ESTs in Plants: Where Are We Heading?

Sameera Panchangam, Nalini Mallikarjuna, and Prashanth Suravajhala

Abstract

Expressed sequence tags (ESTs) are the most important resources for transcriptome exploration. Next-generation sequencing technologies have been generating gigabytes of genetic codes representing genes, partial and whole genomes most of which are EST datasets. Niche of EST in plants for breeding, regulation of gene expression through miRNA studies, and their application for adapting to climatic changes are discussed. Some of the recent tools for analysis of EST exclusive to plants are listed out. Systems biology though in its infancy in plants has influenced EST mapping for unraveling gene regulatory circuits, which is illustrated with a few significant examples. This review throws a glance at the evolving role of ESTs in plants.

Keywords

Expressed Sequence Tags (ESTs) • Plant ESTs • EST analysis pipelines • miRNA • Systems biology

1 Introduction

Bioinformatics has provided us an impetus to learn systems biology. The bioinformatics tools have not only allowed us to understand what systems biology could make use of but also on how it dissects the behavior of complex biological organization and processes in terms of molecular constituents. It involves the study of all genes expressed as messenger RNAs and characterization of the proteins and metabolites under different conditions (Kirschner 2005). Significant advancement in high-throughput (HT) technologies such as microarrays, automated sequencing, and mass spectrometry has generated huge amount of data which can be optimized by various computational tools for accelerated process of discovery. Access to a number of next-generation sequencing (NGS) technologies such as Roche/454, Illumina, and ABI SOLiD has drastically reduced the cost and time of sequencing and increased the length of sequence reads. These NGS technologies are being utilized

S. Panchangam (✉) • N. Mallikarjuna
Department of Cell Biology, International Crop Research, Institute for Semi-Arid Tropics, Patancheru 502319, AP, India
e-mail: sameera.panchangam@gmail.com

P. Suravajhala
Bioclues Organization, IKP Knowledge Park, Picket, Secunderabad 500 009, Andhra Pradesh, India

K.K. P.B. et al. (eds.), *Agricultural Bioinformatics*,
DOI 10.1007/978-81-322-1880-7_9,

for de novo sequencing, genome re-sequencing, whole genome, and transcriptome analysis (Morozova and Marra 2008). Despite these advantages and availability of whole genome sequences of more than 180 organisms (http://www.genomenewsnetwork.org/; http://www.ebi.ac.uk/genomes/), the plethora of datasets constituting umpteen genomes is not fully understood. Therefore, it is believed that expressed sequences tags (EST) especially from unsequenced genomes will continue to play an important role in post genome sequencing and will apply NGS technologies in transcriptome sequencing. "Poor man's genome" as they are known, ESTs are short (200–800 nucleotide bases in length), unedited, randomly selected, single-pass sequence reads derived from cDNA libraries (Adams et al. 1991; Nagaraj et al. 2006). Since the use of ESTs as the primary source of human gene discovery in 1991, there has been manifold growth in the generation and accumulation of EST data for a range of organisms from bacteria to vertebrates (Lee and Shin 2009). In combination with NGS, ESTs have proven to be an extremely valuable resource for high-throughput gene discovery, identification of novel genes, splice variants, gene location, and intron–exon boundaries within genomic sequence assemblies. They are a cost-effective alternative to whole genome sequencing (WGS), for annotation of genes and development of molecular markers in organisms with large genome size and in species which lack draft genome sequences (Dias et al. 2000).

2 Identifying Niche of ESTs for Desired Traits in Plants

Plant breeders constantly strive to develop improved varieties of crops for desirable traits through conventional breeding techniques which are laborious and time-consuming as careful phenotypic and genotypic selection is needed. Most of the traits of interest in plant breeding such as high yield, height, drought resistance, disease resistance in many species, etc., are quantitative, also called polygenic, continuous, multifactorial, or complex traits, which further complicate the breeding program (Semagn et al. 2010). However, advances in genomics and DNA marker technology have helped to develop molecular markers, which are now widely used to track loci and genome regions in several crop-breeding programs. With this molecular markers tightly linked with a large number of agronomic and disease resistance, traits have become available in major crop species (Jain et al. 2002; Gupta and Varshney 2004). Some sequence tagged sites (STS) are also enriched and have potentially been used as markers for PCR (polymerase chain reaction). Most of these markers developed in the past were related to genomic DNA (gDNA) and therefore could belong to either the transcribed region or the non-transcribed region of the genome. These markers were termed as random DNA markers (RDMs) (Andersen and Lübberstedt 2003). As a result, a large number of genes have been identified in the recent past through "wet lab" as well as in silico studies, and a wealth of sequence data have been accumulated in public databases (e.g., http://www.ncbi.nlm.nih.gov; http://www.ebi.ac.uk) in the form of BAC (bacterial artificial chromosome) clones, ESTs, full-length cDNA clones, and genes. The availability of enormous amount of sequence data from complete or partial genes has made it possible to develop molecular markers directly from the parts of genes (Varshney et al. 2007). Genic molecular markers (GMMs) that developed from coding sequences like ESTs or fully characterized genes frequently have been assigned known functions. EST-based markers such as SSRs (simple sequence repeats), RFLPs (restriction fragment length polymorphisms), AFLPs (amplified fragment length polymorphisms), and SNPs (single nucleotide polymorphisms) and novel markers such as expressed sequence tag polymorphisms (ESTPs), conserved orthologous set (COS) markers, etc., have been developed for many crop species (Gupta and Rustgi 2004). Orphan crops like peanut, sorghum and millets, groundnut, cowpea, common bean, chickpea, pigeon pea, cassava, yam, and sweet potato (Varshney et al. 2012) and many other important horticultural and forest species with large and complex genomes whose whole genome sequences are not yet available greatly benefit from the EST data.

For example, genes encoding key enzymes for fatty acid and seed storage protein biosynthesis, bacterial wilt disease, and novel genes discovered in peanut were derived from ESTs belonging to different tissues, different growth stages, and under different abiotic and biotic stresses (Feng et al. 2012).

More recently, microRNAs (miRNA) have received a lot of attention due to their role in regulation of gene expression which finds applications in functional genomics and study of various pathways in organisms. In plants, miRNAs are involved in diverse aspects of growth and development such as leaf morphology and polarity, root formation, transition from embryogenic to vegetative phase, flowering time, floral organ identity, and reproduction (Mallory and Vaucheret 2006; Sun 2012). They are also found to be involved in defense mechanisms, hormone signaling, and abiotic and biotic stress responses (Lu et al. 2008). 21,264 entries representing hairpin precursor miRNAs, expressing 25,141 mature miRNA products, in 193 (>170 plants) species are available (www.mirbase.org/). It is generally accepted that plant miRNAs have extensive complementarity to their targets, and their prediction usually relies on the use of empirical parameters deduced from known miRNA–target interactions. The biogenesis of miRNAs suggests that it is possible to find new miRNAs by homology searching of known miRNAs in ESTs, especially in plants whose whole genome sequence data is unavailable (Sunkar and Jagadeeswaran 2008). Since ESTs represent transcribed sequences, their analyses provide direct evidence for miRNA expression through simple tools for comparative genomics which in turn helps in identification of conserved miRNAs (Zhang et al. 2005). Both experimental methods and computational approaches have been adopted to identify miRNAs in plants, and the latter has been identified as the simplest and most effective method (Sun 2012). Several groups have attempted to identify novel miRNAs and decode their interaction with protein coding transcripts by examining ESTs (Nasaruddin et al. 2007; Das and Mondal 2010; Boopathi and Pathmanaban 2012; Muvva et al. 2012). Despite the tremendous applications of miRNA in plant biotechnology and the growing interest, our knowledge about the regulatory mechanisms and functions of miRNAs remains very limited (Liu et al. 2012). The limited number of experimentally validated miRNA targets, the spatio-temporal specific regulation of miRNA, and the lack of graphical-user interface models without the need for programming skills are major constraints. However, user-friendly software packages, which enable computational identification of miRNA and its target (C-mii), functional annotation of miRNAs (miRFANS), transcription factor–miRNA regulation (TransmiR), PMRD, etc., are now publicly available which are exclusive to plants (Liu et al. 2012; Numnark et al. 2012).

"Climate change," "sustainable agriculture," and "Ecogenomics" are some of the paradigms that have influenced researches of late. Genomics and bioinformatics have great potential in addressing various topics in these areas through approaches such as association mapping, genome scans, transcript profiling, and gene regulatory networks, thus leading to an understanding of the genetic architecture of climate change adaptation (Franks and Hoffman 2012). Gene transcription profiling, in particular, is one important step toward identifying those genes and metabolic pathways that underlie ecologically important traits, and ESTs can bridge genomics and molecular ecology because they can provide a means of accessing the gene space of almost any organism (Bouck and Vision 2007). EST libraries are a cost-effective tool to characterize genes important under particular conditions, as well as the starting point for the development of molecular genetic markers, such as gene-linked microsatellites and single nucleotide polymorphisms (SNP). In marine species, gene-linked microsatellites (EST-SSR = simple sequence repeats) were successfully identified, for example, in the ecologically important sea grass *Zostera marina* (eelgrass) to elucidate the molecular genetic basis of adaptation to environmental extremes. Approximately one-third of the eelgrass genes were characteristic for the stress response of the terrestrial plant model *Arabidopsis thaliana* (Reusch et al. 2008). Similarly, EST-based SSR markers for breeding of drought-resistant durum wheat in Mediterranean

dry lands (Habash et al. 2009), over 400 markers for various traits in important tropical fruits like mango and banana (Arias et al. 2012), and linkage mapping studies and identification of markers for beech bark disease resistance in American beech (Mason et al. 2013) are some recent examples of the potential application of ESTs in varied species for adaptation to climate change.

3 ESTs in Plants: Various Pipelines for EST Analysis

The number of EST entries in GenBank dbEST is 74,186,692 as on January 1, 2013 (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). Handling the huge and ever accumulating data efficiently is an important and daunting task (Pertea et al. 2003). Since ESTs are single-pass reads and represent only a small portion of the mRNA, they are prone to errors and inherent deficiencies. Problems such as low-quality regions within the sequence, redundancy, differentially expressed genes in the host, contaminants like vectors, linkers, chimeric sequences, and natural sequence variations need to be dealt with, before further analysis. Several tools have been developed for each of the steps involved in EST analysis in the past few years (Hotz-Wagenblatt et al. 2003; Mao et al. 2003; Kumar et al. 2004; Conesa et al. 2005). A generic protocol of the different steps in the analysis of EST datasets and a list of various tools has been dealt with in considerable detail by Nagaraj et al. (2006). Some of the steps require the use of intensive computing power and an in-depth knowledge of bioinformatics which is not available to small research groups without access to bioinformatics personnel and advanced computer systems. As rightly pointed out by many researches, an ideal EST analysis tool should possess a few characteristics such as (1) to be fully automated in a pipeline covering all the steps from the input chromatogram files to a clean, annotated web-searchable EST database; (2) to be highly modular and adaptable; (3) to be able to run in parallel in a personal computer (PC) cluster, thus benefiting from the multiprocessing capabilities of these systems; (4) to use third-party freely available programs, in order to ease the incorporation of the improvements made by others programmers; (5) to include a highly configurable and extensible user-friendly interface to perform data mining by combining any search criteria, fitting the final user needs; and (6) to be based on an open-source license to allow a continuous development by a community of users and programmers, as well as its customization for the needs of different projects (Forment et al. 2008). As new tools are being constantly developed and the existing ones being updated to meet the requirements, a few of the most recent tools are listed here (Table 1).

4 Systems Biology and Impact on EST Mining

Structural genomics and, more recently, functional genomics have become the base of sustainable agriculture, forestry, industry, and environment (Campbell et al. 2003; Diouf 2003; Mazur et al. 1999; Somerville and Somerville 1999; Walbot 1999). Much of the efforts were directed toward the identification of markers for agronomic traits and physical and nutritional traits, genes encoding biosynthetic enzymes and production of secondary and intermediary metabolites, and understanding of the biochemical pathways in crop and some forage plants (Girke et al. 2003; Sweetlove et al. 2003; Varshney et al. 2007). Systems biology has created sweeping changes in our approach to genomics and plant biology. The focus now is on the molecular, cellular, and organismic changes in plants such as totipotency (dedifferentiation and regeneration ability), apomixis (vegetative seed production), embryogenesis (somatic, zygotic, and microspore), induction of haploids, heterosis or hybrid vigor, flower development, symbiotic nitrogen fixation, etc. For example, transcriptomic, proteomic, and metabolomic studies have led to a deeper understanding of microspore embryogenesis in barley (*Hordeum vulgare* L.), rapeseed (*Brassica napus* L.), tobacco (*Nicotiana* spp.), wheat (*Triticum aestivum* L.), and maize (*Zea mays*), which are now considered model species to study the mechanisms of stress-induced androgenesis (Maraschin et al. 2005). Analysis of

Table 1 EST analysis tools developed after the year 2006

Name	Description	Category	Reference
EST2uni	Processing, clustering, annotation	F/D	Forment et al. (2008)
ESTPiper	Sequencing, assembly, annotation, probe design	F/W/D	Tang et al. (2009)
ESTPass	Processing, annotation		Lee et al. (2007)
ESMP	EST-SSRs pipeline	F/W	Sarmah et al. (2012)
ParPEST	Parallel computing	RA	D'Agostino et al. (2005)
PESTAS	Processing, assembly, annotation	RA/W	Nam et al. (2009)
SCRAF	Sort and assemble 454-EST sequences	F/W	Barker et al. (2009)
OREST	Analysis, annotation	F/W	Waegele et al. (2008)
ConiferEST	Conifer EST mining, processing, annotation	F/W	
KAIKObase	Silkworm database	F/W	Shimomura et al. (2009)
OrchidBase	Processing, clustering, annotation	F/W/D	Tsai et al. (2013)
GarlicEST	Mining, annotation, expression profiling	F/W	Kim et al. (2009)
TomatoEST	Tomato functional genomics data	F/W	Agostino et al. (2007)
MELOGEN	Melon EST database	RA	González-Ibeas et al. (2007)
bEST-DRRD	Barley ESTs involved in DNA repair and replication	F/W	Gruszka et al. (2012)
MoccaDB	Orthologous markers in Rubiaceae	F/W	Plechakova et al. (2009)

F free, *W* web based, *D* downloadable, *RA* restricted access

20,000 ESTs from fresh and cultured microspores of barley revealed clusters of differentially expressed genes and identification of 16 genes which could serve as markers for induction of androgenesis and progression of microspore embryogenesis (Malik et al. 2007). Strategies with fluorescent-labeled probes for in situ hybridization and immunofluorescence have provided unique images of the spatial and temporal pattern of the expression of genes and proteins and of the subcellular rearrangements that accompany microspore embryogenesis (Testillano and Risueño 2009). Another key trait that has defied scientific unraveling is the phenomenon of heterosis (Bircher et al. 2003). A systems biological approach to define how plant genomes interact to create phenotype is needed to arrive at a final resolution of this phenomenon.

Metabolic engineering and synthetic biology are an integral part of systems biology. From an engineering perspective, synthetic biology insists on standardized parts (e.g., genes, proteins, circuits) that can be assembled using bioinformatics and simulation tools to build functionality (Osbourn et al. 2012). Though they are still at infancy in plant research, the impact of systems biology on plants is ever increasing and well documented (Fernie 2012). Traditionally for gene detection, the two main approaches are EST mapping and computational gene prediction combined with homology-based search methods (Wortman et al. 2003). Cometh systems biology, the combination of two or more approaches, has helped in improved annotation of the genome and identification of novel genes and proteins (Allmer et al. 2006). These technologies provide validation

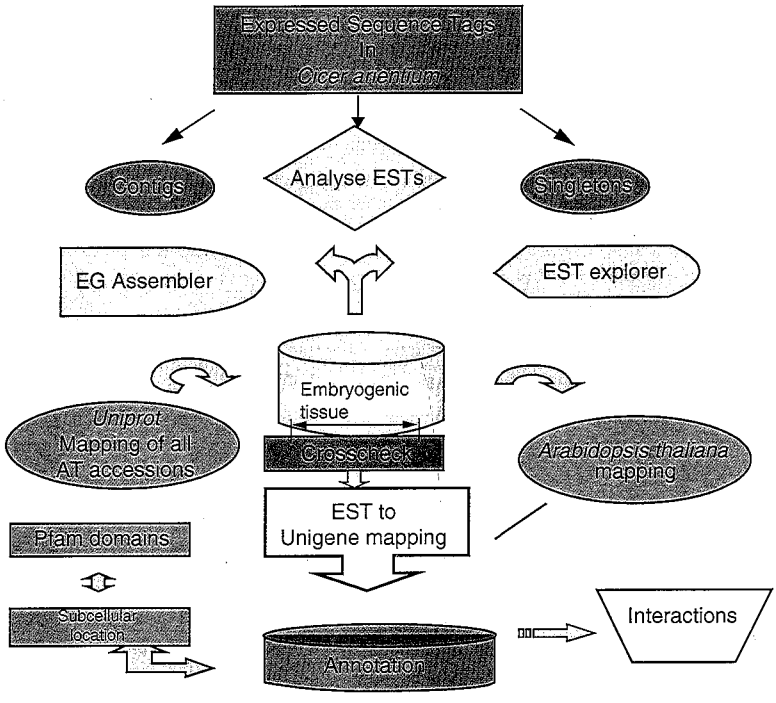

Fig. 1 Steps involved in analysis of Cicer ESTs

of the in silico gene models and enable fast and comprehensive analysis of the molecular plant phenotype (Naumann et al. 2007; Weckwerth 2008) as well as providing complementary means for probing the completeness of genome annotations. A case in example is integrated analysis of the molecular repertoire of *Chlamydomonas reinhardtii,* wherein bioinformatics annotation methods combined with GCxGC/MS-based metabolomics and LC/MS-based shotgun proteomics profiling technologies have been applied to characterize abundant proteins and metabolites, resulting in the detection of 1,069 proteins and 159 metabolites. By integrating genomic annotation information with experimentally identified metabolites and proteins, a draft metabolic network for *Chlamydomonas* was constructed which also provides entry points for further targeted gene discovery or biochemical pathway research (May et al. 2008). Metabolomics integrated with transcriptomic and proteomic studies have led to the identification of key steps involved in response to nitrogen deficiency in maize (Amiour et al. 2012). Yet another example of the application of EST analysis for discerning organization of cells besides predicting biological functions and providing insight into a variety of biochemical processes is the construction of protein interaction networks (PIN) (Guan and Kiss 2008). Despite the availability of advanced methods connecting orthology mapping and comparative approaches for predicting PIN, annotation of those proteins like "predicted" or "similar to" or "hypothetical" poses many challenges. To tackle this, a six-point classification system to validate protein interactions based on diverse features was proposed by Suravajhala and Sundararajan (2012). Using the six-point classification system, the genes related to embryogenesis and apomixis in chickpea were predicted based on the model apomictic plants Poa, Pennisetum, and apomeiotic mutant *Arabidopsis thaliana* (Panchangam et al. 2012). Here, EST analysis pipeline employed for annotation of proteins related to embryogenesis in chickpea is represented as a flowchart (Fig. 1).

Systems biology approach is not limited to crop plants and breeding but is also finding its way into unraveling different metabolic pathways in fruits, vegetables, and aromatic plants. A combined metabolomic, proteomic, and transcriptomic approach

was employed to investigate fruit development in tomato which led to identification of a novel gene regulatory mechanism for ethylene biosynthesis during the post climacteric ripening of the fruit (Van de Poel et al. 2012). A similar study was carried out in apple to obtain proteome information on fruit ripening in response ethylene treatment (Zheng et al. 2013). A database of molecular networks occurring in grapevine was built based on EST datasets, leading to 39,423 unique potential genes and proteins. Among them, 7,265 genes have been assigned to 107 pathways, including 86 metabolic pathways, 3 transporter pathways, 9 genetic information processing pathways, and 9 signal pathways focused mainly on phytohormone signaling (Grimplet et al. 2008). Metabolic pathways occurring in many medicinal and aromatic plants have been reviewed by Khanuja et al. (2012).

5 Conclusion and Future Directions

EST analysis holds an important spot in plant breeding by not only aiding the development of molecular markers for traits and annotation of genes but also providing insights into key developmental processes, regulation of gene expression, and to reveal the complete proteomic repertoire of an organism (Nagaraj et al. 2007). Although EST databases are no substitute for whole genome scaffolds, they certainly played a key role in pre-genome sequencing era and will continue to be promising resources for various in vitro and in silico experiments (Feng et al. 2012). The ability to generate large amount of data has become quick and cheap due to NGS technologies and has transformed various areas of biology which were previously unattainable, particularly for non-model systems that lack extensive genomic resources. Next-generation sequencing has great potential for accurate transcriptome characterization because of the large amount of data obtained at considerably lower costs compared to traditional methods, and with the decreasing costs transcriptome sequencing will be dramatically improved in the near future. EST sequencing along with NGS technologies is revolutionizing applications that revolve around gene expression. With deeper sequencing (e.g., 6–20 plates), researchers attain a level of transcriptome that has never been possible before due to the higher cost of earlier technologies. Not only will these studies sequence more than 90 % of the transcriptome, the coverage per gene will approach traditional sequencing. This should allow researchers to use these genes to identify pathways and determine tissue-specific expression for lowly expressed genes and will be critical for genome annotation (Kerr Wall et al. 2009). In retrospection, ESTs thus do not lose to whole genome sequencing, but coupled with NGS technologies and simulation/computational tools, they have revolutionary applications for both sequenced and unsequenced genomes. The large-scale development of tools for analysis of genes, transcripts, and proteins has generated vast data which holds great promise for revealing novel plant biology. The focus now is a systems perspective with the cumulative "omics" approach (e.g., genomics, epigenomics, transcriptomics, proteomics, metabolomics, interactomics, ionomics, phenomics, etc.) (Liberman et al. 2012). The way to sustainable agriculture in the very near future is to move from genetic manipulation of parts of genomes to more engineering-based approach, combining traditional plant breeding techniques with systems biology and predictive science.

References

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno R, Kerlavage AR, Mccombie WR, Venter JC (1991) Complementary DNA sequencing: expressed sequence tags and the human genome project. Science 252:1651–1656

Allmer J, Naumann B, Markert C, Zhang M, Hippler M (2006) Mass spectrometric genomic data mining: novel insights into bioenergetic pathways in *Chlamydomonas reinhardtii*. Proteomics 6:6207–6220

Amiour N, Imbaud S, Clément G, Agier N, Zivy M, Valot B, Balliau T, Armengaud P, Quilleré I, Cañas R, Tercet-Laforgue T, Hirel B (2012) The use of metabolomics integrated with transcriptomic and proteomic studies for identifying key steps involved

in the control of nitrogen metabolism in crops such as maize. J Exp Bot 63(14):5017–5033
Andersen JR, Lübberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554–559
Arias RS, Borrone JW, Tondo CL, Kuhn DN, Irish BM, Schnell RJ (2012) Genomics of tropical fruit crops. In: Schnell RJ, Priyadarshan PM (eds) Genomics of tree crops. Springer, Dordrecht, p 209
Barker MS, Dlugosch KM, Reddy ACC, Amyotte SN, Rieseberg LH (2009) SCARF: maximizing next-generation EST assemblies for evolutionary and population genomic analyses. Bioinformatics 25 (4):535–536
Bircher JA, Auger DL, Riddle NC (2003) In search of the molecular basis of heterosis. Plant Cell 15:2236–2239
Boopathi N, Pathmanaban R (2012) Additional insights into the adaptation of cotton plants under abiotic stresses by in silico analysis of conserved miRNAs in cotton expressed sequence tag database (dbEST). Afr J Biotechnol 11(76):14054–14063
Bouck A, Vision T (2007) The molecular ecologists' guide to expressed sequence tags. Mol Ecol 16 (5):907–924
Campbell MM, Brunner AM, Jones HM, Strauss SH (2003) Forestry's fertile crescent: the application of biotechnology to forest trees. Plant Biotechnol J 1:141–154
Conesa A, Gotz S, Garcia-Gomez JM et al (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676
D'Agostino N, Aversano M, Chiusano ML (2005) ParPEST: a pipeline for EST data analysis based on parallel computing. BMC Bioinform 6(4):9
D'Agostino N, Aversano M, Chiusano ML (2007) Tomato EST database: *in silico* exploitation of EST data to explore expression patterns in tomato species. Nucleic Acids Res 35:901–905
Das A, Mondal TK (2010) Computational identification of conserved microRNAs and their targets in Tea (*Camellia sinensis*). Am J Plant Sci 1:77–86
Dias NE, Correa RG, Verjovski-Almeida S, Briones MR, Nagai MA, Wilson DS, Zago MA, Bordin S, Costa FF, Goldman GH et al (2000) Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. Proc Natl Acad Sci U S A 97:3491–3496
Diouf D (2003) Genetic transformation of trees. Afr J Biotechnol 2:328–333
Feng S, Wang X, Zhang X, Dang PM, Holbrook CC, Culbreath AK, Wu Y, Guo B (2012) Peanut (*Arachis hypogaea*) expressed sequence tag project: progress and application. Comp Funct Genomic 2012:373768
Fernie AR (2012) Grand challenges in plant systems biology: closing the circle(s). Front Plant Sci 3:35
Forment J, Gilabert F, Robles A, Conejero V, Nuez F, Blanca JM (2008) EST2uni: an open, parallel tool for automated EST analysis and database creation, with data mining web interface and microarray expression data integration. BMC Bioinform 9:5
Franks SJ, Hoffman AA (2012) Genetics of climate change adaptation. Annu Rev Genet 12(46):185–208
Girke T, Ozkan M, Carter M, Raikhel NV (2003) Towards a modelling infrastructure for studying plant cells. Plant Physiol 132:410–414
González-Ibeas D, Blanca J, Roig C, González-To M, Picó B, Truniger V, Gómez P, Deleu W, Caño-Delgado A, Arús P, Nuez F, García-Mas J, Puigdomènech P, Aranda MA (2007) MELOGEN: an EST database for melon functional genomics. BMC Genomics 8:306–312
Grimplet J, Dickerson JA, Victor KJ, Cramer GR, Fennell AY (2008) Systems biology of the Grapevine. In Proceedings of the 2nd annual national viticulture research conference, University of California, Davis, 9–11 July
Gruszka D, Marzec M, Szarejko I (2012) The barley EST DNA Replication and Repair Database (bEST-DRRD) as a tool for the identification of the genes involved in DNA replication and repair. BMC Plant Biol 12:88–94
Guan H, Kiss-Toth E (2008) Advanced technologies for studies on protein interactomes. Adv Biochem Eng Biotechnol 110:1
Gupta PK, Rustgi S (2004) Molecular markers from the transcribed/expressed region of the genome in higher plants. Funct Integr Genomic 4:139–162
Gupta PK, Varshney RK (2004) Cereal genomics: an overview. In: Gupta PK, Varshney RK (eds) Cereal genomics. Kluwer Academic, Dordrecht, p 639
Habash DZ, Kehel Z, Nachit M (2009) Genomic approaches for designing durum wheat ready for climate change with a focus on drought. J Exp Bot 60 (10):2805–2815. doi:10.1093/jxb/erp211
Hotz-Wagenblatt A, Hankeln T, Ernst P et al (2003) ESTAnnotator: a tool for high throughput EST annotation. Nucleic Acids Res 31:3716–3719
Jain SM, Brar DS, Ahloowalia BS (2002) Molecular techniques in crop improvement. Kluwer Academic, Dordrecht
Kerr Wall P, Leebens-Mack J, Chanderbali AS, Barakat A, Wolcott E, Liang H, Landherr L, Tomsho LP, Hu Y, Carlson JE, Ma H, Schuster SC, Soltis DE, Soltis PS, Altman N, dePamphilis CW (2009) Comparison of next generation sequencing technologies for transcriptome characterization. BMC Genomics 10:347
Khanuja SPS, Jhang T, Shasany AK (2012) Medicinal and aromatic plants: a case example of evolving secondary metabolome and biochemical pathway diversity. In: Sharma VP (ed) Nature at work: ongoing saga of evolution. Springer, Dordrecht, p 355
Kim D-W, Jung T-S, Nam S-H, Kwon H-R, Kim A, Chae S-H, Choi S-H, Kim D-W, Kim RN, Park H-S (2009) GarlicESTdb: an online database and mining tool for garlic EST sequences. BMC Plant Biol 9:61–67
Kirschner MW (2005) The meaning of systems biology. Cell 121:503–504
Kumar CG, LeDuc R, Gong G et al (2004) ESTIMA, a tool for EST management in a multi-project environment. BMC Bioinform 5:176

Lee B, Shin G (2009) CleanEST: a database of cleansed EST libraries. Nucleic Acids Res 37:686–689

Lee B, Hong T, Byun SJ, Woo T, Choi YJ (2007) ESTpass: a web-based server for processing and annotating expressed sequence tag (EST) sequences. Nucleic Acids Res 35:159–162

Liberman LM, Sozzani R, Benfey PN (2012) Integrative systems biology: an attempt to describe a simple weed. Curr Opin Plant Biol 15(2):162–167

Liu H, Jin T, Liao R, Wan L, Xu B, Zhou S, Guan J (2012) miRFANs: an integrated database for *Arabidopsis thaliana* microRNA function annotations. BMC Plant Biol 12:68

Lu Y, Gan Q, Chi X, Qin S (2008) Roles of microRNA in plant defense and virus offense interaction. Plant Cell Rep 27(10):1571–1579

Malik MR, Wang F, Dirpaul JM, Zhou N, Polowick PL, Ferrie AMR, Krochko JE (2007) Transcript profiling and identification of molecular markers for early microspore embryogenesis in *Brassica napus*. Plant Physiol 144:134–154

Mallory CA, Vaucheret H (2006) Functions of microRNAs and related small RNAs in plants. Nat Genet 38:31–36

Mao C, Cushman JC, May GD, Weller JW (2003) ESTAP – an automated system for the analysis of EST data. Bioinformatics 19:1720–1722

Maraschin SF, De Priester W, Spaink HP, Wang M (2005) Androgenic switch: an example of plant embryogenesis from the male gametophyte perspective. J Exp Bot 417:1711–1726

Mason ME, Koch JL, Krasowski M, Loo J (2013) Comparisons of protein profiles of beech bark disease resistant and susceptible American beech (*Fagus grandifolia*). Proteome Sci 11:2

May P, Wienkoop S, Kempa S, Usadel B, Christian N, Rupprecht J, Weiss J, Recuenco-Munoz L, Ebenhöh O, Weckwerth W, Walther D (2008) Metabolomics- and proteomics-assisted genome annotation and analysis of the draft metabolic network of *Chlamydomonas reinhardtii*. Genetics 179:157–166

Mazur B, Krebbers E, Tingey S (1999) Gene discovery and product development for grain quality traits. Science 285:372–375

Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. Genomics 92(5):255–264

Muvva C, Tewari L, Aruna K, Ranjit P, Zahoorullah SMD, Matheen KAMD, Veeramachaneni H (2012) In silico identification of miRNAs and their targets from the expressed sequence tags of *Raphanus sativus*. Bioinformation 8:2

Nagaraj SH, Gasser RB, Ranganathan S (2006) A hitchhiker's guide to expressed sequence tag (EST) analysis. Brief Bioinform 8(1):6–21

Nagaraj SH, Deshpande N, Gasser RB, Ranganathan S (2007) ESTExplorer: an expressed sequence tag (EST) assembly and annotation platform. Nucleic Acids Res 35:143–147

Nam S-H, Kim D-W, Jung T-S, Choi Y-S, Kim D-W, Choi H-S, Choi S-H, Park H-S (2009) PESTAS: a web server for EST analysis and sequence mining. Bioinformatics 25(14):1846–1848

Nasaruddin MN, Harikrishna K, Othman YR, Hoon SL, Harikrishna AJ (2007) Computational prediction of microRNAs from Oil Palm (*Elaeis guineensis* Jacq.) expressed sequence tags. Asian Pac J Mol Biol Biotechnol 15(3):107–113

Naumann B, Busch A, Allmer J, Ostendorf E, Zeller M et al (2007) Comparative quantitative proteomics to investigate the remodelling of bioenergetic pathways under iron deficiency in *Chlamydomonas reinhardtii*. Proteomics 7:3964–3979

Numnark S, Mhuantong W, Ingsriswang S, Wichadakul D (2012) C-mii: a tool for plant miRNA and target identification. BMC Genomics 13:7–16

Osbourn AE, O'Maille PE, Rosser SJ, Lindsey K (2012) Synthetic biology. New Phytol 196(3):671–677

Panchangam S, Mallikarjuna N, Suravajhala P (2012) Apomixis in Chickpea: biology and bioinformatics. Poster session presented at VI international conference on legume genetics and genomics, Hyderabad, India

Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F et al (2003) TIGR Gene Indices Clustering Tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19:651–652

Plechakova O, Tranchant-Dubreuil C, Benedet F, Couderc M, Tinaut A, Viader V, De Block P, Hamon P, Campa C, de Kochko A, Hamon S, Poncet V (2009) MoccaDB – an integrative database for functional, comparative and diversity studies in the Rubiaceae family. BMC Plant Biol 9:123–129

Reusch TBH, Reusch AS, Preuss C, Weiner J, Wissler L, Beck A, Klages S, Kube M, Reinhardt R, Bornberg-Bauer E (2008) Comparative analysis of expressed sequence tag (EST) libraries in the seagrass *zostera marina* subjected to temperature stress. Mar Biotechnol 10:297–309

Sarmah R, Sahu J, Dehury B, Sarma K, Sahoo S, Sahu M, Barooah M, Sen P, Modi MK (2012) ESMP: a high-throughput computational pipeline for mining SSR markers from ESTs. Bioinformation 8:4

Semagn K, Bjornstad A, Xu Y (2010) The genetic dissection of quantitative traits in crops. Electron J Biotechnol 13:5

Shimomura M, Minami H, Suetsugu Y, Ohyanagi H, Satoh C, Antonio B, Nagamura Y, Kadono-Okuda K, Kajiwara H, Sezutsu H, Nagaraju J, Goldsmith MR, Xia Q, Yamamoto K, Mita K (2009) KAIKObase: an integrated silkworm genome database and data mining tool. BMC Genomics 10:486–493

Somerville C, Sommerville S (1999) Plant functional genomics. Science 285:380–383

Sun G (2012) MicroRNAs and their diverse functions in plants. Plant Mol Biol 80(1):17–36. doi:10.1007/s11103-011-9817-6

Sunkar R, Jagadeeswaran G (2008) *In silico* identification of conserved microRNAs in large number of diverse plant species. BMC Plant Biol 8(37)

Suravajhala P, Sundararajan VS (2012) A classification schema to validate protein interactors. Bioinformation 8(1):34–39

Sweetlove LJ, Last RL, Fernie AR (2003) Predictive metabolic engineering: a goal for systems biology. Plant Physiol 132:420–425

Tang Z, Choi J-H, Hemmerich C, Sarangi A, Colbourne JK, Dong Q (2009) ESTPiper – a web-based analysis pipeline for expressed sequence tags. BMC Genomics 10:174

Testillano PS, Risueño MC (2009) Tracking gene and protein expression during microspore embryogenesis by confocal laser scanning microscopy. In: Touraev A (ed) Advances in haploid production in higher plants. Springer, Dordrecht, p 339

Tsai WC, Fu CH, Hsiao YY, Huang YM, Chen LJ, Wang M, Liu ZJ, Chen HH (2013) OrchidBase 2.0: comprehensive collection of Orchidaceae floral transcriptomes. Plant Cell Physiol 54(2):7

Van de Poel B, Bulens I, Markoula A, Hertog M, Dreesen R, Wirtz M, Vandoninck S, Oppermann Y, Keulemans J, Hell R, Waelkens E, De Proft MP, Sauter M, Nicolai BM, Geeraerd AH (2012) Targeted systems biology profiling of tomato fruit reveals coordination of the yang cycle and a distinct regulation of ethylene biosynthesis during postclimacteric ripening. Plant Physiol 160:1498–1514

Varshney RK, Mahendar T, Aggarwal RK, Börner A (2007) Genic molecular markers in plants: development and applications. In: Varshney RK, Tuberosa R (eds) Genomics-assisted crop improvement: genomics approaches and platforms, vol 1. Springer, Dordrecht, pp 13–29

Varshney RK, Ribaut J-M, Buckler ES, Tuberosa R, Rafalski JA, Langridge P (2012) Can genomics boost the productivity of orphan crops? Nat Biotechnol 30:1172–1176. doi:10.1038/nbt.2440

Waegele B, Schmidt T, Mewes HW, Ruepp A (2008) OREST: the online resource for EST analysis. Nucleic Acids Res 1(36(Web Server issue)):W140–W144

Walbot V (1999) Genes, genomes, genomics. What can plant biologists expect from the 1998 National Science Foundation plant genome research program? Plant Physiol 119:1151–1155

Weckwerth W (2008) Integration of metabolomics and proteomics in molecular plant physiology: coping with the complexity by data-dimensionality reduction. Physiol Plant 132:176–189

Wortman JR, Haas BJ, Hannick LI, Smith RK Jr, Maiti R, Ronning CM, Chan AP, Yu C, Ayele M, Whitelaw CA, White OR, Town CD (2003) Annotation of the arabidopsis genome. Plant Physiol 132(2):461–468

Zhang BH, Pan XP, Wang QL, Cobb GP, Anderson TA (2005) Identification and characterization of new plant microRNAs using EST analysis. Cell Res 15:336–360

Zheng Q, Song J, Campbell-Palmer L, Thompson K, Li L, Walker B, Cui Y, Li X (2013) A proteomic investigation of apple fruit during ripening and in response to ethylene treatment. J Proteomics. http://dx.doi.org/10.1016/j.jprot.2013.02.006