Application of High Resolution Melt analysis (HRM) for screening haplotype variation in a non-model plant genus: Cyclopia (Honeybush)

Nicholas C Galuszynski [Corresp., 1] , Alastair J Potts [1]

1 Department of Botany, Nelson Mandela University, Port Elizabeth, Eastern Cape, South Africa

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

Aim. This study has three broad aims: to a) develop genus-specific primers for High Resolution Melt analysis (HRM) of members of Cyclopia Vent., b) test the haplotype discrimination of HRM compared to Sanger sequencing, and c) provide an example of using HRM to detect novel haplotype variation in wild *C.* subternata Vogel. populations.

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

Methods. Polymorphic loci were detected through a screening process of sequencing 12 non-coding chloroplast DNA segments across 14 Cyclopia species. Twelve genus-specific primer combinations were designed around variable cpDNA loci, four of which failed to amplify under PCR, the eight remaining were applied to test the specificity, sensitivity and accuracy of HRM. The three top performing HRM Primer combinations were then applied to detect novel haplotypes in wild *C.* subternata populations, and phylogeographic patterns of *C.* subternata were explored.

Results. We present a framework for applying HRM to non-model systems. HRM accuracy varied across the PCR products screened using the genus-specific primers developed, ranging between 56 and 100 %. The nucleotide variation failing to produce distinct melt curves is discussed. The top three performing regions, having 100 % specificity (i.e. different haplotypes were never grouped into the same cluster, no false negatives), were able to detect novel haplotypes in wild *C.* subternata populations with high accuracy (96%). Sensitivity below 100 % (i.e. a single haplotype being clustered into multiple unique groups during HRM curve analysis, false positives) was resolved through sequence confirmation of each cluster resulting in a final accuracy of 100 %. Phylogeographic analyses revealed that wild *C.* subternata populations tend to exhibit phylogeographic structuring across mountain ranges (accounting for 73.8 % of genetic variation base on an AMOVA), and genetic differentiation between populations increases with distance (p < 0.05 for IBD analyses).

Conclusions. After screening for regions with high HRM clustering specificity — akin to the screening process associated with most PCR based markers — the technology was found to be a high throughput tool for detecting genetic variation in non-model plants.

Application of High Resolution Melt analysis (HRM) for screening haplotype variation in a non-model plant genus: *Cyclopia* (Honeybush)

Nicholas C. Galuszynski[1], Alastair J. Potts[1]

Department of Botany, Nelson Mandela University, Port Elizabeth, Eastern Cape, South Africa

Corresponding Author:
Nicholas C. Galuszynski[1]
University Way, Summerstrand, Port Elizabeth, 6011, South Africa
Email address: nicholas.galuszynski@gmail.com

Abstract

Aim. This study has three broad aims: to a) develop genus-specific primers for High Resolution Melt analysis (HRM) of members of *Cyclopia* Vent., b) test the haplotype discrimination of HRM compared to Sanger sequencing, and c) provide an example of using HRM to detect novel haplotype variation in wild *C. subternata* Vogel. populations.

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

Methods. Polymorphic loci were detected through a screening process of sequencing 12 non-coding chloroplast DNA segments across 14 *Cyclopia* species. Twelve genus-specific primer combinations were designed around variable cpDNA loci, four of which failed to amplify under PCR, the eight remaining were applied to test the specificity, sensitivity and accuracy of HRM. The three top performing HRM Primer combinations were then applied to detect novel haplotypes in wild *C. subternata* populations, and phylogeographic patterns of *C. subternata* were explored.

Results. We present a framework for applying HRM to non-model systems. HRM accuracy varied across the PCR products screened using the genus-specific primers developed, ranging between 56 and 100 %. The nucleotide variation failing to produce distinct melt curves is discussed. The top three performing regions, having 100 % specificity (i.e. different haplotypes were never grouped into the same cluster, no false negatives), were able to detect novel haplotypes in wild *C. subternata* populations with high accuracy (96%). Sensitivity below 100 % (i.e. a single haplotype being clustered into multiple unique groups during HRM curve analysis, false positives) was resolved through sequence confirmation of each cluster resulting in a final accuracy of 100 %. Phylogeographic analyses revealed that wild *C. subternata* populations tend to exhibit phylogeographic structuring across mountain ranges (accounting for 73.8 % of genetic variation base on an AMOVA), and genetic differentiation between populations increases with distance ($p < 0.05$ for IBD analyses).

Conclusions. After screening for regions with high HRM clustering specificity — akin to the screening process associated with most PCR based markers — the technology was found to be a high throughput tool for detecting genetic variation in non-model plants.

Introduction

Describing intra-population genetic diversity across a species range requires access to sufficiently variable genetic markers that can be applied to large sample sets in an efficient and cost effective manner. The lack of widely transferable marker systems with these qualities has impeded phylogeographic work in the past, especially in developing countries that harbour much of the planet's biodiversity (Beheregaray 2008). High Resolution Melt analysis (HRM, sometimes acronymed to HRMA) is a high throughput and cost effective means of screening sequence variation post Polymerase Chain Reaction (PCR), offering the unique advantage of providing rapid insights into the levels of sequence variation among samples through melt curve clustering. Having the flexibility to lend itself to a variety of applications, the technology has been widely adopted in clinical (reviewed by Vossen et al. 2009) and crop research (reviewed by Simko 2016). However, despite its apparent benefits, HRM appears to be underutilized for non-model organisms.

The HRM process is briefly described here. The inclusion of a DNA saturating fluorescent dye during PCR produces double stranded DNA molecules with dye bound to each base pair. As

58   such, the presence of double stranded PCR product is measured by its fluorescence. As the
59   PCR products are heated the double stranded DNA molecules dissociate, or melt, releasing the
60   dye, resulting in a decrease in detected fluorescence. The rate at which a DNA fragment melts
61   is dependent on the binding chemistry of the nucleotide sequence of the complementary strands
62   under analysis. Therefore, by plotting the decrease in fluorescence against the steady rate of
63   temperature increase, a melt curve determined by the DNA template under analysis is
64   produced. The resultant melt curve differences (curve shape and melt peak [Tm]) are potentially
65   indicative of sequence variation among PCR products.
66   The genotyping and mutation scanning abilities of HRM have been tested using well described
67   systems in the past, including: artificially generated SNPs (Reed & Wittwer 2004) and loci from
68   the human genome (Ebili & Ilyas 2015; Garritano et al. 2009; Li et al. 2014; Reed & Wittwer
69   2004), where the technology was found to be highly sensitive and specific, with reproducible
70   results. These studies suggest that HRM is capable of detecting single SNP variation with an
71   average sensitivity of 95% (sd=8 %) and specificity of 97% (sd=7%) in amplicons of various
72   lengths (50-1000 bp, Reed & Wittwer 2004; 51-547 bp, Li et al. 2014; and 211-400 bp, Garritano
73   et al. 2009). However, such accuracy is only possible if the starting DNA template is of sufficient
74   quality and quantity (Ebili & Ilyas 2015). Being non-destructive in nature, the PCR products can
75   also be Sanger sequenced post HRM (Vossen et al. 2009). The power of the HRM approach to
76   screen sequence variation is that it helps to avoid redundant sequencing of identical nucleotide
77   motifs (Dang et al. 2012; Vossen et al. 2009), thereby potentially reducing overall sequencing
78   costs of projects where intra-population genetic variation may be low, as in the slow evolving
79   chloroplast genome of plants (Schaal et al. 1998). In addition, HRM has been shown to be more
80   sensitive than traditional gel electrophoresis methods for microsatellite genotyping (Distefano et
81   al. 2012). Fast, reliable and cost effective — HRM appears to be an ideal molecular tool for
82   studies that require the characterization of a large number of samples that are likely to exhibit
83   low nucleotide variation.
84   Despite its apparent utility, HRM has rarely featured in phylogeographic work. Smith et al.
85   (2010) were some of the first to apply HRM to population genetics. By melting short amplicons
86   (40-60 bp) that targeted known SNPs, they successfully genotyped 121 accessions from five
87   wild swordfish (*Xiphias gladius* Bloch, Xiphiidae) populations. Cubry et al. (2015) were
88   successful in applying HRM for the discrimination of four cpDNA haplotypes that corresponded
89   with the geographic structuring of black alder (*Alnus glutinosa* (L.) Gaertn., Betulaceae),
90   screening 154 accessions across 23 populations. These studies, and most others applying
91   HRM to non-model organisms (Dang et al. 2012; Li et al. 2012; Radvansky et al. 2011), set out
92   to develop HRM primers having prior knowledge of the nucleotide variation under analyses.
93   Unfortunately, such knowledge is generally not available for the study of non-model organisms
94   and the application of HRM for detecting and genotyping of novel genetic variation in wild
95   populations is still rare (Nunziata et al. 2019; Sillo et al. 2017). High Resolution Melt analysis
96   appears to be an underutilized resource by phylogeographers.
97   Here we test the application of HRM for non-model taxa, *Cyclopia*, a commercially important
98   plant genus endemic to the CFR. This study: a) develops a set of genus-specific primers for the
99   HRM analysis of non-coding cpDNA loci to test: b) the haplotype discrimination sensitivity,
100   specificity, and accuracy of HRM, and c) the potential application of HRM for haplotype
101   detection in wild *Cyclopia* populations, focusing here on *C. subternata*. This study demonstrates

102   that (when optimized) HRM is a fast, accurate, and cost effective tool for haplotype detection in
103   non-model organisms, successfully describing the geographic structuring of genetic diversity in
104   wild *C. subternata* populations.

105

106   **Materials & Methods**

107   Taxonomic background and sampling

108   This study focuses on members of the genus *Cyclopia* Vent., which is endemic to the Cape
109   Floristic Region (CFR) and consists of 23 described species; two of which are considered
110   extinct (*Cyclopia filiformis* Kies, *Cyclopia laxiflora* Benth.) and various others ranging from
111   critically endangered to vulnerable (SANBI, 2012). *Cyclopia* species and populations tend to
112   exhibit highly localised distributions (Schutte 1997), making them potentially vulnerable to
113   genetic pollution from foreign genotypes translocated for the cultivation of Honeybush tea and
114   associated products (Ellstrand & Elam 1993; Levin et al. 1996; Potts 2017; Schutte 1997) — an
115   increasingly common practice in the CFR (McGregor 2017). The characterization and
116   conservation of wild *Cyclopia* genetic diversity is therefore of high importance.
117   To maximise the amount of genetic variation detected and the transferability of the primers
118   designed across the genus, 14 species (summarized in Table 1, closed circles in Fig 1) were
119   sampled from the full geographic range of the genus. Additionally, eight wild populations (open
120   circles in Fig 1) of *C. subternata* Vogel. were sampled to test the potential application of HRM
121   for haplotype detection using the genus-specific primers generated. Between 10 and 24
122   samples were collected per *C. subternata* population. Fresh leaf material was clipped from the
123   growing tips of wild specimens over the period of 2015-2018 and placed directly into a silica
124   desiccating medium for a minimum of two weeks prior to DNA extraction. All sampling was
125   approved by the relevant permitting agencies, Cape Nature (Permit number: CN35-28-4367),
126   the Eastern Cape Department of Economic Development, Environmental Affairs and Tourism
127   (Permit numbers: CRO 84/ 16CR, CRO 85/ 16CR), and the Eastern Cape Parks and Tourism
128   Agency (Permit number: RA_0185).

129   DNA extraction

130   Whole genomic DNA was extracted from silica-dried leaf material using a CTAB approach
131   modified from Doyle and Doyle (1987), the full extraction protocol is described in Methods S1.
132   Extracted DNA was suspended in 50 µL molecular grade water for PCR amplification with the
133   products sequenced using Sanger sequencing (Sanger et al. 1977). Samples that failed to
134   amplify during PCR, were subject to repeat DNA extracted from new leaf material and then PCR
135   amplified.

136   Developing *Cyclopia* specific HRM primers

137   While HRM has been shown to successfully detect sequence variation in PCR products of
138   various sizes (see introduction), it has been suggested that shorter PCR products are likely to
139   produce more pronounced melt curve differences than larger products with the same nucleotide
140   variation (Dang et al. 2012; Dobrowolski et al. 2009; Li et al. 2014; Liew et al. 2004; Smith et al.
141   2013; Taylor et al. 2011). Universal marker systems, such as those developed by Shaw et al.
142   (2005, 2007) are therefore unlikely to be directly transferable to HRM, as they amplify relatively

143 large PCR products, thus HRM specific primers must be developed to target shorter, variable
144 regions.
145 Developing HRM primers requires prior knowledge of the nucleotide variation of regions across
146 samples. The means of acquiring such data is dependent on the resources available to the
147 researcher and the availability of existing sequence data for the study organisms. Thus template
148 data could range from Next Generation Sequencing derived genomic data to the application of
149 HRM to existing microsatellite markers, or existing data available from international nucleotide
150 sequence databases such as GenBank (https://www.ncbi.nlm.nih.gov/genbank/).
151 For *Cyclopia*, however, existing sequence data (predominantly from the ribosomal ITS region)
152 exhibited low levels of differentiation amongst species (Galuszynski and Potts 2017; Van Der
153 Bank et al. 2002), lacking the variation required for population level analyses. Therefore,
154 polymorphic loci were identified from non-coding cpDNA regions via Sanger sequencing
155 (Sanger et al. 1977) of PCR products amplified using the protocols and universal primers
156 described by Shaw et al. (2005, 2007). A total of 16 non-coding cpDNA regions under went
157 PCR, however four regions failed to amplify (and could not be sequenced). The 12 regions that
158 were sequenced are summarized in Table 1, all necleotide sequence data is availible from
159 GenBank and accession numbers are provided in Table S1.
160 Sequences were assembled using CondonCode Aligner *[v2.0.1]* (Codon Code Corp, http://www.
161 codoncode.com). The PHRED base-calling program (Ewing et al. 1998) was used to assign a
162 quality score for each sequence, then sequences were automatically aligned using ClustalW
163 (Thompson et al. 1994) and visually inspected for quality. All short indels (< 3 bp) occurring in
164 homopolymer repeat regions were considered alignment errors and removed from the
165 alignment. The consensus sequence alignment for polymorphic regions were exported and
166 utilized in HRM primer design.
167 Primer design was guided by two constraining factors: (1) sequences had to contain
168 conservative regions with a high GC content that could form the primer binding template, and
169 (2) these regions had to flank polymorphic sites. Wherever possible, internal HRM primers were
170 designed in a way that would split a region into neighboring loci, as suggested by Dang et al.
171 (2012). This approach allows for adjacent loci to be sequenced in a single run by amplifying the
172 full region, and then during alignment, split the region into the neighboring loci that underwent
173 HRM analysis. This approach reduces the time involved in sequence alignment and number of
174 samples required to be sequenced for HRM clustering verification.
175 High Resolution Melt specific primers were designed using the online resource Primer-Blast
176 (www.ncbi.nlm.nih.gov/tools/primer-blast/). The sub-family Faboideae was used as the
177 reference taxon to check for primer specificity searched against the NCBI Reference Sequence
178 representative genomes (www.ncbi.nlm.nih.gov/refseq/); PCR product size was limited to
179 between 50 and 550 bp (as this falls within the amplicon size predicted to produce the highest
180 levels of genotyping accuracy; Dang et al. 2012; Dobrowolski et al. 2009; Li et al. 2014; Liew et
181 al. 2004; Taylor et al. 2011), primer melting temperature was set at 60 °C (± 3 °C) (as
182 suggested by Taylor et al. 2011) and a maximum of 20 primer pairs were returned per search.
183 The positions of these primers within their respective region alignment were manually evaluated
184 to ensure that they occurred in well conserved sites, i.e. any primers occurring across
185 polymorphic loci were discarded.

186 Eleven genus-specific primer pairs (Table S2) were developed from seven of the twelve non-
187 coding cpDNA regions, of which eight primer pairs successfully amplified PCR products and
188 were thus selected for HRM screening (Table 2). The remaining three were excluded from the
189 analysis due to poor PCR amplification. The primer pairs selected for HRM screening amplified
190 between four and six unique haplotypes each, across five cpDNA regions (nucleotide
191 differences are summarized in Table 3). Primers selected for the evaluation of HRM accuracy
192 are reported in Table 2.

Testing PCR amplification of HRM primers

194 The genomic DNA extracted for samples that amplified unique haplotypes (as determined from
195 the sequence data used to develop HRM primers) was quantified using a NanoDrop 2000c
196 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE19810r Scientific, USA) and 5 ng/L
197 DNA dilutions were made for HRM analysis. High Resolution Melt analysis was conducted for all
198 primer pairs developed, with 16 replicates amplified per sample (haplotype). However, only
199 replicates that produced sufficient PCR products, as determined from PCR amplification curves
200 (see examples in Fig 2 and 3) were included in the evaluation of HRM haplotype discrimination
201 (number of replicated subjected to HRM analysis for each haplotype are reported in Table 3).
202 This PCR amplification screening approach was adopted as the aim of this phase of the study
203 was to test the haplotype discrimination abilities of HRM based on the underlying nucleotide
204 differences between haplotypes and not the quantity and quality of PCR product under analysis
205 (which can vary due to pippetting errors). Regions that failed to produce consistent PCR
206 amplification curves (possibly due to non-specific primer binding), were excluded from
207 subsequent analysis (see examples of PCR and HRM curves excluded from analysis in Figs 2
208 and 3).

PCR and HRM reactions

210 All reactions (PCR amplification and subsequent HRM) took place in a 96 well plate CFX
211 Connect (Bio-Rad Laboratories, Hercules, California, U.S.A.) in 10 µL reaction setups,
212 consisting of 4 µL genomic DNA (5 ng/µL), 1 µL each primer (10 mM) and 5 µL Precision Melt
213 Supermix containing hot-start iTaqTM DNA polymerase, dNTPs, MgCl2, EvaGreen dye (Bio-
214 Rad Laboratories, Hercules, California, U.S.A.).
215 Polymerase Chain Reaction amplification and melt conditions were as per manufacturer's
216 specifications (Table 4) and the annealing temperature set to the primer pair's mean Tm
217 (melting temperature), reported in Table 2. The automated clustering algorithm of the High
218 Precision Melt software™ (Bio-Rad Laboratories, Hercules, California, U.S.A.) was performed
219 on the normalized florescence data and used to group melt curves into clusters that represent
220 putative haplotypes. HRM clustering settings used were ΔTm threshold at 0.05 ℃ and curve
221 shape sensitivity settings and temperature correction, 70% and 20 respectively.

HRM discrimination of sequenced haplotypes

223 Following the descriptions of Altman and Bland (1994), HRM discrimination (sensitivity,
224 specificity and accuracy) was determined for each of the haplotypes amplified by the eight HRM
225 primers that produced sufficient PCR product for HRM analysis. Sensitivity, or the true positive
226 rate, refers to HRM's ability to correctly assign haplotype replicates into the same HRM cluster.
227 $$\text{Sensitivity} = TP/(TP+FN)$$

228 TP=TruePositive FN=FalseNegative

229 Specificity, or true negative rate, is the measure of HRM's ability to correctly discern between
230 haplotypes, grouping them into different HRM clusters.

231 Specificity =TN/(TN+FP)

232 TN=TrueNegative FP=FalsePositive

233 The accuracy of HRM refers to how close haplotype clustering reflects the true identities of the
234 haplotypes and was measured as:

235 Accuracy=(TP+TN) /(TP+FP+TN+FN)

236 Since sensitivity below 100 % will be accounted for during HRM cluster (i.e. putative haplotype)
237 confirmation by sequencing (with a subset of samples from each unique HRM cluster
238 sequenced), all regions with 100 % specificity were included for the detection of novel
239 haplotypes in wild *C. subternata* populations.

The potential for HRM to detect haplotype variation in wild populations

241 Only three regions (MLT S1 - MLT S2, MLT S3 - MLT S4, and MLT U1 - MLT U2) were found to
242 have an HRM clustering specificity of 100% (Fig 4). Thus these regions were screened for
243 haplotype variation across 142 accessions from eight wild *C. subternata* populations.
244 The same approach as Dang et al. (2012) was employed, with each sample run in duplicate and
245 haplotype clustering performed on a single population basis with the intention of reducing errors
246 resulting from variation of PCR product concentration and quality across samples from different
247 population extractions. This was achieved by using the built in well group function in the CFX
248 Manager™ Software (Bio-Rad Laboratories, Hercules, California, U.S.A.), thus multiple
249 populations could be included in a run, but analyses separately for HRM clustering.
250 The cpDNA regions that were used to design the primers used for HRM haplotype detection
251 were amplified and sequenced (following the same protocols as before) to confirm the haplotype
252 identity of HRM clusters. The loci amplified by MLT S1- MLT S2 and MLT S3 - MLT S4 are
253 adjacent to one another and by sequencing the full atpI-atpH intergenic spacer, the sequence
254 identity of both loci could be confirmed with reduced sequencing and alignment effort. Moreover,
255 the position of the loci amplified by the HRM primers occurred near the center of their respective
256 parent regions and unidirectional sequencing using the reverse primers of Shaw et al. (2007)
257 proved sufficient for verifying the sequence motifs under HRM analysis. A minimum of three
258 accessions representing each HRM cluster (i.e. putative haplotype) in each population were
259 sequenced for haplotype verification. Samples whose replicates were classified as two different
260 clusters, thus having uncertain haplotype identity, were also sequenced to ensure they were
261 assigned correctly. A total of 46 and 38 accessions were sequenced for the *atpI-atpH* intergenic
262 spacer and *ndhA* intron, respectively. Haplotype discrimination by HRM was calculated using
263 the *C. subternata* samples sequenced for haplotype confirmation, following the same formula as
264 before.

Phylogeographic analysis of *C. subternata*

266 The haplotypes detected via HRM clustering and confirmed by sequencing (described above)
267 were assembled following the same procedure described under 'Developing *Cyclopia* specific
268 HRM primers'. All wild *C. subternata* samples that underwent HRM analysis were then assigned

269   the haplotype identity of the HRM cluster they belonged using a custom *R* script written by A.J.P
270   (provided as File S1 which includes example files for running the script). The cpDNA regions
271   under investigation (*atpI-atpH* intergenic spacer and *ndhA* intron) are maternally inherited in
272   tandem and not subject to recombination (Reboud and Zeyl, 1994), and were therefore
273   concatenated for subsequent analysis.
274   The genealogical relationships among the concatenated haplotypes were determined from a
275   Statistical Parsimony (SP) network (Fig 5 inset) constructed in TCS *[v1.2.1]* (Clement et al.
276   2000). Two *C. intermedia* E. Mey. individuals with existing sequence data for the *atpI-atpH*
277   intergenic spacer and *ndhA* intron generated during the primer development phase of the study
278   were included as outgroup taxa. Default options were used to build the network and all indels
279   were reduced to single base-pairs as the software treats a multiple base pair gap as multiple
280   mutations. Haplotype distributions were mapped (Fig 5) in QGIS *[v3.2.2]* (QGIS Development
281   Team 2018).
282   The following population genetic differentiation measures were calculated from the
283   concatenated haplotypes: pairwise Gst (Nei 1973), G""st (Hedrick 2005) (both indicators of
284   allele fixation) Jost's D (Jost 2008), which measures allelic differentiation between populations,
285   and Prevosi's dist (Prevosti et al. 1975) a measure of pairwise population genetic distance that
286   counts gaps as evolutionary events (all gaps were reduced to single base pair events). These
287   measures provide insight into current allele distributions without assuming historical gene flow
288   patterns (Jost et al. 2018). Isolation By Distance (IBD) was evaluated among populations testing
289   the correlation between these genetic differentiation measures and pairwise geographic
290   distance using a Mantel test (Wright 1943) with 9999 permutations, as implemented using the
291   *ade4 [v1.7]* library (Dray & Dufour 2007; Kamvar et al. 2014) in *R [v3.5.1]* (R Core Team 2018).
292   In order to account for the possibility of non linear population expansion, relationship between
293   population differentiation measures and the natural logarithm of geographic distance was tested
294   following the same approach (Rousset, 1997). Finally, genetic differentiation across the
295   mountain ranges that populations were sampled from was tested via an Analysis of Molecular
296   Variance (AMOVA) (Excoffier et al. 1992). The mountain ranges included in the AMOVA
297   included: the Tsitsikamma (3 populations, 52 samples), Outeniqua east (2 populations, 31
298   samples), Outeniqua west (2 populations, 35 samples), and Langeberg (1 population, 24
299   samples) ranges, as described in Table 5.

Results

HRM discrimination of sequenced haplotypes

302   High Resolution Melt curve clustering of haplotypes identified via sequencing for primer
303   development produced variable results: sensitivity ranged from 56 % - 100 %, specificity ranged
304   from 27 % - 100 %, and accuracy ranged from 36 % - 100 % (the number of replicates assigned
305   to each HRM cluster is reported in Table 3 and sensitivity, specificity and accuracy is
306   summarized in Fig 4).
307   Nucleotide differences between haplotypes failing to produce distinct melt curves, and thus
308   undifferentiated by HRM clustering, are summarized in Table 6. Of the haplotypes not
309   differentiated by HRM: two haplotypes differ by indels, while the remaining 15 comparisons
310   differ by at least one transversion, and two comparisons differed by a transversion and
311   transition. The haplotypes that did produce distinct melt curves differed by at least a transition

312 (26 cases), or multiple SNPs (16 cases), one haplotype differed by a 19 bp indel, and another
313 by a 6 bp indel. All haplotype sequence variation is summarized in Table 3. As previously
314 stated, the three HRM primer combinations with specificity of 100%, two targeting the *atpI-atpH*
315 intergenic spacer (MLT S1 -S2, MLT S3 - MLT S4) and one targetting *ndhA* intron (MLT U1 –
316 U2), were selected for haplotype discovery in wild *C. subternata* populations.

Detection of haplotype variation in wild populations via HRM

318 High Resolution Melt curve analysis of accessions from wild *C. subternata* populations revealed
319 no variation in the cpDNA haplotypes amplified by the MLT S3 - MLT S4 primer combination,
320 confirmed by sequencing, and the locus was subsequently excluded from further analyses. Five
321 distinct haplotypes were verified by sequencing a subset of samples (ranging from three to eight
322 individuals per population) from each HRM cluster for the remaining two primer combinations.
323 Of the 142 samples less than 29 % were required to be sequenced for haplotype confirmation.
324 Both loci were found to have 100 % specificity, i.e. HRM successfully discriminated among all
325 haplotypes detected in wild *C. subternata* populations. However, haplotype richness was
326 overestimated by HRM (sensitivity of 87.6 % and 95.5 % for MLT S1 - MLT S2 and MLT U1 -
327 MLT U2 respectively), both cpDNA regions had accuracies of 96 %. However, as these
328 additional clusters were sequenced for haplotype confirmation, samples were assigned the true
329 identity of haplotypes resolving any potential issues of low sensitivity.
330 The final cpDNA dataset comprised 561 bp, 217 bp from the *atpI-atpH* intergenic spacer (MLT
331 S1 - MLT S2) and 344 bp from the *ndhA* intron (MLT U1 - MLT U2), with a GC content of 29 %.
332 An additional 310 base pairs (bp) were amplified by MLT S3 - MLT S4, revealing no nucleotide
333 variation. The dataset contained five polymorphic sites; four transversions, one transition, and a
334 seven bp indel (nucleotide differences summarised in Table S3).

Cyclopia subternata phylogeography

336 The SP network revealed a radiation from a central ancestral haplotype, with few mutations
337 separating haplotypes (Fig 5 inset). The ancestral haplotype was present in all populations,
338 except the western most Garcia's Pass population, located in the Langberg Mountains. This
339 population contains a single, unique haplotype. An additional two populations (Kareedouw Pass
340 and Bloukrans Bridge) were also found to contain rare, localized haplotypes and a low
341 frequency haplotype was detected in two populations located in the Tsitsikamma and Outeniqua
342 mountains (Fig 5). Population genetic differentiation measures increased with geographic
343 distance (R^2 = 0.77, 0.74, 0.70, and 0.76 for Gst, G''st, Jost's D and Provesti's dist
344 respectively, $p < 0.05$ for all measures), with significance increasing when tested against log
345 transformed geographic distance (R^2=0.64, 0.67, 0.61, and 0.65 for Gst, G''st, Jost's D and
346 Provesti's dist as before, $p < 0.05$ for all measures). The AMOVA revealed significant ($p <0.05$)
347 structuring across mountain ranges, accounting for 73.8 % of genetic variation (AMOVA results
348 summarised in Table S4).

Discussion

350 A nested framework (Fig 3) was developed to test and apply HRM to non-model organisms,
351 members of the Cape endemic plant genus *Cyclopia*. Polymorphic sites were identified via
352 sequencing 12 non-coding cpDNA regions across 14 *Cyclopia* species. PCR primers for HRM

353   analysis were designed to flank these variable sites, producing 11 HRM primer pairs across 7
354   regions. Eight of these pairs successfully amplified PCR products and were subsequently
355   analysed via HRM. Specificity of 100% was detected for three of the primer pairs, which were
356   then used to detect haplotype variation in wild *C. subternata* populations with a haplotypes
357   detection accuracy of 96 %. Haplotype detection errors were due to false negatives reducing
358   HRM sensitivity. False negatives occur when HRM incorrectly assigns a single haplotype to
359   multiple clustering groups, an issue that is resolved when the haplotype identity of HRM clusters
360   is confirmed by sequencing. Optimized HRM was demonstrated to be a powerful tool for
361   detecting genetic variation in non-model organisms, providing immediate insights into within
362   population genetic variation via automated melt curve clustering and substantially reduced
363   sequencing efforts. The framework provided here offers a straightforward approach to develop
364   and test the potential application of HRM to non-model systems.

365   ## HRM discrimination of sequenced haplotypes

366   Differences in DNA melt curves, as detected by HRM, stem from the effects nucleotide
367   sequence chemistry has on melt peak intensity and curve shape. While HRM is reported to be
368   capable of discriminating between any SNP type, the approach may be constrained by physical
369   and chemical properties of the DNA fragment under melt analysis (Gundry et al. 2008). Some
370   nucleotide variations, namely class 3 (C ↔ G) and class 4 (A ↔ T) SNPS, tend to produce
371   negligible changes in melt behaviour (curve shape and melt peak) and are often poorly detected
372   by HRM (Dang et al. 2012; Gundry et al. 2008; Yamagata et al. 2018). This is likely to be
373   exaggerated when analysing longer PCR products, as shorter PCR products produce more
374   pronounced melt curve differences than longer nucleotide motifs with the same SNP variation
375   (Li et al. 2014; Liew et al. 2004; Taylor et al. 2011; Tindall et al. 2009). Furthermore, nearest
376   neighbour chemistry (the identity of nucleotides directly adjacent to the SNP under investigation)
377   has been shown to impact the melt peak of PCR products, negating any change in melt peak
378   produced by class 3 and 4 SNPs in some cases (Yamagata et al. 2018).
379   Many of these observations are supported by the findings of this study, however some important
380   deviations were detected. Haplotypes that were successfully discriminated by HRM tended to
381   have a class 1 SNP (transitions, C ↔ T and A ↔ G) or multiple SNPs differentiating them.
382   However, seven haplotypes differing by multiple SNPs did not produce distinct melt curves
383   (Table 6), suggesting that some SNPs may potentially counteract one anothers impact on the
384   melt curve. Furthermore, haplotypes that differed by a class 2 (transversions, C ↔ A, G ↔ T)
385   and, as predicted, class 4 SNPs do not appear to have detectable melt curve differences. It is,
386   however, uncertain why in this study some class 2 SNPs produced distinct melt curves in some
387   cases (MLT M1 - MLT M2 and MLT S3 - MLT S4), but not in others (MLT C1-C4 and MLT C3 -
388   C4). Nearest neighbor chemistry does not appear to be provide insights into this as the SNPs
389   had the same neighbouring base pairs across PCR products. Furthermore, a class 2 SNP was
390   differentiated by HRM in a larger PCR product (527 bp) and not in the smaller products (386 bp
391   and 236 bp), indicating that shorter DNA fragments do not necessarily produce more distinct
392   melt curves than larger fragments with the same nucleotide differences.
393   The primer design choices in this study were largely based on the suggestions that nucleotide
394   variation in shorter DNA strands will have a more pronounced impact on melt curve shape and
395   intensity. This appears to have not been the case and larger PCR products performed as well, if

396 not better, than smaller regions, as detected elsewhere (Dang et al. 2012; Dobrowolski et al.
397 2009). Future HRM primer design efforts should possibly explore larger target regions that are
398 more likely to cover multiple SNPs and thus produce more distinct melt curves (Dang et al.
399 2012), such as the products amplified by primer combinations; MLT S1- MLT S2, MLT S3 - MLT
400 S4, and MLT U1- MLT U2. This opens HRM up to exploration of existing universal primers, such
401 as those of Shaw et al. (2005, 2007), but additional PCR optimization may be required prior to
402 being applied to HRM.

403 Detection of haplotype variation in wild *Cyclopia* populations via HRM

404 High Resolution Melt analysis using the two best performing primer pairs that amplified variable
405 regions proved to be a highly accurate (96 % for both regions screened) means of detecting
406 haplotypes variation in wild *Cyclopia* populations with no cases of different haplotypes occurring
407 in the same cluster (specificity = 100 %).
408 A remarkable feature of HRM is its high and rapid throughput. Running samples in duplicate on
409 a 96 well plate allowed for 48 samples to be screened every three hours. As such, all 142 wild
410 *C. subternata* samples were screened across the two cpDNA regions in two days, with
411 immediate insights into the underlying levels of genetic variation (based on HRM clusterings).
412 This rapid data production comes at a minimal cost per sample, which in this study amounted to
413 $ 11.09 including all PCR amplification and sequencing for the phylogeographic analysis of *C.*
414 *subternata*. A costing analysis based on quotes obtained in 2017, for a broader *Cyclopia*
415 research project that employed Anchored Hybrid Enrichment (Lemmon et al. 2012) for
416 nucleotide sequence generation, revealed that, while the cost per bp was not greatly reduced
417 when applying HRM ($ 0.013 /bp) as compared to Sanger sequencing ($ 0.015 /bp), and more
418 costly than high throughput sequencing approaches ($ 0.0005 /bp, excluding library preparation
419 and bioinformatic services). The true value of HRM lies in the ability to screen large numbers of
420 samples, with the cost per sample for HRM being 40 % that of Sanger sequencing and 16 %
421 that of Anchored Hybrid Enrichment.

422 Distribution of *C. subternata* genetic diversity

423 Despite the relatively low genetic differentiation and variation detected across wild *C. subternata*
424 populations, with a widespread haplotype detected in all populations sampled in the
425 Tsitsikamma and Outeniqua mountains, genetic diversity does appear to be spatially structured.
426 Geographically isolated haplotypes were detected in populations in the Tsitsikamma mountains,
427 and complete haplotype turnover was detected in Garcia's Pass population from the Langeberg;
428 possibly a consequence of a genetic bottleneck resulting from a small founding population,
429 facilitating rapid fixation of rare alleles (Klopfstein et al. 2006). These, and an additional low
430 frequency haplotype shared between Langkloof and Outeniqua populations, provided sufficient
431 divergence across mountain ranges to be detected by an AMOVA and roughly coincide with NJ
432 clustering of populations (Fig S1). The transition between mountain ranges represents steps of
433 increased genetic differentiation between populations (supported by significant IBD, Slatkin
434 1993), and the movement of seed and seedlings across these isolating barriers for Honeybush
435 cultivation should be avoided.
436 The population divergence described above is in contrast to that reported for the nuclear
437 genome of *C. subternata* (Niemandt et al. 2018). While Niemand et al. (2018) also detected a
438 genetically unique population (located in Harlem), no *C. subternata* was detected in this area

439   during sampling activities despite assistance from landowners in locating wild C. *plicata* Kies
440   populations (iNaturalist observation 14257580) that have been harvested and traded as *C.*
441   *subternata.* We suggested that additional work be done to describe the *C. plicata* and *C.*
442   *subternata* populations in this area to confirm potential sympatry between these two
443   morphologically and ecologically similar species (Schutte, 1997). No genetic divergence was
444   reported between the two wild *C. subternata* populations (sampled from the Tsitsikamma and
445   Outeniqua mountains) screened and the Agricultural Research Council's (ARC) genebank
446   accessions (Niemandt et al. 2018). Genetic material from this genebank has recently been
447   made commercially available for the establishment of cultivated Honeybush stands, including in
448   the Langeberg that supports the genetically distinct GAR population (Joubert et al. 2011;
449   Niemandt et al. 2018). The effective population size of the *C. subternata* nuclear genome is a
450   scale of magnitude larger than the cpDNA due to the species high ploidy level (hexalpoid, 2n =
451   6x = 54, Motsa et al. 2018; Schutte 1997), as such drift may occur more slowly. Additionally,
452   pollen dispersal by carpenter bees (*Xylocopa* spp) may reduce population divergence through
453   rare long distance dispersal events. Seed, in contrast, is dispersed locally by ants (Schutte
454   1997) and dehiscent seed pods and long distance dispersal is extremely unlikely, unless
455   anthropogenically mediated; this has likely been the case with genetic material actively
456   redistributed across the CFR for the establishment of cultivated populations and breeding
457   trials(Joubert et al. 2011).
458   The geographic distribution of *C. subternata* genetic diversity, as described here, indicates that:
459   a) unique haplotypes occur within populations, and b) these unique haplotypes are spatially
460   structured. These patterns of genetic diversity need to be acknowledged in the management of
461   this economically important species, with seed and seedling not translocated outside of the
462   mountain range that they were sourced from.

Conclusions

464   This study demonstrates that HRM is capable of discerning between cpDNA haplotypes, with
465   variable levels of success. When the top performing HRM regions were applied to screening
466   genetic variation in wild populations of the non-model organism, *C. subternata*, all haplotypes
467   were differentiated. While the framework described herein provides a clear guideline on
468   generating the markers required for applying HRM to non-model systems, some analytical
469   adjustments may be required based on the HRM platform available to the lab in question. The
470   high throughput of HRM offers the molecular ecologist the opportunity to increase
471   intrapopulation sample numbers without increasing project costs, while the automated clustering
472   provides real time insights into the underlying levels of genetic variation. Furthermore, this
473   technology may be particularly well suited to the study of conserved and slow mutating nuclear
474   regions and the chloroplast genome of plants (Schaal et al. 1998) where low intrapopulation
475   genetic variation is predicted and redundant sequencing of the same nucleotide motifs is likely.
476   The *Cyclopia* specific primers developed here provide a starting point for assessing potential
477   issues of genetic pollution associated with the transition to commercial Honeybush cultivation
478   (Potts 2017). However, further resolution may be required for more in depth population studies
479   and additional cpDNA regions as well as low copy nuclear loci should be explored for HRM
480   primer development. Furthermore, the tools produced here, while suitable for phylogeographic
481   work (as demonstrated here), are limited to the maternally inherited chloroplast genome and are
482   not suitable for exploration of interspecific hybrid detection in cultivated Honeybush populations.

483
484 **Acknowledgements**
485 We would like to thank Gillian McGregor and her students, Nicola van Berkel, Dianne Turner, as
486 well as the Agricultural Research Council of South Africa for their assistance in sample
487 collection and for the valuable discussions on Honeybush ecology and industry. We would also
488 like to thank A. Shutte-Vlock for providing us with samples of hard to collect species and for
489 always being willing to assist in identifying species. We thank the agencies that provided the
490 funding for this research, the National Research Fund of South Africa (Grant No. 99034, 95992,
491 114687) and the Table Mountain Fund (Grant no. TM2499) and the various land-owners and
492 conservation bodies that allowed for sample collection. Additionally, we acknowledge Dr J. R.
493 Alvarado-Bremer and one anonymous reviewer for their comments that improved the quality
494 and clarity of this manuscript.
495
496 References
497 Altman, D. G., & Bland, J. M. (1994). Diagnostic tests. 1: Sensitivity and specificity. BMJ
498 (Clinical research ed.), 308(6943), 1552.
499 Beheregaray, L. B. (2008). Twenty years of phylogeography: The state of the field and the
500 challenges for the Southern Hemisphere. Mol. Ecol., 17(17), 3754-3774.
501 Clement, M., Posada, D., & Crandall, K. A. (2000). TCS: A computer program to estimate gene
502 genealogies. Mol. Ecol., 9(10), 1657-1659.
503 Cubry, P., Gallagher, E., O'Connor, E., & Kelleher, C. T. (2015). Phylogeography and
504 population genetics of black alder (*Alnus glutinosa* (L.) Gaertn.) in Ireland: putting it in a
505 European context. Tree Genet. Genomes, 11(5), 99.
506 Dang, X. D., Kelleher, C. T., Howard-Williams, E., & Meade, C. V. (2012). Rapid identification of
507 chloroplast haplotypes using High Resolution Melting analysis. Mol. Ecol. Resour., 12(5), 894-
508 908.
509 Distefano, G., Caruso, M., La Malfa, S., Gentile, A., & Wu, S.-B. (2012). High resolution melting
510 analysis is a more sensitive and effective alternative to gel-based platforms in analysis of SSR:
511 An example in citrus. PLoS One, 7(8), e44202.
512 Dobrowolski, S. F., Hendricks, A. T. M., van den Bosch, B. J. C., Smeets, H. J. M., Gray, J.,
513 Miller, T., & Sears, M. (2009). Identifying sequence variants in the human mitochondrial genome
514 using high-resolution melt (HRM) profiling. Hum. Mutat., 30(6), 891-898.
515 Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh
516 leaf tissue. Phytochem. Bull., 19, 11-15.
517 Dray, S., & Dufour, A.-B. (2007). The ade4 Package: Implementing the duality diagram for
518 ecologists. J. Stat. Softw., 22(4).
519 Ebili, H., & Ilyas, M. (2015). High resolution melt analysis, DNA template quantity disparities and
520 result reliability. Clin. Lab., 61(1-2), 155-159.
521 Ellstrand, N. C., & Elam, D. R. (1993). Population genetic consequences of small population
522 size: Implications for plant conservation. Annu. Rev. Ecol. Evol. Syst., 24(1), 217-242.
523 Ewing, B., Hillier, L., Wendl, M. C., & Green, P. (1998). Base-calling of automated sequencer
524 traces using phred. I. Accuracy assessment. Genome Res., 8(3), 175-185.
525 Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred
526 from metric distances among DNA haplotypes: Application to human mitochondrial DNA
527 restriction data. Genetics, 131(2), 479-491.

528 Galuszynski, N. C., & Potts, A. J. (2017). Barcoding life in the Cape: insights from the
529 phylogeography of a small Cape genus. Genome, 60 (11), 935-936.
530 Garritano, S., Gemignani, F., Voegele, C., Nguyen-Dumont, T., Le Calvez-Kelm, F., De Silva D.,
531 Lesueur F., Landi S., & Tavtigian S.V. (2009). Determining the effectiveness of High Resolution
532 Melting analysis for SNP genotyping and mutation scanning at the TP53 locus. BMC Genet., 10,
533 5.
534 Gundry, C. N., Dobrowolski, S. F., Martin, Y. R., Robbins, T. C., Nay, L. M., Boyd, N., & Teng,
535 D. H. F. (2008). Base-pair neutral homozygotes can be discriminated by calibrated high-
536 resolution melting of small amplicons. Nucleic Acids Res., 36(10), 3401-3408.
537 Hedrick, P. W. (2005). A standardized genetic differentiation measure. Evolution, 59(8), 1633-
538 1638.
539 Jost, L. (2008). GST and its relatives do not measure differentiation. Mol. Ecol., 17(18), 4015-
540 4026.
541 Jost, L., Archer, F., Flanagan, S., Gaggiotti, O., Hoban, S., & Latch, E. (2018). Differentiation
542 measures for conservation genetics. Evol. Appl., 11(7), 1139-1148.
543 Joubert, E., Joubert, M. E., Bester, C., de Beer, D., & De Lange, J. H. (2011). Honeybush
544 (*Cyclopia* spp.): From local cottage industry to global markets: The catalytic and supporting role
545 of research. S. Afr. J. Bot., 77(4), 887-907.
546 Kamvar, Z. N., Tabima, J. F., & Granwald, N. J. (2014). Poppr: An R package for genetic
547 analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ, 2, e281.
548 Klopfstein, S., Currat, M., & Excoffier, L. (2006). The fate of mutations surfing on the wave of a
549 range expansion. Mol. Biol. Evol., 23(3), 482-490.
550 Lemmon, A. R., Emme, S. A., & Lemmon, E. M. (2012). Anchored hybrid enrichment for
551 massively high-throughput phylogenomics. Syst. Biol., 61(5), 727-744.
552 Levin, D. A., Francisco-Ortega, J., & Jansen, R. K. (1996). Hybridization and the extinction of
553 rare plant species. Conserv. Biol., 10(1), 10-16.
554 Li, F., Niu, B., Huang, Y., & Meng, Z. (2012). Application of high-resolution DNA melting for
555 genotyping in lepidopteran non-model species: *Ostrinia furnacalis* (Crambidae). PLoS One,
556 7(1), e29664.
557 Li, M., Zhou, L., Palais, R. A., & Wittwer, C. T. (2014). Genotyping accuracy of high-resolution
558 DNA melting instruments. Clin. Chem., 60(6), 864-872.
559 Liew, M., Pryor, R., Palais, R., Meadows, C., Erali, M., Lyon, E., & Wittwer, C. (2004).
560 Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons.
561 Clin. Chem., 50(7), 1156-1164.
562 McGregor, G. K. (2017). Industry Review: An overview of the honeybush industry. Retrieved
563 from: Department of Environmental Affairs and Development Planning, Cape Town,
564 https://www.westerncape.gov.za/eadp/files/atoms/files/eadp696_an_overview_of_the_honeybus
565 h_industry_may2017_0.pdf.
566 Motsa, M. M., Bester, C., Slabbert, M. M., Hannweg, K., & Booyse, M. (2018). Flow cytometry:
567 A quick method to determine ploidy levels in honeybush (*Cyclopia* spp.). Genet. Resour. Crop
568 Evol., 65(6), 1711-1724.
569 Nei, M. (1973). Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci.,
570 70(12), 3321-3323.

571 Niemandt, M., Roodt-Wilding, R., Tobutt, K. R., & Bester, C. (2018). Microsatellite marker
572 applications in *Cyclopia* (Fabaceae) species. S. Afr. J. Bot., 116, 52-60.
573 Nunziata, A., De Benedetti, L., Marchioni, I., & Cervelli, C. (2019). High throughput measure of
574 diversity in cytoplasmic and nuclear traits for unraveling geographic distribution of rosemary.
575 Ecol. Evol., 9(7), 3728-3739.
576 Potts, A. J. (2017). Genetic risk and the transition to cultivation in Cape endemic crops: The
577 example of honeybush (*Cyclopia*)? S. Afr. J. Bot., 110, 52-56.
578 Prevosti, A., Ocana, J., Alonso, G., Ocaa, J., & Alonso, G. (1975). Distances between
579 populations of *Drosophila subobscura*, based on chromosome arrangement frequencies. Theor.
580 Appl. Genet., 45(6), 231-241.
581 QGIS Development Team (2018). QGIS Geographic Information System. Open Source
582 Geospatial Foundation Project. http://qgis.osgeo.org
583 Radvansky, J., Bazsalovicsova, E., Kralova-Hromadova, I., Minarik, G., & Kadasi, L. (2011).
584 Development of high-resolution melting (HRM) analysis for population studies of *Fascioloides*
585 *magna* (Trematoda, Fasciolidae), the giant liver fluke of ruminants. Parasitol. Res., 108(1), 201-
586 209.
587 R Core Team. R: a language and environment for statistical computing. (2018).
588 Reed, G. H., & Wittwer, C. T. (2004). Sensitivity and specificity of single-nucleotide
589 polymorphism scanning by high-resolution melting analysis. Clin. Chem., 50(10), 1748-1754.
590 Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under
591 isolation by distance. Genetics, 145(4), 1219-1228.
592 SANBI. (2019). Threatened Species Programme: SANBI Red List of South African Plants.
593 Retrieved from: http://redlist.sanbi.org/index.php.
594 Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating
595 inhibitors. Proc. Natl. Acad. Sci. U. S. A., 74(12), 5463-5467.
596 Schaal, B. A., Hayworth, D. A., Olsen, K. M., Rauscher, J. T., & Smith, W. A. (1998).
597 Phylogeographic studies in plants: Problems and prospects. Mol. Ecol., 7(4), 465-474.
598 Schaal, B. A., Hayworth, D. A., Olsen, K. M., Rauscher, J. T., & Smith, W. A. (1998).
599 Phylogeographic studies in plants: Problems and prospects. Mol. Ecol., 7(4), 465-474.
600 Schutte, A. L. (1997). Systematics of the genus *Cyclopia* Vent. (Fabaceae, Podalyrieae).
601 Edinburgh J. Bot., 54(2), 125-170.
602 Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S. B., Liu, W., Miller, J., & Small, R. L. (2005). The
603 tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for
604 phylogenetic analysis. Am. J. Bot., 92(1), 142-166.
605 Shaw, J., Lickey, E. B., Schilling, E. E., & Small, R. L. (2007). Comparison of whole chloroplast
606 genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The
607 tortoise and the hare III. Am. J. Bot., 94(3), 275-288.
608 Sillo, F., Giordano, L., Zampieri, E., Lione, G., De Cesare, S., & Gonthier, P. (2017). HRM
609 analysis provides insights on the reproduction mode and the population structure of
610 *Gnomoniopsis castaneae* in Europe. Plant Pathol., 66(2), 293-303.
611 Simko, I. (2016). High-resolution DNA melting analysis in plant research. Trends Plant Sci.,
612 21(6), 528-537.
613 Slatkin, M. (1993). Isolation by distance in equilibrium and non-equilibrium populations.
614 Evolution, 47(1), 264-279.

615   Smith, B. L., Lu, C. P., & Alvarado Bremer, J. R. (2010). High-resolution melting analysis
616   (HRMA): a highly sensitive inexpensive genotyping alternative for population studies. Mol. Ecol.
617   Resour., 10(1), 193-196.
618   Taylor, S., Scott, R., Kurtz, R., Fisher, C., Patel, V., & Bizouarn, F. (2011). A practical guide to
619   high resolution melt analysis genotyping. Retrieved from: http://www.bio-
620   rad.com/webroot/web/pdf/lsr/literature/Bulletin_6004.pdf.
621   Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity
622   of progressive multiple sequence alignment through sequence weighting, position-specific gap
623   penalties and weight matrix choice. Nucleic Acids Res., 22(22), 4673-4680.
624   Tindall, E. A., Petersen, D. C., Woodbridge, P., Schipany, K., & Hayes, V. M. (2009). Assessing
625   high-resolution melt curve analysis for accurate detection of gene variants in complex DNA
626   fragments. Hum. Mutat., 30(6), 876-883.
627   Van Der Bank, M., Chase, M.W., Van Wyk, B. E., Fay, M. F., Van Der Bank, F. H., Reeves, G.
628   & Hulme, A. (2002), Systematics of the tribe Podalyrieae (Fabaceae) based on DNA,
629   morphological and chemical data. Bot. J, 139, 159-170.
630   Vossen, R. H., Aten, E., Roos, A., & den Dunnen, J. T. (2009). High-resolution melting analysis
631   (HRMA): More than just sequence variant screening. Hum. Mutat., 30(6), 860-866.
632   Wright, S. (1943). Isolation by distance. Genetics, 28(2), 114-138.
633   Yamagata, Y., Yoshimura, A., Anai, T., & Watanabe, S. (2018). Selection criteria for
634   SNP loci to maximize robustness of high-resolution melting analysis for plant breeding.
635   Breed. Sci., 68(4), 488-498.

636

Figure 1

Sample distribution map

Study domain superimposed with the distribution of the CFRs fynbos biome, to which *Cyclopia* is endemic. Inset indicates the position of the study domain in relation to South Africa and the African continent. Distribution of samples included in non-coding cpDNA haplotype screening for HRM primer development are displayed (filled circles) in conjunction with the locations of the *C. subternata* populations included in the phylogeographic analysis (open circles). Closed circles are numbered based on species identity: *1 = C. galioides, 2 = C. genistoides, 3 = C. buxifolia, 4 = C. maculata, 5 = C. sessilifolia, 6 = C. burtonii, 7 = C. aurescens, 8 = C. bolusii, 9 = C. subternata, 10 = C. plicata, 11 = C. alpina, 12 = C. intermedia, 13 = C. longifolia, 14 = C. pubescens.*

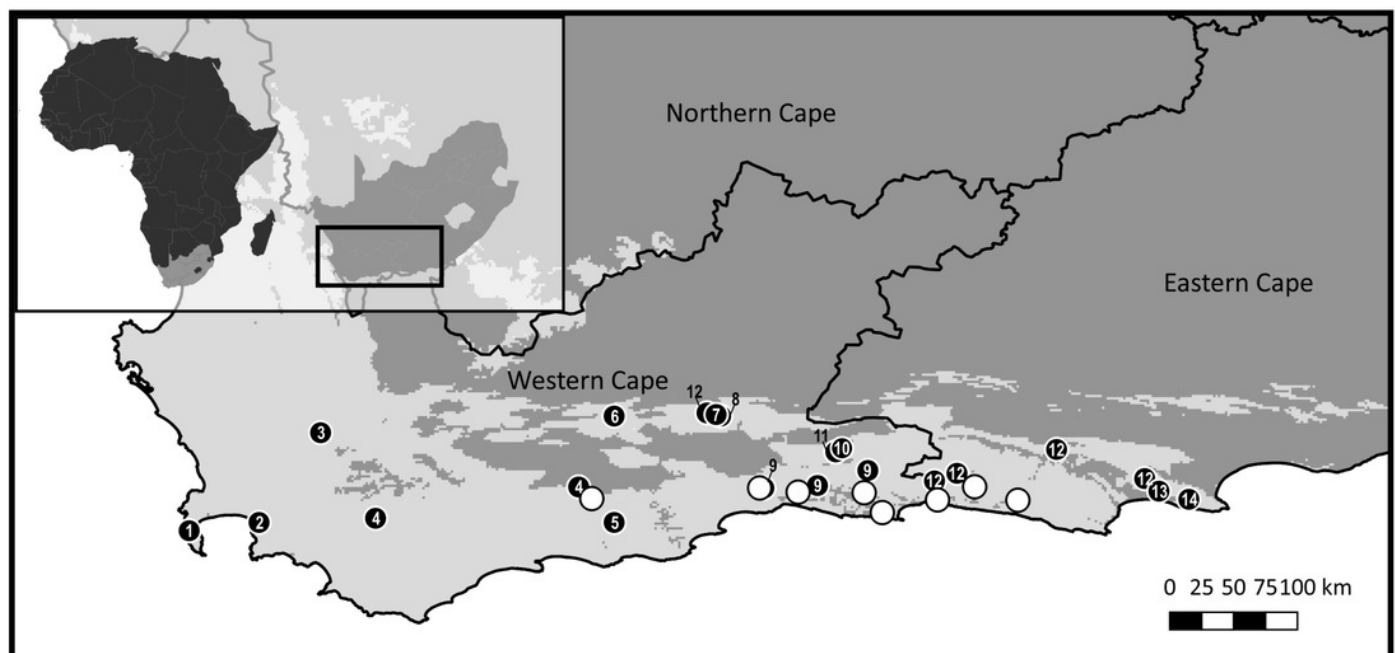



Figure 2

High Resolution Melt curve examples

Melt curves and their difference curves for the PCR products amplified by three of the genus specific primers developed. Curves are ordered in decreasing HRM clustering accuracy and the bottom curves (E,F) were generated using the primer pair MLT T1-MLT T2 (*TrnQ-5'rps16* intergenic spacer) that was excluded from HRM analysis due to poor amplification resulting in inconsistent melt curve production, the details of this primer pair, in addition to all primer pairs that were excluded from HRM haplotype discrimination analysis due to poor PCR amplification, are provided in Table S1. HRM curves (A,C,E), the normalized change in florescence associated with PCR product dissociation when heated. Melt domain identification and melt curve normalization was automated by the HRM software in this study, this process may be required to be performed manually on other platforms. A reference melt curve is selected and used as a baseline to plot melt curve differences across the melt domain, therefore difference curves (B,D,E) have different X axes. HRM clusters are automatically generated and colorised by the HRM software used. Melt curves were generated from the PCR products generated using the primer pairs, (A,B) MLT S1 - MLT S2 (*atpI-atpH* intergenic spacer), (C,D) MLT C3 - MLT C4 (*trnG-trnG2G* intergenic spacer), and (E,F) MLT T1 – MLT T2 (*trnQ-5'rps16* intergenic spacer).

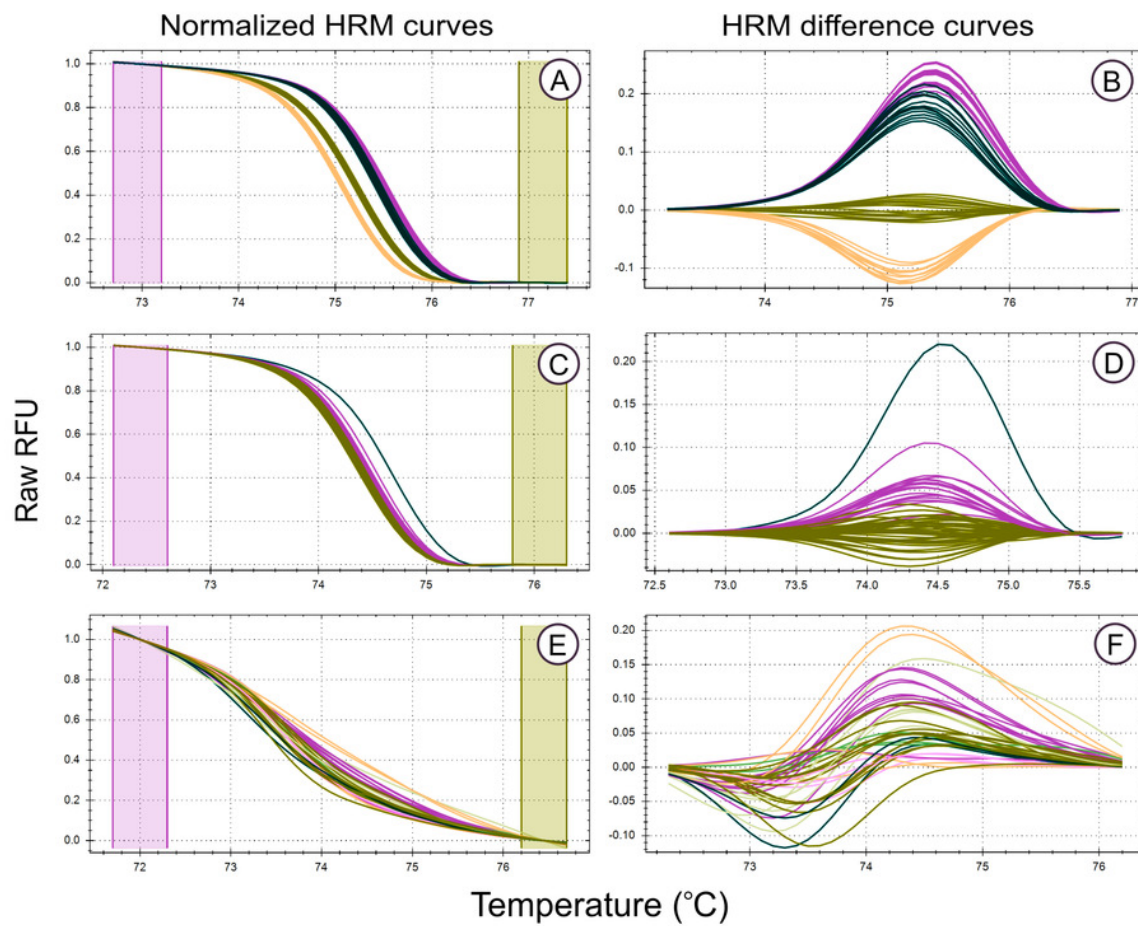



Figure 3

Framework used to developed, test, and apply HRM to the genus *Cyclopia*, a group of non-model organisms.

Step 1: Identify polymorphic sites from sequence data.

Step 2: Design primers that flank polymorphic sites.

Step 3: Test PCR amplification of HRM primers using single individual samples in replicate per haplotype. Post amplification HRM curve generation is performed on all PCR-ed asseccions, but primers consitantly failing to amplify PCR products are excluded from HRM analysis. Samples that amplified but failed to meet the PCR quality requirments occur within the red bar to the right of the amplification curve, only samples in the green were used to test HRM haplotype discrimination.

Step 5: Apply the top performing primer pairs, with the highest specificity and accuracy, to detect genetic variation in non-model organisims. Sequence only a sub set of individuals assigned to a HRM cluster for haplotype confirmation.

N	MLG	Hexp	Gst
94	6	0.20186**	0.7253 (0.1779)***
72	2	0.00456	0.02197 (0.01903)
92	6	0.06870**	0.3316 (0.3439)*
72	3	0.01324	0.02625 (0.02954)
96	2	0.00831	0.05042 (0.05522)
71	1	0	0 (0)

Step 4: Test HRM clustering of sequenced haplotypes using data generated in Step 3 above, only primers that produced consistant amplification curves are included in HRM analysis. Sensitivity, specificity and accuracy was calculated for each primer pair that produced good quality PCR amplification curves.

Figure 4

Summary of the (A) specificity, (B) sensitivity and (C) accuracy for the regions used to test haplotype discrimination by HRM.

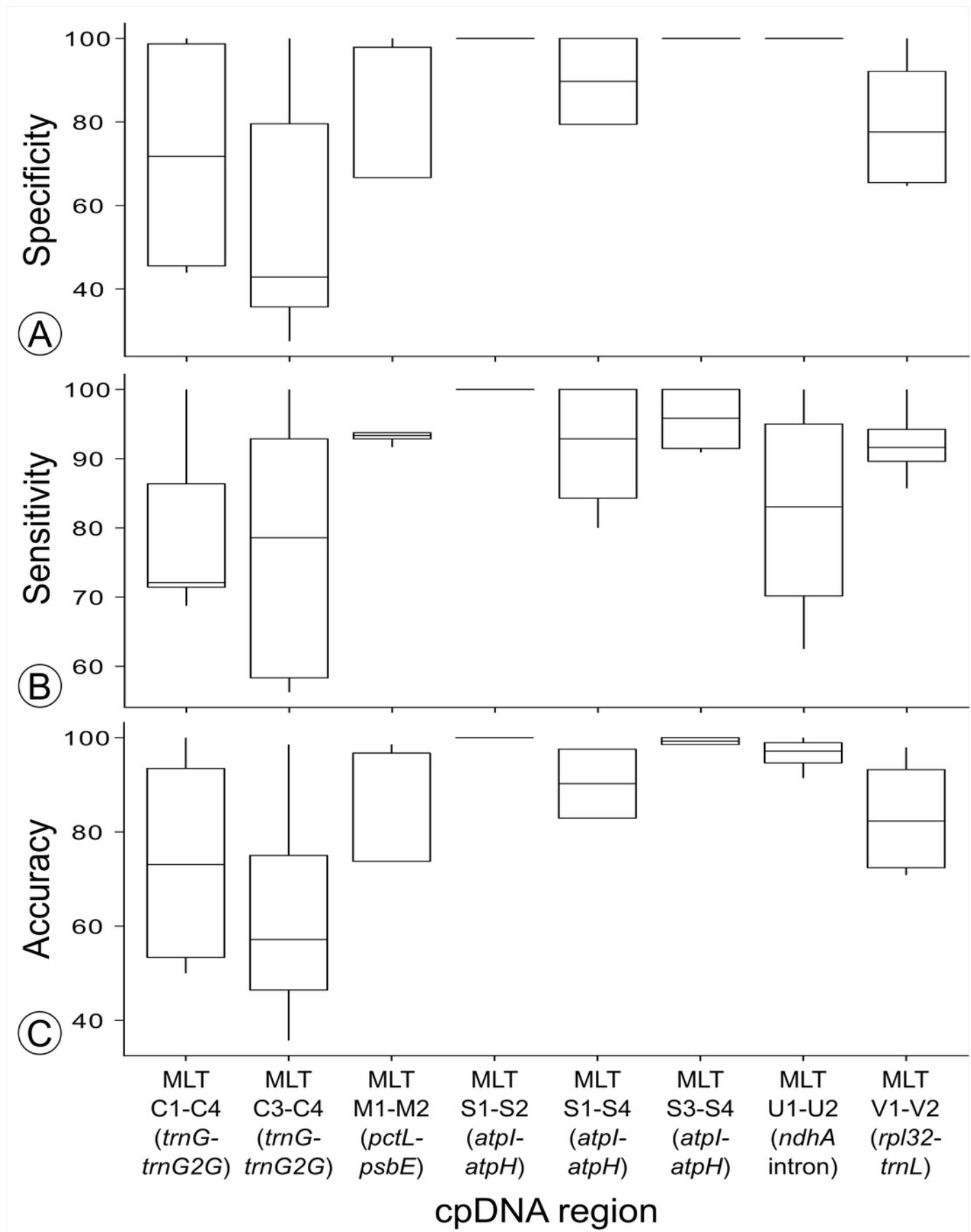
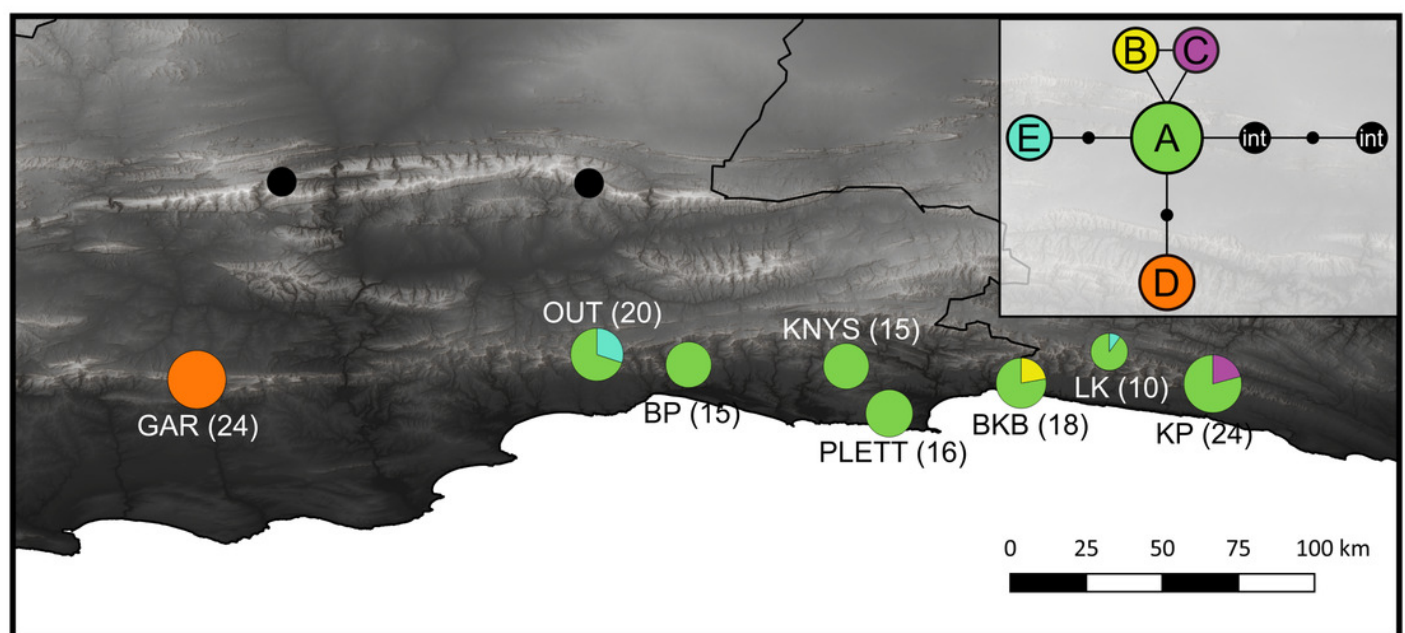



Figure 5

Haplotype distribution and number of accessions for the eight *C. subternata* populations screened via HRM.

Black circles mark *C. intermedia* samples collected from the Swartberg mountains and included as out-group taxa. Inset is the genealogical relationship between haplotypes ascertained using the Statistical Parsimony algorithm. Haplotype frequency is indicated as a proportion of the circles representing each population, with total number of accessions provided in parenthesis. The color-coding in the map corresponds to the SP network. Population naming follows the description in Table 4. GAR = Garcia's Pass located in the Langeberg, OUT = Outeniqua Pass and BP =Bergplaas MTO located in the western Outeniqua mountains, KNYS = Diepwalle Knysna and PLETT = Plettenberg Bay in the eastern Outeniqua mountains, and the BKB = Bloukrans Bridge, LK =Langkloof, and KP = Kareedouw Pass in the Tsitsikamma mountains.

Species and non-coding cpDNA regions screened for HRM primer development for the non-model plant genus *Cyclopia* Vent.

Species	rpl32-trnL intergenic spacer	ndhA intron	trnQ-50rps16 intergenic spacer	atpI-atpH intergenic spacer	petL-psbE intergenic spacer	trnD-psbM intergenic spacer	trnG-trnG2G intergenic spacer	30trnV-ndhC intergenic spacer	TrnfM-trnS	psbJ-petA	psaI-accD	psbD-trnT
				Non-coding cpDNA regions sequenced (dependent on successful amplification)								
C. alpina		X						X	X	X		
C. aurescens		X		X				X	X	X	X	X
C. bolusii		X										
C. burtonii			X	X			X	X				
C. buxifolia		X		X	X		X					
C. galioides				X								
C. genistoides		X		X			X			X		X
C. intermedia	X	X	X	X		X	X	X	X	X	X	X
C. longifolia	X		X	X	X	X	X	X	X	X	X	X
C. maculata		X						X		X	X	X
C. plicata	X	X						X	X	X	X	
C. pubescens	X	X	X	X	X	X	X	X	X	X	X	
C. sessilifolia	X	X		X	X	X	X		X	X	X	
C. subternata	X	X		X	X	X	X		X	X	X	X

Following PCR amplification, only clear, bright bands visualized through gel electrophoresis were selected for sequencing with a maximum of six and a minimum of one sample selected for sequencing per species.

Cyclopia specific primers designed for testing HRM haplotype discrimination

Primers used to screen haplotype variation in wild *C. subternata* populations are indicated in bold. Primer details provided include; non-coding cpDNA region the primers are located in, as well as each primers' annealing temperature (Tm) in degrees Celcius, GC content, and sequence motif.

Region	Primer	Direction	Tm (°C)	GC (%)	Sequence (5'→3')
trnG intron	MLT_C1	F	57.3	43.5	ACTCCTCTTCTATTCATGGGGA
	MLT_C2	R	58.0	50.0	
	MLT_C3	F	61.8	40.9	TCAACGAACGATTCGAGGAATA
	MLT_C4	R	61.1	45.5	TGCTTCAATCTCTCCTACCCAA
pctL-psbE intergenic spacer	MLT_M1	F	58.0	43.5	TGTCGAGAACCCTTATACTCTCA
	MLT_M2	R	58.7	47.6	TACCAAGGGTGTCTTTCGAGT
atpI-atpH intergenic spacer	**MLT_S1**	**F**	64.3	50	ATTACAGATGAAACGGAAGGGC
	MLT_S2	**R**	61.5	45.5	TGGGGGTTTCAAAGCAAAGG
	MLT_S3	**F**	61.5	45.5	CCTTTGCTTTGAAACCCCCA
	MLT_S4	**R**	66.4	36.5	TTCCCGTTTCATTCATTCACATTCA
ndhA intron	**MLT_U1**	**F**	59.1	40.0	AGGTACTTCTGAATTGATCTCATCC
	MLT_U2	**R**	62.2	52.4	GCAGTACTCCCCACAATTCCA
rpl32-trnL intergenic spacer	MLT_V1	F	59.9	60.0	CTCCTTCCCTAAGAGCAGCG
	MLT_V2	R	59.2	40.0	GTTGGAATAATCTGAATTAGCCGGA

Nucleotide differences and clustering results for HRM discrimination of known haplotype

Sample ID of the accessions that were PCR amplified in replicates of 16, the number of replicates that successfully amplified during PCR and subject to HRM analysis is given (N), followed by HRM haplotype discrimination (sensitivity, specificity and accuracy), the grouping of each replicate into a HRM cluster is provided for each haplotype amplified per primer combination (clusters 1 - 11), a summary of the nucleotide differences between haplotypes is also provided.

Primer pair	N	Sen	Spe	Acc	HRM grouping of replicates into cluster 1-cluster 11											Nucleotide difference summary			
					1	2	3	4	5	6	7	8	9	10	11				
MLT C1-MLT C4 (150 bp) (TrnG intron)																19	20	72	205
																T	A	T	A
Haplotype A	14	71	94	88	2	10		2								G	T	G	.
Haplotype B	16	69	44	52	11		4		1							.	.	.	.
Haplotype C	11	91	49	58	10	1										.	.	.	C
Haplotype D	11	73	44	50	8	1	3									.	.	G	.
MLT C3-MLT C4 (236 bp) (TrnG intron)																41	48-55	62	
																T	#	A	
Haplotype A	14	100	43	57	14											.	#	T	
Haplotype B	16	56	27	36	9	6	1									.	-	.	
Haplotype C	12	58	79	75	4	7		1								G	#	.	
Haplotype D	14	79	36	46	11	3										.	#	.	
																# = AAAAATTG			
MLT M1-MLT M2 (170 bp) (pctL-psbE intergenic spacer)																84	88	110	118
																G	G	G	A
Haplotype A	15	93	98	97		14	1									A	.	.	.
Haplotype B	16	94	67	74	15			1								.	.	.	.
Haplotype C	14	93	98	97		1	13									.	T	.	.
Haplotype D	16	94	67	74	15			1								.	.	T	G
MLT S1-MLT S2 (217 bp) (atpI-atpH intergenic spacer)																53	54	62-80	95

								T	A	-	G
Haplotype A	12	100	100	100		12		.	.	-	.
Haplotype B	15	100	100	100	15			.	.	-	A
Haplotype C	11	100	100	100			11	A	C	-	.
Haplotype D	14	100	100	100		14		A	C	#	.

\# = TTCATAGATAACTAGTTAG

MLT S1-MLT S4 (527 bp) (atpI-atpH intergenic spacer)

									75	76	86-104	117	267	281	287	382	477-481
									T	A	-	G	C	C	T	C	*
Haplotype A	14	100	79	83	14				.	.	-	.	T	.	.	.	-
Haplotype B	10	80	100	98			8	2	.	.	-	A	.	.	.	.	-
Haplotype C	14	100	79	83	14				.	.	-	.	T	.	.	.	*
Haplotype D	14	86	100	98		12		2	A	C	#	.	.	T	G	.	*
Haplotype E	16	100	100	100	16				.	.	-	A	.	.	.	A	-
Haplotype F	14	71	100	95			10	2 2	A	C	-	.	.	T	G	.	*

\# = CATAGATAACTAGTTAGTT, * = TTTTC

MLT S3-MLT S4 (310 bp) (atpI-atpH intergenic spacer)

								52	66	72	167	262-266
								C	C	T	C	-
Haplotype A	15	100	100	100	15			T	.	.	.	-
Haplotype B	12	92	100	95		11	1	.	.	.	.	-
Haplotype C	11	91	100	95		10 1		T	.	.	.	#
Haplotype D	16	100	100	100	16			.	T	G	.	#
Haplotype E	14	93	100	95		13	1	.	.	.	A	-

\# = TTTTC

MLT U1-MLT U2 (345 bp) (ndhA intron)

								15	22	47	79	135-141	149	172	183	220	253	289
								T	T	G	C	#	C	A	A	G	T	A
Haplotype A	16	100	100	100	16			.	.	.	.	#	.	.	.	.	.	.
Haplotype B	11	73	100	96		8	3	.	C	.	A	#	A	C	G	.	G	T

Haplotype C	15	93	100	99		14						1	C	.	.	.	#	.	.	.	.	.	.
Haplotype D	16	63	100	91				10	3		2	1		.	.	A	.	#	.	.	.	.	.	.
Haplotype E	12	92	100	99		11				1			.	.	A	.	-	.	.	.	A	.	.

= TATCCCC

MLT V1-MLT V2 (340 bp)
(rpl32-trnL intergenic spacer)

							34-38	56	104
							#	T	T
Haplotype A	14	86	65	71	12	2	-	.	.
Haplotype B	10	91	99	98		10	#	.	.
Haplotype C	12	92	66	73	12		#	A	A
Haplotype D	10	100	89	92		10	#	A	.

= ATTATT

Protocol for PCR amplification and subsequent HRM curve generation. Primer specific annealing temperatures (Tm) are provided in Table 2.

2

Process	Step	Temperature	Time	Number of Cycles
PCR Amplification	Initial Denaturing	95°C	2 min	1
	Denaturing	95°C	10 sec	
	Annealing/Extension + Plate Read	**Primers mean Tm**	30 sec	40
	Extension + Plate Read	72°C	30 sec	
HRM Analysis	Heteroduplex Formation	95°C	30 sec	1
		60°C	1 min	1
	HRM + Plate Read	65–95°C (in 0.2°C increments)	10 sec/step	1

Cyclopia subternata population locations

The geographic co-ordinates, number of accessions screened via HRM, and haplpotype frequencies (as detected by HRM and verified by sequencing) are given for each *C. subternata* population. Nucleotide differences among haplotypes are provided in Table S3.

J

2
3
4
5
6
7
8

Mountain range	Population	Co-ordinates		N	Haplotype				
		S	E		A	B	C	D	E
Langeberg	Garcia's pass (GAR)	-33.96	21.22	24	-	-	-	24	-
Outeniqua W	Outeniqua Pass (OUT)	-33.88	22.40	20	14	-	-	-	6
	Bergplaas MTO (BP)	-33.91	22.67	15	15	-	-	-	-
Outeniqua E	Diepwalle, Knysna (KNYS)	-33.92	23.14	15	15	-	-	-	-
	Plettenberg bay (PLETT)	-34.06	23.26	16	16	-	-	-	-
Tsitsikamma	Bloukrans Bridge (BKB)	-33.97	23.65	18	14	-	4	-	-
	Langkloof (LK)	-33.87	23.91	10	9	-	-	-	1
	Kareedouw pass (KP)	-33.97	24.22	24	19	5	-	-	-

Nucleotide varition not differentiated by HRM

2
3
4
5
6

Primers	Haplotypes	Nucleotide difference	Specificity
MLT C1 - MLT C4 (TrnG intron)	C-D	T ↔ G & C ↔ A	18
	B-C	A ↔ C	20
	B-D	T ↔ G	29
	A-C	GT ↔ TA, G ↔ T & T ↔ A	88
	A-D	G ↔ T & T ↔ A	88
	A-B	GT ↔ TA & G ↔ T	93
MLT C3 - MLT C4 (TrnG intron)	A-D	T ↔ A	11
	A-B	8 bp indel & T ↔ A	22
	B-D	8 bp indel	33
	B-C	T ↔ G & 8 bp indel	65
	C-D	G ↔ T	73
	A-C	T ↔ G & T ↔ A	83
MLT M1 - MLT M2 (pctL-psbE intergenic spacer)	A-C	G ↔ T & A ↔ G	6
	B-D	A ↔ G & G ↔ T	93
MLT S1 - MLT S4 (atpI-atpH intergenic spacer)	A-C	5 bp indel	0
MLT V1 - MLT V2 (rpl32-trnL intergenic spacer)	A-C	6bp indel, T ↔ A & T ↔ A	11
	A-D	6bp indel & T ↔ A	93
	C-D	A ↔ T	96