TITLE: Assessing rarity: genomic insights for population assessments and conservation of the most poorly known Amazonian trees

Ellen J. Quinlan[1], David A. Neill[2], Gonzalo Rivas-Torres[3], and Miles R. Silman[1,4]

1. Department of Biology, Wake Forest University, Winston-Salem, North Carolina, USA
2. Universidad Estatal Amazónica, Puyo, Pastaza, Ecuador
3. Estación de Biodiversidad Tiputini, Universidad San Francisco de Quito, Quito, Ecuador
4. The Sabin Center for Environment and Sustainability, Wake Forest University, Winston-Salem, North Carolina, USA

Corresponding Author: Ellen J. Quinlan, Department of Biology, 1834 Wake Forest Rd., Winston-Salem, North Carolina, USA; quinej18@wfu.edu

ABSTRACT: Tropical forests comprise a few hyperdominant and many rare tree species, but distinguishing the truly rare from those under-sampled remains a challenge for ecology and conservation. Given the vastness of Amazonia (~6 million km^2, ~3.9x10^{11} individual trees), increasing sampling cannot solve this problem. Still, half of all species are known from three or fewer collections, making predicting their abundances and distributions impossible with census data alone. Here, we integrate census data with next-generation genomics to assess the rarity of one of the most poorly known and highly threatened Amazonian trees, *Magnolia yantzazana*. Genetic analyses indicate that while there is relatively high nucleotide diversity among sequences ($\pi > 0.5$), there is also evidence of a loss of heterozygosity ($H_e > H_o$) and inbreeding ($F_{IS} > 0.5$), consistent with a small, isolated population. Demographic reconstructions show population decline since the late Pleistocene, with a predicted effective population size (N_e) of ~10^3 in recent millennia. Together, the low heterozygosity, potential inbreeding, demographic trajectory, and census data suggest *M. yantzazana* is in fact a truly rare species, highly vulnerable to ongoing environmental change and anthropogenic threats in the region, notably mining, and support updating its conservation status to Critically Endangered (CR). This study offers a framework for using genomic tools to advance our understanding of the rarest Amazonian trees and establishing conservation priorities, despite the limited field collections available for most species.

Keywords: Amazon, biodiversity, conservation, conservation genetics, demographic history, rare species, tropical forests

1. INTRODUCTION

The predominant community pattern described for tropical forests from field collections is one of hyperdominance and extreme rarity (e.g., Gentry 1982, Hubbell 2013). Ter Steege et al. (2013) estimated there are approximately 16,000 tree species (≥10 cm dbh) in the Amazon Forest based on forest plot inventory data, finding 227 species (1.4% of the total) so common they account for half of all individual trees ("hyperdominants"). The rarest 11,000 species (~70% of the total) represent just 0.12% of individuals. Though field collections are the best available data on species abundance and distribution, the conclusions that can be drawn from those data have limitations due to spatial autocorrelation and the limited forest area they represent. Moreover, given the large number of species and the exceptionally low predicted abundances of most, increased sampling cannot overcome these limitations. Sampling efforts across the Andes-Amazon system have more than doubled over the last two decades, yet, less than half of all species are known from more than a handful of collections and at least a quarter remain unknown to science (Feeley and Silman 2011a; ter Steege et al. 2013; ter Steege et al. 2016; Guevara Andino et al. 2019).

Under-sampling and the immense numbers of unidentified collections make it difficult to distinguish species that are truly "rare" from those that maintain low local abundances but may be common at broad geographic scales (i.e., Rabinowitz 1981). For example, Pitman et al. (1999) surveyed 19,252 individuals (≥10 cm dbh) from 21 plots (36 ha) in Manu National Park, Peru, and found 31% of the 829 species (or morphospecies) identified represented singletons (1 individual / 36 ha). When extrapolated to the full department of Madre de Dios where Manu National Park is located (78,415 km^2), the estimated abundance for these rare trees is 200,000 stems, with possibilities ranging anywhere from <1000 to >10^6 when smaller stems are included (Pitman et al. 1999). Increasing field sampling efforts alone cannot solve this problem, e.g., ter Steege et al. (2020) showed a 10-fold increase in forest plots would only represent 0.0035% of the Amazonian forest area, capturing <50% of the species richness.

55 Such uncertainty in our knowledge of Amazonian biodiversity affects how we understand species
56 richness, ecosystem function, biogeography, the evolution of Neotropical forests and the population
57 biology of the species that comprise them. Further, estimates of abundance and distribution are
58 essential components of any conservation effort, especially those identifying extinction risks and species'
59 responses to anthropogenic land conversion and climate change (e.g., Hoban et al. 2020). However, we
60 will never sample enough individuals through forest inventories alone to have an accurate account of
61 population sizes and ranges, much less demographic history. Instead, the genomes of individuals already
62 collected can offer information for refining these estimates, as the genome of even a single individual
63 represents a population-level sample of genes and their unique histories. Thus, the integration of
64 modern genomics with field collections can improve population inferences from only a small number of
65 sampled individuals (Nazareno et al. 2017; Lemopoulos et al. 2019) and help identify those that are truly
66 rare and threatened from those simply under-sampled.

67 Advances in next-generation sequencing (NGS), and particularly the development of restriction-site
68 associated DNA sequencing (RAD-seq), have made genome-wide study of non-model organisms such as
69 Neotropical trees possible and cost-effective (Andrews et al. 2016, Parchman et al. 2018). When
70 combined with common metrics of population genomics, such as nucleotide diversity (π), heterozygosity
71 (i.e., H_e, H_o), and inbreeding depression and differentiation (i.e., F_{IS}, F_{ST}), these data offer important
72 insight into the extent, diversity, and interconnectedness of populations (e.g., Linan et al. 2020; Kebaïli et
73 al. 2021, Gautschi et al. 2024), even with low sample sizes (Nazareno et al. 2017; Lemopoulos et al.
74 2019). For example, low H_o compared to H_e can be a signal of inbreeding depression, indicating an
75 increase in homozygosity and loss of genetic diversity. Similarly, positive F_{IS} values indicate an excess of
76 homozygosity compared to what is expected under Hardy-Weinberg equilibrium, a signal of inbreeding
77 and potentially inbreeding depression if deleterious alleles become more frequent. Measures of genomic
78 diversity can also be used to estimate the effective population size (N_e), a value representing how large a

population would need to be to maintain the observed level of diversity under genetic drift alone, calculated from heterozygosity, generation time, and the neutral mutation rate (Hahn 2018). Though N_e can both underestimate and overestimate census population size (N_C), it can be used to set the bounds of possible census sizes, particularly in orders of magnitude, a goal impossible for most Amazonian tree species given inventory data alone. The demographic history of populations can also be reconstructed by estimating changes in N_e through time, calculating changes in the coalescent rate as N_e is inversely proportional to coalescent time. This is possible due to recombination, which results in different regions of the genome having different gene trees, each which contains information on population growth, contraction, and divergence in the variants they carry (Excoffier et al. 2013, Hahn 2018).

The genus *Magnolia* is an ancient clade flowering plants (Angiosperm Phylogeny Group et al. 2016), comprising some 245 species of evergreen or deciduous trees and shrubs distributed in temperate and tropical regions across Southeast and East Asia, the Antilles, and the Americas (Cires et al. 2013). Nearly half of all *Magnolia* species are globally threatened, and the status of at least another third remains unassessed (Rivers et al. 2016). Many *Magnolia* species are valued for their timber, medicinal, and ornamental use. Overexploitation and human disturbance combined with life history traits such as long generation times and slow recruitment have contributed to population declines (Cires et al. 2013). For these reasons, studies have begun to assess the population structure and genetic diversity of a few rare *Magnolia* species (e.g., Isagi et al. 2007 – *M. obovata*, Yang et al. 2022 – M. fistulosa, Budd et al. 2015 – *M. acuminata*, Tamaki et al. 2019 – *M. kobus*, Hernández et al. 2020 – *M. cubensis subsp. acunae*), but to date no studies have included species from Central and South America, an important center of diversity for the genus.

Here, we use *Magnolia yantzazana* F. Arroyo as a case study to explore the application of population genomic and demographic reconstruction methods to understand the population biology and

conservation status of one of the most poorly known and highly threatened Amazonian trees. Specifically, we: (1) test the assumption of extreme rarity in this species from field collections and (2) assess how its effective population size (N_e) has changed through time, with the overarching goal of setting bounds on potential population size estimates. With this study, we offer a framework for advancing understanding of the rarest Amazonian trees and establishing better informed conservation priorities.

2. METHODS

2.1 **Study species and sampling**

Magnolia yantzazana F. Arroyo is an evergreen canopy tree reaching up to 25 m in height (Figure 1b), described from the premontane humid forests on the western slopes of the Cordillera del Cóndor in Zamora-Chinchipe Province, Yantzaza canton, in southeastern Ecuador (Arroyo and Pérez 2013). This species has large ovate leaves and ellipsoid fruits (Figure 1c), and has been found growing on sandstone plateaus from 1400 – 1650 m elevation. While beetles are the predominant pollinator of *Magnolia* flowers, Diptera (flies) and Hymenoptera (bees, etc.) have also been observed, and birds are the principal disperser (Thien et al. 1996). *M. yantzazana* is one of the rarest magnolias in Amazonia, a narrow endemic confined to a small geographic area (~20 km^2; Vázquez-García et al. 2015). Its known range is exclusively within the watershed of the Machinaza River, a tributary of the Zamora River, and is entirely within the "Fruta del Norte" mining concession operated by the Canadian-based company Lundin Gold, Inc. under license from the government of Ecuador (see Supporting Information for more information and an extinction risk assessment; Figure S1).

Originally described from a single collection, the Global Biodiversity Information Facility (GBIF.org) holds 31 records for this species, including 15 unique collections and their duplicates, collected between 2008 and 2024 (Figure S1, GBIF.org). For this study, *M. yantzazana* trees that were

125 previously identified to species and tagged in survey plots within the intact forest for the Lundin Gold
126 mining concession were located, and individuals selected to represent the known geographic distribution
127 and elevation range of the species; a total of five trees were sampled (N=5). Voucher collections were
128 obtained for individuals, in some cases with flowers or fruit as well as leaves, and deposited in the
129 Ecuadorian herbaria ECUAMZ and LOJA (Table 1, Figure 2).

130 **2.2 DNA extraction and sequencing**

131 Leaf tissue was field-collected and stored in silica. High molecular weight DNA was extracted from ~50
132 mg dried tissue using Aboul-Maaty and Oraby's (2019) modified CTAB protocol for non-model plants,
133 with the following modifications: samples were ground with a bead-beater, incubated at -20˚C overnight
134 (and up to 24 hrs), and the DNA pellet was washed 2x with 500 µl 70% ethanol. DNA was checked for
135 quality and quantity on a 0.8% agarose gel and Qubit 2.0 fluorometer dsDNA HS assay kit (Invitrogen)
136 before being sent to Floragenex (Beaverton, OR) for double-digest RAD (ddRAD) library prep and
137 sequencing. ddRAD is a variation of RAD-sequencing that uses two restriction enzymes to create more
138 evenly distributed and predictable cut sites, allowing for better genome coverage, less data loss, and
139 fewer sequencing errors, ideal for non-model species. Samples were ligated with 6 bp barcodes and
140 digested with the PstI/MseI +2 enzyme pair.

141 **2.3 Sequence filtering and assembly**

142 Genomic data were demultiplexed using iPYRAD v. 0.9.95. (Eaton & Overcast 2020) with an average 3.13
143 million ± 1.03 million reads recovered per sample. Reads were assembled both de novo and mapped to
144 the reference genome *Magnolia sinica* (PRJNA774088) in iPYPRAD and the two assemblies were
145 compared. Both assemblies used the default parameters in iPYRAD except the minimum number of
146 samples required per locus was set to 2 due to the small sample size. The final de novo dataset retained
147 3129 SNPs across 2828 loci, with a final concatenated sequence length of 358,502 total sites and 51.34%

148 missing data. The reference assembly retained 1419 SNPs across 2214 loci, with a final concatenated
149 sequence of 269,528 bp and 42.6% missing data.

150 SNPs from both assemblies were filtered in VCFtools v. 0.1.16 (Danecek et al. 2011) for minor allele
151 frequency (-maf) and missing data (-max-missing). The MAF filter was set to 0.05 and three levels of
152 missing data were tested (0.2, 0.4, and 0.6). The missing data filter of 0.4 was selected for downstream
153 analyses as it optimized both the number of sites and stringent filtering, with 3114 sites retained for the
154 de novo assembly and 1339 sites retained for the reference. Finally, sites were thinned to one SNP per
155 locus (--thin 1000) to avoid linkage disequilibrium, with 1192 and 767 SNPs retained in the de novo and
156 reference assemblies, respectively.

157 **2.4 Genomic diversity and structure**

158 The nucleotide diversity (π), expected heterozygosity (H_e), observed heterozygosity (H_o), and inbreeding
159 coefficient (F_{IS}) were calculated using VCFtools for both assemblies individually. Population substructure
160 was assessed using the Bayesian clustering algorithm STRUCTURE v.2.3.4 (Pritchard et al. 2000) and a
161 principal components analysis (PCA), both implemented in iPYRAD. STRUCTURE analyses partition
162 individuals into genetic clusters based on their genotypes, where the clusters represent genetically
163 distinct groups which may correspond to actual populations, subpopulations, or subgroups within
164 populations. The assignment of samples into K distinct genetic groups was assessed using mean log
165 probability and DeltaK. For the PCA, we ran 25 replicate analyses that subsampled different random sets
166 of unlinked SNPs, plotting the centroid of all points for each sample.

167 **2.5 Inference of demographic history**

168 We used Stairway plot 2 (Liu and Fu 2020) to infer the demographic history of *M. yantzazana*, using a
169 mutation rate of 4e -9 (calculated for the reference species *M. sincia*, Yang et al. 2022). Generation time
170 for *M. yantzazana* is unknown, but as generation time for *M. sinica* is estimated to be 10 years from
171 cultivation records (Yang et al. 2022), and other magnolias have known generation times of up to 25

172 years, we ran the model with three different generation times: 10, 25, and 50 years. The folded (de novo
173 assembly) and unfolded (reference assembly) site-frequency-spectrum was generated in easySFS
174 (https://github.com/isaacovercast/easySFS; Gutenkunst et al. 2009) from the thinned SNP matrices,
175 selecting the projection that retained the highest number of segregating sites for each. Demographic
176 history was inferred at the population level, ignoring potential substructure, due to the limited number
177 of individuals sampled.

178 3. RESULTS

179 **3.1 Genomic diversity and structure**

180 The genetic diversity statistics calculated for individuals and the population were similar across the two
181 datasets (de novo and reference, Table 2). Observed heterozygosity (H_o) was low compared to expected
182 heterozygosity (H_e) within individuals and the population. Among individuals, H_o values ranged from
183 0.158–0.232 (de novo) and 0.158–0.254 (reference), and H_e ranged from 0.485–0.509 (de novo) and
184 0.461–0.494 (reference). The estimated population H_o was 0.197 ± 0.03 (de novo) and 0.213 ± 0.03
185 (reference), and H_e was 0.497 ± 0.009 (de novo) and 0.480 ± 0.01 (reference) (Table 2). Nucleotide
186 diversity (π) was relatively high among sequences at 0.523 (de novo) and 0.504 (reference); the
187 inbreeding coefficient (F_{IS}) was also high at 0.604 (de novo) and 0.555 (reference), indicative of a highly
188 inbred population (Table 2).

189 STRUCTURE analyses indicate the most likely number of genetic clusters in the de novo dataset is
190 K = 3, from the estimated mean log probability and DeltaK after running K = 1 to 5 (Figure 3). This
191 substructure was generally supported by the PCA (Figure 3); however, the substructure did not directly
192 correspond to sampling locality (Figure 2). Individual P1 from the Portal site clustered with individuals
193 from the Cantera site (C1 and C2), and had a distinct ancestry from the second individual collected at
194 that site (P2). The P2 individual showed homozygous ancestry under the K = 3 clustering scenario (Figure
195 3, cluster shown in purple), and this cluster was only identified in one other individual from a different

196 locality (C1). The individual sampled from the most distant locality (NAR, ~4.3 km from Cantera and ~5.7
197 km from Porto) was homozygous for a different genetic cluster (Figure 3, cluster shown in orange) and
198 this cluster was identified in all other individuals except P2.

199 **3.2 Demographic history**

200 The models show a general pattern of population collapse over the last several hundred thousand years
201 in both the folded (unmapped) and unfolded (reference mapped) datasets and all generation time
202 scenarios (t_g = 10, 25, 50 years; Figure 4). The demographic history inferred from the folded SFS suggests
203 *M. yantzazana* reached a maximum N_e around 200,000 and began to decline 100,000 years ago (t_g = 10),
204 and 400,000 years ago (t_g = 25, 50), stabilizing with a N_e of approximately 1,000–2,000 between 1,000 (t_g
205 = 10) and 10,000 (t_g = 50) years ago. The demographic histories inferred from the unfolded (reference
206 mapped) SFS showed similar population trends, though they were generally associated with narrower
207 confidence intervals, and the magnitude of the maximum inferred N_e was larger, reaching approximately
208 500,000. N_e began to decline between 100,000 years ago (t_g = 10) and 400,000 years ago (t_g = 50), again
209 stabilizing with a N_e of 4,000–20,000 between 2,000 and 10,000 years ago. Under all scenarios, the
210 models infer at least one prolonged period of stability lasting at least 50,000 years, following the initial
211 population crash. However, the models disagree on the inferred N_e during this stabilizing period, with the
212 folded model predicting a N_e ~5,000 and the unfolded ~10,000.

213 4. DISCUSSION

214 Even in the largest forest inventories, most Amazonian tree species appear as singletons or are
215 completely undetected, hyper-rare and for all practical purposes, invisible to conservation efforts. As of
216 2024, the Amazonian Tree Diversity Network (ATDN) maintains forest inventory plots covering >2000 ha
217 and 5,122 species, less than half of the named Amazonian trees. Thousands more remain completely
218 unknown to science. *M. yantzazana* is one such hyper-rare species, originally described as a singleton
219 and after thorough field investigation, is still known only from a single locality and a handful of

220 individuals. However, using the genomic information stored in these field collections, we can infer
221 population sub-structuring and demographic trends to provide a first conservation assessment.

222 Population genomic statistics suggest that while there is relatively high nucleotide diversity
223 within *M. yantzazana* among sequences ($\pi \approx 0.5$), there is also evidence of a loss of heterozygosity ($H_e >$
224 H_o) and inbreeding ($F_{IS} > 0.5$). This may reflect a historically larger population with diverse ancestry, that
225 due to a population bottleneck now experiences restricted mating with limited gene flow. The values
226 calculated here for *M. yantzazana* are comparable to those obtained for other rare *Magnolias*. For
227 example, Yang et al. (2022) calculated an F_{IS} value of 0.316 for *M. fistulosa*, and Hernandez et al. (2020)
228 found similar values for π (0.504), H_o (0.434), and H_e (0.469) in *M. cubensis subsp. acunae*. Such a high F_{IS}
229 is particularly concerning for conservation, as inbreeding may reduce the fitness of the population and
230 reduce the genetic variation available for natural selection to act upon, leading to inbreeding depression
231 and greater vulnerability to changing conditions (Lewontin 1974; Hoban et al. 2020). STRUCTURE
232 analysis found individuals with ancestry from multiple genetic clusters in both the Cantera and Portal
233 sampling locations, indicating a history of admixture. However, two of the five individuals sampled (P2
234 and NAR) showed ancestry from a single genetic group, further indicating low genetic mixing and
235 inbreeding.

236 The inferred demographic history from both datasets (folded and unfolded) and all generation
237 time scenarios (t_g = 10, 25, 50) suggest that *M. yantzazana* experienced a population collapse in the late
238 Pleistocene. Looking at a global sample of 15 economically important plants with habits including herbs,
239 shrubs, and trees, Patil et al. (2021) found a similar late Pleistocene bottleneck trend across tropical
240 species. Patil et al. (2021) proposed this joint decline in N_e for tropical species globally corresponds to
241 changing environmental conditions, such as prolonged drought and a decrease in CO_2 concentration.
242 However, conditions in the western Amazon became wetter over the same period (Cheng et al. 2013).

More studies on non-model tropical forest trees are required to disentangle the potential drivers and regionality of these perceived late-Pleistocene population bottlenecks.

Perhaps more significant is the dynamism of commonness and rarity through time. *M. yantzazana* was once at least 1–2 orders of magnitude more common than it is today, now seemingly isolated within a single watershed and mining concession (Figure S1; GBIF.org). While some species may recover their populations (as Patil et al. 2021 proposed for *Faidherbia albida*) other once rarer species surely capitalized on these community shifts and are now more common on the landscape. One such example is the "hyperdominant" species complex *Protium heptaphyllum*, which experienced an increase in population size after diverging from its ancestors (ca. 5 mya), followed by several diversification and population expansion events throughout the Pleistocene (Damasco et al. 2021). Understanding how patterns of commonness and rarity change over time and the causes are fundamental questions in ecology, particularly in hyper-diverse tropical forests, that have traditionally been unanswerable given the coarseness of paleo reconstructions. Whether the species that are common today have always been common underpins our most basic understanding of community structure and function, and is critical for both biodiversity conservation and community ecology theory testing.

Demographic reconstructions have inherent uncertainties. N_e is often an underestimate of N_c and has a lag time of several generations before significant changes in N_c may be reflected. Thus, N_e is best used to set the bounds of possible census sizes and as a proxy for genetic erosion, as N_e has an inverse and non-linear relationship with genetic erosion which accelerates as N_e declines (Hahn 2018; Hoban et al. 2020). The demographic histories inferred here, for all models, suggest the bounds of modern estimates of N_e for *M. yantzazana* are ~10^3 individuals (over the last 1000–2000 years). This is a comparably small population for an Amazonian tree, where populations of the most common species have been estimated to be >10^8 and only the rarest 5800 species have population estimates of <1000 (ter Steege et al. 2013). Generation time is one of the greatest uncertainties in these models (Liu & Fu

267 2015), yet is ill-defined across the demographic literature. Further, these values are either unknown or
268 difficult to constrain in rare species, much less common ones, with implications for the biogeographic
269 interpretations of such reconstructions (i.e. Caswell 2009; Figure 4). In this paper, we dealt with this
270 uncertainty for *M. yantzazana* by modeling with three different generation times spanning the range of
271 values estimated for other *Magnolia* species (i.e. Yang et al. 2022) and long-lived tropical forest trees
272 generally (Lieberman et al. 2009)

273 While a larger sample size may increase the informativeness of these models and statistics, such
274 sampling is not always possible, particularly with rare tropical species. At least half of all tree species
275 predicted for the Amazonian basin have fewer than three collections, with 90% having < 90 (ter Steege et
276 al. 2016). Simulations have shown that sampling many SNPs from across the genome (>1000 SNPs) can
277 accurately estimate genomic diversity with small sample sizes, and increasing sample size beyond eight
278 individuals has diminishing returns (Nazareno et al. 2017). As less than half of all Amazonian trees are
279 known from more than a few collections (ter Steege et al. 2016), efforts should focus on leveraging the
280 genomic data stored in samples already in collections, which, when combined with field census data, can
281 help distinguish the truly rare species in need of conservation protection from those simply under-
282 sampled. Here, the combined evidence from genomic diversity statistics, demographic inference, and
283 census data indicates *M. yantzazana* is a truly rare tree rather than an artifact of under-sampling, with
284 an estimated modern population size of ~10^3, though once 1-2 orders of magnitude more common.

285 Ongoing acute threats from human disturbance (notably mining) in the region suggest this
286 species is now of significant concern and unlikely to recover without urgent management intervention.
287 These threats, combined with the data presented here, support an IUCN status assignment of Critically
288 Endangered (CR; IUCN 2012, 2024; S1). Future work may benefit from modeling many species from the

same geographic area together to help to shed light on population trends in the region and highlight specific localities of concern.

5. DATA AVAILABILITY STATEMENT

All genetic data generated for this study are deposited in the GenBank online repository under PRJNA1206534 (www.ncbi.nlm.nih.gov/bioproject/PRJNA1206534).

6. REFERENCES

Aboul-Maaty, N. A. F., & Oraby, H. A. S. (2019). Extraction of high-quality genomic DNA from different plant orders applying a modified CTAB-based method. *Bulletin of the National Research Centre*, *43*(1), 1-10.

Andrews KR, Good JM, Miller MR, Luikart G, and Hohenlohe PA. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17 (2) 81 - 92.

Angiosperm Phylogeny Group, Chase, M. W., Christenhusz, M. J., Fay, M. F., Byng, J. W., Judd, W. S., ... & Stevens, P. F. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, *181*(1), 1-20.

Arroyo, F. & Pérez, A. J. (2013). Three new species of *Magnolia* (Magnoliaceae) from Ecuador. Phytoneuron, 55, 1-6.

Budd, C., Zimmer, E., & Freeland, J. R. (2015). Conservation genetics of Magnolia acuminata, an endangered species in Canada: can genetic diversity be maintained in fragmented, peripheral populations? *Conservation Genetics*, *16*, 1359-1373.

Caswell, H. (2009). Stage, age and individual stochasticity in demography. *Oikos*, *118*(12), 1763-1782.

Cheng, H., Sinha, A., Cruz, F. W., Wang, X., Edwards, R. L., d'Horta, F. M., ... & Auler, A. S. (2013). Climate change patterns in Amazonia and biodiversity. *Nature Communications*, *4*(1), 1411.

Cires, E., De Smet, Y., Cuesta, C., Goetghebeur, P., Sharrock, S., Gibbs, D., ... & Samain, M. S. (2013). Gap analyses to support ex situ conservation of genetic diversity in Magnolia, a flagship group. *Biodiversity and Conservation*, *22*, 567-590.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156-2158.

Eaton, D. A., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, *36*(8), 2592-2594.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, *9*(10), e1003905.

Feeley, K. J., & Silman, M. R. (2011a). The data void in modeling current and future distributions of tropical species. *Global Change Biology*, 17(1), 626-630.

Feeley, K. J., & Silman, M. R. (2011b). Keep collecting: accurate species distribution modelling requires more collections than previously thought. *Diversity and Distributions*, *17*(6), 1132-1140.

Gautschi, D., Heinsohn, R., Ortiz-Catedral, L., Stojanovic, D., Wilson, M., Crates, R., ... & Neaves, L. (2024). Genetic diversity and inbreeding in an endangered island-dwelling parrot population following repeated population bottlenecks. *Conservation Genetics*, 1-13.

Gentry, Alwyn H. "Patterns of neotropical plant species diversity." *Evolutionary Biology: Volume 15*. Boston, MA: Springer US, 1982. 1-84.

Guevara Andino, J. E., Pitman, N. C., Ulloa Ulloa, C., Romoleroux, K., Fernández-Fernández, D., Ceron, C., ... & Ter Steege, H. (2019). Trees of Amazonian Ecuador: a taxonomically verified species list with data on abundance and distribution.

Hahn MW. (2018). Molecular Population Genetics. Oxford University Press.

Hoban, S., Bruford, M., Jackson, J. D. U., Lopes-Fernandes, M., Heuertz, M., Hohenlohe, P. A., ... & Laikre, L. (2020). Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. *Biological Conservation*, *248*, 108654.

Hubbell S. (2013). Tropical rain forest conservation and the twin challenges of diversity and rarity. *Ecology and Evolution*, 3 (10) 3263 – 3274.

IUCN. (2012). IUCN Red List Categories and Criteria, Version 3.1. Second edition. Prepared by the IUCN Species Survival Commission. IUCN, Gland, Switzerland; Cambridge, United Kingdom.

IUCN. (2024). Guidelines for using the IUCN Red List categories and criteria. Version 16. Prepared by the IUCN Standards and Petitions Committee.

Isagi, Y., Tateno, R., Matsuki, Y., Hirao, A., Watanabe, S., & Shibata, M. (2007). Genetic and reproductive consequences of forest fragmentation for populations of Magnolia obovata. *Sustainability and Diversity of Forest Ecosystems: An Interdisciplinary Approach*, 382-389.

Kebaïli C, Sherpa S, Rioux D, and Després L (2021). Demographic inferences and climatic niche modelling shed light on the evolutionary history of the emblematic cold-adapted Apollo butterfly at regional scale. *Molecular Ecology*, 31, 448 – 466.

Lemopoulos, A., Prokkola, J. M., Uusi-Heikkilä, S., Vasemägi, A., Huusko, A., Hyvärinen, P., ... & Vainikka, A. (2019). Comparing RADseq and microsatellites for estimating genetic diversity and relatedness—Implications for brown trout conservation. *Ecology and Evolution*, *9*(4), 2106-2120.

Lewontin, R. C. (1974). The genetic basis of evolutionary change. Columbia University Press

Lieberman, D., Lieberman, M., Hartshorn, G., & Peralta, R. (1985). Growth rates and age-size relationships of tropical wet forest trees in Costa Rica. *Journal of Tropical Ecology*, 1(2), 97-109.

Linan, A. G., Lowry, P. P., Miller, A. J., Schatz, G. E., Sevathian, J. C., & Edwards, C. E. (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*(18), 4520-4537.

Liu, X., & Fu, Y. X. (2015). Exploring population size changes using SNP frequency spectra. *Nature Genetics*, *47*(5), 555-559.

Liu, X., & Fu, Y. X. (2020). Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biology*, *21*(1), 280.

Nazareno, Alison G., et al. "Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species." *Molecular Ecology Resources* 17.6 (2017): 1136-1147.

Parchman, T. L., Jahner, J. P., Uckele, K. A., Galland, L. M., & Eckert, A. J. (2018). RADseq approaches and applications for forest tree genetics. *Tree Genetics & Genomes*, *14*, 1-25.

Patil, A. B., Shinde, S. S., Raghavendra, S., Satish, B. N., Kushalappa, C. G., & Vijay, N. (2021). The genome sequence of Mesua ferrea and comparative demographic histories of forest trees. *Gene*, *769*, 145214.

Pitman, N. C., Terborgh, J., Silman, M. R., & Nuñez V, P. (1999). Tree species distributions in an upper Amazonian forest. *Ecology*, *80*(8), 2651-2661.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959.

Rabinowitz, D. (1981). Seven forms of rarity. Biological aspects of rare plant conservation.

Rivers, M., Beech, E., Murphy, L., & Oldfield, S. (2016). *The Red List of Magnoliaceae: revised and extended*. Richmond, Surrey, UK: Botanic Gardens Conservation International (BGCI).

Tamaki, I., Kawashima, N., Setsuko, S., Lee, J. H., Itaya, A., Yukitoshi, K., & Tomaru, N. (2019). Population genetic structure and demography of Magnolia kobus: variety borealis is not supported genetically. *Journal of Plant Research*, *132*, 741-758.

Ter Steege, H., Pitman, N. C., Sabatier, D., Baraloto, C., Salomão, R. P., Guevara, J. E., ... & Silman, M. R. (2013). Hyperdominance in the Amazonian tree flora. *Science*, *342*(6156), 1243092.

Ter Steege, H., Vaessen, R. W., Cárdenas-López, D., Sabatier, D., Antonelli, A., De Oliveira, S. M., ... & Salomão, R. P. (2016). The discovery of the Amazonian tree flora with an updated checklist of all known tree taxa. *Scientific Reports*, *6*(1), 29549.

Ter Steege H, Prado PI, de Lima RAF, Pos E, et al. (2020). Biased-corrected richness estimates for the Amazonian tree flora. *Scientific Reports,* 10, 10130.

Thien, L. B., Kawano, S., Azuma, H., Latimer, S., Devall, M. S., Rosso, S., ... & Jobes, D. (1996). The floral biology of the Magnoliaceae. In *In: Hunt, David, ed. Magnolias and their allies. Proceedings of an International Symposium. Egham, Surrey, UK: Royal Holloway, University of London: 37-58.* (pp. 37-58).

Yang, F., Cai, L., Dao, Z., & Sun, W. (2022). Genomic data reveals population genetic and demographic history of Magnolia fistulosa (Magnoliaceae), a plant species with extremely small populations in Yunnan Province, China. *Frontiers in Plant Science*, *13*, 811312.

7. TABLES AND FIGURES

Table 1. Voucher numbers and sampling locations for the *Magnolia yantzazana* individuals.

Voucher #	ID	Herbaria	Altitude (m)	Latitude	Longitude
Neill 18868	Cantera 1	LOJA, ECUAMZ	1625	3°46’41” S	78 °30’00” W
Neill 18869	Cantera 2	LOJA, ECUAMZ	1620	3 °46’41” S	78 °30’00” W
Neill 18870	Portal 1	LOJA, ECUAMZ	1424	3 °45’49” S	78 °30’30” W
Neill 18871	Portal 2	ECUAMZ	1425	3 °45’48” S	78 °30’30” W
Neill 18872	NAR	ECUAMZ	1445	3 °45’21” S	78 °32’46” W

Table 2. Summary of population genetic diversity statistics for each assembly.

Assembly	π	$H_o \pm SD$	$H_e \pm SD$	F_{IS}
De novo	0.523	0.197 ± 0.03	0.497 ± 0.009	0.604
Reference	0.504	0.213 ± 0.03	0.480 ± 0.01	0.555

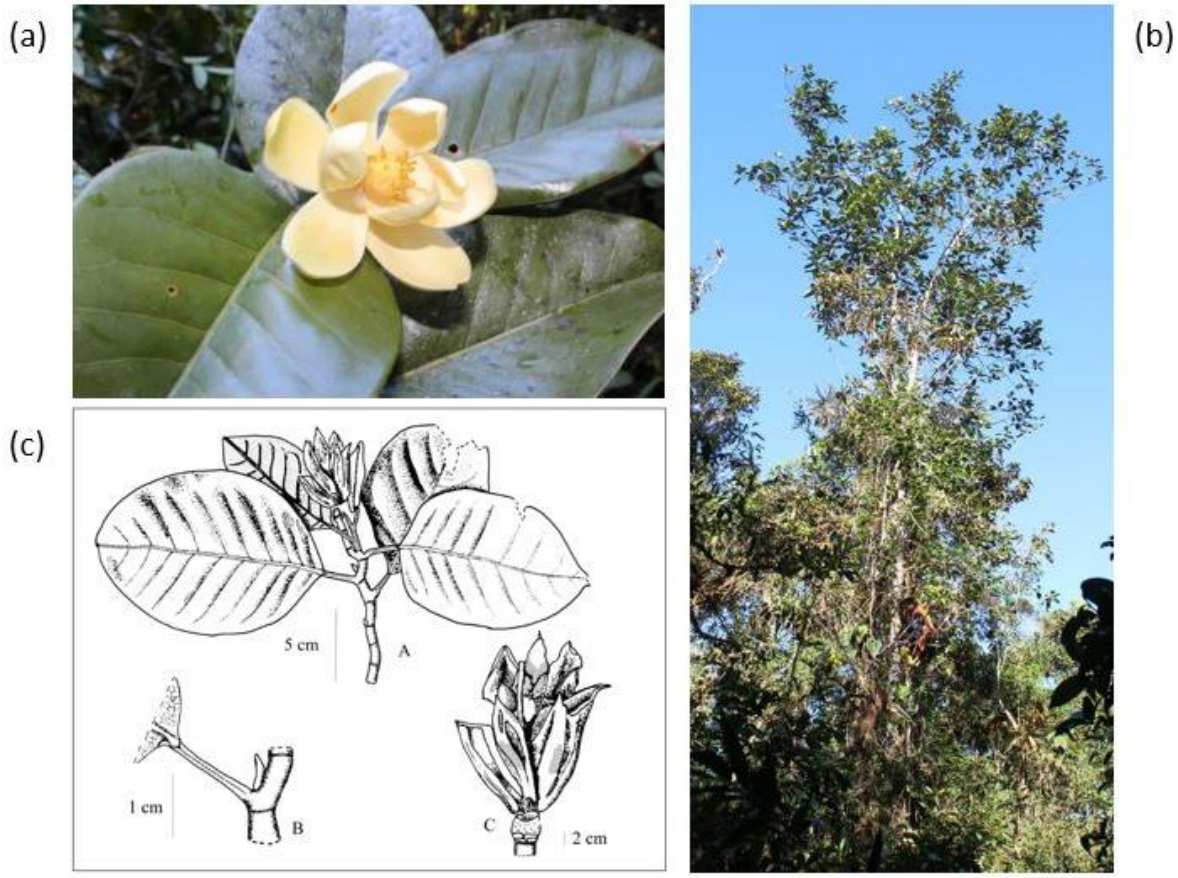

Figure 1. (a) The flower of *Magnolia yantzazana* F. Arroyo (photo by David A. Neill); (b) 25 m tall *M. yantzazana* tree (photo by David A. Neill); (c) *M. yantzazana* holotype showing the fruiting branchlet, detailed petiole, and mature fruit at dehiscence (from original species description in Arroyo and Pérez 2013).

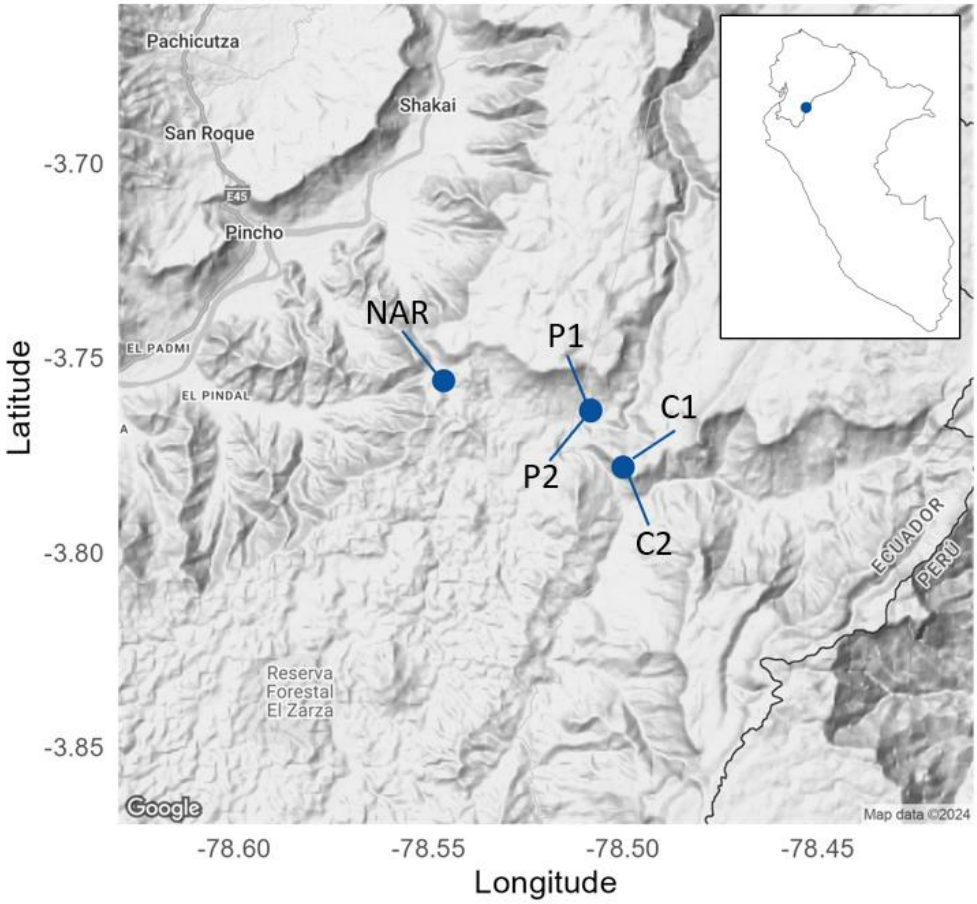

Figure 2. Map of *Magnolia yantzazana* sampling localities in Zamore Chinchipe, Yantzaza canton, Ecuador, within the Ludin Gold Inc. mining concession.

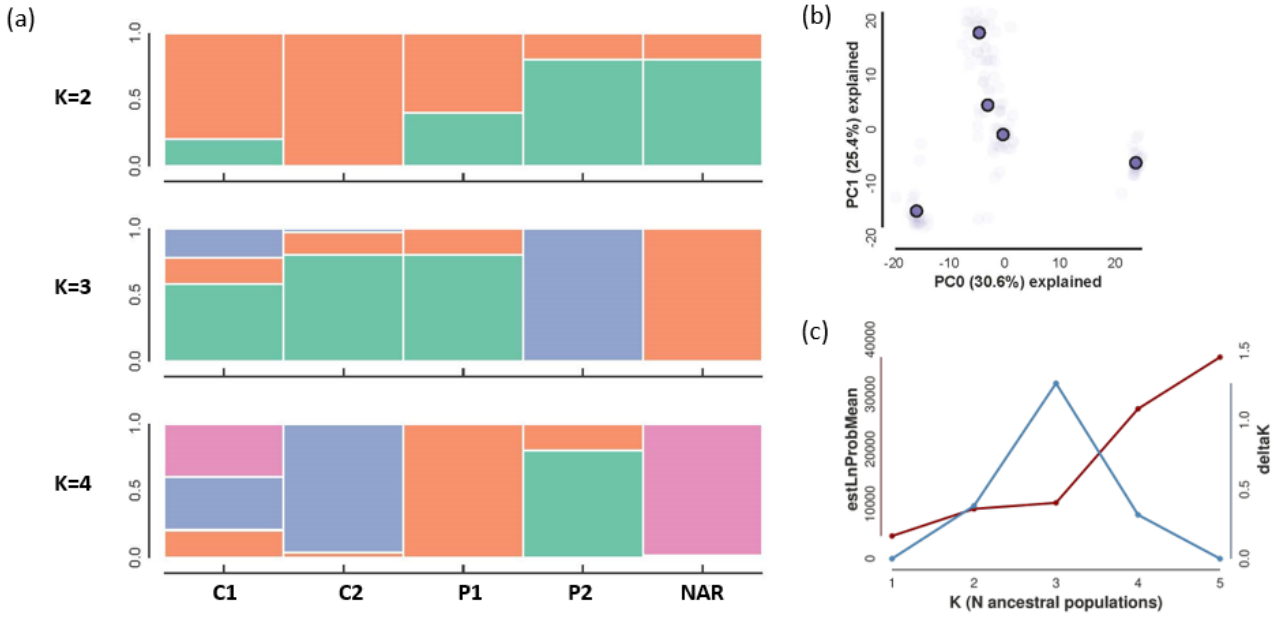

Figure 3. Population structure of *Magnolia yantzazana* F. Arroyo inferred from the de novo assembled data set. (a) STRUCTURE plot results for K=2 to K=4. STRUCTURE analyses partition individuals into genetic clusters based on their genotypes, where the clusters represent genetically distinct groups which may correspond to actual populations, subpopulations, or subgroups within populations. Each column represents an individual, partitioned into segments corresponding to their membership in the inferred clusters; (b) PCA with point clouds showing the 25 replicate analyses of randomly subsampled sets of SNPs — the point in bold is the centroid of all points for each sample; (c) K-value assessment using the mean log probability and DeltaK.

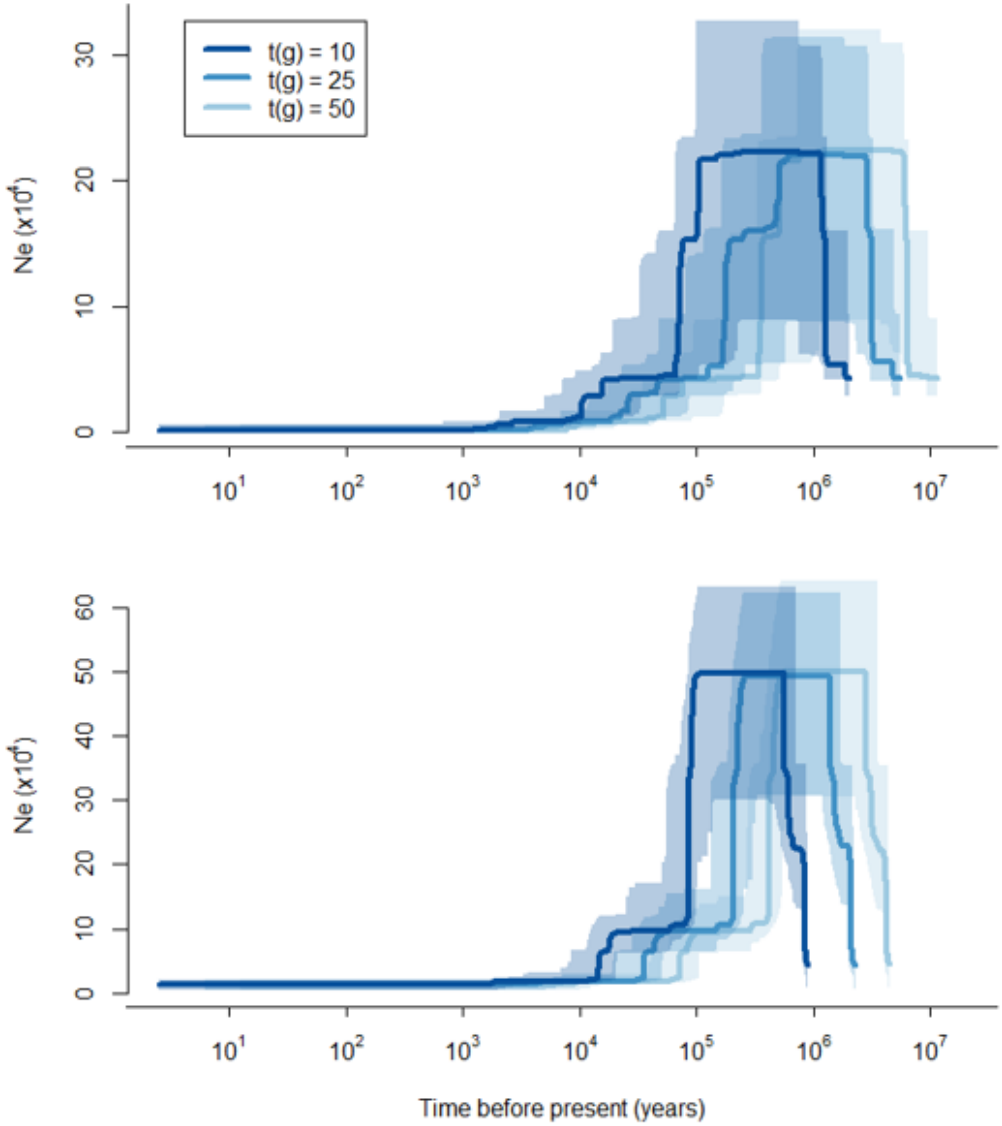

Figure 4. Demographic history of *Magnolia yantzazana* inferred by Stairway plot 2. The three generation time scenarios are represented by different shades of blue with (a) inferenced from the folded SFS and (b) the unfolded (reference mapped) SFS. Lines show predicted change in effective population size N_e over time and shading indicates the 95% confidence interval.