The transition to agricultural cultivation of neo-crops may fail to account for wild genetic diversity patterns: insights from the Cape Floristic Region

Nicholas C Galuszynski [Corresp. 1]

[1] Department of Botany, Nelson Mandela University, Gqebehra, South Africa

Corresponding Author: Nicholas C Galuszynski
Email address: nicholas.galuszynski@gmail.com

Aim: The global increase in the cultivation of native wild plants has raised concerns regarding potential risks associated with translocating genetic lineages beyond their natural range. This study aimed to investigate whether agricultural cultivation of neo-crops a) accounts for the levels of genetic diversity present in wild populations, and whether b) cultivated populations are genetically divergent from wild populations and thus pose a potential threat to wild genetic diversity.

Location: The Cape Floristic Region (CFR), located along the southern Cape of South Africa.

Methods: High Resolution Melt analysis (HRM) coupled with Sanger sequencing was used to screen three non-coding chloroplast DNA loci in *Cyclopia* Vent. (Fabaceae), a CFR endemic neo-crop cultivated for the production of a herbal infusion referred to as Honeybush tea. Wild and cultivated populations for three of three widely cultivated Honeybush species (*C. intermedia, C. longifolia,* and *C. subternata*) were screened. Genetic diversity and differentiation were measured and compared between wild and cultivated groups.

Results: Across all asseccions, a total of 17 haplotypes were detected, four of which were shared between wild and cultivated populations, while the remaining 13 were only detected in wild populations. Genetic diversity and differentiation was significantly higher in wild populations than in cultivated populations.

Conclusions: If no guidelines exist to facilitate the introduction of native wild plant taxa to a cultivated setting, wild genetic diversity patterns are likely to be compromised by cultivated populations. In the case presented here, cultivation represents a genetic bottleneck, failing to account for rare haplotypes, and may have disrupted species boundaries by initiating interspecific hybridization. More empirical work is required to evaluate the extent to which neo-crop cultivation poses a risk to wild genetic resources in the CFR and globally.

The transition to agricultural cultivation of neo-crops may fail to account for wild genetic diversity patterns: insights from the Cape Floristic Region.

Introduction

Commercial trade of wild crop and medicinal plants relies predominantly on material sourced from wild populations. However, consumer demand for 'natural products' and products manufactured from renewable resources has promoted an increase in the domestication and cultivation of wild crop species (Lubbe and Verpoorte 2011), referred to here as neo-crops. If managed effectively, cultivation could facilitate the preservation of these economically important species by safeguarding genetic diversity *ex-situ*. Alternatively, widespread cultivation may reduce incentives to protect natural populations, with natural populations being replaced by cultivated ones. Additionally, the cultivation of commercial strains adjacent to natural populations may increase the risk of exposing wild plants to non-local genetic lineages that may result in hybridization, and thus, genetic erosion (Hammer and Teklu 2008; Laikre et al. 2010).

Gene flow from cultivated to wild populations is particularly common in traditional crop systems (Ellstrand et al. 1999) and can negatively impact wild populations by disrupting local genetic diversity and adaptation (Laikre et al. 2010). Neutral and selective pressures result in the divergence of ecological and genetic traits among geographically separate populations (Hufford and Mazer 2003). When this geographic separation is overcome due to anthropocentric translocation of genetic material, the fate of non-local genes that escape into wild populations is challenging to predict. There have, however, been cases where non-local lineages have invaded native populations, resulting in a loss of local genetic diversity (e.g. *Carex caryophyllea* Latourr (Cyperaceae), Whitlock et al. 2010; and *Phragmites australis* (Cav.) Steud,

25   Chambers et al. 1999). A precautionary approach that limits the distance that genetic material

26   is translocated should, therefore, be adopted during anthropogenic redistribution of genetic

27   material (Byrne and Stone 2011; Galuszynski and Potts 2020a). This is rarely the case, and the

28   evolutionary history of the taxa involved is often not considered. Rather, in many cases, seed is

29   sourced from distant populations or seed lots and have already undergone some form of

30   screening for individuals with commercially favorable traits (Hyten et al. 2006; Schipmann et al.

31   2005; Tembrock et al. 2017; Yuan et al. 2010), that may be detrimental when introduced to

32   natural populations.

33   Cultivated populations are likely to be poor representatives of local genetic diversity. In the

34   case of the stimulant plant qat (*Catha edulis* [Vahl] Forssk. ex Endl., Celastraceae), genotypes

35   sourced from wild populations in Ethiopia were used to establish cultivated populations in

36   Yemen and Kenya, yet both regions support genetically distinct natural qat populations that

37   could have formed the basis for local cultivation (Tembrock et al. 2017). In contrast, cultivated

38   populations of the Chinese skullcap (*Scutellaria baicalensis* Georgi, Lamiaceae) contained a

39   combination of haplotypes sourced from multiple, geographically separate, wild populations,

40   lacking the phylogeographic structuring present in wild populations (Yuan et al. 2010). Despite

41   representing different approaches to collecting genetic material for cultivation, if gene flow

42   were to occur from cultivated to wild populations in either of these cases, the genetic integrity

43   of wild populations would be compromised.

44   The Cape Floristic Region (CFR; Goldblatt 1978), located on the southern coast of South Africa is

45   well known for its species richness, supporting over 9000 species in an area of approximately

46   90 000 km^2 (Goldblatt and Manning 2002), and home to various economically important plant

47   species (Reinten et al. 2011; Scott and Hewett 2008; Turpie et al. 2003). The high floristic

48   diversity of this region has resulted from low extinction rates in a topographically and

49   edaphically heterogeneous landscape, which produces steep ecological gradients and isolate

50   populations over relatively short distances (Barraclough 2006; Cowling et al. 2009; Cowling et

51   al. 2017). Genetic divergence (within and among species) in the CFR is therefore possible over

52   relatively short distances, producing plant populations that exhibit spatially structured genetic

53   diversity (phylogeographic structuring) (Galuszynski & Potts, 2020b; Tolley et al., 2014). The CFR

54 is thus, an ideal system for testing the representation of wild genetic variation in cultivated

55 neo-crop plants.

56 The commercial trade in South African plant products relies predominantly on raw material

57 sourced from wild populations (van Wyk and Prinsloo 2018). However, species used for the

58 production of products with high export value are becoming widely cultivated (Reinten et al.

59 2011; Turpie et al. 2003). This transition to cultivation may pose a threat to the genetic integrity

60 of wild populations of the target species, as the underlying levels and distribution of genetic

61 diversity are not considered during the selection and translocation of commercially important

62 CFR plants (van Wyk 2008). The consequences of this have already led to interspecific

63 hybridization among *Protea* L. species (Proteaceae) (Macqueen and Potts 2018) and possible

64 genetic erosion of wild Rooibos (*Aspalathus linearis* (Burm.f.) R.Dahlgren, Fabaceae) (Malgas et

65 al. 2010). As a result, concerns regarding potential genetic risk associated with a shift to

66 widespread cultivation of Honeybush tea — a herbal infusion produced from members of the

67 CFR endemic genus *Cyclopia* Vent (Fabaceae) — have been raised (Potts 2017).

68 The global increase in demand for Honeybush tea, coupled with local declines in wild

69 populations, prompted interest in cultivation (Joubert et al. 2011). Initial breeding trials

70 included 12 of the 23 *Cyclopia* species and involved the selection of individuals with

71 commercially favorable traits (e.g. bushy and vigorous growth forms), and cross pollination

72 experiments (Joubert et al. 2011). Seed produced from these trials formed the foundation

73 of early Honeybush cultivation, which was promoted in areas that supported natural *Cyclopia*

74 populations (Jacobs 2008; Joubert et al. 2011). This history of domestication trials and

75 translocation is not unique to Honeybush, with a similar approach applied to neo-crops

76 elsewhere in Southern Africa (Akinnifesi et al. 2006), Australia (Ahmed and Johnson 2000), and

77 South Korea (Pemberton and Lee 1996), and may be typical of neo-crop development.

78 This study is the first to describe the levels of genetic diversity among wild and cultivated

79 populations of an endemic crop plant originating from the CFR, focusing specifically on three

80 widely cultivated *Cyclopia* species. High Resolution Melt analysis (HRM, Wittwer et al. 2003)

81 coupled with sequence confirmation is applied to screen variation across two non-coding

82 chloroplast DNA (cpDNA) regions (the *atpI - atpH* intergenic spacer and *ndhA* intron). This study

83 explores the prediction that cultivated *Cyclopia* populations will fail to reflect the spatial

84 distribution and diversity of chloroplast lineages present in the wild. Thus, the current state of

85 honeybush cultivation, and likely neo-crop cultivation in general, may represent a genetic risk

86 to the integrity of wild populations occurring in close proximity to cultivated populations.

Methods and materials

87

Target taxa and sampling

88

89 The species selected for evaluation (*C. intermedia* E.Mey, *C. subternata* Vogel., and *C. longifolia*

90 Vogel.) represent the most widely cultivated Honeybush taxa, cultivated in the Western Cape

91 and Eastern Cape provinces of South Africa (Joubert et al., 2011; McGregor 2017).

92 Consequently, these three species have likely experienced the greatest extent of redistribution

93 outside of their natural range through cultivation (Fig 1; Fig 3 in Joubert et al., 2011) — placing

94 them at high risk of genetic pollution. Additionally, these species represent three distinct

95 distribution and life history patterns (Schutte 1997): a widespread obligate resprouter occurring

96 at altitudes between 500 - 1700 m (*C. intermedia*), a widespread coastal lowland obligate

97 seeder (*C. subternata*) and a critically endangered Eastern Cape endemic, riparian specialist

98 with a mixed post fire response of facultative seeding and sprouting (*C. longifolia*)*.* Since life

99 history traits and range size impact a species' demographic history (Ellegren and Galtier 2016),

100 these three taxa are likely to exhibit different genetic diversity patterns that need to be

101 accounted for during translocation and cultivation.

102 Two of these species (*C. Intermedia* and *C. Subternata)* were subject of recent phylogeographic

103 studies (Galuszynski and Potts 2020a; 2020c) and these previously published data sets were

104 bolstered with additional samples to ensure that 24 individuals were available from each

105 populations for analysis. Samples were collected from four geographically separated wild

106 populations across the natural range of each species. The intention was to a) maximise the

107 genetic variation detected among populations, and b) provide a representative reference of

108   haplotypes to describe the origins of material under cultivation. Cultivated material was

109   sampled from Honeybush farms identified remotely from internet searches rather than relying

110   on existing farmer networks, and were located in different mountain ranges (details of the

111   locations are provided in Fig 1 and Table 1). This approach was employed to avoid potentially

112   redundant sampling of cultivated material originating from seed exchange between farmers

113   located in close proximity to one another, however, the true origin of the commercial seed

114   remains uncertain (see discussion). The cultivators included in the study are situated near to

115   three of the four major Honeybush nurseries reported by Joubert et al. (2011) (Fig 1). Three

116   cultivated populations of each species were sampled. From all populations (wild and

117   cultivated), a total of 24 plants were sampled with a minimum of 5m distance between sampled

118   individuals. The final data set consisted of 504 samples collected across 21 (12 wild, 9

119   cultivated) populations of three commercially important *Cyclopia* species, population locations

120   are mapped in Fig. 1. Fresh leaf material was collected from a healthy growing tip of each

121   individual and placed into silica desiccating medium for a minimum of two weeks prior to DNA

122   extraction. All sampling was approved by the relevant landowners involved and permitting

123   agencies, Cape Nature (Permit number: CN35-28-4367), the Eastern Cape Department of

124   Economic Development, Environmental Affairs and Tourism (Permit numbers: CRO 84/ 16CR,

125   CRO 85/ 16CR), and the Eastern Cape Parks and Tourism Agency (Permit number: RA_0185)

126   **DNA extraction and haplotype detection**

127   The DNA extraction and haplotype detection protocol followed the approach previously

128   described in Galuszynski and Potts (2020c) and a brief overview of the approach is provided

129   here. Whole genomic DNA was extracted using a modified CTAB DNA extraction approach,

130   adapted from Doyle & Doyle (1987). Extracted DNA was quantified using a NanoDrop 2000c

131   Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE19810r Scientific, USA) and

132   diluted to 5 ng/μL for PCR amplification and subsequent HRM analysis.

133   High Resolution Melt analysis involves the gradual heating of PCR products amplified in the

134   presence of a DNA saturating dye. As the double stranded DNA is heated it dissociates at a rate

135   based on the binding strength of the nucleotide sequence under analysis. As such, different

136   nucleotide sequences should produce a distinct melt curve when plotting sample differences in

137   measured fluorescence against change in temperature.

138   Three DNA fragments from two non-coding cpDNA regions (*atpI-atpH* intergenic spacer and

139   *ndhA* intron) were amplified using *Cyclopia* specific primers and subsequently screened for

140   nucleotide variation via HRM curve analysis. Samples were run in duplicates and HRM clustering

141   was conducted on a single population basis following the recommendations of Dang et al.

142   (2012). This was achieved by grouping populations using the 'well group' option in the CFX

143   Manager Software (Bio-Rad Laboratories, Hercules, California, U.S.A.) and running the HRM

144   clustering on these predefined population well groups. All reactions (PCR amplification and

145   subsequent HRM) took place in a 96 well plate CFX Connect (Bio-Rad Laboratories, Hercules,

146   California, U.S.A.). Haplotype melt curve grouping was achieved using the automated clustering

147   algorithm of the High Precision Melt software (Bio-Rad Laboratories, Hercules, California,

148   U.S.A.) (ΔTm = 0.05, curve shape sensitivity = 70%, temperature correction = 20). HRM cluster

149   to haplotype confirmation was achieved by unidirectional sequencing, as described in

150   Galuszynski and Potts (2020c). The chloroplast regions targeted by HRM were sequenced for a

151   subset of individuals per HRM cluster from populations not previously studied by Galuszynski

152   and Potts (2020a; 2020c). A total of 39 individuals for the *atpI-atpH* intergenic spacer and 36

153   individuals for the *ndhA* intron were PCR amplified using the reverse primers and following PCR

154   protocols of (Shaw et al. 2007) and sequenced. The PCR and HRM conditions and details of the

155   primers used in this study, are provided in S1.

156   Sequences were assembled using *CondonCode Aligner [v2.0.1]* (CodonCode Corp, http://www.

157   codoncode.com). Each base-call was assigned a quality score using the PHRED base-calling

158   program (Ewing et al. 1998). Sequences were then automatically aligned using ClustalW

159   (Thompson et al. 1994) and visually inspected. All indels that were difficult to score (due to

160   homopolymer repeats that are prone to alignment errors) were removed. The cpDNA regions

161   under investigation are maternally inherited in tandem and not subject to recombination

162   (Reboud & Zeyl 1994), and were therefore combined for subsequent analysis. A custom R script

163 (provided with a minimum working example online:

164 https://doi.org/10.6084/m9.figshare.12624620.v1) was then used to assign each sample its

165 respective haplotype identity based on HRM clustering.

166 **Haplotype diversity analysis**

167 All analyses were performed in *R [v3.5.1]* (R Core Team 2018). The genealogical relationships

168 among haplotypes were established using a Statistical Parsimony (SP) network (Fig. 2)

169 constructed in *TCS [v1.2.1]* (Clement et al. 2000). As *TCS* treats each base pair in an indel as an

170 evolutionary event, all indels were reduced to a single base pair prior to analysis with default

171 options selected for network construction. Haplotype diversity and differentiation was

172 compared between wild and cultivated individuals grouped by species and origin. Differences in

173 gene diversity (GD) between wild and cultivated populations were tested via Mantel tests

174 (Mantel 1967); using *Hs.test* function implemented in the *adegenet [v2.1.1]* library, (Jombart &

175 Ahmed 2011). Genetic differentiation between wild and cultivated populations were tested via

176 an Analysis of Molecular Variance (AMOVA) (Excoffier et al. 1992) using the *poppr.amova*

177 function with 999 permutations implemented in the *poppr [v2.8.3]* library (Kamvar et al. 2014).

178 Three additional population differentiation measures were calculated: two fixation indices,

179 pairwise Gst (Nei 1973) and G''''st (Hedrick 2005), and a measure of genetic divergence, Jost's D

180 (Jost 2008), using the *pairwise_Gst_Nei, pairwise_Gst_Hendrick* , and *pairwise_D* functions

181 respectively, all from the *mmod [v1.3.3]* library (Winter 2012). Population clustering was

182 inferred from a Neighbor Joining tree constructed using Prevostis pairwise population genetic

183 distance (Prevosti et al. 1975), calculated using the *prevosti.dist* function in the *poppr* library).

184 This distance measure treats alignment gaps as evolutionary events and all gaps were reduced

185 to a single base pair prior to analysis. Support for population clustering was assessed via a

186 bootstrap analysis with 9999 replicates, implemented using the *aboot* function implemented in

187 *poppr* .

188 # Results

Haplotype detection

Of the 504 samples screened for haplotype variation, seven (six wild and one cultivated) failed to PCR amplify despite repeated efforts and the final dataset consisted of 497 samples. High Resolution Melt analysis with haplotype confirmation by sequencing revealed 17 cpDNA haplotypes with 100% specificity for all three loci (i.e. no cases of different haplotypes being grouped into the same HRM cluster were detected). The final concatenated dataset consisted of 794 bp (457 bp from the *atpI-atpH* intergenic spacer and 339 bp from the *ndhA* intron) with an overall GC content of 28.1%. The alignment contained 22 polymorphic sites including nine transitions, ten transversions, and three indels (two of 7 bp and one of 71 bp). Haplotype frequency within populations and nucleotide variation among haplotypes are summarized in Tables 1 and 2, respectively.

Genetic diversity analysis

The SP network (Fig 2) revealed relatively low divergence among haplotypes, with all haplotypes diverging from a central variant. Of the 17 haplotypes detected, only four were present in cultivated populations. These four haplotypes (in addition to a fifth haplotype, detected only in wild *C. subternata* and *C. intermedia* populations from Garcia's Pass, GAR Fig 1) were shared among species. Two of these haplotypes were detected in *C. intermedia* populations (wild and cultivated) and cultivated *C. subternata* populations, but were not present in any wild *C. subternata* populations screened.

Clustering of *Cyclopia* populations, based on pairwise population genetic distance resulted in weak grouping of species. Cultivated populations, however, exhibited little differentiation and generally clustered together based on species (Fig 3). Similarly, wild *C. longifolia* populations exhibited little genetic differentiation and all wild and cultivated populations of this species formed a single group (Fig 3). Wild *C. intermedia* and *C. subternata* populations tended to exhibit higher levels of genetic divergence. All cultivated *C. subternata* populations were clustered with two wild populations originating from the Tsitsikamma mountains (Kareedouw Pass KAR, and Bloukrans Bridge BKB). The remaining wild *C. subternata* populations were more

216   divergent and did not cluster with other *C. subternata* populations. The *C. intermedia* and *C.*

217   *subternata* populations sampled from Garcia's pass (GAR) exhibited no genetic differentiation

218   (both fixed for haplotype F). In general, wild *C. intermedia* populations tended to be genetically

219   distinct, supporting unique haplotypes and formed no clear clusters in the Neighbor Joining

220   population tree (Fig 3). Cultivated *C. intermedia* populations did not exhibit this variability.

221   Rather, all cultivated *C. intermedia* populations form a single cluster with the wild *C. intermedia*

222   population sampled from the Langkloof (LK).

223   In the cases of *C. subternata* and *C. intermedia*, genetic structuring was detected in wild

224   populations, with 60.1% and 83.3% of genetic variation detected among populations for the

225   two species, respectively (associated Fst values significant, $p < 0.005$, Fst = 0.093 and 0.023,

226   respectively). In contrast, no structuring was detected in cultivated populations of these

227   species, with 4.3% and 1.7% of variation structured within cultivated populations for *C.*

228   *intermedia* and *C. subternata,* respectively ($p < 0.005$, Fst = 0.0001 for both species). No genetic

229   structuring was found for wild or cultivated *C. longifolia* populations, as all populations shared

230   the same common haplotype and only two rare haplotypes (N and M, Fig 2) were detected in

231   wild populations. Gene diversity and genetic differentiation followed a similar pattern, with

232   wild *C. intermedia* and *C. subternata* populations having higher mean diversity than cultivated

233   populations ($p < 0.01$), but no differences in gene diversity was detected between wild and

234   cultivated *C. longifolia* populations. Mean genetic differentiation (Gst, G"st and Josts D) was

235   higher in all wild populations than in cultivated populations. All population differentiation and

236   diversity measures are summarized in Table 3.

237   **Discussion**

238   This study set out to explore haplotype diversity patterns in wild and cultivated populations of

239   Honeybush, an endemic neo-crop from the Cape Floristic Region (CFR) of South Africa.

240   Cultivated populations appear to have originated from a small number of founding populations

241   and/or individuals and represent a genetic bottleneck. Thus, cultivated populations tend to lack

242   the genetic diversity and phylogeographic structuring present in wild populations and may

243   represent a genetic threat to wild populations if gene-flow occurs.

244   **Origin of cultivated genetic diversity**

245   Despite initial *Cyclopia* breeding material originating from multiple wild populations (Joubert et

246   al. 2011, Fig 1), screening for individuals with commercially desirable traits has likely removed

247   much of the haplotype richness from commercial breeding stock. The transition to Honeybush

248   cultivation therefore represents a genetic bottleneck — under representing rare haplotypes

249   and homogenizing the cultivated genepool.

250   Based on the NJ clustering of populations (Fig 3), cultivated *C. intermedia* have likely originated

251   from populations located in the Langkloof (LK); *C. subternata* populations from the Tsitsikamma

252   and/or Outeniqua mountains (these wild populations share the same common haplotype found

253   in cultivated populations); while *C. longifolia* originates from its only known wild source, the

254   Van Stadens River system. These findings are consistent with those from recent microsatellite

255   analysis of *C. subternata* (Niemandt et al. 2018). They compared wild populations from the

256   Tsitsikamma and Outeniqua mountains to the Agricultural Resource Council's (ARC) commercial

257   genebank (an important source of commercially traded Honeybush seed), revealing no genetic

258   differentiation between the two wild populations sampled and the genebank accessions. The

259   lack of haplotype diversity detected in cultivated populations is therefore unlikely to be a

260   byproduct of failing to detect variation in the slow evolving chloroplast genome (Schaal et al.

261   1998), as greater divergence was detected among populations originating from the

262   Tsitsikamma or Outeniqua mountains using cpDNA screening via HRM than in the microsatellite

263   based study.

264   The history of the movement of cultivated seed remains speculative and conversations with

265   cultivators during sampling revealed that seed is largely sourced from existing farmer networks

266   in their respective areas (with the initial origin of seed unknown). However, *C. subternata*

267   cultivated in Harkerville (H) was confirmed to have been established from commercial seed and

268   did not differ in dominant haplotypes from other cultivated populations. Furthermore,

269   Harkerville shared a rare haplotype with cultivated material from Uniondale (U) — where

270   putative hybrids were detected (N.C. Galuszynski personal observations, 2018, leaf material

271   from these individuals was collected and is stored at the Nelson Mandela University in Port

272   Elizabeth, South Africa). This rare haplotype was only detected in wild *C. intermedia*

273   populations located in the Swartberg Mountains (SWB) and may be evidence of possible

274   chloroplast capture (Hansen et al. 2003) resulting from interspecific crosses taking place during

275   the initial breeding trials (Joubert et al. 2011) or under field cultivation. The reasoning behind

276   this argument is two fold. First, the chloroplast regions screened exhibit phylogeographic

277   structuring in both species (Galuszynski and Potts 2020a, 2020c) and it is therefore unlikely that

278   a wild *C. subternata* population will support this rare haplotype outside of the Swartberg

279   mountains (where *C. subternata* does not naturally occur). Secondly, *C. intermedia* and *C.*

280   *subternata* have been found to successfully produce hybrid offspring with other members of

281   the genus under experimental conditions, and *C. intermedia* material from the Swartberg was

282   included in the initial Honeybush cultivation trials (Joubert et al. 2011, Fig 1). However,

283   interspecific hybridization should be investigated through additional molecular work targeting

284   the nuclear genome, which is subject to recombination and provides more insight into

285   introgression history.

286   **Potential impacts of cultivated genetic material on wild populations**

287   Higher levels of genetic diversity were detected in wild *Cyclopia* populations than cultivated

288   populations, with multiple cases of near complete haplotype turnover among wild populations

289   (Table 1) — cultivated populations exhibited nearly no differentiation (Table 3). This level of

290   haplotype turnover and genetic structuring of cpDNA in the wild was expected and has been

291   described in detail for *C. intermedia* (Galuszynski and Potts 2020a) and *C. subternata*

292   (Galuszynski and Potts 2020c). This suggests that there may be a tendency for cultivated

293   populations to fail to account for natural phylogeographic patterns in regions where genetic

294   structuring of plant populations occurs, including regions that may play a significant role in the

295   discovery of neo-crops (eg. Mesoamerica, Ornelas et al. 2013; South America, Turchetto-Zolet

296 et al. 2013; Australia, Byrne 2008; the Mediterranean basin, Feliner 2014; and the Cape of

297 South Africa, Galuszynski and Potts 2020b, Tolley et al., 2014).

298 Genetic diversity differed between species (Table 3), suggesting different demographic histories

299 among these closely related taxa (Ellegren and Galtier 2016). Notably, rare and locally endemic

300 taxa are predicted to have low levels of genetic variation due to their restricted distributions

301 and small population sizes (Ellstrand and Elam 1993; Gitzendanner and Soltis 2000; Segarra-

302 Moragues et al. 2012), evident in wild *C. Longifolia* populations. The redistribution of genetic

303 lineages via neo-crop cultivation needs to account for differences in demographic histories.

304 Local endemics, for instance, may be at greater genetic risk from cultivated variants due to

305 naturally low levels of genetic variation being more susceptible to genetic pollution by foreign

306 lineages (Levin et al. 1996; Wolf et al. 2001). Commercial production of narrow endemics may

307 therefore require periodic supplementing of cultivated stands with locally sourced seed

308 material in order to promote the preservation of rare haplotypes.

309 Genetic pollution can only occur if genetic material is able to escape into the wild. Seed

310 dispersal is limited to a few meters in *Cyclopia,* and one would expect the chances of seed

311 escape to be low. However, unmonitored spillover of cultivated seed into adjacent natural

312 habitat does occur (N. C. Galuszynski, pers. obs. 2017) in addition to rare cases of cultivated

313 plants intentionally established in natural vegetation (N. C. Galuszynski pers. obs. 2017; G.

314 McGregor pers. com. 2016; S. Nortje pers. com. 2019). The chloroplast genome is maternally

315 inherited with no recombination in the majority of angiosperms (Mogensen, 1996) and is,

316 therefore, subject to directional selective sweeps. Thus, the introduction of vigorous maternal

317 lineages from commercial seed lots could disrupt local haplotypes diversity in wild crop plants.

318 While the impacts of foreign haplotypes establishing in wild Honeybush populations is

319 unknown, the introduction of foreign *Spartina alterniflora* (Poaceae) haplotypes from the

320 United States to native Chinese populations resulted in a hybrid swarm that disrupted local

321 cpDNA genome frequency and ecological processes due to vigorous growth displacing native

322 plant species (Qiao et al., 2019).

323   Mass flowering of cultivated populations forms a powerful attractant to pollinators, increasing

324   local pollinator density (Holzschuh et al. 2011; Westphal et al. 2003). This facilitates the spread

325   of genetic material from cultivated populations into the wild via pollen flow, particularly in

326   outcrossing species (Ellstrand et al., 1999), such as members of *Cyclopia* (Keon et al., 2020). By

327   altering local allele frequencies through the introduction of large genetically depauperate

328   commercial plantations, a landscape wide genetic bottleneck may result, promoting erosion of

329   wild genetic diversity. The current state of neo-crop cultivation, relying on low genetic diversity

330   breeding stock that is redistributed outside of its natural range (and possibly containing

331   interspecific hybrid taxa as observed in the Honeybush populations in Uniondale, but requiring

332   further research), represents a genetic threat that should be acknowledged and mitigated.

333   Until the genetic risks are better understood, formal guidelines should be developed to

334   facilitate sustainable cultivation of neo-crops. In the case of Honeybush, an ecology-centric

335   approach, as the one outlined by Potts (2017), may be desired due to the conservation value of

336   many *Cyclopia* species; including the commercially important taxa: *C. longifolia* (critically

337   endangered), *C. genistoides* (near threatened), *C. maculata* (near threatened), *C. plicata*

338   (endangered), and *C. sessiliflora* (near threatened). However, it is unlikely that all neo-crops will

339   exhibit the levels of phylogeographic structuring, genetic diversity and, conservation

340   requirements of *Cyclopia,* and more work is required to define the extent to which wild genetic

341   resources are at risk during neo-crop cultivation practices, particularly in species rich

342   landscapes such as the Cape of South Africa.

343   # References

344   Ahmed AK, Johnson KA. 2000. Horticultural development of Australian native edible plants.
345   Australian Journal of Botany 48: 417-426. https://doi:10.1071/BT99042

346   Akinnifesi FK, Kwesiga F, Mhango J, Chilanga T, Mkonda A, Kadu CAC, Kadzere I, Mithofer D,
347   Saka JDK, Sileshi G, Ramadhani T, Dhliwayo P. 2006. Towards the development of Miombo fruit
348   trees as commercial tree crops in Southern Africa. Forests, Trees And Livelihoods 16: 103-121
349   https://doi:10.1080/14728028.2006.9752548

350 Barraclough, T.G., 2006. What can phylogenetics tell us about speciation in the Cape flora?.
351 Diversity & Distributions, *12:* 21–26. https://doi.org/10.1111/j.1366-9516.2006.00208.x

352 Britton, M.N., Hedderson, T.A., Verboom, G.A., 2014. Topography as a driver of cryptic
353 speciation in the high-elevation cape sedge *Tetraria triangularis* (Boeck.) C. B. Clarke
354 (Cyperaceae: Schoeneae). Molecular Phylogenetics and Evolution 77: 96–109.
355 https://doi.org/10.1016/j.ympev.2014.03.024

356 Byrne, M., 2008 Evidence for multiple refugia at different time scales during Pleistocene
357 climatic oscillations in southern Australia inferred from phylogeography. Quaternary Science
358 Reviews 27: 2576-85. https://doi.org/10.1016/j.quascirev.2008.08.032

359 Byrne, M., Stone, L., 2011. The need for 'duty of care' when introducing new crops for
360 sustainable agriculture. Current Opinion in Environmental Sustainability 3: 50–54.
361 https://doi.org/10.1016/j.cosust.2010.11.004

362 Chambers, R.M., Meyerson, L.A., Saltonstall, K., 1999. Expansion of *Phragmites australis* into
363 tidal wetlands of North America. Aquatic Botany 64: 261–273. https://doi.org/10.1016/s0304-
364 3770(99)00055-8

365 Clement, M., Posada, D., Crandall, K.A., 2000. TCS: a computer program to estimate gene
366 genealogies. Molecular Ecology 9: 1657–1659. https://doi.org/10.1046/j.1365-
367 294x.2000.01020.x

368 Cowling, R.M., Bradshaw, P.L., Colville, J.F., Forest, F., 2AD. Levyns law: explaining the evolution
369 of a remarkable longitudinal gradient in Cape plant diversity. Transactions of the Royal Society
370 of South Africa 72:184–201. https://doi.org/10.1080/0035919x.2016.1274277

371 Cowling, R.M., Procheş, Ş., Partridge, T.C., 2009. Explaining the uniqueness of the Cape flora:
372 Incorporating geomorphic evolution as a factor for explaining its diversification. Molecular
373 Phylogenetics and Evolution 51: 64–74. https://doi.org/10.1016/j.ympev.2008.05.034

374 Dang, X.D., Kelleher, C.T., Howard-Williams, E., Meade, C.V., 2012. Rapid identification of
375 chloroplast haplotypes using High Resolution Melting analysis. Molecular Ecology Resources 12:
376 894–908. https://doi.org/10.1111/j.1755-0998.2012.03164.x

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

379 Ellegren, H., Galtier, N., 2016. Determinants of genetic diversity. Nature Reviews Genetics 17:
380 422–433. https://doi.org/10.1038/nrg.2016.58

381 Ellstrand, N.C., Elam, D.R., 1993. Population genetic consequences of small population size:
382 implications for plant conservation. Annual Review of Ecology and Systematics 24: 217–242.
383 https://doi.org/10.1146/annurev.es.24.110193.001245

384 Ellstrand, N.C., Prentice, H.C., Hancock, J.F., 1999. Gene flow and introgression from
385 domesticated plants into their wild relatives. Annual Review of Ecology and Systematics 30:
386 539–563. https://doi.org/10.1146/annurev.ecolsys.30.1.539

387 Ewing, B., Hillier, L., Wendl, M.C., Green, P., 1998. Base-calling of automated sequencer traces
388 using Phred. I. accuracy assessment. Genome Research 8: 175–185.
389 https://doi.org/10.1101/gr.8.3.175

390 Excoffier, L., Smouse, P.E., Quattro, J.M., 1992. Analysis of molecular variance inferred from
391 metric distances among DNA haplotypes: application to human mitochondrial DNA restriction
392 data. Genetics 131: 479–491.

393 Feliner G.N., 2014 Patterns and processes in plant phylogeography in the Mediterranean Basin.
394 A review. Perspectives in Plant Ecology, Evolution and Systematics 16: 265-78.
395 https://doi.org/10.1016/j.ppees.2014.07.002

396 Galuszynski, N.C., Potts, A.J., 2020a. Applied phylogeography of *Cyclopia intermedia* (Fabaceae)
397 highlights the need for 'duty of care' when cultivating honeybush. PeerJ 8e9818.
398 https://doi.org/10.7717/peerj.9818

399 Galuszynski N.C., Potts A.J., 2020b. Do Centres of Endemism provide a spatial context for
400 predicting and preserving plant phylogeographic patterns in the Cape Floristic Region, South
401 Africa? PeerJ 8:e10045 http//doi: 10.7717/peerj.10045

402 Galuszynski, N.C., Potts, A.J., 2020c. Application of High Resolution Melt analysis (HRM) for
403 screening haplotype variation in a non-model plant genus: *Cyclopia* (Honeybush). PeerJ 8e9187.
404 https://doi.org/10.7717/peerj.9187

405 Gitzendanner, M.A., Soltis, P.S., 2000. Patterns of genetic variation in rare and widespread plant
406 congeners. American Journal of Botany 87: 783–792. https://doi.org/10.2307/2656886

407 Goldblatt, P., 1978. An analysis of the flora of Southern Africa: Its characteristics, relationships,
408 and orgins. Annals of the Missouri Botanical Garden: 365-369.
409 https://doi.org/10.2307/2398858

410 Goldblatt, P., Manning, J.C., 2002. Plant diversity of the Cape Region of southern Africa. Annals
411 of the Missouri Botanical Garden: 289-281. https://doi.org/10.2307/3298566

412 Hammer, K., Teklu, Y., 2008. Plant genetic resources: selected issues from genetic erosion to
413 genetic engineering. Journal of Agriculture and Rural Development in the Tropics and
414 Subtropics 109: 15–50

415 Hansen, L.B., Siegismund, H.R., Jørgensen, R.B., 2003. Progressive introgression between
416 *Brassica napus* (oilseed rape) and *B. rapa*. Heredity 91: 276–283.
417 https://doi.org/10.1038/sj.hdy.6800335

418 Hedrick, P.W., 2005. A standardized genetic differentiation measure. Evolution 59: 1633–1638.
419 https://doi.org/10.1111/j.0014-3820.2005.tb01814.x

420 Holzschuh, A., Dormann, C.F., Tscharntke, T., Steffan-Dewenter, I., 2011. Expansion of mass-
421 flowering crops leads to transient pollinator dilution and reduced wild plant pollination.
422 Proceedings of the Royal Society B: Biological Sciences 278: 3444–3451.
423 https://doi.org/10.1098/rspb.2011.0268

424 Hufford, K.M., Mazer, S.J., 2003. Plant ecotypes: genetic differentiation in the age of ecological
425 restoration. Trends in Ecology & Evolution 18: 147–155. https://doi.org/10.1016/s0169-
426 5347(03)00002-8

427 Hyten, D.L., Song, Q., Zhu, Y., Choi, I.Y., Nelson, R.L., Costa, J.M., Specht, J.E., Shoemaker, R.C.,
428 Cregan, P.B., 2006. Impacts of genetic bottlenecks on soybean genome diversity. Proceedings of
429 the National Academy of Sciences 103: 16666–16671.
430 https://doi.org/10.1073/pnas.0604379103

431 Jacobs R. 2008. Honeybush: where can it be cultivated successfully?. Agri-Probe.

432 Jombart, T., Ahmed, I., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP
433 data. Bioinformatics 27: 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

434 Jost, L., 2008. GST and its relatives do not measure differentiation. Molecular Ecology 17: 4015–
435 4026. https://doi.org/10.1111/j.1365-294x.2008.03887.x

436 Joubert, E., Joubert, M.E., Bester, C., de Beer, D., De Lange, J.H., 2011. Honeybush (*Cyclopia*
437 spp.): from local cottage industry to global markets - The catalytic and supporting role of
438 research. South African Journal of Botany 77: 887–907.
439 https://doi.org/10.1016/j.sajb.2011.05.014

440 Kamvar, Z.N., Tabima, J.F., Grünwald, N.J., 2014. Poppr: an R package for genetic analysis of
441 populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2, e281.
442 https://doi.org/10.7717/peerj.281

443 Koen, J., Slabbert, M.M., Booyse, M., Bester, C., 2020. Honeybush (*Cyclopia* spp.) anther-stigma
444 distance and intraspecies cross compatibilities. South African Journal of Plant and Soil: 1-6.

445 Laikre, L., Schwartz, M.K., Waples, R.S., Ryman, N., 2010. Compromising genetic diversity in the
446 wild: unmonitored large-scale release of plants and animals. Trends in Ecology & Evolution 25:
447 520–529. https://doi.org/10.1016/j.tree.2010.06.013

448 Levin, D.A., Francisco-Ortega, J., Jansen, R.K., 1996. Hybridization and the extinction of rare
449 plant species. Conservation Biology 10: 10–16. https://doi.org/10.1046/j.1523-
450 1739.1996.10010010.x

451 Lubbe, A., Verpoorte, R., 2011. Cultivation of medicinal and aromatic plants for specialty
452 industrial materials. Industrial Crops and Products 34: 785–801.
453 https://doi.org/10.1016/j.indcrop.2011.01.019

454 Macqueen, T.P., Potts, A.J., 2018. Re-opening the case of Frankenflora: evidence of
455 hybridisation between local and introduced *Protea* species at Van Stadens Wildflower Reserve.
456 South African Journal of Botany 118: 315–320. https://doi.org/10.1016/j.sajb.2018.03.018

457 Malgas, R.R., Potts, A.J., Oettlé, N.M., Koelle, B., Todd, S.W., Verboom, G.A., Hoffman, M.T.,
458 2010. Distribution, quantitative morphological variation and preliminary molecular analysis of
459 different growth forms of wild rooibos (*Aspalathus linearis*) in the northern Cederberg and on
460 the Bokkeveld Plateau. South African Journal of Botany 76: 72–81.
461 https://doi.org/10.1016/j.sajb.2009.07.004

462 Mantel, N., 1967. Ranking procedures for arbitrarily restricted observation. Biometrics: 65-78.
463 https://doi.org/10.2307/2528282

464 McGregor, G.K., 2017. Industry review: an overview of the Honeybush industry. Department of
465 Environmental Affairs and Development Planning, Cape Town.

466 Mogensen, H.L. 1996. The hows and whys of cytoplasmic inheritance in seed plants. American
467 Journal of Botany 83: 383-404. https://doi.org/10.1002/j.1537-2197.1996.tb12718.x

468 Nei, M., 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National
469 Academy of Sciences 70: 3321–3323. https://doi.org/10.1073/pnas.70.12.3321

470 Niemandt, M., Roodt-Wilding, R., Tobutt, K.R., Bester, C., 2018. Microsatellite marker
471 applications in *Cyclopia* (Fabaceae) species. South African Journal of Botany 116: 52–60.
472 https://doi.org/10.1016/j.sajb.2018.02.408

473 Ornelas, J.F., Sosa, V., Soltis, D.E., Daza, J.M., González, C., Soltis, P.S., Gutiérrez-Rodríguez, C.,
474 de los Monteros, A.E., Castoe, T.A., Bell, C. and Ruiz-Sanchez, E., 2013. Comparative

475   phylogeographic analyses illustrate the complex evolutionary history of threatened cloud
476   forests of northern Mesoamerica. *PloS one*, *8*(2), p.e56283.

477   Pasquet, R.S., Peltier, A., Hufford, M.B., Oudin, E., Saulnier, J., Paul, L., Knudsen, J.T., Herren,
478   H.R., Gepts, P., 2008. Long-distance pollen flow assessment through evaluation of pollinator
479   foraging range suggests transgene escape distances. Proceedings of the National Academy of
480   Sciences 105: 13456–13461. https://doi.org/10.1073/pnas.0806040105

481   Pemberton, R. W., Lee, N. S., 1996. Wild food plants in South Korea; market presence, new
482   crops, and exports to the United States. *Economic Botany 50:* 57-70.
483   https://doi.org/10.1007/BF02862113

484   Potts, A.J., 2017. Genetic risk and the transition to cultivation in Cape endemic crops - The
485   example of honeybush (*Cyclopia*)?. South African Journal of Botany 110: 52–56.
486   https://doi.org/10.1016/j.sajb.2016.09.004

487   Potts, A.J., Hedderson, T.A., Vlok, J.H.J., Cowling, R.M., 2013. Pleistocene range dynamics in the
488   eastern Greater Cape Floristic Region: A case study of the Little Karoo endemic *Berkheya*
489   *cuneata* (Asteraceae). South African Journal of Botany 88:401–413.
490   https://doi.org/10.1016/j.sajb.2013.08.009

491   Prevosti, A., Ocaña, J., Alonso, G., 1975. Distances between populations of *Drosophila*
492   *subobscura*, based on chromosome arrangement frequencies. Theoretical and Applied Genetics
493   45: 231–241. https://doi.org/10.1007/bf00831894

494   Qiao, H., Liu, W., Zhang, Y., Zhang, Y. Y., & Li, Q. Q. (2019). Genetic admixture accelerates
495   invasion via provisioning rapid adaptive evolution. *Molecular ecology*, *28*(17), 4012-4027.
496   https://doi: 10.1111/mec.15192

497   Reboud, X., Zeyl, C., 1994. Organelle inheritance in plants. Heredity 72: 132–140.
498   https://doi.org/10.1038/hdy.1994.19

499   Reinten, E.Y., Coetzee, J.H., van Wyk, B.E., 2011. The potential of South African indigenous
500   plants for the international cut flower trade. South African Journal of Botany 77: 934–946.
501   https://doi.org/10.1016/j.sajb.2011.09.005

502   Schaal, B.A., Hayworth, D.A., Olsen, K.M., Rauscher, J.T., Smith, W.A., 1998. Phylogeographic
503   studies in plants: problems and prospects. Molecular Ecology 7: 465–474.
504   https://doi.org/10.1046/j.1365-294x.1998.00318.x

505   Schipmann, U., Leaman, D.J., Cunningham, A.B., Walter, S., 2005. Impact of cultivation and
506   collection on the conservation of medicinal plants: Global trends and issues. Acta Horticulturae
507   31–44. https://doi.org/10.17660/actahortic.2005.676.3

508 Schutte, A.L., 1997. Systematics of the genus *Cyclopia* Vent. (Fabaceae, *Podalyrieae*). Edinburgh

509 Journal of Botany 54: 125–170. https://doi.org/10.1017/s0960428600004005

510 Scott, G., Hewett, M.L., 2008. Pioneers in ethnopharmacology: The Dutch East India Company

511 (VOC) at the Cape from 1650 to 1800. Journal of Ethnopharmacology 115: 339–360.

512 https://doi.org/10.1016/j.jep.2007.10.020

513 Tembrock, L.R., Simmons, M.P., Richards, C.M., Reeves, P.A., Reilley, A., Curto, M.A., Meimberg,

514 H., Ngugi, G., Demissew, S., Al Khulaidi, A.W., Al-Thobhani, M., Simpson, S., Varisco, D.M., 2017.

515 Phylogeography of the wild and cultivated stimulant plant qat (*Catha edulis*, Celastraceae) in

516 areas of historical cultivation. American Journal of Botany 104: 538–549.

517 https://doi.org/10.3732/ajb.1600437

518 Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of

519 progressive multiple sequence alignment through sequence weighting, position-specific gap

520 penalties and weight matrix choice. Nucleic Acids Research 22: 4673–4680.

521 https://doi.org/10.1093/nar/22.22.4673

522 Tolley, K.A., Bowie, R.C.K., Measey, J.G., Price, B.W., Forest, F., 2014. The shifting landscape of

523 genes since the Pliocene: terrestrial phylogeography in the Greater Cape Floristic Region. In:

524 Allsopp N, Colville JF, Verboom GA, eds. Fynbos: Ecology, Evolution, and Conservation of a

525 Megadiverse Region. Oxford: Oxford University Press. 142-163

526 Turpie, J.K., Heydenrych, B.J., Lamberth, S.J., 2003. Economic value of terrestrial and marine

527 biodiversity in the Cape Floristic Region: implications for defining effective and socially optimal

528 conservation strategies. Biological Conservation 112: 233–251. https://doi.org/10.1016/s0006-

529 3207(02)00398-1

530 Westphal, C., Steffan-Dewenter, I., Tscharntke, T., 2003. Mass flowering crops enhance

531 pollinator densities at a landscape scale. Ecology Letters 6: 961–965.

532 https://doi.org/10.1046/j.1461-0248.2003.00523.x

533 Whitlock, R., Grime, J.P., Burke, T., 2010. Genetic variation in plant morphology contributes to

534 the species-level structure of grassland communities. Ecology 91: 1344–1354.

535 https://doi.org/10.1890/08-2098.1

536 Winter, D.J., 2012. mmod: an R library for the calculation of population differentiation statistics.

537 Molecular Ecology Resources 12: 1158–1160. https://doi.org/10.1111/j.1755-

538 0998.2012.03174.x

539   Wittwer, C.T., Reed, G.H., Gundry, C.N., Vandersteen, J.G., Pryor, R.J., 2003. High-resolution
540   genotyping by amplicon melting analysis using LCGreen.. Clinical chemistry 49: 853–60.
541   https://doi.org/10.1373/49.6.853

542   Wolf, D.E., Takebayashi, N., Rieseberg, L.H., 2001. Predicting the risk of extinction through
543   hybridization. Conservation Biology 15: 1039–1053. https://doi.org/10.1046/j.1523-
544   1739.2001.0150041039.x

545   Yuan, Q.J., Zhang, Z.Y., Hu, J., Guo, L.P., Shao, A.J., Huang, L.Q., 2010. Impacts of recent
546   cultivation on genetic diversity pattern of a medicinal plant, *Scutellaria* (Lamiaceae). BMC
547   Genetics 11: 1-13. https://doi.org/10.1186/1471-2156-11-29

548   R Core Team, 2018. R: a language and environment for statistical computing. R foundation for
549   statistical computing, Vienna, Austria.

550   Segarra-Moragues, J.G., Torres-Díaz, C., Ojeda, F., 2012. Are woody seeder plants more prone
551   than resprouter to population genetic differentiation in Mediterranean-type ecosystems?.
552   Evolutionary Ecology 27: 117–131. https://doi.org/10.1007/s10682-012-9586-y

553   Turchetto-Zolet, A.C., Pinheiro, F., Salgueiro, F., Palma-Silva, C., 2013. Phylogeographical
554   patterns shed light on evolutionary process in South America. Molecular ecology 22: 1193-
555   1213. https://doi.org/10.1111/mec.12164

556   van Wyk, A.S., Prinsloo, G., 2018. Medicinal plant harvesting, sustainability and cultivation in
557   South Africa. Biological Conservation 227: 335–342.
558   https://doi.org/10.1016/j.biocon.2018.09.018

559   van Wyk, B.E., 2008. A broad review of commercially important southern African medicinal
560   plants. Journal of Ethnopharmacology 119: 342–355. https://doi.org/10.1016/j.jep.2008.05.029

Summary of wild and cultivated Honeybush (*Cyclopia*) populations.

Population localities, including population name and abbreviation (used in Fig. 1 and 3), mountain range each population was sampled from, geographic coordinates, number of accessions screened per population (N), number of haplotypes detected per population (H) with haplotypes unique to the population given in parenthesis, and summary of haplotypes found in each population.

Origin	Species	Location		GPS co-ordinates		N	H	Haplotype																
		Population	Mountain	X	Y			A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
Cultivated	*C. intermedia*	George (G)	Outeniqua	-33.93	22.32	24	2 (1)	-	23	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-
		Uniondale (U)	Kammanassie	-33.66	23.14	24	1	-	24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
		Harlem (HAR)	Tsitsikamma	-33.74	23.34	24	2	2	22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	C. longifolia	George (G)	Outeniqua	-33.99	22.36	24	1	24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
		Uniondale (U)	Kammanassie	-33.66	23.14	23	1	23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
		Plettenberg Bay (P)	Tsitsikamma	-33.93	23.48	24	1	24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	C. subternata	Uniondale (U)	Kammanassie	-33.66	23.14	24	3	-	2	21	-	-	-	-	-	-	1	-	-	-	-	-	-	-
		Harkerville (H)	Outeniqua	-34.04	23.23	24	3	1	-	22	-	-	-	-	-	-	1	-	-	-	-	-	-	-
		Plettenberg Bay (P)	Tsitsikamma	-33.93	23.48	24	1	-	-	24	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Wild	*C. intermedia*	Garcia's Pass (GAR)	Langeberg	-33.96	21.22	23	1	-	-	-	-	-	23	-	-	-	-	-	-	-	-	-	-	-
		Swartberg Mountains (SWB)	Swartberg	-33.33	22.04	23	6 (2)	1	3	-	-	-	14	3	1	1	-	-	-	-	-	-	-	-
		Langekloof (LK)	Kouga	-33.78	23.79	24	2 (1)	-	23	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
		Ladyslpper (LS)	Cockscomb	-33.9	25.25	24	3 (2)	-	-	-	-	-	2	-	-	-	-	21	1	-	-	-	-	-
	C. longifolia	Longmore Forest (LMF)	Van Stadens	-33.84	25.09	23	1	23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
		Sand River (SR)	Van Stadens	-33.73	25.09	24	2 (1)	19	-	-	-	-	-	-	-	-	-	-	-	-	5	-	-	-
		Longemore River (LMR)	Van Stadens	-33.81	25.15	23	2 (1)	19	-	-	-	-	-	-	-	-	-	-	-	4	-	-	-	-
		Van Stadens River (VS)	Van Stadens	-33.9	25.21	24	1	24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	C. subternata	Garcia's Pass (GAR)	Langeberg	-33.96	21.22	24	1	-	-	-	-	-	24	-	-	-	-	-	-	-	-	-	-	-
		Outeniqua Pass (OP)	Outeniqua	-33.88	22.4	24	3 (1)	1	-	18	-	-	-	-	-	-	-	-	-	-	-	5	-	-
		Bloukranz Bridge (BKB)	Tsitsikamma	-33.97	23.65	22	2 (1)	-	-	17	-	-	-	-	-	-	-	-	-	-	-	-	5	-
		Kareedou Pass (KAR)	Tsitsikamma	-33.97	24.22	23	2	-	-	19	-	-	-	-	-	-	-	-	-	-	-	-	-	4

Summary of chloroplast DNA nucleotide differences for the three loci screened by HRM.

Haplotype frequency in each population is reported in Table 1.

Position	MLT S1 – MLT S2 *atpI-atpH* intergenic spacer			MLT S3 – MLT S4 *atpI-atpH* intergenic spacer									MLT U1 – MLT U2 *ndhA* intron					
	4-10	47-110	144-150	213	230	294-301	308	314	380	388	421	443	503	597	682	717	731	799
Consensus	1	2a	3	G	C	4	G	T	T	G	T	C	G	T	G	C	G	C
Haplotype																		
A	1	.	.	.	.	4	.	.	.	.	c	.	.	.	.	.	.	.
B	1	.	.	.	.	4	.	.	.	.	.	.	.	.	.	.	.	.
C	1	.	.	.	.	4	.	.	.	a	.	.	.	.	.	.	.	.
D	1	.	.	.	.	4	.	.	g	.	.	.	.	.	.	.	.	.
E	1	2b	.	.	.	4	.	.	.	a	.	.	.	.	.	.	.	.
F	1	.	.	.	.	4	.	.	.	.	.	.	.	.	t	.	t	.
G	1	.	.	.	.	4	.	.	.	.	.	.	a	.	t	.	.	.
H	1	.	.	.	t	4	.	.	.	.	.	.	a	.	t	.	.	.
I	—	.	.	.	.	4	.	.	.	.	.	.	.	.	.	.	.	.
J	1	.	—	.	.	4	a	.	.	.	.	.	.	.	.	.	.	.
K	1	—	.	.	.	4	.	.	.	.	.	.	.	.	.	a	.	a
L	1	2c	.	.	.	4	.	c	.	.	.	.	.	.	.	.	.	.
M	1	.	.	.	.	4	.	.	.	.	c	t	.	.	.	.	.	.
N	1	.	.	.	.	4	.	.	.	.	c	.	.	.	.	a	.	.
O	1	.	.	a	.	-	.	.	.	a	.	.	.	.	.	.	.	.
P	1	2d	.	.	.	4	.	.	.	a	.	.	.	.	.	.	.	.
Q	1	2e	.	.	.	4	.	.	.	a	.	.	.	.	.	.	.	.

1 - tatctaa; 3 - aaaattt; 4 - tatcccc

2a – tacagatgaaaggaagggcttcgtttttttgaatcctatctaaatttacagtaacagggcaaa;

2b – tacagatgaaaggaagggcttcgtttttttgaaaactatctaaatttacagtaacagggcaaa;

2c – tacagatgaaaggaaggggttcgtttttttgaatcctatctaaatttacagtaacagggcaaa;

2d – taaagatgaaaggaagggcttcgtttttttgaatcctatctaaatttacagtaacagggcaaa;

2e – tatagatgaaaggaagggcttcgtttttttgaatcctatctaaatttacagtaacagggcaaa

Genetic diversity, fixation and differentiation measures for wild (W) and cultivated (C) Honeybush (*Cyclopia*) populations.

Significance values are indicated for comparisons of mean genetic diversity, fixation and differentiation between wild and cultivated populations for each species and all species pooled (Total).

Species	Source	N	Hr	GD	Gst (SD)	G''st	Jost's D	Genetic variation (%)
C.intermedia	W	94	6	0.202**	*0.725 (0.178)****	**0.864 (0.111)*****	*0.223 (0.100)****	83.3***
	C	72	2	0.004	0.022 (0.019)	0.043 (0.037)	0.0003 (0.0003)	4.3***
C.subternata	W	93	6	0.069**	**0.332 (0.344)***	**0.493 (0.375)***	**0.049 (0.045)***	60.1***
	C	72	3	0.013	0.026 (0.030)	0.050 (0.057)	0.0005 (0.0006)	1.7***
C.longifolia	W	94	2	0.008	0.050 (0.055)	0.093 (0.102)	0.001 (0.002)	16.3
	C	71	1	0	0 (0)	0 (0)	0 (0)	0
Total	W	281	11	0.169**	**0.645 (0.292)*****	**0.756 (0.298)*****	**0.154 (0.097)*****	85.4***
	C	215	3	0.074	0.262 (0.213)	0.394 (0.245)	0.0334 (0.028)	28.6**

*p < 0.05, ** p < 0.01, ***p < 0.005 (Mantel test, Student t-test, Wilcoxon rank sum test). Genetic variation represents variation between populations determined from AMOVA.

Figure 1

Distribution of *Cyclopia* populations screened for haplotype diversity.

Cyclopia populations indicated by circles, with the colors for wild and cultivated populations for each species indicated. Red stars indicate the locations of Honeybush nurseries, white pentagons indicate the locations of populations initially used for cultivar development for each species respectively. The natural distribution of the three target species is indicated using the same colors used to indicate wild populations and shaded. A) Distribution of *C. intermedia*, in set indicates the study domain in relation to Africa and South Africa; B) distribution of *C. subternata*; and C) distribution of *C. longifolia*, the initial source of *C. longifolia* breeding material is the same location as the Longmore populations (LMF, LMR) and cultivated material was sourced from G, P, and U (in A and B and indicated by a light blue outline). Population naming follows the descriptions in Table 1. Cultivated populations: G = George, U = Uniondale, H = Harkerville, HAR = Harlem, P = Plettenberg Bay. Wild populations: GAR = Garcia's Pass, SWB= Swartberg Mountains, LK = Langkloof, LS = Lady Slipper, OP = Outeniqua Pass, BKB = Bloukrans Bridge, KAR= Kareedouw Pass, LMF = Longmore Forest, LMR = Longmore River, SR = Sand River, VS = Van Stadens River.

⬠ Putative origin of cuiltivated material ★ Commercial Honeybush nursary

A

SWB

GAR

U
HAR
LK

G

LS

C.intermedia ● Wild ○ Cultivated

B

U

GAR

OP

P
BKB

H

KAR

0   25   50   75   100 km

C.subternata ● Wild ○ Cultivated

C

SR

LMF

LMR
VS

C.longifolia
● Wild ○ Cultivated
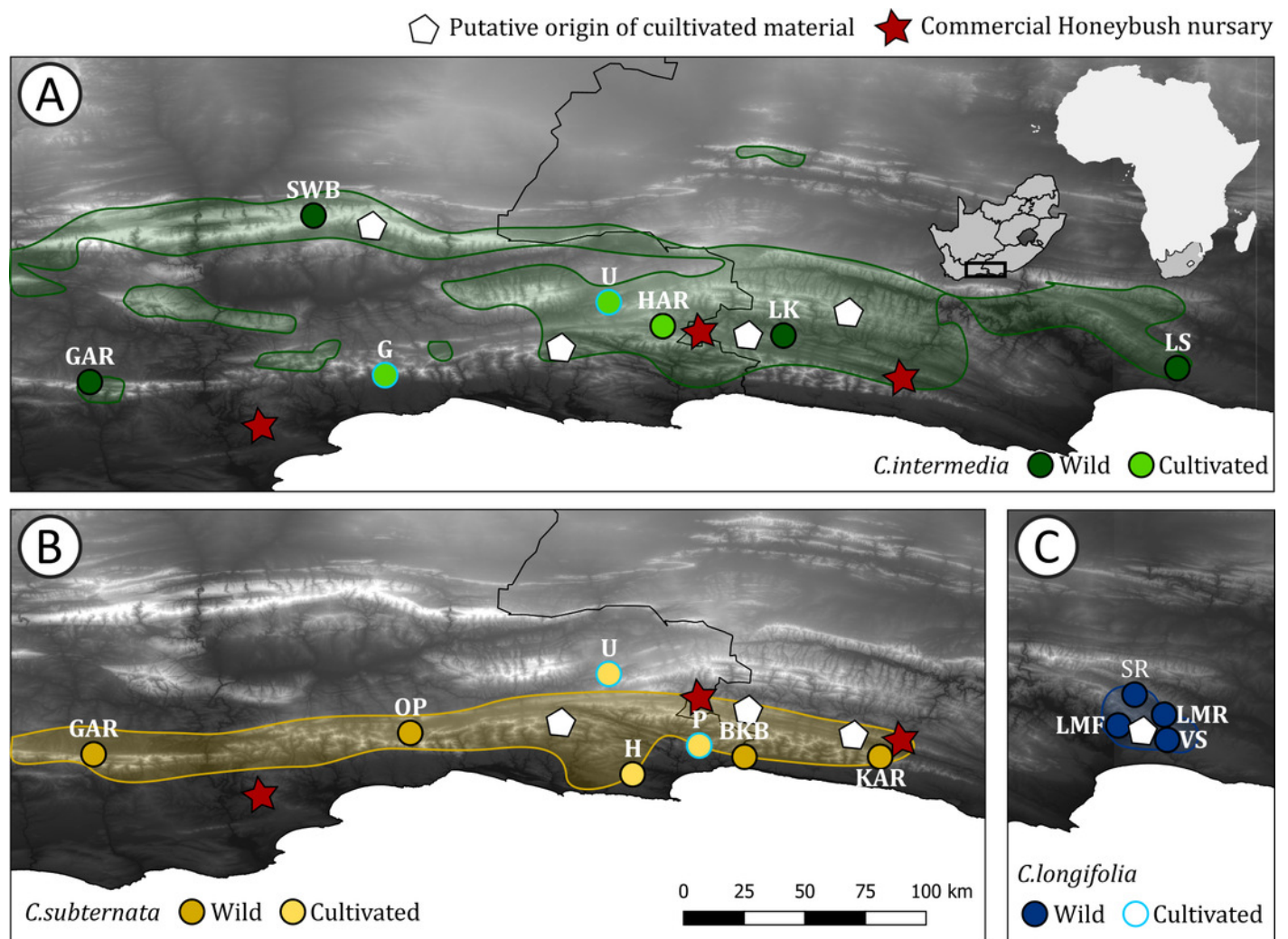
Figure 2

The relationships among haplotypes from the merging chloroplast DNA regions screened via HRM, as inferred from the statistical parsimony algorithm.

Black circles indicate "missing" haplotypes, whist haplotypes connected by a single line differ by a single nucleotide mutation. Areas of circles and numerical labels correspond to the haplotype frequency. Circle colour indicates the species and source of each haplotype (as denoted by the figure key, with the size of each colour segment corresponding to haplotype frequency. Note that haplotypes A and B occur at low frequencies in some species: A occurs once in cultivated and wild *C. subternata* populations, and once in wild and twice in cultivated *C. intermedia* populations, B occurs twice in cultivated *C. subternata* populations. Haplotypes marked with * indicate possible cases of chloroplast capture, as these haplotypes were detected only in cultivated *C. Subternata* populations and wild and cultivated *C. intermedia* populations, but not in any wild *C. subternata* populations. Haplotypes frequencies for each population are given in Table 1 and nucleotide differences among haplotypes are summarized in Table 2.
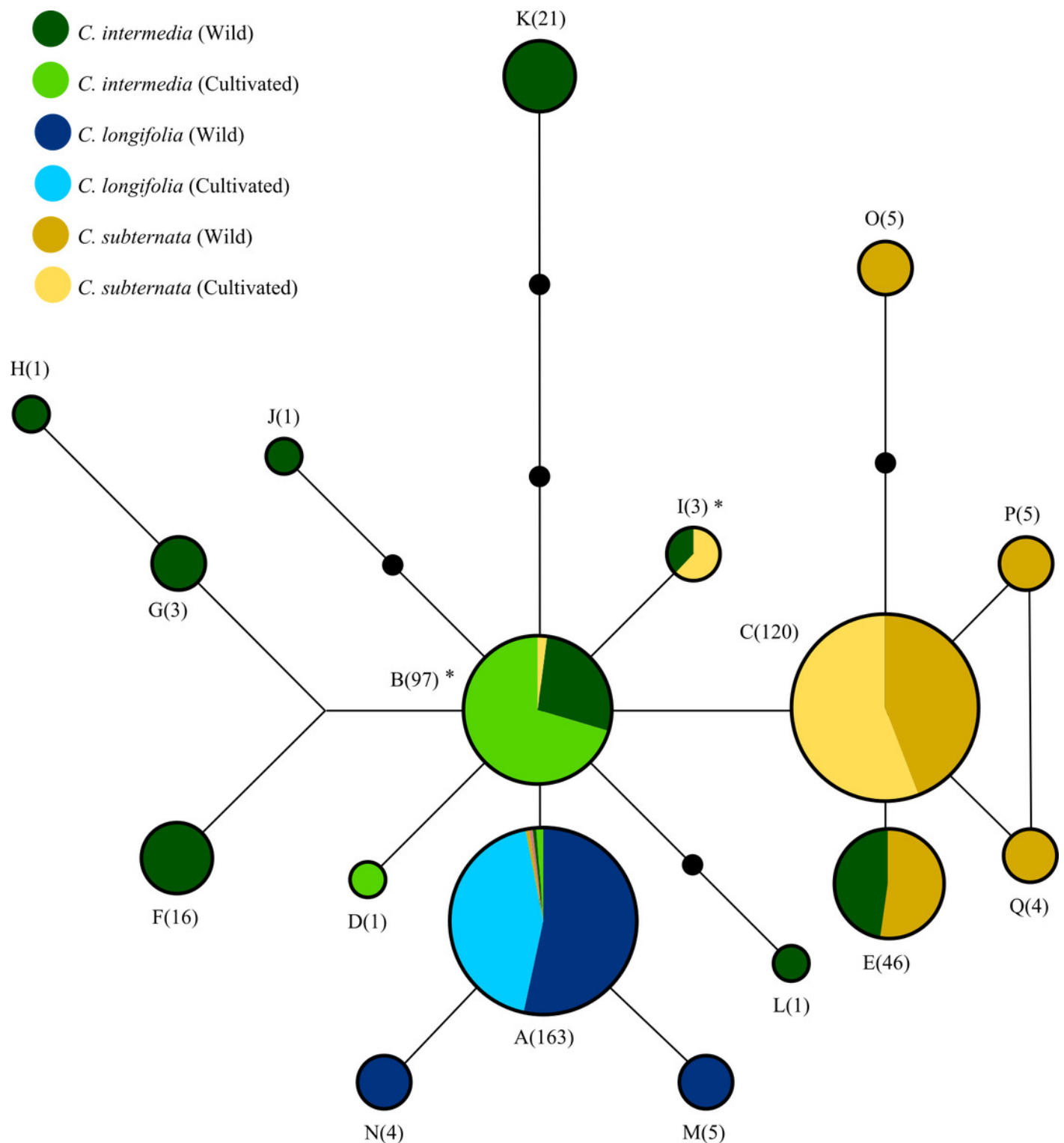
Figure 3

Unrooted Neighbour Joining clustering diagram of *Cyclopia* populations based on pairwise population genetic distance.

Branches with over 50% bootstrap support are labeled. A scale bar of pairwise population genetic distance is provided above the diagram. Branch tips are labeled by species followed by an abbreviated population name following the descriptions Table 1. Open circles indicate wild populations while closed circles indicate cultivated populations. Wild and cultivated populations that group together are indicated by bold type face.

0.15 0.1 0.05 0

● Cultivated populations

○ Wild populations

● *C.longifolia U*

● *C.longifolia G*

● *C.longifolia P*

○ *C.longifolia LMF*

○ *C.longifolia LMR*

○ *C.longifolia VS*

○ *C.longifolia SR*

○ *C.intermedia LK*

● *C.intermedia G*

● *C.intermedia U*

● *C.intermedia HAR*

○ *C.intermedia SWB*

○ *C.subternata KAR*

○ *C.subternata BKB*

● *C.subternata H*

● *C.subternata U*

● *C.subternata P*

○ *C.subternata OP*

○ *C.subternata GAR*

○ *C.intermedia GAR*

○ *C.intermedia LS*

65

60.3

59.6

100

71