A platform for efficient genotyping in *Musa* using microsatellite markers

Pavla Christelová1, Miroslav Valárik1, Eva Hřibová1, Ines Van den houwe2, Stéphanie Channelière3, Nicolas Roux3 and Jaroslav Doležel1*

1 Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Sokolovská 6, Olomouc CZ-77200, Czech Republic
2 Laboratory of Tropical Crop Improvement, Katholieke Universiteit Leuven, Kasteelpark Arenberg 13, B-3001 Leuven, Belgium
3 Bioversity International, Parc Scientifique Agropolis II, Montpellier Cedex 5, 34397, France

Received: 17 March 2011; Returned for revision: 14 May 2011; Accepted: 16 August 2011; Published: 22 August 2011


Abstract

Background and aims

Bananas and plantains (*Musa* spp.) are one of the major fruit crops worldwide with acknowledged importance as a staple food for millions of people. The rich genetic diversity of this crop is, however, endangered by diseases, adverse environmental conditions and changed farming practices, and the need for its characterization and preservation is urgent. With the aim of providing a simple and robust approach for molecular characterization of *Musa* species, we developed an optimized genotyping platform using 19 published simple sequence repeat markers.

Methodology

The genotyping system is based on 19 microsatellite loci, which are scored using fluorescently labelled primers and high-throughput capillary electrophoresis separation with high resolution. This genotyping platform was tested and optimized on a set of 70 diploid and 38 triploid banana accessions.

Principal results

The marker set used in this study provided enough polymorphism to discriminate between individual species, subspecies and subgroups of all accessions of *Musa*. Likewise, the capability of identifying duplicate samples was confirmed. Based on the results of a blind test, the genotyping system was confirmed to be suitable for characterization of unknown accessions.

Conclusions

Here we report on the first complex and standardized platform for molecular characterization of *Musa* germplasm that is ready to use for the wider *Musa* research and breeding community. We believe that this genotyping system offers a versatile tool that can accommodate all possible requirements for characterizing *Musa* diversity, and is economical for samples ranging from one to many accessions.

* Corresponding author’s e-mail address: dolezel@ueb.cas.cz

Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Introduction

The important role of bananas and plantains (Musa spp.) as one of the top world trade commodities and as food security for millions of people, especially in humid tropics, is unquestionable. However, this crop faces serious endangerment by numerous pests and diseases. Breeding efforts are hampered by a high degree of banana sterility and a lack of characterized germplasm as potential parents for breeding. Currently grown banana cultivars are mainly triploid clones, which originated as intraspecific hybrids of Musa acuminata and interspecific hybrids between M. acuminata and Musa balbisiana, with a possible involvement of a few other species within the genus. To set up an efficient strategy for breeding improved banana varieties and support the choice of crossing parents, a solid understanding of the genetic diversity of available resources is needed. Likewise, conservation of existing gene resources is essential, especially when we observe the continuous loss of banana diversity due to indelicate environmental treatment of the rain forests, as well as changed farming practices of smallholders. The main objectives and means for Musa diversity conservation were formulated in the Global Conservation Strategy for Musa (INIBAP 2006) under the scope of GMGC (Global Musa Genomics Consortium). Nevertheless, irrespective of the selected strategy, efficient collection and preservation of banana diversity highly depend on unambiguous sample identification. To avoid problems of duplicates within national, regional and global germplasm collections, an accurate and standardized characterization of newly introduced accessions as well as those already deposited in gene banks would be of great benefit. This rationalization effort will allow Musa accessions to be efficiently conserved.

Traditional classification of Musa species is based on morphological characters and chromosome counts (basic chromosome number; x) (Cheesman 1947; Simmonds and Shepherd 1955). Although a morphotaxonomic system allows for differentiation of specific banana clones (Stover and Simmonds 1987), insufficiencies of this approach start to emerge as the genetic basis of the plants under study gets narrow. Additionally, a small change at the DNA level can cause a large phenotypic manifestation, while sometimes no or minor morphological changes can be observed after extensive genetic changes. Obviously, a classification system that relies exclusively on the phenotypic manifestations of the genome suffers from limited accuracy (Crouch et al. 2000; De Langhe et al. 2005), but can be made robust if supported by molecular-based characterization.

The enormous increase in the availability of various molecular techniques over the past decades has facilitated the classification of new banana cultivars, as well as reassessment of the traditional taxonomy. Among the broad portfolio of molecular tools, some of the markers have gained special attention in terms of their use in diversity studies and molecular characterization of banana genotypes. Most recently, diversity arrays technology was used for the assessment of genetic diversity within Musa spp. (Risterucci et al. 2009). While having the advantage of a high-throughput approach suitable for large numbers of genotypes, its use for a limited number of samples in a short turn-around time would rank it within the more demanding methods in terms of funding support. The same applies to the genotyping by sequencing approach, which has gained special attention recently (Elshire et al. 2011). Other molecular markers applied in Musa diversity studies were RAPDs (random amplified polymorphic DNA; Pillay et al. 2000, 2001; Ruangsuttapha et al. 2007; Venkatachalam et al. 2008) and AFLPs (amplified fragment length polymorphisms; Loh et al. 2000; Wong et al. 2001a; Ude et al. 2002; Wang et al. 2007). Both these markers have a relatively high level of polymorphism, but they are dominant and, in the case of RAPDs, their reproducibility is a serious limitation (Jones et al. 1997). The more advantageous co-dominant markers were also used for Musa, such as RFLPs (restriction fragment length polymorphisms; Gawel et al. 1999; Nwakanma et al. 2003; Ning et al. 2007) and SSRs (simple sequence repeats; e.g. Kaemmer et al. 1997; Grapin et al. 1998; Lagoda et al. 1998; Buhariwalla et al. 2005). While RFLPs perform well in terms of reproducibility, they have a relatively low level of polymorphism and are difficult to use. On the contrary, SSR markers outperform the RFLPs and RAPDs in all the above-mentioned aspects.

Microsatellites (SSRs) are stretches of simple 1- to 6-base-pair-long repeat motifs arranged tandemly within the genomes of prokaryotic and eukaryotic organisms. Their flanking regions, which are usually highly conserved, are suitable for designing locus-specific primers. Simple sequence repeats have been successfully applied in the molecular genotyping of many important crops such as rice (Pessoa-Filho et al. 2007), cereals (Hayden et al. 2007), grapevine (This et al. 2004) or cacao (Zhang et al. 2006). Moreover, the use of SSR markers opens up the possibility of automation and multiplexing, which significantly increases the throughput of the technique.

With the aim of developing a standardized protocol to classify Musa germplasm, we have tested and optimized the use of 22 published SSR markers on a set of banana...
Materials and methods
Plant material and the reference DNA collection
The reference DNA collection, comprising a total of 65 accessions [Additional Information 1], was established to represent genetic diversity within the genus Musa. In vitro plantlets of these accessions are available for distribution from the Bioversity ITC. The genomic DNA of 61 of the 65 accessions is stored in the Genome Resources Centre (http://www.musagenomics.s.org/cetest_firstpage1/genomic_dna.html) and is available for distribution. Out of the 65 accessions, 54 were successfully included in the analysis [Additional Information 1]. To extend the diploid representation of the genotype set, 39 additional diploid accessions were included [Additional Information 2], with three of them being duplicate samples to the Reference DNA collection. These duplicates were included intentionally to test the capability of the genotyping platform to identify sample duplicates. All 39 additional diploid accessions originated from the ITC collection (Leuven, Belgium) as in vitro rooted plants and were maintained in a heated greenhouse after transfer to soil. The DNA of these 39 entries was isolated from young leaf tissue using the Invisorb™ Spin Plant Mini kit (Invitek, Berlin, Germany), following the manufacturer’s instructions.

Polymerase chain reaction amplification and fragment analysis
The 22 SSR loci (Table 1) were amplified using specific primers (Crouch et al. 1998; Lagoda et al. 1998; Hippolyte et al. 2010) that were adjusted by 5’-M13 tails to enable the use of universal fluorescently labelled primer according to Schuelke (2000). Four different fluorophores were used for the primer labelling [6-carboxyfluorescein (6-FAM), VIC, NED and PET; Applied Biosystems, Foster City, CA, USA], allowing for subsequent multiplexing of the reactions (Table 1). The reaction was performed in a final volume of 20 μL containing 10 ng of template genomic DNA, reaction buffer (consisting of 10 mM Tris–HCl (pH 8), 50 mM KCl, 0.1% Triton-X100 and 1.5 mM MgCl₂), 200 μM dNTPs (each), 1 U of Taq polymerase, 8 pmol of the M13-tailed locus-specific forward primer, 6 pmol of the fluorescently labelled universal M13 forward primer and 10 pmol of the locus-specific reverse primer. The cycling conditions were set as follows: initial denaturation step at 94 °C for 5 min, followed by 35 cycles of denaturation (94 °C/45 s), annealing at the temperature corresponding to the locus-specific primer (1 min) and extension (72 °C/1 min). Final extension was allowed for 5 min at 72 °C. The polymerase chain reaction (PCR) products were purified by ethanol/sodium acetate precipitation. Three independent PCR reactions were performed in order to improve the accuracy of allele binning.

For automatic capillary electrophoresis, optimized amounts of amplification products were combined with highly deionized formamide and internal standard (GeneScan™-500 LIZ size standard; Applied Biosystems). After 5 min denaturation at 95 °C, samples were loaded onto the automatic 96-capillary ABI 3730xl DNA Analyzer, and electrophoretic separation and signal detection were carried out with default module settings. In order to reduce the cost and increase the capacity of the genotyping platform, samples were multiplexed for the second and third round of electrophoretic separation. Up to 4-fold multiplexing was applied by combining four PCR products, labelled with different fluorescent dyes (6-FAM, VIC, NED and PET; Table 1) into a single sample for loading. The level of multiplexing could be further increased by combining products of different expected lengths, labelled with the identical fluorescent dyes.

Fragment sizing and data analysis
The resulting data were analysed using GeneMarker® v1.75 (Softgenetics, LLC, State College, PA, USA). Automated scoring of the data was followed by a careful manual check, and low-quality DNA samples were discarded from the analysis. The marker panels were built based on allele calls of the Reference DNA collection sample set and later extended by additional diploid accession allele calls, in order to increase the reference SSR-profiles database. Bins for each allele were set with respect to the allele frequencies and signal strength extracted from the three repeated runs of each sample.

The diploid and triploid accessions were analysed separately, because in the case of polyploid species, the polysonomic inheritance brings the simultaneous occurrence of several alleles of a single SSR. In such a situation, the exact number of copies of individual alleles cannot be determined; therefore, the genotypic data are converted into binary data (coded by 1—presence/0—absence) and analysed as a dominant marker’s record (Weising et al. 2005). Both genotypic and binary data were used to generate genetic similarity matrices based on Nei’s genetic distance coefficient.

genotypes. The goal of the present study was to investigate the potential of this marker set to distinguish individual accessions and to develop a standardized procedure for Musa genotyping that could serve as a basis for molecular characterization of new samples introduced into the global Musa gene bank (International Transit Centre (ITC), Leuven, Belgium) as well as to the wider Musa research and breeding community.
(Nei 1973) in the software PowerMarker v3.25 (Liu and Muse 2005). The unweighted pair-group method with arithmetic mean (UPGMA; Michener and Sokal 1957) was used to assess the relationship between individual genotypes. The results of UPGMA cluster analysis were visualized in the form of a tree using TreeView v1.6.6 (Page 1996). Polymorphism information content (PIC) and heterozygosity of individual markers were estimated in PowerMarker v3.25. The overall probability of identity ($P_{ID}$) of unrelated multilocus genotypes was assessed according to Paetkau et al. (1995), as implemented in the IDENTITY program (Wagner and Sefc 1999).

**Blind test**

In order to verify the reliability of the optimized genotyping platform and its potential as a standardized methodology for molecular characterization of new accessions, a set of anonymous samples was analysed [Additional Information 3]. The genomic DNA was extracted from lyophilized leaf tissue provided by the ITC, and samples were analysed following an identical experimental procedure as for the reference DNA collection. Negative and positive controls (five previously analysed reference genotypes) were included in the blind test to ensure correct allele sizing and control the consistency of the electrophoretic condition. The unknown samples were coded numerically and their true identity was disclosed by our partners only after the data analysis. As revealed subsequently, the blind test sample set contained an additional four samples that were duplicates of the reference DNA collection [see Additional Information 1 and 3].

<table>
<thead>
<tr>
<th>Marker</th>
<th>Fluorophore</th>
<th>Motif</th>
<th>Reference</th>
<th>Accession GenBank</th>
<th>Annealing temperature (this study; °C)</th>
<th>Minimum allele (this study; bp)</th>
<th>Maximum allele (this study; bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mMaCIR01</td>
<td>6-FAM</td>
<td>(GA)20</td>
<td>Lagoda et al. (1998)</td>
<td>X87262</td>
<td>55</td>
<td>241</td>
<td>440</td>
</tr>
<tr>
<td>mMaCIR03</td>
<td>6-FAM</td>
<td>(GA)10</td>
<td>Lagoda et al. (1998)</td>
<td>X87263</td>
<td>55</td>
<td>111</td>
<td>147</td>
</tr>
<tr>
<td>mMaCIR07</td>
<td>NED</td>
<td>(GA)13</td>
<td>Lagoda et al. (1998)</td>
<td>X87258</td>
<td>53</td>
<td>136</td>
<td>195</td>
</tr>
<tr>
<td>mMaCIR08</td>
<td>VIC</td>
<td>(TC)6N24(TC)7</td>
<td>Lagoda et al. (1998)</td>
<td>X87264</td>
<td>55</td>
<td>229</td>
<td>283</td>
</tr>
<tr>
<td>mMaCIR13</td>
<td>PET</td>
<td>(GA)16N76(GA)8</td>
<td>Lagoda et al. (1998)</td>
<td>X90745</td>
<td>53</td>
<td>268</td>
<td>427</td>
</tr>
<tr>
<td>mMaCIR24</td>
<td>PET</td>
<td>(TC)7</td>
<td>Lagoda et al. (1998)</td>
<td>Z85972</td>
<td>48</td>
<td>240</td>
<td>291</td>
</tr>
<tr>
<td>mMaCIR27</td>
<td>PET</td>
<td>(GA)9</td>
<td>Lagoda et al. (1998)</td>
<td>Z85962</td>
<td>58</td>
<td>232</td>
<td>277</td>
</tr>
<tr>
<td>mMaCIR39</td>
<td>VIC</td>
<td>(CA)5GATA(GA)5</td>
<td>Lagoda et al. (1998)</td>
<td>Z85970</td>
<td>52</td>
<td>329</td>
<td>390</td>
</tr>
<tr>
<td>mMaCIR40</td>
<td>6-FAM</td>
<td>(GA)13</td>
<td>Lagoda et al. (1998)</td>
<td>Z85977</td>
<td>54</td>
<td>169</td>
<td>247</td>
</tr>
<tr>
<td>mMaCIR45</td>
<td>6-FAM</td>
<td>(TA)4CA(CTCGA)4</td>
<td>Lagoda et al. (1998)</td>
<td>Z85968</td>
<td>57</td>
<td>274</td>
<td>318</td>
</tr>
<tr>
<td>mMaCIR50</td>
<td>VIC</td>
<td>(CA)10</td>
<td>Hippolyte et al. (2010)</td>
<td>AM950440</td>
<td>54</td>
<td>253</td>
<td>376</td>
</tr>
<tr>
<td>mMaCIR152</td>
<td>6-FAM</td>
<td>(CTT)18,(CT)17,(CA)6</td>
<td>Hippolyte et al. (2010)</td>
<td>AM950442</td>
<td>54</td>
<td>147</td>
<td>195</td>
</tr>
<tr>
<td>mMaCIR164</td>
<td>VIC</td>
<td>(AC)14</td>
<td>Hippolyte et al. (2010)</td>
<td>AM950454</td>
<td>55</td>
<td>256</td>
<td>458</td>
</tr>
<tr>
<td>mMaCIR195</td>
<td>VIC</td>
<td>(GA)11,(GA)6</td>
<td>Hippolyte et al. (2010)</td>
<td>AM950461</td>
<td>54</td>
<td>262</td>
<td>306</td>
</tr>
<tr>
<td>mMaCIR196</td>
<td>NED</td>
<td>(TA)4, (TC)17, (TC)3</td>
<td>Hippolyte et al. (2010)</td>
<td>AM950462</td>
<td>55</td>
<td>163</td>
<td>201</td>
</tr>
<tr>
<td>mMaCIR214</td>
<td>NED</td>
<td>(AC)7</td>
<td>Hippolyte et al. (2010)</td>
<td>AM950480</td>
<td>53</td>
<td>115</td>
<td>238</td>
</tr>
<tr>
<td>mMaCIR231</td>
<td>NED</td>
<td>(TC)10</td>
<td>Hippolyte et al. (2010)</td>
<td>AM950497</td>
<td>55</td>
<td>236</td>
<td>286</td>
</tr>
<tr>
<td>mMaCIR260</td>
<td>PET</td>
<td>(TG)8</td>
<td>Hippolyte et al. (2010)</td>
<td>AM950515</td>
<td>55</td>
<td>204</td>
<td>264</td>
</tr>
<tr>
<td>mMaCIR264</td>
<td>6-FAM</td>
<td>(CT)17</td>
<td>Hippolyte et al. (2010)</td>
<td>AM950519</td>
<td>53</td>
<td>234</td>
<td>383</td>
</tr>
<tr>
<td>mMaCIR307</td>
<td>NED</td>
<td>(CA)6</td>
<td>Hippolyte et al. (2010)</td>
<td>AM950533</td>
<td>54</td>
<td>143</td>
<td>173</td>
</tr>
<tr>
<td>Ma-1-32</td>
<td>NED</td>
<td>(GA)17AA(GA)8AA(GA)2</td>
<td>Crouch et al. (1998)</td>
<td>n/a</td>
<td>58</td>
<td>208</td>
<td>251</td>
</tr>
<tr>
<td>Ma-3-90</td>
<td>PET</td>
<td>(CT)11</td>
<td>Crouch et al. (1998)</td>
<td>n/a</td>
<td>53</td>
<td>147</td>
<td>191</td>
</tr>
</tbody>
</table>

*Excluded from the analysis due to unreproducible amplification.
**Genotyping error handling**

To eliminate genotyping errors, several precautions were employed in the genotyping process, following the recommendations by Bonin et al. (2004). First, to minimize the allelic dropout effect, the multitube approach (Taberlet et al. 1996) was used with three independent reactions for each marker/genotype combination. The error-prone samples with low-quality DNA were discarded from the analysis. Second, the multilocus genotype was examined and accessions differing at a single locus were carefully inspected and reanalysed (if needed) to confirm the difference. Third, to decrease human factor errors, sample preparation was performed by two different people for the replicated reactions. Data evaluation was ruled by strictly pre-set parameters to avoid errors such as misinterpretation of stutter peaks.

**Results**

Twenty-two SSR markers were selected by CIRAD as a set enabling one to distinguish between individuals in the Musa reference DNA collection (Crouch et al. 1998; Lagoda et al. 1998; Hippolyte et al. 2010; Website 1; Table 1). After the initial double-repeated primer test screening using our protocol, 19 markers were selected out of the initial 22 markers set, for their clear reproducible amplification pattern. The three markers that were excluded from the analysis produced extensive stuttering of peaks, disabling the reproducible interpretation of the SSR profiles. All further analyses were performed with the selected 19 SSRs. Altogether, the SSR profiles were collected for 70 diploid and 38 triploid banana accessions. All necessary information on the genotyping methodology as well as the complete allele score files for the analysed genotypes are also available online through http://alomouc.ueb.cas.cz/musa-genotyping-centre.

**Analysis of diploid accessions**

Diploid accessions were underrepresented in the reference DNA collection; therefore, we decided to include additional diploids in the analysis to increase the number of reference SSR profiles [Additional Information 2]. In the resulting set of 70 diploid accessions (including the blind test entries), a total of 292 alleles were scored from the 19 loci, with an average of 15.4 alleles per locus. The observed heterozygosity (the fraction of all individuals who are heterozygous for the observed locus) ranged between 0.179 and 0.714 (mean 0.450). The PIC of the markers used was relatively high (mean 0.827), ranging between 0.625 and 0.936 (see Table 2 for details). The $P_{10}$ (combined over all loci), which represents the probability of observing identical genotypes purely by chance, was $9.44 \times 10^{-29}$, denoting the extremely high resolution power of this marker set.

The UPGMA cluster analysis based on the Nei (1973) genetic distance revealed a relatively clear grouping of genotype groups and subgroups (Fig. 1). The B-genome representatives M. balbisiana including the diploid hybrid cultivars (AB and BB×T) formed a separate cluster (cluster I). The A-genome representatives M. acuminata species were grouped in several clusters depending on their subspecies classification. Musa acuminata ssp. banksii entries grouped within cluster II, M. acuminata ssp. microcarpa grouped together with Musa schizocarpa and AS hybrids within cluster III. The sole representative of errans subspecies, cultivar Agutay, was present at the separate clade related to the above-described M. acuminata clusters. Subcluster VI contained the M. acuminata ssp. zebrina representatives. Subspecies burmannica, burmannicoides and siamea were grouped within cluster VII, sharing their position with several entries from the section Rhodochlamys. Musa acuminata ssp. malaccensis subspecies formed a separate cluster labelled VIII (Fig. 1). Most of the AA cultivars were grouped within cluster IV. The Australimusa section representatives included in the study formed cluster V, together with Musa beccarii (classified under the Callimusa section). Musa coccinea, another representative of the Callimusa section, was separated from all the other groups, resembling the behaviour of an outgroup species. As mentioned before, Rhodochlamys species were partly present in cluster VII (specifically the Musa ornata and Musa manii entries). Musa velutina accessions, another representative of the Rhodochlamys section, formed a separate cluster labelled IX together with a single M. ornata accession (ITC 1330).

**Blind test with diploid accessions**

When the anonymous samples were included in the dataset, the clustering was slightly changed (Fig. 2). The position of accession Agutay (M. acuminata ssp. errans) moved into cluster II containing mostly the M. acuminata ssp. banksii entries. Another alteration could be seen in the position of M. acuminata ssp. zebrina species, which no longer formed a separate subclade (previously labelled with VI), but instead clustered within cluster IV containing the AA cultivars. Finally, cluster VII, although not changed in the content, now showed a different subclustering pattern, with the M. acuminata ssp. burmannica, burmannicoides and siamea species grouped together within one subcluster (VIIa), separated from the Rhodochlamys entries (subcluster VIIb).

Out of the nine anonymous accessions, eight were assessed correctly as the closest related species to the corresponding reference accession (Fig. 2). The only...
exception was blind sample no. 4 (M. acuminata ssp. malaccensis ITC 0250), which did not group together with its reference genotype (the same ITC 0250 accession), but instead clustered together within the M. acuminata ssp. banksii subgroup (clade II). The multilocus genotypes of the blind sample no. 4 (ITC 0250) and the closest related genotype Higa (ITC 0428) differed at a single locus only, suggesting that the blind sample no. 4 belonged very likely to the banksii subspecies.

In order to further investigate this incongruence in the blind test results, we conducted internal transcribed spacer (ITS) locus sequence analysis according to Hřibová et al. (2011) in the problematic malaccensis accessions. This analysis confirmed that the blind sample no. 4 was not identical to the genotype M. acuminata ssp. malaccensis ITC 0250, which was originally received from the ITC and stored in the local greenhouse [see Additional Information 4]. The results are, however, not conclusive about the identity of blind sample no. 4, as only a single representative of the banksii subspecies was used for the ITS analysis in our previous study. Thus, it cannot be explicitly stated whether blind sample no. 4 is a different genotype of M. acuminata ssp. malaccensis or ssp. banksii, or rather a hybrid between malaccensis and banksii subspecies. Only a more detailed sequence analysis would probably provide a definite answer.

### Analysis of triploid accessions

Altogether, 38 triploid accessions were analysed (including the blind test entries). The 19 microsatellite loci scored a total of 267 alleles, ranging between 8 and 24 per locus, with a mean value of 14 alleles per locus. The average PIC of the SSR markers applied on the triploid accessions was 0.850 (Table 3).

The UPGMA analysis majority rule consensus tree showed two main clusters, cluster A and cluster B (Fig. 3). Cluster A contained solely the AAA hybrid accessions, with a separated clade bearing the Lujugira/Mutika subgroup representatives, as well as a

<table>
<thead>
<tr>
<th>Marker</th>
<th>Major allele frequency</th>
<th>Number of unique genotypes observed</th>
<th>Allele number</th>
<th>Observed heterozygosity</th>
<th>PIC*</th>
</tr>
</thead>
<tbody>
<tr>
<td>mMaCIR01</td>
<td>0.125</td>
<td>39</td>
<td>26</td>
<td>0.531</td>
<td>0.936</td>
</tr>
<tr>
<td>mMaCIR03</td>
<td>0.357</td>
<td>13</td>
<td>7</td>
<td>0.400</td>
<td>0.694</td>
</tr>
<tr>
<td>mMaCIR07</td>
<td>0.181</td>
<td>33</td>
<td>21</td>
<td>0.551</td>
<td>0.883</td>
</tr>
<tr>
<td>mMaCIR08</td>
<td>0.231</td>
<td>22</td>
<td>12</td>
<td>0.646</td>
<td>0.830</td>
</tr>
<tr>
<td>mMaCIR13</td>
<td>0.229</td>
<td>28</td>
<td>19</td>
<td>0.543</td>
<td>0.870</td>
</tr>
<tr>
<td>mMaCIR24</td>
<td>0.328</td>
<td>19</td>
<td>15</td>
<td>0.344</td>
<td>0.767</td>
</tr>
<tr>
<td>mMaCIR39</td>
<td>0.200</td>
<td>39</td>
<td>20</td>
<td>0.714</td>
<td>0.893</td>
</tr>
<tr>
<td>mMaCIR40</td>
<td>0.233</td>
<td>29</td>
<td>23</td>
<td>0.534</td>
<td>0.887</td>
</tr>
<tr>
<td>mMaCIR45</td>
<td>0.207</td>
<td>16</td>
<td>8</td>
<td>0.357</td>
<td>0.801</td>
</tr>
<tr>
<td>mMaCIR150</td>
<td>0.328</td>
<td>20</td>
<td>15</td>
<td>0.522</td>
<td>0.797</td>
</tr>
<tr>
<td>mMaCIR152</td>
<td>0.232</td>
<td>19</td>
<td>11</td>
<td>0.250</td>
<td>0.849</td>
</tr>
<tr>
<td>mMaCIR164</td>
<td>0.161</td>
<td>28</td>
<td>22</td>
<td>0.322</td>
<td>0.916</td>
</tr>
<tr>
<td>mMaCIR196</td>
<td>0.250</td>
<td>23</td>
<td>13</td>
<td>0.453</td>
<td>0.855</td>
</tr>
<tr>
<td>mMaCIR214</td>
<td>0.383</td>
<td>12</td>
<td>7</td>
<td>0.313</td>
<td>0.670</td>
</tr>
<tr>
<td>mMaCIR231</td>
<td>0.214</td>
<td>27</td>
<td>14</td>
<td>0.540</td>
<td>0.880</td>
</tr>
<tr>
<td>mMaCIR260</td>
<td>0.329</td>
<td>20</td>
<td>14</td>
<td>0.357</td>
<td>0.765</td>
</tr>
<tr>
<td>mMaCIR264</td>
<td>0.239</td>
<td>35</td>
<td>24</td>
<td>0.522</td>
<td>0.900</td>
</tr>
<tr>
<td>mMaCIR307</td>
<td>0.500</td>
<td>10</td>
<td>6</td>
<td>0.179</td>
<td>0.625</td>
</tr>
<tr>
<td>Ma-3-90</td>
<td>0.167</td>
<td>31</td>
<td>15</td>
<td>0.474</td>
<td>0.893</td>
</tr>
<tr>
<td>Mean</td>
<td>0.258</td>
<td>24.4</td>
<td>15.4</td>
<td>0.450</td>
<td>0.827</td>
</tr>
</tbody>
</table>

*Polymorphism information content.

---

Table 2 Allele number, frequency of the major allele, unique genotypes observed, heterozygosity and informativeness (PIC) of the 19 microsatellite loci applied on the dataset of 70 diploid Musa accessions.

---

---
Fig. 1 Dendrogram showing the results of the UPGMA analysis of diploid accessions dataset. Bootstrap support values higher than 50% are marked below the corresponding branches. The classification of the genotypes into individual sections, species and subspecies of the genus Musa is indicated by the coloured side bars and legends. A complete list of accessions with their taxonomic details can be found in [Additional Information 1 and 2].

Fig. 2 Dendrogram showing the results of the UPGMA analysis of diploid accessions dataset including the blind test samples. Bootstrap support values higher than 50% are marked below the corresponding branches. The anonymous samples included in the blind test are highlighted in red. The classification of the genotypes into individual sections, species and subspecies of the genus Musa is indicated by the coloured side bars and legends. A complete list of accessions with their taxonomic details can be found in [Additional Information 1, 2 and 3].
distinct clade leading to the edible species from the Cavendish and Gros Michel subgroups. Among all the AAA entries included in the analysis, only the accession Pisang Berangan clustered outside the A cluster, sharing a clade (IVa) with the African plantain representatives within the main cluster B. The second main cluster B was split into four subclusters/subclades. While subcluster II was formed exclusively by the AAB hybrid entries, subcluster I also contained an ABB genotype Namwa Khom (Pisang Awak subgroup), as a closest relative of the AAB Figue Pomme Géante accession from the Silk subgroup. Two of the ABB hybrid representatives, Kluai Tiparot and Pelipita, formed the third subclade within the B cluster (III). Most of the ABB hybrids were grouped under IVb, together with an ABB accession Popoulou. The African plantains formed a separate clade IVa with a single AAA representative P. Berangan, as mentioned above.

**Table 3** Major allele frequency, allele number and informativeness (PIC) of the 19 microsatellite loci applied on the dataset of 38 triploid *Musa* accessions.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Major allele frequency</th>
<th>Allele number</th>
<th>PIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>mMaCIR01</td>
<td>0.105</td>
<td>24</td>
<td>0.942</td>
</tr>
<tr>
<td>mMaCIR03</td>
<td>0.237</td>
<td>12</td>
<td>0.839</td>
</tr>
<tr>
<td>mMaCIR07</td>
<td>0.132</td>
<td>17</td>
<td>0.912</td>
</tr>
<tr>
<td>mMaCIR08</td>
<td>0.237</td>
<td>14</td>
<td>0.867</td>
</tr>
<tr>
<td>mMaCIR13</td>
<td>0.342</td>
<td>12</td>
<td>0.804</td>
</tr>
<tr>
<td>mMaCIR24</td>
<td>0.289</td>
<td>12</td>
<td>0.817</td>
</tr>
<tr>
<td>mMaCIR39</td>
<td>0.316</td>
<td>18</td>
<td>0.859</td>
</tr>
<tr>
<td>mMaCIR40</td>
<td>0.289</td>
<td>9</td>
<td>0.817</td>
</tr>
<tr>
<td>mMaCIR45</td>
<td>0.289</td>
<td>12</td>
<td>0.814</td>
</tr>
<tr>
<td>mMaCIR150</td>
<td>0.263</td>
<td>8</td>
<td>0.808</td>
</tr>
<tr>
<td>mMaCIR152</td>
<td>0.263</td>
<td>12</td>
<td>0.850</td>
</tr>
<tr>
<td>mMaCIR164</td>
<td>0.131</td>
<td>18</td>
<td>0.913</td>
</tr>
<tr>
<td>mMaCIR196</td>
<td>0.237</td>
<td>15</td>
<td>0.881</td>
</tr>
<tr>
<td>mMaCIR214</td>
<td>0.263</td>
<td>8</td>
<td>0.788</td>
</tr>
<tr>
<td>mMaCIR231</td>
<td>0.132</td>
<td>16</td>
<td>0.905</td>
</tr>
<tr>
<td>mMaCIR260</td>
<td>0.474</td>
<td>13</td>
<td>0.733</td>
</tr>
<tr>
<td>mMaCIR264</td>
<td>0.158</td>
<td>18</td>
<td>0.913</td>
</tr>
<tr>
<td>mMaCIR307</td>
<td>0.342</td>
<td>8</td>
<td>0.760</td>
</tr>
<tr>
<td>Ma-3-90</td>
<td>0.105</td>
<td>21</td>
<td>0.934</td>
</tr>
<tr>
<td>Mean</td>
<td>0.242</td>
<td>14.1</td>
<td>0.850</td>
</tr>
</tbody>
</table>

**Blind test with triploid accessions**

Six encoded triploid samples were included in the blind test and all of them were assessed correctly as the closest related species to the corresponding reference genotype from identical subgroups, with significant statistical support (Fig. 4). The position of some clades was slightly altered after the inclusion of anonymous samples in the analysis (Fig. 4). Specifically, the UPGMA cluster analysis has now shown an altered position of the clade previously labelled III (ABB accessions Pelipita and Kluai Tiparot) and the subclade of the cluster previously labelled II, bearing the AAB genotypes P. Palembang, P. Rajah and P. Raja Bulu. However, the bootstrap statistical support for nodes leading to these clades was not significantly strong in either dataset, and the position of all the other clades in the consensus tree remained unchanged.

**Identification of duplicate accessions**

One hundred per cent similarity in multilocus genotypes was seen in nine pairs of duplicate accessions [Additional Information 5]. Some of the duplicates were introduced into the accession set intentionally from the local greenhouse (originally coming from the ITC collection) to assess the capability of our genotyping system at spotting the duplicate accessions. Others were introduced through the blind test samples (see Materials and methods). All the duplicates were identified [Additional Information 5], with two exceptions. The *Musa textilis* reference collection DNA sample (ref. 50), which was reported to correspond to the ITC accession ITC 1072, was shown to be identical to another *M. textilis* accession (ITC 0539). This suggests that the reference sample (ref. 50) was mislabelled or its origin was not reported correctly.

Another anticipated duplicate, introduced into the triploid entries through the blind test, was accession blind 12 (Pisang Bakar ITC 1064). Its corresponding reference DNA sample was ref. 19. However, their identity based on the multilocus molecular profile was not approved. Although the two samples differed at 7 out of the 19 scored SSR loci, their closest relationship was revealed after the UPGMA cluster analysis (Fig. 4), suggesting that their mutual subgroup classification (subgr. Ambon) may be correct, but the identity of one of the samples was confused.

Moreover, more than one duplicate accession was reported for both accession ref. 8 (*M. acuminata* ssp. *burmannicoides* ‘Calcutta4’) and ref. 21 (*M. balbisiana* ‘Tani’). The second duplicate for each of the two reference samples was classified under the same species/subspecies [Additional Information 5]. This indicates that
either the marker set used did not have enough resolution power to distinguish these accessions or, more likely, based on the low $P_{ID}$ value mentioned above, these accessions were mislabelled.

**Discussion**

While the use of microsatellite markers to analyse genetic diversity among Musa species is well documented (e.g. Kaemmer et al. 1997; Grapin et al. 1998; Buharwalla et al. 2005; Ning et al. 2007; Venkatachalam et al. 2008; Wang et al. 2010), its application in the form of a standardized platform to serve for genotyping purposes for the wider Musa community is still missing. In this study, we attempted to develop an optimized SSR-based system for molecular characterization of Musa accessions that could be used as the basis for the foundation of the Musa Genotyping Centre (MGC).

Mislabelling of accessions and sample duplications are common problems in germplasm collections (e.g. Virk et al. 1995; Zhang et al. 2006). The resolution of the marker set tested in this study was high enough ($P_{ID} = 9.44 \times 10^{-29}$) to distinguish between different accessions and proved to be powerful enough to identify mislabelled accessions, as documented in the case of the *M. acuminata* ssp. *malaccensis* accession. Similarly, its potential for identifying duplicates was clearly proved on the present dataset. Nevertheless, we wanted to ensure reproducibility of results and minimization of genotyping errors prior to its implementation into practice. When compared with the original data reported by Lagoda et al. (1998) for a subset of markers, the allele size ranges were overlapping, but not identical. Similar problems have been described previously, and most often they were attributed to the method used and the conditions of electrophoretic separation (e.g. Testolin et al. 2000; Creste et al. 2003). Also, the automatic capillary electrophoresis system used in this experiment allows for much higher resolution and run-to-run precision than the previously used gel-based systems. Therefore, the wider range of allele sizes and higher numbers of identified alleles are adding to the resolution power of the marker set rather than restricting the capability of the platform.

Among the common genotyping errors that are responsible for misidentification of a particular genotype, allele dropout and false allele amplification play an important role. Allele dropout is an accidental failure of PCR to amplify one of the alleles present at the heterozygous locus, which produces false homozygous patterns (Pompanon et al. 2005). To deal with this problem, three options have been proposed. The first relies on systematic replication of the genotyping, i.e. a multtube approach, which in most cases would expose
the underlying allelic dropouts or allele shifts due to poor amplification (Taberlet et al. 1996). Another possibility is to allow for a certain level of mismatch tolerance, provided that enough loci are scored. Then based on the multilocus genotype, the differences generated by genotyping errors can be distinguished from those that are actual differences between two genotypes by the low number of mismatches (McKelvey and Schwartz 2004). The third option combines the two former ones, with replicated genotyping only for samples where three or fewer mismatches at different loci were observed. These multilocus genotypes are re-evaluated after the repeated typing to prove that they are different genotypes in reality, but the cost increase by PCR replications is minimized (Zhang et al. 2006). In this pilot study, we adopted the multitube approach with three replicates to ensure maximum precision. However, with many more samples coming to be analysed in the MGC, and thereby increasing the reference database of molecular profiles, the third (combined) option appears to be adequate and is currently being tested.

The grouping revealed by the UPGMA cluster analysis was consistent with the characterization based on the morphotaxonomic classification of accessions (Figs. 2 and 4). The Callimusa section, however, did not form a separate cluster, which reflects its controversial position and agrees with its previously reported close relationship to the Australimusa species (Jarret and Gawel 1995; Wong et al. 2001b, 2002). Also, the close relationship between Rhodochlamys and M. acuminata species (Wong et al. 2002; Bartoš et al. 2005; Li et al. 2010; Liu et al. 2010) was confirmed. The marker set enabled distinction to the level of individual subgroup/subspecies. The degree of polymorphism varied between subgroups and subspecies, and polymorphic sites were still to be found within the subgroups and subspecies. For example, in contrast to the study of Creste et al. (2003) who were not able to find polymorphic loci among the Cavendish subgroup of bananas in their study based on six SSR loci, the marker set used in our study did provide polymorphic loci among the three representatives of the Cavendish subgroup, allowing for their distinction. Obviously, the larger number of loci scored increases the possibility of finding enough polymorphic loci. On the other hand, limitations in the resolution of microsatellite markers become evident when somatic mutants are analysed; as they share the common origin, the genetic variation that is narrowed through
the cycles of vegetative propagation may not be reflected in their SSR molecular profile (Cipriani et al. 1994; Creste et al. 2003; Esselink et al. 2003). As most of the commercial banana cultivars are vegetatively propagated clones, assessment of their genetic variability through the marker set tested in this study may not be successful and is yet to be confirmed. However, it still presents a very useful platform for molecular characterization of unknown samples and assessment of the genetic integrity of the Musa germplasm collections.

Although microsatellites have been used as reliable markers for projects with labour division among laboratories (Bredemeijer et al. 2002; Röder et al. 2002), several pieces of work have shown that there was a significant level of incongruence between the results obtained at different workplaces, thus complicating the transferability and comparability of the data (Jones et al. 1997; Weeks et al. 2002; This et al. 2004; Van Treuren et al. 2010). In the light of this, centralization of genotyping activities in Musa and its standardization as a service to the research community appear to be preferable options. In addition to facile quality control, the core facility would enable the use of other methods to support the genotyping, such as flow cytometric estimation of ploidy level and/or genome size, keeping in mind that the genotyping data treatment differs for the diploid and polyploid accessions (see Materials and methods). Obviously, sample transfer requirements can be minimized if both types of analysis are performed at a single site. Moreover, with every new sample passing through the analysis, the database of reference SSR profiles is enlarged and the probability of identifying the closest relative or exactly matching accession is enhanced.

Based on our results obtained with the SSR markers presented in this work and those of Hříbová et al. (2011) obtained with ITS, as well as the long-term experience in DNA flow cytometry (Doležel 1991; Lysák et al. 1999; Roux et al. 2003; Bartoš et al. 2005; Doleželová et al. 2005), the MGC has been established at the Institute of Experimental Botany in Olomouc (Czech Republic) under the umbrella of Bioversity International (http://olomouc.ueb.cas.cz/musa-genotyping-centre). The Centre serves the whole Musa research and breeding community. Moreover, the genotyping platform has already been included in the pipeline for characterization of newly introduced accessions to the international banana germplasm collection (ITC). In this pipeline, fresh leaf tissue samples for molecular characterization are received at the MGC, where they are subjected to ploidy level measurement via flow cytometry; the DNA is extracted and used for collecting the SSR profiles of the 19 markers as described above. In certain cases, where the results of the SSR genotyping are not conclusive enough to reliably classify the unknown samples, the ITS sequence analysis according to Hříbová et al. (2011) can be applied. Although it is obvious that new high-content, high-throughput, genotyping approaches will gradually replace marker-based systems, we feel confident that the platform described here offers a well-founded and ready-to-use approach, which can be applied immediately and which offers higher flexibility in scaling the analysis with respect to sample size, cost efficiency and turn-around time for results.

Conclusions and forward look

The platform for genotyping of Musa germplasm described here provides a robust and reproducible approach to characterize the genetic variability of this important crop, support the management of germplasm collections and direct genotype selection for breeding improved cultivars. The database of molecular profiles keeps growing with every new sample passing through the analytical pipeline, resulting in stepwise improvement in the grouping, and consequently increasing the chance of finding an exact match for unknown samples. As part of the future plans, a batch of tetraploid accessions will be included in the analysis to make it more versatile and satisfying all possible requirements for molecular characterization of the diverse Musa gene pool.

Additional information

The following additional information is available in the online version of this article –

File 1: Taxonomic details of the reference DNA collection accessions.

File 2: List of additional diploid accessions from the ITC collection (maintained in a local greenhouse) included in the analysis.

File 3: List of encoded accessions included in the blind test.

File 4: Detailed results of the ITS sequence analysis of blind sample no. 4 and its putative corresponding reference accession—M. acuminata ssp. malaccensis (ITC 0250).

File 5: List of duplicates identified among the analysed genotypes.

Sources of funding

This work has been supported by Bioversity International (LOA Cfl. 2009/48 and LoA Cfl. 2010/58), Internal Grant Agency of Palacký University, Olomouc, Czech Republic (grant award no. Prf-2010-001) and by the Ministry of...
Education, Youth and Sports of the Czech Republic and the European Regional Development Fund (Operational Programme Research and Development for Innovations No. CZ.1.05/2.1.00/01.0007).

Contributions by the authors
All authors have contributed to, read and approved the manuscript.

Acknowledgements
We thank our colleague Marie Seifertová for her excellent technical assistance. The work of Pavla Christelová, Miroslav Valárik, Eva Hřibová, Nicolas Roux, Stéphanie Channelière and Jaroslav Doyležel was in the context of IAEA Coordinated Research Project ‘Molecular Tools for Quality Improvement in Vegetatively Propagated Crops Including Banana and Cassava’ (D23027) and the FAO/IAEA Joint Programme.

Conflict of interest statement
None declared.

References
INIBAP. 2006. Global conservation strategy for Musa (Banana and Plantain). A consultative document prepared by INIBAP with the collaboration of numerous partners in the Musa research-and-development community. France: INIBAP.
Li LF, Häkkinnen M, Yuan YM, Hao G, Ge XJ. 2010. Molecular phylo-
geny and systematics of the banana family (Musaceae) inferred from multiple nuclear and chloroplast DNA fragments, with a special reference to the genus Musa. Molecular Phylogenetics and Evolution 57: 1–10.


geny of banana cultivars from Thailand based on HAT-RAPD markers. Genetic Resources and Crop Evolution 54: 1565–1572.

Schuelke M. 2000. An economic method for the fluorescent label-


Virk PS, Newbury HJ, Jackson TM, Ford-Lloyd BV. 1995. The identi-
fication of duplicate accessions within a rice germplasm collec-

Wagner HW, Sefc KM. 1999. IDENTITY 1.0. Vienna: Centre for Applied Genetics, University of Agricultural Sciences.


