Indian Journal of History of Science, 51.2.2 (2016) 355-368 DOI: 10.16943/ijhs/2016/v51i2.2/48449

Origin, Diversity and Genome Sequence of Mango (*Mangifera indica* L.)

Nagendra K Singh[1], Ajay K Mahato[1], Pawan K Jayaswal[1], Akshay Singh[1], Sangeeta Singh[1], Nisha Singh[1], Vandna Rai[1], Amitha Mithra S V[1], Kishor Gaikwad[1], Nimisha Sharma[2], Shiv Lal[2], Manish Srivastava[2], Jai Prakash[2], Usha Kalidindi[2], Sanjay K Singh[2], Anand K Singh[2], Kasim Khan[3], Rupesh K Mishra[3], Shailendra Rajan[3], Anju Bajpai[3], B S Sandhya[4], Puttaraju Nischita[4], K V Ravishankar[4], M R Dinesh[4], Neeraj Kumar[5], Sarika Jaiswal[5], Mir A Iquebal[5], Dinesh Kumar[5], Anil Rai[5] and Tilak R Sharma[1]

(Received 12 February 2016)

Abstract

Mango (*Mangifera indica* L.) is known as the 'king of fruits' for its rich taste, flavor, color, production volume and diverse end usage. It belongs to plant family Anacardiaceae and has a small genome size of 439 Mb (2n = 40). Ancient literature indicates origin of cultivated mango in India. Although wild species of genus *Mangifera* are distributed throughout South and South-East Asia, recovery of Paleocene mango leaf fossils near Damalgiri, West Garo Hills, Meghalaya point to the origin of genus in peninsular India before joining of the Indian and Asian continental plates. India produces more than fifty percent of the world's mango and grows more than thousand varieties. Despite its huge economic significance genomic resources for mango are limited and genetics of useful horticultural traits are poorly understood. Here we present a brief account of our recent efforts to generate genomic resources for mango and its use in the analysis of genetic diversity and population structure of mango cultivars. Sequencing of leaf RNA from mango cultivars 'Neelam', 'Dashehari' and their hybrid 'Amrapali' revealed substantially higher level of heterozygosity in 'Amrapali' over its parents and helped develop genic simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers. Sequencing of double digested restriction-site-associated genomic DNA (ddRAD) of 84 diverse mango cultivars identified 1.67 million high quality SNPs and two major sub-populations. We have assembled 323 Mb of the highly heterozygous 'Amrapali' genome using long sequence reads of PacBio single molecule real time (SMRT) sequencing chemistry and predicted 43,247 protein coding genes. We identified in the mango genome 122,332 SSR loci and developed 8,451 Type1 SSR and 835 HSSR markers for high level of polymorphism. Among the published genomes, mango showed highest similarity with sweet orange (*Citrus sinensis*). These genomic resources will fast track the mango varietal improvement for high productivity, disease resistance and superior end use quality.

Key words: Distribution, Genome sequence, Mango (*Mangifera indica* L.), Origin, Population structure, SSR/SNP markers

[1]ICAR-National Research Centre on Plant Biotechnology, Pusa Campus, New Delhi-110012. Email: nksingh4@gmail.com
[2]ICAR-Indian Agricultural Research Institute, Pusa Campus, New Delhi-110012
[3]ICAR-Central Institute for Subtropical Horticulture, Lucknow-226101
[4]ICAR-Indian Institute of Horticultural Research, Bengaluru-560089
[5]ICAR-Indian Agricultural Statistics Research Institute, New Delhi-110012

1. Origin, Domestication and Geographical Distribution of Mango

Mango (*Mangifera indica* L.) is a member of the plant family Anacardiaceae (cashews family), order Sapindales, class Magnoliopsida and division Tracheophyta (vascular plants). It is a diploid fruit tree with 20 pairs of chromosomes and a small genome size of 439 Mbp [Arumuganathan and Earle (1991)]. There is consensus among the historians and horticulturists that the cultivated mango has originated in India [Hooker (1876); Mukherjee (1951, 1953, 1972); Woodrow (1904), p.13]. Vavilov (1926) has suggested Indo-Burma region as the centre of origin of mango based on the observed level of genetic diversity. Mukherjee (1951) considered origin of genus *Mangifera* probably in the South-East Asia but the origin of cultivated mango in the Assam-Burma region. Scientists of the Birbal Sahni Institute of Palaeobotany, Lucknow, have traced the origin of genus *Mangifera* from 60 million years old fossil compressions of carbonized mango leaves in the Palaeocene sediments near Damalgiri, West Garo Hills, Meghalaya and named it *Eomangiferophyllum damalgiriensis* (Mehrotra et al., 1998). Extensive comparison of the anatomy and morphology of several modern-day species of the genus *Mangifera* with the fossil samples reinforced the view that North-East India is the centre of origin of mango genus, from where it has spread into neighboring areas of South-East Asia after the formation of land connection following collision of the Indian plate with the Asian plate and after that species diversified extensively in the Malaysian and Sumatran rain forests. There are total 72 species of genus *Mangifera* today most of them surviving in the rain forests of Malaysia and Indonesia (Fig. 1). Apart from the widely cultivated mango *Mangifera indica*, there are seven other species cultivated for fresh fruits to a limited extent, viz. *M. sylvatica* Roxb. in

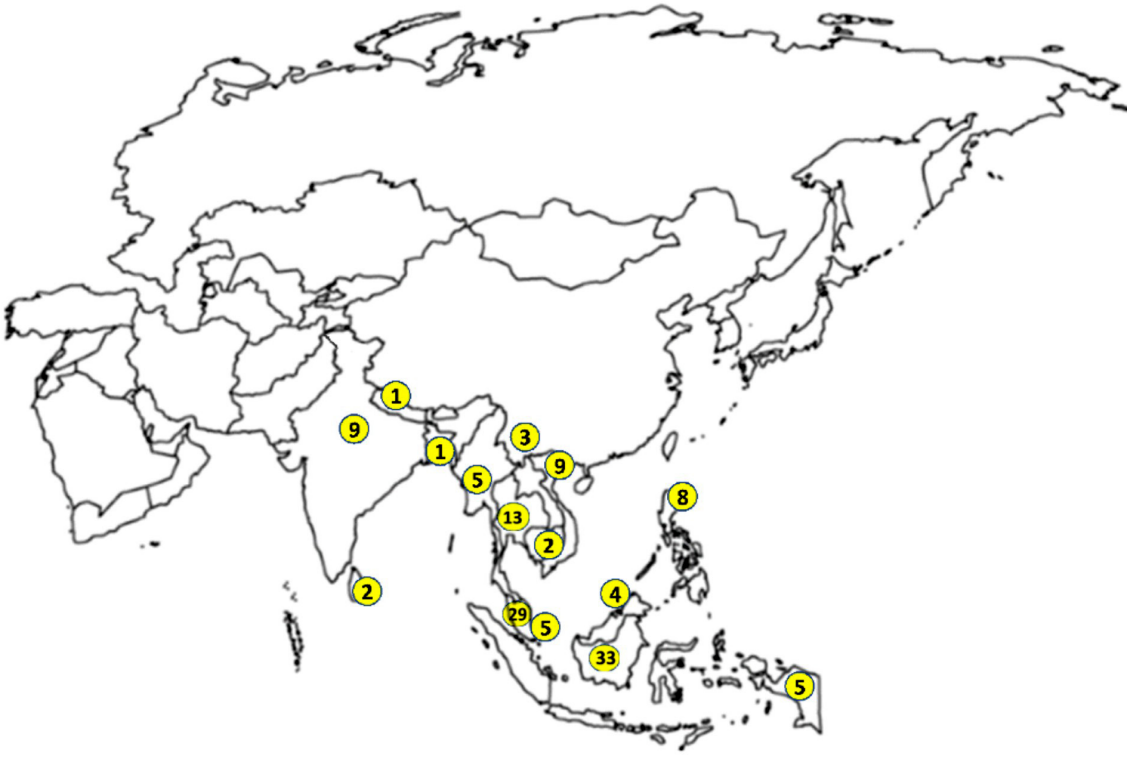

Fig. 1. Spatial distribution of 72 different wild species of geneus *Mangifera* in Asia. The numbers in the circles indicate number of wild species growing naturally in the respective countries.

Andaman, Nepal and Eastern Himalayas; *M. foetida* Lour*, M. caesia* Jack in Malaysia, Philippines and Indonesia*, M. odorata* Griff. in Malaysia and Philippines. Among these species *M. sylvatica* Roxb has the largest tree size of up to 50 meters. Other less popular species of wild mango cultivated by the Malayan villagers include *M. longipetiolata* King, *M. maingayi* Hook f., *M. kemanga* Blume, and *M. pentandra* Hook f. (Bompard, 1992, p.207; Kostermans, 1993; Salma, 2010, p.90; Williams, 2012, p.224).

Today the highest diversity of wild mango species occurs in Malaysia and Indonesia, particularly in peninsular Malaya, Borneo and Sumatra. The natural occurrence of *Mangifera* species extends as far north as 27°N latitude and as far East as the Caroline Islands [Bompard and Schnell (1997), pp. 21-48]. Wild mango species are also found in India, Sri Lanka, Bangladesh, Myanmar, Thailand, Kampuchea, Vietnam, Laos, Southern China, Singapore, Brunei, the Philippines, Papua New Guinea and the Solomon and Caroline Islands (Kole, 2011, p.64; Litz, 2009, p.26; Sreekumar, 1996; Orwa, 2009; Wu, 1979, p.368).

Although mango has been planted in India since time immemorial, earliest written records are present in ancient Sanskrit literature of pre-Buddhist era. Valmiki Ramayan, regarded as the earliest epic poetry after the Vedas, which after a long oral tradition was written down around 500 BC has several references to mango plantations, e.g. "वधू नाटक स्नधैः च संयुक्ताम सर्वतः पुरीम। उद्यान आम्र वणोपेताम महतीम साल मेखलाम।।" (Balkand, 1-5-12), "the city of Ayodhya accommodates groups of danseuses and theatrical personnel, and is surrounded everywhere with the gardens and brakes of mango trees, and her wide fort-wall is like her cincture ornament" (1-5-12). Similarly, Varah Puran (172.39) says that "One who plants one peepal (*Ficus religiosa*), one neem (*Azadirachta indica*), one Banyan (*Ficus benghalensis*), two pomegranates (*Punica granatum*), two orange (*Citrus reticulate*), five mango trees (*Mangifera indica*) and ten flowering plants or creeper shall never go the hell" (Renugadevi, 2012).

English word mango originated from Malayalam "manga" and Tamil "mangai" (http://www.oxforddictionaries.com/definition/english/mango). After its domestication in India more than 4000 years ago, traders, travelers and rulers have taken mango for plantation in different subtropical regions of the world over the last 2,500 years (Fig. 2). During 4-5th centuries BC the Buddhist monks took mango to Malaya Peninsula and East Asia. Mango was first introduced in China from India during middle of the 7th century AD when Chinese traveler Hwen T'sang returned from India to China with the mango (Tang Dynasty; Litz, 2009, p.10; Gao et al., 2011). Further, in the 10th century AD the Persians carried it to East Africa (Purseglove, 1969). During 16th century AD the Portuguese have taken it to West Africa and Brazil (Litz, 2009, p.10). After becoming established in Brazil, the mango was carried to the West Indies, being first planted in Barbados about 1742 and later in the Dominican Republic. It reached Jamaica about 1782 and, early in the 19th Century it reached Mexico from the Philippines and the West Indies (Morton, 1987, pp. 221-239). Mango reached Miami in 1862 or 1863 from the West Indies and it is believed seedling was polyembryonic and from 'No.11' parent (Litz, 2009, p.10). In same decade, about 40 varieties of Mangoes from India were initially planted in 1875 in North Queensland Australia after post-European colonization (Morton, 1987, pp.221-239).

2. Mango Production, Varietal Diversity and Trade

Mango production occupies a close second position after banana among the tropical fruits and is known as the 'king of fruits' for its rich taste, flavor, color, huge variability and varied end usage. Commercial varieties of mango have

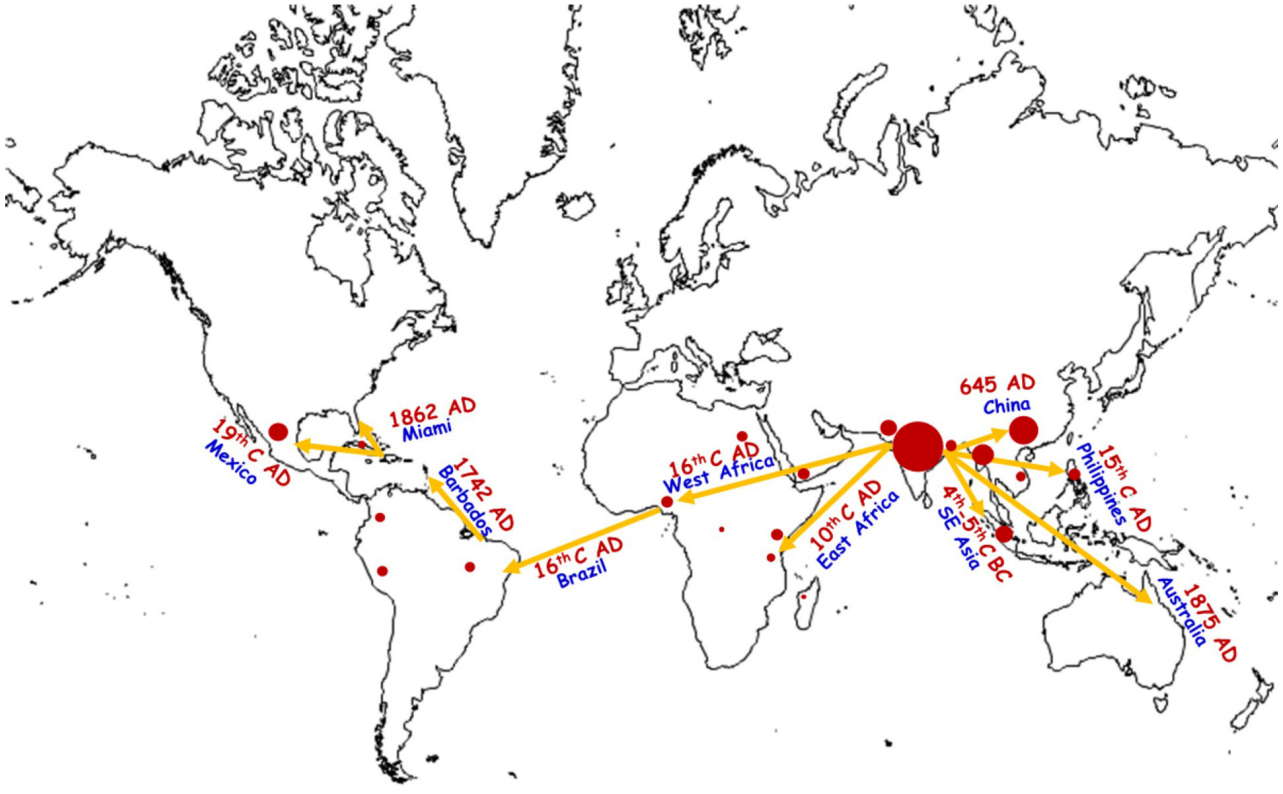

Fig. 2. Production volume, represented by area of the red circles, in the top twenty mango producing countries (FAOSTAT 2013) and adoption of mango cultivation in different regions of the world over the last 2500 years after its origin and domestication in India.

mostly originated from selection of variants resulting from recombination and segregation of characters in the progenies and then spread to the rest of the world. There are hundreds of mango cultivars distributed throughout the world, of which Asia and particularly India has over 1000 varieties of which more than 500 are fully characterized. Perhaps some of these varieties are duplicates with different names, but at least 350 are propagated in commercial nurseries. However, in the Western Hemisphere, a few cultivars derived from a breeding program in Florida are the most popular for international trade, e.g. Irwin, Tomy Atkins. Many mango cultivars are grown locally often as seedling trees as a backyard food source (Rieger, 2001). The Horticulture Research Unit of the U.S. Department of Agriculture and the Agricultural Research and Education Centre of the University of Florida, together maintain a germplasm of 125 mango cultivars as a resource for mango growers and breeders in many countries. Worldwide India is the largest producer of mango with 18 Mmt in 2012-13 contributing about 50% of global production (Handbook on Horticulture Statistics 2014, http://agricoop.nic.in/imagedefault/whatsnew/handbook2014.pdf) from 2.5 Mha of cultivated area (Fig. 2), followed by China and Thailand (4.3 and 2.6 Mmt, FAOSTAT 2011). Information released under National Data Sharing and Accessibility Policy (NDSAP) the major mango producing states in India are Uttar Pradesh (23.85%), Andhra Pradesh (22.14%), Karnataka (11.71%), Bihar (8.78%), Gujarat (5.99%) and Tamil Nadu (5.42%). Apart from the production India's export of mangoes was 203,000 tons (Fresh mango and mango pulp) and the total value was 87327.51 lakhs during 2012-13. India mainly exports fresh mangoes to United Arab

Emirates (23046.65 MT), Kuwait (4601.44 MT), United Kingdom (3381.08 MT), Bangladesh (2899.85 MT) and Saudi Arabia (1721.91 MT) during 2013-14. Apart from fresh fruit, mango is used in the form of pulp, pickle, jam, jelly, chutney, powder, while mango shake is a popular drink all over the world. However, there are only about 25 major commercial cultivars some of which are highly preferred in the international market. According to Agricultural and Processed Food Products Export Development Authority (APEDA) India exported 41,280 tons of mangoes worth around 50.7 million USD during 2013-14. Despite its huge economic significance limited genomic resources are available and genetics of useful horticultural traits are poorly understood.

3. Diversity and Population Structure of Mango Cultivars based on Genome Wide SNP Markers

Molecular diversity analysis and fingerprinting of mango cultivars has been carried out using different types of DNA markers, including random amplified polymorphic DNA (RAPD) by Schnell et al. (1995); Lopez-Valenzuela et al. (1997); Ravishankar et al. (2000); Kumar et al. (2001); Karihaloo et al. (2003); inter simple sequence repeats (ISSR) by Eiadthong et al. 1999; Pandit et al. (2007) and Singh et al. (2007); amplified fragment length polymorphism (AFLP) by Eiadthong et al. (2000); Kashkush et al. (2001) and simple sequence repeats (SSR) by

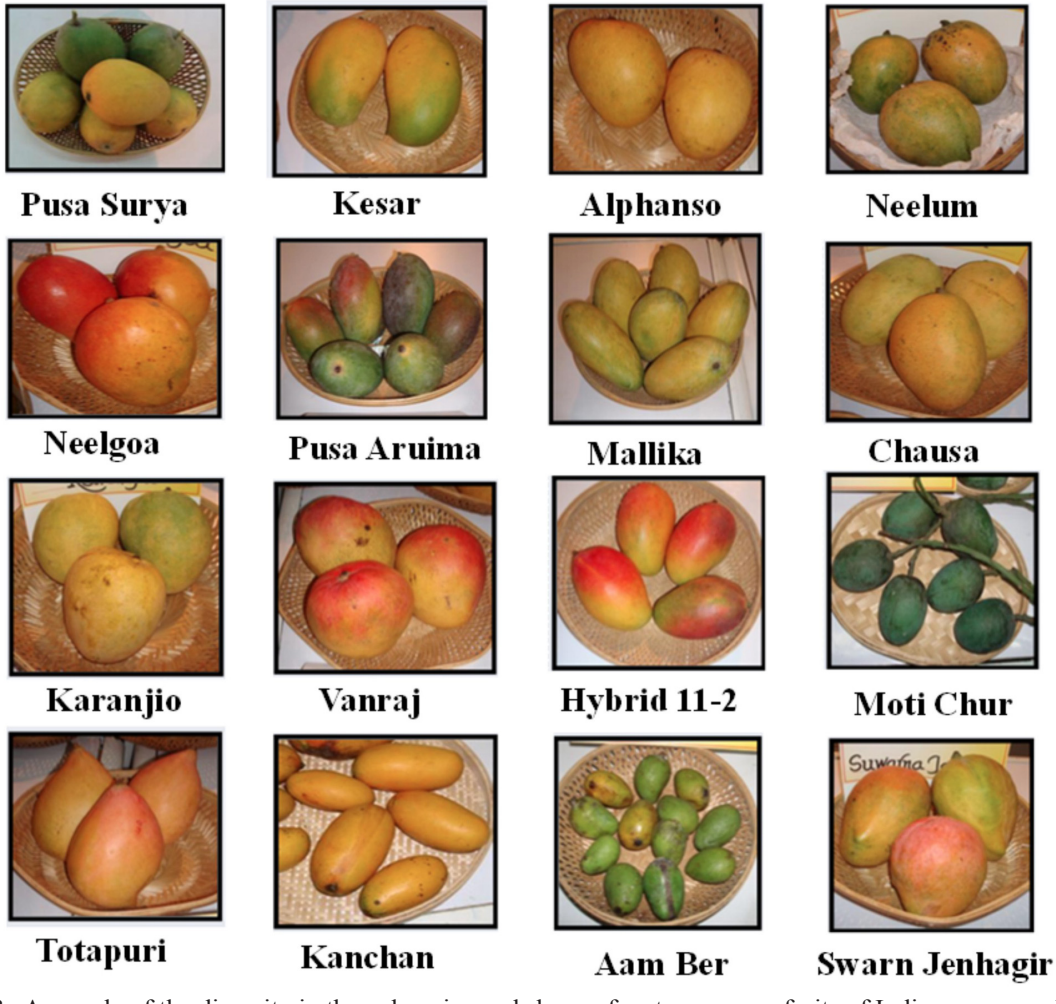

Fig. 3. A sample of the diversity in the color, size and shape of mature mango fruits of Indian mango cultivars

Honsho et al. (2005); Duval et al. (2005); Viruel et al. (2005); Schnell et al. (2006); Singh and Bhat (2008); Ravishankar et al. (2011); Dillon et al. (2013); Malathi et al. (2013). However, most of these studies were carried it on small sets of genotypes. Recently, Ravishankar et al. (2015) using 14 carefully selected SSR markers on a comprehensive set of 367 Indian mango cultivars identified two main sub-populations representing mango cultivars from the North-East and South-West regions of India.

We identified 1.67 million high quality SNPs by sequencing double digested restriction site associated DNA (ddRAD) from 84 diverse mango cultivars from different regions of India and abroad and a database of SNPs was created. The preprocessed RAD-Seq data of 84 varieties of mango were subjected to de novo SNP mining using the STACKS software version 1.29 (Catchen et al., 2013), using denovo_map.pl script. parameters –m taking minimum of three identical raw reads required to create a stack, –M with minimum of two mismatches allowed between loci when processing single individual, –n with minimum of three mismatches allowed between loci when building catalog and –T with fifteen threads to execute. The population program parameters were also passed along with the denovo_map.pl to calculate a number of population genetics statistics across the populations and exporting the resulting SNPs data in vcf, phylip and structure standard output format. These population genetic statistics were used to analyze the genetic diversity and population structure in the mango varieties. Due to random distribution and low genome coverage of the ddRAD sequence reads, the sequence reads for the number of SNP loci common to all the samples was low. There were only 1,159 SNPs common to 74 or more samples and 741 SNPs that were common to all the 84 mango varieties. The population structure in the 84 varieties was then determined based on these 1159 common SNPs using the Bayesian, model based program, STRUCTURE version 2.3.4 (Pritchard et al., 2000); Hubisz et al., 2009). For determining population structure, populations K = 1 to K = 10 were tested. For each K, three replications were run. Each run was implemented with a burn-in period of 100,000 steps followed by 100,000 Monte Carlo Markov Chain replicates derived for each K, setting the admixture model as the ancestry model and allele frequency correlated as the allele frequency model. The results of structure tool were subsequently collated using the Structure harvester tool Web version 0.6.94 (Earl et al., 2012) to derive the ΔK values based on the rate of change in the log probability of data between successive K values and mean of estimated log probabilities of data for each value of K to infer the final population. We used web based STRUCTURE HARVESTOR software (Evanno et al., 2005) to extract the relevant information (http://taylor0.biology.ucla.edu/structureHarvester/) and summarized it using CLUMPP v.1.2.2. (Jakobsson and Rosenberg, 2007) and visualized it with DISTRUCT v.1.1 (Rosenberg, 2004). The analysis revealed an optimum K value of two sub-populations, with a large number of admixed types (Fig. 4A). There were 49 cultivars in sub-population I and 35 cultivars in sub-population II, but unlike the results of Ravishankar et al. (2015) the grouping was not strictly according to South-West and North-East origin of the varieties. Here sub-population I has varieties from North, East and West of the country and subpopulation II has varieties originating from the South but also the cross derivatives varieties of Pusa including Amrapali, Mallika, Pusa Shreshtha, Pusa Lalima, Pusa Peetambar etc. which are derived from cross between Southern and Northern mango genotypes.

To further analyze the genetic diversity and relationships among the 84 diverse mango cultivars their ddRAD sequences were aligned using BioEdit software (Hall et al. 1999).

Phylogenetic tree was constructed using an improved version of the neighbour-joining algorithm, and visualized using FigTree v1.4.043 (Rambaut et al., 2009). The phylogenetic tree grouped the 84 diverse mango cultivars into seven distinct clustered represented by different colors in Fig. 4B. A comparison of the population structure bar plot and the phylogenetic trees showed that the structure sub-population I comprising of 49 cultivars corresponded to phylogenetic clusters 1-6, whereas the sub-population II comprising 35 cultivars corresponded to the phylogenetic cluster 7. There were minor exceptions to this as two of the varieties from sub-population II, namely Banganpalli and Sonatol were grouped in cluster 6 instead of cluster 7, and three of the varieties from sub-population I were grouped in cluster 7 instead of clusters 6, but these exchanges were on the borderline of genetic similarity between

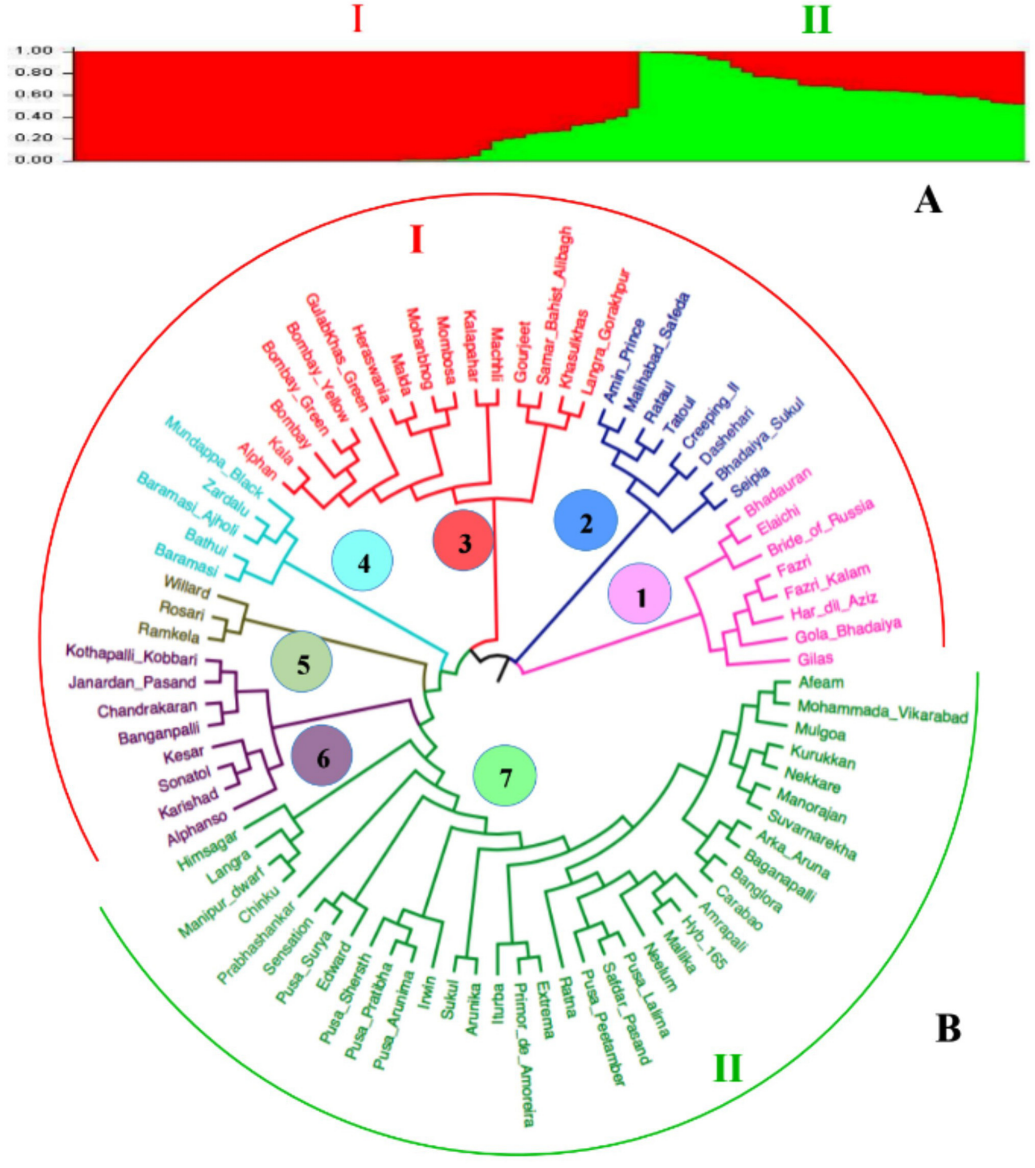

Fig. 4. Population structure (**A**) and an unrooted phylogenetic tree (**B**) of 84 diverse mango cultivars based on 1191 genome wide ddRAD SNPs and their associated DNA sequences, respectively. Correspondence between color-coded phylogenetic clusters 1-7 and STRUCTURE sub-populations I-II are also shown.

phylogenetic clusters 6 and 7. There was some degree of geographical origin based clustering in the sub-population I where clusters 1 and 2 mainly represented the varieties from North, clusters and 3 and 4 representing varieties from East whereas clusters 5 and 6 representing varieties of South, West and East. Exotic mango varieties like Irwin, Edward, Sensation and Willard were closer to the varieties of the South and West.

4. Mango Transcriptome and Genic DNA Markers

There is a need to accelerate the genetic improvement of mango varieties for nutritional, organoleptic, horticultural and commercial value traits using modern tools of molecular breeding as the rate of conventional tree breeding programmes are quite slow. It requires development of highly reliable and practical marker technologies, which can be deployed in the trait based breeding programme making use of available germplasm resources. Different types of DNA markers, e.g. microsatellites or simple sequence repeats (SSR) have immense potential in the characterization of mango germplasm resources, including analysis of genetic diversity, phylogeny, population structure and marker-assisted breeding. RNA sequencing or transcriptome data generated by next generation sequencing is a rapid way of generating genic SSR and SNP markers for practical application. Recently, transcriptome sequencing of different mango tissues are reported by scientists from China, Pakistan and Israel [Azim et al (2014); Dautt-Castro et al (2015); Luria et al (2014); Wu et al (2014)]. Here we summarize our results with the sequencing of leaf transcriptome of hybrid cultivar 'Amrapali' along with its parental lines 'Neelam' and 'Dashehari'.

We have developed large number of genic simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers from RNA sequence data. We generated 60 and 58 million short sequence reads of pooled cDNA libraries of RNA from the leaves of mango varieties 'Neelam' and 'Dashehari' using SOLiD sequencing technology and assembled these into 27,528 and 20,771 transcriptome shotgun assembly (TSA) contigs, respectively. These were further merged into a set of 34,654 non-redundant unigene contigs and used for the identification of genic SNP and SSR. We also produced 4.8 million larger sequenced reads of mango hybrid 'Amrapali' using Illumina Miseq 2x250 pair-end sequencing for a three way identification of SNPs between hybrid 'Amrapali' and its parents 'Neelam' and 'Dashehari' and found 6,831 new heterozygous SNP loci in hybrid Amrapali, which were homozygous in the parents.

Distribution of SSR motifs provides detail composition of the genome and our analysis showed SSR motifs unequally distributed in the genome. Total 3,319 genic-SSR loci were identified in 3,208 contigs, representing 9.3% of the total 34,654 TSA unigene contigs which frequency was similar to other plants like 4.7% in rice, 3.3 % in soyabean and 1.5% in maize. In this study we observed that mononucleotide SSRs were the most abundant accounting 62.82% of total geneic SSRs identified. 560 (16.87%) copies of dinucleotide SSRs the second most abundant type of SSRs motif followed by trinucleotide (16.11%) and pentanucleotide (11.19%) SSRs motif, respectively. Tetra and penta type repeats were present in very less number of geneic SSRs marker in mango. Of these SSR markers 166 were Type I SSR ($n \geq 20$bp) and 100 primer pairs were synthesized and used for wet lab PCR amplification of expected size products and were designated as "validated genic-SSR markers". Out of 100 SSR primers, 43 yielded PCR amplicons of expected size and we designated these as "validated genic-SSR markers", 36 primer pairs amplified multiple products (≥ 3bands), and 21 primer pairs failed to amplify. A large proportion of the genic-SSR was monomorphic in eight mango varieties tested.

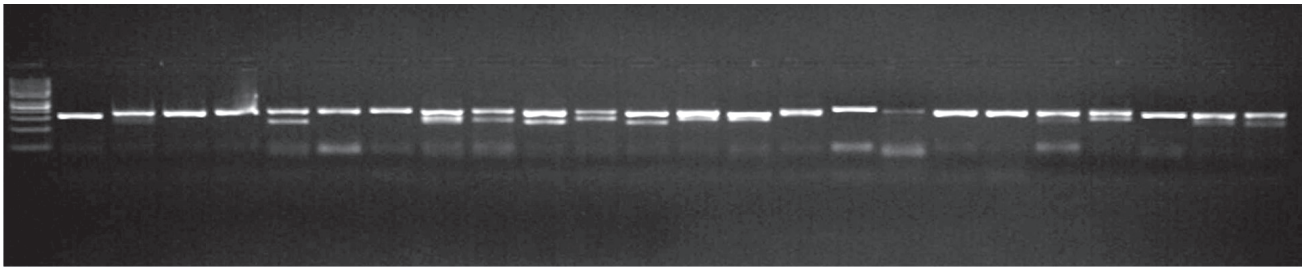

Fig. 5. Segregation pattern of genic-SSR marker MSSR-100 designed and validated from leaf transcriptome sequence data in 25 different mango varieties

Similarly, we have found in 10,571 transcripts total 22,306 SNPs; which were distributed from 1 to 16 SNPs per transcript. We have observed hetrozyosity 64.53%, 49.33 %, 30.19% in Amrapali and its parental lines Neelam and Dashehari respectively. These SNPs markers play very important role in the population diversity analysis and cultivar identification (Emanuelli et al., 2013; Esteras et al., 2013; McNally et al., 2009; Singh et al., 2015). Phylogenetic analysis based on the SNP marker of *Mangifera indica* showed the varieties grouping together in the tree with clustering according to genome origin of the respective varieties. The individual samples which do not group with any other varieties of mango as expected will be investigated further to confirm their origin. We have further identified 1.67 million high quality SNPs by double digestion restriction site associated DNA (ddRAD) sequencing of 84 diverse mango varieties from different zones of India for which a database has been created and population structure of Indian mango varieties is determined. From this data a single-copy gene based 50K SNP chip has been designed for genotyping using Affymetrix platform.

5. Mango Genome Sequence

Mango has a relatively small genome size of 439 Mb that should have made it relatively easy to sequence and assemble the genome using high throughout second generation sequencing technologies. However, high heterozygosity is a real challenge in achieving a high quality reference genome assembly of mango. We presented the first draft genome assembly of mango cultivar 'Amrapali' using Illumina MiSeq overlapping paired-end reads at San Diego PAG meeting (Singh et al., 2014), but the assembled genome size of 492 Mbp in 211,141 contigs was unexpectedly higher than the actual genome size, indicating redundancy in the contig assembly due to high heterozygosity. Unfortunately there are no homozygous inbred or doubled haploid genotypes of mango for haploid genome assembly. Therefore, to facilitate the diploid genome assembly of highly heterozygous 'Amrapali' genome we have resorted to third generation PacBio single molecule real time (SMRT) sequencing with long average read lengths of >3.5 kb. Total 55 SMRT cells of sequence data were generated with P4C2 and P5C3 chemistries with 70.1-fold genome coverage. *De novo* assembly using FALCON experimental PacBio diploid genome assembler resulted in an assembly of 323 Mbp, covering 73.2% of the of mango genome in 9,550 large contigs with the largest contig size of 1.09 Mb and a high N50 value of 98.3 Kb (Singh et al., 2016). *In silico* prediction using FGENESH programme of MOLQUEST software (www.softberry. com) identified 43,247 gene models with average gene size of 894 bps and a range of gene size from 150 to 12,102 bp. Annotation using BLASTX programme found that 33,365 (77.14%) of the predicted genes match with one other entries in the database, while 9,882 (22.86%) genes did not show any match in the

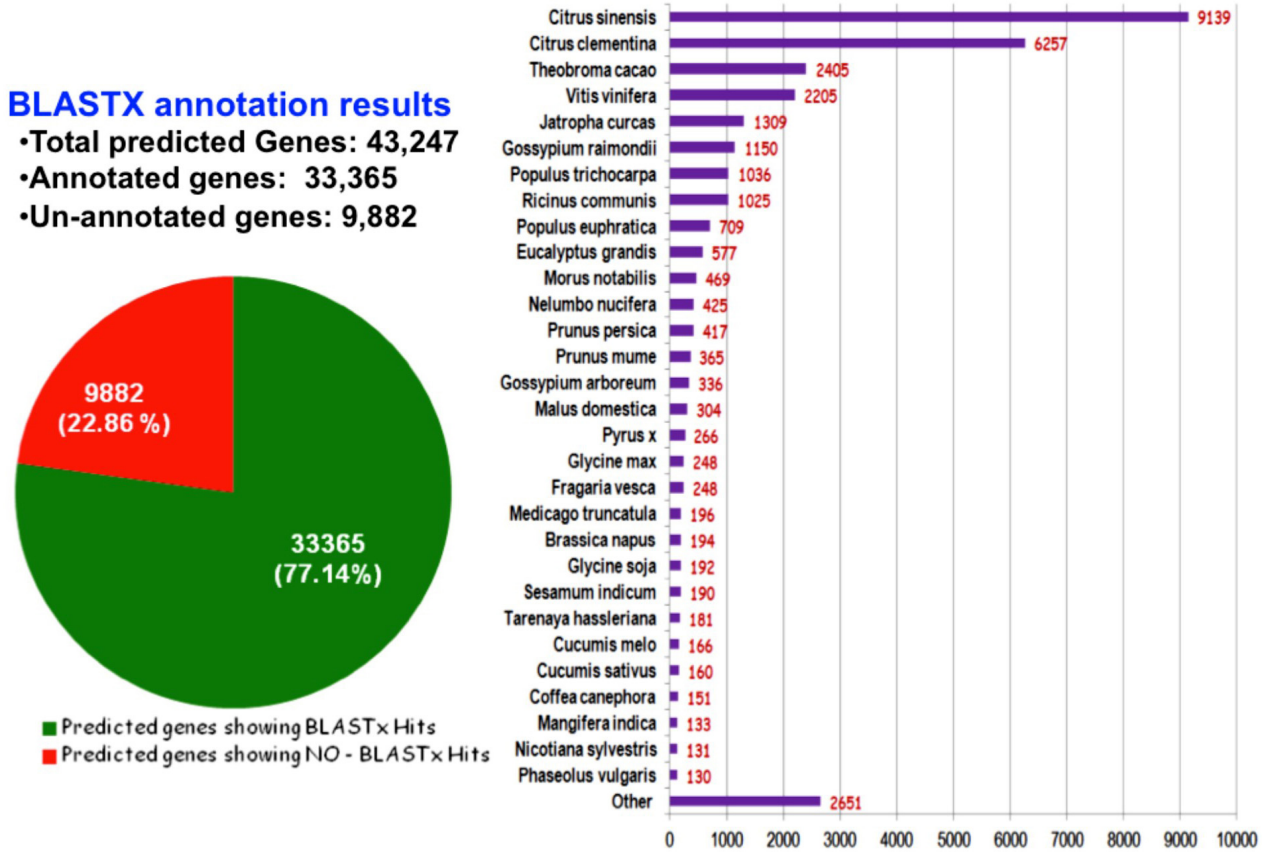

Fig. 6. Sequence similarities with other species of 33,365 annotated genes from the total 43,247 genes predicted in the genome of mango cultivar 'Amrapali'.

NCBI-NR database using optimized search criteria (Singh et al., 2004), hence these genes are unique to *M.indica*. Interestingly, maximum similarities were found with *C. sinensis* and *C. clementina* (Fig. 6). Annotated gene sequences were further classified into various functional categories e.g. physiological, DNA synthesis, disease resistance, defense response, protein synthesis, stress response, TE-related and hypothetical proteins. Those annotated genes not having any pre-defined function in the NCBI-NR database were grouped under unknown category.

We also identified repetitive element (RE) in the mango genome using *de novo* as well as homology based approaches with Repeat Modeler and Repeat Masker software (Benson, 1999; Bao and Eddy, 2002; Lander et al., 2001; Waterston et al., 2002; Wootton and Federhen, 1993). Repeat Masker programme masked and categorized the repetitive element of mango into different categories like SINEs, LINEs, LTR elements, DNA elements, simple and small RNA repeats and maximum number of RE belonged to the unknown or unclassified category which are specific to mango. We have also identified 122,332 genomic SSR loci in the mango genome of which, excluding mononucleotide repeats and complex SSR, 8,451 were type 1 SSR and 835 were hypervariable HSSR markers with high level of detectable polymorphism.

6. Prospects of Mango Genomics

The ancient heritage, huge economic significance and growing international popularity of mango make it imperative to assemble a high quality reference genome of the mango. The

availability of genome will not only help characterize the existing genetic diversity of cultivated mango and its wild relative species but also breeding efforts to further improve mango productivity, resistance to various pests and diseases as well as its nutritional, organoleptic, keeping (shelf life) and processing qualities. The challenges in mango production are many including, availability of planting material of varieties suitable for high-density small tree plantations, irregular bearing, short shelf life, disease like mango malformation, powdery mildew, anthracnose on leaves and fruits, bacterial leaf blight, blossom blight, sooty mould, and insect pests like leaf hoppers, mealy bugs, leaf webber, thrips, shoot borer, scale insect, red ants and termites. In-built genetic resistance in mango against these major diseases and pests is the most economical way to address the problem without the economical, health and environmental costs of using chemical pesticides. Genome wide association mapping, candidate gene based association mapping, marker-assisted breeding and genomic selections help find and deploy useful mango genes, which will work as effectively as the chemical pesticides without the negative effect on the health and environment.

Acknowledgement

This study was carried out under the Network Project on Transgenics in Crops (NPTC) with funding support from Indian Council of Agricultural Research, Ministry of Agriculture and Farmers Welfare, Government of India.

Bibliography

Arumuganathan, K. and Earle, E.D. Nuclear DNA content of some important plant species, Plant Molecular Biology, 9 (1991) 208–218.

Azim, M.K. Khan; I.A. Zhang, Y. Characterization of mango (*Mangifera indica* L.) transcriptome and chloroplast genome, Plant Molecular Biology 85 (2014) 193-208.

Bao, Z. and Eddy, S.R. Automated de novo identification of repeats sequence families in sequenced genomes, *Genome Research* 12 (2002): 1269-1276.

Benson, G. Tandem repeats finder: a program to analyze DNA sequences, *Nucleic Acids Research*, 27 (1999): 573-580.

Bompard, J.M. Mangifera caesia Jack and Mangifera kemanga Blume, In Coronel, R.E. & Verheij, E.W.M. (Eds.): *Plant Resources of South-East Asia*. No. 2: Edible fruits and nuts, Prosea Foundation, Bogor, Indonesia, 1992, pp. 207-209.

Bompard, J.M. and Schnell, R.J. Taxonomy and Systematics. In '*The Mango, Botany, Production and Uses*, Ed. R. Litz, CAB International:Wallingford, 1997.

Catchen, J.; Hohenlohe, P.A.; Bassham, S.; Amores, A.; Cresko, W.A.; Stacks: an analysis tool set for population genomics, *Molecular Ecology*, 22 (2013): 3124-3140.

Dautt-Castro, M.; Ochoa-Leyva, A.; Contreras-Vergara, C.A.; Pacheco-Sanchez, M.A.; Casas-Flores, S.; Sanchez-Flores, A.; Kuhn, D.N.; Islas-Osuna, M.A. Mango (Mangifera indica L.) cv. Kent fruit mesocarp de novo transcriptome assembly identifies gene families important for ripening, *Frontiers in Plant Science*, 6 (2015): 62.

"Definition for mango - Oxford Dictionaries Online (World English)". Oxforddictionaries.com. Retrieved 2015-11-12.

Dillon, N.L.; Baly, I.S.E.; Wright, C.L;. Hucks, L.; Ines, D.J.; Dietzgen, R.G. Genetic diversity of the Australian national mango genebank, *Scientia Horticulturae*, 150 (2013): 213-226.

Duval, M. F.; Bunel, J.; Sitbon, C. and Risterucci, A. M. Development of microsatellite markers for mango (Mangifera indica L.), *Molecular Ecology Notes* 5 (2005): 823-826.

Earl, D.A.; von Holdt, B.M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method, *Conservation Genetics Resources*, 4 (2012): 359-361.

Eiadthong, W.; Yonemori, K.; Sugiura, A.; Utsunomiya, N.; Subhadrabandhu, S.; Identification of mango cultivars of Thailand and evaluation of their genetic variation using the amplified fragments by simple sequence repeat- (SSR-) anchored primers, *Scientia Horticulturae*, 82 (1999): 57-66.

Eiadthong, W.; Yonemori, K.; Kansaki, S.; Sugiura, A.; Utsunomiya, N.; Subhadrabandhu, S. Amplified fragments length polymorphism analysis for studying genetic relationships among Mangifera species in Thailand, *Journal of the American Society for Horticultural Science*, 125 (2000): 160-164.

Emanuelli, F.; Lorenz, S.; Grzeskowiak, L.; Catalano, V.; Stefanini, M.; Troggio, M.; Myles, S.; Martinez-Zapater, J.M.; Zyprian, E.; Moreira, F.M.; Grando, M.S. Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape, *BMC Plant Biology*, 13(2013): 1-17.

Esteras, C.; Formisano, G.; Roig, C.; Díaz, A.; Blanca, J.; Garcia-Mas, J.; Gómez-Guillamón, M.L.; López-Sesé, A.I.; Lázaro, A.; Monforte, A.J.; Picó, B. SNP genotyping in melons: genetic variation, population structure, and linkage disequilibrium, *Theoretical and Applied Genetics*, 126 (2013): 1285-1303.

Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study, *Molecular Ecology*, 14 (2005): 2611-2620.

*Gao, A.; Chen, Y.; Crane, J.H.; Zhu, M.; Huang, J.; Luo, H. Status and Analysis on Mango Production in China, Advances in Biomedical Engineering (International Conference on Agricultural and Biosystems Engineering), 1-2 (2011): 472-476.

Gepts, P. PLB143: Crop of the Day: Mango, *Mangifera indica*, The evolution of crop plants. Department of Plant Sciences, Section of Crop & Ecosystem Sciences, University of California, Davis, 2009.

Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT, *Nucleic Acid Symp*, Ser. 41 (1999) 95-98.

Honsho, C.; Nishiyama, K.; Eiadthong W. and Yonemori K. Isolation and characterization of new microsatellite markers in mango (*Mangifera indica*), *Molecular Ecology Notes*, 5 (2005): 152-154.

Hooker, J.D. The Flora of British India, State For India in Council,: L. Reeve, London 1876.

Hubisz, M.J.; Daniel, F.; Stephens, M. and Pritchard J.K. Inferring weak population structure with the assistance of sample group information, *Mol Ecol Resour*, 9 (2009): 1322-1332.

Jakobsson, M. Rosenberg, N.A. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23 (2007): 1801-1806.

Karihaloo, J.L.; Dwivedi, Y.K.; Archak, S. and Baldev Gaikwad, A. Analysis of Indian mango cultivars using RAPD markers, *The Journal of Horticultural Science and Biotechnology*, 78 (2003): 285-289.

Kashkush, K.; Jinggui, F.; Tomer, E.; Hillel, J.; Lavi, U. Cultivar identification and genetic map of mango (*Mangifera indica*), *Euphytica*, 122 (2001): 129-136.

Kostermans, A.J.G.H. and Bompard, J.M. *The mangoes: Their Botany, Nomenclature, Horticulture, and Utilization*, Academic Press, London, UK, 1993.

Kole, C. *Wild crop relatives: Genomics and Breeding Resources Tropical and Subtropical Fruits*, Springer-Verlag Berlin Heidelberg, 2011.

Kumar, H.; Narayanswamy, P.; Prasad, T.; Mukunda, G.K.; Sondur, S. Estimation of genetic diversity of commercial mango cultivars using RAPD markers, *Journal of Horticultural Science and Biotechnology* 76 (2001): 529-533.

Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; Funke, R.; Gage, D.; Harris, K.; Heaford, A.; Howland, J. et al. Initial sequencing and analysis of the human genome, *Nature*, 409 (2001): 860-921.

Litz, R.E. *The Mango: Botany, Production and Uses*, CABI, U.S.A., 2009, pp.10.

López-Valenzuela, J.A.; Martínez, O.; Paredes-López, O. Geographic Differentiation and Embryo Type Identification in *Mangifera indica* L. Cultivars Using RAPD Markers, *Hort Science*, 32 (1997): 1105-1108.

Luria, N.; Sela, N.; Yaari, M.; Feygenberg, O.; Kobiler, I.; Lers, A.; Prusky, D. De-novo assembly of mango fruit peel transcriptome reveals mechanisms of mango response to hot water treatment, *BMC Genomics*, 15 (2014) 957.

Malathi, S.; Lakshminarayan, R.V.; Begum, H.; Purushotham, R.B.; Neetasri, C.; Nagaraju, J.; Anwar, S.Y.; Siddiq, E.A. Population structure and genetic analysis of different utility types of mango (*Mangifera indica* L.) germplasm of Andhra Pradesh state of India using microsatellite markers. *Plant Systematics and Evolution* 299 (2013): 1215-1229.

McNally, K.L.; Childs, K.L.; Bohnert, R.; Davidson, R.M.; Zhao, K.; Ulat, V.J.; Zeller, G.; Clark, R.M.; Hoen,

D.R.; Bureau, T.E.; Stokowski, R.; Ballinger, D.G.; Frazer, K.A.; Cox, D.R.; Padhukasahasram, B.; Bustamante, C.D.; Weigel, D.; Mackill, D.J.; Bruskiewich, R.M.; Rätsch, G.; Buell, C.R.; Leung, H.; Leach, J.E. Genome-wide SNP variation reveals relationships among landraces and modern varieties of rice, *Proceedings of the National Academy of Sciences*, 106 (2009): 12273-12278.

Mehrotra, R.C.; Dilcher, D.L.; Awasthi, N. A Paleocene Mangifera - Like leaf fossil from India, *Phytomorphology*, 48 (1998): 91-100.

Morton, J. Mango, In: *Fruits of warm climates*, Julia F. Morton, Miami, 1987.

Mukherjee, S.K. Origin of mango, *Indian Journal of Genetics and Plant Breeding*, 11 (1951): 49-56.

Mukherjee, S.K. The mango—its botany, cultivation, uses and future improvement, especially as observed in India, *Economic Botany*, 7 (1953) 130-162.

Mukherjee, S.K. Origin of Mango (*Mangifera indica*), Economic Botany, 26 (1972) 260-264.

Orwa, C.; Mutua, A.; Kindt, R.; Jamnadass, R.; Anthony, S.; Agroforestree Database : a tree reference and selection guide version 4.0 (http://www.worldagroforestry.org/sites/treedbs/treedatabases.asp), 2009.

Pandit, S.S.; Mitra, S.; Giri, A.P.; Pujari, K.H.; Patil, B.P.; Jambhale, N.D. and Gupta, V.S. Genetic diversity analysis of mango cultivars using inter simple sequence repeat markers, *Current Science*, 93 (2007) 1135-1141.

Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data, *Genetics*, 155 (2000) 945-959.

Purseglove, J.W. Some aspects of mango culture in the Western Tropics. *Proc. Int. Symp. On Mango and Mango Culture*, I.A.R.I., New Delhi, 1969.

†Purseglove, W. *Tropical crops: Dicotyledons*, Wiley, New York, 1968, pp. 24-32.

Rambaut, A. Molecular evolution, phylogenetics and epidemiology, http://tree.bio.ed.ac.uk/software/figtree/, 2009.

Ravishankar, K. V.; Anand, L. and Dinesh, M. R. Assessment of genetic relatedness among a few Indian mango cultivars using RAPD markers. *The Journal of Horticultural Science & Biotechnology* 75 (2000): 198-201.

Ravishankar, K.V.; Bommisetty, P.; Bajpai, A.; Srivastava, N.; Mani, B.H.; Vasugi, C.; Rajan, S.; Dinesh, M.R. Genetic diversity and population structure analysis of mango (*Mangifera indica*) cultivars assessed by microsatellite markers, Trees, 29 (2015): 775-783.

Ravishankar, K.V.; Mani, B.H.; Anand, L.; Dinesh, M.R. Development of new microsatellite markers from Mango (*Mangifera indica*) and cross-species amplification, *American Journal of Botany*, 98 (2011): e96-99.

Renugadevi, R. Environmental ethics in the Hindu Vedas and Puranas in India. *African Journal of History and Culture* 4 (2012): 1-3.

Rieger, M. Distance Education versus Classroom Instruction in Horticulture- An Introduction to Fruit Crops – Case Study, Course of Horticulture, University of Georgia, 2001, pp.513-515.

Rosenberg, N.A. Distruct: a program for the graphical display of population structure, *Molecular Ecology Notes*, 4 (2004): 137-138.

Salma, I.; Khadijah, A.; Masrom, H.; Azuan, A.; Raziah, M.L. and Rahman Abdul, M. Distribution and diversity of Mangifera species on farm in Malaysia, *Journal of Tropical Agriculture and Food Science*, 38 (2010): 89-95.

Schnell, R. J.; Brown, J. S.; Olano, C. T.; Meerow, A. W.; Campbell R. J.; and Kuhn D. N. Mango genetic diversity analysis and pedigree inference for Florida cultivars using microsatellite markers, *Journal of the American Society for Horticultural Science*, 131 (2006) 214-224.

Schnell, R.J.; Ronning, C.M. and Knight, R.J. Identification of cultivars and validation of genetic relationships in *Mangifera indica* l. using RAPD markers, *Theoretical and Applied Genetics*, 90 (1995): 269-274.

Sreekumar, P. V.; Veenakumari, K.; Padhye, P. M. *Mangifera griffithii* (Anacardiaceae) - an addition to the Indian mangoes, from Andaman Islands, India.). *Malayan Nature Journal*, 50 (1996): 85-87.

Singh, A. K.; Devanshi, Sharma, P.; Singh, R.; Singh, B.; Koundal, K.R.; Singh, N.K. Assessment of Genetic Diversity in *Ziziphus mauritiana* Using Inter-Simple Sequence Repeat Markers, *Journal of Plant Biochemistry & Biotechnology*, 16 (2007): 35-40.

Singh, N.; Jayaswal, P.K.; Panda, K.; Mandal, P.; Kumar, V.; Singh, B.; Mishra, S.; Singh, Y.; Singh, R.; Rai, V.; Gupta, A.; Sharma, T.R;. Singh, N.K. Single-copy gene

based 50 K SNP chip for genetic studies and molecular breeding in rice, *Scientific Reports*, 5 (2015): 11600.

Singh, N.K.; Raghuvanshi, S.; Srivastava, S.K.; Gaur, A.; Pal, A.K.; Dalal, et al. Sequence analysis of the long arm of rice chromosome 11 for rice–wheat synteny. Functional and Integrative Genomics, 4 (2004): 102-117.

Singh, S.; Bhat, K.V. Molecular Characterization and analysis of geographical differentiation of Indian Mango (*Mangifera indica* L) germplasm, *Acta Horticulturae* (ISHS), 839 (2008): 599-606.

Singh et al. *A Draft Genome of the King of Fruit, Mango* (*Mangifera indica* L.). PAGXIII 2015, PAG 2014.

Singh et al. Assembly of highly heterozygous mango (*Mangifera indica* cv. Amrapali) genome using PacBio long sequence reads (2016), PA 2016.

Vavilov, N.I. Centres of Origin of Cultivated Plants, *Bulletin of Applied Botany of Genetics and Plant-Breeding,* 16 (1926): 1-248.

Viruel, M. A.; Escribano, P.; Barbieri, M.; Ferri M.; and Hormaza J. I. Fingerprinting, embryo type and geographic differentiation in mango (*Mangifera indica* L., Anacardiaceae) with microsatellites, *Molecular Breeding*, 15 (2005): 383-393.

Waterston, R.H. et al. Initial sequencing and comparative analysis of the mouse genome, *Nature*, 420 (2002): 20-62.

Williams, C. J. *Medicinal Plants in Australia* Volume 3: Plants, Potions and Poisons, Rosenberg, Australia, 2012, pp. 224.

Woodrow, G.M. *The Mango; its culture and varieties*, Paisley: Alexander Gardner, Culcutta, 1904.

Wootton J.C.; Federhen, S. Statistics of local complexity in amino acid sequences and sequence databases, *Computers and Chemistry*, 17 (1993): 149-163.

Wu, C.Y. and Ming T.L. Flora *Yunnan*, 2 (1979) p.368.

Wu, H.X.; Jia, H.M.; Ma, X.W.; Wang, S.B.; Yao, Q.S.; Xu, W.T.; Zhou, Y.G.; Gao, Z.S.; Zhan, R.L. Transcriptome and proteomic analysis of mango (*Mangifera indica* Linn) fruits, *Journal of Proteomics*, 105 (2014) 19-30.