

Bioinformatics Database of Some Leguminous Trees in Anand district of Gujarat state in India

Sagar Patel¹, Kalpesh Anjaria², Hetalkumar Panchal³

¹ G. H. Patel Post Graduate Department of Computer Science and Technology, Sardar Patel University, Vallabh Vidyanagar, Dist. Anand, Gujarat, INDIA.

² Department of Biology, Shri R. K. Parikh Arts and Science College, Petlad, Dist. Anand Pin 388450 Gujarat, INDIA.

³ Gujarat Agricultural Biotechnology Institute, Navsari Agricultural University, Surat, Gujarat- 395007, India.

***Corresponding Author**

Name: Sagar S Patel

Email: sgr308@gmail.com

Abstract: 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. Biological Databases are libraries of life sciences information, collected from scientific experiments, published literature, high throughput experiment technology, and computational analyses. Biological databases play a central role in bioinformatics. They offer scientists the opportunity to access a wide variety of biologically data centrally. Authors have tried to generate preliminary bioinformatics data for some leguminous tree 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 leguminous trees which are found in Anand district 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, Multiple Sequence Alignment, Protein sequence information, DNA sequence information, some Leguminous trees

INTRODUCTION

Biological Databases are libraries of life sciences information, collected from scientific experiments, published literature, high throughput experiment technology, and computational analyses [20]. Biological databases play a central role in bioinformatics. They offer scientists the opportunity to access a wide variety of biologically data centrally. As much as possible of a particular type of information should be available in one single place (book, site, and database). Published data may be difficult to find or access and collecting it from the literature is very time-consuming. Bioinformatics is considered as computational branch of molecular biology. This has been used by various people of different fields in their own context. Plant taxonomists are also one of them. This has been considering as one more doors towards modern trends in exploring plant taxonomy [5]. 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 some leguminous tree species found in Anand district of Gujarat state in

India which include total 31 tree species of Leguminosae family (Fig.1) [1-5, 8].

Leguminosae family is further divided into 3 subfamilies, namely, 1) Fabaceae (Papilionoideae), 2) Caesalpinoideae and 3) Mimosoideae [1,18].

METHODS

This database is divided into two parts. First part includes only botanical information of 31 tree species[1-2, 8], where as second part comprises of Bioinformatics section which includes all the information of each tree species whose data are deposited in respective database. Database creation was done with the help of Macromedia- dreamweaver MX – Software and formatting was done with html web page creation feature of this software. In this database, analysis was done with respective data of tree species DNA, PROTEIN sequences which are deposited in various biological databases [15-17]. Analysis was done for all tree species and such 330 results placed in different section.

Information were collected for 31 tree species with respect to its general description, taxonomical

classification, locality, local name, fruit and flower, uses of trees and plant's picture for each tree species (Fig 2) Data were collected from reference books, thesis and other sources [1-2, 5].

Further, out of 31 tree species, they were arranged into subfamilies such as Papilionoideae (Fabaceae) – 8 tree species, Caesalpinoideae – 12 tree species, Mimosoideae – 11 tree species. This database was divided into two parts. First part includes only botanical information for 31 tree species Botanical part contains information of each subfamily and also Key to the Genera & Species of each subfamily.

In second part which is Bioinformatics section, is further divided into 3 more sub-sections where each species was searched in NCBI database [17] and after analysis was done with available species' DNA and PROTEIN [15-17] data which is further described as,

1. Multiple sequence alignment,
2. DNA sequence information & analysis,
3. Protein sequence information & analysis

See, Figure 1 which is flow chart. It describes how this database was created.

Figures

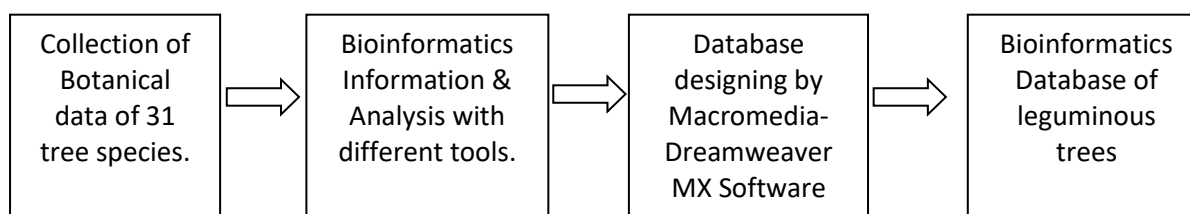


Figure 1. Methods to develop Bioinformatics database

Some Leguminous Trees In Anand District

[Home](#)
[General Information](#)
[Bioinformatics Information](#)
[Glossary](#)
[References](#)
[Tools & Softwares Used](#)
[Contact Us](#)

WHAT IS LEGUMINOUS TREES?

- A legume botanical writing is a plant in the family Leguminosae, or a fruit of these specific plants. A legume fruit is a simple dry fruit that develops from a simple carpel and usually dehisces (opens along a seam) on two sides. A common name for this type of fruit is a pod, although the term "pod" is also applied to a few other fruit types, such as vanilla and radish. Well-known legumes include alfalfa, clover, peas, beans, lentils, lupins, mesquite, carob, soy, and peanuts.
- **Fixation of nitrogen in the soil....**
 1. Legume plants are notable for their ability to fix atmospheric nitrogen, thanks to a mutualistic symbiotic relationship with bacteria (rhizobia) found in root nodules of these plants. The ability to form this mutualism reduces fertilizer costs for farmers and gardeners who grow legumes, and allows legumes to be used in a crop rotation to replenish soil that has been depleted of nitrogen. The nitrogen fixation ability of legumes is enhanced by the availability of calcium in the soil and reduced by the presence of ample nitrogen.
 2. Legume seed and foliage have a comparatively higher protein content than non-legume material, probably due to the additional nitrogen that legumes receive through nitrogen-fixation symbiosis. The high protein content makes them desirable crops in agriculture.
- **Uses by humans.....**
 1. Farmed legumes can belong to many agricultural classes, including forage, grain, blooms, pharmaceutical/industrial, fallow/green manure, and timber species. Most commercially farmed species fill two or more roles simultaneously, depending upon their degree of maturity when harvested.
 2. forage legumes such as Leucaena or Albizia are woody shrub or tree species that are either broken down by livestock or regularly cut by humans to provide livestock feed.
 3. Grain legumes are cultivated for their seeds, and are also called pulses. The seeds are used for human and animal consumption or for the production of oils for industrial uses. Grain legumes include beans, lentils, lupins, peas, and peanuts
 4. Legume species grown for their flowers include lupins, which are farmed commercially for their blooms as well as being popular in gardens worldwide.
 5. Industrially farmed legumes include Acacia species, which are cultivated for natural gum production, respectively.
 6. Fallow/green manure legume species are cultivated to be tilled back into the soil in order to exploit the high levels of captured atmospheric nitrogen found in the roots of most legumes. Numerous legumes farmed for this purpose include Sesbania species.
 7. Various legume species are farmed for timber production worldwide, including numerous Acacia species.
- **Nutritional facts...**
 1. Legumes contain relatively low quantities of the essential amino acid methionine. To compensate, some vegetarian cultures serve legumes along with grains, which are low in the essential amino acid lysine, which legumes contain.
 2. Thus a combination of legumes with grains can provide all necessary amino acids for vegetarians. Common examples of such combinations are dal with rice by Indians, and beans with corn tortillas, tofu with rice, and peanut butter with wheat bread.

Figure 2. Home Page of database.

Some Leguminous Trees In Anand District

[Home](#)
[General Information](#)
[Bioinformatics Information](#)
[Glossary](#)
[References](#)
[Tools & Softwares Used](#)
[Contact Us](#)

Erythrina variegata Linn.

(trop. & Subtrop.)

- A small to medium sized deciduous tree reaching 6-12 m. in height, bark smooth, grey and greenish-ash coloured armed with small conical dark coloured prickles.
- Leaflets broadly ovate or ovate rhomboid. terminal leaflets is the largest, coriaceous, acute or acuminate, base truncate or rhomboidal.
- Flowers appear when the tree is leafless. Flowers are bright-coral red in dense racemes.
- Pods torulose dark, glabrescent. Seeds 4-8 brown, ellipsoid-oblong.

TAXONOMY CLASSIFICATION

Kingdom:	Plantae
(unranked):	Angiosperms
(unranked):	Eudicots
(unranked):	Rosids
Order:	Fabales
Family:	Leguminosae
SubFamily:	Fabaceae
Genus:	Erythrina
Species:	variegata

Locality:	Planted in botanical garden VVN, oftenally found on farm hedegs.
Local name:	Pangaro, Pangara
Flowers :	February-April
Fruits :	March-November

USES

1. Leaves and tender shoots are consumed as potherbs; leaves are also used as cattle fodder.
2. Leaves are considered as diuretic, laxative, anthelmintic, galactagogue and emmenagogue and applied externally on venerel buboes.
3. Bark yields fibers used for cordage. Wood is used for rafts, floats, canoes and catamarans; it is also suitable for paper pulp. Seeds contain fixed oil.

GALLERY



Figure 3. Botanical information of tree species.

Some Leguminous Trees In Anand District

Home	General Information	Bioinformatics Information	Glossary	References	Tools & Softwares Used	Contact Us
----------------------	-------------------------------------	--	--------------------------	----------------------------	--	----------------------------

BIOINFORMATICS INFORMATION & ANALYSIS

[MULTIPLE SEQUENCE ALIGNMENT](#)

[DNA SEQUENCE INFORMATION & ANALYSIS](#)

[PROTEIN SEQUENCE INFORMATION & ANALYSIS](#)

Figure 4 - Bioinformatics section.

Some Leguminous Trees In Anand District

Home	General Information	Bioinformatics Information	Glossary	References	Tools & Softwares Used	Contact Us
----------------------	-------------------------------------	--	--------------------------	----------------------------	--	----------------------------

MULTIPLE SEQUENCE ALIGNMENT & ANALYSIS

[CLUSTAL W](#)

[T-COFFEE](#)

[SEQUENCE LOGO](#)

Figure 5 – Multiple sequence alignment section.

Some Leguminous Trees In Anand District

Home	General Information	Bioinformatics Information	Glossary	References	Tools & Softwares Used	Contact Us
----------------------	-------------------------------------	--	--------------------------	----------------------------	--	----------------------------

DNA SEQUENCE INFORMATION & ANALYSIS

[NCBI](#)

[EMBL](#)

[EMBOSS Transeq](#)

[GENOMATIX](#)

[ORF FINDER](#)

[GENE MARK](#)

[GENE SCAN](#)

Figure 6 – DNA sequence information & analysis section.

Some Leguminous Trees In Anand District

Home	General Information	Bioinformatics Information	Glossary	References	Tools & Softwares Used	Contact Us
----------------------	-------------------------------------	--	--------------------------	----------------------------	--	----------------------------

PLEASE CLICK ON FOLLOWING SPECIES TO GET DNA SEQUENCES..

<i>FABACEAE</i>	<i>CAESALPINIACEAE</i>	<i>MIMOSACEAE</i>
Butea monosperma	Bauhinia purpurea	Acacia farnesiana
Dalbergia sissoo	Cassia fistula	Acacia nilotica
Sesbania grandiflora	Cassia javanica	Acacia Senegal
Sesbania sesban	Cassia siamea	Albizia lebbeck
Gliricidia sepium	Delonix elata	Albizia procera
Erythrina variegata	Delonix regia	Pithecellobium dulce
	Peltophorum pterocarpum	Samanea saman
	Tamarindus indica	Prosopis chilensis
		Prosopis cineraria

Figure 7 – DNA sequence information of species.

Some Leguminous Trees In Anand District

Home	General Information	Bioinformatics Information	Glossary	References	Tools & Softwares Used	Contact Us
----------------------	-------------------------------------	--	--------------------------	----------------------------	--	----------------------------

PROTEIN SEQUENCE INFORMATION & ANALYSIS

[NCBI](#)
[UNIPROT](#)
[EMBOSS -pepinfo](#)
[TMHMM](#)
[ProtParam](#)
[GOR](#)

Figure 8 – PROTEIN sequence information & analysis section.

Some Leguminous Trees In Anand District

Home	General Information	Bioinformatics Information	Glossary	References	Tools & Softwares Used	Contact Us
----------------------	-------------------------------------	--	--------------------------	----------------------------	--	----------------------------

<i>WORD</i>	<i>DESCRIPTION</i>
<i>achene</i>	a dry, one-celled, one-seeded fruit, the product of a simple pistil.
<i>alternate</i>	when a single leaf occurs at a node.
<i>apex</i>	the portion of the leaf furthest from the petiole.
<i>arborescent</i>	having the size and form of a tree.
<i>aril</i>	a single seed surrounded by a thin or fleshy layer.
<i>base</i>	portion of a leaf nearest the petiole.
<i>berry</i>	a fleshy, several-seeded fruit with fleshy inner and outer walls.
<i>bipinnately compound</i>	when the leaflets of a compound leaf are compounded or branched a second time.
<i>blade</i>	expanded portion of a leaf.
<i>buds</i>	contain the growing points of stems.
<i>capsule</i>	a dry fruit which is the product of a compound pistil.

Figure 9- Glossary.

RESULTS

Results include bioinformatics analysis with the help of free online tools and softwares are available, each Result is kept in respective section in this database where user could find their choice of result and

information. In addition we have included information of each tool, like it's basic information and interpretation of that results are also included which

would be very beneficial to people who don't know about this bioinformatics tools and softwares.

Multiple sequence alignment (Fig.5)

Here 3 different tools have been used namely ClustalW2, TCOFFEE and Sequence Logo.

1. ClustalW2- It is a general purpose global multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen. Evolutionary relationships can be seen via viewing Cladograms or Phylograms [22, 9].

2. TCOFFEE- (Tree-based Consistency Objective Function For alignment Evaluation) is multiple sequence alignment software using a progressive approach. It generates a library of pair wise alignments to guide the multiple sequence alignment. It has advanced features to evaluate the quality of the alignments and some capacity for identifying occurrence of motifs (Mocca) [21].

3. Sequence Logo- It is a graphical representation of an amino acid or nucleic acid multiple sequence alignment. Each logo consists of stacks of symbols, one stack for each position in the sequence. The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each amino or nucleic acid at that position. By using each tool, we got total 9 results (3 results of each tool) which include multiple sequence alignment of tree species, Cladogram, Phylogram, Guide tree and logo [23]. Also information of each tools that how this tool is useful to get information and also interpretation of results also included. Tree species placed to their respective subfamily like,

1. Papilionoideae (Fabaceae) - 6 tree species
2. Caesalpinioideae - 7 tree species
3. Mimosoideae - 8 tree species.

This section includes total 9 results from three tools which have been used.

DNA sequence information & analysis (Fig.6)

This section is divided into seven parts where first two sections are of biological databases which are NCBI [17] and EMBL [16].

Here, direct link of these two databases given, so user has to just click on the name of tree species and respective information will be fetched from NCBI and EMBL online databases, user also gets DNA sequence information of tree species (Fig. 7) [17]. DNA sequence information of 23 tree species is available in NCBI and EMBL databases. This section also includes five tools for DNA sequence analysis which are mentioned below. Total 23 results for each tool were shown [24-28].

1. EMBOSS-Transeq - It translates nucleic acid sequences to the corresponding peptide sequence. It can translate in any of the 3 forward or three reverse sense frames, or in all three forward or reverse frames, or in all six frames [24].

2. Genomatix- It is used to get DNA sequence's various statistics like, AT/GC-content, mono-nucleotides, di-nucleotides, tri-nucleotides [25], [10].

3. The ORF Finder (Open Reading Frame Finder) - It is a graphical analysis tool which finds all open reading frames of a selectable minimum size in a user's sequence or in a sequence already in the database. This tool identifies all open reading frames using the standard or alternative genetic codes [26].

4. GeneMark - It determine the protein-coding potential of a DNA sequence (within a sliding window) by using species specific parameters of the Markov models of coding and non-coding regions. This approach allows delineating local variations of coding potential; therefore, the GeneMark graph shows details of the protein-coding potential distribution along a sequence [27].

5. GENSCAN- GenScan is an online program to identify complete gene structures in genomic DNA. It is based on a probabilistic model of genomic sequence composition / gene structure [28].

Total 161 results of all tools are included in this section are in .pdf file format so user could directly save results in .pdf file in his/her computer.

Protein sequence information & analysis (Fig.8)

This section is divided into six parts where first two sections are of biological databases which are NCBI [17] and UNIPROT [15].

Here, direct link of these two databases given, so user has to just click on the name of tree species and respective information will be fetched from NCBI and UNIPROT online databases, and they gets PROTEIN sequence information of tree species . PROTEIN sequence information of 23 tree species is available in databases. This section includes four tools for PROTEIN sequence analysis which are mentioned here under.

1. EMBOSS – pepinfo - It is used to get Hydrophobicity plots and Histogram of the presence of residues with the physico-chemical properties: Tiny, Small, Aliphatic, Aromatic, Non-polar, Polar, Charged, Positive, Negative. This gives various amino acid properties in parallel for an input protein sequence. Total 21 results are shown using this tool [29].

2. TMHMM – It is a membrane protein topology prediction method based on a hidden Markov model, it can discriminate between soluble and membrane proteins with both specificity and sensitivity better than 99%, although the accuracy drops when signal peptides are present. Total 21 results are shown using this tool [31].

3. ProtParam - It computes various physico-chemical properties like the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity of given protein sequence. Total 20 results are shown using this tool [30].

4. GOR: It gives two outputs, one eye-friendly giving the sequence and the predicted secondary structure in rows, H=helix, E=extended or beta strand and C=coil; the second gives the probability values for each secondary structure at each amino acid position. The predicted secondary structure is the one of highest probability compatible with a predicted helix segment of at least four residues and a predicted extended segment of at least two residues. Total 21 results are shown using this tool [32].

Total 129 such results using all above mentioned tools are included in this section which are in .pdf file format so user could directly save results in .pdf file in his/her computer.

Figure 9 describes Glossary section; which includes botanical terms used in this database.

CONCLUSIONS

This particular database served the demands of the present botanical community. So far such information is not available at one platform there by it will serve the purpose of their needs too. Various scattered data on such tree species were placed in such a manner that any botanist or biotechnologist who desire to find information on these particular trees will get at one touch or on mouse click. This particular work spreads over botanical realm to biotechnological realm through bioinformatics tools. As on one hand all botanical information such as scientific classifications, uses, images, distributions, general identification keys and characters are available on other hand any biotechnological information ranging from available kinds of proteins, number of amino acids, kind of amino acids, nucleotide sequences on particular tree species were gathered and placed systematically using bioinformatics tools in database.

Acknowledgements

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

REFERENCES

1. Shah GL; Flora of Gujarat State. Publ. by Sardar Patel University, Vallabh Vidya nagar, Anand, India. 1978.
2. Kalpesh A; Ph. D. Thesis: Floristic studies of Anand District. Submitted to Sardar Patel University, Vallabh Vidyanagar, Anand, India. 2002.
3. Jean-Michel Claverie, Cedric Notredame; Bioinformatics– A Beginner's Guide. Publ. by Wiley Publishing, Inc. USA. 2003.
4. Oza GM; Kishore S, Rajput; Biodiversity of Gujarat Forest Trees. Publ. by INSONA, Vadodara, India, 2006.
5. Patel A, Panchal; Leguminous Trees In Anand District: Collection and Analysis With Bioinformatics Applications. LAP LAMBERT Academic Publishing. 2012.
6. Sagar Patel, Panchal H; Leguminobase: A Tool To Get Information Of Some Leguminosae Family Members From Ncbi Database in Journal of Advanced Bioinformatics Applications and Research, 2013; 4(3):54-59.
7. Sagar Patel, Panchal H, Smart J, Anjaria K; Species Information Retrieval Tool: A Bioinformatics tool for Leguminosae family in International Journal of Bioinformatics and Biological Science, 2013;1(2):187-194.
8. Sagar Patel, Panchal H., Smart J., Anjaria K.; Distribution of Leguminosae family members in Gujarat State of India: Bioinformatics Approach in International Journal of Computer Science and Management Research, 2013; 2(4): 2184-2189.
9. Sagar Patel, Panchal H, Anjaria K; Phylogenetic analysis of some leguminous trees using CLUSTALW2 Bioinformatics Tool. Publication Year: 2012; 917- 921. Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Philadelphia, USA, Oct 4-7, 2012.
10. Sagar Patel, Panchal H, Anjaria K; DNA Sequence analysis by ORF FINDER & GENOMATIX Tool: Bioinformatics Analysis of some tree species of Leguminosae Family; 2012, 922- 926. Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Philadelphia, USA.
11. Available online from www.flickr.com/photos
12. Available online www.amrita.edu/ces/trees/index.html
13. Available online www.plants.usda.gov
14. Available online www.fao.org/ag/AGP/AGPC/doc/Gallery/pic.htm
15. Available online www.uniprot.org/
16. Available online www.ebi.ac.uk/embl/

-
17. Available online www.ncbi.nlm.nih.gov/
 18. Available online www.en.wikipedia.org
 19. Available online www.worldagroforestrycentre.org/sea/Products/AFDbases/af/asp/
 20. Available online www.flowersofindia.net
 21. Available online TCOFFEE - [www.igs.cnrs-mrs.fr Tcoffee/tcoffee_cgi/index.cgi](http://www.igs.cnrs-mrs.fr/Tcoffee/tcoffee_cgi/index.cgi)
 22. Available online ClustalW2- www.ebi.ac.uk/Tools/msa/clustalw2/
 23. Available online Sequence LOGO- weblogo.berkeley.edu/logo.cgi
 24. Available online EMBOSS-Transeq- www.ebi.ac.uk/Tools/emboss/transeq/index.html
 25. Available online Genomatix- www.genomatix.de/cgi-bin/tools/tools.pl
 26. Available online ORFFINDER- www.ncbi.nlm.nih.gov/projects/gorf/
 27. Available online GeneMark- exon.gatech.edu/gmhmm2_prok.cgi
 28. Available online GENSCAN- genes.mit.edu/GENSCAN.html
 29. Available online EMBOSS-pepinfo- www.ebi.ac.uk/Tools/emboss/pepinfo/index.html
 30. Available online ProtParam- expasy.org/tools/protparam.html
 31. Available online TMHMM- www.cbs.dtu.dk/services/TMHMM/
 32. Available online GOR-npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html