ISSN 1479-2621

Plant Genetic Resources: Characterization and Utilization (2019) **17**(6); 536–544
doi:10.1017/S1479262119000376

Genetic diversity and core subset selection in *ex situ* seed collections of the banana crop wild relative *Musa balbisiana*

Yves Bawin[1,2,3,4]*, Bart Panis[5], Samuel Vanden Abeele[1,6], Zhiying Li[7], Julie Sardos[8], Janet Paofa[9], Xue-Jun Ge[10], Arne Mertens[1,11], Olivier Honnay[2] and Steven B. Janssens[1,2]

[1]*Crop Wild Relatives and Useful Plants, Meise Botanic Garden, Meise, Vlaams Brabant, Belgium,* [2]*Plant Conservation and Population Biology, KU Leuven, Leuven, Belgium,* [3]*Flanders Research Institute for Agriculture, Fisheries and Food (ILVO), Melle, Belgium,* [4]*Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium,* [5]*Bioversity International, Heverlee, Belgium,* [6]*Evolutionary Biology and Ecology, Université Libre de Bruxelles, Bruxelles, Belgium,* [7]*Institute of Tropical Crop Genetic Resources, Chinese Academy of Tropical Agricultural Sciences and National Gene Bank of Tropical Crops, Danzhou, China,* [8]*Bioversity International, Parc Scientifique Agropolis II, Montpellier, France,* [9]*PNG National Agricultural Research Institute, Southern Regional Centre – Laloki, Port Moresby, Papua New Guinea,* [10]*South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China and* [11]*Laboratory of Tropical Crop Improvement, KU Leuven, Leuven, Belgium*

Received 11 June 2019; Accepted 18 October 2019 – First published online 29 November 2019

Abstract

Crop wild relatives (CWRs) play a key role in crop breeding by providing beneficial trait characteristics for improvement of related crops. CWRs are more efficiently used in breeding if the plant material is genetically characterized, but the diversity in CWR genetic resources has often poorly been assessed. Seven seed collections of *Musa balbisiana*, an important CWR of dessert and cooking bananas, originating from three natural populations, two feral populations and two *ex situ* field collections were retrieved and their genetic diversity was quantified using 18 microsatellite markers to select core subsets that conserve the maximum genetic diversity. The highest genetic diversity was observed in the seed collections from natural populations of Yunnan, a region that is part of *M. balbisiana*'s centre of origin. The seeds from the *ex situ* field collections were less genetically diverse, but contained unique variation with regards to the diversity in all seed collections. Seeds from feral populations displayed low genetic diversity. Core subsets that maximized genetic distance incorporated almost no seeds from the *ex situ* field collections. In contrast, core subsets that maximized allelic richness contained seeds from the *ex situ* field collections. We recommend the conservation and additional collection of seeds from natural populations, preferentially originating from the species' region of origin, and from multiple individuals in one population. We also suggest that the number of seeds used for *ex situ* seed bank regeneration must be much higher for the seed collections from natural populations.

Keywords: banana, crop wild relatives, *ex situ* seed banking, genetic diversity, *Musa balbisiana*

*Corresponding author. E-mail: yves.bawin@kuleuven.be

Introduction

Crop wild relatives (CWRs) play a key role in breeding by providing beneficial trait characteristics for improvement of related crops. However, the inter- and intraspecific diversity in CWRs is in decline due to global threats such as ecosystem degradation and climate change (Ford-Lloyd *et al.*, 2011). Given the importance of CWRs for agriculture, different *ex situ* conservation strategies have been developed (Heywood *et al.*, 2007). Seed banking has the advantage over other *ex situ* methods in that it allows long-term storage of plant material at a reasonable cost and that it can include a larger part of the gene pool (Li and Pritchard, 2009). Nevertheless, many CWR *ex situ* seed banks are underused because of the absence of genetic diversity information (Schoen and Brown, 2001; Dempewolf *et al.*, 2017). Moreover, the lack of genetic diversity assessments in *ex situ* seed banks may result in the loss of genetic diversity when germplasm is regenerated, because the subset of seeds used for regeneration might not sufficiently reflect the total diversity in the collection (Schoen and Brown, 2001; Fu, 2017). If the genetic diversity in an *ex situ* seed bank is known, genetic resources conservation can be optimized by delineating core collections. Core collections are subsets of accessions that incorporate the maximal amount of genetic diversity present in the original collection (Brown, 1989). Genetic diversity in *ex situ* collections can be maximized by either maximizing allelic richness or genetic distance. A distant subset of widely-adapted accessions is desired by plant breeders, while subsets that include rare alleles are more interesting for taxonomists and geneticists (Marita *et al.*, 2000).

Dessert and cooking bananas (*Musa* spp.) belong to the most prominent tropical and subtropical food commodities in the world (FAO, 2019). The genetic contribution of the CWR *Musa balbisiana* Colla to banana cultivars has been associated with a higher tolerance to banana weevil infestation and drought (Stover and Simmonds, 1987; Thomas *et al.*, 1998; Ocan *et al.*, 2008; Kissel *et al.*, 2015). *M. balbisiana* has a natural geographic range that reaches from India to South China (Perrier *et al.*, 2011) with its centre of origin most likely situated in the northern Indo-Burma region (Janssens *et al.*, 2016). In addition, feral *M. balbisiana* populations are found far outside its natural range (Perrier *et al.*, 2011). *M. balbisiana* seeds can be stored after desiccation without losing their viability, making them suitable for *ex situ* seed bank conservation (Stotzky *et al.*, 1962).

Here, we quantified genetic diversity in seven *M. balbisiana ex situ* seed collections that were separately collected from three natural populations, two feral populations and two *ex situ* field collections (online Supplementary Table S1 and Fig. S1). Our research questions were: (i) how genetically diverse are these *M. balbisiana* seed collections and (ii) which core subsets of seeds maximize genetic distance, allelic richness, or both? Our study contributes to the delineation of a conservation strategy of *M. balbisiana* genetic resources, serving as an example for CWR seed conservation of dessert and cooking bananas.

Materials and methods

Sampling and genotyping

In total, 247 seeds belonging to seven *ex situ* seed collections available at the Bioversity International Musa Germplasm Transit Center (ITC) and Meise Botanic Garden were selected for this study (online Supplementary Table S1). Each seed collection was retrieved from one bunch of bananas, which is common practice in the collection of banana seeds. Three seed collections were obtained from two natural populations in Yunnan and one in Hainan (China). Two seed collections were retrieved from one feral population in Amami (Japan) and one in Lae (Papua New Guinea), while two other seed collections originated from two *ex situ* field collections at the IITA genebank facilities in Kampala (Uganda) and Arusha (Tanzania). An *ex situ* field collection consisted of *M. balbisiana* accessions originating from separate populations in separate regions that were brought together in one collection.

The seed embryo was isolated using embryo rescue and subsequently germinated on a culture medium, substantially increasing the germination rate compared to seeds that are sown in a greenhouse (Afele and De Langhe, 1991). The leaves of the juvenile plants were dried on silica gel for DNA extraction. For the seed collection of Lae, DNA was directly taken from the embryo. DNA from the leaves and embryos was extracted using a modified cetyltrimethylammonium bromide protocol of Doyle and Doyle (1987). Eighteen polymorphic microsatellite markers (online Supplementary Table S2) were selected from previous studies on wild *M. balbisiana* accessions (Ge *et al.*, 2005; Wang *et al.*, 2011; Rotchanapreeda *et al.*, 2016). The reverse primer of each marker was coupled to a universal primer sequence published by Schuelke (2000) and all primer combinations were arranged in four multiplexes using Multiplex Manager v1.2 (Holleley and Geerts, 2009). Microsatellite regions were amplified using the Type-it Microsatellite PCR Kit (Qiagen, Venlo, the Netherlands), following a modified M13-like labelling protocol, which is described in detail in Vanden Abeele *et al.* (2018). Afterwards, 1.5 µl of each polymerase chain reaction (PCR) amplicon was genotyped on an ABI 3730 sequencer (Applied Biosystems, Foster City, California, VS) with 12 µl of HiDi Formamide and 0.3 µl of the MapMarker 500 labelled with the DY-632 size standard

https://doi.org/10.1017/S1479262119000376 Published online by Cambridge University Press

(Eurogentec, Seraing, Belgium). The raw genetic data were scored with Geneious Pro v9.1.7 (Kearse *et al.*, 2012). All microsatellite loci displayed distinct allelic patterns within each multiplex, validating the rearrangement of these markers into new multiplex PCRs.

Data analysis

Genetic diversity variables including the average number of alleles (N_A), the average number of alleles with an allele frequency of at least 5% ($N_{A\geq5\%}$), the number of private alleles (N_{priv}) and observed (H_O) and expected (H_E) heterozygosity were calculated with the GenAlEx v6.5 plug-in in Microsoft Excel (Peakall and Smouse, 2012). Genetic differentiation between seed collections was assessed based on Wright's *F*-statistics (F_{ST}) and visualized by a principal coordinates analysis (PCoA) using the GenAlEx v6.5 plug-in in Microsoft Excel (Peakall and Smouse, 2012). The significance of F_{ST} values was tested with 999 permutations. Genetic clustering was examined using a Bayesian Markov Chain Monte Carlo (MCMC) clustering analysis implemented in STRUCTURE v2.3.4 (Pritchard *et al.*, 2000). A series of independent runs with *K* values ranging from 1 to 10 was run in order to determine the best fitting number of clusters. Subsequently, the probability for each *K* was computed using the median of medians (MEDMEDK), the median of means (MEDMEAK), the maximum of medians (MAXMEDK) and the maximum of means (MAXMEAK) (Puechmaille, 2016) implemented in StructureSelector (Li and Liu, 2018). These statistics were demonstrated to be more robust for large differences in sampling size between populations that are included in the dataset (Puechmaille, 2016). The admixture model with correlated alleles was selected and the burn-in period length and the number of MCMC replicates were set to 150,000 and 200,000, respectively, as these estimates generated stable results for each value of *K*.

Core subset delineation

Five non-redundant accessions of *M. balbisiana* (i.e. core subsets) were selected using three different methods. First, the Maximization strategy (M-strategy) (Schoen and Brown, 1993), implemented in software CoreFinder (Cipriani *et al.*, 2010), was used with an autogenerated random seed number and 99 permutations to delineate a core collection with the highest possible allelic richness. The M-strategy minimizes the sum of probabilities that alleles are not conserved in the core collection when a certain set of accessions is selected. At least one individual of every putative population is included in the final core collection (Schoen and Brown, 1993). Second, a maximum length subtree (MLST) (Perrier *et al.*, 2003) was constructed using DARwin v6 software (Perrier and Jacquemoud-Collet, 2006) to select the genetically most distant individuals in our dataset. The MLST method required the reconstruction of a weighted neighbour-joining tree based on a dissimilarity matrix that was calculated for our dataset. The tree was subsequently pruned in a stepwise manner, each step removing one unit of each unit pair with the minimal length to the external edge. The number of individuals that remained present in the tree was set to be equal to the size of the subset that was determined with the M-strategy. Finally, the R package Corehunter III (De Beukelaer and Davenport, 2018), used in R v3.5.0 (R Core Team, 2018), was applied to maximize the Cavalli-Sforza and Edwards distance (CE distance) and the Shannon diversity index (SH index) through an advanced stochastic local search method. The CE distance is a Euclidean distance parameter that calculates distances between accessions as the square root of the differences between the allele frequencies of two individuals. The SH index reduces the redundancy of alleles in the collections by minimizing allele frequencies (Thachuk *et al.*, 2009). A core collection that contained both a high number of alleles (high SH index) and genetically distant accessions (high CE distance) was constructed as well. The SH index and CE distance contributed in equal weight to the composition of this collection, resulting in a set of accessions that is interesting for both taxonomists, geneticists and plant breeders.

Results

Genetic diversity and differentiation

Eleven out of 18 amplified microsatellite loci were polymorphic in the *ex situ* seed collections. The seed collections from natural populations of *M. balbisiana* carried a higher average number of alleles than those gathered from feral populations (Table 1). Within the group of seed collections from natural populations, the average number of alleles was higher for the seeds of Yunnan (Yunnan-1 = 2.06 ± 0.25, Yunnan-2 = 2.06 ± 0.27) than for the seeds of Hainan (1.72 ± 0.23). Seed collections from natural populations also had a higher number of low-frequency alleles, while the seeds of feral populations had no polymorphic loci if rare alleles were not included (Table 1). Furthermore, the number of private alleles (N_{priv}) was relatively low in all seed collections, but N_{priv} was much higher (0.30 ± 0.14) in the *ex situ* field collection of Kampala than in the other seed collections. The highest heterozygosity levels were observed in the seeds of Yunnan, while the observed and expected heterozygosity were remarkably low in the seed collections of Amami (feral), Lae (feral) and Kampala (*ex situ* field collection) (Table 1).

https://doi.org/10.1017/S1479262119000376 Published online by Cambridge University Press

Table 1. Genetic diversity parameters of seven seed collections of *M. balbisiana* based on 18 microsatellite markers

	$N_A \pm SE$	$N_{A\geq 5\%} \pm SE$	$N_{priv.} \pm SE$	$H_O \pm SE$	$H_E \pm SE$
Amami	1.06 ± 0.06	1.00 ± 0.00	0.06 ± 0.06	0.00 ± 0.00	0.00 ± 0.00
Hainan	1.72 ± 0.23	1.44 ± 0.19	0.00 ± 0.00	0.16 ± 0.06	0.15 ± 0.06
Lae	1.11 ± 0.08	1.00 ± 0.00	0.11 ± 0.08	0.01 ± 0.00	0.03 ± 0.00
Arusha	1.33 ± 0.14	1.33 ± 0.14	0.06 ± 0.06	0.16 ± 0.07	0.13 ± 0.05
Kampala	1.17 ± 0.09	1.17 ± 0.09	0.30 ± 0.14	0.08 ± 0.05	0.08 ± 0.04
Yunnan-1	2.06 ± 0.25	1.78 ± 0.17	0.11 ± 0.08	0.26 ± 0.06	0.22 ± 0.05
Yunnan-2	2.06 ± 0.27	1.56 ± 0.20	0.06 ± 0.06	0.21 ± 0.07	0.18 ± 0.06

N_A, average number of alleles per locus; $N_{A\geq 5\%}$, average number of alleles per locus with an allele frequency higher than 5%; $N_{priv.}$, average number of unique alleles per locus; H_O, observed heterozygosity; H_E, expected heterozygosity; SE, standard error.

Table 2. F_{ST} values calculated between all pairs of seed collections

	Amami	Hainan	Lae	Arusha	Kampala	Yunnan-1	Yunnan-2
Amami	–						
Hainan	0.474	–					
Lae	0.906	0.680	–				
Arusha	0.826	0.599	0.564	–			
Kampala	0.867	0.729	0.869	0.756	–		
Yunnan-1	0.567	0.458	0.476	0.455	0.627	–	
Yunnan-2	0.557	0.445	0.640	0.554	0.617	0.229	–

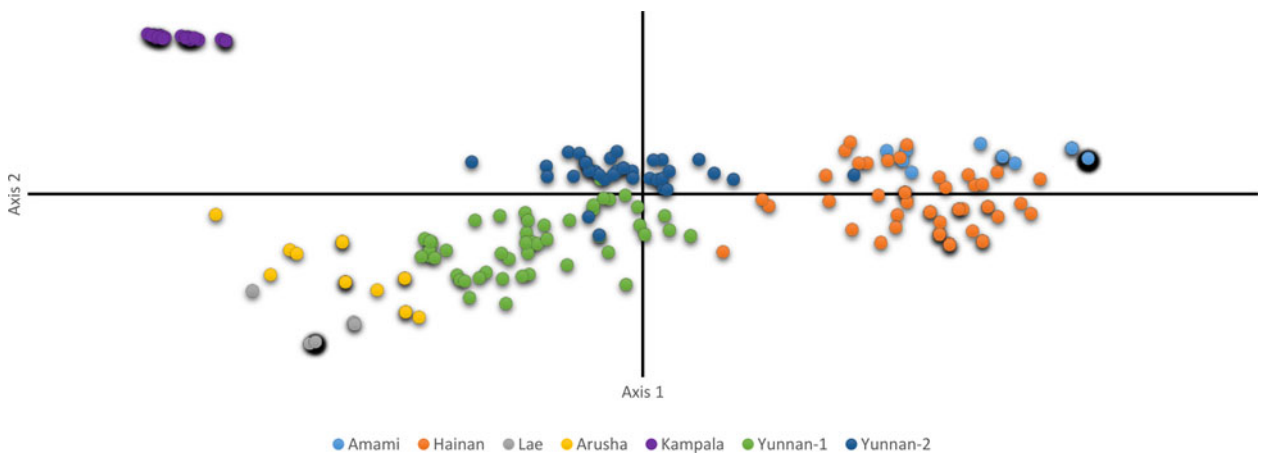

Fig. 1. The results of the PCoA displayed along the first two axes. Individuals of the same seed collection are shown in the same colour. The dark shade behind data points reflects overlapping data points.

All F_{ST} values were very high (>0.4), except for the F_{ST} value between Yunnan-1 and Yunnan-2 (Table 2). The PCoA results showed a clear genetic clustering in the dataset. The Kampala seed collection was positioned in the top left corner of the PCoA graph (Fig. 1), clearly separated from all other collections. Three other clusters were recognized along the first principal axis: one cluster with all seeds from Arusha and Lae, a second cluster that combined the Yunnan seed collections, and a third cluster that contained the seeds of Hainan and Amami. The STRUCTURE analysis for the most optimal value of k ($k = 6$) delineated similar clusters compared to the PCoA results (Fig. 2). The seed collections from natural populations showed some admixture, especially between the collections of

https://doi.org/10.1017/S1479262119000376 Published online by Cambridge University Press

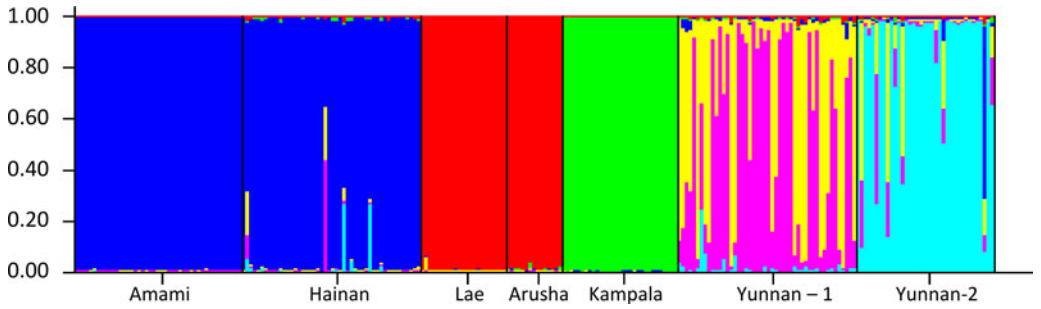

Fig. 2. STRUCTURE bar plot with designation of the seed collection origin for the most optimal value of *K* (*K* = 6).

Table 3. Number (*N*) and proportion (*P*) of seeds selected in core subsets to maximize genetic diversity using the Maximization strategy (M), the Cavalli-Sforza and Edwards distance (CE), the Shannon diversity index (SH) or a combination of CE and SH

	N_M	P_M	N_{CE}	P_{CE}	N_{SH}	P_{SH}	N_{CE+SH}	P_{CE+SH}
Amami	1	0.08	1	0.02	2	0.04	1	0.02
Hainan	1	0.08	10	0.20	7	0.14	10	0.20
Lae	1	0.08	1	0.02	1	0.02	1	0.02
Arusha	1	0.08	2	0.04	7	0.14	6	0.12
Kampala	2	0.17	1	0.02	12	0.24	9	0.18
Yunnan-1	4	0.33	21	0.43	13	0.27	12	0.24
Yunnan-2	2	0.17	13	0.27	7	0.14	10	0.20
Total	12	1.00	49	1.00	49	1.00	49	1.00

Yunnan, and encompassed three clusters that were not found in other seed collections. The feral populations and the *ex situ* field collections were clearly assigned to three clusters, combining the seeds of Amami and Hainan in one cluster and the seeds of Arusha and Lae in a second cluster. The third cluster exclusively consisted of seeds from Kampala (Fig. 2).

Core subset delineation

The two core subsets that were constructed by methods that maximize allelic richness (i.e. the M-strategy and the SH-index) contained many seeds from the Yunnan and Kampala seed collections (Table 3). The core subset that was composed using the M-strategy contained 12 genotypes, but 80% of the allelic diversity in the dataset was found in only four seeds originating from Yunnan-1, Hainan, Kampala and Arusha (online Supplementary Fig. S2). The seed collection from the *ex situ* field collection in Arusha also contributed substantially to the core subset that maximized the Shannon diversity index (Table 3). The two distance-based core subsets (constructed by the CE distance and the MLST method) predominantly included seeds from natural populations in Yunnan and Hainan (Table 3, Fig. 3). When the allelic richness (SH index) and genetic distance (CE distance) were both optimized, the resulting core collection mainly consisted of seeds from natural populations and *ex situ* field collections.

Discussion

Genetic diversity assessment and sampling recommendations

This study assessed the genetic diversity in *M. balbisiana* seed collections retrieved from natural populations, feral populations and *ex situ* field collections. The genetic diversity in all seed collections (N_A and H_O) is low compared to that previously reported in wild *M. balbisiana* populations (Ge *et al.*, 2005; Jayaweera and Samarasinghe, 2016). However, some natural populations in China that were initially described as *M. balbisiana* populations were more recently assigned to another *Musa* species (i.e. *Musa itinerans*), which may partly explain the difference between our results and previously reported findings (Ge *et al.*, 2005). The lower genetic diversity in seed collections may additionally be explained by two factors. First, seeds from the same bunch of bananas have a common maternal ancestry. So each seed collection consists exclusively of half-siblings which are, by definition, genetically less

https://doi.org/10.1017/S1479262119000376 Published online by Cambridge University Press

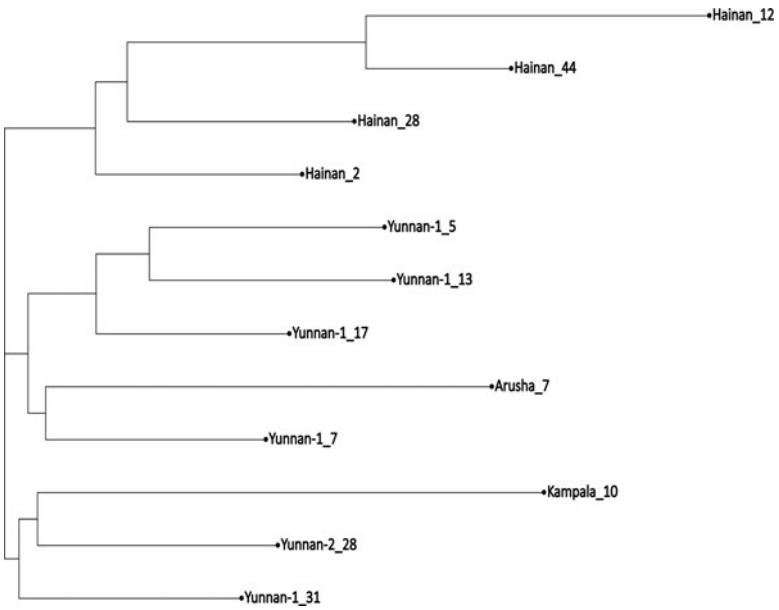

Fig. 3. Maximal length subtree that was derived from a neighbour-joining tree with 200 bootstrap replicates. Each individual is named after its origin and sampling number.

diverse than *M. balbisiana* populations with more distantly related individuals assessed in previous studies. Second, pollen flow might be limited in *M. balbisiana* populations so that one bunch of bananas might only include alleles from a relatively small number of pollen donors. Consequently, the collection of a small number of seeds from different individuals within the same population might be necessary to efficiently conserve the genetic diversity in that population. Unfortunately, seeds of different individuals are in reality hard to find during one prospection, making it difficult to collect seeds from several individuals at once. Besides, *M. balbisiana* is a short-living species that clonally propagates via budding (Ge *et al.*, 2005), two life history traits that are believed to decrease the ability of populations to persist after seed harvest (Meissen *et al.*, 2015). Collecting seeds from multiple individuals might only be possible in large populations and is also preferentially spread through time in order to reduce the impact on population viability and allow for seed sampling from different individuals.

We observed a higher genetic diversity in the two seed collections from Yunnan than in the seed collections from other regions. Previously reported genetic analyses of natural *M. balbisiana* populations from China had the highest diversity in Yunnan as well (Ge *et al.*, 2005; Wang *et al.*, 2007), confirming the high value of populations in its region of origin for conservation. Regional genetic diversity assessments of *M. balbisiana* only found moderate levels of genetic differentiation between populations (Ge *et al.*, 2005; Wang *et al.*, 2007; Jayaweera and Samarasinghe, 2016), which is in accordance with the low genetic differentiation that we observed between the two seed collections from Yunnan. Hence, the collection of seeds from several populations within the same region might not strongly increase the total genetic diversity in the *ex situ* seed bank. In contrast, the high genetic differentiation that was observed between seed collections from different regions rather suggests that gathering seeds from regions that are part of a wide geographical range should result in a higher increase in genetic diversity in the *ex situ* seed bank. These findings align with Rivière and Müller (2017) who provided evidence for common intraspecific sampling gaps in *ex situ* seed collections and argued that a more extensive sampling of the diversity across multiple biogeographic regions is required to fill these gaps. Prioritized sampling locations for seeds of *M. balbisiana* are especially located in its natural distribution area, and more specifically in its region of origin. The conservation of seeds from regions that are absent in the *ex situ* seed bank, such as the northeastern part of India and the northern regions of Laos, Vietnam and Myanmar should be of prior concern. In addition, land use changes have reduced the number of *M. balbisiana* populations in Papua New Guinea and in northern China during the last few decades, urging the conservation of the *M. balbisiana* genepool (Ge *et al.*, 2005; Wang *et al.*, 2007).

Ex situ *seed bank curation*

The delineation of genetically diverse subsets substantially increases the manageability of collections, but the

https://doi.org/10.1017/S1479262119000376 Published online by Cambridge University Press

composition of a subset varies depending on whether a high genetic distance or high allelic richness is preferred. Distance-based methods only select seeds from Yunnan or Hainan, indicating that these collections are especially interesting for plant breeders. However, methods that capture the highest allelic diversity include more seeds from the *ex situ* field collections, which makes these collections more important for taxonomists and conservation biologists (Marita *et al.*, 2000; Thachuk *et al.*, 2009). The high number of private alleles in the collection of Kampala suggests that these seeds contain a different part of the gene pool of *M. balbisiana*. However, the seeds in the *ex situ* field collections are open-pollinated and it cannot be excluded that certain unique alleles in the seeds from these field collections are introgressed from another *Musa* species, such as *M. acuminata*, which occurs in the proximity of *M. balbisiana* accessions. Furthermore, the seed collections from feral populations capture very low amounts of genetic variation, suggesting that the presence of these collections in the *ex situ* seed bank is only of secondary importance. However, these seed collections may serve as safety backups for alleles that are also conserved in other seed collections (van Hintum and Visser, 1995; Milner *et al.*, 2019).

In order to maintain a viable seed collection, it is necessary to regenerate a subset of seeds after a certain period of time. The regeneration of seeds can result in the loss of genetic diversity if the reared seeds do not properly cover the diversity in the entire seed collection or if the size of the regenerated sample pool is not large enough. Our results suggest that the number of seeds that must be used for regeneration to maintain the genetic diversity in a seed collection must be substantially larger for the seed collections from natural populations than for the collections from feral populations. *Ex situ* seed banking of *M. balbisiana* seed collections becomes much more efficient when these differences in genetic diversity between seed collections are taken into account.

The results of this study indicate that the seed collections from natural populations, feral populations and *ex situ* field collections of *M. balbisiana* are three complementary sources of genetic diversity. The seed collections from natural populations, preferably sampled within the centre of origin of the species, include high levels of genetic diversity, and conservation and collection efforts should primarily focus on these regions. We also recommend collecting a relatively small number of seeds from multiple *M. balbisiana* individuals within one population to efficiently conserve genetic diversity in the target population. The seed collections from *ex situ* field collections add unique genetic variation to the *ex situ* seed bank. These collections are also easily accessible and their storage in an *ex situ* seed bank additionally safeguards the diversity present in the *ex situ* field collections. The seed collections from *ex situ* field collections are interesting for genetic or taxonomic research, while our results suggest that the contribution of these seed collections to plant breeding might be limited if plant material from natural populations is available. The seed collections from feral populations provide safety backups for genetic resources in the seed collections from natural populations. A small number of seeds is probably sufficient to conserve the genetic diversity in feral populations. Nevertheless, the number of seed collections available for this study was limited to seven and the collection and characterization of additional seed material is needed to validate our results.

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/S1479262119000376.

Acknowledgements

We acknowledge Matthew Turner for providing the seed collection of Amami and Nicolas Roux for facilitating the transfer of seed collections for genetic analyses. Special thanks also go to the colleagues Kevin Longin and Tom Vanderstraeten at the International Transit Center who performed embryo rescue on the banana seeds and reared all embryos and to Wim Baert and Alexia Semeraro from Meise Botanic Garden for their support in the lab. Finally, we owe many words of gratitude to Professor Olivier Hardy and his research team at the University of Brussels (ULB) for their hospitality and for sharing their experiences in the analysis of microsatellite markers. Part of this work was funded by the Genebank CGIAR Research Program, the CGIAR Research Program on Roots, Tubers and Bananas (RTB) with a contribution of the Belgium government (DGD) through the PhenSeeData project, by Research Foundation - Flanders (FWO) (No. G0D9318N) and by the National Natural Science Foundation of China (No. 31261140366). In addition, this work was supported by the project 'BBTV mitigation: Community management in Nigeria, and screening wild banana progenitors for resistance (2015–2020)', funded by the Bill and Melinda Gates foundation. The funding agencies were not involved in the design of the study, in the provision or analysis of the data or in the writing and submission of the manuscript.

References

Afele JC and de Langhe E (1991) Increasing in vitro germination of *Musa balbisiana* seed. *Plant Cell, Tissue and Organ Culture* 27: 33–36.

Brown AHD (1989) Core collections: a practical approach to genetic resources management. *Genome* 31: 818–824.

Cipriani G, Spadotto A, Jurman I, Di Gaspero G, Crespan M, Meneghetti S, Frare E, Vignani R, Cresti M, Morgante M,

Pezzotti M, Pe E, Policriti A and Testolin R (2010) The SSR-based molecular profile of 1005 grapevine (*Vitis vinifera* L.) accessions uncovers new synonymy and parentages, and reveals a large admixture amongst varieties of different geographic origin. *Theoretical and Applied Genetics* 121: 1569–1585.

De Beukelaer H and Davenport G (2018) corehunter: Multi-Purpose Core Subset Selection. R package version 3.2.1. https://CRAN.R-project.org/package=corehunter.

Dempewolf H, Baute G, Anderson J, Kilian B, Smith C and Guarino L (2017) Past and future use of wild relatives in crop breeding. *Crop Science* 57: 1070–1082.

Doyle JJ and Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

FAO (2019) FAOSTAT. Retrieved on 20 February 2019, from http://www.fao.org/faostat/en/#data/QC.

Ford-Lloyd BV, Schmidt M, Armstrong SJ, Barazani O, Engels J, Hadas R, Hammer K, Kell SP, Kang D, Khoshbakht K, Li Y, Long C, Lu B-R, Ma K, Nguyen VT, Qiu L, Ge S, Wei W, Zhang Z and Maxted N (2011) Crop wild relatives – undervalued, underutilized and under threat? *BioScience* 61: 559–565.

Fu (2017) The vulnerability of plant genetic resources conserved *ex situ*. *Crop Science* 57: 2314–2328.

Ge XJ, Liu H, Wang K, Schaal BA and Chiang TY (2005) Population structure of wild bananas, *Musa balbisiana*, in China determined by SSR fingerprinting and cpDNA. *Molecular Ecology* 14: 933–944.

Heywood V, Casas A, Ford-Lloyd B, Kell S and Maxted N (2007) Conservation and sustainable use of crop wild relatives. *Agriculture, Ecosystems and Environment* 121: 245–255.

Holleley CE and Geerts PG (2009) Multiplex Manager 1.0: a cross-platform computer program that plans and optimizes multiplex PCR. *BioTechniques* 46: 511–517.

Janssens SB, Vandelook F, De Langhe E, Verstraete B, Smets E, Vandenhouwe I and Swennen R (2016) Evolutionary dynamics and biogeography of Musaceae reveal a correlation between the diversification of the banana family and the geological and climatic history of Southeast Asia. *New Phytologist* 210: 1453–1465.

Jayaweera SLD and Samarasinghe WLG (2016) Genetic diversity and population structure of wild banana (*Musa balbisiana*) populations in Sri Lanka. *Acta Horticulturae* 1114: 53–60.

Kearse M, Moir R, Wilson A, Stones-havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P and Drummond A (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)* 28: 1647–1649.

Kissel E, Van Asten P, Swennen R, Lorenzen J and Carpentier SC (2015) Transpiration efficiency *versus* growth: exploring the banana biodiversity for drought tolerance. *Scientia Horticulturae* 185: 175–182.

Li Y and Liu J (2018) Structureselector: a web-based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology Resources* 18: 176–177.

Li D and Pritchard HW (2009) The science and economics of *ex situ* plant conservation. *Trends in Plant Science* 14: 614–621.

Marita JM, Rodriguez JM and Nienhuis J (2000) Development of an algorithm identifying maximally diverse core collections. *Genetic Resources and Crop Evolution* 47: 515–526.

Meissen JC, Galatowitsch SM and Cornett MW (2015) Risks of overharvesting seed from native tallgrass prairies. *Restoration Ecology* 23: 882–891.

Milner SG, Jost M, Taketa S, Mazón ER, Himmelbach A, Oppermann M, Weise S, Knüpffer H, Basterrechea M, König P, Schüler D, Sharma R, Pasam RK, Rutten T, Guo G, Xu D, Zhang J, Herren G, Müller T, Krattinger SG, Keller B, Jiang Y, González MY, Zhao Y, Habekuß A, Färber S, Ordon F, Lange M, Börner A, Graner A, Reif JC, Scholz U, Mascher M and Stein N (2019) Genebank genomics highlights the diversity of a global barley collection. *Nature Genetics* 51: 319–326.

Ocan D, Mukasa HH, Rubaihayo PR, Tinzaara W and Blomme G (2008) Effects of banana weevil damage on plant growth and yield of East African *Musa* genotypes. *Journal of Applied Biosciences* 9: 407–415.

Peakall R and Smouse PE (2012) Genalex 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics* 28: 2537–2539.

Perrier X and Jacquemoud-Collet JP (2006) DARwin software. Retrieved from http://darwin.cirad.fr/darwin.

Perrier X, Flori A and Bonnot F (2003) Data analysis methods. In: Hamon P, Seguin M, Perrier X and Glaszmann JC (eds) *Genetic Diversity of Cultivated Tropical Plants*. Enfield: Science Publishers, pp. 43–76.

Perrier X, De Langhe E, Donohue M, Lentfer C, Vrydaghs L, Bakry F, Carreel F, Hippolyte I, Horry J, Jenny C, Lebot V, Risterucci A, Tomekpe K, Doutrelepont H, Ball T, Manwaring J, de Maret P and Denham T (2011) Multidisciplinary perspectives on banana (*Musa* spp.) domestication. *PNAS* 108: 11311–11318.

Pritchard JK, Stephens M and Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

Puechmaille SJ (2016) The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources* 16: 608–627.

R Core Team (2018) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. URL: https://www.R-project.org/.

Rivière S and Müller JV (2017) Contribution of seed banks across Europe towards the 2020 Global Strategy for Plant Conservation targets, assessed through the ENSCONET database. *Oryx* 52: 464–470.

Rotchanapreeda T, Wongniam S, Swangpol SC, Chareonsap PP, Sukkaewmanee N and Somana J (2016) Development of SSR markers from *Musa balbisiana* for genetic diversity analysis among Thai bananas. *Plant Systematics and Evolution* 302: 739–761.

Schoen DJ and Brown AHD (1993) Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. *PNAS* 90: 10623–10627.

Schoen DJ and Brown AHD (2001) The conservation of wild plant species in seed banks. *BioScience* 51: 960–966.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18: 233–234.

Stotzky G, Cox EA and Goos RD (1962) Seed germination studies in *Musa*. 1. Scarification and aseptic germination of *Musa balbisiana*. *American Journal of Botany* 49: 515–520.

Stover RH and Simmonds NW (1987) *Bananas (Tropical Agricultural Series)*. New York: Longman, p. 468.

Thachuk C, Crossa J, Franco J, Dreisigacker S, Warburton M and Davenport GF (2009) Core Hunter: an algorithm for sampling

genetic resources based on multiple genetic measures. *BMC Bioinformatics* 10: 1–13.

Thomas DS, Turner DW and Eamus D (1998) Independent effects of the environment on the leaf gas exchange of three banana (*Musa* sp.) cultivars of different genomic constitution. *Scientia Horticulturae* 75: 41–57.

Vanden Abeele S, Hardy OJ and Janssens SB (2018) Isolation of microsatellite loci in the African tree species *Staudtia kamerunensis* (Myristicaceae) using high-throughput sequencing. *Molecular Biology Reports* 45: 1539–1544.

van Hintum TJL and Visser DL (1995) Duplication within and between germplasm collections. II. *Genetic Resources and Crop Evolution* 42: 135–145.

Wang X-L, Chiang T-Y, Roux N, Hao G and Ge X-J (2007) Genetic diversity of wild banana (*Musa balbisiana* Colla) in China as revealed by AFLP markers. *Genetic Resources and Crop Evolution* 54: 1125–1132.

Wang J-Y, Huang B-Z, Chen Y-Y, Feng S-P and Wu Y-T (2011) Identification and characterization of microsatellite markers from *Musa balbisiana*. *Plant Breeding* 130: 584–590.