Journal of Advanced Bioinformatics Applications and Research
ISSN 0976-2604.Online ISSN 2278 – 6007
Vol 4, Issue3, 2013, pp54-59
http://www.bipublication.com

LEGUMINOBASE: A TOOL TO GET INFORMATION OF SOME LEGUMINOSAE FAMILY MEMBERS FROM NCBI DATABASE

Sagar S. Patel* **and Hetalkumar J. Panchal**

G. H. Patel Post Graduate Department of Computer Science and Technology,
Sardar Patel University, Vallabh Vidyanagar, Gujarat-388120, India.
***Corresponding author:** Email: sgr308@gmail.com Tel: (M) 09825510747

[Received-18/09/2013, Accepted-05/10/2013]

ABSTRACT:

Bioinformatics is rapidly developing and growing field in current era. It includes the computational analysis of biological data, consisting of the information stored in the form of DNA, Protein and Genome sequences in various biological databases. Leguminosae family is one of the largest families that contain thousands of species of Plants, Herbs, Shrubs and Trees worldwide. There are more than 250 species of this family which are found in Gujarat state of India, out of which we only got information of around 149 species' from NCBI database. There are three subfamilies of Leguminosae family which are Fabaceae (Papilionaceae), Caesalpiniaceae and Mimosaeae. In this paper authors have developed one tool in which user has to select respective option and after click on Submit button it directly fetch various information from NCBI database like it's Species Name, PubMed, Pubmed Central, Nucleotide, SRA, PopSet, Genome, BioProject, Protein and Structure information of particular species of Leguminosae family.

Keywords: Leguminobase, Leguminosae family, Bioinformatics, NCBI, Biological database.

[I] INTRODUCTION

Leguminosae family is one of the important families which consist of diverse numbers of species which are very beneficial to living organisms. Legumes are useful to convert atmospheric nitrogen into nitrogenous compounds. 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 [9]. This enables many legumes to survive in atmosphere with less percentage of nitrogen. Leguminosae family is further classified into three subfamilies; 1. Fabaceae (Papilionaceae), 2. Caesalpiniaceae and 3. Mimosaeae.

1.1. NCBI (The National Center for Biotechnology Information)

The National Center for Biotechnology Information (NCBI) is part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health. The NCBI houses a series of databases relevant to biotechnology and biomedicine. Major databases include GenBank for DNA sequences, Protein, Genome, EST etc. All these databases are available online through the Entrez search engine [12]. http://www.ncbi.nlm.nih.gov/. The National Center for Biotechnology Information (NCBI 2001) defines bioinformatics as:"Bioinformatics is the field of science in which biology, computer science, and information technology merge into a single discipline. There are three important sub-disciplines within bioinformatics: the development of new algorithms and statistics which assess relationships among members of large data sets, the analysis and interpretation of various types of data including nucleotide and amino acid sequences, protein domains and the development and implementation of tools that enable efficient access and management of different types of information."

1.2. PubMed

PubMed comprises more than 23 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites [20].

1.3. PumMed Central (PMC)

PMC is a free full-text archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health's National Library of Medicine (NIH/NLM) [21].

1.4. DNA (Deoxyribonucleic acid) / Nucleotide

The Deoxyribonucleic acid (DNA) is a molecule that encodes the genetic instructions used in the development and functioning of all known living organisms and many viruses. Along with RNA and proteins, DNA is one of the three major macromolecules essential for all known forms of life. Genetic information is encoded as a sequence of nucleotides (guanine, adenine, thymine, and cytosine) recorded using the letters G, A, T, and C. Most DNA molecules are double-stranded helices, consisting of two long polymers of simple units called nucleotides, molecules with backbones made of alternating sugars (deoxyribose) and phosphate groups (related to phosphoric acid), with the nucleobases (G, A, T, C) attached to the sugars. DNA is well-suited for biological information storage, since the DNA backbone is resistant to cleavage and the double-stranded structure provides the molecule with a built-in duplicate of the encoded information [9, 22].

1.5. SRA

The Sequence Read Archive (SRA) stores raw sequencing data from the next generation of sequencing platforms including Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD® System, Helicos Heliscope®, Complete Genomics®, and Pacific Biosciences SMRT® [23].

1.6. Popset

A PopSet is a set of DNA sequences that have been collected to analyse the evolutionary relatedness of a population. The population could originate from different members of the same species, or from organisms from different species [24].

1.7. Genome

The Genome is the entirety of an organism's hereditary information. It is encoded either in DNA or RNA (for many types of viruses). The genome includes both the genes and the non-coding sequences of the DNA/RNA [25].

1.8. BioProject

A BioProject is a collection of biological data related to a single initiative, originating from a single organization or from a consortium. A BioProject record provides users a single place to find links to the diverse data types generated for that project [26].

1.9. Protein

Proteins are large biological molecules consisting of one or more chains of amino acids. Proteins perform a vast array of functions within living organisms, including catalyzing metabolic reactions, replicating DNA, responding to stimuli, and transporting molecules from one location to another. Proteins differ from one another primarily in their sequence of amino acids, which is dictated by the nucleotide sequence of their genes, and which usually results in folding of the protein into a specific three-dimensional structure that determines its activity [9, 27].

1.10. Structure

Three dimensional structures provide a wealth of information on the biological function and the evolutionary history of macromolecules. They can be used to examine sequence-structure-function relationships, interactions, active sites, and more [28].

[II] MATERIALS AND METHODS

In this paper we have considered around 266 species which are found in Gujarat state of India. Further we searched each species in ncbi database and finally found around 149 species' information like Pubmed, Pubmed Central, Nucleotide, SRA, Popset, Genome, Bioproject, Protein and Structure information of leguminosae family. After compiling and collection of each leguminosae family's species we need to create one database for accessing and retrieval of each species data at one platform. So, we designed one database which includes all this information and one tool to retrieve data directly from ncbi database. We have created one database with three tables in it, with the help of XAMPP package by using MySQL database and tool designed with Dreamweaver software and coding done in PHP language.

2.1. XAMPP package

It is a free and open source cross-platform web server solution stack package, consisting of mainly the Apache HTTP Server, MySQL database, and interpreters for scripts written in the PHP and Perl programming languages. We just need to download, extract and start to use. XAMPP is a free and open source cross platform web server package. It is available for Linux, Windows, Solaris and Mac OS X platforms [17].

2.2. 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 [15, 18]. We have written script in PHP language as it has many useful features for current work.

2.3. 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 [16]. We have connected our database in this software and written PHP script to fetch data from database.

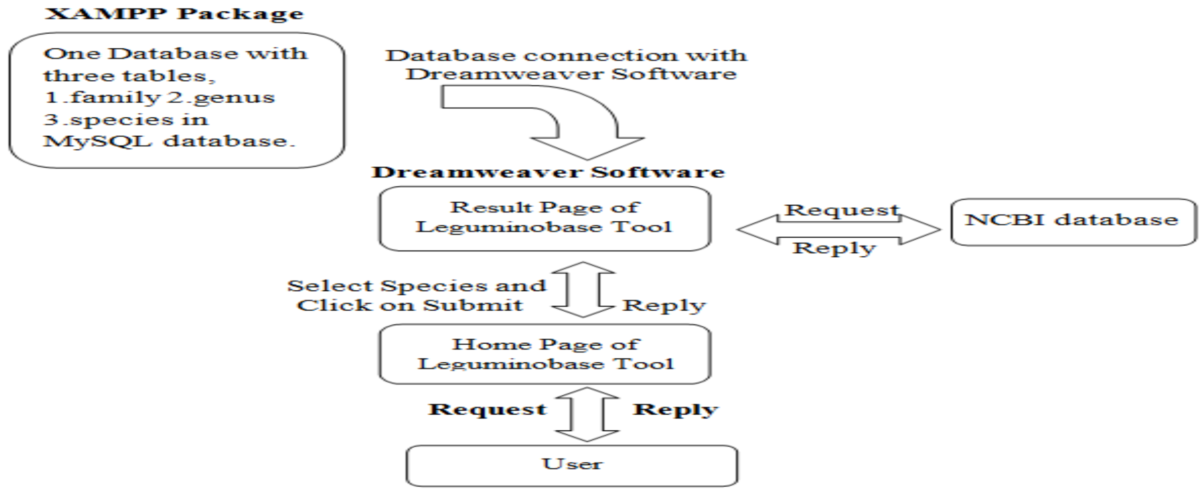

XAMPP Package
One Database with three tables, 1.family 2.genus 3.species in MySQL database.
Database connection with Dreamweaver Software
Dreamweaver Software
Result Page of Leguminobase Tool
Request
Reply
NCBI database
Select Species and Click on Submit
Reply
Home Page of Leguminobase Tool
Request
Reply
User

Figure 1: Flow chart of Leguminobase Tool

[III] RESULTS

3.1. Leguminobase Tool

This tool is based on user's choice, user has to choose one subfamily in first Sub-family Option and genus which is included in the respective subfamily will be seen in next Genus option (Figure 2). User has to select one genus and species related to that genus will be seen in last option which is Species option. Now select your species and then click on Submit (Figure 3). User will get full information of particular species which includes PubMed, Pubmed Central, Nucleotide, SRA, PopSet, Genome, BioProject, Protein and Structure information of Leguminosae family (Figure 4). After click on particular option like DNA, Protein etc, user will directly get information from NCBI database to this result page.

User can get total 149 records for specific species of Leguminosae family (Table 1). **[Table 1]**

Sub family	DNA	Protein	Genome
Fabaceae (Papilionaceae)	104	67	15
Caesalpiniaceae	26	26	-
Mimosaeae	19	14	-
TOTAL	**149**	**107**	**15**

Table 1. Overall Bioinformatics information size:

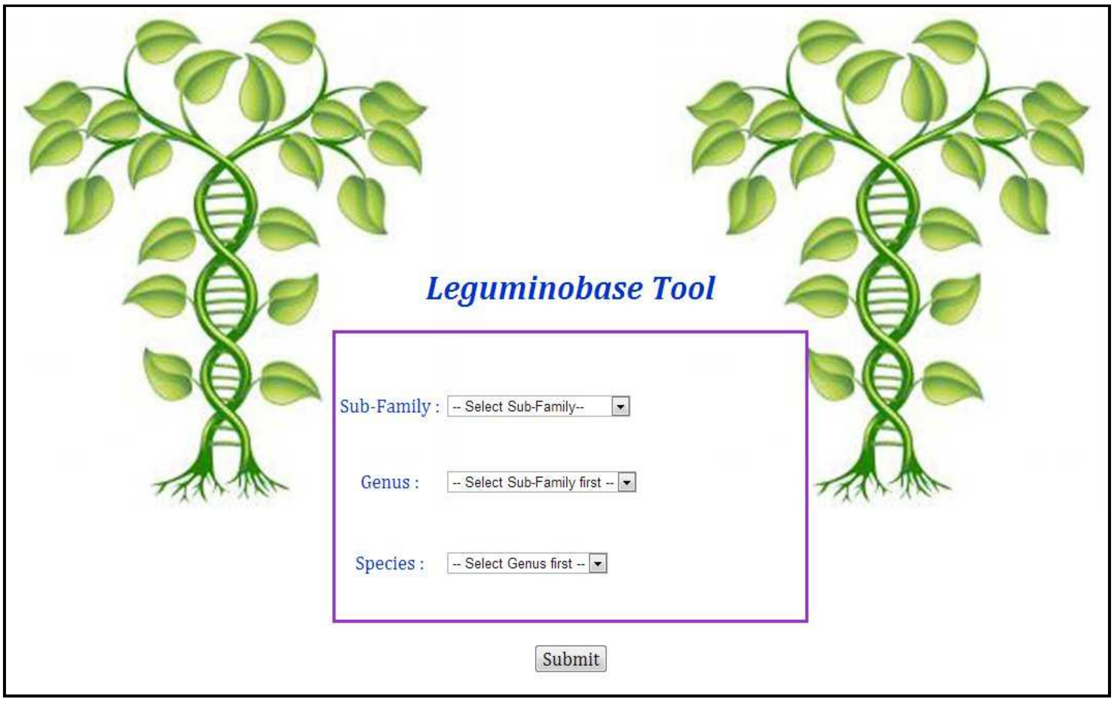
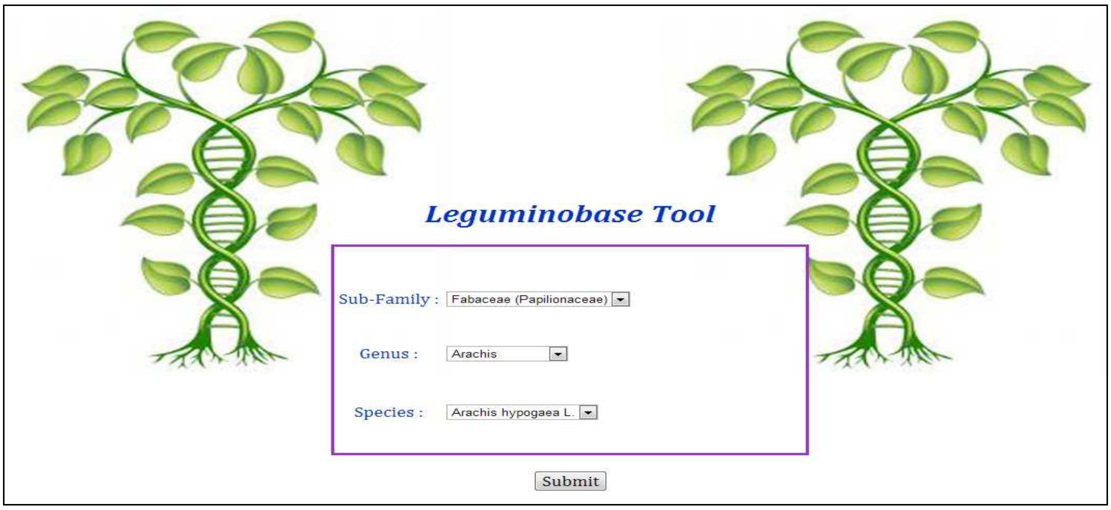


Figure 2: Screen Shot of Leguminobase Tool.

Let's take one example to see how this Tool works (Figure 3).

1. Select Sub-Family: *Fabaceae (Papilionaceae),*
2. Select Genus: *Arachis,*
3. Select Species: *Arachis hypogaea L.,*
4. Click on *Submit* button,



Leguminobase Tool

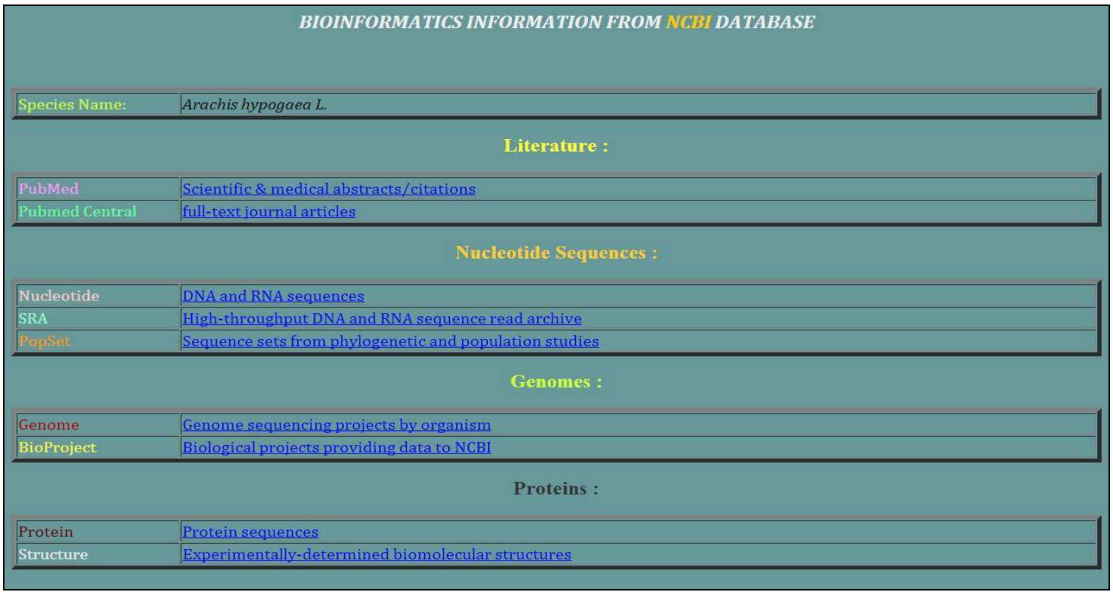
Sub-Family : Fabaceae (Papilionaceae)

Genus : Arachis

Species : Arachis hypogaea L.

Figure 3: Screen Shot of Species Selection.

5. After click on Submit, user will get following output (Figure 4).



BIOINFORMATICS INFORMATION FROM NCBI DATABASE

Species Name:	*Arachis hypogaea L.*

Literature :

PubMed	Scientific & medical abstracts/citations
Pubmed Central	full-text journal articles

Nucleotide Sequences :

Nucleotide	DNA and RNA sequences
SRA	High-throughput DNA and RNA sequence read archive
PopSet	Sequence sets from phylogenetic and population studies

Genomes :

Genome	Genome sequencing projects by organism
BioProject	Biological projects providing data to NCBI

Proteins :

Protein	Protein sequences
Structure	Experimentally-determined biomolecular structures

Figure 4: Result of *Arachis hypogaea L.* species.

[IV] CONCLUSION

There are about 266 Leguminosae Family species are found in Gujarat state then we searched every species in NCBI database we got information of 149 species out of 266 species. The creation of such kind of database and Tool is useful when we centralised all data into one database. Normally in NCBI database we have to write species name in search box in every option like PubMed, Pubmed Central, Nucleotide, SRA, PopSet, Genome, BioProject, Protein and Structure information etc. Here we have done coding in PHP language such that all respective options that described above will directly fetch related information from NCBI database to this result page. This kind of creation of tool will help to get various information of NCBI database at one platform very quickly and also it will reduce search time to get information of particular species of Leguminosae family species which are found in Gujarat state of India.

ACKNOWLEDGEMENT

We would like to thank Sardar Patel University and PURSE program of DST- SPU for their support.

REFERENCES

[1] G. L. Shah (1978): Flora of Gujarat State. Publ. by Sardar Patel University, Vallabh Vidyanagar, Anand, India.
[2] Kalpesh Anjaria (2002) Ph. D. Thesis: Floristic studies of Anand District. Submitted to Sardar Patel University, Vallabh Vidyanagar, Anand, India.
[3] Sagar Patel (2011) Some Leguminous trees in Anand District (M.Sc., Project work) Sardar Patel University, Vallabh Vidyanagar, Gujarat, India.
[4] Jean-Mchel Claverie and Cedric Notredame (2003) Bioinformatics – A Beginner's Guide. Publ. by Wiley Publishing, Inc. USA.
[5] G. M. Oza; Kishore S. Rajput (2006) Biodiversity of Gujarat Forest Trees. Publ. By INSONA, Vadodara, India.
[6] Patel, Anjaria, Panchal (2012) Leguminous Trees In Anand District: Collection and Analysis With Bioinformatics Applications. LAP LAMBERT Academic Publishing, Germany.
[7] 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
[8] Sagar Patel, Panchal H., Smart J., Anjaria K., 2013. Species Information Retrieval Tool: A Bioinformatics tool for Leguminosae family. International Journal of Bioinformatics and Biological Science. Vol.1 Issue.2 June, 2013 Page numbers: 187-194. ISSN 2319-5169.
[9] http://www.en.wikipedia.org
[10] www.theplantlist.org/browse/A/Leguminosae
[11] http://www.ildis.org/
[12] http://www.ncbi.nlm.nih.gov/
[13] http://www.kew.org/
[14] http://www.missouribotanicalgarden.org/
[15] http://www.w3schools.com
[16] www.adobe.com/in/products/dreamweaver.html
[17] http://www.apachefriends.org/en/xampp.html
[18] http://www.php.net
[19] http://www.google.com
[20]http://www.ncbi.nlm.nih.gov/pubmed/
[21]http://www.ncbi.nlm.nih.gov/pmc/
[22]http://www.ncbi.nlm.nih.gov/nuccore/
[23]http://www.ncbi.nlm.nih.gov/sra/
[24]http://www.ncbi.nlm.nih.gov/popset/
[25]http://www.ncbi.nlm.nih.gov/genome/
[26]http://www.ncbi.nlm.nih.gov/bioproject/
[27]http://www.ncbi.nlm.nih.gov/protein/
[28]http://www.ncbi.nlm.nih.gov/structure/