

# Significant Role of Genomics in Crop Breeding: An Overview

Anoushka Kotra<sup>1\*</sup>, Neetu Jabalia<sup>2</sup> and Nidhee Chaudhary<sup>3</sup>

<sup>1,2,3</sup>Amity Institute of Biotechnology, Amity University, Uttar Pradesh, Noida  
E-mail: <sup>1</sup>nushibushi20@gmail.com, <sup>2</sup>njabalia@amity.edu, <sup>3</sup>nchaudhary@amity.edu

**Abstract**—The role of genomics in crop breeding is of great value to agriculture and human kind. Genomes are the entire set of genes present on the chromosome of the individual plant and thus act as the genetic material of the plant. The study and analysis of all the constituents of this genetic material is called genomics. The science of genetics and genomics can be applied to the field of agriculture. It uses various tools. The knowledge of EST's (Expressed Sequence Tag's) are used by researchers for comparison purpose to have further information on gene function. Genetic maps of various crops such as *Glycine max* help various researchers to locate a specific gene on a chromosome. The phenotype and genotype data can be integrated using the QTL (Quantitative Trait Loci). The determination of interaction of thousand of gene combinations can be done using DNA chips. In this overview, we describe significance of genomics that can be used for improvement in crop breeding.

**Keywords:** Genomes, Genomics, DNA chips, EST,

## 1. INTRODUCTION

The organism's genetic material consists of an entire set of DNA that carries all the genes. These are known as genomes.[1]

The study for sequencing, compilation and inspection of properties and composition of genomes that uses r-DNA(recombinant DNA), DNA sequencing procedures and bioinformatics is called as Genomics.[2-3] The field involves genetic mapping and skills for analyzing the DNA sequence of a particular organism.[4] Even the brain that is considered to be the most complicated biological system can be examined with the increasing development in the field of genomics.[5] The entire DNA is not made up of genes. In case of humans, only 3% of the entire DNA are genes. The field of bioinformatics attempts to use tools involving statistics and some data mining methods that help in determination of genes. The molecular composition of a DNA sequence can be unraveled by the efforts made in structural genomics.

Various applications of genomics can be applied to all forms of organisms. It aims to enhance the ability to cure diseases such as sickle cell anemia, cystic fibrosis, etc. In case of plants, it has its application in the field of agriculture such as crop

breeding. Various genetic maps of different crops for instance *Glycine max* have been discovered.[6]

## 2. METHODOLOGY

The genome analysis can be divided into the following components- DNA sequencing, assembly of the sequence and analyzing the representation of the original chromosome obtained from the assembly of the sequence.[4]

**2.1.Sequencing-** Research laboratories having expensive instruments and high tech facilities can perform sequencing in specialized sequencing centers.

The genomic sequencing can be further classified into two broad categories namely shotgun and high throughput sequencing.[7-8]

### 2.1.1 Shotgun sequencing

Shotgun sequencing can also known be as Sanger sequencing. A DNA sequence consisting upto a 1000 base pairs can be analyzed using this method.[9] By the means of fragmentation and sequencing one can easily obtain about a number of reads for the target DNA. These can further be assembled (Fig. 1) into a regular sequencing using various computational tools that utilizes the overlapping ends of the reads obtained.[9-10]

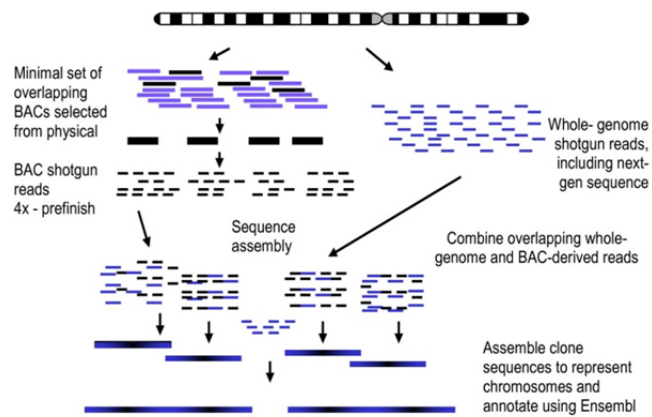


Fig. 1: Shotgun sequencing method

DNA templates, DNA primers, DNA polymerase, deoxynucleotidetriphosphates(dNTP's), and modified nucleotides are of prime importance for the chain termination method.[4] In 48 runs a day of a single batch one can easily sequence upto a100 samples of DNA(approx).[11]

### 2.1.2 High-throughput sequencing-

This sequencing method (Fig. 2) aims to minimize the cost of DNA sequencing. 500,000 sequencing-by-synthesis operations can be run parallel.[12] This further includes illumina sequencing and ion torrent.

**Illumina sequencing:** Reversible dye-terminator technology is the basic principle of this kind of sequencing method. The DNA colonies are formed using this technology by amplification of DNA molecules that are first laid on the slide and amplified using polymerase.[13-14]

**Ion torrent:** The principle of this technology is the standard DNA replication chemistry. A microwell constitutes a DNA template that has a nucleotide. It determines if the nucleotide is complementary to the template causing the release of a hydrogen ion.[15]

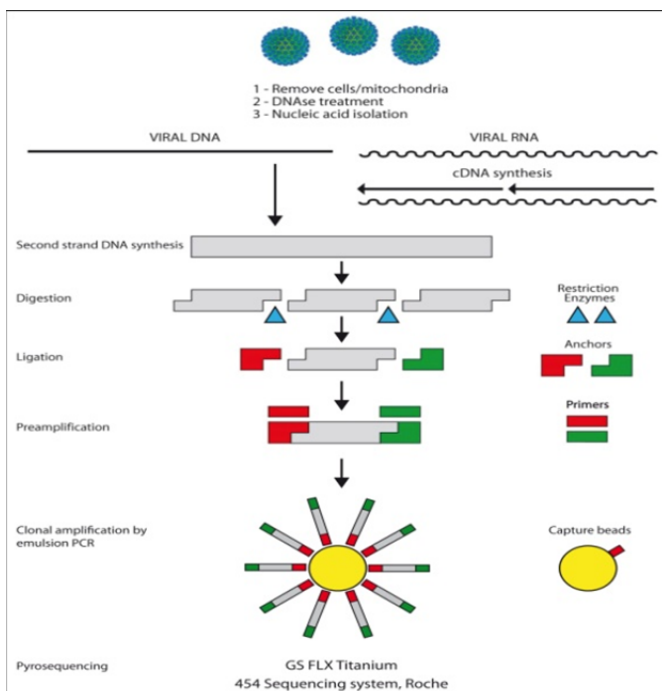


Fig. 2: Highthroughput sequencing

**2.2. Assembly-** The reconstruction of the original sequence can be done by orientation and compilation of fragments of a long DNA sequence. This is done in order to read the sequence in a regular pattern as the current DNA sequencing technology reads the sequence by first breaking it into small pieces of 20-100bases.[4]

### 2.3. Annotation-

The linking of biological information to sequences is referred to as genome annotation. Three important steps are involved in this process-

1. Identification of parts of genomes that code for proteins.
2. Gene prediction which is the identification of the elements on a genome.
3. And finally the attachment of biological information to the sequence.[16]

## 3. APPLICATION OF GENOMICS

**3.1. Metagenomics:** It is applied to micro organisms. The physiology and ecology of micro organisms can be studied by facilitating the technical invention. Bacteriorhodopsin was the first from the bacterial origin that was discovered in the field of metagenomics. The field of metagenomics aims towards connecting genomic analysis with pure culturing studies in order to improve culturing strategies.[17]

### 3.2. Genomic tools in plant breeding:

Various genomic tools can be used for determining the relationship between the genotype and the phenotype of the plant species. The bulk sequencing of genomes and transcriptomes can be done using the Next Generation Sequencing (NGS) technologies. Screening mutants and collections of germplasm for allelic variants in target genes can also be done using some genomic tools such as TILLING and EcoTILLING. High-density genetic maps are also constructed via these tools leading to 'superdomestication' of crops.[18]

### 3.3. Genomics in healthcare

Genetic changes in a tumor can be tested leading to the invention of a drug that can be further used to target at a specific site within that tumor. A specific treatment or diagnosis can be performed by genetically testing the drug. Prognosis, diagnosis, gene stem cell therapy, optimization of dose, etc are among others that can be performed using various tools of genomics.[19]

## 4. BIOINFORMATICS GENOMICS TOOLS USED IN CROP BREEDING

The development in technology provides caliber to examine the plant phenotype in new ways. Variation in metabolites and ions are contributed by the genetic variation so as to study the phenotype at high resolution that can be combined with genetics. The main focus of plant genomics is to improve the crop yeild for the identification of genotypes with optimal agronomic traits.[20]As much as 50% yeild is reduced in crop plants due to abiotic stresses.[21] Among cereals wheat and barley are the main constituents of world food supply. Thus, drought tolerance of cereals is of prime importance. Also recent technological development and challenges faced in the

field of agriculture have led to the emergence of various genomic tools that can be used to explore and exploit the plant genomes for crop improvement.[22]

There are various bioinformatics tools that can be used for genomic analysis in crop breeding. These are:

**NCBI- bioproject-** The information related to various plant specimens and their genomic sequences can be retrieved from the NCBI(National Center for Biotechnology Information) site. About 115 partial sequences and 40,000 EST's(Expressed Sequence Tags) of different plant species are present in the current version of NCBI. Thus, it is important to become familiar with this site as it constitutes some very vital information, data links and resources of many agronomically important crops.

**Phytozome-** The site (Fig. 3) is designed to facilitate and carry out various comparison related studies among different green plants. It includes all the necessary genome sequences and data sets for various crop specimens. It currently consists of about 31 different plant species that are wholly or partially sequenced.

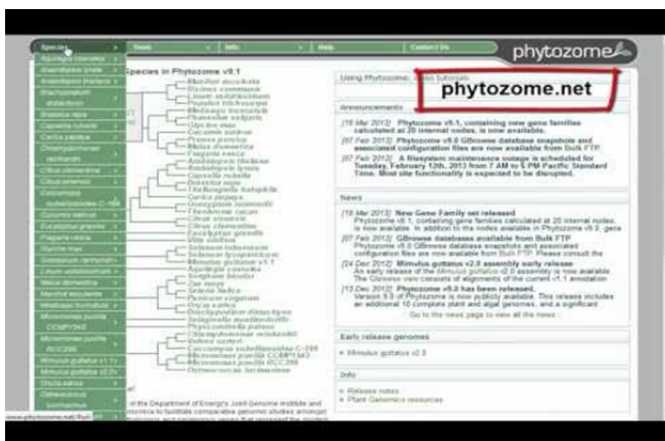


Fig. 3: phytozome site

**NCBI-BLAST-** Familiarism between nucleotide sequences in databanks can be easily determined with the help of this site. Searches such as translation of nucleotide sequences, conservation domains and numerous alignment tools can be applied to all organisms or to a specific plant specie (Fig. 4).

**Graingenes and PlantGDB-** The following database is designed specifically for Triticeae and Avena genes, markers, maps and germplasm. It is valuable in the case of agricultural grain species however it carries limited information than the NCBI that carries information about most plants.

**Molecular (DNA) Markers-** The greater potential gains in breeding for QTL and traits with low heritability for direct selection can be done with the use of DNA markers. The complex nature of multiple alleles are known among crop

breeders since decades. With the invention of molecular markers, genetic markers and other forms of genetic structures in crop breeding can now take place.



Fig. 4: NCBI-BLAST homepage

**Restriction fragment length polymorphism(RFLP)-** The technique of RFLP is capable to provide data of high quality. However, it has certain restrictions on throughput due to requirement of large amounts of data.

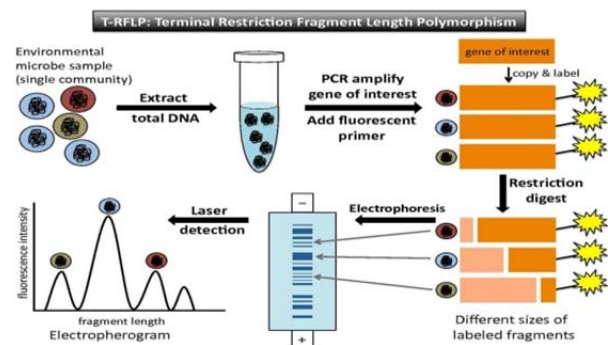


Fig. 5: Technique of RFLP.

**Simple Sequence Repeats(SSR)-** These are the most costly to develop. The markers are species and cultivar specific. In order for these to develop, they often require extensive band sequencing data.

**Single Nucleotide Polymorphism(SNP)-** The technique does not require electrophoresis which is the biggest advantage as far as automation is concerned. However, DNA sequencing is of prime importance but is very expensive.

**EST-** EST is referred to as Expressed Sequence Tag. It involves synthesising of cDNA from RNA. These are the only markers that are based on RNA. The technique can be used when there is extensive sequencing for crop species.

**NCBI-plant markers-** It constitutes various molecular markers such as SNP, SSR, conserved ortholog set cosmid

(COS) markers and primers from a number of plant species in a genetic marker database.[23]

## 5. ROLE OF GENOMICS IN CROP BREEDING

There has been an immense improvement in crop breeding. Over the years many annotated sequences of various crop species along with their genetic maps have been developed. The field of plant genomics mainly aims towards shedding light on the evolution, dynamics and complexity of plant genomes that lead to an immense understanding of biological systems of plants. The need of the hour is for developing plant genomics to improve crop breeding.[24]

**Prediction of gene-to-phenotype relationships-** the quantitative trait loci(QTL) shows caliber to minimize the genotype-phenotype gap in plant species. This QTL information has been incorporated into plant models for analyzing the genotype-by-environment interactions.[25]

**Drought tolerance improvement in crop plants-** there are three main approaches for creating drought resistance crop species.

- i. Plant physiology helps us in understanding the complex network of drought related traits with the development of new tools.
- ii. Molecular genetics on the other hand helps us discover QTL's that affect the yield under drought conditions.
- iii. Molecular biology that helps in providing genes which can be further used as candidate sequences.[26]

**Improving salt tolerance in plants-** The lack of availability of genes that confer salt resistance along with the lack of knowledge in understanding the molecular basis of salt tolerance has restricted the development in breeding for salt tolerant plants. Thus genetics and genomics provide suitable evidence regarding the characteristics and variations in crop species that can be used or exploited effectively to develop and improve salt tolerance in plants.[27]

**Crop improvement through modification of plant's own genome-** a plant derived (P-) DNA fragment was used to replace the Agrobacterium transfer (T-) DNA. The method that was used here could be able to produce multiple marker-free and backbone free potato plants that could display less amount of tuber-specific polyphenol oxidase gene in potato.[28]

## 6. CONCLUSION

Genomics has a very important role to play in improving the crop breeding. The field of genomics focuses on generating some new useful tools for instance, functional molecular markers and informatics. Also it aims towards providing some knowledge about the statistical and inheritance phenomena in crop species for their efficiency and improvement in breeding practices. The designing of genotypes of crop species have become easy by using various tools that help to identify

relative alleles at a loci in a single population. However, currently its expensive to use thus restricting the implementation of genomics assisted crop improvement.[29]

## REFERENCES

- [1] Ridley, M., *Genome*, New York, 2006.
- [2] "National Human Genome Research Institute", *A brief guide to Genomics*, November 8 2010
- [3] *Concepts of genetics*, San Francisco, 10, 2012.
- [4] *Pevsner, Hoboken, N.J, Wiley-Blackwell, Bioinformatics and functional genomics*, 2, 2009.
- [5] "National Human Genome Research Institute", *FAQ About Genetic and Genomic Science*, November 8 2010.
- [6] <http://www.faqs.org/espionage/Fo-Gs/Genomics.html>
- [7] M. Baker, "Nature News Blog", *Benchttop sequencers ship off*, September 14 2014
- [8] Quail, MA, Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P., Gu and Y, "A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers", *BMC Genomics*, 341, 13, 2012.
- [9] Staden and R., *A strategy of DNA sequencing employing computer programs*, *Nucleic Acids Research*, 7, 6, June 1979, pp. 2601-2610.
- [10] Anderson and S., "Shotgun DNA sequencing using cloned DNase I-generated fragments", *Nucleic Acids Research*, 13, 9, July 1981, pp. 3015-3027.
- [11] "An Introduction to Next-Generation Sequencing Technology", USA, February 28 2012.
- [12] Tucker, T., Marra, M., Friedman and J.M., "Massively parallel sequencing: the next big thing in genetic medicine", *American Journal of Human Genetics*, 85, 2, August 2009, pp. 142-154.
- [13] Kawashima, E.H., Farinelli, L., Mayer and P., "Method of nucleic acid amplification", May 12 2005.
- [14] Mardis and E.R., "Next-generation DNA sequencing methods", *Annual Review of Genomics and Human Genetics*, 9, 2008, pp. 387-402.
- [15] Davies and K., "Powering Preventative Medicine": *Bio ITWorld*, 2011.
- [16] Brent and M.R., "Steady progress and recent breakthroughs in the accuracy of automated genome annotation", *Nature Reviews*, 9, 1, January 2008, pp. 62-73.
- [17] *Microbiol Mol Biol Rev.*, 68, 4, December 2004, pp. 669-685.
- [18] *Curr Genomics*, 13, 3, May 2012, pp. 179-195.
- [19] <http://www.geneticseducation.nhs.uk/genomics-in-health>
- [20] Flavell and R., "From genomics to crop breeding", *Nat Biotechnol*, 28, 2010, pp. 144-145.
- [21] F.Qin, K. Shinozaki and K. Yamaguchi, "Achievements and challenges in understanding plant abiotic stress responses and tolerance," *Plant and Cell Physiology*, 52, 9, 2011, pp. 1569-1582.
- [22] file:///C:/Users/hcl/Downloads/i14\_3\_05.pdf
- [23] file:///C:/Users/hcl/Downloads/9781461470274-c1%20(2).pdf
- [24] [http://grzegorz\\_bartoszewski.users.sggw.pl/images/Plant\\_Genomics.pdf](http://grzegorz_bartoszewski.users.sggw.pl/images/Plant_Genomics.pdf)

- 
- [25] Xinyou, YinPaul, C. Struik, Martin and J. Kropff, "Role of crop physiology in predicting gene-to-phenotype relationships", *The Netherlands*, 9,9, September 2004, pp.426-432.
- [26] L. Cattivelli, F. Rizza, F.W. Badeck, E. Mazzucotelli, A.M. Mastrangelo, E. Francia, C. Marè and A. Tondelli, A.M. Stanca, "Drought tolerance improvement in crop plants: An integrated view from breeding to genomics", 105,1-2, January 2 2008, pp. 1-14.
- [27] V. Chinnusamy, A. Jagendorf and J.K. Zhu, "Understanding and Improving Salt Tolerance in Plants", 45,2, pp.437-448.
- [28] C.M. Rommens, J. M. Humara, J. Ye, H. Yan, C. Richael, L. Zhang, R. Perry and K. Swords, "Crop Improvement through Modification of the Plant's Own Genome", *Plant Physiology*, 135, May 7 2004, pp.1421-1431.
- [29] R. K. Varshney, A. Graner and M.E. Sorrells, "Genomics-assisted breeding for crop improvement", 10,12, December 2005, pp.621-630.