International Journal of Proteomics & Bioinformatics

Research Article

Comparative Sequence and Structural Analysis of Lectin Protein in Chickpea (*Cicer arietinum L.*) and their Relationship with *fabaceae* Family - a

Megha Ujinwal, Pooja Archana Sahani and Nisha Singh*

**Department of Computational Sciences (Bioinformatics), Central University of Punjab*

***Address for Correspondence:** Nisha Singh, Department of Computational Sciences (Bioinformatics), Central University of Punjab, Bathinda-151001 Punjab, India, Tel: +91-874-400-6631; E-mail: singh.nisha88@gmail.com; ns868@cornell.edu

Submitted: 01 November 2018; **Approved:** 10 April 2019; **Published:** 17 April 2019

Cite this article: Ujinwal M, Sahani PA, Singh N. Comparative Sequence and Structural Analysis of Lectin Protein in Chickpea (*Cicer arietinum L.*) and their Relationship with *fabaceae* Family. Int J Proteom Bioinform. 2019;4(1): 001-006.

ABSTRACT

Chickpea (*Cicer arietinum* L.) also known as garbanzo beans or Bengal gram, is a most important food legume in India. Chickpea has a class of sugar-binding and cell-agglutinating protein present known as Lectin, serve as recognition molecule within cell and different organisms. It also plays a significant role in biotic stress tolerance in chickpea and aids as defense mechanism like insecticidal and anti-microbial. It is a potent protein that can be utilize in several biological and medical aspects. *In-silico* sequence and structural analysis revealed that there are two lectin proteins present in chickpea, G1K3R9_CICAR and R9TPI6_CICAR. Lectin protein has both anti-nutritional and nutraceutical attributes. In present study, analyzed sequence and structural analysis of lectin protein in chickpea and their relationship with other legume crops. Analysis revealed lectin protein including phylogeny, conserved functional motif and its different application agricultural and pharmaceutical field. Sequence analysis results revealed that it comprises high abundance of these three essential amino acids namely phenylalnine, isoleucine and leucine. Due to its high hydrophobic affinity it helps in protein folding and keeps it in stable state. Phylogenetic analysis depicted lectin protein family of chickpea are nested with *Pisum sativum*, *Vigna unguiculata* and *Gylcine max*. Gene structure analysis revealed that six conserved regions which have lower e-value. In total ten active sites were found in lectin protein which are similar to *Lathyrus odoratus*. The current study reveals the detail of primary, secondary and tertiary structure of lectin protein, its interaction with *Arabidopsis thaliana*, construct a homology model and provide valuable information for further analysis to infer the of role of lectin protein mechanism of action at molecular level and structural and functional relationship. Furthermore, this analysis will be useful to examine and elucidate the therapeutic effects, nutritional benefits, and toxic consequences of lectins.

Keywords: Chickpea, hemagglutinins, lectin, legume, motif, sequence analysis

INTRODUCTION

Chickpea (*Cicer arietinum* L), also known as garbanzo beans or Bengal gram, are the third most important type of legume after dry beans and peas. Chickpeas are the second most important grain legume [1]. Fabaceae family of legume plants is the largest plant families which consist of important plants like beans, peas, alfalfa, clover, and so on. Chickpea is self-pollinated crop and diploid annual grain legume with 16 chromosomes and a 738 Mb genome size [2]. There are 28,256 protein-coding genes identified with a mean coding length of 1,166 bp (Database, 2014).

Most widely cultivated group in India about 90% of total world. The highest ever average yield of chickpea in the country was 975 kg per hectare during 2013-14 [3]. Global chickpea production has increased from 7.68 million tonnes (1961) to 13.73 million tonnes (2014) [4]. There are two types of chickpea named Kabuli and Desi. Grains of desi chickpea are small in size, light to dark brown in color and have a thick seed coat while grains of Kabuli chickpea are bigger in size, have a whitish-cream color and thin seed coat. The desi type is more prominent and accounts up to 80% of global chickpea production [5]. The energy content in Desi grain is 327 kcal/100g while in Kabuli grain is 365kcal/100g. Chickpea grains contain 60-65% carbohydrates, 6% fat, and between 12% and 31% protein higher than any other pulse crop. Chickpea is a good source of protein, carbohydrates and all essential amino acid [5]. Chickpea seed contains 23% protein, 64% total carbohydrates, 47% starch, 5% fat, 6% crude fiber, 6% soluble sugar and 3% ash and maximum digestibility as compared to other dry edible legumes. They are also a good source of calcium, magnesium, potassium, phosphorus, iron, zinc and manganese [6]. The protein content varies from 20.9 - 25.27% [7].

Chickpea proteins had a higher true digestibility, biological value and net protein utilization than those of cowpea and mung bean. The nitrogen fixation nodules produced by chickpea can enhance the soil fertility. It comprises the highest nutritional compositions of any dry edible legume and contains only low levels of some anti-nutritional factors [8]. Legume seed proteins primarily increase the nutritional quality and impart a variety of functional properties, including structure, texture, flavour and colour to food products. Legumes are good source of carbohydrate and have a low Glycemic Index (GI) value it has resistant starch properties. [9,10]. Chickpea have low GI value of 28-32 (nipgr.nes/NGCPCG). Seed lectin protein is group of glycoprotein function as storage and defenses protein [11]. Lectin is carbohydrate binding protein macromolecule that is highly specific for sugar moieties. Lectin performs recognition at cellular and molecular level and plays a numerous role in biological ubiquitously in different plant species and extracted principally from seeds. The rich hydrophobic amino acid regions of lectins allow interactions with other molecules like host–pathogen interactions, cell targeting and cell-cell communications [12]. X-ray characterization of a lectin from chickpea having molecular weight of 43 kDa composed of two identical subunits [13]. It also lower blood the glucose level as they bind with and disrupt intestinal mucosal cells, leading to hindrance in nutrient absorption [14]. It constitutes part in defensive mechanism of the seed but also considered as anti-nutritional factors for the human diet [15].

MATERIAL AND METHODS

Clustering of sequence and primary structure analysis of different lectin protein

Lectin protein sequences of chickpea and other legume seed were retrieved from UniProt database (http://www.uniprot.org/). Under Fabaceae family, we found about 24 species with lectin protein. The number of lectin protein at different locus is varies among species. In *Cicer arietinum*, we found only two lectin protein (*R9TPI6_CICAR, G1K3R9_CICAR*) at different locus. To analyzed amino acid composition, hydrophilicity and hydrophobicity using BioEdit sequence alignment tool [16].

Hydrophobicity and hydrophilicity of protein of different legume seeds estimated by Kyte and Doolittle scale mean hydrophobic scale and Bokyo scale mean hydrophobicity profile method in Bio Edit. Graph is plotted with X-axis and Y-axis. A window of defined size is moved along a sequence, the hydropathy scores are summed along the window, and the average (the sum divided by the window size) is taken for each position in the sequence. To standardize the protein structure, we determine secondary structure of lectin protein from PHD a neural network server (https://npsa-prabi.ibcp.fr/NPSA/npsa_phd.html). A random prediction of secondary structure in three states (helix, strand, rest—here termed loop) yields an overall

per-residue accuracy of 35% [17]. The sequence has to be written into a file according to the format. It used to determine alpha helix, Beta Bridge, random coil, beta turn, extended strand and ambiguous state of lectin protein for all 23 legume crops.

Construct multiple sequence alignment and phylogenetic tree

Multiple sequence alignments of amino acid sequence of lectin protein of chickpea and those of soya bean, cowpea, pigeon pea and pea were performed using Clustal Omega (http://www.clustal.org). MSA was performed to determine protein structure and function prediction, phylogeny inference which were common tasks in sequence analysis. Clustal Omega is use mBED algorithm for calculating guide tree for either large or small numbers of DNA/RNA or protein sequences due [18].

The predicted chickpea lectin protein was classified within group based on sequence alignment and phylogenetic relationship with clearly classified lectin protein from *Pisum sativum, Glycine max, Cajanus cajan, Vigna unguiculata*. The phylogenetic tree was visualized using MEGA (The Molecular Evolutionary Genetics Analysis) available at https://www.megasoftware.net/. Evolutionary genetics analysis was performed by using maximum likelihood, evolutionary distance, and maximum parsimony methods. A bootstrap analysis with 1000 reiterations was conducted to determine the statistical stability of each node. A phylogenetic tree approximates the relationships between taxa (or sequences) and their hypothetical common ancestors [19-21].

Motif identification

For analysis of conserved region within lectin protein of different legume crop were identified using MEME ((Multiple EM for Motif Elicitation) tools (meme-suite.org/tools/meme) across diverse lectin protein family. MEME perform by searching for repeated, un-gapped sequence patterns that occur in the DNA or protein sequences provided. MEME determine the width and number of occurrences of each motif repeatedly in order to minimize the 'E-value' of the motif. E-value is the probability of finding an equally well conserved pattern in random sequences [22]. All results were manually examined to re-confirm the output.

Homology modelling

The three-dimensional (3D) structure of lectin protein with one known protein (template), *Lathyrus odoratus* were generated by intensive protein modelling using MODELLER ((salilab.org/modeller.html) [23] using 'Normal' mode modelling based on alignment to experimentally solved protein structures. for their lectin protein. The protein sequences were aligned to be model having template structure, the atomic co-ordinates of templates and a script file.

RESULTS AND DISCUSSION

Analysis of primary and secondary structure of lectin protein

An amino acid compositions of lectin protein in chickpea were examined the total essential amino acid content (Molar percentage) varies from 12506.97 Daltons to 100511.87 Dalton proteins. The result of primary structure analysis suggested that both lectin proteins of chickpea are more abundant in phenylanaline (8.37%), isoleucine (7.93%) and lysine (8.37%) amino acid. Also the molar percentage difference between chickpea essential and non-essential amino acid i.e., essential amino acid glutamic (number = 10 and mol % = 4.41) and non-essential amino acid phenylalanine (number = 19 mol % = 8.37) show different values as shown in graph (Figure 1).

Kyte and Doolittle method was used to calculate mean hydrophobicity as it can identify surface exposed region and transmembrane region. Hydrophobic moment was calculate for each residue by keeping same window size for all of amino acid. A set of "hydropathy scores" is performing for the 20 amino acids based upon a compilation of experimental data from the literature. Xaxis show position of each amino acid in sequence and Y-axis show mean hydrophobicity and mean hydrophilicity, respectively. Here window size is 13 which are suitable to finding transmembrane domains (Figure 2,3).

Value above zero is considered to have more amino acid at particular position and give its mean hydrophobicity value. Similarly, negative value show less number of amino acid in that particular region also gives its mean hydrophocity value. But if we compare it with hydrophilicity scale, it depicts that it is greater than hydrophilicity. Thus, results signify that lectin protein have high hydrophobicity property than hydrophilicity. Similarly in case of

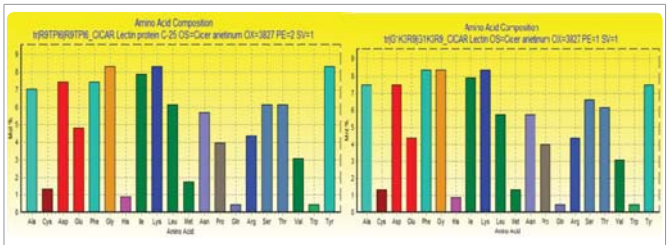

Figure 1: Amino acid composition of two Chickpea lectin proteins (G1K3R9_CICER Lectin, R9TPI6_CICER Lectin) found at two different locus.

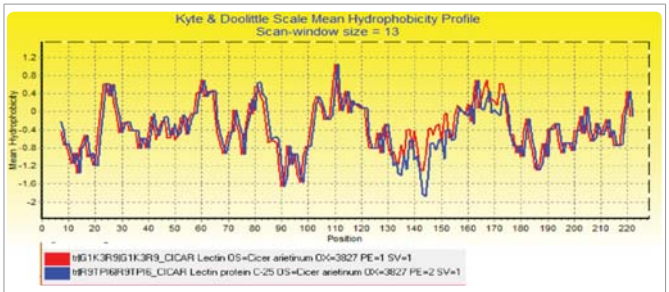

Figure 2: Kyte-Doolitte hydrophobicity of two different chickpea lectin protein on its X-axis and degree of hydrophobicity on its Y-axis.

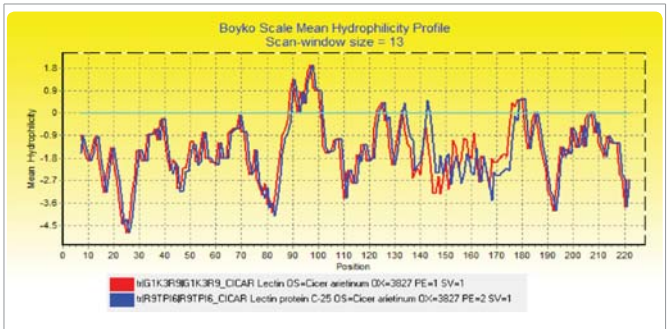

Figure 3: Bokyo plot has amino acid sequence of Chickpea lectin protein on its X-axis and degree of hydrophobicity on its Y-axis.

hydrophilicity, only few of residue shows positive value, means it has low hydrophilicity as shown in graph (Figure 4,5).

Multiple alignment of five legume seeds (*Cicer arietinum*, *Cajanus cajan* , *Pisum sativum*, *Gylcine max*, *Vigna unguiculata*) of lectin protein. Different color represents different amino acid to see locations of similarity and difference among the sequences based on the chemical nature of the amino acid residues. RED (residues AVFPMILW) show Small (small+ hydrophobic (incl. aromatic -Y)), BLUE (residues DE) are Acidic, MAGENTA (residues RK) are Basic-H, GREEN (residues STYHCNGQ) show Hydroxyl, sulfhydryl, amine (G) group. Sequence alignment show similarly among lectin protein of different legume crops. From this interpretation, structure can be predicted by using different modelling tool.

Phylogenetic analysis of lectin protein family across legumes

A strong correlation between phylogenetic analysis with the gene function has been observed in lectin protein indicating that the amino acid based phylogenetic analysis can be used to predict a putative function of the lectin protein. To investigate the phylogenetic relationships and to predict the functionality of the lectin protein we constructed a phylogenetic tree of 32 sequences of lectin protein from five legumes including *Glycine max* (ten), *Cajanus cajan* (ten), *Pisum sativum* (seven), *Cicer arietinum* (two) and *Vigna unguiculata* (three). A phylogenetic trees were constructed to decipher the relationships across different legumes and their relative order of speciation events. In phlyogentic tree, lectin protein were clustered into five distinct groups represented with different color. The percentage of amino acid sequence similarity represents that the sequences of Gylcine max are similar to each other and have common ancestor. There are 2 clade in group I. Three of sequences are considered to be similar in group I. Similarly, in group II, nested with pigeonpea and soyabean lectin proteins. Surprisingly, they both possess high content of lectin then chickpea. However, group III is one of the largest clade shared with chickpea emphasis a great relationship with pea, soyabean, and cowpea it showed great similarity has total 13 lectin protein sequences in which 3 cowpea, 2 chickpea and 8 from pea.

From these evidence for an evolutionary relationship It signifs, lectins have high degree of amino acid sequence identity shared across different legume lectin protein family. Thus, across entire six group chickpea showed more kinship with pea and cowpea in group III. This represent that chickpea share its common ancestor with these two legume seed and have high percentage of amino acid sequence similarity than other. Group IV and group III are diverging from a common node, but don't show similarity. Therefore, it revealed that chickpea lectin protein are more closely related with pea and cowpea than soyabean and pigeonpea (Figure 6).

Conserved residue identification (Motif) in chickpea lectin protein

The motif discovery algorithm is looking for a set of similar short sequences (the needle) in a set of much longer sequences (the haystack). The problem is easier when the motif instances are long and very similar to each other. The amino acid sequences are represented by different colors (Figure 7,8).

We set a parameter for motif identification i.e., identifies ten conserved regions in MEME. In all 10 motifs, we found six are more conserved motif then rest of other. Result depict that six are functional motif and other are non-functional motif. In six functional motifs, first four are more common significant than other motifs. These four

Figure 4: Secondary structure of lectin protein finds alpha-helix (17.14%), extended strand (24.18%) and random coil (58.68%). This depict that lectin protein are highly compose of beta plated sheet and random coil. Lectin protein is beta sheet dominated protein family. Multiple Sequence Alignment (MSA) and phylogenetic analysis.

Figure 5: Alignment of deduced amino acid sequence of chickpea seed lectin with other legume lectins using Clustal Omega.

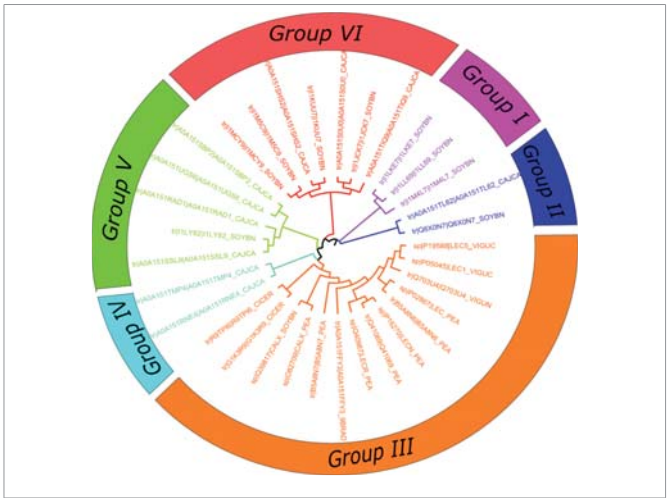

Figure 6: Phylogenetic tree representing the evolutionary relationship among the lectin proteins of chickpea, pea, soyabean, pigeon pea and cow pea.

conserved regions show low E-value (2.6e to 1.0e), number of sites also different then non-functional motif. The smaller the E-value, higher the degree of similarity. Small E-value i.e., low number of hits, but of high quality. Each of four significant regions show likelihood ratio, relative entropy, Bays Threshold value which are nearby to each other. Protein motifs are generally easier to discover remaining to the length of the protein sequence and the chemical similarity with groups of amino acids. This allows shorter motifs to be more statistically significant and makes it easier to distinguish functional motifs from non-functional motif.

Homology modelling of chickpea lectin proteins

In order to understand the structural properties of lectin protein of chickpea, the 3-dimensional (3-D) protein model of two of lectin protein were constructed using MODELLER and result are shown in above figure 9. Both of the 3-D protein models have been constructed with 100% confidence and residue coverage. Hence, the predicted 3-D protein structures are considered highly reliable and offer a preliminary basis for understanding the molecular function of lectin protein in chickpea. Lectin protein contained highly conserved region with less number of alpha-helix and highly composed of beta-plated sheet both parallel and anti-parallel. This pointed out that the homology-based 3-D modelling of plant chickpea can be effectively used for understanding the substrate specificity and molecular function of lectin protein. The chemical composition of chickpea flour is different significantly from the composition of wheat flour [24]. Lectin is cell-agglutinating protein of non-immune origin that bind monosaccaride and oligosaccharide reversibly and with high specificity. Bioactive consequence was shown by lectin protein in human like small intestine hyperplasia, change in intestinal flora immuno-modulating activity and hormone secretion [25].

Figure 7: Number of identified motif in protein sequence.

Figure 8: Visualization of protein motif. The height of a letter indicates its relative frequency at the given position (x-axis) in the motif.

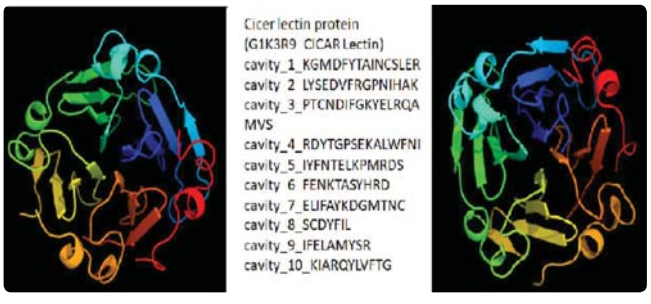

Figure 9: 3-D structure of Lectin protein (R9TPI6_CICAR Lectin [right]) and (G1K3R9_CICAR [left]) of *Cicer arietinum*.

It has great potential in the treatment, prevention and diagnosis of chronic diseases, such as cancer and diabetes. This study signifies the identification and genome wide survey of lectin protein in chickpea with other *fabaceae* family. *In-silico* analysis revealed that lectin protein has high percentage of essential amino acid and high hydrophobicity mean scale. Secondary and tertiary structure depicted that it has more beta-platted sheet both parallel and anti-parallel than alpha-helix and random coil. Phylogenetic analysis reveals that it shares its evolutionary relationship with pea, soya bean and cowpea. It shows a significant interaction with *Arabidopsis thaliana*. From 3-D structure we conclude that lectin protein has ten active site and show similarity with *Lathyrus odoratus* (Sweet pea). Furthermore, lectins from legume resources have allowed the discovery of proteins with a great capacity to distinguish specific glycans from a diversity of glyco-targets found in a variety of organisms.

FUTURE PROSPECTIVE

Lectins from legume resources, specific glycan and their target can be identified in different organisms with great capacity. Useful in lectin-replacement therapy for patients suffering from lectin deficiency defects. Further research, including clinical trials, mechanisms of action at the molecular level, and structure-function relationships, should help researchers continue to examine and elucidate the therapeutic effects, nutritional benefits, and toxic consequences of lectins.

ACKNOWLEDGEMENTS

I would like to thank centre of computational Sciences Department of Bioinformatics, central university of Punjab, for providing me an opportunity and all the necessary facilities for carrying out of the research work.

REFERENCES

1. Acharjee S, Sarmah BK. Biotechnologically generating 'super chickpea' for food and nutritional security. Plant science. 2013; 207: 108-116. https://bit.ly/2Pf8SAU
2. Muehlbauer FJ, Sarker A. Economic importance of chickpea: production, value, and world trade. In The Chickpea Genome. Springer; 2017. p. 5-12
3. Nahatkar S. Production trend and cost-profitability structure of chickpea in central India: The Dynamic Scenario. Other Publications of the Directorate. 2017; 30.
4. FAOSTAT. 2011. https://bit.ly/2PfZuwZ
5. Thudi, K, Roorkiwal M, Kudapa H, Chaturvedi SK, Singh NP, Varshney RK. An overview of chickpea research: from discovery to delivery. Pulse India. 2017; 2: 22-25.
6. Hayriye Ibrikci, Sharon JB Knewtson, Michael A Grusak. Chickpea leaves as a vegetable green for humans: evaluation of mineral composition. Journal of the Science of Food and Agriculture. 2003; 83: 945-950. https://bit.ly/2DfakP6
7. Dhawan K, Malhotra S, Dahiya BS, Singh D. Seed protein fractions and amino acid composition in gram (Cicer arietinum). Plant Foods Hum Nutr. 1991; 41: 225-232. https://bit.ly/2Gv2Dqh
8. Friedman M. Nutritional value of proteins from different food sources. A review. J Agric Food Chem. 1996; 44: 6-29. https://bit.ly/2V61Kwj
9. Foster Powell K, Holt SH, Brand Miller JC. International table of glycemic index and glycemic load values: 2002. Am J Clin Nutr. 2002; 76: 5-56. https://bit.ly/2KNtPV9
10. Johnson SK, Thomas SJ, Hall RS. Palatability and glucose, insulin and satiety responses of chickpea flour and extruded chickpea flour bread eaten as part of a breakfast. Eur J Clin Nutr. 2005; 59: 169-176. https://bit.ly/2v6QukN

11. Esteban R, Dopico B, Munoz FJ, Romo S, Labrador E. A seedling specific vegetative lectin gene is related to development in Cicer arietinum. Physiol Plant. 2002; 114: 619-626. https://bit.ly/2ZiI9YQ

12. Lagarda Diaz I, Guzman Partida AM, Vazquez Moreno L. Legume lectins: proteins with diverse applications. Int J Mol Sci. 2017; 18. 1242. https://bit.ly/2Gs6lRn

13. Katre UV, Gaikwad SM, Bhagyawant SS, Deshpande UD, Khan MI, Suresh CG. Crystallization and preliminary X-ray characterization of a lectin from Cicer arietinum (chickpea). Acta Crystallogr Sect F Struct Biol Cryst Commun. 2005; 61: 141-143. https://bit.ly/2Gtj1HM

14. Gupta RK, Gupta K, Sharma A, Das M, Ansari IA, Dwivedi PD. Health risks and benefits of chickpea (Cicer arietinum) consumption. J Agric Food Chem. 2017; 65: 6-22. https://bit.ly/2v9rtWc

15. Roy F, Boye, JI, Simpson BK. Bioactive proteins and peptides in pulse crops: Pea, chickpea and lentil. Food research international. 2010; 43: 432-442. https://bit.ly/2VXpbow

16. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In Nucleic acids symposium series (Vol. 41, No. 41, pp. 95-98).Information Retrieval Ltd., 1979-2000. 1999

17. Rost B, Sander C. Prediction of protein secondary structure at better than 70% accuracy. J Mol Biol. 1993; 232: 584-599. https://bit.ly/2KKtGC4

18. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011; 7: 539. https://bit.ly/2PeqooX

19. Masatoshi Nei, Sudhir Kumar. Molecular evolution and phylogenetics. Oxford university press; 2000. p.348. https://bit.ly/2Pe6JWk

20. Felsenstein J, Felenstein J. Inferring phylogenies (Vol. 2). Sunderland: Sinauer associates; 2004.

21. Hall BG. Building phylogenetic trees from molecular data with MEGA. Molecular biology and evolution. Mol Biol Evol. 2013; 30: 1229-1235. https://bit.ly/2UEsJQ6

22. Bailey TL, Williams N, Misleh C, Li WW. MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. 2006; 34: 369-373. https://bit.ly/2Xk3aAQ

23. Webb B, Sali A. Protein Structure Modeling with MODELLER. Methods Mol Biol. 2014; 1137: 1-15. https://bit.ly/2Ir20jl

24. Rachwa Rosiak D, Nebesny E, Budryn G. Chickpeas-composition, nutritional value, health benefits, application to bread and snacks: a review. Crit Rev Food Sci Nutr. 2015; 55: 1137-1145. https://bit.ly/2v8ea86

25. Duranti M. Grain legume proteins and nutraceutical properties. Fitoterapia. 2006; 77: 67-82. https://bit.ly/2KFVmb0