Lecture 33: Expressed Sequence Tags

Now a days, genome analysis has employed a rapid analysis tool known as expressed sequence tags. In 1983, SD Putney for the first time demonstrated the use of cDNA in identification of genome. The term expressed sequence tags (ESTs) was coined by Anthony Kerlavage at the Institute for Genomic Research. In 1991, Mark Adams used EST in relation to gene discovery and Human Genome project.

cDNA or complementary DNA is reversely transcribed from mRNA using reverse transcriptase enzyme. cDNA is widely used in cloning of genes in eukaryotes.

Expressed Sequence Tags or ESTs, as the name suggests, are the new generation tools providing new dimension to transcriptome analysis. They are the tiny sequences of cistron randomly selected from genome library and can be used to identify and map the whole genome of any particular species. ESTs are usually 200 to 500 nucleotides long and are generated by sequencing the ends of DNA.

ESTs can be obtained without much expenditure and are quite fast in genomic analysis. The EST sequences can be used to search the homologous organisms in different databases such as NCBI (National Centre for Biotechnology Information). Thus we can collect information on expression patterns of different species. Therefore, they play vital role in discovery of gene and genome analysis.

Cistron (or gene) is a portion of DNA which specifically transcribe to form mRNA and finally helps in synthesis of protein.

Generation of Expressed Sequence Tags: The presence of introns makes gene identification quite difficult. The DNA is firstly transcribed to mRNA which is the key for synthesis of building blocks i.e. proteins by a process called translation. Interestingly, mRNAs do not contain the sequence transcribed from introns. Thus mRNA isolation is the key for ESTs construction. mRNA is quite unstable outside the cell, therefore reverse transcription id performed to convert it to cDNA, which is comparatively stable.

Translation is the process of protein synthesis from mRNA on the surface of ribosomes.

Expressed Sequence tags are generated from cDNA cloned from mRNA of any particular species (refer to the Figure 1). As the cDNA used is complementary to mRNA, the ESTs represent portions of expressed genes.

The ESTs can be generated by following steps:

Transcription of Genomic DNA: Genomic DNA is first transcribed to generate Nascent mRNA followed by splicing of synthesize perfect mRNA.

Reverse transcription of mRNA: mRNA can also be directly isolated from the species by using different kits (e.g. RNAgent Promega). mRNA synthesized undergoes reverse transcription to form cDNA library.

Generation of ESTs: From the cDNA library 5' or 3'-ESTs are generated by cDNA end sequencing. 5' EST is formed from a region of transcript which forms protein whereas the ending portion of cDNA forms 3'EST.

Assembly and organization of ESTs: The constructed ESTs can then be assembled separately in multimember sequence assembly, Bridged sequence assembly and small clusters on the basis of size of ESTs.

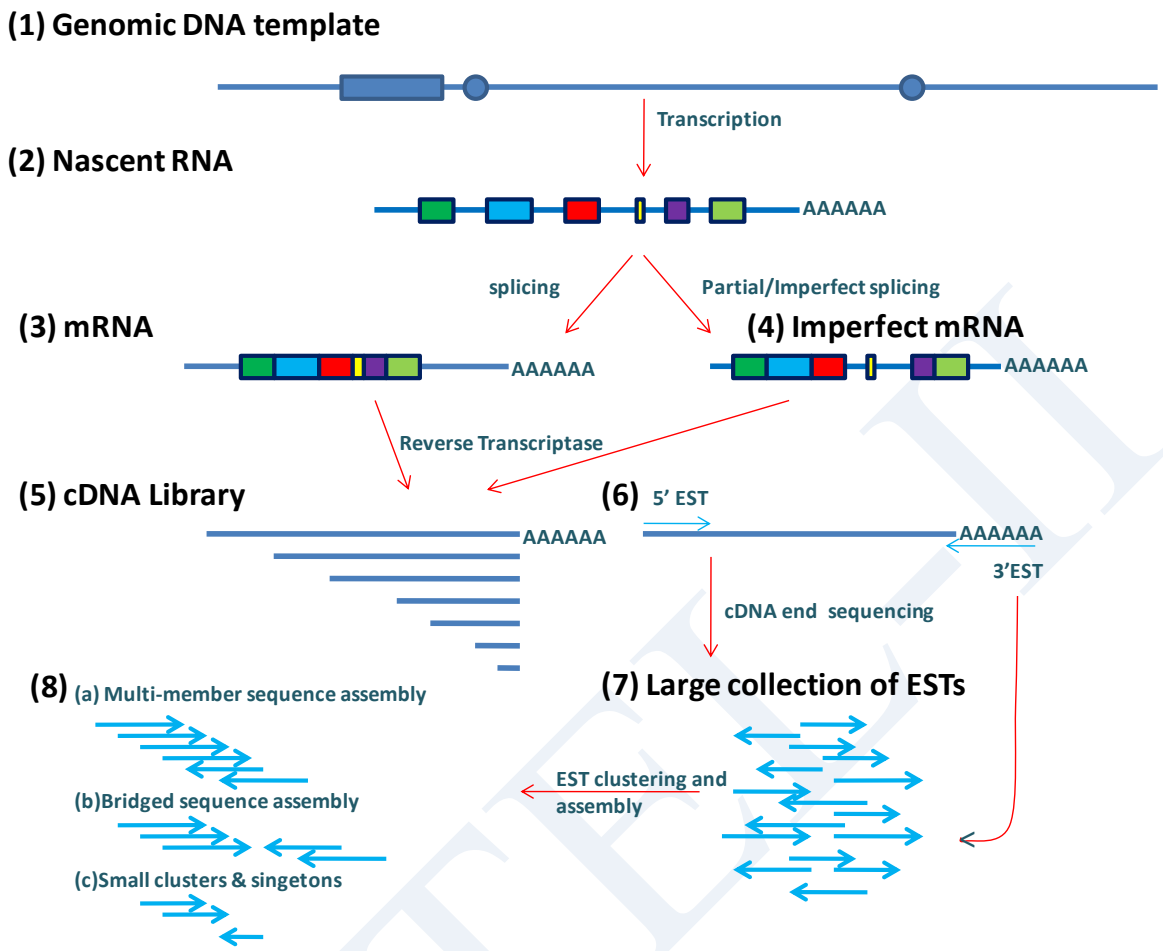

Figure 1 : Method of construction of ESTs from nascent DNA. The process involves transcription of nascent DNA, reverse transcription of mRNA and finally EST synthesis and clustering.

Applications of ESTs

ESTs can be used as functional DNA arrays for the analysis of whole genome of a species.

ESTs as Genome searching tool: Scientists use the genome map to travel through the billions of nucleotide in the genome of a species. ESTs work as landmarks for the genome mapping. Currently, the most powerful technique for genome mapping is Sequence Tagged Site (STS) mapping. 3'-ESTs as a commonly used as a source of STSs because of their uniqueness in an individual's genome (Figure 2)

ESTs are most commonly used to locate a gene rapidly and accurately.

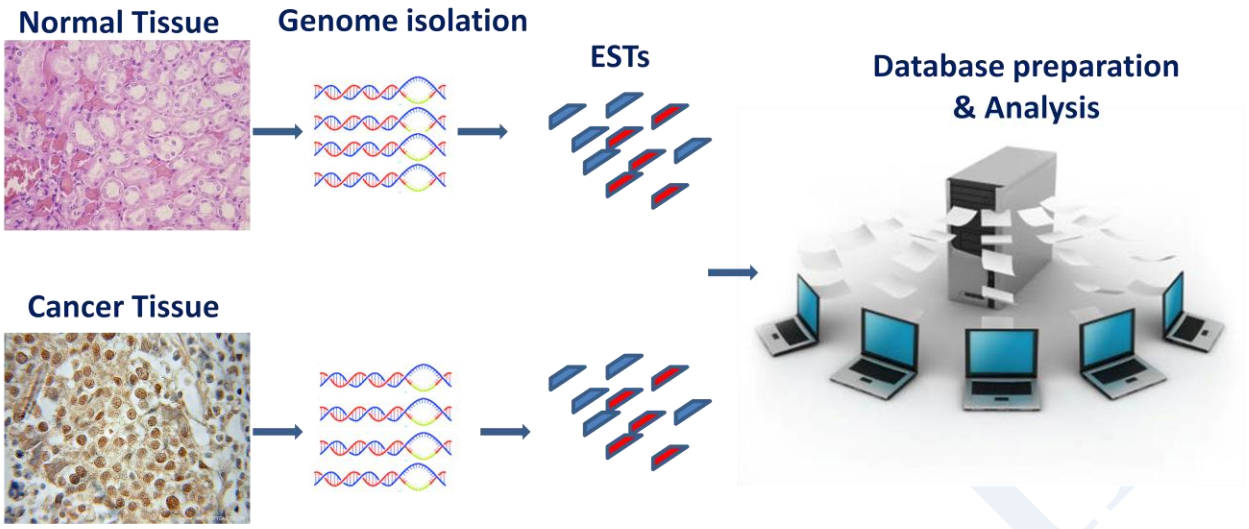

Figure 2: Identification of genes using Expressed sequence Tags: Genome is isolated from normal and diseased tissue, ESTs are used to identify the genetic variations between these two tissue. Finally a database is prepared for further reference.

Identification of position of genes: ESTs are widely and efficiently used to locate an already known gene in the genome of a species. Most commonly used technique for this purpose is called as Sequence Tagged Site (STS) mapping.

To find out gene responsible for disease: This method allows us to identify the genes responsible for any deformity or disease such as Alzheimer's and colon cancer have already been investigated.

ESTs and Human Genome Project: Existence of thousands of genes have been identified in Human Genome Project solely on the basis of ESTs.

Human Genome Project was initiated by U.S. Department of Energy and the National Institutes of Health to sequence the complete genome of Homo sapience and completed in 2003.

Use of ESTs in similarity searches: Due to their putative and fast behavior, NCBI has included millions of ESTs databases for several species. Scientists as well as genome sequencing centers are widely using these ESTs for similarity searching between different species.

BRIEFINGS IN BIOINFORMATICS. VOL 8. NO 1. 6–21
Advance Access publication May 23, 2006
doi:10.1093/bib/bbl015

Further reading and recommended article

A hitchhiker's guide to expressed sequence tag (EST) analysis

Shivashankar H. Nagaraj, Robin B. Gasser and Shoba Ranganathan

Received 24th April 2006

Abstract

Expressed sequence tag (EST) sequencing projects are underway for numerous organisms, generating millions of short, single-pass nucleotide sequence reads, accumulating in EST databases. Extensive computational strategies have been developed to organize and analyse both small- and large-scale EST data for gene discovery, transcript and single nucleotide polymorphism analysis as well as functional annotation of putative gene products.

We provide an overview of the significance of ESTs in the genomic era, their properties and the applications of ESTs. Methods adopted for each step of EST analysis by various research groups have been compared. Challenges that lie ahead in organizing and analysing the ever increasing EST data have also been identified.

The most appropriate software tools for EST pre-processing, clustering and assembly, database matching and functional annotation have been compiled (available online from http://biolinfo.org/EST). We propose a road map for EST analysis to accelerate the effective analyses of EST data sets. An investigation of EST analysis platforms reveals that they all terminate prior to downstream functional annotation including gene ontologies, motif/pattern

Further reading and recommended article

African Journal of Biotechnology Vol. 7 (4), pp. 331-341, 19 February, 2008
Available online at http://www.academicjournals.org/AJB
ISSN 1684–5315 © 2008 Academic Journals

Review

Expressed sequence tags (ESTs) and single nucleotide polymorphisms (SNPs): Emerging molecular marker tools for improving agronomic traits in plant biotechnology

Kwadwo Owusu Ayeh

Department of Plant and Environmental Sciences, The Norwegian University of Life Sciences (UMB), Postboks 5003, 1432, ÅS-NORWAY. E-mail: kwadwo.owusu.ayeh@umb.no. Tel: + 47-64965606.

Accepted 16 January, 2008

Expressed Sequence Tags (ESTs) and Single Nucleotide Polymorphisms (SNPs) are providing in depth knowledge in plant biology, breeding and biotechnology. The emergence of many novel molecular marker techniques are changing and accelerating the process of producing mutations in plant molecular biology research. This coupled with the availability of cheap sequencing techniques and access to a complete genome sequence has been shown to complement traditional marker –based approaches. Expressed Sequence Tags (ESTs) have provided an important source for the study of Single Nucleotide Polymorphisms (SNPs) in plants. SNP markers have become popular, partly because of their high density within the genome and also their ease with which they are characterized. This review also focuses on some methods used in genotyping SNPs.

Further reading and recommended article

Single Nucleotide Polymorphism (SNP) Genotyping Techniques—An Overview

Richard M. Twyman
University of York, York, U.K.

INTRODUCTION

Single nucleotide polymorphisms (SNPs) are individual base positions in the genome that show natural variation in a population. They represent the most abundant form of genetic variation in humans, accounting for more than 90% of all differences between unrelated individuals. SNP patterns are likely to influence many human phenotypes; therefore large-scale association studies based on SNP genotyping are expected to help identify genes affecting complex diseases and responses to drugs or environmental chemicals. SNPs these follow the fate of a label either in real time or at the assay end point. Uniquely, mass spectrometry can be used to detect the allele-specific product of a discrimination assay without the need for a label, by distinguishing the masses of DNA molecules containing alternative bases.

ALLELE DISCRIMINATION METHODS

Allele-Specific Hybridization

http://www.writescience.com/RMT%20PDFs/Elsevier/Twyman%2005%20EMGP.pdf