le Development and Biodiversity 2

wani *Editor*

able Horticultural Systems

hnology and Innovation

ble horticulture movement is growing the past few decades and gaining increas-tion among the scientific and farming community. An understanding of the and relevant adaptation strategies are of foremost importance to sustain the ivity and profitability of horticulture crops in the climate change scenario, ecessitates synthesis of current knowledge to develop strategies for adaptation gation to achieve climate-resilient horticulture. Issues have been covered in chapters to make this book a treasure of knowledge in sustainable horticulture. examples in fruits and vegetables included and various other important aspects ate change, pests and diseases, soil are also covered in the book. The chapters en written by well- known professionals in their research field.

ok contains three sections; Sustainable Horticultural Systems, Biodiversity inable Horticultural Systems and Breeding and Improvement in Sustainable ltural Systems.

k is useful for undergraduates, graduates students, horticulturists, agriculture ers and extension agents, plant scientists as well as horticultural industries. It efore help all scientists, decision-makers, professors, farmers and politicians h to build a safe agriculture, energy and food system for future generations.

nces

978-3-319-06903-6

3319 069036

Nandwani *Ed.*

Sustainable Horticultural Systems

Sustainable Development and Biodiversity 2

Dilip Nandwani *Editor*

Sustainable Horticultural Systems

Issues, Technology and Innovation

Springer

This book series provides complete, comprehensive and broad subject based reviews about existing biodiversity of different habitats and conservation strategies in the framework of different technologies, ecosystem diversity, and genetic diversity. The ways by which these resources are used with sustainable management and replenishment are also dealt with. The topics of interest include but are not restricted only to sustainable development of various ecosystems and conservation of hotspots, traditional methods and role of local people, threatened and endangered species, global climate change and effect on biodiversity, invasive species, impact of various activities on biodiversity, biodiversity conservation in sustaining livelihoods and reducing poverty, and technologies available and required. The books in this series will be useful to botanists, environmentalists, marine biologists, policy makers, conservationists, and NGOs working for environment protection.

More information about this series at http://www.springer.com/series/11920

Dilip Nandwani
Editor

Sustainable Horticultural Systems

Issues, Technology and Innovation

Springer

Editor
Dilip Nandwani
Agricultural and Environmental Sciences
Tennessee State University
Nashville
Tennessee
USA

ISSN 2352-474X ISSN 2352-4758 (electronic)
ISBN 978-3-319-06903-6 ISBN 978-3-319-06904-3 (eBook)
DOI 10.1007/978-3-319-06904-3
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014947972

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To My Mother

Biswas MK, Dutt M, Roy UK, Islam R, Hossain M (2009) Development and evaluation of in vitro Somaclonal variation in strawberry for improved horticultural traits. Sci Horticult 122:409–416
Coudert MJ (1982) Market openings in West Africa for Cowpeas and Bambara groundnuts. Int Trade Technol X:147–241
Dias JS (2011) Biodiversity and vegetable breeding in the light of developments in intellectual property rights, ecosystems biodiversity. Grillo O (ed) ISBN: 978-953-307-417-7, InTech. http://www.intechopen.com/books/ecosystems-biodiversity/biodiversity-and-vegetable-breeding-in-thelight-of-developments-in-intellectual-property-rights
Fitzgerald P (2011) Agrifood Awareness Australia Limited- (2011) http://www.horticulture.com.au/librarymanager/libs/174/Paula%20Fitzgerald%20%20Glocal%20Status%20of%20gene%20tachnology%20in%20horticulture.pdf. Accessed 28 Nov 2013
Food and Agriculture Organization (2012) The State of Food Insecurity in the World 2012. http://www.fao.org/docrep/016/i3027e/i3027e.pdf. Accessed 9 June 2013
Food Security Network (2012) http://www.foodsecuritynews.com/What-is-food-security.htm. Accessed 9 May 2013
Fry WE, Godwin SB (1997) Re-emergence of potato and tomato late blight in the United States. Plant Dis 81(12):1349–1357
João SD (2010) Impact of improved vegetable cultivars in overcoming food insecurity. Euphytica 176:125–136
Kingsbury N (2009) Hybrid: the history and science of plant breeding. The University of Chicago Press, Chicago
Lippman ZB, Zamir D (2007) Heterosis: revisiting the magic. Trends Genet 23:60–66
Milbank J (2007) Globally competitive horticultural production and marketing utilizing renewable energy. Nuffield Australia Project No SCVCT0601
Riboua SB, Douama F, Hamanta O, Frohlich MW, Negrutiua I (2013) Plant science and agricultural productivity: why are we hitting the yield ceiling? Plant Sci 210:159–176
Rodriguez-Enriquez J, Dickinson HG, Grant-Downton RT (2011) MicroRNA misregulation: an overlooked factor generating somaclonal variation? Trends Plant Sci 16:242–248
Shu QY (2009) Induced plant mutations in the genomics era. Food and Agriculture Organization of the United Nations, Rome, pp 425–427
Shu QY, Lagoda, PJL (2007) Mutation techniques for gene discovery and crop improvement. Mol Plant Breed 5:193–195
Sleper DA, Poehlman JM (2006) Breed Field Crops Crop Sci 47:900–901
Sparrow DHB (1979) Special techniques in plant breeding. In: Knight R (ed) Plant breeding. Academy Press, Brisbane, p 163
Spore (2007) Wageningen, The Netherlands: The Technical Centre for Agricultural and Rural Cooperation, CTA, ACP-EU 1-2. No. 128
Wilde DH, Chen Y, Jiang P, Bhattacharya A (2012) Emir. J. Food Agric 24(1):31–41

Transcriptome Analysis of *Musa* and its Applications in Banana Improvement

S Backiyarani, S Uma, M. S. Saraswathi and A Chandrasekar

1 Introduction

Banana and plantains are of great socio economic importance being a staple food for more than 400 million people with a global production of 129.0 million t. Bananas are considered to be a high energy food and also rich in minerals and vitamins. They are one of the earliest domesticated plant species (Denham et al. 2003) originated in South East Asia and Western Pacific and are now widely distributed throughout the tropical and subtropical regions. Although banana and plantains are best known as a food-fruit crop, almost every part of the plant can be used in one way or another, starting from banana sap to seed for the benefit of mankind.

The vast majority of cultivated bananas have evolved either from natural inter/intraspecific hybridization between two wild species, *M. acuminata* and *M. balbisiana*. Cultivated bananas are sterile triploids or diploids, with parthenocarpic fruit development and propagated through vegetative means. Asexual driven evolution has resulted in a narrow genetic base which is currently threatened by major pests and diseases. They include diseases like *Mycosphaerella* spp., Fusarium wilt, bacterial wilt, Moko, viral diseases etc. The pests include pseudostem borers, corm weevils, aphids, nematodes etc. Although management of selected pests and diseases is possible through chemical means, it often leads to environmental pollution and enhanced cost of cultivation. The ultimate solution for these threats is the development of resistant varieties through introgression of resistant genes either through conventional breeding or genetic engineering approaches. Understanding the host-pathogen interaction is essential for developing pest and disease resistance through the identification of signal transduction pathways in bananas. Isolation of trait specific genes and markers and their application in marker assisted selection (MAS) has accelerated both conventional breeding and genetic engineering meth-

S. Uma (✉) · S. Backiyarani · M. S. Saraswathi · A. Chandrasekar
Crop improvement, National Research Centre for Banana, (ICAR), Thogamalai Road, Thayanur Post, Tiruchirapalli, TN 620102, India

ods for crop improvement. In the present day context, identification of trait specific gene(s) has been made easier with the available data on genome sequencing. In banana, team led by French Agronomic Research Centre, CIRAD has successfully completed the whole genome sequencing of double haploid Pahang (W), of *M. acuminata* subsp. *malaccensis* (A genome) (D'Hont et al. 2012). It has genome coverage of 472.2 Mb which represented 70% of the DH-Pahang genome size. Similarly, whole genome sequencing of wild diploid *M. balbisiana* (BB) type 'Pisang Klutuk Wulung (PKW)' has been completed (Davey et al. 2013) with a genome coverage of 341.4 Mb representing 78% of the expected B genome size. The sequence information on the ancestral genomes (A and B) of present day bananas will provide an idea about the genome structure, organization, allelic diversity and regulatory elements of the *Musa* species. Though genome sequencing provides the vital and voluminous information, it fails to give a clear picture about the expression of genes in a particular tissue for a particular physiological condition. To understand this relationship between the genome and the phenotype, it is essential to study the genome products like messenger RNAs. Transcriptome studies has an added advantage over whole genome sequencing with respect to the information on the mRNA variants which are the resultant products of alternative splicing, RNA editing or transcription initiation and termination sites. Thus understanding the transcriptome is essential for interpreting the functional elements of the genome, revealing the molecular constituents of tissues and also for understanding the resistance mechanism for a specific stress. Transcriptome profiling has been made easy with the availability of user friendly bioinformatic tools to understand the molecular mechanism of plants under various physiological conditions. *Musa* improvement program through conventional approach is a slow process owing to its complex genetic structure. Moreover bananas exhibit some unique characteristics such as different ploidy levels, parthenocarpy, sterility, vegetative propagation and biparental cytoplasmic inheritance (Roux et al. 2008). Thus, transcriptome analysis and gene expression studies of *Musa* will be a powerful model for understanding the above mentioned fundamental aspects for which the other two model species (*Arabidopsis,* Rice) use differently.

2 Global Approaches to Gene Expression Analysis

Various technologies have been developed to deduce and quantify the transcriptome, including hybridization or sequence-based approaches. Over the years, several techniques, such as suppression subtractive hybridization (SSH), representational difference analysis (RDA), expressed sequence tags (EST), cDNA fragment fingerprinting (cDNA AFLP), serial analysis of gene expression (SAGE), DNA microarrays etc., which allow high throughput identification and/or quantification of transcriptome components have been developed. The merits and demerits of these techniques are mentioned in Table 1. The aforesaid techniques have been successfully used to identify many differently expressed genes in various tissues against various stresses/developmental stages (Table 2). The details of those tech-

Table 1 Principle, merits and demerits of banana transcriptome techniques

Techniques	Principle	Merits	Demerits	Reference
SSH	PCR-based amplification of only cDNA fragments that differ between a control (driver) and experimental transcriptome	To compare two mRNA populations and obtain cDNAs representing genes that are either over-expressed in one population as compared to other	RNA is required in High quantities	Diatchenko et al.1996 Munir et al. 2004
RDA	To find sequence differences in two genomic or cDNA samples	It simplifies analysis of the subtracted library	Experiment is expensive	Lisitsyn et al. 1993
cDNA AFLP	To distinguish closely related individuals at the sub-species level and can also map genes	Identification of up and downregulated genes in contrast to subtractive hybridization	DNA sequencing and assembly of complex, large, polyploid, and/or repetitive genomes is still technically difficult	Althoff et al. 2007
SAGE	Short sequence tags are sufficient to identify a gene transcript, provided the tags represent a known location within the gene	Data accurate, quantitative and cumulative	More sequencing errors due to the misidentification of the region in the database sequence	Velculescu et al. 1995
EST	The EST sequences are generated by randomly picking the clones from a cDNA library and performing a single sequencing reaction to produce 300–500 bp of sequence per clone	To identify gene transcripts, which are instrumental in gene discovery and gene sequence determination	"EST" to describe genes for which little or no further information exists besides the tag	Bortoluzzi and Danieli 1999
DNA/ cDNA microarray	Hybridization between two DNA strands, the property of complementary nucleic acid sequences to specifically pair with each other by forming hydrogen bonds between complementary nucleotide base pairs	Reduces the bias towards gene sets that are already known	Experimental replication is expensive	Maskos and Southern 1992

Table 2 Transcriptome techniques exploited in banana

Variety/cultivar	Techniques	Tissues	Stage/stress	No. of expressed genes	Reference	Laboratory
M. acuminata, subgroup Cavendish 'Grand Nain'	cDNA AFLP	Fruit	Crown rot disease	10	Lassois et al. 2011	University of Liege, Gembloux, Belgium
M. acuminata ssp. burmannicoides 'Calcutta-4, cv. Rose, FHIA 17, Williams'	cDNA AFLP	Root	*Fusarium*	76	Munro 2008	University of Pretoria, Pretoria, South Africa
M. acuminata ssp. burmannicoides 'Calcutta-4'	SSH	Root	*Fusarium*	83	Swarupa et al. 2013	IIHR, Bangalore, India
M. acuminata cv. Manoranjitham (AAA)	SSH	Leaf	*Eumusae* leaf spot	805	Uma et al. 2011	NRCB, Trichy, India
M. acuminata cv. Karthobiumtham (ABB)	SSH	Root	Nematode	256	Backiyarani et al. 2014	NRCB, Trichy, India
M. acuminata cv. Tuu Gia	SAGE	Leaf		5292	Coemans et al. 2005	Laboratory of Tropical Crop Improvement, Katholieke Universiteit Leuven, Belgium
M. acuminata ssp. burmannicoides 'Calcutta-4'	EST Sequencing	Hot and cold stress exposed leaf tissues	Fruit ripening	220	Santos et al. 2005	EMBRAPA, Brazil
M. acuminata 'Cavendish Williams'	Combination of SSH and microarray	Fruit	Fusarium	79	Van den berg et al. 2007	FABI, University of Pretoria, South Africa
M. acuminata, cv. Grand Nain	Combination of SSH and microarray	Fruit	Early ripening of banana	84	Manrique-Trujillo et al. 2007	Centro de Investigación y de EstudiosAvanzados del IPN, México
M. acuminata AAA group	Combination of SSH and microarray	Fruit	Early ripening of banana	265	Xu et al. 2007	Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, China

Table 2 (continued)

Variety/cultivar	Techniques	Tissues	Stage/stress	No. of expressed genes	Reference	Laboratory
M. acuminata L. AAA group	Combination of SSH and microarray	Fruit	Ethylene biosynthesis initiation in fruits.	22	Jin et al. 2009	Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, China
M. acuminata, Dwarf Cavandish (AAA)	Combination of SSH and microarray	Fruit	Ethylene biosynthesis initiation in fruits	37	Kesari et al. 2007	NBRI, Lucknow, India

2.1 *cDNA-amplified Fragment Length Polymorphism Analysis (cDNA-AFLP)*

cDNA-AFLP is an improvement over traditional differential display techniques. It is a PCR-based method, in which cDNA is digested with two restriction enzymes (four-cutter and a six-cutter) and adapters are ligated to the ends of the fragments. In the first step, only those fragments that are digested by both restriction enzymes are thus have different adapters at the end. In the following amplification, the complex starting mixture of cDNA is fractionated into smaller subsets by selective PCR amplification using primers for the adapters with one or more extra nucleotides. By increasing stringency of PCR amplification (with additional nucleotides to the primers), the sensitivity of the analysis can be increased which allows the detection of genes with low expression levels. Differences in intensity of the bands provide a good measure of the relative difference in the levels of gene expression.

Munro (2008) examined the *Fusarium oxysporum* f. sp. *cubense* (FOC) resistant (cvs. Rose and Calcutta 4 and tolerant hybrid FHIA 17) and susceptible (Cavendish cv. Williams) to elucidate the genetic factors involved in banana defense response against FOC. Transcriptome profile using cDNA-AFLP was carried out on tissues collected at 6 and 72 h post inoculation of these cultivars with FOC 'subtropical' race 4 (VCG 0120). Seventy-six differentially expressed transcript derived fragments (TDFs) were isolated, sequenced and annotation of these sequences resulted in identification of defense-related genes such as S-adenosylmethionine synthase (SAMS) and isoflavone reductase, which are potentially involved in the production of cell wall strengthening compounds like lignin etc.

Crown rot disease, caused by *Colletotrichum musae*, is one of the main postharvest diseases of banana that has a negative impact on the market value of fruits and affects export. Lassois et al. (2008) tried to identify genes influencing the susceptibility to crown rot disease, at distinct physiological growth stages of 'Grande Naine' fruits. The cDNA-AFLP fingerprinting of this study resulted in identification of

genes involved in signal transduction, pathogenesis- related proteins, cellulose synthase gene and dopamine-β-monooxygenase. Delgado et al. (2003) also evidenced that reduction in cellulose synthesis leads to activation of defense responsive genes through various signalling pathways. Identification of differentially expressed genes like dopamine-monooxygenase, which play a key role in catecholamine biosynthesis pathway, suggested the possible involvement of these genes in resistance mechanism (Lassois et al. 2011).

2.2 EST Sequencing

Expressed sequence tags (ESTs) is a quantitative method to measure specific transcripts within a cDNA library (Rudd 2003) and is achieved by single-pass sequencing of a large number of cDNAs traditionally yielding sequence tags of around 300–500 bp in length. In a study to understand the function and regulation of the genes involved in temperature stress, Santos et al. (2005), developed and characterized two full length enriched cDNA libraries from the RNA isolated from cold (5–25 °C) and hot (25–45 °C) stress exposed leaves of *M. acuminata* ssp. *burmannicoides* var. Calcutta 4. The global analysis of these two libraries indicated that only 10% of the cDNAs had common expression, while 42 and 48% of cDNAs were unique to cold and hot libraries respectively. From cold EST library, two genes related to the ABA response to cold acclimation and one low -temperature responsive genes were identified. The data is in public domain of EMBRAPA, Brazil (http://www.cenargen.embrapa.br/index.html) as a part of the DATA_*Musa* database (a genome database for *Musa* spp consisting DNA sequences and sequence information).

2.3 Serial Analysis of Gene Expression (SAGE)

SAGE is a method that generates sequences from cDNA fragments for the discovery of new genes and the quantification of their expression levels. SAGE is a strategy for extensive tag profiling of cDNA over EST sequencing (Velculescu et al. 1995) which is based on counting sequence tags of 10–15 bp from cDNA libraries. Briefly, a short cDNA fragment or tag (13 bp) is extracted from a defined position in each transcript by a series of linker ligations and restriction digestions. Tags are then amplified once by PCR, concatenated, cloned in a vector and sequenced. These tags, in most cases, contain sufficient information for unambiguous annotation to EST sequences. The frequency of each tag directly reflects the abundance of the corresponding mRNA (Velculescu et al. 1995). But annotation of the tags depends largely on the availability of cDNA libraries or EST collections.

For the first time, *M. acuminata* gene expression pattern in youngest unfurled leaf tissue of *M. acuminata* cv. "Tuu Gia" (accession No. ITC.610) was analysed by Coemans et al. (2005) through SAGE. A total of 10,196 tags were generated from 5292 expressed genes and abundantly expressed transcripts were annotated by homology to cDNA or EST sequences. Analysis of the transcript profiles showed that the majority of the abundant transcripts are involved in energy production, mainly photosynthesis. However, the most abundant tag was derived from a type 3 metallothionein transcript, which accounted for nearly 3% of total transcripts analysed. The unknown tags were also identified by applying 3' rapid amplification of cDNA ends (3' RACE) which resulted in the identification of a novel NADPH protochlorophyllide oxidoreductase, the key enzyme in chlorophyll biosynthesis. Limited availability of banana genomic DNA and cDNA/EST sequences at that time restricted the use of classical SAGE yielding short tags.

2.4 Suppression Subtractive Hybridisation (SSH)

The principle of subtractive hybridization is commonly been used to analyze the differences in gene expression between two samples. The selection of differentially expressed sequences is performed by allowing the samples to cross-hybridize, followed by isolation of hybridization products unique for one sample (termed the tester) from sequences that are shared between the samples or unique for the other sample (the driver). Suppression subtractive hybridization was developed for the enrichment of differentially expressed genes of both high and low abundance (Diatchenko et al. 1996).

Swarupa et al. (2013) demonstrated the FOC resistance mechanism involved in *M. acuminata* spp. *burmannicoides* var. Calcutta 4, using SSH approach. From this library, a total 83 unique genes were obtained and homology comparison revealed that nearly 17% of genes were involved in stress and defense while 15% in signal transduction. Based on RT-PCR studies in contrasting cultivars like Calcutta 4 and Kadali for FOC, many genes which are involved in recognition of pathogens like, transmembrane protein, receptor like protein kinase, mannose binding protein etc. were confirmed based on their expression pattern in the resistant cultivar. Ravishankar et al. (2011) has identified some differentially expressed genes (lipoxygenase, rubiscoactivase, glycinedehydrogenease, catalase and ethylene responsive factor) from drought stressed leaf samples of Bee Hee Kela (BB) through SSH approach.

Similarly, Uma et al. (2011) identified a total of 498 unigenes, which consisted of 78 contigs and 420 singletons from the SSH library obtained from *M. eumusae* challenged resistant cultivar (Manoranjitham AAA). Top BLAST hits of BLAST2GO analysis resulted in a total of 89 hits, of which 13 defense related genes were chosen for gene expression studies. It revealed that cytochrome oxidase, lipoxygenase, metallothionein, flavin containing monoxygenase were upregulated only in the resistant cultivar within 48 h post inoculation (HPI) and these genes could be involved in *M. eumusae* leaf spot resistance mechanism. Backiyarani et al. (2014) also identified 256 unique genes from the root lesion nematode (*Pratylenchus coffeae*) challenged resistant cultivar (Karthobiumtham ABB) through SSH approach. Of these, 26.8% were unigenes involved in defense and/or signal transduction including resistant gene homologues, disease resistance response proteins and protein kinase

signaling. 12.8% of the unigenes hit with unknown protein classes. These findings imply that invasion of nematode could trigger multiple signaling pathways. Like in many other methods, SSH also relies on restriction enzymes leading to the loss of transcripts lacking restriction sites. Combining SSH technique with high throughput screening of the harvested clones through the use of cDNA microarrays could greatly reduce the tedious work of northern blot analysis and also the likelihood of false-positive clones enriched via SSH (Yang et al. 1999).

2.5 *Combination of SSH and Microarray*

Combination of the suppressive subtractive hybridization and cDNA microarray techniques has greatly improved the efficiency of fishing out of the differentially expressed clones (Yang et al. 1999; Geschwind et al. 2001). Two possible combinations are (i) using the subtracted cDNA clones as probes printed on chips and (ii) using the labeled and enriched amplicons. In the second method, instead of using the subtracted clones as probes spotted on chips, enriched amplicons were labeled and used as targets to hybridize the pre-made microarray chips for microarray analysis.

This combination technology, allowed not only the identification of genes without extensive sequencing works in the subtracted amplicons regardless of many redundant clones, but also to easily evaluate the subtraction efficiency and specificity. Moreover, the SSH/microarray approach made it possible to conduct a transcriptome-wide identification of differentially expressed genes particularly for those with low expression. It has been reported that the absolute expression level is not a crucial determinant for identifying the genes, while the relative difference in expression levels does impact on whether or not a gene is recovered by subtractive hybridization.

In banana, the differentially expressed genes during fungal and fruit ripening stages were identified by this SSH/microarray approach. Van den berg et al. (2007) constructed banana SSH library with 736 differentially expressed genes in FOC tolerant Cavendish selection 'GCTCV-218' at 6 HPI of FOC race 4 (VCG 0120). Subsequently, this library was screened using glass slide microarray to identify and discard housekeeping and rRNA genes that had escaped subtraction and to select defense response-associated genes. As a result, out of 736, 79 clones were shortlisted and sequenced. Annotation of these sequences revealed that many showed homology to defense-associated genes, including cell wall-strengthening genes. These genes were confirmed through quantitative RT-PCR based on up-regulation and differential expression throughout a time-course, following FOC infection in the tolerant GCTCV-218 compared with susceptible cv. Williams. It was confirmed that genes coding for PR-1, PAE, xylanase inhibitor, peroxidase, catalase 2, metallothionein, response regulator 6 and trypsin inhibitors were differentially expressed in the tolerant cultivar GCTCV-218. This study also proved that tolerance of GCTCV-218 was linked to significantly increased induction of cell wall-associated phenolic compounds.

A better understanding of the mechanism of postharvest ripening is necessary to manipulate postharvest management in banana. Xu et al. (2007) tried to identify the differentially expressed genes during the early stages of fruit ripening through SSH approach. SSH was performed along with cDNA microarray analysis on the day of harvest as the "driver" and cDNA of the fruit, 2 days afters harvest (DAH) as the "tester." A total of 265 readable differentially expressed genes were obtained through forward subtraction from tester and their expression levels were estimated by overlaying images of cDNA microarrays screened by mRNA from 0 to 2 DAH. Their expression level was confirmed through RT-PCR and the results were found to be in accordance with the microarray data. Upregulation of endoglucanase and cellulose synthase genes were generally associated with cellulose biosynthesis of primary and secondary cell walls suggesting the importance of these proteins for the biosynthesis of highly crystalline cellulose. The over expression of alcohol dehydrogenases (Adh) at both early and late stages of fruit ripening suggested that aroma production begins at the beginning of postharvest and continues up to the late phase of ripening. In order to gain an understanding of the genetic components responsible for metabolic shifts associated with banana ripening, Manrique-Trujillo et al. (2007) identified differentially expressed transcripts in a late stage of fruit ripening through SSH approach. The differentially expressed SSH clones were confirmed through hybridization with DNA arrays and genes involved in the processes associated with fruit ripening, such as stress, detoxification, cytoskeleton and biosynthesis of volatile compounds (orcinol O-methyltransferase, putative alcohol dehydrogenase, ubiquitin-protein ligase, chorismatemutase etc.) were identified.

Davey et al. (2009) tried to profile the banana leaf transcriptome to drought stress, pooled RNA of control and drought stressed leaves of the *Musa* cultivar 'Cachaco' were hybridized to Affymetrix Rice Genome Array. This resulted in the identification of 2910 transcripts displaying >2 fold difference in expression levels in response to drought. These drought responsive transcripts included many functional classes associated with plant biotic and abiotic stress responses, as well as regulatory genes like ERF, DREB, MYB, bZIP and bHLH transcription factor families. The outcome of this study proved that cross species (heterologous) microarray studies using gDNA-based probe selection as a highly promising strategy to study complex biological responses in a non-model species.

2.6 *Next Generation Sequencing*

Recent technologies of next generation sequencing (NGS) provide a comprehensive snapshot of the transcriptome. Transcriptome analysis through NGS is called as RNA sequencing or RNA-seq, which enables simultaneous sequencing of millions of molecules without involving any cloning. Thus RNA-seq has become an attractive analytical tool in trancriptomics, due to minimum running cost and the possibility to uncover novel transcriptional-related events (Wang et al. 2011).

- NGS transcriptome analysis is more accurate and highly sensitive over other techniques.
- NGS technique offers true quantification in contrast to microarrays
- Simultaneous data collection for genes with differential expression levels is possible only in transcriptome analysis
- This technique has the ability to distinguish isoforms of the same transcript and to detect new splicing variants.
- Very effective for the discovery of novel splicing sites
- New transcripts could also be detected from NGS transcriptome techniques which are not possible through microarrays.
- Transcriptome sequencing does not necessarily require prior knowledge of the genome sequence whereas designing of microarrays needs a reference sequence. Thus this method is particularly attractive for non-model species.
- NGS transcriptome enables SNP discovery whereas it is not possible through other transcriptome techniques

RNA–Seq techniques involve three major steps namely, library construction, DNA sequencing and sequence data analysis. Library construction includes isolation of high quality RNA samples, fragmentation of RNAs, and reverse transcription of RNA into cDNA followed by second strand synthesis and end modification by ligation of oligonucleotide adapters. Optionally this cDNA library may be amplified through PCR to selectively enrich the adapter molecules ligated to both ends. Then their quality and quantity can be validated by capillary electrophoresis, spectrophotometry and qPCR prior to sequencing.

NGS platforms are ideally suited for the sequencing of RNA-Seq libraries. The term NGS is used to collectively describe technologies other than Sanger sequencing that have the potential to process millions of sequence reads in parallel, rather than 96 at a time. Also, next generation sequence reads are produced from fragment libraries that have not been subjected to the conventional vector based cloning and *E. coli* based amplification stages used in capillary sequencing. The length of next generation sequenced reads is 35–250 bp (depending on the platform) (Table 3) than capillary sequenced reads which is about 650–800 bp. Of many NGS platforms available, Roche 454, ABI SOLid, Illumina (Hiseq, Miseq) are the popular platforms and Ion Proton, PacBio RS, Oxford nanopore are some of the newer platforms.

Million or billion reads obtained through this platforms will be filtered for obtaining high quality reads for further downstream analysis (Fig. 1). Initially, the reads are mapped to the reference genome or transcriptome and the reads that align to the same locus are assembled into transcripts. The expression level of that particular transcript is estimated by counting the number of reads that aligned to this transcript. Mortazavi et al. (2008) proposed the quantification of transcript levels is an accurate method, ie. reads per kilobase of exon model per million mapped measures (RPKM). $RPKM = 10^9 \times C/N.L$ where C is the number of mappable reads that fall on to the gene's exon, N is the total number of mappable reads and L is the total length of exons.

Table 3 Comparison of various NGS platforms

Platform	Chemistry	Read length (bp)	Run time	Gb/Run	Primary error	Base error rates	Advantage	Disadvantage
454 GS Junior (Roche)	Pyro-sequencing	500	8 h	0.04	Indel	1.00%	Long read length	High error rate in homopolymer
454 GS FLX + (Roche)	Pyro-sequencing	700	23 h	0.7	Indel	0.5%	Long read length	High error rate in homopolymer
Hi Seq (Illumina)	Reversible terminator	2*100	2 days (rapid mode)	120 (rapid mode)	Substitution	~1–2% over 100 bp	High throughput/ low cost	Short reads long run time
MiSeq (Illumina)	Reversible terminator	2*150	27 h	1 Gb	Substitution	~1–2% over 100 bp	High throughput/ low cost	Short reads
SOLiD (Life)	Ligation	85	8 days	150	A-T bias	0.06%	Low error rate	Short reads long run time
Ion Proton (Life)	Proton detection	200	2 h	100	Indel	1.2% over 150 bp	Short Run times	Higher error rate
PacBio RS	Real-time sequencing	3000 (up to 15,000)	20 min	3	CG deletion	13.00%	Long read length without PCR	High error rate

To enable comparison of two or more different transcriptome libraries, the values are normalized by taking into account the fact that the number of reads per transcript depends on the transcript length and the total number of reads obtained for a particular library (Mortazavi et al. 2008). Apart from quantification of transcripts, RNA-Seq also permits isoform identification, detection of new splicing sites. Specific softwares are available as free or open source programs for each analysis (Table 4).

In a nutshell, the NGS transcriptomics aims at the detection of transcripts expressed under specific physiological conditions, determination of the transcriptional structure of genes, splicing patterns and other post-transcriptional modifications apart from quantification of varying expression levels of each transcript.

2.6.1 Transcriptome Profiling in Banana Through NGS

Understanding the complexity of the disease resistance will pave way for the development of biotic stress resistant bananas. Many studies have been carried out to understand the molecular mechanism of resistance/tolerance through RNAseq or

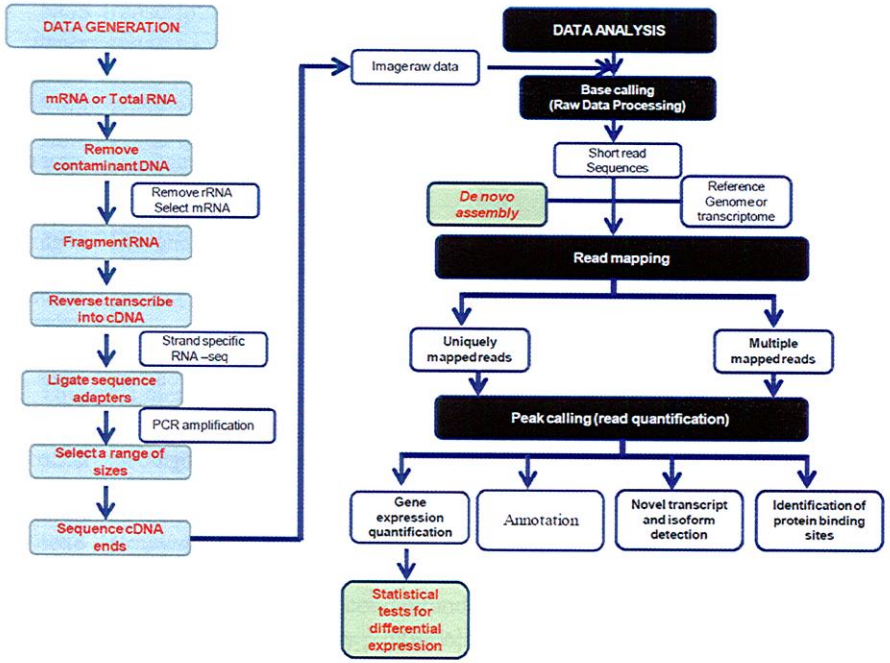

Fig. 1 Methodology involved in transcriptomics using NGS

Fusarium wilt caused by *FOC* is affecting bananas in tropical and subtropical countries (Ploetz 2006). Li et al. (2012) identified defense related transcripts which are involved in *FOC* TR-4 resistance by comparing the root transcriptome profiles of 'Nongke No 1' (resistant mutant of Cavendish) and susceptible variety 'Brazilian'. Study could generate 2691 differentially expressed unigenes including defense and signal transduction in the resistant mutant. Most of the defense related genes and pathways identified in this study differed from those in model plants such as rice and Arabidopsis, suggesting that the mechanism of resistance in banana could be highly variable. Based on this profiling study, they demonstrated that in mutant variety, jasmonic acid (JA) and ethylene (ET) hormone signaling pathways are involved in *FOC* resistance but not salicylic acid (SA) pathway.

Similarly Wang et al. (2012) also indicated that JA biosynthesis is regulated in *FOC* TR-4 resistant cultivar. They performed a combination of DGE (Digital Gene Expression) and RNA seq transcriptome analysis in *M. acuminata* AAA group 'Brazilian' which was challenged with GFP tagged *FOC* TR-4. Some of the genes like peroxidases, 4-coumarate, cinnamate 4- hydroxylases and phenylalanine ammonia lyases involved in phenylpropanoid biosynthesis were reported to be upregulated. Costa et al. (2013) tried to identify the differentially expressed genes in contrasting cultivars BRS Platina (R) and Prata-Ana (S) under Fusarium challenged condition by profiling the global transcriptome. This result is worth mentioning as it enabled

Table 4 Bioinformatics tools for transcriptome analysis

Sl. No	Quality control and pre-processing data		
	Tool name	Description	Website
1	FastQC	Quality control tool for high-throughput sequence data	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
2	FASTX	Toolkit is a set of command line tools to manipulate reads in files FASTA or FASTQ format	http://hannonlab.cshl.edu/fastx_toolkit/
3	RNA-SeQC	Tool with application in experimental design, process optimization and quality control before computational analysis	http://www.broadinstitute.org/cancer/cga/rna-seqc
4	PRINSEQ	Generates statistics of sequence data for sequence length, GC content, quality scores, replicates, complexity, tag sequences, poly-A/T tails, odds ratios	http://edwards.sdsu.edu/cgi-bin/prinseq/prinseq.cgi
5	FastQ Screen	Screens FASTQ format sequences to a set of databases to confirm that the sequences contain what is expected (such as species content, adapters, vectors, etc)	http://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/
Alignment tools			
6	TopHat	Fast splice junction mapper for RNA-Seq reads	http://tophat.cbcb.umd.edu/
7	Bowtie	Fast short aligner using an algorithm based on the Burrows-Wheeler transform and the FM-index	http://bowtie-bio.sourceforge.net/index.shtml
8	BWA	Short aligner using an algorithm based on the Burrows-Wheeler transform	http://bio-bwa.sourceforge.net/
9	MAQ	First aligns reads to reference sequences and after performs a consensus stage	http://maq.sourceforge.net/maq-man.shtml
Transcriptome assemblers (Genome-Guided assemblers)			
10	Cufflinks	Assembles transcripts, estimates their abundances, and tests for differential expression and regulation in RNA-Seq samples	http://cufflinks.cbcb.umd.edu/
11	RNAeXpress	To extract and annotate biologically important transcripts from next generation RNA sequencing data	http://www.rnaexpress.org/
Transcriptome assemblers (Genome-Independent assemblers)			
12	SOAPdenovo	De novo Assembler	http://soap.genomics.org.cn/soapdenovo.html
13	Velvet	Sequence assembler for very short reads	http://www.ebi.ac.uk/~zerbino/velvet/
14	Trinity	Novel method for the efficient and robust de novo reconstruction of transcriptomes from RNA-seq data	http://trinityrnaseq.sourceforge.net/

Table 4 (continued)

Sl. No	Quality control and pre-processing data		
	Tool name	Description	Website
Quantitative analysis and differential expression			
15	Cuffdiff	Measure global de novo transcript isoform expression. It performs assembly of transcripts, estimation of abundances and determines differential expression (Cuffdiff) and regulation in RNA-Seq samples	http://cufflinks.cbcb.umd.edu/
16	DESeq	Bioconductor package to perform differential gene expression analysis based on negative binomial distribution	http://bioconductor.org/packages/release/bioc/html/Deseq.html
17	EdgeR	R package for analysis of differential expression of data from DNA sequencing methods, like RNA-Seq, SAGE or ChIP-Seq data	http://www.bioconductor.org/packages/release/bioc/html/edgeR.html
Workbench (analysis pipeline/integrated solutions)			
18	Avadis NGS	The data mining and visualization platform at the core of all bioinformatics products developed by Strand Scientific Intelligence	http://www.avadis-ngs.com/
19	CLC Genomics Workbench	CLC bio deliver genomics sequencing analytics solution	http://www.clcbio.com/
20	DNASTAR	DNASTAR's Lasergene Genomics Suite includes software that allows users to quickly and easily align next-gen RNA reads against a reference genome, and then perform in-depth analysis of the aligned data	http://www.dnastar.com/t-sub-nextgen-genome-solutions-rnaseq.aspx
21	NextGENe	Perfect analytical partner for the analysis of desktop sequencing data produced by the ION PGM™, Roche Junior, IlluminaMiSeq as well as high throughput systems as the Ion Torrent Proton, Roche FLX, Applied BioSystemsSOLiD™ and Illumina® GA, and HiSeq systems	http://www.softgenetics.com/NextGENe.html
Open source solutions			
22	ArrayExpressHTS	BioConductor package that allows preprocessing, quality assessment and estimation of expression of RNA-Seq datasets	http://bioconductor.org/packages/2.11/bioc/html/ArrayExpressHTS.html
23	easyRNASeq	Count summarization and normalization for RNA-Seq data	http://www.bioconductor.org/packages/2.11/bioc/html/easyRNASEq.html
24	Galaxy:	Galaxy is a general purpose workbench platform for computational biology	http://galaxyproject.org/

Table 4 (continued)

Sl. No	Quality control and pre-processing data		
	Tool name	Description	Website
Functional, network & pathway analysis tools			
25	Blast2GO	Tool for functional annotation of (novel) sequences and the analysis of annotated data	http://www.blast2go.com/b2ghome
Visualization tools			
26	GBrowse	Combination of database and interactive web pages for manipulating and displaying annotations on genomes	http://gmod.org/wiki/GBrowse
27	Artemis	Genome Browser and Annotation Tool	http://www.sanger.ac.uk/resources/software/artemis/
28	Integrated Genome Browser	An application intended for visualization and exploration of genomes and corresponding annotations from multiple data sources	http://www.affymetrix.com/estore/partners_programs/programs/developer/tools/igbsource_terms.affx
RNA-Seq databases			
29	Queryable-rna-seq-database	System is designed to simplify the process of RNA-seq analysis by providing the ability to upload the result data from RNA-Seq analysis into a database, store it, and query it in many different ways	https://github.com/fatPerlHacker/queryable-rna-seq-database
30	RNA-Seq Atlas	A reference database for gene expression profiling in normal tissue by next generation sequencing	http://medicalgenomics.org/rna_seq_atlas
31	SRA	The Sequence Read Archive (SRA) stores raw sequencing data from the next generation sequencing platforms	http://www.ncbi.nlm.nih.gov/sra
SSR identification tools			
32	MISA	MIcroSAtellite identification tool	http://pgrc.ipk-gatersleben.de/misa/misa.html
SNP identification tools			
33	SNPsFinder	Web-based program developed to facilitate the SNPs discovery process	http://snpsfinder.lanl.gov/
Primer designing			
34	BatchPrimer3	Comprehensive web primer design program using Primer3 core program as a major primer design engine to design different types of PCR primers and sequencing primers in a high-through manner	http://probes.pw.usda.gov/batchprimer3/

the identification of unique genes, effectors of the immune response, in resistant cultivar and many of which were validated through RT-qPCR.

Genus *Mycosphaerella* is the most important foliar fungal pathogen on banana. Three species of this genus, such as *M. fijiensis, M. musicola and M. eumusae* cause black leaf streak disease (BLSD), Sigatoka leaf spot (yellow Sigatoka) and *eumusae* leaf spot respectively. Passos et al. (2012) performed Roche 454 pyrosequencing of expressed genes in contrasting genotypes Calcutta 4 (R) and Cavendish cv. Grand Naine (S) for BLSD resistance. Many unigenes which are potentially involved in signal transduction like receptor protein kinase containing LRR repeats, serine/threonine protein kinase, mitogen activated protein kinase kinase, WRKY super family transcription factors, which are typically associated with plant immunity mechanisms were identified. More accumulation transcripts of OXOs (Oxalate oxidase) only in incompatible interactions suggested their possible involvement in ROS (Reactive Oxygen Species) and associated HR (Hypersensitive Response) components. OXO's catalyse the conversion of ROS into hydrogen peroxide (H_2O_2), an important component of hypersensitive response HR in plants. Similarly upregulation and early expression of Glutathione S transferase, metallathionein families in incompatible interactions revealed their involvement in cell protection like ROSscavenging during HR re-sponses.

M. balbisiana, is one of the ancestors of the present day triploid banana cultivars, harbouring many pest and disease resistant and drought tolerant genes (Vanhove et al. 2012; Liu et al. 2010; Uma et al. 2013). Whole transcriptome profiling of *M. balbisiana* has been completed by National Research Centre for Banana (NRCB), India, using Ion-torrent platform. The quality reads obtained from these sequencing were assembled with AA reference genome (*M. acuminata* DH Pahang). The annotated results revealed that approximately 3301 unigenes were found to be homologous to known defense-related genes in other plants. Plant CYC database analysis also stated that these unigenes were significantly enriched in various known resistance related metabolic or signaling pathways. Overlaying of these transcripts with the recently sequenced gene models of B genome (*M. balbisana* type Pisang Klutuk Wulung) is in progress to identify the unique candidate genes, detection of SNPs and SSRs in resistant/defense related genes.

Under the Network project of Transgenic crops—Functional Genomics component, of ICAR, India, NRCB has carried out banana transcriptomic studies for various biotic (*Eumusae* leaf spot and root lesion nematode) and abiotic (water defecit) stresses. The experiment for each stress was carried out by constructing four cDNA libraries in both resistant and susceptible cultivars with challenged and unchallenged conditions. These libraries were subjected to next generation sequencing using Illumina platform and quality reads were assembled with AA reference genome. This project was initiated with an aim to have a near-complete snapshot of transcriptome, including the rare transcripts that have regulatory roles in various biotic and abiotic stresses in compatible and incompatible genotypes. Besides, the DGE analysis of these transcriptome data of the contrasting cultivars under challenged and unchallenged conditions was also carried out to identify the differentially expressed genes in resistant cultivars.

By comparing the expression profiling of *M. eumusae* challenged and unchallenged resistant and susceptible cultivars, 46 unique genes which are involved in resistance related pathways were detected only in resistant genotype (cv. Manoranjitham). Similarly on KEGG analysis, it was observed that among all defense related transcripts, 953 were involved in PTI (PTI is a branch of plant immunity which involves interactions between host pattern recognition receptor-like kinases (PRRs) and pathogen associated molecular patterns (PAMPs)), 816 transcripts in plant hormones biosynthesis and signaling, 88 transcripts in pathogenesis—related and 346 in cell-wall modification. 54 transcripts in effectors triggered immunity (ETI) etc.

Similarly, 16 genes which are expressed only in root lesion nematode challenged resistant cultivar (Karthobiumtham-ABB) were considered as unique genes and in-depth study on these putative genes is in progress to confirm their role in on resistance mechanism. The digital gene expression analysis also revealed that the genes involved in phenylpropanoid pathway, flavanoid and lignin biosynthesis pathway are over expressed in resistant genotypes under challenged condition compared to unchallenged resistant genotype and challenged susceptible genotype (cv. Nendran). Portal et al. (2011) studied the compatible interaction of *M. fijiensis* with banana through the construction of suppression subtractive hybridization. Upon infection, it was found that, jasmonic acid and ethylene signaling transduction dependent defense mechanisms, such as production of membrane damaging PR proteins and GDSL like-lipase were induced only in later stages. However, activation of the JA and ET pathways have not prevented BLSD development in the compatible interaction of cultivar 'Grand Naine', probably because the defense mechanisms induced were repressed again at very late stages of infection. At this stage, plant defense mechanisms appear to be actively suppressed, while the fungal UGPase encoding gene with a possible role in membrane protection as an osmoprotectant is induced.

3 Databases and Tools for *Musa* Transcriptome Analysis

Till date (2013), 55854 *Musa* EST sequences have been submitted in the Expressed Sequence Tags database (dbEST) (http://www.ncbi.nlm.nih.gov/dbEST). The dbEST contains sequence data and other information on single-pass cDNA sequence or Expressed Sequence Tags. The number of ESTs with respect to each *Musa* genomic group submitted in the dbEST is given in Table 5.

The *Musa* transcriptome database maintained by Global *Musa* Genomics Consortium has 91,041 banana cDNA sequences till date in public domain. To support post-genomic efforts, Droc et al. (2013) improved existing systems (eg. web front end, query builder), by integrating available *Musa* data into generic systems (e.g. markers and genetic maps, synteny blocks). They have made inter operable with the banana hub (http://banana-genome.cirad.fr/) and other existing systems containing *Musa* data (e.g. transcriptomics, rice reference genome, workflow manager) and generated new results from sequence analyses (e.g. SNP and polymorphism analysis). Thus study of the banana genome has been enhanced by a number of tools and

Table 5 Number of *Musa* ESTs available in the NCBI public domain

Musa species, hybrids	No. of ESTs
M. acuminata (AA group)	29,525
Musa x paradisiaca	11,154
Musa (ABB Group)	11,070
M. acuminata (AAA Group)	7651
M. balbisiana (BB group)	5289
M. acuminata subsp. burmannicoides (AA group)	2289
Musa (AAB Group)	24
Musa hybrids and cultivars	6
Total	55,854

The Banana Genome Hub is supported by the South Green Bioinformatics Platform (http://southgreen.cirad.fr/), which provides access to original bioinformatics methods and tools to manage genetic and genomic resources of tropical plants. The comparative genomics is critical for the generation of reliable gene function annotations. The Banana Genome Hub relies on a robust comparative genomics database called GreenPhylDB (http://greenphyl.cirad.fr), which is now added with the *Musa* protein coding genes. This database comprises complete proteome sequences from the major plant phyla. The clustering of these proteomes was performed to define a consistent and extensive set of homeomorphic plant families. Based on this, lists of gene families such as plant or species specific families and several tools are provided to facilitate comparative genomics within plant genomes. Comparison involves two main steps; gene family clustering and phylogenomic analysis of the generated families. Once a group of sequences (cluster) is validated, phylogenetic analyses are performed to predict homolog relationships such as orthologs and ultraparalogs. GreenphylDB is linked to Uniprot DB to access *M. acuminata* proteome.

TropGeneDB (http://tropgenedb.cirad.fr) is a database that manages genomic, genetic and phenotypic information about tropical crops. Banana markers, maps, germplasm and genotypes data module are included in this DB. Genotyping 546 accessions of *Musa* germplasm with 22 SSR markers data are included in this DB.

Expressed Sequence Tag Treatment and Investigation Kit (Esttik) (http://esttik.cirad.fr) is a tool for editing and assembling of raw data of cDNA to form unigenes. The pipeline executes a series of programs to assess quality and nucleotides from chromatograms, then edits and assembles the input DNA sequence information into a non-redundant data set. This unigene is then searched for SSRs and SNPs. It is used as input for annotation against public databases including an extraction of Gene Ontology terms (Argout et al. 2007).

HarvEST: *Musa* displays EST sequences from 13 cDNA libraries from *M. acuminata*, *M. balbisiana* and related species. All sequences were derived from trace files received from various sources. Although HarvEST contains BLASTX hits from UniProt it needs updating (Close et al. 2007).

GNPAnnot (http://www.gnpannot.org/) is a project on green genomics which intends to develop a community system of structural and functional annotation supported by comparative genomics and dedicated to plant, insect and fungal genomes

allowing both automatic predictions and manual curations of genomic objects. Community annotation system for *M. acuminata* is also provided in this website.

4 Application of Transcriptome Analysis in Banana Improvement

Gene-derived SSRs can be associated with functional genetic variation, as they are located in transcribed regions potentially influencing the gene function, transcription or translation (Tranbarger et al. 2012). Consequently, these EST SSRs offer potential tool for marker assisted breeding, with markers either originating from a gene for a desirable phenotypic trait, or co-localizing with a particular quantitative trait locus. With such markers isolated from coding regions, conservation is also potentially greater, increasing transferability to related species (Gupta et al. 2003). Amorim et al. (2012) and Wang et al. (2007) used EST-SSR markers for banana germplasm accessions and revealed that these markers are highly informative SSR's which can be used to accelerate the banana breeding programs, improve the studies on genetic diversity and leading to the development of saturated genetic linkage maps.

Passos et al. (2013) tried to develop defense related gene derived SSR loci, selected on the basis of BLAST similarities and KOG (eukaryotic orthologous groups of proteins). A total of 4068 genic-SSR loci were identified in Calcutta 4 and 4095 in Cavendish cv. Grande Naine. A subset of 95 potential defense-related, gene-derived SSRs loci was validated for specific amplification and polymorphism across *M. acuminata* accessions. This result evidenced that only 14.7% of markers showed polymorphism among 20 *M. acuminata* accessions with alleles per polymorphic locus ranging from 3 to 8. In spite of limited diversity observed while using gene derived markers, the polymorphic markers were applicable for association with trait loci and down stream marker assisted selection, as well as evolutionary analysis, parentage assessment and general genotyping applications in breeding programs.

Similarly at NRCB, SSR database containing 5362 SSRs with primer details, product size, annealing temperature etc. has been developed using the *Musa* ESTs available at NCBI. This assisted in the identification of two putative candidate markers which could differentiate the nematode resistant and susceptible banana cultivars.

Based on transferability studies Backiyarani et al. (2012) reported that SSR markers developed from *Musa* ESTs could be used across different genera of the family Zingiberaceae. These transferable *Musa* EST-SSR markers could also be used in cardamom improvement program as the availability of cardamom specific primers are very meager. Fatokun et al. (1992) reported that different plant species often share orthologous genes for very similar functions. This suggested that *Musa* EST-SSR markers have potential to develop conserved orthologous set (COS) markers, which facilitate comparative genomic studies between species and members of the order Zingiberales

Table 6 Number of SSRs and SNPs developed from various cDNA libraries of *Musa* cultivars

Libraries	*M. eumusae*		Nematode (*P. coffeae*)		Water deficit stress	
	SSR	SNP	SSR	SNP	SSR	SNP
Unchallenged resistant	4825	83	11,650	432	9692	451
Challenged resistant	9153	355	3408	287	8128	351
Unchallenged susceptible	5008	79	8250	143	8478	92
Challenged susceptible	9239	271	7512	305	3775	334
Total	28,225	788	30,820	1167	30,073	1228

The transcriptome data on 12 cDNA libraries available at NRCB were subjected to *in silico* analysis to predict the SSR and SNP markers and resulted in the identification of a maximum of 11,650 and 355 SSRs and SNPs respectively (Table 6). *Musa* SNP database is being developed with all associated information.

Wan and Pentecost (2013) suggested that identification of multiple components involved in the chitin-mediated defense pathway (chitin degradation and signal generation; signal perception and amplification and downstream signaling components) is an important strategy to develop resistance to diverse biotic stresses. Hence, efforts were made to identify the chitinase isoforms from the available transcriptome data. DGE analysis of *M. eumusae* challenged resistant cultivar and *P. coffeae* challenged resistant cultivar indicated that the class II chitinase belonging to PR3 protein family was found to be commonly over expressed in both nematode and *M. eumusae* resistant genotypes. This information could be used for selection of genes to develop multiple resistances in banana through transgenic approach.

Shekhawat et al. (2011a) identified a novel SK3 type dehydrin (*Musa*DHN-1) from banana leaf cDNA library (Gene bank accession No. ES 436956) and found that this gene is significantly upregulated in abiotic stresses like cold, drought and high salinity in a robust and hardy cultivar 'Karibale' (Monthan, ABB). This gene is being used to improve drought and salt stress tolerance in cv. Rasthali through genetic transformation. Shekhawat et al. (2011b) also identified a banana EST sequence corresponding to a WRKY gene based on comparative analysis of stressed and non-stressed tissue derived EST data sets obtained from cv. Cachaco, a drought tolerant variety. Shekhawat et al. (2013) developed a transgenic Rasthali by incorporating this novel WRKY gene through transgenic approach and displayed enhanced tolerance towards oxidative and salt stress. Although *Musa*WRKY71 is induced in response to elicitor molecules of biotic stress response pathways like ethylene, salicylic acid, *Musa*WRKY71 over expressing plants were found to be equally susceptible to the infection of *Fusarium oxysporum* f. sp. *cubense* as the untransformed control plants. These results indicated that *Musa*WRKY71 is an important constituent in the transcriptional reprogramming involved in diverse stress responses in banana.

Conclusion

Transcriptome analysis of various tissues exposed to different stresses enables the identification of differentially expressed genes. Among various methods available, genome wide expression profiling would allow the measurement of the activity of majority of genes and to understand the molecular mechanism involved. In 1990s EST sequencing was the best approach for rapid identification of expressed genes, or at least fragments of those genes. Although at that time EST sequencing was considered a very high-throughput technique, both costs and technical limitations prevented the development of a complete transcript catalog. As a consequence, much of our knowledge on the protein-coding region of the transcriptome relied on different computational gene prediction methods. Other technologies like tag based methods (SAGE and MPSS) are the best methods over ESTs as these approaches have the capability to quantify the level of gene expression, but fail to distinguish the isoforms. In addition, most of them are based on traditional Sanger sequencing technology, which is expensive to apply on a large scale. A hybridization-based approach like microarray is one of the inexpensive ways to detect and quantify the transcripts on a large scale. Though this technique has some advantages like high throughput and ability to quantify distinct spliced isoforms, it has some disadvantages too. Because of differences in hybridization strength, cross-hybridization and other experimental variables, microarrays provide a noisy output signal. They can measure only the genes for which the sequence and the precise exon-intron boundaries are known, thus failing to identify novel genes or novel splicing events. Recently, RNA-seq technologies provide unprecedented opportunities for characterizing the set of RNA transcripts produced in tissues. Many platforms are available to sequence the RNAs, the choice of platforms varies with the objectives (de novo sequencing and resequencing) and budget availability. Unlike hybridization-based methods, it is not limited to the detection of known transcripts, but can also measure a broader range of expression levels. Among its other advantages, RNA-seq data has relatively low background noise, achieves base-pair resolution, allows precise identification of exon and intron boundaries, detecta single nucleotide polymorphisms (SNPs) and other variants within transcripts.

In banana, transcriptome analysis has been carried out to understand the molecular mechanisms involved during fruit ripening and various biotic and abiotic stress conditions using different approaches like SSH, cDNA-AFLP, ESTs and microarray. After the development of next generation sequencing, many have been carried out to understand the *Musa*-pathogen interactions and to identify the unique genes responsible for inducing resistance. The generation of enormous amount of sequence data from these transcriptome analyses has resulted in development of genetic markers like EST-SSRs, SNPs etc. These have led to the development of trait specific markers being used in conventional breeding for the selection of parents and progenies economising the time and resources needed for field evaluation. Besides these markers are valuable resources for comparative mapping in evolutionary studies.

It is evidenced from previous studies that there are differences in expression levels across homeologous chromosomes suggesting independent contribution of the two genomes to banana metabolism. This emphasises the need of both A and B genome as reference sequences for the transcriptome mapping of any banana and plantain cultivar irrespective of its genomic group (AAA, AA, AAB, ABB, BB). The fact that nearly 20% of the transcripts do not hit with any known functional genes emphasise the need to study the function of uncharacterized genes either through RNAi and/or knockout gene approaches. This may lead to the identification of novel genes for important traits like parthenocarpy, sterility etc. The outcome of the research will not only be useful for improving banana but also paves way to develop seedlessness in other fruits crops.

References

Althoff DM, Gitzendanner MA, Segraves KA (2007) The utility of amplified fragment length polymorphisms in phylogenetics: acomparison of homology within and between genomes. Syst Biol 56:477–484

Amorim EP, Silva PH, Ferreira CF, Amorim VBO, Santos VJ, Vilarinhos AD, Santos CMR, Souza Júnior MT, Miller RNG (2012) New microsatellite markers for bananas (*Musa* spp). Genet Mol Res 11:1093–1098

Anon (1994) FAO Trade yearbook 1993. FAO, Rome

Argout X, Ruiz M, Fouet O, Lanaud C, Wincker P, Da Silva C, Courtois B (2007) ESTtik: a semi-automatic cDNA sequence analysis and annotation pipeline including SSR and SNP search tools. 15th International Cocoa Research Conference 1:501–506

Backiyarani S, Uma S, Varatharj P, Saraswathi MS (2012) Mining of ESTSSR markers of *Musa* and their transferability studies among the members of order the Zingiberales. Appl Biochem Biotechnol 169(1):228–238

Backiyarani S, Uma S, Arunkumar G, Saraswathi MS, Sundararaju P (2014) Differentially expressed genes in incompatible interactions of *Pratylenchus coffeae* with *Musa* using suppression subtractive hybridization. Physiol Mol Plant Pathol 86:11–18

Banana Market Review and Banana Statistics 2012–2013 Market (FAO statistics), Rome, 2014

Bortoluzzi S, Danieli GA (1999) Towards an in silico analysis of transcription patterns. Trends Genet. 15:118–119

Cano-Delgado A, Penfield S, Smith C, Catley M, Bevan M (2003) Reduced cellulose synthesis invokes lignification and defense responses in *Arabidopsis thaliana*. Plant J 34:351–362

Close TJ, Wanamaker S, Roose ML, et Lyon M (2007) HarvEST. Methods Mol Biol 406:161–177

Coemans B, Matsumura H, Terauchi R, Remy S, Swennen R, Sagi L (2005) SuperSAGE combined with PCR walking allows global gene expression profiling of banana (*Musa acuminata*), a non-model organism. Theor Appl Genet 111:1118–1126

Costa JL, Herai RH, Haddad F, Nebó JF, Nishimura DS, Ferreira CF, Amorim EP, e Figueira AVO (2013) Analysis of differentially expressed genes in banana cultivars infected by *Fusariumoxysporum* f. sp. *cubense*. In: IV Simposio Brasileiro De Genetica Molecular De Plantas, p 33

Davey MW, Graham NS, Vanholme B, Swennen R, May ST, Keulemans J (2009) Heterologous oligonucleotide microarrays for transcriptomics in a non-model species; a proof-of-concept study of drought stress in *Musa*. BMC Genomics 10(436):1–19

Davey MW, Gudimella R, Harikrishna JA, Sin LW, Khalid N, Keulemans J (2013) A draft *Musa balbisiana* genome sequence for molecular genetics in polyploid, inter- and intra-specific *Musa* hybrids. BMC Genomics 14:683

Denham TP, Haberle SG, Lentfer C et al (2003) Origins of agriculture at Kuk Swamp in the Highlands of New Guinea. Science 301:189–193

D'Hont A, Denoeud F, Aury J, Baurens F, Carreel F, Garsmeur O et al (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature 488:213–217

Diatchenko L, Lau YF, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED, Siebert PD (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc Natl Acad Sci 93:6025–6030

Droc G, Lariviere D, Guignon V, Yahiaoui N, This D, Garsmeur O, Dereeper A, Hamelin C, Argout X, Dufayard JF, Lengelle J, Baurens FC, Cenci A, Pitollat B, D'Hont A, Ruiz M, Rouard M, Bocs S (2013) The banana genome hub. Database (Oxford) 2013:1–14

Fatokun CA, Menacio DI, Danesh D, Young ND (1992) Evidence for orthologous seed weight genes in cowpea and mungbean, based upon RFLP mapping. Genetics 132:841–846

Geschwind DH, Ou J et al (2001) A genetic analysis of neural progenitor differentiation. Neuron 29(2):325–339

Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Genet Genomics 270:315–323

Jin ZQ, Xu BY, Liu JH, Su W, Zhang JB, Yang XL, Jia CH, Lia MY (2009) Identification of genes differentially expressed at the onset of the ethylene climacteric in banana. Postharvest Biol Technol 52:307–309

Kesari R, Kumar Trivedi P, Nath P (2007) Ethylene induced ripening in banana evokes expression of defense and stress related genes in fruit tissue. Postharvest Biol Technol 46:136–143

Lassois L, de Lapeyre deBL, Jijakli MH (2008) Biological control of crown rot of bananas with *Pichia anomala* strain K and *Candida oleophila* strain O. Biol Control 45:410–418

Lassois L, Frettinger P, de Lapeyre deBL, Lepoivre P, Jijakli H (2011) Identification of genes involved in the response of banana to crown rot disease. Mol Plant Microbe Interact 24(1):143–153

Li CC, Deng G, Yang J, Jin AVY, Kuang R, Zuo C, Zhi-cheng L (2012) Transcriptome profiling of resistant and susceptible Cavendish banana roots following inoculation with *Fusarium oxysporum* f. sp. *cubense* tropical race 4. BMC Genomics 13:374

Lisitsyn N, Lisitsyn N, Wigler M (1993). Cloning the differences between two complex genomes. Science 259:946–951

Liu HY, Dai JR, Feng DR, Liu B, Wang HB, Wang JF (2010) Characterization of a novel plantain Asr gene, MpAsr, that is regulated in response to infection of *Fusarium oxysporum* f. sp. *cubense* and abiotic stresses. J Integr Plant Biol 52:315–323

Manrique- Trujillo SM, Ramírez LAC, Ibarra Laclette E, Gómez-Lim MA (2007) Identification of genes differentially expressed during ripening of banana. J Plant Physiol 164:1037–1050

Maskos U, Southern EM (1992) Oligonucleotide hybridizations on glass supports: a novel linker for oligonucleotide synthesis and hybridization properties of oligonucleotides synthesised in situ. Nucl Acids Res. 20(7):1679–84

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-SEq. Nat Methods 5:621–628

Munro CL (2008) Identification of defense-related genes in banana against *Fusarium* Wilt. MSc dissertation, University of Pretoria, Pretoria, p 137

Passos MA, de Oliveira CV, Emediato FL, de Camargo TC, Souza MT Jr, Matsumoto T, RennóAzevedo VC, Ferreira CF, Amorim EP, de Alencar Figueiredo LF, Martins NF, de Jesus B, Cavalcante M, Baurens FC, da Silva OB Jr, Pappas GJ Jr, Pignolet L, Abadie C, Ciampi AY, Piffanelli P, Miller RN (2012) Development of expressed sequence tag and expressed sequence tag-simple sequence repeat marker resources for *Musa acuminata*. AoB Plants 2012:1–63

Passos MA, de Cruz VO, Emediato FL, de Teixeira CC, Azevedo VC, Brasileiro AC, Amorim EP, Ferreira CF, Martins NF, Togawa RC, Júnior GJ, da Silva OB Jr, Miller RN (2013) Analysis of the leaf transcriptome of *Musa* acuminata during interaction with *Mycosphaerella musicola*: gene assembly, annotation and marker development. BMC Genomics 14(78):1–17

Ploetz RC (2006) *Fusarium* wilt of banana is caused by several pathogens referred to as *Fusarium* oxysporum f. sp. *cubense*. J Phytopathol 96:653–656
Portal O, Izquierdo Y, De Vleesschauwer D, Sánchez-Rodríguez A, Mendoza-Rodríguez M, Acosta-Suárez M, Ocaña B, Jiménez E, Höfte M (2011) Analysis of expressed sequence tags derived from a compatible *Mycosphaerella fijiensis* banana interaction. Plant Cell Rep 30:913–928
Ravishankar KV, Rekha A, Swarupa V, Savitha G (2011). Gene expression analysis in roots of *Musa* acuminata ssp. burmanniccoides 'Calcutta-4', a banana genotype tolerant to *Fusarium* wilt. Acta Horticult (ISHS) 897:363–370
Roux N, Baurens F, Doležel J, Hřibová E, Harrison PH, Town C, Sasaki T, Matsumoto T, Aert R, Remy S, Souza M, Lagoda P (2008) Genomics of banana and plantain (*Musa* spp.), major staple crops in the tropics. Plant Genet Genomics Crops Models 1:83–111
Rudd S (2003) Expressed sequence tags: alternative or complement to whole genome sequences. Trends Plant Sci 8:321–329
Santos CMR, Martins NF, Horberg HM, de Almeida ERP, Coelho MCF, Togawa RC, da Silva FR, Caetano AR, Miller RNG, Souza MT (2005) Analysis of expressed sequence tags from *Musa acuminata* ssp. *burmanniccoides*, var. Calcutta 4 (AA) leaves submitted to temperature stresses. Theor Appl Genet 110:1517–1522
Shekhawat UKS, Ganapathi TR (2013) *Musa* WRKY71 overexpression in banana plants leads to altered abiotic and biotic stress responses. PLoS ONE 8(10):e75506
Shekhawat UK, Srinivas L, Ganapathi TR (2011a) *Musa* DHN-1, a novel multiple stress-inducible SK(3)-type dehydrin gene, contributes affirmatively to drought- and salt-stress tolerance in banana. Planta 234(5):915–932
Shekhawat UKS, Ganapathi TR, Srinivas L (2011b) Cloning and characterization of a novel stress-responsive WRKY transcription factor gene (*Musa* WRKY71) from *Musa* spp. cv. Karibale-Monthan (ABB group) using transformed banana cells. Mol Biol Rep 38:4023–4035
Swarupaa V, Ravishankar KV, Rekha A (2013) Characterization of tolerance to *Fusarium* xysporumf. sp., cubense infection in banana using suppression subtractive hybridizationand gene expression analysis. Physiol Mol Plant Pathol 83:1–7
Tranbarger TJ, Kluabmongkol W, Sangsrakru D, Morcillo F, Tregear JW, Tragoonrung S, Billotte N (2012) SSR markers in transcripts of genes linked to posttranscriptional and transcriptional regulatory functions during vegetative and reproductive development of *Elaeis guineensis*. BMC Plant Biol 12:1–12
Uma S, Sudhakar B, Thangavelu R, Saraswathi MS, Backiyarani S, Saravanakumar AS, Vasanthi R (2011) EST based probing and isolation of functionally important full length transcripts against sigatoka (*Mycosphaerella musae*) leaf spot and confirmation by semi quantitative real time PCR in *Musa accuminata* cv. Manoranjitham (AAA). In: XVIII International Conference on Plant & Animal Genomes held at Town & Country convention Center, San Diego, CA, 13-18 Jan 2011
Uma S, Backiyarani S, Saraswathi MS, Mustaffa MM (2013) Banana genome to enrich fruit basket. Indian Horticult 58(2):3–6
Van denBN, Hein I, Birch PRJ, Berger DK, Wingfield MJ, Viljoen A (2007) Tolerance in banana to *Fusarium* wilt is associated with early up-regulation of cell wall-strengthening genes in the roots. Mol Plant Pathol 8(3):333–341
Vanhove AC, Vermaelen W, Panis B, Swennen R, Carpentier SC (2012) Screening the banana biodiversity for drought tolerance: can in vitro growth model and proteomics be used as a tool to discover tolerant varieties and understand homeostasis. Front Plant Sci 3:176
Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science 270:484–487
Wan J, Pentecost G (2013) Potential application of chitin signaling in engineering broad spectrum disease resistance to fungal and bacterial pathogens in plants. Adv Crop Sci Tech 1:e103
Wang N, Hu J, Ohsawa R Y O, Ohta M, Fujimura T (2007) Identification and characterization of microsatellite markers derived from expressed sequence tags (ESTs) of radish (Raphanus satvius L.). Mol. Ecol. Notes 7: 503–506

Wang L, Wang X, Wang X, Liang Y, Zhang X (2011) Observations on novel splice junctions from RNA sequencing data. Biochem Biophys Res Commun 409(2):299–303
Wang Z, Zhang J, Jia C, Liu J, Li Y, Yin X, Xu B, Jin Z (2012) *De Novo* characterization of the banana root transcriptome and analysis of gene expression under *Fusarium oxysporum* f. sp. *cubense* tropical race 4 infection. BMC Genomics 13:650
Xu BY, Su W, Liu JH, Wang JB, Jin ZQ (2007) Differentially expressed cDNAs at the early stage of banana ripening identified by suppression subtractive hybridization and cDNA microarray. Planta 226:529–539
Yang GP, Ross DT et al (1999) Combining SSH and cDNA microarrays for rapid identification of differentially expressed genes. Nucleic Acids Res 27(6):1517–1523