Manuscript Click here to download Manuscript manuscript_text.pdf

Isolation by distance and isolation by environment contribute to population differentiation in *Protea repens* (Proteaceae L.), a widespread South African species[1]

R. Prunier [2,7], M. Akman[3], C. T. Kremer[4], N. Aitken[5], A. Chuah[5], J. Borevitz[5], K. E. Holsinger [6]

[1] Manuscript received _____; revision accepted _____.

[2] *Department of Biological and Environmental Sciences, Western Connecticut State University, 190 White Street, Danbury, CT 06810, USA*

[3] *Department of Plant Sciences, University of California Davis, 387 N. Quad Ave, Davis, CA 95616, USA*

[4]*Department of Ecology and Evolutionary Biology, Yale University, P.O. Box 208106, New Haven, CT 06520-8106, USA*

[5] *Research School of Biology, Australian National University, Canberra, ACT 2601, Australia*

[6] *Department of Ecology and Evolutionary Biology, University of Connecticut, 75 N. Eagleville Rd., U-3043, Storrs, CT 06268, USA*

[7]Corresponding author at: Department of Biological and Environmental Sciences, Western Connecticut State University, 190 White Street, Danbury, CT 06810, USA. Tel.: +1 203 837 8233; fax: +1 203 837 8769. *E-mail address:* prunierr@wcsu.edu

Running title: Population differentiation and natural selection in *Protea repens*

ABSTRACT

Premise — The Cape Floristic Region (CFR) of South Africa is renowned for its botanical diversity, but the evolutionary processes that underlie this diversity remain controversial. Both neutral and adaptive processes have been implicated in driving diversification, but population-level studies of plants in the CFR are rare. Here, we investigate the limits to gene flow and potential environmental drivers of selection in *Protea repens* L. (Proteaceae L.), a widespread CFR species.

Methods — We sampled nineteen populations across the range of *Protea repens* and used genotyping by sequencing (GBS) to identify 2066 polymorphic loci in 663 individuals. We used a Bayesian F_{ST} outlier analysis to identify SNPs marking genomic regions that may be under selection and excluded these SNPs from analyses of gene flow and genetic structure and used those SNPs to identify potential drivers of selection.

Results — We identified a pattern of isolation by distance, suggesting that there is limited gene flow between nearby populations. We also determined that the populations of *Protea repens* fall naturally into two or three groupings, which correspond to an east-west split. We also found that differences in rainfall seasonality contribute to diversification in highly divergent loci.

Conclusions — The strong pattern of isolation by distance is in contrast to the findings in the only other widespread species in the CFR that has been similarly studied, while the effects of rainfall seasonality are consistent with well-known patterns. Assessing the generality of these results will require investigations of other CFR species.

Keywords — Cape Floristic Region, F_{ST} outlier, GBS, isolation by distance, isolation by environment, local adaptation, phylogeography, *Protea repens*, Proteaceae

53 INTRODUCTION

54 The Cape Floristic Region (CFR) of southwestern South Africa is renowned for its
55 spectacular botanical diversity (>9000 plant species; Myers et al., 2000; Goldblatt and
56 Manning, 2002), but the evolutionary processes that gave rise to this diversity remain
57 controversial (Linder, 2003). Both adaptive and non-adaptive (i.e., drift and vicariance)
58 processes have been implicated (Adamson, 1958; Linder, 1985; Johnson, 1996; van der
59 Niet and Johnson, 2009; Ellis et al., 2014) because of strong environmental gradients and
60 extreme topographic diversity in the region. In reality, both of these processes are likely
61 to contribute to diversification among populations and species. However, despite a wealth
62 of studies investigating evolutionary processes above the species level (e.g. Goldblatt et
63 al., 2002; Sauquet et al., 2009; Verboom et al., 2009, 2014; Valente et al., 2010; Pirie et
64 al., 2016), very little is known about the evolutionary processes driving differentiation
65 within plant species in the CFR (but see Prunier and Holsinger 2010; Rymer et al., 2010;
66 Lexer et al. 2014).

67 CFR an important environment in which to study the balance adaptive and non-
68 adaptive diversification processes because unlike other hyper diverse regions, diversity in
69 the CFR is primarily the result of high differentiation among sites (beta diversity) rather
70 than high diversity within sites (alpha diversity; Latimer et al., 2005). High beta diversity
71 indicates that there is substantial species turnover from location to location, often along
72 environmental gradients (Whittaker, 1960), and it implies that many species have narrow
73 ranges. Indeed, most CFR species are limited either geographically (e.g., to one section of
74 a mountain range) or environmentally (e.g., to one portion of an environmental gradient).
75 For example, *Protea aurea* (Proteaceae) has a relatively broad geographic range, but is

found only on sites with shale soils (Rebelo, 2001). Other species are restricted to limestone-derived soils (e.g. *Protea obtusifolia*; Cowling et al., 1988). Environmental gradients abound in the CFR. For example, rainfall varies both in amount (60-3300 mm/year) and in timing (entirely in the winter; equally likely all year; or primarily occurring in the summer; Schultze, 2007). Soils are also highly variable and range from acidic nutrient-poor quartz-based soils to richer shale-based soils to calcareous limestone-based soils (Schultze, 2007). Further, the mountains of the Cape Fold Belt (max elev. 2230 meters) create steep elevational gradients along which temperature, rainfall, and solar irradiation vary over short distances (Schultze, 2007).

Arguments favoring adaptation and ecological speciation as the primary mechanism of diversification in the CFR focus on the narrow ranges and ecological specialization of many CFR species along strong climate, topographic, and edaphic (e.g. Linder, 1985; Goldblatt and Manning, 2002; van der Niet and Johnson, 2009). For example in the *Disa draconis* (Orchidaceae) species complex, local adaptation to pollinators seems to be driving reproductive isolation between populations (Johnson and Steiner, 1997). Similarly, in *Pelargonium* section *Hoarea* (Geraneaceae), a diverse group of geophytes, species are ecologically similar, but differ in floral morphology (Gibby et al., 1996), suggesting that speciation is a result of sexual selection by pollinators. In other groups, local adaptation has been identified both within and between species (Ellis and Weis, 2006; Carlson et al., 2011; Prunier et al. 2012), though whether local adaptation contributed to geographical isolation or evolved as a result of geographical isolation is unclear.

98 Local adaptation that contributes to geographical isolation may lead to ecological
99 speciation, and incipient ecological speciation may be revealed by a pattern of isolation
100 by environment (IBE; Nosil et al., 2009). If isolation by environment is detected at loci
101 presumed to be neutral, it suggests that genetic differentiation is occurring outside the
102 "gene islands" responding to selection (Rieseberg and Burke, 2001; Wu, 2001), and that
103 ecological differences are limiting gene flow between populations. Such a pattern might
104 arise through reduced migration or establishment in environments different from the
105 source population (Wang and Bradburd 2014).

106 Non-adaptive differences may also play a significant role in diversification, due to
107 drift and vicariance that occurs when when populations of a widespread species become
108 isolated in pockets of suitable habitat (Kozak and Wiens, 2006). Vicariance has long been
109 hypothesized as an important factor associated with diversification in the CFR (Adamson,
110 1958). In particular, isolation seems to have been important in the sedge genus *Tetraria*
111 (Cyperaceae; Britton et al., 2014; Verboom et al., 2015) and in *Protea* section *Exsertae*
112 (Prunier and Holsinger, 2010). High topographic diversity is associated with vicariant
113 diversification in other areas (e.g., Knowles, 2001), and Verboom et al. (2015) proposed
114 that in the CFR, vicariance plays a larger role in speciation at high elevations because of
115 both isolation and historical climate fluctuations.

116 We investigate patterns of population differentiation in *Protea repens* (Proteaceae L.),
117 a widespread CFR species, using anonymous SNPs identified through genotyping by
118 sequencing (GBS). We identify SNPs marking genomic regions that may be subject to
119 diversifying selection, and we use the remaining SNPs, presumed to be neutral, to
120 understand phylogeographic patterns. We also investigate the relative contributions of

121 isolation by distance and isolation by environment in all SNPs and compare them
122 between those presumed to neutral and those that may be responding to differential
123 natural selection. Specifically, we address four questions. (1) Can we identify genomic
124 regions that may be targets of selection? (2) What genetic groupings are present within *P.*
125 *repens,* and are they associated with geographic boundaries that define distributions in
126 other species? (3) To what extent is genetic differentiation between populations
127 associated with their physical distance and environmental differences? (4) Do these
128 patterns differ between genomic regions that are more or less likely to be targets of
129 selection?
130
131 MATERIALS AND METHODS
132 ***Study system*** — *Protea repens* is a widespread species native to the Cape Floristic
133 Region in southwestern South Africa. This large, sclerophyllous shrub is found across the
134 CFR and into the thicket biome of the Eastern Cape. It is found from sea level to nearly
135 2000 meters (mean elevation 752 meters, Appendix S1, see Supplemental Data with the
136 online version of this article). It is the most abundant member of the genus *Protea*
137 (Rebelo, 2001) and is common throughout its range, either as scattered plants or in dense
138 stands. *Protea repens* is a relatively old species, estimated to have split from its sister
139 group, which contains 48 species, about 12 million years ago (Schnitzler et al., 2011).
140 ***Population sampling*** — We used seeds that were collected from between 5 and 33
141 maternal lines from each of 19 *Protea repens* populations broadly distributed across the
142 species' range (Table 1, Fig. 1). Collaborators germinated the seeds and planted the
143 seedlings into a common garden at Kirstenbosch Botanical Gardens in Cape Town, South
144 Africa (see Akman et al. 2015; Carlson et al. 2015). We then collected leaf samples from

145 717 individuals in the garden, and stored them in a CTAB and sodium chloride buffer for
146 preservation. Vouchers for each population were deposited in CONN.

147 ***Genotyping By Sequencing (GBS)*** —We extracted genomic DNA using a modified
148 CTAB extraction (Doyle and Doyle, 1987), first washing the leaf samples to remove the
149 CTAB and sodium chloride buffer, then grinding with a mortar and pestle. We prepared
150 GBS libraries following the protocol described in Nicotra et al. (2016) using the
151 restriction enzyme PstI (modified from Elshire et al., 2011; Morris et al., 2011). We gave
152 each sample a forward and reverse barcodes (four - eight bp long) such that each sample
153 had a unique combination and could be multiplexed. We used three lanes of 100 bp
154 paired-end Illumina HiSeq 2000 to sequence our samples. The 717 individuals were first
155 split across two lanes. However, there was poor coverage for a majority of the
156 individuals, so we reran 672 individuals on another lane to improve coverage. After
157 sequencing, we joined the barcodes from either end of the paired-end reads to make one
158 complete barcode that we reattached to both reads. Each read was subsequently treated as
159 single end with the TASSEL UNEAK pipeline (version 3.0; Lu et al., 2013). Specifically,
160 reads with a complete barcode, a PstI cut site, and no Ns in the 64 bp following the cut
161 site (Lu et al., 2013) were used.

162 ***SNP Calling —*** We identified SNPs using the TASSEL UNEAK pipeline (version
163 3.0; Lu et al., 2013). Reads with exactly one bp mismatched were considered candidate
164 SNPs. We used the network filter to discard complicated networks of reads that each
165 varied by only one bp mismatch, thus identifying reciprocal read pairs using default
166 settings. Each read contains either zero or one SNP. The UNEAK pipeline identified
167 112064 candidate SNPs, the vast majority of which were represented in only the few

168 individuals that had the most reads (Fig. 1 in Appendix S2). We filtered the candidate
169 SNPs three ways to ensure that they would be informative for population genetic
170 analyses. First, we removed any SNP calls that had lower than 5X coverage (per
171 individual per called locus, Fig. 1 in Appendix S2). Thus, assuming no bias in SNP
172 amplification within an individual, there is less than 1 chance in 32 (approximately 3%)
173 that an individual scored as homozygous is actually a heterozygote for which only one
174 chromosome was sequenced. This lowered the total number of SNP calls from ~3.5
175 million to ~1.25 million calls out of 74.3 million possible calls (if each individual was
176 called at each locus). After removing low coverage loci, we removed individuals that had
177 fewer than 500 total reads (out of more than 57 million reads across all individuals),
178 leaving 663 individuals in the data set. This cutoff is low, leaving some individuals that
179 were only called for a few loci and had few reads, but most of the individuals (631) had
180 more than 10,000 reads and were called at many loci (Table 1 in Appendix S2). Of the
181 remaining SNPs and individuals, we removed SNPs that were not called in at least 20%
182 of the remaining individuals from the analysis, resulting in a final set of 2,066 SNPs in
183 our analysis. Of the 2,066 SNPs, more than half were called for at least 400 individuals,
184 and of the 663 individuals, the vast majority was called for more than 1,000 SNPs (Figs.
185 2, 3 and 4 in Appendix S2).

186 While there may be small amounts of gametic disequilibrium among some of these
187 loci, our statistical analyses depend either on methods relying on genotype counts (F_{ST}
188 outliers and population phylogeny) that are largely unaffected by gametic disequilibrium
189 (Guo et al. 2009) or on methods that implicitly account for gametic disequilibrium
190 (individual assignment and identity by descent).

191 ***Identification of F_{ST} outliers*** —Locus-specific estimates of F_{ST} can be used to
192 identify regions of the genome that may be subject to selection (Beaumont and Balding
193 2004). We used a Bayesian genome scan to identify F_{ST} outlier loci. Although genome
194 scans in humans have identified genes apparently subject to selection even without an
195 association with known phenotypes (e.g., Akey, 2009), widely used methods (BayeScan,
196 FDIST) assume an island model of migration and are prone to false positives (Excoffier et
197 al., 2009; Lotterhos and Whitlock, 2014). To avoid these problems we used the Bayesian
198 method described by Guo et al. (2009).
199 This method is a conservative approach to detecting outliers, both because it uses the
200 observed genomic distribution of F_{ST} as a basis of comparison and thus avoids specific
201 demographic assumptions and because missing data "shrinks" point estimates of locus-
202 specific F_{ST} to the genome wide mean F_{ST} and widens the posterior distributions, causing
203 us to lose power for detecting outliers. We also used a very stringent criterion to identify
204 outliers ($p < 0.0001$). Even with a much less stringent criterion ($p < 0.05$) this method has
205 a low rate of false positives (Guo et al., 2009). As an empirical check on the frequency of
206 false positives, we produced six data sets in which individuals were randomly shuffled
207 among populations. We found no F_{ST} outliers at any of the 2066 loci included in the
208 analysis, which indicates that the upper 95% confidence limit on the false positive rate
209 for this analysis is less than 4.2×10^{-6}. The analysis was implemented in JAGS v3.4.0
210 (http://mcmc-jags.sourceforge.net/). Code is available at (Holsinger, 2016).
211 SNPs identified as high outliers have allele frequencies that differ more than expected
212 between populations, and could reflect diversifying selection. SNPs identified as low
213 outliers have allele frequencies that differ less than expected between populations and

214 could reflect stabilizing selection across populations. However, low outliers might also be
215 artifacts arising from treating fixed heterozygosity in paralogous loci as allelic
216 heterozygosity in orthologous loci. Because of this uncertainty, we do not investigate
217 these low outliers, and we removed them from remaining analyses. We treat loci not
218 identified as outliers as presumably neutral.

219 ***Identifying SNPs in Known Genes*** — To determine whether the SNPs (both the non-
220 outlier and the high F_{ST} outliers) were in known genes, we used BLAST to map the 64 bp
221 long sequences to the draft *Protea repens* leaf transcriptome (Akman et al., 2015). This
222 draft transcriptome was constructed with sequence data from individuals representing the
223 populations used in this study and was annotated using sequence similarity to
224 *Arabidopsis thaliana* EST libraries. We identified the SNPs that mapped to the
225 transcriptome with an e-value less than 10^{-10}.

226 ***Population Structure*** —

227 *A. Distribution of Genetic Diversity* — For each population, we estimated nucleotide
228 diversity and the percentage of loci that are polymorphic in Arlequin 3.5.1.2 (Excoffier
229 and Lischer, 2010) for both the non-outlier and the high F_{ST} outlier loci. We also used
230 AMOVA (implemented in Arlequin 3.5.1.2; Excoffier and Lischer, 2010) to partition
231 genetic diversity within and among populations for both the non-outlier and the high F_{ST}
232 outlier loci.

233 *B. Individual Assignment* —We used an individual-based clustering approach
234 implemented in STRUCTURE (version 2.3.4; Pritchard et al., 2000; Falush et al., 2003) to
235 determine how individuals are grouped into larger genetic clusters. To perform this
236 analysis, we used only the non-outlier SNPs. We performed ten runs of 200,000 iterations

237 with 100,000 iterations discarded as a burn-in at Ks ranging from 1 to 19. We used
238 STRUCTURE HARVESTER (version web 0.6.94; Earl and vonHoldt, 2012) to generate
239 CLUMPP input files and to calculate delta K (Evanno et al., 2005), a statistic often used to
240 identify the number of clusters most useful for interpretation. We used CLUMPP (version
241 1.1.2; Jakobsson and Rosenberg, 2007) to combine the results of our 10 runs at each K
242 and visualized the results using DISTRUCT (version 1.1; Rosenberg, 2003).

243 *C. Multidimensional Scaling* — As a complement to the analysis in STRUCTURE, we
244 used multidimensional scaling of pairwise identity by descent to visualize the
245 relationships among individual genotypes. We approximated the probability of identity
246 by descent between individuals *i* and *j* in our sample as

247 $$f_{ij} = \frac{1}{K}\sum_{k=1}^{k}\left(p_{ik}p_{jk} + \left(1 - p_{ik}\right)\left(1 - p_{jk}\right)\right) ,$$

248 where p_{ik} is the frequency of the major allele at locus *k* in individual *i* (1 for an individual
249 homozygous for the major allele at this locus, ½ for an individual heterozygous at this
250 locus, 0 for an individual homozygous for the minor allele at this locus; compare
251 Patterson et al., 2006). We used cmdscale() in R v3.0 (R Core Development Team, 2014)
252 to perform the multidimensional scaling.

253 *D. Population Phylogeny* — We used TREEMIX (version 1.12; Pickrell and Pritchard,
254 2012; https://bitbucket.org/nygcresearch/treemix/wiki/Home) to estimate the sequence of
255 population splits. TREEMIX is widely used to estimate population relationships in humans
256 in the presence of admixture (e.g., Meyer et al. 2012; Patterson et al. 2012; Lazardis et al.
257 2014; Raghavan et al. 2014). To provide estimates of the support for branches we
258 produced 5,000 bootstrap samples using the –bootstrap option. We report the majority-

259 rule consensus tree and placed the root of the phylogeny between the main groups
260 identified by both STRUCTURE and multidimensional scaling.

261 ***Drivers of Genetic Differentiation — IBD vs. IBE***— To determine how strongly
262 genetic differentiation between populations (isolation) is driven by physical distance
263 (IBD), environmental differences (IBE), and historical barriers to gene flow we examined
264 differentiation at each of the 2,066 loci using Bayesian linear mixed models implemented
265 in the MCMCglmm package (version 2.23; Hadfield, 2010). This locus-by-locus
266 approach allows us to explore how the effect size of each covariate changes across loci,
267 and ultimately to determine if outlier and non-outlier loci respond differently to these
268 environmental gradients. This analysis and those that follow were implemented in in R
269 v3.3.1 (R Core Development Team, 2016). Data and R code are available at
270 http://github.com/ctkremer/P_repens.

271 The response variable in these models is the genetic distance between population
272 pairs. We calculated pairwise genetic distances between populations for each locus
273 individually with the allele frequency differential delta (Gregorius and Roberds, 1986).
274 The predictor variables are physical distance, environmental differences, and whether or
275 not two populations co-occur in the same phytogeographic zone. Because we examined
276 the pairwise genetic distances of populations, observations of our response variable are
277 not independent (they may share a population in common). We account for this lack of
278 independence using a random effect to account for multiple membership (Browne et al.,
279 2001; Fielding and Goldstein, 2006); any population specific effects are absorbed into
280 this random effect. While this approach fails to take into account the covariance in allele
281 frequencies among closely related populations (Coop et al., 2010), because the internal

282 branches of our population phylogeny (Fig. 2B) are extremely short (relative to the
283 terminal branches), the error introduced by this approximation is small.
284 To test for presence and strength of isolation by distance and isolation by
285 environment we calculated the distance between populations in physical and
286 environmental space. Specifically, for isolation by distance we determined the Euclidean
287 physical distances between populations, calculated using the earth.dist() command in the
288 fossil package (version 0.3.7; Vavrek, 2011). For examining isolation by environment,
289 we needed to calculate environmental differences between populations. We based these
290 differences on seven environmental variables that have previously been shown to be
291 associated with physiologically significant trait variation in this and other *Protea* species
292 (Carlson et al., 2011, Prunier et al., 2012, Carlson et al., 2015). These variables are:
293 rainfall concentration (average intra-annual variation in precipitation), mean annual
294 precipitation (MAP), average January maximum temperature, mean annual temperature,
295 altitude, average July minimum temperature, and the amount of rainfall that falls in the
296 summer months. For all known populations, including the 19 featured in thus study, of
297 *Protea repens* (Protea Atlas Project; Rebelo, 2006), we extracted all of the variables
298 except for rainfall concentration from the 2007 Atlas of Climatology and Agrohydrology
299 (Schultze, 2007). Due to boundary artifacts in the rainfall concentration variable in the
300 2007 Shultze dataset, we extracted rainfall concentration from the 1997 Schultze dataset
301 (Schultze, 1997).
302 Next, to formulate orthogonal environmental axes, we performed a PCA using the
303 princomp() function in the base R package (R Core Development Team, 2016) on all
304 seven variables. Across all *Protea repens* populations, the first three axes of the PCA

305 explained 90 percent of the environmental variation. PC1 explained 47% of the variation
306 and was mainly driven by differences in temperature (hereafter PC1-temp); PC2
307 explained 23% of the variation, and was driven by the proportion of rainfall that falls in
308 the winter (hereafter PC2-winter rain); and PC3 explained 20% of the variation and was
309 driven by total rainfall (hereafter PC3-MAP), (Appendix S3). We extracted values for the
310 first three PC axes for the 19 populations studied here, and calculated the distances
311 between the populations on the three PC axes. Both the calculated physical distances and
312 distances along these three PC axes were standardized to have a mean of zero and
313 standard deviation of one, allowing their relative effect sizes to be compared.

314 Finally, to determine the effect of historical barriers to gene flow on current patterns
315 of genetic similarity we scored the pairs of populations for coexistence in the
316 phytogeographic zones defined by Goldblatt and Manning (2002). Phytogeographic
317 zones are regions that share an endemic flora and, as such, might reflect barriers that are
318 not explained by either physical distance or measured environmental differences.
319 Population pairs were either scored as 1 (different phytogeographic zone) or 0 (same
320 zone). This covariate completed the suite of variables we used to examine genetic
321 differentiation between populations.

322 Models were run using the MCMCglmm() function. For all models, the link function
323 was identity and the Gaussian distribution was used for the error distribution. We used
324 the standard priors of $MCMC_{GLMM}$ and a burn-in of 500,000 iterations followed by
325 500,000 iterations thinned every 750 iterations. For a random subset of the loci we
326 visually inspected trace plots of the MCMC chains, observing that they were well mixed.

327 Each model fit provides coefficients that estimate how strongly differences in a locus'
328 allele frequencies between populations are related to differences in five factors (physical
329 distance, location in the same phytogeographic zone, and three environmental variables).
330 We are interested in how these relationships at the locus level differ across loci, and
331 especially between the F_{ST} outlier loci and the non-outlier loci. We used a generalized
332 additive model (GAM) to quantify trends in estimated coefficients as a function of locus-
333 specific F_{ST} values (see identifying F_{ST} outliers above), because GAMs allowed us to
334 avoid making explicit parametric assumptions about the shape of these trends. Rather
335 than run separate GAMs for the outlier and non-outlier loci, we ran a single GAM for all
336 loci. Inferences were based on the relationships with the F_{ST} of each locus, rather than
337 their status as outliers as determined in the F_{ST} outlier analysis above. However, loci with
338 high F_{ST} estimates are predominantly outliers (Figure 5 in Appendix 2).
339 Finally, to determine whether trends revealed by the GAM analysis provide
340 meaningful insight into the mechanisms driving genetic differentiation, or simply emerge
341 by chance, we ran five null simulations. In each simulation, we randomized the observed
342 allele frequencies among populations within each locus, to break any correlations relating
343 the genetic distance between populations to their environmental and physical distance.
344 We then repeated our MCMCglmm() analyses of all loci, and fit GAM regressions to the
345 resulting set of coefficients as a function of the loci's original F_{ST} values, as described in
346 the preceding paragraph. Contrasting the GAM fits of the original and randomized data
347 allowed us to determine when genetic differentiation among populations is more strongly
348 predicted by covariates than expected by chance.
349

350 **RESULTS**

351 ***F_{ST} outliers*** — Mean genome-wide F_{ST} was 0.065 (95% credible intervals: 0.062,
352 0.068). One hundred and nine loci were identified as high outliers and sixty loci were
353 identified as low outliers (Figure 5 in Appendix S2). High outliers have allele frequencies
354 that differ between populations more than expected based on the distribution of locus-
355 specific F_{ST} estimates and could reflect the effects of diversifying selection across
356 populations. Low F_{ST} outliers could be the result of either stabilizing selection or fixed
357 heterozygosity in paralogous loci so the low outliers were removed from the remaining
358 analyses. Of the 109 high outliers, 32 map to the *Protea repens* transcriptome with e-
359 values less than 1^{-10} and 14 of those also map to *Arabidopsis thaliana* genes (Appendix
360 S4). These outliers are not necessarily the gene variants or the genes that are under
361 selection, but they mark regions of the genome that may be subject to diversifying
362 selection. A slightly larger proportion of the non-outlier loci also had hits to the *P. repens*
363 transcriptome (607 of 1897) and to the *A. thaliana* genome (367 of 1897, Appendix S5).

364 ***Distribution of Genetic Diversity*** — Both the non-outlier and high F_{ST}-outlier loci
365 were highly variable within populations; the proportion of polymorphic loci varied
366 between 0.56 and 0.96 for the non-outlier loci, and 0.50 and 0.94 for the high F_{ST} outlier
367 loci (Table 1). Nucleotide diversity was higher in the non-outlier loci than the outlier loci
368 with population specific estimates of nucleotide diversity ranging between 0.116 and
369 0.181 in the non-outlier loci and 0.071 and 0.108 in the F_{ST} outlier loci. Most of the
370 genetic variation in the non-outlier loci was found within populations (~94%, Table 2).
371 As expected, much more of the genetic variation in the outlier loci was between
372 populations, with ~76% of the variation within populations and ~24% between.

373 ***Population Structure*** — The multidimensional scaling (MDS) and individual
374 assignment (STRUCTURE) analyses revealed similar patterns. The MDS analysis shows
375 that the populations are split into an eastern group and a western group (Fig. 3), with the
376 GAR population somewhat isolated within the eastern group and RIV somewhat isolated
377 from the western group. Similarly, the STRUCTURE analysis also supports two groups.
378 Using the Evanno et al. criterion (Appendix S6), K=2 and K=3 received similar support
379 as the models that best represent the underlying genetic structure in the populations. With
380 K=2, we again see a clear delimitation between eastern and western populations (Fig.
381 2A). With K=3, the GAR population is separated from the remaining eastern populations.
382 While STRUCTURE is not well suited to situations in which there is isolation by distance
383 (Pritchard et al., 2000), the similarity in the results between the STRUCTURE analysis, the
384 MDS, and population phylogeny described below, suggests that, in this instance, the
385 results are likely to be reliable.
386 The population phylogeny estimated with TREEMIX reveals a pattern consistent with
387 these analyses. We found strong bootstrap support for eastern and western groups of
388 populations (Fig. 2B). The eastern populations also show a pattern suggestive of stepping
389 stone colonization from west to east in which each population is sister to the remaining
390 populations to the east. Patterns in the western populations are more complex. CDB, one
391 of the northernmost populations, is sister to the remaining western populations, and
392 coastal populations (KLM, POT, RIV and BRD) form a clade, but there is no clear
393 relationship between population history and geographic distribution in the western
394 populations. Consistent with the AMOVA analysis (Table 2), which showed that most of
395 the genetic variation was within populations; the longest branches in the population

396 phylogeny are the terminal branches. Internal branches, even the branch separating the
397 two clades identified in the STRUCTURE and MDS analyses, were much shorter (Fig. 2B).
398 ***Drivers of Genetic Differentiation — IBD vs. IBE –*** Of the five covariates we
399 explored, three revealed patterns that deviated substantially from the expectations of the
400 null models. Together, these provide evidence of both isolation by distance and isolation
401 by environment. Consistent with our expectations, the null models reveal clear patterns of
402 increasing variation in effect sizes as locus F_{ST} increases (loci that are more variable
403 across all individuals are capable of showing more variation in allele frequencies among
404 populations), but the average of these effect sizes across loci is zero (Fig. 1 in Appendix
405 S7). Physical distance had a positive effect on genetic differentiation between populations
406 (Fig. 4A, Fig. 2 and Table 1 in Appendix S7). The effect of distance diverges linearly
407 from the null model expectations even for loci with low F_{ST}, suggesting that isolation by
408 distance drives differentiation across F_{ST} values. Additionally, the magnitude of this
409 effect is in proportion to the F_{ST} of loci; as the amount of variation between populations
410 gets larger, so does the amount of that variation that physical distance explains. Thus,
411 variation in allele frequencies in both non-outlier and outlier loci (which have high F_{ST}) is
412 affected similarly by physical distance (Fig. 4A). There is also evidence of isolation by
413 environment, particularly in the high F_{ST} loci in response to PC2-winter rain. However,
414 this effect is less pronounced than that of physical distance, and deviates from the null
415 models only at higher levels of F_{ST} (Fig. 4C). This suggests that IBE only plays a role in
416 differentiation in the high F_{ST} loci, many of which are outliers. In addition, historical
417 barriers to gene flow have an outsized effect on differences in allele frequencies in high
418 F_{ST} loci (Fig. 4E). Specifically, in these loci, pairs of populations spanning

419 phytogeogephic zones had more divergent allele frequencies. The relationships between
420 the effects of PC1-temp (Fig.4B) and PC3-MAP (Fig. 4C) and locus F_{ST} were not
421 significant compared to our null models. While we were able to detect several significant
422 relationships between F_{ST} and physical and environmental distance, much additional
423 variation between loci remains (Fig. 2 and Table 1 in Appendix S7). This suggests that,
424 despite overall trends, individual loci can respond to these predictor variables
425 idiosyncratically.
426
427 **DISCUSSION**
428 We detected the effects of both non-adaptive and adaptive evolutionary processes in
429 the genome of *Protea repens*, a widespread, mid-elevation CFR species. We found
430 evidence for an east-west split in genetic groupings (Figs. 2 and 3) as well as for a recent
431 expansion into the eastern extent of the range of *P. repens* (Fig. 3A). We also detected a
432 strong pattern of isolation by distance, at loci with both low and high amounts of among-
433 population differentiation (Fig. 4A). Isolation by environment, specifically differences in
434 winter rainfall (Fig. 4C), also contributes to genetic differentiation, but only at loci with
435 high amounts of differentiation. We also found that historical barriers to gene flow, as
436 defined by previously recognized phytogeographic zones, help explain genetic
437 differentiation in loci with high F_{ST} estimates (Fig. 4E).
438 The analyses based on the non-outlier loci, which presumably reflect primarily
439 neutral variation, identified a split between eastern and western populations. This split
440 corresponds to a switch in both the timing of rainfall, from predominantly winter rainfall
441 in the west to a more even rainfall distribution in the east (Shultze, 2007), and to
442 differences in flowering time (Rebelo, 2001; Carlson pers. comm.). Differences in

443 flowering time could hinder gene flow between populations, reinforcing this east/west
444 divide. That said, the very short branch lengths associated with the east-west split in the
445 TREEMIX analysis (Fig. 3) suggest that most of the genetic differentiation is a result of the
446 accumulation of differences within populations and that relatively little of it is associated
447 with this divide.

448 Patterns of variation at the non-outlier loci also suggest that eastern populations of
449 *Protea repens* arose through stepping-stone colonization. This pattern is consistent with
450 both another *Protea* species in the eastern cape (*P. subvestita*; Prunier and Holsinger
451 2010), and historical patterns of climate change in the region. During the Pleistocene, the
452 western CFR had a more constant climate, while the eastern CFR was subject to drying
453 during the glacial periods (Cowling et al., 2009). Species in the eastern CFR may have
454 been forced higher in elevation or extirpated during these dry periods, with the area being
455 recolonized during wetter periods (Cowling et al., 2009).

456 We find a pattern of isolation by distance in both non-outlier and outlier loci. This
457 could arise simply because gene flow is geographically limited in *Protea repens* (Wright
458 1943). However, despite the genetic differentiation between distant populations, *P.*
459 *repens* shows no indication of speciation, likely because gene flow between adjacent
460 populations allows for sufficient genetic connectivity. *Restio capensis* (Restionaceae;
461 Lexer et al. 2014), the only other widespread mid-elevation CFR species for which there
462 is population genomic data, does not exhibit a consistent isolation by distance effect.
463 Instead, in *R. capensis*, environmental differences are the best predictor of genetic
464 differentiation in both non-outlier and outlier loci. While both species are broadly
465 distributed in the CFR, species in Restionaceae are generally sensitive to water

466 availability and soil types (Araya et al., 2010) while species in the genus *Protea* are
467 generally deeply rooted (Richards et al., 1995; Watt and Evans, 1999). Thus, the different
468 patterns observed in these two analyses could arise if *R. capensis* is inherently more
469 likely to differentiate along environmental gradients than *P. repens*. These two studies
470 conflict on the importance of non-adaptive processes to genetic diversification in the
471 CFR.

472 While gene flow in *Protea repens* was most consistently limited by physical distance,
473 we did find evidence that, for highly variable loci, among-population differentiation is
474 associated with differences in winter rainfall. This is consistent with the results of several
475 studies of functional traits in these same populations (Carlson et al., 2015). The
476 association between traits and differences in winter rainfall (Prunier et al 2012, Carlson et
477 al. 2011) as well as differential selection on those traits in response to winter rainfall
478 (Carlson et al. 2011) has also been found in *Protea* section *Exertae*, a group of widely
479 distributed high elevation species. Lexer et al. (2014) also found patterns of isolation by
480 environment, but they did not identify which environmental gradient is driving these
481 differences. These results lend support to the hypothesis that adaptive processes
482 contribute to differentiation in the CFR.

483 Historical barriers to gene flow also contributed to genetic differentiation in *Protea*
484 *repens*, but only in the loci with high amounts of among-population differentiation. We
485 recognize these barriers as the boundaries between phytogeographic zones defined by
486 breaks in the ranges of species from many groups of plants in the CFR (Goldblatt and
487 Manning, 2002). It is curious that we identified this effect only for highly differentiated
488 loci. If the phytogeographic zones were acting as barriers to dispersal, they should affect

489 all loci in a similar manner, like isolation by distance. Instead, the pattern we observed
490 suggests that the boundaries between phytogeographic zones reflect other environmental
491 differences not included in our PCA axes that might different between phytogeographic
492 zones, such as soil type (Linder 2001; Goldblatt and Manning, 2002) or fire return
493 interval (Litsios et al., 2013), or perhaps combinations of environmental factors.

494 Identifying the genes and gene variants responsible for the responses to diversifying
495 selection that have been identified in *Protea repens* (see Carson et al 2015) would
496 provide insight into the physiological processes that allow individuals to survive in
497 varying environments. Genomic regions surrounding the outlier loci we have identified
498 may well contain such genes. Although more complete SNP coverage is necessary before
499 good candidate genes can be confidently identified, some outlier loci in the present
500 analysis have intriguing relationships to the environmental gradients. For example,
501 among-population differences in allele frequency at locus TP26245, which occurs in gene
502 AT5G48930 (involved in auxin transport and growth in *Arabidopsis thaliana*), are
503 correlated with changes in PC2 winter rainfall (Appendix S8). This coincides with a
504 decrease in growth rate growth rate in *P. repens* populations with lower winter rainfall
505 (Merow et al., 2014).

506 ***Conclusions*** — We document that a combination of adaptive and non-adaptive processes
507 are associated with diversification in a widespread CFR species. Our results provide
508 evidence consistent with post-glacial colonization of the eastern part of the region. They
509 also suggest that while physical distance reduces gene flow between distant populations,
510 there is considerable gene flow between nearby populations. Interestingly, historical
511 barriers that have contributed to diversification in other groups (phytogeographic zones)

512 also contribute to divergence in *P. repens* populations, but only in loci showing the
513 largest amounts of among-population differentiation. This work adds to the currently
514 limited number of studies of the micro-evolutionary processes driving diversification in
515 the CFR. Given the differences between the patterns observed in *Protea repens* and
516 *Restio capensis*, additional studies of plants from the CFR are required to discover
517 whether the patterns shown by either species hold broadly in the CFR.
518

519 ACKNOWLEDGEMENTS

520 The authors thank J.E. Carlson for her major contribution collecting seeds, growing
521 seedlings, and for editorial comments. A. Latimer collected leaf samples and Y.
522 Chupalova, K. Varney, A. Kunicki, and M. Oldakowski extracted DNA from hundreds of
523 leaves. K.E. Theiss, A. Royer, A. Latimer, two reviewers and the editor at AJB provided
524 thoughtful feedback on the manuscript. We thank the reserve managers and property
525 owners for access to collection sites. L. Nurrish and T. Rebelo of SANBI generously
526 provided garden access at Kirstenbosch. This work was supported by the National
527 Science Foundation (DEB-1046328). Seeds were collected under Cape Nature permits
528 AAA005-00214-0028 and AAA005-00224-0028 and Eastern Cape Province permit CRO
529 4/11 CR.
530

531 **LITERATURE CITED**

532 Adamson, R.S. 1958. The Cape as an ancient African flora. *The Advancement of Science*
533 58: 1–10.

534 Akey, J.M. 2009. Constructing genomic maps of positive selection in humans: Where do
535 we go from here? *Genome Research* 19: 711–722.

536 Akman, M., J. Carlson, K. Holsinger, and A. Latimer. 2016. Transcriptome sequencing
537 reveals population differentiation in gene expression linked to functional traits and
538 environmental gradients in South African shrub *Protea repens*. *New Phytologist*
539 210: 295–309.

540 Beaumont, M.A., and D.J. Balding. 2004. Identifying adaptive genetic divergence among
541 populations from genome scans. *Molecular Ecology* 13: 969–980.

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

546 Browne, W.J., H. Goldstein, and J. Rasbach. 2001. Statistical modelling. *Statistical*
547 *Modeling* 1: 103–124.

548 Carlson, J.E., C.A. Adams, and K.E. Holsinger. 2015. Intraspecific variation in stomatal
549 traits, leaf traits and physiology reflects adaptation along aridity gradients in a South
550 African shrub. *Annals of Botany* 117: 195–207.

551 Carlson, J.E., K.E. Holsinger, and R. Prunier. 2011. Plant responses to climate in the
552 Cape Floristic Region of South Africa: evidence for adaptive differentiation in the
553 Proteaceae. *Evolution* 65: 108–124. Available at:

554 http://doi.wiley.com/10.1111/j.1558-5646.2010.01131.x.
555 Coop, G., D. Witonsky, A. Di Rienzo, and J.K. Pritchard. 2010. Using environmental
556 correlations to identify loci underlying local adaptation. *Genetics* 185: 1411–1423.
557 Available at: http://www.genetics.org/cgi/doi/10.1534/genetics.110.114819.
558 Cowling, R.M., B.M. Campbell, P. Mustart, D.J. McDonald, M.L. Jarman, and E.J. Moll.
559 1988. Vegetation classification in a floristically complex area: the Agulhas Plain.
560 *South African Journal of Botany*.
561 Cowling, R.M., Ş. Procheş, and T.C. Partridge. 2009. Explaining the uniqueness of the
562 Cape flora: Incorporating geomorphic evolution as a factor for explaining its
563 diversification. *Molecular Phylogenetics and Evolution* 51: 64–74.
564 Doyle, J.J., and J.L. Doyle. 1987. A rapid DNA isolation procedure for small quantities
565 of fresh leaf tissue. *Phytochemistry Bulletin* 19: 11–15.
566 Earl, D.A., and B.M. vonHoldt. 2012. STRUCTURE HARVESTER: a website and program
567 for visualizing STRUCTURE output and implementing the Evanno method.
568 *Conservation Genetics Resources* 4: 359–361. Available at:
569 http://link.springer.com/10.1007/s12686-011-9548-7.
570 Ellis, A.G., G.A. Verboom, T. van der Niet, S.D. Johnson, and H.P. Linder. 2014.
571 Speciation and extinction in the Greater Cape Floristic Region. *In* N. Allsopp, J. F.
572 Colville, G. A. Verboom, and R. M. Cowling [eds.], Fynbos: Ecology, Evolution
573 and Conservation of a Megadiverse Region, 119–141. Oxford University Press,
574 Oxford, UK.
575 Ellis, A.G., and A.E. Weis. 2006. Coexistence and differentiation of "flowering stones":
576 The role of local adaptation to soil microenvironment. *Journal of Ecology* 94: 322–

577 335.
578 Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E.
579 Mitchell. 2011. A robust, simple Genotyping-by-Sequencing (GBS) approach for
580 high diversity species. *PLoS ONE* 6: e19379. Available at:
581 http://dx.plos.org/10.1371/journal.pone.0019379.
582 Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of
583 individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*
584 14: 2611–2620. Available at: http://doi.wiley.com/10.1111/j.1365-
585 294X.2005.02553.x.
586 Excoffier, L., T. Hofer, and M. Foll. 2009. Detecting loci under selection in a
587 hierarchically structured population. *Heredity* 103: 285–298. Available at:
588 http://dx.doi.org/10.1038/hdy.2009.74.
589 Excoffier, L., and H.E.L. Lischer. 2010. Arlequin suite ver 3.5: A new series of programs
590 to perform population genetics analyses under Linux and Windows. *Molecular*
591 *Ecology Resources* 10: 564–567.
592 Falush, D., M. Stephens, and J.K. Pritchard. 2003. Inference of population structure using
593 multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*
594 164: 1567–1587.
595 Fielding, A., and H. Goldstein. 2006. Cross-classified and multiple membership
596 structures in multilevel models- An introduction and review. *University of*
597 *Birmingham Research Reports* 1–69.
598 Gibby, M., S. Hinnah, E.M. Marais, and F. Albers. 1996. Cytological variation and
599 evolution within *Pelargonium* section *Hoarea* (Geraniaceae). *Plant Syst. Evol.* 203:

600 111–142.
601 Goldblatt, P., and J.C. Manning. 2002. Plant diversity of the Cape region of southern
602 Africa. *Annals of the Missouri Botanical Garden* 89: 281–302.
603 Goldblatt, P., V. Savolainen, O. Porteous, I. Sostaric, M. Powell, G. Reeves, J.C.
604 Manning, et al. 2002. Radiation in the Cape flora and the phylogeny of peacock
605 irises *Moraea* (Iridaceae) based on four plastid DNA regions. *Molecular*
606 *Phylogenetics and Evolution* 25: 341–360.
607 Gregorius, H.R., and J.H. Roberds. 1986. Measurement of genetical differentiation
608 among sub-populations. *Theorical and Applied Genetics* 71: 826–834.
609 Guo, F., D.K. Dey, and K.E. Holsinger. 2009. A Bayesian hierarchical model for analysis
610 of single-nucleotide polymorphisms diversity in multilocus, multipopulation
611 samples. *Journal of the American Statistical Association* 104: 142–154.
612 Hadfield, J. 2010. MCMC methods for multi-response generalized linear mixed models:
613 The MCMCglmm R package. *Journal of Statistical Software* 33: 1–22.
614 Holsinger, K.E. 2016. Genotyping by sequencing in *Protea repens*. *Zenodo*. DOI:
615 10.5281/zenodo.54919
616 Jakobsson, M., and N.A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation
617 program for dealing with label switching and multimodality in analysis of
618 population structure. *Bioinformatics* 23: 1801–1806. Available at:
619 http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btm233.
620 Johnson, S.D. 1996. Pollination, adaptation and speciation models in the Cape flora of
621 South Africa. *Taxon* 45: 59–66.
622 Johnson, S.D., and K.E. Steiner. 1997. Long-tongued fly pollination and evolution of

floral spur length in the *Disa draconis* complex (Orchidaceae). *Evolution* 51: 45–53. Available at: http://www.jstor.org/stable/2410959.

Knowles, L.L. 2001. Did the Pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshopprers. *Molecular Ecology* 10: 691–701.

Kozak, K.H., and J.J. Wiens. 2006. Does niche conservatism promote speciation? A case study in North American salamanders. *Evolution* 60: 2604–2621. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.0014-3820.2006.tb01893.x/full.

Latimer, A.M., J.A. Silander, and R.M. Cowling. 2005. Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot. *Science* 309: 1722–1725. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.1115576.

Lazaridis, I., N. Patterson, A. Mittnik, G. Renaud, S. Mallick, K. Kirsanow, P.H. Sudmant, et al. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513: 409–413. Available at: http://dx.doi.org/10.1038/nature13673.

Lexer, C., R.O. Wüest, S. Mangili, M. Heuertz, K.N. Stölting, P.B. Pearman, F. Forest, et al. 2014. Genomics of the divergence continuum in an African plant biodiversity hotspot, I: drivers of population divergence in *Restio capensis* (Restionaceae). *Molecular Ecology* 23: 4373–4386. Available at: http://doi.wiley.com/10.1111/mec.12870.

Linder, H.P. 1985. Gene flow, speciation, and species diversity patterns in a species-rich area: the Cape flora. *In* E. S. Vrba [ed.], Species and Speciation, 53–57. Transvall Museum, Pretoria, RSA.

Linder, H.P. 2001. On areas of endemism, with an example from the african

646 Restionaceae. *Syst. Biol* 50: 892–912.

647 Linder, H.P. 2003. The radiation of the Cape flora, southern Africa. *Biological Reviews*
648 *of the Cambridge Philosophical Society* 78: 597–638.

649 Litsios, G., R.O. Wüest, A. Kostikova, F. Forest, C. Lexer, H.P. Linder, P.B. Pearman, et
650 al. 2014. Effects of a fire response trait on diversification in replicated radiations.
651 *Evolution* 68: 453–465.

652 Lotterhos, K.E., and M.C. Whitlock. 2014. Evaluation of demographic history and
653 neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology*
654 23: 2178–2192. Available at: http://doi.wiley.com/10.1111/mec.12725.

655 Lu, F., A.E. Lipka, J. Glaubitz, R. Elshire, J.H. Cherney, M.D. Casler, E.S. Buckler, and
656 D.E. Costich. 2013. Switchgrass genomic diversity, ploidy, and evolution: novel
657 Insights from a network-based SNP discovery protocol. *PLoS Genetics* 9: e1003215.
658 Available at: http://dx.plos.org/10.1371/journal.pgen.1003215.

659 Merow, C., A.M. Latimer, A.M. Wilson, S.M. Mcmahon, A.G. Rebelo, and J.A. Silander.
660 2014. On using integral projection models to generate demographically driven
661 predictions of species' distributions: Development and validation using sparse data.
662 *Ecography* 37: 1167–1183.

663 Meyer, M., M. Kircher, M.-T. Gansauge, H. Li, F. Racimo, S. Mallick, J.G. Schraiber, et
664 al. 2012. A high-coverage genome sequence from an archaic denisovan individual.
665 *Science* 338: 222 LP – 226. Available at:
666 http://science.sciencemag.org/content/338/6104/222.abstract.

667 Morris, G.P., P.P. Grabowski, and J.O. Borevitz. 2011. Genomic diversity in switchgrass
668 (*Panicum virgatum*): From the continental scale to a dune landscape. *Molecular*

669 *Ecology* 20: 4938–4952.
670 Myers, N., R.A. Mittermeier, C.G. Mittermeier, G.A.B. daFonseca, and J. Kent. 2000.
671 Biodiversity hotspots for conservation priorities. *Nature* 403: 853–858.
672 Nicotra, A.B., C. Chong, J.G. Bragg, C.R. Ong, N.C. Aitken, A. Chuah, B. Lepschi, and
673 J.O. Borevitz. 2016. Population and phylogenomic decomposition via Genotyping-
674 By-Sequencing in Australian *Pelargonium*. *Molecular Ecology* 25: 2000–20014.
675 Available at: http://doi.wiley.com/10.1111/mec.13584.
676 van der Niet, T., and S.D. Johnson. 2009. Patterns of plant speciation in the Cape floristic
677 region. *Molecular Phylogenetics and Evolution* 51: 85–93. Available at:
678 http://linkinghub.elsevier.com/retrieve/pii/S1055790308005381.
679 Nosil, P., D.J. Funk, and D. Ortiz-Barrientos. 2009. Divergent selection and
680 heterogeneous genomic divergence. *Molecular Ecology* 18: 375–402. Available at:
681 http://doi.wiley.com/10.1111/j.1365-294X.2008.03946.x.
682 Patterson, N., P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, et
683 al. 2012. Ancient admixture in human history. *Genetics* 192: 1065 LP – 1093.
684 Available at: http://www.genetics.org/content/192/3/1065.abstract.
685 Patterson, N., A.L. Price, and D. Reich. 2006. Population structure and eigenanalysis.
686 *PLoS Genetics* 2: e190.
687 Pickrell, J.K., and J.K. Pritchard. 2012. Inference of population splits and mixtures from
688 genome-wide allele frequency data. *PLoS Genetics* 8: e1002967.
689 Pirie, M.D., E.G.H. Oliver, A. Mugrabi de Kuppler, B. Gehrke, N.C. Le Maitre, M.
690 Kandziora, D.U. Bellstedt, et al. 2016. The biodiversity hotspot as evolutionary hot-
691 bed: spectacular radiation of Erica in the Cape Floristic Region. *BMC Evolutionary*

692 *Biology* 16: 190. Available at:
693 http://bmcevolbiol.biomedcentral.com/articles/10.1186/s12862-016-0764-3.
694 Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure
695 using multilocus genotype data. *Genetics* 155: 945–959.
696 Prunier, R., and K.E. Holsinger. 2010. Was it an explosion? Using population genetics to
697 explore the dynamics of a recent radiation within *Protea* (Proteaceae L.). *Molecular*
698 *Ecology* 19: 3968–3980. Available at: http://doi.wiley.com/10.1111/j.1365-
699 294X.2010.04779.x.
700 Prunier, R., K.E. Holsinger, and J.E. Carlson. 2012. The effect of historical legacy on
701 adaptation: do closely related species respond to the environment in the same way?
702 *Journal of Evolutionary Biology* 25: 1636–1649. Available at:
703 http://doi.wiley.com/10.1111/j.1420-9101.2012.02548.x.
704 R Core Development Team. 2014. R: A language and environment for statistical
705 computing. Available at: http://www.r-project.org/.
706 R Core Development Team. 2016. R: A language and environment for statistical
707 computing. Available at: http://www.r-project.org/.
708 Rebelo, T.G. 2006. Protea Atlas Project Website. Available at:
709 http://www.proteaatlas.org.za/ [Accessed February 20, 2014].
710 Rebelo, T.G. 2001. Sasol Proteas: A field guide to the proteas of southern Africa. 2nd ed.
711 Fernwood Press, Vlaeberg, South Africa.
712 Raghavan, M., P. Skoglund, K.E. Graf, M. Metspalu, A. Albrechtsen, I. Moltke, S.
713 Rasmussen, et al. 2014. Upper Palaeolithic Siberian genome reveals dual ancestry of
714 Native Americans. *Nature* 505: 87–91. Available at:

715 http://dx.doi.org/10.1038/nature12736.
716 Richards, M.B., W.D. Stock, and R.M. Cowling. 1995. Water relations of seedings and
717 adults of two fynbos species in relation to their distribution patterns. *Functional*
718 *Ecology* 9: 575–583.
719 Rieseberg, L.H., and J.M. Burke. 2001. Commentary: A genic view of species
720 integration. *Journal of Evolutionary Biology* 14: 883–886.
721 Rosenberg, N.A. 2003. DISTRUCT: a program for the graphical display of population
722 structure. *Molecular Ecology Notes* 4: 137–138. Available at:
723 http://doi.wiley.com/10.1046/j.1471-8286.2003.00566.x.
724 Rymer, P.D., J.C. Manning, P. Goldblatt, M.P. Powell, and V. Savolainen. 2010.
725 Evidence of recent and continuous speciation in a biodiversity hotspot: a population
726 genetic approach in southern African gladioli (Gladiolus; Iridaceae). *Molecular*
727 *Ecology* 19: 4765–4782. Available at: http://doi.wiley.com/10.1111/j.1365-
728 294X.2010.04794.x.
729 Sauquet, H., P.H. Weston, N.P. Barker, C.L. Anderson, D.J. Cantrill, and V. Savolainen.
730 2009. Using fossils and molecular data to reveal the origins of the Cape proteas
731 (subfamily Proteoideae). *Molecular Phylogenetics and Evolution* 51: 31–43.
732 Available at: http://linkinghub.elsevier.com/retrieve/pii/S1055790308005885.
733 Schnitzler, J., T.G. Barraclough, J.S. Boatwright, P. Goldblatt, J.C. Manning, M.P.
734 Powell, T. Rebelo, and V. Savolainen. 2011. Causes of plant diversification in the
735 Cape biodiversity hotspot of South Africa. *Systematic Biology* 60: 343–357.
736 Available at: http://sysbio.oxfordjournals.org/cgi/doi/10.1093/sysbio/syr006.
737 Schultze, R.E. 1997. South African Atlas of Climatology and Agrohydrology: WRC

Report TT82/96; ACRU report, 46. Pretoria, RSA.

Schultze, R.E. 2007. South African Atlas of Climatology and Agrohydrology: WRC Report 1489/1/06. Pretoria, RSA.

Valente, L.M., G. Reeves, J. Schnitzler, I.P. Mason, M.F. Fay, T.G. Rebelo, M.W. Chase, and T.G. Barraclough. 2010. Diversification of the African genus *Protea* (Proteaceae) in the cape biodiversity hotspot and beyond: equal rates in different biomes. *Evolution* 64: 745–760. Available at: http://doi.wiley.com/10.1111/j.1558-5646.2009.00856.x.

Vavrek, M.J. 2011. Fossil: palaeoecological and palaeogeographical analysis tools. *Palaeontologia Electronica* 14: 1T.

Verboom, G.A., J.K. Archibald, F.T. Bakker, D.U. Bellstedt, F. Conrad, L.L. Dreyer, F. Forest, et al. 2009. Origin and diversification of the Greater Cape flora: Ancient species repository, hot-bed of recent radiation, or both? *Molecular Phylogenetics and Evolution* 51: 44–53. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1055790308000365.

Verboom, G.A., N.G. Bergh, S.A. Haiden, V. Hoffmann, and M.N. Britton. 2015. Topography as a driver of diversification in the Cape Floristic Region of South Africa. *New Phytologist* 207: 368–376. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25708902\nhttp://doi.wiley.com/10.1111/nph.13342.

Verboom, G.A., H.P. Linder, F. Forest, V. Hoffmann, N.G. Bergh, and R.M. Cowling. 2014. Cenozoic assembly of the Greater Cape flora. *In* N. Allsopp, J. F. Colville, G. A. Verboom, and R. M. Cowling [eds.], Fynbos: Ecology, Evolution and Conservation of a Megadiverse Region, 93-118. Oxford University Press, Oxford,

761 UK.

762 Wang, I.J., and G.S. Bradburd. 2014. Isolation by environment. *Molecular Ecology* 23:

763 5649–5662. Available at: http://doi.wiley.com/10.1111/mec.12938.

764 Watt, M., and J.R. Evans. 1999. Proteoid roots. Physiology and development. *Plant*

765 *Physiology* 121: 317–323.

766 Whittaker, R.H. 1960. Vegetation of the Siskiyou mountains, Oregon and California.

767 *Ecological Monographs* 30: 279–338.

768 Wright, S. 1943. Isolation by distance. *Genetics* 28: 114–138.

769 Wu, C.I. 2001. The genic view of the process of speciation. *Journal of Evolutionary*

770 *Biology* 14: 851–865.

771 Table 1. Spatial coordinates, sample sizes, and estimates of genetic diversity in non-outlier and high outlier nuclear GBS loci for 19 populations of

772 *Protea repens* studied in the CFR of South Africa.

population	longitude	latitude	individuals sampled	maternal lines sampled	non-outlier loci % polymorphic	non-outlier loci nucleotide diversity	high outlier loci % polymorphic	high outlier loci nucleotide diversity
1 - VanRhynsdorp	19.0225	-31.37083	35	21	0.923	0.181 ± 0.087	0.853	0.108 ± 0.055
2 - Cederberg	19.10676	-32.4064	22	14	0.853	0.154 ± 0.074	0.78	0.087 ± 0.046
3 - Banghoek	18.64454	-32.72662	8	5	0.562	0.116 ± 0.058	0.5	0.089 ± 0.049
4 - Riverlands	18.55111	-33.51604	36	22	0.901	0.162 ± 0.077	0.843	0.104 ± 0.054
5 - Ceres	19.276	-33.36542	41	24	0.923	0.166 ± 0.08	0.861	0.095 ± 0.049
6 - Kleinmond	19.03363	-34.33109	39	23	0.907	0.164 ± 0.079	0.692	0.101 ± 0.053
7 - Montagu	20.10137	-33.7839	12	8	0.833	0.168 ± 0.083	0.724	0.094 ± 0.05
8 - Riviersonderend	19.99314	-34.10228	42	25	0.941	0.163 ± 0.078	0.899	0.096 ± 0.05
9 - Bredasdorp	20.0386	-34.54593	46	23	0.929	0.164 ± 0.078	0.899	0.089 ± 0.046
10 - DeHoop	20.60545	-34.41555	28	15	0.896	0.166 ± 0.08	0.778	0.092 ± 0.048
11 - Anysberg	20.70409	-33.49369	52	28	0.961	0.167 ± 0.08	0.945	0.096 ± 0.049
12 - Klein Swartberg	21.28479	-33.45617	47	25	0.931	0.149 ± 0.071	0.908	0.083 ± 0.043
13 - Garcia's Pass	21.21982	-33.96789	73	33	0.948	0.171 ± 0.081	0.881	0.099 ± 0.051

14 - Swartberg	22.04591	-33.34953	58	29	0.949	0.154 ± 0.074	0.917	0.087 ± 0.045
15 - Uniondale	23.05253	-33.72581	30	9	0.903	0.173 ± 0.083	0.844	0.098 ± 0.051
16 - Baviaanskloof	23.63356	-33.49336	36	15	0.914	0.161 ± 0.077	0.907	0.096 ± 0.049
17 - Kareedouw	24.07328	-33.87468	24	18	0.853	0.145 ± 0.07	0.717	0.073 ± 0.039
18 - Loerie Dam	25.0423	-33.85452	22	12	0.874	0.164 ± 0.079	0.759	0.106 ± 0.055
19 - Alicedale	26.09839	-33.24081	12	6	0.644	0.135 ± 0.066	0.59	0.071 ± 0.039

775 Table 2. Analysis of molecular variance (AMOVA) of 19 populations of *Protea repens* in in South Africa estimated for non-outlier and high F_{ST}

776 outlier loci. *** indicates F_{ST} estimate significantly different from zero.

777

778

Genome Fraction	source of variation	d.f.	Variance Components	% of Variation	F_{ST}
RADseq/ non-outlier	Among	18	9.74121	6.29	0.06292***
	Within	1307	145.07112	93.71	
RADseq/ outliers	Among	18	1.35714	23.87	0.23867***
	Within	1305	4.32908	76.13	

779

780

781 APPENDICES

782 Appendix 1. Voucher specimens for each *Protea repens* population included in this study all are

783 deposited at the CONN herbarium.

784

785 Population code; CONN accession number

786

787 ALC 228495; ANY 228480; BAN 228482; BAV 228483; BRD 228488; CDB 228484; CER

788 228492; GAR 228490; KAR 228479KLM 228489; KSW 228486; LOE 228491; MGU 228485;

789 POT 228496; RIV 228493; RND 228484; SWA 228477; UNI 228481; VAN 228487;

790

791 FIGURE LEGENDS

792 Fig. 1. Sampling locations of *Protea repens* (indicated by stars) and boundaries of the
793 phytogeographic regions (black outlines) of the Cape Floristic Region drawn from Goldblatt and
794 Manning (2002). Inset is a map of Africa, with the enlarged South African provinces demarcated
795 by the red outline.

796 Fig. 2. First two multidimensional scaling (MDS) axes drawn from pairwise identities by descent
797 between *Protea repens* individuals. Individuals are indicated by small dots and population mean
798 values by large dots. Populations are coded by color corresponding to the STRUCTURE analysis
799 (Fig. 3A); eastern populations are yellow and western populations are green. The shading
800 indicates the degree to which each population is east or west with centrally located populations
801 darker than peripheral populations.

802 Fig. 3. (A) Results of the STRUCTURE analyses for K=2 and K=3. Each *Protea repens* individual
803 (grouped by population) is represented by a vertical bar. The proportion of the bar in each color
804 corresponds to the average posterior likelihood that the individual is assigned to the cluster
805 indicated by that color. Populations are separated by black lines and are displayed approximately
806 from west to east. (B) TREEMIX consensus tree with maximum likelihood branch-lengths based
807 on population level allele frequencies at the 1897 non-outlier SNP loci. Bootstrap values above
808 50 are indicated below the nodes.

809 Fig. 4. Both IBD and IBE contribute to genetic divergence between pairs of *Protea repens*
810 populations. The loci we examined have a wide range of global F_{ST}s. Loci with a high global F_{ST}
811 may occur in genomic regions that experience divergent selection. To explore the factors driving
812 population differentiation for each locus, we examined how strongly differences in allele
813 frequencies between pairs of populations relate to their: (A) differences in physical distance, three
814 environmental axes ((B) PC1-temp, (C) PC2-winterrain, (D) PC3-MAP), and (E)
815 phytogeographic zone. The strength of these relationships (effect size) varied substantially among
816 different explanatory factors (panels) and across individual loci (arranged by their global F_{ST}

817 within panels). For clarity, we present here only the trends in effect size with 95% confidence
818 bands (as quantified using GAM fits). Fits to our original data are in blue, while the results of five
819 null model randomizations are in gray (for more detail, see Methods and Appendix SX). Physical
820 distance (A) was the most informative factor, showing positive effect sizes across a wide range of
821 locus F_{ST}s that deviate from the null model patterns. PC2- winter rain (C) and phytogeographic
822 zone (E) also display significant positive effects, but only at loci with high F_{ST} values. PC1-temp
823 (B) and PC3-MAP (E) show no deviation from the null model. These results suggest that all loci
824 show the effects of IBD, independent of their F_{ST}, whereas only the more variable outlier loci
825 (high F_{ST}) exhibit IBE and population differentiation that we attribute to historical forces.
826

Figure 1

Click here to download Figure fig1_phytoregions_nonames.pdf

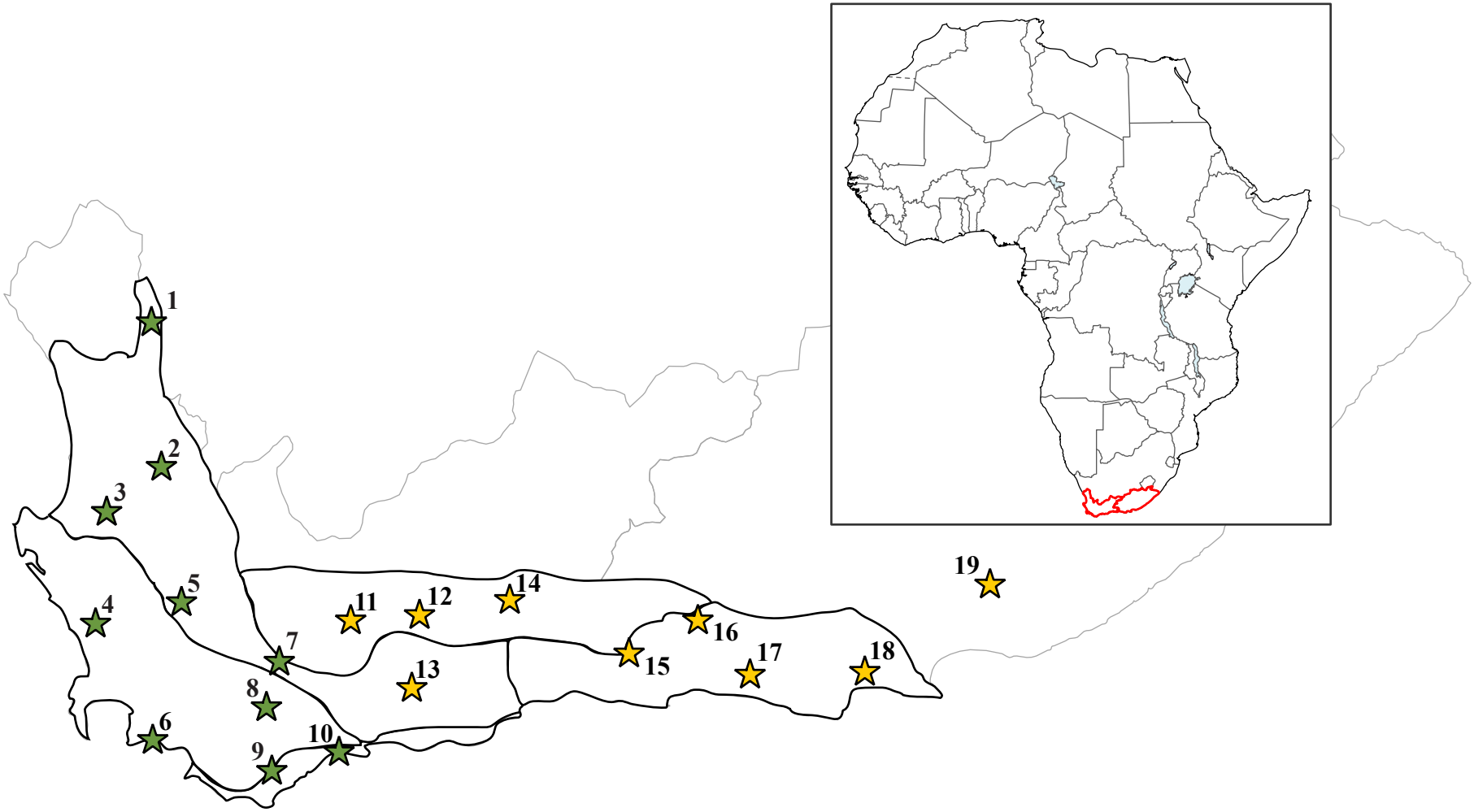


Figure 2

Click here to download Figure fig2_new_structure.pdf

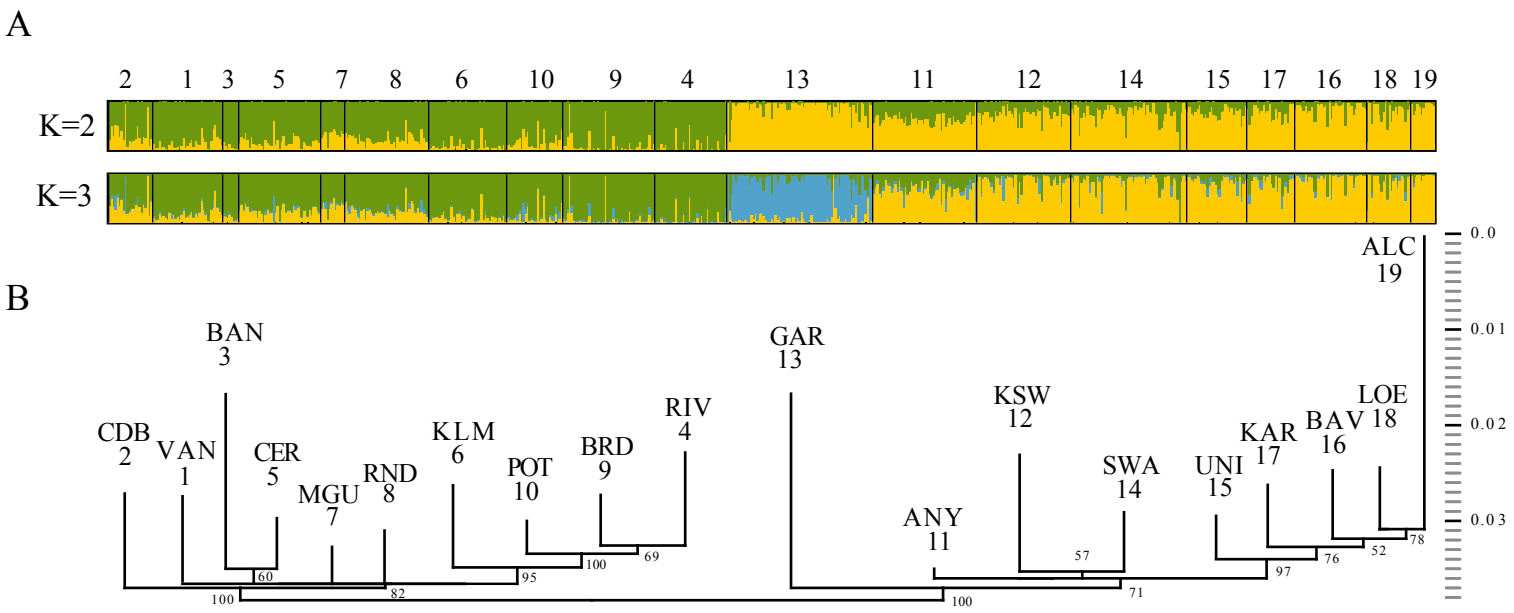


Figure 3

Click here to download Figure fig3_MDS-f-dist-repens_greenv2_nonames.pdf

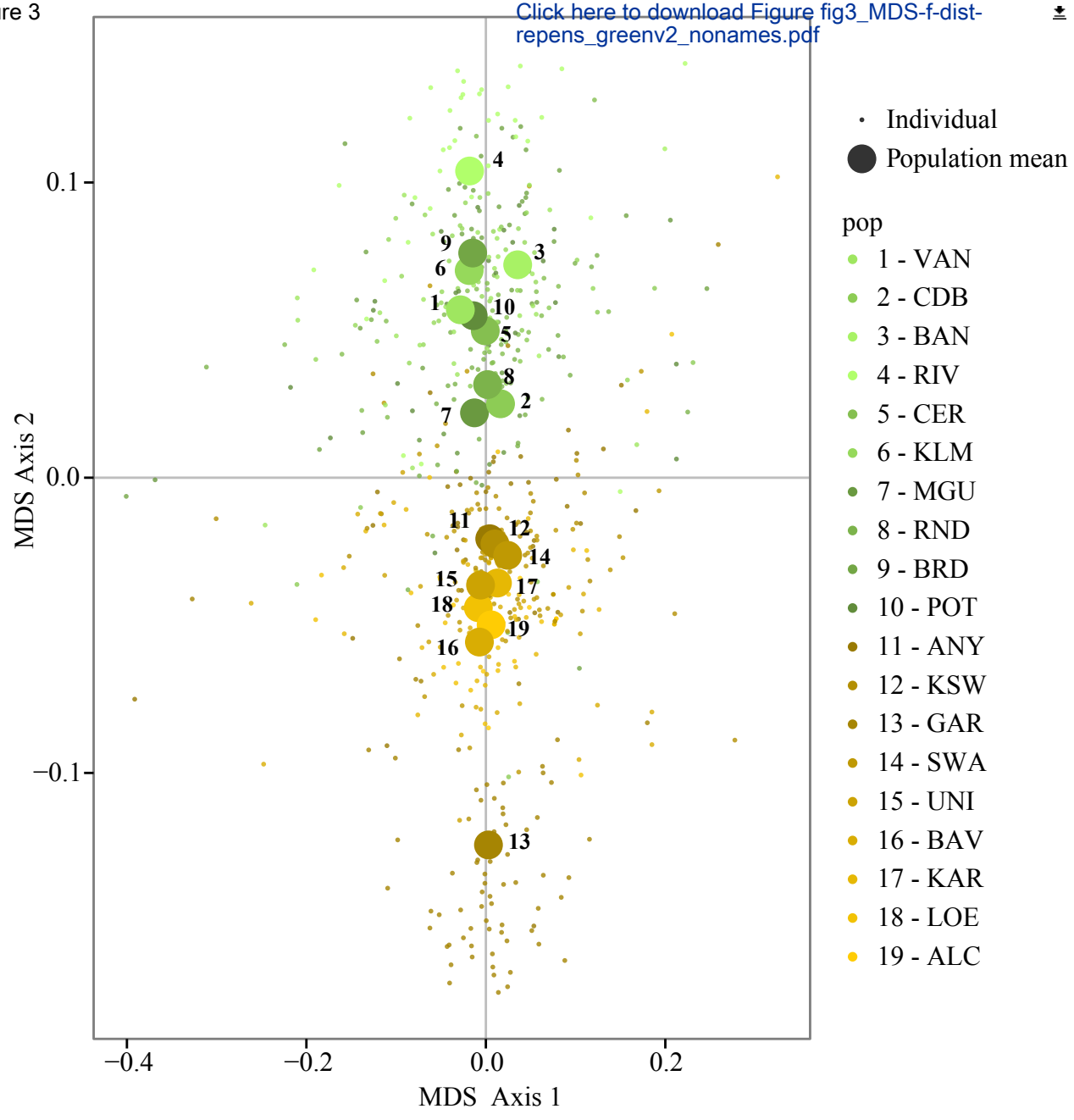
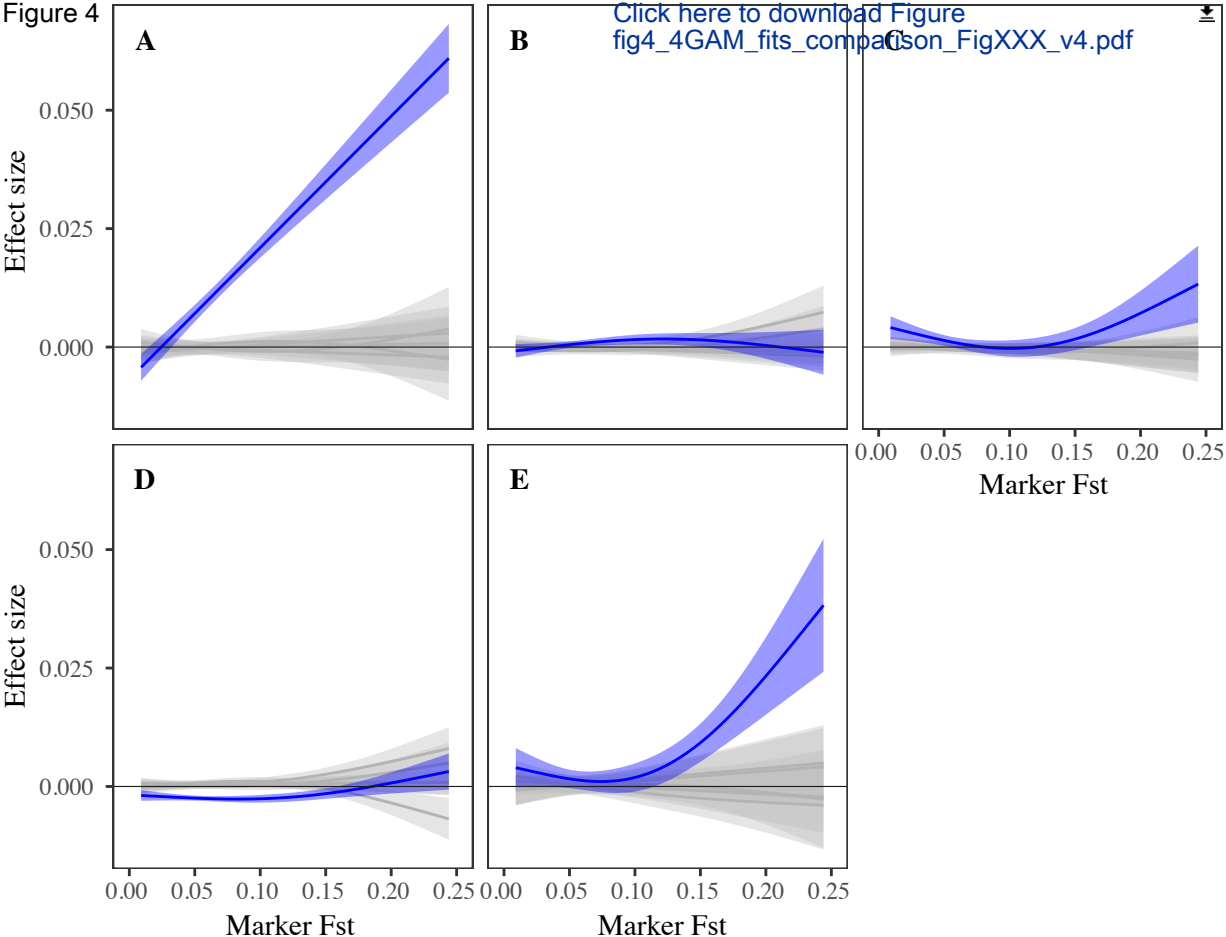


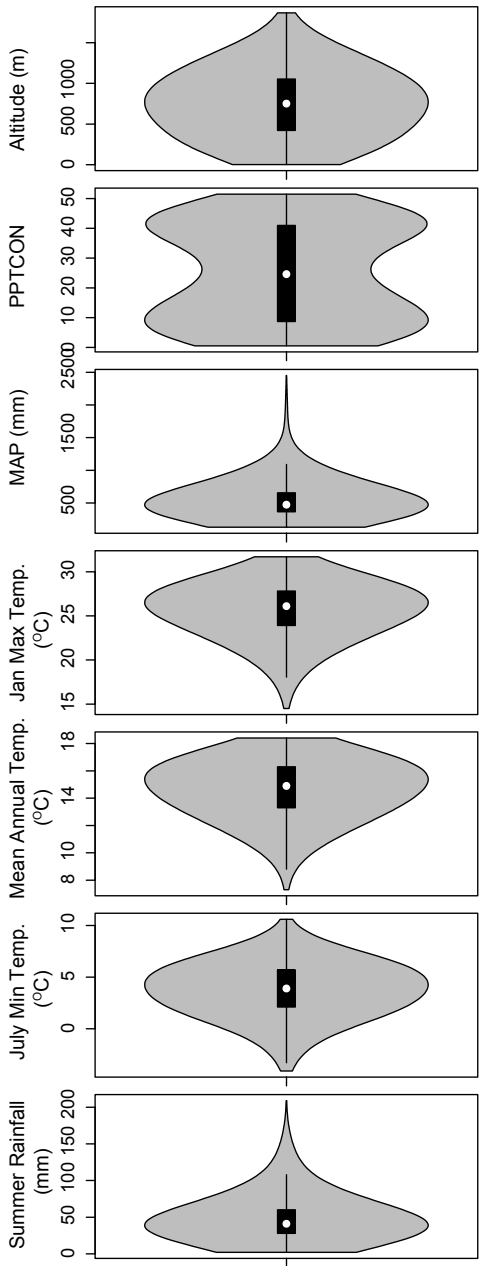
Figure 4

Click here to download Figure
fig4_4GAM_fits_comparison_FigXXX_v4.pdf



Online Supplemental 1 Click here to download Online Supplemental Appendix_S1.docx

Appendix S1. Violin plots showing variation in the seven environmental variables drawn from Schultze 2007 that were used to make the principal component axes. The mean is indicated by the white dot and the first and third quartiles by the black rectangle. The kernel density is indicated by the gray shape.



Online Supplemental 2

Click here to download Online Supplemental Appendix_S2_data.docx

Appendix S2. Summary of SNP processing

Figure 1: Histograms of SNP calls per individual before (blue) and after (red) the calls based on fewer than five reads were removed. Note the y-axis is in log scale for readability.

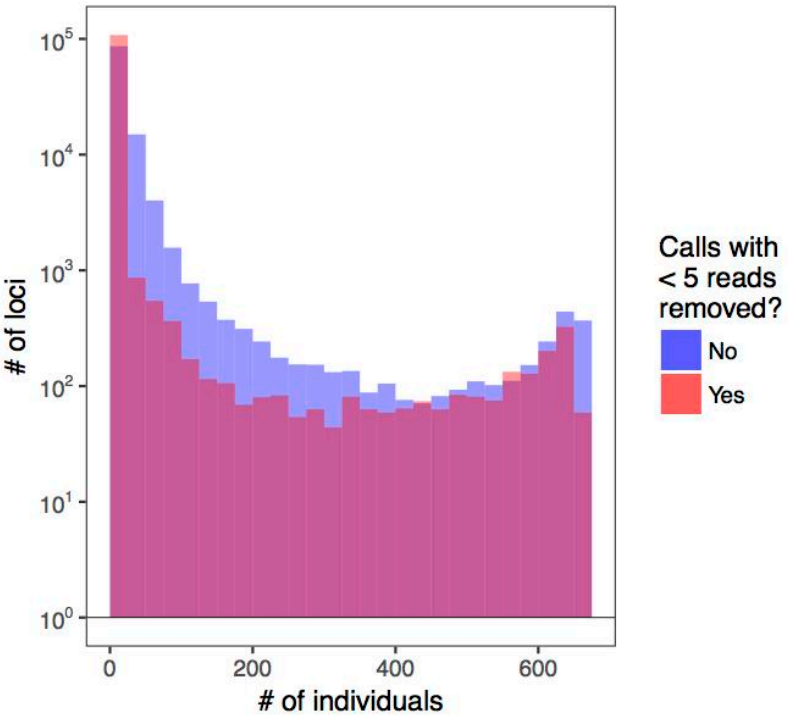


Table 1: Summary of number of calls per individual, after the calls with fewer than 5 reads were removed. Note that most individuals have many reads.

Read count range	number of individuals
0-500	53
500-1000	2
1000-5000	16
5000-10000	9
10000-50000	133
50000-100000	293
>100000	210

Figure 2: Distribution of the number of SNPs for which we have different numbers of individuals. Many SNPs are called for most of the individuals

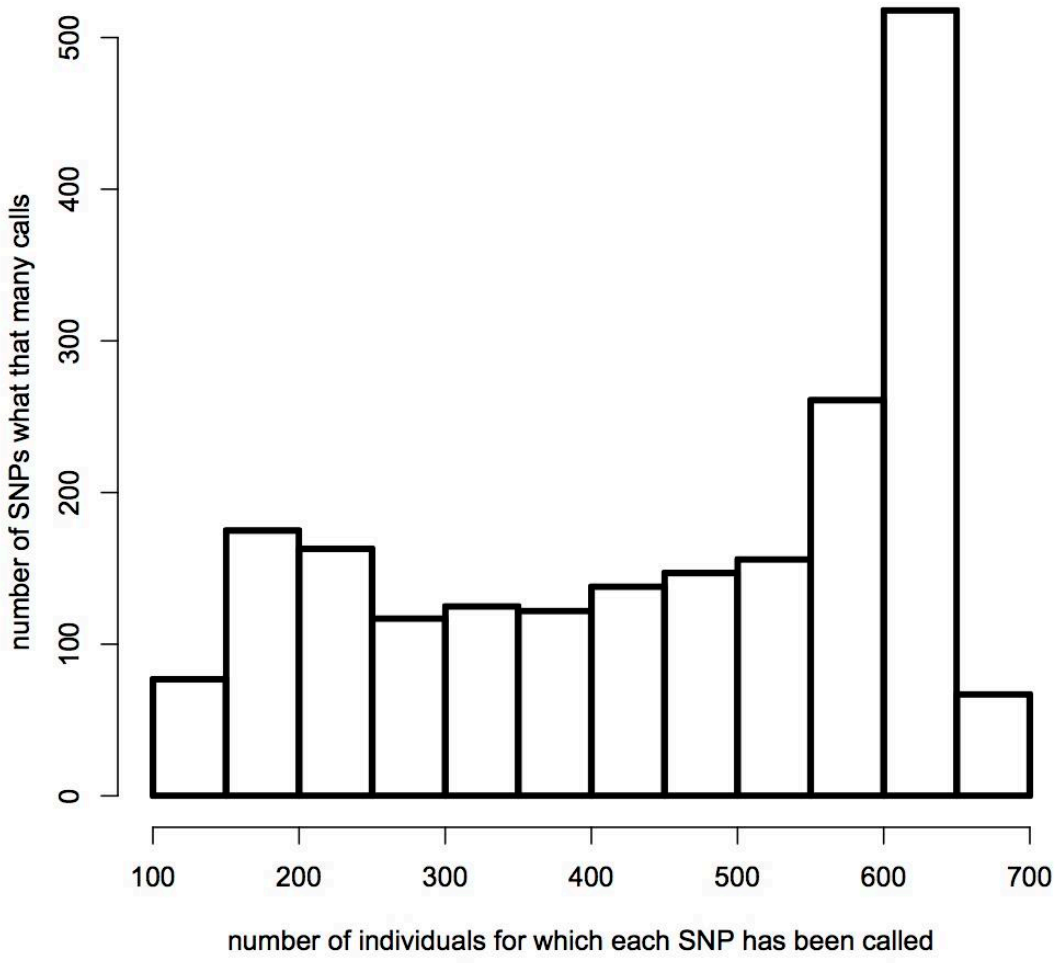


Figure 3: Distribution of the number of individuals for which we have different numbers of SNPs called. Most individuals are called for more than 1000 SNPs.

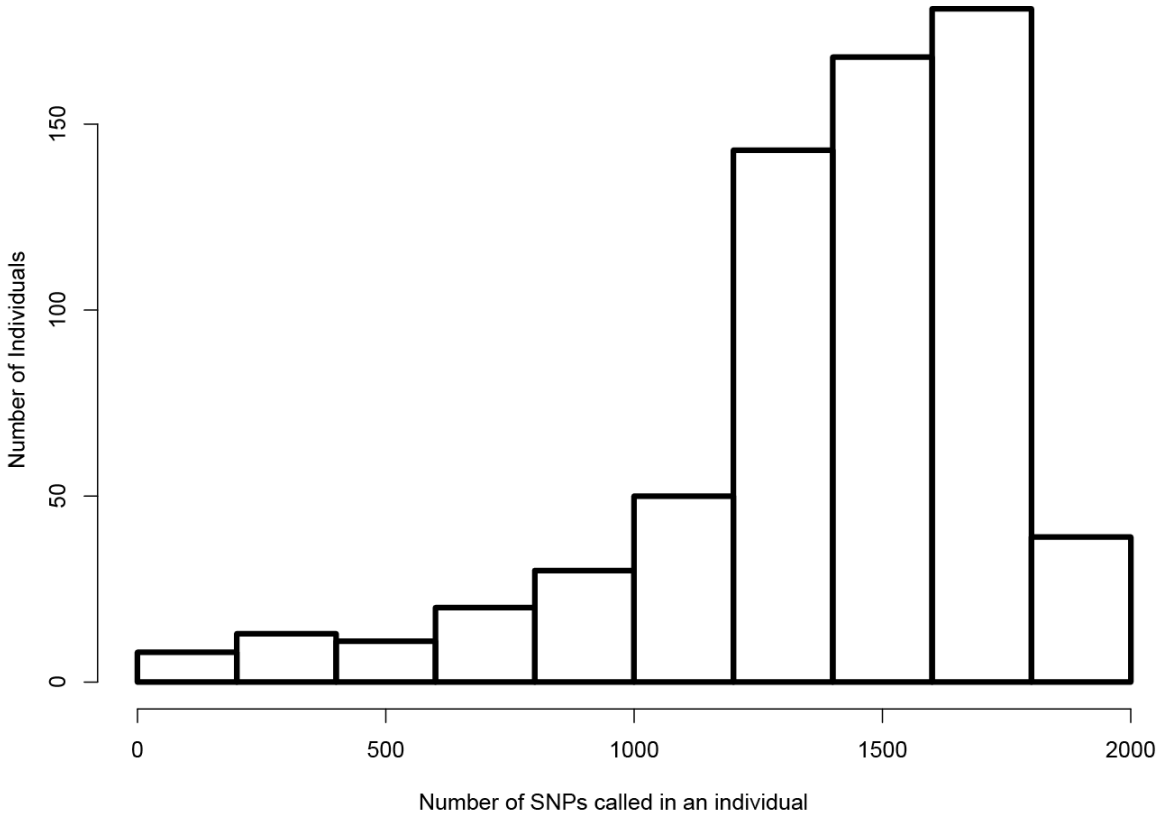


Figure 4: Percent coverage of SNPs in each population. Most SNPs have high coverage in each population.

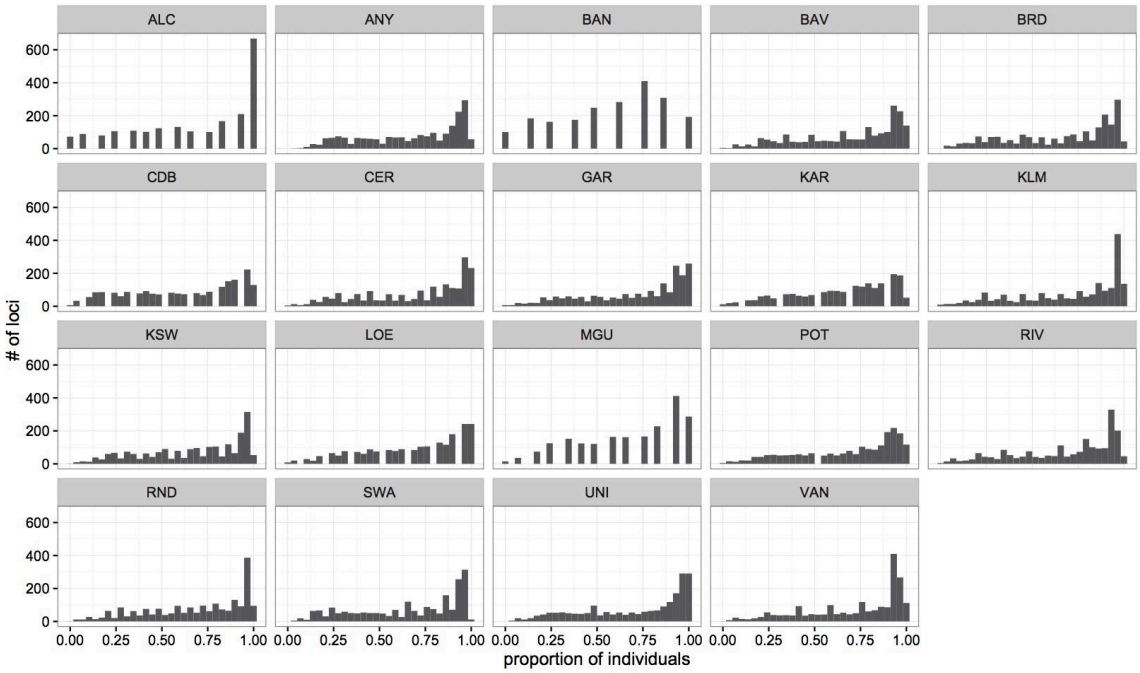
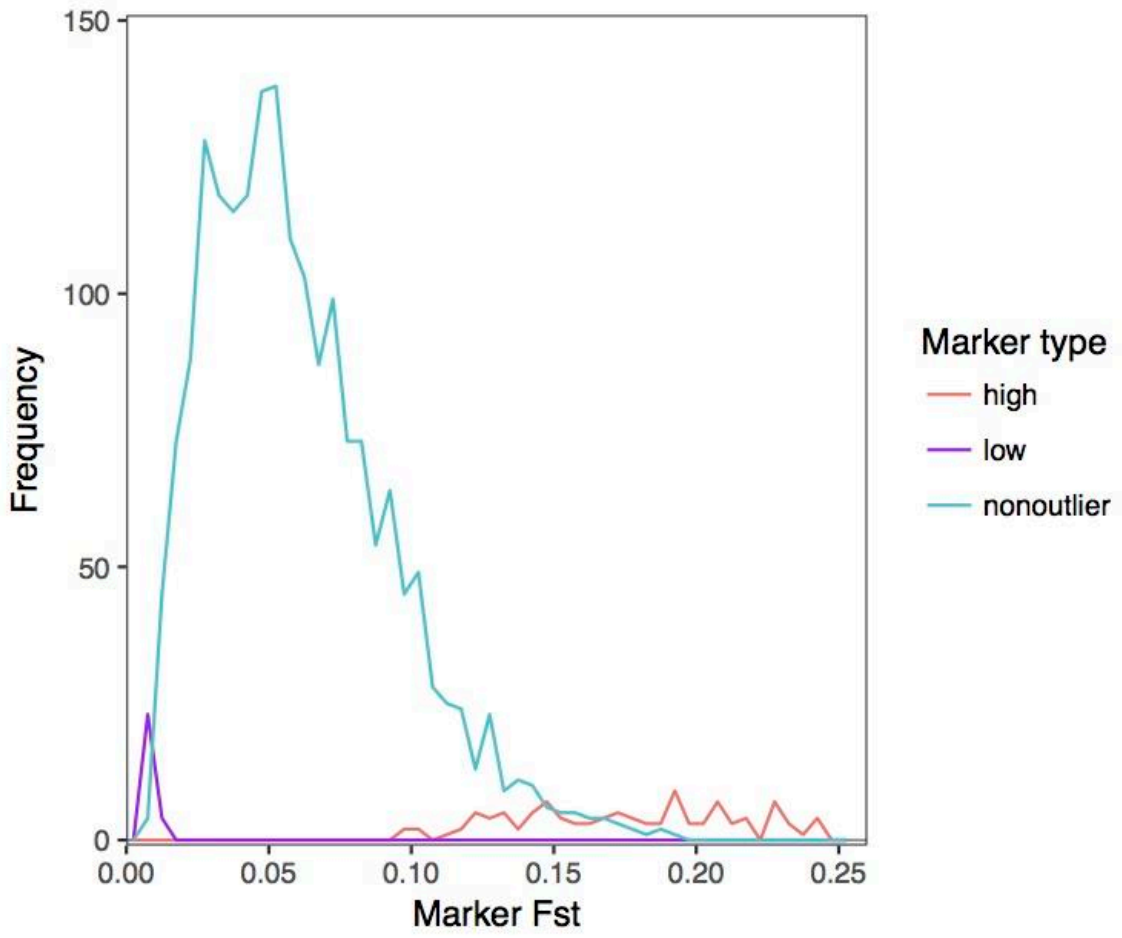


Figure 5: Distribution of low outlier, non-outlier, and high outlier SNPs from the final set of 2066 loci.



Online Supplemental 3 Click here to download Online Supplemental Appendix_S3.docx

Appendix S3. Loadings of environmental variables of all known *Protea repens* populations (Rebelo, 2006) onto the principal component axes. Variables with large loadings are bolded. Below the line, the proportion of the variance explained by each axis and the cumulative proportion of variance explained are displayed.

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5	Axis 6	Axis 7
Rainfall Seasonality		**0.72**	0.292	-0.236	-0.221	0.533	
Mean Annual Precipitation	-0.166		**0.779**	-0.256		**-0.538**	
Max. Temp. January	**-0.428**		-0.295	**-0.732**			-0.434
Mean Annual Temp.	**-0.534**		-0.187		0.129		**0.808**
Altitude	**0.509**			-0.378	**0.734**	0.17	0.171
Min. Temp. July	**-0.474**	0.244		0.44	**0.619**	0.126	-0.34
Summer Rainfall proportion	-0.143	**-0.643**	0.423			**0.616**	
Proportion of variance	0.471	0.228	0.199	0.072	0.017	0.011	0.002
Cumulative Proportion	0.471	0.699	0.898	0.970	0.987	0.998	1.000

Online Supplemental 4 Click here to download Online Supplemental Appendix_S4_blasted_outliers.docx

Appendix S4. The high F_{ST} outliers which blasted to the *Protea repens* transcriptome, the contig to which it blasted (see Akman et al 2015) and the e-value for the blast hit, the *Arabidopsis thaliana* gene to which the outlier marker blasted, and a short annotation of the function of the *A. thaliana* gene drawn from TAIR (https://www.arabidopsis.org)

Marker Name	transcriptome blast	evalue	AT gene	short description AT gene
TP97882	contig52281	1.00E-12	AT4G14690	assembly of photosynthetic machinery
TP26245	contig80099	2.00E-08	AT5G48930	auxin transport and growth
TP62904	contig90023	3.00E-26	AT5G48930	auxin transport and growth
TP73820	contig90023	4.00E-24	AT5G48930	auxin transport and growth
TP102588	contig32440	3.00E-26	AT3G15790	regulatory protein
TP104681	contig100792	1.00E-23	AT2G21710	seed dormancy
TP76683	contig109003	2.00E-16	AT2G17260	stomatal closure
TP61066	contig93861	2.00E-15	AT1G08960	sodium-calcium exchanger
TP106423	contig90234	1.00E-24	AT2G41900	zinc finger
TP6429	contig50538	5.00E-23	AT3G17380	TRAF, cell survival
TP28934	contig111156	6.00E-22	AT3G27260	housekeeping
TP13583	contig19992	3.00E-25	AT4G17880	jasmonic acid pathway
TP72119	contig4451	1.00E-24	AT4G34500	protein kinase
TP53116	contig115744	1.00E-24	AT5G08610	helicase activity
TP66086	contig100284	2.00E-22		
TP21716	contig51049	1.00E-24		
TP28454	contig52494	1.00E-24		
TP111004	contig66148	4.00E-12		
TP29697	contig68943	2.00E-16		
TP116767	contig70715	2.00E-14		
TP92897	contig71220	2.00E-15		
TP41985	contig73758	0.075		
TP22349	contig76599	1.00E-24		
TP7698	contig87039	4.00E-12		
TP5893	contig9040	1.00E-24		
TP63099	contig93726	4.00E-24		

Online Supplemental 5

Click here to download Online Supplemental Appendix_S5_blasted_non_outliers.xlsx

Appendix S5. The non-outlier loci which blasted to the Protea repens transcriptome, the contig to which it blasted (see Akman et al 2015) and the e-value for the blast hit, the Arabidopsis thaliana gene to which the outlier marker blasted

marker name	Trascriptome Contig	evalue	A. thaliana gene
TP15266	contig112859	2.00E-16	AT1G01320
TP19471	contig34561	1.00E-24	AT1G01320
TP111743	contig29656	3.00E-26	AT1G01860
TP65811	contig83175	1.00E-23	AT1G02730
TP22651	contig93161	1.00E-24	AT1G04960
TP93734	contig93161	1.00E-24	AT1G04960
TP6544	contig60639	1.00E-24	AT1G05180
TP9380	contig77952	1.00E-18	AT1G05940
TP29259	contig85705	5.00E-23	AT1G07710
TP7382	contig82301	7.00E-15	AT1G07710
TP129244	contig31812	1.00E-24	AT1G08070
TP969	contig84596	3.00E-26	AT1G09280
TP128806	contig2191	3.00E-26	AT1G09710
TP123127	contig77060	1.00E-24	AT1G10200
TP81633	contig7448	2.00E-15	AT1G10760
TP19861	contig6928	1.00E-24	AT1G11915
TP129782	contig15137	3.00E-26	AT1G12770
TP129216	contig85345	3.00E-13	AT1G13770
TP93752	contig85345	4.00E-24	AT1G13770
TP100210	contig95878	9.00E-14	AT1G14410
TP12114	contig117946	1.00E-11	AT1G14410
TP9470	contig95878	3.00E-26	AT1G14410
TP17294	contig43142	1.00E-24	AT1G16080
TP45567	contig43142	9.00E-14	AT1G16080
TP4614	contig108663	3.00E-26	AT1G17650
TP123887	contig98925	6.00E-22	AT1G17750
TP9911	contig22867	1.00E-24	AT1G18740
TP109153	contig87905	1.00E-24	AT1G19220
TP1048	contig11355	3.00E-26	AT1G20830
TP18429	contig15796	3.00E-26	AT1G21410
TP80448	contig62295	1.00E-24	AT1G23180
TP107335	contig94679	3.00E-26	AT1G25320
TP66321	contig63842	3.00E-26	AT1G28700
TP15162	contig61908	8.00E-21	AT1G29880
TP17037	contig69525	3.00E-26	AT1G30330

TP57672	contig69525	1.00E-24	AT1G30330
TP45308	contig56016	1.00E-17	AT1G30470
TP80041	contig10781	1.00E-24	AT1G30500
TP49646	contig24038	1.00E-24	AT1G30570
TP17707	contig45917	1.00E-24	AT1G31730
TP34314	contig94528	3.00E-25	AT1G32150
TP94299	contig41019	3.00E-26	AT1G33330
TP107183	contig90942	1.00E-23	AT1G35510
TP31668	contig114183	1.00E-24	AT1G36180
TP41935	contig96324	3.00E-13	AT1G47710
TP105025	contig81546	1.00E-24	AT1G48760
TP51408	contig81546	1.00E-24	AT1G48760
TP66681	contig81546	3.00E-26	AT1G48760
TP93817	contig81546	1.00E-24	AT1G48760
TP16765	contig53108	1.00E-24	AT1G49590
TP91048	contig29999	3.00E-26	AT1G50030
TP1860	contig49699	1.00E-24	AT1G50120
TP76133	contig117142	4.00E-18	AT1G50130
TP71643	contig42662	9.00E-14	AT1G51040
TP24368	contig40704	1.00E-24	AT1G51060
TP88152	contig40704	1.00E-24	AT1G51060
TP3502	contig59243	9.00E-14	AT1G51160
TP28247	contig14460	5.00E-23	AT1G51340
TP119400	contig89590	3.00E-26	AT1G51550
TP71578	contig94851	9.00E-20	AT1G51760
TP127290	contig73369	1.00E-24	AT1G53050
TP54006	contig73369	3.00E-26	AT1G53050
TP119261	contig21178	3.00E-26	AT1G53710
TP125588	contig84275	1.00E-23	AT1G55090
TP32387	contig39161	3.00E-26	AT1G55530
TP95509	contig39161	1.00E-24	AT1G55530
TP29442	contig64134	3.00E-26	AT1G55620
TP9386	contig7353	3.00E-19	AT1G55760
TP29765	contig96971	3.00E-26	AT1G56550
TP29768	contig96971	2.00E-16	AT1G56550
TP128131	contig48713	3.00E-26	AT1G59750
TP107433	contig110542	1.00E-23	AT1G60080
TP64839	contig565	1.00E-24	AT1G60490
TP4406	contig7101	1.00E-24	AT1G61040
TP113691	contig82870	1.00E-24	AT1G62400
TP3781	contig58110	1.00E-23	AT1G63610

TP46680	contig99439	4.00E-18	AT1G64450
TP130168	contig59202	3.00E-26	AT1G65020
TP53834	contig53802	3.00E-26	AT1G67510
TP70187	contig112649	3.00E-26	AT1G67590
TP4763	contig12850	2.00E-15	AT1G70130
TP6581	contig12850	2.00E-14	AT1G70130
TP25826	contig58528	2.00E-15	AT1G70320
TP3741	contig55961	1.00E-24	AT1G73350
TP15380	contig18440	1.00E-23	AT1G74260
TP19171	contig97438	2.00E-15	AT1G74260
TP94094	contig39350	1.00E-24	AT1G74990
TP19246	contig16020	1.00E-24	AT1G76490
TP29345	contig110236	3.00E-25	AT1G76540
TP59385	contig110236	1.00E-11	AT1G76540
TP111647	contig7274	4.00E-11	AT1G76850
TP22815	contig74383	9.00E-14	AT1G76940
TP77824	contig74383	1.00E-24	AT1G76940
TP92991	contig7931	1.00E-18	AT1G77120
TP92992	contig7931	1.00E-18	AT1G77120
TP53293	contig14612	1.00E-23	AT1G77670
TP53294	contig14612	1.00E-23	AT1G77670
TP96564	contig79561	3.00E-26	AT1G78900
TP16718	contig55714	9.00E-14	AT1G79650
TP127077	contig12103	3.00E-26	AT1G79690
TP29744	contig99097	3.00E-26	AT1G80260
TP15810	contig92908	3.00E-20	AT1G80760
TP22276	contig82298	3.00E-26	AT1G80930
TP44533	contig48960	4.00E-11	AT1G80930
TP41084	contig92015	1.00E-11	AT2G01610
TP43013	contig59248	4.00E-11	AT2G15790
TP42951	contig37174	1.00E-24	AT2G16230
TP6254	contig47416	2.00E-15	AT2G16370
TP92890	contig79864	8.00E-21	AT2G16370
TP6454	contig45015	3.00E-26	AT2G18040
TP24025	contig36417	4.00E-11	AT2G20760
TP87630	contig9741	6.00E-16	AT2G21280
TP30645	contig7047	3.00E-26	AT2G21385
TP56419	contig7048	1.00E-24	AT2G21385
TP26579	contig92584	1.00E-24	AT2G21390
TP28513	contig92119	1.00E-23	AT2G21410
TP125620	contig65464	6.00E-22	AT2G22125

TP27239	contig104454	1.00E-24	AT2G22800
TP46971	contig13340	1.00E-23	AT2G22840
TP47613	contig17776	2.00E-14	AT2G24060
TP1546	contig11691	1.00E-24	AT2G24100
TP74414	contig8331	1.00E-11	AT2G24260
TP28655	contig39838	2.00E-22	AT2G25800
TP109362	contig68709	1.00E-23	AT2G25840
TP19170	contig68709	3.00E-26	AT2G25840
TP113703	contig62634	2.00E-15	AT2G25970
TP44126	contig69236	3.00E-19	AT2G26780
TP50551	contig24446	3.00E-26	AT2G27210
TP125070	contig13745	1.00E-24	AT2G31410
TP29856	contig81581	3.00E-26	AT2G32950
TP38119	contig50276	1.00E-23	AT2G34040
TP78175	contig42106	3.00E-19	AT2G35050
TP37053	contig35868	5.00E-17	AT2G35410
TP61757	contig17891	1.00E-24	AT2G35630
TP7341	contig17891	3.00E-26	AT2G35630
TP114021	contig72357	7.00E-15	AT2G36380
TP118532	contig2096	3.00E-26	AT2G37310
TP125192	contig2096	1.00E-24	AT2G37310
TP7356	contig23722	5.00E-23	AT2G38680
TP24042	contig41171	9.00E-14	AT2G38760
TP120305	contig48623	3.00E-26	AT2G39740
TP66281	contig48623	1.00E-24	AT2G39740
TP123583	contig44949	2.00E-15	AT2G40540
TP28390	contig83407	9.00E-14	AT2G40610
TP106298	contig75811	3.00E-26	AT2G41940
TP119198	contig75811	1.00E-24	AT2G41940
TP30825	contig57796	1.00E-24	AT2G42590
TP14657	contig8592	3.00E-25	AT2G43250
TP28449	contig8592	2.00E-22	AT2G43250
TP92629	contig8592	8.00E-21	AT2G43250
TP129048	contig77438	4.00E-12	AT2G45630
TP97913	contig45498	3.00E-26	AT2G45670
TP60370	contig79526	4.00E-12	AT2G46020
TP1820	contig92731	1.00E-24	AT2G46280
TP47589	contig17992	4.00E-11	AT2G46550
TP130802	contig24796	3.00E-26	AT2G47070
TP61348	contig24796	1.00E-25	AT2G47070
TP5949	contig7522	1.00E-24	AT2G47440

TP28452	contig56627	3.00E-26	AT2G48070
TP105408	contig77445	8.00E-21	AT3G06060
TP78403	contig53327	2.00E-15	AT3G07660
TP69519	contig9224	1.00E-23	AT3G11820
TP117936	contig43709	3.00E-26	AT3G12280
TP14796	contig33256	3.00E-26	AT3G12530
TP106698	contig14503	4.00E-18	AT3G13300
TP32829	contig57001	3.00E-26	AT3G13900
TP128021	contig65944	1.00E-24	AT3G14240
TP29311	contig51146	1.00E-23	AT3G14930
TP119729	contig107182	1.00E-24	AT3G15220
TP127274	contig54811	3.00E-26	AT3G15260
TP51202	contig52576	9.00E-14	AT3G16490
TP71362	contig57037	8.00E-21	AT3G16630
TP92578	contig104416	1.00E-23	AT3G18040
TP17681	contig820	3.00E-26	AT3G19490
TP102597	contig12957	1.00E-24	AT3G19900
TP18270	contig88988	1.00E-24	AT3G20150
TP109712	contig46527	6.00E-16	AT3G21690
TP124343	contig109068	2.00E-14	AT3G21690
TP90857	contig3723	1.00E-24	AT3G23580
TP101549	contig15709	8.00E-21	AT3G24350
TP101550	contig15709	3.00E-19	AT3G24350
TP119673	contig83548	3.00E-26	AT3G24506
TP119450	contig81576	4.00E-18	AT3G24660
TP30590	contig83506	1.00E-23	AT3G25110
TP129026	contig59984	1.00E-12	AT3G25120
TP67254	contig89830	1.00E-24	AT3G26640
TP72560	contig82649	3.00E-26	AT3G28345
TP2794	contig28696	3.00E-26	AT3G45090
TP13918	contig4934	3.00E-26	AT3G45190
TP79684	contig83236	3.00E-26	AT3G45242
TP106295	contig106924	2.00E-22	AT3G47850
TP109386	contig31350	1.00E-24	AT3G48050
TP26171	contig89255	8.00E-21	AT3G48200
TP2057	contig19712	1.00E-24	AT3G50110
TP47841	contig84216	3.00E-26	AT3G50920
TP53622	contig98124	3.00E-25	AT3G51010
TP16015	contig9570	3.00E-26	AT3G51040
TP125580	contig34556	3.00E-26	AT3G51670
TP32098	contig44543	6.00E-22	AT3G52140

TP43512	contig44543	3.00E-26	AT3G52140
TP3642	contig90326	3.00E-19	AT3G52200
TP68996	contig78410	1.00E-24	AT3G52370
TP99558	contig43048	3.00E-26	AT3G52390
TP105982	contig39255	1.00E-24	AT3G54230
TP93064	contig6294	1.00E-24	AT3G54230
TP22066	contig40681	3.00E-26	AT3G54790
TP62868	contig67102	1.00E-12	AT3G54820
TP104715	contig63439	1.00E-24	AT3G54890
TP106688	contig62910	4.00E-18	AT3G55220
TP4153	contig116002	2.00E-21	AT3G55220
TP30704	contig24640	3.00E-26	AT3G55320
TP32580	contig45894	4.00E-12	AT3G55590
TP60551	contig45271	3.00E-26	AT3G55760
TP110946	contig25807	3.00E-25	AT3G56430
TP118517	contig42986	2.00E-15	AT3G57150
TP117733	contig96829	3.00E-25	AT3G59300
TP66015	contig52220	9.00E-20	AT3G60750
TP18913	contig8164	3.00E-26	AT3G60920
TP78553	contig97119	3.00E-26	AT3G62110
TP20758	contig57595	1.00E-24	AT3G63500
TP120581	contig78753	8.00E-21	AT4G01400
TP5919	contig78753	3.00E-26	AT4G01400
TP29550	contig99562	3.00E-25	AT4G01880
TP113209	contig10086	3.00E-26	AT4G02070
TP50509	contig5750	1.00E-24	AT4G02320
TP124478	contig100566	3.00E-26	AT4G02560
TP20227	contig26658	1.00E-18	AT4G02580
TP43745	contig20961	1.00E-24	AT4G06599
TP119429	contig67861	1.00E-24	AT4G10960
TP130387	contig31971	3.00E-26	AT4G12710
TP107175	contig29380	3.00E-26	AT4G12790
TP62963	contig6621	3.00E-26	AT4G14300
TP31611	contig52280	1.00E-23	AT4G14690
TP12386	contig73655	3.00E-26	AT4G14740
TP124967	contig73655	1.00E-24	AT4G14740
TP19356	contig73655	2.00E-14	AT4G14740
TP62086	contig73655	1.00E-25	AT4G14740
TP113177	contig29809	1.00E-18	AT4G15560
TP19499	contig77075	1.00E-24	AT4G17330
TP68901	contig16965	1.00E-24	AT4G17350

TP69409	contig16965	2.00E-14	AT4G17350
TP9989	contig90576	1.00E-18	AT4G17915
TP79113	contig72065	2.00E-21	AT4G18050
TP62172	contig25265	1.00E-24	AT4G18570
TP90918	contig22185	1.00E-23	AT4G20400
TP129867	contig46590	3.00E-20	AT4G21720
TP17225	contig86330	3.00E-19	AT4G22758
TP54927	contig86330	2.00E-15	AT4G22758
TP28327	contig93180	2.00E-14	AT4G24180
TP100482	contig47996	3.00E-25	AT4G24330
TP63455	contig89028	3.00E-26	AT4G24690
TP4491	contig37140	1.00E-24	AT4G24800
TP50290	contig73621	1.00E-24	AT4G25880
TP47262	contig91814	9.00E-14	AT4G27010
TP23253	contig15776	9.00E-14	AT4G27340
TP23254	contig15776	9.00E-14	AT4G27340
TP113263	contig79003	3.00E-19	AT4G28080
TP125204	contig37687	4.00E-12	AT4G28470
TP80297	contig16604	1.00E-24	AT4G28910
TP77182	contig22819	3.00E-19	AT4G29160
TP26290	contig43609	1.00E-24	AT4G29210
TP42768	contig35166	3.00E-26	AT4G29790
TP27702	contig93942	3.00E-25	AT4G30220
TP100348	contig25249	2.00E-16	AT4G32400
TP47176	contig86386	3.00E-13	AT4G33350
TP43006	contig6489	1.00E-11	AT4G33410
TP61469	contig4401	3.00E-26	AT4G34320
TP126798	contig4451	3.00E-26	AT4G34500
TP119403	contig30684	3.00E-13	AT4G35100
TP97002	contig74327	3.00E-26	AT4G35800
TP88079	contig7494	1.00E-24	AT4G36945
TP3530	contig96566	3.00E-26	AT4G37200
TP80716	contig18059	1.00E-24	AT4G37250
TP128751	contig58021	1.00E-24	AT4G37510
TP93116	contig74949	6.00E-22	AT4G37640
TP1352	contig96464	1.00E-24	AT4G37880
TP106795	contig50770	8.00E-21	AT4G38200
TP27483	contig23683	4.00E-11	AT4G38650
TP32865	contig23683	3.00E-26	AT4G38650
TP7547	contig60815	3.00E-26	AT4G39420
TP29933	contig87256	1.00E-24	AT4G39690

TP54762	contig87256	3.00E-25	AT4G39690
TP102341	contig750	6.00E-22	AT5G01330
TP8804	contig750	1.00E-24	AT5G01330
TP50256	contig28516	3.00E-26	AT5G01410
TP108485	contig67922	3.00E-26	AT5G03280
TP3912	contig67922	1.00E-24	AT5G03280
TP69147	contig61013	2.00E-22	AT5G03450
TP64984	contig113467	3.00E-26	AT5G05660
TP36646	contig68215	3.00E-26	AT5G08050
TP30250	contig58356	1.00E-24	AT5G09380
TP31776	contig88810	3.00E-26	AT5G09950
TP113100	contig667	3.00E-26	AT5G10030
TP53668	contig667	1.00E-24	AT5G10030
TP122062	contig36225	1.00E-24	AT5G10050
TP127239	contig49223	5.00E-17	AT5G10370
TP26168	contig77102	1.00E-24	AT5G10490
TP4437	contig77102	1.00E-24	AT5G10490
TP41344	contig42851	1.00E-24	AT5G10920
TP72716	contig42851	1.00E-24	AT5G10920
TP685	contig2784	3.00E-26	AT5G11490
TP99855	contig2553	2.00E-15	AT5G12200
TP46649	contig110642	3.00E-26	AT5G13000
TP111726	contig49654	2.00E-15	AT5G13870
TP18894	contig12482	3.00E-13	AT5G14550
TP119759	contig117795	3.00E-26	AT5G16690
TP22363	contig117795	3.00E-26	AT5G16690
TP124791	contig35253	3.00E-26	AT5G17310
TP105557	contig11025	1.00E-23	AT5G19050
TP709	contig97692	3.00E-26	AT5G19760
TP1510	contig73065	5.00E-23	AT5G20680
TP113363	contig16268	9.00E-14	AT5G20730
TP1265	contig977	1.00E-24	AT5G23110
TP93364	contig977	3.00E-26	AT5G23110
TP74447	contig75894	1.00E-24	AT5G23535
TP14227	contig52941	3.00E-25	AT5G24060
TP43455	contig33537	1.00E-24	AT5G24320
TP9775	contig36199	3.00E-26	AT5G25110
TP23117	contig84756	1.00E-24	AT5G25940
TP99424	contig94268	4.00E-12	AT5G35320
TP69514	contig16795	3.00E-26	AT5G36210
TP47223	contig111173	1.00E-24	AT5G36880

TP69594	contig53746	2.00E-15	AT5G36890
TP10257	contig11462	1.00E-24	AT5G39320
TP33602	contig57426	1.00E-24	AT5G39650
TP65162	contig57426	3.00E-26	AT5G39650
TP129482	contig91809	1.00E-24	AT5G40010
TP45274	contig34956	2.00E-22	AT5G40020
TP71702	contig34613	1.00E-24	AT5G42080
TP15149	contig51587	3.00E-26	AT5G42180
TP15980	contig81128	3.00E-26	AT5G44050
TP60347	contig85885	9.00E-14	AT5G46250
TP96133	contig28334	2.00E-22	AT5G47390
TP2938	contig9677	1.00E-23	AT5G48385
TP7773	contig35518	5.00E-17	AT5G49830
TP128302	contig18599	3.00E-26	AT5G50890
TP16960	contig93652	2.00E-16	AT5G51940
TP95570	contig54239	3.00E-25	AT5G54580
TP15245	contig59089	1.00E-24	AT5G54630
TP129827	contig84339	3.00E-26	AT5G54940
TP71941	contig99575	1.00E-24	AT5G55860
TP125373	contig72130	1.00E-24	AT5G55920
TP25726	contig70816	4.00E-24	AT5G56750
TP126842	contig13098	8.00E-21	AT5G58300
TP78789	contig46280	3.00E-26	AT5G59910
TP100128	contig39093	3.00E-26	AT5G60710
TP99529	contig89954	1.00E-24	AT5G60710
TP104875	contig46769	4.00E-18	AT5G60750
TP118426	contig46769	2.00E-21	AT5G60750
TP78071	contig27562	3.00E-26	AT5G61500
TP73371	contig79587	9.00E-14	AT5G61730
TP123118	contig95639	3.00E-26	AT5G61930
TP9737	contig95639	3.00E-26	AT5G61930
TP37891	contig91125	3.00E-26	AT5G63950
TP13121	contig34032	3.00E-26	AT5G64050
TP1503	contig11074	2.00E-15	AT5G64780
TP110519	contig9535	1.00E-24	AT5G64940
TP32668	contig9535	1.00E-24	AT5G64940
TP4771	contig88462	3.00E-26	AT5G65110
TP99509	contig77550	1.00E-17	AT5G65470
TP25877	contig34374	6.00E-22	AT5G65730
TP93001	contig988	1.00E-24	AT5G65940
TP18902	contig26236	3.00E-26	AT5G66030

TP29557	contig26231	1.00E-24	AT5G66030
TP71396	contig26236	6.00E-22	AT5G66030
TP126160	contig58522	3.00E-26	ATCG00350
TP100091	contig73358	1.00E-24	#N/A
TP100310	contig68091	3.00E-26	#N/A
TP102051	contig21185	3.00E-26	#N/A
TP102124	contig27640	1.00E-24	#N/A
TP102230	contig68451	1.00E-24	#N/A
TP102333	contig90102	9.00E-20	#N/A
TP104190	contig42869	2.00E-15	#N/A
TP104213	contig78217	1.00E-24	#N/A
TP104224	contig68681	1.00E-24	#N/A
TP104952	contig110497	1.00E-24	#N/A
TP106093	contig94841	1.00E-12	#N/A
TP106221	contig51741	2.00E-21	#N/A
TP106439	contig51742	2.00E-15	#N/A
TP106487	contig52713	3.00E-26	#N/A
TP107352	contig91983	1.00E-24	#N/A
TP10828	contig81950	1.00E-24	#N/A
TP108484	contig71825	4.00E-12	#N/A
TP10851	contig96757	3.00E-26	#N/A
TP108873	contig112569	4.00E-18	#N/A
TP109135	contig14853	4.00E-18	#N/A
TP113088	contig47630	1.00E-24	#N/A
TP113700	contig73813	1.00E-24	#N/A
TP114474	contig35176	1.00E-23	#N/A
TP114669	contig76925	3.00E-19	#N/A
TP116059	contig6538	1.00E-24	#N/A
TP11619	contig86370	1.00E-24	#N/A
TP116940	contig80079	2.00E-16	#N/A
TP117179	contig93400	3.00E-19	#N/A
TP117796	contig88808	8.00E-21	#N/A
TP117971	contig118413	3.00E-26	#N/A
TP118838	contig36624	3.00E-26	#N/A
TP118886	contig64847	1.00E-24	#N/A
TP120101	contig109392	3.00E-26	#N/A
TP121692	contig35172	1.00E-25	#N/A
TP123069	contig13156	3.00E-26	#N/A
TP123196	contig117605	3.00E-20	#N/A
TP123468	contig73927	2.00E-22	#N/A
TP123896	contig55051	1.00E-24	#N/A

TP124121	contig44074	1.00E-24	#N/A
TP124236	contig25342	5.00E-23	#N/A
TP124265	contig64686	3.00E-26	#N/A
TP124315	contig71764	3.00E-26	#N/A
TP124438	contig50746	1.00E-24	#N/A
TP124505	contig89711	8.00E-21	#N/A
TP124727	contig78928	3.00E-26	#N/A
TP124768	contig106471	1.00E-24	#N/A
TP125246	contig7081	3.00E-26	#N/A
TP125363	contig87052	3.00E-26	#N/A
TP125628	contig99966	1.00E-24	#N/A
TP125639	contig56913	3.00E-26	#N/A
TP125803	contig63403	3.00E-26	#N/A
TP125945	contig37704	7.00E-15	#N/A
TP127094	contig10232	3.00E-26	#N/A
TP127876	contig113289	1.00E-24	#N/A
TP127985	contig89274	5.00E-23	#N/A
TP128065	contig69267	1.00E-24	#N/A
TP128411	contig88808	9.00E-14	#N/A
TP128926	contig77648	1.00E-25	#N/A
TP129362	contig69394	1.00E-24	#N/A
TP129449	contig46217	1.00E-24	#N/A
TP129550	contig42076	3.00E-26	#N/A
TP129732	contig69043	1.00E-24	#N/A
TP129860	contig16029	3.00E-26	#N/A
TP130335	contig49381	6.00E-22	#N/A
TP130519	contig83291	3.00E-26	#N/A
TP130546	contig73352	1.00E-24	#N/A
TP130605	contig93014	3.00E-26	#N/A
TP13084	contig100944	9.00E-14	#N/A
TP13711	contig82151	1.00E-24	#N/A
TP1467	contig96485	6.00E-16	#N/A
TP14730	contig84673	1.00E-11	#N/A
TP14731	contig54972	9.00E-20	#N/A
TP14771	contig54972	9.00E-14	#N/A
TP14773	contig51742	9.00E-14	#N/A
TP14787	contig14853	3.00E-13	#N/A
TP14911	contig35172	1.00E-25	#N/A
TP14930	contig112222	4.00E-12	#N/A
TP15193	contig93565	3.00E-26	#N/A
TP16282	contig83982	1.00E-18	#N/A

TP16421	contig65701	4.00E-11	#N/A
TP17298	contig78792	1.00E-11	#N/A
TP17437	contig71825	9.00E-20	#N/A
TP17685	contig100389	3.00E-26	#N/A
TP1770	contig77074	3.00E-26	#N/A
TP17892	contig88069	3.00E-26	#N/A
TP1899	contig56996	1.00E-24	#N/A
TP19117	contig90645	3.00E-26	#N/A
TP19228	contig55781	3.00E-26	#N/A
TP19574	contig33893	1.00E-24	#N/A
TP1982	contig30552	6.00E-16	#N/A
TP2021	contig23569	3.00E-26	#N/A
TP20804	contig34950	4.00E-24	#N/A
TP20943	contig80290	1.00E-24	#N/A
TP21343	contig112222	1.00E-24	#N/A
TP21726	contig69653	3.00E-26	#N/A
TP22222	contig99748	2.00E-14	#N/A
TP22611	contig104269	1.00E-23	#N/A
TP22992	contig105392	3.00E-26	#N/A
TP23133	contig32521	3.00E-26	#N/A
TP23148	contig71181	3.00E-26	#N/A
TP23733	contig106491	6.00E-22	#N/A
TP2376	contig110497	3.00E-26	#N/A
TP23787	contig13364	2.00E-15	#N/A
TP23863	contig99591	1.00E-25	#N/A
TP24059	contig119793	3.00E-26	#N/A
TP243	contig90243	1.00E-24	#N/A
TP24385	contig83297	9.00E-14	#N/A
TP25375	contig91983	6.00E-22	#N/A
TP25400	contig75690	2.00E-15	#N/A
TP25568	contig69817	3.00E-26	#N/A
TP26302	contig80582	3.00E-26	#N/A
TP26670	contig100389	3.00E-20	#N/A
TP26956	contig58771	8.00E-21	#N/A
TP27508	contig32108	3.00E-26	#N/A
TP27681	contig17428	3.00E-26	#N/A
TP28938	contig4203	3.00E-26	#N/A
TP29047	contig106041	5.00E-23	#N/A
TP29161	contig13158	1.00E-17	#N/A
TP29420	contig88318	1.00E-24	#N/A
TP29433	contig52503	8.00E-21	#N/A

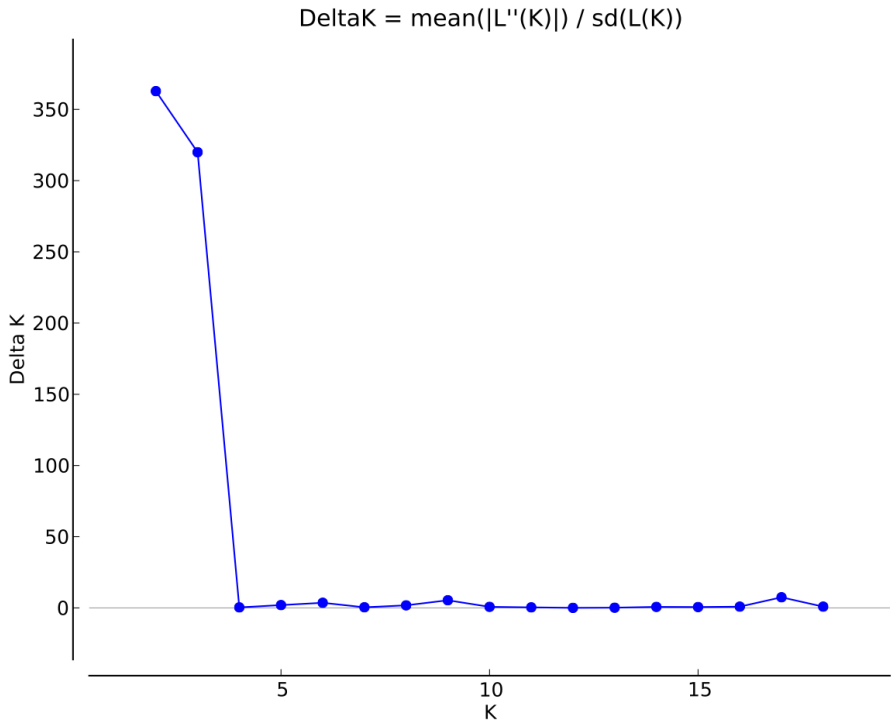
TP29470	contig41284	3.00E-26	#N/A
TP29672	contig110428	1.00E-23	#N/A
TP29674	contig110428	8.00E-21	#N/A
TP30053	contig111404	3.00E-26	#N/A
TP30328	contig83592	1.00E-12	#N/A
TP30456	contig63800	3.00E-26	#N/A
TP30631	contig20010	1.00E-24	#N/A
TP30708	contig64138	1.00E-24	#N/A
TP3096	contig74625	3.00E-26	#N/A
TP31499	contig72412	8.00E-21	#N/A
TP32280	contig114243	1.00E-24	#N/A
TP32597	contig80581	1.00E-24	#N/A
TP32802	contig19718	2.00E-14	#N/A
TP32988	contig5996	1.00E-24	#N/A
TP33023	contig86538	1.00E-24	#N/A
TP33281	contig23636	5.00E-23	#N/A
TP33597	contig104151	1.00E-24	#N/A
TP34584	contig91682	3.00E-26	#N/A
TP34627	contig109664	1.00E-24	#N/A
TP36173	contig109392	3.00E-26	#N/A
TP36660	contig2153	9.00E-20	#N/A
TP37572	contig28929	1.00E-24	#N/A
TP38360	contig68579	2.00E-14	#N/A
TP39236	contig50711	3.00E-26	#N/A
TP39291	contig81732	3.00E-26	#N/A
TP39433	contig93400	1.00E-17	#N/A
TP41341	contig98612	1.00E-24	#N/A
TP41524	contig110629	1.00E-24	#N/A
TP41983	contig42151	3.00E-26	#N/A
TP42461	contig15969	3.00E-26	#N/A
TP43029	contig100443	3.00E-26	#N/A
TP43736	contig82221	3.00E-26	#N/A
TP44461	contig82871	3.00E-26	#N/A
TP46600	contig15737	1.00E-24	#N/A
TP4681	contig65213	3.00E-26	#N/A
TP47135	contig50711	1.00E-24	#N/A
TP47808	contig114600	3.00E-26	#N/A
TP48049	contig60429	8.00E-21	#N/A
TP48428	contig87355	7.00E-15	#N/A
TP493	contig97032	8.00E-21	#N/A
TP4963	contig94836	1.00E-11	#N/A

TP49905	contig112980	1.00E-23	#N/A
TP51458	contig77801	1.00E-11	#N/A
TP52268	contig93400	2.00E-14	#N/A
TP525	contig51439	1.00E-24	#N/A
TP53432	contig96054	1.00E-23	#N/A
TP57771	contig92759	2.00E-15	#N/A
TP58098	contig117672	1.00E-17	#N/A
TP60143	contig9066	1.00E-25	#N/A
TP6226	contig3808	7.00E-15	#N/A
TP6426	contig104151	1.00E-24	#N/A
TP64873	contig49498	1.00E-24	#N/A
TP65331	contig20434	1.00E-24	#N/A
TP65409	contig93400	9.00E-14	#N/A
TP68260	contig70466	1.00E-24	#N/A
TP68336	contig82374	1.00E-23	#N/A
TP68535	contig75017	1.00E-24	#N/A
TP6894	contig95557	3.00E-26	#N/A
TP693	contig64686	3.00E-26	#N/A
TP6958	contig78566	4.00E-18	#N/A
TP6960	contig78566	7.00E-15	#N/A
TP6966	contig78566	8.00E-21	#N/A
TP70025	contig109691	3.00E-26	#N/A
TP70414	contig93499	1.00E-24	#N/A
TP70830	contig10525	5.00E-23	#N/A
TP71003	contig89802	4.00E-12	#N/A
TP7142	contig78566	4.00E-11	#N/A
TP71911	contig101273	1.00E-11	#N/A
TP72218	contig66468	3.00E-26	#N/A
TP72422	contig10431	3.00E-26	#N/A
TP72482	contig75093	1.00E-25	#N/A
TP72485	contig80582	1.00E-24	#N/A
TP7312	contig59070	3.00E-26	#N/A
TP74159	contig114439	1.00E-24	#N/A
TP74923	contig87052	3.00E-26	#N/A
TP74976	contig25342	1.00E-24	#N/A
TP76049	contig44785	1.00E-24	#N/A
TP76403	contig111595	3.00E-26	#N/A
TP77188	contig73895	1.00E-24	#N/A
TP77537	contig96921	2.00E-14	#N/A
TP77813	contig112024	3.00E-13	#N/A
TP78029	contig114600	1.00E-24	#N/A

TP78188	contig89711	6.00E-22	#N/A
TP78700	contig99058	3.00E-26	#N/A
TP79076	contig69394	3.00E-26	#N/A
TP79172	contig97032	6.00E-22	#N/A
TP79501	contig79732	1.00E-17	#N/A
TP79558	contig8467	1.00E-23	#N/A
TP7985	contig56245	4.00E-12	#N/A
TP80052	contig2106	1.00E-17	#N/A
TP80197	contig91242	1.00E-24	#N/A
TP80247	contig67036	1.00E-24	#N/A
TP80427	contig87345	9.00E-20	#N/A
TP80450	contig120389	1.00E-24	#N/A
TP825	contig51476	6.00E-16	#N/A
TP82703	contig21651	1.00E-23	#N/A
TP8500	contig16641	3.00E-19	#N/A
TP86444	contig94973	3.00E-26	#N/A
TP86774	contig95525	3.00E-26	#N/A
TP86904	contig118182	2.00E-15	#N/A
TP88062	contig79320	2.00E-14	#N/A
TP88208	contig47109	3.00E-26	#N/A
TP91020	contig20399	2.00E-14	#N/A
TP9131	contig70466	3.00E-26	#N/A
TP92280	contig94097	4.00E-12	#N/A
TP92640	contig80581	1.00E-24	#N/A
TP93462	contig107147	2.00E-14	#N/A
TP93773	contig69394	1.00E-24	#N/A
TP93785	contig109691	3.00E-26	#N/A
TP9400	contig14195	3.00E-26	#N/A
TP95130	contig39726	1.00E-24	#N/A
TP95729	contig42075	3.00E-26	#N/A
TP95744	contig9889	6.00E-16	#N/A
TP96889	contig83982	7.00E-15	#N/A
TP97301	contig98342	6.00E-16	#N/A
TP97667	contig81930	1.00E-24	#N/A
TP9890	contig80976	3.00E-26	#N/A
TP99082	contig73813	3.00E-26	#N/A
TP99359	contig71595	1.00E-24	#N/A
TP99837	contig78566	1.00E-12	#N/A

Online Supplemental 6 Click here to download Online Supplemental Appendix_S6.docx

Appendix S6: Delta K (Evanno et al. 2005), the second order increase in likelihood for each K tested in the individual assignment Structure analysis of individuals from 19 populations of *Protea repens*.



Online Supplemental 7

Click here to download Online Supplemental Appendix_S7_IBD_vs_IBE.docx

Appendix S7. Summary of MCMCglmm IBD vs IBE analysis

Figure 1: The IBD vs IBE results for one simulation of the data in which allele frequencies were randomized between populations. We show both the effect sizes for all simulated loci and the trends in effect size with 95% confidence bands (as quantified using GAM fits). This (together with four other simulations) represents the null expectation for a relationship between global FST and a each predictor variable.

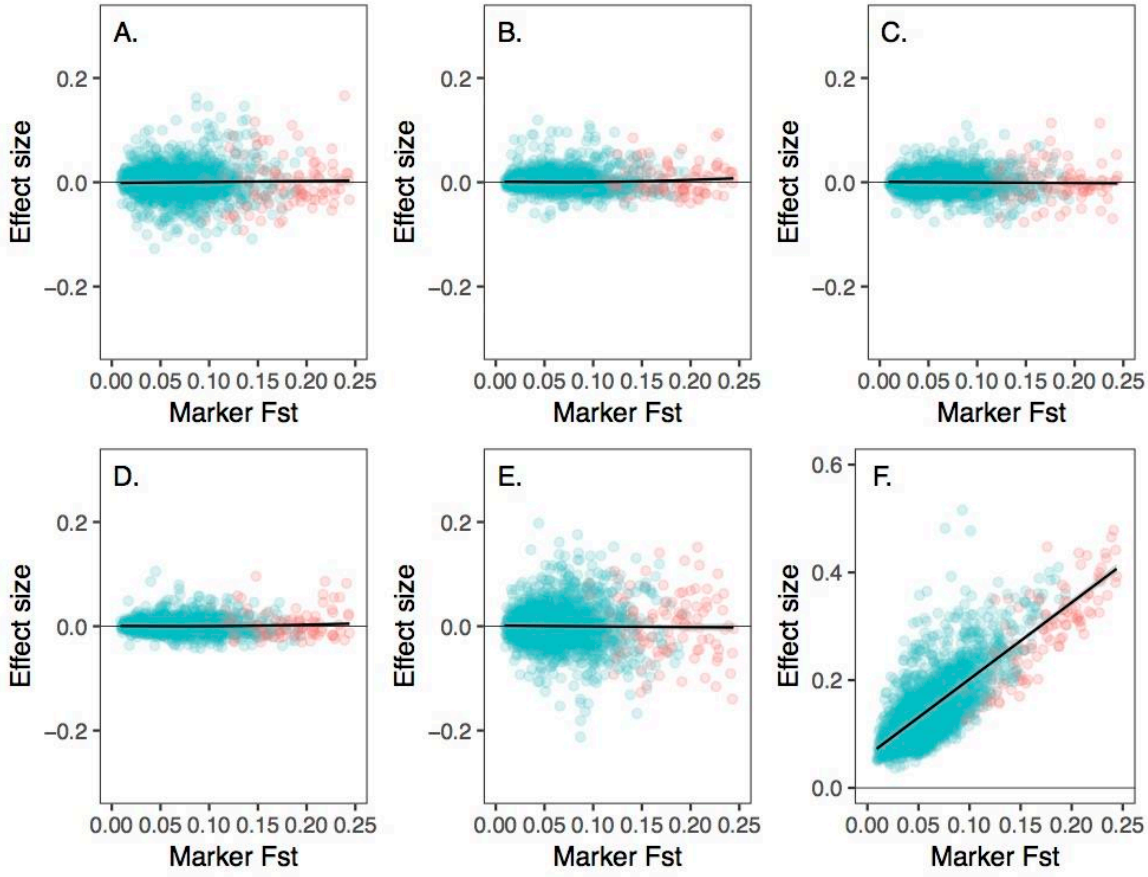


Figure 2: IBD and IBE contribute to genetic divergence between pairs of populations. The loci we examined have a wide range of genome-wide F_{ST}s. Loci with a high genome-wide F_{ST} may occur in genomic regions that experience divergent selection. To explore the factors driving population differentiation for each locus, we examined how strongly differences in allele frequencies between pairs of populations relate to their: **A.** differences in physical distance, three environmental axes (**B.** PC1-temp, **C.** PC2-winterrain, **D.** PC3-MAP), and **E.** phytogeographic zone. The strength of these relationships (effect size) varied substantially among different explanatory factors (panels) and across individual loci (arranged by their global F_{ST} within panels). **We present here both the effect sizes for all loci and the trends in effect size with 95% confidence bands (as quantified using GAM fits).** While there is considerable variation in effect size, physical distance (Panel A) was the most informative factor, showing positive effect sizes across a wide range of marker F_{ST} that deviate from the null model patterns. PC2-winter rain (panel C) and phytogeographic zone (panel E) also display significant positive effects, but only at loci with high F_{ST} values. PC1-temp (panel B) and PC3-MAP (panel E) show no deviation from the null model. These results suggest that all loci show the effects of IBD, independent of their F_{ST}, whereas only the more variable outlier loci (high F_{ST}) exhibit IBE and population differentiation that we attribute to historical forces.

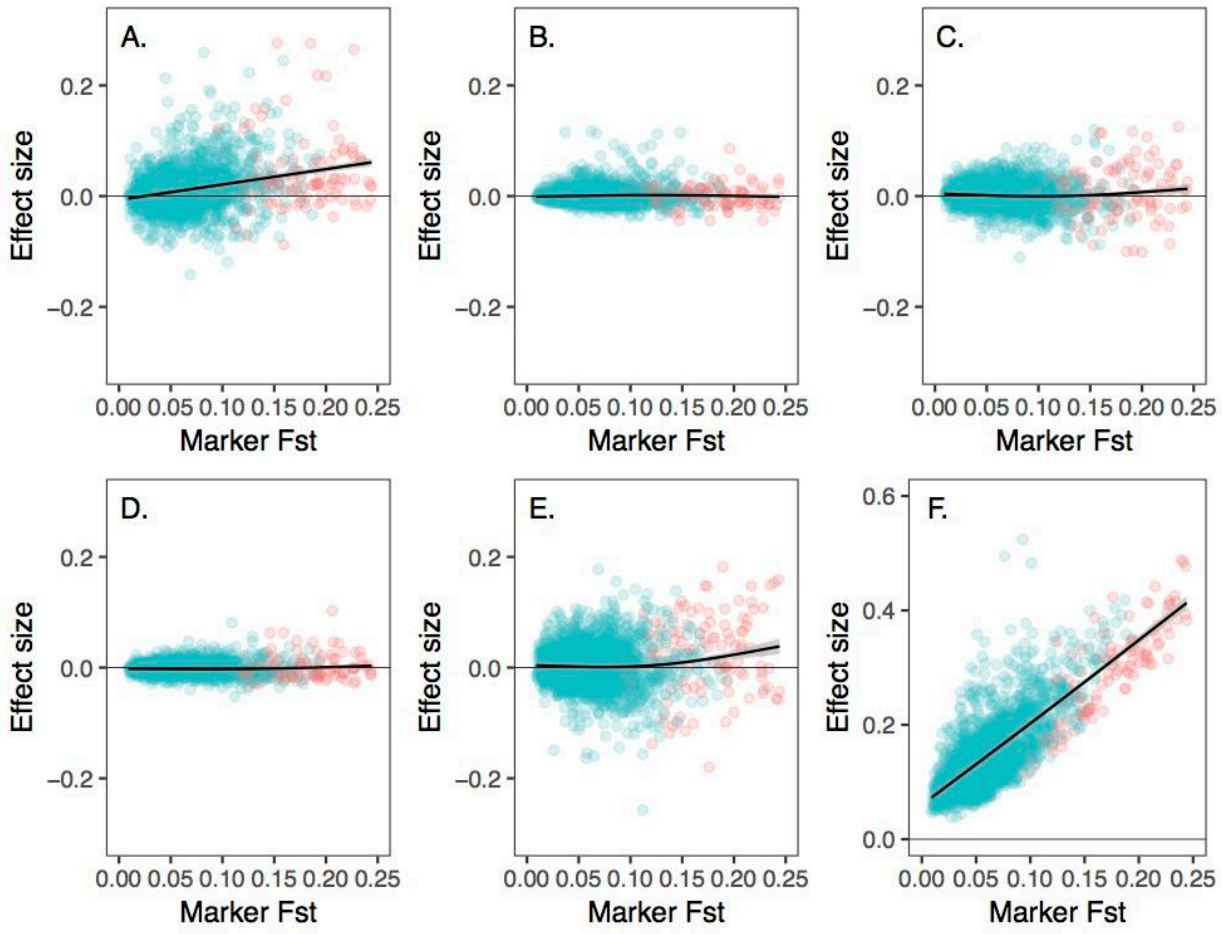


Table 1. Statistical output for GAM fits relating effect sizes of five coefficients to F_{st}, for original dataset and five randomizations of the original data. Sample size for all regressions was n = 2005, after removing low F_{st} outliers and a single observation with F_{st} > 0.29, which exerted excessive leverage on the regression.

Run	**Coefficient**		**Parametric intercept term**				**Smoother term (for F_{st})**		
		R^2	Estimate	Std. Error	t-value	p-value	Estimated df	F-value	p-value
Original	Intercept	0.635	0.1555	0.0010	154.10	<0.0001	1.3575	2178.7	<0.0001
Original	Phytogeography	0.012	0.0031	0.0009	3.3265	0.0009	1.9210	11.811	<0.0001
Original	Physical distance	0.084	0.0118	0.0008	14.009	<0.0001	1.0000	185.90	<0.0001
Original	PC1	0.002	0.0006	0.0003	1.9733	0.0486	1.7710	2.9856	0.0828
Original	PC2	0.005	0.0015	0.0005	2.8523	0.0044	1.9210	5.5966	0.0037
Original	PC3	0.003	-0.0023	0.0003	-8.9804	<0.0001	1.8824	3.6515	0.0207
Simulation 1	Intercept	0.616	0.1550	0.0010	148.97	<0.0001	1.0000	3213.0	<0.0001
Simulation 1	Phytogeography	0.000	0.0003	0.0009	0.3609	0.7182	1.0000	0.4597	0.4978
Simulation 1	Physical distance	0.000	-0.0001	0.0006	-0.1691	0.8657	1.0000	1.2329	0.2670
Simulation 1	PC1	0.002	0.0007	0.0004	1.8366	0.0664	1.7653	2.3135	0.0713
Simulation 1	PC2	0.000	0.0000	0.0004	0.1076	0.9143	1.0000	1.0740	0.3002
Simulation 1	PC3	0.002	0.0004	0.0003	1.2673	0.2052	1.7323	1.7714	0.1244
Simulation 2	Intercept	0.618	0.1551	0.0010	149.02	<0.0001	1.0000	3241.9	<0.0001
Simulation 2	Phytogeography	0.000	0.0007	0.0009	0.7313	0.4647	1.0000	1.1567	0.2823
Simulation 2	Physical distance	0.000	-0.0003	0.0006	-0.5123	0.6085	1.0000	0.4171	0.5185
Simulation 2	PC1	0.001	-0.0006	0.0004	-1.6373	0.1017	1.6328	1.2421	0.2020
Simulation 2	PC2	0.000	0.0000	0.0004	-0.0122	0.9903	1.0000	0.4616	0.4970
Simulation 2	PC3	0.006	0.0005	0.0003	1.7273	0.0843	1.7262	5.6603	0.0024
Simulation 3	Intercept	0.618	0.1550	0.0010	149.32	<0.0001	1.0000	3238.8	<0.0001
Simulation 3	Phytogeography	0.000	0.0001	0.0009	0.0717	0.9428	1.3374	0.1671	0.7764
Simulation 3	Physical distance	0.000	-0.0001	0.0007	-0.1713	0.8640	1.6082	0.7803	0.4655
Simulation 3	PC1	0.000	-0.0001	0.0004	-0.3262	0.7443	1.0000	0.4578	0.4987
Simulation 3	PC2	0.000	-0.0004	0.0004	-0.8802	0.3788	1.3778	0.2324	0.6310
Simulation 3	PC3	0.000	0.0003	0.0003	1.0997	0.2716	1.0000	0.2392	0.6248
Simulation 4	Intercept	0.612	0.1551	0.0011	147.27	<0.0001	1.0000	3166.6	<0.0001
Simulation 4	Phytogeography	0.000	0.0004	0.0010	0.4648	0.6421	1.0000	0.8371	0.3603
Simulation 4	Physical distance	0.000	-0.0001	0.0007	-0.1610	0.8721	1.0000	0.0922	0.7615
Simulation 4	PC1	0.000	-0.0001	0.0004	-0.3761	0.7069	1.0000	1.3474	0.2459
Simulation 4	PC2	0.000	0.0000	0.0004	0.0178	0.9858	1.0000	0.9939	0.3189
Simulation 4	PC3	0.000	0.0006	0.0003	1.9381	0.0528	1.0000	0.0192	0.8898
Simulation 5	Intercept	0.616	0.1550	0.0010	149.31	<0.0001	1.0000	3221.8	<0.0001
Simulation 5	Phytogeography	0.001	0.0001	0.0009	0.0720	0.9426	1.2061	1.9280	0.2127
Simulation 5	Physical distance	0.000	0.0003	0.0007	0.4633	0.6432	1.5792	0.8361	0.4786
Simulation 5	PC1	0.000	0.0002	0.0004	0.6289	0.5295	1.0000	0.1304	0.7181
Simulation 5	PC2	0.000	-0.0006	0.0004	-1.4311	0.1526	1.5400	0.7546	0.5307
Simulation 5	PC3	0.005	0.0005	0.0003	1.7206	0.0855	1.9210	5.2312	0.0044

Online Supplemental 8 Click here to download Online Supplemental Appendix_S8.docx

Appendix S8: summary of MCMCglmm model output for outlier marker TP262045. Significant effects are bolded.

predictor	posterior mean of effect size	pMCMC
(Intercept)	**0.33**	**0.001**
physical distance	-0.002	0.945
PC1 - temp	-0.018	0.381
PC2 - winter rain	**0.058**	**0.027**
PC3 - MAP	0.009	0.654
phytogeozone	-0.055	0.369