"PLANT EXOMICS IN CEREAL IMPROVEMENT"

PITAMBARA[1]; PRACHI SINGH[2] & POONAM KANANI[3]

[1,3]Department of Agricultural Biotechnology, Anand Agricultural University, Anand, Gujarat, India

[2]Department of Mycology and Plant Pathology, Institute of Agricultural Sciences,

Banaras Hindu University, Varanasi, Uttar Pradesh, India

ABSTRACT

Exome refers to entire protein coding sequence in an organism i.e. exon. Although exome represent almost a hundredth of whole genome. It don't provide complete picture of gene regulation they represent all sequences that code for proteins and regulate phenotype. It has advantages of producing sequencing data in a quicker way, since exome comprises only a small portion of genome. It is usually used to identify mutations in protein coding genes. Exome sequencing data are now becoming widely available for secondary uses. Many of the functional elements located outside exonic region is not sequenced (about 50% of SNPs are located in the intronic region). Exomics plays an important role in exploring biodiversity, to understand host pathogen interaction and investigating evolution of cereals.

KEYWORDS: EXOME, Intronic, SNPs and Organism

INTRODUCTION

Exomics pertains to the study of exons, which code for the protein region in whole genome of an organism. It includes a number of methods that are grouped broadly as template preparation, sequencing and imaging, and data analysis:

- **Template Preparation: Target-Enrichment:** Template preparation requires robust methods that produce a representative, non-biased source of nucleic acid material from the genome under investigation. Current techniques for targeted enrichment can be categorized according to the nature of their core reaction principle: 1. Polymerase-mediated capture 2. Hybrid capture (Solid-phase hybridization, Liquid-phase hybridization) 3. Molecular inversion probes (MIP) mediated capture.

- **Sequencing and Imaging:** After selection of target region (exons), sequencing is done using one of the new platforms. 1. Pyrosequencing: Roche/454 2. Reversible terminator-based sequencing: Illumina / Solexa Genome Analyzer 3. Sequencing-by-ligation: SOLiD

- **Data Analysis:** After acquisition of the image data, these recorded signals have to be converted into nucleotide bases. Furthermore, statistical models provide a measure of certainty of each base call in addition to the nucleotide itself. Then Bioinformatical analysis is done by alignment and recalibration, variant calling and variant annotation.

APPLICATIONS OF PLANT EXOMICS IN CEREALS

Exomics for Exploring Biodiversity

Brozynska *et al*. (2014) compared chloroplast genome sequences of different rice taxa and used the variant present between them to determine the relationships between the wild taxa and cultivated rice. It was concluded that the genetic difference between Taxon A and *O. rufipogon* (125) is comparable with that between Taxon A and *O. sativa japaonica* (125) and that between *O. rufipogon* and *O. sativa japonica* (118).

Exomics to Study Host–Pathogen Interactions

Thakur *et al*. (2015) evaluated 92 rice lines for blast disease occurrence and based on disease assessment scale found that 72 lines were resistant and the rest 20 lines were susceptible. It was also determined that the nucleotide diversity at the *Pi54* locus in rice lines and identified presence of a total of 197 single-nucleotide polymorphisms (SNPs).

Exomics for Investigating the Natural Evolution of Crops

Simons *et al*. (2006) screened M_2 generation of a population of Chinese Spring EMS mutants for the speltoid phenotype and found that mutant mq194 had a single base substitution in an *AP2* DNA binding domain (exon 5) that resulted in the change of amino acid from cysteine to tyrosine. It was also concluded that Q allelic polymorphism was due to one conserved nucleotide difference at position 329 where all Q containing genotypes possessed an isoleucine amino acid while all q containing genotypes possessed a valine amino acid.

CASE STUDIES

Shao *et al*. (2012) genotyped a set of 516 fragrant rice accessions and found that over 80% of them carried the *badh2.7* allele. Further from re-sequenced *Badh2* region it was found that the major genetic bases of fragrance lie in the 8 bp deletion in *Badh2* exon 7 and in the 7 bp deletion in *Badh2* exon 2.

Winfield *et al*. (2012) captured and characterized 56.5 Mb of genomic DNA of eight different UK allohexaploid wheat varieties and identified more than 500000 putative (SNPs) and found that out of total SNPs 80% were homeologous.

Mascher *et al*. (2013) developed and employed an in-solution hybridization-based sequence capture platform for barley to selectively enrich 61.6 Mb coding sequence target and showed that this exome capture platform provided a clear path towards a broader and deeper understanding of the natural variation residing in the mRNA-coding part of the barley genome which will thus constitute a valuable resource for genetic diversity analysis.

Udomchalothorn *et al*. (2014) identified the changes in the rice genome due to somaclonal variation that possibly lead to salt and/or draught tolerance characteristics and comparatively studied genomes of LPT123 and LPT123-TC171 and detected that mutations within the genes responding to both draught and salt stress were present in 493 positions while mutations within the genes responding to only salt stress were found in 100 positions.

Muraya *et al.* (2015) captured and enriched 29 Mb genomic DNA of 21 diverse inbred maize lines (7 flints, 14 dents) and sequenced it using the 454 NGS platform to 19.6-fold average depth coverage, and performed a broad evaluation of read alignment and variant calling methods to select optimal procedures for variant discovery. By sequence alignment with the B73 reference and de novo assembly, 383,145 putative single nucleotide polymorphisms (SNPs) were identified.

Jia *et al*. **(2016)** performed exome sequencing of mutant line 146 which was generated by gamma irradiation of B73 maize reference line and identified one deletion (4.8 kb) on chromosome 4 which contain 10 exons (SUGARY-1 gene (SU1)) and suggested that this deletion was responsible for increased sucrose concentration, decreased concentration of amylopectin and shrunken kernel phenotype in mutant line 146.

LIMITATIONS OF EXOMICS

- Some useful region of the genome is uncovered in exomics like: MicroRNAs (miRNAs), other noncoding RNAs, 3'- and 5'-UTR regions and pseudogenes

- Exon- exon boundaries are sometime not captured.

- Unknown introns interfere with the hybridization efficiency and capture success.

CONCLUSIONS

- Most cost-efficient sequencing approaches to conduct in less time period.

- Less data is generated which is easy to assemble and analyse.

- Exome data set from any candidate species can be utilized to classify organisms.

- Provides new opportunities to decode the complexity of host pathogen interactions.

- Efficient and unbiased identification of genes involved in crop evolution (tb1, Waxy, Q).

- Important role in understanding natural variation residing in the mRNA-coding part of cereal genome which can be used for genetic diversity analyses.

- Plays an important role in identifying the variants in cereal genome possibly leading to different stress resistance, quality and yield related characters.

FUTURE THRUST

- Requires integration of bioinformatics tools and multiple-omics platforms.

- Not designed specifically for the off target reads.

- Capture efficiency should be increased.

- Reference genome should be increased in order to sequence capture smaller functional gene.

- Robust computer algorithms/software solutions for large-scale genome sequence data.

- User-friendliness of data analysis tools is still questionable.

- Improvement of alignment tools to handle massive amounts of short reads.

- The next generation/ next-next generation sequencing technologies also need improvements.

REFERENCES

1. Brozynska, M., Omar, E. S. and Henry, R. J. (2014). *Tropical Plant Biol*, **7**:111–120

2. Jia, S., Li, A., Morton, K., and Holding, D. (2016). *Genes Genome Genetics*, **6**: 2385-2395

3. Mascher, M., Richmond, T. A., Gerhard, D. J. and Stein, N. (2013). *The Plant Journal*, **76**: 494–505

4. Muraya, M. M., Schmutze, T. and Altmann, T. (2015). *PLOS ONE*, **10**(1371):1-34

5. Samuels, D. C., Han, L., Li, J. and Clark, T.A. (2013). *Trends in Genetics*, **29**(10): 593-599

6. Shao, G., Tang, S. and Chen, M. (2013). *Genomics*, **101:** 157–162

7. Simons, K. J., Fellers, J. P. and Faris, J. D. (2006). *Genetics* **172**: 547–555

8. Singh, D., Singh, P. K., Chaudhary, S. and Kumar, S. (2012). *Advances in Genetics*, **79:** 87-121

9. Thakur, S., Singh, P. K. and Sharma,T.R. (2015). *Frontiers in Plant Science,* **6**(345):1-16

10. Udomchalothorn, T., Plaimas, K. and Chadchawan, S. (2014). *The plant genome*, **7**(3):1-11

11. Winfield, M. O., Wilkinson, P. A. and Edwards, K. J. (2012). *Plant Biotechnology Journal,* pp. 1–10