Theor Appl Genet (2016) 129:547–561
DOI 10.1007/s00122-015-2647-1

ORIGINAL ARTICLE

The triploid East African Highland Banana (EAHB) genepool is genetically uniform arising from a single ancestral clone that underwent population expansion by vegetative propagation

Mercy Kitavi[1,2] · Tim Downing[3] · Jim Lorenzen[2,4] · Deborah Karamura[5] · Margaret Onyango[6] · Moses Nyine[2] · Morag Ferguson[2] · Charles Spillane[1]

Received: 24 August 2015 / Accepted: 30 November 2015 / Published online: 8 January 2016

Abstract

Key message **All East African Highland Banana varieties are genetically uniform having arisen from a single clone introduced to Africa.**

Abstract East African Highland bananas (EAHBs) are a subgroup of triploid (AAA genome) bananas of importance to food security in the Great Lakes region of Africa. Little is known about their genetic variation, population structure and evolutionary history. Ninety phenotypically diverse EAHB cultivars were genotyped at 100 SSR microsatellite markers to investigate population genetic diversity, the correlation of genetic variability with morphological classes, and evolutionary origins since introduction to Africa. Population-level statistics were compared to those for plantain (AAB) and dessert (AAA) cultivars representing other *M. acuminata* subgroups. EAHBs displayed minimal genetic variation and are largely genetically uniform, irrespective of whether they were derived from the distinct Ugandan or Kenyan germplasm collections. No association was observed between EAHB genetic diversity and currently employed morphological taxonomic systems for EAHB germplasm. Population size dynamics indicated that triploid EAHBs arose as a single hybridization event, which generated a genetic bottleneck during foundation of the EAHB genepool. As EAHB triploids are sterile, subsequent asexual vegetative propagation of EAHBs allowed a recent rapid expansion in population size. This provided a basis for emergence of genetically near-isogenic somatic mutants selected across farmers and environments in East Africa over the past 2000 years since EAHBs were first introduced to the African continent.

Communicated by A. H. Schulman.

M. Onyango: Deceased (RIP).

Electronic supplementary material The online version of this article (doi:10.1007/s00122-015-2647-1) contains supplementary material, which is available to authorized users.

✉ Charles Spillane
charles.spillane@nuigalway.ie

[1] Genetics and Biotechnology Lab, Plant and AgriBiosciences Research Centre (PABC), School of Natural Sciences, C306 Aras de Brun, National University of Ireland Galway, University Road, Galway, Ireland

[2] International Institute for Tropical Agriculture (IITA), Biosciences Eastern and Central Africa (BecA-ILRI), P.O. Box 30709-00100, Nairobi, Kenya

[3] School of Mathematics, Statistics and Applied Mathematics, National University of Ireland Galway, University Road, Galway, Ireland

[4] Bill and Melinda Gates Foundation, 500 5th Ave N, Seattle, WA 98102, USA

[5] Bioversity International, PLOT 106, Katalima Road, P.O. Box 24384, Kampala, Uganda

[6] Kenya Agricultural Research Institute (KARI), KARI, Kisii Centre, P.O. Box 523-40200, Kisii, Kenya

Introduction

Bananas (genus *Musa*, family *Musaceae*) are monocotyledons from the *Zingiberales*, a sister group of the *Poales*

and of major economic importance (e.g., food security and income) in many tropical and subtropical countries (Martin et al. 2013). Though not well understood, the *Musa* domestication process is considered to have started some 7000 years ago in Southeast Asia (D'Hont et al. 2012). It is speculated that spontaneous hybridization and triploidization events involving almost all diploid cultivars lead to the formation of cultivated triploids, including pure *M. acuminata* varieties (AAA) and interspecific *M. acuminata* × *M. balbisiana* varieties (AAB, ABB) (Till et al. 2010; Perrier et al. 2011; Pachuau et al. 2014). Indeed, it is likely that sterility and parthenocarpy could have been selected for in triploids to generate seedless fruits with increased pulp mass, which gave rise to the current predominantly sterile triploid cultivars (Perrier et al. 2009, 2011; de Jesus et al. 2013).

The East African Highland bananas (EAHB; also known as the 'Lujugira-Mutika' subgroup) have a triploid A genome (AAA). The EAHBs consist of many farmer-selected landrace cultivars that dominate the Great Lakes region of East Africa (Karamura 1998) with no clear analogs globally (Pillay et al. 2001). The current diversity of this subgroup is believed to have resulted from the accumulation of somatic mutations that consistently recur in the crop—a condition that is still not well understood (De langhe 1961, 1964; Crane and Laurence 1956; Shepherd 1957; Ude et al. 2003). Based on the phenotypic diversity of the EAHBs, the Great Lakes region is considered as a secondary center of banana diversity (Tugume et al. 2003).

The edible diploid *M. acuminata* subspecies zebrina and banksii are considered to have a recent common ancestry with EAHB (Li et al. 2013). The *M. acuminata* subspecies banksii is also related to the AAB Plantains, which have a "B" genome ancestry from *M. balbisiana* (Perrier et al. 2011). Lack of historical records and archeological evidence has limited understanding of the movement of the EAHB genepool outwards from their center of origin (South East Papua New Guinea). However, the geographical ranges of extant *Musa acuminata* diploid parents and triploid hybrids support a scenario of numerous ancient migrations both before (in diploids) and after hybridization (in triploids). Current evidence suggests a single introduction of AA or AAA varieties 'Lujugira-Mutika' into Africa about 2500 years ago, which was distinct from introduction(s) contributing to the origins of AAB Plantains (Perrier et al. 2011).

Like most cultivated bananas, the EAHBs are strictly vegetatively propagated via suckers planted by farmers. The absence of seed propagation involving sexual recombination has resulted in cultivated bananas having a limited potential for producing genetically diverse offspring. Yet, morphological variation does exist within the EAHB subgroup which is typically considered to have arisen through the accumulation of somaclonal mutations leading to somaclonal variants (Perrier et al. 2011). In this regard, approximately 120 EAHB cultivars are recognized across Uganda, Kenya, western Tanzania, Rwanda, Burundi and eastern Congo (Karamura 1998; Ssebuliba et al. 2005), although about 200 different names are used as a number of local names may exist for the same cultivar.

Even though the morphotaxonomic system has provided a coherent classification basis for this banana subgroup (Simmonds and Shepherd 1955; Shepherd 1957; Karamura 1998), this system has inadequate resolution for populations within species (Christelova et al. 2011). Many EAHB cultivars cannot be easily distinguished on the basis of their morphology, especially if they are closely related (Pillay et al. 2001). Such morphological diversity measures are often inaccurate (Abdullah et al. 2012), where they can under- or over-estimate the extent of genetic variability (Resmi et al. 2011). An improved understanding of molecular genetic diversity of extant EAHB population can enable more informed and improved approaches to EAHB breeding (Ganapathy et al. 2012).

Pillay et al. (2001) used RAPDs as molecular markers to analyze the genetic diversity of 29 EAHBs and concluded that, while they could be distinguished, they were closely related with a narrow genetic base (Cooper et al. 2000). Using AFLP, Tugume et al. (2003) reported low levels of DNA diversity of 115 EAHB cultivars and his molecular classification somehow matched the morphological characterization of Karamura (1998). Other published studies have included EAHB cultivars to represent out-group taxa or as a genomic group (AAA) and have corroborated the above findings (Ude et al. 2003; Christelova et al. 2011; Changadeya et al. 2012; Hippolyte et al. 2012; de Jesus et al. 2013). However, contradictory results regarding the diversity of the EAHB have also been reported based on 24 SSR markers (Buwa 2009).

While morphological variation exists between the different EAHB varieties, many such varieties have the potential to be genetically very similar, particularly if such varieties have arisen via accumulation of somatic mutations. Given that EAHBs are a starchy staple food for over 80 million people in the Great Lakes region of Africa, the potential genetic vulnerability of EAHBs (e.g., to neo-virulent diseases and changing environmental pressures) due to a narrow genetic base is a major consideration (Cooper et al. 2000). In addition, the complex and lengthy breeding cycles associated with EAHBs pose a challenge for broadening the genetic base of EAHBs (Li et al. 2013).

Of major concern is the genetic vulnerability of bananas to *Fusarium* wilt (or panama) disease caused by strains of a soil fungus *F. oxysporum cubense* (Foc). The original Foc strain of *Fusarium* wilt previously eliminated the Gros Michel banana cultivar, the most common export banana

from the nineteenth century until the 1950s. Gros Michel was replaced with a Cavendish variety, which was resistant to the original Foc strain. However, the Cavendish is susceptible to a new Foc Tropical Race 4 (Foc-TR4) strain, and is likely to succumb in the same manner as the Gros Michel variety. The Foc-TR4 variant was first detected in Southeast Asia in the 1990s and has since been spreading to other banana growing regions. More recently, the Foc-TR4 variant has been reported in Jordan (García-Bastidas et al. 2014) and Mozambique (Butler 2013), where its continuing spread is considered as a threat to banana-based food and livelihood security in East Africa.

Inadequate knowledge of the underlying basis of observed morphological variation, genetic diversity, and population history of the EAHB germplasm genepool is an impediment to the improvement of EAHB, despite the socioeconomic importance of this staple crop. Here, we present a comprehensive analysis on the genetic diversity and population history of the EAHB subgroup of bananas. Using SSR markers, we have: (1) investigated associations between genetic diversity and morphological diversity of cultivars in the EAHB subgroup; (2) related EAHB genetic variability to that of other *Musa* subgroups and; (3) investigated support for recent population size changes among the EAHB since triploidization and subsequent domestication.

Materials and methods

EAHB samples and DNA extraction

Based on 73 morphological traits, the EAHBs have been classified into groups (referred to as clone sets) to reflect their variation in vegetative structures, bunch, fruit and male bud (Table 1). These are Nfuuka, Nakitembe, Musakala and Nakabululu (Karamura 1998). Sometimes cultivars of the cooking type become astringent and are no longer used for cooking but making beer. These are referred to as Mbidde (Table 1).

Ninety cultivars representing phenotypic diversity within and between the EAHB clone sets and including the beer types were obtained from field germplasm collections of EAHBs in Uganda and Kenya that are considered to represent the primary gene pool of EAHBs. These cultivars consisted of 48 cultivars from the Ugandan genebank and 42 cultivars from the Kenyan genebank of the triploid East African Highland banana (EAHB) as well as 4 AAB (African plantain) and 2 AAA (dessert) samples used as outgroups in this study (Supplementary Table 1). All 90 cultivars had unique phenotypic characteristics based on the IPGRI Musa descriptors used for germplasm management. All the EAHB specimens were classified in clone sets (Musakala, Nakitembe, Nfuuka, Nakabululu and Mbidde) according to the conventional criteria and fruit use (Karamura 1998). Genomic DNA was extracted from leaf samples of each specimen following modifications of two protocols; Dellaporta et al. (1983) and Mace et al. (2003). DNA was eluted in 100 µl low salt Tris–EDTA (TE; 1.0 mM EDTA, 10 mM Tris–HCL) buffer and diluted to 20 ng/µl working stocks based on spectrophotometric measurements.

DNA amplification of SSR and AFLP loci

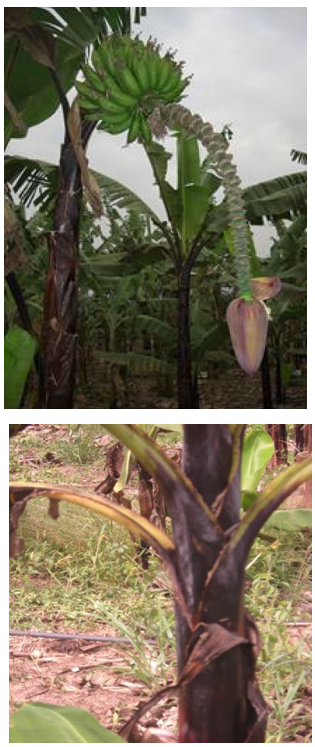
Simple Sequence Repeats (SSRs) are a type of microsatellite markers that can be used for investigating population structure and history (Spencer et al. 2000; Amos and Hoffman 2010; Mariette et al. 2010; Galov et al. 2013; Hoban et al. 2013). Previously developed SSRs derived from banana ESTs (Crouch et al. 1997; Hippolyte et al. 2010; Mbanjo et al. 2012) (Supplementary Table 1) were used in this study. In total, 250 SSR markers were screened for polymorphism and a final set of 100 SSRs were selected for this study (Supplementary Table 2). The final 100 SSRs selected for this study were chosen based on the following criteria; (a) polymorphism; with more than one allele, (b) Multiallelic; and (c) clear scorable band with no stutters. PCR was used to amplify the SSRs. The 10 µl volume PCRs contained 1× standard *Taq* buffer with $MgCl_2$; 0.2 mM dNTP mix; 0.5 units/µl *Taq* polymerase (New England Biolabs); 30 ng/µl genomic DNA template; and 0.3 µM fluorescently labeled primer. Amplification steps followed (1) initial denaturation at 95 °C for 3 min; (2) 40 cycles at 95 °C for 0.30, 1 min at 52–61 °C annealing temperature (primer pair specific), and 72 °C for 2 min; and (3) final extension of 20 min at 72 °C. All loci were individually amplified and the post-PCR primer products were multiplexed based on the dye and expected size of the fragment prior to capillary electrophoretic separation (ABI 3730xl DNA Analyzer), sizing (GeneScanTM-500 LIZ internal size standard) and manual verification of allele calling (Genemapper v4.1). The multiallelic information at each SSR locus was treated as binary data, so each SSR allele was treated as a separate marker (Christelova et al. 2011; de Jesus et al. 2013). We assessed the six outgroup cultivars representing other *Musa acuminata* subgroups separately from the set of 90 Kenyan and Ugandan EAHB.

A set of 13 AFLPs were also amplified in the same 90 Kenyan and Ugandan EAHBs as well as the six subspecies. AFLP analysis followed the procedure of Vos et al. (1995) modified to accommodate fluorescent visualization and using the restriction enzyme pair *EcoR*I/*Mse*I (Supplementary Table 3). Total genomic DNA (1 µg) was simultaneously digested and ligated in a 40 µl reaction that included 5 units each of *EcoR*I, *Mse*I, and T4 DNA ligase (New England Biolabs), 30 pmol of each *EcoR*I and *Mse*I

Table 1 Main morphological characteristics used to group the EAHB cultivars into clone sets

Nakabululu clone set: Sub-horizontal to oblique bunches and male rachis, bunches are very compact with short fruits below 5 cm, and almost perpendicular to rachis. The rachis usually has no persistent floral remains on the fruits and rachis, although some clones have semi-persistent floral remains on the rachis	
Nfuuka clone set: Have oblique to pendulous bunches and male rachis, bunches compact with fruits of medium size (15–20 cm), not strongly recurved toward the rachis. The persistent floral remains are absent on fruit tips and rachis at maturity; male bud not imbricated with an intermediate shaped apex	
Musakala clone set: Have pendulous bunches and male rachises, long fruits above 20 cm with bottle necked fruits, recurved toward the rachis	

Table 1 continued

Nakitembe clone set: Oblique to pendulous bunches and male rachis, compact bunches, fruits of medium size (15–20 cm), not strongly recurved toward the rachis; persistent floral remains on fruit tips and rachis, male bud imbricated with apex intermediate between acute and obtuse shape	
Beer (Mbidde/Mbiire): There are not many morphological characters which have defined this clone set, although majority are completely black along the length of their pseudostems. However, the clones in this clone set share characters with members of other clone sets in such a way that clones of other clone sets (Musakala, Nakabululu, Nakitembe and Nfuuka) have corresponding Mbidde clones. Such characters include: Fruits are bitterish and astringent due to the higher tannin content than the cooking types Pulp with brown sticky excretions/or latex Plenty of very sticky fast flowing sap or milky latex Majority of heavily pigmented stems and upper sheaths If cooked, food has a lumpy texture, poor brown color Uncooked fruits used for beer	

double-stranded DNA adaptor, 50 ng/μl BSA, and 50 mM NaCl in T4 Ligase buffer (New England Biolabs). Following complete digestion and ligation at room temperature, initial PCR enrichment of a subset of fragments (preamplification) used 5 μl of the digestion–ligation product as template: 0.5 μM of the *Eco*RI + 0/*Mse*I + 0 primers (IDT, Coralville, IA, USA) and 0.25 U Taq DNA polymerase (New England Biolabs) in 20 μl deionized distilled water. PCR conditions were 2 min at 94 °C; 26 cycles of 94 °C for 1 min, 56 °C for 1 min; 72 °C for 1 min; followed by 5 min at 72 °C. The pre-amplification product was then diluted tenfold with deionized distilled water. Selective amplification reactions were similar to pre-amplifications, with 2 μl of diluted pre-amplification product used as template and substituting the 50 pM of the appropriate FAM/NED labeled *Eco*RI + 3 *Mse*1 + 3 selective AFLP primers. Thirteen selective primer combinations were used on all samples (Supplementary Table 4). PCR conditions were 2 min at 94 °C, 15 cycles of 30 s at 94 °C, 30 s at 65 °C dropping 1 °C per cycle, 1 min at 72 °C; then 23 cycles of 30 s at

94 °C, 1 min at 56 °C, 2 min at 72 °C; followed by 10 min at 72 °C. AFLP genotypes were electrophoresed and visualized with an ABI 3730 DNA analyzer. Bins within the range of 100–500 bp were manually generated for the amplified fragments (GeneMapper version 4.1; SoftGenetics LLCR, State College, PA, USA). We manually checked the quality of each AFLP fingerprint and bin using the method described by Whitlock et al. (2008) with slight modifications and restricted our analyses to fragments with relative florescence units greater than 100 to reduce background noise. We scored AFLP fragments manually for their presence (denoted as 1) and absence (denoted as 0).

Intra-EAHB population genetic variation

Levels of genetic diversity in the EAHB population were evaluated by calculating the average polymorphic information content (PIC) for each SSR locus as $\mathrm{PIC}_i = 2f_i(1 - f_i)$ where i is the information of the ith marker, f_i is the frequency of the amplified allele (the presence of a band) and $(1 - f_i)$ is the frequency of null alleles that have no band (Botstein et al. 1980; de Jesus et al. 2013). Variability was also assessed based on the average number of alleles per locus, percentage of alleles identical by state, the population allele frequency, and the expected heterozygosity with Powermarker v3.25. Confidence intervals were estimated by non-parametric bootstrapping.

To assess genetic differences between cultivars, we calculated the average number of alleles in each cultivar and private alleles per cultivar (de Jesus et al. 2013). Pairwise genetic distance between cultivars was calculated as the shared allele distance (D_{AS}) with Powermarker v3.25: D_{AS} values are linearly related to the time since common ancestry for a stepwise mutation model (SMM) (Goldstein et al. 1995). It was calculated as $D_{\mathrm{AS}} = \sum_{j=1}^{m} \left(\sum_{t=1}^{k} \min\left(p_{ij}, q_{ij}\right) \right)$ where p_{ij} and q_{ij} are frequencies of ith allele at the jth locus, while k is the number of alleles at the jth locus, and m is the number of loci examined (Liu and Muse 2005).

The correlation of allele frequencies in the 90 EAHB was examined using a Bayesian framework with Structure v2.3.3 (Pritchard et al. 2000) that probabilistically assigned cultivars to genetically distinct clusters (K) and estimated admixture proportions for each cultivar in a population-free manner independent of a mutation model. The proportion of membership for each cluster permitted incomplete membership to minimize overfitting (Falush et al. 2007). To determine the most likely number of clusters, a range of values were tested ($1 \leq K \leq 10$). Analyses assumed an admixture model with correlated allele frequencies for a burn-in period of 10^5 steps prior to a run length of 10^5 with three independent iterations per K to confirm chain convergence (Pritchard et al. 2000). The second-order rate of change of the likelihood function (ΔK) was used to determine the most likely number of clusters (Evanno et al. 2005) with Structure Harvester (Earl 2011).

Population and clone set variability in the context of outgroups

To determine the genetic variability of cultivars within and among the two EAHB populations and five morphological clone sets, measures of genetic diversity were calculated: Shannon's Information Index (Lewontin 1972), Nei's genetic diversity (Nei 1973), the percentage of polymorphic bands, and the number of private alleles. Heterozygosity was defined as the mean number of polymorphic alleles. Genetic diversity was partitioned within (Hs) and among (Hb) groups for populations and clone sets. The population metrics were compared with the outgroup cultivars to provide a context for the level of population variation.

To investigate the proportion of the total variance among and within clone sets, we performed an analysis of molecular variance (AMOVA). Clone set pairwise PhiPT values were calculated to investigate the distribution of genetic differences among and within clone sets and to identify deviations from expected heterozygosity (Peakall and Smouse 2012). PhiPT was computed as Nei's genetic diversity within clone sets divided by that within and among clone sets (equivalent to an F_{ST} value). The number of migrants among clone sets in each generation (Nm) (Slatkin 1985) was calculated as $\mathrm{Nm} = (1 - \mathrm{PhiPT})/(2\mathrm{PhiPT})$.

The population structure of the EAHB set of 90 and the six outgroup cultivars was investigated using principal coordinates analysis (PCA) of pairwise SSR D_{AS} values using R (http://www.r-project.org). This was repeated for the pairwise AFLP D_{AS} values. The pairwise SSR D_{AS} values were also visualized using phylogenetic networks of neighbor-net uncorrected p distances with SplitsTree v4 (Huson and Bryant 2006) to relate the population-level variation to that of the outgroups.

Inference of historical population sizes

To gain an insight into the evolutionary history of the EAHB, we compared a neutral hypothesis of a constant recent population size compared to an alternative one of a population expansion during domestication. Genetic signatures of a bottleneck associated with clonal propagation from a small founder population may be diminished during the subsequent recovery phase during which the population expands. Consequently, distinguishing a population in post-bottleneck recovery from one that was expanding after a small ancestral population size may not be possible for inbred groups with a single recent origin.

Table 2 Minimal genetic variation in the 90 EAHB was found for the majority of the 100 SSRs used by grouping the SSRs based on the number of alleles: 84 corresponding SSRs had only one to three alleles per locus

Groups	Number of alleles per SSR	Number of corresponding SSRs	Total alleles
A	1	14	14
B	2	36	72
C	3	34	102
D	4	6	24
E	5	6	30
F	6	3	18
G	7	1	7
Total		100	267

To gain an insight into the evolutionary history of the 90 samples, we investigated variation in the mutation rate across SSRs so that we could subsequently estimate the ancestral and recent effective population size (N_e) using the estimated mutation rates with Beast v1.8 and Tracer v1.5 (Drummond 2005; Drummond and Rambaut 2007). We examined the parsimony-informative SSRs (27) that had at least three taxa per SSR as diploid data within the set of 90 EAHB with repeat lengths ranging from 1 to 115 for a one-phase site model (two-phase models may not be significantly superior, Sainudiin et al. 2004). A single representative from genetically identical taxa was used, meaning nine samples were omitted. The SSRs were modeled as unlinked microsatellites so that the mutation rate for each depended on the SSR length, mutation copy number change, and likely copy number mutation direction (Wu and Drummond 2011). Mutation rates were estimated using data for the full set of 96 (EAHB with outgroup cultivars) for these 27 SSRs assuming a constant population size for 9.5×10^7 MCMC iterations after a burn-in of 9.5×10^6 steps to ensure the effective sample size >100 for each SSR. This was repeated for the vegetatively propagated set of 90 EAHBs (assuming a constant population size for 8.6×10^7 MCMC iterations after a burn-in of 8.6×10^6 steps).

This comparison indicated the set of 96 provided a more accurate mutation rate calibration because of the higher total number of mutations, and so these mutation rate estimates were used as the SSR clock rates for an extended Bayesian skyline plot to infer N_e for the 90 EAHB (5.8×10^7 MCMC iterations after a burn-in of 5.8×10^6) and the set of 96 with outgroups (6.1×10^7 iterations with a burn-in of 6.1×10^6) (Heled and Drummond 2008). Posterior density intervals were calculated to compute the

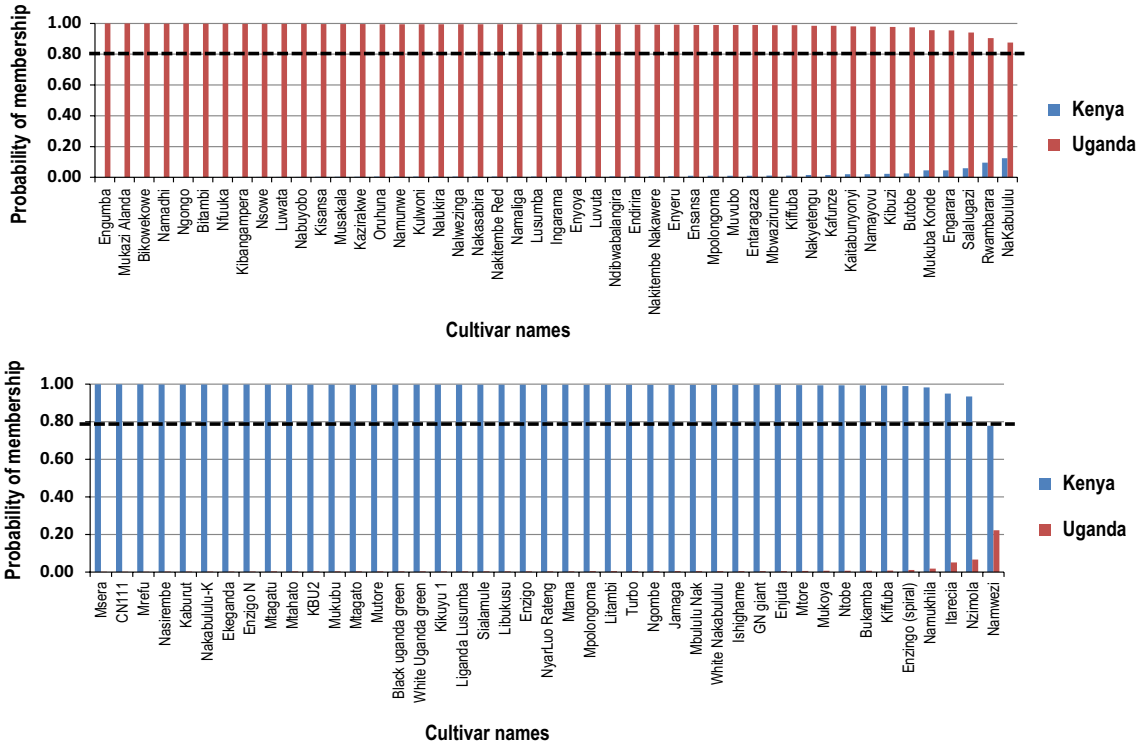


Fig. 1 Population membership for 90 East African Highland Banana cultivars from the Kenyan ($n = 42$, *green*) and Ugandan ($n = 47$, *red*) germplasm collections at $K = 2$, the most likely number of groups based on Structure classification. The *blue arrows* indicate cultivars that are known to originate from other countries. Although the SSRs distinguished the samples from the Kenyan and Ugandan collections, their total diversity was low in the context of the six subspecies. The *horizontal line* indicates an 80 % probability of membership. The *light green coloring* in the *bars* in **a**, and the *blue coloring* in the *bars* in **b** indicate cultivars with a lower probability of group membership (color figure online)

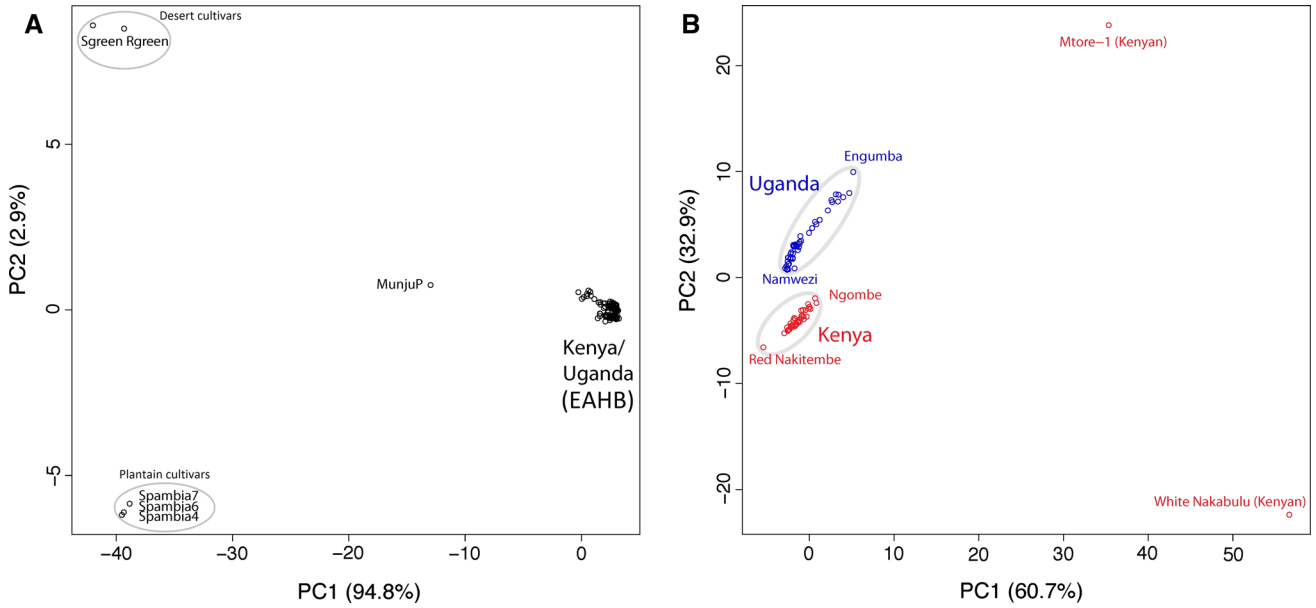


Fig. 2 Principal coordinates analysis (PCA) of the EAHB populations and six outgroup cultivars shows that population structure exists in the East African highland banana population. The EAHB-Uganda and EAHB-Kenya are genetically close, whereas the plantains (AAB) and dessert (AAA) are from genetically different *Musa* groups, though MunjuP retains an intermediate classification. **a** shows the EAHBs and the outgroup banana cultivars. **b** shows only the EAHBs

Table 3 Mean pairwise Nei's genetic diversity (Nei 1973), PhiPT and the number of migrants among clone sets per generation did not differentiate the five morphological clone sets (Mbidde, Musakala, Nakabululu, Nakitembe, Nfuuka)

	Mbidde	Musakala	Nakabululu	Nakitembe
Nei's genetic diversity				
Musakala	0.995			
Nakabululu	0.995	0.993		
Nakitembe	0.992	0.996	0.996	
Nfuuka	0.997	0.994	0.998	0.995
PhiPT				
Musakala	0.074			
Nakabululu	0.080	0.107		
Nakitembe	0.125	0.040	0.061	
Nfuuka	0.024	0.090	0.011	0.071
Migrants per generation				
Musakala	6.296			
Nakabululu	5.746	4.155		
Nakitembe	3.508	11.965	7.648	
Nfuuka	20.058	5.084	43.360	6.559

PhiPT was computed as Nei's genetic diversity within clone sets divided by that within and among clone sets (equivalent to an F_{ST} value). The number of migrants among clone sets per generation (Slatkin 1985) was calculated as Nm = (1 − PhiPT)/(2PhiPT)

relative changes in N_e, estimated using the mean mutation rate across the complete set of 27 SSRs, over time scaled using generations.

An expanding population should have lower heterozygosity (Cornuet and Luikart 1996; Piry et al. 1999) caused by a higher incidence of rare alleles (Luikart et al. 1998) at selectively neutral loci like SSRs. In contrast, a population bottleneck would distort the allele frequency distribution such that low-frequency alleles (<0.1) would be lost more rapidly than ones at higher frequencies (Maruyama and Fuerst 1985; Luikart et al. 1998). Notably for triploid specimens, heterozygosity is robust to complex ploidy states or examining inbred samples (DeGiorgio et al. 2011). We examined the heterozygosity of each SSR relative to the observed number of alleles and sample size with Bottleneck v1.2.02 using coalescent simulations under two mutation models: stepwise (SMM) (Di Rienzo et al. 1994) and a two-phase model (TPM). For the TPM, the SSR alleles were geometrically distributed (Fu and Chakraborty 1998) with 90 % following an SMM and with a variance of 30 % for the 10 % of non-stepwise mutations (Amos and Hoffman 2010). Evidence for an expansion was examined using a standardized differences statistic (T_2) and standardized sign test (Cornuet and Luikart 1996).

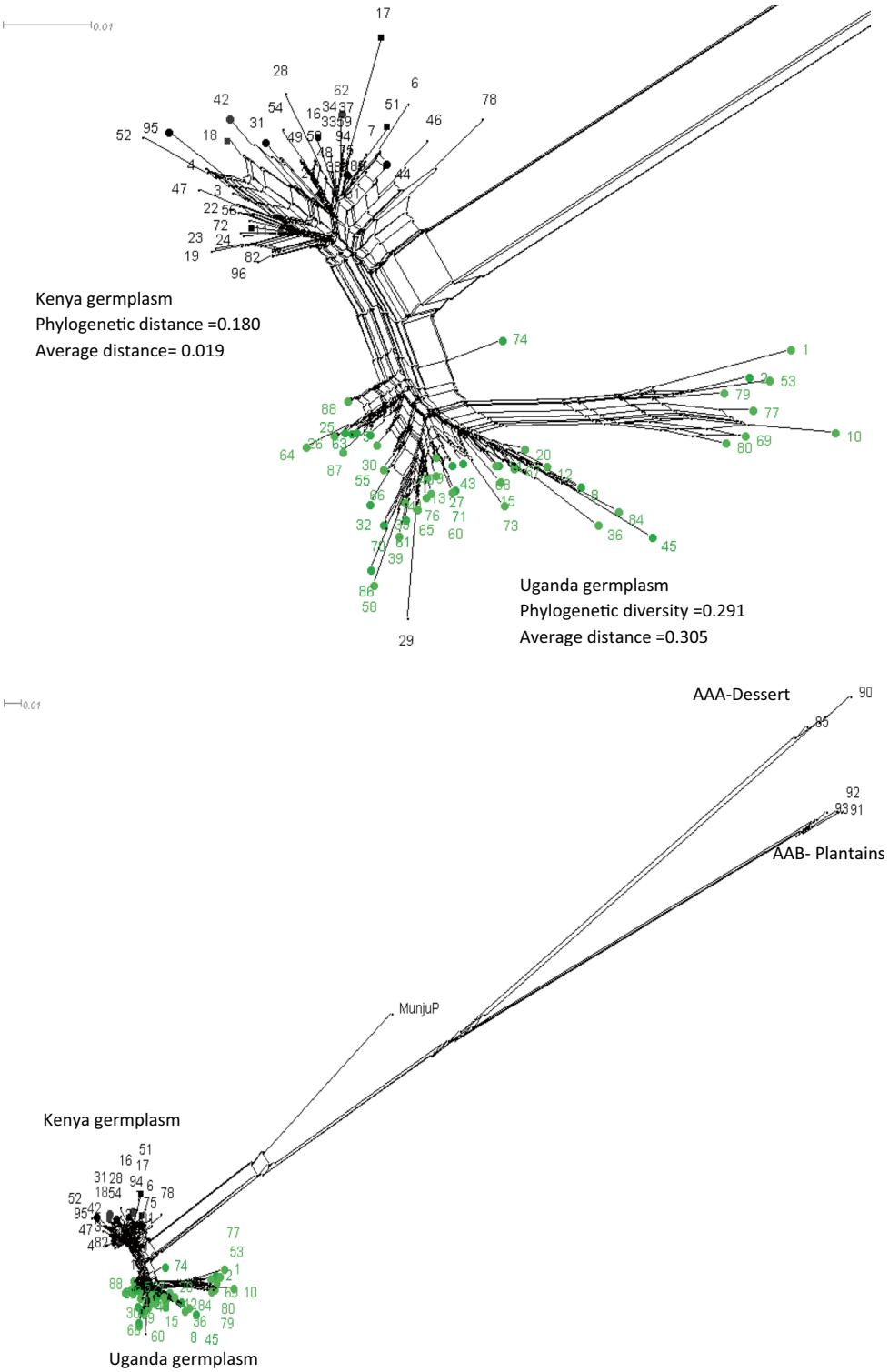


Fig. 3 A phylogenetic network of Neighbor-net uncorrected *p* distances (Bryant and Moulton 2004) of 100 SSRs of shared allele distances (D_{AS}) for the **a** set of 90 EAHB cultivars and **b** full set of 90 EAHB and six outgroups: plantains (Spambia 4, 6 and 7) and AAA-dessert (*Somatic green* and *Red green*). One sample (MunjuP) was genetically intermediate between the plantains, dessert varieties, and the EAHB. Sample numbers correspond to those in Supplementary Table 1 (color figure online)

Results

To determine the extent of genetic variation across the EAHB genepool, we amplified 100 SSRs across a population of 90 EAHBs (sourced from Kenyan and Ugandan germplasm collections) and across six African plantain and dessert bananas.

The East African Highland banana population is genetically uniform

Low levels of genetic variability were observed in the 90 EAHB cultivars compared to the outgroup cultivars. In total, 475 SSR alleles were scored in 100 SSR, 267 were discovered in the set of 90 EAHB cultivars, while 43.78 %

Table 4 The 90 EAHB showed a lower than expected heterozygosity consistent with a population expansion (Cornuet and Luikart 1996; Piry et al. 1999)

Mutation model	Expected SSRs with heterozygosity excess	Observed SSRs with heterozygosity excess	Observed SSRs with heterozygosity deficiency	Wilcoxon sign test
Stepwise (SMM)	50.95	32	78	0.00021
Two-phase (TPM)	48.21	32	78	0.00048

Results were assessed with Bottleneck v1.2.02 using coalescent simulations under two mutation models: stepwise (SMM) and two-phase (TPM). The SSR alleles were geometrically distributed with 90 % of mutations following an SMM and with a variance of 30 % for the TPM. The significant heterozgosity deficit was supported by Wilcoxon standardized sign tests and negative standardized differences statistic values (T_2)

of total were exclusively scored in the outgroup. Due to the triploid nature of the EAHBs, three alleles at each SSR locus were expected. In this study, the allelic pattern detected ranged from one to a maximum of six alleles. The 100 SSRs could be grouped into seven categories based on the number of alleles scored per locus (Table 2): 84 out of 100 SSRs had three alleles or less. Allelic homogeneity was observed in 86 % of the total alleles scored in EAHBs where cultivars showed identical allelic profiles, which we termed as the main allele pattern. In 24 out of the 90 cultivars, a few alleles showed allelic variations (deviations from the main allele pattern) where they occurred either as a loss of a main allele or new allele(s) (in addition or as substitution to the main allele) and cultivars either showed one or a combination of both variations. Emergence of new alleles in addition or as substitution to the main allele was more common (observed in 27 alleles) than the loss of main allele (observed in 15 alleles).

The set of 90 EAHBs had low genetic diversity as evidenced by the mean PIC (0.058) and gene diversity (0.070) values. This low level of genetic variability was further highlighted by a lack of genetic differentiation between pairs ($D_{AS} = 0.071$, range 0.000–0.176); 81.3 % of the 90 cultivars had $D_{AS} < 0.1$; while the mean minor allele frequency was just 0.05. Moreover, 58 % of alleles were identical by state, even though an average of 209.3 alleles per cultivar was sampled (from a minimum of 205 for the cultivar NyarLuo Ratong to a maximum of 214 for the cultivar Nakitembe Red).

Population-free clustering with Structure identified two genetically distinguishable groups (Evanno et al. 2005) which corresponded to the sub-populations obtained from the Kenyan and Ugandan germplasm collections (Fig. 1; Supplementary Table 4), with population membership probabilities greater than 0.8 (Pritchard et al. 2000).

Principal coordinates analysis (PCA) of the population of 90 and six outgroup cultivars (plantain and dessert) analyzed by SSRs showed little differentiation between the two EAHB populations (Kenya and Uganda, PC2 with 1.9 % of total variation) (Fig. 2). In contrast, the six outgroup cultivars were distributed across PC1 accounting for 94.2 % of diversity. This lack of resolution was also evident for the AFLP analysis (Supplementary Figure S1). Though much of the diversity was within subpopulations (Hs = 0.056; $p < 0.0001$) compared to between subpopulations (Hb = 0.052; $p < 0.0001$), the Uganda EAHB subpopulation was more diverse than the Kenyan one. The Ugandan EAHB subpopulation had a higher expected heterozygosity (0.065 vs 0.049) and proportion of polymorphic loci (13.5 vs 9.4 %), possibly suggesting that the Kenyan EAHB subpopulation may be derived from the Ugandan one. However, this may be complicated by genetic drift distinguishing the populations at few loci, and novel variants unique to certain cultivars (White Nakabululu and Mtore; Supplementary Figure S5).

No genetic differentiation of morphological groups of EAHBs

There was much higher genetic variation observed within the five morphological groups (clone sets) than between the five morphological clone sets (Hw = 0.0925 vs Hb = 0.022): 96 % of the variation was within clone set groups and only 4 % of the variation was between groups (Supplementary Table 6). This was supported by a lack of differentiation between clone sets (mean PhiPT = 0.036, $p = 0.01$ with a range of 0.011 and 0.125, Table 3). Furthermore, the maximum D_{AS} value between clone set pairs was very low at 0.004 (Table 3).

Low genetic diversity within EAHB

Diversity within the two EAHB subpopulations was much lower compared to the outgroup cultivars. This was illustrated by PCA (Fig. 2) and was further highlighted in a phylogenetic network that partitioned the data in a similar manner: the set of 90 EAHBs form two closely related groups that are markedly homogeneous compared to the highly diverse six outgroups (Fig. 3; Supplementary Table 5). These results also demonstrated the differentiation of the EAHB from the plantains (AAB) and dessert (AAA) bananas—though MunjuP retained an intermediate classification suggesting progenitor(s) related to both cultivar sets.

Molecular evidence of a recent EAHB population expansion

In this study, we performed the first estimation of SSR mutation rates in bananas. Considerable variation in the estimated SSR mutation rates across 27 parsimony-informative SSRs was observed, with a mean of 0.00166 substitutions per SSR per generation, ranging from 0.00027 (locus 44) to 0.0030 (locus 47). This 11-fold magnitude of variation illustrated the heterogeneity associated with STR/SSR mutation rates (Scarcelli et al. 2013) (Supplementary Figure S3). In addition, somatic mutation rates estimated for the set of 90 EAHBs differed: these had a mean value of 0.00804 substitutions per SSR per generation, ranging from 0.00065 (locus 44) to 0.07620 (locus 70). Compared to the data for the entire set of 96, this produced an average excess of 4.8-fold, but this varied from 0.6 (locus 27) to 85.0-fold (locus 70), highlighting potential mutation rate differences between vegetatively propagated samples of different *Musa acuminata* parental origin.

We also investigated evidence for recent population size changes in the EAHBs. Using the mutation rates estimated at 27 SSRs for the set of 96, extended Bayesian skyline plots indicated a historical low constant N_e with 95 % highest posterior density (HPD) values <4.9 individuals for both. Present generation N_e estimates indicated a dramatic recent increase in N_e for the sets of 90 EAHB (1435.7 with a 95 % HPD range of 1175–12,234) and 96 containing the outgroups (2868.6, 95 % HPD range 2352–35,301) (Supplementary Figure S4). Despite this recent jump in N_e, no statistically significant evidence of population size change was inferred using this dataset.

Our analysis detected a significant heterozygote deficiency for both SMM and TPM mutation rates, characteristic of a population expansion (Wilcoxon one-tailed $p < 0.0005$, Table 4) (Cornuet and Luikart 1996) or recovery after a decrease in size (McEachern et al. 2011). In addition, a significantly negative standardized differences value (T_2) was observed (SMM: $T_2 = -3.774$, $p = 0.00008$; TPM: $T_2 = -2.668$, $p = 0.00381$). Allele frequency distributions of the 43 polymorphic SSRs showed a shifted mode from the normal 'L-shaped' incompatible with a recent bottleneck (Supplementary Figure S2).

Discussion

The domestication of crop species from crop wild relatives has typically resulted in a genetic bottleneck and associated founder effect on genetic diversity in the primary cultivated genepool. For many crops, post-domestication breeding and artificial selection have allowed the introduction of new alleles and haplotype combinations through sexual recombination. However, the domestication of triploid crops can be associated with sterility, which generates a reproductive barrier to the introduction of novel germplasm (e.g., from wild relatives) and acts as a major constraint for sexual recombination based breeding programs. This is particularly the case for the triploid East African Highland bananas, which can be vegetatively propagated despite their seed sterility.

The majority of the current EAHB cultivars grown by almost 100 million farmers in East Africa have not arisen from formal breeding programs. Instead such varieties represent vegetatively propagated clones that have been transmitted over generations and exchanged between farmers for centuries. The phenotypic diversity for EAHBs that exists in the Great Lakes region of East Africa has led to the proposition that this region is a secondary center of diversity for EAHBs. To capture the diversity of the EAHB primary genepool, both the field genebank EAHB germplasm collections in Uganda and Kenya were established in the 1990s from germplasm collected from farmers fields across East Africa. It is possible that the morphological variation between the ~120 morphologically distinct cultivars of EAHB in these germplasm collections has arisen due to somaclonal mutations which were selected and propagated as somaclonal mutants by farmers.

Despite their importance for food and livelihood security, little is known about the genetic diversity and history of the EAHB (AAA) subgroup of bananas. In this study, we used 100 microsatellite markers (and AFLPs) to investigate the genetic diversity, population structure and evolutionary history of 90 phenotypically diverse EAHB cultivars (representative of the EAHB primary genepool) obtained from Ugandan and Kenyan germplasm collections. Our study included six cultivars (four plantains and two dessert samples) representing different *Musa acuminata* subgroups.

The cultivated EAHB primary genepool is genetically uniform

Although substantially reduced genetic diversity has been reported within other *Musa* groups (Creste et al. 2003; Hippolyte et al. 2012), our study indicates that EAHBs exhibit significantly lower genetic variability than other *Musa* groups (El-Khishin et al. 2009; Opara et al. 2010; Christelova et al. 2011; Resmi et al. 2011; Abdullah et al. 2012; Shaibu 2012; de Jesus et al. 2013). In addition, while EAHBs are classified into four clone sets based on morphological characteristics, our study indicated that the morphological classification system was not reflected at the level of genetic diversity and relatedness. Our population classification and phylogenetic analysis showed minimal differentiation of the cultivars in the Ugandan and Kenyan collections, suggesting that the cultivars in both these germplasm collections have a recent single origin, more likely to be ancestral to the Ugandan set on the basis of its comparatively higher genetic variability.

No association between morphological classification systems and genetic diversity of EAHB cultivars

In this study, we have investigated whether there is an association between genetic diversity and morphological diversity for EAHBs and find no evidence for such an association. This raises the question of the underlying basis for the morphological variation observed between the EAHB cultivars and clone sets (Table 1). It is possible that the morphological differences observed between EAHBs represent rare somatic mutations (e.g., point mutation, transposon insertions/excisions) that could not be directly detected by low-density markers such as SSRs or AFLPs) that affect the morphological trait used for classification (e.g., fruit size, fruit length, stem color, etc.). Where they exist, the identification of genomic loci and mutations underlying the morphological traits used as descriptors for EAHBs would require experimental designs involving mapping populations, bulked segregant and/or genome wide association approaches. In addition, studies on heritability of the morphological traits that differ across EAHBs would be necessary to inform such approaches. The advent of higher density genotyping by sequencing (GBS) approaches may facilitate the identification of loci underlying heritable morphological differences between EAHBs. Alternatively (or in addition), the morphological variation in EAHBs could be arising due to mitotically heritable epigenetic variation (arising from environmental and/or human selection effects) across vegetative propagation cycles.

Irrespective of the underlying mechanistic basis of the morphological traits used to classify EAHBs, our study demonstrates that the cultivated EAHB genepool is genetically uniform. As a result, EAHB cultivation, consumption and associated food security in East Africa are likely to be vulnerable to neo-virulent pests and pathogens of EAHBs.

The broadening of the genetic base of EAHB production (while maintaining and improving quality and yield characteristics of EAHBs) remains an important priority. Despite the inherent difficulties for polyploid breeding in EAHBs, EAHB breeding efforts in East Africa are engaged in efforts to introduce new genetic variation into EAHBs (Ortiz and Swennen 2014). *Musa* variation is ultimately derived from four wild species: *M. acuminata* (A genome), *M. balbisiana* (B), *M. schizocarpa* (S) and *M. textilis* (T)—of which the latter two show extensive genetic but not phenotypic variation (de Jesus et al. 2013), and among whom only *M. acuminata* is parthenocarpic (Heslop-Harrison and Schwarzacher 2007; Kennedy 2008).

EAHBs likely arose from a single ancestral clone followed by a recent population expansion by vegetative (asexual) propagation by farmers

Banana domestication may have occurred up to 10,000 years ago (Perrier et al. 2011). However, movement from primary centers of diversity, such as from Papua New Guinea to Africa, may have occurred much recent (e.g., about 2000–2500 years ago) (Perrier et al. 2011). It is currently not known whether the triploid origin of EAHBs occurred in Papua New Guinea or in Africa post-introduction. In either scenario, the generation of the initial triploid ancestors was most likely followed by ongoing human selection of new clones (as somatic mutants that were vegetatively propagated).

Our simulations of the recent historical effective population size and expected heterozygosity are symptomatic of a recent EAHB population expansion (Cornuet and Luikart 1996). Although low allelic diversity may be caused by a genetic bottleneck (Gebremedhin et al. 2009), the EAHB sample set did not support this directly, consistent with previous work (Li et al. 2013)—though the ancestral population size was extremely small. This genetic pattern may be accentuated by a Meselson effect of heterozygous alleles persisting in an asexual reproductive system (Butlin 2002). A recent expansion of EAHBs could be due to human-mediated establishment of EAHBs in new agro-environments across East Africa (Schoebel et al. 2014).

The intermediate phylogenetic placement of the cultivar MunjuP between the plantains, dessert and EAHBs could suggest ancient interbreeding events (Bryant and Moulton 2004), which is supported by previous work on African plantains and EAHBs, which may share at least one maternal A genome derived from *M. acuminata* subspecies *banksii* (Kennedy 2008). The authors recognize that while common descent does not, sensu *strictu*, imply interbreeding events, it does suggest that these genomes had a common genome donor parent. The observation of historical mixing events across subspecies suggests that deeper molecular investigation of these subspecies can help to determine the hybrid origins of extant cultivars, for which the number and timing ancient migrations both before (in diploids) and after hybridization (in triploids) remain unknown (de Jesus et al. 2013).

The differences between the 90 EAHB cultivars across geographical regions (Kenya and Uganda) were minimal relative to the higher diversity in the plantain and dessert cultivars. Introgression of genetically distinct cultivars at low levels would have erased the genetic homogeneity in the EAHBs (Miller et al. 2012). Yet, no evidence for multiple origins or admixture was observed in our study. Our

results support our hypothesis of a single seed origin for EAHBs, consistent with a hybridization event, followed by selection and subsequent clonal propagation by farmers over centuries and across environments.

Our study strongly supports the hypothesis that the EAHB arose from genetically monomorphic clones, likely selected during domestication. We consider it likely that wild and cultivated diploids should harbor an array of genetically diverse clones. For further studies, a finer delineation of the events associated with the triploidization and domestication of EAHB will be achieved by comparing genetic variation in EAHB (AAA), plantain (AAB) and dessert (AAA) subspecies, to the genetic variation in *M. acuminata* wild and edible diploid (AA) and triploid (AAA) samples. Overall, our SSR and AFLP study conclusively demonstrates that the cultivated EAHB is genetically uniform, most likely arising from a single recent triploid origin followed by somaclonal vegetative reproduction by farmers across East Africa.

Author contribution statement CS and JL designed and supervised the research. MK collected samples and conducted lab work. Sample collection assistance was provided by DK, MN and MO. Data analysis and manuscript drafting was conducted by MK, TD, JL, MF and CS. Sample collection assistance was provided by DK and MO. Ph.D. research supervision of MK by JL, MF and CS.

Acknowledgments This research was conducted at Biosciences Eastern and Central Africa (BecA)-ILRI and in the Genetics and Biotechnology Lab at the Plant and AgriBiosciences Research Centre (PABC), School of Natural Sciences, National University of Ireland Galway. We thank Dr Kassa Semagn for his invaluable help in data analysis. This research was supported by funding from Irish Aid and the International Institute for Tropical Agriculture (IITA).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Abdullah NAP, Bin Saleh G, Putra ETS, Bin Wahab Z (2012) Genetic relationship among Musa genotypes revealed by microsatellite markers. Afr J Biotechnol 11(26):6769–6776

Amos W, Hoffman JIW (2010) Evidence that two main bottleneck events shaped modern human genetic diversity. Proc Biol Sci 277(1678):131–137

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32(3):314–331

Bryant D, Moulton V (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol 21(2):255–265

Butler D (2013) Fungus threatens top banana: fears rise for Latin American industry as devastating disease hits leading variety in Africa and Middle East. Nature 504:195–196

Butlin R (2002) The costs and benefits of sex: new insights from old asexual lineages. Nat Rev Genet 3:311–317

Buwa R (2009) Using SSR markers to fingerprint the East African Highland banana cultivars. M.Sc., Makerere University

Changadeya W, Kaunda E, Ambali AJD (2012) Molecular characterization of Musa cultivars in Malawi using SSR. Afr J Biotechnol 11(18):4140–4157

Christelova P, Valarik M, Hribova E, Van den Houwe I, Channeliere S, Roux N, Dolezel J (2011) A platform for efficient genotyping in Musa using microsatellite markers. AoB Plants 2011:plr024

Cooper H, Spillane C, Hodgkin T (2000) Broadening the genetic bases of crop production. IPGRI, FAO, CABI, Rome (Italy), Wallingford (UK)

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting population bottlenecks from allele frequency data. Genetics 144:2001–2014

Crane MB, Lawrence WJC (1956) The genetics of garden plants, 4th edn. Macmillan, London

Creste S, Neto AT, de Oliveira Silva S, Figueira A (2003) Genetic characterization of banana cultivars (*Musa* spp.) from Brazil using microsatellite markers. Euphytica 132:259–268

Crouch JH, Vuylsteke D, Ortiz R (1997) Microsatellite markers for molecular breeding of Musa. InfoMusa 6(1):5–6

de Jesus ON, Silva Sde O, Amorim EP, Ferreira CF, de Campos JM, Silva Gde G, Figueira A (2013) Genetic diversity and population structure of Musa accessions in ex situ conservation. BMC Plant Biol 13:41

De Langhe E (1961) La taxonomie du bananier plantain en Afrique Équatoriale. Journal d'agriculture tropicale et de botanique appliquée 8:417–449

De Langhe EA (1964) The origin of variation in the plantain banana. Mededelingen van de Landbouwhogeschool en de Opzoekingsstations van de Staat te Gent 39:45–80

DeGiorgio M, Degnan JH, Rosenberg NA (2011) Coalescence-time distributions in a serial founder model of human evolutionary history. Genetics 189(2):579–593

Dellaporta S, Wood J, Hicks J (1983) A plant DNA minipreparation: version II. Plant Mol Biol Rep 1:19–21

D'Hont A, Denoeud F, Aury JM, Baurens FC, Carreel F, Garsmeur O, Noel B, Bocs S, Droc G, Rouard M et al (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature 488(7410):213–217

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci 91(8):3166–3170

Drummond AJ (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 22(5):1185–1192

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214

Earl DA (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4(2):359–361

El-Khishin D, Belatus EL, El-Hamid AA, Radwan KH (2009) Molecular characterization of banana cultivars (*Musa* spp.) from Egypt using AFLP. Res J Agric Biol Sci 5(3):272–279

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14(8):2611–2620

Falush D, Stephens S, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol Ecol Notes 7:574–578

Fu Y-X, Chakraborty R (1998) Simultaneous estimation of all the parameters of a stepwise mutation model. Genetics 150:487–497

Galov A, Byrne K, Gomerčić T, Duras M, Arbanasić H, Sindičić M, Mihelić D, Kovačić A, Funk SM (2013) Genetic structure and admixture between the Posavina and Croatian coldblood in contrast to the Lippizan horse from Croatia. Czech J Anim Sci 58(2):71–78

Ganapathy KN, Gomashe SS, Rakshit S, Prabhakar B, Ambekar SS, Ghorade RB, Biradar BD, Saxena U, Patil JV (2012) Genetic diversity revealed utility of SSR markers in classifying parental lines and elite genotypes of sorghum (*Sorghum bicolor* L. Moench.). Aust J Crop Sci 6(11):1486–1493

García-Bastidas F, Ordóñez N, Konkol J, Al-Qasim M, Naser Z, Abdelwali M, Salem N, Ploetz RC, Kema GHJ (2014) First report of *Fusarium oxysporum* f. sp. cubense tropical race 4 associated with Panama disease of banana outside Southeast Asia. Plant Disease 98:694

Gebremedhin B, Ficetola GF, Naderi S, Rezaei HR, Maudet C, Rioux D, Luikart G, Flagstad Ø, Thuiller W, Taberlet P (2009) Combining genetic and ecological data to assess the conservation status of the endangered Ethiopian walia ibex. Anim Conserv 12(2):89–100

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics 139(1):463–471

Heled J, Drummond AJ (2008) Bayesian inference of population size history from multiple loci. BMC Evol Biol 8:289

Heslop-Harrison JS, Schwarzacher T (2007) Domestication, genomics and the future for banana. Ann Bot 100(5):1073–1084

Hippolyte I, Bakry F, Seguin M, Gardes L, Rivallan R, Risterucci AM, Jenny C, Perrier X, Carreel F, Argout X et al (2010) A saturated SSR/DArT linkage map of *Musa acuminata* addressing genome rearrangements among bananas. BMC Plant Biol 10:65

Hippolyte I, Jenny C, Gardes L, Bakry F, Rivallan R, Pomies V, Cubry P, Tomekpe K, Risterucci AM, Roux N et al (2012) Foundation characteristics of edible Musa triploids revealed from allelic distribution of SSR markers. Ann Bot 109(5):937–951

Hoban S, Gaggiotti OE, Bertorelle G (2013) The number of markers and samples needed for detecting bottlenecks under realistic scenarios, with and without recovery: a simulation -based study. Mol Ecol 22:3444–3450

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23(2):254–267

Karamura D (1998) Numerical taxonomic studies of the East African highland bananas (Musa AAA- East Africa) in Uganda. Doctor of Philosophy, PGRI, Reading

Kennedy J (2008) Pacific Bananas: complex origins, multiple dispersals? Asian Perspect 47:75–94

Lewontin R (1972) The apportionment of human diversity. Evol Biol 6:381–398

Li L, Wang HY, Zhang C, Wang XF, Shi FX, Chen WN, Ge XJ (2013) Origins and domestication of cultivated banana inferred from chloroplast and nuclear genes. PLoS One 8(11):e80502

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinform Appl Note 21(9):2128–2129

Luikart G, Allendorf FW, Cornuet JM, Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. Am Genet Assoc 89(3):238–247

Mace E, Buhariwalla K, Buhariwalla H, Crouch J (2003) A high-throughput DNA extraction protocol for tropical molecular breeding programs. Plant Mol Biol Rep 21:459–460

Mariette S, Tavaud M, Arunyawat U, Capdeville G, Millan M, Salin F (2010) Population structure and genetic bottleneck in sweet cherry estimated with SSRs and the gametophytic self-incompatibility locus. BMC Genet 11:77

Martin G, Baurens FC, Cardi C, Aury JM, D'Hont A (2013) The complete chloroplast genome of banana (*Musa acuminata*, Zingiberales): insight into plastid monocotyledon evolution. PLoS One 8(6):e67350

Maruyama T, Fuerst PA (1985) Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. Genetics 111(3):675–689

Mbanjo EGN, Tchoumbougnang F, Mouelle AS, Oben JE, Nyine M, Dochez C, Ferguson ME, Lorenzen J (2012) Development of expressed sequence tags-simple sequence repeats (EST-SSRs) for Musa and their applicability in authentication of a Musa breeding population. Afr J Biotechnol 11(71):13546–13559

McEachern MB, Van Vuren DH, Floyd CH, May B, Eadie JM (2011) Bottleneck and rescue effects in fluctuating population of golden mantle ground squirrels. Conserv Genet 12:285–296

Miller M, Haig SM, Mullins Thomas D, Poppe KJ, Green M (2012) Evidence for population bottlenecks and subtle genetic structure in the Yellow Rail. The Condor 114(1):100–112

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

Opara UL, Jacobson D, Al-Saady NA (2010) Analysis of genetic diversity in banana cultivars (*Musa* cvs.) from the South of Oman using AFLP markers and classification by phylogenetic, hierarchical clustering and principal component analyses. J Zhejiang Univ Sci B 11(5):332–341

Ortiz R, Swennen R (2014) From crossbreeding to biotechnology-facilitated improvement of banana and plantain. Biotechnol Adv 32(1):158–169

Pachuau L, Atom AD, Thangjam R (2014) Genome classification of Musa cultivars from northeast India as revealed by ITS and IRAP markers. Appl Biochem Biotechnol 172(8):3939–3948

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28(19):2537–2539

Perrier X, Bakry F, Carreel F, Jenny C, Horry JP, Lebot V, Hippolyte I (2009) Combining biological approaches to shed light on evolution of edible bananas. Ethnobot Res Appl 7:199–216

Perrier X, De Langhe E, Donohue M, Lentfer C, Vrydaghs L, Bakry F, Carreel F, Hippolyte I, Horry JP, Jenny C et al (2011) Multidisciplinary perspectives on banana (*Musa* spp.) domestication. Proc Natl Acad Sci 108(28):11311–11318

Pillay M, Ogundiwin E, Nwakanma D, Ude G, Tenkouano A (2001) Analysis of genetic diversity and relationships in East African banana germplasm. Theor Appl Genet 102(6–7):965–970

Piry S, Luikart G, Cornuet J-M (1999) Bottleneck: a computer program for detecting recent reductions in the effective population size using allele frequency data. J Hered 90(4):502–503

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2):945–959

Resmi L, Kumari R, Bhat KV, Nair AS (2011) Molecular Characterization of genetic diversity and structure in South Indian Musa cultivars. Int J Bot 7(4):274–282

Sainudiin R, Durrett RT, Aquadro CF, Nielsen R (2004) Microsatellite mutation models: insights from a comparison of humans and chimpanzees. Genetics 168:383–395

Scarcelli N, Marie Couderc M, Baco MN, Egah J, Vigouroux Y (2013) Clonal diversity and estimation of relative clone age: application to agrobiodiversity of yam (*Dioscorea rotundata*). BMC Plant Biol 13:178

Schoebel CN, Stewart J, Gruenwald NJ, Rigling D, Prospero S (2014) Population history and pathways of spread of the plant pathogen *Phytophthora plurivora*. PLoS One 9(1):e85368

Shaibu AA (2012) Genetic diversity analysis of Musa species using amplified fragment length polymorphism and multivariate statistical technique. Int J Biochem Biotechnol 1(6):175–178

Shepherd K (1957) Banana cultivars in East Africa. Trop Agric 34:277–286

Simmonds NW, Shepherd K (1955) The taxonomy and origins of the cultivated bananas. J Linn Soc Lond Bot 55(359):302–312

Slatkin M (1985) Gene flow in natural populations. Annu Rev Ecol Syst 16:393–430

Spencer CC, Neigel JE, Leberg PL (2000) Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks. Mol Ecol 9(10):1517–1528

Ssebuliba RN, Rubaihayo R, Tenkouano A, Makumbi D, Talegera D, Magambo M (2005) Genetic diversity among East African Highland bananas for female fertility. Afr Crop Sci J 13(1):13–26

Till BJ, Jankowicz-Cieslak J, Sagi L, Huynh OA, Utsushi H, Swennen R, Terauchi R, Mba C (2010) Discovery of nucleotide polymorphisms in the Musa gene pool by Ecotilling. Theor Appl Genet 121(7):1381–1389

Tugume AK, Lubega GW, Rubaihayo PR (2003) Genetic diversity of East African Highland bananas using AFLP. Infomusa 11(2):28–32

Ude G, Pillay M, Ogundiwin E, Tenkouano A (2003) Genetic diversity in an African plantain core collection using AFLP and RAPD markers. Theor Appl Genet 107(2):248–255

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M et al (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23(21):4407–4414

Whitlock R, Hipperson H, Mannarelli M, Butlin RK, Burke T (2008) An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. Mol Ecol Resour 8(4):725–735

Wu CH, Drummond AJ (2011) Joint inference of microsatellite mutation models, population history and genealogies using transdimensional Markov Chain Monte Carlo. Genetics 188(1):151–164