LSeqDB: a repository for annotation and storage of biological sequences of legumes

José Gomes Lopes Filho, Luiz Augusto Macedo Morais, Franciandro Dantas dos Santos, José Wilker Lima Silva, Carlos Eduardo Alves Soares

Universidade Estadual da Paraíba, Universidade Estadual da Paraíba, Universidade Federal de Campina Grande, Universidade Estadual da Paraíba, Universidade Federal de Campina Grande

Abstract

Background: The Fabaceae is one of the most important Botany families. Economic, medicinal, food and environmental are some aspects that we can say about importance of Leguminoseae for human beings. Nowadays, other legume interesting issue is a noticeable increasing of DNA and protein sequences of this botanical group in public biological databases. On the other hand, a Brazilian web specific repository produced to storage and analyze legume sequences by Bioinformatics tools is not idealized yet. The main goal of our work was produce a new approach web database with searching, automating the signature collection, distribution and analyzing of Fabaceae data. Considering the growing of this specific biological information, it has been an useful initiative that stimulated the development of techniques and other resources for organize and store these kind of data for academic or research uses. So our web resource called LSeqDB (Legume Sequence Data Base) could be used as a basis for further studies and consultations. The aim of this study was to develop a database (DB) as a Bioinformatics tool for the web for storage and annotation of biological sequences of legumes. Results: Our DB was also created for users who want study Bioinformatics with Leguminosae Sequences. The DB was built using MySQL, the language Python and its frameworks Django and Biopython. For analysis sequences the user also could perform BLAST options (BLASTp, BLASTn, tBLASTn and tBLASTx). LegSeqDB repository model was based on three tables: Organism, Sequence and User. Each of them with their restrictions. For example, to add an organism in the LSeqDB, the user has to link the table ID organism to its respective sequence in fasta format (fasta was chose as the unique format option for working). For more entrances, we developed a script Python to optimize the velocity of adds by user. To accomplish the annotation of sequences in the database, the website is divided in two sections: one for insertion and the other for query of sequences. The languages HTML, CSS and Javascript were also used in this web construction. Conclusion: Finally, LegSeqDB was made to be a first Brazilian repository web with Bioinformatics tools for data mining in legume biological information.

Keywords:

Biological Sequences, Databases, Data Mining