Available online at www.ijpab.com

DOI: http://dx.doi.org/10.18782/2320-7051.5922

ISSN: 2320 – 7051

Int. J. Pure App. Biosci. **5 (6):** 1650-1659 (2017)

International Journal of Pure & Applied Bioscience

Review Article

OPEN ACCESS

Genome Sequencing and its Role in Plant Breeding

U. Mehraj[1*] and M. Ahmad[2]

[1]Division of genetics and Plant Breeding, Wadura, Sopore;[2]Mountain Agriculture Research and Extension Station, Kargil; Sheri- Kashmir University of Agriculture Science and Technology Shalimar Kashmir, India

*Corresponding Author E-mail: uzmamehrajskuastk@rediffmail.com

Received: 23.10.2017 | Revised: 30.11.2017 | Accepted: 4.12.2017

ABSTRACT

Plant breeding can be accomplished through many different techniques ranging from simply selecting plants with desirable characteristics for propagation, to more complex molecular techniques. Nowadays, the availability of genomic tools and resources is leading to a new revolution of plant breeding, as they facilitate the study of the genotype and its relationship with the phenotype, in particular for complex traits. Gene sequencing is the process of determining the precise order of nucleotides within a DNA molecule. The entire information lies in its base sequence and the determination of this is of central importance. Gene sequencing includes several methods or technologies that are used to determine the order of the four bases—adenine, guanine, cytosine, and thymine—in a strand of DNA. These genomics-based approaches are having a profound influence on the improvement of crops. Identification of all genes within a species allows an understanding of how important agronomic traits are controlled. It is a prerequisite resource for fully understanding the role of genes in development and efficiently exploiting the natural and induced genetic diversity of an organism. Following the publication of the first plant genome sequence, Arabidopsis (Arabidopsis thaliana), in 2000 the sequencing of large and complex genomes of crop species, facilitated by new sequencing technologies and bioinformatic approaches, has provided new opportunities for crop improvement. With fast development and wide applications of next-generation sequencing (NGS) technologies, genomic sequence information is within reach to aid the achievement of goals to decode life mysteries, make better crops, detect pathogens, improve life qualities and allows the mass sequencing of genomes and transcriptomes, which is producing a vast array of genomic information. The analysis of NGS data by means of bioinformatics allows discovering new genes and makes available large collections of molecular markers.

Key words: *Gene sequencing, Next-generation sequencing (NGS), Genome, Molecular markers.*

INTRODUCTION

Ever since the beginnings of the domestication of plants, some 10,000 years ago, plant breeding has been extremely successful in developing crops and varieties[1]. Application of conventional pre-genomics scientific breeding methodologies has led to the development of modern cultivars, which have contributed to the dramatic improvement of yield of most major crops since the middle of the 20th century.

Cite this article: Mehraj, U. and Ahmad, M., Genome Sequencing and its Role in Plant Breeding, *Int. J. Pure App. Biosci.* **5(6):** 1650-1659 (2017). doi: http://dx.doi.org/10.18782/2320-7051.5922

Copyright © Nov.-Dec., 2017; IJPAB

The success of plant breeding in the last century has relied in the utilization of natural and mutant induced genetic variation and in the efficient selection, by using suitable breeding methods, of the favorable genetic combinations. In this respect, the evaluation and identification of genetic variants of interest as well as the selection methodologies used have largely been based in the phenotypic evaluation. Nowadays, genomics, dealing with the collection and characterization of genes and analysis of the relationships between gene activity and cell function, provides breeders with a new set of tools and techniques that allow the study of the whole genome, and which represents a paradigm shift, by facilitating the direct study of the genotype and its relationship with the phenotype[2]. While classical genetics revolutionized plant breeding at the beginning of the 20th century, genomics is leading to a new revolution in plant breeding at the beginning of the 21th century. The field of genomics and its application to plant breeding are developing very quickly. The combination of conventional breeding techniques with genomic tools and approaches is leading to a new genomics-based plant breeding. One of the main pillars of genomic breeding is the development of Genome sequencing technologies. Genome sequencing is determining the precise order of nucleotides within the DNA molecule. The nucleotides include four bases i,e Adenine, Guanine, Thymine (Uracil) and cytosine. Advent of rapid DNA sequencing method has greatly accelerated biological and medical research and discovery. Knowledge of DNA sequencing has been indispensable for basic biological research, agriculture research and in numerous applied fields like Diagnostics, Biotechnology, forensic biology and biological systematic. The method was founded in 1975 by Frederick Sanger. The rapid speed of sequencing attained with modern DNA sequencing technology has been instrumental in the sequencing of complete DNA sequences, genomes of numerous types and species of life including human genome and other complete DNA sequences of many animals, plants and microbial species. The first full DNA genome to be sequenced was that of bacteriophage Φx174 in 1977. Historically there are two main methods of DNA sequencing or First Generation Sequencing Platforms. Several new methods for DNA sequencing were developed in the mid to late 1990s. These techniques comprise the "next – generation" sequencing methods or Second Generation Sequencing Platforms and includes 1) Solexa; based on whole genome sequencing. 2) SOLiD; based on ligation and detection. 3) 454; based on pyrosequencing and 4) Helicose; based on single molecule sequencing. Modern sequencing equipment uses the principles of the Sanger technique. With the increasing usage and new modification in next generation sequencing, the third generation sequencing or 'next-next' generation sequencing (NNGS) is coming out with new insight in the sequencing and Single-molecule real-time (SMRT) is the third-generation sequencing method.

Next Generation Sequencing (NGS), also known as high- throughput sequencing, allows mass sequencing of genomes and transcriptomes much more quickly and cheaply than the previously used sanger sequencing. In case of sanger sequencing speed of data generation is very low with high cost and also it involves cloning. Third-generation sequencing has two main characteristics. First, PCR is not needed before sequencing, which shortens DNA preparation time for sequencing. Second, the signal is captured in real time, which means that the signal, no matter whether it is fluorescent or electric current, is monitored during the enzymatic reaction of adding nucleotide in the complementary strand.

Application of Gene Sequencing in Plant Breeding

The field of genomics and its application to plant breeding are developing very quickly. The combination of conventional breeding techniques with genomic tools and approaches is leading to a new genomics-based plant breeding. In this new plant breeding context, genomics will be essential to develop more

efficient plant cultivars, which are necessary, according to FAO, for the new 'greener revolution' needed to feed the world's growing population while preserving natural resources. One of the main pillars of genomic breeding is the development of high-throughput DNA sequencing technologies, collectively known as next generation sequencing (NGS) methods. These and other technical revolutions provide genome-wide molecular tools for breeders (large collections of markers, high-throughput genotyping strategies, high density genetic maps, new experimental populations, etc.) that can be incorporated into existing breeding methods. These new technologies have reduced the cost of sequencing by more than one thousand times compared to Sanger technology, by avoiding time-consuming and tedious traditional cloning steps and making possible to perform millions of sequencing reactions in parallel[3]. Recent advances in genomics are producing new plant breeding methodologies improving and accelerating the breeding process in many ways (e.g., association mapping, marker assisted selection, 'breeding by design', gene pyramiding, genomic selection, etc.)[4].

The main application of gene sequencing in plant breeding are the following:

Genome and Transcriptome Sequencing

The availability of the whole genome sequence of a crop is of great utility for plant breeding. Until recently it has been an unachievable goal for most crops. This privilege was restricted to a reduced number of model species with small genomes and to projects with an important investment in time and resources, but now has extended to an increasing number of crops. However, it is also true that for important cultivated species with large and complex genomes such as wheat, sugarcane, or coffee, the sequencing of the whole genome is very challenging and may take several years before a draft is completed. Due to high cost of whole genome sequencing, transcriptome sequencing has been a cheaper alternative. The cDNA sequences (expressed sequence tags, ESTs) provide relevant information about the genes expressed in a certain tissue or organ, at a given stage of development and under particular environmental conditions. ESTs sequencing projects do not provide information about non-coding sequences and, even using diverse libraries, it is difficult to identify all genes and transcripts variants. Despite these limitations, ESTs collections have been very useful for breeders. The Sanger technology has been used to sequence several genomes as well as many transcriptomes. The first international collaborative project resulted in the whole genome sequence of the model plant *Arabidopsis thaliana.* After that reference genomes of selected genotypes were completed in a limited number of crops such as rice, maize, sorghum, populous, grapevine, papaya or soybean. The transcriptomes of most major crops, to a greater or lesser extent, were also sequenced. A global view of the genomes and transcriptomes sequenced can be obtained from the Gene Index Project or in the NCBI Unigene database.

Due to NGS technologies field of genomics has changed. These new technologies have reduced the cost of sequencing by more than one thousand times compared to Sanger technology, by avoiding time-consuming and tedious traditional cloning steps. By using these NGS technologies, the sequencing capacity of laboratories is continuously increasing. For instance, one High-Seq 2000 Illumina Sequencer is able to generate 55 Gb per day, which is roughly eighteen times the size of the human genome. Moreover, new, "third generation" platforms are being developed and incorporated to sequencing projects. Nowadays, it is feasible to sequence most crop genomes by combining Sanger with NGS technologies. A fully sequenced genome provides useful tools for the breeders, as it allows the discovery of genes, determining their position and function, as well as the development of large marker collections and high resolution maps. Many transcriptomes have also been sequenced, Sweet potato, squash[5], pigeonpea or buckwheat[6] represent some examples. These assays are showing the great complexity of plant transcriptomes,

allowing the identification of rare transcript variants that are being used to improve gene annotation and our knowledge of gene function and regulation.

RNA-seq (RNA Sequencing)

RNA-seq (RNA Sequencing) is also called Whole Transcriptome Shotgun Sequencing (WTSS), is a technology that uses the capabilities of next-generation sequencing to reveal a snapshot of RNA presence and quantity from a genome at a given moment in time. The transcriptome of a cell is dynamic it continually changes. The recent developments of next-generation sequencing (NGS) allow for increased base coverage of a DNA sequence, as well as higher sample throughput. This facilitates sequencing of the RNA transcripts in a cell, providing the ability to look at alternative gene spliced transcripts, post-transcriptional modifications, gene fusion, mutations/SNPs and changes in gene expression[7]. In addition to mRNA transcripts, RNA-Seq can look at different populations of RNA to include total RNA, small RNA, such as miRNA, tRNA, and ribosomal profiling. RNA-Seq can also be used to determine exon/intron boundaries and verify or amend previously annotated 5' and 3' gene boundaries. RNA-Seq research includes observing cellular pathway alterations during infection, and gene expression level changes in cancer studies. Before NGS, transcriptomics and gene expression studies were done with expression microarrays, which contain thousands of DNA sequences (probes) that potentially match complementary sequences in the sample, making available a profile of all transcripts being expressed. This was later done with serial analysis of gene expression (SAGE).

Genotyping-by-sequencing (GBS)

Genotyping-by-sequencing (GBS) is highly multiplexed genotyping system involving DNA digestion with different enzymes and the construction of a reduced representation library, which is sequenced using an NGS platform. It enables the detection of thousands of SNPs in large populations or collections of lines that can be used for mapping, genetic diversity analysis, and evolutionary studies[8]. The rapid development of next-generation sequencing techenique has enabled the use of sequencing for routine genotyping across a range of genetics studies and breeding applications.

Mining for Genes of Agronomic Importance

Knowledge of the gene underlying a trait enables the transfer of the trait between cultivars and even species using genetic modification. Alternatively, the gene conferring the favourable trait may be incorporated into a cultivar by marker-assisted selection (MAS) breeding. While many of the simple traits have been well characterized at the genome level, there are many other traits which are poorly understood. This is particularly true for complex traits which are controlled by interacting gene networks. Although many aspects of a complex trait, such as yield, may be characterized individually, it is unlikely that the genetic basis underlying all components of yield heritability will be understood in the near future. The identification of all the genes for a crop is only one step towards understanding the inheritance of agronomic traits. The functions of many of the genes identified by genome sequencing remain unknown and the genetic control of the majority of agronomic traits has yet to be determined. Producing a finished genome sequence for crop is an important first step and is becoming feasible for an increasing number of crop species (Imelfort and Edwards,in press).

Bioinformatics

NGS technologies brought new challenges, as millions of short DNA reads have to be analysed and assembled[3]. Also, genetic maps, genotypes, or expression information at a genomic scale have to be processed so as to obtain the relevant biological information. Therefore, it is necessary to develop new bioinformatics tools (algorithms and software), which allow the analyses of huge amounts of genome-wide data. The field of sequence analysis has a long tradition and has enabled the assembly of many genome sequences

obtained by Sanger sequencing. The huge amount of sequence reads generated by NGS and the low quality of individual reads requires new software tools and algorithms that allow dealing with these data in an efficient way. Two of the most common analyses carried out on these NGS reads are genome assembly and annotation and mapping.

Molecular marker discovery

Single nucleotide polymorphisms now dominate molecular marker applications, because of recent advances in DNA sequence technology enabling their discovery, and the development of high throughput assays. As with most molecular markers, the factor limiting the implementation of SNP is the initial cost of their development[9]. SNP discovery involves finding differences between two sequences. Traditionally this has been performed through PCR amplification of genes ∕ genomic regions of interest from multiple individuals selected to represent diversity in the species or population of interest, followed by either direct sequencing of these amplicons or the more expensive method of cloning and sequencing. Sequences are then aligned and polymorphisms is identified. However, large quantities of sequence data are being generated by the latest second generation sequencing technologies and these provide a valuable resource for the mining of molecular markers[10]. While the large volume of next generation sequencing data is generally produced at the expense of sequence quality, the oversampling of genome data enables the differentiation between true SNP and sequence error. Whole genome sequencing is the most robust method to identify the great variety of genetic diversity in a population and gain a greater understanding of the relationship between the inherited genome and observed heritable traits. The continued rapid advances in genome sequencing technology will likely lead to whole genome sequencing becoming standard method for genetic polymorphism discover.

DNA Sequencing and Crop Protection

The rapid, accurate and reliable plant pathogen/pest detection, identification and quantification is one of the crucial step in crop protection. It allows to control the spread of the diseases/pests by screening vegetal propagative material and to implement quarantine regulations. The pathogen/pest detection and identification are fundamental for epidemiological studies and for the design of new control strategies. Traditionally, the most used approach to identify plant pathogens relied with visual inspection of symptoms usually followed by laboratory analyses based on morphological identification using microscopy and isolation and culturing of the organisms. In some cases, these methods are still used, but actually these conventional methods require skilled and specialized expertise which often takes many years to acquire, are laborious, time consuming and not always sensitive and specific enough. Moreover, closely related organisms may be difficult to discriminate on morphological characters alone, symptoms are not always specific and not all the microorganisms are culturable in vitro. So efforts have been devoted to the development of novel methods, mainly nucleic acid based molecular approaches, for detecting and identifying plant pathogens and pests[11]. The use of DNA-based methods derives from the premise that each species of pathogen carries unique DNA or RNA signature that differentiates it from other organisms. Knowing the pathogen/pest nucleic acid sequence enables scientists to construct oligos to detect them. The improvement of the sequencing technology and in particular the advent of the NGS together with the evolution of bioinformatics enable large-scale sequencing projects permitting a new comparative genomics approach and thus it will facilitate the identification of candidate genes to be used in crop protection. The possibility to down-regulate or over express genes in plants or in plant pathogenic organisms is a very powerful tool, within pest management and agriculture in general, to meet farmer and consumer demands.

Mutant and Germplasm Collections in the Genomics Era: TILLING and EcoTILLING:

In order to facilitate the identification of mutation and hence the accessions of interest in these collections, a genetic reverse approach has been used. Targeting Induced Local Lesions in Genomes (TILLING)[12] is able to identify all allelic variants of a DNA region present in an artificial mutant collection. A similar procedure called Ecotype TILLING (EcoTILLING) can be used to identify allelic variants for targeting genes in natural collections. These two methods are based on the use of endonucleases, such as CEL I or Endo I, that recognize and cut mismatches in the double helix of DNA[12]. Since the TILLING and EcoTILLING techniques identify all allelic variants for a certain genomic region, the phenotypic characterization effort can be concentrated in a reduced number of accessions with different variants. The success of the identification of variation useful for breeding programmes will depend on the right selection of target genes. The availability of sequences coming from NGS sequencing projects and the information provided by gene expression studies is significantly increasing the number and quality of candidates for TILLING and EcoTILLING studies.

These procedures have been successfully used in many crops. For example, TILLING has been applied to *Arabidopsis,* Lotus, barley, maize, pea, and melon. Rice was the first crop for which EcoTILLING was applied. Subsequently, EcoTILLING has been used in other crops and wild relatives, like barley, wheat, or the wild peanut *Arachis duranensis*.

Re-Sequencing for SNPs Discovery and Use in Genotyping Platforms

One of the most interesting applications of NGS for plant breeders is the discovery of genetic variation. Now it is possible to sequence rapidly multiple individuals within a species with limited technical expertise and at minimal cost. The parallel development of computational pipeline tools is greatly accelerating the accurate mining of these sequences for genetic variants that can be converted into genetic markers, mainly microsatellites or simple sequence repeats (SSRs) and SNPs[13]. SSRs and SNPs are now the predominant markers in plant genetic analysis. SNPs are more abundant, stable, amenable to automation, and increasingly cost-effective, thus are fast becoming the marker system of choice in modern genomics research[14].

The genome-wide SNPs discovery by massive re-sequencing has been performed in model species with small genomes, such as *Arabidopsis thaliana*, where the 1001 Genomes project (http://www.1001genomes.org)[15] is attempting to unveil the whole-genome sequence variation in this reference plant. Similar re-sequencing efforts are becoming possible in rice, maize, grape, soybean, poplar etc. by sequencing sets of related genotypes, individually or pooled, within each species (elite cultivars, breeding lines, ecotypes, landraces, and/or weedy and wild relatives of a crop). The higher complexity in architecture and repeat content of these genomes has made necessary the use of approaches for genomic complexity reduction that also reduce sequencing cost. Identification of SNPs is also very challenging in species with high levels of heterozygosity and with complex ploidy levels. Both Roche 454 and Illumina GA have been mostly used for genome re-sequencing.

Construction of High Density Genetic Maps

One of the main applications of genomic advances is the development of high density genetic maps. The high-density map construction involves the location of hundreds or even thousand markers in the different linkage groups. In these maps the coverage should be very high and no large gaps must be present. The newly developed maps, enriched in sequence-based markers are facilitating comparative mapping. Recent examples are

high density SNPs maps of barley compared with wheat and rice[16].

Genome-Wide Genetic Diversity Studies

One of the main challenges in agricultural genetics is to access and use the tremendous genetic variation present in germplasm collections and in the wild relatives. A significant part of this variation remains untapped because of the difficulties in effectively identifying genetic differences in large collections. Some traits, with high heritability and of simple characterization, are easy to select for. However, desirable allelic variants and genetic combinations for complex traits are difficult to identify. Recent advances in genotyping are enabling genome-wide diversity studies capable of better capturing the spectrum of variability in natural and breeding populations. This study allowed the discovery of regions with highly suppressed recombination that appear to have influenced the effectiveness of selection during maize inbred development and may be a major component of heterosis. Also, highly differentiated regions were found that probably contained loci that are key to geographic adaptation.

Genome-wide survey of genetic diversity is useful to elucidate the causative genetic differences that give rise to observed phenotypic variation providing a foundation for dissecting complex traits through genome-wide association studies. However, its utility is limited if phenotypic data are not available. Not just genomics and transcriptomics, but the other 'omics' disciplines, like proteomics and metabolomics, are useful to understand how the changes in the genotype lead to differences in the final phenotype. Phenomics, which uses high-throughput technologies to characterize germplasm, is being developed and will help to deal with this issue[17].

Identification of Molecular Markers Linked to Single Genes and QTL

NGS and high-resolution maps have led to a significant improvement in the identification of molecular markers linked to specific genes and to QTLs. The most important advantage comes from the dense genome coverage, which allows the identification of markers closely linked to any target genomic region, with the advantages that this tight linkage provides. For example, a fine genetic mapping of the single dominant scab resistance gene (*Ccu*) in RILs of cucumber (*Cucumis sativus*) has been conducted[18]. The resistant cucumber genome was sequenced with Solexa/Illumina NGS and compared with the susceptible cucumber genome. As a result, three insertion/deletion (indel) markers closely linked to the *Ccu locus* where obtained. A detailed study of the region delimited by markers revealed four resistance gene analogs as possible candidates for *Ccu*. QTL detection has traditionally been conducted by linkage mapping. NGS technologies are significantly contributing to increase accuracy in detection of QTLs. They allow increases in many orders of magnitude of the number of markers mapped, ensuring high mapping resolution, and also aid in the development of mapping populations, such as RILs, near isogenic lines (NILs), and more appropriated for QTLs detection.

Marker Assisted Backcross Selection

Marker assisted selection (MAS) is an indirect process where selection is carried out on the basis of a marker instead of the trait itself. The successful application of MAS relies on the tight association between the marker and the major gene or QTL responsible for the trait. MAS, which includes marker-assisted back-crossing (MABC) uses molecular markers that map within specific genes or QTLs known to be associated with target traits or phenotypes to select individuals that carry favourable alleles for traits of interest (and/or to discard those that do not)[8]. The basis of a marker-assisted backcrossing (MAB) strategy is to transfer a specific allele at the target locus from a donor line to a recipient line while selecting against donor introgressions across the rest of the genome. Since development of markers is one of the important application of sequencing, number of Simple sequence repeats SSR’s have been developed using sequencing information of rice. One of the case study by Neeraja *et al*[19] who developed

Copyright © Nov.-Dec., 2017; IJPAB

the submergence tolerance variety Swarna sub1 through a MAB approach. A major QTL (sub 1) was identified and fine mapped on chromosome 9, in the submergence tolerant cultivar FR13A, and through MAB the major QTL was transferred at the target locus from a donor line to a recipient line. The *Sub1* locus was monitored by markers shown to be closely linked with the gene[20]. Using tightly linked (RM464A) and flanking (RM219, RM316) markers, as suggested by Hospital and Charcosset [21], ensured efficient foreground and recombinant selection.

Table 1: Progress in crop genome sequencing

Plant Genomes Project

Species name	Genome size (Mb)	Ploidy	Chr. no.	Gene no.	Transposable elements (%)	Sequence strategy	Publication date
Arabidopsis thaliana	125	2n=10	5	25498	7.8	BAC physical map.	2000
Oryza sativa (rice)	389	2n=2x=24	12	40000	35	BAC to BAC and Sanger sequencing	Aug 2005
Sorghum bicolour (sorghum)	700	2n=2x=20	10	2800	63	WGS, Sanger sequencing	Jan 2009
Zea mays	2300	2n=2x=20	10	33000	85	BAC physical map.	Nov 2009
Glycine max (soybean)	1,115	2n=2x=40	20	42,094	59	WGS, Sanger sequencing	Jan 2010
Malus domestica (apple)	750	2n=2x=34	17	57386	42.4	WGS, Sanger, Roche 454	Oct 2010
Theobroma cacao	430	2n=2x=20	10	29000	24	WGS, Sanger,	Dec 2010
Brassica rapa (cabbage)	485	2n=2x=20	10	41018	40	WGS, Sanger sequencing	Aug 2011
Cajanus cajan (pigeonpea)	833	2n=2x=22	11	47,004	52	WGS	Jan 2012
Triticum aestivum (bread wheat)	1700	2n= 6x = 42	7	108000	80	WGS, Roche 454	Nov 2012
Hordeum vulgare (barley)	5100	2n = 2x = 14	7	24287	50	WGS, BAC physical map.	Nov 2012
Cicer arietinum (chickpea)	738	2n = 2x =16	8	27571	45.64	WGS	Jan 2013
Saccharum offi cinarum (sugar cane)	>15000	2n = 80	10		54	WGS	In progress

The 1000 Plant Genomes Project

A new initiative launched in November 2008 will acquire gene sequence information for 1000 plant species. Our mandate includes everything from algae to land or aquatic plants, with a particular focus on plants that make valuable bioproducts. The project is led from Alberta by Gane Ka-Shu Wong and Michael Deyholos, and the sequencing will be done at BGI-Shenzhen. An international multidisciplinary consortium has been formed to participate in this research. All of our sequence data will be released to the public upon publication, specifically through GenBank and other open access websites. This project will begin what we hope is a longer term effort by the research community to study the vast biodiversity that to date has barely been touched by genomics. Not only will this lead to great science, but also, we believe it will lead to commercialization opportunities.

1001 Arabidopsis Genome Project

The 1001 Genomes Project, launched at the beginning of 2008, has as a goal to discover the whole-genome sequence variation in 1001 strains (accessions) of the reference plant Arabidopsis thaliana. The resulting information is paving the way for a new era of genetics that identifies alleles underpinning phenotypic diversity across the entire genome and the entire species. Each of the accessions in the 1001 Genomes project is an inbred line with seeds that are freely available from the stock centre to all our colleagues.

1,000 Plant and Animal Reference Genomes Project

In January 2010, BGI launched the 1,000 Plant & Animal reference genomes project and called for collaboration from around the world. The goal of the project is to generate reference genomes for 1,000 economically and scientifically important plant/animal species. Together with our collaborators, so far we

have initiated 505 plant and animal genome projects, completed genome maps for over 100 species and finished the sequencing of about 200 species. The completed projects include rice, silkworm, cucumber, panda, camel, oyster, ant, grouper, goose, crested ibis, potato genomes, and more. Many other genomes are in process of active sequencing. We look for more collaboration to complete the tree of reference genomes. Table 1 listed the Progress in crop genome sequencing**.**

CONCLUSION

The genome sequences of organisms are fundamentally important for understanding the functions of individual genes and defining evolutionary relationships. The identification of genes and molecular markers underlying agronomic traits will help to accelerate the breeding process and lead to improved varieties with improved yield and quality, tolerance to unfavourable environmental conditions and resistance to diseases. DNA sequencing is a functional assay, and as it gets faster and cheaper, there will be more and more applications and uses for it in plant breeding Next-generation sequencing has revolutionized our ability to study the variations occurring in whole genomes of organisms in a very short period of time at far lesser costs. Sequencing of crops provides valuable information on genome structure and organization. It opens up an excess of opportunities for research related to the life-sciences including evolutionary biology, developmental biology, biochemistry, genetics and molecular biology. In recent years, agricultural sciences have been in the middle of a second technological revolution in DNA sequencing.

REFERENCES

1. Fedoroff, N.V., The past, present and future of crop genetic modification. *New Biotechnology.* **27:** 461–465 (2010).
2. Perez-de-Castro, A.M., Vilanova, S., Canizares, J., Pascual, L., Blanca, J.M., Díez, M.J., Prohens, J., and Picó, B., Application of Genomic Tools in Plant Breeding. *current genomics.* **13(3):** 179–195 (2012).
3. Metzker, M., Sequencing technologies-the next generation. *Nature Reviews Genetics*, **11:** 31 (2010).
4. Varshney., R.K. and Tuberosa, R., Genomics-Assisted Crop Improvement. *Genomics Approaches and Platforms*. **1:** 1–12 (2007).
5. Blanca, J., Canizares, J., Roig, C., Ziarsolo, P., Nuez, F, and Pico, B. Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae). BMC Genomics. **12:** 104 (2011).
6. Logacheva, M.D., Kasianov, A.S., Vinogradov, D.V., Samigullin, T.H., Gelfand, M.S., Makeev, V.J. and Penin, A.A., *De novo* sequencing and characterization of floral transcriptome in two species of buckwheat (*Fagopyrum*). BMC Genomics, **12:** 30 (2011).
7. Maher, C.A., Kumar-S.C., Cao, X., Kalyana, S., Han, B., Jing, X., Sam, L., Barrette, T., Palanisamy, N. and Chinnaiyan, A.M., Transcriptome sequencing to detect gene fusions in cancer. Nature, **5(7234):** 97-101 (2009).
8. Varshney, R.K., Terauchi, R. and McCouch, S.R., Harvesting the Promising Fruits of Genomics: Applying Genome Sequencing Technologies to Crop Breeding. *PLOS Biology*, **12(6):** (2014).
9. Duran, C., Appleby, N., Edwards, D. and Batley, J., Molecular genetic markers: discovery, applications, data storage and visualisation. *Current Bioinformatics*, **4:** 16–27 (2009).
10. Imelfort, M. and Edwards, D., Next generation sequencing of plant genomes. *Brief. Bioinformatics*, in press.
11. Lievens, B. and Thomma, B.P.H.J., Recent developments in pathogen detection arrays: Implications for fungal plant pathogens and use in practice. *Phytopathology.* **95(12):** 1374-1380 (2005).
12. Till, B.J., Reynolds, S.H., Greene, E.A., Codomo, C.A., Enns, L.C., Johnson, J.E., Burtner, C., Odden, A.R., Young, K., Taylor, N.E., Henikoff, J.G., Comai, L.

and Till, B.J., Burtner, C., Comai, L. and Henikoff, S., Mismatch cleavage by single-strand specific nucleases. *Nucleic Acids Research*, **32:** 2632–2641 (2004).

13. Deschamps, S. and Campbell, M.A., Utilization of next-generation sequencing platforms in plant genomics and genetic variant discovery. *Molecular Breeding*, **25:** 553–570 (2010).
14. Ganal, M.W., Altmann, T. and Roder, M.S., SNP identification in crop plants. *Current Opinion in Plant Biology*, **12:** 211–217 (2009).
15. Weigel, D. and Mott, R., The 1001 Genomes Project for *Arabidopsis thaliana. Genome Biology*, **10:** 107 (2009).
16. Sato, K., Nankaku, N. and Takeda, K., A high-density transcript linkage map of barley derived from a single population. *Heredity*, **103:** 110–117 (2009).
17. Finkel, E.I., With 'phenomics,' plant scientists hope to shift breeding into overdrive. *Science*. **325:** 380–381 (2009).
18. Kang, H.X., Weng, Y.Q., Yang, Y.H., Zhang, Z.H., Zhang, S.P., Mao, Z.C., Cheng, G.H. and Huang, S.W., Fine genetic mapping localizes cucumber scab resistance gene *Ccu* into an R gene cluster. *Theoretical and Applied Genetics* **122:** 795–803 (2011).
19. Neeraja, C.N., Maghirang, R.R., Pamplona, A., Heuer, S., Collard, B.C.Y., Septiningsih, E.M., Vergara, G., Sanchez, D., Xu , K., Ismail, A.M. and Mackill, D. J., A marker-assisted backcross approach for developing submergence-tolerant rice cultivars. *theoretical applied genetics*, **115:** 767–776 (2007).
20. Xu, K.N., Deb, R. and Mackill, D.J., A microsatellite marker and a codominant PCR-based marker for marker-assisted selection of submergence tolerance in rice. *Crop Science*, **44:** 248–253 (2004).
21. Hospital, F. and Charcosset, A., Marker-assisted introgression of quantitative trait loci. *Genetics*, **147:** 1469–1485 (1997).