction of the tertiary structure of the proteins by the Phyre2 server, the lectin sequences of *A. sativum* and *A. ursinum*, representative of two evolutionary lines of the genus *Allium*, were selected (Figure 5).

After refinement of the 3D structures by the KiNG software and the Yasara Force Field Server, the models were evaluated by the MOLPROBITY server for error recognition in the tertiary structures. The refined models presented 93.29% and 94.12% of amino acid residues in the favored regions by the Ramachandran plot analysis for the species *A. sativum* and *A. ursinum* respectively. The Z-score values evaluated by the ProSa server were -4.94 and -5.36 for *A. sativum* and *A. ursinum* respectively (Figure 6). Z-scores of protein structures are widely used because they characterize the conformational energy of proteins, and indicate overall model quality. A negative Z-score value is expected for a good model (Zhang et al., 1998), but more importantly, the black dot represented in the graph is within the blue region.

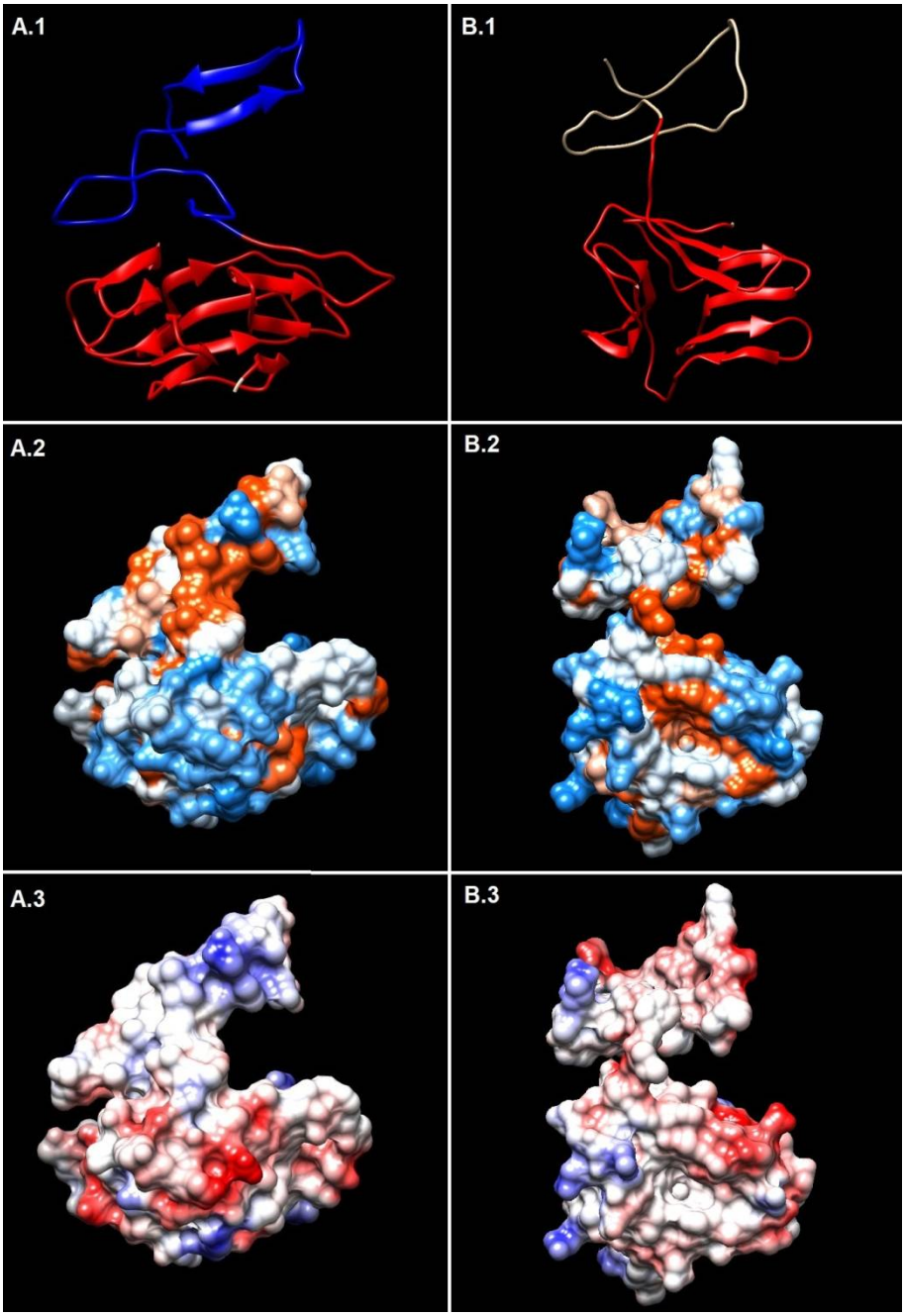

Figure 5. Three-dimensional structure predicted by the Phyre2 server for mannose-specific lectin of *Allium sativum* (A1) and *Allium ursinum* species. (A2). Hydrophobicity represented as a color gradient, the blue being the most hydrophilic and reddish orange for the most hydrophobic species *Allium sativum* and (B2) *Allium ursinum* (B1). Electrostatic surface represented as a color gradient, from domains negatively charged (red) to the most positively charged (blue) of the species *A. sativum* (A3) and *A. ursinum* (B3).

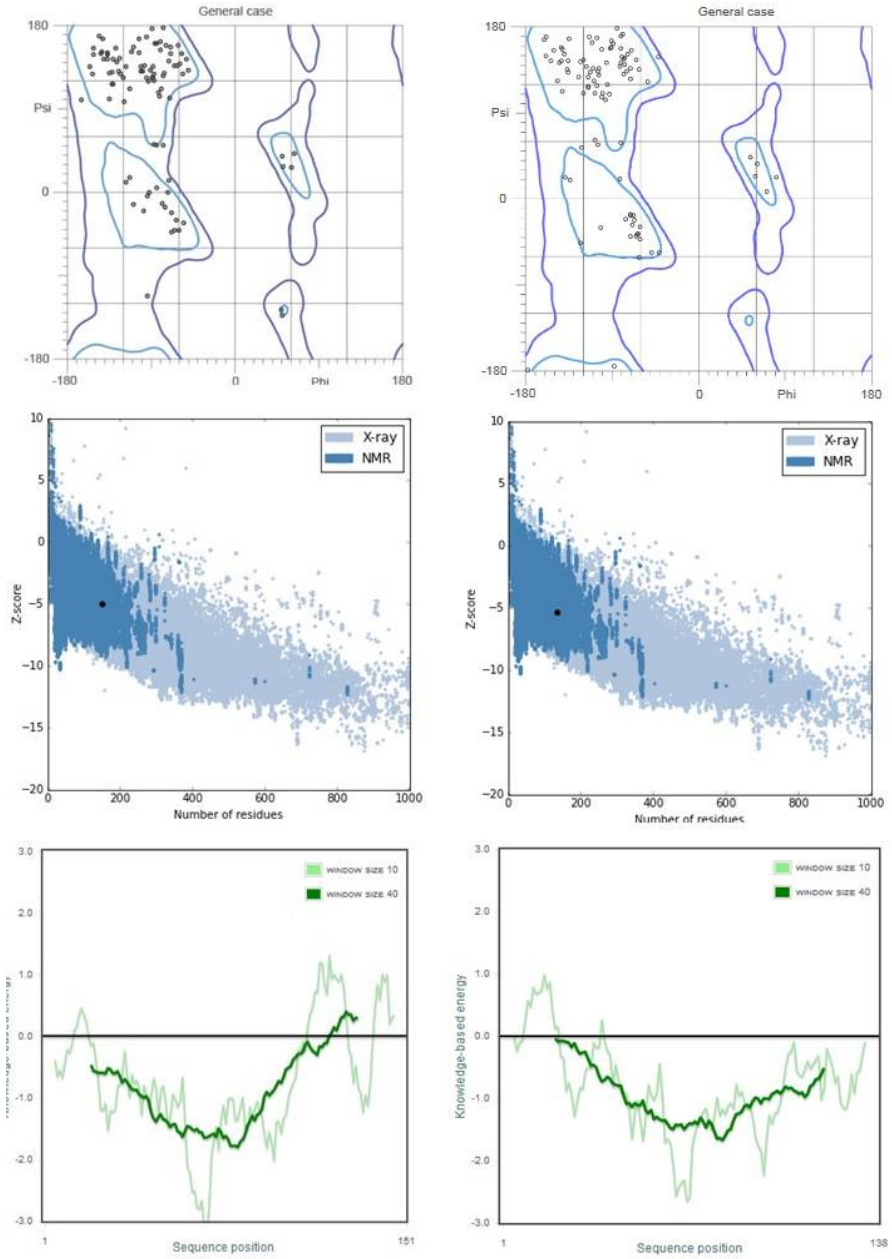

Figure 6. Ramachandran Plot Analysis generated by the MolProbity server, from the three-dimensional model built by the server Phyre2 for *A. sativum* (A1). E *Allium ursinum* (B1). Z-score value (black dot) by ProSA-web for *Allium. sativum* (A2) and *Allium ursinum* (B2). Use of ProSA-web for mannose-specific lectin showing the energy scatter chart of residues from the native protein structure for *Allium sativum* (A3) and *Allium ursinum* (B3).

MBLs plays an important role in the defense of plants against herbivores and pathogens, (Murdock et al., 2002) being toxic to some aphids and sucking insects (Rahbe et

al., 1995). The insecticidal activity of plant lectins against a large variety of insect species belonging to the orders Coleoptera, Hemíptera, Diptera and Lepidoptera is well documented (Powell et al., 1995). They bind to glycoproteins in the peritrophic matrix or other membranous tissues of the insect midgut to paralyze digestive processes and nutrient uptake. This property points to a potential use of plant lectins as a natural insecticide against a number of harmful pests. With this anti-insecticidal characteristic, some plant lectins are strong postulants for the engineering of plants with resistance to insects (Dutta et al., 2005).

In the Amaryllidaceae family, the most studied anti-insect lectin is the mannose-binding lectin of the species *Galanthus nivalis*, which is known to cause diarrhea when fed to the brown rice leafhopper (*Nilaparvata lugens*) by means of artificial diets (Powell et al., 1995). By means of immunostaining it was found that the lectin was able to cross the epithelial barrier of the midgut and pass into the circulatory system of the insect, resulting in a systemic toxic effect. Electron microscopy studies showed morphological changes in the midgut region of grasshoppers fed, which indicate the possibility of using MBLs of species of the *Allium* genus as a potent control agent to design cultivated plants for insect resistance. The understanding of MBL functions in this process is of great importance for studies on plant-host interaction.

Lectins are reported as pollen germination acceleration factors, reducing pollen tube emergence time and also acting in pollen-pistil recognition (Southworth, 1975; Sousa, 2015). Lectins are described by acting on cell membrane polarization and regulation of specific ion channels and may also be involved in cases of genetic incompatibility (Carvalho, 1990).

The 3D models proposed in this work demonstrate significant differences in the structure of the *A. sativum* and *A. ursinum* lectins observed in Figure 4. It is observed the presence of beta-sheets in the PD585253 region of the *A. sativum* model, that are absent in *A. ursinum*. The absence of the beta-sheets may explain the absence of this conserved domain in *A. ursinum*. The PD330654 domain, however, presents similar conformation, suggesting that this domain presents the same function in the two species.

Protein functions are related to their three-dimensional configuration and this configuration, in turn, is established by its amino acid sequence. Despite the complexity of the protein sequences, homology computational modeling is a viable option for the generation of 3D models of proteins with structure still unknown, based on homologous proteins with already determined structures.

To our knowledge, this work presents the first 3D models of MBLs for the *Allium* genus. 3D modeling and comparative analysis of this protein will provide valuable insights into its evolution and molecular functions in species of this genus.

CONCLUSIONS

The phylogenetic trees based on the MBL sequences are not consistent with the proposed infrageneric classification presented in the literature for the *Allium* genus. However, the methods used were efficient for separation up to the level of tribes within the family Amaryllidaceae.

The use of *in silico* methods proved to be feasible for the construction of a 3D model of MBLs by means of homology modeling, allowing the analysis of their tertiary structures and molecular functions.

ACKNOWLEDGMENTS

"This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001" Jimenez HJ was supported by a PhD fellowship CAPES, RM Moraes Filho was supported by a postdoctoral fellowship PNPD-CAPES.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Andrade MA, O'Donoghue SI and Rost B (1998). Adaptation of protein surfaces to subcellular location. *J. Mol. Biol.* 276: 517-525. https://doi.org/10.1006/jmbi.1997.1498.

Andrade LF, Nahum LA and Avelar LGA (2011). Eukaryotic protein kinases (ePKs) of the helminth parasite Schistosoma mansoni. *BMC Genomics.* 12: 215. https://doi.org/10.1186/1471-2164-12-215.

Barre A, Hervé C, Lescure B and Rougé P (2002). Lectin receptor kinases in plants. *Plant Sci*. 21: 379-399.

Benkeblia N (2003). Postharvest technology of onions. In: Driss R, Niskanen R, Mohan–Jain S. (Ed.) Crop management and post harvest handling of horticultural products. Enfield Science Publishers. Chap. 5: p. 107-137.

Carneiro JGM (1997). Características Funcionais de Concentrados Protéicos de Soro de Leite de Cabras. Campinas. Ph.D. Thesis. Faculdade de Engenharia de Alimentos, Universidade Estadual de Campinas. http://repositorio.unicamp.br/handle/REPOSIP/254345.

Carvalho HF (1990). Aspectos moleculares e biológicos das lectinas. *Cien. Cult.* 42: 884-893.

Chen VB, Arendall WB, Headd JJ and Keedy DA (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* 66: 12-21. http:doi.org/10.1107/S0907444909042073.

Chen VB, Davis IW and Richardson DC (2009). KING (Kinemage, Next Generation): a versatile interactive molecular and scientifc visualization program. *Protein Sci*. 18: 2403-2409. http://dx.doi.org/10.1002/pro.250.

Dias RO, Machado LS, Migliolo L and Franco OL (2015). Insights into Animal and Plant Lectins with Antimicrobial Activities. *Molecules.* 20: 519-541. https:// doi.10.3390/molecules20010519.

Dutta I, Saha P, Majumder P, Sarkar A, et al. (2005). The efficacy of a novel insecticidal protein, *Allium sativum* leaf lectin (ASAL), against homopteran insects monitored in transgenic tobacco. *Plant Biotechnol. J.* 3: 601–611. https://doi: 10.1111/j.1467-7652.2005.00151.

Follmer C, Wassermann GE and Carlini CR (2004). Separation of Jack (*Canavalia ensiformis*) Urease isoforms by immobilized metal affinity chromatography and characterization of insecticidal properties unrelated to ureolytic activity. *Plant Sci*. 167: 241-246. https://doi.org/10.1016/j.plantsci.2004.03.019.

Franz H (1988). The ricin story. *Adv. Lectin Res.* 1: 10–25.

Fritsch RM, Blattner FR and Gurushidze M (2010). New classification of *Allium* L. subg. Melanocrommyum (Webb & Berthel) Rouy (Alliaceae) based on molecular and morphological characters. *Phyton-Int. J. Exp. Bot*. 49: 145–220. https://www.cabdirect.org/cabdirect/abstract/20103124428.

Fritsch RM and Astanova SB (1998). Uniform karyotypes in different sections of *Allium* L. subgenus *Melanocrommyum* (Webb & Berth.) Rouy from Central Asia. *Feddes Repert*. 109: 539–549.

Friesen N and Klaas M (1998). Origin of some minor vegetatively propagated *Allium* crops studied with RAPD and GISH. *Genet. Resour. Crop. Ev*. 45: 511–523. https://doi.org/10.1023/A:100864770.

Gasteiger E, Hoogland C, Gattiker A and Duvaud S (2005). Protein identification and analysis tools on the ExPASy server. In: The proteomics protocols handbook (Walker JM, ed.). 571-607. https://doi.org/10.1385/1-59259-890-0:571.

Ghazarian H, Idoni B and Oppenheimer SBA (2011). glycobiology review: carbohydrates, lectins and implications in cancer therapeutics. *Acta Histochem*. 113(3): 236-247. https:// doi:10.1016/j.acthis.2010.02.004.

Harris JC, Cottrell SL, Plummer S and Lloyd D (2001). Antimicrobial properties of *Allium sativum* (garlic). *Appl. Microbiol. Biotechnol*. 57(3): 282-286. https://www.ncbi.nlm.nih.gov/pubmed/11759674.

Hecht M, Bromberg Y and Rost B (2015). Better prediction of functional effects for sequence variant. *BMC Genomics*. 16: S1. https://doi.org/10.1186/1471-2164-16-S8-S1.

He XJ (1999). Studies on evolutionary biology of the genus *Allium* L. in China. PhD thesis, Sichuan University, Chengdu.

Herre J, Willment JA, Gordon S and Brown GD (2004). The role of Dectin-1 in antifungal immunity. *Crit. Rev. Immunol*. 24: 193–203. https://doi: 10.1615/CritRevImmunol.v24.i3.30.

Kasap S and Benkrid K (2010). High performance phylogenetic analysis with maximum parsimony on reconfigurable hardware. IEEE Transactions on Very Large Scale Integration VLSI Systems. 99: 1–13. https:// doi: 10.1109/TVLSI.2009.2039588.

Khan F, Ahmad A and Khan MI (2007). Purification and characterization of a lectin from endophytic fungus Fusarium solani having complex sugar specificity. *Arch. Biochem. Biophysics*. 457: 243–251. https://doi.org/10.1016/j.abb.2006.10.019.

Kelley LA, Mezulis S, Yates CM and Wass MN (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc*. 10: 845-858. https://doi.org/10.1038/nprot.2015.053.

Kiraga J, Mackiewicz P, Mackiewicz D, Kowalczuk M, et al. (2007). The relationships between the isoelectric point and: length of proteins, taxonomy and ecology of organisms. *BMC Genomics*. 8: 163-10. https://doi.org/10.1186/1471-2164-8-163.

Krieger E, Joo K, Lee J, Raman S, et al. (2009). Improving physical realism, stereochemistry and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. *Proteins*. 77: 114-122. http://doi:10.1002/prot.22570.

Kruse J (1992). Growth form characters and their variation in *Allium* L. In: Hanelt P, Hammer K, Knupffer H. eds. The genus *Allium*: taxonomic problems and genetic resources. Proceedings of an International Symposium held at Gastersleben. p. 173-179.

Kumar S, Stecher G and Tamura K (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol*. 33(7): 1870-4. https://doi.org/10.1093/molbev/msw054.

Li QQ, Zhou SD, He XJ, Yu Y, et al. (2010). Phylogeny and biogeography of *Allium* (Amaryllidaceae: Allieae) based on nuclear ribosomal internal transcribed spacer and chloroplast rps16 sequences, focusing on the inclusion of species endemic to China. *Ann. Bot. London*. 106(5): 709-733. http://doi:10.1093/aob/mcq177.

Liu B, Li J, Ao Y, Qu J, et al. (2012). Lysin motif-containing proteins LYP4 and LYP6 play dual roles in peptidoglycan and chitin perception in rice innate immunity. *Plant Cell*. 24(8): 3406-3019. http://dx. doi:10.1105/tpc.112.102475.

Mo H, Van Damme EJ, Peumans WJ and Goldstein IJ (1993). Purification and characterization *of a mannose-specific lectin from Shallot (Allium ascalonicum) bulbs. Arch. Biochem. Biophys*. 306(2): 431-8.

Moraes Filho RM and Martins LSS (2016). In silico comparative analysis of tylenchid nematodes pectate lyases. *Genet. Mol. Res. GMR*.15038402. http://dx.doi.org/10.4238/gmr.15038402.

Murdock R, Larry L and Shade C (2002). Lectins and protease inhibitors as plant defenses against insects. *J. Agr. Food Chem*. 50(22): 6605–6611. http://dx.doi.org 10.1021/jf020192c.

Nandi S, Mehra N, Lynn AM, Bhattacharya A, et al. (2006). Comparison of theoretical proteomes: identification of COGs with conserved and variable pI within the multimodal pI distribution. *BMC Genomics*. 6: 116-10. http://doi:1186/1471-2164-6-116.

Nelson DL and Cox MM (2005). Lehninger: Principles of Biochemistry. Company Overview of W. H. Freeman and Company, New York.

Peumans WJ and Van Damme JM (1995). Lectins as plants defense proteins. *J. Plant Physiol*. 109: 2347-352.

Pettersen EF, Goddard TD, Huang CC, Couch GS, et al. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem*. 25(13): 1605-12. https://doi.org/10.1002/jcc.20084.

Powell KS, Gatehouse AMR, Hidler VA and Gatehouse JA (1995). Different antimetabolic effects of related lectins towards nymphal stages of *Nilaparvata lugens*. *Entomol. Exp. Appl*. 75: 51-59.

Rahbe Y, Sauvion N, Febvay G, Peumans WJ, et al. (1995). Toxicity of lectins and Processing of ingested proteins in the pea aphid *Acyrthosiphon pisum*. *Entomol. Exp. Appl*. 76: 143-155.

Rahn K (1998). Alliaceae. In: Kubitzki, K. (ed.). The families and genera of flowering plants. Ebook. Series Ed. 3: 70-78p.

Rubinstein N, Ilarregui JM, Toscano MA and Rabinovich GA (2004). The role of galectins in the initiation, amplification and resolution of the inflammatory response. *Tissue Antigens*. 64: 1–12. https://doi: 10.1111/j.0001-2815.2004.00278.

Saito K, Misaki A and Goldstein IJ (1997). Purification and characterization of a new mannose-specific lectin from *Sternbergia lutea* bulbs. *Glycoconjugate J*. 14(8): 889-96.

Servant F, Bru C, Carrère S, Courcelle E, et al. (2002). ProDom: Automated clustering of homologous domains. *Brief. Bioinform*. 3: 246-251.

Sharon N and Lis H (2004). History of lectins: from hemagglutinins to biological recognition molecules. Glycobiology. 14(11): 53-62. https://doi.org/10.1093/glycob/cwh122.

Smeets K, Van Damme E and Peumans WJ (1997). Developmental regulation of lectin and alliinase synthesis in garlic bulbs and leaves. *Plant Physiol*. 113: 765–771.

Sousa AS (2015). Viabilidade e ação de lectinas na germinação in vitro de grãos de pólen de dendezeiro (*Elaeis guineensis* jacq. – arecaceae) feira de Santana. Ph.D. Thesis. - Faculdade de Engenharia de Alimentos, Universidade Estadual de Campinas.

Souza VC and Lorenzi H (2012). Botânica sistemática: Guia ilustrado para identificação das famílias de Fanerógamas nativas e exóticas no Brasil, baseado em APGIII. 3.ed. São Paulo: Editora Plantarum. 768.

Southworth D (1975). Lectins stimulate pollen germination. *Nature*. 258: 600-602.

Troegeler A, Mercier I and Cougoule C (2017). C-type lectin receptor DCIR modulates immunity to tuberculosis by sustaining type I interferon signaling in dendritic cells. *P. Natl. Acad. Sci. U.S.A.* 114(4): 540-549. https:doi:10.1073/pnas.1613254114.

Tsirigos KD, Peters C, Shu N and Käll L (2015). The TOPCONS web server for combined membrane protein topology and signal peptide prediction. *Nucleic. Acids. Res*. 43: 401-407. https://doi.org/10.1093/nar/gkv485.

Van Damme EJ, Goldstein IJ and Peuman WJ (1991). A comparative study of mannose-binding lectins from the Amaryllidaceae and Alliaceae. Phytochemistry. 30 (2): 509-514.

Van Damme EJM, Peumans WJ, Barre A and Rougé P (1998). Plant lectins: a composite of several distinct families of structurally and evolutionarily related proteins with diverse biological roles. *Crit. Rev. Plant Sci*. 17: 575-692.

Van Holle S, De Schutter K, Eggermont L and Tsaneva M (2017). Comparative Study of Lectin Domains in Model Species: New Insights into Evolutionary Dynamics. Komatsu S, ed. *Int. J. Mol. Sci*. 18(6): 1136. https://doi:10.3390/ijms18061136.

Vandenborre G, Smagghe G, Gesquière B, Menschaert, et al. (2011). Diversity in protein glycosylation among insect species. *Plos O*ne. 6(2): 1-9. https:// doi: 10.1371/journal.pone.0016682.

Vasconcelos J, Vieira JGP and Vieira EPP (2009). Plantas Tóxicas: Conhecer para Prevenir. *Rev. Cient. UFPA*. 7(1): 1-10.

Vojdani F (1996). Solubility. In: Hall, G.M. Ed., Methods of Testing Protein Functionality.London:St. Edmundsbury Press. 11-60.

Watson JD, Laskowski RA and Thornton JM (2005). Predicting protein function from sequence and structural data. *Curr. Opin. Struc. Biol*. 15: 275–284. https:// doi: 10.1016/j.sbi.2005.04.003.

Wiederstein M and Sippl MJ (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic. Acids. Res*. 35: 407-410. https://doi.org/10.1093/nar/gkm290.

Zhang Z (1998). Electron transfer by domain movement in cytochrome bc1. *Nature*. 392: 677-84.

Zhou SD, He XJ, Yu Y and Xu JM (2007). Karyotype studies on twenty-one populations of eight species in *Allium* section *Rhiziridium*. *Acta Phytotaxon. Sin*. 45: 207–216. https://doi.10.1360/aps06023.

Zuo Z, Fan H, Wang X, Zhou W, et al. (2012). Purification and characterization of a novel plant lectin from *Pinellia ternate* with antineoplastic activity. *Springer Plus*. 1: 13. https://doi.org/10.1186/2193-1801-1-13.