Received: 13 October 2023 | Accepted: 23 February 2024

DOI: 10.1002/aps3.11589

APPLICATION ARTICLE

Applications in Plant Sciences

Capturing and managing genetic diversity in ex situ collections of threatened tropical trees: A case study in *Karomia gigas*

Alexander G. Linan[1] | **Roy E. Gereau**[1] | **Rebecca Sucher**[1] | **Fandey H. Mashimba**[2] | **Burgund Bassuner**[1] | **Andrew Wyatt**[1] | **Christine E. Edwards**[1]

[1]Missouri Botanical Garden, 4344 Shaw Blvd., St. Louis, Missouri 63110, USA

[2]Tanzania Forest Service Agency, Directorate of Tree Seed Production, Box 40832, Nyerere Road, Mpingo House, Dar es Salaam, Tanzania

Correspondence

Alexander G. Linan, Missouri Botanical Garden, 4344 Shaw Blvd., St. Louis, Missouri 63110, USA.
Email: alinan@mobot.org

This article is part of the special issue "From Theory to Practice: New Innovations and Their Application in Conservation Biology."

Abstract

Premise: Although ex situ collections of threatened plants are most useful when they contain maximal genetic variation, the conservation and maintenance of genetic diversity in collections are often poorly known. We present a case study using population genomic analyses of an ex situ collection of *Karomia gigas*, a critically endangered tropical tree from Tanzania. Only ~43 individuals are known in two wild populations, and ex situ collections containing 34 individuals were established in two sites from wild-collected seed. The study aimed to understand how much diversity is represented in the collection, analyze the parentage of ex situ individuals, and identify efficient strategies to capture and maintain genetic diversity.

Methods: We genotyped all known individuals using a 2b-RADseq approach, compared genetic diversity in wild populations and ex situ collections, and conducted parentage analysis of the collections.

Results: Wild populations were found to have greater levels of genetic diversity than ex situ populations as measured by number of private alleles, number of polymorphic sites, observed and expected heterozygosity, nucleotide diversity, and allelic richness. In addition, only 32.6% of wild individuals are represented ex situ and many individuals were found to be the product of selfing by a single wild individual.

Discussion: Population genomic analyses provided important insights into the conservation of genetic diversity in *K. gigas*, identifying gaps and inefficiencies, but also highlighting strategies to conserve genetic diversity ex situ. Genomic analyses provide essential information to ensure that collections effectively conserve genetic diversity in threatened tropical trees.

KEYWORDS

botanic garden, conservation genetics, ex situ, genetic diversity, *Karomia*, parentage analysis, Tanzania

Biodiversity is experiencing a global crisis. Although extinction is a natural process that occurs across all branches of life, a recent study found that the current extinction rate in plants is up to 500 times greater than the background rate before 1900 (Humphreys et al., 2019), with the greatest rate found in tropical and subtropical shrubs and trees. Nic Lughadha et al. (2020) found that around 42.8% of plant species are threatened with extinction, primarily due to anthropogenic threats such as habitat loss due to urbanization and commercial activities, direct exploitation, competition with invasive species, and climate change. As a result of these pressures, some plant species have been reduced to just a few individuals, requiring urgent conservation action to prevent their extinction.

Ideally, the conservation of a threatened plant species occurs in its natural habitat with protection and proper management; however, ex situ conservation (e.g., conserving germplasm off-site in a protected area, seed bank, or botanical garden) is a complementary approach commonly used in conjunction with in situ conservation efforts. The primary goals of ex situ conservation are to safeguard

against the loss of individuals and populations of a threatened species and to provide source material for population augmentations and reintroductions (Cohen et al., 1991; Falk and Holsinger, 1991; Kramer et al., 2011; Guerrant et al., 2014; Center for Plant Conservation, 2019). Ex situ conservation is particularly important for safeguarding species experiencing threats in their native habitat, including widespread habitat conversion, adverse land use practices, or reproductive failure (Pritchard et al., 2012). For critically endangered species with only a small number of remaining individuals, ex situ conservation is critical for safeguarding against catastrophic, irrecoverable losses of these individuals, which represent most or all of the genetic diversity remaining in the species.

Most ex situ plant collections (hereafter referred to as collections) consist of either living plants or banked germplasm, the latter usually consisting of seeds stored in seed banks specifically for conservation purposes (Kramer et al., 2011). However, some species do not reliably produce seeds or have seeds with low viability, in which case propagation through tissue culture or other approaches may be the only way to develop a collection. Furthermore, some seeds do not remain viable using traditional storage techniques (i.e., recalcitrant or exceptional species), including those of many tropical species, and must be germinated shortly after field collection and maintained as living plants in collections. Collections of whole living plants can be useful for captive breeding programs to generate new plants for augmentations and (re)introductions. However, whole living plants are resource-intensive to maintain, and approaches to ensure efficiency in the resources invested in these collections are of considerable interest to the plant conservation community.

It is widely recognized that collections of plants are most valuable when they encompass as much of a species' genetic and geographic variation as possible (Schoen and Brown, 1993; Petit et al., 1998; De Souza et al., 2015). Because genetic diversity is important for maintaining the resiliency and adaptive potential of a species, the goal of ex situ conservation efforts for critically endangered plant species with only a few remaining individuals is often to safeguard all of the remaining genetic diversity (e.g., Mashburn et al., 2023). However, collections may not represent the full genetic diversity for a variety of technical reasons. Often, ex situ conservation efforts occur in response to a crisis, forcing conservationists to collect whatever tissue is available. Conservationists may lack permission or be physically unable to access some individuals (Griffith et al., 2020). Some plants may not produce sufficient seed or seeds may have low germination rates, limiting the ability to represent some individuals in collections when vegetative propagation or tissue culture is not possible. Although some plants that do not produce seeds may contribute to an ex situ collection by serving as pollen donors, the paternal parents of seeds in collections are rarely known, limiting our knowledge of which individuals are represented. Thus, analyses that help clarify parentage of collections can provide important insight into the wild individuals represented ex situ and those that should be targets for additional collection efforts, when possible.

Another concern related to ex situ collections is that if they are founded with only a few individuals, mating within the population can lead to declines in genetic diversity over time due to inbreeding and genetic drift (Lacy, 1987; Willoughby et al., 2015; Foster et al., 2022). This issue has been of particular interest for collections of animals in zoos because they are often founded with only a few individuals and there may be a limited ability to source new individuals from wild populations. Previous studies found that one of the most effective strategies to maintain genetic diversity in a closed, captive animal population is to maintain as large an effective population size as possible by selecting breeding individuals that are the most genetically distant from the rest of the population (Lacy, 2012). This is accomplished by selecting individuals with the lowest mean kinship (Fernandez et al., 2004), which is the probability that an allele sampled from an individual is identical by descent with an allele at the same locus sampled at random from the population (Willoughby et al., 2015). Mean kinship in zoos is frequently assessed and managed across multiple sites as a unit (i.e., metapopulations) through careful pedigree tracking or by employing genome-wide molecular marker data (Weir and Goudet, 2017; Goudet et al., 2018), which are used to coordinate breeding efforts across institutions. Collections of critically endangered plants may be susceptible to declines in genetic diversity due to small founder population sizes, inbreeding, and drift (e.g., Foster et al., 2022; Diaz-Martin et al., 2023), but most do not have established ex situ breeding programs to manage their genetic diversity.

Given the importance of maintaining genetic diversity in collections and the fact that most ex situ conservation efforts of critically endangered plant species are hindered by a lack of insight into how much diversity is conserved, additional work is needed to ensure that collections effectively and efficiently conserve genetic diversity. One important way to do so is to use population genomic approaches to compare the genetic diversity present in wild populations and ex situ collections and assess the parentage of ex situ individuals, which can then be used to develop a strategy to ensure that as much as possible of the wild diversity is conserved (e.g., Diaz-Martin et al., 2023; Mashburn et al., 2023). Population genomic data can also be used to assess mean kinship among individuals in a collection to develop breeding strategies to maintain genetic diversity. However, the use of genomic approaches to conserve and manage genetic variation in collections of critically endangered plants is not commonplace, even though this information is crucial for ensuring that essential genetic diversity is safeguarded against irrecoverable losses.

In this study, we used an ex situ conservation program for *Karomia gigas* (Faden) Verdc. (Lamiaceae) as a case study to explore the use of population genomic approaches to evaluate and manage genetic diversity in a critically endangered, tropical tree species. *Karomia gigas* is a large tree that occurs in two widely separated coastal forest reserves in Tanzania

(ca. 100 km apart), where it is threatened by land conversion to agriculture, logging, and charcoal operations. In 2020, only 43 known wild individuals remained in two populations (Litipo and Mitundumbea). Due to the small number of individuals, the threat of habitat loss, and the risk of logging, *K. gigas* was recently listed by the International Union for Conservation of Nature (IUCN) as critically endangered (Mashimba and Shaw, 2022). Given its particularly high risk of extinction, safeguarding all of the remaining genetic diversity of the species is crucial for maintaining its viability. Thus, beginning in 2017, the Tanzania Forest Service (TFS) and Missouri Botanical Garden (MBG) began working together to establish an ex situ collection of *K. gigas*. In 2018, attempts to propagate the species from seed and cuttings resulted in successful seed germination. The seedlings were used to establish two collections: one at MBG in St. Louis, Missouri, USA, containing 28 individuals, and one at TFS in Morogoro, Tanzania, containing six individuals. However, the amount of genetic diversity remaining in the wild, the degree to which the wild diversity is safeguarded in collections, and the optimal strategy to develop and maintain genetic diversity in collections in the future are unknown.

Here, we used a population genomic approach to genotype all known individuals of *K. gigas*. We compared genetic diversity between wild populations and collections, conducted parentage analysis of the individuals in the collections, and analyzed mean kinship within the collections. Our goals were: (1) to understand how much genetic diversity is represented in the collections relative to the wild populations, (2) to analyze the contribution of wild individuals to the parentage of ex situ individuals, and (3) to devise a strategy to conduct genetically informed controlled crosses among ex situ individuals to maintain genetic diversity in the collections. We discuss the implications of the results for devising strategies to conserve maximal genetic diversity in *K. gigas* and to conduct genetically informed breeding to maintain levels of genetic diversity in the collections.

METHODS

Study species

Karomia gigas is a large tropical tree with small purple and white zygomorphic flowers and large, oval, papery fruits (Figure 1). It was first discovered in Kenya in 1977, at which time only one mature tree and one sapling were found. When the species was described (as *Holmskioldia gigas* Faden), Faden (1988) reported that the two original plants had been cut down and speculated that the species may be extinct. Intensive

FIGURE 1 Images of *Karomia gigas* depicting (A) Tanzania Forest Service staff collecting fruits from a mature, wild individual of *K. gigas* for ex situ propagation, (B) flower and flower buds, (C) developing fruits, (D) seeds, and (E) ex situ individuals grown at the Missouri Botanical Garden. Photo credits: Fandey Mashimba (A, C); Rebeca Sucher (B); Andrew Wyatt (D); Christine Edwards (E).

searches in coastal Kenya in the late 1980s and early 1990s failed to locate additional plants, and the species was again reported to be possibly extinct (Verdcourt, 1992; Beentje, 1994). Unexpectedly, a survey conducted by the Frontier-Tanzania Coastal Forest Research Programme discovered a single fruit of *K. gigas* at the edge of Ngarama Forest Reserve in Tanzania, over 600 km away from its type locality, which was verified using the type and one other herbarium specimen from the type locality (Clarke et al., 2011). Additional searches for *K. gigas* identified two populations occurring in seasonally dry tropical forests within Mitundumbea and Litipo Forest Reserves in southern coastal Tanzania (Clarke et al., 2011; Gereau et al., 2022) (Figure 2). In their published IUCN Red List assessment, Mashimba and Shaw (2022) reported only 43 known wild individuals, including 21 mature trees and 22 juveniles (i.e., young trees 5–9 m tall; no wild seedlings have been observed). The calculated area of occupancy of 16 km^2 (using 2 × 2-km grid cells) is probably an overestimate given the small number of individuals in each subpopulation. Litipo Forest Reserve has a surface area of only 10 km^2, whereas Mitundumbea, with its 15 individuals grouped into three geographically distinct clusters, has an area of 87 km^2.

Propagation of ex situ collections

We attempted to propagate *K. gigas* through seed, cuttings, air layering, and tissue culture. Although most attempts were

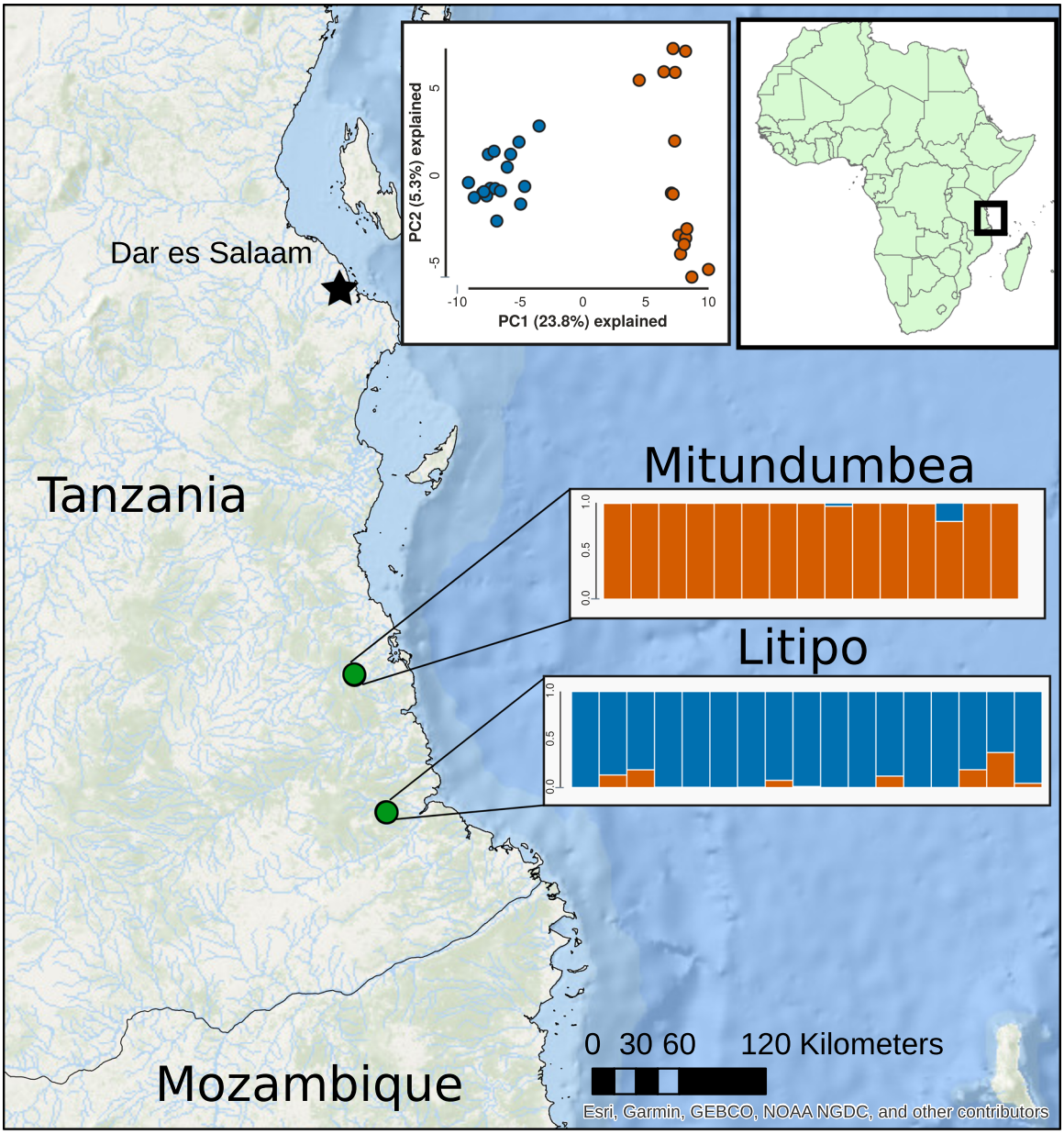

FIGURE 2 Collection localities and genetic structure of wild populations of *Karomia gigas*. Barplots depict admixture proportions inferred by STRUCTURE ($K = 2$). PCA plot of the two wild populations, Litipo (blue) and Mitundumbea (orange).

unsuccessful, propagation through seed achieved some success (Appendix S1; see Supporting Information for detailed treatments). For the MBG collection, fruits of *K. gigas* were collected from six wild maternal individuals found at Litipo and Mitundumbea. Fruits were directly collected from each individual by tree climbers. Of the more than 24,000 seeds received at MBG in 2018, only 133 appeared to be potentially viable; of these, 29 germinated and one individual died post-germination, resulting in 28 individuals at MBG. For the TFS collection, fruits of *K. gigas* were collected from two wild maternal individuals. The number of seeds used for germination attempts at TFS was not recorded but resulted in a collection of six individuals.

Sampling for genetic analyses

We collected leaf tissue for genetic analysis in 2020–2021 from a total of 77 individuals, representing all known living individuals of *K. gigas*. For the wild populations, we collected 28 individuals from Litipo Forest Reserve and 15 from Mitundumbea Forest Reserve in 2020 (Figure 2). Ex situ samples included six individuals planted outside the TFS office in Tanzania and 28 individuals from MBG's collection in the United States (collected in 2021). Each of the ex situ plants in MBG's living collections is associated with the 9-digit accession number of the wild maternal trees from which seeds were collected as documented in MBG's Living Collections Management System (https://www.livingcollections.org). Voucher specimens for the two wild populations are stored at the University of Dar es Salaam herbarium (DSM). All leaf tissue was dried and preserved in silica gel at room temperature until DNA extraction.

DNA extraction, library preparation, and sequencing

DNA extraction and high-throughput DNA sequencing library preparation were conducted in the Conservation Genetics lab at MBG. DNA was extracted using a modified cetyltrimethylammonium bromide (CTAB) protocol with an additional sorbitol wash (Doyle and Doyle, 1987; Štorchová et al., 2000). The DNA concentration of each sample was quantified using a Qubit fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Libraries were prepared using a 2b-RADseq approach (Wang et al., 2012) to generate high-quality single-nucleotide polymorphism (SNP) data across the genome following the protocol described in Linan et al. (2021). We digested 250–500 ng of DNA with the type-IIb restriction enzyme *Bcg*I (New England Biolabs, Ipswich, Massachusetts, USA), which generates 36-bp DNA fragments. Unique adapters were ligated to samples in each column of a 96-well plate. Ligated DNA samples were amplified using High Fidelity Phusion PCR mix (New England Biolabs) for 14 PCR cycles, while incorporating a uniquely barcoded PCR primer for each row, resulting in 77 uniquely dual-indexed barcoded samples per plate. The final PCR was amplified for 15 cycles, and amplification was checked using agarose gel electrophoresis. The resulting bands were excised and purified using a MinElute Gel Extraction Kit (QIAGEN, Hilden, Germany). The amount of DNA in each excised band was quantified, normalized, and pooled into one sample with a final concentration of 10 nM, which was sequenced at Northwestern University Center for Genetic Medicine on one lane of an Illumina HiSeq 4000 (Illumina, San Diego, California, USA) using 1 × 50-bp reads.

Locus assembly

Raw reads were inspected with FASTQC version 0.11.5 (Andrews, 2010), demultiplexed using a custom script from the Matz lab (https://github.com/z0on/2bRAD_denovo), and quality filtered using FASTX toolkit version 0.0.14 (options: -Q 33 -p 90). Demultiplexed reads were assembled de novo using STACKS version 2.60 (Catchen et al., 2013). Following Paris et al. (2017), we determined optimal STACKS settings for the minimum stack depth parameter (-m) and maximum distance between stacks (-M) using the R package RADstackshelpR version 0.1.0 (DeRaad, 2021). The final parameters were m = 3, M = 1, and a maximum distance between catalog loci of $n = 3$ with gapped alignments disabled. We removed loci that were absent in >50% of samples along with those with heterozygosity >0.75 to remove potentially paralogous loci (Paris et al., 2017) and retained one random SNP per locus (to avoid linkage among loci). The resulting data set that contained all 77 samples was used to analyze population genetic diversity, parentage, and internal relatedness (IR; see description of analyses below). For STRUCTURE and principal component analysis (PCA; see below), which are sensitive to missing data, we created a reduced data set in which we removed samples with >40% missing data, resulting in the removal of 14 samples, leaving 63 samples.

Genetic diversity of wild and ex situ populations

Genetic diversity was examined by grouping samples in two ways. First, we compared genetic diversity measures between individuals grouped into wild (containing both Litipo and Mitundumbea samples; $n = 43$) and ex situ (containing both MBG and TFS samples; $n = 34$) populations. We then repeated the analysis, grouping individuals into the four populations: Litipo ($n = 28$), Mitundumbea ($n = 15$), MBG ($n = 28$), and TFS ($n = 6$). Using STACKS, we calculated observed heterozygosity (H_o), expected heterozygosity (H_e), nucleotide diversity (π), private alleles, and Wright's inbreeding coefficient (F_{IS}; Wright, 1922) based on the SNP data set, which contains only variable sites. Although nucleotide diversity is traditionally analyzed using entire loci containing both invariant and variant sites, our

analyses of nucleotide diversity were based on SNP data and were used to compare genetic diversity among wild and ex situ populations in this study; thus, the estimates contained in the present study should not be used to compare genetic diversity with other taxa. To correct for differences in population sample size when comparing estimates of allelic diversity, we calculated rarified allelic richness (A_R) using HP-Rare version 1.1 (Kalinowski, 2005). Finally, we calculated Tajima's *D* statistic (Tajima, 1989) for wild *K. gigas* populations to test for signatures of a genetic bottleneck using VCFtools version 0.1.15 (Danecek et al., 2011) as implemented in the R package r2vcftools version 0.0.0.9 (https://rdrr.io/github/nspope/r2vcftools/), with 1000 simulations of allele frequency spectrum under the neutral model to test for significance.

Assessing the genetic representation of wild individuals in the ex situ collection

We performed parentage analysis in Cervus version 3.0.7 to identify the parents of each ex situ individual (Marshall et al., 1998; Kalinowski et al., 2007). All wild individuals were used as candidate parents in a simulation of parentage analysis using the "parent pair, sexes unknown" option in Cervus, with the following settings: 10,000 offspring simulated, 43 candidate parents (wild populations), "Proportion sampled" = 1, "Proportion loci typed" = 0.5, "Minimum typed loci" set to 50% of all loci, and all other parameters set to the defaults. Parentage analysis was conducted using the "Parent pair, sexes unknown" option using default parameters, including "the two most likely parents" for each offspring from the list of candidate parents. This analysis can select the same candidate parent twice if the individual is the result of selfing (or biparental inbreeding). Confidence in parentage assignment was assessed via the likelihood ratio of each candidate parent (taking account of possible typing errors), in the form of a natural log of the likelihood ratio (LOD) score. We also inspected the results to identify samples with a large number of mismatching loci and non-significant tri-loci scores compared to their most likely parents, which may represent individuals in which one parent may not be in the database, potentially indicating undiscovered wild individuals.

To confirm the status of progeny identified as being the product of selfing/biparental inbreeding, we calculated IR, a measure of the proportion of homozygous loci in an individual, using the R function GENHET version 3.1 (Coulon, 2010). Any sample resulting from a selfing event is expected to exhibit a significant increase in homozygosity compared to parental individuals. To test for differences in IR among selfed collections and wild populations, we conducted one-way ANOVA analyses with post-hoc Tukey honest significant difference (HSD) pairwise comparisons. The results of parentage and IR analysis were used to generate lists of wild individuals serving as parents contributing to each population and those not represented in collections. To assess the contribution of the parental genotypes to the collections populations, we: (1) calculated the proportion of each population derived from each pair of parents (i.e., full-sib groups), and (2) calculated the proportion of each population derived from each unique genotype, where a selfed individual is a contribution of 1 from the parental genotype, and an outcrossed individual is a contribution of 0.5 from each parental genotype.

To validate results of parentage analysis and assess how genetic variation is structured within and among wild populations and collections, we analyzed the data using PCA and Bayesian clustering in STRUCTURE version 2.3.4 (Pritchard et al., 2000). For both analyses, we analyzed patterns of genetic structure in two data sets: (1) only wild individuals, and (2) both wild and ex situ individuals. PCA and STRUCTURE were performed and visualized using the analysis toolkit in iPyrad version 0.9.92 (Eaton and Overcast, 2020). In STRUCTURE, samples were assigned to genetic clusters, with assignments to two or more clusters indicating admixture between genetic clusters. We ran five separate runs at each *K* from 1 to 4 using an admixture model and correlated allele frequencies, with a burn-in of 100,000 generations and a run length of 300,000 generations. We used both delta *K* and plots of the -Ln likelihood values at each *K* in STRUCTURE HARVESTER (Earl and vonHoldt, 2012) to estimate the optimal number of clusters in the data set.

To help guide decisions about controlled crosses in the collections, we estimated pairwise kinship (coancestry) between all sampled individuals using $\hat{\beta}_{jj'}$ ("beta"; Weir and Goudet, 2017), which estimates the kinship coefficient between a pair of samples relative to the average kinship of all pairs of samples. The kinship coefficient was calculated using the "beta.dosage" function in the R package HIERFSTAT version 0.5-11 (Goudet, 2005).

RESULTS

Sequencing and locus assembly

Sequencing returned 235,118,486 total reads with a Phred score greater than 33, with each sample averaging 3,014,340 reads (range 133,082 to 8,923,992 reads per sample). The STACKS assembly produced a matrix containing 77 individuals genotyped at 1183 loci with 16.7% missing data. For STRUCTURE and PCA, which are sensitive to missing data, we created a reduced data set in which we removed 14 samples with >40% missing data (11 samples from Litipo, one from TFS, and two from MBG), leaving 63 individuals.

Genetic diversity measures across wild and ex situ populations

When comparing genetic diversity between all wild populations (Litipo + Mitundumbea) and ex situ (MBG + TFS) collections, wild populations displayed greater levels of

TABLE 1 Genetic diversity metrics calculated for each *Karomia gigas* population and for populations combined into "wild" and "ex situ" groups.

Population	Private alleles	A_R	Sites	Polymorphic sites	% Polymorphic loci	H_o	H_e	π	F_{IS}
Litipo	184	1.35	1182	739	62.5	0.157	0.149	0.155	0.008
Mitundumbea	144	1.36	1182	745	63.0	0.160	0.156	0.163	0.018
MBG	78	1.34	1182	757	64.0	0.128	0.136	0.139	0.054
TFS	15	1.28	1177	301	25.6	0.143	0.099	0.123	−0.037
Ex situ (combined)	95	1.61	1182	812	68.7	0.128	0.139	0.142	0.057
Wild (combined)	370	1.73	1182	1087	92.0	0.154	0.167	0.170	0.068

Note: A_R = allelic richness; H_o = observed heterozygosity; H_e = expected heterozygosity; π = nucleotide diversity; F_{IS} = inbreeding coefficient; MBG = Missouri Botanical Garden; TFS = Tanzania Forest Service.

diversity across all metrics (Table 1). Notably, wild populations had a greater number of private alleles (370 vs. 95), polymorphic sites (1087 vs. 812), H_o (0.154 vs. 0.128), H_e (0.167 vs. 0.139), π (0.170 vs. 0.142), and A_R (1.73 vs. 1.61) than collections, respectively. When comparing genetic diversity metrics among populations, the two wild populations had nearly identical values for A_R, polymorphic loci, H_o, H_e, π, and F_{IS}, although Litipo had a greater number of private alleles than Mitundumbea (184 vs. 144; Table 1), possibly due to Litipo's larger population size. MBG's collection had a greater number of polymorphic loci than Litipo or Mitundumbea (757 vs. 739 and 745, respectively), higher F_{IS} values (0.054 vs. 0.008 and 0.018), similar values of A_R (1.34 vs. 1.35 and 1.36), and lower H_o (0.128 vs. 0.157 and 0.160) and H_e (0.136 vs. 0.149 and 0.156) values than Litipo or Mitundumbea, respectively (Table 1). TFS's collection had the lowest diversity across all genetic diversity metrics, likely due to its small number of individuals (six), but also had the lowest inbreeding coefficient (F_{IS}; Table 1). Finally, Tajima's D was 0.15 ($P = 0.004$) for wild populations (Litipo + Mitundumbea), indicating a population bottleneck.

Parentage analysis of ex situ individuals

We successfully assigned parents to 33 of the 34 ex situ individuals; the sole individual for which parentage analysis was unsuccessful had >50% missing data. However, three additional individuals, all from the same accession and mother tree (LC_2018_1457_1_MT1, LC_2018_1457_2_MT1, LC_2018_1457_3_MT1), had non-significant LOD scores, indicating lower confidence in the inference of their parents (Table 2). Because parentage was not confidently assigned, we removed these individuals from all subsequent analyses. Of the 30 individuals with confident parentage assignments, 29 were inferred to be the result of reproduction within the same population and only one individual was inferred to be a cross between wild populations (Table 2, Figure 3), indicating little gene flow between the wild populations.

Results of parentage analysis showed that a large proportion of individuals in the collection were the result of self-fertilization. Of the 30 individuals with confident matches, parentage analysis indicated that 16 ex situ individuals (53.3%) were the result of outcrossing, whereas 14 (46.7%) were the result of self-fertilization (selfed). Thirteen of the 14 selfed ex situ individuals occurred in the MBG population and originated from seed collected from one individual at Mitundumbea (MT03), whereas only one TFS individual was produced via selfing and originated from Litipo (Table 2). To validate the results of parentage, we compared the IR of individuals that were found to be selfed to that of wild individuals and outcrossed individuals in the collection. As expected, the average IR of selfed individuals (average −0.078) was significantly greater (i.e., more homozygous) than that of wild individuals (average −0.167) (Table 2; Appendices S2, S3).

Parentage analyses showed that only a small proportion of wild individuals served as parents and that one parent contributed disproportionately to each of the two collections. In the 30 ex situ individuals with confident parentage matches, only 14 of the 43 wild individuals (32.6%) served as parents, whereas 29 (67.4%) had no offspring represented in the collections (Appendix S2). Of the 14 parents, 10 occur in Litipo and four occur in Mitundumbea (Figure 3). In the MBG collection, individuals originated from parents in both Litipo and Mitundumbea (Table 2, Figure 3). Eight wild genotypes served as parents to the 24 MBG ex situ individuals, forming 11 unique full-sib groups (Figure 3). In MBG, the majority of individuals (13 of 24) were derived from selfing by MT03, such that genotype MT03 was by far the most common parent, serving as a parent for 58% of the MBG population (Figure 3). Based on the maternal individuals identified by parentage analysis, the MBG ex situ collection likely originated from just four maternal trees, even though collection records stated that the seeds originated from six. The TFS population originated only from parents in Litipo. Five wild genotypes served as a parent to the six TFS ex situ individuals, forming five unique full-sib groups (Figure 3). However, genotype LT01 was over-represented, serving as a parent in 50% of the TFS population (Figure 3).

We next assessed the extent to which the parent reported in collection records corresponded to the results of parentage

TABLE 2 Results of parentage analysis for individuals in the ex situ collections at Missouri Botanical Garden (MBG) and the Tanzania Forest Service (TFS), along with internal relatedness (IR) measures.

Sample ID[a]	Population	Inferred Parent 1	Inferred Parent 2	Fertilization method	Trio loci compared[b]	Trio loci mismatching[b]	Trio LOD score[c]	IR
LC_2018_1449_LT1	MBG	LT1	LT21	outcrossed	1580	31	97.6*	−0.127
LC_2018_1450_1_LT7	MBG	LT1	LT6	outcrossed	444	12	13.7*	0.007
LC_2018_1450_2_LT7	MBG	LT1	LT8	outcrossed	1544	26	84.9*	−0.007
LC_2018_1450_3_LT7	MBG	LT1	LT21	outcrossed	1425	17	107.8*	−0.176
LC_2018_1452_3_MT4	MBG	MT3	MT8	outcrossed	1641	43	25.2*	−0.078
LC_2018_1452_4_MT4	MBG	MT3	MT3	selfed	1426	33	90.9*	−0.039
LC_2018_1452_5_MT4	MBG	MT3	MT3	selfed	1399	40	68.4*	−0.136
LC_2018_1452_6_MT4	MBG	MT3	MT3	selfed	1517	45	75.3*	−0.026
LC_2018_1452_7_MT4	MBG	MT3	MT3	selfed	1371	30	126.4*	−0.044
LC_2018_1452_8_MT4	MBG	MT3	MT3	selfed	1473	40	78.6*	−0.089
LC_2018_1452_9_MT4	MBG	MT3	MT3	selfed	1447	34	134.0*	−0.107
LC_2018_1452_10_MT4	MBG	MT3	MT3	selfed	1476	42	71.2*8	−0.24
LC_2018_1452_11_MT4	MBG	MT3	MT3	selfed	1355	41	90.6*	−0.013
LC_2018_1452_12_MT4	MBG	MT3	MT3	selfed	1331	30	104.1*	−0.168
LC_2018_1454_1_MT2	MBG	MT2	MT6	outcrossed	1629	36	125.5*	−0.067
LC_2018_1454_2_MT2	MBG	LT18	MT2	outcrossed	1471	27	36.0*	−0.112
LC_2018_1454_3_MT2	MBG	MT2	MT8	outcrossed	1668	52	3.7*	−0.182
LC_2018_1455_1_MT1	MBG	LT1	LT17	outcrossed	1255	18	116.0*	−0.215
LC_2018_1455_2_MT1	MBG	LT1	LT14	outcrossed	847	15	36.1*	−0.076
LC_2018_1456_1_MT3	MBG	MT3	MT3	selfed	1146	27	99.9*	0.021
LC_2018_1456_2_MT3	MBG	MT2	MT3	outcrossed	1488	32	35.2*	−0.07
LC_2018_1457_1_MT1	MBG	MT1	MT9	outcrossed	1392	51	−24.5	−0.208
LC_2018_1457_2_MT1	MBG	LT18	MT1	outcrossed	1307	39	−81.3	−0.221
LC_2018_1457_3_MT1	MBG	LT18	MT1	outcrossed	1202	35	−57.9	−0.202
LC_2018_1458_1_MT3	MBG	MT3	MT3	selfed	1436	29	129.2*	−0.067
LC_2018_1458_2_MT3	MBG	MT3	MT3	selfed	1258	40	18.4*	−0.038
LC_2018_1458_3_MT3	MBG	MT3	MT3	selfed	1040	31	38.9*	−0.096
SD57_FM46	TFS	LT1	LT2	outcrossed	1299	44	22.1*	−0.173
SD58_FM47	TFS	LT1	LT20	outcrossed	770	15	42.2*	−0.018
SD59_FM46	TFS	LT2	LT27	outcrossed	1384	39	49.3*	−0.124
SD60_FM47	TFS	LT1	LT1	selfed	1125	22	124.3*	−0.045
SD61_FM48	TFS	LT1	LT6	outcrossed	555	23	18.4*	−0.043
SD77_FM48	TFS	LT1	LT6	outcrossed	933	17	86.3*	−0.058

Note: LOD = natural log of the likelihood ratio.

[a]Sample LC_2018_1452_1_MT4 was not genotyped at enough loci to infer parents and was excluded from parentage analysis. Individuals that did not show high confidence in parentage analysis are indicated in italics.

[b]"Trio loci compared" are the number of shared loci that were analyzed between the sample (Sample ID) and inferred parents, while "Trio loci mismatching" are number of mismatches in loci between inferred parents and offspring (Sample ID).

[c]Significant LOD scores are indicated by an asterisk.

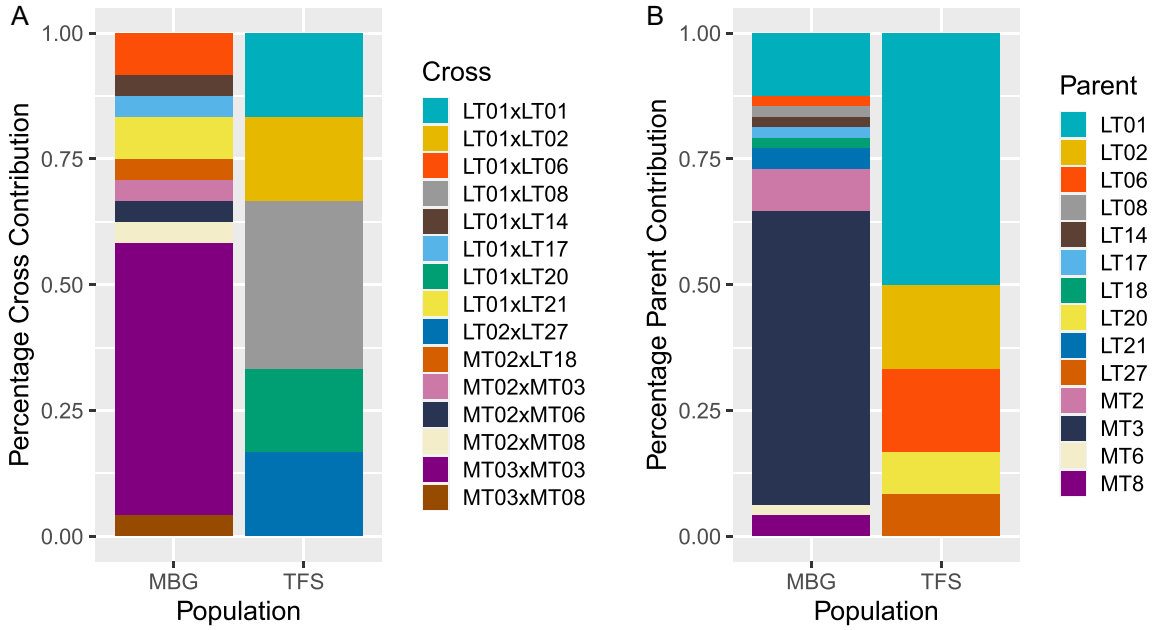

FIGURE 3 Genotypic and parental contribution to the ex situ collection. The relative contribution of (A) each full-sib cross and (B) each parent to the Missouri Botanical Garden (MBG) and Tanzania Forest Service (TFS) ex situ populations.

analysis for the ex situ individuals from MBG, with the assumption that at least one of the genetically inferred parents should match the maternal individual reported in collection records. Of the 24 individuals in the collection at MBG for which we were able to assign parentage, 15 showed a mismatch between the recorded maternal parent in collection records and the parents identified through parentage analysis. Thirteen of the 15 mismatches correctly identified the population of origin but misidentified the maternal individual that produced the seeds within the population; for example, seeds recorded as originating from MT4 were likely collected from MT3 based on parentage analysis. Both the source population and the maternal plant were misidentified for two trees (LC_2018_1455_1_MT1 and LC_2018_1455_2_MT1) that were recorded as originating from Mitundumbea but for which both of the most likely parents occur in Litipo.

Analyses of genetic structure

The results of the STRUCTURE analysis for the wild populations (Litipo and Mitundumbea) suggested an optimal K of 2 based on both the Evanno method (ΔK) and the K value at which log likelihood values (ln Pr[X | K]) reached a plateau (Appendix S4). The two populations each formed distinct genetic clusters both in STRUCTURE and in PCA, with little admixture inferred between them, although the Litipo population contained a few individuals that showed a small amount of ancestry from the Mitundumbea genetic cluster (Figure 2). In the PCA, PC1 explained 23.8% of the variation and divided samples into two distinct groups corresponding to the Litipo and Mitundumbea populations. PC2, which explained 5.3% of the variation, separated Mitundumbea into two groups (Figure 2).

The STRUCTURE analysis including both wild and ex situ individuals (Figure 4) also showed $K = 2$ as the optimal value of K, with wild individuals grouped by population of origin as described above. The placement of ex situ individuals into STRUCTURE clusters generally agreed with the results of parentage analysis. Ex situ individuals clustered with one of the two wild populations, indicating their origin; the MBG population had individuals that clustered with both wild populations, whereas the TFS population clustered only with Litipo individuals. The assignment of ex situ individuals to source populations largely agreed with collection records except for two individuals at MBG in which the source population was likely mislabeled (LC_2018_1455_1_MT1 and LC_2018_1455_2_MT1). The placement of ex situ individuals in the PCA analysis also corresponded closely with the results of parentage analysis; the ex situ individuals identified as being the result of self-fertilization were placed in close proximity to the parent identified by parentage analysis (e.g., the tight group of ex situ individuals clustering with one wild individual in the lower left corner of Figure 4B), whereas outcrossed individuals were generally placed in an intermediate position between the two parents identified by parentage analysis (e.g., the ex situ individuals in the center-left of Figure 4B).

Analysis of relatedness for captive breeding

Finally, we analyzed kinship coefficients, as measured by beta, between pairs of samples to generate strategies for a captive breeding strategy to cross the most genetically

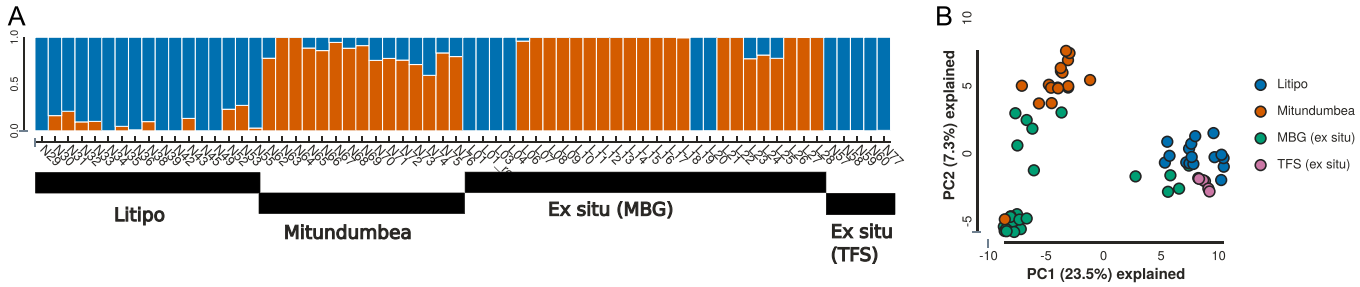

FIGURE 4 Analysis of genetic structure in *Karomia gigas* wild and ex situ populations. (A) STRUCTURE plot ($K = 2$), organized by population. (B) PCA plot depicting the four populations assessed. Note that the large cluster of individuals in the extreme lower left corner of the PCA shows the numerous selfed offspring in MBG collections that originated from a wild individual from Mitundumbea.

divergent individuals to maintain genetic diversity. As expected, the mean kinship coefficients within each of the four populations (MBG, TFS, Mitundumbea, and Litipo) showed similarly positive values ranging from 0.417 (TFS) to 0.483 (MBG), whereas comparisons between individuals from different populations were lower, reflecting lower kinship (Appendix S5). When comparing individuals within the MBG population, we found the lowest pairwise kinship coefficients between individuals originating from different wild populations (−0.2 to 0.08), indicating low kinship.

DISCUSSION

Here, we used the ex situ conservation program for *Karomia gigas* as a case study to evaluate the utility of population genomic assessments for managing the genetic diversity in collections of critically endangered tropical trees. To identify inefficiencies in collections and pinpoint priorities for increasing the ex situ representation of wild individuals, we compared genetic diversity between wild populations and ex situ collections and conducted parentage analysis to understand the contribution of wild individuals to the ex situ collection. We also analyzed kinship among individuals in collections to develop a captive breeding program. The study revealed insights into the genetic composition of an ex situ collection, with important implications for ensuring that extant genetic diversity is conserved and maintained in ex situ collections through a captive breeding program.

One goal of our study was to understand how much genetic diversity is represented in the collections of *K. gigas* relative to wild populations. The wild populations showed a signature of a genetic bottleneck, which was expected given the large reduction in the species' range. The two wild populations exhibited comparable levels of diversity despite differences in population size, but the two collections exhibited lower diversity than either wild population (Table 1). When comparing all wild vs. all ex situ individuals, the wild populations displayed slightly higher levels of genetic diversity than the collections, as has been reported in previous studies (Hoban et al., 2020; Diaz-Martin et al., 2023; Mashburn et al., 2023). This is unsurprising given that the collections of *K. gigas* were established from seed collected from only a few wild individuals. However, despite having lower diversity, we found that the collections conserved about 75% of the polymorphic loci found in the wild population (Table 1), suggesting that a fair amount of the wild genetic diversity is represented in the collection as a whole. Two factors that have likely facilitated the maintenance of genetic diversity in the ex situ conservation efforts are that both wild populations are represented in the collections and that more than 50% of the individuals were produced from outcrossing. Interestingly, another recent study also showed only a small reduction in genetic diversity in a collection of an open-pollinated tropical tree established from seed (Diaz-Martin et al., 2023). These results suggest that establishing collections from seed can be an effective way to conserve a moderate amount of the genetic diversity of open-pollinated tropical trees, even when they are established from a subset of wild individuals.

For critically endangered species, however, the goal for ex situ collections is often to represent all of the genetic diversity found in known wild individuals; we therefore conducted parentage analysis to identify which wild individuals contributed to the collection. Parentage analysis revealed that individuals from both Litipo and Mitundumbea (wild populations) are currently represented in collections, but only a small subset (32.6%) of the total known wild individuals of *K. gigas* are represented ex situ. The small proportion of wild individuals represented in collections is likely due to issues such as a lack of reproductive maturity for some individuals (which represents ca. 50% of the wild population) and difficulty collecting fruits from a large canopy tree (Figure 1). To increase the representation of wild individuals and increase genetic diversity in collections, we recommend prioritizing for additional seed collection efforts any mature individuals that have not previously served as a parent for the collections, as well as those with the lowest kinship coefficients with the existing parents of the collections (Appendix S5). Monitoring of the wild juvenile trees for reproductive maturity will also be necessary so that their seeds can be added to the ex situ collections as they begin to produce fruit.

Because a collection of large tropical trees is difficult to maintain in most botanical gardens due to space constraints, and particularly so when the botanical garden is located in a temperate region and the collection must be maintained in greenhouses, we investigated how to maximize genetic diversity in the most efficient way possible in *K. gigas*. Parentage analysis in *K. gigas* revealed a disproportionately large contribution of some wild parents to the ex situ collection; for example, the majority of individuals in the MBG population (13 of 24) were produced through selfing by MT3, and one individual in TFS was derived from selfing by LT1. Because MT3 and LT1 were also parents to other outcrossed individuals in the collections, these selfed individuals add nothing in terms of representing wild parents or adding unique genetic diversity, revealing some inefficiency in the collections. Given the few individuals remaining in the species, these selfed individuals are still valuable, but some could be distributed to other botanical gardens located in tropical areas where they would be easier to maintain, which would also serve as an additional safeguard against the extinction of the species. Furthermore, over the long term, given the large size of *K. gigas* trees, we will likely need to use a metacollection strategy paired with tissue culture to maintain the genetic diversity of *K gigas* in ex situ collections. These results illustrate how genetic analysis can help identify inefficiencies in an ex situ collection, potentially saving space, time, and money that could be used to improve conservation capacity. Although previous studies employed a resampling analysis to understand whether a subset of individuals in collections can capture maximal wild genetic diversity in the most efficient way possible (i.e., following Namoff et al., 2010; Griffith et al., 2015, 2021; Wei and Jiang, 2020), this is not currently needed in *K. gigas* given the small number of individuals currently in the collections. However, such analyses may be necessary in the future if additional ex situ individuals are added to the collection.

Another interesting result is the possibility that at least one unknown mature tree may have contributed to the ex situ collection. Parentage could not be confidently assigned to three individuals at MBG, all originating from the same maternal individual (MT1) and fruit. The genetic mismatch between these individuals and their most likely parents suggests that one parent may be an unknown or unsampled wild individual. Supporting this hypothesis is the relatively large number of private alleles in the MBG population given its status as a collection (Table 1), which may be attributable to the unique alleles found in the collection that were not sampled in the wild. Additional searches are necessary to identify additional wild individuals of *K. gigas*, illustrating how parentage analysis of collections can provide unexpected insights into wild populations, such as identifying the possible existence of unknown wild individuals.

Another goal of the study was to determine how to conduct genetically informed controlled crosses among ex situ individuals to maximize genetic diversity in subsequent generations. A recent study found that because trees take a long time to reach reproductive maturity and individuals in collections may not become reproductively mature at the same time, open pollination in collections of slow-growing, self-compatible tropical trees can lead to high rates of selfing, causing dramatic declines in genetic diversity and a high risk of inbreeding depression in the next generation (i.e., the captive-born generation; Diaz-Martin et al., 2023). The collection of *K. gigas* currently contains many individuals produced by selfing and many individuals that are half or full sibs (Figure 3); thus, allowing passive, open pollination would likely lead to high inbreeding and declines in genetic diversity in subsequent generations, as was found previously in ex situ collections of other tropical plants (Foster et al., 2022; Diaz-Martin et al., 2023). Careful planning of suitable crosses is therefore necessary to avoid inbreeding. We calculated pairwise kinship among all *K. gigas* individuals (Appendix S5); as ex situ individuals become reproductively mature, they will be crossed with those with the lowest kinship to help maintain genetic diversity and minimize inbreeding. At MBG, captive breeding efforts have already begun and, so far, have involved crossing individuals originating from different wild populations. For TFS, whose collection originated only from Litipo, captive breeding may include hand pollinations with pollen collected from individuals at Mitundumbea. Given the limited gene flow between wild populations, any offspring resulting from these crosses could be used for augmentations to increase genetic diversity in natural populations ("assisted gene flow"; Sgrò et al., 2011; Aitken and Whitlock, 2013).

However, one concern with the approach of minimizing kinship between individuals is that crossing individuals from different source populations could result in outbreeding depression, which can occur when two locally adapted individuals cross and produce offspring that are poorly adapted to both parental environments. However, trees in general are predominantly outcrossing and have several features that facilitate high rates of gene flow (Petit and Hampe, 2006), making outbreeding depression somewhat unlikely. Regardless, it will be useful in the future to conduct crosses between individuals with the lowest kinship originating both from the same source population and from different source populations to see if either shows lower fitness. Monitoring the fitness of offspring derived from all of the types of crosses will be important to help fine-tune captive breeding and reintroduction efforts in the future.

Finally, results from parentage analysis highlight why using provenance alone (without genetic analysis) may be problematic for devising a captive breeding strategy for plants. In the present study, we found a mismatch between the recorded maternal individual and the parents identified using parentage analysis for a significant proportion of the collection. One of the main causes for these mismatches likely arose due to issues with relocating individuals between collecting trips to sample seeds and DNA. When the seed was collected in 2018, the locations of the

individuals from which seeds were collected were recorded with a low-resolution GPS, but individuals were not physically tagged in the field. When researchers returned to collect DNA samples in 2020, the low resolution of the GPS coordinates led to difficulty in understanding which tree corresponded to the previous GPS points, such that researchers did their best to assign a name to each tree while sampling. In 2020, the GPS coordinates were re-recorded along with the inferred identity of the tree. However, our results suggest that the identity of several individuals was misassigned in 2020. Because the GPS coordinates taken in 2020 for DNA allow us to relocate each tree and the parentage results allow us to identify which wild individuals gave rise to the collection, we will update the parents recorded in our collections management system to match those identified by parentage analysis. Generally, however, these issues with provenance data are not uncommon (e.g., Diaz-Martin et al., 2023) and highlight why genetic analysis is so important for informing conservation efforts of critically endangered species.

Furthermore, even in cases where the maternal parent is accurately identified, we recommend conducting parentage analysis to devise captive breeding strategies for ex situ collections because the pollen donor is usually unknown, such that it is unknown whether ex situ individuals are the result of selfing or outcrossing (Diaz-Martin et al., 2023). Because making crosses based on erroneous or incomplete source data could result in inbreeding, we therefore advocate the use of marker-based kinship estimators to inform crosses (Goudet et al., 2018), which provides all necessary information to ensure that captive breeding efforts in collections of critically endangered trees maintain genetic diversity.

CONCLUSIONS

This case study provides important insights into why population genomics and parentage analysis are important for managing genetic diversity in collections of long-lived tropical trees. Parentage analysis based on population genomic data can help identify individuals that are not currently represented ex situ, and therefore become targets for future seed collection efforts. It may also identify individuals with a parent not yet identified in the wild, providing information about whether searches for additional wild individuals may be necessary. Parentage analysis can also identify individuals that are the product of selfing or those that represent the same full-sib groups, which could be used to streamline collections, improving the efficiency of ex situ conservation efforts. Finally, the information provided by population genomic data regarding both the parentage and the kinship of individuals in a collection can provide important insights into the most suitable crosses to make in a captive breeding program to maintain genetic diversity. Because of the small number of *K. gigas* individuals, our genomic analysis was not especially expensive, but the genetic approach used here significantly improves our ability to conserve genetic diversity in this critically endangered species; we therefore recommend its use for improving the efficiency and effectiveness of conserving genetic diversity in ex situ conservation programs for a wide range of critically endangered plant species.

AUTHOR CONTRIBUTIONS

C.E.E., A.W., R.S., F.H.M., and R.E.G. planned and designed the research. F.H.M., R.S., A.W., and R.E.G. developed the ex situ collection. B.B. generated the genomic data, and A.G.L. and C.E.E. analyzed the data and wrote the first draft. All authors edited and approved the final version of the manuscript.

ACKNOWLEDGMENTS

The authors thank Botanic Gardens Conservation International (BGCI) and Kirsty Shaw for organizing initial fieldwork; the staff of the Tanzania Forestry Service (TFS), including Lenin Festo, for facilitating the Material Transfer Agreement between Tanzania and the Missouri Botanical Garden (MBG); Yahya Abeid for assistance in fieldwork; and Mzee Jangwa for monitoring the phenology of individuals in the Mitundumbea population. The Conservation Genetics Program at MBG is supported by private donations from Phillip and Sima Needleman, Stephen and Camilla Brauer, and the Bellwether Foundation.

DATA AVAILABILITY STATEMENT

Data are available on the National Center for Biotechnology Information (NCBI)'s Sequence Read Archive under BioProject ID PRJNA1027492.

ORCID

Alexander G. Linan http://orcid.org/0000-0002-8281-1346
Christine E. Edwards http://orcid.org/0000-0001-8837-4872

REFERENCES

Aitken, S. N., and M. C. Whitlock. 2013. Assisted gene flow to facilitate local adaptation to climate change. *Annual Review of Ecology, Evolution, and Systematics* 44: 367–388.

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. Website: http://www.bioinformatics.babraham.ac.uk/projects/fastqc [accessed 29 April 2024].

Beentje, H. J. 1994. Kenya trees, shrubs and lianas. National Museums of Kenya, Nairobi, Kenya.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: An analysis tool set for population genomics. *Molecular Cell* 22: 3124–3140.

Center for Plant Conservation. 2019. CPC best plant conservation practices to support species survival in the wild. Center for Plant Conservation, Escondido, California, USA.

Clarke, G. P., N. D. Burgess, F. M. Mbago, C. Mligo, B. Mackinder, and R. E. Gereau. 2011. Two 'extinct' trees rediscovered near Kilwa, Tanzania. *Journal of East African Natural History* 100: 133–140.

Cohen, J. I., J. T. Williams, D. L. Plucknett, and H. Shands. 1991. Ex situ conservation of plant genetic resources: Global development and environmental concerns. *Science* 253: 866–872.

Coulon, A. 2010. Genhet: An easy-to-use R function to estimate individual heterozygosity. *Molecular Ecology Resources* 10: 167–169.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.

DeRaad, D. 2021. RADStackshelpR. Website: https://devonderaad.github.io/RADstackshelpR/index.html [accessed 29 April 2024].

De Souza, L. M., V. Le Guen, C. B. M. Cerqueira-Silva, C. C. Silva, C. C. Mantello, A. R. O. Conson, J. P. G. Vianna, et al. 2015. Genetic diversity strategy for the management and use of rubber genetic resources: More than 1,000 wild and cultivated accessions in a 100-genotype core collection. *PLoS ONE* 10: e0134607.

Diaz-Martin, Z., J. Fant, K. Havens, W. Cinea, J. M. Tucker Lima, and M. P. Griffith. 2023. Current management practices do not adequately safeguard endangered plant species in conservation collections. *Biological Conservation* 280: 109955.

Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

Earl, D., and B. VonHoldt. 2012. Structure Harvester: A website and program for visualizing Structure output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359–361.

Eaton, D. A. R., and I. Overcast. 2020. Ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics* 36: 2592–2594.

Faden, R. B. 1988. *Holmskioldia* (*Verbenaceae*), a genus new to the flora of Tropical East Africa. *Kew Bulletin* 43: 659–662.

Falk, D. A., and K. E. Holsinger. 1991. Genetics and conservation of rare plants. Oxford University Press, New York, New York, USA.

Fernandez, J., M. Toro, and A. Caballero. 2004. Managing individuals' contributions to maximize the allelic diversity maintained in small, conserved populations. *Conservation Biology* 18: 1358–1367.

Foster, J. A., S. K. Walsh, K. Havens, A. T. Kramer, and J. B. Fant. 2022. Supporting long-term sustainability of ex situ collections using a pedigree-based population management approach. *Applications in Plant Sciences* 10: e11491.

Gereau, R. E., Y. S. Abeid, F. H. Mashimba, K. Shaw, R. Sucher, and A. Wyatt. 2022. Saving Tanzanian trees from extinction: An integrated approach. *In* Y. Harvey-Brown [ed.], Diversity and conservation of African plants: Challenges and opportunities, 39–40. 22nd AETFAT Congress Abstracts 27 June – 1 July 2022, Livingstone, Zambia.

Goudet, J. 2005. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes* 2: 184–186.

Goudet, J., T. Kay, and B. S. Weir. 2018. How to estimate kinship. *Molecular Ecology* 27: 4121–4135.

Griffith, M. P., M. Calonje, A. W. Meerow, F. Tut, A. T. Kramer, A. Hird, T. M. Magellan, and C. E. Husby. 2015. Can a botanic garden cycad collection capture the genetic diversity in a wild population? *International Journal of Plant Sciences* 176: 1–10.

Griffith, M. P., T. Clase, P. Toribio, Y. E. Piñeyro, F. Jimenez, X. Gratacos, V. Sanchez, et al. 2020. Can a botanic garden metacollection better conserve wild plant diversity? A case study comparing pooled collections with an ideal sampling model. *International Journal of Plant Sciences* 181: 485–496.

Griffith, M. P., F. Cartwright, M. Dosmann, J. Fant, E. Freid, K. Havens, B. Jestrow, et al. 2021. Ex situ conservation of large and small plant populations illustrates limitations of common conservation metrics. *International Journal of Plant Sciences* 182: 263–276.

Guerrant, E. O., K. Havens, and P. Vitt. 2014. Sampling for effective ex situ plant conservation. *International Journal of Plant Sciences* 175: 11–20.

Hoban, S., T. Callicrate, J. Clark, S. Deans, M. Dosmann, J. Fant, O. Gailing, et al. 2020. Taxonomic similarity does not predict necessary sample size for ex situ conservation: A comparison among five genera. *Proceedings of the Royal Society B: Biological Sciences* 287: 20200102.

Humphreys, A. M., R. Govaerts, S. Z. Ficinski, E. Nic Lughadha, and M. S. Vorontsova. 2019. Global dataset shows geography and life form predict modern plant extinction and rediscovery. *Nature Ecology & Evolution* 3: 1043–1047.

Kalinowski, S. T. 2005. hp-rare 1.0: A computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* 5: 187–189.

Kalinowski, S. T., M. L. Taper, and T. C. Marshall. 2007. Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16: 1099–1106.

Kramer, A. T., A. Hird, K. Shaw, M. Dosmann, and R. Mims. 2011. Conserving North America's threatened plants: Progress report on Target 8 of the Global Strategy for Plant Conservation. Botanic Gardens Conservation International, Glencoe, Illinois, USA.

Lacy, R. C. 1987. Loss of genetic diversity from managed populations: Interacting effects of drift, mutation, immigration, selection, and population subdivision. *Conservation Biology* 1: 143–158.

Lacy, R. C. 2012. Extending pedigree analysis for uncertain parentage and diverse breeding systems. *Journal of Heredity* 103: 197–205.

Linan, A. G., P. P. Lowry, A. J. Miller, G. E. Schatz, and J. S. Christine. 2021. RAD-sequencing reveals patterns of diversification and hybridization, and the accumulation of reproductive isolation in a clade of partially sympatric, tropical island trees. *Molecular Ecology* 30: 4520–4537.

Marshall, T. C., J. Slate, L. E. B. Kruuk, and J. M. Pemberton. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* 7: 639–655.

Mashburn, B., R. Jhangeer-Khan, A. Bégué, V. Tatayah, K. M. Olsen, and C. E. Edwards. 2023. Genetic assessment improves conservation efforts for the critically endangered oceanic island endemic *Hibiscus liliiflorus*. *Journal of Heredity* 114: 259–270.

Mashimba, H. M., and K. Shaw. 2022. *Karomia gigas* (amended version of 2021 assessment). The IUCN Red List of Threatened Species. 2022: e.T35173A213968411. https://doi.org/10.2305/IUCN.UK.2022-1.RLTS.T35173A213968411.en

Namoff, S., C. E. Husby, J. Francisco-Ortega, L. R. Noblick, C. E. Lewis, and M. P. Griffith. 2010. How well does a botanical garden collection of a rare palm capture the genetic variation in a wild population? *Biological Conservation* 143: 1110–1117.

Nic Lughadha, E., S. P. Bachman, T. C. C. Leão, F. Forest, J. M. Halley, J. Moat, C. Acedo, et al. 2020. Extinction risk and threats to plants and fungi. *Plants, People, Planet* 2: 389–408.

Paris, J. R., J. R. Stevens, and J. M. Catchen. 2017. Lost in parameter space: A road map for stacks. *Methods in Ecology and Evolution* 8: 1360–1373.

Petit, R. J., A. El Mousadik, and O. Pons. 1998. Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* 12: 844–855.

Petit, R. J., and A. Hampe. 2006. Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution, and Systematics* 37: 187–214.

Pritchard, J. K., M. Stephens, and P. Donelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

Pritchard, D. J., J. E. Fa, S. Oldfield, and S. R. Harrop. 2012. Bring the captive closer to the wild: Redefining the role of ex situ conservation. *Oryx* 46: 18–23.

Schoen, D. J., and A. H. D. Brown. 1993. Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. *Proceedings of the National Academy of Sciences, USA* 90: 10623–10627.

Sgrò, C. M., A. J. Lowe, and A. A. Hoffmann. 2011. Building evolutionary resilience for conserving biodiversity under climate change. *Evolutionary Applications* 4: 326–337.

Štorchová, H., R. Hrdličková, J. Chrtek, M. Tetera, D. Fitze, and J. Fehrer. 2000. An improved method of DNA isolation from plants collected in the field and conserved in saturated NaCl/CTAB solution. *Taxon* 49: 79–84.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.

Verdcourt, B. 1992. Verbenaceae. *In* R. M. Polhill [ed.], Flora of Tropical East Africa. A.A. Balkema, Rotterdam, Netherlands.

Wang, S., E. Meyer, J. K. McKay, and M. V Matz. 2012. 2b-RAD: A simple and flexible method for genome-wide genotyping. *Nature Methods* 9: 808–810.

Wei, X., and M. Jiang. 2020. Meta-analysis of genetic representativeness of plant populations under ex situ conservation in contrast to wild source populations. *Conservation Biology* 35: 12–23.

Weir, B. S., and J. Goudet. 2017. A unified characterization of population structure and relatedness. *Genetics* 206: 2085–2103.

Willoughby, J. R., M. Sundaram, B. K. Wijayawardena, S. J. A. Kimble, Y. Ji, N. B. Fernandez, J. D. Antonides, et al. 2015. The reduction of genetic diversity in threatened vertebrates and new recommendations regarding IUCN conservation rankings. *Biological Conservation* 191: 495–503.

Wright, S. 1922. Coefficients of inbreeding and relationship. *The American Naturalist* 56: 330–338.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. Protocols used for seed germination of *Karomia gigas*.

Appendix S2. List of each individual included in the study, including its origin, sample ID, whether missing data exceeded the 40% threshold, whether the individual served as a parent in the ex situ collection, and its internal relatedness.

Appendix S3. Comparison of internal relatedness (IR) values between selfed and outcrossed collections and wild populations as inferred by parentage analyses. Boxplots with different letters denote statistically significant differences from one another based on Tukey's HSD comparison of means ($\alpha = 0.05$).

Appendix S4. Analysis of *K* estimation, showing the delta *K* (blue) and the probability of the data (red).

Appendix S5. Pairwise kinship matrix of all individuals in the study as measured by beta, calculated using the "beta.dosage" function in the R package HIERFSTAT v.0.5-11.

How to cite this article: Linan, A. G., R. E. Gereau, R. Sucher, F. H. Mashimba, B. Bassuner, A. Wyatt, and C. E. Edwards. 2024. Capturing and managing genetic diversity in ex situ collections of threatened tropical trees: A case study in *Karomia gigas*. *Applications in Plant Sciences* 12(3): e11589.
https://doi.org/10.1002/aps3.11589