DOI: 10.15740/HAS/AU/9.3/347-353
Agriculture Update
Volume **9** | Issue 3 | August, 2014 | 347-353 |
Visit us : *www.researchjournal.co.in*
e ISSN-0976-6847

Research Article

Multiple sequence alignments analysis of *Phaseolin gene* from *Phaseolus vulgaris* (Common bean)

ROSY RANI, JITENDER SINGH, PANKAJ KUMAR, PRADEEP SHUKLA AND PRAGATI MISRA

ARTICLE CHRONICLE :
Received :
15.05.2014;
Revised :
13.06.2014;
Accepted :
28.06.2014

SUMMARY : The common bean *Phaseolus vulgaris*, contains major seed storage protein is known as the phaseolin. The phaseolin is the salt-soluble glycoprotein that account of some 50 per cent of the total protein in mature bean seeds. The computational packages and online servers are the current tools used in the nucleotide or protein sequence analysis and characterization. In the present experiment the full length phaseolin gene was sequenced and submitted to NCBI Gene bank. This nucleotide sequence was BLAST against NCBI nucleotide database and resulted ten selected sequences retrieved and were used for multiple sequence alignments analysis, which was performed by using Bioedit software to investigate relationship with retrieved nucleotide sequences.

How to cite this article : Rani, Rosy, Singh, Jitender, Kumar, Pankaj, Shukla, Pradeep and Misra, Pragati (2014). Multiple sequence alignments analysis of *Phaseolin gene* from *Phaseolus vulgaris* (Common Bean). *Agric. Update*, **9**(3): 347-353.

KEY WORDS :
Phaseolus vulgaris, *Phaseolin* gene, BLAST analysis, Multiple sequence alignment

Author for correspondence :
PRAGATI MISRA
Sam Higginbottom Institute of Agriculture, Technology and Sciences, ALLAHABAD (U.P.) INDIA
Email: pragatimisra3@rediffmail.com

See end of the article for authors' affiliations

BACKGROUND AND OBJECTIVES

The common bean, *Phaseolus vulgaris*, is an herbaceous annual plant domesticated independently in ancient Mesoamerica and the Andes and now grown worldwide for its edible bean, popular both dry and as a green bean. Historically, this species has been an important model for the study of seed storage proteins (Vitale and Bollini, 1995). The leaf is occasionally used as a leaf vegetable and the straw is used for fodder. It is a staple food for many people due to its energy, protein, dietary fibre and minerals content (Haytowitz *et al*., 1981; Norton *et al*., 1985).

The common bean seeds have valuable nutritional properties due to the fact that they are an important source of fibre, minerals and vitamins, as well as to their low content of fat and sodium (Sgarbieri and Whitaker, 1982). The major seed storage protein (40–50%) of common bean is phaseolin, which belongs to the 7S vicilin class (Bollini and Vitale, 1981; Gepts and Bliss,1986). Phaseolin contributes, therefore, to the nutritional value of common bean seed proteins in an important way and Phaseolin is the salt-soluble glycoprotein of the common bean (*Phaseolus vulgaris* L.) that account for some 50 per cent of the total protein in mature bean seeds. It was one of the first plant proteins to be translated *in vitro* from mRNA and one of the first plant genes isolated. It was also the first developmentally regulated plant gene to be expressed in a heterologous plant species through *Agrobacterium*-mediated transformation.

Sequence alignment is a standard technique in bioinformatics for identifying the evolutionarily or structurally related positions in a collection of nucleotide or amino acid sequences (Notredame, 2002). An important goal of research in computational and molecular biology is to reach a better understanding of biological systems at the gene level. Therefore, multiple sequence alignment (MSA) analysis is a powerful means for analyzing structural and functional similarities and differences and for

finding historical and evolutionary relationships between nucleotide or protein sequences. MSA involves computing the alignment of three or more sequences and is a computationally expensive method. A large number of methods have been developed for fast and accurate multiple sequence alignments. Most modern programmes for constructing multiple sequence alignments (MSAs) consist of two components: an objective function for assessing the quality of a candidate alignment of a set of input sequences, and an optimization procedure for identifying the highest scoring alignment with respect to the chosen objective function (Notredame, 2002). In this paper we conduct the analysis based on multiple sequence alignment of identified phaseolin gene to find out the relationship with the other related sequences.

RESOURCES AND METHODS

The obtained full length nucleotide sequence of *phaseolin* gene through sequencing from YVR life sciences, Ghaziabad was subjected to BLAST analysis (BLASTX, National Center for Biotechnology Information - NCBI) with the homologous nucleotide sequences. A BLAST search enables a researcher to compare a query sequence with a library or database of sequences and identify library sequences that resemble the query sequence above a certain threshold. The sequences homologues to the query sequence were retrieved on the bases of query coverage, maximum identity and e-value. These retrieved homologous nucleotide sequences were used for the multiple sequence analysis. The multiple sequence alignment and editing of the *phaseolin gene* of *P. vulgaris* were performed with ClustalW (Thompson *et al.*, 1997) in BioEdit software (Hall, 1999). In multiple sequence alignment (MSA), the input set of query sequences are assumed to have an evolutionary relationship by which they share a lineage and are descended from a common ancestor.

OBSERVATIONS AND ANALYSIS

The multiple alignments give higher accuracy alignments than pair wise alignments of the same sequences. This is because when multiple sequences are aligned residues that are conserved through evolution for structural or functional reasons are highlighted within the alignment profile. The multiple alignments are built up progressively by a series of pair wise alignments of the query nucleotide sequence. The nucleotide sequence of *Phaseolin* gene was identified and sequenced, which represents 1935 bp nucleotide sequence. This nucleotide sequence was submitted to NCBI gene bank. The ten sequences were selected from BLAST results on the basis of maximum identity and query coverage. These sequences showed maximum identity between 79-100 per cent and query coverage in the range between 27-98 per cent with low e-value. The query sequence of phaseolin gene was aligned with these ten sequences selected from BLAST analysis results. Given the sequences of a set of nucleotides to be compared, an alignment displays the residues for each nucleotide on a single line, with gaps ("-") inserted such that "equivalent" residues appear in the same column. The precise meaning of equivalence is generally context dependent for the phylogenetic analysis, equivalent residues have common evolutionary ancestry and for the structural analysis, equivalent residues correspond to analogous positions belonging to homologous folds in a set of nucleotides. The sequences *Phaseolus vulgaris* mRNA for beta-type phaseolin precursor (accession no. GenBank: X03004.1), *Phaseolus vulgaris* mRNA for alpha-type phaseolin precursor (accession no. GenBank: X02980.1), *Phaseolus vulgaris* Sanilac clone 2-13 phaseolin (Phs) mRNA, complete cds (accession no. GenBank: U01132.1), *Phaseolus vulgaris* Sanilac clone 1-12 phaseolin (Phs) mRNA, complete cds (accession no. GenBank: U01131.1), *Phaseolus vulgaris* beta-type phaseolin storage protein gene, complete cds (accession no GenBank: J01263.1), *Phaseolus vulgaris* gene for alpha-phaseolin (accession no. GenBank: X52626.1), Part of the gene for phaseolin in *Phaseolus vulgaris* (bean) (accession no. GenBank: V01163.1) *Phaseolus lunatus* phaseolin (Phs) mRNA, complete cds (accession no. GenBank: U01121.1), *Vigna radiate* 8S globulin beta isoform precursor mRNA, complete CDS (accession no. GenBank: DQ538335.1) and *Vigna unguiculata* partial vicilin gene, exons 1-6 (accession no. GenBank: AM905848.1) were selected for multiple sequence alignment. The sequence *Phaseolus vulgaris* gene for alpha-phaseolin (accession no. GenBank: X52626.1) and *Phaseolus vulgaris* beta-type phaseolin storage protein gene, complete cds (accession no GenBank: J01263.1) were showed maximum similarity, while sequence *Vigna radiate* 8S globulin beta isoform precursor mRNA, complete CDS (accession no. GenBank: DQ538335.1) was showed minimum similarity on alignment. The sequence *Phaseolus vulgaris* mRNA for beta-type phaseolin precursor (accession no. GenBank: X03004.1), *Phaseolus vulgaris* mRNA for alpha-type phaseolin precursor (accession no. GenBank: X02980.1), *Phaseolus vulgaris* Sanilac clone 2-13 phaseolin (Phs) mRNA, complete cds (accession no. GenBank: U01132.1), *Phaseolus vulgaris* Sanilac clone 1-12 phaseolin (Phs) mRNA, complete cds (accession no. GenBank: U01131.1), Part of the gene for phaseolin in *Phaseolus vulgaris* (bean) (accession no. GenBank: V01163.1), *Phaseolus lunatus* phaseolin (Phs) mRNA, complete cds (accession no. GenBank: U01121.1), and *Vigna unguiculata* partial vicilin gene, exons 1-6 (accession no. GenBank: AM905848.1) were showed average similarity on multiple sequence alignment (Fig.1). Sequence alignment of phaseolin

Fig. 1: Contd.....

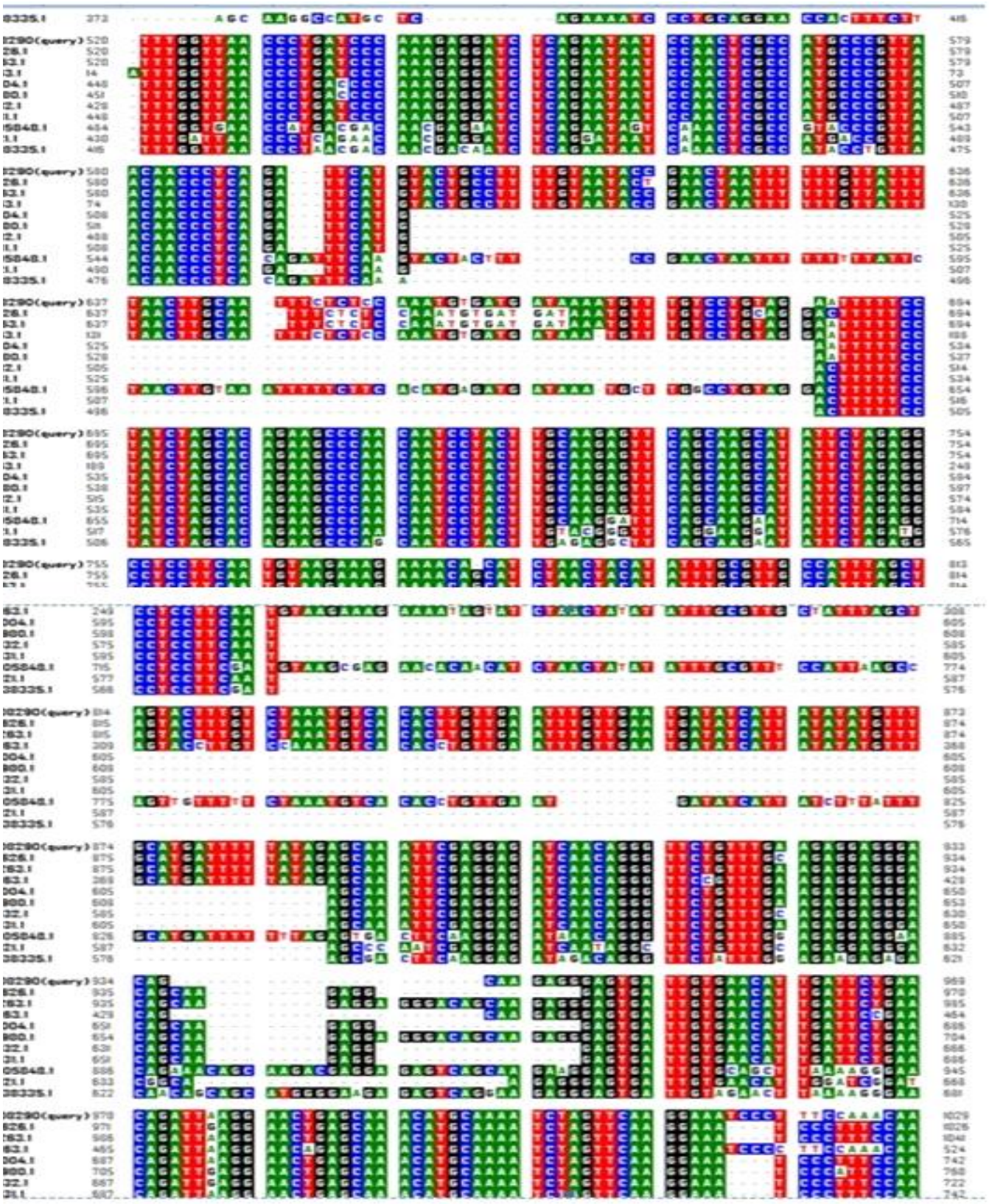

Fig. 1: Contd.....

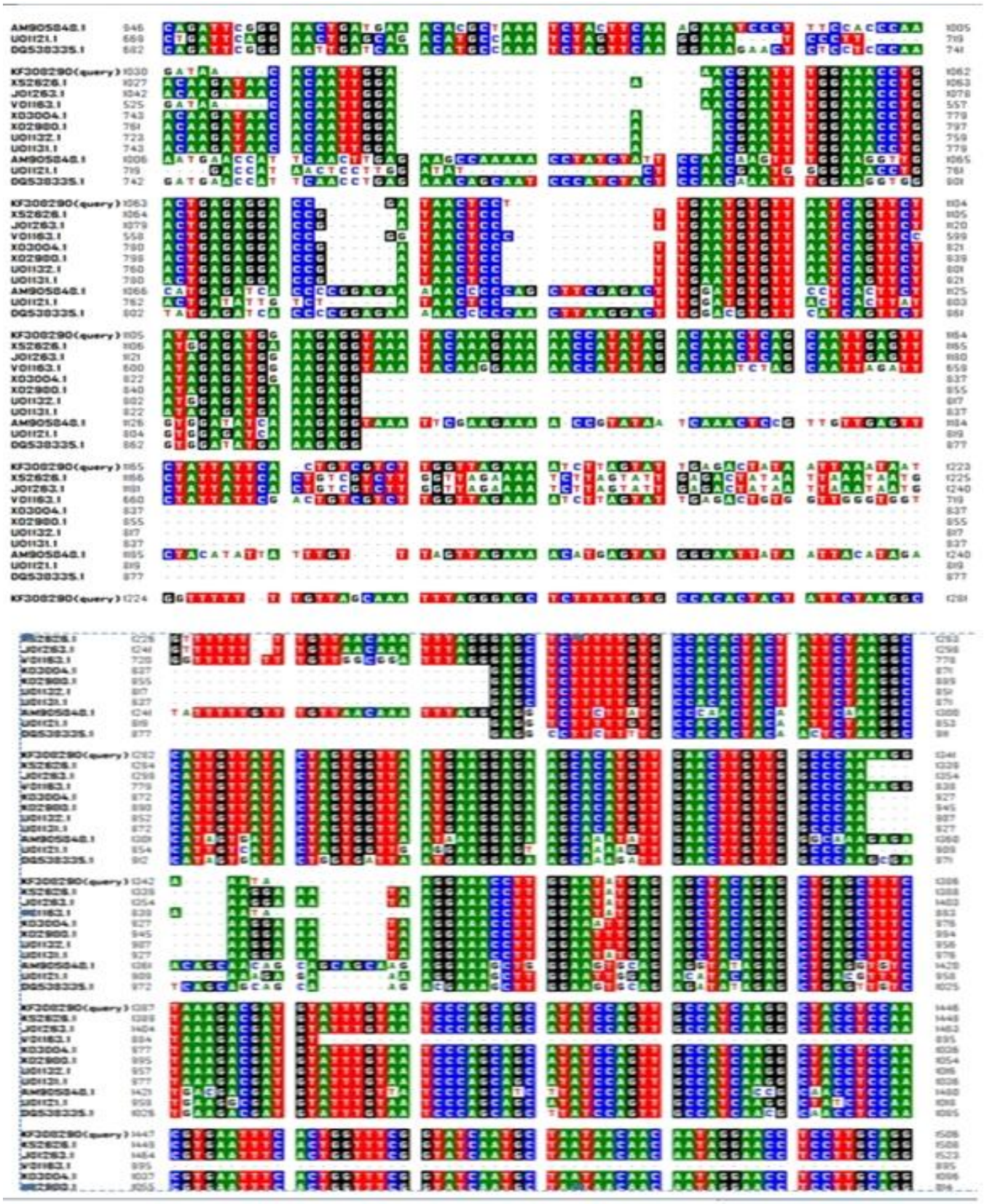

Fig. 1: Contd.....

Fig. 1: Multiple sequence alignment of phaseolin gene with other submitted genes

nucleotide sequence (Fig.1) reveals level of nucleotide similarity and their evolutionary relationships to the earlier reported phaseolin gene sequences. Such an alignment can show incidence of duplications, nucleotide base substitution and deletions as revealed by gaps. These alternations may be related to altered functional properties by presence of different amino acids in protein. The nucleotide sequence for phaseolin and *Vigna radiate* 8S globulin beta isoform precursor mRNA, complete CDS (accession no. GenBank: DQ538335.1) demonstrate their evolutionary difference by the high rate of sequence substitution and presence of unique gap2 providing a distinctive characteristic of phaseolin from *Vigna radiate*, the other additional gaps in phaseolin are showing divergence from other nucleotide sequences (Fig. 1). Aligning different sequences using traditional multiple sequence alignment methods disregards the phylogenetic implications of gap patterns that they create and infers systematically biased alignments with excess deletions and substitutions. The Bioedit tool is based on an algorithm that prevents these systematic errors by recognizing insertions and deletions as distinct evolutionary events, improving the quality of sequence alignments and downstream analyses over a wide range of realistic alignment problems (Löytynoja and Goldman, 2008).

Authors' affiliations :

ROSY RANI AND PRADEEP SHUKLA, Sam Higginbottom Institute of Agriculture, Technology and Sciences, ALLAHABAD (U.P.) INDIA

JITENDER SINGH AND PANKAJ KUMAR, College of Biotechnology, Sardar Vallabhbhai Patel University of Agriculture and Technology, MEERUT (U.P.) INDIA

REFERENCES

Bollini, R. and Vitale, A. (1981). Genetic variability in charge microheterogeneity and polypeptide composition of phaseolin, the major storage protein of *Phaseolus vulgaris* and peptide maps of its three major subunits, *Physiol. Plant.,* **52** : 96–100.

Gepts, P. and Bliss, F.A. (1986). Phaseolin variability among wild and cultivated common beans (*Phaseolus vulgaris*) from Colombia. *Econ. Bot.,* **40**(4): 469-478.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis programme for Windows 95/98/NT. *Nucl. Acids Symp. Ser.,* **41** : 95–98.

Haytowitz, D.B., Marsh, A.C. and Matthews, R.H. (1981). Content of selected nutrients in raw, cooked and processed legumes. *Food Tech.,* **35**: 73-74.

Löytynoja, A. and Goldman, N. (2008). A model of evolution and structure for multiple sequence alignment. *Philos. Trans. R. Soc. Lond B. Biol. Sci.,* **363** : 3913–3919.

Norton, G., Bliss, F.A. and Brezan, R. (1985). Biochemical and nutritional attributes of grain legumes. In R. J. Summerfield and E. H. Roberts (Eds.), Grain Legume Crops (pp. 73-114). LONDON, UNITED KINGDOM.

Notredame, C. (2002). Recent progress in multiple sequence alignment: a survey. *Pharmacogenomics,* **3** : 131–144.

Sgarbieri, V.C. and Whitaker, J.R. (1982). Physical, chemical and nutritional properties of common bean (Phaseolus) Proteins. In: Advances in food research, Vol. **28**, C.O. Chichester, E.M. Mrak, B.S. Schweigert (Eds.), Academic Press, New York, NY, USA pp. 93–166.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmouginand, F. and Higgins, D.G. (1997). The CLUSTAL_X windows interface: Xexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acids Res.,* **25** : 4876–4882.

Vitale, A. and Bollini, R. (1995). Legume storage proteins. In: Seed development and germination. Edited by: Kigel J, Galili G. New York: Marcel Dekker. pp. 73-102.