Genes, Genomes and Genomics ©2010 Global Science Books

Mining for Simple Sequence Repeats from Expressed Sequence Tag Libraries of Banana

Aykkal Riju[1] • Vadivel Arunachalam[2*]

[1] Aikkal House, Kannur, Kerala - 670564, India
[2] Molecular Biology and Bioinformatics Laboratory, Central Plantation Crops Research Institute, Kudlu, Kasaragod, Kerala- 671 124, India
Corresponding author: * v.arundevi@gmail.com, vadivelarunachalam@yahoo.com

ABSTRACT

The purpose of this study was to mine the simple sequence repeats (SSRs) in expressed sequence tags (ESTs) of banana (*Musa* spp.). We retrieved 31,066 EST sequences of *Musa* spp. belonging to different tissues and conditions from dbEST of National Centre for Biotechnology Information. These were made into 4,526 contigs as well as 9596 singletons using sequence assembly program CAP3. SSRs were located by using a MISA perl script, and were classified based on size of repeat and type of the repeating motif. A total of 2,659 SSR sites were observed and 2308 primer pairs were designed. Mononucleotide repeats (44%) were found to be more abundant followed by dinucleotide (23%) and trinucleotide (21%). We annotated the EST contigs and found skp1, putative 60S ribosomal protein L39, DET2, histone H1, GSK1 (gsk3/shaggy-like protein kinase 1), heat shock protein 70, phagocytosis and cell motility protein ELMO1, universal stress protein family as functions encoded. A database was constructed and made available at http://www.riju.byethost31.com/banana/ giving the results of SSRs and primers from the study.

Keywords: bioinformatics, database, expressed sequence tags, fruit, microsatellite, *Musa*

INTRODUCTION

The objective of the present study is to mine the expressed sequence tags (ESTs) for the presence of simple sequence repeats (SSRs) and to design primers flanking these sites. We also intend to develop an online database using the results of the study.

Musa species (Zingiberaceae, Zingiberales) including bananas and plantains are collectively the fourth most important crop after rice, wheat and maize (Arias *et al.* 2003) in terms of gross value of production. It is produced in 130 countries in tropical and sub-tropical regions of the world of mostly developing economies (UNCTAD). Globally, banana and plantain are cultivated in 8.834 million ha with a total production of 100 million tones (FAOSTAT). It is a staple food, a source of fiber and edible plate and an export commodity and contributes to the food security of millions of people of the developing world. The crop is also traded in local markets providing income and employment to rural populations. Each and every part of the banana plant is useful in one form or other. It is a nutrient dense crop with high calorific value and medicinal properties. Bananas are the world's most exported fresh fruit in terms of volume and value (UNCTAD 2006). *Musa* is a member of the monocot order Zingiberales, a Commelinid lineage that diverged from the line leading to rice (Poales) in the mid-cretaceous period over 100 million years ago (Janssen and Bremer 2004; Sanderson *et al.* 2004).

The *Musa* species *M. acuminata* (AA genome) and *M. balbisiana* (BB genome), both with 2n=22 chromosomes, are the two main progenitors of cultivated banana varieties (Lysak *et al.* 1999). Cultivated bananas range in their ploidy level from diploids, triploids and few tetraploids. The haploid genome of *Musa* species is estimated to vary between 560 to 600 Mb in size (Lysak *et al.* 1999; Kamate *et al.* 2001), just four times larger than that of the model plant *Arabidopsis* (125 Mb) and 30% larger than that of rice (390Mb). Genetic maps have been developed for *Musa* (Fauré *et al.* 1993) and recently, bacterial artificial chromosome (BAC) resources were generated for both *M. acuminata* (Vilarinhos *et al.* 2003; Ortiz-Vazquez *et al.* 2005) and *M. balbisiana* (Safar *et al.* 2004). ESTs provide researchers data on gene expression and regulation and are a time- and cost-effective means for discovering new genes. Recently, EST-SSRs have received much attention as the increasing amounts of ESTs being deposited in databases for many crops such as rice, wheat, etc. (Scott *et al.* 2000; Kantety *et al.* 2002; Varshney *et al.* 2002). EST-SSRs can be rapidly developed from EST database by data mining tools at low cost, and due to their existence in transcribed region of genome, they can lead to the development of genic marker-based maps which help to identify candidate genes and increase the efficiency of marker-assisted selection (MAS) (Gupta and Rustgi 2004). In addition, EST-SSRs show a higher level of transferability to closely related species than genomic SSR markers (Scott *et al.* 2000; Saha *et al.* 2004) and can be served as anchor markers for comparative mapping and evolutionary studies (Varshney *et al.* 2005). SSRs are one of the most widely used molecular markers in crop breeding, population genetics, mapping and marker assisted selection (MAS). Being gene specific functional markers, they provide an efficient tool to link phenotypic and genotypic variation (Powell *et al.* 1996; Gupta and Varshney 2000). Based on the length of SSR tracts and their potential as genetic markers, they are categorized into two groups – Class I or hypervariable markers (length of SSR $\geq$ 20bp) and Class II (length of SSR vary 12 bp $\geq$ and $<$ 20 bp) or potentially variable markers (Temnykh *et al.* 2001). Despite the importance of the crop and availability of adequate sequences, there are few studies (Wang *et al.* 2008) on using EST derived molecular markers in *Musa*.

MATERIALS AND METHODS

ESTs of *M. acuminata* from various tissues such as fruit, leaves etc. and under various conditions e.g. BSV activation, *Musa para-*

Received: 1 June, 2009. Accepted: 1 April, 2010.

Original Research Paper

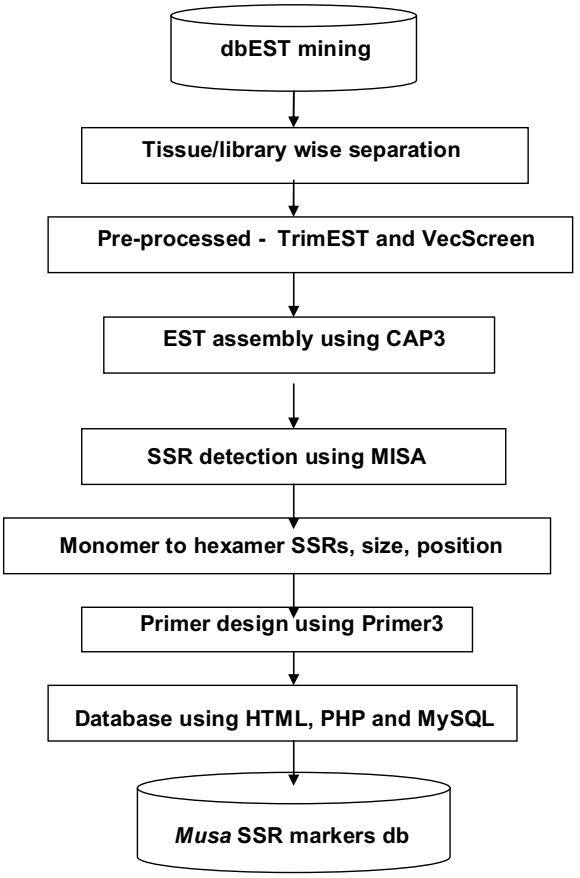

Fig. 1 Flowchart of *in silico* banana SSR discovery and database development.

disiaca, cold stress and hot stress were retrieved from dbEST of NCBI (http://www.ncbi.nlm.nih.gov/). These sequences are processed and analyzed to remove redundancy as well as vector contaminations. The poly A/T removal program trimest (http://mobyle.pasteur.fr/cgi-bin/portal.py?form=trimest) was used to eliminate poly A/T tails. CAP3 (Huang and Madan 1999) was used to remove the redundancy among the EST sequences and assemble the sequences to contigs and singletons. Perl script, MISA (Thiel *et al.* 2003) was used to detect all possible mono- to hexanucleotide repeats and compound repeats. We identified two types of SSRs: (i) perfect SSRs, with a exact repeat of any of motif with length 12bp or more, e.g. $(AT)_{15}$; (ii) compound repeats, combinations of two or more repeated motifs with length 20 or more, e.g. $(CA)_{16}(TC)_{10}$. For mononucleotides, although A, T, C and G are possible, A and T are grouped into a single category, since an A repeat on a strand is the same as a T repeat on the opposite strand. C on a strand is the same as a G on the opposite strand, resulting in 2 unique classes of mononucleotides, A/T and C/G (Katti *et al.* 2001). Similarly, all dinucleotide motifs were grouped into the 4 following unique classes: (*i*) AT/TA; (*ii*) AG/GA/CT/TC; (*iii*) AC/CA/TG/GT; and (*iv*) GC/CG. The trinucleotide repeats are grouped into 10 unique classes as per the SSR classification (Jurka *et al.* 1995; Katti *et al.* 2001).

Putative orthologues and functions of the contigs were identified using NCBI - BLASTX (www.ncbi.nlm.nih.gov). Primers pairs were designed flanking the detected SSR sites using primer3 tool (Rozen and Skaletsky 2000). An online database with user friendly options was created using MySQL to display the results of the study such as type of motif and repeat size along with list of primers (**Fig. 1**).

RESULTS AND DISCUSSION

A total of 31,066 EST sequences of banana plant were used for the study. It is observed that 5746 sequences contained suspected Poly(A/T) tails with minimum length of 4bp or more, and were removed and processed with contig assembly program CAP3 (Hung and Madan 1999). The CAP3 analysis resulted in identification of 14,122 unigene (1,02,63,544 bp), corresponding to 4,526 contigs and 9,596 singletons. A total of 180 singletons of size below 100 bp were removed and 13,942 remaining unigene sequences were used as good quality sequences. The average length of the sequence was found to be 735 bp. Using MISA tool, we found a total of 2659 SSR sites including 2389 perfect and 270 compound repeats. Among the perfect SSRs, 1109 belonged to the class I and 1280 belonged to the class II. We found the SSRs to occur in the banana genome at a frequency of 1 SSR per 3.9 Kb.

Microsatellites are abundant in genomes of *Onion* and *Pinus*, which display 38.9 and 41.5 SSRs per Mb respectively (Von Stackelberg *et al.* 2006). The frequency of SSR in cacao ESTs was 1 per 26.9 Kb (Riju *et al.* 2009). Frequency of SSRs in plants could be one in 2.5 Kb in eucalyptus (Rabello *et al.* 2005) to one in 28.3 Kb in maize (Cardle *et al.* 2000). By comparing frequency information from earlier reports (Cardle *et al.* 2000; Gao *et al.* 2003), *Musa* genome display higher density of microsatellites than maize, rice, soybean, wheat but lower than eucalyptus, onion and pines. In previous EST-SSR reports, authors have used different definition for SSR, varies by size and type of repeat and some authors do not consider monomer repeats as SSR. Many authors target class II SSRs with size of 10 bp and above. Few authors consider SSR only if the repeat motif larger than 20 bp (Varsheny *et al*. 2002). Hence comparison of SSR size and type of repeat are difficult to discuss. So the comparison of observed frequency information to the existing report will not be accurate and it should be noted that the frequency estimates in this study might not reflect the exact picture for *Musa* genome considering the limited number of ESTs covering only fraction of the genome (10.3 Mb).

A total of 1,180 mononucleotide (44.37%), 611 di-nucleotide repeats (23%), 568 tri-nucleotides repeats (21%), 19 tetra-nucleotide repeats, 6 penta-nucleotide repeats and 5 hexa-nucleotide repeats were also detected (**Table 1**). We found most abundant SSR type as mononucleotide repeats followed by dinucleotide and trinucleotide repeats. Among mononucleotides, the repeat A/T occurred 1175 times (99.6% of the mononucleotide). AG/GA/TC/CT was observed to be much abundant motif while GC/CG group was found to be very rare. AAG/TTC (26% of trinucleotide) followed by AGG/TCC (24.6%) were the abundant motifs among the trinucleotide repeats. Frequency distribution of perfect SSRs and motif types is given in **Table 2**.

Wang *et al.* (2008) also reported trinucleotide repeats as the most abundant SSR motif in banana using 2,284 unigenes. By using more number of unigene sequences, we found mononucleotide repeats as the most frequent repeat class in banana ESTs. A/T motifs were abundant among the monomer repeats in *Arabidopsis* (Cardle *et al.* 2000; Lawson and Zhang 2006), *Eucalyptus* (Rabello *et al*. 2005) and many crop species (Gao *et al.* 2003).

Among the dimeric repeat motifs, the AG/GA/TC/CT type was more abudant than other three motifs. The most common plant repeat motif was AA/TT followed by AT/TA and CT/GA (Lagercrantz *et al*. 1993). Many plants studied to date exhibit the AG motif as their dominant EST-SSR dinucleotide repeat as seen in *Arabidopsis*, tomato, poplar, and cotton (Cardle *et al.* 2000), barley (Kantety *et al.* 2002; Varshney *et al.* 2002), rice (Cardle *et al.* 2000; Temnykh *et al.* 2001; Kantety *et al.* 2002; Varshney *et al.* 2002; Gao *et al.* 2003), maize (Cardle *et al.* 2000; Kantety *et al.* 2002; Varshney *et al.* 2002; Gao *et al.* 2003), soybean (Cardle *et al.* 2000; Gao *et al.* 2003), sorghum (Kantety *et al.* 2002), wheat (Kantety *et al.* 2002; Varshney *et al.* 2002; Gao *et al.* 2003), oats and rye (Varshney *et al.* 2002), eucalyptus (Rabello *et al*. 2005), apple (Newcomb *et al.* 2006) and coffee (Aggarwal *et al.* 2007). But the other class of dinucleotide repeat AC/CA/TG/GT was found abundant in primitive plant forms such as mosses, algae and conifers (Von Stackelberg *et al.* 2006) and AT/TA in loblolly pine and

Table 1 Distribution of motifs length in SSRs in *Musa* species.

Motif	Number of total SSRs	Frequency (%)
Mono	1180	44.38
Di	611	22.98
tri	568	21.36
tetra	19	0.71
Penta	6	0.23
Hexa	5	0.19
Compound	270	10.15

Table 2 Distribution of motifs in perfect SSRs of EST-*Musa* species.

Motifs	No. of SSRs	Frequency (based on perfect SSR)
A/T	1175	50%
G/C	5	0%
TA/AT	70	3%
GC/CG	2	0%
AG/GA/TC/CT	504	21%
AC/CA/TG/GT	35	1%
AAC/ACA/CAA/GTT/TTG/TGT	12	1%
AAG/AGA/GAA/CTT/TTC/TCT	146	6%
AAT/ATA/TAA/ATT/TTA/TAT	8	0%
ACC/CCA/CAC/TGG/GGT/GTG	49	2%
ACG/CGA/GAC/TGC/GCT/CTG	43	2%
ACT/CTA/TAC/TGA/GAT/ATG	37	2%
AGC/GCA/CAG/TCG/CGT/GTC	64	3%
AGG/GGA/GAG/TCC/CCT/CTC	139	6%
AGT/GTA/TAG/TCA/CAT/ATC	11	0%
CCG/CGC/GCC/CGG/GGC/GCG	59	3%

Table 3 Tetramer and above repeats in expressed sequences of *Musa species*.

Sequence Id	Repeat motif	Repeat type	SSR length	Start	End
ES434872	(GCCGTC)6	Hexamer	36	595	630
ES432295	(CCCTT)5	Pentamer	25	1	25
ES434390	(GTCGA)5	Pentamer	25	142	166
ES435444	(CTGCA)6	Pentamer	30	573	602
FL667931	(ATGCC)7	Pentamer	35	593	627
ES432040	(GAAA)5	Tetramer	20	297	316
ES432808	(CCTT)5	Tetramer	20	550	569
ES434403	(TTCT)5	Tetramer	20	331	350
ES435316	(AGCG)5	Tetramer	20	32	51
FF558957	(TTCA)5	Tetramer	20	536	555
FL649719	(GATG)5	Tetramer	20	533	552
FL662380	(CGAT)5	Tetramer	20	79	98
FL664589	(ATAA)5	Tetramer	20	221	240
FL646580	(CTGC)6	Tetramer	24	632	655
FL665499	(GAAA)6	Tetramer	24	1	24
FF561074	(ATCT)9	Tetramer	36	73	108

Table 4 Tetramer and above repeats in EST-contig sequences of *Musa* species

Contig Id	Repeat motif	Repeat type	SSR length	Start	End
Contig3536	(GAAGAG)5	Hexamer	30	258	287
Contig532	(AGATGA)5	Hexamer	30	228	257
Contig928	(TCTCCA)5	Hexamer	30	150	179
Contig4219	(GAACCG)6	Hexamer	36	424	459
Contig224	(AGTGC)5	Pentamer	25	3	27
Contig819	(GAGAG)5	Pentamer	25	26	50
Contig1781	(CTAT)5	Tetramer	20	779	798
Contig3222	(TCGA)5	Tetramer	20	83	102
Contig3393	(TGCT)5	Tetramer	20	587	606
Contig3485	(AAGG)5	Tetramer	20	2	21
Contig407	(CGAT)5	Tetramer	20	657	676
Contig3427	(GGAT)6	Tetramer	24	1129	1152
Contig995	(CCTT)7	Tetramer	28	797	824
Contig982	(ATAG)9	Tetramer	36	364	399

spruce (Berube *et al.* 2007). The GT/CA motif is the most abundant dinucleotide repeat in mammals but was found to be considerably less frequent in plants (Lagercrantz *et al.* 1993).

Among the trimeric repeats, AGG/GGA/GAG/TCC/CCT/CTC motif was found to be the most frequent SSR. It was also most abundant trimeric repeat motif type in rice, barley, soybean (Gao *et al.* 2003) and in *Arabidposis* ESTs (Cardle *et al.* 2000) and many dicot plants (Kumpatla and Mukhopadhyay 2005). But the CCG/CGC/GCC/CGG/GGC/GCG type of trimerics is the most abundant motif in cereals (Varshney *et al.* 2002). Hence, trinucleotide repeats are expected to be the most abundant SSR class found in ESTs. The distribution and frequency of SSRs is a result of many factors such as mutation, selection and DNA repair mechanisms (Sreenu *et al.* 2007). We found few tetramer and above repeats in EST sequences (**Table 3**) and contigs (**Table 4**) repeats and 270 compound repeats. These rare repeats have potential as markers in detecting polymorphism and mapping studies.

So far, only few hundreds of SSR loci have been identified in *Musa* genome includes 79 microsatellite loci from various *M. acuminata* subspecies and *accessions*, such as *M. acuminata* ssp. *Malaccensis* (Ridl.) Simmonds; *M. accuminata* cv. Gobusik and *M. accuminata* ssp. *burmaniccoides* 'Calcutta 4' de Langhe & Devreux (Kaemmer *et al.* 1997; Crouch *et al.* 1998; Lagoda *et al.* 1998). Additional 25 microsatellite loci have been obtained from *M. balbisiana* (Buhariwalla *et al.* 2005) and 23 microsatellite loci from *M. acuminate* (Creste *et al.* 2006).

Recently TropGENE-DB - http://tropgenedb.cirad.fr (Ruiz *et al.* 2004) of CIRAD has listed 126 SSR markers developed for banana. Wang *et al.* (2008) reported additional set of 122 SSRs from 2,284 unigene sequences. Here we report a total of additional 2308 SSR markers, which could be used as potential markers to analyzing *Musa* genome.

Annotation of putative functions

After assembling the sequences to contigs, the unigenes were used for putative gene annotation by using BLASTX (http://blast.ncbi.nlm.nih.gov/Blast.cgi)program with an e-value cutoff of $1e^{-5}$ against the non-redundant protein sequence database of Viridiplantae (taxid:33090) among the entries from GenBank CDS translations, PDB, SwissProt, PIR and PRF. Putative functions such as shaggy-like kinase etha, GSK1 (GSK3/SHAGGY-LIKE PROTEIN KINASE 1), BIN2 (BRASSINOSTEROID-INSENSITIVE 2), heat shock protein 70, putative 60S ribosomal protein L39, DET2, histone H1, phagocytosis and cell motility protein ELMO1, universal stress protein family protein and most of hit was hypothetical protein or unnamed protein product. The result of the study is compiled as a public domain database http://www.riju.byethost31.com/banana/. This database contains the details of contigs and singleton sequences, SSR regions with start and end positions, 2308 primer pairs and information on additional primers, and putative genes annotated.

CONCLUSIONS

This study gives an insight into the frequency, type and distribution of *Musa* EST-SSRs. We also highlight the putative genes encoded by the ESTs under study. The markers designed by this study could enrich the current genomic resources for the *Musa* and related species. After polymerase chain reaction validation, the markers have practical implications in gene mapping, MAS, and genetic diversity studies in plantains and bananas.

REFERENCES

Aggarwal RK, Prasad SH, Varshney RK, Prasanna RB, Krishnakumar V, Lalji S (2007) Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. *Theoretical and Applied Genetics* **114**, 359-372

Arias P, Dankers C, Liu P, Pilkauskas P (2003) The world banana economy 1985–2002. FAO. Available online: http://www.fao.org/documents/show_cdr.asp?url_file=/docrep/007/y5102e/y5

102e00.htm

Bérubé Y, Zhuang J, Ralph S, Rungis D, Bohlmann J, Ritland K (2007) Characterization of EST-SSRs in loblolly pine and spruce. *Tree Genetics and Genomics* **3**, 251-259

Buhariwalla HK, Jarret RL, Jayashree B, Crouch JH, Ortiz R (2005) Isolation and characterization of microsatellite markers from *Musa balbisiana. Molecular Ecology Note* **5**, 327-330

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics* **156**, 847-854

Creste S, Benatti TR, Orsi MR, Risterucci AM, Figueira A (2006) Isolation and characterization of microsatellite loci from a commercial cultivar of *Musa acuminata. Molecular Ecology Note* **6**, 327-330

Crouch HK, Crouch JH, Jarret RL, Cregan PB, Ortiz R (1998) Segregation of microsatellite loci in haploid and diploid gametes of *Musa*. *Crop Science* **38**, 211-217

FAO Crops statistics database, http://faostat.fao.org/

Fauré S, Noyer J, Horry J, Bakry F, Lanaud C, Gonzalez D, Leon D (1993) A molecular marker-based linkage map of diploid bananas (*Musa* acuminata). *Theoretical and Applied Genetics* **87**, 517-526

Gao LF, Tang JF, Li HW, Jia JZ (2003) Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Molecular Breeding* **12**, 245-261

Gupta PK, Rustgi S (2004) Molecular markers from the transcribed/expressed region of the genome in higher plants. *Functional and Integrative Genomics* **4**, 139-162

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* **113**, 163-185

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Research* **9**, 868-77

Janssen T, Bremer K (2004) The age of major monocot groups inferred from 800 + rbcL sequences. *Botanical Journal of the Linnean Society* **146**, 385-398

Jurka J, Pethiygoda C (1995) Simple repetitive DNA sequences from primates: Compilation and analysis. *Journal of Molecular Evolution* **40**, 120-126

Kaemmer D, Fischer D, Jarret RL, Kaemmer D, Fischer D, Jarret RL, Baurens F-C, Grapin A, Dambier D, Noyer JL, Lanaud C, Kahl G, Lagoda PJL (1997) Molecular breeding in the genus *Musa*: a strong case for STMS marker technology. *Euphytica* **96**, 49-63

Kamate K, Brown S, Durand P, Bureau JM, De Nay D, Trinh TH (2001) Nuclear DNA content and base composition in 28 taxa of *Musa*. *Genome* **44**, 622-627

Kantety RV, La Rota M, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum, and wheat. *Plant Molecular Biology* **48**, 501-510

Katti MV, Ranjekar PK, Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Molecular Biology and Evolution* **18**, 1161-1167

Kumpatla SP, Mukhopadhyay S (2005) Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome* **48**, 985-998

Lagercrantz U, Ellegren H Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Research* **21**, 1111-1115

Lagoda PJL, Noyer JL, Dambier D, Baurens FC, Grapin A, Lanaud C (1998) Sequence tagged microsatellite site (STMS) markers in the *Musaceae*. *Molecular Ecology* **7**, 657-666

Lawson MJ, Zhang L (2006) Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. *Genome Biology* **7**, R14

Lysak M, Dolezelova M, Horry J, Swennen R, Dolezel J (1999) Flow cytometric analysis of nuclear DNA content in *Musa*. *Theoretical and Applied Genetics* **98**, 1344-1350

Newcomb RD, Crowhurst RN, Gleave AP, Rikkerink EHA, Allan AC, Beuning LL, Bowen JH, Gera E, Jamieson KR, Janssen BJ, Laing WA, McArtney S, Nain B, Ross GS, Snowden KC, Souleyre EJF, Walton EF, Yauk YK (2006) Analysis of expressed sequence tags from apple. *Plant Physiology* **14**, 147-166

Ortiz-Vazquez E, Kaemmer D, Zhang HB, Muth J, Rodriguez-Mendiola M, Arias-Castro C, James A (2005) Construction and characterization of a plant transformation-competent BIBAC library of the black Sigatoka-resistant banana *Musa acuminata* cv. Tuu Gia (AA). *Theoretical and Applied Genetics* **110**, 706-713

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. *Trends in Plant Science* **1**, 215-222

Rabello E, Nunes de Souza A, Saito D, Tsai SM (2005) *In silico* characterization of microsatellites in *Eucalyptus* spp.: Abundance, length variation and transposon associations. *Genetics and Molecular Biology* **28**, 582-588

Riju A, Rajesh MK, Fasila Sherin PTP, Chandrasekar A, Apshara SE, Arunachalam V (2009) Mining of expressed sequence tag libraries of cacao for microsatellite markers using five computational tools. *Journal of Genetics* **88 (2)**, 217-225

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (Eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology.* Humana Press, Totowa, NJ, pp 365-386

Ruiz M, Rouard M, Raboin LM, Lartaud M, Lagoda P, Courtois B (2004) TropGENE-DB, a multi-tropical crop information system. *Nucleic Acids Research* **32**, D364-D367

Safar J, Noa-Carrazana JC, Vrana J, Bartos J, Alkhimova O, Sabau X, Simkova H, Lheureux F, Caruana ML, Dolezel J, Piffanelli P (2004) Creation of a BAC resource to study the structure and evolution of the banana (*Musa balbisiana*) genome. *Genome* **47**, 1182-1191

Saha MC, Mian MA, Eujayl I, Zwonitzer JC, Wang L, May GD (2004) Tall fescue EST-SSR markers with transferability across several grass species. *Theoretical and Applied Genetics* **109**, 783-791

Sanderson MJ, Thorne JL, Wikström N, Bremer K (2004) Molecular evidence on plant divergence times. *American Journal of Botany* **91**, 1656-1665

Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grape ESTs. *Theoretical and Applied Genetics* **100**, 723-726

Sreenu VB, Kumar P, Nagaraju J, Nagarajaram HA (2007) Simple sequence repeats in mycobacterial genomes. *Journal of Biosciences* **32**, 3-15

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. *Genome Research* **11**, 1441-1452

Thiel T, Michalek V, Graner A (2003) Exploiting EST data-bases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* **106**, 411-422

UNCTAD (2006) http://www.unctad.org/infocomm/anglais/banana/market.htm

Varshney RK, Sigmund R, Börner A, Korzun V, Stein N, Sorrells ME, Langridge P, Graner A (2005) Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. *Plant Science* **168**, 195-202

Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002) In *silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cellular and Molecular Biology Letters* **7**, 537-46

Vilarinhos AD, Piffanelli P, Lagoda P, Thibivilliers S, Sabau X, Carreel F, D'Hont A (2003) Construction and characterization of a bacterial artificial chromosome library of banana (*Musa acuminata* Colla). *Theoretical and Applied Genetics* **106**, 1102-1106

Von Stackelberg M, Rensin SA, Reskhi R (2006) Identification of genic moss SSR markers and a comparative analysis of 24 algal and plant gene indices reveal specific rather group specific characteristics of microsatellites. *BMC Plant Biology* **6**, 9

Wang JY, Chen YY, Liu WL, Wu YT (2008) Development and application of EST-derived SSR markers for bananas (*Musa nana* Lour.). *Hereditas* **30**, 933-940