

REVIEW

doi: 10.5455/medarh.2017.71.288-292

MED ARCH. 2017 AUG; 71(4): 288-292

RECEIVED: JUN 12, 2017 | ACCEPTED: AUG 04, 2017

¹Department of Genetics and Bioengineering, International Burch University, Sarajevo, Bosnia and Herzegovina

²Verlab Ltd, Sarajevo, Bosnia and Herzegovina

³Faculty of Agriculture and Food Science, University of Sarajevo, Bosnia and Herzegovina

⁴Technical Faculty Bihac, University of Bihac, Bosnia and Herzegovina

Corresponding author: Jasin Hodzic. Department of Genetics and Bioengineering, International Burch University, Sarajevo, Bosnia and Herzegovina. E-mail: jasin.hodzic@stu.ibu.edu.ba.

Overview of Next-generation Sequencing Platforms Used in Published Draft Plant Genomes in Light of Genotypization of Immortelle Plant (*Helichrysum Arenarium*)

Jasin Hodzic¹, Lejla Gurbeta^{1,2}, Enisa Omanovic-Miklicanin^{1,3}, Almir Badnjevic^{1,2,4}

ABSTRACT

Introduction: Major advancements in DNA sequencing methods introduced in the first decade of the new millennium initiated a rapid expansion of sequencing studies, which yielded a tremendous amount of DNA sequence data, including whole sequenced genomes of various species, including plants. A set of novel sequencing platforms, often collectively named as “next-generation sequencing” (NGS) completely transformed the life sciences, by allowing extensive throughput, while greatly reducing the necessary time, labor and cost of any sequencing endeavor. **Purpose:** of this paper is to present an overview NGS platforms used to produce the current compendium of published draft genomes of various plants, namely the Roche/454, ABI/SOLiD, and Solexa/Illumina, and to determine the most frequently used platform for the whole genome sequencing of plants in light of genotypization of immortal plant. **Materials and methods:** 45 papers were selected (with 47 presented plant genome draft sequences), and utilized sequencing techniques and NGS platforms (Roche/454, ABI/SOLiD and Illumina/Solexa) in selected papers were determined. Subsequently, frequency of usage of each platform or combination of platforms was calculated. **Results:** Illumina/Solexa platforms are by used either as sole sequencing tool in 40.42% of published genomes, or in combination with other platforms - additional 48.94% of published genomes, followed by Roche/454 platforms, used in combination with traditional Sanger sequencing method (10.64%), and never as a sole tool. ABI/SOLiD was only used in combination with Illumina/Solexa and Roche/454 in 4.25% of publications. **Conclusions:** Illumina/Solexa platforms are by far most preferred by researchers, most probably due to most affordable sequencing costs. Taking into consideration the current economic situation in the Balkans region, Illumina/Solexa is the best (if not the only) platform choice if the sequencing of immortal plant (*Helichrysum arenarium*) is to be performed by the researchers in this region.

Keywords: plant genome, high-throughput nucleotide sequencing, base sequence.

1. INTRODUCTION

The beginning of the new millennium was marked with yet another achievement in the field of genetics, specifically plant genomics - the sequencing of *Arabidopsis thaliana*. While this breakthrough set foundations to numerous developments in plant genomics and fostered better understanding of the plant genome, the methodology of sequencing remained a major setback to this field of scientific research. Namely, the sequence was obtained through the conventional Sanger method, which is rather simple, but requires extensive time, labor and finances when sequencing a whole genome. It would remain as such until 2004, when 454 Life Sciences marketed a

paralleled version of pyrosequencing which revolutionized sequencing technology, albeit still requiring previous amplification of the sample (1).

Year 2006 was marked by emergence of Illumina, a company which introduced a sequencing-by-synthesis approach that is even today a staple of whole genome sequencing (2). All sequencing approaches that followed the Sanger method are contemporarily named as next-generation sequencing (NGS), and are known for allowing extensive throughput, while greatly reducing the necessary time, labor and cost of any sequencing endeavor (3).

NGS indeed triggered a revolution in life sciences, labeled by numerous publications of genomes of dif-

© 2017 Jasin Hodzic, Lejla Gurbeta, Enisa Omanovic-Miklicanin, Almir Badnjevic

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ferent organisms, despite the fact that NGS had several drawbacks such as large genome size, high CG content, emergence of homopolymers etc. Different strategies for overcoming these obstacles have been eventually developed, various sequencing methods introduced or further refined, creating a scientific trend that continues as we speak. These high-throughput sequencing methods fostered numerous valuable discoveries in fields of molecular and evolutionary biology, medicine, forensics, agriculture etc. When it comes to application of NGS to plant species, the broadest and most prominent is whole genome sequencing (WGS) in quest for revealing the full sequence of plant genomes, their genetic make-up as well as genetic background of desirable traits in agricultural production. WGS is especially useful for producing draft genomes of plants that are sequenced for the first time (3). The researchers are predominantly interested in staple cereal, vegetable and fruit species, for obvious economic gain such research may potentially bring. However, many species of interest have incredibly large and complex genomes which make *de novo* sequencing of the whole genome labor-intensive and occasionally completely impracticable. Due to this, strategies such as sequencing of only one or several chromosomes of interest, transcriptome sequencing, or exome sequencing are some of the alternatives researchers embrace in quest of specific plant genome data (4). Nevertheless, currently there is several dozens of published draft plant genomes, including a number of plants of great agricultural and industrial significance (Table 1).

Common name	Species	Genome size (Mb)	Year	Reference
Bread wheat	<i>Triticum aestivum</i>	17,000	2014	[28]
Canola	<i>Brassica rapa</i>	516	2011	[52]
Cassava	<i>Manihot esculenta</i>	770	2012	[31]
Cotton	<i>Gossypium raimondii</i>	880	2012	[26]
Flax	<i>Linum usitatissimum</i>	373	2012	[22]
Potato	<i>Solanum tuberosum</i>	844	2011	[21]
Sugar beet	<i>Beta vulgaris</i>	758	2014	[33]
Tobacco	<i>Nicotiana tabacum</i>	4,500	2014	[47]

Table 1. Published Draft Genomes of several Agriculturally and Industrially Significant Plants

Apart from direct use of various cultivars for production of food and beverages, animal feed, fabric, ropes etc., pharmaceuticals and cosmetics are significant areas of industrial production that involve numerous plants whose genomics have not been yet elucidated. One such plant is immortelle (*Helichrysum arenarium*), a perennial plant widely spread on the Adriatic coast of Croatia, Bosnian and Herzegovina and Montenegro (5, 6, 7). Due to high market value of essential oil of immortelle, and low investments required for its production, agricultural production of this plant has expanded rapidly in the aforementioned regions, especially in in-land Herzegovina, where producers were able to achieve average annual revenues of 19,715.58 BAM (approximately 10,000 EUR) per 1 ha of immortelle (8). Elucidating the genome of this plant in quest of better understanding of its de-

sirable traits for agricultural production would undoubtedly have a great impact on the economy of the region.

This paper focuses on comparing and contrasting three commercial technologies most commonly applied for WGS of plants, namely the Roche/454, ABI/SOLiD, and Solexa/Illumina, and presenting the applications areas of each in terms of WGS of plant genomes in light of genotypization of immortelle plant.

2. MATERIALS AND METHODS

This article has descriptive character and present systematic review of literature focusing on sequencing technologies utilized in sequencing of plant genomes. For the purpose of this study, we examined the published draft genomes of plants. Since publicly available data only cover researches published up to year 2014, any more recent research was not included in the study. Criteria for paper selection were: a) the paper must introduce a draft genome of a plant species; b) the paper must be in English; c) sequencing method utilized is disclosed in a freely available form; d) sequencing method is either Roche/454, ABI/SOLiD, Illumina/Solexa or any combination of those (combinations with traditional Sanger method were also included); e) the results must indicate the size of obtained draft genome.

Platform	Roche/454	ABI/SOLiD	Solexa/Illumina
Sequencing mechanism	Pyrosequencing	Sequencing by ligation	Sequencing by synthesis
Read length	700 bp	75 bp	2x 101 bp
Accuracy	99.9%	99.94%	98%
Reads	1 M	1.2-1.4 G	3 G
Output data	0.7 Gb	120 Gb	600 Gb
Run time	24 h	14 d	3-10 d
Major advantage	Read length and speed	Accuracy	High throughput
Major disadvantage	Error rate	Short read assembly	Short read assembly

Table 2. Technical properties of the three sequencing platforms

Platform	Roche/454	ABI/SOLiD	Solexa/Illumina
Instrument price	Instrument 500,000 USD, 7000 USD per run	Instrument 495,000 USD, 15,000/100 Gb USD	Instrument 690,000 USD, 6000 USD/(30x) human genome
Automation of library preparation	Available	Available	Available
Cost/million bases	10 USD	0.13 USD	0.07 USD

Table 3. Cost of the three sequencing platforms

Applying these criteria, 45 papers (9-53) were selected (with 47 presented plant genome draft sequences), and frequency of application of each NGS platform determined. The papers and corresponding plant genomes are listed in Appendix 1. The frequencies were subsequently brought in correlation with technical characteristics and cost of sequencing, in order to deduce the best approach to sequencing immortelle (*Helichrysum arenarium*). For

comparison purposes technical properties (Table 2) and sequencing cost data (Table 3) obtained from Liu, et al. (2) were used. Most frequently used platforms or platform combinations were then analyzed in terms of size of genomes deduced through their application.

3. RESULTS

Papers included in the study were firstly examined in terms of utilized NGS platform (or combination of platforms, including combining with traditional Sanger method), as shown in Table 4 and Figure 1 below.

Platform	Number of papers	Number of genomes	Percentage of total number of genomes
Illumina/Solexa	18	19	40.42%
Combination of Sanger sequencing, Roche/454 and Illumina/Solexa	10	10	21.28%
Combination of Sanger sequencing and Illumina/Solexa	5	6	12.77%
Combination of Sanger sequencing and Roche/454	5	5	10.64%
Combination of Roche/454 and Illumina/Solexa	5	5	10.64%
Combination of Roche/454, ABI/SOLiD and Illumina/Solexa	2	2	4.25%

Table 4. Frequency of used NGS platforms among draft genomes of plants published until 2014

Illumina/Solexa is by far most frequently used platform (with different models of the technology) either as a sole NGS platform used (in 19 studies, or 40.42%), or in combination with other platforms and techniques, such as Roche/454, ABI/SOLiD and Sanger sequencing (for a total of 42 draft plant genomes published using Illumina/Solexa platform to some extent). Roche/454 has been utilized in combination with Sanger sequencing,

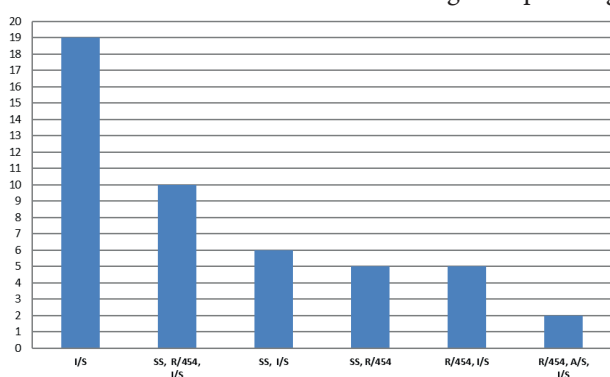


Figure 1. Most frequently used NGS platforms, or platform combinations (including combinations with Sanger sequencing) in draft plant genome publications until 2014. Illumina/Solexa (I/S) has been solely utilized for sequencing the largest number of genomes, a total of 19. (I/S=Illumina/Solexa; R/454=Roche/454; A/S=ABI/SOLiD; SS=Sanger sequencing)

and other two NGS platforms in 25.53% of the published plant genome sequencing projects. Interestingly, none of the examined papers involves sole usage of Roche/454 or

Genome size (Mb)	I/S	SS, R/454 and I/S	SS and I/S	SS and R/454	R/454 and I/S	R/454, A/S and I/S
0-99		1				
100-199					1	
200-299	1				1	1
300-399	3	1	3			
400-499	3	3				
500-599	1	1				
600-699	1				1	
700-799	1	1		2		
800-899	2	1	1			
900-999		1			1	
1,000-1,999		1	1	2		
2,000-2,999	3					1
3,000-3,999	2			1		
4,000-4,999	2				1	
17,000	1					

Table 5. Sizes of plant genomes published using different NGS platforms (or combinations of platforms)

ABI/SOLiD platforms. ABI/SOLiD was only used in two publications, in combination with Roche/454 and Illumina/Solexa, and has by far the lowest overall inclusion in plant genome sequencing projects, 4.25%.

The results demonstrate that the combining of different platforms is a very common practice, as 59.58% of published studies where the sequence was obtained through combining several (up to three) different sequencing approaches. Combining the platforms appears

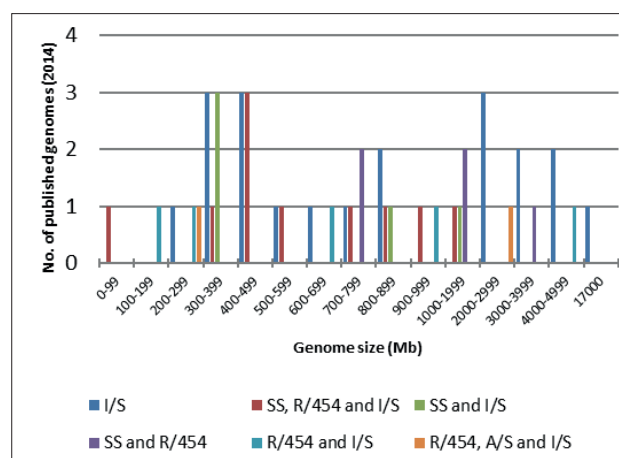


Figure 2. Size and number of genomes obtained using different NGS platforms. Illumina/Solexa (I/S) has been solely utilized for sequencing the broadest range of genome sizes

to be due to the practical and infrastructural reasons, and employed on wide range of genome sizes, however, it is undoubtedly helpful in mitigating some of the setbacks that each platform on its own has. Although the focus of this study is on the NGS platforms, it is important to point out that we can still find a significant portion of sequencing projects that involve traditional Sanger sequencing approach. We have found that a total of 21 studies utilize Sanger sequencing (and always in combination with NGS platforms). Accordingly, 44.69% of published draft genomes of plant have been elucidated using Sanger sequencing to some extent.

Genome size is another parameter taken into consideration when selecting sequencing techniques for a research. Table 5 and Figure 2 present the most frequently used NGS platforms (or platform combinations) in terms of size of the genomes deduced through their usage.

Figure 2 clearly demonstrates that the size of the genome, although an important consideration when assembling the sequence (54), is not effective to the choice of the sequencing platform by researchers. Illumina/Solexa platform (marked dark blue in Figure 2), being the most frequently used, is also used on the broadest range of genome sizes, from as low as 200-299 megabases to massive 17,000-megabase genome of bread wheat. Next in the line of broadness of application in terms of the size of sequenced genomes is the combination of traditional Sanger sequencing, Roche/454 and Illumina/Solexa, with the range of 0-99 to 1000-1999 Mb. Overall, no particular trend can be detected when comparing the size of the genome and sequencing platform researchers opt for.

Since genome size of *Helichrysum arenarium* has not been reported yet in the literature as of the end of 2016, we will rely on the 2014 study of several other species from the *Helichrysum* genus, published by Azizi et al. Since species displayed various degrees of polyploidy, genome size range for genus *Helichrysum* can be, based on the available data, confined to roughly 8,000-18,000 Mb. Based on this estimation and the results of our study of the NGS platform usage in plant genome sequencing, it is clear that Illumina/Solexa is the most viable choice, especially if we consider the financial weight of such sequencing project. If we refer back to the Table 3, the calculated price range for the estimated *Helichrysum arenarium* genome size range is 560-1,260 USD for a single run using Illumina/Solexa platform, 1,040-2,340 USD using ABI/SOLiD and 80,000-180,000 USD using Roche/454. However, additional run costs including the necessary chemicals, utensils, labor etc. must be taken into consideration. Although the results of this paper also indicate that combining of different platforms is a widespread practice, making aforementioned price estimations somewhat incomplete, it does not take away from the fact that sole usage of Illumina/Solexa is by far the most affordable approach. Another fact that adds to such a conclusion is that Illumina/Solexa is the most widespread sequencing platform, meaning that the majority of research institutions most likely possess a device manufactured by this company. Additionally, this makes Illumina/Solexa platforms most readily reachable to the

researchers that do not have a sequencer at disposal in institutions they are affiliated with.

4. DISCUSSION

Examination of technical properties can be helpful for proper selection of a sequencing platform. Since longer reads are preferable for accurate assembling and for interpreting repetitive sequences, such as those of many plants of interest, the Sanger method would be the most suitable. However, it is avoided due to high cost, time and labor requirements. Sanger sequencing is often used in combination with NGS platforms, for library sequencing, sequencing of genome portions that are improperly sequenced by NGS, subsequent proofreading of certain genome portions etc. Turktas, et al. suggest that Roche/454 technology, offering the longest read-length capacity and highest speed among NGS platforms, appears as the method of choice for plant WGS without considering the total sequencing cost (3). This study shows, however, that this is not the case, and that Illumina/Solexa platforms are by far most preferred by researchers, either on their own, or in combination with other platforms. In fact, Roche/454 has been used only in combination with Sanger sequencing on 5 projects, and never as a sole sequencing tool, presumably due to the high cost. Another indicator of the cost being the key consideration is the fact that researchers favored Illumina/Solexa platforms in a widest range of genome sizes (Table 5, Figure 2). The failure rates of Illumina/Solexa platforms are generally compensated for by their deep coverage, albeit it is not sufficient for avoiding gap generation when repetitive sequence is longer than the read length. Shatz, et al. suggest using paired-end sequencing (54), and Illumina/Solexa platforms are all capable of performing it. This may be another indicator why researchers favor Illumina/Solexa to the extent presented in this study. Additionally, it has been reported on several occasions that in recent years the Illumina sequencing platform has been the most successful platform in terms of market share and widespreadness, to the point of near monopoly, which only adds to the fact that it is the most popular choice of scientists performing the plant genome sequencing (55, 56). Although Azizi et al. reported genome sizes of several species of genus *Helichrysum* (56), there are no exact data regarding the size of *H. arenarium* genome. Hence, we relied on estimated size range when calculating the estimated price range of a single sequencing run using Illumina/Solexa platform. Clearly, karyotype and genome size analysis for *H. arenarium* would provide valuable information prior to the sequencing and reduce the risk of project failure due to financial reasons.

5. CONCLUSION

Among 47 published draft plant genomes, 19 were obtained through sole usage of Illumina/Solexa platforms (40.42%), while additional 23 sequences were obtained through combining Illumina/Solexa with other platforms and techniques (48.64%). Usage of Illumina/Solexa platforms also encompasses nearly entire range of sizes

of published genomes. Since Illumina/Solexa platforms are the most affordable in terms sequencing cost, this appears to be the key determinant in nearly all published sequencing projects. Taking into consideration the current economic situation in the Balkans region, as well as the estimated range for the size of immortal genome, Illumina/Solexa is the best (if not the only) platform choice if the sequencing of immortal plant (*Helichrysum arenarium*) is to be performed by the researchers in this region.

• Conflict of interest: none declared.

REFERENCES

- Margulies M, Egholm M, Altman WE, et al. Genome Sequencing in Open Micro-fabricated High Density Pico-liter Reactors. *Nature*. 2005; 437(7057): 376-80.
- Liu L, Li Y, Li S, Hu N, He Y, Pong R et al. Comparison of Next-Generation Sequencing System. *Journal of Biomedicine and Biotechnology*. 2012; 251364.
- Turktas M, Kurtoğlu KY, Dorado G, Zhang B, Hernandez P. Sequencing of plant genomes - a review. *Turkish Journal of Agriculture and Forestry*. 2005 Sep; 39: 361-76.
- Unver T, Dorado G, Zhang B. Genomics and transcriptomics for utilization of plant genetic resources and crop improvement. *Turkish Journal of Agriculture and Forestry*. 2015; 39(3): 1.
- Grinev VS, Shirokov AA, Navolokin NA, Polukonova NV, Kurchatova MN, Durnova NA, et al. Polyphenolic Compounds of a New Biologically Active Extract from Immortelle Sandy Flowers (*Helichrysum arenarium* (L.) Moench). *Russ J Bioorg Chem*. 2016; 42: 770.
- Mekinić IG, Skroza D, Ljubenkov I, Krstulović L, Možina SS, Katalinić V. Phenolic Acids Profile, Antioxidant and Antibacterial Activity of Chamomile, Common Yarrow and Immortelle (Asteraceae). *Nat Prod Commun*. 2014; 9(12): 1745-8.
- Kladar NV, Anačkov GT, Rat MM, Srđenović BU, Grujić NN, et al. Biochemical Characterization of *Helichrysum italicum* (Roth) G. Don subsp. *italicum* (Asteraceae) from Montenegro: Phytochemical Screening, Chemotaxonomy, and Antioxidant Properties. *Chem Biodivers*. 2015 Mar; 12(3): 419-31.
- Čagalj M, Barbarić M, Ivanković M. Financijska analiza ekološke proizvodnje smilja u Bosni i Hercegovini. Internet.. Hrvatska znanstvena bibliografija. 2017 cited 16 April 2017.. Available from: https://bib.irb.hr/datoteka/862542.Cagalj_Immortelle.pdf
- Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, et al. The genome of the cucumber, *Cucumis sativus* L. *Nat Genet*. 2009; 41: 1275-81.
- International Brachypodium Initiative. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*. 2010; 463: 763-8.
- Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaró A, Kalyanaraman A, et al. The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nat Genet*. 2010; 42: 833-9.
- Sato S, Hirakawa H, Isobe S, Fukai E, Watanabe A, Kato M, et al. Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. *DNA Res*. 2010; 18: 65-76.
- Dassanayake M, Oh DH, Haas JS, Hernandez A, Hong H, Ali S, Y et al. The genome of the extremophile crucifer *Thellungiella parvula*. *Nat Genet*. 2011; 43: 913-8.
- Ahmad R, Parfitt D, Fass J, Ogundiwin E, Dhingra A, Gradziel T, et al. Whole genome sequencing of peach (*Prunus persica* L.) for SNP identification and selection. *BMC Genomics*. 2011; 12: 569.
- Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, et al. The genome of woodland strawberry (*Fragaria vesca*). *Nature Genet*. 2011; 43: 109-16.
- Argout X, Salse J, Aury JM, Guittinan MJ, Droc G, Gouzy J, et al. The genome of *Theobroma cacao*. *Nat Genet*. 2011; 43: 101-8.
- Young ND, Debelle F, Oldroyd GED, Geurts R, Cannon SB, Udvardi MK, et al. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature*. 2011; 480: 520-4.
- Al-Dous EK, George B, Al-Mahmoud ME, Al-Jaber MY, Wang H, Salameh YM, et al. De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nat Biotech*. 2011; 29: 521-7.
- Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, et al. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat Biotechnol*. 2012; 30: 83-9.
- Van Bakel H, Stout J, Cote A, Tallon C, Sharpe A, Hughes T, et al. The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol*. 2011; 12: R102.
- Xu X, Pan S, Cheng S, Zhang B, Mu D, Ni P, et al. Genome sequence and analysis of the tuber crop potato. *Nature*. 2011; 475: 189-95.
- Wang Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, et al. The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads. *Plant J*. 2012; 72: 461-73.
- Zhang Q, Chen W, Sun L, Zhao F, Huang B, Yang W, et al. The genome of *Prunus mume*. *Nat Commun*. 2012; 3: 1318.
- Zhang G, Liu X, Quan Z, Cheng S, Xu X, Pan S, et al. Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nature Biotechnol*. 2012; 30: 549-54.
- D'Hont A, Denoeud F, Aury JM, Baurens FC, Carreel F, Garsmeur O, et al. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature*. 2012; 488: 213-7.
- Wang K, Wang Z, Li F, Ye W, Wang J, Song G, et al. The draft genome of a diploid cotton *Gossypium raimondii*. *Nat Genet*. 2012; 44: 1098-1103.
- Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*. 2012; 485: 635-41.
- International Wheat Genome Sequencing Consortium. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*. 2014; 345(6194): 1251788.
- Bombarely A, Rosli HG, Vrebalov J, Moffett P, Mueller LA, Martin GB. A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. *Mol Plant Microbe In*. 2012; 25: 1523-30.
- Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G, González VM, et al. The genome of melon (*Cucumis melo* L.). *P Natl Acad Sci USA*. 2012; 109: 11872-7.
- Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, Mohiuddin M, et al. The cassava genome: current progress, future directions. *Trop Plant Biol*. 2012; 5: 88-94.
- Staton SE, Bakken BH, Blackman BK, Chapman MA, Kane NC, Tang S, et al. The sunflower (*Helianthus annuus* L.) genome reflects a recent history of bi-accumulation of transposable elements. *Plant J*. 2012; 72: 142-53.
- Dohm JC, Minoche AE, Holtgrawe D, Capella-Gutierrez S, Zakrzewski F, Tafer H, et al. The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature*. 2014; 505: 546-9.
- Xu Q, Chen LL, Ruan X, Chen D, Zhu A, Chen C, et al. The draft genome of sweet orange (*Citrus sinensis*). *Nature Genet*. 2013; 45: 59-66.
- Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, et al. The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat Genet*. 2013; 45: 51-8.
- Wu J, Wang Z, Shi Z, Zhang S, Ming R, Zhu S, et al. The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res*. 2013; 23: 396-408.
- Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, et al. Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat Biotechnol*. 2013; 31: 240-6.
- Peng Z, Lu Y, Li L, Zhao Q, Feng Q, Gao Z, et al. The draft genome of the fast-growing non-timber forest species moso bamboo (*Phyllostachys heterocycla*). *Nat Genet*. 2013; 45: 456-61.
- Rahman AYA, Usharraj A, Misra B, Thottathil G, Jayasekaran K, Feng Y, et al. Draft genome sequence of the rubber tree *Hevea brasiliensis*. *BMC Genomics*. 2013; 14: 75.
- Jia J, Zhao S, Kong X, Li Y, Zhao G, He W, et al. *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature*. 2013; 496: 91-5.
- Ling HQ, Zhao S, Liu D, Wang J, Sun H, Zhang C et al. Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature*. 2013; 496: 87-90.
- He N, Zhang C, Qi X, Zhao S, Tao Y, Yang G, et al. Draft genome sequence of the mulberry tree *Morus notabilis*. *Nat Commun*. 2013; 4: 2445.
- Singh R, Ong-Abdullah M, Low ETL, Manaf MAA, Rosli R, Nookiah R, et al. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature*. 2013; 500: 335-9.
- Chen J, Huang Q, Gao D, Wang J, Lang Y, Liu T, et al. Whole-genome sequencing of *Oryza brachyantha* reveals mechanisms underlying *Oryza* genome evolution. *Nat Commun*. 2013; 4: 1595.
- Sierro N, Battey JN, Ouadi S, Bovet L, Goepfert S, Bakaher N, Pet al. Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. *Genome Biol*. 2013; 14: R60.
- Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee HA, et al. Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. *Nat Genet*. 2014; 46: 270-8.
- Sierro N, Battey JND, Ouadi S, Bakaher N, Bovet L, Willig A, et al. The tobacco genome sequence and its comparison with those of tomato and potato. *Nature Commun*. 2014; 5: 3833.
- Zhang J, Liu J, Ming R. Genomic analyses of the CAM plant pineapple. *J Exp Bot*. 2014; 65: 3395-3404.
- Bolger A, Scossa F, Bolger ME, Lanz C, Maumus F, Tohge T, et al. The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nat Genet*. 2014; 46: 1034-8.
- Ming R, VanBuren R, Liu Y, Yang M, Han Y, Li LT, et al. Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.). *Genome Biol*. 2013; 14: R41.
- Ibarra-Laclette E, Lyons E, Hernández-Guzmán G, Pérez-Torres CA, Carretero-Paulet L, Chang TH, et al. Architecture and evolution of a minute plant genome. *Nature*. 2013; 498: 94-8.
- Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, et al. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*. 2014; 345: 950-3.
- Wu GA, Prochnik S, Jenkins J, Salse J, Hellsten U, Murat F, et al. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nat Biotechnol*. 2014; 2: 656-62.
- Schatz MC, Delcher AL, Salzberg SL. Assembly of large genomes using second-generation sequencing. *Genome Res*. 2010; 20: 1165-11.
- Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. *Genomics*. 2016; 107(1):1-8.
- Azizi N, Sheidai M, Mozaffarian V, Nourmohammadi Z. Karyotype and genome size analyses in species of *Helichrysum* (Asteraceae). *Acta Bot. Bras*. 2014; 28(3):367-75.