Journal of Pharmacognosy and Phytochemistry

Available online at www.phytojournal.com

E-ISSN: 2278-4136
P-ISSN: 2349-8234
JPP 2020; 9(1): 547-552
Received: 22-11-2019
Accepted: 24-12-2019

Rahul Kumar
Ph.D. Scholar, Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi, India

J Jorben
Ph.D. Scholar, Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi, India

Raju Ratan Yadav
Senior Research Fellows, Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi, India

Manoranjan Senapati
Ph.D. Scholar, Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi, India

Corresponding Author:
Rahul Kumar
Ph.D. Scholar, Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi, India

Genomics and its application in crop improvement

Rahul Kumar, J Jorben, Raju Ratan Yadav and Manoranjan Senapati

DOI: https://doi.org/10.22271/phyto.2020.v9.i1i.10488

Abstract
Using traditional methods and methodologies, plant breeding has been very effective in producing improved varieties. Nowadays, the availability of genomic tools and resources leads to a new plant breeding revolution as they promote the study of the genotype and its relationship with the phenotype, particularly for complex traits. Genomics (coined in 1986 by Tom Roderick) is an interdisciplinary research field that focuses on the study of any organism's genome. Genomics field revolves around gene analysis and gene working. It is generally classified as structural genomics, functional genomics, comparative genomics, epigenomics, translational genomics and pharmacogenomics. Functional genomics links the organism's genome to its function and phenotype. It could be further subdivided into proteomics, transcriptomics and metabolomics and could be studied by reverse genomic techniques like RNAi and mutagenesis. Genomics or decoding the plant genome sequence using high throughput techniques will allow scientific community to access agronomically important genes and will speed up the breeding programs for the development of superior varieties with higher yield and stress tolerance.

Keywords: Gene, genomics, plant breeding, next generation sequencing

Introduction
Since the beginning of plant domestication, some 10,000 years ago, plant breeding has been extremely successful in the development of crops and varieties that have contributed to the development of modern societies and successively beat (neo-) Malthusian predictions. The application of traditional scientific breeding methodologies for pre-genomics has led to the development of modern cultivars, which since the middle of the 20th century have contributed to the dramatic improvement of the yield of most major crops. In the last century, the success of plant breeding was based on the use of natural and mutant-induced genetic variation and efficient selection of suitable genetic combinations using effective breeding methods. In this regard, the phenotypic assessment was largely based on the examination and detection of genetic variants of interest as well as the selection methodologies used. Nowadays, by enabling the direct analysis of the genotype and its relationship with the phenotype, genomics provides breeders with a new set of tools and techniques that allow the study of the entire genome and reflect a paradigm shift. At the beginning of the 20th century, though classical genetics revolutionized plant breeding, genomics led to a new revolution in plants breeding in the early 21st century. The genomics area and its application to the breeding of plants are developing rapidly. The combination of conventional breeding techniques and genomic tools and approaches leads to a new plant breeding based on genomics. In this new context of plant breeding, genomics will be important for the production of more productive plant cultivars, which, according to the FAO, are crucial for the current' green revolution' needed to feed the rising population of the world while maintaining natural resources.
Thomas Roderick used the word genomics for the first time in 1986. It concerns the analysis of the structure and function of a living organism's entire genome. Genomics is now being developed as a genetic sub-discipline that is dedicated to genome mapping, sequencing, and functional analysis. A genome is described as the whole genetic make-up and it includes all of an individual organism's hereditary information known as genes. It is known that some genes are pleiotropic. Dissection of gene activity pathways is one of many research programs main goals. Plant design aims at enabling them to withstand biotic and abiotic stress, to withstand pathogens attacks, to improve the yields and nutritional values. This is where genomics comes into picture to provide molecular insights in order to understand and improve these characteristics. Plant genomics has vast potential in crop breeding since it offers very valuable information that can be used to enhance the varieties that we currently rely on in our breeding programs (Siva, Ramesh, Ibrokhim, & Jafar, 2012).

Evolution of Genomics

Below are the following major events that have influenced the history of genomics.

1953: James Watson and Francis Crick discover the double helix structure of DNA with contributions from Rosalind Franklin and Maurice Wilkins.

1977: Frederick Sanger introduces a technique in DNA sequencing that he and his team use to sequence the first complete genome–that of the phiX174 virus.

1990: Launch of the Human Genome Project. The project is aimed at sequencing all 3 billion human genome letters in 15 years.

1995: The first genome sequence of bacteria (*Haemophilus influenza*) has been completed.

1996: An international team completes the sequencing of the yeast (*Saccharomyces cerevisiae*) genome. The first cloned cow, Dolly the Sheep, was raised at the University of Edinburgh's Roslin Institute.

1998: John Sulston and Bob Waterston published the nematode worm genome, C. eleganzs.

1999: Chromosome 22 is the first human chromosome in the Human Genome Project to be sequenced.

2000: Full genome sequence of the *Drosophila melanogaster* model organism (fruit fly).

2001: First draft of the published sequence of human genomes.

2002: The mouse is the first animal to complete its entire genome sequence. The research is being undertaken by the International Consortium of Mouse Genome Sequencing. The genome of the mouse is 14% smaller than the human genome, but more than 95% of the genome of the mouse is identical to ours. The International Hap Map Project is launched with the objective of producing a' catalogue' of common genetic variations in humans and where they are found in the genome.

2003: The Human Genome Project has been completed and confirms that some 20,000–25,000 genes are available to humans. The human genome is sequenced, 2 years ahead of schedule, to 99.99 percent accuracy.

2005: Report published in Nature on Hap Map (Map of Human Genetic Variation). The genome of chimpanzees is full.

2008: 1,000 Genomes Project began–the first project aimed at sequencing a large number of people's whole genomes (2,500).

2012: The ENCODE project publishes 30 research papers identifying the active human genome regions, including confirmation that there are 20,687 protein-coding genes in the human genome.

2018: On 17 August 2018, the IWGSC published a detailed description and review of the bread wheat genome reference sequence in the international journal Science, the world's most widely cultivated crop.

Genomics in Crop plants

Work on plant genomics began in December 2000 with the publication of the entire genome sequence of the Arabidopsis thaliana model plant species. This has resulted in a significant improvement in our understanding of the molecular basis of both plant growth and environmental stimulus response. Having the genome sequence enabled genomic approaches aimed at assigning a function to each of the 26,000 genes that were predicted. Knockout mutation generated by RNAi technique insertion mutagenesis or gene silencing indicates a role for a gene if it is possible to link the mutants to phenotype. Significant developments include the August 2005 release of a high-quality rice genome sequence, September 2006 draft poplar genome, 2007 full genome sequence of two grapevine genotypes, 2008 transgenic papaya, and 2018 full wheat genome.

Classification of Genomics

The genomics discipline consists of three sections that are Structural Genomics, Comparative Genomics and Functional Genomics. These are described as the following:

Structural Genomics

It deals with the study of the structure of entire genome of a living organism. In other words, it deals with the study of the genetic structure of each chromosome of the genome. This specifies the size of a species ' genome in [Mb] mega-bases as well as the genes found in a species ' entire genome. It focuses on building genomic sequence data, finding and locating gene and building gene maps. It uses DNA Sequencing technology and program software to produce, store and analyze information about the genomic sequence.

Functional Genomics

It is involved in the research of the function of all genes found in the entire genome. This deals with the proteome and transcriptome. The transcriptome refers to a complete set of genome-transcribed RNAs and refers to a complete set of genome-encoded proteins. Functions must be allocated to all genes in the series after genome sequencing is annotated. Some of the genes may already allocate functions using a conventional method Mutagenesis and linkage mapping. Some may not have assigned functions-use homology searches. Functional approaches to genomics predominantly use methodologies based on sequence or hybridization. Expressed Sequence Tags (ESTs), Serial Analysis of Gene Expression (SAGE) and Massively Parallel Signature Sequencing (MPSS) are analyzed sequence-based approaches. Hybridization-based approaches are array-based techniques that use target DNA hybridization with cDNA or oligonucleotide samples attached to a surface for expression assessment. These array-based approaches are targeted; that is prior transcript information to be studied, either sequence or clones, is a prerequisite for designing samples. Targeting Induced Local Lesions In Genomes (TILLING) primarily allows high-throughput analysis of large numbers of mutants, a modified technique, called Eco TILLING, has been developed to detect natural polymorphisms, similar to TILLING-assisted induced mutation detection. Editing of T-DNA, RNAi and Genome editing also used to assign Gene function.

Comparative genomics

It is a biological research field in which the genome sequences of different human species, mouse and a wide variety of other organisms are compared from bacteria to chimpanzees. Through comparing the genomic sequences of different organisms, researchers can explain what distinguishes different forms of life from each other at the molecular level. Comparative genomics also provides a powerful tool for researching evolutionary changes among organisms, helping to recognize retained or common genes among species, as well as genes that give unique characteristics to each organism. This compares the sequences of genes to explain the relationship is functional or evolutionary. Comparative genomics is a promising method for species with largely unexplored genomes to gain

information by using conservation among closely related plant species. Comparative genomics has made a significant contribution to the development of the idea of "genome zipper," which makes it possible to establish a virtual gene order in a partially sequenced genome. Genome zippers associate completely sequenced and annotated rice and sorghum genomes with different data sources from less well-studied species, including genomic sample sequences and genetically mapped markers. In these organisms, predict the gene order and organization. Nevertheless, depending on synteny means that this strategy cannot investigate newly evolved genes and small-scale rearrangements.

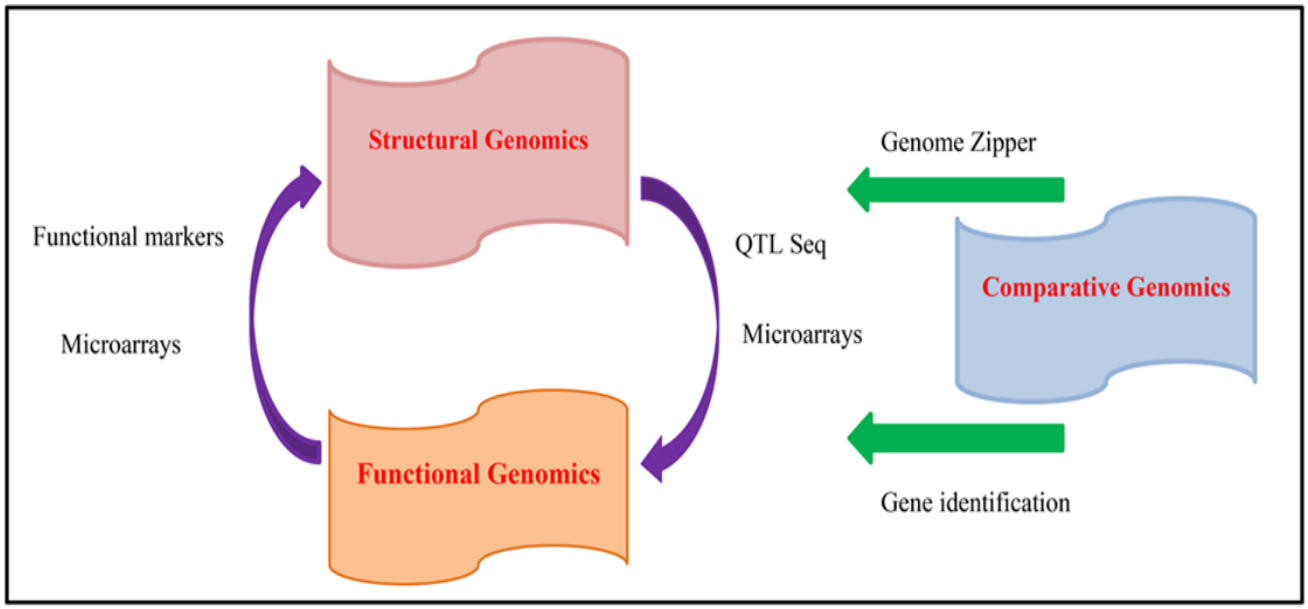

Fig. Relation between structural, functional and comparative genomics

Role of Genomics in Crop Improvement

Genomics has a range of practical uses to boost crops. In a number of ways, genome mapping is useful. It is useful or provides information on genome size, gene number, gene mapping, gene sequencing, crop plant evolution, gene cloning, DNA marker recognition, marker assisted selection, transgenic breeding, linkage map creation and QTL mapping.

i. Genome size: Genome mapping is a very useful technique for assessing the size of the genome in different species of plants. The largest genome size was recorded in Wheat (14.5 Gb) and the smallest in Arabidopsis thaliana (120 Mb) in the plant species studied so far.

ii. Gene Number: Genome mapping provides information on a species ' gene number. The maximum number of genes reported in Wheat that is 107,891 (Table 1) was reported in crop plants studied so far.

iii. Gene mapping: Genome research is very useful in mapping / tagging genes on a genome's different chromosomes. In other words, it helps to discover new genes in a genome on a large scale.

iv. Gene Sequencing: Genome mapping helps to establish the chromosome gene order. The gene order is determined on each genome chromosome.

v. Evolution: Genome mapping provides information on various organisms ' evolution. This tests the interaction between different genomes and thus provides information on crop plants ' association or evolutionary biology.

vi. Gene Cloning: Genome research is very useful in making multiple gene copies and moving the same from genotype to genotype. It therefore assists in the precise transfer of genes.

vii. DNA marker identification: genome mapping techniques are useful to classify DNA markers that can be used in molecular breeding, i.e. marker assisted selection. From the mapping groups, For DNA markers, inter-specific crosses are extremely polymorphic than those representing populations resulting from intra-specific crosses.

viii. Marker Assisted Selection: Marker Assisted Selection refers to indirect selection based on the band pattern of related DNA markers for a specific phenotype. Using such selection, crop improvement is called molecular breeding. Similar DNA markers used for this purpose are RFLP, AFLP, SSR, etc. DNA marker results are associated with morphological markers and then it's picked for a particular characteristic. Selection based on DNA markers is more effective because environmental factors do not affect DNA markers.

ix. Transgenic breeding: In gene cloning, genome mapping is important. The gene of interest can be cloned and used in the development of transgenic plants. Transgenic breeding enables the direct transfer of genes bypassing the sexual cycle.

x. Linkage maps construction: Genome mapping helps to create linkage classes. Based on gene mapping and gene sequencing details, the linkage groups can be constructed.

xi. QTL Mapping: The techniques of genome mapping are commonly used for quantitative trait loci (QTL) mapping. Through conventional methods, i.e. recombination mapping and deletion mapping techniques, mapping of QTL or polygenic traits is not feasible.

Table 1: Genome size, Gene No. and Variety used for sequencing of major crops

	Botanical name	Genome size (Mb)	Gene No.	var. sequenced
Field crops				
Arabidopsis	*Arabidopsis thaliana*	125	25,498	Columbia
Rice	*Oryza sativa*	389	37,544	Indica'93-11' , Japonica 'Nipponbare'
Corn	*Zea mays*	2,300	32,000	B73
Wheat	*Triticum aestivum*	14.5 Gb	107,891	Chinese spring
Sorghum	*sorghum bicolor*	730	34,496	Moench'BT×623'
Barley	*Hordeum vulgare*	5100	26,159	Morex
Pigeonpea	*Cajanus cajan*	833	48,680	Asha
Soybean	*Glycine max*	1115	46,430	Williams 82
Tomato	*Solanum lycopersicum*	900	34,727	Heinz 1706
Fruit crops				
Apple	*Malus domestica*	742	57,386	Golden Delicious
Papaya	*Carica papaya*	372	28,629	Sun Up
Grapes	*Vitis vinifera*	500	26,346	PN40024
Banana	*Musa acuminata*	523	36,542	DH-Pahang

Genomic Tools and Resources for Plant Breeding

Genome and Transcriptome Sequencing

The availability of a crop's entire genome sequence is very useful for plant breeding, given the high cost of sequencing the entire genome; transcriptome sequencing was a cheaper alternative. The cDNA sequences (expressed sequence tags, ESTs) provide relevant information on the genes expressed in a particular tissue or organ, at a particular stage of development and under specific environmental conditions. With the advent of NGS technologies, the genomics landscape has shifted. Such new technologies have lowered sequencing costs by more than 1,000 times compared to Sanger technology, reducing time consuming and repetitive Modern steps in cloning and allowing millions of simultaneous sequencing reactions. The platforms 454 (Roche, http:/www.454.com) and Illumina (Illumina, http:/www.illumina.com) are already widely used to sequence crop species among the "second generation" technologies. Some, such as Solid (Applied Biosystems, http:/www. Applied biosystems. com / technologies), were less exploited in plants. Moreover, new, "third generation" platforms are being developed and incorporated to sequencing projects, such as PacBio RS (Pacific Biosciences), Helico (Helicos, http:/www.helicos bio.com), or Ion Torrent (Life Technologies, http:/www.iontorrent. com). NGS typically deposits the sequence collected in the NCBI Sequence Read Archive (http:/www.ncbi.nlm.nih.gov /unigene).

Bioinformatics

The sequence analysis field has allowed the assembly of various sequences of genomes obtained through Sanger sequencing. The assembly of genome is a complex task that requires powerful computers and professional bio informaticians. Roche's 454 assembler, Celera Assembler, and Mira are some of the most widely used assemblers. Once there is a reference genome in the population, its variability is widely studied. Instead of assembler software, mapper software is widely used for this. A mapper attempts to match each read with the reference genome. This method is much easier than assembly and much quicker. The code specifications are typically less rigorous in this case and the limiting factor could be the capacity for storage. Bowtie, BWA, and Top Hat are some widely used mappers. Once the reads are matched, the use of the SAM tools or the GigaBayes SNP callers will detect single nucleotide polymorphism (SNPs).

Expression Studies, from Microarrays to RNA-Seq

The expansion and acceleration of gene expression studies are also of interest to new genomic tools. Studies of gene expression were initially based on the classical Northern blot method which only permitted the simultaneous quantification of tens of genes. The qRT PCR is a more efficient and quantitative method, but there is also a limited number of genes examined through the procedure. All Differential display and cDNA amplified fragment length polymorphisms (cDNA-AFLPs) were approaches that enabled the study of thousands of genes. Such techniques, however, are not really quantitative and are constrained by the ability of the existing libraries to collect transcripts of low abundance. The serial study of gene expression (SAGE) and massively parallel signature sequencing (MPSS) are other approaches that solve some of these problems. Nevertheless, hybridization-based systems or microarrays are currently the most used tools for analyzing transcript profiling. There are several advantages of expression arrays when compared to other studies. They can measure tens of thousands of different transcripts in the same reaction; if they are expressed in a given number, they are semi-quantitative and responsive to low-abundance transcripts.

Mutant and Germplasm Applications in the Age of Genomics: Tilling and Eco Tilling

A reverse genetic approach has been used to promote the detection of interest accessions in these applications. Targeting Induced Local Genome Lesions (TILLING) is capable of identifying all the allelic variants of a DNA region found in a series of artificial mutants. A similar procedure called ecotype TILLING (Eco TILLING) can be used to identify allelic variants for targeting genes in natural

collections. Such two approaches are based on the use of endonucleases in the double helix of DNA, such as CEL I or Endo I, which identify and break mismatches. Since the techniques of TILLING and Eco TILLING classify all the allelic variants for a certain genomic region, the effort of phenotypic characterization can be concentrated in a small number of accessions with different variants. The availability of sequences from NGS sequencing projects and knowledge from gene expression studies greatly increases the number and efficiency of TILLING and Eco TILLING studies candidates. Arabidopsis, lotus, wheat, corn, pea and melon were used for TILLING. Rice was the first crop to be used for Eco TILLING. Subsequently, using both gene bank collections and natural populations, Eco TILLING was used in other crops and wild relatives, such as barley. Many studies focused on the detection of allelic variants in genes most closely related to organoleptic quality or resistance to disease.

Re-Sequencing for SNPs Discovery and Use in Genotyping Platforms

The detection of genome-wide SNPs by large resequencing has been carried out in model organisms with small genomes, such as Arabidopsis thaliana, where the project 1001 Genomes (http:/www.1001genomes.org) aims to reveal the variability of the entire genome sequence in this reference plant. Different re-sequencing efforts are made possible by sequencing sets of similar genotypes, individually or pooled, within each species (elite cultivars, breeding lines, ecotypes, landraces, and/or weedy and wild relatives of a crop) in rice, maize, grape, soybean, poplar, etc. In genome re-sequencing, both Roche 454 and Illumina GA were mostly used. Because most of these genomes, small SSR and SNP samples, were available from early resequencing prior to the advent of NGS, but recent genome-wide re-sequencing extends the SNP pools and renders them more representative of the natural variation spectrum.

Genomics based Plant Breeding

Genome-wide genetic diversity studies

Most of the genotyping platforms described above are used in the corresponding crops for studies of diversity and population structure. By using representative diversity panels, polymorphism levels are calculated for individual SNP markers, minor allele frequencies (MAFs), etc., enabling the collection of these biologically and highly polymorphic SNPs in the various groups. For example, the Infinium arrays produced in some of these crops are used to construct haplotype maps for large collections of germplasms, such as the USDA soybean germplasm collection's 18,000 accessions.

Molecular marker recognition Linked to single genes and QTLs

NGS and high-resolution maps have resulted in significant advances in the detection of molecular markers linked to specific genes and QTLs. The most important advantage is the broad coverage of the genome, which enables markers to be identified that are closely linked to any target genomic region, with the benefits that this offers a close link. Methods that were already used in the pre-genomics era to facilitate the identification of single loci markers, such as bulked segregant analysis (BSA), are now being optimized. A Golden Gate assay, for example, has in conjunction with BSA; the mapping of the dominant resistance locus to soybean rust Rpp3 was greatly accelerated. There are a growing number of reports on the use of NGS technologies to classify molecular markers that are closely linked to major genes in this regard. For example, in cucumber RILs (*Cucumis sativus*), a fine genetic mapping of the single dominant scab resistance gene (Ccu) was performed. Detection of QTL has historically been done by mapping linkages. NGS technologies contribute significantly to the precision of QTL detection. They can increase the number of markers in many orders of magnitude mapped, ensuring high mapping resolution, and also aid in the development of mapping populations, such as RILs, near isogenic lines (NILs), and CSSLs, more appropriated for QTLs detection.

Marker Assisted 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 probability of recombination, given the close relation between the marker and the gene, restricts the use of MAS. Using intragenic markers can help to overcome this limitation, also called functional markers. Projects for NGS sequencing produce large sets of usable markers. Such markers boost the actual gene supported reproduction, thus reducing the risk of loss of desirable feature of recombination. This is now feasible in many crop species where sequencing of NGS cDNA is performed. Some of these studies carry out profiling of expression, recognizing candidates and specific markers associated with the gene. MAS are also often used in the sense of backcrossing programs to perform background selection. For example, background selection integrated with foreground selection of bacterial blight resistance (xa13 and Xa21 genes), amylose content (waxy gene) and fertility restore gene was performed to identify top lines with maximum recovery of Basmati rice genome along with quality characteristics and minimal non-target genomic introgressions of donor chromosomes.

Genomic Selection

The method was first identified as an attempt to exploit knowledge created from emerging technologies of genotyping in 2001. Genomic selection is based on simultaneous estimation of the phenotype effects of all available loci, haplotypes, and markers. The disparity between Many MAS approaches are based on the fact that no previous collection of phenotype-related markers has been established. Genomic selection requires the availability for the reference population of phenotypic and genotypic data. This data set will allow the model's parameters to be calculated so that the markers analyzed explain the differences at the phenotype level. Once the model is developed, application to breeding populations allows the determination of each individual's genomic value, i.e., the predicted phenotype based on genotypic data. The prerequisite is that adequate molecular markers are available to provide good genome coverage. The utility of genomic selection applied to an initial cross between an adapted line and exotic germplasm was demonstrated through simulation experiments using maize. It was possible to obtain good selection response after 7-8 generations with 512 markers and a reference population of 288 F_2 plants evaluated in six different environments. Simulations also found that for genomic selection, the response to selection was 18 to 43 percent larger than for recurrent selection supported by markers. The response obtained by phenotypic selection when using genomic selection may be lower than the response. Nevertheless, due to early MAS, the reduction in cycle duration results in an increase in gain per unit of time. Species with a long generation period, such as tree species, are

accused of this reduction even more. The development of phenotypic databases for different crops has made it possible to equate the genotypic value predictions produced using genomic selection with the actual genotypic value as shown by the trait's phenotypic manifestation.

Conclusions

For some major crops, it will be difficult to maintain the rate observed in the 20th century for genetic yield gains and other complex traits if only current pre-genomics technologies are used. Nevertheless, plant breeding is a complex science and, luckily, resources and methods for genomics are already available and are helping to make another quantitative leap in plant breeding. There are already many developments in this regard, and super-domestication, i.e., "processes leading to domestication with dramatically increased yield that could not be selected in natural environments from naturally occurring variation without recourse to new technologies", will require the combination of conventional breeding with crop genomics. Genomic methods and strategies will help conventional breeding make significant progress in crop breeding that has either remained orphaned or ignored from the point of view of genetic improvement. Thus, though traditional plant breeding pre-genomics has been, is and will be effective in improving our crops, the application of genomic tools and resources to realistic plant breeding will drive forward the genetic gains obtained through breeding programs. Advanced Technologies are already being produced, should make it easier for breeders to obtain new cultivars with improved features, either by encouraging selection or by increasing the variation available to breeders through the use of specific breeding approaches. The current and emerging methods of genomics, in particular, are of great value for the genetic dissection and breeding of complex traits.

References

1. Barbazuk WB, Emrich SJ, Hsin D, Chen L, Li P, Schnable PS. SNP discovery via 454 transcriptome sequencing. Plant J. 2007; 51:910-918.
2. Barkley NA, Wang ML. Application of Tilling and Eco Tilling as reverse genetic approaches to elucidate the function of genes in plants and animals. Curr. Genomics. 2008; 9:212-226.
3. Bernardo R. Genomewide selection for rapid introgression of exotic germplasm in maize. Crop Sci. 2009; 49:419-425.
4. Bernardo R, Yu J. Prospects for genomewide selection for quantitative traits in maize. Crop Sci. 2007; 47:1082-1090.
5. Colbert T, Till BJ, Tompa R, Reynolds S, Steine MN, Yeung AT *et al*. Highthroughput screening for induced point mutations. Plant Physiol. 2001; 126:480-484.
6. Collard BCY, Mackill DJ. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Phil. Transact. Royal Soc. B. 2008; 363:557-572.
7. Fedoroff NV. The past, present and future of crop genetic modification. New Bio technol. 2010; 27:461-465.
8. Heffner EL, Sorrells ME, Jannink JL. Genomic selection for crop improvement. Crop Sci. 2009; 49:1-12.
9. Horner DS, Pavesi G, Castrignanò T, Meo PDOD, Liuni S, Sammeth M *et al*. Bioinformatics approaches for genomics and post genomics applications of nextgeneration sequencing. Brief. Bio inform. 2009; 2:181-197.
10. Jannink JL, Lorenz AJ, Iwata H. Genomic selection in plant breeding: from theory to practice. Brief. Funct. Genomics. 2010; 9:166-177.
11. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al*. Genome Project Data P: The Sequence Alignment/Map format and SAM tools. Bio informatics, 2009; 25:2078-2079.
12. Logacheva MD, Kasianov AS, Vinogradov DV, Samigullin TH, Gelfand MS, Makeev VJ *et al*. De novo sequencing and characterization of floral transcriptome in two species of buckwheat (Fagopyrum). BMC Genomics. 2011; 12:30.
13. Metzker M. Sequencing technologies-the next generation. Nat. Rev. Genet. 2010; 11:31.
14. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001; 157:1819-1829.
15. Morgante M, Salamini F. From plant genomics to breeding practice. Curr. Opin. Bio technol. 2003; 14:214-219.
16. Perry JA, Wang TL, Welham TJ, Gardner S, Pike JM, Yoshida S *et al*. A Tilling reverse genetics tool and a web-accessible collection of mutants of the legume Lotus japonicus. Plant Physiol. 2003; 131:866-871.
17. Pop M, Salzberg SL. Bioinformatics challenges of new sequencing technology. Trends Genet. 2008; 24:142-149.
18. Varshney RK, Glaszmann JC, Leung H, Ribaut JM. More genomic resources for less-studied crops. Trends Bio technol. 2010; 28:452-460.
19. Varshney RK, Tuberosa R. Genomics-assisted crop improvement vol. 1: Genomics approaches and platforms. Springer; New York, 2007a.
20. Varshney RK, Tuberosa R. Genomics-assisted crop improvement vol. 2: Genomics applications in crops. Springer, New York, 2007b.
21. Wang Z, Fang B, Jingyi C, Zhang X, Luo X, Huang L *et al*. De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweetpotato (Ipomoea batatas). BMC Genomics. 2011; 11:726.
22. You FM, Huo N, Deal KR, Gu YQ, Luo MC, McGuire PE *et al*. Annotation-based genome-wide SNP discovery in the large and complex *Aegilops tauschii* genome using next-generation sequencing without a reference genome sequence. BMC Genomics. 2011; 12:59.