Review Article

Modern molecular approaches for walnut:a review

Abstract

Walnut is the nut tree belongs to jugland aceace family. The walnut is nutrient-dense with protein and essential fatty acids. Walnut is cultivated throughout the temperate regions. The availability of sequenced walnut genome and has opened a new era of research. The genome denote the complete set of chromosomal and extra chromosomal genes and extra chromosomal genes present in an organism, including virus whereas, genomics includes sequencing of genomes, determination of complete set of proteins encoded by an organism, and the functioning of genes and metabolic pathways in an organism. Molecular markers play very important role in genomics. They are very important for studying genome of any organism. DNA-based molecular markers have facilitated the development of marker-based gene tags, map-based cloning of agronomically important genes, diversity analysis, phylogenetic analysis, linkage mapping and marker-assisted selection of desirable genotypes. This review makes an attempt to conclude the genomics work done in walnut.

Keywords: walnut, genomics, molecular markers, biodiversity, gene editing

Shailja, Rajinder Kaur

Department of Biotechnology, Dr. Yashwant Singh Parmar University of Horticulture and Forestry, India

Correspondence: Shailja, Department of Biotechnology, Dr. Yashwant Singh Parmar University of Horticulture and Forestry, India, Tel +918988086845, Email shailjabiotech@gmail.com

Introduction

Walnut (*Juglans regia L.*) is an important nut belonging to family *Juglandaceae*. In India it is commonly known as 'Akhrot', almost all parts of which are used in one way or the other. *Juglans regia* - Persian walnut or English walnut is an indigenous species in Eurasia which is cultivated throughout the temperate regions of world for its high quality wood and edible nuts. Persian walnut is monoecious and heterodichogamous, with 2n chromosome number= 32. *Juglans regia* is economically important species in Europe, Asia and North America. The worldwide production of walnuts has been increasing rapidly in recent years, with the largest increase coming from Asia. Persian walnuts are the most common, their nutrient density and profile are significantly different from those of black walnuts. Unlike most nuts that are high in monounsaturated fatty acids, walnut oil is composed largely of polyunsaturated fatty acids, particularly alpha-linolenic acid and linoleic acid. They also contain triglycerides effective in reducing the risk of cardiovascular diseases. Compared to certain other nuts, such as almonds, peanuts and hazelnuts; walnuts contain the highest spectrum of antioxidants, including free antioxidants and the antioxidants bound to fiber. There are no standard walnut cultivars under cultivation in India. Most of the walnut produce comes from trees of seedling origin grown in semi-wild state. Like elsewhere in India, walnut plantations in Himachal Pradesh, too, comprise of genetically diverse seedling trees of unknown origin, there by constituting a vast gene pool. Accurate estimation, of distances between different genotypes of the germplasm, can assist the breeders in crop improvement programmes. For genetic characterization of a plant species, the molecular markers are considered superior to morphological and biochemical traits.

Genetic diversity analysis

Various molecular markers are used to study the genetic diversity in walnut. Juglans populations,[1–4] to determine the impact of different timberharvest scenarios on residual levels of genetic diversity.[3] Pijut et al.[5] reviewed technological applications of molecular markers used on several temperate hardwood tree species. Modern genomics-based tools are currently being used in several walnut species. Erturk et al. (2011) studied that RAPD analysis was used to study the characterization and grouping of walnut genotypes.A total of forty five primers were used out of which 37 primers were polymorphic. Vischi M, et al.[6] studied genetic diversity of walnut (*Juglansregia L.*) in the eastern italian alps. They studied two hundred and fifteen wild accessions native to the area were sampled, georeferenced, and genotyped with 20 microsatellite loci that were selected from the literature. Mahmoodi R et al.[7] assayed genetic diversity among 16 accessions and 5 cultivars of *Juglans regia*. Nine SSR markers and morphological traits were used to study the genetic diversity. They concluded that high diversity among the genotypes was used for breeding programmes. Najafi et al.[8] isolated and characterized 13 microsatellite markers of *Juglans regia L*. All 13 primer pairs could amplify 36 genotypes of *Juglans regia*. The number of polymorphic alleles ranged from 2 to 4 (with an average of 4.35). The polymorphic information content values ranged from 0.47 to 0.88 (with an average of 0.69). TC/AG and GAA/CTT class of repeats were the most abundant di-nucleotide and tri-nucleotide repeats, respectively. Out of these TC/AG di-nucleotide was more abundant.

Salieh et al.[9] assessed genetic relationship among 12 walnut genotypes using ten RAPD and nine SSR primers. RAPD primers produced 85 bands, out of which 36 (42.35%) were polymorphic with an average of 3.6 bands per primer. In SSR analysis, nine primers produced 26 bands, out of which 23 (88.16%) were polymorphic, with an average of 2.3 bands per primer. Genetic similarities were calculated and ranged from 0.4 to 0.93 for RAPD data and 0.27 to 1.00 for SSR data. Cluster analysis by RAPD and SSR markers revealed clear distinct diversity between genotypes. The dendrogram constructed from the RAPD and SSR markers grouped the 12 genotypes into four major clusters. This study demonstrated that RAPD and SSR analysis can be used for the characterization and grouping of walnut genotypes. Zhang et al.[10] used publicly available walnut (*Juglans regia*) EST database to develop SSR markers and used them to study the genetic analysis of the widespread *Juglans nigra*, Carya cathayensis and an endangered species Annamocarya sinensis. A total of 7262 unigenes, including 1911 contigs and 5351 singletons, were obtained from 13,559 ESTs retrieved from the NCBI database. They

further reduced 7262 unigenes to 706 EST-SSR sequences containing 805 SSR loci. Then, they randomly designed 309 EST-SSR primers. The 77 primers were transferable among species namely, J. regia, J. nigra, C. cathayensis, Carya dabieshanensis, and A. sinensis. 13 highly polymorphic EST-SSRs were further used for genetic analysis in these five species. The genetic relatedness of 82 walnut genotypes by using 13 SSR and 20 RAPD primers was reported by Ahmed et al.[11] A high level of genetic diversity was observed within populations with the number of alleles per locus ranging from one to five and two to six in SSR and RAPD primers, respectively. The proportion of polymorphic loci was 100 % and similarity ranged from 12 % to 79 % with an average of 49 %. Dendrogram showed that all the accessions formed four main clusters with various degree of sub-clustering within the clusters.

Cosmulescu & Botu[12] evaluated *Juglans regia L.* germplasm from the Oltenia region, located in the South Western part of Romania and determined the variability in walnut germplasm. Variability found in nut weight was between 6.8-18.4 g, in kernel weight between 1.7-8.79 g, in weight kernel/weight nut ratio between 23.6-71.7%. Nut size, bud breaking time, nut maturity time and phenological characteristics were also evaluated. The data obtained have indicated that walnut trees studied in this region have high variations in fruit characteristics that indicate the higher potential in selecting new genotypes of material under study. A total of 5,025 walnut ESTs (covering 16.41 Mb) were retrieved from the National Center for Biotechnology Information database.[10] They analyzed the SSR motifs by SSR Hunter software. They obtained 398 SSRs, dinucleotide repeat motifs accounted for 69.85%, followed by trinucleotide with a frequency of 27.64%, while low frequency (2.51%) of tetranucleotide to hexanucleotide was observed. Subsequently, a total of 123 primer pairs were designed from the non-redundant SSR-containing unigenes with the selection threshold of SSR length set to 10 bp or more. The efficiency of markers was examined by seven DNA pools of walnut that were collected from geographically different accessions. This study demonstrated that 41 SSR primer sets produced high polymorphic amplification products. Pollegioni et al.[13] characterized Juglans sp. using SSR markers. Natural hybrids between the two species, known as Juglans X intermedia (Carr), are valued for timber production. They tested ten nuclear microsatellite markers to identify new J. X intermedia hybrids and then characterized their parental species-*Juglans regia* and *Juglans nigra* to detect *Juglans nigra* genotypes with a spontaneous crossing ability with *Juglans regia* in a mixed Italian population. They also studied the transferability of ten black walnut SSR loci to Persian walnut. Ten microsatellites amplified in both species, producing fragments of variable sizes. Indices of genetic diversity revealed high level of variability. Total sample set of 112 alleles divided into three main groups: *Juglans nigra*, *Juglans regia* and J. X intermedia hybrids. They performed the microsatellite fingerprinting, and identified a triploid hybrid plant with two genome parts of *Juglans nigra* and one part of *Juglans regia*. The sequence of amplified fragments confirmed the cross-species amplification of SSR. Similarity coefficient based on 112 alleles divided the studied genotype into distinct four groups.

The genetic diversity among 16 'Sorrento' plants grown in Caserta (ten originated from seeds and six fromgrafts), and 26 grafted 'Sorrento' clones grown in the Sorrento peninsula was evaluated by Foroni et al.[14] They compared their genotypes along with six other walnut cultivars using 12 SSR markers. A total of 66 putative alleles were detected, 16 of which were unique to one individual.

Two loci namely WGA9 and WGA71 were useful for distinguishing Caserta samples from Sorrento peninsula clones. The phylogenetic and structure analysis highlighted the genetic distance between the Sorrento peninsula and Caserta groups. The samples were grouped into two different clusters (or populations) corresponding closely, but not perfectly, to each sample's geographic origin. Ruiz et al.[15] characterized a collection of 57 common walnut cultivars using 32 microsatellite markers. The 19 selected microsatellite markers could discriminate the studied cultivars, with a total of 97 alleles and an average of five alleles per locus, confirming that these markers are more suitable tools for walnut identification than other molecular markers studied previously. The genetic similarity estimated from the molecular data clearly separated the Spanish walnuts from the Californian genotypes. UPGMA analysis grouped 57 walnut cultivars into two clusters based on similarity coefficient.

Foroni et al.[14] analyzed ten plants that originated from 'Sorrento' seeds, six grafted 'Sorrento' clones and compared them with six other walnut cultivars using SSR DNA markers. These primers amplified the alleles at six SSR loci in Persian walnut which were derived from *Juglans nigra*. A total 33 putative alleles were detected, nine of which were unique to one genotype. Two loci namely WGA5 and WGA27 were useful for distinguishing walnut varieties. Cluster analysis clarified the relatively large genetic distance between most of the 'Sorrento' plants and some genotypes labeled as 'Sorrento'. The genetic diversity found among seed and vegetatively propagated plants. Jia et al.[16] studied 77 samples grown in China, including 14 introduced cultivars, 12 domestic seedling breeding cultivars, and 49 fine pecan plants, together with Carya cathayensis and *Juglans nigra*. A total of 77 ISSR and 19 SSR primers were prescreened, out of which ten ISSR and eight SSR primers were selected for further studies. ISSR markers yielded a total of 94 amplified bands (100% polymorphic) in the range of 140–950 bp while SSR markers ampified 70 bands (100% polymorphic) in the range of 50–350 bp. They analysed genetic diversity which indicated that Chinese-grown pecan cultivars and fine plants had significant diversity at the DNA level. The dendrogram constructed with ISSR, SSR or combined data were very similar, but showed very weak grouping association with morphological characters. However, the progeny was grouped with the parents. This study demonstrated that ISSR and SSR techniques both were suitable for genetic diversity analyses and the identification of pecan cultivars.

Fingerprinting

DNA fingerprinting technique to plant genome analysis was first introduced in 1988. Ryskov et al.[17] reported first time DNA fingerprint in plants. They reported that DNA fragment pattern differences between two varieties of barley (*Hordeum vulgare*), following Southern blot hybridization of Hae III-digested DNA samples with the M13 probe. F He et al. identified various walnut cultivars with the help of AFLP (Amplified Fragment Length Polymorphism) fingerprinting. They used AFLP markers to identify 5 *Juglans regia L.* cultivars. They identified that with 5 pairs of E+3/M+3 primers 28 specific bands were produced. The five cultivars were discriminated successfully with the help of DNA fingerprints. Therefore, DNA fingerprint techniques provided an effective tool for the identification of the cultivars. Jan Dvorak et al.[18] fingerprinted the individuals using eighteen microsatellite loci to confirm the hybridity and parentage. One hundred thirty-five F1 individuals were fingerprinted from a cross between 'Chandler' x 'Idaho'. This experiment resulted in the

identification of six out-crossed individuals possessing alleles other than the ones present in the parents. The results were analyzed to verify the pattern of inheritance of molecular markers. About half of this F1 population is already in bearing stage and phenological data, bearing habit, and fruit and nut traits are being collected. The remaining half will come into bearing in the coming years.

Next generation sequencing

The scientific revolution that started with the human genome sequencing project carried out with first generation sequencing technology, has initiated other sequencing projects, including those for plant species. High throughput genome sequencing strategies are used for sequencing whole genome of the organisms. Two basic methods used to sequence large genomes are explained below:

1. Whole genome shotgun method- It is one of the efficient methods mainly used in sequencing prokaryotic genomes. The genomes which are less repetitive in nature can be easily sequenced with WGS method and assembled by using computational tools (Figure 1).
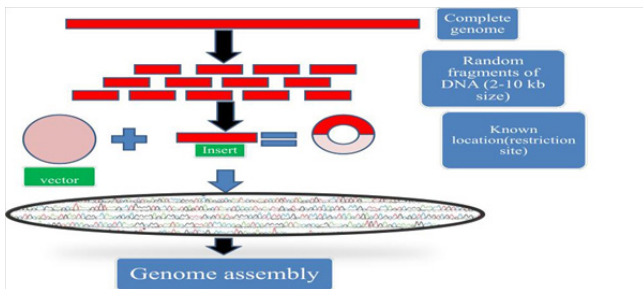

Figure 1 Diagrammatical representation of whole genome shotgun method.

2. Hierarchial sequencing method - This method is also known as clone-by-clone approach of sequencing large genomes. In this method construction of genetic linkage map of an organism is a pre-requisite. So the availability of genomic resources in terms of DNA markers and BAC Libraries are very important to construct maps of the plants (Figure 2).

Different technologies have been developed together with the second and third generation sequencing platformscalled next generation sequencing whereas, determining the number and order of nucleotides that make up a givenmolecule of DNA is called as sequencing Different techniques were used to sequence the genome of different crops. (Figure 3 representing genome sequencing) The Illumina Genome Analyzer, a groundbreaking platform for geneticanalysis and functional genomics.

Genome sequencing

Figure 3 Diagramatical representation of Genome sequencing method (Source of pic –Google) Illumina Genome Analyzer : This technology was commercialized by Illumina, U.S.A, in 2007 in the form of Solexa (Genome Analyzer), which was followed by Illumina Genome Analyzer 1 Gb, and the recent platform High Seq. 600 Gb.The Illumina system can be used for de novo genome sequencing, resequencing and transcript profiling (Figure 4). Walnut genome sequence was revealed by Martínez García Pedro J et al.[19] They used Illumina sequencing technique for sequencing the genome of walnut. Illumina sequencing of paired end reads from short- and large-

fragment libraries, resulted in 500 million reads and 120X genome coverage. Developments in sequencingtechnology make it possible to obtain large-scale sequence data in a short time, the assembly and analysis of sequences remains a challenging task.
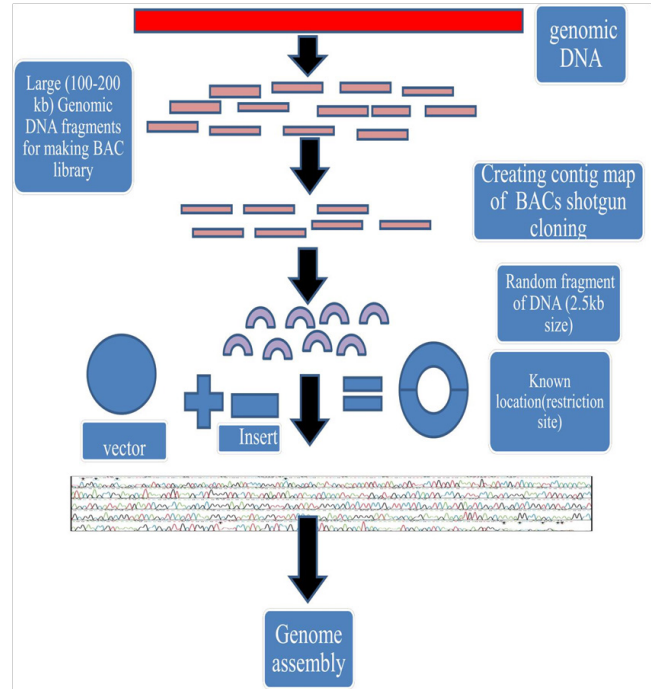

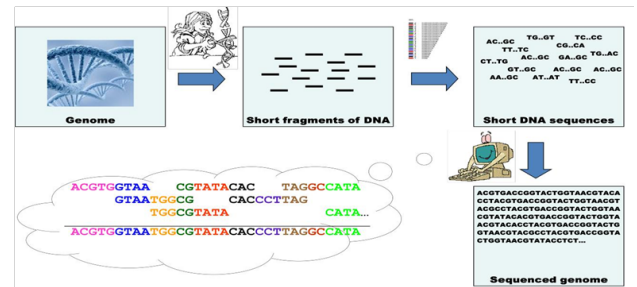**Figure 2** Diagrammatical representation of hierarchical sequencing method.

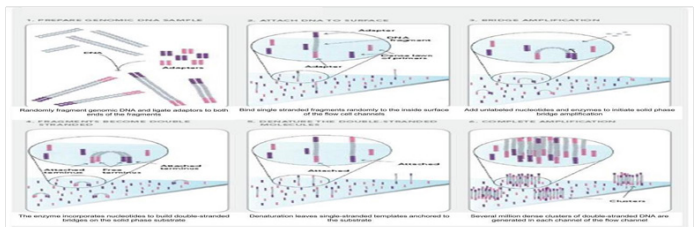**Figure 3** Diagrammatical representation of genome sequencing method (Source of pic –Google).

Figure 4 Diagrammatical representation of liumina genome analyzer Genome sequencing method (Source of 200 pic –Google).

Figure 5 Diagramatical representation of IIumina Genome Analyzer Genome sequencing method (Source of pic–Google) (cont…part of Figure 4) It is concluded that NGS diagnostics - shifted towards data analysis rather than the technical component.NGSinfrastructures

must consist of appropriate expertise and computational hardware. Unprecedented amounts of medical data and various processing algorithms necessitate adequate tools for Data management. Many new denovo and resequenced plant genomes are expected in the near future for plants in general and crop species in particular, using second and mostly third generation sequencing platforms.
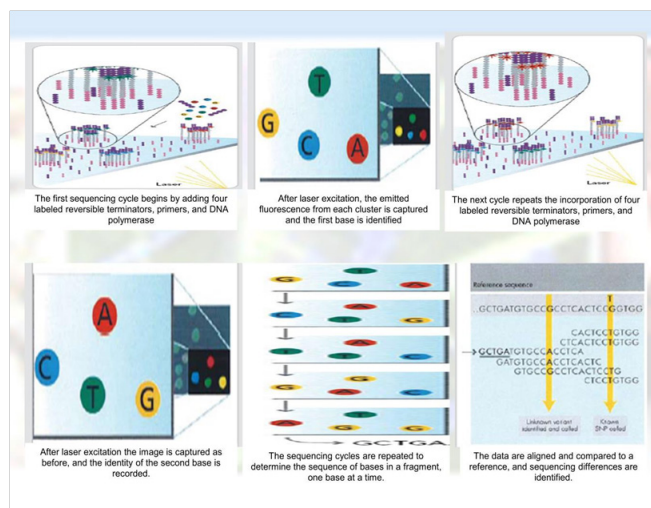

Figure 5 Diagrammatical representation of liumina genome analyzer genome sequencing method (Source of pic–Google) (cont…part of Figure 4).

Genome editing

In general Genome editing is a combination of two word i.e Genome + Editing means there is some editing in the genome. It is considered as the type of genetic engineering in which DNA is inserted, deleted or replaced in the genome of an organism using engineered nucleases. These nucleases create site-specific double-strandbreaks at desired locations in the genome. There are currently four families of engineered nucleases being used:

1. Meganucleases.
2. Zinc finger nucleases (ZFNs).
3. Transcription activator-like effector-based nucleases (TALEN).
4. CRISPR-Cassystem.

Currently no successful reports of using genome editing in tree fruits are available. But in future efforts have been undertaken.

Conclusion

Markers are powerful tools in the field of genomics. The use of these markers would reduce the cost and therefore facilitate cultivar identification, genetic distance assessments, gene mapping and possible marker-assisted selection. Fruit and nut crops have received considerable attention in genomics. Genomic selection in fruit tree crops is expected to enhance breeding efficiency. These tools complement, but do not replace, classical plant breeding. The variation in agronomic traits of walnut populations would be of great interest for future cultivar improvement. The combination of genomics and breeding results in the improvement of walnut fruit.

Conflict of interest

The authors declare that there is no conflict of interest among them.

References

1. Hoban SM, Borkowski DS, Brosi SL, et al. Range wide distribution of genetic diversity in the North American tree Juglans cinerea: a product of range shifts, not ecological marginality or recent population decline. *Mol Ecol*. 2010;19(22):4876–4891.

2. Karimi R, Ershadi A, Vahdati K, et al. Molecular characterization of Persian walnut populations in Iran with microsatellite markers. *HortScience*. 2010;45(9):1403–1406.

3. Robichaud RL, Glaubitz JC, Rhodes OE, et al. Genetic consequences of harvest in a mature second-growth stand of black walnut (Juglans nigra L.). *Annals of Forest Science*. 2010;67(7):702.

4. Victory ER, Glaubitz JC, Rhodes OE, et al. Genetic homogeneity in *Juglans nigra* (Juglandaceae) at nuclear microsatellites. *American Journal of Botany*. 2006;93(1):118–126.

5. Pijut PM, Woeste KE, Vengadesan G, et al. Technological advances in temperate hardwood tree improvement including breeding and molecular marker applications. *In Vitro Cellular & Developmental Biology - Plant*. 2007;43(4):283–303.

6. Vischi M, Chiabà C, Raranciuc S, et al. Genetic Diversity of Walnut *(Juglans Regia L.)* in the Eastern Italian Alps. *Forests*. 2017;8(3):81.

7. Mahmoodi R, Rahmani F, Rezaee R. Genetic diversity among *Juglans regia L.* genotypes assessed by morphological traits and microsatellite markers. *Spanish Journal of Agricultural Research*. 2013;11(2):431–437.

8. Najafi F, Mardi M, Fakheri B, et al. Isolation and characterization of novel microsatellite markers in walnut (*Juglans regia L.*). *AJPS*. 2014;5(3):409–415.

9. Salieh HMF, Tahir RAN, Faraj JM. Assessment of genetic relationship among some Iraqi walnut genotypes (*Juglans regia L.*) in Sulaimani region using RAPD and SSR molecular markers. *JJAS*. 2013;9(3):351–362.

10. Zhang R, Zhu AD, Wang XJ, et al. Development of *Juglans regia* SSR markers by data mining of the EST database. *Plant Molecular Biology Reporter*. 2010;28(4):646.

11. Ahmed N, Mir JI, Mir RR, et al. SSR and RAPD analysis of genetic diversity in walnut (*Juglans regia L.*) genotypes from Jammu and Kashmir, India. *Physiol Mol Biol Plants*. 2012;18(2):149–160.

12. Cosmulescu S, Botu M. Walnut biodiversity in South-Western Romania-resource for perspective cultivars. *Pak J Bot*. 2012;44(1):307–311.

13. Pollegioni P, Major A, Bartoli S, et al. Application of microsatellite and dominant molecular markers for the discrimination of species and interspecific hybrids in genus Juglans. *Acta Horticulturae*. 2008;705:191–197.

14. Foroni I, Woeste K, Monti LM, et al. Identification of 'Sorrento' walnut using simple sequenc repeats (SSRs). Genetic Resources and Crop Evolution. 2007;54(5):1081–1089.

15. Ruiz GL, Lopez OA, Fuentes D et al. Identification of a walnut (*Juglans regia L.*) germplasm collection and evaluation of their genetic variability by microsatellite markers. *SJAR*. 2011;9(1):179–192.

16. Jia XD, Wang T, Zhai M, et al. Genetic diversity and identification of chinese- grown pecan using ISSR and SSR markers. *Molecules*. 2011;16(12):10078–10079.

17. Ryskov AP, Jincharadze AG, Prosnyak MI, et al. M13 phage DNA as a universal marker for DNA fingerprinting of animals, plants and microorganisms. *FEBS Lett*. 1988;233(2):388–392.

18. Jan Dvorak, Ming Cheng Luo, Mallikarjuna Aradhaya, et al. Walnut Genome Analysis. Walnut Research Reports. 2007;39–45.

19. Martínez García Pedro J, Marc W Crepeau, Daniela Puiu, et al. The walnut (Juglans regia) genome sequence reveals diversity in genes coding for the biosynthesis of nonstructural polyphenols. *Plant J*. 2016;87(5):507–32.