



## High Through-Put Sequencing Technologies and Application in Crop Improvement with Big Data Analysis

Mahnoor Patel\*

\*Department of Molecular Biology and Genetic Engineering, School of Bioengineering and Biosciences, Lovely Professional University, Punjab, India.

### Abstract

Big data analysis is the most confirmatory platforms for the implementation of large data analytics task which can encompass the way for finding unidentified connections, hidden patterns, including other indispensable data from the widespread dispersed dataset. For observing disseminated dataset of an expensive crop deals by applying data clustering can be done for crop planning, which may furthermore lead to an increase within agriculture planning in India. It is also important in the prediction about the crop sowing schedule for deciding which crop should be sown in which particular season. Input quality is an essential factor for the analysis of yield quality. Therefore, obtainability and convenience of right particular inputs to farmers is the most important key for the empowerment of farmers. It can also predict crop prices in additional years of the season; it also helps the farmers for adopting the cultivation of crop planning. Because farmer acts as a backbone to any nation and the economy.

**Keywords:** Application, Big Data Analysis, Crop Improvement, High Through-Put Sequencing, Next Generation Sequencing.

### Corresponding author: Mahnoor Patel

Department of Molecular Biology and Genetic Engineering, School of Bioengineering and Biosciences, Lovely Professional University, Punjab, India.

E-mail: [patelmahnoor786@gmail.com](mailto:patelmahnoor786@gmail.com)

ORCID ID: <https://orcid.org/0000-0003-4073-5952>

**Citation:** Mahnoor Patel. (2020), High Through-Put Sequencing Technologies and Application in Crop Improvement with Big Data Analysis Int J Biotech & Bioeng. 6:7

**Copyright:** © 2020 Mahnoor Patel. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

**Received:** November 11, 2020

**Accepted:** November 16, 2020

**Published:** XXXX XX, 2020

### 1. Introduction:

#### 1.1: Big Data Analysis:

Government of every country spends a lot on gathering the data of agriculture. Data is much quicker than the speed of computation. Sales data for the crops can be used to signify the data for the crops. Government dynamically and continuously collecting a dataset of crop sales, but dataset size is a big data in a real-world that's why analysis is a hard problem to be solved. For the analysis of big data, statistical techniques and data mining needs to be expanded under the equivalent and dispersed computing platforms having large storage

amount with computation time for the management of enormous dataset. Big data is real-time in nature as compared to traditional applications<sup>[1]</sup>.

#### 1.2: Big Data Platforms:

**1.2.1: Big Data:** It is generally compared with the small data. Data bigger than subsequent data needs more diverse and improved systems for solving social issues. Big data analysis is a custom method for accumulating the large data volumes, organization and analysis of huge facts set for finding patterns including alternate cooperative information<sup>[1]</sup>.

**1.2.2: Big Data Traits:** Volume – the scale of data, quantity which needs to be processed is rapidly growing, including the exponential boom in composed/created data. Variety – data which is in much more forms, it can apply to produce/collect different forms for extracting knowledge; it also needs to relate with each other. Velocity – data which is in motion that is streaming facts analysis which makes possible decisions in a fraction of a second, also information can be collected quickly. Veracity – uncertainty in data, because of the discrepancy and reliability, uncertainties, dormancy, deception and approximation in the model<sup>[1]</sup>.

**1.2.3: Big Data can handle any Data:** Structured Data – any kind of data which can be stored, retrieved and processed in a mounted format known as structured data. Unstructured Data – data without any predefined format. Data can be in any kind of format which needs to be stored. Semi-Structured Data – data having unknown kind or the shape which is classified as unstructured data<sup>[1]</sup>.

### 1.3: The New Era of Agriculture with Big Data:

For increasing effectiveness of an entire agriculture chain which enables financial enclosure of farmers with the help of data-driven insights for the improvement of the efficient marketplace and supply chain for the agriculture producers which includes traders and policymakers. For that unique risk management solution creation which can use the multi-sensor data like satellites, drones and the algorithms of artificial intelligence for actionable intelligence development [1].

## 2. Methodology:

First of all, crop sales data needs to be collected from the computers for running data collection program which can collect data from internet and store that in an open database of agriculture including two parameters year and crop type. Special attributes can also use inside the clustering method, as time and type of crop. File clustering and distribution is useful requirements. Submitting, extracting and storing processes can be repeated according to the crop numbers which are there in the list of crops [1].

### 2.2: Framework of Big Data Analysis:

The framework includes crop sales which will be analyzed by applying data clustering. Data clustering algorithm needs to run inside the platform which offers parallel processing with allotted distributed characteristics of computers to the process. Prediction of the grouping of data which is based on patterns found previously can aid with help of decision tree [1,2].

## 3. Next Generation Sequencing:

Next-generation sequencing (NGS) technology used for the generation of whole-genome sequencing for a wider range of crop species. When it associates with the precise methods of phenotyping, it provides commanding and quick tools for the identification of the genetic basis of the agriculturally vital traits for prediction of breeding value of individuals in crop breeding population [2-4].

An increase in the world's population creating a global problem for food. For feeding the growing population raising a problem for supplying the quality food, that is why food production needs to increase approximately 70% till 2050 [5-9]. Increasing global warming problems and climate changes creating more problems in food production. Therefore, an imperative need for acceleration and improvements in crop breeding by implementing new strategies for the management for achieving sustainable yield increasing without further expands in farming land or environment damage [10]. To overcome these challenges, scientist across the globe are developing new and efficient breeding strategies by increasing genomic technologies and high throughput phenotyping methods for the better utilization of natural and induced genetic variation. Rapid development in next-generation sequencing from the last decades helps more and more for the exploration of new opportunities between genotype and phenotype [11]. Nowadays, the cost for sequencing is also decreasing, so plant breeders have started using it for increasing regularities for large scale sequencing of plant genome [12,13].

With the help of manipulation of next-generation sequencing (NGS) technology, many species of economically important crops, subjected to de-novo assembly and resequencing whole-genome sequencing.

NGS technologies deliver improved opportunities for crop genomics and post-genomic study more and more closely which includes transcriptomics, proteomics and metabolomics. With the help of molecular markers which are generated by NGS and other types of analysis, the scientist can able to explore genetic diversity with crop evolution including whole genome sequencing between crop species and its accessions within other crop species. The increasing availability and reduced cost of high throughput technology have encouraged genomic sequencing of few model plant species to the sequencing of any crop which is important for food security and production for the economic purpose [14].

Nowadays, production of crops showed continuous stable growth, still, improvement is needed as the hunger and malnutrition are faced by particular portions of the world's population. Also, the increasing demand for high-quality food and wealth levels creating an imbalance in food purchasing. It also creates problems in food prices doubling with massive production of biofuels, changes in climate and urbanization which makes greater competitions for water resources, land, energy sources, even biodiversity and natural ecosystems needs protection.

To feed the increasing population of the world, agriculture biotechnology development will be the key source for improvement in crop species which are environmentally friendly and socially sustainable. NGS which is the most advanced technology for sequencing genome is becoming the main tool for the novel molecular markers development and identification of important genes in agriculture. Before the development of these methods, time-consuming clone by clone method was performed for sequencing genome including the strategy of least redundant overlapping clone's identification. Because physical crop genetic map to be sequenced needs to provide prior for performing these types of labour-intensive time-consuming experiments. By selecting the sequencing technology which can generate tremendous amounts of data in a short period is very important [14-17].

## 4. Status of Crop Genome Sequencing:

After switching from the Sanger sequencing method to NGS, the sequencing of plant genome increases dramatically. *Arabidopsis thaliana* was the first model plant to be sequenced completely. With the low cost including short time consumption, whole-genome sequencing quality has been improved. It also helps in understanding the strong relationship between the complex genotype of the plant and its evolution [18-20].

## 5. Whole Genome Sequencing Technology:

### 5.1: De-novo Assembly:

As the reference genome was not available previously, the discovery of crop genes was completely dependent on unassembled genome sequences including expressed sequence tags. From these kinds of sequences single nucleotide polymorphism and simple sequence repeat, molecular markers could be developed in the absence of a reference genome. That's why sufficient data of orphan or less studied plants can be obtained by this strategy. The earlier strategy of clone by clone sequencing was employed as a physical map by using cosmid or bacterial artificial chromosome clones that were the typical de novo

assembly for whole-genome sequencing before the advancement in NGS project. Nowadays, with the advancement in whole-genome sequencing strategy, inserts of different sizes from genomic DNA will be constructed and sequenced from both the ends. With the rapid expansion in NGS as compared to other sequencing methods, hybrid methods are developed which is based on pyrosequencing and will become a current strategy for de novo assembly. It is also very fast technology for sequencing as compared to other conventional techniques and helped in the development of many de novo assembly algorithms [21-23].

## 6. Applications in Crop Improvement:

### 6.1: Whole Genome Analysis:

With the rapid advancement in bioinformatics tools, sequences which are generated by NGS will be analyzed at the genome level. Data which are generated from NGS will be sequence align, array hybridization, mapping of the genome and genotyping study for the analysis of similarities and differences between genomes. The whole-genome will be magnified and analyzed on the global scale by discovering whole-genome profiles of different chromosomes within the species or between two genomes [24-27].

### 7. Future Directions:

Next-generation sequencing technology offers a golden opportunity for crops biological systems understanding. As compared to Sanger sequencing which is a traditional method, the cost of NGS is radically decreased with the increase in advancement in the NGS technology which is becoming more and more feasible for the researchers who are engaging themselves for crop genome sequencing for economical purposes. Because of lack in marker information and sequences, some crops considered as the orphan or the less-studied crops. But nowadays, with the help of de novo assembly employment, whole-genome sequences of that kind of less studied or orphan crops also becoming more and more feasible for the improvement of crops. It also includes the development of more molecular markers like SNP and InDels which have been developing rapidly at the low cost and is also applicable for the crop breeding program [28].

### 8. Big Data in Agriculture and Food:

Nowadays, farming is experiencing a digital revolution. Generally, farming has been an empirically determined for the century ago, but data were not collected in digital form. As compared to the historic gathering of data, big data is a different form of collecting data when it comes to volume and analytics which includes contemporary digital technologies. It also promises of precision data, storage of information, it's processing, analyzing which was impossible previously because of the limitations of that technologies. In comparison with a notebook which is important for a farmer to keep the information about crop performance with digital phone "app" used for predicting and directing practices of future production. That application will use the logging information for efficient work such as voice recognition and information volume by which farmer can access. Tools of agricultural management provide admittance for interaction with the datasets which can stretch way outside the separate farm. Till now, nature of traditional historic farming monitoring and the big data is remaining an open question in research, scientifically digital revolution tracing

for agriculture and registering affordance as well big data-driven limitations is the vast goal in research for food and agriculture realm [29].

With the help of breeding, crop improvement can be done and it is a sustainable approach for increasing yield and its stability without fertilizers and pesticides use. The latest advancement in bioinformatics and genomics is important for providing new opportunities in acceleration in growth and improvement of crops. With the advancement in third-generation sequencing technologies can overcome challenges included in the genomic assembly of plants which can cause by polyploidy or frequent element repetitions. Because of that, crop reference genomes of the highest quality are available increasingly, it also benefited in a downstream analysis like variant calling with mapping association for identification of breeding targets presents in genomes [30]. For the identification and analysis of genomic regions in agronomic value by facilitations in a functional annotation in the genome for enabling high-throughput real-time phenotyping of important traits in agronomy presents in the glasshouse and available also in the field with the help of machine learning. Moreover, databases of crops which can integrate the genotype and phenotype data in greater volume deliver valuable assets for crop breeders with an opportunity to mine the data approaches for uncovering candidate genes associated with the novel trait. With the expansion of crop genetics knowledge, selection and editing in the genome provide great promises for the disease-resistant breeding and stress-tolerant crops yielding.

For the processing and analyzing genomic datasets at a large scale to get the functional insights of crop genomics, bioinformatics plays a crucial role. When it comes to data analysis at the genomic level, assembly of genome, sequence alignment with variant calling known as standard tools of bioinformatics. There are different algorithms which important and required for analysis based on different computational approaches including unique biases. When it comes to alignment of third-generation sequencing data, many more tools are developed just for the performing of short reads and its alignment with long reads. The greatest challenge is to combine the data collected from research and its assembly with alignment, that data includes short and long reads for reducing the impact on the bases of unique biases. Comparative genomic analysis is also known as downstream analysis and variant calling provides complete information for the facilitation of improvement in the crop. Variant callers are also different in the ability to develop call InDels including their biases near heterozygous with reference variants. Because of all these details, it is very necessary to use the right bioinformatics tools and software for the analysis is very crucial. There are wide ranges of tools available for the genome mining and variant data analysis, processing, but still, it is a leading challenge which will be faced by the researchers working on crop genomics for the improvement. It still needs to adapt the available databases and its improvement for the processing the data generated by the research on crop improvement, so that; the breeders can use that data for the improvement of crop variety. It also needs to accessible for every crop breeders. For an interdisciplinary approach needs to be developed for plant breeding program in the 21st century for the identification and breeding challenges to improve the production of crops. With the help of development in novel glasshouse technologies for the acceleration of plants development, bioinformatics including genomics plays a crucial role in increasing the improved crop production rate. However,

when the genotypic and phenotypic data generates in a vast amount, it needs to be integrated for the diverse data output for the breeding purposes and it will be a great challenge. Integration of phenotypes, bioinformatics and genomics resources including tools which are available in the public and private pipelines of breeding can address the challenges for the breeding targets delivery [31].

### 9. Crop Genome Assembly Improvement with the Help of Third-Generation Sequencing:

After the introduction of NGS technology for sequencing and resequencing became standard tools in disciplines of plant biotechnology. Though, there are few limitations of this sequencing technology, like the ambiguous alignment of the repetitive elements and inherent biases which is needed for the assembly of high fragmented genome drafting and its complications of the hidden InDels study with structural variants. The development of the third generation sequencing technologies has facilitated the generation and production of the long read of more and more precise and contiguous genomic assemblies. It also supports for the generation assemblies of de novo highest quality whole genome sequencing, with the help of regions of reads spanning complex like high-level repetitive sequence and also sheds lights on the repeat sequence complex which remains with the other structural variants. Furthermore, sequencing of full-length transcripts also produced by the use of third-generation sequencing technologies which allows the accurate study of exons, alternatively spliced regions including splice sites for improvement of genomic annotations. These kinds of the completely assembled well-annotated genome of the crops permit breeders to discover agronomic traits related important genes, also helps in the determination of the location of particular genes and its functions as well genome-wide molecular markers development [31].

### 10: Integrated Crop Database:

With the development and advancement in third-generation sequencing technologies including other "Omics" approach emerging, data which are available in the large amount will be helpful for the investigation of crop traits to population-level from genes level. A sequence repository which includes Genbank, PlantGDB, and European Molecular Biological Laboratory plays an important role for the storing and management of genome data without the integration of variant or phenotypic data from all sources. But this makes the data much more challenging for breeders and plant biotechnologist for linking of genotype with appropriate phenotype. For the creation of integrative crop database with the help of combining annotated genome sequences, interaction networks, gene functions and the phenotype traits is much more challenging because of applicable data which are dispersed in various databases in different forms and formats with diverse coverage and quality. Data mining perceptively from large crop databases needs merging of the complex resources of data which allows the discovery of gene and crop improvements. With the integration of crop database to advance mining technologies helps plant breeders for understanding complex traits in a better way including identification of trait-associated applicant gene important for benefits in improvements of crops [25-31].

## 11: Integrative Genomics Application for the Trait Discovery and Improvements of Crops:

### 11.1: Mining of Quantitative Trait Loci Studies:

Analysis of Quantitative Trait Loci i.e. QTL permits the assessment of genetic regions which is linked with Quantitative Phenotypic Traits, for connecting the gap of field and genomics. Though, as the cumulative number of QTL studies which being directed and reported for particular plant, new challenges rises for identification of highest quality applicant loci with the improvement of crops breeding and integrates information of diverse QTL study. In these kinds of studies, the meta-analysis tool for pooling the outcomes in a variety of studies ranges and prediction of QTL locations with more accuracy rather than individual studies is essential for full utilization of existing resources. For proficiently carrying out analysis of meta-QTL, a wide range of bioinformatics tools are available. Such as the consensus model with statistics, MetaQTL computational package help to reduce the length of QTL confidence interval which leads to an accurate assessment of QTL location correctly and effectively [32].

### 11.2: Applying Machine Learning in Crop Breeding:

Machine learning permits algorithms for data interpretation with the help of patterns learning through experience. For the largest, formless and diverse datasets, like which are generated by the sequencing or photo imaging, machine learning will provide considerable advantages over another analytical approach with the assistance of Machine Learning i.e. ML, plant breeders can effectively phenotype the particular plants and also mine diverse patterns datasets like the association between genes and its related phenotypes and traits.

### 11.3: High-Throughput Crop Phenotypes:

Measurement of structural or functional traits from cellular to organismal level is a vital character for the crop improvement and association studies with the help of plant phenotyping. Intensively developments of genomic research with sequencing technologies, there will an increase in the demand of plant phenotypes for the understanding of genomic data. As the conventional phenotyping is frequently a bottleneck because it is error-prone, subjective, time-consuming and labour intensive with the limited numbers of traits including environments and plants which can be sampled. With the advancement in the measuring technologies i.e. high throughput imaging with automatic sensors including machine learning permits establishment of the high throughput robotic phenotyping, which allows overcoming the human-based traditional phenotyping methods and also allows rapid generations of phenotypical features in the large populations. There are four different elements in the high throughput phenotyping includes detection with the help of sensors or imaging, classification of phenotypical data, quantification of features and specific models or the algorithms based prediction. Phenotyping by using machine learning also applied in disease monitoring and stress phenotyping. Real timing machine learning-based high phenotyping methods were developed for the assessment of relentlessness of iron deficiency chlorosis. Though nowadays machine learning has been applied successfully for the crops genomics and phenotyping, many challenges need to be solved till now. It necessitates big datasets for the model construction as well for training. When it comes to the small training set, statistically it is insignificant and leads to many problematic false predictions. And the big datasets is more time consuming with lots of cost for the acquisition including crop

measurement which can only take one time per its growth cycle. Also, high throughput machine learning approach based phenotyping limits to certain leading research institutes which include few prominent commercial companies. It also needs to reduce the purchase cost with the operating cost of the technology to make available the machine learning approach based widely used phenotype in the future farming [29-32].

## 12: Machine Learning Approach in Crop Genomics Research:

Machine learning plays a crucial role in genomic research which includes genomic assembly also. It can be used for the improvement of assemblies of polyploid genomes with the redundancy of complex genome. For an understanding of the plant gene functions and structures, whole-genome assembly with annotation is the groundwork to track variations in genetics within plant species. This is a very crucial pipeline discovery of crop traits. Highly redundant genome assembling is the challenge for non-machine learning-based assembling approaches which use a linear algorithm for assembling regions of repetitive sequences. Assuming the relationship between the genes and elements is an auspicious field for previously unknown candidate's identification for improvement of crops. Co-expression based only on the gene level, in-silico regulatory network constructed is limited just because of association between the genes which may not reflect shared genes regulation accurately. Therefore, machine learning-based methods which can incorporate different kinds of regulatory signal from the different sources of data now become popular for interactive exploration of gene regulation networks [32].

## 13: Conclusion:

As agriculture faces significant challenges in connecting the cascade genomic data of different origin and diverse format for improvement of crops. For overcoming these challenges, novel methods of breeding need to develop with the advancement in bioinformatics tools for translation of genomic data for crop yielding acquisition and its stability. For acceleration of robust gene trait detection, the researcher needs to apply the meta-QTL analysis. Although, genome editing provides a valuable approach for rapid introduction of beneficial mutants into elite cultivars. Machine learning algorithms also take an advantage of high through-put phenotyping with genomic data to automate gene discovery pipeline like genome annotation with image interpretation which remains predominantly challenges. With the application of novel technologies and diverse methods, future plants breeding this can achieve improvement of crops rate requirement for confirmation of food security.

## 14: References:

1. Sekhar C., Udaykumar J., Kumar B.K. and Sekhar C. (2018). "Effective Use of Big Data Analytics in Crop Planning to Increase Agriculture Production in India." *International Journal of Advanced Science and Technology*. Vol. 113.
2. Varshney R.K., Terauchi R. and McCouch S.R. (2014). "Harvesting the Promising Fruits of Genomics: Applying Genome Sequencing Technologies to Crop Breeding." *PLoS Biology*. Vol. 12 (6).

3. Van K., Rastogi K., Kim K.H. and Lee S.H. (2013). "Next-Generation Sequencing Technology for Crop Improvement." *SABRAO Journal of Breeding and Genetics*. Vol. 45 (1).
4. Wang W., Cao X.H., Mičlus M., Xu J. and Xiong W. (2017). "The Promise of Agriculture Genomics." *International Journal of Genomics*. Vol. 2017.
5. Zvinavashe A.T., Lim E., Sun H. and Marelli B. (2019). "A Bioinspired Approach to Engineer Seed Microenvironment to Boost Germination and Mitigate Soil Salinity." *PNAS*.
6. Bronson K. and Knezevic I. (2016). "Big Data in Food and agriculture." *Big Data and Society*.
7. Hu H., Scheben A. and Edwards D. (2018). "Advances in Integrating Genomics and Bioinformatics in the Plant Breeding Pipeline." *Agriculture*. Vol. 8 (75).
8. Tester M. and Langridge P. (2010). "Breeding Technologies to Increase Crop Production in a Changing World." *Science*. Vol. 327.
9. Varshney R.K., Bansal K.C., Aggarwal P.K. Datta S.K. and Craufurd P.Q. (2011). "Agricultural Biotechnology for Crop Improvement in a variable Climate: Hope or Hype?" *Trends Plants Science*. Vol. 16.
10. Godfray H.C., Crute I.R., Haddad L., Lawrence D. and Muir J.F. (2010). "The Future of the Global Food System." *Philos Trans R Soc Lond B Biol Sci*. Vol. 365.
11. Edwards D. and Batley J. (2010). "Plant Genome Sequencing: Applications for Crop Improvement." *Plant Biotechnology Journal*. Vol. 8.
12. Krzywinski M., Schein J., Birol I., Connors J., Gascoyne R., Horsman D., Jones S.J. and Marra M.A. (2009). "Circos: An Information Aesthetic for Comparative Genomics." *Genome Res*. Vol. 19.
13. Smit E., Nieto F.J., Crespo C.J., and Mitchell P. (1999). "Estimates of animal and Plant Protein Intake in US Adults: Results from the Third National Health and Nutrition Examination Survey." *J. Acad. Nutr. Diet*. Vol. 99.
14. Ulijaszek S.J. (1991). "Human Dietary Change." *Trans. R. Soc. Lond. B. Biol. Sci*. Vol. 334.
15. Tilman D., Balzer C., Hill J. and Befort B.L. (2011). "Global Food Demand and the Sustainable Intensification of Agriculture." *Proc. Natl. Acad. Sci*. Vol. 108.
16. Godfray H.C.J., Beddington J.R., Crute I.R., Haddad L., Lawrence D., Muir J.F., Pretty J., Robinson S., Thomas S.M. and Toulmin C. (2010). "Food Security: The Challenge of Feeding 9 Billion People." *Science*. Vol. 327.

17. Perez-de-Castro A.M., Vilanova S., Canizares J., Pascual L., M Blanca J., J Diez M., Prohens J. and Pico B. (2012). "Application of Genomic Tools in Plant Breeding." *Curr. Genom.* Vol. 13.
18. Abberton M., Batley J., Bentley A., Bryant J., Cai H., Cockram J., De Oliveira A.C., Cseke L.J., Dempewolf H. and De Pace C. (2016). "Global Agricultural Intensification During Climate Change: A Role for Genomics." *Plant Biotechnology Journal.* Vol. 14.
19. Edwards D., Batley J. and Snowdon R.J. (2013). "Accessing Complex Crop Genomes with Next-Generation Sequencing." *Theor. Appl. Genet.* Vol. 126.
20. IRGSP. (2005). "The Map-Based Sequence of the Rice Genome." *Nature.* Vol. 436.
21. Velasco R., Zharkikh A., Troglio M., Cartwright D.A., Cestaro A., Pruss D., Pindo M., Fitzgerald L.M., Vezzulli S. and Reid J. (2007). "A High Quality Draft Consensus Sequence of the Genome of a Heterozygous Grapevine Variety." *PLoS One.* Vol. 2.
22. Huang S., Li R., Zhang Z., Gu X., Fan W., Lucas W.J., Wang X., Xie B. and Ni P. (2009). "The Genome of the Cucumber, *Cucumis sativus* L." *Nat. Genet.* Vol. 41.
23. Goodwin S., McPherson J.D. and McCombie W.R. (2016). "Coming of Age: Ten Years of Next-Generation Sequencing Technologies." *Nat. Rev. Genet.* Vol. 17.
24. Michael T.P. and Jackson S. (2013). "The First 50 Plant Genomes." *Plant Genome.* Vol. 6.
25. Mousavi-Derazmahalleh M., Bayer P.E., Hane J.K., Babu V., Nguyen H.T., Nelson M.N., Erskine W., Varshney R.K., Papa R. and Edwards D. (2018). "Adapting Legume Crops to Climate Change Using Genomic Approaches." *Plant Cell Environ.*
26. Dwivedi S.L. Scheben A., Edwards D., Spillane C. and Ortiz R. (2017). "Assessing and Exploiting Functional Diversity in Germplasm Pools to Enhance Abiotic Stress Adaptation and Yield in Cereals and Food Legumes." *Front. Plant Sci.* Vol. 8.
27. Moore J.H. Asselbergs F.W. and Williams S.M. (2010). "Bioinformatics challenges for Genome-Wide Association Studies." *Bioinformatics.* Vol. 26.
28. Batley J. and Edwards D. (2016). "The Application of Genomics and Bioinformatics to Accelerate Crop Improvement in a Changing Climate." *Curr. Opin. Plant Biol.* Vol. 30.
29. Lawrence M., Huber W., Pages H., Aboyou P., Carlson M., Gentleman R., Morgan M.T. and Carey V.J. (2013). "Software for Computing and Annotating Genomic Ranges." *PLoS Comput. Biol.* Vol. 9.
30. Pabinger S., Dander A., Fischer M., Snajder R., Sperk M., Efremova M., Krabichler B., Speicher M.R., Zschocke J. and Trajanoski Z. (2014). "A Survey of Tools for Variant Analysis of Next-Generation Genome Sequencing Data." *Brief Bioinform.* Vol. 15.
31. Hwang S., Kim E., Lee I. and Marcotte E.M. (2015). "Systematic Comparison of Variant Calling Pipelines Using Gold Standard Personal Exome Variants." *Sci. Rep.* Vol. 5.
32. Sedlazeck F.J., Lee H., Darby C.A. and Schatz M.C. (2018). "Piercing the Dark Matter: Bioinformatics of Long-Range Sequencing and Mapping." *Nat. Rev. Genet.* Vol. 6.