

## Conseq Tool: A Tool to Find Conserved Region in Protein Sequences of Leguminosae Family

Sagar S. Patel<sup>1\*</sup>, Dipti B Shah<sup>1</sup> and Hetalkumar J. Panchal<sup>2</sup>

<sup>1</sup>G. H. Patel Post Graduate Department of Computer Science and Technology,  
Sardar Patel University, Vallabh Vidyanagar, Gujarat-388120, India.

<sup>2</sup>Gujarat Agricultural Biotechnology Institute,  
Navsari Agricultural University, Surat, Gujarat- 395007, India.

\*Corresponding author: Email: sgr308@gmail.com Tel: (M) 09825510747

[Received-12/06/2014, Accepted-01/08/2014]

### ABSTRACT:

Bioinformatics is currently emerging and fast growing field of applied sciences. Leguminosae family is one of the largest families that contain thousands of species of Plants, Herbs, Shrubs and Trees worldwide. There are three subfamilies of Leguminosae family which are Fabaceae (Papilionaceae), Caesalpiniaceae and Mimosaceae. In this paper authors have developed a tool in which user has to provide protein sequences of rbcL or matK as input and the tool as a result will provide, user's sequence followed by its length, Conserved sequence found in that sequence and based on this conserved sequence the species may be included in one of three subfamilies of Leguminosae family and is it of rbcL or matK protein sequence. The result of this tool is based on few criteria based on our previous study and after we found few conserved sequence in each subfamily which is described in this paper.

**Keywords:** Leguminosae family, Bioinformatics, rbcL, matK.

### [I] INTRODUCTION

#### 1.1. Leguminosae Family

Leguminosae family contains species of Plants, Herbs, Shrubs, and Trees. Legumes are used as crops, forages and green manures; they also synthesize a wide range of natural products such as flavours, drugs, poisons and dyes<sup>[17]</sup>. Legumes are able to convert atmospheric nitrogen into nitrogenous compounds useful to plants. This is achieved by the presence of root nodules containing bacteria of the genus *Rhizobium*. These bacteria have a symbiotic relationship with

Legumes, fixing free nitrogen for the plants; in return legumes supply the bacteria with a source of fixed carbon produced by photosynthesis. This enables many legumes to survive and compete effectively in nitrogen poor conditions. Leguminosae family is further classified into three subfamilies; Fabaceae (Papilionaceae), Caesalpiniaceae and Mimosaceae<sup>[12],[13]</sup>.

#### 1.2 rbcL gene

The most common gene used for plant phylogenetic analyses is the plastid-encoded rbcL

gene. This single copy gene is approximately 1430 base pairs in length and is free from length mutations except at the far 3' end. It has fairly conservative rate of evolution. The function of the *rbcL* gene is to code for the large subunit of ribulose 1, 5 bisphosphate carboxylase/oxygenase (RUBISCO or RuBPCase). *rbcL* gene is involved in following pathways, Glyoxylate and dicarboxylate metabolism, Carbon fixation in photosynthetic organisms and Carbon metabolism [2],[3],[12].

### 1.3 matK gene

The *matK* gene, formerly known as *orfK*, is emerging as yet another gene with potential contributions to plant molecular systematics and evolution. The gene, ~1500 base pairs (bp), is located within the intron of the chloroplast gene *trnK*, on the large single-copy section adjacent to the inverted repeat. Further, the molecular information generated from *matK* has been used to resolve phylogenetic relationships from shallow to deep taxonomic levels [2],[3],[12].

### 1.4. PHP Language

PHP is an open-source server-side scripting language designed for Web development to produce dynamic Web pages. It is one of the first developed server-side scripting languages to be embedded into an HTML source document rather than calling an external file to process data. The code is interpreted by a Web server with a PHP processor module which generates the resulting Web page. We have written script in PHP language as it has many features. PHP is an object-oriented language - In PHP you can define your own reusable data structures called objects as well as define their attributes (properties) and things they can do (methods). You can also create relationships between various objects and data structures [12],[14],[15].

### 1.5. Dreamweaver Software

It is a web design and development application developed by Adobe Systems that provides a visual WYSIWYG editor and a code editor with standard features such as syntax highlighting, code

completion, and code collapsing as well as more sophisticated features such as real-time syntax checking and code introspection for generating code hints to assist the user in writing code. We have connected our database in this software and written PHP script to fetch data from database. As Word is to paper documents, Dreamweaver is specifically for creating high quality web sites, and supplies you with a variety of tools to do so. And as with any new application, Dreamweaver has its own set up and tools [12],[16].

## [II] MATERIALS AND METHODS

This tool is based on the study which focuses on species of Leguminosae Family which are restricted within Gujarat state in India [1],[4],[6].

After compiling data, authors got useful data of both *rbcL* and *matK* gene which could be further used for analysis purposes [4]. After analysis by Multiple Sequence Alignment of *rbcL* and *matK* gene sequences, authors found that there are few amino acids which are found conserved in particular subfamily as shown in Table 1 and 2.

**Table 1.** Information of *rbcL* Conserved sequences.

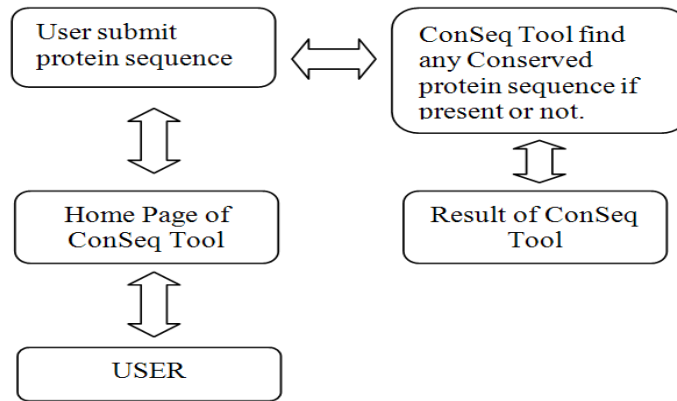
Sub family	<i>rbcL</i> gene <i>Conserved Sequence</i>
Fabaceae	ASKWSPELAAACEVWK
Caesalpinaceae	GFKAGVKDYK
Mimosaceae	RGGLDFTKDDENVNSQPFMR

**Table 2.** Information of *matK* Conserved sequences.

Sub family	<i>matK</i> gene <i>Conserved Sequence</i>
Fabaceae	FSLLVKRL
Caesalpinaceae	IRS
Mimosaceae	QTLRYWVKDA

As shown in Table 1 and 2 that each subfamily has conserved protein sequences so that one tool designed to find these conserved sequences in unknown *rbcL* or *matK* protein sequence. ConSeq Tool is designed in php language and web page designing done in Dreamweaver Software.

**Figure 1.** Flow chart of ConSeq Tool



### [III] RESULTS

#### 3.1. Home Page of ConSeq Tool

Figure 2 shows the homepage of ConSeq Tool. User has to paste *rbcl* or *matK* protein sequence in the box and click on Submit button as shown in Figure 3.

**ConSeq Tool**

**About this tool:**

- This tool is based on the conserved sequence which is present in *rbcl* or *matK* protein sequence.
- User has to paste *rbcl* or *matK* protein sequence and click on submit.
- If your sequence has conserved sequence which is present in all other Leguminosae Family members then your species may be fall near by particular Leguminosae Family's one of three sub family.
- This tool also gives you length of your sequence and particular conserved sequence which is found in your sequence if any.
- If your sequence don't have any conserved sequence then it will show that your sequence may not be similar to any Leguminosae Family's species.

*Note: [Please read this document before you use tool.](#)*

Paste your sequence :

**Figure 2.** Home Page of ConSeq Tool.

### ConSeq Tool

#### About this tool:

- This tool is based on the conserved sequence which is present in *rbcl* or *matK* protein sequence.
- User has to paste *rbcl* or *matK* protein sequence and click on submit.
- If your sequence has conserved sequence which is present in all other Leguminosae Family members then your species may be fall near by particular Leguminosae Family's one of three sub family.
- This tool also gives you length of your sequence and particular conserved sequence which is found in your sequence if any.
- If your sequence don't have any conserved sequence then it will show that your sequence may not be similar to any Leguminosae Family's species.

*Note: Please read this document before you use tool.*

Paste your sequence :

```
GVKDYKLYTYTPDYKTKDIDLAAFRVTPQGPVPEEAGA AAVAESSTGTWTTVWTDGLTSLDRYKGRCY
HIEPVAGEESQFIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRIPNAYVKTQGGPHGI
QVERDKLNKYGRPLLGTCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENVNSQPFMRWRDRFLCAEL
YKAQAEETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHDYLTGGFTANTSLSHYCRDNGLLHHIHR
AMHAVIDRQKNHGMHFRVLAKALRLSGGDHHSVTVGKLEGEREITLGFVDLLRDDFVEKDRSRGIYFT
QDWVSLPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLGHPWGNAPGAVANRVALEACVQARNEGR
DLASEGNQIIREASKWSPELAAACEVWKEIKFE
```

Figure 3. User has to paste protein sequence in ConSeq Tool.

### 3.2. ConSeq Tool Result

Figure 4 shows the result of ConSeq Tool. If your protein sequence has any conserved protein sequence then it shows the result as shown in figure 4 which includes your submitted sequence followed by Length of your sequence, Conserved Sequence, and as per the presence of conserved sequence it may be included to particular one of three subfamilies of Leguminosae Family as this conserved sequence is similar which is found in that particular one of three subfamilies of Leguminosae Family and if your sequence do not has any conserved sequence then it shows the results as shown in figure 5.

### ConSeq Tool - RESULT

- **Your Sequence is:**  
 GVKDYKLYTYTPDYKTKDIDLAAFRVTPQGPVPEEAGA AAVAESSTGTWTTVWTDGLTSLDRYKGRCYHIEPVAGEESQFIAYVAYPLDLFEEGSVTN  
 MFTSIVGNVFGFKALRALRLEDLRIPNAYVKTQGGPHGIQVERDKLNKYGRPLLGTCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENVNSQPFMRWR  
 DRFLCAEALYKAQAEETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHDYLTGGFTANTSLSHYCRDNGLLHHIHRAMHAVIDRQKNHGMHFRVLA  
 KALRLSGGDHHSVTVGKLEGEREITLGFVDLLRDDFVEKDRSRGIYFTQDWVSLPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLGHPWGNAP  
 GAVANRVALEACVQARNEGRDLASEGNQIIREASKWSPELAAACEVWKEIKFE
- Length of Your Sequence is: **453**
- **Your Species' Conserved Sequence is:** ASKWSPELAAACEVWK
- Your Species may be included into **Fabaceae** subfamily as it contains ASKWSPELAAACEVWK which is found conserved in this particular **Fabaceae** subfamily.
- Your Sequence is of **rbcl** gene protein sequence.

Figure 4. Result of ConSeq Tool.



**Figure 5.** Result of ConSeq Tool if no conserved sequence found.

#### [IV] DISCUSSION

This study based on the species of Leguminosae family which are found in Gujarat state, India. Further only rbcL and matK protein sequence data considered for analysis and after Multiple sequence alignment analysis few conserved sequence found in each Leguminosae Family's subfamily which are; For Fabaceae subfamily 38%, for Caesalpinaceae 60% and for Mimosaceae 54% species found conserved sequences as shown in Table 1 and 2. There is no such tool which finds conserved sequence in rbcL or matK protein sequence of Leguminosae Family species and as per our study creation of this tool is necessary to get more information about conserved sequences of rbcL and matK protein sequence. This ConSeq Tool would be very useful to many people to find conserved sequence in Leguminosae Family species.

#### ACKNOWLEDGEMENT

We are heartily thankful to Prof. (Dr.) P.V. Virparia, Director, GDCST, Sardar Patel

University, Vallabh Vidyanagar, for providing us facilities for the research work.

#### REFERENCES

1. G. L. Shah (1978): *Flora of Gujarat State*. Publ. by Sardar Patel University, Vallabh Vidyanagar, Anand, India.
2. Hayashi K. Kawano S. 2000. Molecular systematics of Liliaceae and allied genera (Liliaceae): phylogenetic relationships among Liliaceae and related genera based on the rbcL and matK gene sequence data. *Plant Species Biology* 15: 73-93.
3. Hilu KW, Liang H: The matK gene: sequence variation and application in plant systematics. *American Journal of Botany* 1997, 84:830-839.
4. Sagar Patel, and Hetal Kumar Panchal. "Bioinformatics Information of Leguminosae Family in Gujarat State." *International Journal of Agriculture, Environment & Biotechnology* 7.1 (2014).
5. Sagar Patel, Hetalkumar Panchal., 2014. Evolutionary studies of few species belonging to Leguminosae family based on RBCL gene. *Discovery*, 2014, 9(22), 38-50.
6. Sagar Patel, Panchal H., 2013. Leguminobase: A Tool To Get Information Of Some Leguminosae Family Members From NCBI Database in *Journal*

- of Advanced Bioinformatics Applications and Research: Vol 4, Issue3, 2013, Pages. 54-59. ISSN 0976-2604. Online ISSN 2278 – 6007.
7. Sagar Patel, Panchal H., Anjaria K., DNA Sequence analysis by ORF FINDER & GENOMATIX Tool: Bioinformatics Analysis of some tree species of Leguminosae Family. Publication Year: 2012, Page(s): 922- 926. Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Philadelphia, USA, Oct 4-7, 2012; E-ISBN: 978-1-4673-2744-2; Print ISBN: 978-1-4673-2746-6.
  8. Sagar Patel, Panchal H., Anjaria K., Phylogenetic analysis of some leguminous trees using CLUSTALW2 Bioinformatics Tool. Publication Year: 2012, Page(s): 917- 921. Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Philadelphia, USA, Oct 4-7, 2012; E-ISBN: 978-1-4673-2744-2; Print ISBN: 978-1-4673-2746-6.
  9. Sagar Patel, Panchal H., Smart J., Anjaria K., 2013. Distribution of Leguminosae family members in Gujarat State of India: Bioinformatics Approach in International Journal of Computer Science and Management Research, Pages- 2184-2189 Vol 2 Issue 4 April 2013, ISSN 2278-733X
  10. Sagar Patel, Panchal H., Smart J., Anjaria K., 2013. Species Information Retrieval Tool: A Bioinformatics tool for Leguminosae family in International Journal of Bioinformatics and Biological Science: Vol.1 n.2 Pages.187-194 June, 2013 Print ISSN 2319-5169
  11. <http://www.ncbi.nlm.nih.gov/>
  12. <http://www.en.wikipedia.org>
  13. <http://www.theplantlist.org/browse/A/Leguminosae>
  14. <http://www.w3schools.com>
  15. <http://www.apachefriends.org/en/xampp.html>
  16. <http://www.adobe.com/in/products/dreamweaver.html>
  17. <http://www.ildis.org/>