Available online at www.scholarsresearchlibrary.com

Extended Abstract

Journal of Computational Methods in Molecular Design, 2021, 11(5) **ISSN 2231-3176**
https://www.scholarsresearchlibrary.com/journals/journal-of-computational-methods-in-molecular-design/

Insilco studies of mango genome cultivars and development of analysis tool

Rabia Faizan, Sir Syed University of Engineering & Technology Karachi, Pakistan
E-mail: rabiatabassumkhi@gmail.com

Mango is one of the famous and fifth most important subtropical/tropical fruit crop worldwide with the production centred in India and South-East Asia. Mango breeding is still mostly the selection of trees with superior fruit traits rather than hybrid production and evaluation due to the difficulty of specific pollination. Recently, there has been a worldwide interest in mango genomic to produce tools for Market Assisted Selection and trait association. There are five transcriptomes produces respectively in Israel, India, Mexico, Pakistan and the US. Our project "Mango Genome Database" is constructed upon the completion of the Pakistan leaf transcriptome and chloroplast genome of Mangifera Indica L. specie sequencing with the aim of providing the scientific community with an accurate data and annotation of the mango genome sequence. The project is consisting of a website which contains the information of predicted genes of the whole genome of Mango and the unigenes annotated in other species, GO (Gene Ontology) terms providing a glimpse of the pathways and traits in which they are involved. The annotated and analysed sequence data then can be browsed through Gene search page or queried using various categories in the search sites. The whole genome sequences and genetic results can also be accessed and downloaded through different modules present in a website including (Blast, Data Downloads and BI Tools). The online analysis tools involve the Blast server for the Mango datasets, a sequence gene retrieval module, Phylogeny, plotting and other alignment tools. As a member of the family Anacardiaceae, mango (Mangifera Indica Linn.) ranks second among 3 tropical fruit crops after banana due to its rich sensational taste, color, aroma and huge 4 economics significance (Litz RE, 2009; Srivastava S, 2016). According to Food and Agriculture Organization of the United Nations (FAO), India holds the 1st position in mango production followed by China, whereas Pakistan and Mexico rank 5th and 6th position respectively. In the Ayurvedic and indigenous medical systems, it counts for an important herb for nearly over 4000 years (KA Shah, 2010). The substances in mango have high therapeutic potential. According to Ayurveda, mango tree comprises of different medicinal properties could be used to treat tumor, rabid dog, piles, toothache, diarrhea, cough, insomnia, hypertension, asthma, anemia, hemorrhage, dysentery, heat stroke, blisters, miscarriage, liver disorders, tetanus, wounds in the mouth and excessive urination. Phytomedicines should be sufficiently regulate based on this knowledge (Kalita, 2014). Mango has a small genome size of 439 Mb (2n = 40) and a diploid fruit tree with 20 pairs of chromosomes (Arumuganathan, 1991). There are total 72 species of genus Mangifera from which most of them surviving in the rain forests of Malaysia and Indonesia (Nagendra KSingh, 2016). Mango has been widely cultivated in India and Southeast Asia for thousands of years. It has now grown throughout the world (tropical and subtropical) in 99 countries with a total fruit production of 34.3 million tons of fruit per annum (Galán Saúco, 2013). Asia is the largest producer of mangoes, 76% of world production comes from Asia with the second and third largest producers Americas (12%), and Africa (11.8%) (Galán Saúco, 2013). Mango is a rich source of vitamins (vitamin A and vitamin C) and minerals and popular for its attractive color, texture and juicy flavor covers the total area of 2,297,000 ha in 2010 to 2011 (Hiwale1, 2015). The chemical analysis of mango pulp evinces that it has a relatively high content of calories (60 Kcal/100 g fresh weight) and is an important source of potassium, fiber, and vitamins (Marianna Lauricella, 2017). The nutritive value of mango is listed in (Table 1) reporting some data from the National Nutrient Database for Standard Reference (United States Department of Agriculture) (USDA, 2016). This project aims to build a successful response to the difficulties currently faced by fruit science agencies in meeting the need of gradually increasing fruit demands. The project intended to contribute to the development of a framework designed to support the analysis of the vast data for the availability of mango genomic information. It will also provide a platform for genetic information of different varieties of Mangifera Indica from different countries to make the estimated data available for the scientific community and general public. Data Mangifera Indica Linn. usually known as mango, is a species of the flowering plant in the sumac and poison ivy family Anacardiaceous. In this project four mango cultivars cv. Langra from Pakistan, cv. Zill from China, cv. Kent from Mexico and cv. Shelly from Israel transcriptomic sequences or the unigenes which are generated by processing the RNA-seq reads and transcriptome de novo assembly reported in 2014 (Azim MK, 2014) are retrieved, analyzed and presented in a database. We obtained four unigenes datasets corresponding to four mango cultivars from JamilurRahman Center for Genome Research, Dr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi. The Blast Analysis or sequence comparison of unigenes of four mango cultivars were done against two databases: non–redundant (NR) protein sequence and Swiss–Prot. First, all unigenes of Pakistan dataset (cv. Langra) was aligned against non–redundant (NR) protein sequence database using BLASTX (E value cutoff ≤ 1e−3 and 10 hits per sequence) in Blast2GO java application to retrieve common transcripts. Obtained results were then subjected to mapping and annotation for functional annotation and Gene Ontology (GO) assignments. This analysis was done only with cv. Langra from Pakistan for the Gene search module of the Website. Secondly, the four unigenes datasets are filtered for redundant sequences using CD-HIT (Cluster Database at High Identity with Tolerance) (Fu L, 2012) at 90% identity threshold. The resultant non-redundant unigenes were further analyzed for coding regions using TransDecoder. Obtained coding sequences (CDS) were then subjected to homology search and BLASTed against Swiss–Prot databases by BLASTP with an E value cutoff ≤ 1e−3 to find the common and unique transcripts.

Bottom Note: *This work is partly presented at International Conference on BIOINFORMATICS & SYSTEM BIOLOGY* March 20-21, 2019 | Singapore City, Singapore