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

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 has been employed 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, happy mapping, 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.
ESTs as a Resource for Gene Discovery and Population Genetic Analysis of Crop Plants

[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.

Acknowledgement

The author S. Pandian (UGC order no: E25-1/2014-15 (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).

References