Assembling genetic structure of Gardenia remyi, a critically endangered tree endemic to the Hawaiian Islands

Hansen, Kristina Egholm; Opgenorth, Mike; Flynn, Tim; Kennedy, Barbara; Rønsted, Nina; Olofsson, Jill K.; Barnes, Christopher J.

Published in:
Conservation Science and Practice

DOI:
10.1111/csp2.13011

Publication date:
2023

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Assembling genetic structure of *Gardenia remyi*, a critically endangered tree endemic to the Hawaiian Islands

Kristina Egholm Hansen | Mike Opgenorth | Tim Flynn
Barbara Kennedy | Nina Rønsted | Jill K. Olofsson | Christopher J. Barnes

1Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark
2Section for GeoGenetics, Globe Institute University of Copenhagen, Copenhagen, Denmark
3National Tropical Botanical Garden, Hāna, Hawaii, USA
4National Tropical Botanical Garden, Kalaheo, Hawaii, USA
5Bishop Museum, Honolulu, USA

**Correspondence**
Mike Opgenorth, National Tropical Botanical Garden, 650 Ulaino Road, Hāna, Maui, HI 96713, USA.
Email: mopgenorth@ntbg.org

**Present address**
Jill K. Olofsson, Section for Forest, Nature and Biomass, Department of Geoscience and Natural Resource Management, University of Copenhagen, Frederiksberg C, Denmark.

**Funding information**
Foundation Franklinia, Grant/Award Number: 2020-4; National Tropical Botanical Garden

**Abstract**
Conservation and restoration planning of extremely rare species relies on an understanding of the genetic diversity and population dynamics within a species to overcome potential inbreeding depression. Nānū or Nāʻū (*Gardenia remyi* H. Mann.) is an endemic tree native to the Hawaiian Islands and is one of more than 200 endangered plant species in Hawaiʻi with less than 50 individuals remaining in the wild. Efforts to understand the genetic diversity and connectivity between wild populations are foundational to conservation management plans, however little is known of the population structure of the species. In this study we utilize double digest restriction-site associated sequencing (ddRADSeq) on both historical herbarium specimens and samples from living ex situ collections to: (1) Test the hypothesis that we can capture genetic diversity in herbarium material of *G. remyi* using ddRADSeq, and (2) test the hypothesis that there are genetically distinct populations or subpopulation units among different Hawaiian islands. Usable sequencing data from thirty-seven samples of herbarium specimens collected between 1952 and 2017 and twenty wild sourced living collection samples were obtained representing all four islands where *G. remyi* is known to occur. Phylogenetic and population structure analysis revealed a monophyletic ingroup and a clear division between *G. remyi* samples of the northern island of Kauaʻi and those from the more southeastern younger islands of Molokai, Maui and Hawaiʻi islands. The Kauaʻi samples were further split into a subpopulation from Southern Kauaʻi and the subpopulations from Northern Kauaʻi. Some admixed samples were detected. Our results are consistent with subpopulations of *G. remyi*, which needs to be considered in future conservation planning and breeding efforts to minimize inbreeding depression.

**KEYWORDS**
conservation genomics, genetic diversity, herbaria, museomics
INTRODUCTION

The long-term survival of a species is dependent on genetic diversity for the species to be able to adapt to the rate and degree of environmental changes. In both common and rare species, erosion of intraspecific genetic diversity is a pressing issue as a result of climatic changes and especially in small and isolated populations which are more prone to lower genetic diversity (Diaz et al., 2020; Hoban, Bruford, et al., 2020; Leigh et al., 2019). Furthermore, rare species tend to suffer from loss of habitat, decreased population sizes, and reduced potential for genetic exchange. In plants, these conditions can promote self-pollination and genetic drift, leading to inbreeding and loss of founder lines (Foster et al., 2022). This can decrease the adaptive potential of a species in its natural habitat which leaves it at a higher risk of extinction (Willi et al., 2006). Together, these issues have led to increased recognition of the importance of a thorough understanding of the population structure of rare species to inform management decisions (Frankham, 2015; Hoban, Bruford, et al., 2020). These advances in sequencing technology have made genetic management a valuable tool in the conservation planning of both animal and plant species (Robinson et al., 2018; Van Rossum et al., 2020; Willi et al., 2022). Furthermore, applying population genetics to improve demographic viability is particularly pertinent for establishing robust conservation plans for small, isolated populations that might be especially impacted by immediate threats and adverse environmental changes (Frankham et al., 2017; Hoban et al., 2021).

A strategy that has proven especially useful for the conservation of very rare plant species is to utilize ex situ collections where specimens are grown in conservation collections and stored in seed banks (Backs et al., 2021; Foster et al., 2022; Wolkis et al., 2021). However, a major challenge for ex situ collections is a lack of information about wild source populations as well as capturing the genetic variation of their origin (Backs et al., 2021; Maunder et al., 2001). Previous studies provide examples of critically endangered Hawaiian endemic plants, such as Brighamia insignis A. Gray, with the main ex situ conservation collection missing genetic diversity found in other ex situ collections elsewhere around the world. For collections maintained for generations, this increases the chance for inbreeding depression within certain collections (Foster et al., 2022; Walsh, 2015; Walsh et al., 2019). In such cases, herbarium specimens are a valuable source of genetic data (Bieker & Martin, 2018; Jafari et al., 2018) which can provide a window into the historical levels of genetic diversity and the changes that have occurred due to human-induced habitat changes (Cozzolino et al., 2007). This information can then be compared to current wild and ex situ populations to inform conservation management (Rocchetti et al., 2021; Wolkis et al., 2021). There are challenges to acquiring robust genetic data from historic samples, namely due to high levels of degraded DNA (Bieker & Martin, 2018; Graham et al., 2015). However, advances in DNA extraction methods and techniques for DNA sequencing have made it possible to cost-effectively obtain viable genome-wide genetic data from non-model species with minimal and degraded DNA (Jordon-Thaden et al., 2020; Lang et al., 2020).

The isolated oceanic islands of the Hawaiian archipelago are characterized by one of the highest levels of endemism in the world. Approximately 90% of its 1,367 native plant taxa are endemic, many occurring on a single-island or valley (Rønsted, Walsh, et al., 2022; Sakai et al., 2002). The Hawaiian flora is currently subjected to high risk of widespread and rapid extinction (Humphreys et al., 2019; Kier et al., 2009), which is in part due to small population sizes of many of the island’s plant species as well as their prolonged evolutionary isolation. This resulted in unique characteristics such as narrow habitat range and highly specialized niches (Caujapé-Castells et al., 2010; Frankham, 1998). There is evidence that at least 9% of the endemic Hawaiian plant taxa are now extinct in the wild (Sakai et al., 2002; Wood et al., 2019) and 31% of the native Hawaiian vascular flora is listed as threatened or endangered by the US Fish and Wildlife Service (USFWS) under the US Endangered Species Act (USFWS, 2022). Ongoing threats such as competition with invasive species, spread of infectious diseases, and habitat loss and degradation have resulted in loss of both populations and entire species of native Hawaiian flora (Rønsted, Walsh, et al., 2022). The challenges in preventing rapid extinction rates for Hawaiian plant species and ensuring robust recovery of vulnerable plant populations are further complicated by the pressing issue of climate change which calls for urgent measures in conservation and restoration (Rønsted, Campbell, et al., 2022).

Gardenia remyi H. Mann, a flowering tree native to the Hawaiian Islands, is a focus of restoration efforts in Hawai‘i due to its status as endangered (Grave et al., 2020; USFWS, 2016; USFWS, 2021a; USFWS, 2021b) (Figure 1). It is recognized as one of three endemic Hawaiian taxa in the genus Gardenia J. Ellis that share the vernacular Hawaiian name “nānu‘u” or ‘nā‘u‘u’. The species previously occurred on the islands of Kaua‘i, Moloka‘i, Maui, and Hawai‘i island (Wagner et al., 1999). Today, there is some uncertainty in the number of extant plants and subpopulation units. The most recent assessment for the IUCN Red List of Threatened Species lists this plant as “Critically Endangered” with a decline to approximately 82 trees remaining in 18 extant subpopulations (Grave et al., 2020; IUCN, 2022). However, a recent species report from the US Fish and
Wildlife Service estimates that only 30–34 individuals may remain in 11 extant wild populations and that as many as 52 different subpopulation units may have existed in the past (USFWS, 2021a). The most recent estimate through personal correspondence with the Department of Land and Natural Resources (DLNR) Hawai‘i State Botanist estimated that approximately 40 plants remain in the wild across 14 subpopulation units and that there are few observations of fruits or seedlings (Matthew Kier, pers. communication, 2022).

Considering the endangered status and continued rate of decline of G. remyi populations in the wild, the identification and maximization of genetic diversity within ex situ collections is a matter of urgency. The loss of genetic diversity within living plant collections can potentially be mitigated by developing a robust breeding program that...
incorporates genetic data from founding individuals into breeding decisions and ex situ management (Fant et al., 2016, 2019; Wood et al., 2020). Furthermore, utilizing methods focused on equalizing founder contribution and minimizing inbreeding can be used when propagating new plants to replenish aging accessions (Foster et al., 2022). However, a fundamental understanding of the historic range and genetic diversity is necessary to effectively connect monitoring of intraspecific genetic variation to the curation of seed banks and botanical gardens, and potential reintroduction to the wild from these ex situ collections.

In this study, we utilize double digest restriction-site associated sequencing (ddRADSeq), a simple and inexpensive method for creating reduced representation libraries for single-nucleotide-polymorphism (SNP) discovery and genotyping, from both historical herbarium specimens and samples from living ex situ collections of wild provenance to test the hypotheses that: (a) we can capture genetic diversity in herbarium material of *G. remyi* using ddRADSeq; and (b) that there are genetically distinct subpopulations or subpopulations among different Hawaiian islands.

## 2 | MATERIALS AND METHODS

### 2.1 | Sampling strategy

Sampling was aimed at including all extant and possibly extinct wild subpopulations as well as representing the diversity within subpopulations (Rocchetti et al., 2021; Wolks et al., 2021). The term “population” is difficult to define for Hawaiian plants, especially when current distributions reflect extirpation, fragmentation, and habitat destruction (Laukahi, 2020). Contrary to the common use of the biological concept of “population,” The Standards and Petitions Committee of the IUCN Species Survival Commission defines “population” as the total number of individuals of a species, and “subpopulations” as geographically or otherwise distinct groups in the population between which there is little demographic or genetic exchange (IUCN, 2022). Recognizing the systematic challenge to define distinct units, a working definition of a subpopulation was adopted by the Hawaii Rare Plant Restoration Group (HRPRG) as: “A group of conspecific individuals that are in close spatial proximity to each other (i.e., less than 1,000 m apart, or less if other factors such as barriers to dispersal or gene flow are also present), and are presumed to be genetically similar and naturally capable of sexual (recombinant) reproduction” (Laukahi, 2020). In this study, we use the same HRPRG population classification along with the Geographic Reference Area (GRA) map created by the HRPRG drawn along ahupua’a (traditional Hawaiian watershed) boundaries and natural topography. Other codes were added for large single-landowner properties and botanical gardens (Laukahi, 2020). Hawaiian place names were consolidated following Soehren (2010).

We included both historical herbarium specimens and samples from living conservation collections in the sampling scheme. A total of 20 accessions of known wild provenance were sampled as young leaves. Of these, 19 were from living collections of *G. remyi* in the conservation nursery of the National Tropical Botanical Garden (NTBG) on Kaua’i and one sample was obtained from the Olinda Rare Plant Facility on Maui. In addition, 51 herbarium samples of *G. remyi*, collected between 1909 and 2017, were sourced from the two main herbarium collections in Hawai’i. Of these, 12 specimens were from the NTBG herbarium (PTBG) and 39 were provided by Bernice P. Bishop Museum herbarium (BISH). As DNA is generally expected to be degrading over time, we evaluated age effects on data quality in downstream analysis.

A total of 32 of the herbarium samples did not provide DNA of sufficient quality for ddRAD-sequencing and were excluded from downstream processes. In total, the sequencing pool included 39 *G. remyi* samples of both herbarium and living collection origin, with the oldest sample collected in 1952 (Table S1). These samples represent 14 subpopulation units across the four islands Kaua’i, Moloka’i, Maui, and Hawai’i island. (Table 1). As an out-group, we also included four samples of *G. brighamii*, one of the two other native Hawaiian *Gardenia* species (three from the NTBG living collection and one PTBG herbarium specimen).

### 2.2 | DNA extraction

All lab work for this project was conducted at the molecular biology labs for contemporary DNA at GLOBE Institute, University of Copenhagen. To minimize risk of contamination, lab work was done in Airstream® Class II Biological Safety Cabinets (ESCO Lifesciences, Singapore) whenever possible in dedicated clean-labs or post-PCR labs as suitable.

Genomic DNA was extracted from up to 20 mg (Table S1) of dried leaf material using the QIAGEN DNeasy Plant Mini Kit (QIAGEN, Germany), following manufacturer’s protocol with minor adjustments to ensure a high yield from herbarium samples with low DNA content (see Appendix A). The DNA concentration was measured with Invitrogen’s Qubit® 3.0 Fluorometer using the dsDNA HS assay kit (ThermoFisher Scientific, USA).
2.3 ddRADseq

Samples were processed with the ddRADseq protocol (Peterson et al., 2012) with small modifications as outlined in Olofsson et al. (2019). Briefly, 3 μL digestion master mix containing EcoRI and MseI restriction enzymes was added to 7 μL (25 ng–280 ng) of DNA extract and incubated for eight hours at 37°C with a heated lid in a GeneAmp® thermocycler. A double stranded common adaptor and double stranded barcoded EcoRI adapters from Parchman et al. (2012) were ligated to the digested DNA by incubation for six hours at 16°C. Each of the EcoRI adapters contain unique barcodes of 8–10 bp and differ by at least four bases to account for sequencing errors (Davey et al., 2011; Parchman et al., 2012).

The digest-ligation product was diluted up to 50 μL with 0.1x TE and standard Illumina sequencing primers were added to the diluted digest-ligation product through PCR. Primers as well as details on amplification conditions optimized for samples with low DNA concentration can be found in Appendix A. This resulted in successful amplification of 43 libraries which were verified with gel electrophoresis and 5300 Fragment Analyzer System (Agilent Technologies, USA). Finally, the library pool was paired-end sequenced (150 bp) at Novogene (UK) on the NovaSeq (Illumina Inc., UK) generating 90 G sequencing data, corresponding to roughly 2 G of data per sample.

2.4 De novo assembly of reads and SNP calling

Removal of adaptor and primer sequences as well as quality filtering of raw reads was done using fastp v.0.20.1 for paired end sequencing data (Chen et al., 2018). The settings used for quality filtering can be found Appendix A. The cleaned reads were further processed and demultiplexed using the process_radtags script from Stacks v2.2 (Catchen et al., 2013). All cleaned reads were truncated at 75 bp to allow for analyses with the Stacks pipeline, and the option to remove all reads with uncalled bases was enabled. De novo assembly of putative loci and subsequent SNP calling was done using the denovo_map.pl script which executes the Stacks pipeline (ustacks, cstacks, sstacks, tsv2bam, and gstacks). The Stacks populations program was used to create summary statistics and export the Stacks data to different formats. Only loci present in a minimum of 50% of the individuals of a population and in

<table>
<thead>
<tr>
<th>Island</th>
<th>Geographic reference subpopulation unit</th>
<th>Extant or extirpated in the wild</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawai‘i</td>
<td>Hakalau Nui</td>
<td>Extant</td>
<td>1</td>
</tr>
<tr>
<td>Hawai‘i</td>
<td>‘Ola’a to Kea‘au</td>
<td>Extirpated</td>
<td>1</td>
</tr>
<tr>
<td>Kaua‘i</td>
<td>Hule‘ia</td>
<td>Extant</td>
<td>14</td>
</tr>
<tr>
<td>Kaua‘i</td>
<td>Kalalau</td>
<td>Extant</td>
<td>5</td>
</tr>
<tr>
<td>Kaua‘i</td>
<td>Limahuli</td>
<td>Extant</td>
<td>3</td>
</tr>
<tr>
<td>Kaua‘i</td>
<td>Lumahai</td>
<td>Extant</td>
<td>2</td>
</tr>
<tr>
<td>Kaua‘i</td>
<td>Wainiha</td>
<td>Extirpated</td>
<td>1</td>
</tr>
<tr>
<td>Maui</td>
<td>Honolulu</td>
<td>Extant</td>
<td>2</td>
</tr>
<tr>
<td>Maui</td>
<td>Kauku‘ai</td>
<td>Extant</td>
<td>2</td>
</tr>
<tr>
<td>Maui</td>
<td>Waiehu</td>
<td>Extirpated</td>
<td>1</td>
</tr>
<tr>
<td>Moloka‘i</td>
<td>Mana‘olu‘olu - Kalua‘aha</td>
<td>Extant</td>
<td>2</td>
</tr>
<tr>
<td>Moloka‘i</td>
<td>Manuahi ridge</td>
<td>Extirpated</td>
<td>1</td>
</tr>
<tr>
<td>Moloka‘i</td>
<td>Hālawa</td>
<td>Extant</td>
<td>1</td>
</tr>
<tr>
<td>Moloka‘i</td>
<td>Kukuinui ridge</td>
<td>Extant</td>
<td>1</td>
</tr>
</tbody>
</table>
at least 14 populations out of the 15 sequenced, including the outgroup population, were retained for further analyses. The data analyses were further restricted to one random SNP per locus and the minimum minor allele frequency required to process a nucleotide site at a locus was specified to 0.05. Two samples (KH-59 and KH-76) with more than 50% missing data were removed from the final analyses. This resulted in a total of 37 ingroup samples and four outgroup samples from both herbarium and living collections in the final analyses (Table S1).

2.5 | Phylogeny and population genetic structure

RAxML v8.2.11 (Stamatakis, 2014) was used to create a maximum likelihood phylogeny for the nuclear SNP alignment of all 41 samples using a GTR+G substitution model. One hundred rapid bootstraps were performed to evaluate node support (BS: bootstrap support). Subsequent analyses were conducted excluding three admixed samples to increase resolution among subpopulations (Figure 3).

The population structure was evaluated by principal coordinates analysis (PCO) performed with the R package adegenet v2.1.5 (Jombart & Ahmed, 2011) and admixture analysis with the program fastSTRUCTURE v1.0 (Raj et al., 2014). FastSTRUCTURE uses a Baysian model based approach to assigning K putative populations, also referred to as genetic or population clusters. The optimal number of K, 1–10 population clusters, for the admixture analysis was estimated by the fastSTRUCTURE chooseK.py script. The admixture analysis was done twice, one run with the outgroup included and one with only G. remyi samples. Missing data in the PCO was handled by replacing it with the mean of the allele frequency at each SNP among individuals. Preliminary analyses found that KH-12 (herbarium sample from Kalalau, North Kaua‘i) and KH-62, (living collection sample from Hule‘ia, South Kaua‘i), showed distant clustering from all the rest of the G. remyi samples and these samples were therefore excluded from the PCOs. Genetic relatedness parameters including expected heterozygosity, inbreeding coefficient and relatedness were calculated using VCFtools (Table S2).

FIGURE 2 Map showing sampling included in the final analysis. Sample size indicated by black circles. DarkGray areas represent extant subpopulations and light gray areas represent possibly extirpated historical populations. Map produced by Ben Nyberg, National Tropical Botanical Garden.
3 | RESULTS

After quality control, including removal of two samples with more than 50% missing data, the final sampling for analyses included 3375 retained SNPs from four outgroup and 37 ingroup samples representing all four islands and 15 valleys or populations (Figure 2, Tables 1 and S1). This included four herbarium samples representing populations

---

**FIGURE 3** Unrooted circle diagram and phylogenetic tree from RAxML analysis of *Gardenia remyi* samples. (a) Circle diagram including *G. brighamii* outgroup (gray). (b) Phylogenetic tree excluding outgroup. A clear division is shown between samples of Kaua‘i and those from the more southeastern younger islands of Moloka‘i, Maui and Hawai‘i. Colors represent the Hawaiian Islands as also linked to the map. Orange/yellow: Hawai‘i Island, Blue: Moloka‘i, Green: Maui, Red: Northern Kaua‘i subpopulations, Maroon: Southern Kaua‘i Hulē‘ia subpopulation. Gray designates the outgroup *G. brighamii*. 
that are considered extinct in the wild, one from each island, Kaua’i (Wainiha, KH-27), Moloka’i (Pelekunu, KH-3), Maui (Waihe’e, KH-2), and Hawai’i island (Kulani, KH-41). Together, this validated the approach to capture genetic information of modern and historic herbarium samples of *Gardenia remyi* using ddRADSeq.

The initial phylogenetic analysis showed a strongly supported (BS = 100%) *G. remyi* ingroup. However, three

---

**Figure 4** Principal Coordinates Analysis (PCO) of *Gardenia remyi* genetic diversity based on ddRadSeq data. Numbers designate KH-xx sample number and colors designate island/subpopulation (a–d) Subpopulations grouped by island. Populations of Kaua’i in red and maroon, Maui in blue, Moloka’i in green and Hawai’i in yellow. Herbarium samples are designated with triangles and living collection samples with circles. (a) PCO of PC1 and PC2 including the outgroup, (b) PCO of PC1 and PC2 without the outgroup, (c) PCO of PC3 and PC2 without the outgroup. (d) PCO including only samples from Maui, Moloka’i and Hawai’i (e, f) Subpopulations of Kauai grouped by valley. Samples from Hulē’ia in maroon, Kalalau in purple, Limahuli in pink, Lumahai in cyan and Wainiha in orange (e) PCO showing PC1 and PC2(f) PCO showing PC3 and PC1 with collection year shown in parentheses for herbarium samples.
presumed admixed samples (KH-62, KH-12, and KH-06) resulted in long branches with low support, and placement between clades (Figure S1). Sample KH-12 from Kalalau (North Kaua‘i) is positioned as sister to a clade including all samples from Maui, Moloka‘i, and Hawai‘i. Sample KH-06 is sister to all samples except the Hulē‘ia samples and sample KH-62 from Hulē‘ia appeared as sister to the remainder of the ingroup samples. The subsequent analysis (Figure 3) excluding the three admixed samples resulted in three clades representing Kaua‘i, Hawai‘i and Maui Nui. The Kaua‘i clade is further split into two monophyletic groups, one including samples from Hulē‘ia located in South Kaua‘i and one including samples from four populations located in North Kauai, Kalalau, Limahuli, Lumahai and Wainiha.

Principal coordinates analysis with both G. remyi samples and outgroup samples revealed that 85.0% of the variance in the dataset is explained by the first three principal coordinates (Figure 4a). The first principal coordinate (PC1; 63.2%) splits G. remyi from the outgroup G. brighamii (black) while PC2 (18.5%) splits samples from Kaua‘i (red and maroon) and samples from Moloka‘i, Maui and Hawai‘i (green, blue, and orange; Figure 4a). In the PCO excluding outgroup samples, 67.8% of the variance among the samples can be explained by the first three principal coordinates (Figure 4b,c). Here, PC1 (55.1%) splits samples from Kaua‘i and samples from Moloka‘i, Maui, and Hawai‘i, while PC2 (9.7%) splits samples from northern Kaua‘i (red) and samples from southern Kaua‘i (maroon; Figure 4b). Sample KH-06 lies between the two Kaua‘i clusters (Figure 4b). Finally, PC3 (3.0%) clusters samples from Moloka‘i (green), Maui (blue), and Hawai‘i (orange) into three groups corresponding to the islands’ geographical location (Figure 4c). Sample KH-54 from the Halawa valley on Moloka‘i lies more towards the Maui cluster than the Moloka‘i (Figure 4d).

In the PCO with samples solely from Kaua‘i (Figure 4e,f), 46.4% of the variance among samples can be explained by the first three principal coordinates. Here, PC1 (32.3%) again splits the samples into a North and a South group (Figure 4e). However, PC2 (8.6%) now shows clustering corresponding to valleys of North Kaua‘i (Figure 4e). Finally, PC3 (5.5%) divides samples of North Kaua‘i collected before and after 2002 with sample KH-17 (collected in 2021) being an outlier (Figure 4f).

The fastSTRUCTURE analyses with both in group and outgroup samples suggests that the optimal number of population clusters is three, which divides the samples into one cluster representing the outgroup G. brighamii (black) and two groups of G. remyi corresponding to the northern island of Kaua‘i (red) and one cluster for the southeastern islands of Maui, Moloka‘i and Hawai‘i (purple; Figure 5). Sample KH-62 (living collection, Hulē‘ia) appears to be admixed between the Hulē‘ia G. remyi population and the outgroup species G. brighamii, while KH-12 (herbarium, Kalalau) appears to be admixed between the Kaua‘i and Maui-Moloka‘i-Hawai‘i populations (Figure 5).

The distant clustering in the preliminary PCA analyses of these samples is likely explained by their admixed status. To test that the admixed profile of samples KH-12 and KH-62 were not due to sample contamination, we mapped cleaned reads to the chloroplast genome of G. jasminoides J. Ellis to test for a higher proportion of heterozygous variants on the chloroplast. The above-mentioned samples do not have a higher proportion of heterozygous sites than any of the other samples included in the study (results not shown), which along with a high proportion of admixture makes contamination unlikely. The admixture analysis was repeated excluding the outgroup samples, which resulted in a similar pattern as described above. (Figure S2). While the choose.py fastSTRUCTURE program suggest that the optimal K populations would be two, increasing the number of population clusters to four also shows a biologically relevant pattern where the Kauaʻi population is split into two distinct clusters corresponding to northern Kauaʻi (orange) and southern Kauaʻi (brown). This supports the findings in the PCO and phylogenetic analyses (Figure S2).

Analysis of heterozygosity and inbreeding found most G. remyi and G. brighamii samples had a much higher expected heterozygosity than was observed and for all but three samples the inbreeding coefficient was higher than 0.4 (Table S2). One of these was an outgroup sample (KH-13), while the other two were ingroup samples KH-12 and KH.62. KH-62 had a negative inbreeding coefficient which corroborates the admixed profile of this sample in the fastSTRUCTURE analysis (Figure 5).

4 | DISCUSSION

4.1 | Applicability of ddRADSeq for herbarium samples

A primary objective in this study was to test that the ddRADseq pipeline could be applied to herbarium material to capture genetic data comparable to that from samples from living collections. We find that this approach is valuable in cases of extremely rare species, such as G. remyi, with few remaining extant individuals. Additionally, the approach might provide valuable information on the historical diversity of the species where certain populations are now extinct.

Libraries were initially obtained from 13 of 54 herbarium samples, but through adjusting the amount of PCR amplification cycles, 21 samples could be included in the sequencing pool. While an increase in sample size is preferable, the additional PCR steps can introduce biases, such as early PCR cycle mutations (Andrews et al., 2016; Davey et al., 2011). Herbarium samples generally had a lower number of reads and a higher percentage of missing data, after quality filtering and RADTag processing, compared to samples from living collections (Figure S3). A number of different factors can give rise to missing RADseq data, however, working with degraded samples such as herbarium specimens, the high amount of missing data is likely caused by fragmentation bias of the DNA template in the library preparations and thus uneven PCR enrichment and sequencing coverage (Graham et al., 2015). Indeed, we observed a significant difference between sequencing coverage for herbarium and living collection samples \( p < 0.05 \); Figure S3). While the sequencing coverage was lower for herbarium samples than samples from living collections, we ultimately found that using standard genetic clustering analyses of the ddRAD data from both sample types showed clearly discernible population structure (Figure 5).

We were unable to obtain ddRADseq libraries from herbarium samples collected prior to 1952, many of which were samples from Molokaʻi, Maui, and Hawaiʻi. To allow for inclusion of older samples, future studies may explore the use of other Reduced-Representation Library approaches, such as sequence capture (Baker et al., 2021; Brewer et al., 2019). Alternative methods could be the Angiosperm353 probe kit (Johnson et al., 2019) or Genome Skimming approach (Bakker, 2017; Nevill et al., 2020), however, these would generate new datasets that are not directly combinable with the ddRADseq data and would thus need to include re-sequencing and/or resampling of all specimens. More broadly, our study demonstrates a need for further development of methods for historical or other highly degraded samples.

4.2 | Population structure mirrors island biogeography and limited dispersal ability

Both phylogenetic (Figure 3) and principal coordinate analyses (Figure 4) of SNP data included in this study showed a division between samples of North Kauaʻi, South Kauaʻi, and those from the more southeastern islands of Molokaʻi, Maui, and Hawaiʻi. PCO results further indicated some clustering of the samples from the southeastern (and younger) islands into three groups corresponding to the three islands, Molokaʻi, Maui, and Hawaiʻi (Figure 4). The spread of G. remyi is presumed to be characterized by rare seed dispersal events between islands (founder events) as is observed for many angiosperms on the Hawaiian Islands (Price & Wagner, 2004). Genetic exchange between subpopulations is also considered rare with most remaining subpopulation units being represented by fewer than five trees (Grave et al., 2020; USFWS, 2021a).
Little is known about the reproductive biology of *G. remyi*, but preliminary observations suggest this species has a dioecious or sub-dioecious breeding system (USFWS, 2021a; Wood, 2012). Observations of plants in the wild and self-pollinated ex-situ suggest *G. remyi* are self-incompatible (Wood, De Motta, NTBG, personal communication) and pollination mechanisms have not been observed. However, morphological characteristics such as tubular, fragrant white flowers suggest pollination by insects such as moths and a potential dispersion of seeds by birds that would be attracted to the fleshy fruits (Sakai et al., 1995; Wood, 2012; Figure 1). The combination of rare seed dispersal and the potential of reduced pollination between a dwindling number of flowering plants results in limited gene flow between the individual islands and valleys. Together, these are considered driving factors of the observed genetic population structure in *G. remyi*. The observed population structure might also be influenced by other processes such as genetic drift, inbreeding and selection favoring adaptive traits in response to local ecological factors.

In this study we found little genetic differentiation between samples from Moloka’i and Maui (Figures 3 and 5). This could be due to the relatively few samples from these populations and restrictive filtering of the RADseq data, but there could also exist real biological explanations for this pattern. The phylogeography showed the Moloka’i samples as being part of a clade with the Maui samples (Figure 3). Maui and Moloka’i were once part of a larger island referred to as Maui Nui which did not separate into two units until around 400,000 years ago (Sakai et al., 2002). It is thus possible that *G. remyi* colonized Maui Nui before the split and that the Moloka’i and Maui subpopulation units stem from the same founder population. Therefore, it is expected there would have been greater gene flow between these populations before the islands separated. We recommend further investigation using fossil records to date the split of Maui Nui into separate islands (Cole & Morden, 2021).

The admixture analyses (Figure 5 and Figure S2) are generally consistent with the population clustering observed in the PCO analysis (Figure 4) however, the resolution was not high enough to separate these populations to the same degree. Sample KH-12 (Kalalau) appears to be admixed with the Moloka’i-Maui-Hawai‘i population (Figure 5). Furthermore, sample KH-62 (Hulē’ia) shows admixture with the outgroup *G. Brighamii* (Figure 5). The hypothesis that these two samples appear admixed due to contamination was tested by mapping cleaned reads to the chloroplast genome of *G. jasminoides*. Heterozygous variants called from reads mapping to the chloroplast can be caused by plastid-nuclear gene transfers (Martin, 2003), however, samples with a high proportion of heterozygous chloroplast variants outside the expected range could indicate sample contamination. The above-mentioned samples do not have a higher proportion of heterozygous sites than any of the other samples included in the study. Along with a high proportion of admixture (Figure 5) this indicates that contamination during the lab work of the project is unlikely. Excluding contamination as the source of the admixture, the observed admixture for sample KH-12 can be an intraspecific hybridization. However, in terms of geography the observed admixture seems less likely than admixture between populations on the same islands, or islands that are located close to each other. For KH-62, a possible explanation could be interspecific hybridization between *G. remyi* and *G. Brighamii*. Such hybridization is possible in natural plant populations but is potentially more likely if the species are grown in close proximity in ex situ living collection (Volis, 2017; Zhang et al., 2010). Hybrid gardenia plants have also become widely established in ornamental landscapes over the last two decades, with a well-known hybrid of *G. brighamii × G. taitensis* commonly being advertised as a robust form of *G. brighamii* (James, 2008), which in some locations is advertised outright as *G. brighamii*. This *G. brighamii × G. taitensis* hybrid is more common to encounter in a landscape setting than a true Hawaiian gardenia, and is generally less challenging to cultivate. Care should be taken to maintain provenance of plants reared for conservation efforts and to prevent hybrids from inadvertently cross pollinating with remaining wild plants.

### 4.3 Opportunistic sampling biases

The sampling initially covered 26 historically observed subpopulations of *G. remyi* from all four islands. The over-representation of samples from Kaua‘i reflects extensive monitoring and collection efforts by NTBG based on Kaua‘i over the last decades. For this study, we focused on the two main herbaria holding collections of Hawaiian plants, PTBG and BISH (Table S1). Few collections of *G. remyi* exist in other herbaria, primarily in the USA (Smithsonian Institution), and are mostly older duplicates, which were not acquired for this study.

The non-random and irregular acquisition of samples is a general limitation in the use of specimens from historical collections (Bieker & Martin, 2018; Renstedt, Grace & Carine, 2020). This overrepresentation is also reflected in the number of samples from Kaua‘i populations particularly the Hulē’ia subpopulation of South Kaua‘i (Table S1). Had these not been included, however, there would have been only one herbarium sample from the Hulē’ia subpopulation (KH-06), minimizing the
range of *G. remyi* on Kaua`i. It was also of interest to include living collection samples to explore to what extent they represent the historical genetic diversity of distinct *G. remyi* subpopulation units.

Another potential sampling bias is reflected in the span of years in which the included herbarium specimens were collected (Table S1). Here only specimens from Kaua`i, Moloka`i, and Maui were collected after 2000, whereas the most recent sample from Hawai`i island in this study was from 1977 (KH-40). The oldest successfully sequenced sample was KH-41, collected in 1952 (Table S1). However, this sample appears to be an outlier, as all the other herbarium specimens that underwent successful library preparation were sampled between 1977 to present day.

Finally, as is common for studies on threatened species, the access to plant material for sequencing was limited for this study. While the lowest amount of input leaf material which yielded a verifiable ddRADseq library was 1 mg (KH-25) the majority of samples with such a small amount of leaf material did not produce sequencing libraries. Thus, both low amounts of starting DNA and age (>50 years) are limiting factors to process herbarium materials with ddRADseq protocols, the latter likely associated with DNA degradation (Bieker & Martin, 2018). However, handling of the samples, from collection to storage condition in the herbarium, can also influence the quality of DNA (Dean et al., 2018). Drying conditions have been found to be critical as slow drying typically hastens DNA degradation (Dean et al., 2018). However, despite all of these challenges with including older herbarium samples, we believe it is worth considering this opportunity to increase sample size and represent lost geographical and genetic diversity, when attempting to understand historical and current genetic variation for extremely rare and critically endangered species, such as *G. remyi*. While many studies have discussed and increased methodology for working with historical samples often in the context of taxonomic representation in phylogenetic studies, this approach has not been well represented in conservation population genomics yet.

### 4.4 Implications for conservation planning

While the exact number of extant trees is unclear, the USFWS (2021b) recommends that the first steps in recovering *Gardenia remyi* are through management of threats in situ (e.g. removal of invasive plants, fencing to prevent ungulate damage, and traps for rodent removal). In addition, a minimum of three in situ (sub)populations should be established anywhere throughout the islands of Hawai`i, Maui, Moloka`i, and Kaua`i where they now occur or occurred historically and each of these populations must be naturally reproducing with a minimum of 50 mature, reproducing individuals per (sub)population. In addition, 50 individuals (or the total number of individuals if fewer than 50 exist) from each of three (sub)populations should be represented in an ex situ collection (USFWS, 2021b). This will move the conservation status from the “Preventing Extinction” phase to the “Interim stage.”

This study identified between two to four discernable phylogenetic clades or main clusters. At the lowest resolution the two populations are between Kaua`i and the southeastern Islands of Moloka`i, Maui, and Hawai`i Island. At higher resolution, analyses illustrate two subpopulations on Kaua`i, one from Hawai`i island, and one representing a mixture of samples from Maui and Moloka`i. To maximize availability of remaining genetic diversity we recommend cultivating plants sourced from these four clusters as high priority for both in situ and ex situ conservation.

As we suspect inbreeding (Table S2) may be ongoing due to the small remaining population size, we further recommend extending ex situ conservation to include all four islands and both southern and northern Kaua`i to maximize the remaining genetic variation of extant *G. remyi* ex situ. Once plants from each cluster are represented in ex situ collections, identifying subject plants for potential cross pollination should be conducted to reduce the risk of inbreeding depression and outbreeding depression while contributing to increased genetic diversity that can be candidates for introductions in situ. This can be done by implementing a pedigree-based approach to managing reproduction of plants ex situ, which will slow the inevitable loss of genetic diversity and in turn, result in healthier collections (Foster et al., 2022).

Strategies for cross-pollinating between subpopulation units led by pedigree-based approaches for rare taxa should be established in order to determine the circumstances in which mixing subpopulation or island-specific genetics are necessary to reduce inbreeding depression and encourage plant fitness (Foster et al., 2022). Given that phenology, pollination, and dispersal patterns vary significantly across taxa, each species requires individualistic planning that allows conservation practitioners to determine when cross-pollination between subpopulation units should occur. In the case of *Gardenia remyi*, we recommend a two-tiered approach that first crosses between subpopulation units of identified clusters in this study, which most closely aligns with the species natural ability to disperse. Second, these efforts could be followed by conducting inter-island cross-pollination attempts across...
phylogenetic clades identified in this study to determine if greater plant fitness can be achieved. With this information, conservation managers may determine if crosses are adequate for outplanting in habitat.

In addition, it is important to ensure ex situ facilities such as botanical gardens and seed banks maintain genetic diversity, and to exchange knowledge among collection holding institutions to establish optimal storage, propagation, and cultivation methods in preparation for future major outplanting campaigns. Finally, pollination studies for Hawaiian gardenia will increase understanding related to the ability for natural genetic exchange potential that remains within and between subpopulation units of Gardenia remyi. Investigating pollination should be accompanied by research on dispersal mechanisms for gardenia fruits and/or seeds. As a result, conservationists will better understand the range for which new plants may become naturally established, therefore guiding conservation management activities and decision-making. Should pollination and dispersal studies find fruit, pollen, or seeds can still move between islands, this will contribute to the understanding of subpopulation unit dynamics.

**DATA AVAILABILITY STATEMENT**

**Data accessibility:** Details of materials used and voucher information is listed in Table S1. Raw ddRADSeq reads will be deposited in the NIH sequence read archive (SRA) bionumber PRJNA1010505 during the review process.

**Benefits generated:** This research addresses a priority concern by providing the first population genetic data for a critically endangered species. The research was initiated and conducted in close collaboration with a local conservation organization. The subpopulation scheme used follows the local conservation assessment in consultation with the responsible local authorities enabling direct transfer of results to conservation planning.

**ORCID**

Mike Oppenorth | https://orcid.org/0000-0002-0741-2038
Nina Rønsted | https://orcid.org/0000-0002-2002-5809
Jill K. Olofsson | https://orcid.org/0000-0002-9527-6573
Christopher J. Barnes | https://orcid.org/0000-0001-6800-4233

**REFERENCES**


Academy of Sciences of the United States of America, 100(15), 8612–8614. https://doi.org/10.1073/pnas.1633606100


Willi, Y., Buskirk, J. V., & Hoffmann, A. A. (2006). Limits to the adaptive potential of small populations. *Annual Review of...
APPENDIX A: EXPANDED METHODS

A.1 DNA extraction

The adjustments were made to the QIAGEN DNAeasy Plant Mini Kit (QIAGEN, Germany) extraction protocol. 50 μL AE elution buffer was used rather than 100 μL to ensure a higher DNA concentration of the elution. Furthermore, the samples were incubated at 37°C rather than RT for 5 min prior-to elusion as prior work in the lab had shown this could give a high yield. The elution step was repeated placing the spin column in a new microcentrifuge tube to prevent lowering the DNA concentration of the first elution.

A.2 ddRADseq

A PCR master mix for the ddRADseq pipeline was prepared with forward primer: 5’AATGATACGG CGACCACCGATCTACACACTTTCCACACGAG CGTCTTCCGATCT’, and reverse primer (index in bold): 5’CAAGCAGAAGACGGCATACGAGATCGTGATGTCAGCACGCT CAGATGTGACGGGACACGTG'. A total of 16.15 μL master mix was added to 4 μL of the diluted restriction-fragment products and the following amplification conditions (GeneAmp® PCR System 9700 thermocycler) were used: (1) initial denaturation at 98°C for 30 s; (2) 16 cycles at 98°C for 20s, 60°C for 30 s and 72°C for 40 s; and (3) final extension at 72°C for 10 min. Initial trials of library preparation with 16 amplification cycles showed that some samples had no or very faint and short smear on gel electrophoresis, indicating either absent or very short and degraded DNA fragments. To optimize the protocol to these samples, qPCR was conducted, and the number of PCR cycles was increased to 30 for these samples. This resulted in successful amplification of eight additional libraries verified with gel electrophoresis, indicating either absent or very short and degraded DNA fragments. To optimize the protocol to these samples, qPCR was conducted, and the number of PCR cycles was increased to 30 for these samples. This resulted in successful amplification of eight additional libraries verified with gel electrophoresis and stretched with 5300 Fragment Analyzer System (Agilent Technologies, USA) to have fragments around 300 bp that were not overlapping with primer dimer artifacts.

A.3 De novo assembly of reads and SNP calling

Adapter and primer sequences were removed using fastp v.0.20.1 for paired end sequencing data (Chen et al., 2018) by specifying a fasta file of expected primer and adapter sequences to be trimmed was specified. Furthermore, polyG tail trimming was enabled and the cleaned reads were filtered by length (minimum 36 bp) and stretches of bases of low complexity (minimum 30%). Quality filtering was enabled by default, with default phred quality for a base at Q15 and percentages of bases allowed to be unqualified at 40%.

SUPPORTING INFORMATION

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

How to cite this article: Hansen, K. E., Opgenorth, M., Flynn, T., Kennedy, B., Ronsted, N., Olofsson, J. K., & Barnes, C. J. (2023). Assembling genetic structure of Gardenia remyi, a critically endangered tree endemic to the Hawaiian Islands. Conservation Science and Practice, 5(10), e13011. https://doi.org/10.1111/csp2.13011


A.3.2. Parameters for the denovo_map.pl script stacks v2.2

The denovo_map.pl script which executes the stacks pipeline (ustacks, cstacks, sstacks, tsv2bam and gstacks) was used for De novo assembly of putative loci and SNP calling. For assembly, the minimum depth of coverage to create a stack (putative alleles) was set to three. The number of mismatches allowed between stacks within individuals was set to two and the number of mismatches allowed between stacks between individuals was set to one.