



# From Genomes to GENE-omes: Exome Sequencing Concept and Applications in Crop Improvement

*Parampreet Kaur and Kishor Gaikwad\**

*National Research Centre on Plant Biotechnology, New Delhi, India*

## OPEN ACCESS

### Edited by:

Hikmet Budak,  
Montana State University,  
United States

### Reviewed by:

Martin Mascher,  
Leibniz-Institut für Pflanzengenetik  
und Kulturpflanzenforschung (IPK),  
Germany  
Charu Lata,  
National Botanical Research Institute  
(CSIR), India

### \*Correspondence:

Kishor Gaikwad  
kish2012@nrpcb.org

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

**Received:** 19 September 2017

**Accepted:** 08 December 2017

**Published:** 19 December 2017

### Citation:

Kaur P and Gaikwad K (2017) From  
Genomes to GENE-omes: Exome  
Sequencing Concept  
and Applications in Crop  
Improvement.  
*Front. Plant Sci.* 8:2164.  
doi: 10.3389/fpls.2017.02164

Exome sequencing represents targeted capture and sequencing of 1–2% of ‘high-value genomic regions’ (subset of the genome) which are enriched for functional variants and harbors low level of repetitive regions. We discuss here an overview of exome sequencing, ways to approach plant exomes, and advantages and applicability of this powerful approach in deciphering functional regions of genomes. Though initially this approach was developed as an alternative to whole genome sequencing (WGS), but the multitude of benefits conferred by sequence capture via hybridization approaches created a niche for itself to solve many of biological riddles, particularly for resolving phylogenetic distances. The technique has also proved to be successful in understanding the basis of natural and induced molecular variation, marker development and developing genomic resources for complex, wild and non-model species, which are still intractable for WGS efforts. Thus, with profound applications of this powerful sequencing strategy, near future is expected to witness a collective expansion of both techniques, i.e., sequence capture via hybridization for evolutionary and ecological research and WGS approaches for its universal accessibility.

**Keywords:** exome, whole genome sequencing, whole exome sequencing, crop improvement, genomics

## INTRODUCTION

Completion of an Arabidopsis plant genome sequence in 2000 marked a major breakthrough in the sequencing era. Further progress in deciphering of large and complex genomes of crop species is a reflection of the advancement of sequencing technologies that are commonly referred to as NextGen sequencers along with the evolution of more powerful data analysis tools.

Polyploidization has played a critical role during evolution in shaping the complex genomes of important crop species viz. Wheat, Brassica etc., and has resulted in large genome sizes with exceptionally high repeat content, heterozygosity and presence of closely related homeologous genes in large gene families. This poses significant challenge for whole genome sequencing (WGS) and resequencing in many such plant lineages. Several strategies such as reduction of genome complexity, sequencing of diploid progenitors and sequencing of sorted and purified chromosome arms are thus deployed to gain chromosome specific insight into genomes of complex polyploids (Bevan et al., 2017). One important way to achieve reduction in genome complexity analysis is referred to as Sequence Capture or Targeted Sequencing that includes either specific genes of interest or targets within genes or the entire protein coding region of genome, i.e., exome and could be achieved using either of the three modes: hybridization based sequence capture, PCR-based amplification, and selective circularization (Dahl et al., 2005; Hodges et al., 2007;

Gnirke et al., 2009). Among these, hybrid capture mode allows several megabases of DNA to be analyzed for efficient sequencing of the complete 'exome' and represents its current major application. Additionally, the technique has advantage of being extraordinarily quick, simple, inexpensive, and requires small amount of input DNA (<1–3 µg) (Mertes et al., 2011).

Exome sequencing restricts attention only to the genomic fraction that encodes for mRNA and eventually a phenotype, and is thus considered to be adequate to explain the molecular origin of genetic variation (induced/natural). Coding sequences represents only 1–2% of genome depending on species (Warr et al., 2015) and comprises high level of functional variants and low repeat content. For example, wheat exome constitute 170–340 Mb (Paux et al., 2006) of size in comparison to massively large genomes of tetraploid (10 Gb) and hexaploid (17 Gb) wheat. Considering the manageable size of many crop exomes and by emphasizing on 'high-value genomic regions' (Hodges et al., 2007), exome analysis permits deep sequencing for large number of samples to identify useful variants for incorporation in molecular breeding strategies.

## HOW TO APPROACH PLANT EXOME

Two technological alternatives available for hybridization based exome capture are: array/chip- based capture (Hodges et al., 2007; Okou et al., 2007) and In solution capture (Gnirke et al., 2009). Both approaches relies on specifically designed probes or baits for target enrichment from sequencing library, but solution captures requires more concentration of baits over DNA library in contrast to array capture that requires excess of library over probes to perform enrichment.

Exome sequencing could be divided into two phases, first phase constitutes probe hybridization for selection of subset of DNA that encodes for a protein, i.e., target enrichment and second phase constitutes its high throughput DNA sequencing (Figure 1). Certain factors such as mode and quality of processing of input DNA sample, number of targets, coverage depth for each target, probe design and GC content (Zhou and Holliday, 2012), expected enrichment efficiency, sequencing technology used and biological system studied etc., should be considered while designing an exome capture experiment (Mertes et al., 2011; Grover et al., 2012). Of all these factors, depth of the sequence coverage is important in achieving good reliability of sequencing of exome capture experiments and requires coverage of at least 30X or higher to successfully validate the variation identified (Winfield et al., 2012). Coverage depth in turn is governed by probe specificity, homeology and heterozygosity level, ploidy level, genome size, presence of orthologs and paralogs as well as by characteristics of probes such as genomic regions from where the probe is designed, i.e., whether conserved or unique (Grover et al., 2012). Well-established strategies are available for construction of exome libraries along with ready to use exome kits (NimbleGen, Affymetrix) and their user dependent customizations are available for many crop species such as soybean, wheat, barley, and maize. Hybridization capture approaches are known to suffer from off target capture and

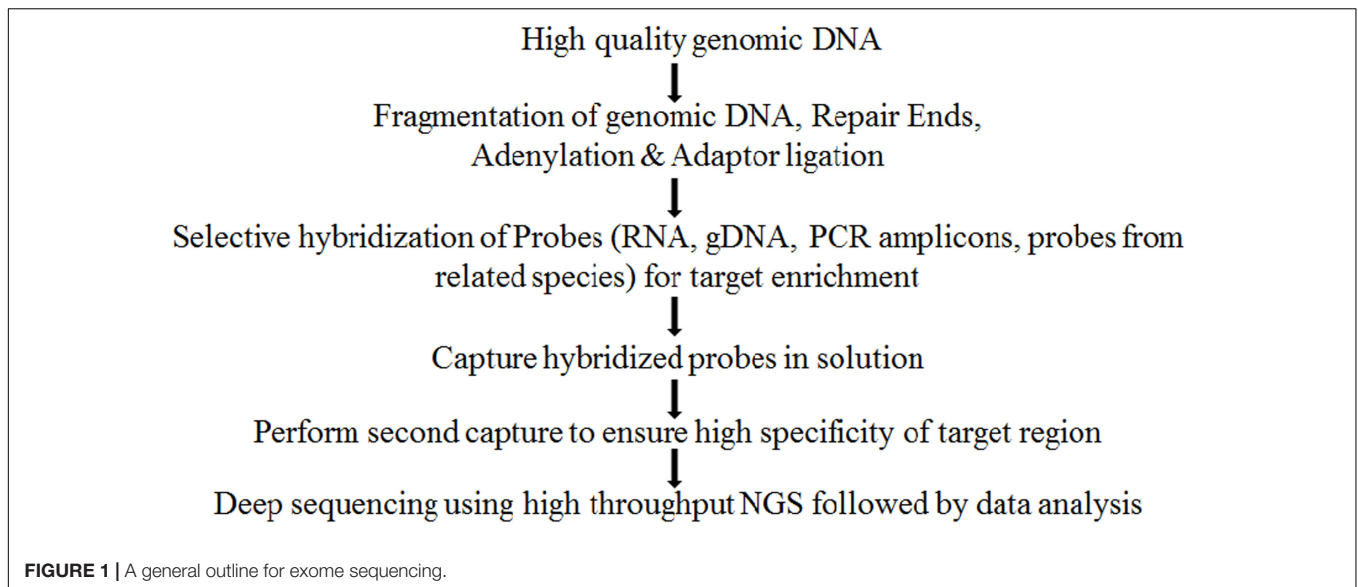
capture of highly repetitive sequences. Recently, a low cost PCR based method has been designed that utilizes multiplex ligation dependent probe amplification (MLPA probes) of enriched library followed by capillary electrophoresis to validate exome library in terms of enrichment efficiency to prevent sequencing of unsuccessfully enriched libraries (Klonowska et al., 2016).

## EXOME SEQUENCING: AN EDGE OVER OTHER APPROACHES

Whole exome sequencing (WES) holds certain advantages over its comparable approaches, like RNA sequencing, that is sometimes limited by the bias of transcript abundance and also dependent on tissue and stage while exome sequencing permits exploration of the genes and alleles. Similarly, sequencing of gene rich regions by methylation filtration method is limited by the minimum control a researcher can execute on specific target regions especially during genotyping of diverse germplasm in parallel. While exome analysis is a probe-based enrichment approach that targets specific regions rather than random euchromatic regions. Other methodologies viz., EST sequencing, high Cot DNA selection (Barbazuk et al., 2005) are less efficient in delivering specific sequences in targeted manner (Fu et al., 2010).

Exome sequencing serves a great advantage when WGS is either not practical or necessary and as an adjunct to later. It allows affordable dissection of subregions of complex genomes and reduction of non-pertinent repetitive sequences to confer several benefits over WGS: (a) multiplexing of more samples for a given sequencing space, (b) sequencing of targeted informative regions reduces the complexity of data analysis, (c) identification of functional molecular markers, (d) useful platform to collect genomic data at population level for evolutionary and phylogeny studies, (e) existing databases provides functional context for exome identified SNPs through transcript/exon annotation in contrast to SNPs identified outside coding regions through genotyping based sequencing which are not easily annotated (Scheben et al., 2017), and (f) provides high coverage for identification of low frequency sequence variants (Klonowska et al., 2016) for complex traits (Kiezun et al., 2012). Exome sequencing represents an established method of targeted resequencing of gene space (Bamshad et al., 2011) especially for phylogenetics and divergence studies by targeting evolutionary more conserved regions, i.e., specific genes or genomic regions and thus provides improved sequencing depth of targeted regions rather than by surveying the entire genome.

Though high-depth WGS is considered as a gold standard for sequencing and re-sequencing purposes as it can access and interrogate all regions of genomes, exome sequencing still forms an alternative method of choice for the trait or target specific studies. It has demonstrated its potential in generating genome wide data for species such as black cotton wood (Zhou and Holliday, 2012), pine (Neves et al., 2013), and sugarcane (Song et al., 2016) etc., which are intractable for WGS studies. Since hybridized captured fragments are longer than probes used to retrieve them, exome sequencing provides insights into



coding, intronic, UTRs and putative regulatory regions for a comprehensive view of gene regulation.

## EXOME SEQUENCING AND PHYLOGENETIC STUDIES

Phylogenetic studies rely on the information content of homologous genomic regions across diverse taxonomic groups. Due to general conservation of exons of protein coding sequences as well as their inherent variability in third codon position, probes designed from the same ensures the efficacy of target capture in reference species and also holds potential to extend their capture across diverse taxas to resolve topologies at moderate to deep evolutionary time scales. Introns are generally targeted to analyze phylogeny at or below the species level. Apart from probe design, loci under consideration, target number, size and its properties, population size and time scale under study are other important determinants to consider during phylogeny analysis using sequence capture.

Relative to other approaches viz. use of restriction enzymes, PCR amplification, genome skimming or WGS approaches, phylogenetic analysis using sequence capture via hybridization offers certain advantages, such as, (a) ability to enrich for degraded DNA samples obtained from museums or herbariums; (b) reduced sequencing of contaminant DNA from paleontological samples; (c) targeted approach to reduce genome complexity relative to anonymous loci produced by restriction enzymes based approaches; (d) multiplexing offers broader sampling and deep coverage relative to WGS to target phylogenetically informative loci to resolve topologies with increased resolution in a cost effective manner without involving complexities of data analysis and storage at whole genome scale; and (e) deep coverage of targeted and unlinked loci allows to accommodate for more variation in probe designs to generate family wide sets of probes that ensures high

resolution and strong bootstrap values for phylogenetic trees (de Sousa et al., 2014). These features form the forte for wide applicability of sequence capture tools through probe hybridization in phylogenetic studies (Table 1) (Jones and Good, 2016). Additionally, targeted enrichment of a sufficient number of informative loci prior to sequencing ensures phylogenetic accuracy to deduce broader and high resolution topologies. Whereas, selection of loci post-sequencing could cause systematic errors during analysis of the enormous amount of data generated at whole genome scale, that apart from targeted information also includes low quality, unusable data or data that cannot be effectively modeled. Since, the science of evolution and ecology is not dependent on sequencing of each base of the whole genome, rather it relies on deep sequencing of informative loci with desired evolutionary rates to produce well-resolved topologies.

Though, initially most of the targeted sequencing effort to define phylogenetic relations were directed toward plastome or ribosomal DNA sequences, recently, Mandel et al. (2014) utilized conserved orthologous set (COS) loci from the nuclear genome of Asteraceae (Chapman et al., 2007) to deduce the phylogenetic relationship across its 15 species and also demonstrated usability of 763 COS loci to facilitate taxonomic reconstruction in consistency with known phylogenetic relationships. Weitemier et al. (2014) further combined sequence capture with genome skimming and performed target enrichment for both low copy nuclear genes and genes from organelle genome to analyze phylogeny among 12 milkweed genotypes. The technique was named as Hyb-Seq and utilized enrichment probes designed from taxon specific genomic and transcriptomic data. Recently, Comer et al. (2015) compared the efficacy of hybrid gene capture and Long range PCR to describe phylogeny across 31 members of Arcoideae tribe and 5 outgroup taxas. de Sousa et al. (2014) utilized 50 low copy number (LCN) genes to describe Medicago phylogenetic. LCN genes are characterized by properties that are sufficiently informative for their use in phylogenetic analysis,

**TABLE 1** | Crop species targeted for exome sequencing.

Features	Crop	Reference	
Phylogenetic analysis	Hordeum	Brassac and Blattner, 2015	
	Arecoideae	de Sousa et al., 2014	
	Asteraceae	Mandel et al., 2014	
	Milkweed	Weitemier et al., 2014	
Genomic diversity	Maize	Fu et al., 2010; Liu et al., 2012; Muraya et al., 2015	
	Wheat	Saintenac et al., 2011; Winfield et al., 2012	
	Soybean	Haun et al., 2011	
	Sugarcane	Song et al., 2016	
	Barley	Mascher et al., 2013	
	Brassica	Clarke et al., 2013	
	Black Cottonwood	Zhou and Holliday, 2012	
Monitoring alien introgression	Wheat	Winfield et al., 2016	
	Barley	Wendler et al., 2014	
Mutation characterization	Soybean	Bolon et al., 2011	
	Rice	Henry et al., 2014	
	Wheat	King et al., 2015; Krasileva et al., 2017	
	Wheat	Steuernagel et al., 2016	
Mutagenesis + R gene enrichment	Wheat	Steuernagel et al., 2016	
Trait mapping	Antioxidant metabolism	Tomato	Ruggieri et al., 2016
	Wood property traits	Eucalyptus	Dasgupta et al., 2015
	Flowering time regulatory gene	Brassica	Schiessl et al., 2014
	Many nodded dwarf phenotype	Barley	Mascher et al., 2014
	Early maturation	Barley	Pankin et al., 2014
Biomass production	Maize	Muraya et al., 2015	
	Wheat	Allen et al., 2013	
Sequence based genetic mapping	Brassica	Galvão et al., 2012; Clarke et al., 2013	
	Pine	Neves et al., 2014	
	Cassava	Pootakham et al., 2014	
	Wheat	Allen et al., 2013	

such as high evolutionary rate, biparental inheritance (Small et al., 1998; Hughes et al., 2006), stable copy number, and provides higher resolution in comparison to ribosomal DNA or Plastome. Similarly, Brassac and Blattner (2015) analyzed 12 nuclear single copy loci and 1 chloroplast loci to infer phylogeny relations of *Hordeum* taxa and reported utility of multilocus species level phylogeny without sequencing whole genomes. Thus, the approach permits to target variation in much broader gene pool by accommodating a large number of genomes to be sequenced with less complexity in data analysis, in comparison to resequencing of whole genomes of same number of genotypes.

## APPLICATIONS OF EXOME SEQUENCING FOR CROP IMPROVEMENT

Apart from phylogenetic studies, WES also holds a wide applicability for crop improvement (Table 1) namely in (a) investigation of the molecular basis of natural variation and induced mutation, (b) identification of new genetic markers and rare variants, (c) and as an effective alternative for genomics studies of non-model and related wild species.

One of the first successful application of sequence capture was to enrich for 2.3 Mb chromosome interval and set of

dispersed 43 genes in maize (Fu et al., 2010) and subsequently whole exome capture of maize to validate genetic diversity between cultivars (Liu et al., 2012) leading to identification of variation affecting biomass production (Muraya et al., 2015). Exome capture has been explored to identify natural variation and genetic diversity between wild and cultivated durum wheat accessions (Saintenac et al., 2011), hexaploid wheat varieties (Winfield et al., 2012), identification of novel co-dominant SNP markers (Allen et al., 2013) and monitoring of alien introgressions to ensure linkage drag reduction (Winfield et al., 2016). Exome capture in soybean confirmed the presence of unwanted intracultivar heterogeneity influencing observed variation in cv. Williams 82 (Haun et al., 2011). Barley exome was characterized to identify natural variation (Mascher et al., 2013), mutation involved in early maturation (Pankin et al., 2014), gene causative of many nodded dwarf phenotype (Mascher et al., 2014) and to differentiate between markers of *H. vulgare* and *H. bulbosum* for alien introgressions (Wendler et al., 2014). Exome sequencing has played a crucial role in accessing genomic variation of important perennials, like miscanthus<sup>1</sup>, black cotton wood (Zhou and Holliday, 2012) and pine (Neves et al., 2013) so as to better link their genotypes to phenotypes for agricultural improvement. Similarly, eucalyptus exome has been analyzed

<sup>1</sup><https://pag.confex.com/pag/xxi/webprogram/Paper5474.html>



to identify mutation for wood property traits, wherein targeted resequencing for 94 genes functioning in different steps of secondary xylem formation served as an efficient strategy to identify SNPs and SNVs to develop mapping pedigrees (Dasgupta et al., 2015). Exome capture has been successfully utilized to generate high density gene rich genetic map in pine (Neves et al., 2013) and also for sequence based genetic mapping in brassicacea (Galvão et al., 2012). Thus, it is a highly relevant approach to examine genome synteny, identification of QTLs and candidate regulatory genes as well as to assist in genome assembly of species without reference genome. Sequence variation of 378 genes and associated regulatory regions involved in antioxidant metabolism was analyzed for 44 tomato landraces identified 4000 rare variants with >40X coverage (Ruggieri et al., 2016), thus demonstrating the capacity of exome sequencing to extend the knowledge of the genetic base (Gasc et al., 2016). Similarly, Pootakham et al. (2014) utilized target capture to genotype 100 F1 progeny of cassava mapping population segregating for starch viscosity phenotypes. Clarke et al. (2013) performed exome capture to target variation in 47 genomic regions of 10 brassica genotypes and characterized 589367 SNPs in QTL regions associated with nutritional or agronomically important traits such as yield, seedling vigor, seed quality, blackleg disease resistance, etc. Efficacy of sequence capture to analyze variation between homologs for flowering time regulatory gene in *B. napus* (Schiessl et al., 2014) open avenues for its usage to decode other important pathways in polyploid crops.

Cross capturing of exomes using probes of related species have been effectively applied for less resourced species. Extensive collinearity shared between sorghum and sugarcane genic regions (Wang et al., 2010) was exploited to enrich 5.8 Mb region of two sugarcane genotypes (Bundock et al., 2012). Till date, no reference genome sequence is available for sugarcane considering its genome structure (Zhang et al., 2012), but lately exome of 12 different accessions of *Saccharum* complex has been captured successfully (Song et al., 2016) to characterize natural allelic variation and haplotype pattern of 406 candidate genes, a first report on haplotype identification in polyploid sugarcane genome.

Suitability of exome capture to identify induced variants through mutagenesis to investigate gene function has been successfully demonstrated in many crops. Exome resequencing of four selected soybean mutants arose through fast neutron radiation confirmed presence of deletions (Bolon et al., 2011). Exon capture of mutagenized lines of rice (Henry et al., 2014)

and wheat (King et al., 2015) has been done to identify novel mutations. MutRenSeq, i.e., exome capture of EMS lines in combination with RenSeq (R gene enrichment sequencing) was performed in hexaploid wheat for cloning Sr22 and Sr25 stem rust resistance genes (Steuernagel et al., 2016). More than 10 million mutations were cataloged in exome of wheat EMS lines and recessive mutation at *vrn2* locus was identified in association with spring growth habit (Krasileva et al., 2017).

## CONCLUSION AND FUTURE PROSPECTS

Exome sequencing via hybridization, holds high potential for its applicability in phylogenetic studies and genomic studies that require resequencing of multiple individuals of species with large and complex genome sizes. Few considerations have been raised for WES in terms of benefits conferred by WGS approaches, i.e., more consistent genome coverage, uniform distribution of sequencing quality parameters, identification of a large number of variants and absence of reference sequence bias generated by probes usage. But even in the near future, when prices of WGS will become at par with that of sequence capture via hybridization, the latter approach is much anticipated to sustain given its practical utilities for solving specific biological queries that rely on target sequencing. Resequencing of genomes of individuals in a mapping population/multiple accessions or cultivars of plants with large and complex genomes cannot be routinely employed and poses technical challenges in handling of robust informatics data. To its rescue, sequencing of sub-genomic regions is an effective alternate in terms of cost and efforts to achieve an equivalent depth of coverage, particularly for studies that do not require access to whole genomes.

## AUTHOR CONTRIBUTIONS

PK and KG have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## ACKNOWLEDGMENT

Help provided by ICAR-NRCPB and ICAR-NASF is duly acknowledged.

## REFERENCES

- Allen, A. M., Barker, G. L. A., Wilkinson, P., Burrige, A., Winfield, M., Coghill, J., et al. (2013). Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.). *Plant Biotechnol. J.* 11, 279–295. doi: 10.1111/pbi.12009
- Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A., et al. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* 12, 745–755. doi: 10.1038/nrg3031
- Barbazuk, W. B., Bedell, J. A., and Rabinowicz, P. D. (2005). Reduced representation sequencing: a success in maize and a promise for other plant genomes. *Bioessays* 27, 839–848. doi: 10.1002/bies.20262
- Bevan, M. W., Uauy, C., Wulff, B. B. H., Zhou, J., Krasileva, K., and Clark, M. D. (2017). Genomic innovation for crop improvement. *Nature* 543, 346–354. doi: 10.1038/nature22011
- Bolon, Y.-T., Haun, W. J., Xu, W. W., Grant, D., Stacey, M. G., Nelson, R. T., et al. (2011). Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean. *Plant Physiol.* 156, 240–253. doi: 10.1104/pp.110.170811
- Brassac, J., and Blattner, F. R. (2015). Species level phylogeny and polyploid relationships in *Hordeum* (Poaceae) inferred by next-generation sequencing and *In-Silico* cloning of multiple nuclear loci. *Syst. Biol.* 64, 792–808. doi: 10.1093/sysbio/syv035
- Bundock, P. C., Casu, R. E., and Henry, R. J. (2012). Enrichment of genomic DNA for polymorphism detection in a non-model highly polyploid

- crop plant. *Plant Biotechnol. J.* 10, 657–667. doi: 10.1111/j.1467-7652.2012.00707.x
- Chapman, M. A., Chang, J. C., Weisman, D., Esseli, R. V., and Burke, J. M. (2007). Universal markers for comparative mapping and phylogenetic analysis in the Asteraceae (Compositae). *Theor. Appl. Genet.* 115, 747–755. doi: 10.1007/s00122-007-0605-2
- Clarke, W. E., Parkin, I. A., Gajardo, H. A., Gerhardt, D. J., Higgins, E., Sidebottom, C., et al. (2013). Genomic DNA enrichment using sequence capture microarrays: a novel approach to discover Sequence Nucleotide Polymorphisms (SNP) in *Brassica napus* L. *PLOS ONE* 8:e81992. doi: 10.1371/journal.pone.0081992
- Comer, J. R., Zomlefer, W. B., Barrett, C. F., Davis, J. I., Stevenson, D. W. M., Heyduk, K., et al. (2015). Resolving relationships within the palm subfamily Arecoideae (Arecaceae) using plastid sequences derived from Next-Generation sequencing. *Am. J. Bot.* 102, 888–899. doi: 10.3732/ajb.1500057
- Dahl, F., Gullberg, M., Stenberg, J., Landegren, U., and Nilsson, M. (2005). Multiplex amplification enabled by selective circularization of large sets of genomic DNA fragments. *Nucleic Acids Res.* 33, e71. doi: 10.1093/nar/gni070
- Dasgupta, M. G., Dharanishanthi, V., Agarwal, I., and Krutovsky, K. V. (2015). Development of genetic markers in Eucalyptus species by target enrichment and exome sequencing. *PLOS ONE* 10:e0116528. doi: 10.1371/journal.pone.0116528
- de Sousa, F., Bertrand, Y. J. K., Nylinder, S., Oxelman, B., Eriksson, J. S., and Pfeil, B. E. (2014). Phylogenetic properties of 50 nuclear loci in Medicago (Leguminosae) generated using multiplexed sequence capture and Next-Generation sequencing. *PLOS ONE* 9:e109704. doi: 10.1371/journal.pone.0109704
- Fu, Y., Springer, N. M., Gerhardt, D. J., Ying, K., Yeh, C. T., Wu, W., et al. (2010). Repeat subtraction-mediated sequence capture from a complex genome. *Plant J.* 62, 898–909. doi: 10.1111/j.1365-313X.2010.04196.x
- Galvão, V. C., Nordström, K. J. V., Lanz, C., Sulz, P., Mathieu, J., Posé, D., et al. (2012). Synteny-based mapping-by-sequencing enabled by targeted enrichment. *Plant J.* 71, 517–526. doi: 10.1111/j.1365-313X.2012.04993.x
- Gasc, C., Peyretailade, E., and Peyret, P. (2016). Sequence capture by hybridization to explore modern and ancient genomic diversity in model and nonmodel organisms. *Nucleic Acids Res.* 44, 4504–4518. doi: 10.1093/nar/gkw309
- Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., et al. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* 27, 182–189. doi: 10.1038/nbt.1523
- Grover, C. E., Salmon, A., and Wendel, J. F. (2012). Targeted sequence capture as a powerful tool for evolutionary analysis. *Am. J. Bot.* 99, 312–319. doi: 10.3732/ajb.1100323
- Haun, W. J., Hyten, D. L., Xu, W. W., Gerhardt, D. J., Albert, T. J., Richmond, T., et al. (2011). The composition and origins of genomic variation among individuals of the soybean reference cultivar Williams 82. *Plant Physiol.* 155, 645–655. doi: 10.1104/pp.110.166736
- Henry, I. M., Nagalakshmi, U., Lieberman, M. C., Ngo, K. J., Krasileva, K. V., Vasquez-Gross, H., et al. (2014). Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing. *Plant Cell* 26, 1382–1397. doi: 10.1105/tpc.113.121590
- Hodges, E., Xuan, Z., Balija, V., Kramer, M., Molla, M. N., Smith, S. W., et al. (2007). Genome-wide in situ exon capture for selective resequencing. *Nat. Genet.* 39, 1522–1527. doi: 10.1038/ng.2007.42
- Hughes, C. E., Eastwood, R. J., and Bailey, C. D. (2006). From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction. *Philos. Trans. R. Soc. B Biol. Sci.* 361, 211–225. doi: 10.1098/rstb.2005.1735
- Jones, M. R., and Good, J. M. (2016). Targeted Capture in evolutionary and ecological genomics. *Mol. Ecol.* 25, 185–202. doi: 10.1111/mec.13304
- Kiezun, A., Garimella, K., Do, R., Stitzel, N. O., Neale, B. M., McLaren, P. J., et al. (2012). Exome sequencing and the genetic basis of complex traits. *Nat. Genet.* 44, 623–630. doi: 10.1038/ng.2303
- King, R., Bird, N., Ramirez-Gonzalez, R., Coghill, J. A., Patil, A., Hassani-Pak, K., et al. (2015). Mutation scanning in wheat by exon capture and next-generation sequencing. *PLOS ONE* 10:e0137549. doi: 10.1371/journal.pone.0137549
- Klonowska, K., Handschuh, L., Swiercz, A., Figlerowicz, M., and Kozłowski, P. (2016). MTTE: an innovative strategy for the evaluation of targeted/exome enrichment efficiency. *Oncotarget* 7, 67266–67276. doi: 10.18632/oncotarget.11646
- Krasileva, K. V., Vasquez-Gross, H. A., Howell, T., Bailey, P., Paraiso, F., Clissold, L., et al. (2017). Uncovering hidden variation in polyploid wheat. *Proc. Natl. Acad. Sci. U.S.A.* 114, E913–E921. doi: 10.1073/pnas.1619268114
- Liu, S., Ying, K., Yeh, C. T., Yang, J., Swanson-Wagner, R., Wu, W., et al. (2012). Changes in genome content generated via segregation of non-allelic homologs. *Plant J.* 72, 390–399. doi: 10.1111/j.1365-313X.2012.05087.x
- Mandel, J. R., Dikow, R. B., Funk, V. A., Masalia, R. R., EvanStaton, S., Kozik, A., et al. (2014). A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the compositae. *Appl. Plant Sci.* 2:1300085. doi: 10.3732/apps.1300085
- Mascher, M., Jost, M., Kuon, J.-E., Himmelbach, A., Aßfalg, A., Beier, S., et al. (2014). Mapping-by-sequencing accelerates forward genetics in barley. *Genome Biol.* 15:R78. doi: 10.1186/gb-2014-15-6-r78
- Mascher, M., Richmond, T. A., Gerhardt, D. J., Himmelbach, A., Clissold, L., Sampath, D., et al. (2013). Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *Plant J.* 76, 494–505. doi: 10.1111/tj.12294
- Mertes, F., ElSharawy, A., Sauer, S., van Helvoort, J. M. L. M., van der Zaag, P. J., Franke, A., et al. (2011). Targeted enrichment of genomic DNA regions for next-generation sequencing. *Brief. Funct. Genomics* 10, 374–386. doi: 10.1093/bfpg/elr033
- Muraya, M. M., Schmutzer, T., Ulpinnis, C., Scholz, U., and Altmann, T. (2015). Targeted sequencing reveals large-scale sequence polymorphism in maize candidate genes for biomass production and composition. *PLOS ONE* 10:e0132120. doi: 10.1371/journal.pone.0132120
- Neves, L. G., Davis, J. M., Barbazuk, W. B., and Kirst, M. (2013). Whole-exome targeted sequencing of the uncharacterized pine genome. *Plant J.* 75, 146–156. doi: 10.1111/tj.12193
- Neves, L. G., Davis, J. M., Barbazuk, W. B., and Kirst, M. (2014). A high-density gene map of loblolly pine (*Pinus taeda* L.) Based on Exome Sequence Capture Genotyping. *G3* 4, 29–37. doi: 10.1534/g3.113.008714
- Okou, D. T., Steinberg, K. M., Middle, C., Cutler, D. J., Albert, T. J., and Zwick, M. E. (2007). Microarray-based genomic selection for high-throughput resequencing. *Nat. Methods* 4, 907–909. doi: 10.1038/nmeth1109
- Pankin, A., Campoli, C., Dong, X., Kilian, B., Sharma, R., Himmelbach, A., et al. (2014). Mapping-by-sequencing identifies HvPHYTOCHROME C as a candidate gene for the early maturity 5 locus modulating the circadian clock and photoperiodic flowering in barley. *Genetics* 198, 383–396. doi: 10.1534/genetics.114.165613
- Paux, E., Roger, D., Badaeva, E., Gay, G., Bernard, M., Sourdille, P., et al. (2006). Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. *P. J.* 48, 463–474. doi: 10.1111/j.1365-313X.2006.02891.x
- Pootakham, W., Shearman, J. R., Ruang-Areerate, P., Sonthirod, C., Sangsakru, D., Jomchai, N., et al. (2014). Large-scale SNP discovery through RNA Sequencing and SNP genotyping by targeted enrichment sequencing in cassava (*Manihot esculenta* Crantz). *PLOS ONE* 9:e116028. doi: 10.1371/journal.pone.0116028
- Ruggieri, V., Anzar, I., Paytavi, A., Calafiore, R., Cigliano, R. A., Sansaverino, W., et al. (2016). Exploiting the great potential of Sequence Capture data by a new tool, SUPER-CAP. *DNA Res.* 24:dsw050. doi: 10.1093/dnares/dsw050
- Saintenac, C., Jiang, D., and Akhunov, E. D. (2011). Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. *Genome Biol.* 12:R88. doi: 10.1186/gb-2011-12-9-r88
- Scheben, A., Batley, J., and Edwards, D. (2017). Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. *Plant Biotechnol. J.* 15, 149–161. doi: 10.1111/pbi.12645
- Schiessl, S., Samans, B., Hüttel, B., Reinhard, R., and Snowdon, R. J. (2014). Capturing sequence variation among flowering-time regulatory gene homologs in the allopolyploid crop species *Brassica napus*. *Front. Plant Sci.* 5:404. doi: 10.3389/fpls.2014.00404
- Small, R. L., Ryburn, J. A., Cronn, R. C., Seelanan, T., and Wendel, J. F. (1998). The tortoise and the hare: choosing between noncoding plastome and nuclear Adh sequences for phylogeny reconstruction in a recently diverged plant group. *Am. J. Bot.* 85, 1301–1315.

- Song, J., Yang, X., Resende, M. F. R., Neves, L. G., Todd, J., Zhang, J., et al. (2016). Natural allelic variations in highly polyploidy saccharum complex. *Front. Plant Sci.* 7:804. doi: 10.3389/fpls.2016.00804
- Steuernagel, B., Periyannan, S. K., Hernández-Pinzón, I., Witek, K., Rouse, M. N., Yu, G., et al. (2016). Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. *Nat. Biotechnol.* 34, 652–655. doi: 10.1038/nbt.3543
- Wang, J., Roe, B., Macmil, S., Yu, Q., Murray, J. E., Tang, H., et al. (2010). Microcollinearity between autopolyploid sugarcane and diploid sorghum genomes. *BMC Genomics* 11:261. doi: 10.1186/1471-2164-11-261
- Warr, A., Robert, C., Hume, D., Archibald, A., Deeb, N., and Watson, M. (2015). Exome sequencing: current and future perspectives. *G3* 5, 1543–1550. doi: 10.1534/g3.115.018564
- Weitemier, K., Straub, S. C. K., Cronn, R. C., Fishbein, M., Schmickl, R., McDonnell, A., et al. (2014). Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. *Appl. Plant Sci.* 2:1400042. doi: 10.3732/apps.1400042
- Wendler, N., Mascher, M., Nöh, C., Himmelbach, A., Scholz, U., Ruge-Wehling, B., et al. (2014). Unlocking the secondary gene-pool of barley with next-generation sequencing. *Plant Biotechnol. J.* 12, 1122–1131. doi: 10.1111/pbi.12219
- Winfield, M. O., Allen, A. M., Burrridge, A. J., Barker, G. L. A., Benbow, H. R., Wilkinson, P. A., et al. (2016). High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol. J.* 14, 1195–1206. doi: 10.1111/pbi.12485
- Winfield, M. O., Wilkinson, P. A., Allen, A. M., Barker, G. L. A., Coghill, J. A., Burrridge, A., et al. (2012). Targeted re-sequencing of the allohexaploid wheat exome. *Plant Biotechnol. J.* 10, 733–742. doi: 10.1111/j.1467-7652.2012.00713.x
- Zhang, J., Nagai, C., Yu, Q., Pan, Y. B., Ayala-Silva, T., Schnell, R. J., et al. (2012). Genome size variation in three Saccharum species. *Euphytica* 185, 511–519. doi: 10.1007/s10681-012-0664-6
- Zhou, L., and Holliday, J. A. (2012). Targeted enrichment of the black cottonwood (*Populus trichocarpa*) gene space using sequence capture. *BMC Genomics* 13:703. doi: 10.1186/1471-2164-13-703

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Kaur and Gaikwad. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.