Computational Molecular Biology

An integration of experimental molecular and genome biology with computational technology

2014
Vol.4
Computational Molecular Biology is an open access, peer reviewed journal publishing original research papers of general interest involving in the computational biology at the molecular level. The Journal is publishing all the latest and outstanding research articles, letters, methods, and reviews in all areas of Computational Molecular Biology, covering new discoveries in molecular biology, from genes to genomes, using statistical, mathematical, and computational methods as well as new development of computational methods and databases in molecular and genome biology.

The papers published in the journal are expected to be of interests to computational scientists, biologists, and teachers/students/researchers engaged in biology, as well as are appropriate for R & D personnel and general readers interested in computational technology and biology.

Computational Molecular Biology is published independently by BioPublisher. It is committed to publishing and disseminating significant original achievements in the related research fields of molecular biology.

All the articles published by BioPublisher are Open Access, and are distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

BioPublisher uses CrossCheck service to identify academic plagiarism through the world's leading anti-plagiarism tool, iThenticate, and to protect the original authors' work.
Latest Content

**PlantSecKB: the Plant Secretome and Subcellular Proteome KnowledgeBase**
Gengkon Lum, John Meinken, Jessica Orr, Stephanie Frazier, Xiang Min
Computational Molecular Biology, 2014, Vol.4, No.1

**GC2 Biology Dictates Gene Expressivity in *Camellia sinensis***
Supriyo Chakraborty, Prosenjit Paul
Computational Molecular Biology, 2014, Vol.4, No.2

**Association Rules for Diagnosis of Hiv-Aids**
Anubha Dubey
Computational Molecular Biology, 2014, Vol.4, No.3

**In Silico Proteomic Functional Re-annotation of *Escherichia coli* K-12 using Dynamic Biological Data Fusion Strategy**
Ramesh Gopal, Subazini Thankaswamy Kosalai, Rajadurai Chinnasamy Perumal, Palani Kannan Kandavel
Computational Molecular Biology, 2014, Vol.4, No.4

**Long Non-coding RNAs: key players in brain and central nervous system development**
Jie Lv, Hui Liu, Hongbo Liu, Qiong Wu, Yan Zhang
Computational Molecular Biology, 2014, Vol.4, No.5

**Phylogeny in Few Species of Leguminosae Family Based on *matK* Sequence**
Sagar Patel, Dipti B. Shah
Computational Molecular Biology, 2014, Vol.4, No.6
Phylogeny in Few Species of Leguminosae Family Based on \textit{matK} Sequence

Sagar S. Patel, Diptri B. Shah

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

Abstract In this paper, few species of Leguminosae family considered for phylogenetically analyses which are found in Gujarat state in India and \textit{matK} gene sequence data from NCBI database are considered for evolutionary analysis. The sequence data of the \textit{matK} gene are more accurate than \textit{rbcL} sequence data in the reconstruction of phylogenies throughout the seed plants. Leguminosae family is one of the largest families that contain thousands of species of Plants, Herbs, Shrubs and Trees worldwide. This study shows that species of Leguminosae family which is further classified into Fabaceae (Papilionaceae), Mimosaceae and Caesalpiniaceae; based on morphological characters has different members and the based on the DNA and protein \textit{matK} sequence data analysis, few species are not related with each other as per morphological classification. We conclude that few species are related with each other as per botanical or morphological classification of Leguminosae family but evolutionary results shows that based on DNA and protein \textit{matK} sequence data some species are not related with morphological or taxonomical classification.

Keywords Leguminosae family; Bioinformatics; NCBI; \textit{matK}

1 Introduction

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. The legume family is the third largest family of angiosperms (Mabberley, 1997) with approximately 730 genera and over 19,400 species worldwide (Lewis et al., in press). 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. The predilection of legumes for semi-arid to arid habitats is related to a nitrogen-demanding metabolism, which is thought to be an adaptation to climatically variable or unpredictable habitats whereby leaves can be produced economically and opportunistically (McKey, 1994), (Wojciechowski et al). Leguminosae family is further classified into three subfamilies; Fabaceae (Papilionaceae), Caesalpiniaceae and Mimosaceae (http://en.wikipedia.org).

1.1 \textit{matK} gene

The \textit{matK} gene, formerly known as orfK, is emerging as yet another gene with potential contributions to plant molecular systematics and evolution (Johnson and Soltis, 1994, 1995; Steele and Vilgalys, 1994; Liang and Hilu, 1996; Gadek, Wilson, and Quinn, in press). The gene, ~1500 base pairs (bp), is located within the intron of the chloroplast gene \textit{trnK}, on the large single-copy section adjacent to the inverted repeat (Figure 1). Further, the molecular information generated from \textit{matK} has been used to resolve phylogenetic relationships from shallow to deep taxonomic levels (Johnson and Soltis, 1994; Hayashi and Kawano, 2000; Hilu et al., 2003; Cameron, 2005).
1.2 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 (http://www.ncbi.nlm.nih.gov).

1.3 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 (http://en.wikipedia.org). 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 (http://www.ncbi.nlm.nih.gov/nuccore/).

1.4 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 (http://en.wikipedia.org). 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 (http://en.wikipedia.org), (http://www.ncbi.nlm.nih.gov/protein/).

2 Materials and Methods

In this paper we have considered around 266 species which are found in Gujarat state of India (Sagar Patel et al., 2013). Further we searched each species in NCBI database and finally found around 149 species’ information like DNA, Protein and other useful information of Leguminosae family (Sagar Patel et al., 2014). Further we have only considered matK gene sequences of DNA and Protein sequences. Evolutionary analysis done in MEGA software by Maximum Likelihood method (Bootstrap method) (Tamura et al., 2011) as shown in Figure 2.

3 Results

3.1 Result of DNA matK gene sequences

As shown in above Figure 3 which is result of DNA matK Sequences by Maximum Likelihood method (bootstrap method), starting from top we observed that species are placed in subfamily wise; like first, Fabaceae (Papilionaceae), Mimosaceae followed by Caesalpiniaceae but First and last species is from Fabaceae (Papilionaceae) subfamily, so species of Mimosaceae and Caesalpiniaceae subfamilies are included within Fabaceae (Papilionaceae). Starting from top species of Fabaceae (Papilionaceae) subfamily are present in which species of genus Medicago, Crotoraria, Sesbania, Vigna, Tephrosia, Butea and Trigonella genus are related as per morphological characters or botanical classifications except Medicago lupulina, Vigna radiata, Vigna
Phylogeny in Few Species of Leguminosae Family Based on matK Sequence

...unguiculata which is distantly related to their respective genus and each species of genus Lathyrus, Vicia and Vigna is distantly related with other species.

Figure 3 Result of Maximum Likelihood (Bootstrap Method) of matK DNA sequences of Leguminosae Family

After then species of Mimosaceae subfamily are there in which species of genus Albizia and Acacia genus are related as per morphological characters or botanical classifications and species of Prosopis genus are distantly related with each other.

Then after there are species which belongs to Caesalpiniaceae subfamily; in that species of Cassia genus are related as per morphological characters or botanical classifications and species of Caesalpinia, Delonix and Bauhinia genus are distantly related with each other.

3.2 Result of protein matK gene sequences

As shown in Figure 4 which is result of Protein matK Sequences by Maximum Likelihood method (bootstrap method), starting from top we observed that species are placed in subfamily wise; like first, Caesalpiniaceae, Mimosaceae followed by Fabaceae (Papilionaceae). Starting from top, there are species which belongs to Caesalpiniaceae subfamily; in that species of Cassia, Delonix and Bauhinia genus are related as per morphological characters or botanical classifications and species of Caesalpinia genus are distantly related with each other (Figure 4).

Figure 4 Result of Maximum Likelihood (Bootstrap Method) of matK Protein sequences of Leguminosae Family

After then species of Mimosaceae subfamily are there in which species of genus Albizia and Acacia genus are related as per morphological characters or botanical classifications except Acacia senegal which found between species of Albizia species.

After then species of Fabaceae (Papilionaceae) subfamily are present in which species of genus Medicago, Crotonaria, Canavalia, Sesbania, Tephrosia, Vicia, Butea and few species of Vigna are related as per morphological characters or botanical classification except species of genus Lathyrus and Trigonella are distantly related with their nearby species of same genus.

4 Discussion

In this study we observed that species belongs to Leguminosae Family; which is further classified into Fabaceae (Papilionaceae), Mimosaceae, Caesalpiniaceae are as per the botanical classification classified differently based on their morphological features like species’ flower color, size and shape, types and arrangements of Stipules, size of plant etc. But this study focus on evolutionary relationship of Leguminosae Family species based on DNA & Protein sequences of matK sequences with Multiple sequence alignment by Maximum likelihood where
we observed that in matK protein sequences result; some species belonging to same genus are fall very nearly as per botanical classification which is correct as per both botanical and evolutionary relationship but we observed in matK DNA sequence result that it really differs and it is not related with morphologically or botanical classification and further we observed that few species are distantly related even if they are from same genus. Further conserved matK protein sequences could be model and functional annotation may give accurate information regarding to evolution as structural proteins are more accurate in evolution study which gives accurate details regarding to study.

As per literature review we come to know that matK sequences are more accurate than rbcL sequences which are normally used for phylogeny reconstruction and after this analysis we also recommend that matK sequences are more accurate than rbcL gene sequences and we suggest from our study that especially matK protein sequences gives more accurate result on evolutionary or phylogeny study than matK DNA sequences.

Acknowledgement
We would like to thank Sardar Patel University.

References


Martin F. Wojciechowski,matt Lavin, michael J. Sanderson. A Phylogeny Of Legumes (Leguminosae) Based On Analysis Of The Plastid Matk Gene Resolves Many Well-supported Subclades Within The Family


Neuhaus H. Link G. 1987. The chloroplast tRNALys (UUA) gene from mustard (Sinapis alba) contains a class II intron potentially coding for a maturase-related polypeptide. Current Genetics 11: 251-257 http://dx.doi.org/10.1007/BF00355398

Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 5 (Tamura, Peterson, Stecher, Nei, and Kumar 2011)

Sagar Patel, and Hetal Kumar Panchal. 2014, Bioinformatics Information of Leguminosae Family in Gujarat State, International Journal of Agriculture, Environment & Biotechnology 7.1

Sagar Patel, and Hetalkumar Panchal, 2014, Evolutionary studies of few species belonging to Leguminosae family based on RBCL gene. Discovery, 9(22): 38-50


Reasons to publish in BioPublisher

A BioScience Publishing Platform

★ Peer review quickly and professionally
☆ Publish online immediately upon acceptance
★ Deposit permanently and track easily
☆ Access free and open around the world
★ Disseminate multilingual available

Submit your manuscript at: http://bio.sophiapublisher.com