

RESEARCH ARTICLE

Open Access



Maximizing genetic representation in seed collections from populations of self and cross-pollinated banana wild relatives

Simon Kallow^{1,2,3*} , Bart Panis^{1,4} , Dang Toan Vu⁵ , Tuong Dang Vu⁵ , Janet Paofa⁶, Arne Mertens^{1,2,7} , Rony Swennen^{1,8}  and Steven B. Janssens^{2,7} 

Abstract

Background: Conservation of plant genetic resources, including the wild relatives of crops, plays an important and well recognised role in addressing some of the key challenges faced by humanity and the planet including ending hunger and biodiversity loss. However, the genetic diversity and representativeness of ex situ collections, especially that contained in seed collections, is often unknown. This limits meaningful assessments against conservation targets, impairs targeting of future collecting and limits their use.

We assessed genetic representation of seed collections compared to source populations for three wild relatives of bananas and plantains. Focal species and sampling regions were *M. acuminata* subsp. *banksii* (Papua New Guinea), *M. balbisiana* (Viet Nam) and *M. maclayi* s.l. (Bougainville, Papua New Guinea). We sequenced 445 samples using suites of 16–20 existing and newly developed taxon-specific polymorphic microsatellite markers. Samples of each species were from five populations in a region; 15 leaf samples from different individuals and 16 seed samples from one infructescence ('bunch') were analysed for each population.

Results: Allelic richness of seeds compared to populations was 51, 81 and 93% (*M. acuminata*, *M. balbisiana* and *M. maclayi* respectively). Seed samples represented all common alleles in populations but omitted some rarer alleles. The number of collections required to achieve the 70% target of the Global Strategy for Plant Conservation was species dependent, relating to mating systems. *Musa acuminata* populations had low heterozygosity and diversity, indicating self-fertilization; many bunches were needed (> 15) to represent regional alleles to 70%; over 90% of the alleles from a bunch are included in only two seeds. *Musa maclayi* was characteristically cross-fertilizing; only three bunches were needed to represent regional alleles; within a bunch, 16 seeds represent alleles. *Musa balbisiana*, considered cross-fertilized, had low genetic diversity; seeds of four bunches are needed to represent regional alleles; only two seeds represent alleles in a bunch.

Conclusions: We demonstrate empirical measurement of representation of genetic material in seeds collections in ex situ conservation towards conservation targets. Species mating systems profoundly affected genetic representation in seed collections and therefore should be a primary consideration to maximize genetic representation. Results are applicable to sampling strategies for other wild species.

Keywords: Conservation strategy, Sampling, Crop wild relatives, Seed bank, Genetic diversity

* Correspondence: S.Kallow@Kew.org

¹Department of Biosystems, Katholieke Universiteit Leuven, Willem de Croylaan 42, 3001 Leuven, Belgium

²Meise Botanic Garden, Nieuwelaan 38, 1860 Meise, Belgium

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Conservation of crop wild relatives (CWRs), wild plant species related to crops, is increasingly recognized as a vital component of both sustainable development for food security (Target 2.5 of the Sustainable Development Goals) [1] and biodiversity conservation (Target 9 of the Global Strategy for Plant Conservation) [2, 3]. Importantly, this should include targeting conservation at the intraspecific level [4], essential for the functioning and flourishing of species, ecosystems [5] and crop breeding [6]. Associated with policy recognition is the need for assessments against indicators or targets. However, assessment of conservation at the genetic level is often lacking and poorly understood [4, 7].

Conservation of CWRs should complementarily include both in situ and ex situ approaches [8]. Ex situ seed conservation can maintain numerous genotypes with minimal input [9]. However, knowledge of the genetic representativeness in ex situ seed collections, the proportion of alleles of wild populations also present in ex situ collections, has only been studied for a very small number of species. In fact, a recent meta-analysis of ex situ and in situ genetic comparisons only found six studies to include from seed bank collections [10]. Only two of these were of wild rather than cultivated species: a Mediterranean aquatic [11], and a temperate dioecious European tree species [12]. There is clearly, therefore, an evidence gap for reporting on the genetic representation of species in ex situ seed collections.

Presently, for seed collectors to maximise genetic capture in collections, sampling guidance is often broad, encompassing all species, or inferred from taxonomically or ecologically related species [13–15]. The general nature of such protocols does not always account for several key factors that shape genetics of populations and seeds, such as the existing genetic diversity of populations, the spatial distribution of plants in the environment [16–18], and species' reproductive systems [14]. It is therefore important to increase evidence-based sampling strategies to inform targeted future seed collections. Such evidence also provides valuable ecological information for in situ conservation and increases the value of seed collections, as it improves the selection and targeting of seed samples in breeding or phenotyping experiments.

Seed conservation of banana CWRs (*Musa* L.) is a case in point. Bananas, together with related plantains (both are *Musa*), are the most important fruit and among the most important crops in the world [19]. Global production is estimated to be 116 million tonnes annually, worth \$31 billion (average of 2017–19) [19]. Worryingly, several biotic threats, such as by Fusarium Wilt Tropical Race 4 and Banana Bunchy Top Virus, threaten banana production. The small genepool of bananas make them,

and the many millions of people who rely on them, particularly vulnerable [20, 21]. There are around 80 species in the genus *Musa* [22]. They are tall herbaceous monocarpic monocotyledons native to tropical and subtropical Asia and the western Pacific. Most cultivated bananas and plantains derive from two species: *M. acuminata* subsp. and *M. balbisiana* [23–26]. The Fe'i bananas of Pacific regions are a distinct cultivated group, deriving from *M. maclayi* [27]. The focal species included in our study (Table 1), are therefore of interest to breeders (e.g. [40]).

Conservation of banana CWRs is increasingly important because they are under threat. In a recent study [41], 15% of species were provisionally assessed as endangered and an additional 19% vulnerable to extinction. Furthermore, 95% of *Musa* species were assessed as insufficiently conserved ex situ [41]. There are only 163 genotype accessions of 35 species maintained in genebanks as living plants [42]. Additionally, there are 131 seed accessions, multiple seeds collected from the same individual or population, of 10 species, stored at the Millennium Seed Bank, UK [43]. Many *Musa* species are therefore not represented in genebanks at all or are represented with little or as yet unknown representativeness. An evaluation of the genetic representation of present collections will help target future conservation efforts.

The objectives of the present study are to assess and compare the genetic capture in seed collections compared to their source populations, at both regional and local scales, for three focal species; to provide guidance about how to maximize genetic capture for future seed collections; and to provide direction for seed distribution on how to provide representative seed samples.

Results

Representation of populations in seeds at the regional level

Allelic richness of seeds as a proportion of populations was 51, 81 and 93% (respectively *M. acuminata*, *M. balbisiana*, and *M. maclayi*, Table 2). Allelic richness (*AR*) of populations and seeds of *M. maclayi* was much higher than the other two species. In general, populations had many alleles that were private (*PA*); seeds had a few *PA* - indicating that some pollination occurred by plants not present in population samples. Only two alleles in *M. acuminata* seeds were private.

For all species, populations were characterized by having more rare alleles than seeds; and seeds had more common alleles (Fig. 1a). Seeds, therefore, captured most of the common alleles of populations, yet less so the rarer alleles. *Musa maclayi* had a notably high number of rare alleles in both populations and seeds.

Table 1 Taxonomic and ecological information on focal species

	<i>Musa acuminata</i> subsp. <i>banksii</i> (F.Muell.) N.W. Simmonds	<i>Musa balbisiana</i> Colla	<i>Musa maclayi</i> s.l. F.Muell.
Abbreviated name	<i>M. acuminata</i>	<i>M. balbisiana</i>	<i>M. maclayi</i>
Section (previous section)	Musa (Eumusa)	Musa (Eumusa)	Callimusa (Australimusa)
Chromosomes	2n = 22	2n = 22	2n = 20
CWR group [28]	1b (bananas and plantain)	1b (bananas and plantain)	1b (Fe'i bananas)
Distribution [22]	New Guinea, N.E. Queensland, Samoa	Sikkim to Papuaia	Papuasias
Habitat	Common in disturbed sites, rainforests and roadsides [29]	Forests, ravines and farm edges [30]	Alluvial gravels, lower montane habitats including forests and seral sites such as old gardens and landslips [29]
Elevation	To 1300 m [31]	To 1200 m [30]	To 1000 m [29]
Observed pollinator	Chiropterophily: <i>Syconycteris australis</i> , the Common blossom bat [32] Ornithophily: <i>Nectarinia jugularis</i> , the Yellow-bellied sunbird [32] Entomophily: Butterflies and small bees	Chiropterophily: fruit bats [33, 34] Entomophily: bumblebees	Ornithophily: Sunbirds (Nectariniidae) and Honeyeaters (Meliphagidae) [35]
Disperser	Not reported	Rodents [34]	Opossums, birds of paradise, hornbills, fruit bats
Flower form	Pendent [35]	Pendent [36]	Erect [36]
Notes on flower	Basal flowers hermaphrodite, male fertile [31, 37]	Basal flowers female with variable development of staminodes, hermaphrodite flowers never observed [29]	Basal flowers female only [29]
Country of collection	Papua New Guinea	Viet Nam	Autonomous Region of Bougainville (Papua New Guinea)
Ecoregion [38]	Northern New Guinea lowland rain and freshwater swamp forests	Northern Indochina subtropical forests	Solomon Islands rain forests
Annual mean temperature (°C) [39]	24.3 ± 2.1	21.4 ± 1.6	26.3 ± 1.3
Temperature annual range (°C) [39]	10.1 ± 2.4	18.7 ± 1.2	8.4 ± 0.4
Annual precipitation (mm) [39]	2561.0 ± 163.8	1613.8 ± 57.5	3534.2 ± 282.8
Precipitation seasonality (mm) [39]	28.4 ± 12.3	86.1 ± 6.6	15.6 ± 6.3

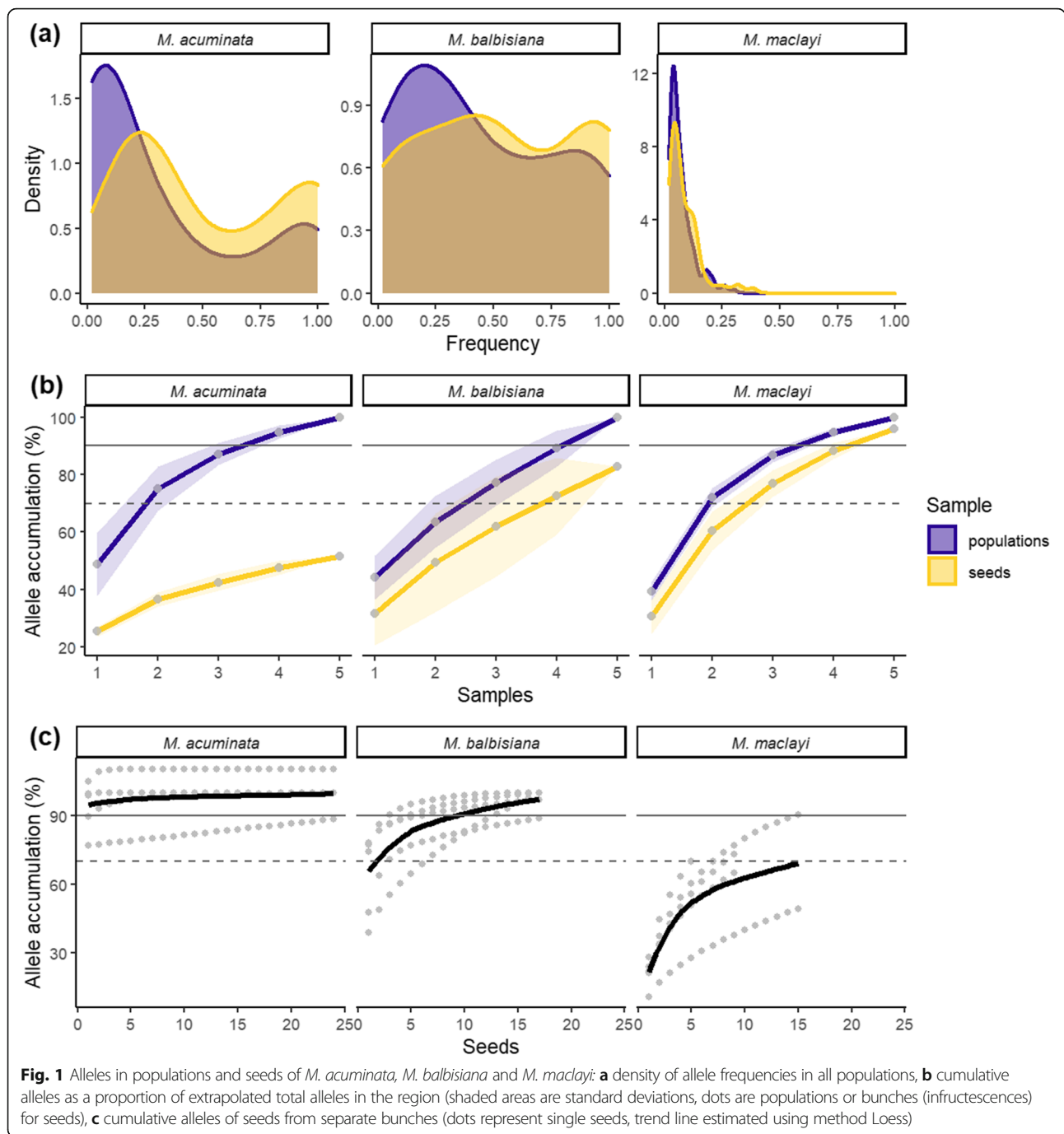
Diversity of *M. acuminata* seeds was significantly lower than that of populations (Shannon-Weiner diversity index (H'): $t = 4.595$, $df = 4.310$, $p = 0.008$; Simpson's index (λ): $t = 3.163$, $df = 4.018$, $p = 0.034$). *Musa balbisiana* seed

diversity was also significantly lower than populations (H' : $t = 2.890$, $df = 6.192$, $p = 0.0267$; λ : $t = 3.291$, $df = 7.908$, $p = 0.0112$). However, diversity of seeds and populations of *M. maclayi* was not significantly different (H' : $t = 0.716$, $df =$

Table 2 Diversity indices of populations and seeds pooled at the regional level^a

Species	Sample	N	AR	PA	MLG	H'	λ	E_5	H_{exp}	H_o
<i>M. acuminata</i>	Population	76	73	33	70	4.214	0.984	0.948	0.290	0.045
<i>M. acuminata</i>	Seeds	104	37	2	25	2.704	0.905	0.684	0.238	0.023
<i>M. balbisiana</i>	Population	83	48	16	82	4.402	0.988	0.993	0.314	0.336
<i>M. balbisiana</i>	Seeds	73	39	6	60	4.029	0.981	0.919	0.227	0.083
<i>M. maclayi</i>	Population	57	242	27	55	3.994	0.981	0.981	0.897	0.723
<i>M. maclayi</i>	Seeds	52	224	8	52	3.951	0.981	1.000	0.872	0.750

^aKey: N number of samples, AR rarefied allelic richness, PA private alleles, MLG number of detected multilocus genotypes, H' Shannon-Weiner diversity index, λ Simpson's index, E_5 evenness index, H_{exp} Nei's gene diversity (expected heterozygosity) [44], H_o observed heterozygosity



5.566, $p = 0.503$; λ : $t = 0.817$, $df = 5.216$, $p = 0.450$). Observed heterozygosity (H_o) was also significantly lower in seeds compared to populations for *M. acuminata* ($t = 251$, $df = 186.17$, $p = 0.026$), and *M. balbisiana* ($y = 4.720$, $df = 84.176$, $p < 0.001$), but not for *M. maclayi* ($t = 1.216$, $df = 137.97$, $p = 0.226$).

Species level differences

Genetic profiles were characteristically distinct according to species. Loci polymorphism for *M. acuminata* was on

average 3.84, for *M. balbisiana* 6.18 and for *M. maclayi* 18.64 (Table S3). In all cases H_o was less than expected heterozygosity (Nei’s gene diversity, H_{exp}) apart from *M. balbisiana* populations where they were approximately equal. The disparity was especially evident for *M. acuminata* population and seeds, and *M. balbisiana* seeds. Evenness (E_5) was high (> 0.9) across all sampling groupings but less so for *M. acuminata* seeds ($E_5 = 0.684$). *M. acuminata* seeds and populations had the lowest diversity in all sampling groups. Fewer multilocus genotypes

were observed in *M. acuminata* populations and, to a greater extent, seeds, compared to the number of actual samples - suggesting clonality or self-fertilization of homozygous mother plants; this was also observed in *M. balbisiana* seeds but to a lesser extent. Heterozygosity was very low in *M. acuminata* and *M. balbisiana*, but high for *M. maclayi*. There was no evidence of null allele excess, large allele drop out and error due to stuttering in *M. balbisiana* or *M. maclayi*. Six out of the 19 loci for *M. maclayi* showed potential null allele excess, possibly inflating homozygosity beyond predicted values, however, homozygosity was high across all loci (Table S4).

Representation of populations in seeds at the local level

Allelic richness of local seeds as a proportion of the local populations from where they were collected was $56 \pm 20\%$, $76 \pm 42\%$ and $78 \pm 18\%$ (mean and standard deviation, *M. acuminata*, *M. balbisiana* and *M. maclayi* respectively, Table 3).

M. acuminata seed collections had very low H_o including two bunches (infructescences) where H_o was zero (Table 3). H_{exp} varied considerably in local populations of *M. acuminata*. Nuru was the most diverse ($H_{exp} = 0.36$), notably seeds from this population were completely homozygous ($H_o = 0$). Sandaun was the least diverse *M. acuminata* population ($H_{exp} = 0.06$). Inbreeding coefficient (F_{is}) was high for all *M. acuminata* populations, Vanimo being the most inbred ($F_{is} = 0.97$), its seeds were also completely homozygous. The Ramu population was the least inbred ($F_{is} = 0.41$), and had seeds with the highest heterozygosity ($H_o = 0.05$). *Musa balbisiana* populations are also characterized by a low degree of diversity, yet inbreeding coefficients were much lower compared to *M. acuminata*. By contrast, populations of *M. maclayi* were characterized by a high level of heterozygosity and genetic diversity and low inbreeding coefficients. Populations of *M. balbisiana* and seeds of *M. maclayi* had negative F_{is} meaning an excess of heterozygotes.

Targeting local seed collections

Most variance found in population samples was within local populations rather than between local populations, according to AMOVA (*M. acuminata* 70%, *M. balbisiana* 75%, *M. maclayi* 82%). This means that in order to maximize genetic capture in seed collections, targeting local populations is less important than increasing the number of bunches collected. This also reflects the real experience in seed collecting because the genetic structure of subpopulations is usually not known at the time of collecting. Knowing how many local collections to make is therefore more informative.

Table 3 Diversity indexes of populations and seeds pooled at the local level^a

Species	Name	Sample	AR	H_o	H_{exp}	F_{is}
<i>M. acuminata</i>	Shungol	Population	35	0.033	0.225	0.880
<i>M. acuminata</i>	Nuru	Population	50	0.102	0.359	0.720
<i>M. acuminata</i>	Ramu	Population	48	0.076	0.268	0.417
<i>M. acuminata</i>	Sandaun	Population	25	0.014	0.056	0.603
<i>M. acuminata</i>	Vanimo	Population	40	0.008	0.267	0.976
<i>M. acuminata</i>	Shungol	Seeds	22	0.037	0.047	0.098
<i>M. acuminata</i>	Nuru	Seeds	19	0.000	0.000	0.000
<i>M. acuminata</i>	Ramu	Seeds	21	0.053	0.053	-0.009
<i>M. acuminata</i>	Sandaun	Seeds	22	0.007	0.007	0.000
<i>M. acuminata</i>	Vanimo	Seeds	19	0.000	0.000	0.000
<i>M. balbisiana</i>	Can Cau	Population	20	0.298	0.228	-0.318
<i>M. balbisiana</i>	Khe Ngau	Population	32	0.389	0.356	0.068
<i>M. balbisiana</i>	Muong Cau	Population	22	0.252	0.209	-0.116
<i>M. balbisiana</i>	Na Bo	Population	23	0.312	0.253	-0.195
<i>M. balbisiana</i>	Seo Leng	Population	27	0.392	0.332	-0.189
<i>M. balbisiana</i>	Can Cau	Seeds	30	0.183	0.230	0.118
<i>M. balbisiana</i>	Khe Ngau	Seeds	15	0.075	0.118	0.216
<i>M. balbisiana</i>	Murong Cau	Seeds	15	0.102	0.116	0.193
<i>M. balbisiana</i>	Na Bo	Seeds	14	0.021	0.053	0.462
<i>M. balbisiana</i>	Seo Leng	Seeds	15	0.070	0.072	0.155
<i>M. maclayi</i>	Aropa	Population	117	0.725	0.862	0.161
<i>M. maclayi</i>	Boku	Population	92	0.673	0.753	0.090
<i>M. maclayi</i>	Kangu	Population	113	0.737	0.847	0.133
<i>M. maclayi</i>	Kurai	Population	105	0.705	0.817	0.136
<i>M. maclayi</i>	Panguna	Population	112	0.737	0.805	0.089
<i>M. maclayi</i>	Aropa	Seeds	86	0.793	0.741	-0.073
<i>M. maclayi</i>	Boku	Seeds	91	0.722	0.680	-0.046
<i>M. maclayi</i>	Kangu	Seeds	109	0.737	0.698	-0.066
<i>M. maclayi</i>	Kurai	Seeds	64	0.829	0.746	-0.113
<i>M. maclayi</i>	Panguna	Seeds	69	0.699	0.665	-0.058

^aKey: AR rarefied allelic richness, H_o is observed heterozygosity, H_{exp} is genetic diversity (expected heterozygosity), F_{is} is the inbreeding coefficient [45]

To assess the cumulative addition of local seed collections as proportion of total allelic richness, total allelic richness of populations and seeds were estimated, and then the mean and standard error of each local population and bunch was added cumulatively. To capture 70% of alleles estimated to be present in the region, at least four bunches need to be collected for *M. balbisiana* (Fig. 1b); for 90% of alleles, at least five bunches are necessary. For *M. maclayi* three bunches are needed to sample 70% of regional AR, and four bunches for 90%, despite the much higher AR. Allelic sampling for *M. acuminata* had a different profile. It was not possible to collect

even 70% of regional *AR*, and for each bunch collected there was minimal gain in allelic capture. If many more bunches were sampled it may be possible to capture up-to 70% regional *AR*, but this would probably require >15 bunches, based on extrapolation (Fig. 1b).

Alleles in seeds of *M. acuminata* are largely shared by all bunches and by the regional population (Fig. 2). Each cumulative bunch adds only a few alleles. Populations with minimal shared alleles, and therefore maximum coverage, include combinations of Ramu and Vanimo, and Ramu and Nuru. *Musa balbisiana* displayed a similar overlapping pattern of shared alleles to *M. acuminata*, with very little gain each time a bunch was added. Seeds from the Can Cau population had the least shared alleles. Bunches from *M. maclayi* had less overlap, suggesting genetic structure and isolation by distance. No alleles in bunches were shared by all bunches, a large proportion of regional alleles was covered by bunch ellipses.

Genetic differentiation of local populations was detected using a permutation test on the AMOVA of population samples (*M. acuminata* $\phi = 0.29$, $p = 0.001$, *M. balbisiana* $\phi = 0.25$, $p = 0.001$, *M. maclayi* $\phi = 0.182$, $p = 0.001$). Genetic distance [46] was calculated pairwise between all local populations and local seeds (Fig. 3). For *M. acuminata* genetic distance was low between all samples. Populations and seeds from Vanimo were most distant from other samples. Seeds clustered with their respective populations for Vanimo and Sandaun, but not for other populations of *M. acuminata*. For *M. balbisiana*, seeds from Na Bo and Seo Leng were most distant. Notably the outlier population of Khe Ngau was not more distant from other populations and seeds. Several populations and seed pairs of *M. maclayi* clustered together. There were two broad clusters with samples from the North West (Boku and Panguna) having greater distance from the other three populations. Isolation by distance was evident in *M. maclayi* populations and seeds (Mantel test, 999 permutations; populations, $p = 0.03$; seeds $p = 0.013$), but not the other species.

Selecting seeds from the same local collection

Accumulation of *AR* of seeds per bunch was estimated (Fig. 1c). For *M. acuminata*, a single seed contained over 90% of the *AR* of the whole bunch. The allele accumulation curve is virtually flat, seeds are therefore more-or-less genetically identical. For *M. balbisiana* over 70% of alleles are found in only two seeds, and 90% of alleles in 10 seeds. For *M. maclayi*, 70% of the estimated total alleles in the bunch are captured by 16 seeds. To achieve 90% of the total the accumulation trend line must be extrapolated considerably beyond the data to around 35–50 seeds.

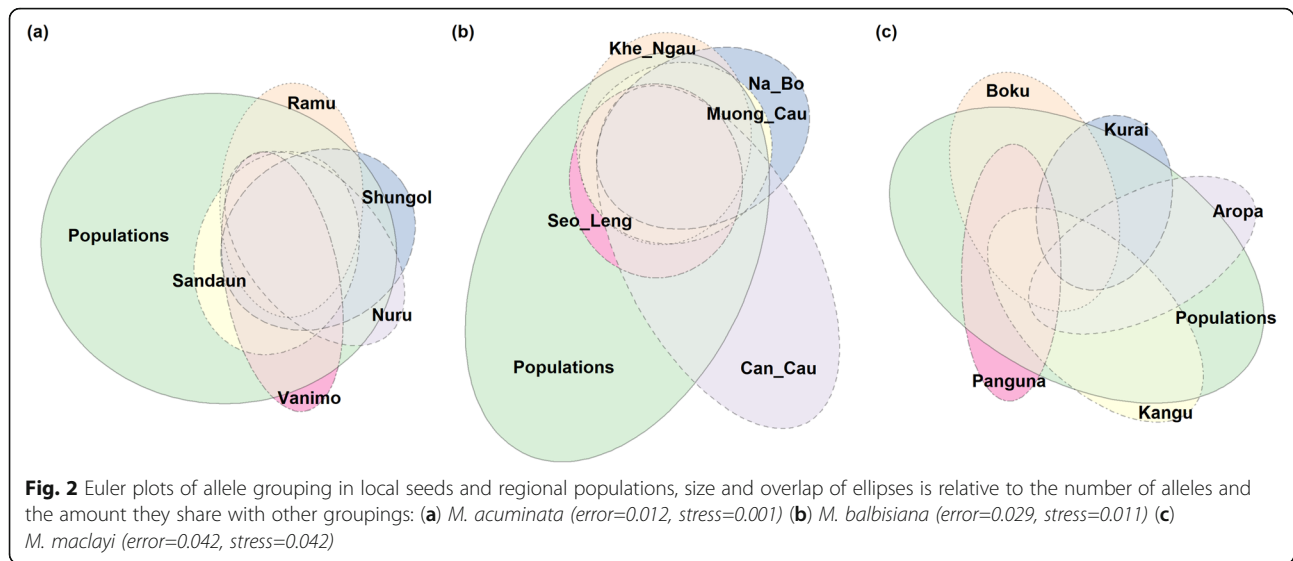
Discussion

Genetic capture in seed collections compared to their source populations, for three focal species

Allelic richness of seeds as a proportion of populations met the conservation target of 70% from the Global Strategy for Plant Conservation [2] for two out of the three focal species (*M. balbisiana* 81% and *M. maclayi* 93%). *M. acuminata* only achieved 51% proportional allelic richness. In all cases several seed collections were required from different local populations to maximize genetic capture. The number of collections necessary to achieve the 70% target was species dependent. For *M. acuminata* it was >15 local seed collections, for *M. balbisiana* four and *M. maclayi* three.

All common alleles of populations were included in seed collections, but the level of representation was lower for rare alleles, with some rare alleles missing from seed collections (Fig. 1a). Brown and Marshall's sampling strategy [15], used by about two thirds of leading seed conservation institutions [13], advocate sampling 30 individuals for out-crossing and 59 for selfing species, with the aim of having a 95% chance of capturing alleles with frequency <0.05. The results of the present study show that collecting from a much lower number of mother plants (a total of five) resulted in relatively high genetic capture, even including alleles rarer than the threshold set by Brown and Marshall. Furthermore, against the key success criteria proposed by Brown and Marshall [15] - capturing locally common alleles (because globally common alleles are easily collected in any sample and globally and locally rare alleles are ultimately limited by the sample size) - our results show that seed collections of *M. maclayi* and, to a lesser extent, *M. acuminata* and *M. balbisiana*, were successful (Figs. 2 and 3).

A high degree of homozygosity and low level of diversity, apparent in populations and seed collections of *M. acuminata* subsp. *banksii* in our study, as well as that of Christelova et al. [47], corroborate a typical genetic signal that is associated with self-fertilization [48]. Unlike cultivated and most wild banana species (including other *M. acuminata* species), seed bearing *M. acuminata* subsp. *banksii* are characterized by self-compatible hermaphroditic flowers, particularly in the upper hands on the inflorescence [31, 37], these likely self-pollinate by autogamy prior to flower bract opening [49]. Similar floral morphology, and therefore probable self-pollination, is also observed in *M. acuminata* var. *chinensis* [50], *M. boman* [29], *M. jackeyi* (interestingly closely related to *M. maclayi*) [31], *M. ingens* [29], *M. rubinea* [51], *M. schizocarpa* [29], *M. yunnanensis* [50] and *M. zainfui* [52]. Seed collections derived by self-fertilization are naturally more representative of the mother plant than the population. Therefore, to capture



the genetic diversity in populations of self-pollinating *Musaceae* species, many mother plants must be sampled.

By contrast, populations and seeds of *M. maclayi* were characterized by higher levels of heterozygosity and diversity, consistent with cross-fertilization [53]. Male and female flowers of *M. maclayi* are temporarily and physically isolated on the same inflorescence. Female flowers are produced first, followed by male flowers, as the peduncle grows [29]. Genetic capture in *M. maclayi* seeds therefore represent both the mother plant and pollen donors within the population. As a result, less bunches need to be collected to represent the population compared to *M. acuminata*. The level of diversity evident in the small number of populations of the present study was somewhat surprising, because of the narrow distribution and relatively recent diversification and dispersal of the former *Australimusa* group to which *M. maclayi* belongs [54]. This demonstrates the strong effect of mating system on genetic diversity in populations and seeds.

Populations of *M. balbisiana*, in our results, had low heterozygosity and diversity, this is in keeping with several previous studies [24, 34, 47, 55, 56]. Moreover, the heterozygosity of seed batches was much lower than within populations. Our results were similar to those found by Bawin et al. [57] for *M. balbisiana* seeds collected from ex situ field collections or feral populations, but our seeds were less diverse than those collected from native populations in Yunnan (China). Even though *M. balbisiana* basal flowers are functionally female [29] and do not produce seeds when pollinators are excluded [35], flowers may effectively be selfed from a different flower of the same genotype on the same mat or from vegetatively

reproduced or planted neighbouring plants [58]. Furthermore, apomictic seed development has been described in *M. acuminata* [59] and induced in *Ensete superbum* with pollen from *M. balbisiana* [60], and may additionally explain levels of homozygosity and apparent clonality in *M. acuminata* and *M. balbisiana* seeds observed.

The low diversity in populations of *M. balbisiana*, in our results, may be caused by a genetic bottleneck and/or founder effect. This hypothesis was also proposed by Ge et al. [34] and Shepherd [61] and is in keeping with *Musa* ecology: being early successional or disturbance-adapted [62, 63]. Additionally, the intensive deforestation and reforestation that has occurred in Viet Nam over the past 50 years [64] may also be causal. Indeed, according to a recent study [65], the ecological traits of *M. balbisiana* makes them particularly vulnerable to genetic erosion from anthropogenic disturbance. Furthermore, as *M. balbisiana* has many uses by local communities [66], plants are often planted or encouraged in vacant land. Finally, seed collections may indeed result from introgression from neighbouring cultivated bananas, as perhaps evident in the Can Cau population. These possibilities illustrate some of the challenges associated with conservation of CWRs by seed.

Variation in genetic capture of different species of the *Musa* genus demonstrates the profound effect of mating system on genetic capture in seed collection. Taxonomic relatedness, therefore, is not a good proxy for a sampling strategy [67]. In support of our results, a recent study by Hoban et al. [68] found species in the same genus required on average 50% more individuals to reach desired levels of capture compared to others. Furthermore, depending on mating system,

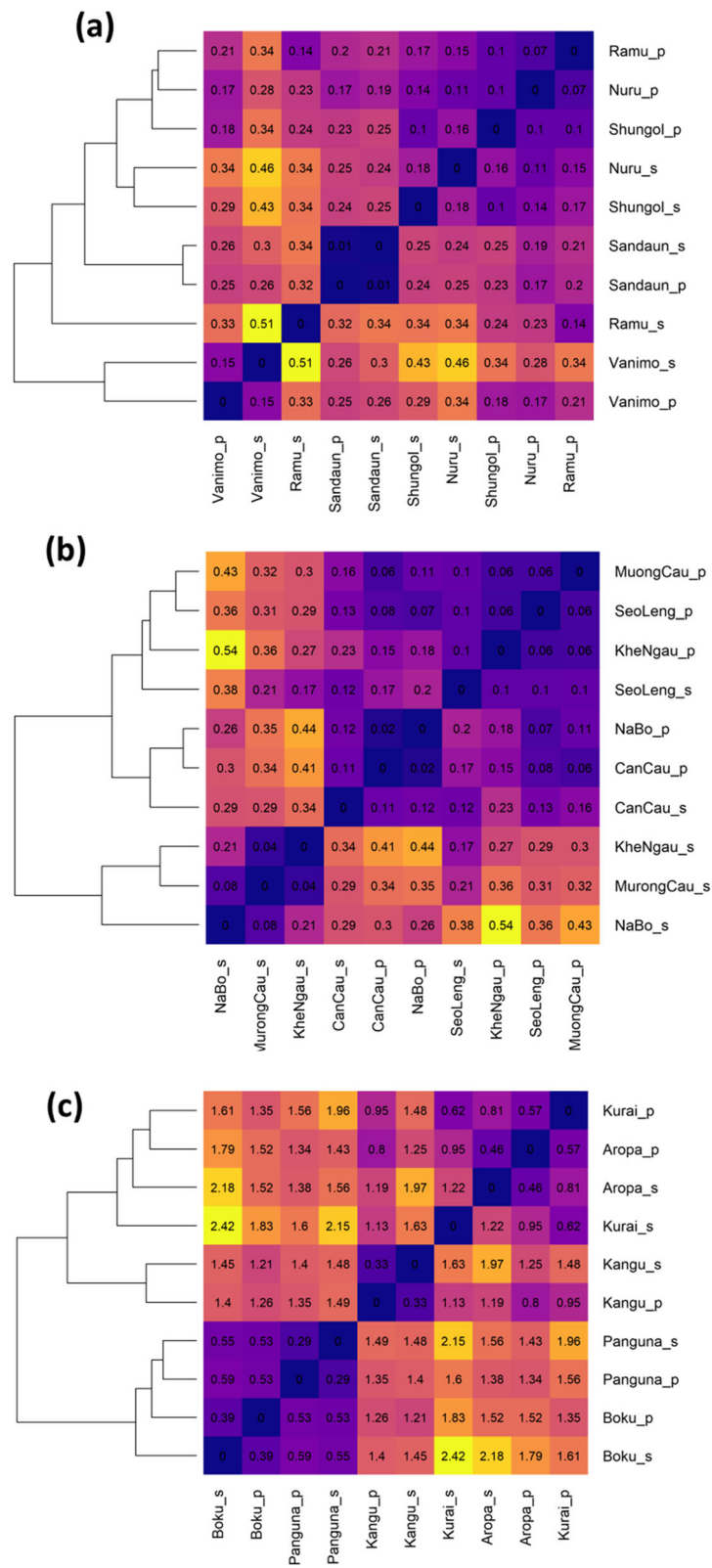


Fig. 3 Pairwise genetic distance (Nei, 1972) of populations and seeds, and clustered dendrograms using hierarchical clustering: **(a)** *M. acuminata* **(b)** *M. balbisiana* **(c)** *M. maclayi*

dispersal distance, life cycle and the sampling strategy employed - up to 5 times as many individuals may need to be sampled for the same level of genetic capture [14].

Guidance about how to maximize genetic capture for future seed collections

To maximize genetic capture in *Musa* seed collections, firstly, we recommend that species mating systems should be considered to inform sampling strategies. Our results are therefore in support of Brown and Marshall's sampling strategy discussed above [15].

For self-pollinated *Musa* species, as many mother plants should be sampled from as possible. For species with wide distributions, populations should be spatially dispersed; however this is less important than increasing the number of plants collected from. Collecting seeds from many individuals of adequate quality for long term storage is highly challenging; it is not straightforward to find mature seeds in the forest suitable for storage [69, 70]. It would certainly not be possible to collect from the 59 individuals proposed by Brown and Marshall [15], or even the 15 proposed here, in one collecting trip. As bananas produce fruit throughout the year, seed collections may therefore require repeated temporal sampling from populations.

To target collections of fully out-crossing species, fewer collections are required to represent regional alleles. We recommend collections should be focussed on increasing the number of local populations collected from rather than the number of mother plants in a population. Local populations should be spatially dispersed to maximize genetic capture. This will also allow for locally distributed alleles to be captured [15]. The amount of both rare and locally distributed alleles therefore depends on resources for collection, but there are diminishing returns associated with such effort.

For all species, but especially for out-crossing species, it is also important to target collections that are far from agriculture and human interference. Large and well established populations should be prioritised [65]. This will likely maximize genetic diversity in source populations [71], and avoid unwanted introgression from cultivated forms [72].

Direction for seed distribution on how to provide representative seed samples

To ensure enough seeds are conserved, self-pollinated species only require one or two seeds from a bunch to be part of a core collection. There is also very little point in using many samples of self-pollinated seeds in experiments. This contrasts with fully out-crossed seeds, where more seeds should be conserved in core collections per bunch or used as samples in experiments. For *M.*

maclayi 16 seeds represent 70% of alleles, and 35–50 seeds represent 90% of alleles. Even so, these numbers of seeds are easily achieved, for most *Musa* species at least, where a bunch can contain hundreds to thousands of seeds. However, for some species we have collected (e.g. *M. ingens*), only a few seeds were found in a bunch, perhaps due to inadequate pollination. Additionally, these findings mean that despite low levels of survival in storage of some collections [62, 69], population genetic diversity can be protected in a few seeds.

Limitations

The present study was constrained in that only one mother plant was used per local population, and only 5 per region. It was therefore not possible to test the effect of additional local seed collections on genetic capture. This was because accessing bunches at the right level of maturity for germination and storage is one of the key challenges for seed conservation of banana CWRs [69]; often mature bunches are not to be found in a forest population. Furthermore, in the present study we compared genetic capture in seed collections at the regional level. This does not account for the full level of diversity across species distributions which may be much wider than that sampled here, particularly in the case of *M. balbisiana* (Table 1). Further research should be done to assesses isolation by distance of source populations and seed genetic capture to optimise sampling strategies that use species distributions across ecozones as sampling strategies (e.g. [73]). Additionally, sampling did not consider temporal effects in sampling, such as collecting from the same populations at different time points, this may prove important, at least for cross-fertilized species.

It is also important to emphasize that whilst broad comparisons between species are of interest, direct comparison between species from our results should be cautioned because different taxon-specific microsatellite markers were employed. Observed allelic variation may indeed be resultant of specific markers used, rather than actual differences, meaningful at species level. However, as we used suites of 16–19 markers per species in the present study this effect is minimised, despite this, any comparative interpretation should be taken with caution. Importantly, direct comparisons between species was not our primary purpose, rather, our aims were to assess genetic capture in seed collections compared to their source populations for three focal species.

Conclusions

Seed banks are efficient ways of conserving genetic diversity present in wild populations and making it available for future use in breeding programmes or conservation. However, because very little is known about both population and seed genetic diversity the

representativeness and therefore the value and use of seed collections is limited. We have demonstrated the measurement of genetic capture in seed collections of three of the most important wild relatives of the most important fruit crop in the world. We have shown how targeted seed sampling should be species specific and genetically informed; notably, species mating systems and evolutionary history (whether natural or anthropogenic) have a profound effect on the level of genetic diversity in seed collections. The results of the present study may be applied in sampling strategies of other wild species, in that species mating systems should be a primary consideration to maximize genetic representation in seed collections.

Methods

Focal species

We focused on three wild *Musa* species: *M. acuminata* subsp. *banksii* (F.Muell.) N.W. Simmonds, *M. balbisiana* Colla and *M. maclayi* s.l. F.Muell. (termed *M. acuminata*, *M. balbisiana* and *M. maclayi* see Table 1). In this study *M. maclayi* includes closely related *M. bukensis* and *M. maclayi* subsp. *maclayi* taxa that occur on the island of Bougainville. Based on the description of both taxa and personal observations there is evidence of introgression between the two taxa on the island, and it is unclear whether they are two different, or one single, species [29].

Study region and populations

Natural populations of focal species in their respective native ranges were sampled during several collecting missions that took place between 2016 and 2019 (Fig. 4; Table 1). Collection of *M. acuminata* was carried out in Papua New Guinea (PNG) in June 2017 and May 2019 [69, 74]. *M. balbisiana* was collected in Viet Nam during November 2018 and April 2019. *Musa maclayi* was collected on the island of Bougainville (PNG) in October 2016 [75, 76].

Plant material

Leaf and seed samples were collected from wild natural populations. All seeds, leaves and data were collected and transferred according to local legislation, with permission and supplied for non-commercial use and research under the Standard Material Transfer Agreement in accordance with the International Treaty on Plant Genetic Resources for Food and Agriculture. None of the species included in the present study are CITES listed. Formal field identification was carried out by Steven B. Janssens (Meise Botanic Garden, Belgium). Leaf samples were collected randomly from 5 local populations per species (Fig. 4, Table S1). From each population, leaves from 15 plants on average were sampled and

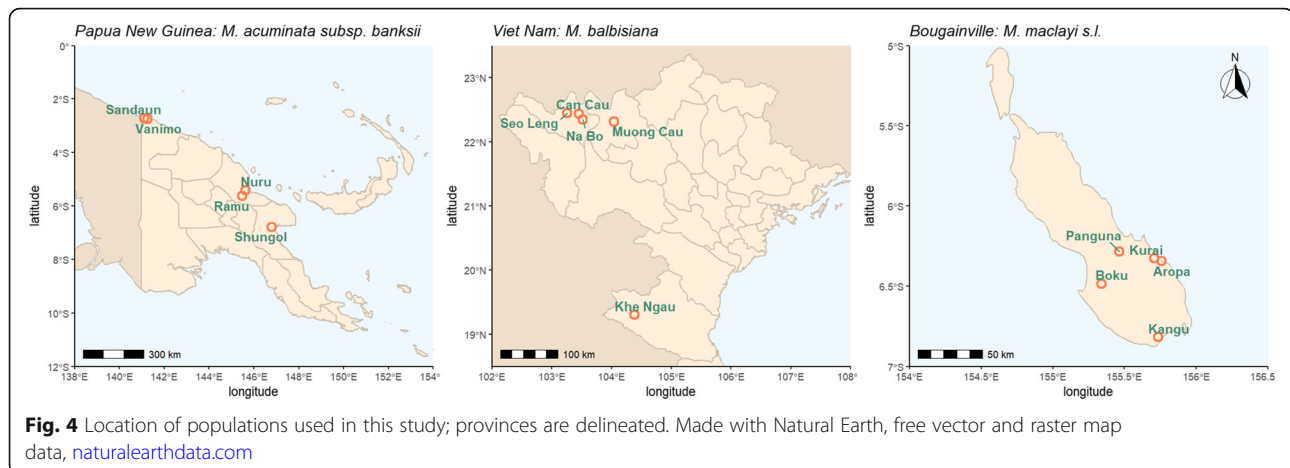
further used in this study. Dried leaf samples were taken to the laboratory following the field mission for DNA extraction. A single seed containing bunch (infructescence) was also collected from each population. Groups of fruits (hands) from the former clusters of flowers subtended by one bract, were separated and processed separately after shipping to Meise Botanic garden as described by Kallow et al. [69]. Bunches collected in Viet Nam were not separated by hand and were processed in a similar way in the laboratory of Plant Resource Center (Ha Noi, Viet Nam). In both cases, seeds were stored at 15% relative humidity and -20°C prior to germination and DNA extraction.

To overcome barriers associated with low and unpredictable *in vivo* germination, seeds were germinated by embryo rescue as described by Kallow et al. [6]. Seeds were selected randomly from two to three hands per bunch, or, for three bunches of *M. acuminata* and all bunches of *M. balbisiana*, from pooled seeds from the whole bunch. Due to low seed numbers and viability of *M. balbisiana* accessions DNA was extracted directly from their embryos. An average of 16 seeds per bunch were used in this study.

For each population, exact coordinates were recorded with a Garmin GPS device. Detailed taxonomic field notes, and notes on geography and ecology, were recorded for each sample. Photographs of mother plants (the plant from which the bunch was taken) and of bunches were taken. Seed samples from PNG and Bougainville were accessioned into the Meise Botanic Garden seed bank (Meise, Belgium). Seeds from Viet Nam were accessioned into the seed bank of Plant Resources Center (Ha Noi, Vietnam).

Microsatellite PCR

We isolated DNA using a method adapted from Doyle and Doyle [77] and then sequenced samples using a suite of taxon specific polymorphic microsatellite markers arranged in multiplexes (Table S2). For *M. acuminata* we developed multiplexes from previous studies [34, 78–83]. A total of 86 primer pairs were tested for amplification individually and then arranged in a total of 15 multiplexes using Multiplex Manager [84] and Multiple Primer Analyzer [85] with 12 *M. acuminata* samples. From this, 20 markers arranged in four multiplexes were selected. For *M. balbisiana*, we used the multiplex arrangement of Bawin et al. [57]. These included 18 SSR markers organized into four multiplexes. For *M. maclayi*, a total of 16 specific SSR markers were newly developed and optimized by Genoscreen (Lille, France) and arranged in four multiplexes. We used an M13 labelling protocol [86] to arrange multiplexes. We used the Type-it Microsatellite PCR Kit (Qiagen, Venlo, the Netherlands) to amplify microsatellite regions. We then



sequenced the resultant PCR product on an ABI 3730 sequencer (Applied Biosystems, Foster City, California, US). See [Supplementary Methods](#) for detailed methodology.

Data analysis

Fragment length analysis and quality check

We analyzed microsatellite fragment lengths using Geneious v 8.1.9 software. Loci and samples with more than 25% missing data were excluded from the analysis to allow for missing-ness to be similar for seeds and populations. This resulted in excluding from the data one locus used for *M. acuminata* data, eight for *M. balbisiana* and two for *M. maclayi* (Table S3). Several loci were missing from *M. balbisiana* presumably because of low DNA concentrations resultant of extraction from embryos rather than leaves. Resultant missing data was 3.9% for *M. acuminata*, 8.1% for *M. balbisiana* and 4.7% for *M. maclayi*. We then assessed allele data for null allele excess, large allele drop out and error due to stuttering using the Microchecker software [87].

Genetic assessment

Genetic assessment was carried out at two levels: the regional level whereby samples were pooled by either all local populations or all seeds per species; and the local level whereby samples were not pooled but kept separate from each local population and each bunch per species. At the regional level we calculated several indices to represent genetic diversity of populations and seeds. All computations were carried out in the R environment [88]. As a broad estimate of the amount of genetic material present, we determined *AR* rarefied to equal sample size [89], using the *pegas* package [90]. We counted *PA*, present in populations and not seeds and vice versa using the *poppr* package [91]; and, in order to assess the rarity of alleles in samples, assessed the relative

frequency of alleles, computed in the *adegenet* package [92]. To represent the genotypic diversity of samples and to assess inbreeding we calculated H_{exp} [44]. We also measured the H_o to assess population genotypic diversity and inbreeding. The number of MLG was computed, as an indicator of clonality. Several commonly used diversity indices were also calculated using the *poppr* package [91]: H' , λ and E_5 . At the local level, we repeated calculations of *AR*, H_o , H_{exp} and additionally calculated F_{IS} [45] in the *hierfstat* package [93]. Indices were compared using two-sample t tests.

Cumulative proportional allelic richness

We assessed the level of genetic variance between and within local populations by performing AMOVA on population samples. As most variance was within populations rather than between, we considered genetic capture could primarily be maximized by increasing the number of local seed collections made. We therefore measured how many bunches are required to capture 70% (based on Target 9 of the Global Strategy of Plant Conservation) [2] and 90% (an arbitrary but sometimes used threshold) of alleles in the region per species. We did this firstly by calculating the total regional population *AR* using bootstrap resampling [94]. Then the *AR* of local populations and bunches were estimated and the mean and standard deviations of these was added cumulatively using bootstrapping, separately for local populations and bunches. Estimates were normalised as percentages of total extrapolated regional *AR*.

A similar approach was employed to estimate proportional cumulative *AR* of seeds per bunch. For this, total *AR* of bunches was extrapolated [95], and then seeds were added cumulatively as described. Computations were made in the *vegan* package [96]. Trend lines were plotted using the loess method in *ggplot2* [97].

Allele groupings and genetic structure

An initial assessment of population differentiation was made by carrying out a permutation test (999 permutations) on the AMOVA described above. Secondly, We made allele groupings for local seeds and regional populations (<http://bioinformatics.psb.ugent.be/webtools/Venn>) and plotted them as Euler diagrams using the *eulerr* package [98]. Thirdly, we calculated pairwise genetic distance of local populations and seeds [46], and produced a heat map with dendrogram using complete linkage hierarchical clustering. Finally, we assessed isolation by distance by comparing Euclidean distances of coordinates and population matrices of seeds and populations (separately) using the Mantel test.

Abbreviations

AR: Allelic richness; CITES: Convention on International Trade in Endangered Species of Wild Fauna and Flora (1973); CWRs: Crop wild relatives; df: degrees of freedom; DNA: DeoxyriboNucleic Acid; E_s: Evenness index; F_{is}: Inbreeding coefficient; GPS: Global positioning system; H': Shannon-Weiner diversity index; H_{exp}: Nei's gene diversity (expected heterozygosity); H_o: Observed heterozygosity; MLG: Multilocus genotypes; *M. acuminata*: *M. acuminata* subsp. *banksii*; *Musa maclayi*: *Musa maclayi* s.l.; N: Number of samples; PA: Private alleles; PCR: Polymerase chain reaction; PNG: Papua New Guinea; Species: Species and infraspecifics; A: Simpson's index

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-021-03142-y>.

Additional file 1: Supplementary Methods.

Additional file 2: Table S1. Populations and samples used. **Table S2.** Microsatellite markers and multiplexes used. **Table S3.** Microsatellite summary statistics per locus. Number of alleles; 1-D, Simpson's index; H_{exp}, Nei's 1978 gene diversity; Evenness; species. **Table S4.** Null allele analysis for *Musa acuminata* loci.

Acknowledgements

We gratefully acknowledge those who carried out the seed and leaf collections. In PNG and Bougainville: J. Pilon, I. Nabo, B. Pitalai, G. Savi, P. Daur, E. Yabu, S. Itau, J. Lapiu, T. Kunou, S. Kambase, J. Guaf, N. Sinokor, S. Carpentier, J. Sardos, G. Sachter-Smith and D. Eyland. In Viet Nam: Le Thi Loan, Ngo Duc The. Thank you also to T. Vanderstraeten, K. Longin, N. Fanega Slezialek and H. Krohn for carrying out embryo rescue and DNA isolation. We are grateful for technical laboratory work and helpful advice from W. Baert, P. Asselman, Y. Bawin, A. Heylen and S. Vanden Abeele at Meise Botanic Gardens. We also gratefully acknowledge S. Hoban and R. Gargiulo for advice on the analysis and manuscript and J. Sardos for the basis of Table 1.

Authors' contributions

SK: Conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, writing – original draft, writing review & editing. BP: conceptualization, funding acquisition, project administration, resources, supervision, writing – review & editing. DTV: data curation, funding acquisition, resources. TDV: data curation, resources. JP: resources. AM: data curation, resources, validation, writing – review & editing. RS: validation, writing – review & editing. SBJ: conceptualization, funding acquisition, methodology, project administration, resources, supervision, validation, writing – review & editing. All authors have read and approved the final manuscript.

Funding

This work was funded as a sub-grant from the University of Queensland from the Bill & Melinda Gates Foundation project 'BBTV mitigation: Community management in Nigeria, and screening wild banana progenitors for

resistance' [OPP1130226]. This study was also funded by a bilateral grant between the Research Foundation - Flanders (FWO) and the Viet Nam National Foundation for Science and Technology Development (NAFOSTED) [FWO.106-NN.2017.02]. In addition, this study received funding from the Genbank CGIAR Research Program, from Research Foundation - Flanders (FWO) [G0 D9318 N]. The collection mission in PNG was funded by the Global TRUST foundation project "Crop wild Relatives Evaluation of drought tolerance in wild bananas from Papua New Guinea"[GS15024]. The authors thank all donors who supported this work also through their contributions to the CGIAR Fund (<http://www.cgiar.org/funders>), and in particular to the CGIAR Research Program Roots, Tubers and Bananas (RTB-CRP). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

All seeds, leaves and data were collected and transferred according to local legislation and supplied for non-commercial use and research under the Standard Material Transfer Agreement in accordance with the International Treaty on Plant Genetic Resources for Food and Agriculture. None of the species included in the present study are CITES listed.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Biosystems, Katholieke Universiteit Leuven, Willem de Croylaan 42, 3001 Leuven, Belgium. ²Meise Botanic Garden, Nieuwelaan 38, 1860 Meise, Belgium. ³Royal Botanic Gardens Kew, Millennium Seed Bank, Wakehurst, Ardingly, Sussex RH17 6TN, UK. ⁴Bioversity International, Willem de Croylaan 42, 3001 Leuven, Belgium. ⁵Plant Resources Center, An Khanh, Hoaiduc, Hà Noi, Viet Nam. ⁶National Agricultural Research Institute, Laloki, Port Moresby 121, Papua New Guinea. ⁷Department of Biology, Katholieke Universiteit Leuven, Kasteelpark Arenberg 31, 3001 Leuven, Belgium. ⁸International Institute of Tropical Agriculture, Plot 15B Naguru East Road, Upper Naguru, 7878 Kampala, Uganda.

Received: 11 March 2021 Accepted: 6 July 2021

Published online: 09 September 2021

References

- UN General Assembly. Transforming our world: the 2030 Agenda for Sustainable Development. 2015.
- Convention on Biological Diversity. Global strategy for plant conservation: 2011–2020. Richmond: Botanic Gardens Conservation International; 2012.
- Convention on Biological Diversity. Strategic plan for biodiversity 2011–2020, including Aichi biodiversity targets. 2010.
- Hoban SM, Bruford MW, D'Urban Jackson J, Lopes-Fernandes M, Heuertz M, Hohenlohe PA, et al. Genetic diversity targets and indicators in the CBD post-2020 global biodiversity framework must be improved. *Biol Conserv.* 2020;248:108654. <https://doi.org/10.1016/j.biocon.2020.108654>.
- Hughes AR, Inouye BD, Johnson MTJ, Underwood N, Vellend M. Ecological consequences of genetic diversity. *Ecol Lett.* 2008;11(6):609–23. <https://doi.org/10.1111/j.1461-0248.2008.01179.x>.
- Fernie AR, Tadmor Y, Zamir D. Natural genetic variation for improving crop quality. *Curr Opin Plant Biol.* 2006;9(2):196–202. <https://doi.org/10.1016/j.pbi.2006.01.010>.
- Garner BA, Hoban S, Luikart G. IUCN red list and the value of integrating genetics. *Conserv Genet.* 2020;21(5):795–801. <https://doi.org/10.1007/s10592-020-01301-6>.
- Engels JMM, Thormann I. Main challenges and actions needed to improve conservation and sustainable use of our crop wild relatives. *Plants.* 2020;9(8):968.

9. Li DZ, Pritchard HW. The science and economics of *ex situ* plant conservation. *Trends Plant Sci.* 2009;14(11):614–21. <https://doi.org/10.1016/j.tplants.2009.09.005>.
10. Wei X, Jiang M. Meta-analysis of genetic representativeness of plant populations under *ex situ* conservation in contrast to wild source populations. *Conserv Biol.* 2020;35:12–23.
11. Coppi A, Lastrucci L, Carta A, Foggi B. Analysis of genetic structure of *Ranunculus baudotii* in a Mediterranean wetland. Implications for selection of seeds and seedlings for conservation. *Aquat Bot.* 2015;126:25–31. <https://doi.org/10.1016/j.aquabot.2015.06.002>.
12. Gargiulo R, Saubin M, Rizzuto G, West B, Fay MF, Kallow S, et al. Genetic diversity in British populations of *Taxus baccata* L.: is the seedbank collection representative of the genetic variation in the wild? *Biol Conserv.* 2019;233:289–97. <https://doi.org/10.1016/j.biocon.2019.01.014>.
13. Hoban S, Way MJ. Improving the sampling of seeds for conservation. *Samara.* 2016;29:8–9.
14. Hoban S, Strand A. *Ex situ* seed collections will benefit from considering spatial sampling design and species' reproductive biology. *Biol Conserv.* 2015;187:182–91. <https://doi.org/10.1016/j.biocon.2015.04.023>.
15. Brown AHD, Marshall DR. A basic sampling strategy: theory and practice. In: Guarino L, Rao VR, Reid R, editors. *Collecting plant genetic diversity: technical guidelines.* Wallingford: CAB International; 1995. p. 75–91.
16. Di Santo LN, Hamilton JA. Using environmental and geographic data to optimize *ex situ* collections and preserve evolutionary potential. *Conserv Biol.* 2020;35:733–44.
17. Kashimshetty Y, Pelikan S, Rogstad SH. Effective seed harvesting strategies for the *ex situ* genetic diversity conservation of rare tropical tree populations. *Biodivers Conserv.* 2017;26:1311–31.
18. Hoban S, Schlarbaum S. Optimal sampling of seeds from plant populations for *ex-situ* conservation of genetic biodiversity, considering realistic population structure. *Biol Conserv.* 2014;177:90–9. <https://doi.org/10.1016/j.biocon.2014.06.014>.
19. FAO. FAOSTAT. Rome: Food and Agriculture Organisation of the United Nations; 2020.
20. Fones HN, Bebbler DP, Chaloner TM, Kay WT, Steinberg G, Gurr SJ. Threats to global food security from emerging fungal and oomycete crop pathogens. *Nat Food.* 2020;1(6):332–42. <https://doi.org/10.1038/s43016-020-0075-0>.
21. Altendorf S. Medium-term Outlook: Prospects for global production and trade in bananas and tropical fruits 2019 to 2028. Rome: FAO Food Outlook; 2019.
22. World Checklist of Musaceae. Royal Botanic Gardens, Kew. 2006 [cited 9th Jan 2020]. Available from: <http://wcp.science.kew.org/>.
23. De Langhe E, Perrier X, Donohue M, Denham T. The original banana split: multi-disciplinary implication of the generation of African and Pacific plantains in island Southeast Asia. *Ethnobot Res Appl.* 2015;14:299–312. <https://doi.org/10.17348/era.14.0.299-312>.
24. Hippolyte I, Jenny C, Gardes L, Bakry F, Rivallan R, Pomies V, et al. Foundation characteristics of edible *Musa* triploids revealed from allelic distribution of SSR markers. *Ann Bot.* 2012;109(5):937–51. <https://doi.org/10.1093/aob/mcs010>.
25. Martin G, Cardi C, Sarah G, Ricci S, Jenny C, Fondi E, et al. Genome ancestry mosaics reveal multiple and cryptic contributors to cultivated banana. *Plant J.* 2020;102(5):1008–25. <https://doi.org/10.1111/tpj.14683>.
26. Perrier X, Jenny C, Bakry F, Karamura D, Kitavi M, Dubois C, et al. East African diploid and triploid bananas: a genetic complex transported from South-East Asia. *Ann Bot.* 2019;123(1):19–36. <https://doi.org/10.1093/aob/mcy156>.
27. Ploetz RC, Kepler AK, Daniells J, Nelson SC. Banana and plantain—an overview with emphasis on Pacific island cultivars. *Spec Prof Pacific Island Agroforest.* 2007;1:21–32.
28. Maxted N, Ford-Lloyd BV, Jury S, Kell S, Scholten M. Towards a definition of a crop wild relative. *Biodivers Conserv.* 2006;15(8):2673–85. <https://doi.org/10.1007/s10531-005-5409-6>.
29. Argent GCG. The wild bananas of Papua New Guinea. *Notes Royal Botanic Gardens.* 1976;35:77–114.
30. Wu D, Larsen K. *Zingiberaceae, Flora of China, vol. 24.* Beijing and St. Louis: Science Press and Missouri Botanical Gardens Press; 2000.
31. Simmonds NW. Botanical results of the banana collecting expedition, 1954–5. *Kew Bull.* 1956;11(3):463–89. <https://doi.org/10.2307/4109131>.
32. Australian Government. The biology of *Musa* L. (banana): Department of Health. Canberra: Office of the Gene Technology Regulator; 2016.
33. Liu M, Ge X, Wang W, Hsu T, Schaal B, Chiang T. Pollen and seed dispersal of *Musa balbisiana* in South China. *Conserv Quart.* 2004;47:9–24.
34. Ge XJ, Liu MH, Wang WK, Schaal BA, Chiang TY. Population structure of wild bananas, *Musa balbisiana*, in China determined by SSR fingerprinting and cpDNA PCR-RFLP. *Mol Ecol.* 2005;14(4):933–44. <https://doi.org/10.1111/j.1365-294X.2005.02467.x>.
35. Nur N. Studies on pollination in Musaceae. *Ann Bot.* 1976;40(166):167–77. <https://doi.org/10.1093/oxfordjournals.aob.a085120>.
36. Boe A, Bortnem R, Johnson PJ. Changes in weight and germinability of black medic seed over a growing season, with a new seed predator. In: 101st Annual Meeting of South-Dakota-Academy-of-Science. Sioux Falls and Pierre: Univ Sioux Falls and South Dakota Acad Science; 2016. 2016 Apr 08–09.
37. Brewbaker JL, Gorrez DD. Classification of Philippine *Musa* III. (a) *Saguing matsing (Musa banksii* FvM); (b) *Alinsanay*, a putative hybrid of *M.texilis* and *M.banksii*. *Phil Agric.* 1956;40:258–68.
38. Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GV, Underwood EC, et al. Terrestrial ecoregions of the world: a new map of life on earth: a new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience.* 2001;51(11):933–8. [https://doi.org/10.1641/0006-3568\(2001\)051\[0933:TEOTWA\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2).
39. Fick SE, Hijmans RJ. Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol.* 2017;25:1965–78.
40. Zuo C, Deng G, Li B, Huo H, Li C, Hu C, et al. Germplasm screening of *Musa* spp. for resistance to *fusarium oxysporum* f. sp. *cubense* tropical race 4 (Foc TR4). *Eur J Plant Pathol.* 2018;151(3):723–34. <https://doi.org/10.1007/s10658-017-1406-3>.
41. Mertens A, Swennen R, Rønsted N, Vandeloek F, Panis B, Sachter-Smith G, et al. Conservation status assessment of banana crop wild relatives using species distribution modelling. *Divers Distrib.* 2021;27:729–46.
42. van den Houwe I, Chase R, Sardos J, Ruas M, Kempenaers E, Guignon V, et al. Safeguarding and using global banana diversity: a holistic approach. *CABI Agricult Biosci.* 2020;1(1):15. <https://doi.org/10.1186/s43170-020-00015-6>.
43. Genesys global portal on plant genetic resources. 2020 [cited 15th September 2020]. Available from: <https://www.genesys-pgr.org/>. Accessed 22 Jan 2020.
44. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics.* 1978;89(3):583–90. <https://doi.org/10.1093/genetics/89.3.583>.
45. Nei M. *Molecular evolutionary genetics.* New York: Columbia University Press; 1987.
46. Nei M. Genetic distance between populations. *Am Nat.* 1972;106(949):283–92. <https://doi.org/10.1086/282771>.
47. Christelová P, De Langhe E, Hřibová E, Čížková J, Sardos J, Hušáková M, et al. Molecular and cytological characterization of the global *Musa* germplasm collection provides insights into the treasure of banana diversity. *Biodivers Conserv.* 2017;26(4):801–24. <https://doi.org/10.1007/s10531-016-1273-9>.
48. Charlesworth D. Evolution of plant breeding systems. *Curr Biol.* 2006;16(17):R726–R35. <https://doi.org/10.1016/j.cub.2006.07.068>.
49. Amah D, Turner DW, Gibbs J, Waniale A, Gram G, Swennen R. Overcoming the fertility crisis in bananas (*Musa* spp.). In: Kema GH, Drenth A, editors. *Achieving sustainable cultivation of bananas. Volume 2: germplasm and genetic improvement.* Cambridge: Burleigh Dodds; 2020.
50. Häkkinen M, Hong W. New species and variety of *Musa* (Musaceae) from Yunnan, China. *Novon.* 2007;17(4):440–6. [https://doi.org/10.3417/1055-3177\(2007\)17\[440:NSAVOM\]2.0.CO;2](https://doi.org/10.3417/1055-3177(2007)17[440:NSAVOM]2.0.CO;2).
51. Häkkinen M, Teo CH. *Musa rubinea*, a new *Musa* species (Musaceae) from Yunnan, China. *Folia Malaysiana.* 2008;9(1):23–33.
52. Häkkinen M, Wang H. *Musa zaifui* sp nov (Musaceae) from Yunnan, China. *Nordic J Botany.* 2008;26(1–2):42–6. <https://doi.org/10.1111/j.0107-055X.2008.00267.x>.
53. Hamrick JL, Godt MJW. Effects of life history traits on genetic diversity in plant species. *Philos Trans R Soc Lond B Biol Sci.* 1996;351(1345):1291–8.
54. Janssens SB, Vandeloek F, De Langhe E, Verstraete B, Smets E, Vandenhoeve I, et al. 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 Phytol.* 2016;210(4):1453–65. <https://doi.org/10.1111/nph.13856>.
55. Jesus ON, Silva SD, Amorim EP, Ferreira CF, Campos JM, Silva GD, et al. Genetic diversity and population structure of *Musa* accessions in *ex situ* conservation. *BMC Plant Biol.* 2013;13(41):1–22 (12 March 2013).

56. Wang XL, Chiang TY, Roux N, Hao G, Ge XJ. Genetic diversity of wild banana (*Musa balbisiana* Colla) in China as revealed by AFLP markers. *Genet Resour Crop Evol.* 2007;54(5):1125–32. <https://doi.org/10.1007/s10722-006-9004-9>.
57. Bawin Y, Panis B, Vanden Abeele S, Li Z, Sardos J, Paofa J, et al. Genetic diversity and core subset selection in ex situ seed collections of the banana crop wild relative *Musa balbisiana*. *Plant Genet Res.* 2019;17(6):536–44. <https://doi.org/10.1017/S1479262119000376>.
58. Fortescue JA, Turner D. Reproductive biology. In: Pillay M, Tenkouano A, editors. *Banana breeding: progress and challenges*. Boca Raton: Taylor & Francis; 2011. p. 145–79.
59. Okoro P, Shaibu AA, Ude G, Oluokolu BA, Ingelbrecht I, Tenkouano A, et al. Genetic evidence of developmental components of parthenocarpy in apomictic *Musa* species. *J Plant Breed Crop Sci.* 2011;3(8):138–45.
60. Ravishanker KV, Ajitha-Kumar R, Malarvizhu M, Ambika DMS. Apomictic seed development in *Ensete superbum* induced by pollen of wild banana sp. *Musa balbisiana*. *Curr Sci.* 2011;101(4):493–5.
61. Shepherd K. Cytogenetics of the genus *Musa*. Montpellier: International Network for the Improvement of Banana and Plantain; 1999.
62. Kallow S, Davies R, Panis B, Janssens SB, Vandeloek F, Mertens A, et al. Regulation of seed germination by diurnally alternating temperatures in disturbance adapted banana crop wild relatives (*Musa acuminata*). *Seed Sci Res.* 2021;30:1–11.
63. Zhang G, Tang J, Shi J, Bai K. Niche dynamics of dominant populations of *Musa acuminata* Colla pioneer community in Xishuangbanna, SW China. *J Plant Resour Environ.* 2000;9(1):22–6.
64. Cochard R, Nguyen VHT, Ngo DT, Kull CA. Vietnam's forest cover changes 2005–2016: veering from transition to (yet more) transition? *World Dev.* 2020;135:105051. <https://doi.org/10.1016/j.worlddev.2020.105051>.
65. Almeida-Rocha JM, Soares LASS, Andrade ER, Gaiotto FA, Cazetta E. The impact of anthropogenic disturbances on the genetic diversity of terrestrial species: a global meta-analysis. *Mol Ecol.* 2020;29(24):4812–22. <https://doi.org/10.1111/mec.15688>.
66. Borborah K, Borthakur SK, Tanti B. *Musa balbisiana* Colla - taxonomy, traditional knowledge and economic potentialities of the plant in Assam, India. *Indian J Tradition Knowl.* 2016;15(1):116–20.
67. Griffith MP, Calonje M, Meerow AW, Francisco-Ortega J, Knowles L, Aguilar R, et al. Will the same ex situ protocols give similar results for closely related species? *Biodivers Conserv.* 2017;26(12):2951–66. <https://doi.org/10.1007/s10531-017-1400-2>.
68. Hoban S, Callicrate T, Clark J, Deans S, Dosmann M, Fant J, et al. Taxonomic similarity does not predict necessary sample size for ex situ conservation: a comparison among five genera. *Proc R Soc B-Biol Sci.* 2020;287(1926):9.
69. Kallow S, Longin K, Slezciak NF, Janssens SB, Vandeloek F, Dickie J, et al. Challenges for ex situ conservation of wild bananas: seeds collected in Papua New Guinea have variable levels of desiccation tolerance. *Plants.* 2020;9(9):1243. <https://doi.org/10.3390/plants9091243>.
70. Singh S, Agrawal A, Kumar R, Thangjam R, Joseph JK. Seed storage behavior of *Musa balbisiana* Colla, a wild progenitor of bananas and plantains - implications for ex situ germplasm conservation. *Sci Hortic.* 2021;280:109926. <https://doi.org/10.1016/j.scienta.2021.109926>.
71. Broadhurst LM, Lowe A, Coates DJ, Cunningham SA, McDonald M, Vesk PA, et al. Seed supply for broadscale restoration: maximizing evolutionary potential. *Evol Appl.* 2008;1(4):587–97. <https://doi.org/10.1111/j.1752-4571.2008.00045.x>.
72. Andersson MS, de Vincent MC. Gene flow between crops and their wild relatives. Baltimore: The John Hopkins University Press; 2010.
73. Khoury CK, Carver D, Barchenger DW, Barboza GE, van Zonneveld M, Jarret R, et al. Modelled distributions and conservation status of the wild relatives of Chile peppers (*Capsicum* L.). *Divers Distrib.* 2020;26(2):209–25. <https://doi.org/10.1111/ddi.13008>.
74. Eyland D, Breton C, Sardos J, Kallow S, Panis B, Swennen R, et al. Filling the gaps in gene banks: collecting, characterizing and phenotyping wild banana relatives of Papua New Guinea. *Crop Sci.* 2020;61:137–49.
75. Sachter-Smith G, Paufa J, Rauka G, Sardos J, Janssens S. Bananas of the autonomous region of Bougainville: a catalog of banana diversity seen on the islands of Bougainville and Buka, Papua New Guinea; 2017.
76. Sardos J, Christelova P, Cizkova J, Paofa J, Sachter-Smith GL, Janssens SB, et al. Collection of new diversity of wild and cultivated bananas (*Musa* spp.) in the autonomous region of Bougainville, Papua New Guinea. *Genet Resour Crop Evol.* 2018;65(8):2267–86. <https://doi.org/10.1007/s10722-018-0690-x>.
77. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 1987;19(1):11–5.
78. Lagoda PJJ, Dambier D, Grapin A, Baurens FC, Lanaud C, Noyer JL. Nonradioactive sequence-tagged microsatellite site analyses: a method transferable to the tropics. *Electrophoresis.* 1998;19(2):152–7. <https://doi.org/10.1002/elps.1150190204>.
79. Hippolyte I, Bakry F, Seguin M, Gardes L, Rivallan R, Risterucci AM, et al. A saturated SSR/DaIT linkage map of *Musa acuminata* addressing genome rearrangements among bananas. *BMC Plant Biol.* 2010;10:18.
80. Miller RN, Passos MA, Menezes NN, Souza MT, do Carmo Costa MM, Azevedo VC, et al. Characterization of novel microsatellite markers in *Musa acuminata* subsp. burmannicoides, var. Calcutta 4. *BMC Res Notes.* 2010;3(1):148.
81. D'Hont A, Denoeuf F, Aury JM, Baurens FC, Carreel F, Garsmeur O, et al. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature.* 2012;488(7410):213–20. <https://doi.org/10.1038/nature11241>.
82. Passos M, Cruz V, Emediato F, Teixeira C, Souza M, Matsumoto T, et al. Development of expressed sequence tag and expressed sequence tag-simple sequence repeat marker resources for *Musa acuminata*. *Aob Plants.* 2012.
83. Rotchanapreeda T, Wongniam S, Swangpol SC, Chareonsap PP, Sukkaewmanee N, Somana J. Development of SSR markers from *Musa balbisiana* for genetic diversity analysis among Thai bananas. *Plant Syst Evol.* 2016;302(7):739–61. <https://doi.org/10.1007/s00606-015-1274-2>.
84. Holleley CE, Geerts PG. Multiplex manager 1.0: a crossplatform computer program that plans and optimizes multiplex PCR. *Bio Techniques.* 2009;46(7):511–7.
85. ThermoFisher. Multiple Primer Analyzer 2020. Available from: <https://www.thermofisher.com/be/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html>.
86. Schuelke M. An economic method for the fluorescent labelling of PCR fragments. *Nat Biotechnol.* 2000;18(2):233–4. <https://doi.org/10.1038/72708>.
87. Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes.* 2004;4(3):535–8. <https://doi.org/10.1111/j.1471-8286.2004.00684.x>.
88. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2019.
89. Hurlbert SH. The nonconcept of species diversity: a critique and alternative parameters. *Ecology.* 1971;51:577–86.
90. Paradis E. Pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics.* 2010;26(3):419–20. <https://doi.org/10.1093/bioinformatics/btp696>.
91. Kamvar ZN, Tabina JF. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *Peer J.* 2014;2:e281.
92. Jombart T. Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics.* 2008;24(11):1403–5. <https://doi.org/10.1093/bioinformatics/btn129>.
93. Goudet J, Jombart T. Hierfstat: estimation and tests of hierarchical F-statistics v 0.5–7. Wein: CRAN; 2020.
94. Smith EP, van Belle G. Nonparametric estimation of species richness. *Biometrics.* 1984;40(1):119–29. <https://doi.org/10.2307/2530750>.
95. Chao A. Estimating the population size for capture-recapture data with unequal catchability. *Biometrics.* 1987;43(4):783–91. <https://doi.org/10.2307/2531532>.
96. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGinn D, et al. Vegan: community ecology package. R package version 2.3–6. Wein: CRAN; 2019.
97. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag; 2016. <https://doi.org/10.1007/978-3-319-24277-4>.
98. Larsson J. Eulerr: Area-proportional euler and venn diagrams with ellipses. R package v 6.1.0; 2020.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.