Short Communication
Volume 12:1, 2021

Journal of Molecular Biomarkers & Diagnosis

ISSN: 2155-9929

Open Access

Achievements in Genome Sequencing of Major Oilseeds

Aditya Pratap Singh*

Department of Genetics and Plant Breeding, Bidhan Chandra KrishiViswavidyalaya, Nadia, West Bengal, India

Short Communication

With the ever increasing human population around the world, food security and climate change are the major thurst areas of research for both public and private sectors alike. Having said that, the most resilient way to combat both issues is to develop high yielding yet stress tolerant crop varieties. Since traditional methods are time consuming, molecular breeding is the chosen one. Molecular breeding essentializes thorough knowledge of plants' genome information. This can be achieved with various ways like gene mapping and cloning which are in turn dependent on genome sequence data.

The recent development of genome-sequencing technology has enabled the effective identification of SNPs and other genetic variations in a large genetic population [1] and novel method for SNP detection, mapping, and quantifying transcriptomes in one run, with the potential to overshadow limitations of traditional marker systems. RNA-seq technology has been applied to plant species, opening the entire transcriptional landscape of gene activity in a high-throughput and quantitative manner in diploid species [2]. For example, large-scale resequencing projects in the human genome have effectively identified millions of SNPs, and used these to identify many genes associated with various genetic diseases [3].

Whole-genome sequencing approaches can be employed using various methods for specific purposes. The RAD-seq method for example has been attempted for a number of species and can be efficient for producing large numbers of SNPs, but has not been used routinely yet in many polyploid species, nor may be as cost-efficient for marker-assisted breeding of large populations. High read depth can be achieved using Illumina short read technology, which has the potential to represent a significant part of the transcriptome and even identify transcription factors, which have important roles in stress response. High read depth also provides an opportunity to identify variants with high confidence, which could be potentially used as markers for selection of varieties.

Keeping in view these, an attempt has been made here to concisely review the significant achievements in genome sequencing of major oilseeds (number sequence of crops are not indicative of their importance).

Significance of genome sequencing in oilseed crops

Oilseeds are gaining importance for production of edible oils with health benefits, biofuels and in pharmaceutical industries. Sequencing of genomes helps us to study the genes and intergenic regions, various transposons in the crop and their evolution from wild types to cultivated ones and their contribution towards genetic diversity of the crop [4]. It also contributes towards identification of the genes involved in production of various components that are present in oil and their respective effects. Through sequencing we can also get information on various marker trait associations and also for development of co-dominant markers [5]. Complete sequence is required while developing a species-specific vector for transfer of desirable genes. One such development is done in sesame for gene transformation using chloroplast genome vector [6].

Achievements

Various methods of sequencing were adopted by scientists for sequencing of either whole genome or a part of genome or DNA containing organelles in different crops (Table 1).

Sesame: Illumina Sequencing was adapted to sequence a contig N50 of 52.2 kb and a scaffold N50 of 2.1 Mb [2]. GS-FLX pyrosequencing was used to sequence the entire chloroplast DNA which shows that it contains 155 contigs comprising of 257,427 base pairs.

Safflower: High-throughput Illumina sequencing was adopted for characterization of miRNA transcriptomes as they play an important role in plant development and adaptation to various stresses [7]. Illumina Solexa sequencing technology was utilized for identification of transcripts involved in the biosynthesis of flavonoids such as carthamone, safflor yellow A, hydroxysafflor yellow A etc. and also studied genes responsible for oleosin production [8]. The component oleosin is found to prevent degradation of oil body present in seed during seed desiccation [9]. Using Illumina sequencing method, genes and pathways for secondary metabolites were identified in safflower tubular flower tissue transcriptome. Illumina sequencing was used to study the genome sequence for developing microsatellite markers and study their cross-species transferability [5]. This resulted in designing of 5716 novel microsatellite primers, of which 325 were validated and out of 325, 93 were found to be polymorphic in nature. A complete choloplast genome was presented using Illumina sequencing which shows that is contains 127 genes of which, 89 are protein encoding genes, 30 tRNA and 8 rRNA genes [10].

Groundnut: Illumina sequencing has been deployed to sequence the transcriptome of four market class cultivars namely OLin, New Mexico Valencia C, Tamrun OL07 and Jupiter and identified potential SNPs [11]. Genotyping-by-sequencing method of NGS was used for whole genome re-sequencing of 38 accessions and transcriptome sequencing of 3 accessions and developed a high-density SNP array that consists of 58,233 unique and informative SNPs. Again, genotyping-by-sequencing method was used to sequence RIL population and generated a genetic map consisting of 585 SNPs [12].

Sunflower: In *Helinthus annus* species different approaches of de novo assembling sequence reads were obtained by NGS procedures (Illumina and 454) to gain a comprehensive characterization of the repetitive component by Natali et al. [13].

Celik et al. used Genome Analyzer II next generation sequencing platform (Illumina Inc. San Diego, CA) and observed a total of 271,445,770 sequence reads were generated. From these reads 1,208,784 tags were generated. While 29.2 % (353,304) of the sequence tags were uniquely aligned to the sunflower genome, 14.2 % (172,196) of the tags were aligned to multiple positions [14].

Castor: The castor bean genome, which is distributed across ten chromosomes, is estimated by flow cytometry to be ~320 Mb, a draft sequence of the castor bean genome by producing ~2.1 million high-quality sequence reads from plasmid and fosmid libraries [15]. High-throughput Illumina sequencing was adopted for tissue-specific whole transcriptome sequencing in Castor for understanding triacylglycerol lipid biosynthetic pathways [16].

Rapeseed-Mustard: Illumina sequencing method was used for the complete chloroplast DNA (cpDNA) sequencing of *Brassica napus* and found

---

***Address for Correspondence:** Singh AP, Department of Genetics and Plant Breeding, Bidhan Chandra KrishiViswavidyalaya, Nadia, West Bengal, India, E-mail: adityapratapbckv@gmail.com*

Received 17 November 2020; **Accepted** 22 January 2021; **Published** 29 January, 2021

Table 1. Major oilseeds and strategies used for their genome sequencing [2,5-21].

Crop	Strategy Used	Achievements
Sesame	GS-FLX pyrosequencing method and Genome Sequencer FLX system	Complete chloroplast DNA sequencing results in 155 contigs that gives 257,427 bp
	Illumina HiSeq 2000	Sequenced a contig N50 of 52.2 kb and scaffold N50 of 2.1 Mb
Safflower	Illumina high-throughput sequencing	Characterization of miRNA transcriptomes
	Illumina Solexa sequencing	Studied de novo transcriptomes and identified transcripts that are involved in flavonoids biosynthesis and oleosin production
	Illumina sequencing	Identified genes and pathways for secondary metabolites
	Illumina HiSeq 2000	Developed SSR markers and assessed their cross-species transferability
	Illumina sequencing	Presented complete chloroplast genome
Groundnut	Illumina sequencing	Analysed the sequences of four cultivars and identified potential SNPs
	Genotyping-by-sequencing (NGS)	Analysed 38 accessions for whole genome re-sequencing and 3 for transcriptome sequencing for identification of SNPs
	Genotyping-by-sequencing	Generate SNP data and develop genetic map
Rapeseed-Mustard	Illumina sequencing	Complete chloroplast genome was sequenced and SSRs were identified
	454 pyro-sequencing	Complete mitochondrial genome was sequenced of CMS hybrid
	Illumina sequencing	Conducted re-sequencing of whole genome and studied the genetic variation of the diversity in ecotypes
Castor	Flow cytometry	Draft sequence of the castor bean genome by producing ~2.1 million high-quality sequences reads from plasmid and fosmid libraries.
	Illumina sequencing by synthesis technology	Tissue-Specific Whole Transcriptome Sequencing in Castor for understanding Triacylglycerol Lipid Biosynthetic Pathways
Sunflower	llumina or 454	Final assembly of the sequences belonging to the six databases produced a whole genome set of 283,800 contigs.
	Genome Analyzer II next generation sequencing platform (Illumina Inc. San Diego, CA).	A total of 271,445,770 sequence reads were generated. From these reads, 1,208,784 tags were generated. While 29.2% (353,304) of the sequence tags were uniquely aligned to the sunflower genome, 14.2% (172,196) of the tags were aligned to multiple positions.

that it consists of 152,860 base pairs of nucleotides that contained a pair of 26,035 bp inverted repeat sequences. The gene coding region occupied the major portion of the cpDNA i.e., 56.4% and the average AT content was found to be 63.7% in *Brassica napus*. Also, a total of 86 SSRs were also identified [17]. Complete sequencing of heterogenous composition mitochondrial genome of Ogura-cms cybrid (oguC) rapeseed was carried out with the help of 454 pyro-sequencing methods. It was found to be composed of 258,473 bp which contain 33 protein coding genes, 23 tRNA genes and 3 rRNA genes. The genome of all the 5 species of genus *Brassica* i.e., *B. napus*, *B. rapa*, *B. oleracea*, *B. carinata* and *B. juncea* was compared and found six identical regions and concluded that they have a stable chloroplast genome source [18]. By using illumine sequencing, polymorphism in the rapeseed genome in its ecotypes was studied by re-sequencing of the worldwide accessions and provides insights into the evolution of rapeseed and flowering time diversity in the three ecotypes of rapeseed [19-21].

References

1. Xiongming Du, Gai Huang, Shoupu He and Zhaoen Yang, et al. "Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits." *Nat Genet* 50 (2018): 796-802.
2. Linhai Wang, Sheng Yu, Chaobo Tong and Yingzhong Zhao, et al. "Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis." *Genome Biol* 15 (2014): 1-13.
3. Genomes Project Consortium. "A map of human genome variation from population-scale sequencing." *Nature* 467 (2010): 1061-1073.
4. David Edwards and Xiaowu Wang. "Genome sequencing initiatives." Genetics, Genomics and Breeding of Oilseed Brassicas (1stedn), 2011.
5. Heena Ambreen, Shivendra Kumar, Murali Tottekkad Variath and Gopal Joshi, et al. "Development of genomic microsatellite markers in Carthamus tinctorius L.(safflower) using next generation sequencing and assessment of their cross-species transferability and utility for diversity analysis." *PLoS One* 10 (2015): e0135443.
6. Dong-Keun Yi and Ki-Joong Kim. "Complete chloroplast genome sequences of important oilseed crop Sesamum indicum L." *PloS one* 7 (2012): e35872.
7. Haiyan Li, Yuanyuan Dong, Yepeng Sun and Erle Zhu, et al. "Investigation of the microRNAs in safflower seed, leaf, and petal by high-throughput sequencing." *Planta* 233 (2011): 611-619.
8. Haiyan Li , Yuanyuan Dong, Jing Yang and Xiuming Liu, et al. "De novo transcriptome of safflower and the identification of putative genes for oleosin and the biosynthesis of flavonoids." *PloS one* 7 (2012): e30987.
9. Dominic J Lacey, Nikolaus Wellner, Frederic Beaudion and Johnathan A. Napier, et al. "Secondary structure of oleosins in oil bodies isolated from seeds of safflower (Carthamus tinctorius L.) and sunflower (Helianthus annuus L.)." *Biochem J* 334 (1998): 469-477.
10. Chaolong Lu, Qi Shen, Jun Yang and Bo Wang, et al. "The complete chloroplast genome sequence of Safflower (Carthamus tinctorius L.)." *Mitochondrial DNA A* 27 (2016): 3351-3353.
11. Ratan Chopra, Gloria Burow, Andrew Farmer and Joann Mudge, et al. "Next-generation transcriptome sequencing, SNP discovery and validation in four market classes of peanut, Arachis hypogaea L." *Mol Genet Genom* 290 (2015): 1169-1180.
12. Manish K Pandey, Gaurav Agarwal, Sandip M. Kale and Josh Clevenger, et al. "Development and evaluation of a high density genotyping 'Axiom_ Arachis' array with 58 K SNPs for accelerating genetics and breeding in groundnut." *Sci Rep* 7 (2017): 1-10.
13. Lucia Natali, Rosa Maria Cossu, Elena Barghini and Tommaso Giordani, et al. "The repetitive component of the sunflower genome as shown by different procedures for assembling next generation sequencing reads." *BMC Genomics* 14 (2013): 1-14.
14. Ibrahim Celik, Sabahattin Bodur, Anne Frary and Sami Doganlar. "Genome-wide SNP discovery and genetic linkage map construction in sunflower (Helianthus annuus L.) using a genotyping by sequencing (GBS) approach." *Mol Breed* 36 (2016): 1-9.
15. Agnes P. Chan, Jonathan Crabtree, Qi Zhao and Hernan Lorenzi, et al. "Draft genome sequence of the oilseed species Ricinus communis." *Nat Biotechnol* 28 (2010): 951-956.
16. Adrian P Brown, Johan TM Kroon, David Swarbreck and Melanie Febrer, et al. "Tissue-specific whole transcriptome sequencing in castor, directed at understanding triacylglycerol lipid biosynthetic pathways." *PLoS One* 7 (2012): e30100.
17. Zhi-Yong Hu, Wei Hua, Shun-Mou Huang and Han-Zhong Wang. "Complete chloroplast genome sequence of rapeseed (*Brassica napus* L.) and its evolutionary implications." *Genet Resour Crop Evol* 58 (2011): 875-887.
18. Juan Wang, Jinjin Jiang, Xiaoming Li and Aimin Li, et al. "Complete sequence

of heterogenous-composition mitochondrial genome (*Brassica napus*) and its exogenous source." *BMC Genomics* 13 (2012): 1-9.

19. Dezhi Wu, Zhe Liang, Tao Yan and Ying Xu, et al. "Whole-genome resequencing of a worldwide collection of rapeseed accessions reveals the genetic basis of ecotype divergence." *Mol Plant* 12 (2019): 30-43.
20. Huang Lulin, Yang Xiao, Sun Pei and Tong Wen, et al. "The first Illumina-based de novo transcriptome sequencing and analysis of safflower flowers." *PloS one* 7 (2012): e38653.
21. Sneha M. Dodia, Binal Joshi, Sunil S. Gangurde and Polavakkalipalayam P. Thirumalaisamy, et al. "Genotyping-by-sequencing based genetic mapping reveals large number of epistatic interactions for stem rot resistance in groundnut." *Theor Appl Genet* 132 (2019): 1001-1016.

How to cite this article: Aditya Pratap Singh. “Achievements in Genome Sequencing of Major Oilseeds.” *J Mol Biomark Diagn* 12 (2021): 449.