

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.

For more information visit www.intechopen.com



Genomics Era for Plants and Crop Species – Advances Made and Needed Tasks Ahead

Ibrokhim Y. Abdurakhmonov

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62083>

Abstract

Historically, unintentional plant selection and subsequent crop domestication, coupled with the need and desire to get more food and feed products, have resulted in the continuous development of plant breeding and genetics efforts. The progress made toward this goal elucidated plant genome compositions and led to decoding the full DNA sequences of plant genomes controlling the entire plant life. Plant genomics aims to develop high-throughput genome-wide-scale technologies, tools, and methodologies to elucidate the basics of genetic traits/characteristics, genetic diversities, and by-product production; to understand the phenotypic development throughout plant ontogenesis with genetic by environmental interactions; to map important loci in the genome; and to accelerate crop improvement. Plant genomics research efforts have continuously increased in the past 30 years due to the availability of cost-effective, high-throughput DNA sequencing platforms that resulted in fully sequenced 100 plant genomes with broad implications for every aspect of plant biology research and application. These technological advances, however, also have generated many unexpected challenges and grand tasks ahead. In this introductory chapter, I aimed briefly to summarize some advances made in plant genomics studies in the past three decades, plant genome sequencing efforts, current state-of-the-art technological developments of genomics era, and some of current grand challenges and needed tasks ahead in the genomics and post-genomics era. I also highlighted the related book chapters contributed by different authors in this book.

Keywords: Plant genome sequencing, genetical genomics, genomic selection, 1KP, 1001 plant genomes, GEEN

1. Introduction

The Plant Kingdom is a key of the food chain in our planet. Plant domestication by humankind occurred in early societal development, and subsequent agricultural practice and uninten-

tional and intentional plant breeding led to developing productive crop species that provided food and feed products for all living organisms, including humans [1, 2]. Plant species are very diverse and there are about 300,000 plant species in the world [3]. Humankind presently grows ~2000 plant species [4] in the agriculturally suitable land of 15.5 million square kilometers to fulfill the human diet. Crop domestication with subsequent breeding and farming has created 15 priority crop species, which provide more than 90% of food products [1, 5]. Besides feeding properties, plants supply clothing and housing materials, balance agrobiosenosis and earth ecology, provide medicines and treatment for many diseases, produce energy and biofuels, and have many other key properties and usages to understand life in our planet [6–10].

Plant domestication, coupled with the need and desire to get more food and feed products, has resulted in continuous development of breeding and genetics efforts [2, 4]. Early primitive selection attempts have subsequently developed the methods of shuffling traits/characteristics between plant genotypes via controlled sexual crosses that discovered the genetics of key characteristics of crops. Furthermore, the development of biological sciences and understanding of the Mendelian and quantitative genetics of phenotypic variations in plant genotypes, equipped with optimized, targeted, and efficient selection, phenotyping, and statistical methods as well as advanced agrochemical technologies of the past centuries, have revolutionized crop breeding efforts. These advances have resulted in the development of superior crop genotypes that have helped to increase agricultural production [11]. Thanks to the “Green Revolution” [11, 12], the efficient exploitation of plant genetic diversity and plant germplasm resources, novel cultivar development, and better and suitable agrochemical technologies for the past 50 years, the world average cereal crop yield has increased 2.6 times (1.35–3.51), whereas there was 5-fold increase in maize production [11]. There are many such examples of successful conventional breeding efforts. Despite this, food deficiency and human starvation still exist widely and will become even worse with an increase of global human population to ~9 billion by 2050 [13], whereby ~1 billion people may suffer hunger [14]. There is a desire and need to feed the increasing human population, sustain agricultural production, and overcome newly emerging biosecurity issues in the era of global climate change with ever worsening environmental conditions on earth, and societal globalization and technological advances [15, 16].

These prompted the plant research community to enrich and power the conventional plant breeding and genetics methods with precise tools beyond conventional hybridization, selection, and cultivation/farming practices. This is also dictated by the long duration of conventional breeding and crop improvement, impacted by the limitations in phenotypic evaluations, masking the effect of the environment, polygenic nature of many key traits with many unnoticed minor genetic components [11], negative genetic correlations between important agronomic traits [15, 17, 18], linkage drags, and distorted segregation issues in hybridization between diverse genotypes [15, 17–19].

To address all these, plant researchers have attempted to decipher the molecular basis of genetic diversities by cloning and sequencing the genes encoding the trait of interest and utilize them in plant breeding as tools in vertical or even via revolutionizing horizontal gene transfers [11]. Progress made toward this goal has elucidated plant genome composition and led to decoding the entire DNA sequences of plant genomes conditioning plant ontogenesis. Here comes “genomics” that was derived from the use of the term “genome”—a haploid set of

chromosomes—coined by Winkeler in 1920. First used in 1986, genomics defined “the enterprise that aimed to map and sequence the entire human genome” [20]. Similarly, “plant genomics” is a discipline of plant sciences targeting to decode, characterize, and study the genetic (DNA/RNA) compositions, structures, organizations, and functions as well as molecular genetic interactions/networks of a plant genome [20–29]. Plant genomics aims to develop large-scale high-throughput technologies and efficient tools and methodologies to elucidate the basics of genetic traits/characteristics, genetic diversities, and by-product production; to understand the phenotypic development throughout plant ontogenesis with genetic by environmental interactions; to map important loci throughout the genome; and to accelerate the crop breeding and selection in a genome-wide scale.

Plant genomics research efforts have continuously increased in the past 30 years. The numbers of scientific publications on plant genomics research have drastically increased and reached 17,210 scientific publications in 2015, as indexed in the PubMed database [30], with its first increase in 2000/2001, following a significant peak after 2010 (Figure 1). The first fully sequenced plant genome was the model plant *Arabidopsis*, which was published in 2000. Since then, almost 50 plant genomes were fully decoded by 2013 [31] and the plant sciences community has finished more than 100 plant genomes by 2015 [32]. Furthermore, the plant sciences community extendedly portrayed a sequencing vision of 1001 *Arabidopsis* accessions [33, 34] and sequencing 1000 plant species [35] that “will have broad implications for areas as diverse as evolutionary sciences, plant breeding and human genetics” while generating many unexpected challenges and grand tasks ahead.

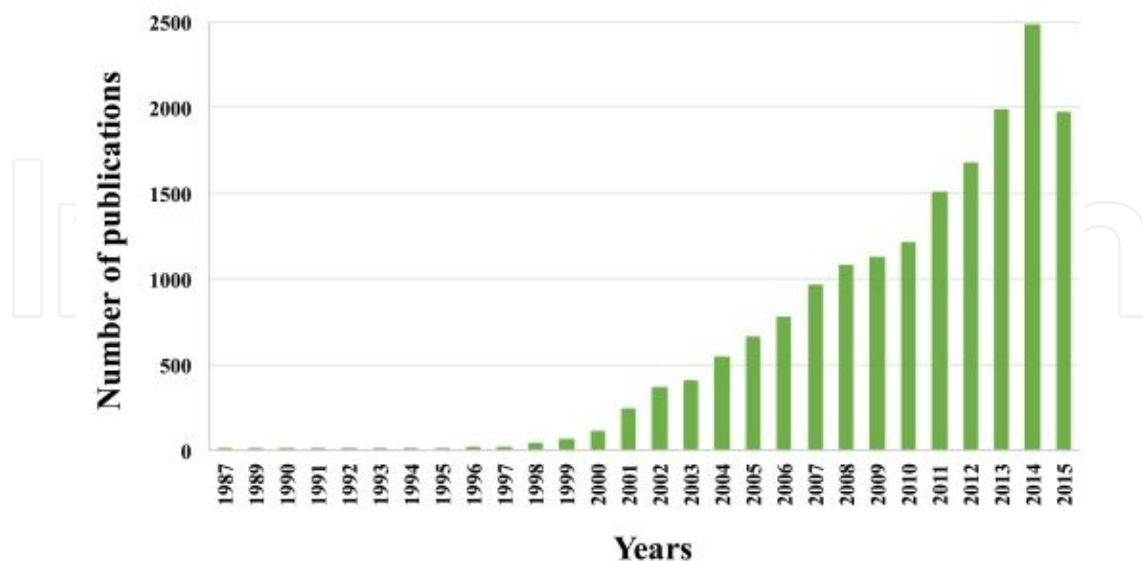


Figure 1. Dynamics of “plant genomics” keyword-retrieved scientific publications in the past three decades. Source: PubMed [30].

2. Genome of plants and crop species

2.1. Challenges and advantages

Compared to other eukaryotic systems, plant genomes are more complex, which create challenges to study its DNA compositions. First of all, the extraction of high-quality DNA from plant tissues, abundantly enriched with phenolic and other metabolic compounds with high affinity to DNA, is conventionally challenging. This interferes with efficient library preparation for whole-genome sequencing [1], although researchers have optimized methodologies to overcome existing issues [36].

Furthermore, plant genomes have widely different chromosome numbers, transposon/retrotransposon transcript retention property, and highly varied ploidy levels with many supergenes, pseudogenes, and repetitive elements including low-, medium-, and high-copy number DNA sequences such as transcribed genes, rRNA genes, and retro-elements or short repetitive sequences, respectively. As a result, plant genomes can be 100 times larger in sizes when compared to animal or other model eukaryotic genomes [1] and may contain many paralogous DNA sequences that make sequencing and genome assemblies difficult, which often will generate false-positive errors [37]. For instance, one of the largest examples of sequenced plant genomes, sugarcane (12 Gbs) and hexaploid wheat genome with 17 Gbs in size, represents 80% repetitive elements [1, 32].

Moreover, these massive repetitive “junk” DNA sequences, organized as a simple tandem repeat, repeat single-copy interspersion, inverted repeats, and compound tandem array arrangements, somewhat mask functionally vital single-copy genes, which create a challenge to characterize and clone important individual genes [32, 37].

Open pollinated, self-pollinated, and clonally propagated plant species have a high level of nucleotide diversity. This can be exemplified by the nucleotide diversity of maize, barley, and grape genomes, where maize genome, for instance, has 10-fold (up to 13%) more polymorphic sites between individual genotypes compared to humans with similar genome size [32, 37]. These polymorphism sites create a challenge in sequence assembly due to the higher rates of nucleotide mismatches to the reference genome.

Lastly, plants tend to have abundant copies of chloroplast genome with two inverted repeat organizations as well as large inversions in some plants with some exchanged regions between nuclear genomes. This creates another challenge in the assembly of repetitive and exchanged regions of chloroplast genomes [32]. The same issue exists in the case of mitochondrial genomes, although it is common for animal genomes as well. All these challenges and complications mentioned above may result in generating fragmented, isolated, and incorrect assemblies in the background of high-copy repeats and paralogous sequences.

However, some specific methodologies and bioinformatics tools have been developed to minimize these challenges. These include the optimized DNA isolation from difficult plant materials [36], use of high-density linkage maps, identification and sorting out of paralogous alleles using local patterns of linkage disequilibrium, and sequencing diploid relatives or

ancestor-like genomes of polyploid plants [37]. The use of laser capture microdissection techniques can isolate individual cell types or chromosome or its arm that could minimize the ploidy or paralogy complexities [27]. Moreover, the use of third-generation single-molecule sequencing approaches [1] and novel assembly methods such as optical mapping and long-range Hi-C interactions can also minimize some of challenging cases with the plant genomes mentioned here, which have been well addressed and covered in detail in a chapter by Deschamps and Llaca in this book.

At the same time, along with these challenges and complexities, plants also offer advantages [37] in genome analyses over other eukaryotic systems. This is due to the clonal propagation and indefinite seed storage properties, which create an opportunity for repeated collection of the same DNA samples for sequencing and studying its phenotype multiple times in many generations across replicated environments [37]. There are no ethical issues associated with the multiple use of plant materials, as it is a sensitive issue for animal cases. The possibility of self-pollination or forced crosses advantageously helps to create highly homozygous samples to reduce existing heterozygosity. There is an opportunity of obtaining double haploid plant genomes [37]. Plant genomes tend to have large chromosomal segments conserved across a large number of taxa in closely related plant species. The collinearity and synteny of plant genomes are very useful to use reference genomes of model species to study homologous and orthologous genes from yet unsequenced genomes [20].

2.2. Sequenced plant genomes

The ability to sequence DNA molecules, which was made possible in the 1970s with the introduction of the “plus and minus” sequencing technique of Sanger and Coulson [38] and Maxam and Gilbert [39], is generally considered to be the starting point of genomics sciences. Later, the simple, long-read chain-terminating dideoxynucleotide DNA sequencing method [40] has become a method of choice to decode genetic sequences. Its eventual automation [41] had extended the capacity and power of this chain-termination sequencing methods to decode the entire genome sequences of living organisms. Because of technological advances and automated sequencing instrumentations [27], a large-scale sequencing of cDNA libraries made it possible to perform serial analysis of gene expression (SAGE) and expressed sequences tags (ESTs). These were the first genomics technologies for all organisms, including plant genomes [42]. Furthermore, these advances powered by microarray tools routinely used by many individual laboratories worldwide have helped to identify the genome structures and functional and regulatory elements across genomes [27] and have facilitated to develop high-throughput reliable molecular markers for genome/trait mapping studies.

The development and generation of massively parallel sequencing technologies [44] provided cost-effective, new-generation sequencing (NGS) platforms that have helped to completely decode the entire genome of many different organisms within a short period. For instance, in plants, the first sequenced genome was a model plant *Arabidopsis thaliana* with 125 Mbs in size, 25,489 individual genes, and 14% repetitive elements, which was published in 2000 [5]. Further, more than 109 plant genomes have been fully sequenced by 2015 [32], including 21 monocots and 83 eudicots, 10 model and 15 non-model plant genomes, five non-flowering plants, and

69 crop species with 6 crop model genomes and 15 wild crop relatives [32]. Following the *Arabidopsis* model, several rice (*Oryza sativa*) genomes in 2002 to 2005, black cottonwood (*Populus trichocarpa*) genome in 2006, and grape (*Vitis vinifera*) genome in 2008 were fully sequenced. Sequencing whole plant genomes has increased in subsequent years, and 10 plant genomes had been sequenced in 2011. About 80% of sequenced genomes were accomplished in the past 3 years (2012–2014; Figure 2).

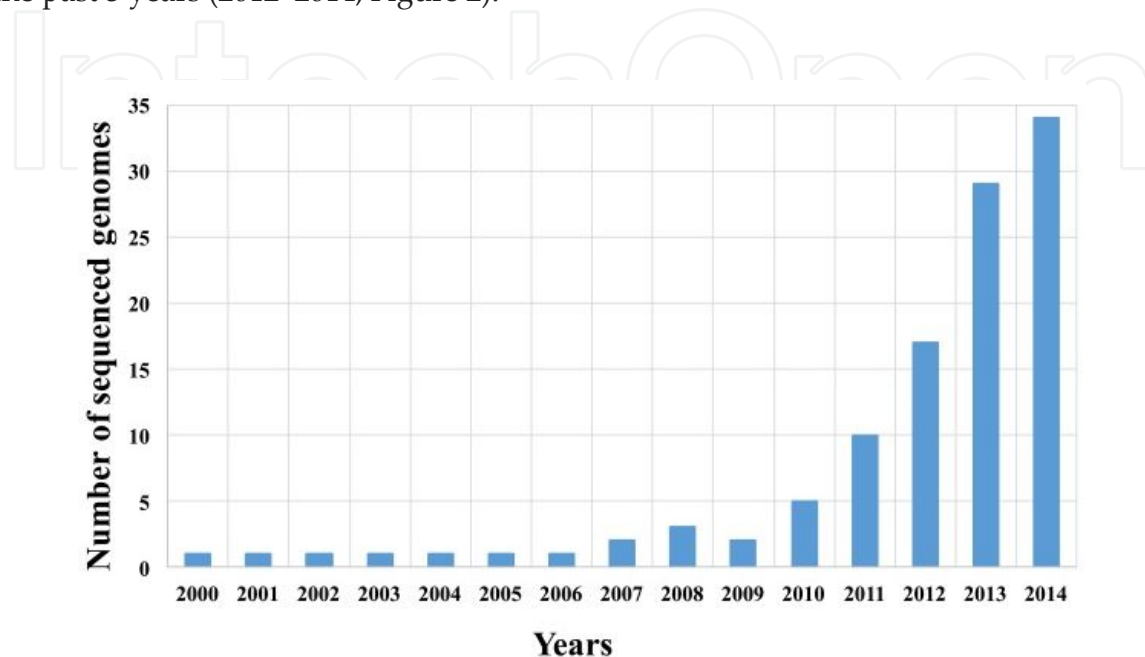


Figure 2. A number of sequenced plant genomes from 2000 to 2014. Source: Ref. [32].

The smallest plant genomes sequenced so far [32] are the two eudicot plants: corkscrew (*Genlisea aurea*) with 64 Mbs genome size and 17,755 genes [45] and bladderwort (*Utricularia gibba*) with 77 Mbs genome size and 28,500 genes [46]. In contrast, the largest genomes sequenced are from gymnosperm plants, including Norway spruce (19,600 Mbs) [47], white spruce (20,000 Mbs) [48], and loblolly pine (23,200 Mbs) [49]. The largest genome sequenced from crop species is the hexaploid wheat (*Triticum aestivum*) with a genome size of 17,000 Mbs [50]. An average size for all published plant genomes is 1850 Mbs. Per published plant genome data [32], the gene numbers of the smallest to largest genomes are within the range of 17,755 (corkscrew) to 124,201 (hexaploid wheat) with an average of 40,738 genes for all sequenced genomes. Repetitive elements are highly variable among published genomes that varied from 3% (bladderwort) to 85% (maize, *Zea mays*) with an average estimate of 46% per genome. These sequenced plant genomes not only provided an updated knowledge on structural compositions and complexities of plant genomes but also elucidated the evolution of gymnosperm and angiosperm plants and specific gene families contributing to the radiation of flowering plants. We learned some direct correlation between genome sizes and gene numbers/repetitive elements, although it does not strictly follow the rules, which was evidenced by several exceptions. For example, one of the largest genomes, Norway spruce, has ~28,000 genes, which is similar to the smallest genome bladderwort. Moreover, medium-sized maize genome (2300 Mbs) or wild tomato (1200 Mbs) contains

more or approximately the same (<80%) contents of repetitive elements compared to the largest sequenced genome of loblolly pine (23,200 Mbs) [32].

In this book, the chapter by Galla et al. (Section 2) presents the results of the first draft of the full genome sequence and assembly of a fresh salad plant leaf chicory (*Cichorium intybus* subsp. *intybus* var. *foliosum* L., $2n=2x=18$, and 1.3 Gbs genome size), named as Radicchio in Italian. The results of decoding the full genome of leaf chicory will “extend the current knowledge of the genome organization and gene composition of leaf chicory, which is crucial for developing new tools and diagnostic markers useful for our breeding strategies in Radicchio” and will be an important addendum to the list of sequenced plant genomes.

2.3. Sequencing “1001 genotypes” and “1000 plant species”

The availability of a few whole reference genomes limits our full understanding of ecotypic variations that affect the function and adaptive evolution of plant species in various climatic conditions. It reduces the power of genome-wide tagging of biologically meaningful natural variations. In other words, the general perceptions are that “a single reference genome is not enough” for plant biology to explain and understand the existing natural variations in particular plant species and its populations [33, 34]. It also limits the development of efficient tools for genome analyses. To address this, as mentioned above, the Arabidopsis plant research community has developed a vision of sequencing a larger number of Arabidopsis genotype accessions, including various ecotypic and experimental population samples. As of today, the “1001 genome sequencing project of Arabidopsis accessions” has completed the full genome sequencing of 1100 Arabidopsis accessions [33, 34] “to record the genetic variation in the entire genome of many strains of the reference plant *Arabidopsis thaliana*” and with the future objective to develop efficient genome analysis tools and software [33].

To understand the tree of life of the Plant Kingdom and study its evolutionary aspects in comparison to other life forms, the international multi-disciplinary consortium of “The 1000 Plants (oneKP or 1KP) Initiative” has generated a large-scale gene sequencing data for more than 1000 various plant species [35]. Rather than concentrating on single species accessions as in the 1001 Arabidopsis whole-genome sequencing project [33, 34], the “1KP” project targeted 1000 distinct plant species with the objective of generating only functionally expressed (i.e., transcriptome) gene sequences. The plant species selected for the project had no restriction, and the samples were “chosen to represent every species known to science, across the Plant Kingdom, at some phylogenetically or taxonomically defensible levels” [35]. The 1KP sample list consists of 1328 entries [51] broadly grouped by phylogenetically (angiosperm, non-flowering, and green algae species) and by application (agriculture, medicine, biochemistry, and extremophytes). Most of these species have been sequenced for the first time (Table 1).

To date, an average of 2000 Mbs transcriptome sequence data have been generated for these 1KP plant species using 28 Illumina Genome Analyzer next-generation DNA sequencing machines at the Beijing Genomics Institute (BGI-Shenzhen, China) [35]. Ultimately, the obtained genomic sequence data will be used to analyze the phylogenetic, taxonomic, and evolutionary relationships of plant species, to study plant speciation, and to determine the timing of gene duplications during speciation events [35, 52]. However, the biggest limitation

is associated with sequencing only transcriptomes rather targeting the whole genome, which limits obtaining many non-coding and repetitive portions of genomes. The results of “1001” and “1KP” sequencing efforts will undoubtedly open a new paradigm for plant genomics and its above-mentioned sub-disciplines. The results should not only accelerate crop improvement and boost the agricultural and medicine production worldwide but also help to understand the basics of plant life, evolution, speciation, and plant adaptations to the extreme environments in the era of global climate change and technological advancements.

Plant species	*Number of samples
Phylogenetic groups	
Angiosperms	830
Angiosperms: Onagraceae samples	50
Non-flowering plants	257
Green algae	241
Application groups	
Medicinal samples	142
Medicine - Alkaloid samples	30
Medicine - Chemotherapeutic samples	12
Biochemistry - Lipid Biosynthesis samples	15
Agriculture - C3/C4 samples	93
Agriculture - Weeds	25
Extremophyte samples	31
Halophytes samples	18

*The number of samples overlaps among groups. Source: Ref. [51].

Table 1. Plant species samples chosen for the “1KP” plant genome sequencing project.

In this book, we have presented several chapters targeting to review and discuss the strategies for sequencing and assembly challenges (by Deschamps and Llaca), new-generation sequencing platforms for comparative genomics of cereal crops (by Sikhakhane et al.) and non-model cactus plant Nopal (*Opuntia* spp.; by Alonso-Herrada et al.), and characterization of small RNA world of plant genomes (Hernández-Salazar et al.). These chapters describe the current advances and future needs on these topics.

3. Crop improvement in the genomics and post-genomics era

3.1. Genomics-assisted selection or genomic selection

At present, the reference genomes for many agricultural plants including specialty crops have been sequenced, as reviewed by Michael and VanBuren [32], which created a new paradigm for modern crop breeding. Crop breeding, which is powered and enriched by molecular

markers, genetic linkage maps, QTL mapping, association mapping, and marker-assisted selection methods in the past century [37, 53], has now greatly accelerated and become ever productive and efficient in the plant genomics era [26]. This is due to the (1) availability of large-scale transcriptome and whole-genome reference sequences [32]; (2) high-throughput SNP marker collection and cost-effective, automated, and high-throughput genotyping platforms (HTP) and technologies (e.g., genotyping by sequencing or GBS), allowing breeders to screen multiple genotypes within a short time [23, 26]; (3) identification and use of expression QTLs (genetical genomics) in breeding [22]; and (4) opportunity to perform genome-wide selection (i.e., genomic selection) [26].

The biggest driving force for genomics-assisted crop breeding in the plant genomics era has been the inexpensive sequencing and re-sequencing opportunity for population individuals of genetic crosses and breeding lines. This helps to precisely identify and link genetic variations to the phenotypic expressions, taking into account the rare and private allelic variations that are abundant in crop line population or germplasm resources [26, 53, 54]. Furthermore, the availability of SNP marker collections and automated genotyping platforms provided a better genome converge to perform genome-wide genotype-to-phenotype associations (GWAS) [11, 37]. Also, when whole-genome sequences are not available and SNP markers are present in a limited number, the breeders using GBS and HTS platforms can readily genotype their mapping population and can provide genomic selections for the targeted crops of interest [23, 26, 54]. Although it was first applied for animal breeding [55], recently genomic selection has been successfully applied to a number of plant species [56–62], including studies using GBS in the context of genomic selection [26]. Most importantly, the application of available genomics tools and a large number of high-throughput DNA markers and new-generation genotyping platforms have made the “breeding by design” [63] possible and have developed “virtual breeding” approaches [64] for efficient crop improvement. Several chapters in this book have covered the advances toward plant resistance genomics and molecular breeding against bacterial diseases in ryegrasses (see the chapter by Dr. Takahashi) as well as biotic/abiotic stress tolerance in agriculture crops (see the chapters by Onaga and Wydra, and Rao et al.).

The availability of genome sequences and a large number of SNP marker collections also provided the analysis of copy number variations (CNVs) in crop genomes, and their links to the key traits have greatly enhanced the crop improvement programs [11, 22, 23, 26, 37]. Furthermore, although challenges are evident, the opportunity provided by post-genome sequencing advances has help to integrate and enrich genomic selection with key proteome and metabolome markers. This significantly fostered and powered up the breeding of complex traits [22] of crops. Consequently, the knowledge gained through plant genomics coupled with proteomic and metabolomic advances has facilitated the emergence of an innovative approach of “personalized” agriculture through the utilization of chemical genomics [21]. This requires the translation of knowledge and expertise of the pharmaceutical industry on the development of “personalized medicine” to treat each person based on its reaction to the medical drugs into the agriculture. Because of high-throughput genome analysis, it is possible to date that many plant compounds, including herbicides, growth regulators and phytohormones, elicitors, low molecular metabolites (e.g., salicylic acids), and/or synthetic hybrid chemicals, can be screened for genetic response of individual crop genotypes and to study their mechanism of actions

contributing to agricultural productivity. Once identified, highly genotype-specific chemical compounds can be developed that impact better than traditionally applied “fit for all” chemicals/growth stimulators and fertilizers. A combination of such chemical genomics approach, proteomics and metabolomics with genetic engineering, and genomic selection will further provide a way for “personalized” agriculture that sustains crop production (for detailed discussions, see a review by Stokes and McCourt [21]).

3.2. Novel transgenomics tools and biotech crops

Crop improvement is also greatly impacted by novel transgenomics and genome editing technologies developed as a result of plant genome characterization and understanding in the era of plant genomics. In the past two decades, a variety of novel transgenomics technologies have been developed to replace or enrich the traditional transgenesis-based genetic engineering and plant molecular biotechnology [65]. These novel technologies include antisense, RNA interference (RNAi), artificial microRNA expression (amiR), virus-induced gene silencing (VIGS), zinc-finger nuclease (ZFN), transcription activator-like effects nucleases (TALENs), oligonucleotide-directed mutagenesis (ODM) of Cibus Rapid Trait Development System (RTDS), and clustered regularly interspaced short palindromic repeats/Cas9 (CRISPR/Cas9) technologies [65, 66]. These novel transgenomics technologies including genome-editing tools, the latter also referred to as genome editing with engineered nucleases (GEEN), are widely developed and utilized to investigate the gene function and apply to solve problems in medicine and agriculture. They have become methods of choice for major functional genomics and biotechnological studies [67]. RNA-mediated genome manipulation (RNAi) tools down-regulate the target genes due to gene silencing effects at transcriptional (TGS) or post-transcriptional (PTGS) levels, whereas GEEN systems help to insert, replace, or remove specific regions of DNA from a genome using artificially engineered nucleases that are referred to as “molecular scissors” [68–70]. For a detailed description of RNAi, readers are suggested to read a chapter by Ricano-Rodriguez et al. in this book as well as to the recently published “RNA Interference” book by InTech Open.

The potential application of RNA-mediated gene silencing methods for crop improvement, including RNAi in plant biotechnology, is huge and the technology has already generated many successful examples in a wide range of technical, food, and horticulture crops. For example, RNAi was used to improve crop yield, food/fiber quality [18, 71–75], resistance to pests, and biotic/abiotic stresses [76, 77], which are being considered for commercialization or are already in commercial production [78]. Employing ODM-mediated single nucleotide editing in *Arabidopsis*, targeting the BFP gene, has demonstrated a precise edition of CAC to TAC, converting histidine (H66) to tyrosine (Y66) in GFP protein that offered a non-transgenic breeding tool for crops [66]. Similarly, GEEN tools have also provided a new strategy for “trait stacking,” whereby several desired traits are physically linked to ensure their co-segregation during the breeding processes [79]. The examples include *A. thaliana* [80–82] and *Z. mays* [83], where ZFN-assisted gene targeting has helped to heritably insert herbicide-resistant genes (SuRA/SuRB and PAT) into the targeted sites in the genome [83]. Although other GEEN technologies such as TALEN [84–92] and CRISPR/Cas9 [93] are just picking its application in plants, their utilization in *Arabidopsis* [84], maize [85], rice [86–88], potato [89, 90], wheat [65],

barley [91], and plum [92] holds a great promise and potential for non-transgenic crop genome modification and improvement [65, 94].

4. Grand tasks ahead

The revolutionizing advances made in the past three decades in plant genomics and its sub-disciplines provided a mass of novel opportunities with easy-solution applications and high-throughput, cost-effective, and time-effective technologies. Plant genomics era increased our understanding of the basis of complex life processes/traits in plants and crop species, and it paved a way for effective improvement of plants to fulfill our diet and other needs. However, it also piled up challenging grand tasks ahead for current genomics and post-genomics era. Several chapters of this book have discussed some aspects of these challenges, and I tried to briefly summarize some of them here.

As mentioned above, tremendous achievements have been made toward sequencing more than hundreds of plant genomes including major crop species and specialty, model/non-model, wild, vascular, flowering, and polyploid plants [31, 32]. There are ongoing and fascinating consortia projects of sequencing “1001 genotypes of Arabidopsis” and “1000 various plant species” [33–35, 51, 52]. However, the first current and future task ahead is to extend such large-scale, multiple accession genome sequencing initiatives for each priority agricultural and specialty crop species including their wild relatives and ancestor-like genome representatives. Although it sounds largely ambitious, this task will be mandatory and important for the next plant genome sequencing phase. This is to effectively use all variations existing among plant/crop germplasm resources and its ecotypic populations and to design efficient GWAS analysis and consequent genomic selections as well as tools/software programs for better analyzing plant genomes and improving genome assembly issues [33–35]. This is especially needed for polyploidy crops [24, 32, 37] because the sequencing of many polyploids and their subgenomes would increase our understanding of the complexity of polyploidy, gene silencing, epigenetics, and biased retention and expression of genes after polyploidization [24, 95–97]. Furthermore, it also helps to discover all natural variations and lost genes during crop domestication that should be useful to restore the key agriculturally important traits in the future.

Sequencing the entire genome of 1KP samples, rather concentrating on only transcriptome/exome, is also the necessary task ahead that would elucidate many important noncoding sequences from these plant species. Results would be useful for plant evolutionary, speciation and taxonomy studies. There are ongoing planning and targets toward this goal, and it should not cause much trouble in the land of experiences gained and inexpensive high-throughput sequencing technologies [1, 27, 32].

Although high-throughput DNA sequencing instrumentation exists and keeps evolving to better versions year-to-year, the consequent task is still to improve the sequence length that would solve many incorrect sequence sites and genome assembly challenges that plant genomics faces currently [1, 32]. Some of the currently ongoing efforts and possible solution with the advent of third-generation sequencing platforms and genome assembly tools and methodologies highlighted herein have been discussed by several book chapters in this book.

A consequent grand task and challenge with the completion of the above-highlighted tasks is the handling, organizing, systematizing, and visualizing a huge amount of plant genome sequencing (“Big Data”) data that require urgent attention, effort, collaborative work, and investment. There is an urgent need to develop more efficient bioinformatics platforms to handle plant genome data due to challenges, specificities, complexities, and sizes of currently available and future sequenced plant genomes mentioned herein [1, 98]. Funding this aspect of plant genomics and bioinformatics research is a necessary key step [1] for future advances on this task ahead.

Furthermore, the most important current and future post-genomics grand task ahead is to link the sequence variation(s) with phenotype(s), trait expression, and epigenetic and adaptive features of plants to their living environment and extreme conditions. The successful completion of this task will require the combined approaches of genomics with bioinformatics, proteomics, metabolomics, phenomics, genomic selections, genetical genomics, reverse genomics, system biology, etc. [11, 21–29, 64, 65, 98]. In other words, there is a need to make sequenced genomes “functional” [31] and biologically meaningful [29, 37]. This also requires the integration of all available genomic and phenotypic data to identify key networks that also require downstream effort of integration of specific networks to networks of other systems in order to connect heterogeneous data [29]. There are suggested thoughts and tasks for plant genomics that should target to develop plant genome-specific “Encyclopedia of DNA Elements (ENCODE)” [31, 32], which will be an important achievement in the next phases of development. There is a need to use molecular phenotyping (i.e., using molecular process such as protein-RNA interactions, translation rates, etc.) in QTL mapping [23] that would help to precisely link the sequence variation(s) to its phenotype(s). There is a task for the development and translation of the concept “personalized agriculture” [21] that requires an attention as an unexplored area in crops with the availability of sequenced genomes and high-throughput genotype, proteome, metabolome, and phenotype profiling platforms and rapid crop line development tools such as genomic selection and new-generation genome-editing tools mentioned above. All these will help to minimize the current challenges with improved crop line development costs through efficient breeding [11, 22, 23, 26]. These particular grand tasks further highlight a need for extended effort and work on the development of inexpensive high-throughput plant phenotyping [25, 26] and plant proteome and metabolome profiling tools and instrumentation [27, 28] by utilizing small amount single-cell-derived samples [27–29].

A parallel grand task to the above-outlined needs is to have concentrated efforts on the timely application of novel transgenomics and genome-editing tools for all types of plants and to optimize it for routine large- and short-scale biotechnology industry usage. There are grandest tasks to (1) utilize the complex effects of plant developmental genes (e.g., core microRNA/RNAi machinery) to simultaneously improve the key traits and overcome negative trait correlations [15, 18] and (2) optimize and better design novel transgenomics and genome-editing technologies for the key priority crops and plant by-product production. In addition, there are needs to (3) identify the appropriate choice of plant tissues for genome editing, (4) reduce or eliminate side effects and off-target toxicity and mutagenesis of application of novel genome modification technologies, and (5) develop reliable screens for the detection of edited genome samples [99]. The revolutionizing effects of these novel genome-editing/manipulation

technologies and genome-edited organisms (GEOs) as well as their safer nature compared to conventional transgenesis are evident. However, without objective or proper regulatory policies, providing understanding and removing confusion of regulatory agencies and stakeholders [94], “these technologies may not live up to their full potential” [64] if they are regulated as genetically modified organisms bearing foreign genes [64, 94]. Therefore, this is one of the most important grand tasks ahead in the front of plant sciences research community in the era of plant genomics and post-genomics.

Finally, the grandest task is a preparation of well-qualified next-generation scientists capable of continuing plant genomics tasks highlighted herein with the understanding of conventional plant biology, ecology, plant breeding, evolution, taxonomy, modern “omics” disciplines, and cross-related scientific disciplines (e.g., mathematics, computing, and modeling) [1, 98]. Importantly, they are required to have a capability to utilize modern computing and instrumentation platforms and bioinformatics knowledge [29]. For instance, there is a huge need for a new generation of molecular breeders [100] with full knowledge and appreciation of conventional plant breeding aspects including the understanding of agrotechnology methodologies, genetic diversity of crop germplasm, and randomized multi-environmental field trails. These breeders also need to have abilities to handle, work, and utilize the sequenced genomes, high-throughput genotyping, and phenotyping platforms. This is a bottleneck for plant genomics at present, which requires urgent awareness, attention, and investment.

5. Conclusions

Thus, in the past three decades, plant genomics has evolved from the enrichment and advances made in conventional genetics and breeding, molecular biology, molecular genetics, molecular breeding, and molecular biotechnology in the land of high-throughput DNA sequencing technologies powering the plant research community to sequence and understand the genetic compositions, structures, architectures, and functions of full plant genomes. The technological and instrumentation advancements as well as the desire and need to feed the increasing human population, overcome biosecurity issues, and sustain agricultural production in the era of global climate change, the societal globalization, and technological advancements have been the main driving forces for plant genomics development. These led to sequence and assemble entire plant genomes including very complex polyploid plants, annotate gene functions, link the sequence variation(s) to the phenotype(s), and exploit sequence variation(s) in plant/crop improvement in genome-wide scale or through targeted native modification of plant genomes in a highly sequence-specific manner.

To date, more than 100 plant genomes including a large number of crops as well as flowering, non-flowering, crop wild relative, model and non-model, and specialty plants have been fully sequenced. As a result, it expanded our knowledge and understanding of many aspects of plant biology, genetics, breeding, and crop evolution and domestication, which contributed to the development of analytical and breeding tools, resulting in accelerated crop improvement programs. To look even deeper scales, more than 1100 *Arabidopsis* accessions from various

eco-geographic origin and experimental populations have been fully sequenced, which will equip plant researchers with better analysis tools and help in tagging and exploiting the biologically meaningful variations. Furthermore, transcriptome profiling of 1000 distinct plant species with agricultural, medicinal, biochemical, and evolutionary utilization has a great value and will be “a gold mining” opportunity for plant biology to explain the evolution of tree of life and Plant Kingdom speciation. All of these successes have significantly accelerated crop improvement using novel genomic selections and new-generation genome-editing and manipulation technologies.

These advances, briefly highlighted herein, also have generated a number of grand challenges and mandatory tasks ahead in plant genomics and post-genomics era. There are many tasks ahead for the plant genomics community, which require more collaborations, integrated approaches, better computing capacity and analytical tools, accelerated training and education of well-qualified researchers, and larger investments. In this book, the authors tried to highlight some updates on current plant genomics efforts with future perspectives. We trust that the next phase of plant genomics efforts and development will be more exciting and help to solve current and future issues in front of humanity.

Acknowledgements

Plant genomics research in Uzbekistan is being jointly funded by the basic science (FA-F5-T030), applied (FA-A6-T081 and FA-A6-T085), and innovation (I-2015-6-15/2 and I5-FQ-0-89-870) research grants of the Academy of Sciences of Uzbekistan and Committee for Coordination Science and Technology Development of Uzbekistan. I thank the Office of International Research Programs (OIRP) of the U.S. Department of Agriculture (USDA) Agricultural Research Service (ARS) and the U.S. Civilian Research & Development Foundation (CRDF) for international cooperative grants P120, P120A, P121, P121B, UZB-TA-31016, UZB-TA-31017, and UZB-TA-2992, which were devoted to cotton genomics, including cotton gene characterization, germplasm analysis, genetic mapping, plant disease resistance, marker-assisted selection, and biotechnology. I also thank Dr. Din-Pow Ma, Mississippi State University, for the critical reading of this introductory chapter.

Author details

Ibrokhim Y. Abdurakhmonov*

*Address all correspondence to: ibrokhim.abdurakhmonov@genomics.uz; genomics@uzsci.net

Center of Genomics and Bioinformatics, Academy of Sciences of Uzbekistan, Tashkent, Uzbekistan

References

- [1] Schatz MC, Witkowski J, McCombie WR. Current challenges in de novo plant genome sequencing and assembly. *Genome Biol.* 2012;13:243. DOI:10.1186/gb-2012-13-4-243
- [2] Kossmann J. Grand challenges in plant biotechnology. *Front Plant Sci.* 2012;3:61. DOI: 10.3389/fpls.2012.00061
- [3] Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S. Plant DNA barcoding: from gene to genome. *Biol Rev Camb Philos Soc.* 2015;90:157-66. DOI:10.1111/brv.12104.
- [4] Abdurakhmonov IY, editor. *Plant breeding.* Rijeka: InTech; 2012. 352 p. DOI: 10.5772/1389
- [5] United Nations Food and Agriculture Organization. *Dimensions of Need - An atlas of food and agriculture. Staple foods: What do people eat* [Internet]. 2015. Available from: <http://www.fao.org/docrep/u8480e/u8480e07.htm> [Accessed: 2015-11-10]
- [6] Falkowski P, Scholes RJ, Boyle E, Canadell J, Canfield D, Elser J, Gruber N, Hibbard K, Hogberg P, Linder S, Mackenzie FT, Moore B, Pedersen T, Rosenthal Y, Seitzinger S, Smetacek V, Steffen W. The global carbon cycle: a test of our knowledge of earth as a system. *Science.* 2000;290:291–296. DOI: 10.1126/science.290.5490.291
- [7] Pimentel D, Harvey C, Resosudarmo P, Sinclair K, Kurz D, McNair M, Crist S, Shpritz L, Fitton L, Saffouri R, Blair R. Environmental and economic costs of soil erosion and conservation benefits. *Science.* 1995;267:1117–1123. DOI: 10.1126/science.267.5201.1117
- [8] Fairley P. Introduction: Next generation biofuels. *Nature.* 2011;474:S2–5. DOI: 10.1038/474S02a
- [9] Mann J. Natural products in cancer chemotherapy: past, present and future. *Nat Rev Cancer.* 2002;2:143–148. doi: 10.1038/nrc723
- [10] Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, Carrington JC, Doerge RW, Colot V, Martienssen R. Role of transposable elements in heterochromatin and epigenetic control. *Nature.* 2004;430:471–476. DOI: 10.1038/nature02651
- [11] Prohens J. Plant Breeding: A Success Story to be Continued Thanks to the Advances in Genomics. *Front Plant Sci.* 2011;2:51. DOI:10.3389/fpls.2011.00051
- [12] Jain HK. *The Green Revolution: History, Impact and Future.* Houston: Studium Press LLC. 2010. 256 p.
- [13] Hubert BM, Rosegrant van Boekel MAJS, Ortiz R. The Future of Food: Scenarios for 2050. *Crop Sci.* 2010;50:S-33–50. DOI:10.2135/cropsci2009.09.0530

- [14] Higgins TJ. Time to modify the GM debate [Internet]. 2011. Available from: <http://theconversation.edu.au/time-to-modify-the-gm-debate-210>. [Accessed from 2015-11-10]
- [15] Abdurakhmonov IY. Role of genomic studies in boosting yield. In: Proceedings of International Cotton Advisory Board (ICAC); 20 September-4 October 2013; Cartagena. Colombia. p. 7-22
- [16] Sharma SB, You Y, Varaprasad RS. Biosecurity – an integral part of global food security strategy. *Indian Journal of Plant Protection*. 2008;2:165-172.
- [17] Selection for increased lint yield and correlated responses in Upland cotton, *Gossypium hirsutum* L. *Crop Sci*. 1967;7:637–640. DOI:10.2135/cropsci1967.0011183X000700060024x
- [18] Abdurakhmonov IY, Buriev ZT, Saha S, Jenkins JN, Abdukarimov A, Pepper AE. Cotton *PHYA1* RNAi enhances major fiber quality and agronomic traits of cotton (*Gossypium hirsutum* L). *Nature Communications*. 2014;4:3062; DOI:10.1038/ncomms4062
- [19] Saha S, Stelly DM, Raska DA, Wu J, Jenkins JN, McCarty JC, Makamov A, Gotmare V, Abdurakhmonov IY, Campbell BT. Chromosome substitution lines: concept, development and utilization in the genetic improvement of Upland cotton. In Abdurakhmonov IY, editor. *Plant Breeding*. Rijeka: InTech; 2012. p. 107-128. DOI: 10.5772/35585
- [20] Varma A, Shrivastava N. The role of plant genomics in biotechnology, In: Doelle, HW, DaSilva EJ, editors. *Biotechnology. Encyclopedia of Life Support Systems (EOLSS)*, Developed under the Auspices of the UNESCO. Oxford: Eolss Publishers. Available from: <http://www.eolss.net/sample-chapters/c17/e6-58-07-15.pdf> [Accessed: 2015-11-11]
- [21] Stokes ME, McCourt P. Towards personalized agriculture: what chemical genomics can bring to plant biotechnology. *Front Plant Sci*. 2014;5:344. DOI: 10.3389/fpls.2014.00344
- [22] Joosen RV, Ligterink W, Hilhorst HW, Keurentjes JJ. Advances in genetical genomics of plants. *Curr Genomics*. 2009;10:540-9. DOI: 10.2174/138920209789503914.
- [23] Jimenez-Goimez JM. Next generation quantitative genetics in plants. *Front Plant Sci*. 2011;2:77. DOI:10.3389/fpls.2011.00077
- [24] Song Q, Chen ZJ. Epigenetic and developmental regulation in plant polyploids. *Curr Opin Plant Biol*. 2015;24:101-9. DOI: 10.1016/j.pbi.2015.02.007.
- [25] Fahlgren N, Gehan MA, Baxter I. Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Curr Opin Plant Biol*. 2015;24:93-9. DOI:10.1016/j.pbi.2015.02.006

- [26] Poland J. Breeding-assisted genomics. *Curr Opin Plant Biol.* 2015;24:119-24. DOI: 10.1016/j.pbi.2015.02.009
- [27] Deal RB. Grand Challenge: Accelerating Discovery through Technology Development. *Front Plant Sci.* 2011;2:41. DOI: 10.3389/fpls.2011.00041
- [28] Heazlewood JL. The green proteome: challenges in plant proteomics. *Front Plant Sci.* 2011;2:6. DOI:10.3389/fpls.2011.00006
- [29] Fernie AR. Grand Challenges in Plant Systems Biology: Closing the Circle(s). *Front Plant Sci.* 2012;3:35. DOI:10.3389/fpls.2012.00035
- [30] PubMed database [Internet]. 2015. Available from: <http://www.ncbi.nlm.nih.gov/pubmed> [Accessed from 2015-09-06].
- [31] Michael TP, Jackson S. The First 50 Plant Genomes. *The Plant Genome.* 2013;6:1-7. DOI:10.3835/plantgenome2013.03.0001.
- [32] Michael TP, VanBuren R. Progress, challenges and the future of crop genomes. *Curr Opin Plant Biol.* 2015;24:71-81. DOI: 10.1016/j.pbi.2015.02.002
- [33] 1001 genomes: A Catalog of *Arabidopsis thaliana* Genetic Variation [Internet]. Available from: <http://1001genomes.org/> [Accessed; 2015-11-10]
- [34] Weigel D, Mott R. The 1001 genomes project for *Arabidopsis thaliana*. *Genome Biol.* 2009;10:107. DOI: 10.1186/gb-2009-10-5-107
- [35] Leebens-Mack J. OneKp Capstone Wiki [Internet]. 2015. Available from: <https://pods.ipplantcollaborative.org/wiki/display/iptol/OneKP±Capstone±Wiki> [Accessed: 2015-11-10]
- [36] Healey A, Furtado A, Cooper T, Henry RJ. Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods.* 2014;10:21. DOI: 10.1186/1746-4811-10-21
- [37] Morrell PL, Buckler ES, Ross-Ibarra J. Crop genomics: advances and applications. *Nat Rev Genet.* 2011;13:85-96. DOI: 10.1038/nrg3097
- [38] Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol.* 1975;94:441-8. DOI: 10.1258/jrsm.99.2.62
- [39] Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci U SA.* 1977;74:560-4. DOI:10.1073/pnas.74.2.560
- [40] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences USA.* 1977;74:5463-7. DOI: 10.1073/pnas.74.12.5463
- [41] Prober JM, Trainor GL, Dam RJ, Hobbs FW, Robertson CW, Zagursky RJ, Cocuzza AJ, Jensen MA, Baumeister K. A system for rapid DNA sequencing with fluorescent

- chain-terminating dideoxynucleotides. *Science*. 1987;238:336-41. DOI:10.1126/science.2443975
- [42] Briggs SP. Plant genomics: more than food for thought. *Proceedings of the National Academy of Sciences USA*. 1998;95:1986-8. DOI:10.1073/pnas.95.5.1986
- [43] Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA. Genome-wide location and function of DNA binding proteins. *Science*. 2000;290:2306-9. DOI:10.1126/science.290.5500.2306
- [44] Morozova O, Marra MA. Applications of next-generation sequencing technologies in functional genomics. *Genomics*. 2008;92:255-64. DOI:10.1016/j.ygeno.2008.07.001
- [45] Leushkin EV, Sutormin RA, Nabieva ER, Penin AA, Kondrashov AS, Logacheva MD. The miniature genome of a carnivorous plant *Genlisea aurea* contains a low number of genes and short non-coding sequences. *BMC Genomics*. 2013;14:476. DOI: 10.1186/1471-2164-14-476
- [46] Ibarra-Laclette E, Lyons E, Hernández-Guzmán G, Pérez-Torres CA, Carretero-Paulet L, Chang TH, Lan T, Welch AJ, Juárez MJ, Simpson J, Fernández-Cortés A, Arteaga-Vázquez M, Góngora-Castillo E, Acevedo-Hernández G, Schuster SC, Himmelbauer H, Minoche AE, Xu S, Lynch M, Oropeza-Aburto A, Cervantes-Pérez SA, de Jesús Ortega-Estrada M, Cervantes-Luevano JI, Michael TP, Mockler T, Bryant D, Herrera-Estrella A, Albert VA, Herrera-Estrella L. Architecture and evolution of a minute plant genome. *Nature*. 2013;498:94-8. DOI: 10.1038/nature12132
- [47] Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, Vicedomini R, Sahlin K, Sherwood E, Elfstrand M, Gramzow L, Holmberg K, Hällman J, Keech O, Klasson L, Koriabine M, Kucukoglu M, Käller M, Luthman J, Lysholm F, Niittylä T, Olson A, Rilakovic N, Ritland C, Rosselló JA, Sena J, Svensson T, Talavera-López C, Theißen G, Tuominen H, Vanneste K, Wu ZQ, Zhang B, Zerbe P, Arvestad L, Bhalerao R, Bohlmann J, Bousquet J, Garcia Gil R, Hvidsten TR, de Jong P, MacKay J, Morgante M, Ritland K, Sundberg B, Thompson SL, Van de Peer Y, Andersson B, Nilsson O, Ingvarsson PK, Lundeberg J, Jansson S. The Norway spruce genome sequence and conifer genome evolution. *Nature*. 2013;497:579-84. DOI: 10.1038/nature12211
- [48] Birol I, Raymond A, Jackman SD, Pleasance S, Coope R, Taylor GA, Yuen MM, Keeling CI, Brand D, Vandervalk BP, Kirk H, Pandoh P, Moore RA, Zhao Y, Mungall AJ, Jaquish B, Yanchuk A, Ritland C, Boyle B, Bousquet J, Ritland K, Mackay J, Bohlmann J, Jones SJ. Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. *Bioinformatics*. 2013;29:1492-7. DOI: 10.1093/bioinformatics/btt178
- [49] Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D, Crepeau MW, Cardeno C, Koriabine M, Holtz-Morris AE, Liechty JD, Martínez-García PJ, Vasquez-Gross HA, Lin BY, Zieve JJ, Dougherty WM, Fuentes-Soriano S, Wu LS, Gilbert D, Marçais G,

- Roberts M, Holt C, Yandell M, Davis JM, Smith KE, Dean JF, Lorenz WW, Whetten RW, Sederoff R, Wheeler N, McGuire PE, Main D, Loopstra CA, Mockaitis K, deJong PJ, Yorke JA, Salzberg SL, Langley CH. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biol.* 2014;15:R59. DOI: 10.1186/gb-2014-15-3-r59
- [50] International Wheat Genome Sequencing Consortium (IWGSC). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science.* 2014;345:1251788. DOI: 10.1126/science.1251788
- [51] List of sequenced plant samples [Internet]. 2015. Available from: <http://www.onekp.com/samples/list.php> [Accessed: 2015-11-10]
- [52] Matasci N, Hung LH, Yan Z, Carpenter EJ, Wickett NJ, Mirarab S, Nguyen N, Warnow T, Ayyampalayam S, Barker M, Burleigh JG, Gitzendanner MA, Wafula E, Der JP, dePamphilis CW, Roure B, Philippe H, Ruhfel BR, Miles NW, Graham SW, Mathews S, Surek B, Melkonian M, Soltis DE, Soltis PS, Rothfels C, Pokorny L, Shaw JA, DeGironimo L, Stevenson DW, Villarreal JC, Chen T, Kutchan TM, Rolf M, Baucom RS, Deyholos MK, Samudrala R, Tian Z, Wu X, Sun X, Zhang Y, Wang J, Leebens-Mack J, Wong GK. Data access for the 1,000 Plants (1KP) project. *Gigascience.* 2014;3:17. DOI: 10.1186/2047-217X-3-17
- [53] Abdurakhmonov IY, Abdurakarimov A. Application of association mapping to understanding the genetic diversity of plant germplasm resources. *International Journal of Plant Genomics.* 2008;574927. DOI:10.1155/2008/574927
- [54] Kumpatla SP, Buyyarapu R, Abdurakhmonov IY, Mammadov JA. Genomics-assisted plant breeding in the 21st century: technological advances and progress. In: Abdurakhmonov IY, editor. *Plant Breeding.* Rijeka: InTech; 2012. p 134-184. DOI: 10.5772/37458
- [55] Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 2001;157:1819-29
- [56] Cros D, Denis M, Bouvet JM, Sánchez L. Long-term genomic selection for heterosis without dominance in multiplicative traits: case study of bunch production in oil palm. *BMC Genomics.* 2015;16:651. DOI: 10.1186/s12864-015-1866-9
- [57] Longin CF, Mi X, Würschum T. Genomic selection in wheat: optimum allocation of test resources and comparison of breeding strategies for line and hybrid breeding. *Theor Appl Genet.* 2015;128:1297-306. DOI:10.1007/s00122-015-2505-1
- [58] Iwata H, Ebana K, Uga Y, Hayashi T. Genomic prediction of biological shape: elliptic Fourier analysis and kernel partial least squares (PLS) regression applied to grain shape prediction in rice (*Oryza sativa* L.). *PLoS One.* 2015;10:e0120610. DOI: 10.1371/journal.pone.0120610
- [59] Spindel J, Begum H, Akdemir D, Virk P, Collard B, Redoña E, Atlin G, Jannink JL, McCouch SR. Genomic selection and association mapping in rice (*Oryza sativa*): ef-

- fect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genet.* 2015;1:e1004982. DOI: 10.1371/journal.pgen.1004982
- [60] Cros D, Denis M, Sánchez L, Cochard B, Flori A, Durand-Gasselín T, Nouy B, Omoré A, Pomiès V, Riou V, Suryana E, Bouvet JM. Genomic selection prediction accuracy in a perennial crop: case study of oil palm (*Elaeis guineensis* Jacq.). *Theor Appl Genet.* 2015;128:397–410. DOI: 10.1007/s00122-014-2439-z
- [61] Beaulieu J, Doerksen TK, MacKay J, Rainville A, Bousquet J. Genomic selection accuracies within and between environments and small breeding groups in white spruce. *BMC Genomics.* 2014;15:1048. DOI: 10.1186/1471-2164-15-1048
- [62] Lipka AE, Lu F, Cherney JH, Buckler ES, Casler MD, Costich DE. Accelerating the switchgrass (*Panicum virgatum* L.) breeding cycle using genomic selection approaches. *PLoS One.* 2014;9:e112227. DOI: 10.1371/journal.pone.0112227
- [63] Peleman JD, van der Voort JR. Breeding by design. *Trends Plant Sci.* 2003;8:330-4. DOI:10.1016/S1360-1385(03)00134-1
- [64] Andersen SB. Virtual breeding. In: Abdurakhmonov IY, editor. *Plant Breeding*. Rijeka: InTech; 2012. p. 107-128. DOI: 10.5772/29802
- [65] Ricroch AE, Henard-Damave MC. Next biotech plants: new traits, crops, developers and technologies for addressing global challenges. *Crit Rev Biotechnol.* 2015: Early online:1-16. DOI:10.3109/07388551.2015.1004521
- [66] Sauer NJ, Mozoruk J, Miller RB, Warburg ZJ, Walker KA, Beetham PR, Schöpke CR, Gocal GF. Oligonucleotide-directed mutagenesis for precision gene editing. *Plant Biotechnol J.* 2015 (in press). DOI: 10.1111/pbi.12496
- [67] McMahon MA, Rahdar M, Porteus M. Gene editing: not just for translation anymore. *Nat Meth.* 2011;9:28-31. DOI:10.1038/nmeth.1811
- [68] Esvelt, KM.; Wang, HH. Genome-scale engineering for systems and synthetic biology. *Mol Syst Biol.* 2014;9:641. DOI:10.1038/msb.2012.66
- [69] Tan WS; Carlson DF; Walton MW, Fahrenkrug SC; Hackett PB. Precision editing of large animal genomes. *Adv Genet* 2012;80:37–97. DOI:10.1016/B978-0-12-404742-6.00002-8
- [70] Puchta H, Fauser F. Gene targeting in plants: 25 years later. *Int. J. Dev. Biol* 2013;57:629–637. DOI:10.1387/ijdb.130194hp
- [71] Saurabh S, Vidyarthi AS, Prasad D. RNA interference: concept to reality in crop improvement. *Planta.* 2014;239: 543–564. DOI:10.1007/s00425-013-2019-5
- [72] Sunilkumar G, Campbell L, Puckhaber L; Stipanovic R, Rathore K. Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. *Proc Natl Acad Sci USA.* 2006;103: 18054–9. DOI:10.1073/pnas.0605389103

- [73] Siritunga D, Sayre R. Generation of cyanogen-free transgenic cassava. *Planta*. 2003;217:367–73. DOI:10.1007/s00425-003-1005-8
- [74] Le L, Lorenz Y, Scheurer S, Fötisch K, Enrique E, Bartra J, Biemelt S, Vieths S, Sonnewald U. Design of tomato fruits with reduced allergenicity by dsRNAi-mediated inhibition of ns-LTP (Lyc e 3) expression. *Plant Biotechnol J*. 2006;4:231–42. DOI: 10.1111/j.1467-7652.2005.00175.x
- [75] Niggeweg R, Michael A, Martin C. Engineering plants with increased levels of the antioxidant chlorogenic acid. *Nat Biotechnol*. 2004;22:746–54. DOI:10.1038/nbt966
- [76] Zadeh A, Foster G. Transgenic resistance to tobacco ringspot virus. *Acta Virol*. 2004;48:145–152
- [77] Mao YB, Cai WJ, Wang JW, Hong GJ, Tao XY, Wang LJ, Huang YP, Chen XY. Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. *Nat Biotechnol*. 2007;25:1307–1313. DOI: 10.1038/nbt1352
- [78] Heinemann JA, Agapito-Tenfen SZ, Carman JA. A comparative evaluation of the regulation of GM crops or products containing dsRNA and suggested improvements to risk assessments. *Environ Int*. 2013;55:43–55. DOI: 10.1016/j.envint.2013.02.010
- [79] Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet*. 2010;11:636–646. DOI:10.1038/nrg2842
- [80] Townsend JA, Wright DA, Winfrey RJ, Fu F, Maeder ML, Joung JK, Voytas DF. High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature*. 2009;459:442–445. DOI:10.1038/nature07845
- [81] Zhang F, Maeder ML, Unger-Wallace E, Hoshaw JP, Reyon D, Christian M, Li X, Pierick CJ, Dobbs D, Peterson T, Joung JK, Voytas DF. High frequency targeted mutagenesis in *Arabidopsis thaliana* using zinc finger nucleases. *Proceedings of the National Academy of Sciences USA*. 2010;107:12028–33. doi: 10.1073/pnas.0914991107
- [82] Osakabe K, Osakabe Y, Toki S. Site-directed mutagenesis in *Arabidopsis* using custom-designed zinc finger nucleases. *Proceedings of the National Academy of Sciences USA*. 2010;107:12034–9. DOI: 10.1073/pnas
- [83] Shukla VK, Doyon Y, Miller JC, DeKolver RC, Moehle EA, Worden SE, Mitchell JC, Arnold NL, Gopalan S, Meng X, Choi VM, Rock JM, Wu YY, Katibah GE, Zhifang G, McCaskill D, Simpson MA, Blakeslee B, Greenwalt SA, Butler HJ, Hinkley SJ, Zhang L, Rebar EJ, Gregory PD, Urnov FD. Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature*. 2009;459:437–41. DOI: 10.1038/nature07992
- [84] Forner J, Pfeiffer A, Langenecker T, Manavella PA, Lohmann JU. Germline-transmitted genome editing in *Arabidopsis thaliana* Using TAL-effector-nucleases. *PLoS One*. 2015;10:e0121056. DOI: 10.1371/journal.pone.0121056

- [85] Char SN, Unger-Wallace E, Frame B, Briggs SA, Main M, Spalding MH, Vollbrecht E, Wang K, Yang B. Heritable site-specific mutagenesis using TALENs in maize. *Plant Biotechnol J*. 2015;13:1002-10. DOI: 10.1111/pbi.12344
- [86] Zhang H, Gou F, Zhang J, Liu W, Li Q, Mao Y, Botella JR, Zhu JK. TALEN-mediated targeted mutagenesis produces a large variety of heritable mutations in rice. *Plant Biotechnol J*. 2015 (in press). DOI: 10.1111/pbi.12372
- [87] Wang M, Liu Y, Zhang C, Liu J, Liu X, Wang L, Wang W, Chen H, Wei C, Ye X, Li X, Tu J. Gene editing by co-transformation of TALEN and chimeric RNA/DNA oligonucleotides on the rice OsEPSPS gene and the inheritance of mutations. *PLoS One*. 2015;10:e0122755. DOI: 10.1371/journal.pone.0122755
- [88] Shan Q, Zhang Y, Chen K, Zhang K, Gao C. Creation of fragrant rice by targeted knockout of the OsBADH2 gene using TALEN technology. *Plant Biotechnol J*. 2015;13:791-800. DOI: 10.1111/pbi.12312
- [89] Nicolia A, Proux-Wéra E, Åhman I, Onkokesung N, Andersson M, Andreasson E, Zhu LH. Targeted gene mutation in tetraploid potato through transient TALEN expression in protoplasts. *J Biotechnol*. 2015;204:17-24. DOI:10.1016/j.jbiotec.2015.03.021
- [90] Clasen BM, Stoddard TJ, Luo S, Demorest ZL, Li J, Cedrone F, Tibebu R, Davison S, Ray EE, Daulhac A, Coffman A, Yabandith A, Retterath A, Haun W, Baltes NJ, Mathis L, Voytas DF, Zhang F. Improving cold storage and processing traits in potato through targeted gene knockout. *Plant Biotechnol J*. 2015 (in press). DOI: 10.1111/pbi.12370
- [91] Budhagatapalli N, Rutten T, Gurushidze M, Kumlehn J, Hensel G. Targeted Modification of Gene Function Exploiting Homology-Directed Repair of TALEN-Mediated Double-Strand Breaks in Barley. *G3 (Bethesda)*. 2015;5:1857-63. DOI: 10.1534/g3.115.018762
- [92] Ilardi V, Tavazza M. Biotechnological strategies and tools for Plum pox virus resistance: trans-, intra-, cis-genesis, and beyond. *Front Plant Sci*. 2015;6:379. DOI: 10.3389/fpls.2015.00379
- [93] Jain M. Function genomics of abiotic stress tolerance in plants: a CRISPR approach. *Front Plant Sci*. 2015;6:375. DOI: 10.3389/fpls.2015.00375
- [94] Wolt JD, Wang K, Yang B. The Regulatory Status of Genome-edited Crops. *Plant Biotechnol J*. 2015 (in press). DOI: 10.1111/pbi.12444
- [95] Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, Wendel JF. Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*). *Genetics*. 2009;182:503-17. DOI:10.1534/genetics.109.102608
- [96] Renny-Byfield S, Gallagher JP, Grover CE, Szadkowski E, Page JT, Udall JA, Wang X, Paterson AH, Wendel JF. Ancient gene duplicates in *Gossypium* (cotton) exhibit

- near-complete expression divergence. *Genome Biol Evol.* 2014;6:559-71. DOI: 10.1093/gbe/evu037
- [97] Yoo MJ, Szadkowski E, Wendel JF. Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity (Edinb).* 2013;110:171-80. DOI: 10.1038/hdy.2012.94
- [98] Sinha NR. Plant developmental biology in the post-genomic era. *Front Plant Sci.* 2011;2:11. DOI:10.3389/fpls.2011.00011
- [99] Puchta, H. & Hohn, B., Breaking news: Plants mutate right on target. *Proc Natl Acad Sci U S A.* 2010;107:11657-8. DOI: 10.1073/pnas
- [100] Moose SP, Mumm RH. Molecular plant breeding as the foundation for 21st century-crop improvement. *Plant Physiol.* 2008;147:969-77. DOI:10.1104/pp.108.118232

