

Microsatellite analysis of populations of the endangered tree *Gomortega keule* suggests pre-Columbian differentiation

Journal:	<i>New Zealand Journal of Botany</i>
Manuscript ID	NZJB-2016-0034.R3
Manuscript Type:	Research Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Munoz-Concha, Diego; Universidad Catolica del Maule, Escuela de Agronomia Davey, Michael; University of Nottingham, Plant and Crop Sciences Division Ribas, Gracia; University of Nottingham, Plant and Crop Sciences Division Mayes, Sean; University of Nottingham - Malaysia Campus, Crops for the Future Research Centre
Keywords:	<i>Gomortega keule</i> , population genetics, fragmentation, genetic structure, megafaunal extinction, relict populations

SCHOLARONE™
Manuscripts

1 **Microsatellite analysis of populations of the endangered tree *Gomortega***
2 ***keule* suggests pre-Columbian differentiation**

3
4 ¹ Diego Muñoz-Concha *

5 ² Michael R Davey

6 ² Gracia Ribas

7 ^{2,3} Sean Mayes

8

9 ¹ Departamento de Ciencias Agrarias, Universidad Católica del Maule, Curicó,
10 Chile

11 ² Plant and Crop Sciences Division, University of Nottingham, Loughborough,
12 LE12 5RD, UK

13 ³ Crops for the Future Research Centre, Jalan Broga, Semenyih, 43500
14 Selangor, Malaysia

15

16 * Corresponding author. Email: dmunoz@ucm.cl

17

18

Style Definition: Balloon Text: English (New Zealand)

Style Definition: Comment Text: English (New Zealand)

19

20 **ABSTRACT**

21

22 Temperate forests have been affected extensively by human activities, resulting
23 in land cover changes and population fragmentation. However, these
24 anthropogenic effects can be superimposed onto the natural history of species,
25 making it difficult to determine which effect is more important for a particular
26 species. *Gomortega keule* is an endangered tree that is found in one of the
27 world's biodiversity hotspots in central-south Chile. Human activities have
28 significantly impacted on the original habitat in this region in recent years and
29 are commonly considered to be the main cause of the scarcity of this species.
30 However, aspects of the natural history of this evergreen tree may also help
31 explain its present-day genetic structure. Therefore, in this study, we undertook
32 microsatellite genotyping of the two southernmost populations of *G. keule*,
33 which are 7.5 km apart from each other and well isolated from other
34 populations. We found that there was genetic differentiation between these
35 populations, suggesting that they exhibited at least some differentiation before
36 becoming isolated, most likely before human activities first impacted the region
37 some two centuries ago. Molecular estimates of their divergence time supported
38 a more ancient differentiation of the populations than would be explained by
39 human activities alone. Therefore, it is possible that their isolation may have
40 followed the extinction of megafaunal seed dispersers around 12,000 years
41 before present in this region, as indicated by fruit characteristics, the absence of
42 recruitment by seedlings and the existence of clonal trees.

43

44 **Keywords:** *Gomortega keule*, population genetics, fragmentation, genetic
45 structure, megafaunal extinction, relict populations

46

47

48 **INTRODUCTION**

49

50 Temperate forests have been significantly affected by human activities,
51 particularly agriculture and forestry, resulting in land use changes and the
52 fragmentation of existing species. However, the life history traits and natural
53 history of species can also be key elements that affect population isolation and
54 thus the population genetics of trees (Bacles and Jump 2011).

55

56 The coastal forests of central-south Chile lie within one of the 25 biodiversity
57 hotspots on Earth (Myers et al. 2000). This region served as a refugium during
58 the Quaternary glacial period, with many species surviving as a result of the
59 temperatures being moderated by the ocean (Villagrán & Armesto 2005).
60 Indeed, the distributions of several woody species are currently restricted to
61 latitudes 36°–40°, which has been attributed to this natural history (Villagrán &
62 Armesto 2005). However, over the last two centuries, this region has also been
63 extensively affected by fires, timber exploitation, land clearance and the
64 introduction of exotic forestry plantations (Armesto et al. 2010), and
65 consequently the population fragmentation and reproductive isolation that is
66 observed in many native woody species in these forests is commonly attributed
67 to anthropogenic disturbance.

68

69 *Gomortega keule* (Mol.) Baillon is an endangered species that is found in
70 central-south Chile. It exhibits a distribution pattern of small, isolated
71 populations over a narrow geographical area near to the coast (Fig. 1) and
72 occurs in only one of the national protected areas in Chile. San Martín &
73 Sánchez (2000) provided details of 22 locations where the species exists, most

74 of which have less than 100 individuals. Wood extraction activities and land
75 clearance for agriculture occurred on a small scale before the arrival of
76 Europeans in central Chile (Armesto et al. 2010), and indigenous people had
77 some impact on the native forests, particularly in the main river basins.
78 However, it is unlikely that *G. keule* populations were strongly affected by
79 these activities because this evergreen tree mainly grows in mountainous areas,
80 often in ravines and on pronounced slopes, where the impacts of indigenous
81 human populations were restricted. Over the last two centuries, however, the
82 areas in which *G. keule* occurs have been considerably affected by the logging
83 of trees and especially the setting of fires for land clearance (Serra et al. 1986;
84 Armesto et al. 2010), as well as the establishment of exotic timber plantations
85 in more recent years, which has led to a reduction in native forest cover from
86 21% in AD 1975 to 7% in AD 2000 in the region where the species occurs
87 (Armesto et al. 2010). Furthermore, individual *G. keule* trees have been found
88 as stump shoots, indicating that the parental trees had been damaged or logged
89 (Serra et al. 1986) and suggesting the loss of individuals in recent decades. In a
90 genetic study of 11 populations using inter simple sequence repeat (ISSR)
91 markers, García-González et al. (2008) concluded that human activities caused
92 the isolation of populations of *G. keule*. However, this assertion should be
93 considered with caution since this is a long-lived species that occurs in a habitat
94 that has experienced only relatively recent anthropogenic disturbance.

95

96 The ecology of *G. keule* has some interesting features. The flowers are
97 pollinated by syrphid flies (Lander et al. 2009). Although barochory has been
98 suggested (Le Quesne & Stark 2006), there have been no studies on seed
99 dispersion. Interestingly, the fruit are among the largest in the Chilean flora and
100 are edible. Janzen & Martin (1982) proposed the idea of 'evolutionary

101 anachronism' for several Central American species, such as *Gymnocladus*
102 *dioicus* (Zaya & Howe 2009), and this could also apply to *Gomortega keule*. If
103 this was the case, the fruit of *G. keule* would have been dispersed by animals
104 that once existed in central Chile, such as *Paleolama*, *Antifer*, *Equus*,
105 *Milodontidae* and *Gomphoteriidae* (Labarca & López 2006). Indeed, the edible
106 pulp and very hard endocarp that encloses the seed (Fig. 2) appear to fit the
107 megafaunal dispersion model (Guimarães et al. 2008). Today, the fruit of these
108 trees rot on the forest floor in the absence of animals that are able to eat the
109 complete fruit and transport the seed. In some regions, domestic animals (cows,
110 pigs and goats) eat the flesh of the fruit but leave the seed. Although small
111 seedlings can be found, they do not grow further, possibly due to foraging by
112 domestic or wild (rabbits, hares) animals, or because of the ecological
113 conditions that occur near settlements (humidity, light, litter alteration). Thus,
114 with the extinction of the megafauna in South America around 12,000 years
115 before present (yr BP) (Barnosky & Lindsey 2010), *G. keule* may have lost one
116 class of its natural seed dispersers, causing it to become confined to areas that
117 are close to the stands and populations where it remains today.

118

119 In this study, we investigated whether the isolation of populations of *G. keule*
120 occurred much longer ago than could be explained by human activities in the
121 last 200 years. To do this, we evaluated the genetic differentiation of two
122 spatially close but distinct populations based on the notion that if these
123 populations showed significant genetic differentiation, a process of genetic
124 differentiation must have already been underway before any superimposed
125 human activities in recent centuries. This could argue for a more complex
126 situation than suggested by García-González et al. (2008), with the megafaunal
127 extinction around 12,000 yr BP potentially contributing to this process.

128

129 The two southernmost populations of *G. keule* were selected for this study,
130 which are only 7.5 km apart and in the same valley, but nearly 70 km from the
131 next nearest population. Importantly, there is a clear geographical separation
132 between these populations today, with no *G. keule* trees between. While most
133 populations of *G. keule* are generally of a similar size to the study populations
134 (60–80 individuals over <1 hectare), we chose these particular populations
135 because they occur in an area with a more limited degree of recent
136 anthropogenic influence, being relatively distant from current human
137 settlements. By contrast, populations near big cities (e.g. Concepción) or
138 agricultural areas (northern area of distribution) are likely to have been affected
139 for a longer period and more deeply by human activities.

140

141

142 MATERIALS AND METHODS

143

144 Sample populations

145

146 Plant materials were obtained from the Maule and Bio-Bio regions of Chile
147 (Table 1) in AD 2007. In population A (Fig. 1; $n = 75$), six *G. keule* trees (trunk
148 diameters >20 cm) were growing from old stumps, while the remainder were
149 young trees or shoots (trunk diameters <10 cm) growing under the canopy in a
150 plantation of *Pinus radiata*. These young individuals were derived from the
151 stumps of old trees that had been cut before or at the time of plantation
152 establishment and so would be expected to have the same genotypes as the
153 original logged trees. No seedlings were seen during collection. By contrast,
154 population B ($n = 63$) mainly comprised very old trees, some of which were the

155 oldest representatives of this species we found (trunk diameters ca. 2 m). As for
156 population A, no seedlings were observed in association with this population.
157 Samples were collected from all of the *G. keule* trees that could be found in
158 each of these populations, with only one sample being collected from each
159 group of shoots or trunks derived from an individual tree stump in population
160 A. In addition, 14 individuals were sampled from a population in the northern
161 area of the species' distribution (190 km away from populations A and B) as an
162 outgroup (group C).

163

164 **Sample collection and DNA extraction**

165

166 Young leaves were collected from each tree or alternatively the cambium was
167 sampled if leaves could not be found for a particular tree. The collected tissues
168 were dried in sealable plastic jars containing self-indicating silica gel crystals,
169 which were renewed as required, and stored at -20°C.

170

171 DNA extraction followed Lander et al. (2007), whereby 5 mg
172 diethyldithiocarbamic sodium salt, 10 mg PVP-40000 and 5 µl β-
173 mercaptoethanol were added per 1 ml of cetyltrimethylammonium bromide
174 (CTAB) solution that had been pre-heated to 65°C. This solution (500 µl) was
175 then added to each sample for DNA extraction.

176

177 **Microsatellite genotyping**

178

179 Each sample was genotyped using seven microsatellite primer pairs. Five of
180 these (Gk-1, Gk-30, Gk-31, Gk-35 and Gk-39) were developed by Lander et al.
181 (2007), while the remaining two (CS2 and CS8) were developed by S. Mayes

182 following the method of Haddrill et al. (2002). The procedure to develop
183 primers CS2 and CS8 involved the digestion and PCR amplification of genomic
184 DNA, hybridisation to filters with artificial SSR repeat oligonucleotides,
185 elution, and amplification. Rather than using cloning, a proportion of the
186 enriched library was pyrosequenced using a 1/16 run (non-titanium reagents;
187 454 Life Sciences, Connecticut, USA).

188

189 To visualise the DNA, 1% agarose gels were prepared in 1× TAE buffer and
190 ethidium bromide was added to a final concentration of 0.5 µg/ml before
191 casting the gels. The PCR products with fluorescently labelled primers were
192 then visualised on 2% agarose gels in 0.5× TBE buffer that were stained with
193 ethidium bromide to a final concentration of 0.1 µg/ml.

194

195 The forward primers were directly labelled with D2, D3 or D4 fluorescent dye
196 (WellRED; Sigma Aldrich, St. Louis, Missouri, USA). The optimum annealing
197 temperature for PCR for each labelled primer was then determined in a Px2
198 Thermal Cycler using an annealing temperature gradient from 45°C to 60°C
199 across the block. PCR for each sample was performed using a 20-µl reaction
200 volume containing 2 µl template DNA, 2.5 µl buffer, 0.2 µl deoxynucleotide
201 (dNTP), 0.05 µl bovine serum albumin (BSA; 10 µg/µl), 0.02 µl of each
202 forward (labelled) and reverse primer (100 µM), 15 µl autoclaved H₂O, and 0.2
203 µl Taq polymerase. Master mixes for multiple reactions were made wherever
204 possible. The reaction conditions were 94°C for 3 min, 30 cycles of 94°C for 1
205 min, 1 min at the specific annealing temperature for each primer, 72°C for 1
206 min, and final extension at 72°C for 20 min. The annealing temperatures were
207 60.9°C (Gk-1), 65.0°C (Gk-30, Gk-31 and Gk-35), 56.8°C (Gk-39) and
208 52.7°C (Gk-44, CS2 and CS8).

209

210 A total of 152 genotypes were analysed for seven loci using capillary
211 electrophoresis (CEQ 8000 Genetic Analysis System; Beckman Coulter,
212 California, USA). The PCR products were pooled according to their different
213 fluorescent tags, with a maximum of three different coloured samples being
214 analysed per well.

215

216 **Microsatellite data analysis**

217

218 The peaks that were produced by capillary electrophoresis were thoroughly
219 examined for each locus and assigned to classes according to Amos et al.
220 (2007) using a software tool developed by the same authors (FlexiBin).

221

222 For each sample, matrices were generated using allele size to score co-dominant
223 markers at each locus. Polymorphic loci (Laurentin 2009) were used to generate
224 unbiased estimates of heterozygosity (Nei 1978) and investigate population
225 structure.

226

227 **Population genetic analysis**

228

229 Samples that were considered to have originated from the same individual were
230 excluded from the population analysis. This included any trees that were both
231 <10 m apart and had identical genotypes in the binary matrix. In total, 19
232 distinct genotypes were identified among 42 samples, resulting in 23 samples
233 (42 minus 19) being excluded. Of the remaining 129 samples that were
234 included in the population analysis, 62 were from population A, 53 from B and
235 14 from C, all of which were assessed for all seven loci.

236

237 The number of alleles per population and per locus, and the allelic richness and
238 private allelic richness per population were calculated with HP-Rare
239 (Kalinowski 2005), using the rarefaction method to account for unequal sample
240 sizes. The observed (H_O) and expected (H_E) heterozygosity for each locus and
241 population were then computed with Arlequin 3.5 (Excoffier et al. 2005) using
242 formulas from Nei (1987), and values of H_O and H_E for the combined
243 populations were calculated with GDA using formulas from Nei (1978).

244

245 According to Laurentin (2009), co-dominant markers can be used to assess the
246 diversity between groups with methodologies based on allele frequencies, such
247 as Wright's F-statistics and Nei's parameters. Therefore, the inbreeding
248 coefficient (F_{IS}) for each locus and population was computed using FSTAT
249 (Goudet 1995) according to Weir & Cockerham (1984). The pairwise fixation
250 index (Rho_{ST} or ρ_{ST}) was calculated for each population using Genepop 4.0.5.3
251 (Rousset 2008), following which a second version of the fixation index (F_{ST})
252 was calculated under the stepwise mutation model using Arlequin.

253

254 Exact tests for deviations from Hardy-Weinberg equilibrium (HWE) were
255 performed with Arlequin and Genepop using the Markov chain method
256 (forecasted chain length: 1,000,000; dememorisation steps: 100,000). Linkage
257 disequilibrium was tested separately for each population using Arlequin and a
258 Bonferroni correction was applied to the P values. The frequency of null alleles
259 was calculated assuming simultaneous inbreeding for each locus and population
260 with the software INEst (Chybicki & Burczyk 2009) using either the population
261 or individual inbreeding model (1,000,000 iterations).

262

263 The similarity index of Nei & Li (1979) was used to evaluate the genetic
264 relationship among individuals, and cluster analysis (UPGMA) and principal
265 coordinate analysis (PCoA) with Euclidean distances were used to visualise
266 these relationships, both of which are appropriate for this type of study
267 (Laurentin 2009). PCoA makes no assumptions about the distribution of the
268 data or population genetics and the Euclidean distance does not consider the
269 common absence of an allele as a shared characteristic (Kloda et al. 2008).

270

271 Cluster analysis and PCoA were performed using the MVSP software
272 (Multivariate Statistical Package version 3.13; Kovach Computing Services,
273 Anglesey, UK). A scatter plot was generated with the two largest eigenvalues
274 along the first two axes and a dendrogram was produced with UPGMA using
275 Nei & Li's index (Nei & Li 1979) as a measure of similarity. All 152 samples,
276 which included those samples that were considered to have originated from the
277 same individual, were included in the PCoA and cluster analysis.

278

279 Population structure was assessed using a Bayesian cluster analysis with the
280 software Structure 2.3.4 (Pritchard et al. 2000). Analyses were performed using
281 the admixture model with independent allele frequencies (20 runs), with burn-in
282 and simulation lengths of 300,000 and 1,000,000 iterations, respectively. The
283 optimal value for K was estimated by calculating the statistic ΔK (Evanno et al.
284 2005). These procedures were initially followed for all samples and then for all
285 samples except group C in an independent set of runs.

286

287 The divergence time of populations A and B was estimated with the software
288 IMA2 (Hey & Nielsen 2007), which estimates population-genetic parameters by
289 calculating posterior probabilities in a Bayesian sampling framework. A total of

290 100,000 genealogies were sampled to estimate the joint posterior probability
291 distribution of the divergence time parameter. Generation times of 25, 100 and
292 175 years were used, as suggested by Jones et al. (2013) for long-lived trees,
293 and a mutation rate of 0.000316 was used, as proposed by Tamaki et al. (2016)
294 for trees. More than 100 runs were made to adjust the priors (values of $q = 20$,
295 $m = 500$ and $t = 2$), allowing for burn-in periods of 127–340 million steps. For
296 each generation time, 13 final estimates were made using different seed
297 numbers, and the mean values of the divergence time (yr BP) and the lower and
298 upper bounds of the estimated 95% highest posterior density intervals were
299 calculated.

300

301

302 **RESULTS**

303

304 **Microsatellite genotyping**

305

306 The number of alleles that was found for each locus is presented in Table 2. All
307 seven markers were polymorphic (major allele frequency <0.95) according to
308 Laurentin (2009). The allelic richness and private allelic richness (after
309 rarefaction) were lower in populations A (2.19 and 0.34, respectively) and B
310 (2.13 and 0.24) than in C (2.63 and 1.61) (Table 2). Private alleles were found
311 for all loci and their occurrence was associated with the presence of rare alleles,
312 as many of them showed low frequencies. Although sample size can affect the
313 number of alleles when a sampling approach is used, to the best of our
314 knowledge populations A and B represented the entire populations.

315

316 The level of heterozygosity that was identified in *G. keule* (Table 3) is similar
317 to or lower than that reported for other endangered trees (Tamaki et al. 2008;
318 Finger et al. 2011; Shepherd & Perrie 2011). The largest difference between H_O
319 and H_E (>0.2) occurred at locus *Gk-31* for population A and locus *Gk-35* for
320 group C (Table 3), with exact tests for deviation from HWE being significant
321 (Bonferroni corrected $P < 0.002$) at each of these loci. The inbreeding
322 coefficient for each locus and population ranged from -0.233 to 0.874 (Table 4),
323 with the highest values ($F_{IS} > 0.8$) being observed for locus *Gk-31* in population
324 A and locus *Gk-35* in group C.

325

326 The pairwise population fixation index ρ_{ST} was 0.045 for A–B, 0.882 for A–C
327 and 0.882 for B–C, while F_{ST} was 0.061, 0.407 and 0.382 for the same pairwise
328 comparisons, respectively. Significant linkage disequilibrium after Bonferroni
329 correction was present in one pairwise comparison for population A and one for
330 population B. The estimated frequency of null alleles was significantly different
331 from zero ($P < 0.001$) for locus *Gk-31* in population A using the population
332 inbreeding model (PIM) and the individual inbreeding model (IIM) with INEst
333 (Chybicki & Burczyk 2009). Similarly, the frequency was also significantly
334 different from zero for locus *Gk-35* in group C ($P < 0.001$ using PIM and $P <$
335 0.01 using IIM). Although the presence of null alleles could have some
336 influence on the accuracy of the statistical analyses, it was not considered likely
337 that it had a major influence on the overall results.

338

339 **Population structure**

340

341 There was some genetic differentiation among the populations studied.
342 Moderate genetic differentiation ($F_{ST} = 0.051$ – 0.150 ; Yeh 2000) was found

343 between populations A and B ($F_{ST} = 0.061$), while very great genetic
344 differentiation ($F_{ST} > 0.25$; Yeh 2000) was found between populations A and C
345 ($F_{ST} = 0.41$), and B and C ($F_{ST} = 0.38$).

346

347 The PCoA analysis (Fig. 3) showed that there was some differentiation between
348 populations A and B, which is consistent with their geographical location: these
349 populations are only 7.5 km apart (Fig. 1) and each occupies <1 hectare.

350

351 The dendrogram (Fig. 4) supported the PCoA analysis, with all samples except
352 those from group C falling into a separate cluster, and samples from
353 populations A and B generally being grouped into two partially overlapping
354 clusters.

355

356 The Bayesian analysis with the software Structure (Fig. 5) appeared to confirm
357 the groups that were identified by the PCoA and the dendrogram analysis. The
358 graphical output of Structure for $K = 2$ indicated some degree of differentiation
359 between populations A and B. When the analysis included group C, the highest
360 ΔK corresponded to two clusters, placing populations A and B in the same
361 cluster. However, the output for $K = 3$ also suggested some differentiation
362 between populations A and B that is biologically meaningful.

363

364 The estimated divergence times for populations A and B and the 95% highest
365 posterior density intervals were 890 (395–2908) yr BP for a generation time of
366 25 years, 3956 yr BP (1978–12,263) for 100 years and 6784 yr BP (3185–
367 20,951) for 175 years.

368

369

370

371 **DISCUSSION**

372

373 **Megafaunal dispersal and natural history of *G. keule***

374

375 It has been proposed that the seeds of *G. keule* require megafaunal dispersal
376 (Muñoz-Concha & Davey 2011), and so it is possible that populations A and B
377 have become increasingly isolated following the extinction of megafauna from
378 Chile, leading to their current level of differentiation. This idea that population
379 differentiation was already underway before human activities affected these
380 populations is supported by a number of observations from the present study.
381 The data suggest that populations A and B, which are geographically very
382 close, have a moderate level of genetic differentiation but are clearly more
383 closely related to each other than to group C, which could argue for the
384 previous existence of a more continuous population that is now fragmented.
385 Since populations A and B occur in the same valley, fragmentation as a result
386 of physical barriers alone can be ruled out. The minimum divergence times
387 calculated in the present study are greater than 200 yr BP (minimum 395 yr BP
388 with 95% confidence for a 25-year generation time), which would argue for
389 pre-existing differentiation before any superimposed effect of human activities.

390

391 The PCoA analysis showed a clearer separation of the populations than was
392 found by García-González et al. (2008), who used ISSR to analyse the genetic
393 structure of 11 populations of *G. keule*, but only sampled some of the
394 individuals from populations A and B (29 and 7, respectively). In addition to
395 the larger number of individuals that were sampled in the present study, the use

396 of co-dominant microsatellite markers improved the resolution of a number of
397 the analyses and also allowed heterozygosity-based indices to be calculated.

398

399 Differentiation between the examined populations of *G. keule* is demonstrated
400 by the fixation index values, which lie within the range of those assessed with
401 microsatellites for other endangered tree species, such as *Magnolia stellata*
402 (Tamaki et al. 2008), *Dalbergia monticola* (Andrianoelina et al. 2009),
403 *Medusagyne oppositifolia* (Finger et al. 2011) and *Pseudopanax ferox*
404 (Shepherd & Perrie 2011). Furthermore, similar levels of genetic differentiation
405 have also been observed among isolated populations of *Myrtus nivellei*, a relict
406 species that is currently experiencing geographic range contraction (Migliore et
407 al. 2013).

408

409 It has previously been argued that clonality plays a role in maintaining
410 unaltered genotypes over a long period of time (even thousands of years) in
411 other woody species with relict and isolated populations, as discussed by Baks
412 et al. (2015), Bradbury et al. (2016) and Migliore et al. (2013). Although there
413 have been no reported studies on the role of clonality as a mechanism to
414 maintain heterozygosity and prevent inbreeding, and no current evidence for the
415 age of clones or trees of *G. keule*, the ability to maintain a population through
416 off-shoots could be a key factor that explains the current level of variation in
417 these populations and so warrants further investigation.

418

419 The generation times of trees can be greatly extended through vegetative
420 (clonal) reproduction. Indeed, clones of relict tree species have been dated back
421 to 3000–11,000 yr BP in *Eucalyptus absitia* (Bradbury et al. 2016) and over
422 13,000 yr BP in *Quercus palmeri* (May et al. 2009). Therefore, since clones of

423 *G. keule* were observed in the present study and by Lander et al. (2010), it is
424 possible that this species has an extended generation time, making a generation
425 time of 25 years unlikely.

426

427 The estimated time for the divergence of populations A and B is over 800 yr
428 BP, placing the population differentiation process well before two centuries
429 ago. Therefore, the isolation of these two populations of *G. keule* was probably
430 well progressed before extensive anthropogenic disturbances first impacted the
431 area around 200 yr BP (Armesto et al. 2010), indicating that population
432 differentiation may have started in glacial times, well before human-mediated
433 fragmentation of the area. This situation was also discussed for the endemic tree
434 *Nothofagus alessandrii* by Torres-Díaz et al. (2007), who concluded that there
435 was probably no gene flow between populations of this species before they
436 were impacted by relatively recent human activities.

437

438 A long regeneration cycle suggests that the study populations have been
439 isolated for a long time. The rare occurrence or complete absence of successful
440 sexual regeneration as evidenced by a lack of seedling establishment in the *G.*
441 *keule* populations sampled has also been observed in other relict tree species
442 such as *Myrtus nivellei* (Migliore et al. 2013), *Q. palmeri* (May et al. 2009), *Q.*
443 *hinckleyi* (Bucks et al. 2015) and *E. absita* (Bradbury et al. 2016). The failure of
444 seedlings to establish and develop into new trees may indicate that the effective
445 generation time is very long, but could also reflect the cessation of seedling
446 recruitment due to the more recent introduction of grazing animals such as
447 hares.

448

449 The fruit of *G. keule* matches the definition of megafaunal fruits that was
450 introduced by Guimarães et al. (2008), being 4–7 cm in diameter and yellow,
451 with a small number of large seeds (Fig. 2). The species also matches the
452 predictions of those authors, showing clumped spatial patterns, reduced
453 geographical ranges and high levels of among-population structuring. Other
454 traits that suggest that *G. keule* may have megafaunal seed dispersal
455 characteristics include the lack of modern dispersal agents, large and well
456 protected seed, edible fruit pulp and a strong vegetative propagation ability, as
457 previously discussed for *Gymnocladus dioicus* by Zaya & Howe (2009). In
458 addition, the present-day rare occurrence or complete absence of recruitment by
459 seedlings and the isolated distribution of individuals and populations further
460 support the idea of megafaunal dispersal of *G. keule*.

461

462 A number of questions still need to be answered to better understand the
463 population dynamics of *G. keule*. These include the juvenility and longevity of
464 individuals, the ecological conditions that are required for seedling survival,
465 and the minimum viable population size.

466

467 **Sampling strategy**

468

469 Some of the samples that were found to be genetically identical were clearly
470 shoots from an old tree that had since disappeared, as reported previously (San
471 Martín & Donoso 1996). However, some trees of *G. keule* that were growing
472 <2 m apart were found to be genetically different, allowing us to reject the
473 assumption that close physical proximity indicates that individuals are clonally
474 related. The intricate way in which the trees occupy space on the forest floor is
475 also well illustrated by the case where a root (10 cm diameter) that was sampled

476 from population B was found to be different from three individuals growing <7
477 m away but identical to a large tree growing 15 m away. Therefore, developing
478 an understanding of the spatial relationships within the population will be an
479 important aspect of any future conservation planning.

480

481 The findings of the present study support the idea that vegetative propagation
482 (i.e. regeneration by sprouting) is currently important in natural populations of
483 *G. keule*, as stated by San Martín & Donoso (1996). Some individuals that were
484 physically close to each other had the same genetic profile and so were
485 probably produced from a single tree. Although we did not undertake extensive
486 sampling of shoots that were very close together, the observation of shoots with
487 the same genetic profile may indicate that clonality has played an important role
488 in the recent survival of *G. keule* trees.

489

490 Clonal production is a relatively common event in trees with reduced
491 populations such as *Eucalyptus absita* (Bradbury et al. 2016) and *Myrtus*
492 *nivellei* (Migliore et al. 2013), with some remarkably old individuals being
493 found, as seen in *Quercus palmeri* (May et al. 2009) and *Q. hinckleyi* (Bucks et
494 al. 2015). Since clonality effectively extends the reproductive cycle time by
495 maintaining the original genetic combinations, the estimated divergence times
496 for *G. keule* may be underestimates.

497

498 **Implications for conservation**

499

500 Although the occurrence of small and increasingly isolated populations of *G.*
501 *keule* seems to have preceded human history in the area, anthropogenic effects
502 remain very important for the future conservation status of this species. It is

503 apparent that there has been extensive destruction of the habitat, populations
504 and trees of this species, as evidenced by the fact that most individuals are
505 recently re-grown trees derived from old tree stumps with signs of destruction
506 by humans (logging, fires).

507

508 Long-term isolation and genetic differentiation among populations may have
509 important implications for conservation management strategies for *G. keule* and
510 other species that occur in disjunct populations, particularly where the natural
511 dispersal agents are no longer present. Conservation efforts should first be
512 directed towards the protection of remnant populations and individuals,
513 including the exclusion of cattle and hares from protected zones, and the careful
514 management of light through canopy maintenance, particularly for individuals
515 that are re-growing in current or former forestry plantations. Given the very
516 limited levels of seedling survival at present, it will be difficult to maintain high
517 genetic diversity through natural seed production. Therefore, seeds or
518 vegetative material should be sampled from as many populations as possible to
519 maximise the amount of genetic diversity that is captured for immediate *ex situ*
520 conservation. However, the risks of outbreeding depression should also be
521 evaluated.

522

523 The level of population differentiation that was observed in *G. keule* may offer
524 a model for the conservation of other Chilean species for which the dispersal
525 agents may be extinct, particularly *Pitavia punctata* and *Jubaea chilensis*, with
526 a need to sample from all extreme populations of each of these. The southern
527 populations of *G. keule* are very important from an agricultural and forestry
528 perspective, as each isolated population contains valuable genetic resources for

529 the future domestication and genetic breeding of the species, and may contain
530 genetic information that is not found among other extant populations.

531

532 **ACKNOWLEDGEMENTS**

533

534 We would like to thank Fernando Campos (Corporación Nacional Forestal de
535 Chile) for field assistance, and Roberto Muñoz (Forestal Celco S.A.) and
536 Forestal Tierra Chilena S.A. for providing permission for field visits. We also
537 acknowledge suggestions for laboratory protocols from Dr. Tonya A. Lander,
538 and Cristian Echeverría, Pablo San Martín and Joselyn San Juan for providing
539 data essential for creating the map in Fig. 1. This research was part of Diego
540 Muñoz-Concha's PhD programme, funded by Universidad Católica del Maule
541 and Comisión Nacional de Investigación Científica y Tecnológica – Gobierno
542 de Chile.

543

544

545 **CONFLICT OF INTEREST**

546

547 The authors declare no conflict of interest.

548

549

550 **REFERENCES**

551

552 Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS 2007.
553 Automated binning of microsatellite alleles: problems and solutions.
554 *Molecular Ecology Notes* 7: 10–14.

555

- 556 Andrianoelina O, Favreau B, Ramamonjisoa L, Bouvet JM 2009. Small effect
557 of fragmentation on the genetic diversity of *Dalbergia monticola*, an
558 endangered tree species of the eastern forest of Madagascar, detected
559 by chloroplast and nuclear microsatellites. *Annals of Botany* 104:
560 1231–1242.
561
- 562 Armesto JJ, Manuschevich D, Mora A, Smith-Ramírez C, Rozzi R, Abarzúa
563 AM, Marquet PA 2010. From the Holocene to the Anthropocene: a
564 historical framework for land cover change in southwestern South
565 America in the past 15000 years. *Land Use Policy* 27: 148–160.
566
- 567 Backs JR, Terry M, Klein M, Ashley MV 2015. Genetic analysis of a rare
568 isolated species: a tough little West Texas oak, *Quercus hinckleyi* C.H.
569 Mull. *The Journal of the Torrey Botanical Society* 142(4): 302–313.
570
- 571 Bacles CFE, Jump AS 2011. Taking a tree's perspective on forest
572 fragmentations genetics. *Trends in Plant Science* 16: 13–18.
573
- 574 Barnosky AD, Lindsey EL 2010. Timing of Quaternary megafaunal extinction
575 in South America in relation to human arrival and climate change.
576 *Quaternary International* 217: 10–29.
577
- 578 Bradbury D, Grayling PM, MacDonald B, Hankinson M, Byrne M 2016.
579 Clonality, interspecific hybridisation and inbreeding in a rare mallee
580 eucalypt, *Eucalyptus absita* (Myrtaceae), and implications for
581 conservation. *Conservation Genetics* 17: 193–205.
582

- 583 Chybicky IJ, Burczyk J 2009. Simultaneous estimation of null alleles and
584 inbreeding coefficients. *Journal of Heredity* 100: 106–113.
585
- 586 Evanno G, Regnaut S, Goudet J 2005. Detecting the number of clusters of
587 individuals using the software Structure: a simulation study.
588 *Molecular Ecology* 14: 2611–2620.
589
- 590 Excoffier L, Laval G, Schneider S 2005. Arlequin ver. 3.0: an integrated
591 software package for population genetics data analysis. *Evolutionary
592 Bioinformatics Online* 1: 47–50.
593
- 594 Finger A, Kettle CJ, Kaiser-Bunbury CN, Valentin T, Doudee D 2011. Back
595 from the brink: potential for genetic rescue in a critically endangered
596 tree. *Molecular Ecology* 20: 3773–3784.
597
- 598 García-González R, Carrasco B, Peñailillo P, Letelier L, Herrera R, Lavandero
599 B, Moya M, Caligari PDS 2008. Genetic variability and structure of
600 *Gomortega keule* (Molina) Baillon (Gomortegaceae) relict
601 populations: geographical and genetic fragmentation and its
602 implications for conservation. *Botany* 86: 1299–1310.
603
- 604 Goudet J 1995. Fstat version 1.2: a computer program to calculate F-statistics.
605 *Journal of Heredity* 86: 485–486.
606
- 607 Guimarães PR, Galetti M, Jordano P 2008. Seed dispersal anachronisms:
608 rethinking the fruits extinct megafauna ate. *PLoS One* 3: e1745 (doi:
609 10.1371/journal.pone.0001745).

610

611 Haddrill PR, Majerus MEN, Mayes S 2002. Isolation and characterization of
612 highly polymorphic microsatellite loci in the 2-Spot Ladybird, *Adalia*
613 *bipunctata*. *Molecular Ecology Notes* 2: 316–319.

614

615 Hey J, Nielsen R 2007. Integration within the Felsenstein equation for
616 improved Markov chain Monte Carlo methods in population genetics.
617 *Proceedings of the National Academy of Sciences* 104: 2785–2790.

618

619 Janzen DH, Martin PS 1982. Neotropical anachronisms: the fruits the
620 Gomphotheres ate. *Science* 215: 19–27.

621

622 Jones FA, Cerón-Souza I, Hardesty BD, Dick CW 2013. Genetic evidence of
623 Quaternary demographic changes in four rain forest tree species
624 sampled across the Isthmus of Panama. *Journal of Biogeography* 40:
625 720–731.

626

627 Kalinowski ST 2005. HP-Rare 1.0: a computer program for performing
628 rarefaction on measures of allelic richness. *Molecular Ecology Notes*
629 5: 187–189.

630

631 Kloda JM, Dean PDG, Maddren C, MacDonald DW, Mayes S 2008. Using
632 principle component analysis to compare genetic diversity across
633 polyploidy levels within plant complexes: an example from British
634 Restharrow (*Ononis spinosa* and *Ononis repens*). *Heredity* 100: 253–
635 260.

636

- 637 Labarca RO, López PG 2006. Los mamíferos finipleistocénicos de la formación
638 Quebrada Quereo (IV Región Chile): biogeografía, bioestratigrafía e
639 inferencias paleoambientales. *Mastozoología Neotropical* 13: 89–101.
640
- 641 Lander TA, Boshier DH, Harris SA 2007. Isolation and characterization of
642 eight polymorphic microsatellite loci for the endangered, endemic
643 Chilean tree *Gomortega keule* (Gomortegaceae) *Molecular Ecology*
644 *Notes* 7: 1332–1334.
645
- 646 Lander TA, Boshier DH, Harris SA 2010. Fragmented but not isolated:
647 contribution of single trees, small patches and long-distance pollen
648 flow to genetic connectivity for *Gomortega keule*, an endangered
649 Chilean tree. *Biological Conservation* 143: 2583–2590.
650
- 651 Lander TA, Harris SA, Boshier DH 2009. Flower and fruit production and
652 insect pollination of the endangered Chilean tree, *Gomortega keule* in
653 native forest, exotic pine plantation and agricultural environments.
654 *Revista Chilena de Historia Natural* 82: 403–412.
655
- 656 Laurentin H 2009. Data analysis for molecular characterization of plant genetic
657 resources. *Genetic Resources and Crop Evolution* 56: 277–292.
658
- 659 Le Quesne C, Stark D 2006. *Gomortega keule* (Mol.) Baillon. In: Donoso C ed.
660 Las especies arbóreas de los bosques templados de Chile y Argentina,
661 Autoecología. Valdivia, Marisa Cuneo Ediciones. Pp 277–284.
662

- 663 May MR, Provance MC, Sanders AC, Ellstrand NC, Ross-Ibarra J 2009. A
664 Pleistocene clone of Palmer's Oak persisting in Southern California.
665 PLoS ONE 4: e8346.
666
- 667 Migliore J, Baumel A, Juin M, Fady B, Roig A, Duong N, Médail F 2013.
668 Surviving in mountain climate refugia: new insights from the genetic
669 diversity and structure of the relict shrub *Myrtus nivellei* (Myrtaceae)
670 in the Sahara Desert. PLoS ONE 8: e73795.
671
- 672 Muñoz-Concha D, Davey MR 2011. *Gomortega keule*, the neglected and
673 endangered Chilean fruit tree. European Journal of Forest Research
674 130: 677–693.
675
- 676 Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GAB, Kent J 2000.
677 Biodiversity hotspots for conservation priorities. Nature 403: 853–
678 858.
679
- 680 Nei M 1978. Estimation of average heterozygosity and genetic distance from a
681 small number of individuals. Genetics 89: 583–590.
682
- 683 Nei M 1987. Molecular evolutionary genetics. New York, Columbia University
684 Press. 512 p.
685
- 686 Nei M, Li WH 1979. Mathematical model for studying genetic variation in
687 terms of restriction endonucleases. Proceedings of the National
688 Academy of Sciences 76: 5269–5273.
689

- 690 Pritchard JK, Stephens M, Donnelly P 2000. Inference of population structure
691 using multilocus genotype data. *Genetics* 155: 945–959.
692
- 693 Rousset F 2008. Genepop'007: a complete reimplementation of the Genepop
694 software for Windows and Linux. *Molecular Ecology Resources* 8:
695 103–106.
696
- 697 San Martín J, Donoso C 1996. Floristic structure and human impact on the
698 Maulino Forest of Chile. In: Armesto JJ, Villagrán C, Arroyo MK eds.
699 *Ecología de los bosques nativos de Chile*. Santiago, Editorial
700 Universitaria. Pp. 153–168.
701
- 702 San Martín J, Sánchez A 2000. The remnant communities of *Gomortega keule*
703 (Gomortegaceae, Magnoliopsida) in central Chile. *Anales del Jardín*
704 *Botánico de Madrid* 57: 317–326.
705
- 706 Serra M, Gajardo R, Cabello A 1986. Ficha técnica de especies amenazadas:
707 *Gomortega keule* (Mol.) Baillon “Keule” (Gomortegaceae), especie en
708 peligro. Santiago, Corporación Nacional Forestal. 18 p.
709
- 710 Shepherd LD, Perrie LR 2011. Microsatellite DNA analyses of a highly
711 disjunct New Zealand tree reveal strong differentiation and imply a
712 formerly more continuous distribution. *Molecular Ecology* 20: 1389–
713 1400.
714
- 715 Tamaki I, Setsuko S, Tomaru N 2008. Genetic variation and differentiation in
716 populations of a threatened tree, *Magnolia stellata*: factors influencing

- 717 the level of within-population genetic variation. *Heredity* 100: 415–
718 423.
- 719
- 720 Tamaki I, Setsuko S, Tomaru N 2016. Genetic diversity and structure of
721 remnant *Magnolia stellata* populations affected by anthropogenic
722 pressures and a conservation strategy for maintaining their current
723 genetic diversity. *Conservation Genetics* 17: 715–725.
- 724
- 725 Torres-Díaz C, Ruiz E, González F, Fuentes G, Cavieres L 2007. Genetic
726 diversity in *Nothofagus alessandrii* (Fagaceae), an endangered
727 endemic tree species of the coastal Maulino forest of central Chile.
728 *Annals of Botany* 100: 75–82.
- 729
- 730 Villagrán C, Armesto JJ 2005. Fitogeografía histórica de la Cordillera de la
731 Costa de Chile. In: Smith C, Armesto JJ, Valdovinos C eds.
732 Biodiversidad y ecología de los bosques de la Cordillera de la Costa
733 de Chile. Santiago, Editorial Universitaria. Pp 99–116.
- 734
- 735 Weir BS, Cockerham CC 1984. Estimating F-statistics for the analysis of
736 population structure. *Evolution* 38: 1358–1370.
- 737
- 738 Yeh FC 2000. Population genetics. In: Young A, Boshier D, Boyle T (eds)
739 *Forest conservation genetics*. CSIRO Publishing: Collingwood, pp
740 21–37.
- 741
- 742 Zaya DN, Howe HF 2009. The anomalous Kentucky coffeetree: megafaunal
743 fruit sinking to extinction?. *Oecologia* 161: 221–226.

744
745
746
747

For Peer Review Only

748

749 **Table 1** Geographical locations and altitudes of the *Gomortega keule*
 750 populations sampled.

Population /group	Locality	Latitude South	Longitude West	Altitude (metres above sea level)	Number of individuals sampled
A	Predio Carmávida, Bosques Arauco S.A. (private forestry farm)	37° 41'	73° 18'	300	75
B	Predio Pino Huacho, Forestal Tierra Chilena Ltda. (private forestry farm)	37° 40'	73° 13'	450	63
C	Reserva Nacional Los Queules (national protected area)	35° 58'	72° 40'	500	11
	Predio Ralbún, Forestal Celco S.A. (private forestry farm)	36° 03'	72° 38'	540	2

751

752

753

754 **Table 2** Number of alleles, allelic richness and private allelic richness per locus and population/group in *Gomortega*755 *keule*.

Locus	Allele size (bp)	Number of alleles				Allelic richness			Private allelic richness		
		A	B	C	Total	A	B	C	A	B	C
<i>Gk-1</i>	224–240	5	4	4	7	2.49	1.92	2.96	0.43	0.26	1.76
<i>Gk-30</i>	182–190	3	4	4	5	2.11	3.35	2.49	0.03	1.07	0.38
<i>Gk-31</i>	208–228	4	3	5	8	1.74	1.52	3.35	0.66	0.06	1.99
<i>Gk-35</i>	224–252	4	2	3	7	2.42	1.99	2.21	0.54	0.11	2.21
<i>Gk-39</i>	133–201	2	2	3	5	1.97	1.83	1.84	0.16	0.03	1.84
<i>CS2</i>	104–119	2	3	3	4	1.98	2.11	2.76	0.01	0.02	1.66
<i>CS8</i>	210–220	3	3	4	5	2.64	2.23	2.83	0.53	0.16	1.40
Mean		3.29	3.00	3.71	5.86	2.19	2.13	2.63	0.34	0.24	1.61

756

757

758

759

760 **Table 3** Observed (H_O) and expected (H_E) heterozygosity for each locus and
 761 population/group of *Gomortega keule*.

Locus	A		B		C		All populations	
	H_O	H_E	H_O	H_E	H_O	H_E	H_O	H_E
<i>Gk-1</i>	0.468	0.451	0.283	0.308	0.786	0.643	0.426	0.500
<i>Gk-30</i>	0.339	0.391	0.735	0.730	0.500	0.585	0.519	0.616
<i>Gk-31</i>	0.032	0.255	0.094	0.159	0.643	0.712	0.124	0.298
<i>Gk-35</i>	0.468	0.512	0.566	0.503	0.071	0.500	0.465	0.614
<i>Gk-39</i>	0.469	0.475	0.358	0.343	0.286	0.262	0.403	0.535
<i>CS2</i>	0.500	0.498	0.491	0.523	0.571	0.648	0.504	0.569
<i>CS8</i>	0.597	0.588	0.453	0.431	0.429	0.558	0.519	0.534
Mean	0.410	0.453	0.426	0.428	0.469	0.558	0.423	0.524

762

763

764

765 **Table 4** Inbreeding coefficient (F_{IS}) for each locus and population/group of
 766 *Gomortega keule*.

Locus	A	B	C	All
<i>Gk-1</i>	-0.036	0.083	-0.233	-0.033
<i>Gk-30</i>	0.134	-0.008	0.150	0.058
<i>Gk-31</i>	0.874	0.410	0.100	0.533
<i>Gk-35</i>	0.086	-0.126	0.862	0.083
<i>Gk-39</i>	0.014	-0.046	-0.095	-0.013
<i>CS2</i>	-0.005	0.063	0.122	0.039
<i>CS8</i>	-0.015	-0.050	0.239	0.002
All	0.095	0.006	0.164	0.069

767 F_{IS} values >0 indicate a deficit and <0 indicate an excess of heterozygotes.

768

769

770

771 **Figure legends**

772

773 **Fig. 1** Geographic distribution of *Gomortega keule*. **A**, Map of Chile with an
774 enlargement of the collection area (arrow). **B**, Populations of *G. keule* (green
775 dots) and the sampled populations (yellow). The distance between populations
776 A and B is 7.5 km.

777

778 **Fig. 2** Fruit of *Gomortega keule*. **A**, Fruit showing the endocarp that encloses
779 the seed. **B**, Representation of extinct megafauna (Gomphoteriidae) eating the
780 fruit of *G. keule*.

781

782 **Fig. 3** Principal coordinate analysis for three populations of *Gomortega keule*
783 using microsatellites for seven loci. Populations A ($n = 75$) and B ($n = 63$) are
784 the southernmost populations of this species, while group C ($n = 14$) comprises
785 individuals found in the northern area of the species' natural distribution. The
786 graph was generated using MVSP 3.13 with the first two axes accounting for
787 30.5% of the variance (Axis 1, 19.4%; Axis 2, 11.1%).

788

789 **Fig. 4** Phenetic dendrogram for populations of *Gomortega keule* generated by
790 cluster analysis with UPGMA using the index of Nei & Li (1979) and the
791 software MVSP 3.13.

792

793 **Fig. 5** Structure analysis of samples of *Gomortega keule* excluding group C
794 showing two clusters (top); and of all samples showing two (middle) and three
795 (bottom) clusters. A Bayesian analysis was conducted using the software
796 Structure 2.3.4 with the admixture model.

797

For Peer Review Only

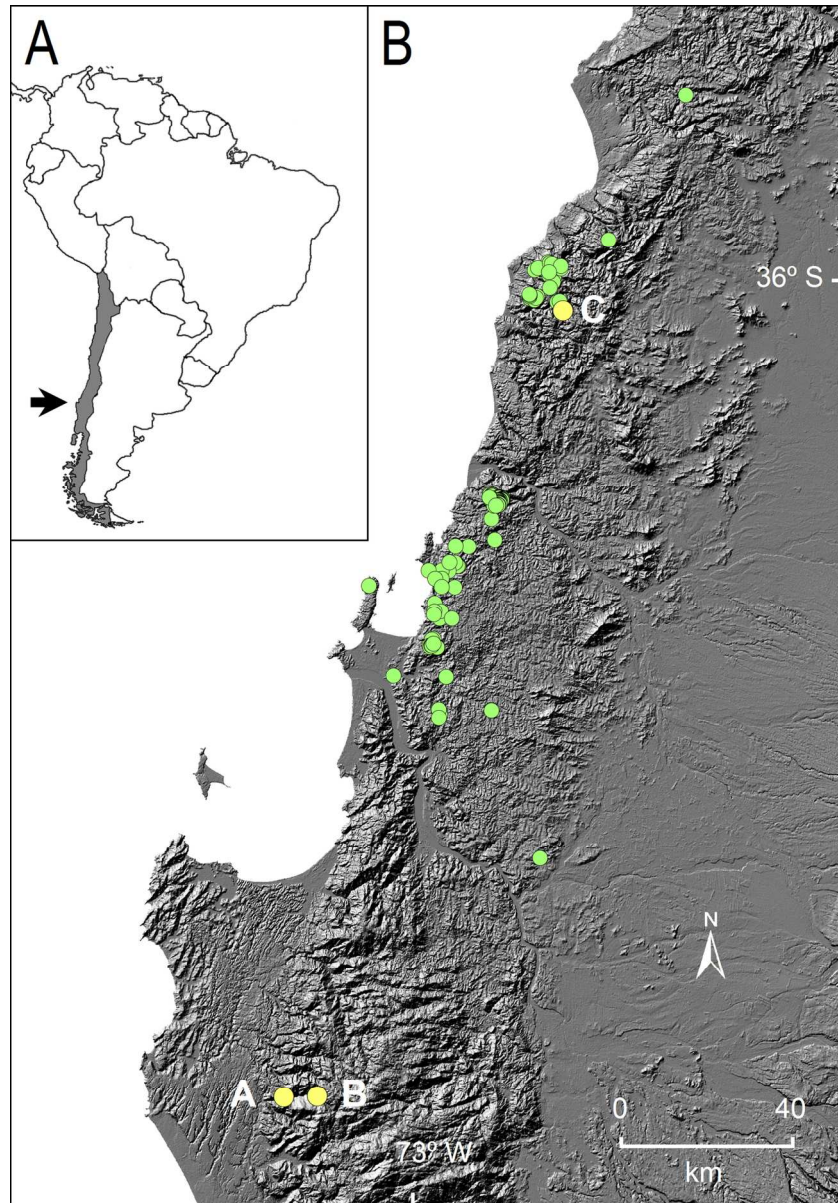


Fig. 1 Geographic distribution of *Gomortega keule*. A, Map of Chile with an enlargement of the collection area (arrow). B, Populations of *G. keule* (green dots) and the sampled populations (yellow). The distance between populations A and B is 7.5 km.

124x179mm (300 x 300 DPI)

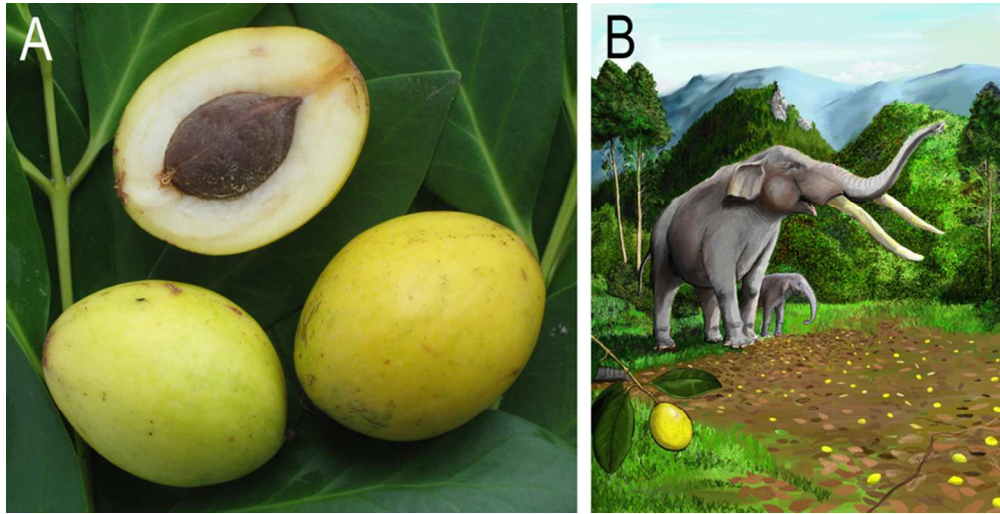


Fig. 2 Fruit of *Gomortega keule*. A, Fruit showing the endocarp that encloses the seed. B, Representation of extinct megafauna (*Gomphoteriidae*) eating the fruit of *G. keule*.

78x39mm (300 x 300 DPI)

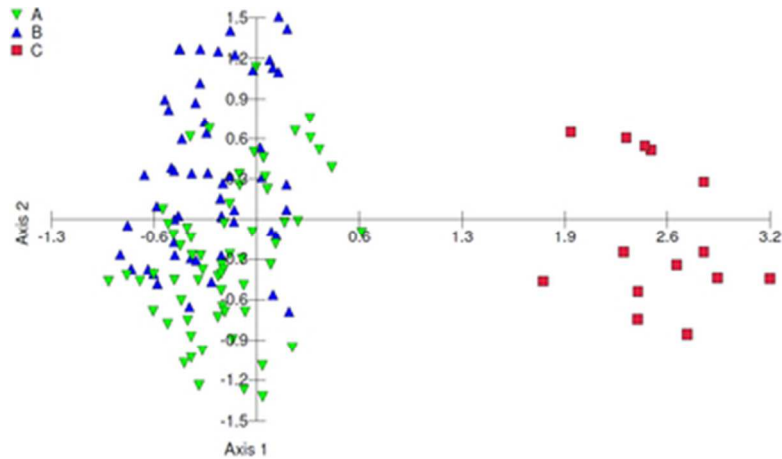


Fig. 3 Principal coordinate analysis for three populations of *Gomortega keule* using microsatellites for seven loci. Populations A ($n = 75$) and B ($n = 63$) are the southernmost populations of this species, while group C ($n = 14$) comprises individuals found in the northern area of the species' natural distribution. The graph was generated using MVSP 3.13 with the first two axes accounting for 30.5% of the variance (Axis 1, 19.4%; Axis 2, 11.1%).

33x19mm (300 x 300 DPI)

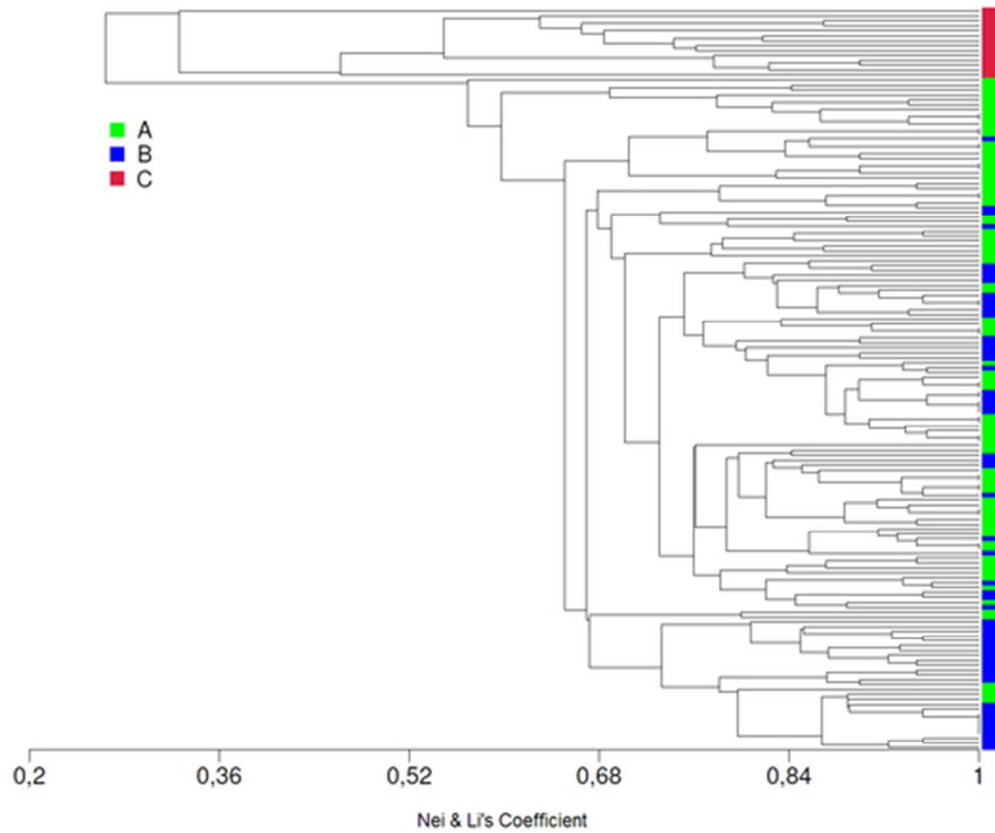


Fig. 4 Phenetic dendrogram for populations of *Gomortega keule* generated by cluster analysis with UPGMA using the index of Nei & Li (1979) and the software MVSP 3.13.

44x37mm (300 x 300 DPI)

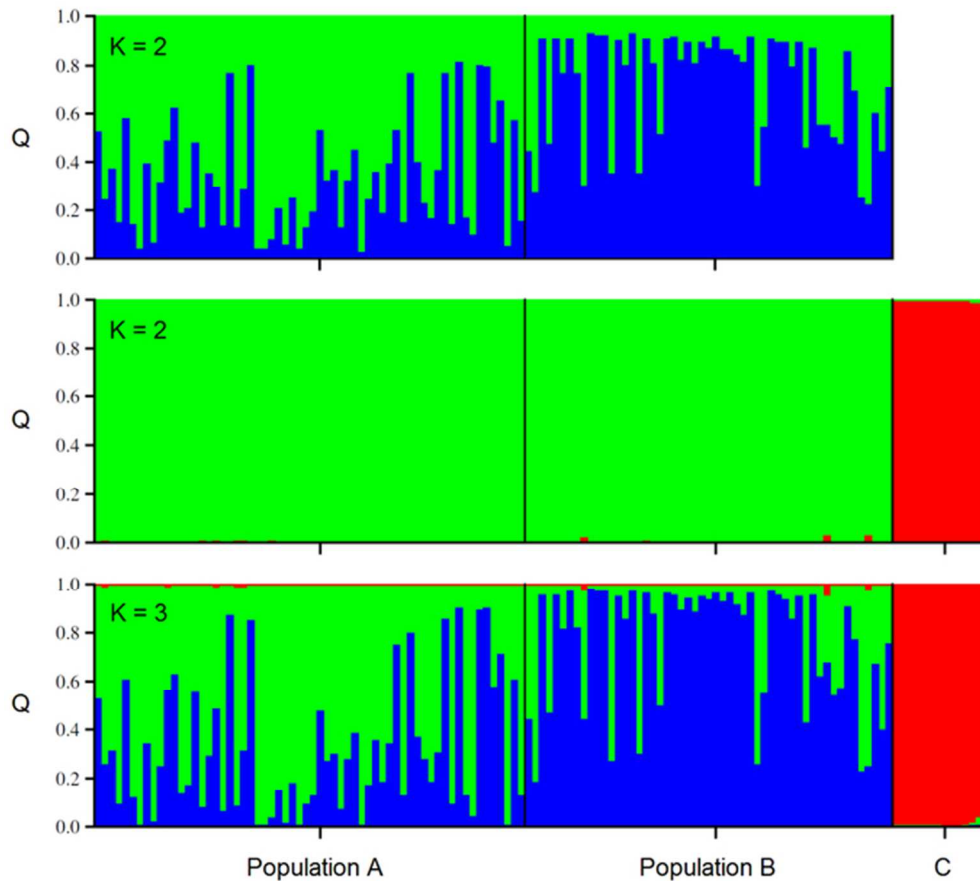


Fig. 5 Structure analysis of samples of *Gomortega keule* excluding group C showing two clusters (top); and of all samples showing two (middle) and three (bottom) clusters. A Bayesian analysis was conducted using the software Structure 2.3.4 with the admixture model.

58x52mm (300 x 300 DPI)