# Genomic assisted breeding in fruit crops

- Author: Dr Shiv Lal

Fruit science is one of the important sectors of horticulture which consists of cultivation, production and improvement of different types of fruit plants. The importance of fruit crops is widely acknowledged in many aspects of innovation, production, quality maintenance, for uplifting economic condition of farmers, entrepreneurs and in providing nutritional security to the people. With the growing population, demand for fruits as such and its products are gradually increasing. But there are several constraints to increased sustained fruit production and its improvement. This limitation is especially evident for tree fruits and nuts, products of breeding cycles that can extend to a dozen years or more. Like other sectors of horticulture, the fruit and nut industries face highly dynamic situations arising from such factors as decreasing labor availability, increasing environmental concerns, cost of energy, climate change and epidemics of new and invasive insects and diseases. In addition, the inability to re-program tree form and phenology quickly limits deployment of highly efficient production technologies. The generally reactive, rather than proactive, nature of response to these factors translates to the release of new cultivars only after such pressures have accumulated significant impact on production.

In recent years significant technological advancement has been observed (concepts, approaches and applications) in fruit improvement which has been revolutionized the fruit production (Janick 2012). In the last two decades, the availability of many genomic resources like genome sequences, high-throughput analysis of gene expression, sufficient numbers of molecular markers, express sequence tags (ESTs) and high-density genetic maps has paved the way to the genetic engineering and molecular breeding of fruit plants for crop improvement. The application of these modern biotechnological techniques to fruit plants can contribute efficiently to solve or reduce the problems faced by the fruit industries. In the last decade, the emphasis on crop improvement using novel genomic tools has shifted toward the identification and functional analysis of miRNAs, one of the hottest research fields in plant sciences (Sun, 2012; Sun *et al*., 2012a). However, as compared to cereals and other crops, the progress of genomic studies in horticultural crops is relatively slow. Availability of next generation sequencing (NGS) technologies like FLX-454, Illumina, SOLiD and Helicose have brought hopes to generate genomic resources for many more horticultural crops in few years time. Therefore, the horticulture breeders should equip themselves to make use of this extensive genome information in their varietal development programmes. The objectives of this review were to take stock of availability of genomic resources in horticultural crops, compile this information at one place, makes sequence information useful for breeders and to identify potential future challenges which one can face while making proper use of genomic resources. The term 'genomics' is coined by Dr. Tom Roderick, a geneticist at the Jackson Laboratory (Bar Harbor, ME) in 1986. Genomics is a discipline in genetics that applies recombinant DNA, DNA sequencing methods, and bioinformatics to sequence, assemble, and analyze the function and structure of genomes (the *complete* set of DNA within a single cell of an organism). The field includes efforts to determine the entire DNA sequence of organisms and fine-scale genetic mapping. The field also includes studies of intragenomic phenomena such as heterosis, epistasis, pleiotropy and other interactions between loci and alleles within the genome.

# Types of genomics

**Structural genomics**: is the study to characterize the structure of the genome.
**Functional genomics**: is a field of molecular biology that attempts to make use of the vast wealth of data produced by genomic projects (such as genome sequencing projects) to describe gene (and protein) functions and interactions.

**Comparative genomics**: is a field of biological research in which the genomic features of different organisms are compared. The genomic features may include the DNA sequence, genes, gene order, regulatory sequences, and other genomic structural landmarks.

**Needs of Genomic studies**

1. To know sequence of whole genome
2. To locate genes on chromosomes & transcribed regions
3. To locate promoters and function of motif
4. Genome manipulation & MAS
5. Evolutionary studies
6. To discover and use of genetic variation

**Advanced genomic approaches for fruit breeding**

1. NGS platforms
2. Advanced Molecular markers & Marker assisted breeding (MAB)
3. GWAS, QTL mapping and Association mapping
4. Comparative genomics
5. Genomics assisted selection

## The detail approaches and its utilization in fruit breeding as follows:

### 1. Genome sequencing projects in fruit crops

Since the publication of the first plant genome sequence, that of Arabidopsis thaliana (Arabidopsis Genome Initiative, 2000), technology based on plant genomics has advanced substantially and revolutionized our understanding of plant biology by unraveling basic mechanisms in plant development, tolerance to biotic and abiotic stresses and adaptation (Feuillet et al., 2011; Hamilton & Buell, 2012). With the significant advancement and achievement in DNA sequencing technology in recent years, efforts have been made to release the draft sequences of many fruit plant species and many more sequencing projects for fruit plants are in progress. Among fruit plants, grapevine was the first crop genome to be sequenced and the first draft with 8 times high-quality sequence was released by the International Grape Genome Project (Jaillon *et al*., 2007). Recently, Velasco *et al*. (2010) reported a high-quality draft genome sequence of the domesticated apple (*Malus domestica*). The putative gene content in apple (57,386) was higher compared to the model plant, *A. Thaliana* (Arabidopsis Genome Initiative, 2000), or other fruit plants of the family Rosaceae such as the pear (Wu *et al*., 2013). Recently Azim *et al*., 2014 has been characterized mango leaf transcriptome and chloroplast genome using next generation DNA sequencing. Blast searching against nonredundant nucleotide databases and several Viridiplantae genomic datasets annotated 24,593 mango unigenes (80 % of total) and identified *Citrus sinensis* as closest neighbour of mango with 9,141 (37 %) matched sequences. The draft mango cp genome size is 151,173 bp with a pair of inverted repeats of 27,093 bp separated by small and large single copy regions, respectively. Out of 139 genes in mango cp genome, 91 found to be protein coding. Sequence analysis revealed cp genome of *C. sinensis* as closest neighbour of mango.

## 2. Advanced molecular markers

A DNA marker is typically derived from a small region of DNA that shows sequence polymorphism between individuals within a species. Thousands of phenotypically neutral, random DNA markers (RDMs) can be generated for any species and have been successfully used in many studies to represent genomes in biodiversity studies (Sunnucks 2000) or to map trait genetic linkage between a specific RDM and a target locus allele, established by QTL studies, for example, can be broken by genetic recombination; this limits the use of RDMs as a diagnostic tool (Rafalski & Tingey 1993). Recently, projects on structural and functional genomics have been established for several crop species (e.g. http://www.grasp-euv.dk/; http://www.maizegenetics.net/).The knowledge generated in these projects will allow systematic development of functional markers (FMs), which are derived from polymorphic sites within genes causally affecting phenotypic trait variation. So far, DNA markers have been compared primarily on their technical properties (Gupta *et al*., 2002). By contrast, RDMs, gene targeted markers (GTMs) and FMs are here defined based on the level of functional characterization of the polymorphisms monitored by these marker types. RDMs are derived at random from polymorphic sites in the genome, whereas GTMs are markers derived from polymorphisms within genes. Both RDMs and GTMs can be developed independently of their relationship to any phenotypic characters. By contrast, FMs are derived from polymorphic sites within genes causally involved in phenotypic trait variation. DNA markers derived from functionally defined sequences are mentioned in the context of plant breeding (Eujayl *et al*., 2002) and biodiversity studies. Terms such as 'functional', 'gene targeted' and 'diagnostic' markers have been used (Hackauf & Wehling , 2002) but have not been clearly defined.

## 3. Expressed sequence tagged (EST)

In recent years, the development of faster and less-expensive sequencing technologies has generated an exponential growth in the number of ESTs for many plant species (Parkinson & Blaxter, 2009). Earlier, the significant fruit EST resources were widely used to identify genes likely to be involved in fruit ripening, flavor development, control of color, the synthesis of chemicals and the generation of aroma (Crowhurst *et al*., 2008; Schaffer *et al*., 2007; Vecchietti *et al*., 2009). In addition, ESTs are valuable resources of simple sequence repeats (SSRs) and single-nucleotide polymorphisms (SNPs), which are useful markers for the genetic diversity analysis, phylogenetic studies and creating genetic maps in plants. Publically available EST sequences have become particularly attractive resource for the identification of SSR markers in a number of fruit plants, for instance, apple, pineapple and grapevine. The main advantage of EST–SSRs (also known as genic microsatellite marker; Varshney *et al*., 2007) over genomic SSRs is the detection of variation in the expressed portion of the genome (Kalia *et al*., 2011). Hence, they are expected to have higher chance to be associated with phenotypes (Wang et al., 2012a). In many fruit crops, sequence data for many fully characterized genes and full-length cDNA clones have been generated and these sequences are available in online databases in the public domain (Sonah *et al*., 2011). Among all fruit crops, grapevines have the largest number of ESTs in databases followed by apple and citrus. EST resources for many major fruit crops, such as apple, grapes, citrus, papaya, banana, are increasing with great pace.

## 4. Marker assisted breeding

Mapping and tagging of agriculturally important genes have been greatly facilitated

by an array of molecular markers in crop plants. Marker-assisted selection (MAS) is gaining considerable importance as it would improve the efficiency of plant breeding through precise transfer of genomic regions of interest (foreground selection) and accelerate the recovery of the recurrent parent genome (background selection). MAS have been widely used for simple inherited traits than for polygenic traits, although there are few success stories in improving quantitative traits through MAS. They are been used to monitor DNA sequence variation in and among the species and create new sources of genetic variation by introducing new and favourable traits from landraces, wild relatives and related species and to fasten the time taken in conventional breeding, germplasm characterization, genetic mapping, gene tagging and gene introgression from exotic and wild species. The success of MAS depend on many critical factors such as the number of target genes to be transferred, the distance between the target gene and the flanking markers, number of genotypes selected in each breeding generation, the nature of germplasm and the technical options available at the marker level. The power and efficiency of genotyping are expected to improve with the advent of markers like single nucleotide polymorphisms (SNP). Although genetic maps have been developed for most important fruit and vegetables species and a number of horticultural important gene loci have been tagged, only a few are reported. New, easy to perform allele testing methods are needed to bridge this large gap between marker development and application. This review discusses the basic requirements and the potential applications of MAS and the significance of integrating MAS into conventional plant breeding programmes.

5. **Mapping Genomes**

Assigning/locating of a specific gene to particular region of a chromosome and determining the relative distances.

**Physical mapping**: The DNA is cut by a restriction enzyme. Once cut, the DNA fragments are separated by electrophoresis. The resulting pattern of DNA migration (i.e., its genetic fingerprint) is used to identify what stretch of DNA is in the clone. By analyzing the fingerprints, contigs are assembled by automated (FPC) or manual means (Pathfinders) into overlapping DNA stretches. Now a good choice of clones can be made to efficiently sequence the clones to determine the DNA sequence of the organism under study (seed picking). Once the map is determined, the clones can be used as a resource to efficiently contain large stretches of the genome. This type of mapping is more accurate than genetic maps. Genes can be mapped prior to the complete sequencing by independent approaches like in situ hybridization.

**Genetic mapping**: A representation of the location of genes or genetic markers with respect to each other, based on recombination frequencies. A genetic map is linear, and distances between loci are measured in recombination percentage = map units = centimorgans. The farther apart two loci are, the more likely that a crossover will occur between them. Conversely, if two loci are close together, a crossover is less likely to occur between them. Recombination can only be detected between two loci, both of which are heterozygous. The dominant/recessive relationships must allow for detection of recombinants. The most useful systems involve codominant alleles. Efficient mapping requires polymorphic loci, i.e. loci with two or more common alleles. Loci that have a single common allele are described asmonomorphic. Any variations in DNA, whether in coding regions of genes or in noncoding regions, can be used as genetic markers, i.e. as a label for a particular point on a chromosome.

**Map-based or Positional Cloning**: using the genetic relationship between a gene and a marker as the basis for beginning a search for a gene.
*The steps in positional cloning as follows:*

- Identify a marker tightly linked to your gene in a "large" mapping population
- Find a YAC or BAC clone to which the marker probe hybridizes
- Create new markers from the large-insert clone and determine if they co-segregate with your gene
- If necessary, re-screen the large-insert genomic library for other clones and search for co-segregating markers
- Identify a candidate gene from large-inset clone whose markers co-segregate with the gene
- Perform genetic complementation (transformation) to rescue the wild-type phenotype
- Sequence the gene and determine if the function is known

**6. Association mapping**

Also known as association genetics or linkage disequilibrium mapping is a method of mapping quantitative trait loci (QTLs) that takes advantage of historic linkage disequilibrium to link phenotypes (observable characteristics) to genotypes (the genetic constitution of organisms). Association mapping is based on the idea that traits that have entered a population only recently will still be linked to the surrounding genetic sequence of the original evolutionary ancestor, or in other words, will more often be found within a given haplotype, than outside of it. It is most often performed by scanning the entire genome for significant associations between a panel of SNPs (which, in many cases are spotted onto glass slides to create "SNP chips") and a particular phenotype. These associations must then be independently verified in order to show that they either (a) contribute to the trait of interest directly, or (b) are linked to/ in linkage disequilibrium with a quantitative trait locus (QTL) that contributes to the trait of interest.

*Advantages of AM over linkage mapping:*
1. Much higher mapping resolution.
2. Greater allele number and broader reference population.
3. Possibility of exploiting historically measured trait data.
4. Less research time in establishing an association

**Genome wide association mapping (GWAS)**: also known as whole genome association study (WGA study, or WGAS), is an examination of many common genetic variants in different individuals to see if any variant is associated with a trait. GWAS typically focus on associations between single-nucleotide polymorphisms (SNPs) and traits like major diseases.

**Why are GWAS possible now?**
Improved Genotyping Technology:
- Now have ability to type millions of SNPs in one reaction on a "SNP chip." The cost can be as low as $200-$300 per person.
- Two primary platforms: Affymetrix and Illumina.

Design and analysis:
- Availability of SNP databases, HapMap, and other resources to identify the SNPs and improve chip design/capacity.
- Faster computers to carry out the millions of calculations make analyses more amenable.

Types of GWAS

- Population-based: Large cohort with genome profile and measurements on quantitative traits for each individual.
- Case-Control:
  Case cohort: Sample with a diagnose (e.g. a disease)
  Control cohort: Neutral sample

*In a summarized form the following types of population association study exists:*

1. Candidate causative polymorphism
   – SNP (single nucleotide polymorphism), deletion, duplication
2. Candidate causative gene (5-50 marker SNPs)
   – evidence from linkage study or function
3. Candidate causative region (100s of marker SNPs)
   – evidence from linkage study
4. Genome-wide – GWAS (>300,000 marker SNPs)
   – no prior evidence required

## 7. Genomic assisted selection

The past two decades have seen a striking decrease in the cost and effort needed to read a plant's DNA sequence, and qualitative improvements in the infrastructure needed to store and analyze these data. These advances have provided a means both to reduce the number of cycles in horticultural breeding programs and to increase the precision and efficiency of new cultivar development. Among the most rapidly developing approaches is genome-wide selection (GWS). GWS makes use of genomic estimated breeding values (GEBVs) as selection parameters, rather than the estimated breeding values (EBVs) traditionally used by fruit breeders. GEBVs are derived for individuals in a phenotyped training population using dense genome-wide single-nucleotide polymorphism (SNP) markers, to establish marker effects on complex phenotypes controlled by a large number of genetic loci. Individuals in breeders' selection populations are then screened and GEBVs of individuals calculated based on genetic marker information, in order to identify outstanding 'elite' individuals (Figure 3). These may then be used to advance generations, or evaluated in the field as potential cultivars.

## Summery

Ultimately, the goal of the breeder will be to assay the genetic makeup of individual plants rapidly and to select desirable genotypes in breeding populations. The construction of 'graphical genotypes' of each plant or progeny row would allow the breeder to determine which chromosome sections are inherited from each parent to facilitate the selection process and perhaps to reduce the need for extensive field tests . A logical extension of whole genome selection for the breeder would be to design the superior genotypes in silico, an approach described as 'breeding by design'. Thus, in the post-genomics era, high-throughput approaches combined with automation, increasing amounts of sequence data in the public domain and enhanced bioinformatics techniques will contribute to genomics research for crop improvement. However, the costs of applying genomics strategies and tools are often more than is available in commercial or public breeding programmes, particularly for inbreeding crops or crops that are only of regional importance. Nevertheless, marker-assisted breeding or marker-assisted selection will gradually evolve into 'genomics-assisted breeding' for crop improvement. Newly developed genetic and genomics tools will enhance, but not replace, the conventional breeding and evaluation process. The ultimate test of the value of a genotype is its performance in the target environment and acceptance by farmers.

**References (if any)**

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, **408**: 796–815.

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

Crowhurst R, Gleave A, MacRae E, *et al*. 2008. Analysis of expressed sequence tags from Actinidia: applications of a cross species EST database for gene discovery in the areas of flavor, health, color and ripening. *BMC Genomics*, **9**: 351

Eujayl, I. *et al*. (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. *Theor. Appl. Genet*., **104**: 399–407.

Feuillet C, Leach JE, Rogers J, *et al*. 2011. Crop genome sequencing: lessons and rationales. *Trends Plant Sci*, **16**: 77–88.

Gupta, P.K. *et al*. 2002. Molecular markers: principles and methodology. In Molecular Techniques in Crop Improvement (Jain, S.M. *et al*., eds), pp. 9–54, Kluwer Academic Publishers.

Hackauf, B. and Wehling, P. 2002. Identification of microsatellite polymorphisms in an expressed portion of the rye genome. *Plant Breed*. 121, 17–25.

Hamilton, J.P., Buell, C.R. 2012. *Advances in plant genome sequencing. Plant J*, **70**: 177–90.

Jaillon, O., Aury, J.M., Noel, B. *et al*. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, **449**: 463–7.

Janick, J. 2012. Fruit breeding: Past, Present, and Future. XXII Congresso Brasilerio de. Fruiticultura, Bento Goncalves-RS, 22a 26 de outtubro de, pp 1-22.

Kalia, R.K., Rai, M.K., Kalia, S, *et al*. 2011. Microsatellite markers: an overview of the recent progress in plants. *Euphytica*, **177**: 309–34.

Parkinson, J., Blaxter, M. 2009. Expressed sequence tags: An overview. In: Parkinson J, ed. Expressed sequence tags (ESTs): generation and analysis. Springer, Berlin; *Humana Press*, **533**: 1–12.

Rafalski, J.A. and Tingey, S.V. 1993. Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. *Trends Genet*. **9**: 275–280.

Schaffer, R.J., Ellen, N., Friel, E.N., *et al*. 2007. A genomics approach reveals that aroma production in apple is controlled by ethylene predominantly at the final step in each biosynthetic pathway. *Plant Physiol*, **144**: 1899–912.

Singh, N.K., Mahato, A.K., Sharma, N., Gaikwad, K., Srivastava, M., Tiwari, K.,

Dogra,V., Rawal S.C., Rajan, S., Singh A.K. & Sharma, T.R. 2014. A draft genome of the king of fruit, mango (*Mangifera indica* L.). Plant and Animal Genome XXII Conference, San Diego, CA.

Sonah H, Deshmukh RK, Singh VP, *et al*. (2011). Genomic resources in horticultural crops: Status, utility and challenges. *Biotechnol Adv*, **29**: 199–209.

Sunnucks, P. 2000. Efficient genetic markers for population biology. *Trends Ecol. Evol*. 15: 199–203.

Varshney, R.K., Mahendar T. & Aggarwal, R.K. 2007. Genomics-Assisted Crop Improvement: (Ed) Varshney, R.K. & Tuberosa, R. Vol. 1: Genomics Approaches and Platforms, Springer,13–29.

Vecchietti, A., Lazzari, B., Ortugno, C., *et al*. 2009. Comparative analysis of expressed sequence tags from tissues in ripening stages of peach (*Prunus persica* L.). *Tree Genet Genomes*, **5**: 377–91.

Velasco, R., Zharkikh, A., Affourtit, J., *et al*. 2010. The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nat Genet*, **42**: 833–9.

Wang, X.,C., Guo, L., Shangguan, L.F., *et al*. 2012a. Analysis of expressed sequence tags from grapevine flower and fruit and development of simple sequence repeat markers. *Mol Biol Rep*, **39**: 6825–34.

Wu, J., Wang, Z., Shi Z, *et al*. 2013. The genome of pear (*Pyrus bretschneideri* Rehd.). *Genome Res*, **23**: 396–408.

## Author accepts the following terms for publication on BiotechArticles.com

- This form is for genuine submissions related to biotechnology topics only.
- You should be the legal owner and author of this article and all its contents.
- If we find that your article is already present online or even containing sections of copied content then we treat as duplicate content - such submissions are quietly rejected.
- If your article is not published within 3-4 days of emailing, then we have not accepted your submission. Our decision is final therefore do not email us enquiring why your article was not published. We will not reply. We reserve all rights on this website.
- Do not violate copyright of others, you will be solely responsible if anyone raises a dispute regarding it.
- Similar to paper based magazines, we do not allow editing of articles once they are published. Therefore please revise and re-revise your article before sending it to us.
- Too short and too long articles are not accepted. Your article must be between 500 and 5000 words.
- We do not charge or pay for any submissions. We do not publish marketing only articles or inappropriate submissions.
- Full submission guidelines are located here: http://www.biotecharticles.com/submitguide.php
- Full Website terms of service are located here: http://www.biotecharticles.com/privacy.php

As I send my article to be published on BiotechArticles.com, I fully agree to all these terms and conditions.

---