The transition to agricultural cultivation of neo-crops may fail to account for wild genetic diversity patterns: insights from the Cape Floristic Region (#56405)

First submission

Guidance from your Editor

Please submit by **1 Feb 2021** for the benefit of the authors (and your $200 publishing discount) .

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Custom checks
Make sure you include the custom checks shown below, in your review.

Raw data check
Review the raw data.

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files

Download and review all files from the materials page.

3 Figure file(s)
3 Table file(s)
1 Raw data file(s)
1 Other file(s)

Custom checks

DNA data checks

- Have you checked the authors data deposition statement?
- Can you access the deposited data?
- Has the data been deposited correctly?
- Is the deposition information noted in the manuscript?

Field study

- Have you checked the authors field study permits?
- Are the field study permits appropriate?

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- All underlying data have been provided; they are robust, statistically sound, & controlled.
- Speculation is welcome, but should be identified as such.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

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 Botany, Nelson Mandela University, Port Elizabeth, South Africa

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

Aim: The global increase in neo-crop cultivation 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 Honeybush (Cyclopia Vent., Fabaceae), a CFR endemic neo-crop. Wild and cultivated populations for three of the most 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 asseccion, 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 guide the introduction of neo-crop taxa to a cultivated setting, wild genetic diversity patterns are likely to be compromised by cultivated populations. In the case of Honeybush, cultivation represents a genetic bottleneck, failing to account for rare haplotypes, and may have disrupted species boundaries. More empirical work is required to evaluate the extent to which genetic diversity is represented in cultivated neo-crop populations.

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

Abstract

Aim: The global increase in neo-crop cultivation 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 Honeybush (*Cyclopia* Vent., Fabaceae), a CFR endemic neo-crop. Wild and cultivated populations for three of the most 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 asseccion, 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 guide the introduction of neo-crop taxa to a cultivated setting, wild genetic diversity patterns are likely to be compromised by cultivated populations. In the case of Honeybush, cultivation represents a genetic bottleneck, failing to account for rare haplotypes, and may have disrupted species boundaries. More empirical work is required to

evaluate the extent to which genetic diversity is represented in cultivated neo-crop populations.

Keywords: Population Genetics, Applied Phylogeography, Wild genetic resources, Honeybush, Genetic Risk

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 domestication and cultivation of novel wild crop species (Lubbe and Verpoorte 2011). If managed effectively, cultivation could facilitate the preservation of these economically important species by safeguarding genetic diversity. Alternatively, widespread cultivation could potentially reduce incentives to protect natural populations and increase the risk of exposing them to non-local genetic lineages (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). The fate of non-local genes that escape into wild populations is challenging to predict, however, there have 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, Chambers et al. 1999).

A precautionary approach that limits the distance that genetic material is translocated should, therefore, be adopted during anthropogenic redistribution of genetic material (Byrne and Stone 2011; Galuszynski and Potts 2020a). This, however, is rarely the case, and the evolutionary history of the taxa involved is often not considered. Rather, in many cases, seed is sourced from distant populations or seed lots and have already undergone some form of screening for

50 individuals with commercially favorable traits (Hyten et al. 2006; Schipmann et al. 2005;
51 Tembrock et al. 2017; Yuan et al. 2010).

52 Cultivated populations are, therefore, likely to be poor representatives of local genetic
53 diversity. In the case of the stimulant plant qat (*Catha edulis* [Vahl] Forssk. ex Endl.,
54 Celastraceae), genotypes sourced from wild populations in Ethiopia were used to establish
55 cultivated populations in Yemen and Kenya, yet both regions support genetically distinct
56 natural qat populations that could have formed the basis of local cultivation (Tembrock et al.
57 2017). In contrast, cultivated populations of the Chinese skullcap (*Scutellaria baicalensis* Georgi,
58 Lamiaceae) represented elements of genetic diversity from all the wild populations screened,
59 lacking the phylogeographic structuring present in wild populations (Yuan et al. 2010). Despite
60 representing different cultivation histories, if gene flow were to occur from cultivated to wild
61 populations in either of these cases, the genetic integrity of wild populations would be
62 compromised.

63 The Cape Floristic Region (CFR; Goldblatt 1978), located on the southern coast of South Africa is
64 well known for its species richness, supporting over 9000 species in an area of around 90 000
65 km^2 (Goldblatt and Manning 2002), and home to various economically important plant species
66 (Reinten et al. 2011; Scott and Hewett 2008; Turpie et al. 2003). The high floristic diversity of
67 this region has resulted from low extinction rates in a topographically and edaphically
68 heterogeneous landscape, which produces steep ecological gradients and isolate populations
69 over relatively short distances (Barraclough 2006; Cowling et al. 2009; Cowling et al. 2017).
70 Genetic divergence (within and among species) is therefore able to occur over relatively short
71 distances in the CFR, producing plant populations that exhibit spatially structured genetic
72 diversity (Galuszynski & Potts, 2020b; Tolley et al., 2014). The CFR is thus, an ideal system for
73 testing the representation of wild genetic variation in cultivated neo-crop populations.

74 The commercial trade in South African plant products relies predominantly on raw material
75 sourced from wild populations (van Wyk and Prinsloo 2018). However, species used for the
76 production of products with high export value are becoming widely cultivated (Reinten et al.
77 2011; Turpie et al. 2003). This transition to cultivation may pose a threat to the genetic integrity

of wild populations of the target species, as the underlying levels and distribution of genetic diversity are not considered during the selection and translocation of commercially important CFR plants (van Wyk 2008). The consequences of this have included interspecific hybridization among *Protea* L. species (Proteaceae) (Macqueen and Potts 2018) and possible genetic erosion of *Aspalathus linearis* (Burm.f.) R.Dahlgren (Fabaceae) (Malgas et al. 2010). As a result, concerns regarding potential genetic risk associated with a shift to widespread cultivation of Honeybush tea — a herbal infusion produced from members of the CFR endemic genus *Cyclopia* Vent (Fabaceae) — have been raised (Potts 2017).

This study is the first to describe the levels of genetic diversity among wild and cultivated populations of an endemic crop plant originating from the CFR, focusing specifically on three widely cultivated *Cyclopia* species. High Resolution Melt analysis (HRM, Wittwer et al. 2003) coupled with sequence confirmation (Sanger et al. 1977) is applied to screen variation across two non-coding chloroplast DNA (cpDNA) regions (the *atpI* - *atpH* intergenic spacer and *ndhA* intron). This study explores the prediction that neo-crop populations will have experienced a genetic bottleneck, due to a small number of founding individuals used to produce commercial seed, and thus fail to represent the haplotype diversity and structuring of wild populations.

Methods and materials

Target taxa and sampling

The species selected for evaluation (*C. intermedia* E.Mey, *C. subternata* Vogel., and *C. longifolia* Vogel.) represent the most widely cultivated Honeybush taxa, cultivated in the Western Cape and Eastern Cape provinces of South Africa (McGregor 2017). Consequently, these three species have likely experienced the greatest extent of redistribution through cultivation — placing them at high risk of genetic pollution. Additionally, these species represent three distinct distribution and life history patterns (Schutte 1997): a widespread resprouter occurring at altitudes between 500 - 1700 m (*C. intermedia*), a widespread coastal lowland obligate seeder (*C. subternata*) and a critically endangered Eastern Cape endemic, riparian specialist

with a mixed post fire response of facultative-resprouting and seeding (*C. longifolia*). Since life history traits and range size impact a species demographic history (Ellegren and Galtier 2016), these three taxa are likely to exhibit different genetic diversity patterns that need to be accounted for during translocation and cultivation.

Samples were collected from four geographically separated wild populations across the natural range of each species. The intention of this was to a) maximise the amount of inter-population genetic variation detected, and b) provide a representative reference of haplotypes to describe the origins of haplotypes found in cultivated populations. Cultivated material was sampled from cultivated Honeybush populations identified from internet searches rather than relying on existing farmer networks. This approach was employed to avoid potentially redundant sampling of cultivated material originating from the same seed lot. The cultivators included in the study are situated near to three of the four major Honeybush nurseries reported by Joubert et al. (2011) (Fig 1). Three cultivated populations of each species were sampled. From each population (wild and cultivated), a total of 24 plants were sampled with a minimum of 5m distance between sampled individuals. Population locations are mapped in Fig. 1. Fresh leaf material was collected from a healthy growing tip of each individual and placed into silica desiccating medium for a minimum of two weeks prior to DNA extraction. All sampling was approved by the relevant permitting agencies, Cape Nature (Permit number: CN35-28-4367), the Eastern Cape Department of Economic Development, Environmental Affairs and Tourism (Permit numbers: CRO 84/ 16CR, CRO 85/ 16CR), and the Eastern Cape Parks and Tourism Agency (Permit number: RA_0185).

DNA extraction and haplotype detection

Whole genomic DNA was extracted using a modified CTAB DNA extraction approach, adapted from the methods outlined by Doyle & Doyle (1987). Extracted DNA was quantified using a NanoDrop 2000c Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE19810r Scientific, USA) and diluted to 5 ng/µL for PCR amplification and subsequent HRM analysis.

High Resolution Melt analysis involves the gradual heating of PCR products amplified in the presence of a DNA saturating dye. As the double stranded DNA is heated it dissociates at a rate based on the binding strength of the nucleotide sequence under analysis. As such, different nucleotide sequences should produce a distinct melt curve when plotting sample differences in measured fluorescence against change in temperature. The application of HRM for haplotype detection in *Cyclopia* has been demonstrated elsewhere (Galuszynski & Potts 2020a, 2020c), and only a brief overview of the approach is provided here.

Three DNA fragments from two non-coding cpDNA regions (*atpI-atpH* intergenic spacer and *ndhA* intron) were amplified using *Cyclopia* specific primers and subsequently screened for nucleotide variation via HRM curve analysis. Samples were run in duplicates of two and HRM clustering was conducted on a single population basis following the recommendations of (Dang et al. 2012). This was achieved by grouping populations using the well group option in the CFX Manager Software (Bio-Rad Laboratories, Hercules, California, U.S.A.) and running the HRM clustering on these predefined population well groups. All reactions (PCR amplification and subsequent HRM) took place in a 96 well plate CFX Connect (Bio-Rad Laboratories, Hercules, California, U.S.A.). Haplotype melt curve grouping was achieved using the automated clustering algorithm of the High Precision Melt software (Bio-Rad Laboratories, Hercules, California, U.S.A.) (ΔTm = 0.05, curve shape sensitivity = 70%, temperature correction = 20). HRM cluster to haplotype confirmation was achieved by unidirectional sequencing (Sanger et al. 1977), as described in Galuszynski and Potts (2020c). The chloroplast regions targeted by HRM were sequenced for a subset of individuals per HRM cluster in each population (110 individuals for the *atpI-atpH* intergenic spacer and 111 individuals for the *ndhA* intron) using the reverse primers and following PCR protocols of (Shaw et al. 2007). The PCR and HRM conditions and details of the primers used in this study, are provided in S1.

Sequences were assembled using *CondonCode Aligner [v2.0.1]* (CodonCode Corp, http://www.codoncode.com). Each base-call was assigned a quality score using the PHRED base-calling program (Ewing et al. 1998). Sequences were then automatically aligned using ClustalW (Thompson et al. 1994) and visually inspected. All indels that were difficult to score (due to

158 homopolymer repeats that are prone to alignment errors) were removed. The cpDNA regions
159 under investigation are maternally inherited in tandem and not subject to recombination
160 (Reboud & Zeyl 1994), and were therefore combined for subsequent analysis. A custom R script
161 (provided with a minimum working example online: Galuszynski 2020) was then used to assign
162 each sample its respective haplotype identity based on HRM clustering.

163 **Haplotype diversity analysis**

164 The genealogical relationships among haplotypes were established using a Statistical Parsimony
165 (SP) network (Fig. 2) constructed in *TCS [v1.2.1]* (Clement et al. 2000). As *TCS* treats each base
166 pair in an indel as an evolutionary event, all indels were reduced to a single base pair prior to
167 analysis with default options selected for network construction. Haplotype diversity and
168 differentiation was compared between wild and cultivated individuals grouped by species and
169 origin. Differences in gene diversity (GD) between wild and cultivated populations were tested
170 via Mantel tests (Mantel 1967); using *Hs.test* function implemented in the *adegenet [v2.1.1]*
171 library, (Jombart & Ahmed 2011). Genetic differentiation between wild and cultivated
172 populations were tested via an Analysis of Molecular Variance (AMOVA) (Excoffier et al. 1992)
173 using the *poppr.amova* function with 999 permutations implemented in the *poppr [v2.8.3]*
174 library (Kamvar et al. 2014). Three additional population differentiation measures were
175 calculated: two fixation indices, pairwise Gst (Nei 1973) and G""st (Hedrick 2005), and a
176 measure of genetic divergence, Jost's D (Jost 2008), using the *pairwise_Gst_Nei,*
177 *pairwise_Gst_Hendrick* , and *pairwise_D* functions respectively, all from the *mmod [v1.3.3]*
178 library (Winter 2012). Population clustering was inferred from a Neighbour Joining tree
179 constructed using Prevostis pairwise population genetic distance (Prevosti et al. 1975),
180 calculated using the *prevosti.dist* function in the *poppr* library). This distance measure treats
181 alignment gaps as evolutionary events and all gaps were reduced to a single base pair prior to
182 analysis. Support for population clustering was assessed via a bootstrap analysis with 9999
183 replicates, implemented using the *aboot* function implemented in *poppr* . All analyses were
184 performed in *R [v3.5.1]* (R Core Team 2018).

Results

Haplotype detection

A total of 504 samples were collected across 21 (12 wild, 9 cultivated) populations of three commercially important *Cyclopia* species; however, seven samples (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, only four were found in cultivated populations. These four haplotypes (in addition to a fifth haplotype, found only in wild *C. subternata* and *C. intermedia* populations from the Langeberg) were shared among species. Two of these haplotypes were found in *C. intermedia* populations (wild and cultivated) and cultivated *C. subternata* populations, but not in wild *C. subternata* populations.

Clustering of *Cyclopia* populations, based on pairwise population genetic distance resulted in weak clustering of species. Cultivated populations generally exhibited little differentiation and clustered together based on species. 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, however, 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 divergent and did not cluster with other *C. subternata* populations. The *C. intermedia* and *C. subternata* populations sampled from Garcias pass (GAR) exhibited no genetic differentiation (both fixed for haplotype F). In general, wild *C. intermedia* populations tended to be genetically distinct, forming no clear clusters. Cultivated *C. intermedia* populations did not exhibit this variability. Rather, all cultivated *C. intermedia* populations form a single clustered with the wild *C. intermedia* population sampled from the Langkloof (LK).

In the cases of *C. subternata* and *C. intermedia*, genetic structuring was detected in wild populations, with 60.1% and 83.3% of genetic variation detected among populations for the two species respectively (associated Fst values significant, $p < 0.005$, Fst = 0.093 and 0.023, respectively). In contrast, no structuring was detected in cultivated populations of these species, with 4.3% and 1.7% of variation structured within cultivated populations for *C. intermedia* and *C. subternata,* respectively ($p < 0.005$, Fst = 0.0001 for both species). No genetic structuring was found for wild or cultivated *C. longifolia* populations, as all populations shared the same common haplotype and only two rare haplotypes (N and M, Fig 2) were detected in wild populations. Gene diversity and genetic differentiation followed a similar pattern, with wild *C. intermedia* and *C. subternata* populations having higher mean diversity than cultivated populations ($p < 0.01$), but no differences in gene diversity was detected between wild and cultivated *C. longifolia* populations. Mean genetic differentiation (Gst, Gst and Josts D) was higher in all wild populations than in cultivated populations. All population differentiation and diversity measures are summarized in Table 3.

Discussion

This study set out to explore haplotype diversity patterns in wild and cultivated populations of Honeybush, an endemic neo-crop from the CFR of South Africa. Cultivated populations appear to have originated from a small number of founding populations and/or individuals and

represent a genetic bottleneck. Thus, cultivated populations tend to lack the genetic diversity and phylogeographic structuring present in wild populations.

Origin of cultivated genetic diversity

Despite initial *Cyclopia* breeding material originating from multiple wild populations (Joubert et al. 2011, Fig 1), screening for individuals with commercially desirable traits has likely removed much of the haplotype richness from commercial breeding stock. The transition to Honeybush cultivation therefore represents a genetic bottleneck — under representing rare haplotypes and homogenizing the cultivated genepool.

Based on the NJ clustering of populations (Fig 3), cultivated *C. intermedia* have likely originated from populations located in the Langkloof (LK); *C. subternata* populations from the Tsitsikamma and/or Outeniqua mountains (these wild populations share the same common haplotype found in cultivated populations); while *C. longifolia* originates from its only known wild source, the Van Stadens River system. These findings are consistent with those from recent microsatellite analysis of *C. subternata* (Niemandt et al. 2018). They compared wild populations from the Tsitsikamma and Outeniqua mountains to the Agricultural Resource Council's (ARC) commercial genebank (an important source of commercially traded Honeybush seed), revealing no genetic differentiation between the two wild populations sampled and the genebank accessions. The lack of haplotype diversity detected in cultivated populations is therefore unlikely to be a byproduct of failing to detect variation in the slow evolving chloroplast genome (Schaal et al. 1998), as greater divergence was detected among populations originating from the Tsitsikamma or Outeniqua mountains using cpDNA screening via HRM than in the microsatellite based study.

The history of the movement of cultivated seed remains speculative and conversations with cultivators during sampling revealed that seed is largely sourced from existing farmer networks in their respective areas (with the initial origin of seed unknown). However, *C. subternata* cultivated in Harkerville (H) was confirmed to have been established from commercial seed and did not differ in dominant haplotypes from other cultivated populations. Furthermore,

Harkerville shared a rare haplotype with cultivated material from Uniondale (U) — where putative hybrids were detected (N.C. Galuszynski personal observations, 2018, leaf material from these individuals was collected and is stored at the Nelson Mandela University in Port Elizabeth, South Africa). This rare haplotype was only detected in wild *C. intermedia* populations located in the Swartberg Mountains and may be evidence of possible chloroplast capture (Hansen et al. 2003) resulting from interspecific crosses taking place during the initial breeding trials (Joubert et al. 2011) or under field cultivation. The reasoning behind this argument two fold. First, the chloroplast regions screened exhibit phylogeographic structuring in both species (Galuszynski and Potts 2020a, 2020c) and it is therefore unlikely that a wild *C. subternata* population will support this rare haplotype outside of the Swartberg mountains (where *C. subternata* does not naturally occur). Secondly, *C. intermedia* and *C. subternata* have been found to successfully produce hybrid offspring with other members of the genus under experimental conditions, and *C. intermedia* material from the Swartberg was included in the initial Honeybush cultivation trials (Joubert et al. 2011, Fig 1). However, interspecific hybridization should be investigated through additional molecular work targeting the nuclear genome, which is subject to recombination and provides more insight into introgression history.

Potential impacts of cultivated genetic material on wild populations

Higher levels of genetic diversity were detected in wild *Cylopia* populations than cultivated populations, with multiple cases of near complete haplotype turnover among wild populations (Table 1) — cultivated populations exhibited nearly no differentiation (Table 3). This level of haplotype turnover and genetic structuring of cpDNA in the wild was expected and has been described in detail for *C. intermedia* (Galuszynski and Potts 2020a) and *C. subternata* (Galuszynski and Potts 2020c). This suggests that there may be a tendency for cultivated populations to fail to account for natural phylogeographic patterns in regions where genetic strcturing of plant populations occurs, including regions that may play a significant role in the discovery of neo-crops (eg. Mesoamerica, Ornelas et al. 2013; South America, Turchetto-Zolet

et al. 2013; Australia, Byrne 2008; the Mediterranean basin Feliner 2014; and the Cape of South Africa, Galuszynski and Potts 2020b, Tolley et al., 2014).

Genetic diversity differed between taxa (Table 3), suggesting different demographic histories among these closely related taxa (Ellegren and Galtier 2016). Notably, rare and locally endemic taxa are predicted to have low levels of genetic variation due to their restricted distributions and small population sizes (Ellstrand and Elam 1993; Gitzendanner and Soltis 2000; Segarra-Moragues et al. 2012), evident in *C. longifolia*. The redistribution of genetic lineages via neo-crop cultivation needs to account for differences in demographic histories. Local endemics, for instance, may be at greater genetic risk from cultivated variants due to naturally low levels of genetic variation being more susceptible to genetic pollution by foreign lineages (Levin et al. 1996; Wolf et al. 2001). Commercial production of narrow endemics may therefore require periodic supplementing of cultivated stands with locally sourced seed material in order to promote the preservation of rare haplotypes.

Genetic pollution can only occur if genetic material is able to escape into the wild. Seed dispersal is limited to a few meters in *Cyclopia,* and one would expect the chances of seed escape to be low. However, unmonitored spillover of cultivated seed into adjacent natural habitat does occur (N. C. Galuszynski, pers. obs. 2017) in addition to rare cases of cultivated plants intentionally established in natural vegetation (N. C. Galuszynski pers. obs. 2017; G. McGregor pers. com. 2016; S. Nortje pers. com. 2019).

Mass flowering of cultivated populations forms a powerful attractant to pollinators, increasing local pollinator density (Holzschuh et al. 2011; Westphal et al. 2003). This facilitates the spread of genetic material from cultivated populations into the wild via pollen flow, particularly in outcrossing species (Ellstrand et al., 1999) such as *Cyclopia* (Keon et al., 2020). By altering local allele frequencies through the introduction of large genetically depauperate commercial plantations, a landscape wide genetic bottleneck may result, promoting erosion of wild genetic diversity. The current state of neo-crop cultivation, relying on low genetic diversity breeding stock that is redistributed outside of its natural range (and possibly containing interspecific

hybrid taxa as observed in the Honeybush populations in Uniondale, but requiring further research), represents a genetic threat that should be acknowledged and mitigated.

Until the genetic risks are better understood, formal guidelines must be developed to facilitate sustainable cultivation of neo crops. In the case of Honeybush, an ecology-centric approach, as the one outlined by Potts (2017), may be desired due to the conservation value of many *Cyclopia* species. Including the commercially important taxa: *C. longifolia* (critically endangered), *C. genistoides* (near threatened), *C. maculata* (near threatened), *C. plicata* (endangered), and *C. sessiliflora* (near threatened). However, this is unlikely to be the case for all neo-crops and more work is required to identify the extent to which wild genetic resources are at risk during neo-crop cultivation practices.

References

Barraclough, T.G., 2006. What can phylogenetics tell us about speciation in the Cape flora?. Diversity & Distributions, *12*(1), 21–26. https://doi.org/10.1111/j.1366-9516.2006.00208.x

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

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

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

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

Clement, M., Posada, D., Crandall, K.A., 2000. TCS: a computer program to estimate gene genealogies. Molecular Ecology 91657–1659. https://doi.org/10.1046/j.1365-294x.2000.01020.x

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

Cowling, R.M., Procheş, Ş., Partridge, T.C., 2009. Explaining the uniqueness of the Cape flora: Incorporating geomorphic evolution as a factor for explaining its diversification. Molecular Phylogenetics and Evolution 5164–74. https://doi.org/10.1016/j.ympev.2008.05.034

Dang, X.D., Kelleher, C.T., Howard-Williams, E., Meade, C.V., 2012. Rapid identification of chloroplast haplotypes using High Resolution Melting analysis. Molecular Ecology Resources 12894–908. https://doi.org/10.1111/j.1755-0998.2012.03164.x

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

Ellegren, H., Galtier, N., 2016. Determinants of genetic diversity. Nature Reviews Genetics 17422–433. https://doi.org/10.1038/nrg.2016.58

Ellstrand, N.C., Elam, D.R., 1993. Population genetic consequences of small population size: implications for plant conservation. Annual Review of Ecology and Systematics 24217–242. https://doi.org/10.1146/annurev.es.24.110193.001245

Ellstrand, N.C., Prentice, H.C., Hancock, J.F., 1999. Gene flow and introgression from domesticated plants into their wild relatives. Annual Review of Ecology and Systematics 30539–563. https://doi.org/10.1146/annurev.ecolsys.30.1.539

Ewing, B., Hillier, L., Wendl, M.C., Green, P., 1AD. Base-calling of automated sequencer traces using Phred. I. accuracy assessment. Genome Research 8175–185. https://doi.org/10.1101/gr.8.3.175

Excoffier, L., Smouse, P.E., Quattro, J.M., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131479–491.

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

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

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

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

Gitzendanner, M.A., Soltis, P.S., 2000. Patterns of genetic variation in rare and widespread plant congeners. American Journal of Botany 87783–792. https://doi.org/10.2307/2656886

Goldblatt, P., 1978. An analysis of the flora of Southern Africa: Its characteristics, relationships, and orgins. Annals of the Missouri Botanical Garden 65369. https://doi.org/10.2307/2398858

Goldblatt, P., Manning, J.C., 2002. Plant diversity of the Cape Region of southern Africa. Annals of the Missouri Botanical Garden 89281. https://doi.org/10.2307/3298566

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

Hansen, L.B., Siegismund, H.R., Jørgensen, R.B., 2003. Progressive introgression between *Brassica napus* (oilseed rape) and *B. rapa*. Heredity 91276–283. https://doi.org/10.1038/sj.hdy.6800335

Hedrick, P.W., 2005. A standardized genetic differentiation measure. Evolution 591633–1638. https://doi.org/10.1111/j.0014-3820.2005.tb01814.x

Holzschuh, A., Dormann, C.F., Tscharntke, T., Steffan-Dewenter, I., 2011. Expansion of mass-flowering crops leads to transient pollinator dilution and reduced wild plant pollination. Proceedings of the Royal Society B: Biological Sciences 2783444–3451. https://doi.org/10.1098/rspb.2011.0268

Hyten, D.L., Song, Q., Zhu, Y., Choi, I.Y., Nelson, R.L., Costa, J.M., Specht, J.E., Shoemaker, R.C., Cregan, P.B., 2006. Impacts of genetic bottlenecks on soybean genome diversity. Proceedings of the National Academy of Sciences 10316666–16671. https://doi.org/10.1073/pnas.0604379103

Jombart, T., Ahmed, I., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 273070–3071. https://doi.org/10.1093/bioinformatics/btr521

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

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

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

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

418 Laikre, L., Schwartz, M.K., Waples, R.S., Ryman, N., 2010. Compromising genetic diversity in the
419 wild: unmonitored large-scale release of plants and animals. Trends in Ecology & Evolution 25:
420 520–529. https://doi.org/10.1016/j.tree.2010.06.013

421 Levin, D.A., Francisco-Ortega, J., Jansen, R.K., 1996. Hybridization and the extinction of rare
422 plant species. Conservation Biology 10: 10–16. https://doi.org/10.1046/j.1523-
423 1739.1996.10010010.x

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

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

430 Malgas, R.R., Potts, A.J., Oettlé, N.M., Koelle, B., Todd, S.W., Verboom, G.A., Hoffman, M.T.,
431 2010. Distribution, quantitative morphological variation and preliminary molecular analysis of
432 different growth forms of wild rooibos (*Aspalathus linearis*) in the northern Cederberg and on
433 the Bokkeveld Plateau. South African Journal of Botany: 7672–81.
434 https://doi.org/10.1016/j.sajb.2009.07.004

435 Mantel, N., 1967. Ranking procedures for arbitrarily restricted observation. Biometrics 2365.
436 https://doi.org/10.2307/2528282

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

439 Nei, M., 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National
440 Academy of Sciences 703321–3323. https://doi.org/10.1073/pnas.70.12.3321

Niemandt, M., Roodt-Wilding, R., Tobutt, K.R., Bester, C., 2018. Microsatellite marker applications in *Cyclopia* (Fabaceae) species. South African Journal of Botany 11652–60. https://doi.org/10.1016/j.sajb.2018.02.408

Ornelas, J.F., Sosa, V., Soltis, D.E., Daza, J.M., González, C., Soltis, P.S., Gutiérrez-Rodríguez, C., de los Monteros, A.E., Castoe, T.A., Bell, C. and Ruiz-Sanchez, E., 2013. Comparative phylogeographic analyses illustrate the complex evolutionary history of threatened cloud forests of northern Mesoamerica. *PloS one*, *8*(2), p.e56283.

Pasquet, R.S., Peltier, A., Hufford, M.B., Oudin, E., Saulnier, J., Paul, L., Knudsen, J.T., Herren, H.R., Gepts, P., 3AD. Long-distance pollen flow assessment through evaluation of pollinator foraging range suggests transgene escape distances. Proceedings of the National Academy of Sciences 10513456–13461. https://doi.org/10.1073/pnas.0806040105

Potts, A.J., 2017. Genetic risk and the transition to cultivation in Cape endemic crops - The example of honeybush (*Cyclopia*)?. South African Journal of Botany 11052–56. https://doi.org/10.1016/j.sajb.2016.09.004

Potts, A.J., Hedderson, T.A., Vlok, J.H.J., Cowling, R.M., 2013. Pleistocene range dynamics in the eastern Greater Cape Floristic Region: A case study of the Little Karoo endemic *Berkheya cuneata* (Asteraceae). South African Journal of Botany 88401–413. https://doi.org/10.1016/j.sajb.2013.08.009

Prevosti, A., Ocaña, J., Alonso, G., 1975.) Distances between populations of *Drosophila subobscura*, based on chromosome arrangement frequencies. Theoretical and Applied Genetics 45231–241. https://doi.org/10.1007/bf00831894

Reboud, X., Zeyl, C., 1994. Organelle inheritance in plants. Heredity 72132–140. https://doi.org/10.1038/hdy.1994.19

Reinten, E.Y., Coetzee, J.H., van Wyk, B.E., 2011. The potential of South African indigenous plants for the international cut flower trade. South African Journal of Botany 77934–946. https://doi.org/10.1016/j.sajb.2011.09.005

Sanger, F., Nicklen, S., Coulson, A.R., 1AD. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences 745463–5467. https://doi.org/10.1073/pnas.74.12.5463

Schaal, B.A., Hayworth, D.A., Olsen, K.M., Rauscher, J.T., Smith, W.A., 1998. Phylogeographic studies in plants: problems and prospects. Molecular Ecology 7465–474. https://doi.org/10.1046/j.1365-294x.1998.00318.x

473 Schipmann, U., Leaman, D.J., Cunningham, A.B., Walter, S., 2005. Impact of cultivation and
474 collection on the conservation of medicinal plants: Global trends and issues. Acta Horticulturae
475 31–44. https://doi.org/10.17660/actahortic.2005.676.3

476 Schutte, A.L., 1997. Systematics of the genus *Cyclopia* Vent. (Fabaceae, Podalyrieae). Edinburgh
477 Journal of Botany 54125–170. https://doi.org/10.1017/s0960428600004005

478 Scott, G., Hewett, M.L., 2008. Pioneers in ethnopharmacology: The Dutch East India Company
479 (VOC) at the Cape from 1650 to 1800. Journal of Ethnopharmacology 115339–360.
480 https://doi.org/10.1016/j.jep.2007.10.020

481 Tembrock, L.R., Simmons, M.P., Richards, C.M., Reeves, P.A., Reilley, A., Curto, M.A., Meimberg,
482 H., Ngugi, G., Demissew, S., Al Khulaidi, A.W., Al-Thobhani, M., Simpson, S., Varisco, D.M., 2017.
483 Phylogeography of the wild and cultivated stimulant plant qat (*Catha edulis*, Celastraceae) in
484 areas of historical cultivation. American Journal of Botany 104538–549.
485 https://doi.org/10.3732/ajb.1600437

486 Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of
487 progressive multiple sequence alignment through sequence weighting, position-specific gap
488 penalties and weight matrix choice. Nucleic Acids Research 224673–4680.
489 https://doi.org/10.1093/nar/22.22.4673

490 Tolley, K.A., Bowie, R.C.K., Measey, J.G., Price, B.W., Forest, F., 2014. The shifting landscape of
491 genes since the Pliocene: terrestrial phylogeography in the Greater Cape Floristic Region. In:
492 Allsopp N, Colville JF, Verboom GA, eds. Fynbos: Ecology, Evolution, and Conservation of a
493 Megadiverse Region. Oxford: Oxford University Press. 142-163

494 Turpie, J.K., Heydenrych, B.J., Lamberth, S.J., 2003. Economic value of terrestrial and marine
495 biodiversity in the Cape Floristic Region: implications for defining effective and socially optimal
496 conservation strategies. Biological Conservation 112:233–251. https://doi.org/10.1016/s0006-
497 3207(02)00398-1

498 Westphal, C., Steffan-Dewenter, I., Tscharntke, T., 2003. Mass flowering crops enhance
499 pollinator densities at a landscape scale. Ecology Letters 6961–965.
500 https://doi.org/10.1046/j.1461-0248.2003.00523.x

501 Whitlock, R., Grime, J.P., Burke, T., 2010. Genetic variation in plant morphology contributes to
502 the species-level structure of grassland communities. Ecology 91:1344–1354.
503 https://doi.org/10.1890/08-2098.1

Winter, D.J., 2012. mmod: an R library for the calculation of population differentiation statistics. Molecular Ecology Resources 12:1158–1160. https://doi.org/10.1111/j.1755-0998.2012.03174.x

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

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

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

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

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

Turchetto-Zolet, A.C., Pinheiro, F., Salgueiro, F., Palma-Silva, C., 2013. Phylogeographical patterns shed light on evolutionary process in South America. Molecular ecology 22:1193-213. https://doi.org/10.1111/mec.12164

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

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

Table 1(on next page)

Summary of wild and cultivated Honeybush 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.

1

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

2

3

Table 2(on next page)

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

Haplotype frequency in each population is reported in Table 1.

1

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.

2
3

Table 3 (on next page)

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

1

	MLT S1 – MLT S2 *atpI-atpH* intergenic spacer			MLT S3 – MLT S4 *atpI-atpH* intergenic spacer									MLT U1 – MLT U2 *ndhA* intron					
Position	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 – tacagatgaaaggaagggcttcgttttttgaatcctatctaaatttacagtaacagggcaaa;
2b – tacagatgaaaggaagggcttcgttttttgaaaactatctaaatttacagtaacagggcaaa;
2c – tacagatgaaaggaaggggttcgttttttgaatcctatctaaatttacagtaacagggcaaa;
2d – taaagatgaaaggaagggcttcgttttttgaatcctatctaaatttacagtaacagggcaaa;
2e – tatagatgaaaggaagggcttcgttttttgaatcctatctaaatttacagtaacagggcaaa

2
3

Figure 1

Distribution of *Cyclopia* populations screened for haplotype diversity.

Cultivated populations indicated by closed circles, wild populations by open circles, stars indicate the locations of Honeybush nurseries, crosses indicate the locations of populations initially used for cultivar development. The natural distribution of the three target species is indicated with broken lines. 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). 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.

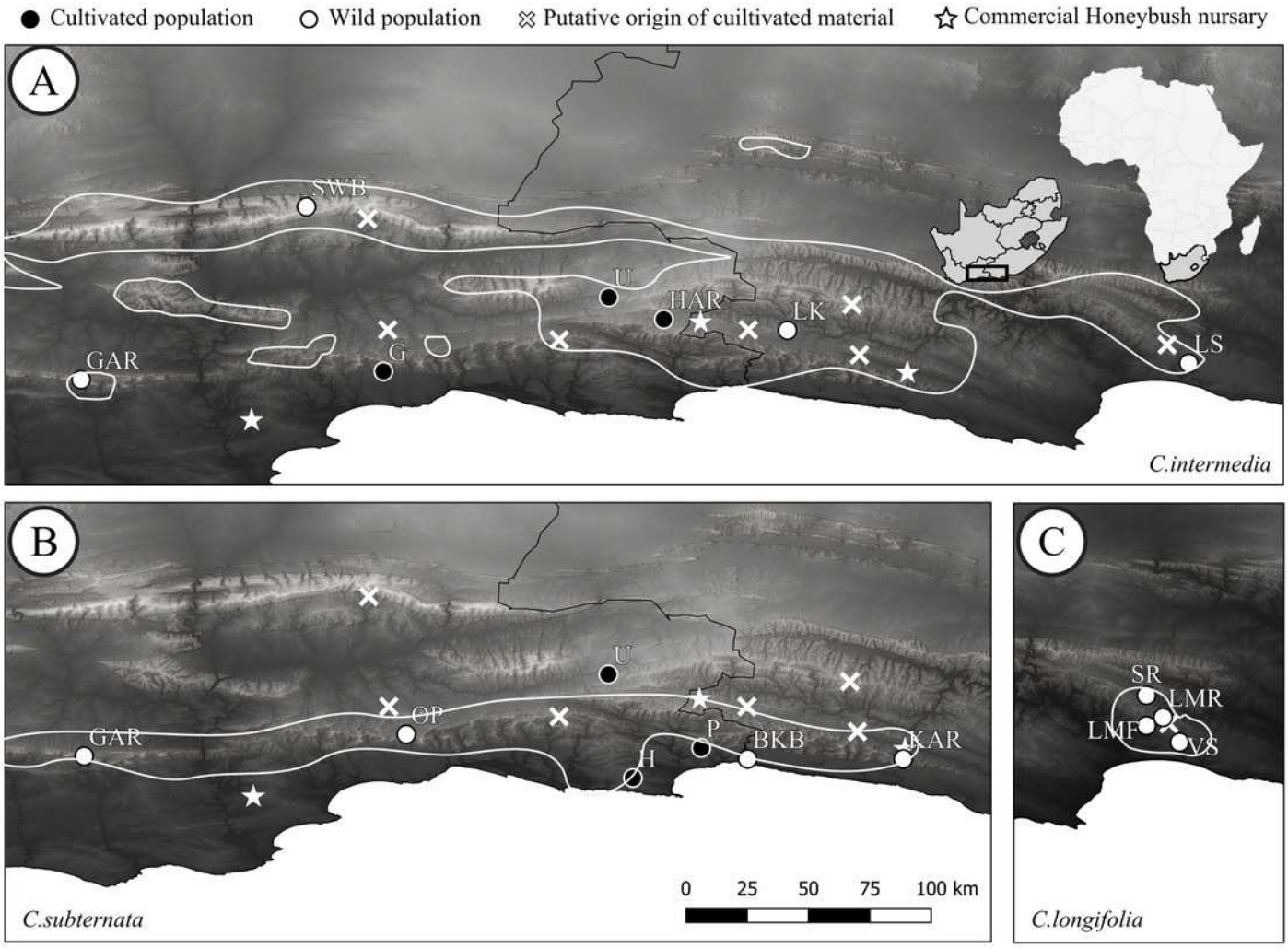

Cultivated population
Wild population
Putative origin of cuiltivated material
Commercial Honeybush nursary
A
SWB
U
HAR
LK
LS
G
GAR
C.intermedia
B
U
OP
GAR
P
BKB
KAR
H
0 25 50 75 100 km
C.subternata
C
SR
LMR
LMF
VS
C.longifolia

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.

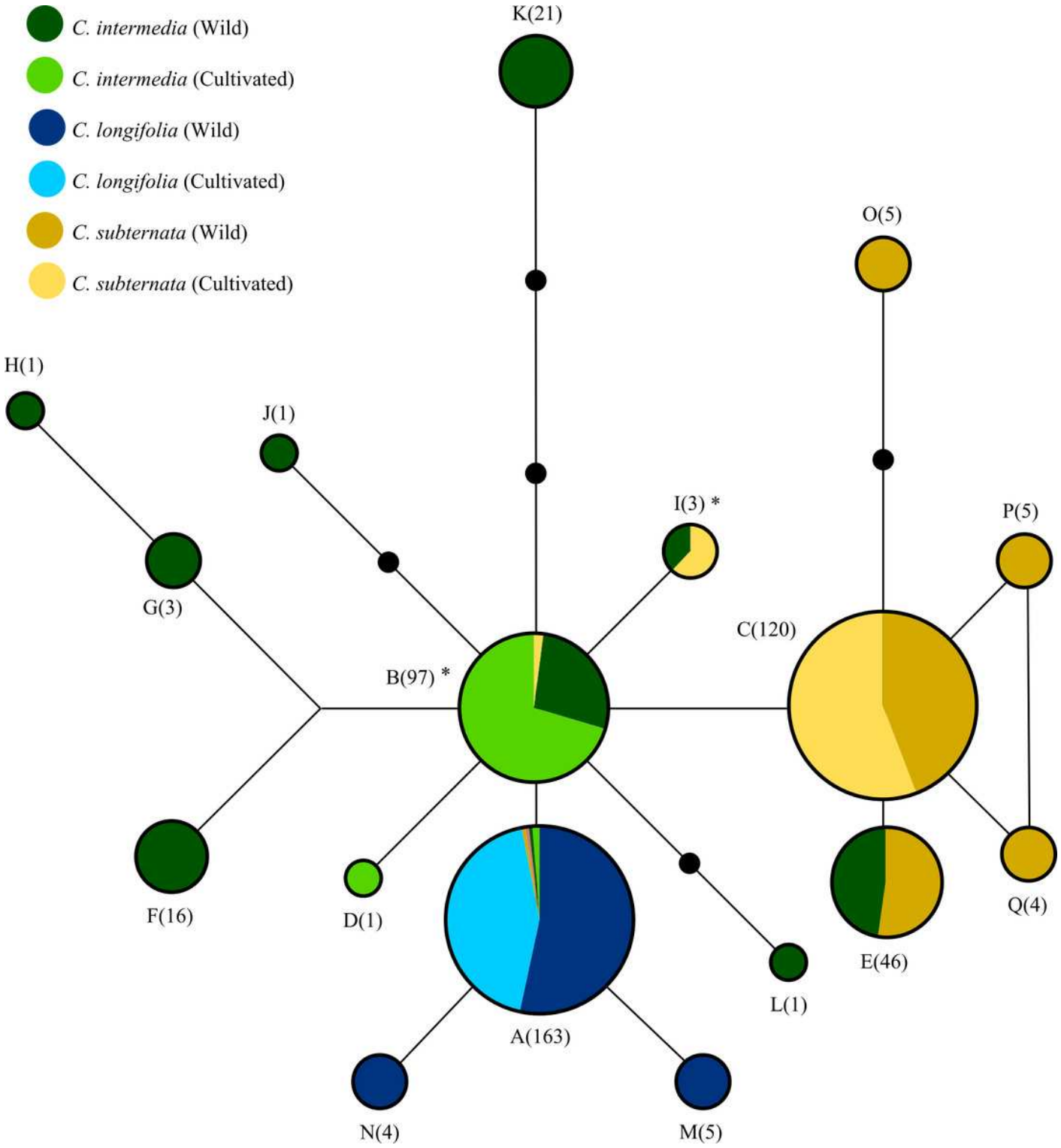
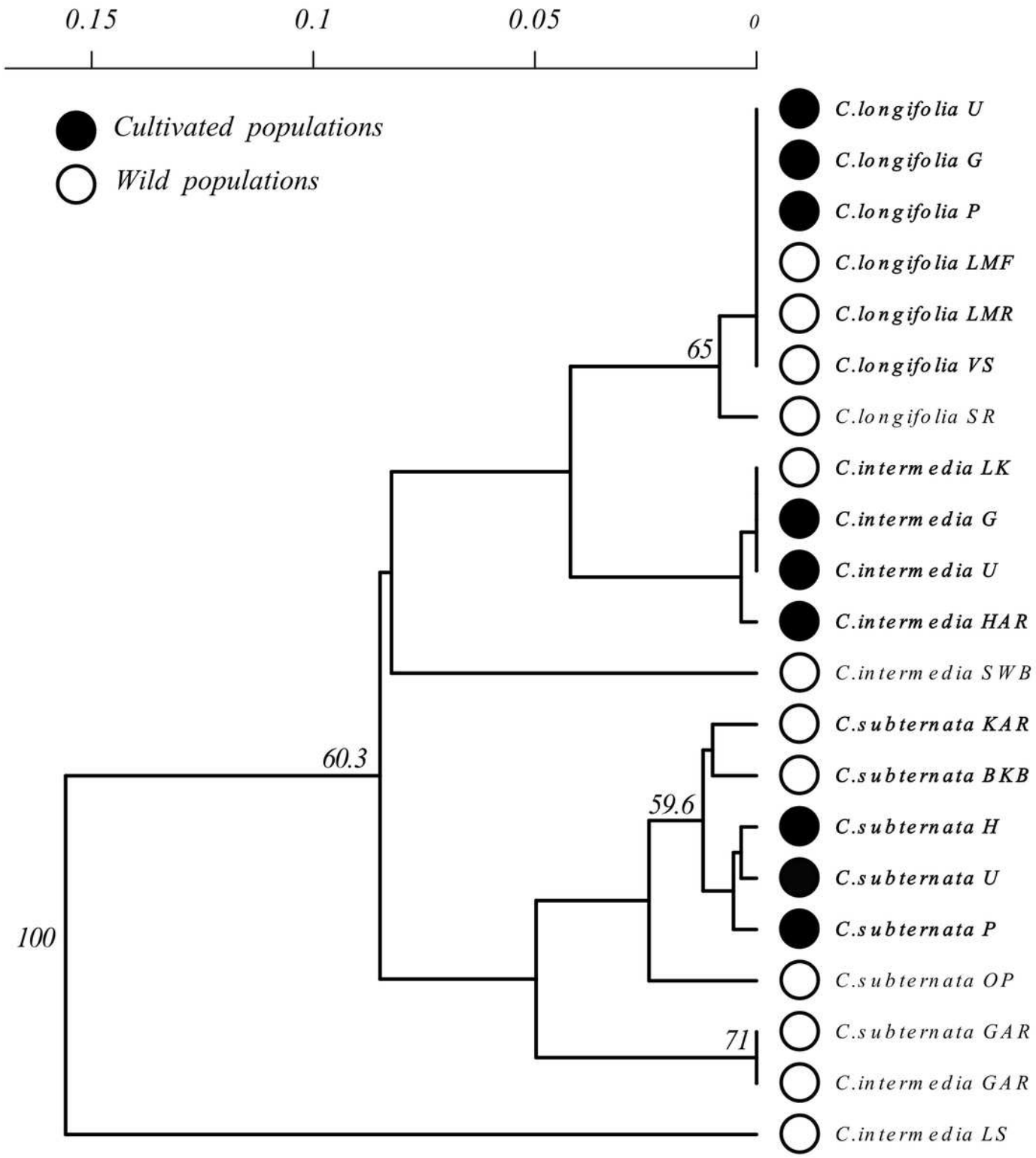

C. intermedia (Wild)
C. intermedia (Cultivated)
C. longifolia (Wild)
C. longifolia (Cultivated)
C. subternata (Wild)
C. subternata (Cultivated)
K(21)
O(5)
H(1)
J(1)
I(3) *
P(5)
G(3)
C(120)
B(97) *
F(16)
D(1)
Q(4)
E(46)
L(1)
A(163)
N(4)
M(5)

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