

ESTs as a resource for gene discovery and population genetic analysis of crop plants

Abstract

In the advent of the genomic era, it has resulted in the vast production of publicly available DNA sequence data, including large collections of expressed sequence tags (ESTs) from a variety of different taxa. ESTs can be useful for gene identification, and with the involvement of *in silico* methodologies and the pathway-based analyses help to elucidate the biological functions and interactions of genes. ESTs encoding putative transcription factors (TFs) were identified by sequence comparison to known TF gene families of crop plants in the Plant Transcription Factor Database. Comparative mapping analysis of ESTs with other crops genome could reveal the phylogenetic relationships of several important genes. The utilization of these ESTs with computational studies will increase the understanding of genetic basis of the crop plants. Then the molecular markers based on ESTs such as EST-SSR, EST-SNPs providing the good source for the population genetics studies and useful for the marker-assisted breeding. These studies will make a great impact in crop improvement programs and keep the global food security.

Keywords: crop plants, expressed sequence tags, gene discovery, genomics, molecular markers, population genetics

Volume 4 Issue 4 - 2017

Subramani Pandian, Pandiyan
Muthuramalingam, Manikandan Ramesh
Department of Biotechnology, Science campus, Alagappa
University, India

Correspondence: Manikandan Ramesh, Department of
Biotechnology, Science campus, Alagappa University, India,
Email mrbiottech.alu@gmail.com

Received: September 20, 2017 | **Published:** October 04, 2017

Abbreviations: EST, expressed sequence tags; SSR, simple sequence repeats; SNPs, single nucleotide polymorphism; TFs, transcription factors

Introduction

In recent years, next-generation sequencing technologies have emerged as a cutting-edge approach for high-throughput sequence determination, providing the powerful and cost-efficient tools for advanced research in many areas, including genome sequencing; genome-wide profiling of epigenetic marks and chromatin structure using ChIP-seq, methyl-seq and DNase-seq; metagenomics studies; and *de novo* transcriptome sequencing for non-model organisms.¹⁻³ These developments made easy the crop improvement by providing detailed information about the particular crop species. Rice, wheat and sorghum are the mostly consumed by the people throughout the world. Therefore, the researchers gave importance to these crops most widely.⁴ High throughput omics technologies also provide the identification of novel genes and metabolic pathways involving the expression of particular genes. Progress in novel gene discovery and molecular breeding in crop plants was under processing with the advent of high throughput technologies and the utilization of bioinformatics methodologies.^{5,6}

Transcriptome sequencing has, however caused a significant upshot in the expressed sequence tags (ESTs) collections in several plant species, including the non-model plant species.⁷ The ESTs, which are generated by large-scale single pass sequencing of randomly picked cDNA clones, have been cost-effective and valuable resource for efficient and rapid identification of novel genes and development of molecular markers.^{8,9} Further, ESTs have been employed in *in silico* analyses to identify the genes that are differentially expressed in various tissues, cell types, or developmental stages of the same or different genotypes.^{10,11} The collection of ESTs reflects the level and complexity of gene expression in the sampled tissue.

Publicly available datasets are useful for future crop research, such as elucidating the molecular mechanisms of specific traits and understanding the complexity of the transcriptome, gene expression regulation, and gene networks, gene annotation and discovery,¹²⁻¹⁴ comparative genomics,⁹ development of molecular markers,^{15,16} and population genomics studies of genetic variation associated with adaptive traits.^{6,17} One can map ESTs to specific chromosome locations using physical mapping techniques, such as radiation hybrid mapping, *in situ* hybridization, or FISH. Alternatively, if the genome of the organism that originated the EST has been sequenced, one can align the EST sequence to that genome using a computer.

Molecular markers play an important role in many aspects of plant breeding, such as identification of the genes responsible for desirable traits. Molecular markers have been widely used to map important genes and assist with the breeding of crops.^{18,19} In molecular breeding the ESTs are play important roles the simple sequence repeats (SSRs) which derived from ESTs are known as EST-SSR. With the availability of large numbers of ESTs and other DNA sequence data, development of EST-based SSR markers through data mining has become fast, efficient, and relatively inexpensive compared with the development of genomic SSRs.^{16,20,21} This is due to the fact that the time-consuming and expensive processes of generating genomic libraries and sequencing of large numbers of clones for finding the SSR containing DNA regions are not needed in this approach.²² The studies based on the ESTs are necessary to develop and to find simple way to unraveling the molecular crosstalk of the crop plants.

Conclusion

After genome sequencing and assembly, genome annotation is the most critical task in the characterization of the genetic blueprint of an organism. This method of using the ESTs with Bioinformatics will provide a timely tool for the annotation of the crop genome, using abundant ESTs from other cereals and plants. And giving importance

to EST based molecular markers significantly reveals the molecular base of a particular crop plant.

Acknowledgements

The author S. Pandian (UGC order no: F.25-1/2014-15 (BSR)/7-326/2011/BSR) thank the University Grants Commission, New Delhi, India for financial support in the form of fellowship. The authors gratefully acknowledge the computational and Bioinformatics facility provided by the Alagappa University Bioinformatics Infrastructure Facility (funded by the Department of Biotechnology, Government of India; Grant No.BT/BI/25/001/2006).

Conflict of interest

The author declares no conflict of interest.

References

- Metzker ML. Sequencing technologies—the next generation. *Nat Rev Genet.* 2010;11(1):31–46.
- Ray S, Satya P. Next generation sequencing technologies for next generation plant breeding. *Front Plant Sci.* 2014;5:367.
- Reuter JA, Spacek DV, Snyder MP. High-throughput sequencing technologies. *Mol Cell.* 2015;58(4):586–597.
- Varshney RK, Terauchi R, McCouch SR. Harvesting the promising fruits of genomics: Applying genome sequencing technologies to crop breeding. *PLoS Biol.* 2014;12(6):e1001883.
- Zhuang J, Zhu B. Analysis of Brassica napus ESTs: gene discovery and expression patterns of AP2/ERF-family transcription factors. *Mol Biol Rep.* 2014;41(1):45–56.
- Wang JX, Lu C, Yuan CQ, et al. Characterization of ESTs from black locust for gene discovery and marker development. *Genet Mol Res.* 2015;14(4):12684–12691.
- Unamba CI, Nag A, Sharma RK. Next generation sequencing technologies: the doorway to the unexplored genomics of non-model plants. *Front Plant Sci.* 2015;6:1074.
- Sreenivasulu N, Kavikishor PB, Varshney RK, et al. Mining functional information from cereal genomes—the utility of expressed sequence tags. *Curr Sci.* 2002;83(8):965–973.
- Kumari K, Muthamilarasan M, Misra G, et al. Development of eSSR-markers in *Setaria italica* and their applicability in studying genetic diversity, cross-transferability and comparative mapping in millet and non-millet species. *PLoS One.* 2013;8(6):e67742.
- Ogihara Y, Mochida K, Nemoto Y, et al. Correlated clustering and virtual display of gene expression patterns in the wheat life cycle by large-scale statistical analyses of expressed sequence tags. *Plant J.* 2003;33(6):1001–1011.
- Ronning CM, Stegalkina SS, Ascenzi RA, et al. Comparative analyses of potato expressed sequence tag libraries. *Plant Physiol.* 2003;131(2):419–429.
- Emrich SJ, Barbazuk WB, Li L, et al. Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Res.* 2006;17(1):69–73.
- Bouck A, Vision T. The molecular ecologist's guide to expressed sequence tags. *Mol Ecol.* 2007;16(5):907–924.
- Muthuramalingam P, Krishnan SR, Pothiraj R, et al. Global Transcriptome Analysis of Combined Abiotic Stress Signaling Genes Unravels Key Players in *Oryza sativa* L.: An In silico Approach. *Front Plant Sci.* 2017;8:759.
- Barbazuk WB, Emrich SJ, Chen HD, et al. SNP discovery via 454 transcriptome sequencing. *Plant J.* 2007;51(5):910–918.
- Babu BK, Pandey D, Agrawal PK, et al. In-silico mining, type and frequency analysis of genic microsatellites of finger millet (*Eleusine coracana*): a comparative genomic analysis of NBS-LRR regions of finger millet with rice. *Mol Biol Rep.* 2014;41(5):3081–3090.
- Namroud MC, Beaulieu J, Juge N, et al. Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Mol Ecol.* 2008;17(16):3599–3613.
- Muthamilarasan M, Venkata Suresh B, Pandey G, et al. Development of 5123 intron-length polymorphic markers for large-scale genotyping applications in foxtail millet. *DNA Res.* 2013;21(4):41–52.
- Satish L, Shilpha J, Pandian S, et al. Analysis of genetic variation in sorghum (*Sorghum bicolor*) genotypes with various agronomical traits using SPAR methods. *Gene.* 2016;576(1 Pt 3):581–585.
- Gupta PK, Rustgi S, Sharma S, et al. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Mol Genet Genom.* 2003;270(4):315–323.
- Nirgude M, Babu BK, Shambhavi Y, et al. Development and molecular characterization of genic molecular markers for grain protein and calcium content in finger millet (*Eleusine coracana*). *Mol Biol Rep.* 2014;41(3):1189–1200.
- Eujayl I, Sledge MK, Wang L, et al. Medicago truncatula EST-SSRs reveal cross-species genetic markers for Medicago spp. *Theor Appl Genet.* 2004;108(3):414–422.