

DOI: 10.15740/HAS/AU/9.3/347-353

Agriculture Update Volume 9 | Issue 3 | August, 2014 | 347-353 |

Visit us : www.researchjournal.co.ir



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 **O**BJECTIVES

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

EF300290(query) X52826.1 J01263.1 X03004.1 X03004.1 X03004.1 X0300.1 U01132.1 U01132.1 U01131.1 AM905540.1 U01131.1 D0530325.1				AT CAT CACAC AT CAT CACACA AT CAT CACACA			CT GGGAATC CT GGGAATC	50 50 ± 50 0 67 0 6
KF300290 (query) X52828.1 J0(283.1 V01163.1 X02900.1 U01132.1 U01132.1 U01132.1 U01131.1 AM905940.1 U01121.1 U01123.1	(2) (2) (2) (2) (2) (2) (2) (2) (2) (2)		A CAC C A CAC C C A CAC C C A CAC C C A CAC C C C C	SCCTCATC SCCTCATC SCCTCATC SCCTCATC SCCTCATTC SCCTCATTC SCCTCATTC SCCTCATCS SCCTCATCS SCCTCATCS SCCTCATCS SCCTCATCS SCCTCATCS SCCTCATCS	CCACT CACT CCACT CACT			19929 - 19929 - 1992 - 19929 - 1992 - 19929 - 1992 - 1992
KF300290(query) X52828.1 J01253.1 V01153.1 X02900.1 X02900.1 U01132.1 U01132.1 U01132.1 U01132.1 U01132.1 U0112.1 D012533335.1				AACTCTGACA AACTCTGACA AACTCTGACA AACTCTGACA AACTCTGACA AACTCTGACA AACTCTGACA AACTCTGACA AACTCTGACA AACTCTGACA	ACT CCT GGAA ACT CCT GGAA GGT GGT T CCA	CACTCTATTC CACTCTATTC CACTCTATTC CACTCTATTC CACTCTATTC CACTCTATTC CACTCTATTC CACTCTATTC CACTCTATTC CACTCTATTC		100 100 100 100 100 100 100 147 105
KF300290 (query) X52626.1 J01263.1 V01163.1 X03004.1 X02980.1 U01132.1 U01131.1 AM905040.1 U01121.1 D0123.1 00123.1	153 153 153 153 153 153 153 157	A GG CACA A GG CACA		CAGAGGTTCG CAGAGGGTTCG CAGAGGGTTCG CAGAGGGTTCG CAGAGGGTTCG CAGAGGGTTCG CAGAGGGTTCG CAGAGGGTTCG CAGAGGTTCG			CAGAATCIIG CAGAATCIIG CAGAATCIIG CAGAATCIIG CAGAATCIIG CAGAATCIIG CAGAATCIIG CAGAATCIIG CAGAATCIIG CAGAATCIIG	240 240 240 240 240 240 207 240 207 240 225
KF308290(query) X57828.1 J0(263.1 V0(163.1 K03004.1	241 241 241 1 241	AAGACTACCG AAGACTACCG AAGACTACCG	TCTTCTGGAG	TTCAGGTCCA			CCTCASCASS CCTCASCASS CCTCASCASS CCTCASCASS	300 300 1 300
U01132.1 U01131.1 AM905840.1 U01121.1 D05383335.1	221 248 208 248 248 227	A A G A C T A C C G A A G A C T A C C G A A M A C T A C C G A A G A C T A C C G A A G A C T A C C G	CTTSTSSAS TCTTSTSSAS TCTTSTSSAS TCTTSTSSAS TCTTSTSSAS	TTCAGETCCA TTCAGETCCA TTCAGETCCA TTCATETCCA TTCATETCCA		CC1 CC 1 TC 1 T CC1 CC 1 CC 1 TC 1 T CC1 CC 1 CC	CCTCAGCAGE CCTCAGCAGE CCTCAGCAGE CCTCAGCAGE CCTCAGCAGE	280 300 267 300 295
KF3003390(query) X52626.1 J01262.1 V01163.1 X02900.1 U01132.1 U01132.1 U01132.1 U01132.1 U01132.1 U01131.1 D053203335.1	308 308 308 8 308 308 308 258 308 258 308 258 308 257			G G CCC A G G CCC A A G G CCC A A A G G CCC A A G G CCC A A G G CCC A A A A A				244 244 244 1 234 234 234 234 234 227 234 220
KF308290(query) K52626.1 J0(282.1 V0183.1 K02900.1 K02900.1 K02900.1 U01132.1 U01132.1 U01132.1 U01132.1 U01132.1 005383-25.1	345 345 1 334 234 234 234 234 234 329 334 320			TCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC				403 403 1 334 334 334 335 334 335 334 335 334 320
K5308290(query) X52628.1 J0(262.1 ¥01963.1 ¥00963.1 X02908.1 U01132.1 U01132.1 U01132.1 U01132.1 U01132.1 D05303:35.1	404 404 404 1 334 334 334 334 335 334 335 334 320		A ACT CG CT A ACT CG CT			A GA GA GA GTA CT A GA GA GA CT A GA GA GTA CT	C C C AC C C C AC AC C C C AC AC C C C AC C C C AC AC C C C AC C C C C	463 463 463 1 391 391 391 391 391 391 391 391 391
KF300290(query) X52626.1 J0(262.1 V0(163.1) X03004.1 X02900.1 U0(132.1) U0(132.1)	454 454 454 1 392 392 372 292			CT CT CA CA CT CT CA CA CT CT CA CA CT CT CA CA CA CA CA CA CA CA CA CA CA CA CA CA CA CA C	CCAGAAAATC CCAGAAAATC CCAGAAAATC CCAGAAAATC CCAGAAAATC CCAGAAAATC CCAGAAAATC	C C T G C A GG A A C C T G C A GG A A		520 520 520 520 445 445 445 445 445 445 445 445 445 44

Fig. 1: Contd.....

³⁴⁹ Agric. Update, **9**(3) Aug., 2014 : 347-353 Hind Agricultural Research and Training Institute

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

8335.1	272		The sectore c	0 G	AGAAAA	CETECAGGAA	CCACINATION	415
1290(query) 28.1 13.1 24.1 104.1 100.1 2.1 11 15840.1 5840.1 1.1 8335.1	520 520 520 14 451 451 420 440 440 405							579 579 579 73 507 510 487 507 543 409 475
1290(query) 26.1 12.1 2.1 04.1 00.1 2.1 1.1 5048.1 1.1 8335.1	500 500 500 74 508 508 508 508 508 544 400 544 400 476		GA CAT GA CAT GA CAT GA CAT GA CAT GA CAT GA CAT GA CAT GA GATT CAA GA GATT CAA CAGAT CAA					\$28 \$25 \$25 \$25 \$25 \$25 \$25 \$25 \$25 \$25 \$25
1290(query) 26.1 12.1 25.1 25.1 25.1 25.1 5040.1 5040.1 1,1 80205.1	637 637 537 535 525 525 525 525 525 525 525 525 525							804 804 804 804 804 804 804 804 804 804
1290(query) 25.1 13.1 2.1 2.1 04.1 10.1 12.1 1.1 5040.1 1.1 5040.1 1.1 8025.1	695 695 695 535 535 535 535 535 535 535 537 506			CAA CCTAC CAA CCTAC				754 754 248 584 584 584 714 585 585
1290(query) 26.1 17.1	755 755 765						SEATHASE!	013 014 014
621 004.1 800.1 32.1 31.1 05040.1 21.1 28.1 38235.1	249 595 598 575 595 715 577 568		GTARGE GRO	AAAA143143				208 605 608 585 605 774 587 576
102390 (query 125.1 153.1 504.1 160.1 22.1 21.1 05040.1 21.1	3 014 815 805 809 809 809 809 809 809 805 805 805 775 587			CACCO GA				873 874 874 388 605 605 825 587
38335.1 16290(query 126.1 183.1 63.1 104.1 100.1 32.1 21.1 05040.1 21.1 21.1 21.1 20325.1	\$79 \$174 \$75 \$65 \$608 \$405 \$405 \$405 \$405 \$405 \$405 \$405 \$405	SCATGATITI SCATGATITI SCATGATITI SCATGATITI						575 933 934 924 429 850 853 850 853 832 821
10790 (guery 826.1 152.1 63.1 1004.1 800.1 32.1 21.1 05048.1 21.1 21.1 21.1 21.1 21.1 21.1 21.1 2	334 335 335 335 335 335 335 335 335 335					GT GAACA GT GAACA	T G A T C T G A A T G A T T C T G A A T G A T T C T G A A T G A T C T G A A	969 970 985 454 585 585 585 585 585 585
6230 County 626.1 63.1 63.1 504.1 800.1 322.1 21.1	971 505 455 587 705 587 587							NU29 1026 1041 524 742 760 722 742

Fig. 1: Contd.....

Agric. Update, **9**(3) Aug, 2014 : 347-353 Hind Agricultural Research and Training Institute

AM905848.1 U01121.1 DQ530335.1	946 669 682						1005 719 741
KF308290(query) X52826.1 J01263.1 V01163.1 X02980.1 X02980.1 U01132.1 U01132.1 U01131.1 AN905846.1 U01121.1 001528335.1	x030 x027 x042 525 743 767 723 767 723 x006 719 742		A C A A TT G G A A C A A TT G G A T C A A C TT G G A T C A A C TT G G A T C A A C TT G G A	AAG <mark>CCAAAAAA</mark> ATAT AAACAGCAAT			106.2 1063 1078 557 779 797 797 797 797 797 797 797 797
K5308290(query) X52828.1 J01263.1 V01853.1 X03004.1 X02900.1 U01132.1 U01132.1 U01131.1 AN905846.1 U01121.1 00530335.1	1063 1064 1079 558 790 790 1066 760 760 1066 762 802	A CTGAGAGGA A CTGAGAGGA A CTGAGAGGA A CTGAGAGGA A CTGAGAGGA A CTGAGAGGA A CTGAGAGAGGA A CTGAGAGAGGA A CTGAGAGAGGA C A GAGATCA TA		TAACTCC TAACTCC TAACTCC TAACTCC TAACTCC TAACTCC TAACTCC TAACTCC TAACTCC TAACTCC TAACTCC TAACTCC TAACTCC	CTTCGAGAC	AA CAG CI AA CAG CI AA CAG CI AA CAG CI AA CAG CC AA CAG CI AA CAG CI AA CAG CI CAG CAG CI CAG CAG CI CAG CAG CI CAG CAG CI	1104 1105 1120 5399 821 839 839 821 823 823 823 823 823 823 825 803 864
KF308290(query) X52828.1 J01263.1 V0163.1 X03004.1 X03004.1 X02900.1 U01132.1 U01132.1 U01132.1 U01132.1 U01123.1 U01123.1	105 105 121 600 822 840 802 822 822 126 804 862					CAAT GAGT CAAT GAGT CAAT GAGT CAAT GAGT CAAT GAGT	1854 1855 1800 8559 837 855 817 184 819 877
KF305290 (query) X52526.1 J01263.1 V0163.1 X03004.1 X02900.1 U01132.1	165 166 161 660 837 855 87		CTGTCGTCT CTGTCGTCTT CTGTCGTCTT ACTGTCGTCT	CGTTAGAAA CGTTAGAAAA CGTTAGAAAA TGGTTAGAAAA	AT CT TAGTAT		1223 1225 1240 719 837 855 817
AM905848.1 U01121.1 DQ538335.1 K5308290(every)	8027 805 819 877						1240 819 8777
R2828.1 40152.1 40152.1 402500.1 50152.1 50152.1 50152.1 50152.1 50152.1 50152.1 505540.1 505540.1 505540.1 505540.1	024 024 725 8255 827 827 827 827 827 827 827 827 827 827						C193 C296 778 875 855 855 855 855 853 853 853 853
B22828.1 J02282.1 J02282.1 J02282.1 K02300.4 S02802.1 S02900.1 S01322.1 J02123.1 J02123.1 J02123.1 J02123.1 J02123.1 J02123.1 J02123.1 J02132.1	2026 C24 7257 817 227 815 817 224 819 817 204 819 77 817 204 204 204 204 204 204 817 775 817 204 817 204 817 204 817 204 817 204 817 204 817 204 817 205 817 204 817 205 817 20 81 81 81 81 81 81 81 81 81 81 81 81 81						C153 C258 C258 778 878 878 878 678 878 878 1000 888 1000 888 1000 888 1000 888 1000 800 8
B22828.1 J07282.1 V07182.1 K02900.1 S071322.1 S071322.1 S071322.1 S07132.1 AM00504.0.1 S07182.1 V07182.1 V07182.1 V07182.1 V07182.1 V07182.1 V07182.1 S02000.1 S07122.1 S0712.1 S	C228 C248 C248 B27 C288 B27 C248 C248 B27 C248 B27 C248 B27 C248 C248 C248 C248 C248 C248 C248 C248						C153 C153 271 825 871 825 871 825 875 853 853 853 827 1088 825 1088 825 1088 825 1088 825 1088 825 1088 825 1088 825 1088 825 1088 825 1088 825 1088 825 1088 855 1088 855 1088 855 1088 855 1088 855 1088 855 1088 855 1088 855 1088 855 1088
SE2828.1 JOC282.1 VOIRE2.1 KO2000.1 SIDIUZ2.1 SIDI							C1233 773 825 871 825 872 875 825 875 825 825 825 825 827 845 845 845 845 845 845 845 845 845 845

Fig. 1: Contd.....

351 *Agric. Update*, **9**(3) Aug., 2014 : 347-353 Hind Agricultural Research and Training Institute ROSY RANI, JITENDER SINGH, PANKAJ KUMAR, PRADEEP SHUKLA AND PRAGATI MISRA

U01122.1 U01131.1 AM905040.1 U01121.1 D0520335.1	107 1027 1451 1015 1015		ACTES TCS ACTES TCS ACTES TCS ACTEC TCS TTTSC TCS	GIATCAATGC GIATCAATGC GIATCAATGC GIATCAATGC GIATCAATGC	A ATAGGAACC A ATAGGAACC CAGAGGAACT TATAGGATCC CAGAGGAATT	CC SCAGE CC SCAGE CC SCAGE CC SCAGE CC SCAGE CC SCAGE	1076 1095 1540 1075 1140
KF305250(query X52526.1 30256.1 V01652.1 KD3004.1 KD2900.1 U01132.1 U01132.1 U01132.1 U01132.1 U01132.1 U01132.1 U01132.1 U01132.5 U01132.5 U01132.5	1507 1509 1524 995 1097 1097 1547 1097 1547 1097 1549 1079 1640						1588 1583 1583 805 1098 105 1098 1098 1098 1098 1098 1098 1098 1098
853032904 (guerry 852628.1 J01262.1 802004.5 802900.3 802900.3 802900.3 802900.3 802900.3 802900.3 900121.1 8001504.0 1 900121.3 02530335.1	1587 1589 1584 995 1098 1098 1098 1098 1590 1590 1590 840		GGTAACG			CACAATGTCA GACAATGTCA GACAATGTCA GACAATGTCA GACAATGTCA GACAATGTCA GACAATGTCA GACAATGTCA GACAATGTCA GACAATGTCA GACAATGTCA GACAATGTCA GACAATGTCA	**************************************
KF3092904 query K526263 J0(252.1 V0(152.1 K02900.1 U0(122.1 U0(122.1 U0(121.1 AM905240.1 U0(121.5 U0(121.5) U0(121.5) U0(121.5)	827 829 844 855 85 85 85 85 85 85 85 85 85 85 85 85	AAGCAGCA AAGCAGCA AAGCAGCA AAGCAGCA AAGCAGCA AAGCAGCA GAGCGAGA AAGCAGCAGCA AAGCAGCA AAGCAGCAGCA AAGCAGCAGCA AAGCAGCAGCA AAGCAGCAGCAGCA AAGCAGCAGCA AAGCAGCAGCAGCA AAGCAGCAGCAGCAGCA AAGCAGCAGCAGCAGCA AAGCAGCAGCAGCAGCAG AAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC		CISCACSCA CISCACSCACSCA CISCACSCACSCA CISCACSCACSCA CISCACSCACSCAC CISCACSCACSCAC CISCACSCACSCAC CISCACSCACSCAC CISCACSCACSCACSCACSCAC CISCACSCACSCACSCAC CISCACSCACSCACSCAC CISCACSCACSCACSCACSCAC CISCACSCACSCACSCAC CISCACSCACSCACSCAC CISCACSCACSCACSCACSCACSCACSCACSCACSCACSC			2535 1538 1733 895 973 895 973 893 853 873 853 873 858 89 1727
KF20022804 every K52825.1 J07252.1 V0162.1 K02000.1 K02900.1 U01122.1 U01122.1 U01122.1 U01121.1	1687 1689 1704 095 174 192 154 174 169 182	C C C C G A C G A C C C C G A C G A C C C C G A C G A C C C C G G A C G A C C C C G G A C G A C C C C G A G A C C C C G A C G A C G C G A G A C C C C C C A C C A C C C C C C A C C A C C C C		C CA CAACA C CA CAACA	A CG AC A CG AC	G C G A T G G A C G C G G A T G G A G C G G A T G G A G C G G A T G G A G C G A T G G A G C G A T G G A G C G A T G G A A C G G A T G G A A	1746 1747 1762 955 032 050 120 120 122 1748 8221
KF300290(query) KF30258.1 J0/253.1 V0/852.1 KD3004.1 KD3900.1 U0/822.1 U0/822.1 U0/822.1 AM9055048.1 U0/823.1 D05292375.1	1747 1747 1752 895 1258 1258 1258 1258 1258 1258 1258 125		ACAGGAACAG TCACCAACAG TCACCAACAG TCACCAACAG TCACCAACAG TCACCAACAG GCACCAACAG GCAACAG GCAACAG GCAACAG GCAACAG				1005 1795 100 835 1055 100 1055 1055 1073 1073 1073
KF300290(query) KF300290(query) J0263.1 V01163.1 K02900.1 K02900.1 U01122.1 U01122.1 U01122.1 U01122.1 U01122.1	1807 1798 1814 895 1257 1302 1254 1257 1773 1275		ACCOLOCATI ACCOLOCATI ACCOLOCATI ACCOLOCATI ACCOLOCATI ACCOLOCATI				1854 1853 1872 435 105 105 105 105 105 105 105 105 105 10
KF300290(query) K525261 J0/2611 V0/0621 K0200041 K02900.1 U0/0221 U0/0221 U0/0221 U0/0221 U0/0221 U0/0221 U0/0221 U0/0221	1057 1050 1050 105 105 105 105 105 105 105 1						1925 195 1933 995 1273 149 100 1072 1773 1093 1429
KF300290(query) X52626.1 J0(263.1	1927	OT THANDAA	925				

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.

