Computational Molecular Biology 2014, Vol. 4, No. 11 1-13
http://cmb.biopublisher.ca

Research Article **Open Access**

Gujarat State Leguminosae Family Database (GLDB): Bioinformatics Database of Leguminosae Family present in Gujarat state of INDIA

Sagar S. Patel[1], Dipti B. Shah[1], Hetalkumar J. Panchal[2]
1. G. H. Patel Post Graduate Department of Computer Science and Technology, Sardar Patel University, Vallabh Vidyanagar, Gujarat-388120, India
2. Gujarat Agricultural Biotechnology Institute, Navsari Agricultural University, Surat, Gujarat- 395007, India
✉ Corresponding author email: sgr308@gmail.com

Computational Molecular Biology, 2014, Vol.4, No.11 doi: 10.5376/cmb.2014.04.0011
Received: 10 Dec., 2014
Accepted: 30 Dec., 2014
Published: 07 Jan., 2015

Preferred citation for this article:
Patel et al., 2014, Gujarat State Leguminosae Family Database (GLDB): Bioinformatics Database of Leguminosae Family present in Gujarat state of INDIA, Computational Molecular Biology, Vol.4, No.11, 1-13 (doi: 10.5376/cmb.2014.04.0011)

Abstract Biological databases play a central role in bioinformatics. They offer scientists the opportunity to access a wide variety of biologically data centrally. In present days molecular data are available for many plant species which can be analyzed in relation to taxonomic or evolutionary or affinity context for different plant species. Authors have tried to generate preliminary Bioinformatics data for Leguminosae family species. As much as possible of a particular type of information should be available in one plat form. In short we have made a database which includes all information of Leguminosae family which are found in Gujarat state of India, which contains Botanical information of each species and Bioinformatics information with analysis at one platform. The creation of such kind of database is reflection of interdisciplinary approach for current era where bioinformatics can play major role for botanical database. Which is further extended by developing bioinformatics database and placed all such relevant information on a wide platform for community in general as well as for scientific society in particular.

Keywords Bioinformatics; Database; Leguminosae family

Introduction

Bioinformatics has evolved into a full-fledged multidisciplinary subject that integrates developments in information and computer technology as applied to Biotechnology and Biological Sciences. Bioinformatics uses computer software tools for database creation, data management, data warehousing, data mining and global communication networking.

Bioinformatics comprises of annotation, storage, analysis, and searching/retrieval of nucleic acid sequence (genes and RNAs), protein sequence and structural information. This includes databases of the sequences and structural information as well methods to access, search, visualize and retrieve the information.

Bioinformatics concern the creation and maintenance of databases of biological information whereby researchers can both access existing information and submit new entries. Function genomics, biomolecular structure, proteome analysis, cell metabolism, biodiversity, downstream processing in chemical engineering, drug and vaccine design are some of the areas in which Bioinformatics is an integral component.

Leguminosae family is one of the largest family that contains thousands of species of Plants, Herbs, Shrubs, Trees worldwide. This family contains more than 250 species which are found in Gujarat state. There are 3 subfamilies of Leguminosae family which are, Fabaceae (Papilionaceae), Mimosaceae and Caesalpiniaceae.

1 Methods

This database includes Leguminosae family's each species' information like its description, taxonomic classification, Locality in Gujarat state, its local name, Pictures, Uses etc. This database also includes Distribution of Leguminosae Family Members in Gujarat state section where different regions of Gujarat state included with Google map followed by list of species which are found in that particular region.

The information contain in this database is from NCBI database and Bioinformatics Analysis of five RNA-Seq Sequences is also done. Data provided in this database is such way that it will be very useful to many people for further study or analysis purpose.

To carry out detailed study of any plant species requires a centrally available data of it. One of the ways to centralize the data is to create a database which could be centrally available and updated regularly according to future requirement.

1.1 Collection of Data

To create any database we need information which is called as Data. Information of all Leguminosae family's species like it's description, taxonomic classification, locality, local name, details of fruit and flowers, locality in Gujarat state, uses and pictures were collected from various resources like many Ph.D. thesis, books etc. After compiling and collection of each Leguminosae family's species we need to create one database for accessing and retrieval of each species data.

Creation of database is necessary while we are dealing with this kind of data. So, we have created database with the help of XAMPP, Dreamweaver and on scripting side PHP language.

2 Results

2.1 Species Information Retrieval Tool

As a part of research, one tool called "Species Information Retrieval Tool" is designed and implemented to get information of species of Leguminosae family which are found in Gujarat state at one click. User has to click on Botanical Information option on left side in GLDB database. Figure 1 is the Home page of GLDB database and Figure 2 is the Home page of Species Information Retrieval Tool, in which species is selected and after click on Submit button it will give full result of that selected particular species (Figure 3).

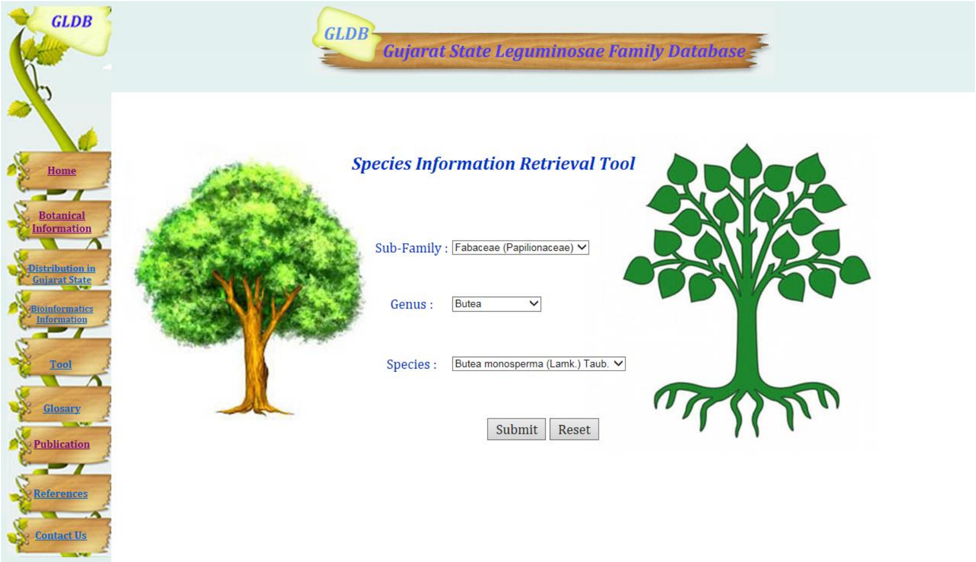
2.2 Distribution of Leguminosae Family Members in Gujarat state

The state is divided into sub zones like Kutch, Saurashtra, North Gujarat, Central Gujarat, South Gujarat and Other region (which includes species found in forests, garden) etc. In addition, Region wise web pages created which gives information of those species which are found in that particular region.

About Leguminosae Family

- Leguminosae family is one of the largest families that contain thousands of species of Plants, Herbs, Shrubs, and Trees worldwide. 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.
- 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.
- This family contains more than 250 species which are found in Gujarat state; even all pulses are covered in this family. We have collected data of each species and made database which contains all information in one platform, so user will get all information of species at one click.
- There are 3 subfamilies of Leguminosae family which are, *Fabaceae (Papilionaceae), Caesalpiniaceae and Mimosaeae.*
- A database is collection of information stored in a computer in a systematic way, such that a computer can consult it to answer questions. A biological database is a large, organised body of persistent data, usually associated with computerized software designed to update, query and retrieve components of the data stored within the system. A simple database might be a single file containing many records, each of which includes the same set of information.

Figure 1 Home page of GLDB database

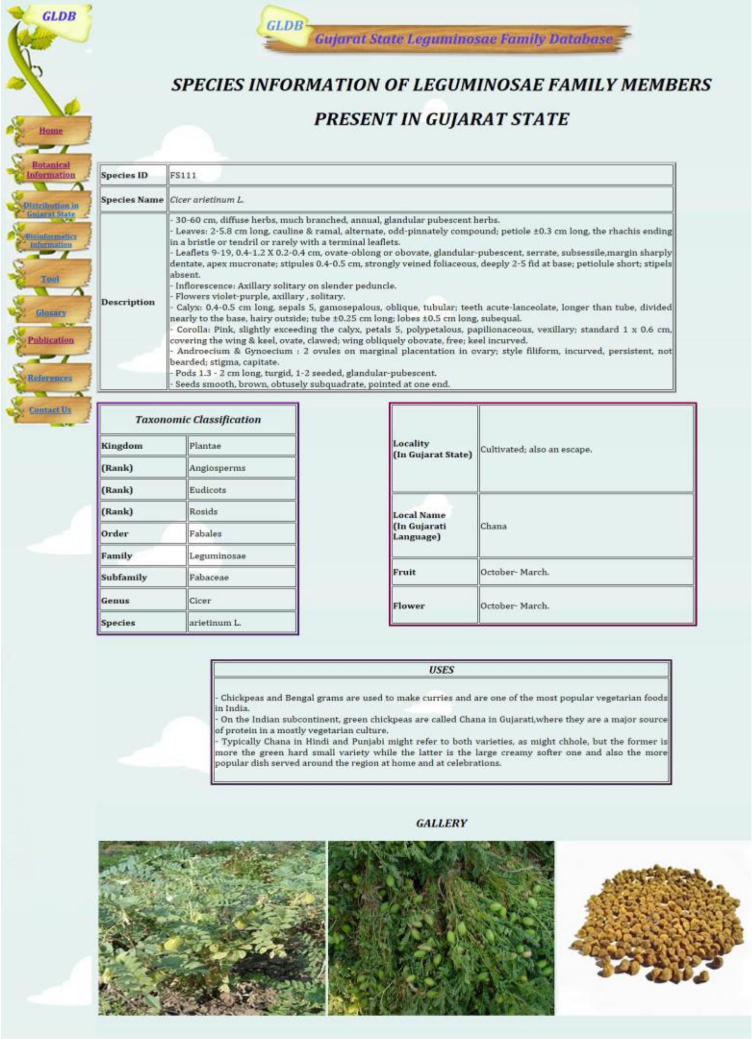


Sub-Family : Fabaceae (Papilionaceae)

Genus : Butea

Species : Butea monosperma (Lamk.) Taub.

Figure 2 Home page of Species Information Retrieval Tool



SPECIES INFORMATION OF LEGUMINOSAE FAMILY MEMBERS PRESENT IN GUJARAT STATE

Species ID	FS111
Species Name	*Cicer arietinum L.*
Description	- 30-60 cm, diffuse herbs, much branched, annual, glandular pubescent herbs. - Leaves: 2-5.8 cm long, cauline & ramal, alternate, odd-pinnately compound; petiole ±0.3 cm long, the rhachis ending in a bristle or tendril or rarely with a terminal leaflets. - Leaflets 9-19, 0.4-1.2 X 0.2-0.4 cm, ovate-oblong or obovate, glandular-pubescent, serrate, subsessile,margin sharply dentate, apex mucronate; stipules 0.4-0.5 cm, strongly veined foliaceous, deeply 2-5 fid at base; petiolule short; stipels absent. - Inflorescence: Axillary solitary on slender peduncle. - Flowers violet-purple, axillary , solitary. - Calyx: 0.4-0.5 cm long, sepals 5, gamosepalous, oblique, tubular; teeth acute-lanceolate, longer than tube, divided nearly to the base, hairy outside; tube ±0.25 cm long; lobes ±0.5 cm long, subequal. - Corolla: Pink, slightly exceeding the calyx, petals 5, polypetalous, papilionaceous, vexillary; standard 1 x 0.6 cm, covering the wing & keel, ovate, clawed; wing obliquely obovate, free; keel incurved. - Androecium & Gynoecium : 2 ovules on marginal placentation in ovary; style filiform, incurved, persistent, not bearded; stigma, capitate. - Pods 1.3 - 2 cm long, turgid, 1-2 seeded, glandular-pubescent. - Seeds smooth, brown, obtusely subquadrate, pointed at one end.

Taxonomic Classification	
Kingdom	Plantae
(Rank)	Angiosperms
(Rank)	Eudicots
(Rank)	Rosids
Order	Fabales
Family	Leguminosae
Subfamily	Fabaceae
Genus	Cicer
Species	arietinum L.

Locality (In Gujarat State)	Cultivated; also an escape.
Local Name (In Gujarati Language)	Chana
Fruit	October- March.
Flower	October- March.

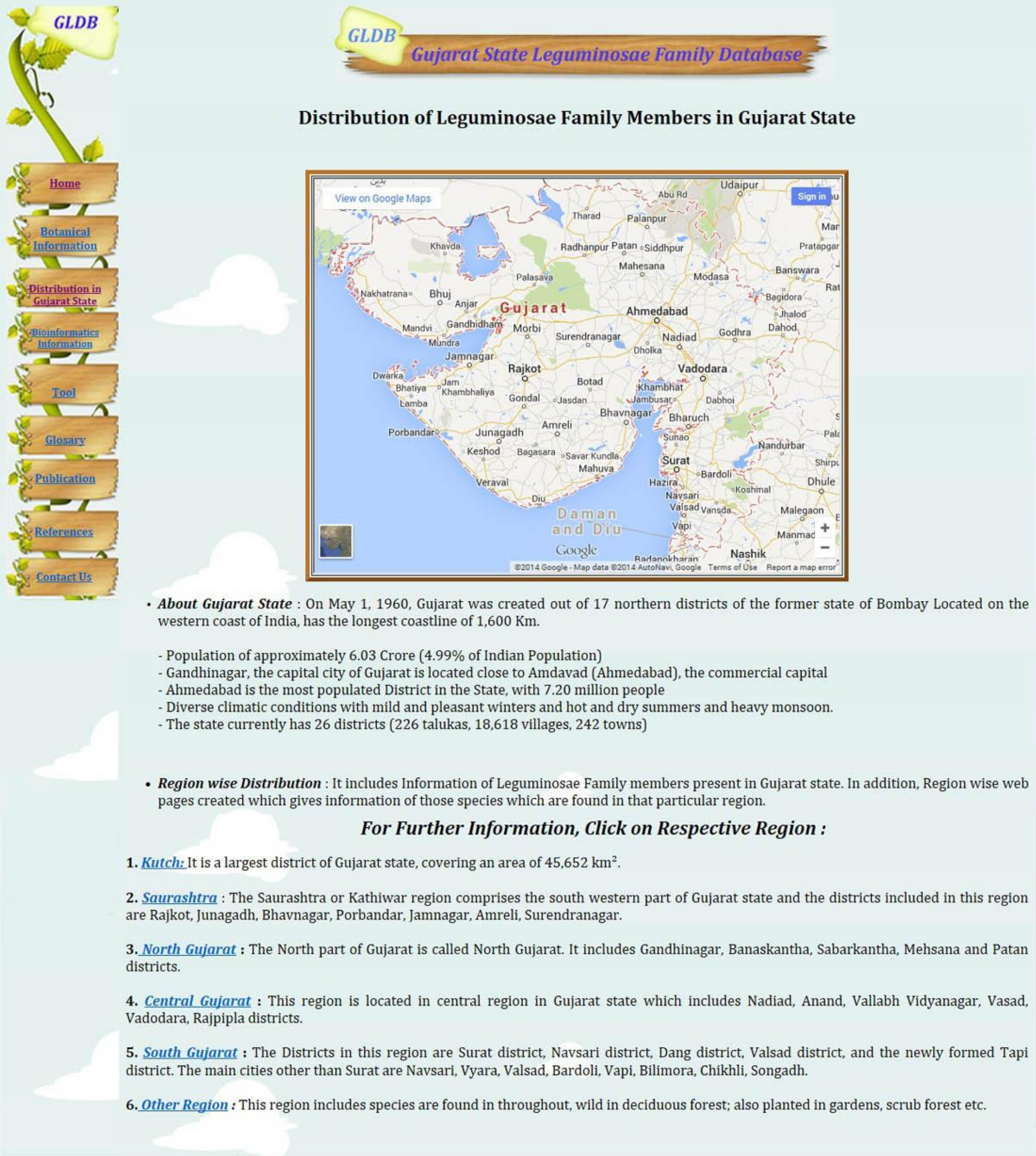
USES

- Chickpeas and Bengal grams are used to make curries and are one of the most popular vegetarian foods in India.
- On the Indian subcontinent, green chickpeas are called Chana in Gujarati,where they are a major source of protein in a mostly vegetarian culture.
- Typically Chana in Hindi and Punjabi might refer to both varieties, as might chhole, but the former is more the green hard small variety while the latter is the large creamy softer one and also the more popular dish served around the region at home and at celebrations.

GALLERY

Figure 3 Result page of Species Information Retrieval Tool

User has to click on Distribution in Gujarat state option on left side in GLDB database. In addition this data are developed by using Bioinformatics approach, so that user will also get location of each region in Google map, as per choice and after click on respective species it will show species' full description. Figure 4 is Home Page which shows information of Gujarat State and Leguminosae Family with region wise distribution.



Distribution of Leguminosae Family Members in Gujarat State

- ***About Gujarat State*** : On May 1, 1960, Gujarat was created out of 17 northern districts of the former state of Bombay Located on the western coast of India, has the longest coastline of 1,600 Km.

  - Population of approximately 6.03 Crore (4.99% of Indian Population)
  - Gandhinagar, the capital city of Gujarat is located close to Amdavad (Ahmedabad), the commercial capital
  - Ahmedabad is the most populated District in the State, with 7.20 million people
  - Diverse climatic conditions with mild and pleasant winters and hot and dry summers and heavy monsoon.
  - The state currently has 26 districts (226 talukas, 18,618 villages, 242 towns)

- ***Region wise Distribution*** : It includes Information of Leguminosae Family members present in Gujarat state. In addition, Region wise web pages created which gives information of those species which are found in that particular region.

For Further Information, Click on Respective Region :

1. ***Kutch:*** It is a largest district of Gujarat state, covering an area of 45,652 km^2.

2. ***Saurashtra*** : The Saurashtra or Kathiwar region comprises the south western part of Gujarat state and the districts included in this region are Rajkot, Junagadh, Bhavnagar, Porbandar, Jamnagar, Amreli, Surendranagar.

3. ***North Gujarat*** **:** The North part of Gujarat is called North Gujarat. It includes Gandhinagar, Banaskantha, Sabarkantha, Mehsana and Patan districts.

4. ***Central Gujarat*** **:** This region is located in central region in Gujarat state which includes Nadiad, Anand, Vallabh Vidyanagar, Vasad, Vadodara, Rajpipla districts.

5. ***South Gujarat*** **:** The Districts in this region are Surat district, Navsari district, Dang district, Valsad district, and the newly formed Tapi district. The main cities other than Surat are Navsari, Vyara, Valsad, Bardoli, Vapi, Bilimora, Chikhli, Songadh.

6. ***Other Region*** **:** This region includes species are found in throughout, wild in deciduous forest; also planted in gardens, scrub forest etc.

Figure 4 Home Page of Gujarat State and Leguminosae Family with region wise distribution

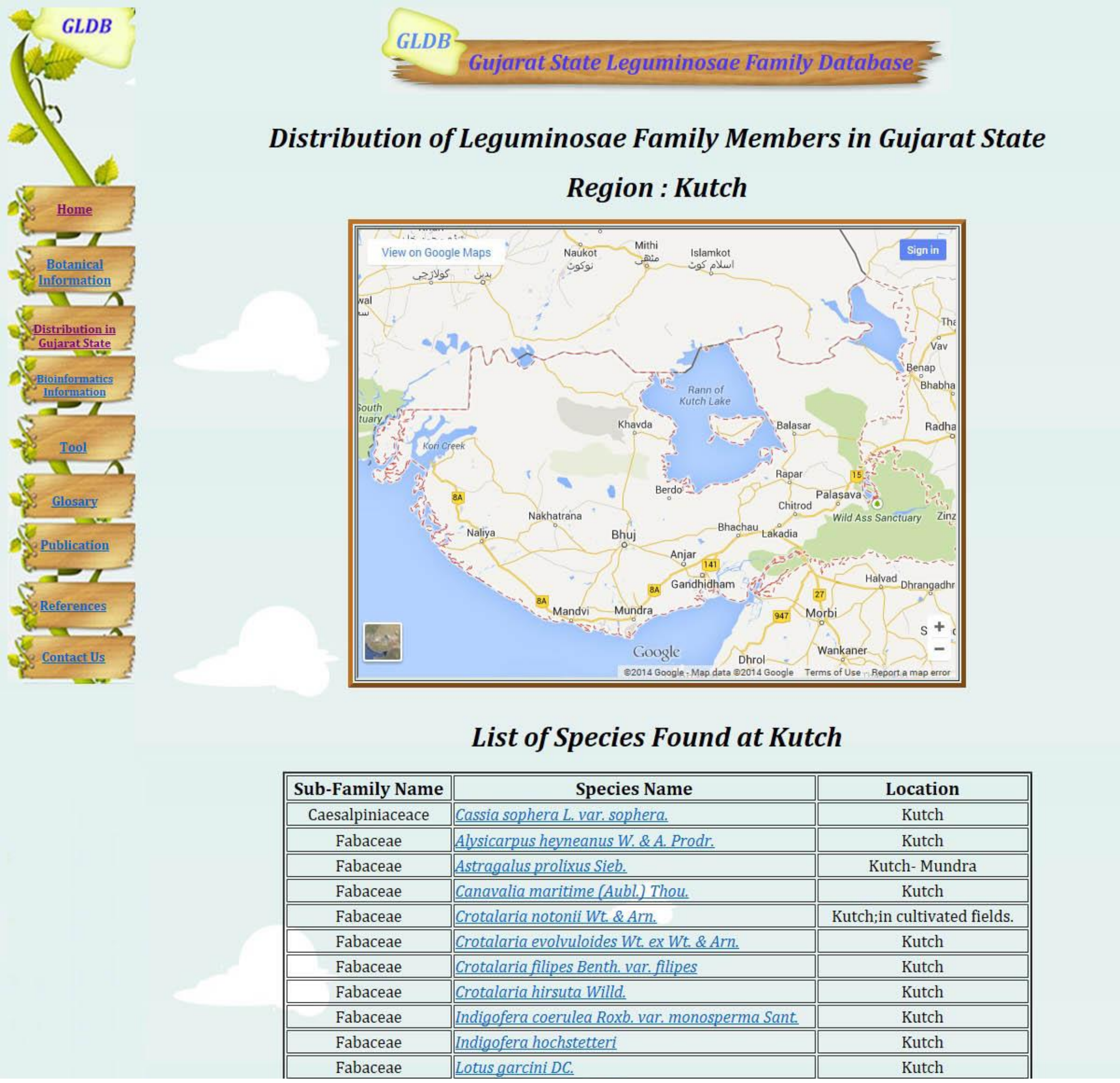
Figure 5 is the information of Kutch region and if user click on any species then it will show result page as shown in Figure 6.

2.3 Data and Analysis of Leguminosae Family

This Section introduces Bioinformatics data analysis of Leguminosae family species by tools and data analysis (Figure 7).

One tool called "Leguminobase Tool" developed to get information like DNA, PROTEIN, Genome etc of each Leguminosae family species which are found in Gujarat state.

One tool called "ConSeq Tool" also developed to find out any conserved sequence found in submitted sequence.



Distribution of Leguminosae Family Members in Gujarat State

Region : Kutch

List of Species Found at Kutch

Sub-Family Name	Species Name	Location
Caesalpiniaceace	*Cassia sophera L. var. sophera.*	Kutch
Fabaceae	*Alysicarpus heyneanus W. & A. Prodr.*	Kutch
Fabaceae	*Astragalus prolixus Sieb.*	Kutch- Mundra
Fabaceae	*Canavalia maritime (Aubl.) Thou.*	Kutch
Fabaceae	*Crotalaria notonii Wt. & Arn.*	Kutch;in cultivated fields.
Fabaceae	*Crotalaria evolvuloides Wt. ex Wt. & Arn.*	Kutch
Fabaceae	*Crotalaria filipes Benth. var. filipes*	Kutch
Fabaceae	*Crotalaria hirsuta Willd.*	Kutch
Fabaceae	*Indigofera coerulea Roxb. var. monosperma Sant.*	Kutch
Fabaceae	*Indigofera hochstetteri*	Kutch
Fabaceae	*Lotus garcini DC.*	Kutch

Figure 5 Kutch region Information

This chapter also introduces RNA-Seq data analysis of five Leguminosae family species; which is described as “De novo RNA-Seq” data analysis (Table 1 and 2).

Table 1 Information of Distribution of Leguminosae family

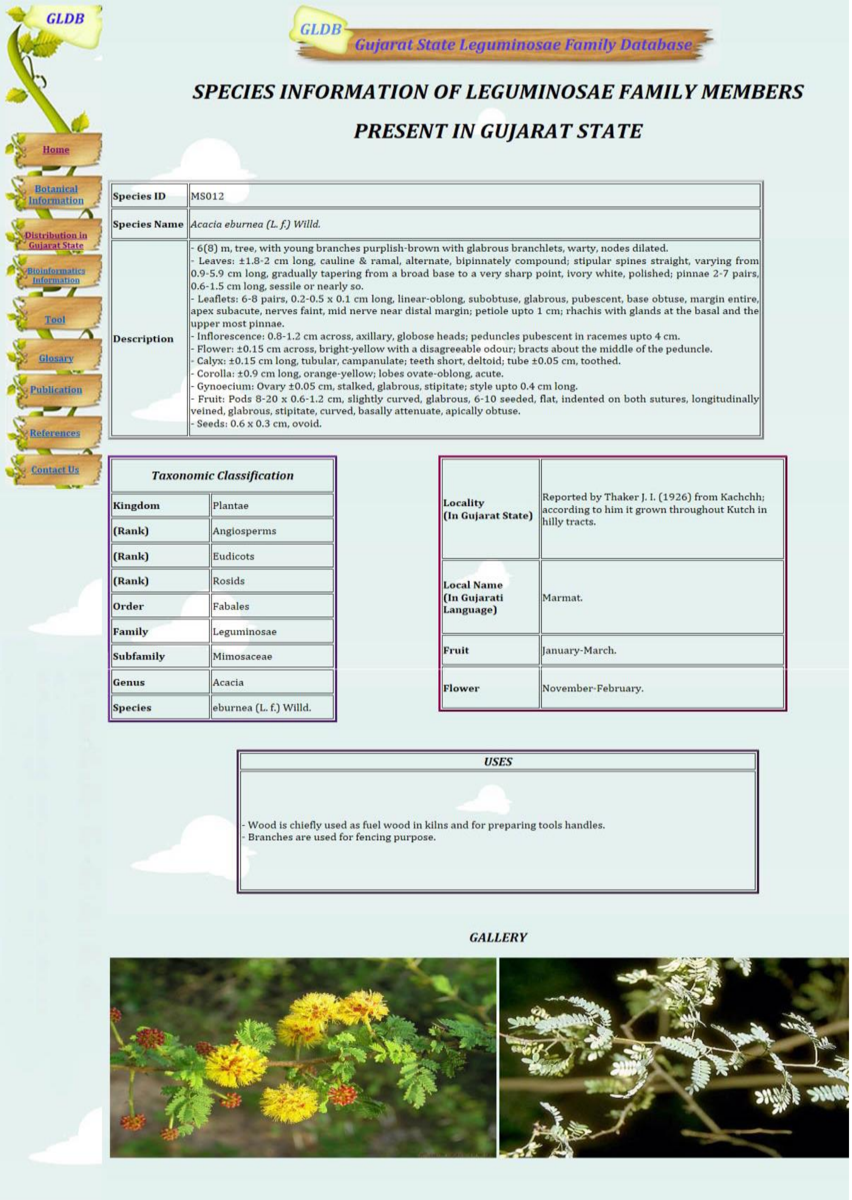
Location	Region specific Species
Kutch	25
Saurashtra	60
North Gujarat	25
Central Gujarat	61
South Gujarat	82
Other Region (Throughout)	115

Table 2 Information of Leguminobase tool

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

2.3.1 Leguminobase Tool

There are more than 250 species of this family which are found in Gujarat state of India, out of which information of around 149 Leguminosae family species from NCBI database has been collected and complied. There are three subfamilies of Leguminosae family which are Fabaceae (Papilionaceae), Caesalpiniaceae and Mimosaeae.



SPECIES INFORMATION OF LEGUMINOSAE FAMILY MEMBERS PRESENT IN GUJARAT STATE

Species ID	MS012
Species Name	*Acacia eburnea (L. f.) Willd.*
Description	- 6(8) m, tree, with young branches purplish-brown with glabrous branchlets, warty, nodes dilated. - Leaves: ±1.8-2 cm long, cauline & ramal, alternate, bipinnately compound; stipular spines straight, varying from 0.9-5.9 cm long, gradually tapering from a broad base to a very sharp point, ivory white, polished; pinnae 2-7 pairs, 0.6-1.5 cm long, sessile or nearly so. - Leaflets: 6-8 pairs, 0.2-0.5 x 0.1 cm long, linear-oblong, subobtuse, glabrous, pubescent, base obtuse, margin entire, apex subacute, nerves faint, mid nerve near distal margin; petiole upto 1 cm; rhachis with glands at the basal and the upper most pinnae. - Inflorescence: 0.8-1.2 cm across, axillary, globose heads; peduncles pubescent in racemes upto 4 cm. - Flower: ±0.15 cm across, bright-yellow with a disagreeable odour; bracts about the middle of the peduncle. - Calyx: ±0.15 cm long, tubular, campanulate; teeth short, deltoid; tube ±0.05 cm, toothed. - Corolla: ±0.9 cm long, orange-yellow; lobes ovate-oblong, acute. - Gynoecium: Ovary ±0.05 cm, stalked, glabrous, stipitate; style upto 0.4 cm long. - Fruit: Pods 8-20 x 0.6-1.2 cm, slightly curved, glabrous, 6-10 seeded, flat, indented on both sutures, longitudinally veined, glabrous, stipitate, curved, basally attenuate, apically obtuse. - Seeds: 0.6 x 0.3 cm, ovoid.

Taxonomic Classification	
Kingdom	Plantae
(Rank)	Angiosperms
(Rank)	Eudicots
(Rank)	Rosids
Order	Fabales
Family	Leguminosae
Subfamily	Mimosaceae
Genus	Acacia
Species	eburnea (L. f.) Willd.

Locality (In Gujarat State)	Reported by Thaker J. I. (1926) from Kachchh; according to him it grown throughout Kutch in hilly tracts.
Local Name (In Gujarati Language)	Marmat.
Fruit	January-March.
Flower	November-February.

USES
- Wood is chiefly used as fuel wood in kilns and for preparing tools handles. - Branches are used for fencing purpose.

GALLERY

Figure 6 Information of species found in Kutch region

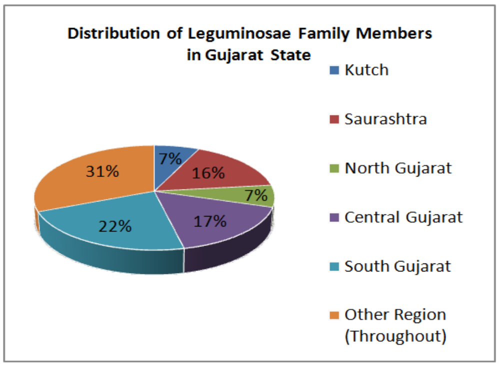
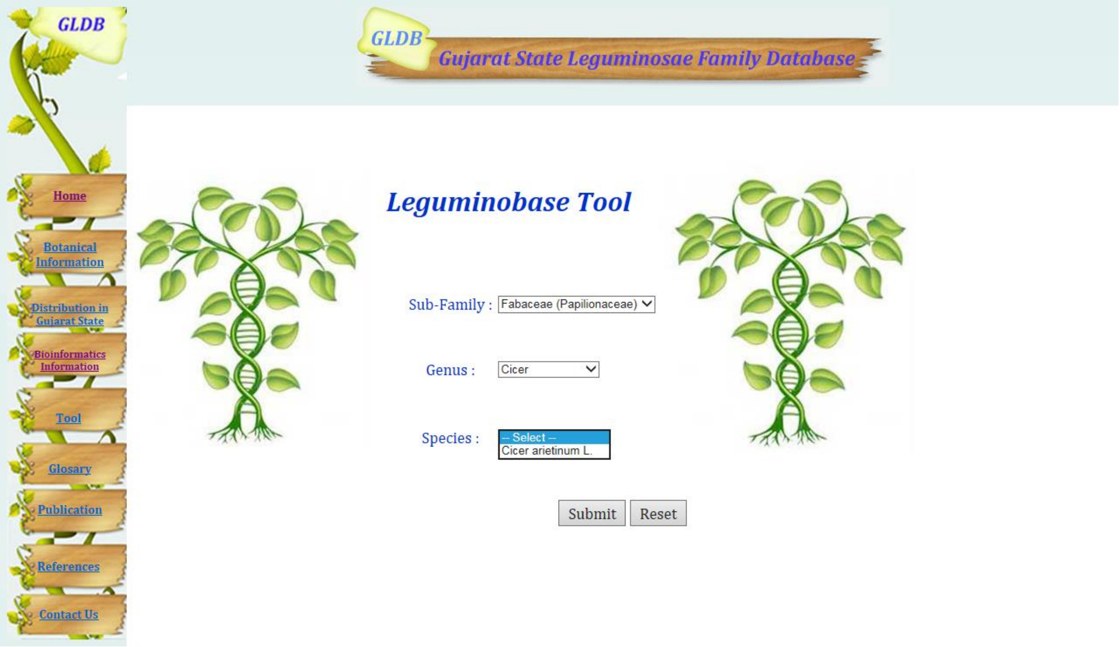


Figure 7 Information of Distribution of Leguminosae family as pie chart

In this "Leguminobase Tool" user has to select respective option and after clicking on Submit button it will 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.

Use has to select Bioinformatics Information option on left side in GLDB database and click on Leguminobase Tool. Figure 8 shows the one species which is selected and after click on submit button, it gives various options

(Figure 9) and after click on any option, it will fetch data from NCBI database into this GLDB database. Figure 10 shows one of the options which shows the Pubmed option of that particular species.



Sub-Family : Fabaceae (Papilionaceae)

Genus : Cicer

Species : -- Select --
Cicer arietinum L.

Figure 8 Leguminobase tool species selection

BIOINFORMATICS INFORMATION FROM NCBI DATABASE

Species Name:	*Cicer arietinum 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

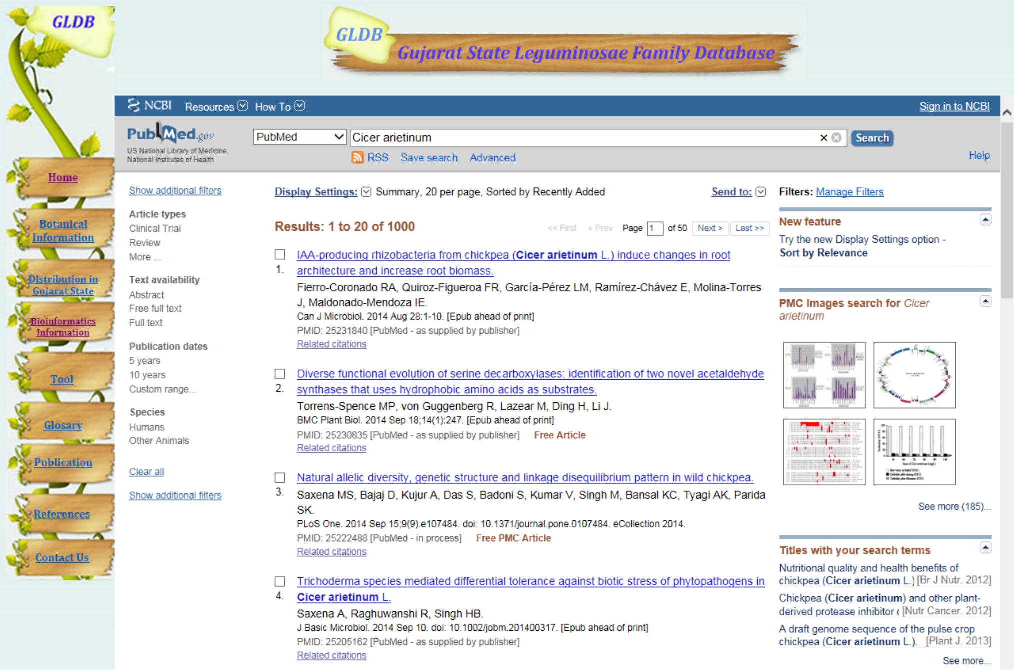
Genomes :

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

Proteins :

Protein	Protein sequences
Structure	Experimentally-determined biomolecular structures

Figure 9 Result page of Leguminobase tool



Display Settings: Summary, 20 per page, Sorted by Recently Added

Results: 1 to 20 of 1000

☐ 1. IAA-producing rhizobacteria from chickpea (**Cicer arietinum** L.) induce changes in root architecture and increase root biomass.
Fierro-Coronado RA, Quiroz-Figueroa FR, García-Pérez LM, Ramírez-Chávez E, Molina-Torres J, Maldonado-Mendoza IE.
Can J Microbiol. 2014 Aug 28:1-10. [Epub ahead of print]
PMID: 25231840 [PubMed - as supplied by publisher]
Related citations

☐ 2. Diverse functional evolution of serine decarboxylases: identification of two novel acetaldehyde synthases that uses hydrophobic amino acids as substrates.
Torrens-Spence MP, von Guggenberg R, Lazear M, Ding H, Li J.
BMC Plant Biol. 2014 Sep 18;14(1):247. [Epub ahead of print]
PMID: 25230835 [PubMed - as supplied by publisher] **Free Article**
Related citations

☐ 3. Natural allelic diversity, genetic structure and linkage disequilibrium pattern in wild chickpea.
Saxena MS, Bajaj D, Kujur A, Das S, Badoni S, Kumar V, Singh M, Bansal KC, Tyagi AK, Parida SK.
PLoS One. 2014 Sep 15;9(9):e107484. doi: 10.1371/journal.pone.0107484. eCollection 2014.
PMID: 25222488 [PubMed - in process] **Free PMC Article**
Related citations

☐ 4. Trichoderma species mediated differential tolerance against biotic stress of phytopathogens in **Cicer arietinum** L.
Saxena A, Raghuwanshi R, Singh HB.
J Basic Microbiol. 2014 Sep 10. doi: 10.1002/jobm.201400317. [Epub ahead of print]
PMID: 25205162 [PubMed - as supplied by publisher]
Related citations

New feature
Try the new Display Settings option - **Sort by Relevance**

PMC Images search for *Cicer arietinum*

Titles with your search terms
Nutritional quality and health benefits of chickpea (**Cicer arietinum** L.) [Br J Nutr. 2012]
Chickpea (**Cicer arietinum**) and other plant-derived protease inhibitor ([Nutr Cancer. 2012]
A draft genome sequence of the pulse crop chickpea (**Cicer arietinum** L.). [Plant J. 2013]

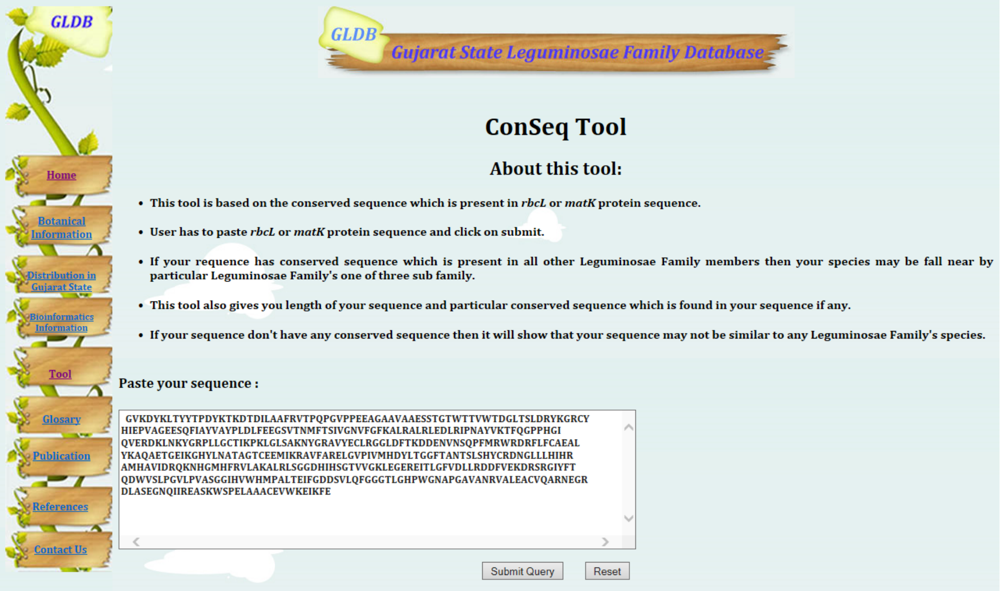
Figure 10 Screenshot of Pubmed information from NCBI database in GLDB database

2.3.2 ConSeq Tool

In this ConSeq tool, 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 is 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.

User has to click on Tool option on left side in GLDB database which redirects to home page of ConSeq Tool (Figure 11). If any conserved region found in that sequence then it will show result as shown in Figure 12.

But if no conserved region found in sequence then ConSeq tool shows output as shown in Figure 13.



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 requence 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.

Paste your sequence :

```
 GVKDYKLTYYTPDYKTKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTTVWTDGLTSLDRYKGRCY
HIEPVAGEESQFIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRIPNAYVKTFQGPPHGI
QVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENVNSQPFMRWRDRFLFCAEAL
YKAQAETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHDYLTGGFTANTSLSHYCRDNGLLLHIHR
AMHAVIDRQKNHGMHFRVLAKALRLSGGDHIHSGTVVGKLEGEREITLGFVDLLRDDFVEKDRSRGIYFT
QDWVSLPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLGHPWGNAPGAVANRVALEACVQARNEGR
DLASEGNQIIREASKWSPELAAACEVWKEIKFE
```

Figure 11 Screen shot of ConSeq Tool

ConSeq Tool - RESULT

- **Your Sequence is:**

```
MEEFQGYLELDRYQQHDFLYPLIFREYIYALAHDHGLNRSILLDNVGYDTKYSLLIIKRLISRMYQQNHLIISANDSNQNKFFGYNKNLYSQMMSEGFAV
IVEIPFSLRLVSSLEATEIVKSYNLRSIHSIFPFLEDKFPHLNYVSDVLIPYPIHLEILVQTLRYWVKDPSSLHLLRLLLHEYYNWNSLITTKKVIFSKS
NPRLFLLLYNSHVCEYESILLFLRNQSSHLRLTSSGIFFERIHFYEKKKYPVEEVFVNDFPAAILWFFKDPFMHYVRYQGKSILSSKDTPLLMNKWKYYL
VNLWQCHSYVWSQPGRIYINQLSKHSLDFLGYFSSMRPNLSVVRSQMLENSFLMDNAMKKLDTLVPIIPLIGSLAKVKFCNALGHPISKSTWADSSDFDI
IDRFVHICRNLSHYYSGSSRKKSLYRIKYILRLSCVKTLARKHKSTVRTFLKRLGYKLLDEFFTEEEQILSLIFPRASYTLKKFYRGRIWYLDIFCINDL
VNHE
```

- Length of Your Sequence is: **504**
- **Your Species Sequence may not contain Conserved sequence which is found in Leguminosae Family Species.**

Figure 12 Result page of ConSeq Tool

ConSeq Tool - RESULT

- **Your Sequence is:**

```
MEEFQGYLELDRYQQHDFLYPLIFREYIYALAHDHGLNRSILLDNVGYDTKYSLLIIKRLISRMYQQNHLIISANDSNQNKFFGYNKNLYSQMMSEGFAV
IVEIPFSLRLVSSLEATEIVKSYNLRSIHSIFPFLEDKFPHLNYVSDVLIPYPIHLEILVQTLRYWVKDPSSLHLLRLLLHEYYNWNSLITTKKVIFSKS
NPRLFLLLYNSHVCEYESILLFLRNQSSHLRLTSSGIFFERIHFYEKKKYPVEEVFVNDFPAAILWFFKDPFMHYVRYQGKSILSSKDTPLLMNKWKYYL
VNLWQCHSYVWSQPGRIYINQLSKHSLDFLGYFSSMRPNLSVVRSQMLENSFLMDNAMKKLDTLVPIIPLIGSLAKVKFCNALGHPISKSTWADSSDFDI
IDRFVHICRNLSHYYSGSSRKKSLYRIKYILRLSCVKTLARKHKSTVRTFLKRLGYKLLDEFFTEEEQILSLIFPRASYTLKKFYRGRIWYLDIFCINDL
VNHE
```

- Length of Your Sequence is: **504**
- **Your Species Sequence may not contain Conserved sequence which is found in Leguminosae Family Species.**

Figure 13 Result page of ConSeq Tool when no conserved sequence found

2.3.3 RNA-Seq Data analysis

De novo means assembling short reads to create full-length (sometimes novel) sequences.

De novo sequencing involves sequencing a novel genome for the first time, and requires specialized assembly of sequencing reads. The unique combination of read length, read depth, and flexible paired-end insert sizes makes Illumina sequencers ideal for de novo sequencing. Unparalleled raw read accuracy enables confident and efficient production of high quality, long contig assemblies. Data of both Illumina and Roche 454 are analyzed with five different species.

The detailed analyses of the data set has provided several important features of five species such as GC content, conserved genes across legumes and other plant species, assignment of functional categories by GO terms and identification of SSRs by MISA tool.

It is noted that this study of five different legume species which are *Arachis hypogaea* L., *Cicer arietinum* L., *Phaseolus vulgaris* L., *Trigonella foenum-graecum* L. and *Vicia sativa* L. will be useful for further functional genomics studies as it includes useful information of each species with full annotation.

Figure 14 shows the home page of De novo RNA-Seq data page. User has to select first platform either Roche 454 or Illumina then select species in second option and finally select contig. After click on Submit button, it will give full information of that contig as shown in Figure 15. Figure 15 the result of contigs number 10017 of *Arachis hypogea* L.

De novo Transcriptome Analysis

Select Platform : Illumina

Select Species : Arachis hypogaea L.

Select Contig: SRR1212866fastq_illu_contig_10017

Figure 14 Home page for De novo RNA-Seq data retrieval

The result shows various different information of the one contigs from various databases and useful information like, Species Name, Name of contigs, Fasta Sequence, Sequence Length, Blast E-value Min, Blast Similarity Mean, Blast GO number, Top-Hit Species, Blast Hit Description (HSP), Blast Hit Gene Name, Blast Hit Accession, Blast Hit E-value, Blast Hit Length, Blast Hit Align Length, Blast Hit Positives, Blast Hit Similarity, Blast Hsp/Hit, Blast Hsp/Query, Blast Hit Query Frame, Blast Hit Uniprot, Blast Hit Score, Blast Hit GOs, GO Accession, GO Names, Enzyme codes, InterPro Ids, InterPro GO Acession, InterPro GO Names, InterPro Motif Detail, InterPro Motif Matches, Number of Blast Hits, Enzymes, KeggMaps followed by KEGG pathway image if any contigs is involved in any pathway.

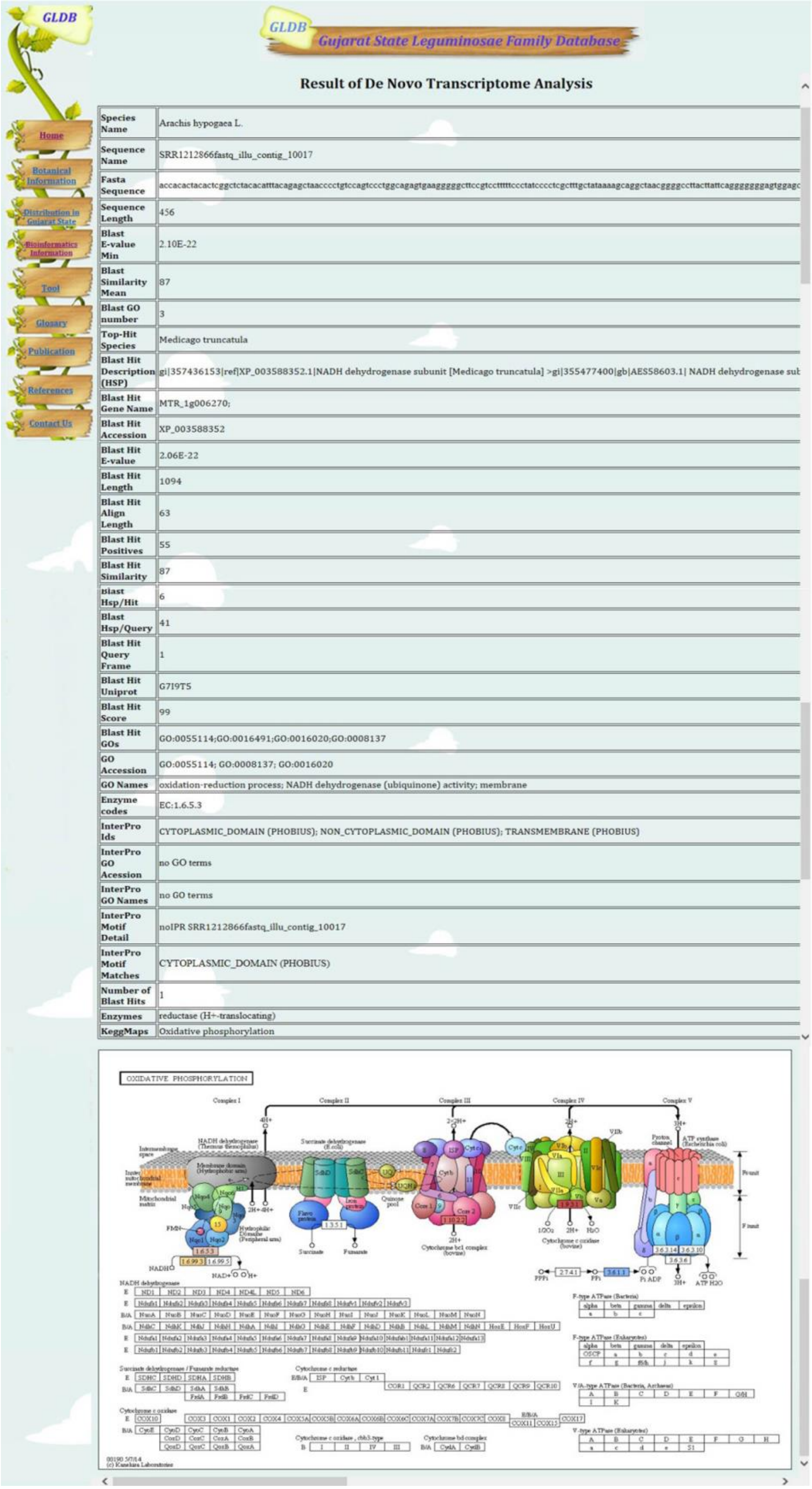
There are total 82,505 records of contigs are inserted, which includes 10824 contigs of *Arachis hypogaea* L., 34678 contigs of *Cicer arietinum* L., 6999 contigs of *Phaseolus vulgaris* L., 7256 contigs of *Trigonella foenum-graecum* L. and 22748 contigs of *Vicia sativa* L.

3 Glossary

It discusses various botanical terms which are useful to identify any plant species. User has to select Glossary option which is on left side in GLDB database.

4 Publication

This section contains list of Publication which are outcome of this database.



Result of De Novo Transcriptome Analysis

Species Name	Arachis hypogaea L.
Sequence Name	SRR1212866fastq_illu_contig_10017
Fasta Sequence	accacactacactcggctctacacatttacagagctaacccctgtccagtccctggcagagtgaagggggcttccgtcctttttccctatcccctcgctttgctataaaagcaggctaacggggccttacttattcagggggggagtggagc
Sequence Length	456
Blast E-value Min	2.10E-22
Blast Similarity Mean	87
Blast GO number	3
Top-Hit Species	Medicago truncatula
Blast Hit Description (HSP)	gi|357436153|ref|XP_003588352.1|NADH dehydrogenase subunit [Medicago truncatula] >gi|355477400|gb|AES58603.1| NADH dehydrogenase sub
Blast Hit Gene Name	MTR_1g006270;
Blast Hit Accession	XP_003588352
Blast Hit E-value	2.06E-22
Blast Hit Length	1094
Blast Hit Align Length	63
Blast Hit Positives	55
Blast Hit Similarity	87
Blast Hsp/Hit	6
Blast Hsp/Query	41
Blast Hit Query Frame	1
Blast Hit Uniprot	G7I9T5
Blast Hit Score	99
Blast Hit GOs	GO:0055114;GO:0016491;GO:0016020;GO:0008137
GO Accession	GO:0055114; GO:0008137; GO:0016020
GO Names	oxidation-reduction process; NADH dehydrogenase (ubiquinone) activity; membrane
Enzyme codes	EC:1.6.5.3
InterPro Ids	CYTOPLASMIC_DOMAIN (PHOBIUS); NON_CYTOPLASMIC_DOMAIN (PHOBIUS); TRANSMEMBRANE (PHOBIUS)
InterPro GO Acession	no GO terms
InterPro GO Names	no GO terms
InterPro Motif Detail	noIPR SRR1212866fastq_illu_contig_10017
InterPro Motif Matches	CYTOPLASMIC_DOMAIN (PHOBIUS)
Number of Blast Hits	1
Enzymes	reductase (H+-translocating)
KeggMaps	Oxidative phosphorylation

Figure 15 Result page of de novo RNA-Seq

5 References

This section contains list of papers, online site , books etc which were considered for creating this database.

6 Contact Us

This section contain contact us form in which user can send any questions to us.

7 Conclusion

Database designing of Leguminosae family members in Gujarat state was taken up with following objectives keeping in mind. Like,

✓ To bridge the botanical information with the Bioinformatics information and analysis.

✓ To utilize various tools of Bioinformatics for analytical purpose for the Leguminosae family species.

✓ To generate secondary information from above work with the help of various tools and software available.

✓ To provide Bioinformatics information for general public in the form of database.

A comprehensive database for Leguminosae family titled as “Gujarat state Leguminosae family Database (GLDB)” was created with useful information of each species of Leguminosae family.

There are many inbuilt tools developed in this database to get information of particular species like it’s full botanical information along with Distribution of each species of Leguminosae family in Gujarat state with Google map.

While Bioinformatics section comprises of many tools like one tool was designed to get particular species’ DNA, PROTEIN, GENOME etc. information from NCBI database and ConSeq Tool designed to find out any conserved sequences. RNA-Seq data analysis of five Leguminosae family done with de novo sequence assembly & annotation also done.

This particular database of Leguminosae family has served the demands of the present botanical scientific community. So far such information at one platform is not available there by it will serve the purpose of their needs too. Various scattered data on such Leguminosae family species were placed in such a manner that any botanist or person who desire to find information on these particular trees will get at one touch or on mouse click.

References

G. L. Shah (1978): Flora of Gujarat State. Publ. by Sardar Patel University, Vallabh Vidyanagar, Anand, India

G. M. Oza; Kishore S. Rajput (2006) Biodiversity of Gujarat Forest Trees.Publ. By INSONA, Vadodara, India

Harborne, J.B. 1994. Phytochemistry of the Leguminosae. In Phytochemical Dictionary of the Leguminosae, eds Bisby,F.A. et al. London: Chapman & Hall

Heywood, V.H.(ed) 1993. Flowering Plants of the World. London: B T Batsford

Hickey, M. & King, C. 1997. Common Families of Flowering Plants. Cambridge: Cambridge University Press

J. L. Collins, J. P. Biggs, C. Voelckel and S. Joly, 2008, An approach to transcriptome analysis of non-model organisms using short-read sequences, Genome Informatics 21:3-14 http://dx.doi.org/10.1142/9781848163324_0001

Jean-Mchel Claverie and Cedric Notredame (2003) Bioinformatics – A Beginner’s Guide. Publ. by Wiley Publishing, Inc. USA

Jianan Zhang, Shan Liang, Jialei Duan, Jin Wang, Silong Chen, Zengshu Cheng, Qiang Zhang, Xuanqiang Liang and Yurong Li, 2012, De novo assembly and Characterisation of the Transcriptome during seed development, and generation of genic-SSR markers in Peanut (Arachis hypogaea L.), BMC Genomics 2012 13:90 http://dx.doi.org/10.1186/1471-2164-13-90

Kalpesh Anjaria (2002) Ph. D. Thesis: Floristic studies of Anand District. Submitted to Sardar Patel University, Vallabh Vidyanagar, Anand, India

Libault, M., Joshi, T., Benedito, V.A., Xu, D., Udvardi, M.K., and Stacey, G., 2009, Legume Transcription Factor Genes: What makes legumes so special?. Plant Physiology 151: 991-1001 http://dx.doi.org/10.1104/pp.109.144105

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 5(7): 621-8 http://dx.doi.org/10.1038/nmeth.1226

Ness, R.W., Siol, M., and Barrett S.C.H., 2011, De novo sequence assembly and characterization of the floral transcriptome in cross and self-fertilizing plants, BMC Genomics 12: 298 http://dx.doi.org/10.1186/1471-2164-12-298

Patel RK, Jain M, 2012, NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data, PLoS ONE 7(2): e30619. doi:10.1371/journal.pone.0030619 http://dx.doi.org/10.1371/journal.pone.0030619

Patel, Anjaria, Panchal (2012) Leguminous Trees In Anand District: Collection and Analysis With Bioinformatics Applications. LAP LAMBERT Academic Publishing, Germany

Polhill, R.M. & Raven, P.H. (eds) 1981. Advances in Legume Systematics. Royal Botanic Gardens, Kew

Rohini Garg, Ravi K. Patel, Akhilesh K. Tyagi, and Mukesh Jain., 2011, De Novo Assembly of Chickpea Transcriptome Using Short Reads for Gene Discovery and Marker Identification, DNA RESEARCH 18, 53–63; doi:10.1093/dnares/dsq028 http://dx.doi.org/10.1093/dnares/dsq028

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

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

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

Sagar Patel, Shah D., Panchal H., Comparative study of five Legume species based on De Novo Sequence Assembly and Annotation, Computational Molecular Biology, Vol.4, No.9, 1-6 (doi: 10.5376/cmb.2014.04.0009)

Sagar Patel, Shah D., Panchal H., Conseq Tool: A Tool to Find Conserved Region in Protein Sequences of Leguminosae Family. Journal of Advanced Bioinformatics Applications and Research, Vol 5, Issue3, 2014, pp134-139, ISSN 0976-2604.Online ISSN 2278–6007

Sagar Patel, Shah D., Panchal H., De Novo RNA Seq Assembly and Annotation of Cicer arietinum L. (SRR627764). Legume Genomics and Genetics, 2014, Vol. 5, No. 6. (doi: 10.5376/lgg.2014.04.0006)

Sagar Patel, Shah D., Panchal H., De Novo RNA Seq Assembly and Annotation of Phaseolus vulgaris L. (SRR1283084), Genomics and Applied Biology, Vol.5, No.5, 1-6 (doi: 10.5376/gab.2014.05.0005)

Shi, C.Y., Yang, H., and Wei, C.L., 2011, Deep sequencing of the Camellia sinensis transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds, BMC Genomics 12: 131
http://dx.doi.org/10.1186/1471-2164-12-131

Smartt, J. & Simmonds, N.W. (eds) 1995. Evolution of Crop Plants. Harlow: Longman Scientific & Technical

Vaidya K., Ghosh A., Kumar V, Chaudhary S, Srivastava N, Katudia K, Tiwari T and Chikara K., 2012, De novo transcriptome sequencing in Trigonella foenum-graecum to identify genes involved in the biosynthesis of diosgenin. The Plant Genome:doi: 10.3835/lantgenome2012.08.0021
http://dx.doi.org/10.3835/plantgenome2012.08.0021

Wang, X.W., Luan, J.B., Li, J.M., Bao, Y.Y., Zhang, C.X., and Liu, S.S., 2010, De novo characterization of a whitefly transcriptome and analysis of its gene expression during development, BMC Genomics 11: 400
http://dx.doi.org/10.1186/1471-2164-11-400

Wang, Z., Gerstein, M., and Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics, Nat Rev Genet. 10(1): 57-63
http://dx.doi.org/10.1038/nrg2484

http://plantnet.rbgsyd.nsw.gov.au/iopi/iopihome.htm
http://www.blast.ncbi.nlm.nih.gov/Blast.cgi
http://www.blast2go.com/b2ghome
http://www.clcbio.com/products/clc-genomics-workbench/
http://www.genome.jp/kegg/pathway.html
http://www.gujaratindia.com/state-profile/demography.htm
http://www.ildis.org/
http://www.kew.org/
http://www.missouribotanicalgarden.org/
http://www.ncbi.nlm.nih.gov/
http://www.ncbi.nlm.nih.gov/genome/
http://www.ncbi.nlm.nih.gov/nuccore/
http://www.ncbi.nlm.nih.gov/protein/
http://www.nipgr.res.in/ngsqctoolkit.html
http://www.pgrc.ipk-gatersleben.de/misa/misa.html
http://www.plantgrn.noble.org/PlantTFcat/
http://www.theplantlist.org/browse/A/Leguminosae
http://www.trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software
http://www.wego.genomics.org.cn