Short Note

Journal of AgriSearch 2(4): 314-318
Peer Reviewed Quarterly

ISSN : 2348-8808 (Print), 2348-8867 (Online)

FabElm_BarcodeDb: barcode database of legumes

JESHIMA KHAN YASIN‡* BHARAT KUMAR MISHRA‡, SAKSHI CHAUDHARY,
NIDHI VERMA, ANIL KUMAR SINGH1
Division of Genomic Resources, ICAR-NBPGR, PUSA campus, New Delhi, India

ARTICLE INFO

Received on :	15.09.2015
Accepted on :	27.11.2015
Published online :	07.12.2015

ABSTRACT

DNA barcoding is aevolving tool usingof chloroplast *rbcL* and *matK* regions exploited as standard molecular barcodes for species identification. Though it is mainly accepted by plant scientists still there is gap in available resources preventing them to use as a barcode. Though this tool is widely used, there is no ready reference source available for barcodes so far.With an objective of developing a database for ready reference, we tried to assemble legume barcode datasets available for legumes. This will facilitate further analyses of these available resources and to resolve the gaps in existing knowledge. This will further facilitate wider use of the available technology.From online resources, barcode loci sequences were retrievedand constructed into a database for ready reference. The database FabElm_BarcodeDb made available at http://jsure.org.in was constructed using available sequence resources.This resource will be of immense value to scientific as well as commercial organisations.

Keywords : Database; DNA Barcoding; identification; Leguminosae; taxononmy

INTRODUCTION

Leguminosae, one among the largest angiosperm family of 730 genera with 19,400 species is of commercial and economicaluse (Mabberley, 1997). Legumes are crops of food, fodder, fibre, timber, fuel, oils, chemicals and medicines contributing to food and nutritional security (Ramya *et al.*, 2013; Singh *et al.*, 2009; Yasin, 2015). Legumes grow in a wider habitat from tropical to temperate regions around the globe (Rundel, 1989; Yasin *et al.*, 2014); known to fix atmospheric nitrogen (Singh *et al.*, 2012). Legumes are terrestrial nitrogen fixing agents (Sprent, 1994).

Maturase (*matK)* functions as a splicing agent (Neuhaus and Link, 1987). The Internal Transcribed Spacer (ITS) section of nuclear ribosomal DNA is used inphylogeny investigations of plants. Several nuclear, chloroplast and mitochondrial loci were exploited to correlate with species identification. Amongst these, *rbcL, matK and trnK* has were used in systematics (Neuhausand Link, 1987).Whereas, *matK*is systematically linked in evolution of land plants and in embryophytes (Turmel*et al.*, 2006). *MatK*is a standard molecular marker for phylogenetic studies and has been amplified from tens of thousands of plant species (CBOL, 2009). The splicing gene, *matK* is absent in parasitic species of the genus Cuscuta (Funk *et al.*, 2007; McNeal *et al.*, 2009) and orchid (Delannoy *et al.*, 2011 indicating flaws in usage of this loci as a barcode. Hence, assembling all these sequences in a database format will further facilitate in depth analyses to findout whether these loci are suitable for further usage as barcodes.

A biological database is a collection of data that is organized so that its contents can easily be accessed, managed, and updated. Bioinformatics is the application of information technology to store, organize and analyse the vast amount of biological data which is available in the form of sequences and structures of proteins and nucleic acids. Databases in general are classified as primary, secondary and composite databases. A primary database contains information about the sequence or structure. A secondary database contains information which is derived from the primary databases. The information can be like conserved sequence, signature sequence and active site residues. Composite database combines the information from various primary databases. Biological databases can contain information related to sequence, structure pathway and tools for further analysis. Here with an objective of developing a primary database a legume barcode database presented (Ioannis *et al.*, 2002;

[1] ICAR-Research Complex for Eastern Region Patna, India
‡authors contributed equally
*corresponding author Email : yasinlab.icar@gmail.com

Susumu *et al.*,2002; Hsu *et al.*,2011).

MATERIALS AND METHODS

DNA sequence data collection

The complete set of reported Leguminosae species list were retrieved from National Centre for Biotechnology information (NCBI) database. A total of 10337 nucleotide sequences and popsets were retrieved from NCBI (https://www.ncbi.nlm.nih.gov/nuccore/?term=fabaceae+matk) to construct a database. The sequences were analysed by Minimum Evolution Tree method through MEGA (Tamura *et al.,* 2013). Statistical parameters were set to default for universal phylogenetic reconstruction. Additionally, the evolutionary clock time for *matK* sequences of different legumes was estimated by MEGA.

RESULTS AND DISCUSSION

With an objective to developing a database for ready reference barcode loci sequences accessible from free resources to construct a knowledge base in the present investigation and constructed into a database. Of the total expected >19,000 sequences for known species of Fabaceae, only non-redundant sequences were available from the database indicating the larger gap in available sequence resources. The database is periodically reviewed and improved further. Few images from the database view have been enclosed in this work. From our initial analyses we found the acacia group more vulnerable with lot of gaps and errors

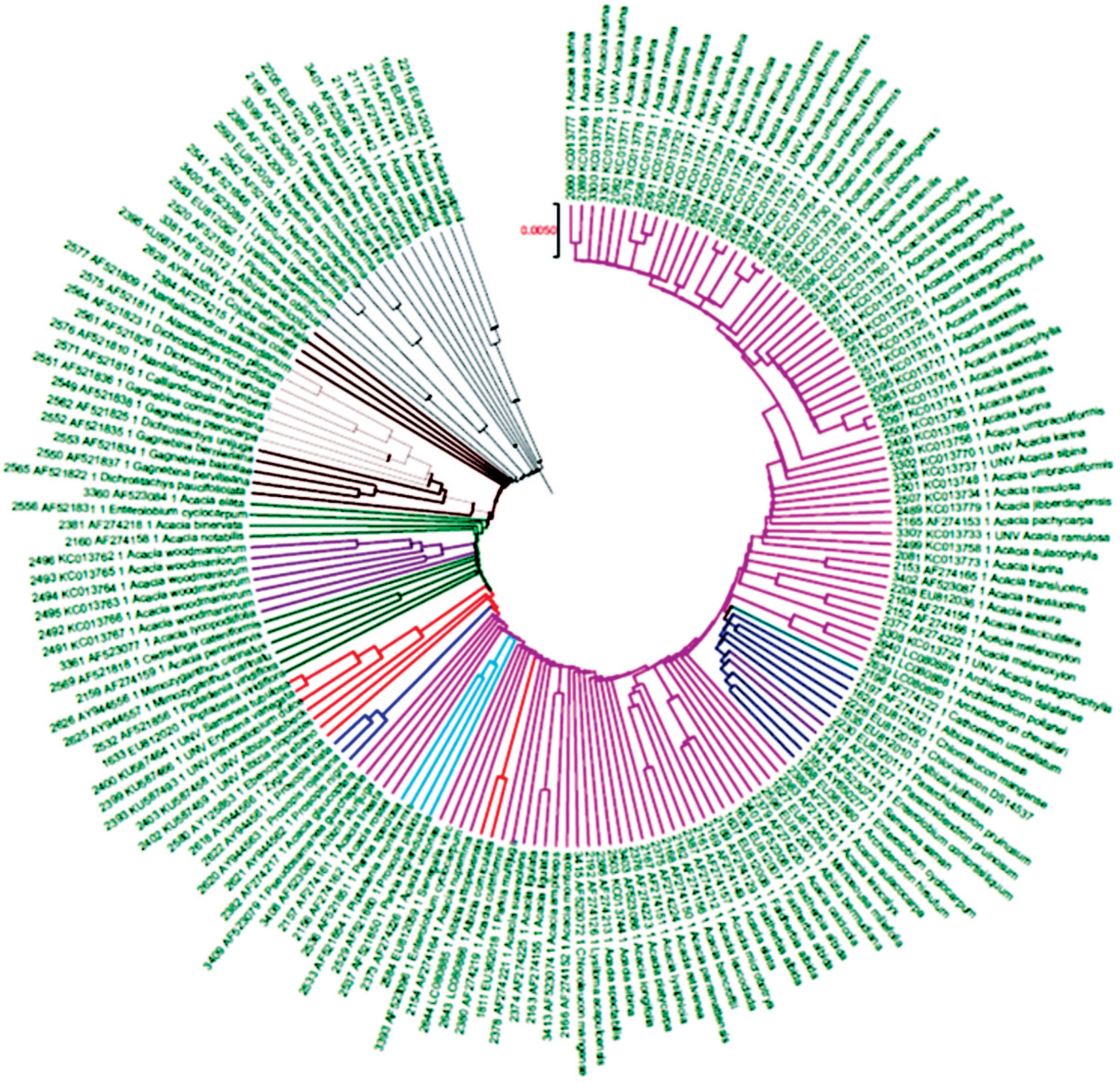

Fig 1: *Acacia* group : The clustering analyses of *Acacia spp.*using single barcode *matK* loci depicting the gap in available information. The single loci may not be sufficient to resolve this genus. Different colour nodes indicates presence of different genus within the cluster

found in the available resources (Fig. 1). Further this has been supported by clustering of diverse genus within Astragalus group of plants (Fig. 2).

We represent FabElm_BarcodeDb (Fig. 3), a barcode database for scrupulousidentificationof legumes using short DNA sequence. These barcodes will help in identifying the unknown plant samples, phylogenetic analyses of selected sequences and retrieval of barcode of expected plant species. FabElm_BarcodeDb can be accessed through user friendly web interface (will be made available at https://jsure.org.in/) that provide search options like genus, species, sequence id etc. Such a standardised identification method will be useful for mapping, species identification and sequence analysis.

Plant barcodes are typically used in an integrative approach with additional information for identification of new species or unknown samples. In earlier studies, unexpected sequence divergence has led to re-examination of morphological/ecological variation, which has then resulted in formal detection of new taxa. As such there is lack of specific database available for legume barcodes. The present approach will maximize the utility of the database for researchers involved in diverse aspects of legume barcode and ensures access to the most relevant high-quality datasets. The user friendly database will be having options to add new sequences uploaded in major sequence submission portals.

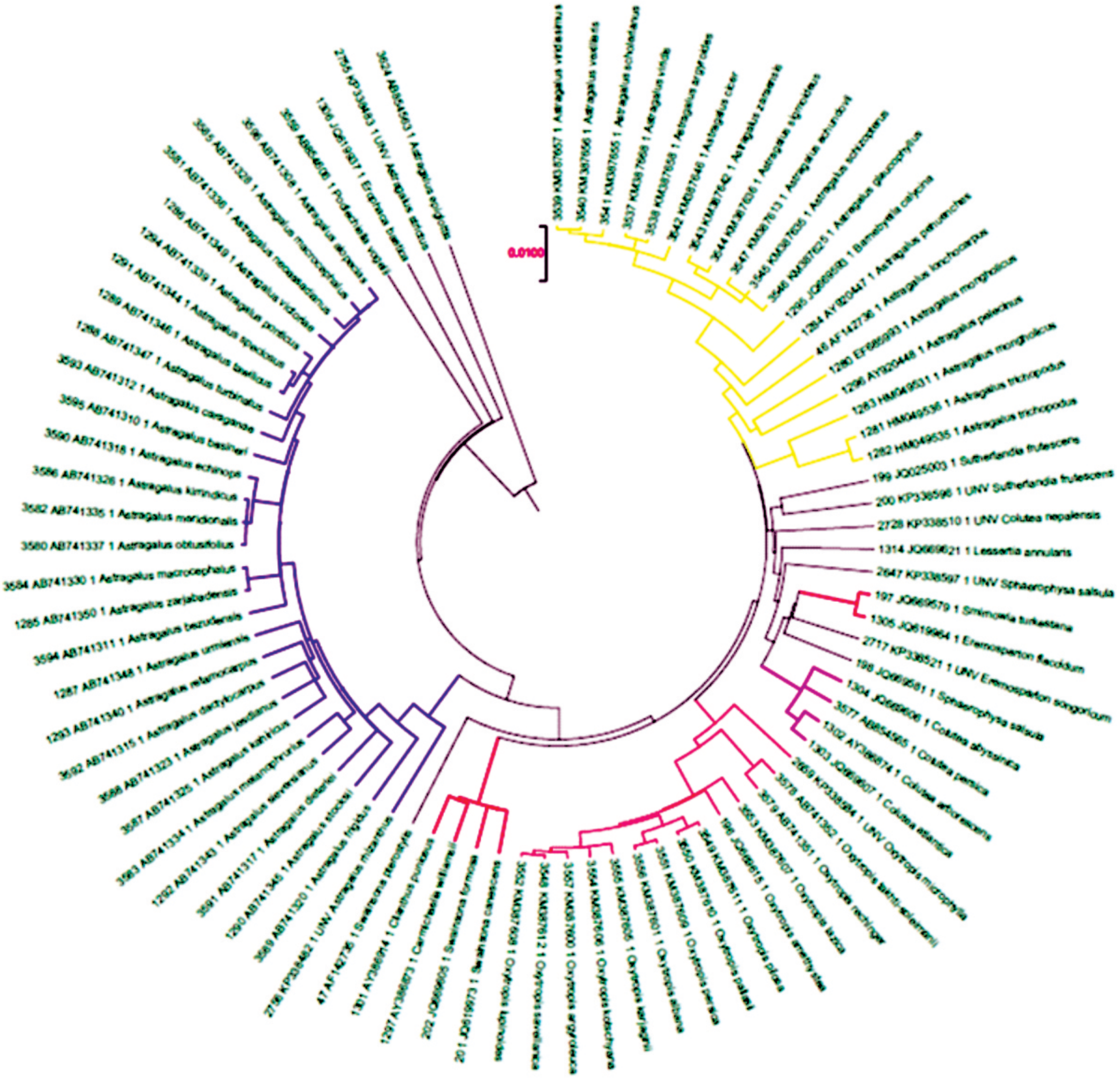
0.0100
3539 KM387657 1 Astragalus viridissimus
3540 KM387656 1 Astragalus vestitus
3541 KM387655 1 Astragalus schuleanus
3537 KM387659 1 Astragalus viridis
3538 KM387658 1 Astragalus argyroides
3542 KM387646 1 Astragalus cicer
3543 KM387642 1 Astragalus zaraensis
3544 KM387636 1 Astragalus sigmoideus
3547 KM387613 1 Astragalus achundovi
3545 KM387635 1 Astragalus schizopterus
3546 KM387625 1 Astragalus glaucophyllus
1295 JQ669580 1 Barnebyella calycina
1284 AY920447 1 Astragalus penduliflorus
48 AF142736 1 Astragalus lonchocarpus
1280 EF685993 1 Astragalus mongholicus
1296 AY920448 1 Astragalus pelecinus
1283 HM049531 1 Astragalus mongholicus
1281 HM049536 1 Astragalus trichopodus
1282 HM049535 1 Astragalus trichopodus
199 JQ025003 1 Sutherlandia frutescens
200 KP338596 1 UNV Sutherlandia frutescens
2728 KP338510 1 UNV Colutea nepalensis
1314 JQ669621 1 Lessertia annularis
2647 KP338597 1 UNV Sphaerophysa salsula
197 JQ669579 1 Smirnowia turkestana
1305 JQ619964 1 Eremosparton flaccidum
2717 KP338521 1 UNV Eremosparton songoricum
198 JQ669581 1 Sphaerophysa salsula
1304 JQ669606 1 Colutea abyssinica
3577 AB854565 1 Colutea persica
1302 AY386874 1 Colutea arborescens
1303 JQ669607 1 Colutea atlantica
2669 KP338594 1 UNV Oxytropis microphylla
3578 AB741352 1 Oxytropis lambertii
3579 AB741351 1 Oxytropis rechingeri
3553 KM387607 1 Oxytropis lazica
196 JQ669615 1 Oxytropis amethystea
3549 KM387611 1 Oxytropis pilosa
3550 KM387610 1 Oxytropis pallasii
3551 KM387609 1 Oxytropis persica
3556 KM387601 1 Oxytropis albana
3555 KM387605 1 Oxytropis karjagini
3554 KM387606 1 Oxytropis kotschyana
3557 KM387600 1 Oxytropis argyroleuca
3548 KM387612 1 Oxytropis savellanica
3552 KM387608 1 Oxytropis lupinoides
201 JQ619973 1 Swainsona canescens
202 JQ669605 1 Swainsona formosa
1297 AY386873 1 Carmichaelia williamsii
1301 AY386914 1 Clianthus puniceus
47 AF142735 1 Swainsona pterostylis
2756 KP338482 1 UNV Astragalus mizanthus
3589 AB741320 1 Astragalus frigidus
1290 AB741345 1 Astragalus stocksii
3591 AB741317 1 Astragalus diederichii
1292 AB741343 1 Astragalus semnanensis
3583 AB741334 1 Astragalus melanophrurius
3587 AB741325 1 Astragalus kahiricus
3588 AB741323 1 Astragalus jesdianus
3592 AB741315 1 Astragalus dactylocarpus
1293 AB741340 1 Astragalus retamocarpus
1287 AB741348 1 Astragalus urmiensis
3594 AB741311 1 Astragalus bezudensis
1285 AB741350 1 Astragalus zarjabadensis
3584 AB741330 1 Astragalus macrocephalus
3580 AB741337 1 Astragalus obtusifolius
3582 AB741335 1 Astragalus meridionalis
3586 AB741326 1 Astragalus kirrindicus
3590 AB741318 1 Astragalus echinops
3595 AB741310 1 Astragalus basineri
3593 AB741312 1 Astragalus caraganae
1288 AB741347 1 Astragalus turbinatus
1289 AB741346 1 Astragalus tawilicus
1291 AB741344 1 Astragalus speciosus
1294 AB741339 1 Astragalus ponticus
1286 AB741349 1 Astragalus victoriae
3581 AB741336 1 Astragalus neomozaffarianus
3585 AB741328 1 Astragalus macrocephalus
3596 AB741308 1 Astragalus alopecias
3559 AB854606 1 Podlechiella vogelii
1306 JQ619937 1 Erophaca baetica
2755 KP338483 1 UNV Astragalus strictus
3624 AB854563 1 Astragalus epiglottis

Fig. 2: *Astragalus* group: the clustering of astragalus with other genus indicates

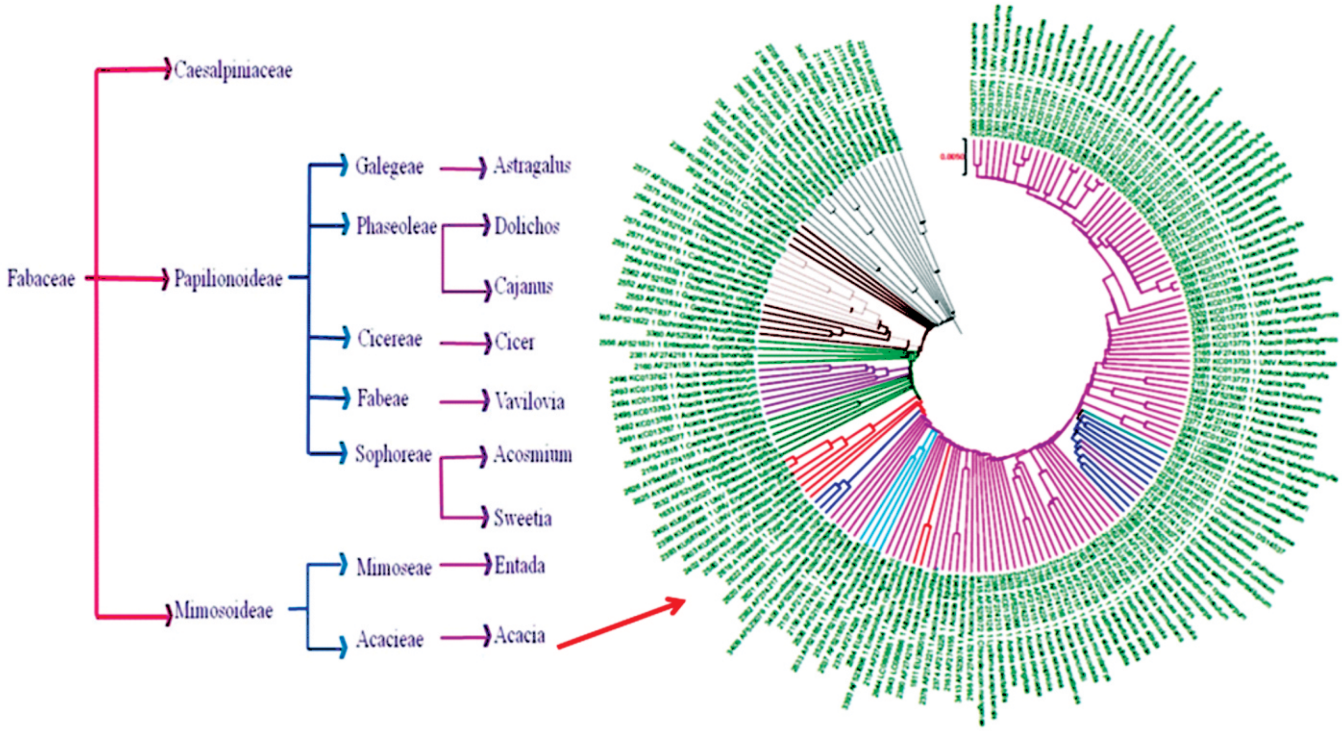

Fig.3: Overview from database homepage – The homepage view of the database; Of the total sequences we started with analyses of the depicted groups. The cluster diagram indicates the gap in Acacia

ACKNOWLEDGEMENTS

We are grateful to Director, ICAR-NBPGR for providing facilities and Indian Council of Agricultural Research, New Delhi India, for financial support to conduct this research work.

REFERENCES

Delannoy E, Fujii S, des Francs-Small CC, Brundrett M and Small I. 2011. Rampant gene loss in the underground orchid Rhizanthellagardneri highlights evolutionary constraints on plastid genomes. *Molecular Biology and Evolution* **8**: 2077-2086.

Funk HT, Berg S, Krupinska K, Maier UG and Krause K.2007. Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, Cuscutareflexa and Cuscutagronovii. *BMC Plant Biology*.**7**:45.

Hsu SD, Lin FM, Wu WY, Liang C, Huang WC, Chan WL, Tsai WT, Chen GZ, Lee CJ, Chiu CM, Chien CH , Wu MC, Huang CY, Tsou AP and Huang HD.2011. miRTarBase: a database curates experimentally validated microRNA–target interactions. *Nucleic Acids Research* **39**(1):D163-9.

Ioannis X, Lukasz S, Xiaoqun JD, Patrick H, Sul-Min K and David E. 2002. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Research* **30**(1):303-5.

Tamura K, Stecher G, Peterson D,Filipski A, and Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0 *Molecular Biology and Evolution.* **30**(12): 2725–2729.

Mabberley DJ.1997. The plant-book: a portable dictionary of the vascular plants. Cambridge university press.

McNeal JR, Kuehl JV, Boore JL and Leebens-Mack J. 2009. Parallel loss of plastid introns and their maturase in the genus Cuscuta. *PLoS One* **4**:e5982.

Neuhaus H and Link G.1987. The chloroplast tRNA Lys (UUU) gene from mustard (Sinapisalba) contains a class II intron potentially coding for a maturase-related polypeptide. *Current Genetics* **11**:251-257.

Ramya KT, Fiyaz RA and Yasin JK. 2013. SMART agriculture for nutritional security. *Current Science*.**105**:1458.

Rundel PW.1989. Ecological success in relation to plant form and function in the woody legumes.Advances in Legume Biology Monogr. Syst. Bot. *Missouri Bot. Gard*.**29**:377-98.

Singh AK, Dimree S K, Khan MA and Upadhyaya Ashutosh.2009.Agronomic Evaluation of Faba Bean (*Vicia faba* L.) Performance under Impending Climate Change Situation. National Symposium on Recent Global Developments in the Management of Plant Genetic Resources, 17-18 December 2009. Souvenir and Abstracts. Indian Society of Plant Genetic Resources, New Delhi.p185

Singh AK, Bhatt BP, Sunram PK, Kumar S, Bharati RC, Chandra N and Rai M.2012. Study of Site Specific Nutrients Management of Cowpea Seed Production and Their Effect on Soil Nutrient Status. *Journal of Agril. Sci.* **4**(10): 191-198.

Sprent JI.1994. Nitrogen acquisition systems in the Leguminosae. In: Sprent JI, McKey D, editors. Advances in Legume Systematics 5. The Nitrogen Factor: Royal Botanic Gardens, Kew 1–16.

Susumu G, Yasushi O, Masahiro H, Takaaki N and Minoru K.2002. LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acids Research* **30**(1):402-4.

Turmel M, Otis C and Lemieux C.2006. The chloroplast genome sequence of Chara vulgaris sheds new light into the closest green algal relatives of land plants. *Molecular Biology and Evolution* **23**:1324-38.

Yasin JK,Nizar MA, Rajkumar S, Verma M, Verma N, Pandey S, Tiwari SK, Radhamani J. 2014. Existence of alternate defense mechanisms for combating moisture stress in horse gram [Macrotyloma uniflorum (Lam.) Verdc.]. *Legume Research*.**37**(2):145-154. DOI: 10.5958/j.0976-0571.37.2.022

Yasin JK. 2015. “SMART” to success: not just combinations and permutations. http://comments.sciencemag.org/content/10.1126/*science*.1254135.

Citation:

Yasin JK, Mishra BK, Chaudhary S, Verma N and Singh AK. 2015. FabElm_BarcodeDb: barcode database of legumes. *Journal of AgriSearch*2 (4): 314-318