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

5,800
Open access books available

142,000
International authors and editors

180M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

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

Sequencing of Non-model Plants for Understanding the Physiological Responses in Plants

Jannette Alonso-Herrada, Ismael Urrutia, Tania Escobar-Feregrino, Porfirio Gutiérrez-Martínez, Ana Angélica Feregrino-Pérez, Irineo Torres-Pacheco, Ramón G. Guevara González, Sergio Casas-Flores and Andrés Cruz-Hernández

Additional information is available at the end of the chapter

Abstract

From a genomic point of view, plants are complex organisms. Plants adapt to the environment, by developing different physiological and genetic properties, changing their genomic and expression profiles of adaptive factors, as exemplified by polyploidy studies. These characteristics along with the presence of duplicated genes/genomes make sequencing with early low-throughput DNA sequencing technologies in plants a challenging task. With the development of new technologies for molecular analysis, including transcriptome, proteome or microarray profiling, a new perspective in the genomic analysis was open, making possible to programs in species without genomic maps. The opportunity to extend molecular studies from laboratory model scale toward naturally occurring plant populations made it possible to precisely answer the longstanding important ecological and evolutionary questions. Some plant species have unique properties that could help to understand their adaptability to environment, crop production, pest protection or other biological processes. Molecular studies on non-model plants, including algae, mosses, ferns and plants with very specific characteristics are ongoing.

Keywords: Genome size, NGS, polyploidy, transcriptome, wild materials

1. Introduction

The first wave of plant genome sequencing has passed, and now new era has started in plant genomics research with new-generation sequence (NGS) strategies, which require a mixture of economic and scientific needs. Until now, several crops have been sequenced and some

others crop´s sequencing is underway, which will greatly help to elucidate unknown biological processes and the phylogenetic relationship among crop plants. Furthermore, the genomic data analysis and its integration of the biological systems will help to establish fundamental models to understand the evolution, development, and adaptability of the plants.

A genomic sequence is an important information for the basic research and understanding of plant evolution and development. It serves as a tool for engineering new genotypes [1]. Different plant species have different amounts of DNA [2]. The DNA of most plants includes from 100 million to the largest example of 150 billion base pairs (designed as alphabets), organized into 20,000 to 50,000 genes [1]. The most important contribution in the field of plant genome analysis is the discovery that many higher plants share a blueprint gene content. As distantly related plant taxa, monocots and dicots, which diverged from a common ancestor about 200 million years ago, retain some common gene order along the genome [1].

The development of new strategies and technologies for genome sequencing, can lead to development of programs to get a partial (transcriptome) or a complete DNA sequence (whole genome sequencing (WGS)) for non-model plants. The costs for these projects are now accessible. The first plant genome sequencing project, represented an effort of several years and millions of dollars. Now, the costs for the sequencing are in the order of thousands of dollars and there are new bioinformatics tools available for the analysis of generated sequence data.

The world's population depends on a few crops such as rice, wheat, maize, and potato for their food. In the following decades the world will face the tremendous challenge of feeding the global population [3]. The study of plant genomics in non-model plants will help to reveal the genetic factors and biochemical pathways involved in many processes such as flowering, nutrition, disease, and pest resistance, as well as tolerance of plants to abiotic stresses.

Model organisms are important for biological and agricultural approaches. The research in model organisms has generated a huge amount of important information on different molecular factors that contribute to plant growth and development, however it has some limitations [4]. The study of wild plants will help to overcome these limitations. Wild plants are well adapted to extreme conditions, and resistant to plant pathogens.

Model organisms, have a limited number of uniform narrow-based genotypic samples or variability limited to a number of specific plants. The study on how some plants survive in extreme conditions may also provide some clues about the mechanisms of plant response to biotic or abiotic stresses. Some non-model organisms are an extraordinary source of plant secondary metabolites.

The wild plant genotypes sometimes does not look attractive for breeding programs because of their morphology; however, they are the repository of ancestral genes and very important sources for the rescue of specific traits.

In Mexico, there are a large number of wild plant populations fit for breeding or sequencing programs are yet to be identified. Wild plant populations are genetically diverse and are source of genes that encode proteins potentially used for health, industrial, or ecological purposes.

2. Non model plants as model for environmental adaptation

The cactus plants of Cactaceae family are an example for a plant that can adapt to several environmental conditions. One of these plants is Nopal (*Opuntia* spp.), which belongs to the genera Opuntia and Nopalea [5]. This is an endemic plant found in semiarid areas in Mexico, but it grows along the American continent, from Canada to La Patagonia in Argentina, where environmental conditions are different from each area. Recently, Nopal plants have become the world´s interesting alternative fruit and forage crop. Only few varieties of nopal fruit originated from the Mexican nopal germplasm, are available in the market.

The history of first Nopal use in Mexico dates back to the ancient Mesoamerican civilizations; people used to collect cladodes and fruits from wild materials, for their nutritional qualities and medicinal purposes. The Spanish conquerors spread Nopal in America and Europe; now it is cultivated in Italy, Morocco, Tunisia, Greece, Israel, India, Philippines, China, Australia, South Africa, Brazil, Argentina, Colombia and the United States [6-8].

Although nopal is propagated asexually for commercial purposes, seed propagation is essential for breeding. Nopal apomixis makes the screening of individual crops obtained from crosses difficult and complicates the genetic studies [5]. Although no genomic map exists for this multipurpose plant, several efforts have been made to get some genomic approaches, and it has been included in the 1000 genomes sequencing program. To date, extensive efforts on cDNA microarrays, microRNAs (miRNAs) microarrays, mRNA deep sequencing and molecular markers studies have been made.

To study the genes associated with crassulacean acid metabolism (CAM), an expressed sequence tag (EST) database of different developmental stages of various tissues was created [9]. Sequences were assembled and compared with the available plant and genetic databases; genes involved in circadian regulation and CAM were identified in plants grown under a long day regime. Three kinds of expression profiles were found: transcripts oscillated with a 24-h periodicity; transcripts of the light-active genes adapted to cycles of 12-h periodicity; and

arrhythmic accumulation patterns. Some genes were scored best to a 12-h rhythm, suggesting a difference with Arabidopsis at level of circadian clock gene interactions. The results indicate that changes at the CAM metabolism are the result of modified circadian regulation at the transcriptional and posttranscriptional levels [9].

In addition, the gene regulation trough miRNAs has been explored [10]. miRNAs are a class of small non-coding RNAs that regulate gene expression. A combination of Northern blot and tissue print hybridization was used, to identify conserved miRNAs expressed during nopal (*Opuntia ficus indica*) fruit development. A comparative analysis detected 34 miRNAs expressed differentially. These miRNA were clustered different groups and associated with the different phases of fruit development. Gradual expression of several miRNAs was observed during fruit development. The work provided the evidence of miRNA expression in the cactus fruit and the basis for future research on miRNAs in Opuntia [10].

One transcendental work is related to the analysis of genomic content in 23 Opuntia species by flow cytometry [11]. A main interest on Opuntia genomes was related to the DNA content

because; almost all the genotypes have a ploidy level of 4x, 8x or 12x; of their genetic complexity. In four different ploidy levels having 2C-DNA amounts, DNA content ranged from 3.75 Giga base pairs (Gb) (*Opuntia incarnadilla* Griffiths) to 5.87 Giga base pairs (Gb) (*Opuntia heliabravoana* Scheinvar) among the samples analyzed.

The 2C DNA content when compared with other species; such as maize, shows that genome of Opuntia is less complex than that of maize (Table 1), which makes Opuntia suitable for a genomic sequencing program.

Common name	Scientific name	Family	Ploidy level	2C genome size (pg)	2C genome size (Gb)
Arabidopsis	*Arabidopsis thaliana*	Brassicaceae	2x	0.30	0.29
Soy bean	*Glycine max*	Leguminosae	2x	2.31	2.25
Maíze	*Zea mays*	Gramineae	2x	13.49	13.19
Tuna charola	*Opuntia streptacantha*	Cactaceae	8x	4.64	4.53
Tuna	*Opuntia megacantha*	Cactaceae	8x	5.01	4.89
Tuna blanca	*Opuntia ficus indica*	Cactaceae	8x	4.90	4.79
Tuna robusta	*Opuntia robusta*	Cactaceae	8x	4.98	4.87
Xoconoxtle	*Opuntia joconostle*	Cactaceae	8x	4.7	4.59

Table 1. Opuntia DNA content and other plant species

The nopal and its products need a more deeply analysis to maximize the real value of this crop. It is a multipurpose plant that is very important in the life of the people because it impact on the economy, nutrition, medicinal practices, and fuel production. Two main aspects are now in the focus for increasing its crop value:

1. Some crops have been sequenced and some others are in progress, however the nopal is waiting to be sequenced. Once sequenced, it would help to understand several mechanisms of plant adaptation to different environments, and will give us clues about controlling the process for adaptation to extreme conditions in other plants [12].

2. A new important aspect involves miRNAs, which are thought to be fine-tuning mechanisms in gene regulation [10]. Wrong expression of miRNAs can produce pleiotropic effects on development. It would be no surprise to discover that several events related to plant adaptation were under the control of miRNA expression. In the future, the expression of miRNA and siRNA will serve as tools for the generation of new *Opuntia* phenotypes. In these experiments the role of different molecules or pathways involved in seed formation, ripening delay or fruit development could be revealed.

3. Non model plants as source of industrial solutions

Development of modern society has led to an increased emission of pollutants into the environment, from industrial and domestic activities, as well as from mining, agriculture and crafting [13]. These compounds are a threat to all the organisms; therefore, numerous methods have been developed to reduce the impact caused by pollution. Conventional methods for the removal of pollutants in soil and water are often costly and can irreversibly affect the properties of the soil, as well as the organisms that inhabit those places [14].

Bioremediation is a tool used to clean pollutants in soil and water, and it is referred to the chemical transformation of pollutants through the use of microorganisms and plants [15]. The genomic content of plants for remediation has been calculated by different methods, and the sizes are included in Table 2. As shown in Table 2, there are no genomic complex organisms and some have been sequenced already.

It is important to consider that there is a great diversity of plants grow under different climates, which belong to different families. This allows researchers to have a wide variety of candidate plants that fits the scientific needs. Some plants of the Asteraceae, Brassicaceae and Solanaceae families have been found as tolerant to different pollutants. According to Lopez et al. [15], plants use a mechanism to alleviate the environmental stresses, by the following three phases:

(1) Absorption, excretion and detoxification of pollutants; (2) the distribution of pollutants throughout the plant and their excretion via volatilization; and (3) detoxification of pollutants by phytoremediation, by any one of the following processes: phytoextraction, rhizofiltration, phytoestimulation, phytostabilization, phytovolatilization or phytodegradation [15,16].

Phytostabilization allow to reduce the bioavailability and mobility of contaminants, avoiding underground transport layers or the atmosphere [15,16]. This process is less expensive than other methods, is easy to apply and aesthetically pleasing.

Phytodegradation is the transformation of organic pollutants in simpler molecules. In certain instances, degradation products will serve to accelerate plant growth, and other cases the contaminants are biotransformed. The phytodegradation has been employed for the removal of explosives, such as TNT, halogenated hydrocarbons, Bis-phenol, PAHs and organochlorine and organophosphorus pesticides [14].

In phytovolatilization, plants absorb water along with the soluble organic and inorganic pollutants (As, Se and Hg). Some of the contaminants can reach the leaves and get evaporated or volatilized into the atmosphere. Plants such as *Bigelovii Salicornia*, *Brassica juncea*, *Astragalus bisulcatus* and *Chara canescens* have been used for bioremediation of Se pollution and *Arabidopsis thaliana* has been used for bioremediation of Hg [14].

Rhizofiltration uses plants to remove contaminants from water environment through the root. In rhizofiltration, these plants are grown in hydroponic way. When the root system is well developed, the plants are introduced into polluted water with metals, where the roots absorb and accumulate. Numerous aquatic plants have the ability to accumulate pollutants, and some examples of these are as follows: *Scirpus lacustris*, *Lemna gibba*, *Azolla caroliniana*, *Elatine*

trianda, *Wolffia papulifera*, *Polygonum punctatum*, *Myriophylhum aquaticum*, and *Mentha palustris* (for Al, As, Au, Cd, Cr, Cu, Cr, Fe, Hg, Mg, Mn, Ni, Pb, Se, Sr, Zn,) [14,15].

Function	Species	Family	1C (Gb)	1C (pg)	Sequencing year	Reference
Phytoestabilization Pb, Zn, Cd, As, Cu, Mn	*Hordeum vulgare*	Gramineae	5.1	5.5	2012	International. Barley Genome Consortium [17]
Mercury Phytovolatilization	*Arabidopsis thaliana*	Cruciferae	0.125	0.16	2000	Arabidopsis Genome Initiative [18]
Phytoextraction Cd, Zn, Pb, Ni, Ag, Cr, Cu, Hg,	*Brassica juncea*	Cruciferae	1.49	1.092	--	Johnston et al. [19].
	Helianthus annuus	Compositae	3.5	2.43	2012	Staton et al. [20]
	Brassica napus	Cruciferae	1.12	1.15	2014	Boulos et al. [21]
Petroleum Contaminants Degradation	*Sorghum bicolor*	Gramineae	1.68	0.835	2009	Paterson et al. [22]
	Medicago sativa	Leguminosae	1.75	0.86	2011	Young et al. [23]
Elimination Cd, Pd, Zn, Cu, Ni, Cr	*Brassica nigra*	Cruciferae	0.632	0.647	--	Johnston et al. [19]
	Helianthus annuus	Compositae	3.5	2.43	2012	Staton et al. [20]
Insecticide Accumulation	*Cucumis sp*	Cucurbitaceae	0.68	0.66	2009	Huang et al [24]
	Cucurbita sp	Cucurbitaceae	0.34	0.33	--	Šisko et al. [25]
Phytoextraction Zn, HgNO$_3$	*Helianthus annuus*	Compositae	3.5	2.43	2012	Staton et al. [20]

Table 2. DNA content in Plants used for bioremediation

Phytoextraction or absorption is carried out by the plant roots and accumulation of polluting metals in the stems and leaves. Some plants used for this approach are: *Thlaspi caerulescens*; *Sedum alfredii*, *Viola* and *Vertiveria baoshanensis*; *Alyssum murale*, *Trifolium nigriscens*, *Psychotria douarrei*, *Pruinosa geissois*, *Homalium guillainii*, *Hybanthus floribundus*, *Sebertia acuminata*, *Stackhousia tryonii*, *Pimelea leptospermoides*, *Aeollanthus biformifolius*; *Haumaniastrum robertii*; *Brassica juncea, Helianthus annuus, Sesbania drummondii* and *Brassica napus* (for Ag; Cd, Cr, Cu, Hg, Ni, Pb, and Zn) [14,15].

Phytodegradation in plants and microorganisms is associated with, degradation of organic pollutants into harmless products and, mineralization into CO_2 and H_2O. Plants such as *Populus* spp. are introduced to absorb the contaminants in soil pores and prevent leaking to other soil layers [15,26].

4. Perspectives for genome sequencing and genome information from non-model plants to plant breeding

Next-generation sequencing (NGS), include several and different technologies which has its own set of characteristics. NGS generates huge amounts of sequence data in a very cost-effective way [27].

The increased number of WGS projects means that more organisms, are becoming important genetic models [28]. At the same time, many molecular studies are focusing on natural variation and adaptation in classical genetic model species, or close relatives of these, such as *Arabidopsis*, thus closing the gap between model and non-model organisms.

For example, the assembled genomes could be used as a reference sequence for further transcriptome analysis or re-sequencing and surveys of genetic variation. They may also be used to develop other genomic tools, such as proteomics and microarrays hybridization [29].

After the novel transcriptome has been annotated using a genomic reference species, it can be used as a starting point for more detailed functional characterizations of desired organisms, using gene ontology databases.

With RNA-seq protocols, or longer sequence reads will also improve applications because large haplotype blocks including several linked polymorphisms will become available. Wherein hundreds of genes are analyzed simultaneously. Some of these may be involved in important phenotypic variation, and this is relevant from the conservation point of view because such variation may be important to maintain within the population.

In the future, the bottleneck is more likely to be at the bioinformatics rather than in producing the sequences [30] because a huge number of biologists are trying to order

the genomic data with biological sense. New approaches for data storage and processing will be needed, because currently available databases might be unable to cope up with the rapid generation of new sequencing data [31].

5. Conclusions and prospects

Plants provide food for all living organisms, and just 15 crop plants provide 90% of the world's food intake [32]. Plant species are responsible for maintaining the balance of the carbon cycle, for developing and maintaining soil from erosion, and plant products are used as human medicines [33, 34]. For these reasons, there is great interest in sequencing plant genomes, but

so far relatively few plant species have been sequenced compared with the hundreds of thousands of species around the world.

Non-model plants are becoming very attractive sources for different purposes, for their ability to adaptation to extreme environments and to produce specific metabolites that can be used for food and medicinal purposes. The materials must be characterized at molecular level to develop any strategy for the generation of genetic data, that is molecular markers, cDNA sequencing, and cDNA microarrays, to have reference data to compare with model organisms.

Large complex plant genomes remain a particularly difficult challenge for *de novo* assembly for various biological, bioinformatics, and biomolecular reasons. Plant genomes can be nearly 100 times larger than the sequenced mammalian genomes [35]. The next frontier for plant genomics is to characterize the diversity of genomic variations across large populations, deeply annotate their functional elements, and develop predictive quantitative models relating genotype to phenotype.

Acknowledgements

Author details

Jannette Alonso-Herrada[1], Ismael Urrutia[1], Tania Escobar-Feregrino[2],
Porfirio Gutiérrez-Martínez[4], Ana Angélica Feregrino-Pérez[1], Irineo Torres-Pacheco[1],
Ramón G. Guevara González[1], Sergio Casas-Flores[5] and Andrés Cruz-Hernández[1,3*]

*Address all correspondence to: andrex1998@hotmail.com

1 Engineering Faculty, Autonomus University of Querétaro, Circuito Universitario, Cerro de las Campanas s/n, C.P. Santiago de Querétaro, Querétaro, México

2 Natural Sciences Faculty, Autonomus University of Querétaro, Circuito Universitario, Cerro de las Campanas s/n, C.P. Santiago de Querétaro, Querétaro, México

3 Chemistry Faculty, Autonomus University of Querétaro, Circuito Universitario, Cerro de las Campanas s/n, C.P. Santiago de Querétaro, Querétaro, México

4 Technological Institute of Tepic. Food Science Postgrade, Integral Laboratory on Food Science and Biotechnology Research. C.P. Tepic, Nayarit, México

5 IPICYT, Molecular Biology Division, C.P. 78216, San Luis Potosí, México

References

[1] Paterson A, Richardson C. Toward a cactus genome project: utilizing information from well studied model plants to map *Opuntia* chromosomes. J Prof Assoc Cactus Dev.1996;1:106–109.

[2] Aramuganathan K, Earle ED. Nuclear DNA content of some important plant species. Plant Mol Biol Rep.1991;9:208–218. DOI: doi.org/10.1007/BF02672069

[3] Food and Agriculture Organization. Our priorities: The FAO strategic objectives. Version 1.1–4. Food and Agriculture Organization of the United Nations. Rome. 2013. www.fao.org/docrep/018/mi317e/mi317e.pdf

[4] de Magalháes JP. The big, the bad and the ugly. EMBO Rep. 2015;16:771-776. DOI: 10.15252/embr.201540606

[5] Mondragón–Jacobo, C. Verification of the apomictic origin of cactus pear (*Opuntia* spp. *Cactaceae*) seedlings of open pollinated and crosses from central Mexico. J Prof Assoc Cactus Dev. 2001;3:49–56. DOI: 10.1016/S0140-1963(5)80016-3

[6] Flores–Valdés C, de Luna JM, Ramírez P.P. Nopalito world market. ASERCA–UACh–CIESTAAM. Chapingo, México. 1995; pp. 16–75.

[7] Barbera G, Inglese P, Pimienta–Barrios E. Agroecology, Cultivation and Uses of Cactus Pear. FAO Plant Production and Protection. Food and Agriculture Organization of the United Nations. Rome. 1995; Paper 132.

[8] Pimienta–Barrios E. Prickly pear (*Opuntia* spp.): a valuable fruit crop for the semiarid lands of México. J Arid Environ. 1994;28:1–11. DOI: 10.1016/S0140-1963(05)80016-3

[9] Mallona I, Egea-Cortines M, Weiss J. Conserved and divergent rhythms of crassulacean acid metabolism-related and core clock gene expression in the cactus *Opuntia ficus-indica*. Plant Physiol; 2011;156:1978–1989. DOI: 10.1104/pp.111.179275

[10] Rosas-Cárdenas FF, Caballero-Pérez J, Gutiérrez-Ramos X, Marsch-Martínez N, Cruz-Hernández A, de Folter S. miRNA expression during prickly pear cactus fruit development. Planta. 2014. DOI: 10.1007/s00425-014-2193-0

[11] Segura S, Scheinvar L, Olalde G, Leblanc O, Filardo S, Muratalla A, Gallegos C, Flores C. Genome sizes and ploidy levels in Mexican cactus pear species *Opuntia* (Tourn.) Mill. series *Streptacanthae* Britton et Rose, *Leucotrichae* DC., *Heliabravoanae* Scheinvar and *Robustae* Britton et Rose. Genet Res Crop Evol. 2007;54:1033–1041. DOI: 10.1007/s10722-006-9196-z

[12] Jackson S, Rounsley S, Purugganan M. Comparative sequencing of plant genomes: choices to make. Plant Cell. 2006;18:1100–1104. DOI: 10,1105/tpc.106,042192

[13] Wei S, Zhou Q, Saha UK. Hyperaccumulative characteristics of weed species to heavy metals. Water Air Soil Pollut. 2008;192:173–181. DOI: 10.1007/ s11270-008-9644-9

[14] Delgadillo LAE, González RCA, Prieto GF, Villagómez IJR, Acevedo SO. Fitorremediation: an alternative to eliminate contamination. Tropical Subtropical Agroecosyst. 2011;14:597- 612

[15] López MS, Gallegos MM, Pérez FL, Gutiérrez RM. Mechanisms of fitorremediation of contaminated soils with organic xenobiotic molecules. Rev Int Contam Ambient. 2005; 2:91-100.

[16] Cartman EP, Crossman TL. *In situ* Treatment Technology. Chapter 9: Phytoremediation. Second Edition. Lewis Publishers. USA. 2001

[17] The International Barley Genome Sequencing Consortium. A physical, genetic and functional sequence assembly of the barley genome. Nature. 2012:491:711-717. DOI: 10.1038/nature11543

[18] The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis Thaliana*. Nature. 2000;408:796-815. DOI: 10.1038/35048692

[19] Johnston JS, Pepper AE, Hall AE, Chen ZJ, Hodnet G, Drabek J, Lopez, Price J. Evolution of Genome Size in Brassicaceae. Ann Bot. 2005;95:229-235. DOI: 10.1093/aob/ mci016

[20] Staton SE, Bakken BH, Blackman BK, Chapman MA, Kane NC, Tang SU, Mark C, Knapp SJ, Rieseberg LH, Burke JM. The sunflower (*Helianthus annuus* L.) genome reflects a recent history of biased accumulation of transposable elements. Plant J. 2012;72:142–153. DOI: 10.1111/j.1365-313X.2012.05072.x

[21] Boulos C, Denoeud F, Liu S, Parkin I, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Corréa M, Da Silva C, Just J, Falentin C, Koh C, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao M, Edger P, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnel N, Le Paslier M, Fan G, Renault V, Bayer P, Golicz A, Manoli S, Lee T, Thi V, Chalabi S, Hu Q, Fan C, Tollenaere R, Lu Y, Battail C, Shen J, Sidebottom C, Wang X, Canaguier A, Chauveau A, Bérard A, Guan M, Liu Z, Sun F, Lim Y, Lyons E, Town E, Bancroft I, Wang X, Meng J, Ma J, Pires J, King G, Brunel D, Delourme R, Renard M, Aury J, Adams K, Batley J, Snowdon R, Tost J, Edwards D, Zhou Y, Hua W, Sharpe A, Paterson A, Guan C, Wincker P. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science. 2014;345:950-953. DOI: 10.1126/ science.1253435

[22] Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Mahe CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang

L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Maye K, Messing J, Rokhsar DS. The *Sorghum bicolor* genome and the diversification of grasses. Nature. 2009;457:551-556. DOI:10.1038/nature07723

[23] Young N, Debelle F, Oldroyd GE, Geurts R, Cannon S, Udvardi M, Benedito V, Mayer K, Gouzy J, Schoof H, Van de Peer Y, Proost S, Cook D, Meyers B, Spannagl M, Cheung F, De Mita S, Krishnakumar V, Gundlach H, Zhou S, Mudge J, Bharti A, Murray J, Naoumkina M, Rosen B, Silverstein K, Tang H, Rombauts S, Zhao P, Zhou P, Barbe V, Bardou P, Bechner M, Bellec A, Berger A, Berges H, Bidwell S, Bisseling T, Choisne N, Couloux A, Denny R, Deshpande S, Dai X, Doyle J, Dudez A, Farmer A, Fouteau S, Franken C, Gibelin C, Gish J, Goldstein S, González A, Green P, Hallab A, Hartog M, Hua A, Humphray S, Jeong D, Jing Y, Jocker A, Kenton S, Kim D, Klee K, Lai H, Lang C, Lin S, Macmil S, Magdelenat G, Matthews L, McCorrison J, Monaghan E, Jeong-HwanMun, Najar F, Nicholson C, Noirot C, O'Bleness M, Paule C, Poulain J, Prion F, Qin B, Qu C, Retzel E, Riddle C, Sallet E, Samain S, Samson N, Sanders I, Saurat O, Scarpelli C, Schiex T, Segurens B, Severin A, Sherrier D, Shi R, Sims S, Singer S, Sinharoy S, Sterck L, Viollet A, Wang B, Wang K, Wang M, Wang X, Warfsmann J, Weissenbach J, White D, White J, Wiley G, Wincker P, Xing Y, Limei Y, Yao Z, Ying F, Zhai J, Zhou L, Zuber A, Denarie J, Dixon R, May G, Schwartz D, Rogers J, Quetier F, Town C, Roe B. Medicago genome provides insight into the evolution of rhizobial symbioses. Nature. 2011;480: 796-810. DOI: 10.1038/nature10625

[24] Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas W, Wang X, Xie B, Ni P, Ren Y, Zhu H, Li J, Lin K, Jin W, Fei Z, Li G, Staub J, Kilian A, van der Vossen E, Wu Y, Guo J, He J, Jia Z, Ren Y, Tian G, Lu Y, Ruan J, Qian W, Wang M, Huang Q, Li B, Xuan Z, Cao J, Asan, Wu Z, Zhang J, Cai Q, Bai Y, Zhao B, Han Y, Li Y, Li X, Wang S, Shi Q, Liu S, Cho W, Kim J, Xu Y, Heller-Uszynska K, Mia H, Cheng Z, Zhang S, Wu J, Yang Y, Kang H, Li M, Liang H, Ren X, Shi Z, Wen M, Jian M, Yang H, Zhang G, Yang Z, Chen R, Liu S, Li J, Ma L, Liu H, Zhou Y, Zhao J, Fang X, Li G, Fang L, Li Y, Liu D, Zheng H, Zhang Y, Qin N, Li Z, Yang G, Yang S, Bolund L, Kristiansen K, Zheng H, Li S, Zhang X, Yang H, Wang J, Sun R, Zhang B, Jiang S, Wang J, Du Y, Li S. The genome of the cucumber, *Cucumis sativus* L. Nature Gen. 2009; 41:1275-1283. DOI: 10.1038/ng.475

[25] Šisko M, Ivancic A, Bohanec B. Genome size analysis in the genus *Cucurbita* and its use for determination of interspecific hybrids obtained using the embryo-rescue technique. Plant Sci. 2003;165: 663–669. DOI: 10.1016/S0168-9452(03)00256-5

[26] Jankaitė A, Vasarevičius S. Use of *Poaceae f.* species to decontaminate soil from heavy metals. Ekologija. 2007;53:84–89.

[27] Gilad Y, Pritchard JK, Thornton K. Characterizing natural variation using next-generation sequencing technologies. Trends Genet. 2009;25:463–471. DOI: 10.1016/j.tig.2009.09.003

[28] Mitchell-Olds T, Feder M, Wray G. Evolutionary and ecological functional genomics. Heredity. 2008;100:101–102. DOI: 10.1038/sj.hdy.6801015

[29] Harr B, Turner LM. Genome-wide analysis of alternative splicing evolution among Mus subspecies. Mol Ecol. 2010;19(s1):228–239.DOI: 10.1111/j.1365-294X.2009.04490.x

[30] Schuster SC. Next-generation sequencing transforms todays biology. Nat Meth. 2008;5:16–18.DOI: 10.1038/nmeth1156

[31] Ekblom R, Galindo J. Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity. 2011;107:1–15. DOI: 10.1038/hdy.2010.152

[32] United Nations Food and Agriculture Organization: Dimensions of Need - An atlas of food and agriculture. Staple foods: What do people eat. Available from http://www.fao.org/docrep/u8480e/u8480e07.htm

[33] 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

[34] Mann J. Natural products in cancer chemotherapy: past, present and future. Nat Rev Cancer. 2002;2:143-148. DOI: 10.1038/nrc723

[35] Schatz S, Witkowski J, McCombie R. Current challenges in *de novo* plant genome sequencing and assembly. Genome Biol. 2012;13:243-249. DOI: 10.1186/gb-2012-13-4-243