Theor Appl Genet (2009) 119:1093–1103
DOI 10.1007/s00122-009-1111-5

ORIGINAL PAPER

Development and assessment of Diversity Arrays Technology for high-throughput DNA analyses in *Musa*

Ange-Marie Risterucci · Isabelle Hippolyte · Xavier Perrier · Ling Xia · Vanessa Caig · Margaret Evers · Eric Huttner · Andrzej Kilian · Jean-Christophe Glaszmann

Received: 26 February 2009 / Accepted: 13 July 2009 / Published online: 20 August 2009

Abstract Diversity Arrays Technology (DArT) is a DNA hybridisation-based molecular marker technique that can detect simultaneously variation at numerous genomic loci without sequence information. This efficiency makes it a potential tool for a quick and powerful assessment of the structure of germplasm collections. This article demonstrates the usefulness of DArT markers for genetic diversity analyses of *Musa* spp. genotypes. We developed four complexity reduction methods to generate DArT genomic representations and we tested their performance using 48 reference *Musa* genotypes. For these four complexity reduction methods, DArT markers displayed high polymorphism information content. We selected the two methods which generated the most polymorphic genomic representations (*Pst*I/*Bst*NI 16.8%, *Pst*I/*Taq*I 16.1%) to analyze a panel of 168 *Musa* genotypes from two of the most important field collections of *Musa* in the world: Cirad (Neufchateau, Guadeloupe), and IITA (Ibadan, Nigeria). Since most edible cultivars are derived from two wild species, *Musa acuminata* (A genome) and *Musa balbisiana* (B genome), the study is restricted mostly to accessions of these two species and those derived from them. The genomic origin of the markers can help resolving the pedigree of valuable genotypes of unknown origin. A total of 836 markers were identified and used for genotyping. Ten percent of them were specific to the A genome and enabled targeting this genome portion in relatedness analysis among diverse ploidy constitutions. DArT markers revealed genetic relationships among *Musa* genotype consistent with those provided by the other markers technologies, but at a significantly higher resolution and speed and reduced cost.

Communicated by H. Nybom.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-009-1111-5) contains supplementary material, which is available to authorized users.

A.-M. Risterucci (✉) · J.-C. Glaszmann
UMR DAP, Cirad, TA A96/03, Avenue Agropolis,
34398 Montpellier, Cedex 5, France
e-mail: ange.risterucci@cirad.fr; risterucci@cirad.fr

I. Hippolyte · X. Perrier
UR 75, Cirad, TA A75/03, Avenue Agropolis,
34398 Montpellier, Cedex 5, France

L. Xia · V. Caig · M. Evers · E. Huttner · A. Kilian
Diversity Arrays Technology, PO Box 7141,
Yarralumla, ACT 2600, Australia

Introduction

Banana and plantain (*Musa* spp.) are vegetatively propagated crops of great importance to subsistence farming in Sub-Saharan Africa, South and Central America, Asia and Pacific areas. They are among the most important starchy crops in the humid tropical regions. About 87% of the production comes from small-scale farmers and is used for local consumption. Production is divided in two major groups, sweet (or dessert) bananas and cooking bananas. All are derived from intra- or inter-specific crosses between two diploid species *M. acuminata* and *M. balbisiana* (Simmonds 1995). There are two distinct genomes (A from *M. acuminata* and B from *M. balbisiana*) whose discrimination is possible through GISH (D'Hont et al. 2000).

Most of the cultivated bananas are triploid, highly sterile, and have been maintained through vegetative propagation for several centuries. The number of clones used for breeding is limited and the F1 population size, due to the high sterility, is low thus reducing accurate selection.

Noteworthy is the case of the plantains, a group with high phenotypic diversity but probably oligo or monoclonal origin (Noyer et al. 2004). Genetic diversity and phylogeny within *Musa* has been assessed using different molecular markers: intergenic spacers (Lanaud et al. 1992), restriction fragment length polymorphisms (RFLPs) (Jarret et al. 1992; Gawel et al. 1992; Carreel et al. 1994, 2002; Raboin et al. 2005), random amplified polymorphic DNA markers (RAPDs) (Howell et al. 1994; Pillay et al. 2001), inter-simple sequence repeats (ISSRs) (Godwin et al. 1997), microsatellites (SSR) (Grapin et al. 1998; Tenkouano et al. 1999; Creste et al. 2003, 2004), amplified fragment length polymorphism (AFLP) (Crouch et al. 1999; Wong et al. 2001; Ude et al. 2003), inter-retroelement amplified polymorphism (IRAP) (Nair et al. 2005; Teo et al. 2005), and methylation sensitive amplification polymorphism (MSAP) (Noyer et al. 2004). An important component of the diversity in *Musa* is known to be due to structural genome rearrangements; the *M. acuminata* complex has been divided into seven "translocation groups" following roughly a division in six subspecies (see Shepherd 1999, for a review). The current map of *Musa*, based on RFLPs, could not address this aspect (Fauré et al. 1993); having a more efficient marker system could facilitate mapping in diverse backgrounds and thus help resolve translocation polymorphisms.

Diversity Arrays Technology markers were developed to enable whole-genome profiling of crops without the need for sequence information. It is based on DNA/DNA hybridization (currently using a microarray platform) revealing the presence or absence of individual fragments in genomic representations as described by Jaccoud et al. (2001). DArT markers in barley and cassava were shown to exhibit high Polymorphism Information Content (PIC) values, high call rates and very high scoring reproducibility (Wenzl et al. 2004, Xia et al. 2005). They have been used for whole-genome profiling of *Eucalyptus grandis* (Lezar et al. 2004), wheat (Akbari et al. 2006; White et al. 2008), pigeon pea (Yang et al. 2006), *Sorghum bicolor* (Mace et al. 2008), and sugarcane (submitted). Excellent data quality was obtained also for *Arabidopsis* where SNP accounted for nearly all the DArT markers discovered (Wittenberg et al. 2005). In wheat, DArT markers were used together with other markers (SSR and AFLP) for a linkage mapping project (Semagn et al. 2006). In barley, DArT markers could be integrated with SSR, RFLP, and STS to create a high density consensus map (Wenzl et al. 2006). In both the studies the DArT markers were well distributed both within and among the linkage groups.

The objective of this study is to develop DArT markers for banana. This involved the testing of four different complexity reduction methods with a small array of candidate clones, followed by the development of two expanded arrays. These arrays were applied to the evaluation of genetic relationships between cultivated genotypes and ancestral A genome genotypes, by typing a panel of genotypes chosen between wild and cultivated diploids and cultivated triploids.

Materials and methods

Plant material and DNA extraction

Musa genotypes were selected from two of the major field germplasm collections of wild and cultivated *Musa* in the world: International Institute of Tropical Agriculture (IITA, Onne Station Port Harcourt Nigeria and Cirad; Station de Recherches Fruitières de Neufchâteau Capesterre Belle Eau, Guadeloupe, France). In most cases these accessions—which originated from different regions of the world—are also held in the ITC reference collection in Leuven.

The sampling strategy aimed to cover wild *M. acuminata*, *M. balbisiana*, and other *Musa* species genotypes as well as cultivated hybrid diploid and triploid genotypes. It made use of existing information for various types of biochemical and molecular markers, including flavonoids and isozymes (Horry and Jay 1988), ribosomal DNA (Lanaud et al. 1992), nuclear and cytoplasmic RFLPs (Carreel et al. 2002), and SSRs (Perrier et al. in press; Hippolyte, personal communication), as well as cytogenetic and morphological descriptions (Jenny et al. 2002).

A first sample of 53 accessions (listed in Table 1) was used for making the representation in order to develop the DArT arrays (see below). A broader sample was chosen for the diversity survey, aiming to cover wild *M. acuminata* and *M. balbisiana* genotypes as well as the main cultivated hybrid diploid and triploid genotypes and subgroups of related clones within these groups. The AAA triploid genotypes, for example, were chosen among Cavendish, Gros Michel, Mutika, Orotava, Ambon, Red, and Ibota subgroups; and the triploid AAB genotypes were chosen among Pome, Laknao, Silk, Iholena Plantain, and Popoulou subgroups. All these subgroups are commonly referenced in the *Musa* germplasm information system (http://www.crop-diversity.org/banana/). The DNA was extracted from fresh leaves according to Risterucci et al. (2000).

Development of genotyping arrays

Microarray preparation

DNA samples from 53 individual plants (Table 1) were processed as described by Jaccoud et al. (2001). Each sample (100 ng) was restricted in a buffer-containing

Table 1 *Musa* genotypes used for the DArT representations, selected in Cirad (Neufchateau, Guadeloupe) and IITA (Ibadan, Nigeria) field collections of wild and cultivar *Musa* genotypes

No.	MGIS code	Accession name	Ploidy	Group
1-1	ITC0582	Lady Finger	AAB	Nadan
2-1	ITC0649	Foconah	AAB	Pome
3-1	ITC0962	Prata Ana	AAB	Pome
4-1	ITC1063	Pisang Klutuk Wulung, IDN 056	BB	Balbisiana
5-1	ITC1156	Pisang Batu, IDN 080	BB	Balbisiana
6-1	ITC0623	Banksii 623	AA wild	Banksii
7-1	ITC0253	Borneo	AA wild	Microcarpa
8-1	ITC0249	Calcutta 4	AA wild	Burmanicoides
9-1	ITC0652	K. Tiparot	ABB	unknown
10-1	ITC1325	Orishele	AAB	Plantain
11-1	ITC0472	Pelipita	ABB	Pelipita
12-1	ITC0767	Dole	AAB	Bluggoe
13-1	NEU017	Grande naine	AAA	Cavendish
14-1	ITC0420	Pisang kayu	AAA	Orotava
15-1	ITC1028	Agutay	AA	Acuminata/ Errans
16-1	ITC0660	Khae phrae	AA wild	Siamea
17-1	ITC0769	Figue pomme géante	AAB	Figue pomme
18-1	ITC1138	Saba	ABB	Saba
19-1	ITC1064	Pisang Bakar	AAA	Ambon
20-1	ITC0046	Monthan	AAB	Monthan
21-1	ITC1120	Tani	BB wild	BB wild
22-1	ITC0283	Long tavoy	AA wild	Burmanica
23-1	ITC0245	Safet Velchi	AB cv	AB cv
24-1	ITC0654	Petite Naine	AAA	Cavendish
26-1	NEU016	Poyo	AAA	Cavendish
27-1	ITC0335	Popoulou	AAB	Popoulou
28-1	NEU035	Simili Radjah	ABB	Peyan
19-2	ITC0484	Gros michel	AAA	Cavendish
30-1	ITC1152	Wompa, PNG 063	AS	AS cv
31-1	ITC1034	Kunan	AB cv	AB cv
9-2	ITC0842	Pisang Kelat	AAB	AAB ind
104-2	ITC0248	Singapuri	BB wild	Balbisiana
156-2	ITC0334	Nzizi	ABB	Peyan
194-2	ITC0282	Banane serpent	AAB	Plantain
31-2	ONN190	IC 2	AAAA	AAAA unknown
32-1	ITC0312	Pisang Jari Buaya/ BS312	AA cv	Microcarpa
33-1	ITC0653	P. mas/Figue Sucrée	AA cv	Microcarpa
34-1	ITC0843	Pisang Raja Bulu	AAB	P Rajah
35-1	ITC0277	Leite	AAA	Rio
36-1	ITC0020	Ice Cream	ABB	Ney mannan
37-1	ITC1177	Zebrina	AA wild	Zebrina
38-1	ITC1187	Tomolo, (PNG023)	AA cv	AA cv
39-1	ITC0247	Honduras	BB	Balbisiana
39-2	ITC0462	Monjet	AA cv	Banksii
40-1	NEU0051	Lal Velchi	BB	Balbisiana
41-1	ITC0659	Namwa Khom, THA011	ABB	P awak
43-1	ITC0082	Intokatoke	AAA	Mutika lujigura
44-1	ITC1123	Yangambi KM5	AAA	Ibota
45-1	ITC1140	Red Yade	AAB	Plantain
46-1	ITC0243	P. Rajah	AAB	P Rajah
47-1	ITC1238	Yawa 2	ABBT	Cavendish
48-1	ITC1441	P. Ceylan	AAB	Mysore
49-1	ITC0307	Gabah-Gabah	AA cv	AA cv

10-mM Tris-OAc, 50-mM KOAc, 10-mM $Mg(OAc)_2$ 5-mM DTT with 3-U *Pst*I and one of the following frequently cutter restriction enzymes: *BstN*I, *Hae*III, *Mse*I, and *Taq*I (NEB, Beverly, USA). The restriction product was then ligated to a *Pst*I adapter (5′-GTTCAGTCAGTTAGATGGTGCA-3′ annealed with 5′-CCATCTAACTTGACTG-3′) with 2-U T4 DNA ligase (NEB). One-microliter aliquot of each 10-fold diluted reaction was used as a template in a 50-µl PCR amplification using a primer complementary to the adapter sequence. Cycling conditions were as in Jaccoud et al. (2001). A small sample (5 µl) from each amplification product was loaded on a 1.2% agarose gel for quality control and a pool of all the amplified products was assembled for each of the four genomic representations.

One library was prepared from each pool: 2 µl of the amplification product were cloned into a PCR 2.1-TOPO vector using a T/A cloning kit (Invitrogen, Carlsbad, California, USA), then transformed into *Escherichia coli* TOP 10F′ and platted on LB medium containing 50-µg/ml ampicillin. For each of the four libraries, 768 individual clones were grown in 384-well plates containing 100-µg/ml ampicillin; aliquots of the culture were used as templates to amplify inserts according to Jaccoud et al. (2001). The amplicons were dried, dissolved in 50% DMSO and spotted on polylysine-coated microarray slides (Erie Scientific, Portsmouth, NJ, USA) using a Microgrid II arrayer (Biorobotics, Cambridge, UK). After printing, slides were heated to 80°C for 2 h, incubated in 95°C water and dried by centrifugation.

Complexity reduction testing procedure and hybridization to genotyping arrays

Targets generated from each of the 48 samples with either of the four different complexity reduction methods were hybridized to matching arrays as follows:

Amplification products were precipitated in 1 volume of isopropanol, washed with 70% ethanol, dried and suspended in 2 µl of water. After 3-min denaturation at 95°C, labeling of

the fragments was carried out using the Klenow fragment of DNA polymerase I (Roche, Mannheim, Germany). Each labeling reaction contained 1× EcoPol buffer (NEB), 0.2 mM of each dNTP, random decamer mix (Roche), 0.1-mM Cy3-dUTP (Amersham, NJ, USA) and 5-U Klenow enzyme. The reaction was incubated for 3 h at 37°C. Labeled representations were mixed to 50 µl of a 50:5:1 mixture of ExpressHyb buffer (Clontech, CA, USA), 10 g l^{-1} herring sperm DNA, and the FAM labeled polylinker fragment of the TOPO vector as a reference. After denaturing, each sample was hybridized onto a microarray, covered with a glass coverslip, and incubated in a humid chamber at 65°C overnight. Then slides were washed in 1× SSC + 0.1% SDS for 5 min, in 1× SSC for 5 min, in 0.2× SSC for 2 min, and in a 0.02× SSC for 30 s. Slides were spun-dried at 200 g for 7 min and kept in a desiccator until scanning.

Genotyping of DNA samples

The two libraries with the highest level of polymorphism (*Pst*I/*Taq*I and *Pst*I/*Bst*NI) were expanded to 6,114 clones. One hundred seventy-nine *Musa* genotypes including AA, AAA, BB, AB, AAB, ABB, AS, AAS ploidy levels were analyzed using the two complexity reduction method used for library construction and hybridized on both arrays.

Polymorphism scoring

Slides were scanned using a fluorescent microarray scanner Tecan LS300 (Grödig, Salzburg, Austria). Individual slides were analyzed and the score of each marker was calculated for each sample using DArTsoft 7.4 (Diversity Arrays Technology P/L, Canberra, Australia). Markers were scored "1" for presence in genomic representation of the sample, "0" for absence, and "X" when the clustering algorithm deployed in DArTsoft was unable to score the sample with sufficient confidence. DArTsoft automatically analyzed batches of 96 slides to identify and score polymorphic markers as described by Wenzl et al. (2004). Three quality parameters were computed for each marker: the percentage of scored DNA samples (call rate), the between-cluster ("0" vs. "1") variance of the relative (denominated by the reference) target hybridization intensity as a percentage of the total relative intensity variance (*P* value) and the multivariate equivalent of the *P* parameter (*Q* value) (Storey and Tibshirani 2003). The Polymorphism Information Content (PIC) was calculated according to Anderson et al. (1993).

Data analysis

Analyses were performed using the DARwin software (http://darwin.cirad.fr/darwin) developed at Cirad. Both modalities (0,1) were given equal weight to calculate Sokal and Michener dissimilarity index d_{ij} between pairs of accessions (Sokal and Michener 1958) as the proportion of unmatching markers:

$$d_{ij} = \frac{n_{10} + n_{01}}{n_{11} + n_{10} + n_{01} + n_{00}} = \frac{n_{10} + n_{01}}{N}$$

where n_{11} is the number of fragments "present" in both *i* and *j*, n_{01} and n_{10} the number of fragments "present" in one accession but "absent" in the other, and n_{00} the number of fragments "absent" in both *i* and *j*.

As DArT markers were scored in a dominant manner, the dissimilarities between two diploids or two triploids and between a diploid and a triploid do not have the same information content. Diversity trees were built for each ploidy level using the weighted Neighbor-Joining (NJ) algorithm (Saitou and Nei 1987) that relaxes the assumption of equal mutation rate, in space and time, which underlies other clustering methods like UPGMA. One thousand bootstraps were calculated for each tree.

For bananas, the triploids are known to result from unreduced 2*n* gametes produced by sexually disturbed diploids (selection for sterility) and normal *n* gametes. The question of relations between diploids and triploids amounts to the question of the contribution of parental diploids to triploids. To answer this question, diversity trees for triploids were built under constraint of diploid tree topology (Perrier et al. 2003). This modified version of the NJ algorithm forces the exhibited structure of diversity to comply with the a priori known diversity structure of one subset of accessions, in this case the diploids. That can be seen as a projection of the triploids on the tree diversity of diploids, assigning groups of triploids to their closest diploid groups. A multidimensional principal co-ordinate analysis (PCO) was also used with specific A genome markers to elucidate relationships among genotypes.

Results

Evaluation of the complexity reduction method

The choice of an effective complexity reduction method for preparing a DArT genomic representation is a key step to reveal a large amount of polymorphism. Four restriction enzymes *Bst*NI, *Hae*III, *Mse*I, and *Taq*I in combination with *Pst*I were selected. Among the 53 genotypes used for building the representation, we choose 48 accessions to have a representation of all ploidy levels and sub-groups, these accessions were used to test the frequency of polymorphic clones on the arrays containing 3,072 clones, 768 derived from each of the four reduction methods. The *Pst*I/*Bst*NI library showed the highest level of polymorphism

(16.8%) against these 768 markers, followed by the *Pst*I/*Taq*I at 16.1% (Table 2). The *Pst*I/*BstN*I complexity reduction method was previously found to generate the most polymorphic representations in barley (Wenzl et al. 2004), while *Pst*I/*Taq*I was most effective in wheat (Akbari et al. 2006). Data analyses (not shown) on these 48 samples showed a clear differentiation between A and B genomes for the four complexity reduction methods.

Development and performance of the two full-size arrays

We, then, extended two libraries to 6144 DArT clones using the best two complexity reduction methods, based on restriction enzymes *Pst*I as a primary cutter and *Taq*I or *BstN*I as a secondary cutter. For each representation, two runs of hybridization including, respectively, 96 and 83 different banana genotypes were performed. For a more efficient polymorphism discovery, the design for each of the two 96-well plates of samples included randomly distributed genotypes representing a full spectrum of diversity.

Between 500 and 700 polymorphic markers per plate and hybridization run were identified in each of these libraries. For *Pst*I/*BstN*I, 463 common markers were identified as polymorphic into the two plates. For *Pst*I/*Taq*I, 373 common polymorphic markers were found.

Two representation comparisons

These data sets were used to perform different types of analyses. First, the results obtained using the two representations including all genotypes and ploidy levels were compared. Then, diploid and triploid samples were analyzed separately. Ultimately, the relationships between genotypes and their origin were analyzed. The overall structures obtained with the two representations (not shown), including all types of genotypes and ploidy levels were very similar and showed that the majority of the markers separated the A and B genomes. The 53 genotypes used for building the array were distributed in all the components of the global structure. At the individual level, results for the two representations were also very close. Those results validate the use of a joined set of all the 836 non-redundant markers obtained from the two complexity reduction methods for diversity studies and relationship between genotypes. Within the total marker set, we could detect 86 "A markers" (absent from all *M. balbisiana* genotypes). Finally, these sets of 836 and 86 markers were used for analyses on 179 *Musa* genotypes, between them some tetraploid clones were not included in the analyses. The frequency of marker presence showed a normal distribution, with the 10 most frequent markers present in 144–152 of the 179 genotypes and the 10 least frequent markers present in only 8–15 genotypes.

Variation among the primary diploids

The analysis of variation among the primary diploids AA and BB, including two AB, four AS (hybrid from *M. acuminata* and *M. schizocarpa*) and eight other *Musa* species was refined using an NJ tree (Fig. 1). This revealed the differentiation between *M. acuminata* and *M. balbisiana* genotypes, with a strong opposition between AA and BB genotypes. Two AB cultivars were found in an intermediate position. Other *Musa* species, which had not been represented in the array constitution, appeared scattered along the A–B differentiation. Within the *M. acuminata* pool, two separate groups were found. The first AA group included mostly wild genotypes (*malaccensis*, *zebrina*, *siamea*, and *burmanica*) and only three cultivars from Indonesia; whereas the second one included cultivars, some wild *M. acuminata* subspecies, namely *banksii*, *errans*, and *microcarpa* as well as three AS genotypes from *M. schizocarpa*. Within this second cluster, a homogeneous subcluster could be identified, which included only cultivars with the exception of one supposedly AS genotype. Genotypes from the *microcarpa* group, defined mainly on the basis of seed size, were found in two clusters with wild *microcarpa* (Microcarpa holotype and Borneo) were found very close to some wild and landrace *banksii*; landrace *microcarpa* (Datil, Kirun, Colatina Ouro) were found near some hybrid landrace *banksii* genotypes.

Table 2 Number of polymorphic clones and their average quality features for four different complexity reduction methods tested on 48 banana samples hybridized with 768 clones for each method, spotted on the same slides

Complexity reduction method	Number of polymorphic clones	Frequency of polymorphic clones (%)	Average call rate	Average *P* quality	Average PIC value
*Pst*I/*Taq*I	124	16.1	91.6	81.4	0.43
*Pst*I/*Mse*I	113	14.7	92.2	83.3	0.45
*Pst*I/*Hae*III	92	11.9	92.4	83.0	0.42
*Pst*I/*BstN*I	129	16.7	92.6	83.6	0.43

The polymorphism is detected by DArTsoft program using standard settings

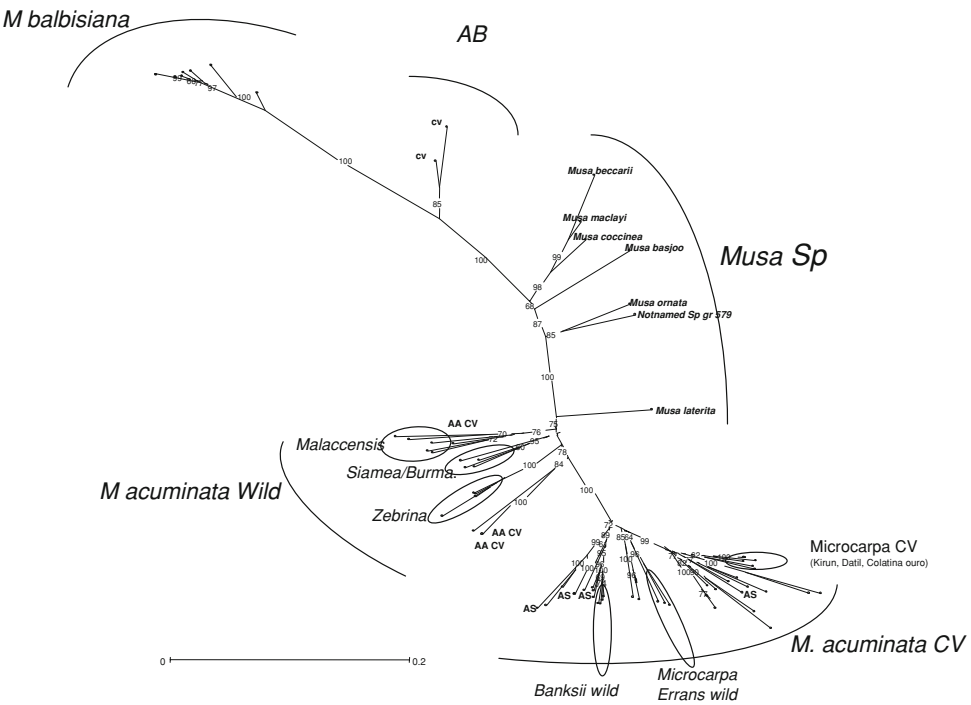

Fig. 1 Neighbor-Joining tree of 74 diploids based on Sokal and Michener dissimilarity index calculated with 836 DArT markers. Bootstrap support value (1,000 replicates) are shown if greater than 60%

Variation among the triploid forms

Variation among the triploid forms was analyzed using an NJ tree under constraint of the diploid tree (Fig. 2) and a PCO using only A genome markers (Fig. 3). The main components of structure in Fig. 2 were determined by the genome composition, with three main triploid clusters corresponding to AAA, AAB, or ABB. The outliers highlighted earlier were found again. It was also possible to separate clusters corresponding to known morphotaxonomic subgroups. As expected, within the AAA genotypes, Cavendish, Gros Michel, Ibota, Red, Orotava, Ambon and East African were clearly separated. Similarly the variation illustrated the differentiation between the Plantain, Popoulou, and Silk (Figue Pomme) groups among the AAB forms, and between the Ney Mannan, Bluggoe, and Pisang Awak groups among the ABB forms. The variation was close to nil in several groups often considered as clones with only somatic variation, such as the Cavendish group or the Plantain group.

The relative distribution of the diploids and triploids highlighted the similarity between Mlali Comorian genotypes and the Cavendish and Gros Michel groups, as well as between Zebrina wild diploids and the Mutika AAA group from East Africa and between the only two AB genotypes and the Silk AAB group.

The large number of DArT markers enables analysis of particular genotype sub groups. To remove the major effect of *M. acuminata* and *M. balbisiana* allele origin, we performed an analysis with the 86 "A" markers only. These were particularly useful to analyze the distribution of the A genome diversity in all forms, including secondary triploids and hybrids (Fig. 3). The PCO axes accounted for 40.6% for axis 1 and 10.9% for axis 2 of the total variation. Three clear clusters were found. The first one included the wild AA genotypes Malaccensis, Zebrina, Siamea, Burmanica, and the only Truncata genotype included in this study, the polyploids of the AAA Mutika group from East African, the AAB Silk group, the two AB genotypes, and all the ABB genotypes. The second cluster included wild and landrace AA *banksii*, AAB Plantain, Popoulou, and Laknao groups, *M. schizocarpa* AS genotypes, and the single AAS genotype included in this study. The third cluster included AA genotypes the Mlali Comorian group and some hybrid genotypes from the *banksii* and *microcarpa* groups, among the AAA genotypes the Cavendish, Gros Michel, Ibota, Red, and Orotava groups, and among the AAB genotypes the Pome, Misore, and Pisang Rajah groups.

Discussion

Our results demonstrate that DArT markers are of good quality and perform well on banana. The four complexity reduction methods initially tested and the two methods used in this study gave a high polymorphism rate with a good marker quality as revealed by their high PIC values, call rates and *P* values.

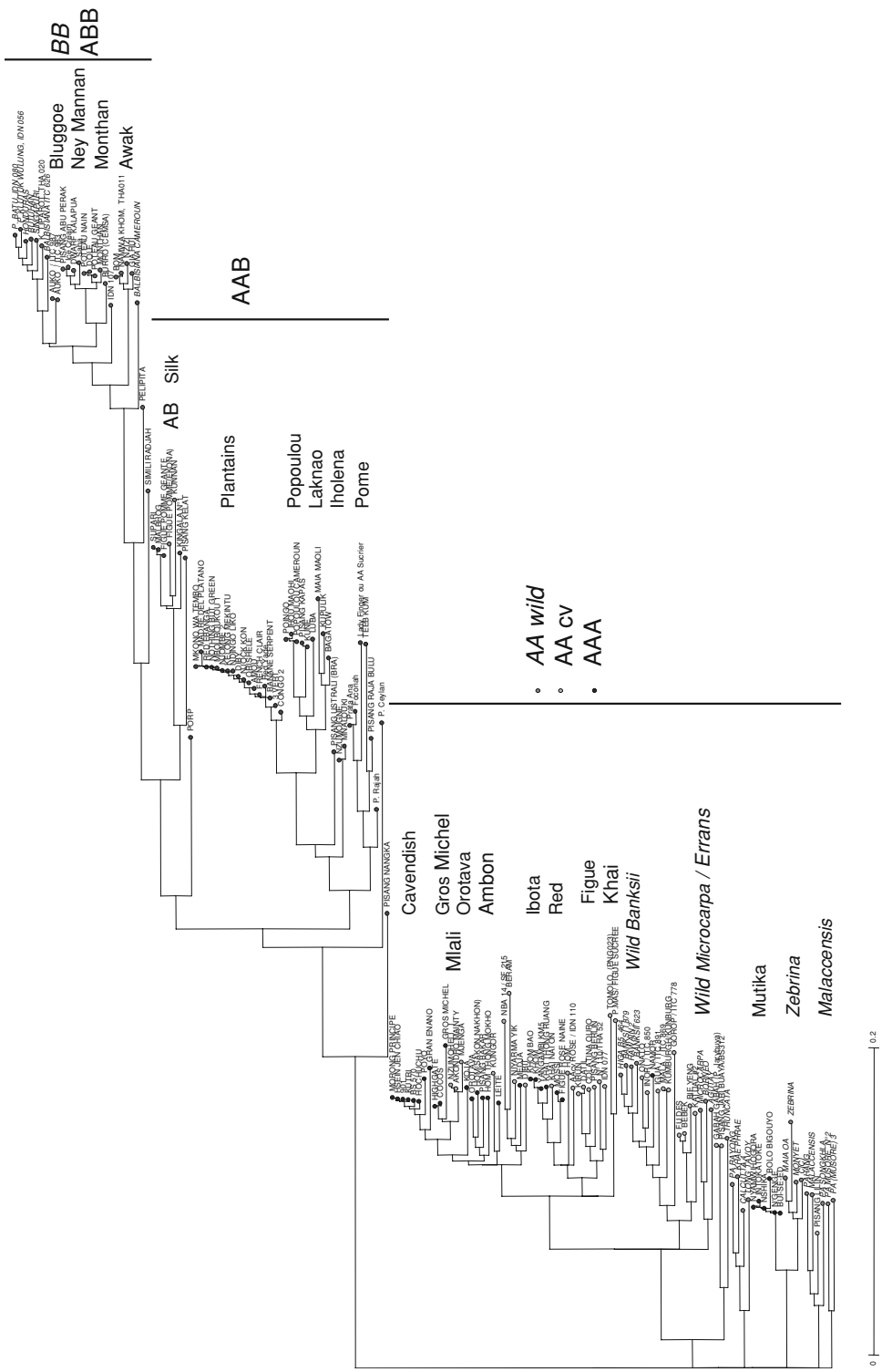

Fig. 2 Neighbor-Joining tree of 93 triploids under constraint of the AA, AB, and BB diploid genotypes tree, based on Sokal and Michener dissimilarity index calculated with 836 DArT markers

This step has benefitted from a choice of 53 founder reference *Musa* genotypes which made use of several rounds of diversity studies conducted with various marker types as well as cytogenetic and morphological descriptions.

The DArT markers provide important information regarding the genetic relationships among a large diversity level ranging from section to subspecies.

On all diploid genotypes a clear differentiation was found between A and B genome. This result demonstrates

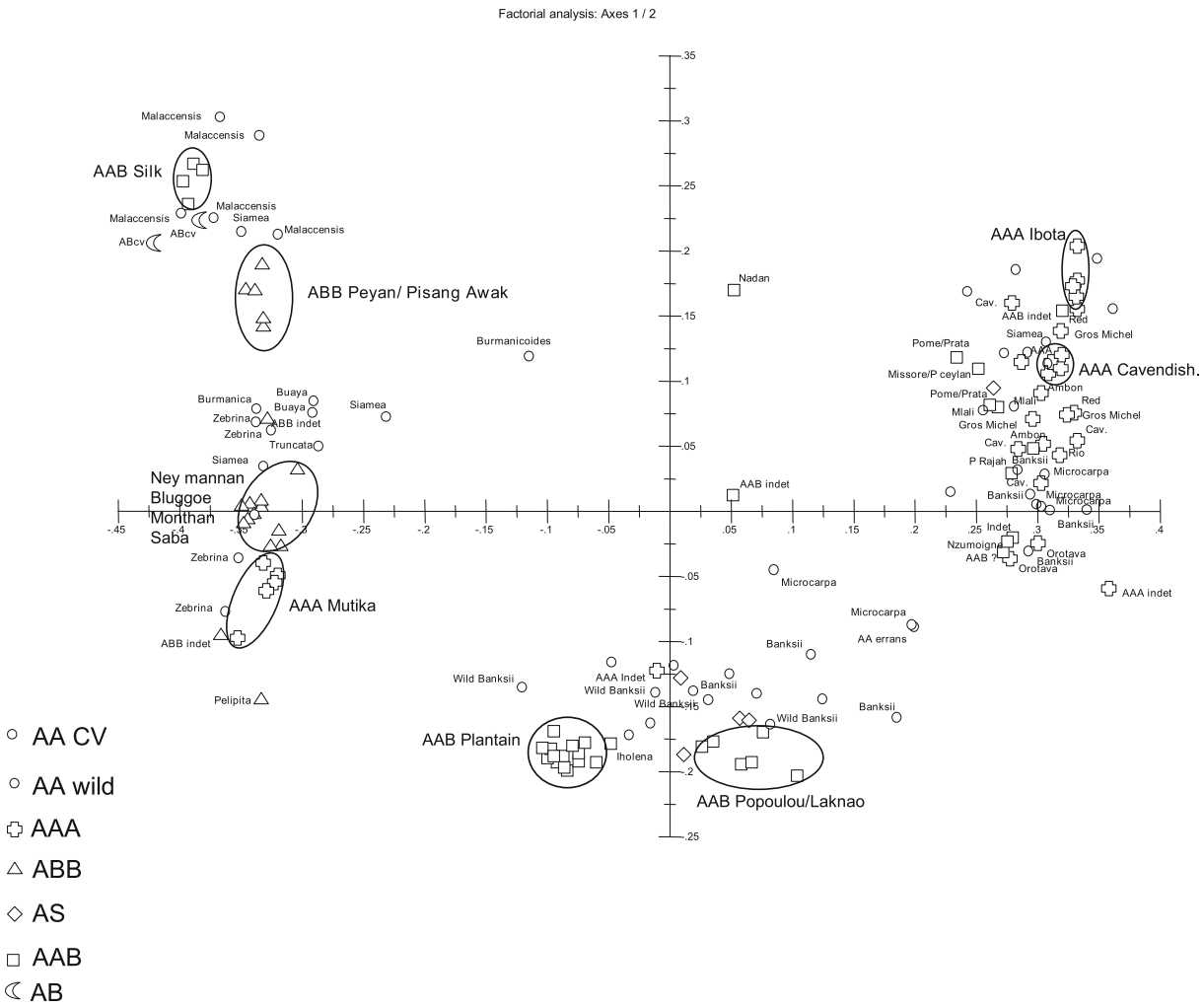

Fig. 3 Principal Co-ordinate Analysis of the variation for 153 genotypes using 86 markers totally absent from the *M. balbisiana* species

the DArT markers capability to label ancient speciation events. Divergence between *M. acuminata* and *M. balbisiana* was estimated to go back to 4 million years (Lescot et al. 2008). At the section level, genotypes from section with $n = 10$ such as *Australimusa* (*M. beccarii*, *M. maclayi*) and *Calimusa* (*M. coccinea*) were separated from genotypes from sections with $n = 11$ such as *Eumusa* (*M. basjoo*, *M. acuminata*, and *M. balbisiana*) and *Rhodochlamys* (*M. laterita*, *M. ornata*). *M. laterita* genotype from section *Rodochlamys* were closer to *M. acuminata* than to other *Eumusa*. Such results are in agreement with those found by Wong et al. (2002) using AFLP analysis and with the observation of Simmonds (1962) that *M. laterita* was closely related to *M. acuminata.*

Within the subspecies *M. acuminata*, a clear distinction was observed between the subspecies *M. a. banksii* from Papua New Guinea, subspecies *M. a. zebrina* and *M. a. burmanica* from Indonesia islands, and *M. a. malaccensis* from the northern part of the *M. acuminata* area distribution. Among the AA cultivars derived from wild *M. acuminata* the Papua New Guinea genotypes were closely related to wild *M. a. banksii*. Our results confirm the hypothesis of Carreel et al. (2002) on a contribution of the subspecies *M. a. banksii* to other AA cultivars from Indonesia, Malaysia and the Mlali group from East African islands found in the same subcluster. Among AA wild genotypes our results confirm those found by Grapin et al. (1998) by microsatellite analysis, notably for Agutay the only *M a. errans* genotype clearly found near the two wild *M. a. microcarpa* genotypes Borneo and Microcarpa holotype and the four wild *M. a. banksii* genotypes.

DArT markers resolve relationship between wild and triploid genotypes.

Our results reveal a very strong structure among triploid genotypes, as summarized in the NJ phenetic tree obtained using the 836 DArT markers set, which relates to the genomic constitutions (AAA, AAB, and ABB). Atypical ABB accessions Pelipita and Simili Rajah fall between ABB and AAB forms; these accessions may be aneuploids, as was identified using GISH for Pelipita, which has 8 A

chromosomes and 25 B chromosomes (D'Hont et al. 2000). A similar situation may apply to Pisang Nangka, which falls between AAB and pure A forms.

At a finer scale, the results clearly revealed that the triploid Cavendish and Gros Michel AAA groups had the Mlali Comorian diploid AA genotypes in their ancestry, corroborating the RFLP results of Raboin et al. (2005). The DArT data also revealed a low but significant diversity in supposedly clonal triploid AAA groups such as Cavendish or Gros Michel. Creste et al. (2003) found a total genetic uniformity of Cavendish varieties using microsatellite markers analyses. The higher number of DArT markers and the possible differentiation due to methylation variation, made possible by the use of the *Pst*I restriction enzyme, may contribute to explain this intra-clonal variation. However, divergence recently found within the chloroplastic genome of Cavendish types by Boonruangrod (personal communication) provides other evidence of genetic diversity within this group and invite further phylogenetic analyses at a finer scale in large successful groups of cultivated forms.

Mutika East African AAA genotypes showed very low intra-group polymorphism, consistent with results obtained by Pillay et al. (2001) using RAPD markers. They were clearly related to wild AA *zebrina* genotypes, consistent with Jenny et al. (2002).

The analysis involving "A" markers only underlined the wild *M. a. banksii* contribution to some AAB genotypes, to African Plantains, to Popoulou (also named Pacific Plantain) as well as to the Laknao group from the Philippines. Limited diversity was found within the plantain group. This result confirmed the very narrow genetic base of this cultivar group, as found by Noyer et al. (2004) through SSR, AFLP, and MSAP markers analyses. Our analysis with the "A" markers set also highlighted evidence of the *M. a. banksii* origin of the A genome in the Plantain oligoclonal group (Carreel et al. 2002), even though only two wild *M. a. banksii* genotypes have been used for representation building.

Finally, at the level of individual samples, DArT data corroborate suspected mislabeling in the germplasm collections; Pa Songkhla is a *malaccensis* genotype and not a *siamea*, and Cici is a *zebrina* genotype and not a *malaccensis* one.

Our results also highlight precautions to be taken when using DArT markers for diversity analysis. The method used for identification of polymorphic marker tends to select against unbalanced frequencies of presence versus absence because they are associated with lower statistical power. Yet such markers can be very important for discriminating minority groups. This is exemplified by the subspecies *M. a. malaccensis* and its contribution to Silk AAB cultivars of Indian origin as well as to AB genotypes close to the Silk group, whose cytoplasmic marker pattern is the same as the Silk group (Carreel et al. 2002). Therefore, it is important to adapt the marker screening so that rare markers are also allowed to be taken into account.

The representativeness of the genotypes selected for the initial representations is obviously very important; it determines the axes of variation that the resulting array will be able to help monitor. This can induce apparent bias in the resolution of the parental origin of hybrid forms. For example, Mutika AAA genotypes were identified as hybrids between a wild *M. a. zebrina* from Java and a wild *M. a. banksii* from Papua New Guinea using cytoplasmic markers (Carreel et al. 2002). SSR nuclear markers show Mutika genotypes at an intermediate position between *zebrina* and *banksii* (Perrier et al. in press); DArT markers place them clearly closer to *zebrina* than to *banksii* (Fig. 3). This trend to emphasize one of the two parental origins can be useful to turn into an advantage for the confirmation of suspected contribution; it is the case for the subspecies *M. a. malaccensis* contribution to AAB Peyan and ABB Pisang Awak genotypes, which was suspected using microsatellite markers and is clearly confirmed by our DArT data.

Altogether, the current DArT resources permit to conduct targeted genetic studies on *Musa* germplasm with speed and efficiency. They are already used through the Generation Challenge Programme as part of its Genotyping Support Service. The first illustration is given by Amorim et al. (2009), who used DArTs to analyze the distribution of banana accessions with high carotenoid content in the fruit among a representation of the *Musa* germplasm collection in Brazil, to optimize hybridization schemes for varietal improvement of fruit nutritional quality.

Conclusion

In this study, we have shown that DArT can be used for genome-wide analysis and establishment of relationships between *Musa* genotypes. The power of this fingerprinting method lies in its ability to compare different genomes at a large number of loci in a single assay. The DArT markers allow the identification of genomic regions shared between related genotypes; they effectively complement other molecular marker technologies for genetic diversity studies, genomics, and breeding knowledge. The analyses cluster genotypes consistently with the accepted classification knowledge. Arrays can enable the resolution of diversity within each genome and the study of the structure in *Musa* complex. Furthermore, the method may differentiate between vegetatively propagated triploids, such as Cavendish. In this group, we identified some differences between accessions not observed with other markers. The

sequence of the polymorphic DArT clones may explain the differences in DArT marker behavior in this group of very limited diversity.

The current set of markers will be applied to improve the genetic map of banana. The easy access to DArT marker sequences gave a great capacity to integrate diversity information with genetic and physical maps and ultimately genome sequence assembly.

Acknowledgment We thank the Generation Challenge Programme for financial support.

References

Akbari M, Wenzl P, Vanessa C, Carling J, Xia L, Yang S, Uszynski G, Mohler V, Lehmensiek A, Kuchel H, Hayden MJ, Howes N, Sharp P, Rathmell B, Vaughan P, Huttner E, Kilian A (2006) Diversity Arrays Technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor Appl Genet 113:1409–1420

Amorim EP, Vilarinhos AD, Cohen KO, Amorim VBO, Santos-Serejo JA, Silva SO, Pestana KN, Santos VJ, Paes NS, Monte DC, Rei RV (2009) Genetic diversity of carotenoid-rich bananas evaluated by Diversity Arrays Technology (DArT). Genet Mol Biol 32:96–103

Anderson JA, Churchill GA, Autrique JE, Tanskley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. Genome 396:181–186

Carreel F, Fauré S, Gonzalez de Leon D, Lagoda PLJ, Perrier X, Bakry F, Tezenas du Montcel H, Lanaud C, Horry JP (1994) Evaluation de la diversité génétique chez les bananiers diploïdes (*Musa* sp). Genet Sel Evol 26(Suppl):125s–136s

Carreel F, Gonzalez de Leon D, Lagoda P, Lanaud C, Jenny C, Horry J, Tezenas du Montcel H (2002) Ascertaining maternal and paternal lineage within Musa by chloroplast and mitochondrial DNA RFLP analyses. Genome 45:679–692

Creste S, Tulmann Neto A, de Oliveira Silva S, Figueira A (2003) Genetic characterization of banana cultivars (*Musa* spp.) from Brazil using microsatellite markers. Euphytica 132:259 258

Creste S, Tulmann Neto A, Vencovsky R, de Oliveira Silva S, Figueira A (2004) Genetic diversity of Musa diploid and triploid accessions from the Brazilian banana breeding program estimated by microsatellite markers. Genet Res Crop Evol 51:723–733

Crouch JH, Crouch HK, Constandt H, Van Gysel A, Breyne P, Van Montagu M, Jarret RL, Ortiz R (1999) Comparison of PCR-based molecular marker analyses of Musa breeding populations. Mol Breeding 5:233–244

D'Hont A, Paget-Goy A, Escoute J, Carreel F (2000) The interspecific genome structure of cultivated banana, Musa spp. revealed by genomic DNA in situ hybridization. Theor Appl Genet 100:177–183

Fauré S, Noyer JL, Horry JP, Bakry F, Lanaud C, Gońzalez de León D (1993) A molecular marker-based linkage map of diploid bananas (*Musa acuminata*). Theor Appl Genet 87:517–526

Gawel NJ, Jarret RL, Whittemore AP (1992) Restriction fragment length polymorphism (RFLP)-based phylogenetic analysis of *Musa*. Theor Appl Genet 84:286–290

Godwin ID, Aitken EAB, Smith LW (1997) Application of inter simple sequence repeat (ISSR) markers to plant genetics. Electrophoresis 18:1524–1528

Grapin A, Noyer JL, Carreel F, Dambler D, Baurens FC, Lanaud C, Lagoda PJL (1998) Diploid *Musa acuminata* genetic diversity assayed with sequence tagged microsatellite sites. Electrophoresis 19:1374–1380

Horry J, Jay M (1988) Distribution of anthocyanins in wild and cultivated banana varieties. Phytochemistry 27:2667–2672

Howell EC, Newbury HJ, Swennen RL, Withers LA, Ford-Lloyd BV (1994) The use of RAPD for identifying and classifying *Musa* germplasm. Genome 37:328–332

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity Arrays: a solid-state technology for sequence information independent genotyping. Nucleic Acids Res 29:e25

Jarret RL, Gawel NJ, Whittemore A, Sharrock S (1992) RFLP-based phylogeny of *Musa* species in Papua New Guinea. Theor Appl Genet 84:579–584

Jenny C, Carreel F, Kodjo T, Perrier X, Dubois C, Horry JP, Tézenas du Montcel H (2002) Banana. In: Hamon P, Seguin M, Perrier X, Glaszmann JC (eds) Genetic diversity of cultivated tropical plants. Science Publishers Inc USA and CIRAD France, pp 99–124

Lanaud C, Tezenas du Montcel H, Jolivot MP, Glaszmann JC, Gonzalez De Leon D (1992) Variation of ribosomal gene spacer length among wild and cultivated banana. Heredity 68:147–156

Lescot M, Piffanelli P, Ciampi AY, Ruiz M, Blanc G, Leebens-Mack J, da Silva FR, Santos CMR, D'Hont A, Garsmeur O, Vilarinhos AD, Kanamori H, Matsumoto T, Ronning CM, Cheung F, Haas BJ, Althoff R, Arbogast T, Hine E, Pappas GJ, Sasaki T, Souza MT, Miller RG, Glaszmann JC, Town CD (2008) Insights into the *Musa* genome: syntenic relationships to rice and between Musa species. BMC Genomics 9:58

Lezar S, Myburg AA, Berger DK, Wingfield MJ, Wingfield BD (2004) Assessment of microarray-based DNA fingerprinting in *Eucalyptus grandis*. Theor Appl Genet 109:1329–1336

Mace ES, Xia L, Jordan DR, Halloran K, Parh DK, Huttner E, Wenzl P, Kilian A (2008) DArT markers: diversity analyses and mapping in *Sorghum bicolor*. BMC Genomics 9:26

Nair AS, Teo CH, Schwarzacher T, Heslop-Harisson JS (2005) Genome classification of banana cultivars from South India using IRAP markers. Euphytica 114:285–290

Noyer JL, Causse S, Tomekpe K, Bouet A, Baurens FC (2004) A new image of plantain diversity assessed by SSR, AFLP and MSAP markers. Genetica 124:61–69

Perrier X, Flori A, Bonnot F (2003) Data analysis methods. In: Hamon P, Seguin M, Perrier X, Glaszmann JC (eds) Genetic diversity of cultivated tropical plants. Science Publishers Inc USA and CIRAD France, pp 43–76

Perrier X, Bakry F, Carreel F, Jenny C, Horry JP, Lebot V, Hippolyte I (in press) Combining biological approaches to highlight the evolution towards edible bananas Sixth World Archaeological Congress Dublin. Ethnobotany Research and Application

Pillay M, Ogundiwin E, Nwakanma DC, Ude G, Tenkouano A (2001) Analysis of genetic diversity and relationships in East African banana germplasm. Theor Appl Genet 102:965–970

Raboin LM, Carreel F, Noyer JL, Baurens FC, Horry JP, Bakry F, Tezenas du Montcel H, Ganry J, Lanaud C, Lagoda PJL (2005) Diploid ancestors of triploid export banana cultivars: molecular identification of 2n restitution gamete donors and n gamete donors. Mol Breed 16:333–341

Risterucci AM, Grivet L, N'Goran JAK, Pieretti I, Flament MH, Lanaud C (2000) A high density linkage map of *Theobroma cacao* L. Theor Appl Genet 101:948–955

Saitou N, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Semagn K, Bjornstad A, Skinnes H, Maroy AG, Tarkegne Y, William M (2006) Distribution of DArT, AFLP, and SSR markers in a genetic linkage map of a doubled-haploid hexaploid wheat population. Genome 49:545–555

Shepherd K (1999) Cytogenetics of the genus Musa. Ipgri-Inibap, Rome

Simmonds NW (1962) The evolution of the bananas. Longmans, London

Simmonds NW (1995) Bananas *Musa* (Musaceae). In: Smartt J, Simmonds NW (eds) Evolution of crop plants. Longman Scientific and Technical, London, pp 370–375

Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. Univ Kansas Sci Bull 28:1409–1438

Storey J, Tibshirani R (2003) Statistical significance for genomewide studies. Proc Nat Acad Sci USA 100:9440–9945

Tenkouano A, Crouch JH, Crouch HK, Vuylsteke D, Ortiz R (1999) Comparison of DNA marker and pedigree-based methods of genetic analysis of plantain and banana (Musa spp.) clones. I. Estimation of genetic relationships. Theor Appl Genet 98:62–68

Teo CH, Tan SH, Ho CH, Faridah QZ, Othman YR, Heslop-Harrison JS, Kalendar R, Schulman AH (2005) Genome constitution and classification using retrotransposon based markers in the orphan crop banana. J Plant Biol 48:96–105

Ude G, Pillay M, Ogundiwin E, Tenkouano A (2003) Genetic diversity in an African plantain core collection using AFLP and RAPD markers. Theor Appl Genet 107:248–255

Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A, Kilian A (2004) Diversity Arrays Technology (DArT) for whole-genome profiling of barley. Proc Nat Acad Sci USA 101:9915–9920

Wenzl P, Li H, Carling J, Zhou M, Raman H, Paul E, Hearnden P, Maier C, Xia L, Caig V, Ovesna J, Cakir M, Poulsen D, Wang J, Raman R, Smith KP, Muehlbauer GJ, Chalmers KJ, Kleinhofs A, Huttner E, Kilian A (2006) A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and phenotypic traits. BMC Genomics 7:206

White J, Law JR, Mackay I, Chalmers KJ, Smith JSC, Kilian A, Powell W (2008) The genetic diversity of UK, US and Australian cultivars of *Triticum aestivum* measured by DArT markers and considered by genome. Theo Appl Genet 116:439–453

Wittenberg A, van der Lee T, Cayla C, Kilian A, Visser R, Schouten H (2005) Quality assessment of the Diversity Array Technology (DArT) using the model plant *Arabidopsis thaliana*. Mol Gen Genomics 274:30–39

Wong C, Kiew R, Loh JP, Gan LH, Lee SK, Ohn S, Lum S, Gan YY (2001) Genetic diversity of the wild banana *Musa acuminata* Colla in Malaysia as evidenced by AFLP. Ann Bot 88:1017–1025

Wong C, Kiew R, Argent G, Set O, Lee SK, Ohn S, Gan YY (2002) Assessment of the validity of section in *Musa* (*Musaceae*) using AFLP. Ann Bot 90:231–238

Xia L, Peng K, Yang S, Wenzl P, de Vicente MC, Fregene M, Kilian A (2005) DArT for high-throughput genotyping of cassava (*Manihot esculenta*) and its wild relatives. Theor Appl Genet 110:1092–1098

Yang S, Pang W, Ash G, Harper J, Carling J, Wenzl P, Huttner E, Zong X, Kilian A (2006) Low level of genetic diversity in cultivated Pigeonpea compared to its relatives is revealed by diversity arrays technology. Theor Appl Genet 113:585–595