RESEARCH ARTICLE

AMERICAN JOURNAL OF BOTANY

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

Rachel Prunier[2,7], Melis Akman[3], Colin T. Kremer[4], Nicola Aitken[5], Aaron Chuah[5], Justin Borevitz[5], and Kent E. Holsinger[6]

PREMISE OF THE STUDY: The Cape Floristic Region (CFR) of South Africa is renowned for its botanical diversity, but the evolutionary origins of 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 19 populations across the range of *P. repens* and used genotyping by sequencing to identify 2066 polymorphic loci in 663 individuals. We used a Bayesian F_{ST} outlier analysis to identify single-nucleotide polymorphisms (SNPs) marking genomic regions that may be under selection; we used those SNPs to identify potential drivers of selection and excluded them from analyses of gene flow and genetic structure.

RESULTS: A pattern of isolation by distance suggested limited gene flow between nearby populations. The populations of *P. repens* fell naturally into two or three groupings, which corresponded to an east-west split. Differences in rainfall seasonality contributed to diversification in highly divergent loci, as do barriers to gene flow that have been identified in other species.

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.

KEY WORDS Cape Floristic Region; F_{ST} outlier; GBS; isolation by distance; isolation by environment; local adaptation; phylogeography; Proteaceae; *Protea repens*

The Cape Floristic Region (CFR) of southwestern South Africa is renowned for its spectacular botanical diversity (>9000 plant species; Myers et al., 2000; Goldblatt and Manning, 2002), but the evolutionary origins of this diversity remain controversial (Linder, 2003). Because of strong environmental gradients and extreme topographic diversity in the region, both adaptive and nonadaptive processes likely contribute to diversification among populations and species (Adamson, 1958; Linder, 1985; Johnson, 1996; van der Niet and Johnson, 2009; Ellis et al., 2014). Despite a wealth of studies investigating evolutionary processes above the species level (e.g., Goldblatt et al., 2002; Sauquet et al., 2009; Verboom et al., 2009, 2014; Valente et al., 2010; Pirie et al., 2016), very little is known about the evolutionary processes driving differentiation within plant species in the CFR (but see Prunier and Holsinger, 2010; Rymer et al., 2010; Lexer et al., 2014).

The CFR is an important location in which to study the balance of adaptive and nonadaptive diversification processes because, unlike in other hyperdiverse regions, species diversity in the CFR is primarily the result of high differentiation among sites (beta diversity) rather than high diversity within sites (alpha diversity; Latimer et al., 2005). High beta diversity indicates that there is substantial species turnover from location to location, often along environmental gradients (Whittaker, 1960), and it implies that many species have narrow ranges. Indeed, most CFR species are limited either geographically

[1] Manuscript received 17 June 2016; revision accepted 3 April 2017.

[2] Department of Biological and Environmental Sciences, Western Connecticut State University, 190 White Street, Danbury, Connecticut 06810, USA;
[3] Department of Plant Sciences, University of California Davis, 387 N. Quad Avenue, Davis, California 95616, USA;
[4] Department of Ecology and Evolutionary Biology, Yale University, P.O. Box 208106, New Haven, Connecticut 06520-8106, USA;
[5] Research School of Biology, Australian National University, Canberra, ACT 2601, Australia; and
[6] Department of Ecology and Evolutionary Biology, University of Connecticut, 75 N. Eagleville Road, U-3043, Storrs, Connecticut 06268, USA
[7] Author for correspondence (e-mail: prunierr@wcsu.edu)

doi:10.3732/ajb.1600232

(e.g., to one section of a mountain range) or environmentally (e.g., to one portion of an environmental gradient). 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., *P. obtusifolia*; Cowling et al., 1988). The environmental contrasts that accompany these patterns of species turnover abound in the CFR. For example, rainfall varies both in amount (60–3300 mm/yr) 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 (maximum elevation 2230 m) 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 these strong climatic, topographic, and edaphic gradients (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 a diverse group of geophytes in *Pelargonium* section *Hoarea* (Geraneaceae), ecologically similar species 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 geographic isolation or evolved as a result of it is unclear.

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

Vicariance may also play a significant role in diversification, due to drift that occurs when populations of a widespread species become isolated in pockets of suitable habitat (Kozak and Wiens, 2006). Vicariance has long been hypothesized as an important driver of diversification in the CFR (Adamson, 1958), with examples recently described in *Tetraria* (Cyperaceae; Britton et al., 2014; Verboom et al., 2015) and *Protea* section *Exsertae* (Prunier and Holsinger, 2010). High topographic diversity is associated with vicariant diversification in other regions of the world (e.g., Knowles, 2001), and Verboom et al. (2015) proposed that in the CFR, vicariance plays a larger role in speciation at high elevations because of both isolation and historical climate fluctuations.

We investigated patterns of population differentiation in *Protea repens* (Proteaceae L.), a widespread CFR species, using anonymous single-nucleotide polymorphisms (SNPs) identified through genotyping by sequencing (GBS). We identified SNPs marking genomic regions that may be subject to diversifying selection, and we used the remaining SNPs, presumed to be neutral, to understand phylogeographic patterns. We also investigated how the SNPs presumed to be neutral and those that may be responding to differential natural selection were affected by both isolation by distance (driven by vicariance) and isolation by environment (driven by divergent natural selection). Specifically, we addressed four questions. (1) Can we identify genomic regions that may be targets of selection? (2) What genetic groupings are present within *P. repens*, and are they associated with geographic boundaries that define distributions in other species? (3) To what extent is genetic differentiation associated with the physical distance and environmental differences between populations? (4) Do these patterns differ between genomic regions that are more or less likely to be targets of selection?

MATERIALS AND METHODS

Study system—*Protea repens* is a widespread species native to the Cape Floristic Region in southwestern South Africa. This large, sclerophyllous shrub is found across the CFR and into the thicket biome of the Eastern Cape. It is found from sea level to nearly 2000 m (mean elevation = 752 m; Appendix S1, see Supplemental Data with this article). It is the most abundant member of the genus *Protea* (Rebelo, 2001) and is common throughout its range, either as scattered plants or in dense stands. Schnitzler et al. (2011) estimated that *P. repens* split from its sister group, which contains 48 species, about 12 mya.

Population sampling—We used seeds that were collected from 5–33 maternal lines from each of 19 *P. repens* populations broadly distributed across the species' range (Table 1 and Fig. 1). Collaborators germinated the seeds and planted the seedlings into a common garden at Kirstenbosch Botanical Gardens in Cape Town, South Africa (see Carlson et al., 2015; Akman et al., 2016). We then collected leaf samples from 717 individuals in the garden and stored them in a CTAB and sodium chloride buffer for preservation. Vouchers for each population were deposited in the CONN herbarium (see Appendix 1 for voucher information).

Genotyping by sequencing—We extracted genomic DNA using a modified CTAB extraction (Doyle and Doyle, 1987), first washing the leaf samples to remove the CTAB and sodium chloride buffer, then grinding with a mortar and pestle. We prepared GBS libraries following the protocol described in Nicotra et al. (2016), using the restriction enzyme PstI (modified from Elshire et al., 2011; Morris et al., 2011). We gave each individual a forward and a reverse barcode 4–8 bp long (one at the 5′ end and one at the 3′ end), such that each individual had a unique barcode combination and could be multiplexed. We used three lanes of 100 bp, paired-end Illumina HiSEq. 2000 to sequence our samples. The 717 individuals were first split across two lanes. However, there was poor coverage for a majority of the individuals, so we reran 672 individuals on another lane to improve coverage. Our subsequent analyses required that each read carry a complete barcode, but the two sequences of a paired-end read each contained only half of an individual's barcode. Therefore, in silico, we recreated each full barcode from paired-end reads by combining the forward and reverse barcodes. We then substituted the full barcode for the forward and reverse barcodes. Each half of each paired-end read was then treated as a single-end read and analyzed using the Tassel UNEAK pipeline version 3.0 (Lu et al., 2013). Specifically, we used reads with a complete barcode, a PstI cut site, and no Ns in the 64 bp following the cut site (Lu et al., 2013).

TABLE 1. Spatial coordinates, sample sizes, and estimates of genetic diversity in non-outlier and high-F_{ST}-outlier loci for 19 populations of *Protea repens* studied in the CFR of South Africa.

					Non-outlier loci		High-outlier loci	
Population	Longitude	Latitude	Individuals sampled	Maternal lines sampled	Percent polymorphic	Nucleotide diversity	Percent polymorphic	Nucleotide diversity
1 VanRhynsdorp (VAN)	19.0225	−31.37083	35	21	0.923	0.181 ± 0.087	0.853	0.108 ± 0.055
2 Cederberg (CDB)	19.10676	−32.4064	22	14	0.853	0.154 ± 0.074	0.780	0.087 ± 0.046
3 Banghoek (BAN)	18.64454	−32.72662	8	5	0.562	0.116 ± 0.058	0.500	0.089 ± 0.049
4 Riverlands (RIV)	18.55111	−33.51604	36	22	0.901	0.162 ± 0.077	0.843	0.104 ± 0.054
5 Ceres (CER)	19.276	−33.36542	41	24	0.923	0.166 ± 0.08	0.861	0.095 ± 0.049
6 Kleinmond (KLM)	19.03363	−34.33109	39	23	0.907	0.164 ± 0.079	0.692	0.101 ± 0.053
7 Montagu (MGU)	20.10137	−33.7839	12	8	0.833	0.168 ± 0.083	0.724	0.094 ± 0.05
8 Riviersonderend (RIV)	19.99314	−34.10228	42	25	0.941	0.163 ± 0.078	0.899	0.096 ± 0.05
9 Bredasdorp (BRD)	20.0386	−34.54593	46	23	0.929	0.164 ± 0.078	0.899	0.089 ± 0.046
10 DeHoop (POT)	20.60545	−34.41555	28	15	0.896	0.166 ± 0.08	0.778	0.092 ± 0.048
11 Anysberg (ANY)	20.70409	−33.49369	52	28	0.961	0.167 ± 0.08	0.945	0.096 ± 0.049
12 Klein Swartberg (KSW)	21.28479	−33.45617	47	25	0.931	0.149 ± 0.071	0.908	0.083 ± 0.043
13 Garcia's Pass (GAR)	21.21982	−33.96789	73	33	0.948	0.171 ± 0.081	0.881	0.099 ± 0.051
14 Swartberg (SWA)	22.04591	−33.34953	58	29	0.949	0.154 ± 0.074	0.917	0.087 ± 0.045
15 Uniondale (UNI)	23.05253	−33.72581	30	9	0.903	0.173 ± 0.083	0.844	0.098 ± 0.051
16 Baviaanskloof (BAV)	23.63356	−33.49336	36	15	0.914	0.161 ± 0.077	0.907	0.096 ± 0.049
17 Kareedouw (KAR)	24.07328	−33.87468	24	18	0.853	0.145 ± 0.07	0.717	0.073 ± 0.039
18 Loerie Dam (LOE)	25.0423	−33.85452	22	12	0.874	0.164 ± 0.079	0.759	0.106 ± 0.055
19 Alicedale (ALC)	26.09839	−33.24081	12	6	0.644	0.135 ± 0.066	0.590	0.071 ± 0.039

SNP calling—We identified SNPs using the Tassel UNEAK pipeline. Reads with exactly 1 bp mismatched were considered candidate SNPs. We used the network filter to discard complicated networks of reads that each varied by only 1 bp mismatch, thus identifying reciprocal read pairs using default settings. Each read contains either zero or one SNP. The UNEAK pipeline identified 112,064 candidate SNPs, the vast majority of which were represented in only the few individuals that had the most reads (Appendix S2a). We filtered the candidate SNPs in three ways to ensure that they would be informative for population genetic analyses. First, we removed any SNP calls that had less than 5× coverage (per individual per called locus; Appendix S2a). Thus, assuming no bias in SNP amplification within an individual, there is less than one chance in 32 (~3%) that an individual scored as homozygous is actually a heterozygote for which only one chromosome was sequenced. This lowered the total number of SNP calls from ~3.5 million to ~1.25 million calls out of 74.3 million possible calls (if each individual was called at each locus). After removing low-coverage loci, we removed individuals that had <500 total reads (out of >57 million reads across all individuals), leaving 663 individuals in the data set. This cutoff is low, leaving some individuals that were only called for a few loci and had few reads, but most of the individuals (636) had >10,000 reads and were called at many loci (Appendix S2b). We removed the remaining SNPs that were not called in ≥20% of the remaining individuals from the analysis, resulting in a final set of 2066 SNPs in our analysis. Of the 2066 SNPs, more than half were called for ≥400 individuals, and of the 663 individuals, the vast majority were called for >1000 SNPs (Appendices S2c, S2d, S2e). There may be small amounts of gametic disequilibrium among some of these loci, but our statistical analyses depend either on methods that rely on genotype counts (F_{ST} outliers and population phylogeny) that are largely unaffected by gametic disequilibrium (Guo et al., 2009) or on methods that implicitly account for gametic disequilibrium (individual assignment and identity by descent).

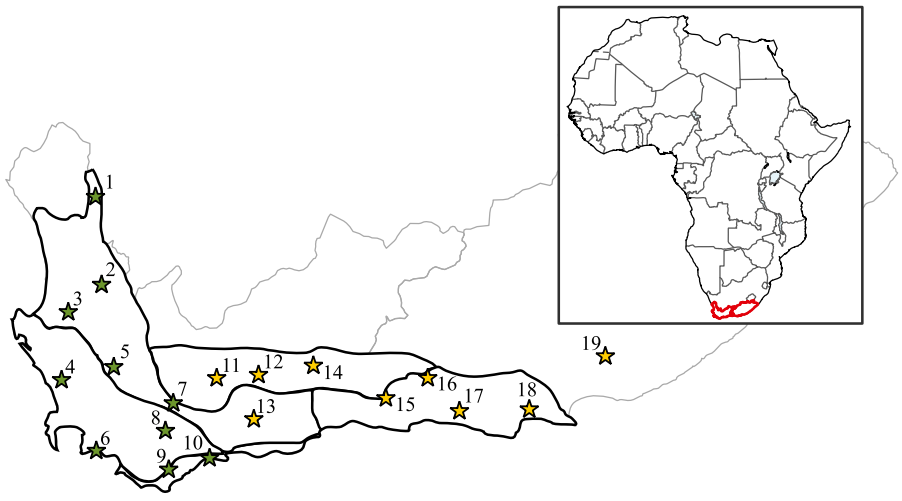

FIGURE 1 Sampling locations of *Protea repens* (indicated by stars) and boundaries of the phytogeographic regions (black outlines) of the Cape Floristic Region (drawn from Goldblatt and Manning, 2002). Inset is a map of Africa, with the enlarged South African provinces demarcated by the red outline. Stars are colored according to membership in the genetic groups identified in the MDS and Structure analyses; eastern populations are yellow and western populations are green.

Identification of outliers—Locus-specific estimates of F_{ST} can be used to identify regions of the genome that may be subject to selection (Beaumont and Balding, 2004). We used a Bayesian genome scan to identify F_{ST} outlier loci. Although genome scans in humans have identified genes apparently subject to selection even without an association with known phenotypes (e.g., Akey, 2009), widely used methods (BayeScan, FDIST) assume an island

model of migration and are prone to false positives (Excoffier et al., 2009; Lotterhos and Whitlock, 2014). To avoid these problems, we used the Bayesian method described by Guo et al. (2009).

This method is a conservative approach to detecting outliers, both because it uses the observed genomic distribution of F_{ST} as a basis of comparison and thus avoids specific demographic assumptions and because missing data "shrink" point estimates of locus-specific F_{ST} to the genome-wide mean F_{ST} and widen the posterior distributions, resulting in a loss of power to detect outliers. We also used a very stringent criterion to identify outliers ($P < 0.0001$). Even with a much less stringent criterion ($P < 0.05$), this method has a low rate of false positives (Guo et al., 2009). As an empirical check on the frequency of false positives, we produced six data sets in which individuals were randomly shuffled among populations. We found no F_{ST} outliers at any of the 2066 loci included in the analysis, which indicates that the upper 95% confidence limit on the false positive rate for this analysis is less than 4.2×10^{-6}. The analysis was implemented in JAGS version 3.4.0 (http://mcmc-jags.sourceforge.net/). Code is available at Holsinger (2016).

SNPs identified as high-F_{ST} outliers have allele frequencies that differ more than expected between populations and could reflect diversifying selection. SNPs identified as low-F_{ST} outliers have allele frequencies that differ less than expected between populations and could reflect stabilizing selection across populations. However, low outliers might also be artifacts arising from treating fixed heterozygosity in paralogous loci as allelic heterozygosity in orthologous loci. Because of this uncertainty, we did not investigate these low outliers, removing them from the remaining analyses. We treat loci not identified as outliers as presumably neutral.

Identifying SNPs in known genes—To determine whether the SNPs (both the non-outliers and the high-F_{ST} outliers) were in known genes, we used BLAST to map the 64 bp sequences to the draft *P. repens* leaf transcriptome (Akman et al., 2016). This draft transcriptome was constructed with sequence data from individuals representing the populations used in the present study and was annotated using sequence similarity to *Arabidopsis thaliana* EST libraries. We identified the SNPs that mapped to the transcriptome with an E-value $<10^{-10}$.

Population structure—*A. Distribution of genetic diversity*—We used Arlequin version 3.5.1.2 (Excoffier and Lischer, 2010) to estimate nucleotide diversity and the percentage of loci that are polymorphic for both the non-outlier and the high-F_{ST}-outlier loci in each population. We also used analysis of molecular variance (AMOVA) implemented in Arlequin to partition genetic diversity within and among populations for both the non-outlier and the high-F_{ST}-outlier loci.

B. Multidimensional scaling—We used multidimensional scaling (MDS) of pairwise identity by descent to visualize the relationships among individual genotypes. We approximated the probability of identity by descent between individuals i and j in our sample as

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

where p_{ik} is the frequency of the major allele at locus k in individual i (1 for an individual homozygous for the major allele at this locus, 1/2 for an individual heterozygous at this locus, 0 for an individual homozygous for the minor allele at this locus; compare Patterson et al., 2006). To perform this analysis, we used only the non-outlier SNPs. We used cmdscale() in R version 3.0 (R Foundation for Statistical Computing, Vienna, Austria) to perform the MDS.

C. Individual assignment—As a complement to the MDS analysis, we used an individual-based clustering approach implemented in Structure version 2.3.4 (Pritchard et al., 2000; Falush et al., 2003) to determine how individuals are grouped into larger genetic clusters. To perform this analysis, we used only the non-outlier SNPs. We performed 10 runs of 200,000 iterations, with 100,000 iterations discarded as burn-in at K values ranging from 1 to 19. We used Structure Harvester version web 0.6.94 (Earl and vonHoldt, 2012) to generate CLUMPP input files and to calculate delta K (Evanno et al., 2005), a statistic often used to identify the number of clusters most useful for interpretation. We used CLUMPP version 1.1.2 (Jakobsson and Rosenberg, 2007) to combine the results of our 10 runs at each K and visualized the results using Distruct version 1.1 (Rosenberg, 2003).

D. Population phylogeny—We used TreeMix version 1.12 (Pickrell and Pritchard, 2012; https://bitbucket.org/nygcresearch/treemix/wiki/Home) to estimate the sequence of population splits. TreeMix is widely used to estimate population relationships in humans in the presence of admixture (e.g., Meyer et al., 2012; Patterson et al., 2012; Lazaridis et al., 2014; Raghavan et al., 2014). To provide estimates of the support for branches, we produced 5000 bootstrap samples using the bootstrap option. We report the majority-rule consensus tree and placed the root of the phylogeny between the main groups identified by both Structure and MDS. We used only the non-outlier SNPs to perform this analysis.

Drivers of genetic differentiation—To determine how strongly genetic differentiation between populations is driven by physical distance, environmental differences, and historical barriers to gene flow, we examined differentiation at each of the 2066 loci using Bayesian linear mixed models implemented in MCMCglmm version 2.23 (Hadfield, 2010). This locus-by-locus approach allows us to explore how the effect size of each covariate changes across loci and, ultimately, to determine whether outlier and non-outlier loci respond differently to these environmental gradients. This analysis and those that follow were implemented in R version 3.3.1. Data and R code are available at http://github.com/ctkremer/P_repens.

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

(Fig. 2B) are extremely short (compared to the terminal branches), the error introduced by this approximation is small.

To test for the presence and strength of isolation by distance and isolation by environment, we calculated the distance between populations in physical and environmental space. For isolation by distance, we determined the Euclidean physical distances between populations, calculated using the earth.dist() command in the fossil package version 0.3.7 (Vavrek, 2011). For isolation by environment, we calculated environmental differences between populations based on seven variables that are associated with physiologically significant trait variation in this and other *Protea* species (Carlson et al., 2011; Prunier et al., 2012; Carlson et al., 2015): rainfall concentration (average intra-annual variation in precipitation), mean annual precipitation (MAP), average January maximum temperature, mean annual temperature, altitude, average July minimum temperature, and the amount of rainfall that falls in the summer months. For all known *P. repens* populations (Protea Atlas Project; Rebelo, 2006), including the 19 featured in the present study, we extracted all of the variables except for rainfall concentration from the 2007 "Atlas of Climatology and Agrohydrology" (Schultze, 2007). Due to boundary artifacts in the rainfall concentration variable in the 2007 Shultze data set, we extracted rainfall concentration from the 1997 Schultze data set (Schultze, 1997).

Next, to formulate orthogonal environmental axes, we performed a principal component analysis (PCA) using the princomp() function in the base R package on all seven variables. The first principal component (PC) explained 47% of the variation and was mainly driven by differences in temperature (hereafter "PC1-temp"); PC2 explained 23% of the variation and was driven by the proportion of rainfall that falls in the winter (negative loading of summer rainfall proportion; hereafter "PC2-winter rain"); and PC3 explained 20% of the variation and was driven by total rainfall (hereafter "PC3-MAP") (Appendix S3). We extracted values for the first three PC axes for the 19 populations studied here and calculated the distances between the populations on the three PC axes. Both the calculated physical distances and distances along these three PC axes were standardized to have a mean of 0 and a standard deviation of 1, allowing their relative effect sizes to be compared.

Finally, to determine the effect of historical barriers to gene flow on current patterns of genetic similarity, we scored the pairs of populations for coexistence in the phytogeographic zones defined by Goldblatt and Manning (2002). Phytogeographic zones are regions that share an endemic flora and, as such, might reflect barriers that are not explained by either physical distance or measured environmental differences. Population pairs were scored as either 1 (different phytogeographic zone) or 0 (same zone).

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

Each model fit provides coefficients that estimate how strongly differences in a locus's allele frequencies between populations are related to differences in five factors (physical distance, three environmental variables, and location in the same phytogeographic zone). We are interested in how these relationships differ across loci, and especially between the F_{ST} outlier loci and the non-outlier loci. We used a generalized additive model (GAM) to quantify trends in estimated coefficients as a function of locus-specific F_{ST} values (see identifying F_{ST} outliers above), because GAMs allowed us to avoid making explicit parametric assumptions about the shape of these trends. Rather than run separate GAMs for the outlier and non-outlier loci, we ran a single GAM for all loci. Inferences were based on the relationships with the F_{ST} of each locus, rather than their status as outliers as determined in the F_{ST} outlier analysis above. However, loci with high F_{ST} estimates are predominantly outliers (Appendix S2f).

To determine whether trends revealed by the GAM analysis provide meaningful insight into the mechanisms driving genetic differentiation or simply emerge by chance, we ran five null simulations.

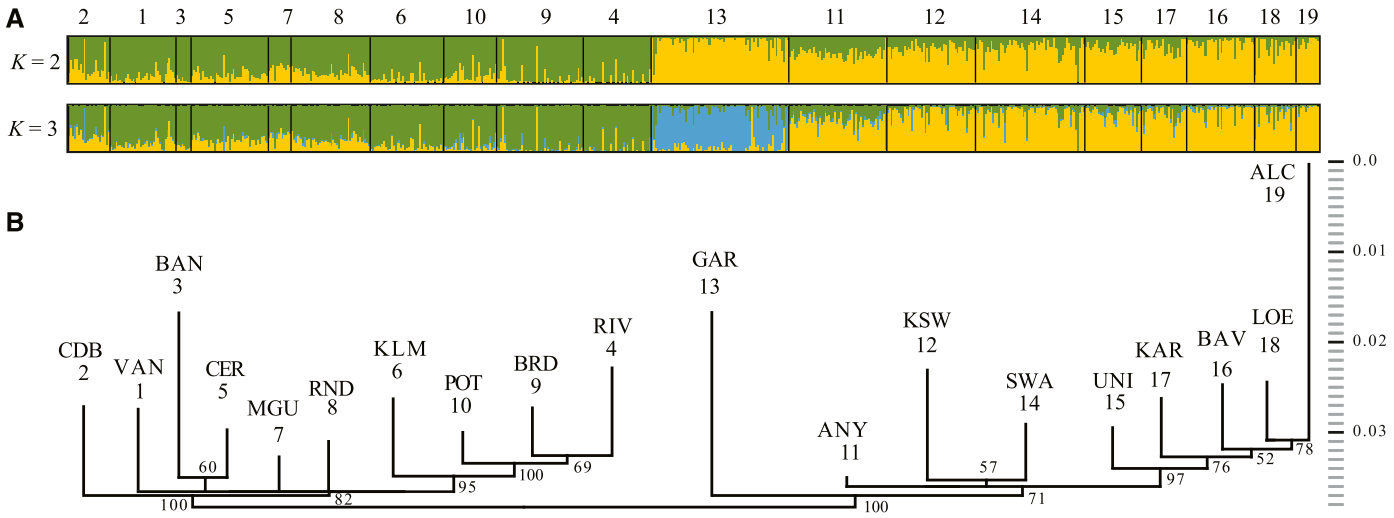

FIGURE 2 (A) Results of the Structure analyses for $K = 2$ and $K = 3$. Each *Protea repens* individual (grouped by population) is represented by a vertical bar. The proportion of the bar in each color corresponds to the average posterior likelihood that the individual is assigned to the cluster indicated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

In each simulation, we randomized the observed allele frequencies among populations within each locus to break any correlations relating the genetic distance between populations to their environmental and physical distance. We then repeated our MCMCglmm() analyses of all loci, and we fit GAM regressions to the resulting set of coefficients as a function of the locus's original F_{ST} values, as described in the preceding paragraph. Contrasting the GAM fits of the original and randomized data allowed us to determine when genetic differentiation among populations is more strongly predicted by covariates than expected by chance.

RESULTS

Outliers—Mean genome-wide F_{ST} was 0.065 (95% credible interval: 0.062–0.068). We identified 109 loci as high outliers (mean high-outlier F_{ST} was 0.24) and 60 loci as low outliers (Appendix S2f). High outliers have allele frequencies that differ between populations more than expected based on the distribution of locus-specific F_{ST} estimates and could reflect the effects of diversifying selection across populations. Low-F_{ST} outliers could be the result of either stabilizing selection or fixed heterozygosity in paralogous loci, so the low outliers were removed from the remaining analyses. Of the 109 high outliers, 26 map to the *P. repens* transcriptome with E-values $<1^{-10}$, and 14 of those also map to *A. thaliana* genes (Appendix S4). These outliers are not necessarily the gene variants or the genes that are under selection, but they mark regions of the genome that may be subject to diversifying selection. A slightly larger proportion of the non-outlier loci also had hits to the *P. repens* transcriptome (607 of 1897) and to the *A. thaliana* genome (367 of 1897; Appendix S5).

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

Population structure—The MDS and individual assignment (Structure) analyses revealed similar patterns. The MDS analysis shows that the populations are split into an eastern group and a western group (Fig. 3), with the GAR population somewhat isolated within the eastern group and RIV somewhat isolated from the western group. Similarly, the Structure analysis also supports two groups. Using Evanno et al.'s (2005) criterion (Appendix S6), $K = 2$ and $K = 3$ received similar support as the models that best represent the underlying genetic structure in the populations. With $K = 2$, we again see a clear delimitation between eastern and western populations (Fig. 2A). With $K = 3$, the GAR population is separated from the remaining eastern populations. While Structure is not well suited to situations in which there is isolation by distance (Pritchard et al., 2000), the similarity in the results between the Structure analysis, the MDS, and population phylogeny described below suggests that, in this instance, the results are likely to be reliable.

The population phylogeny estimated with TreeMix reveals a pattern consistent with these analyses. We found strong bootstrap support for eastern and western groups of populations (Fig. 2B). The eastern populations also show a pattern suggestive of stepping-stone colonization from west to east in which each population is sister to the remaining populations to the east. Population 13 (GAR) is sister to the remaining eastern populations and populations 11, 12, and 14 (ANY, KSW, and SWA) are sister to the remaining eastern populations. It then appears that *P. repens* spread east and north from population 15 (UNI). Patterns in the western populations are more complex. CDB, one of the northernmost populations, is sister to the remaining western populations, and coastal populations (KLM, POT, RIV, and BRD) form a clade, but there is no clear relationship between population history and geographic distribution in the western populations. Consistent with the AMOVA (Table 2), which showed that most of the genetic variation was within populations, the longest branches in the population phylogeny are the terminal branches. Internal branches, even the branch separating the two clades identified in the Structure and MDS analyses, were much shorter (Fig. 2B).

Drivers of genetic differentiation—Our null model simulations revealed two important findings about the expected relationships between our predictor variables and genetic differentiation between populations. First, randomizing allele frequencies among populations removes the relationships between locus-specific F_{ST} and the effect sizes of our covariates (Fig. 4; gray lines are flat on average). Second, the average amount of genetic differentiation among populations necessarily increases with locus-specific F_{ST}. In the absence of informative covariates, this increase appears in the intercept terms, which rise with locus-specific F_{ST} (Appendix S7a).

Against these null patterns, the analyses of the actual observations revealed significant effects of physical distance, PC2-winter rain, and phytogeographic zone, but not PC1-temp or PC3-MAP (Fig. 4; see also panel A of Appendix S7b and Appendix S7c). Physical distance had positive effects on genetic differentiation between populations across a wide range of locus-specific F_{ST} values (Fig. 4A). This suggests that isolation by distance affects most loci. There is also evidence of isolation by environment, driven by PC2-winter

TABLE 2. Analysis of molecular variance of 19 populations of *Protea repens* in South Africa estimated for non-outlier and high-F_{ST}-outlier loci (*** indicates F_{ST} estimate significantly different from zero).

Genome fraction	Source of variation	df	Variance components	Percentage of variation	F_{ST}
Non-outlier loci	Among	18	9.74121	6.29	0.06292***
	Within	1307	145.07112	93.71	
High-F_{ST}-outlier loci	Among	18	1.35714	23.87	0.23867***
	Within	1305	4.32908	76.13	

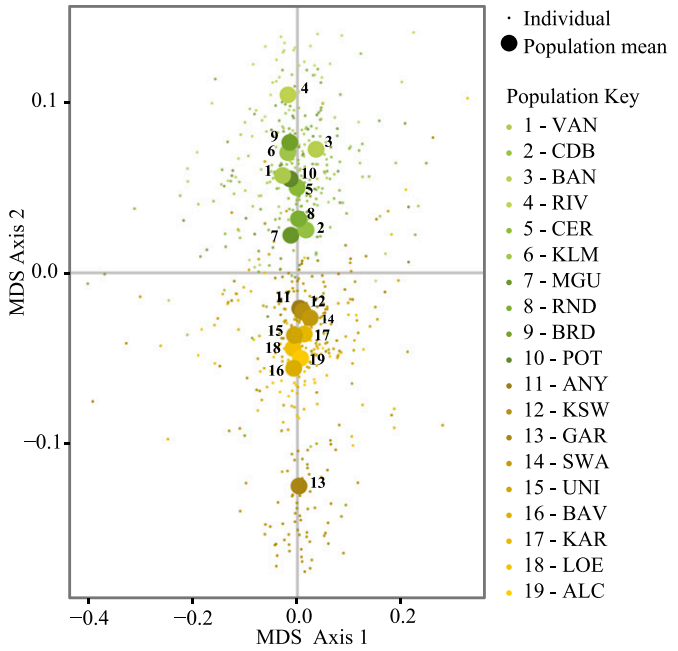

FIGURE 3 First two multidimensional scaling (MDS) axes drawn from pairwise identities by descent between *Protea repens* individuals. Individuals are indicated by small dots and population mean values by large dots. Populations are coded by color, based on the groupings identified in this and the Structure analysis; eastern populations are yellow and western populations are green. The shading indicates the degree to which each population is east or west, with centrally located populations darker than peripheral populations.

rain (Fig. 4C), although this effect is weaker than the effect of physical distance. Lastly, historical barriers to gene flow had large effects on differentiation, such that pairs of populations in different phytogeographic zones had more divergent allele frequencies (Fig. 4E).

Interestingly, the effects of isolation by distance, isolation by environment, and phytogeographic zone were not equally important at all levels of locus-specific F_{ST}. The effect of physical distance increased in proportion to the F_{ST} of the locus: as the proportion of genetic variation between populations increased, so did the amount of that variation that physical distance explained (Fig. 4A). Thus, physical distance has similar effects on variation in allele frequencies in both non-outlier and outlier loci (which have high F_{ST}). By contrast, PC2-winter rain and phytogeographic zone only diverged from the null models for loci with $F_{ST} > 0.15$ (see Fig. 4C, E). This suggests that isolation by environment only affects differentiation in the high-F_{ST} loci, many of which are outliers. Lastly, while we detected several significant relationships between F_{ST} and physical and environmental distance, much additional variation between loci remains (Appendices S7b and S7c). This suggests that, despite these overall trends, individual loci respond to these predictor variables idiosyncratically.

DISCUSSION

We detected the effects of both nonadaptive and adaptive evolutionary processes in the genome of *P. repens*, a widespread, mid-elevation CFR species. We found evidence of an east-west split in genetic groupings (Figs. 2 and 3) as well as of an expansion into the eastern extent of the range of *P. repens* (Fig. 2B). We also detected a strong pattern of isolation by distance, at loci with both low and high amounts of among-population differentiation (Fig. 4A). Isolation by environment, specifically differences in winter rainfall (Fig. 4C), also contributes to genetic differentiation, but only at loci with high amounts of differentiation. We also found that historical barriers to gene flow, as defined by previously recognized phytogeographic zones, help explain genetic differentiation in loci with high F_{ST} estimates (Fig. 4E).

The analyses based on the non-outlier loci, which presumably reflect primarily neutral variation, identified a split between eastern and western populations. This split corresponds to a switch in both the timing of rainfall, from predominantly winter rainfall in the west to a more even rainfall distribution in the east (Schultze, 2007), and to differences in flowering time in *P. repens* (Rebelo, 2001; Heelemann et al., 2008; J. E. Carlson, personal communication). Differences in flowering time could hinder gene flow between populations, reinforcing this east-west divide. That said, the very short branch lengths associated with the east-west split in the TreeMix analysis (Fig. 2B) suggest that little of the genetic differentiation is associated with the east-west difference. Instead, the long terminal branches suggest that most of the genetic differentiation is a result of the accumulation of differences that uniquely identify populations regardless of whether they are in the west or the east.

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

Our analyses of the drivers of genetic differentiation of both the non-outlier and outlier loci identified a strong pattern of isolation by distance. This pattern affects loci with both low and high F_{ST} and indicates that gene flow is geographically limited in *P. repens* (Wright, 1943). While gene flow is limited by distance, we do see some connectivity between nearby populations. This is likely caused by pollen movement by the cape sugarbird (*Promerops cafer*), which pollinates *P. repens* (Rebelo, 2001; Schmid et al., 2015). Individual sugarbirds can travel hundreds of kilometers in a year, migrating between food sources (Hockey et al., 2005). The isolation by distance we observed is in contrast to what was found in *Restio capensis* (Restionaceae; Lexer et al., 2014), the only other widespread CFR species for which there is population genomic data. *Restio capensis* does not exhibit a consistent pattern of isolation by distance. Instead, in *R. capensis*, environmental differences are the best predictor of genetic differentiation in both non-outlier and outlier loci, which Lexer et al. (2014) interpreted as an indication of incipient ecological speciation, perhaps in response to differences in water availability and soil type.

We too found evidence that among-population genetic differentiation is associated with environmental differences, but only in loci with high F_{ST} and only in response to differences in winter rainfall. This is consistent with other studies on *Protea* species; winter rainfall is associated with differences in functional traits in these same

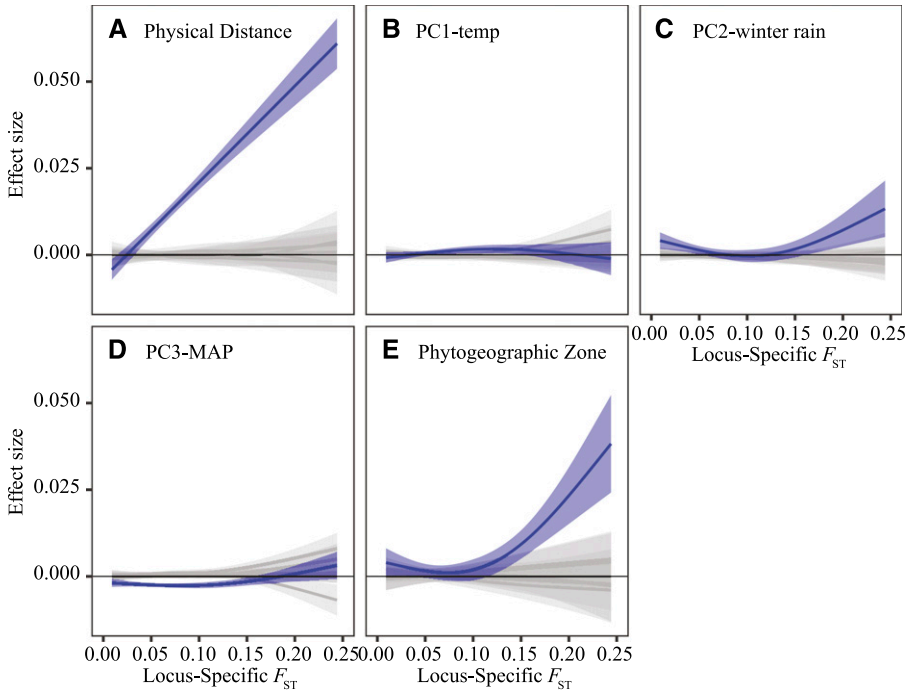

FIGURE 4 Both isolation by distance and isolation by environment contribute to genetic divergence between pairs of *Protea repens* populations. This figure shows how the ability of five covariates to explain genetic differentiation among populations changes with the F_{ST} of loci. In order, these covariates are (A) differences in physical distance; three environmental variables—(B) PC1-temp, (C) PC2-winter rain, and (D) PC3-MAP; and (E) phytogeographic zone. Solid lines indicate trends in effect sizes revealed by GAM fits (see text) and are surrounded by 95% confidence bands. Observed patterns (blue) contrast with null model simulations (gray), revealing significant effects of physical distance across locus-specific F_{ST} values (A), as well as effects of PC2-winter rain (C) and phytogeographic zone (E) at loci with high F_{ST}. By contrast, PC1-temp (B) and PC3-MAP (D) do not deviate from the null models.

populations of *P. repens* (Carlson et al., 2015) as well as in *Protea* section *Exsertae*, a group of six *Protea* species (Carlson et al., 2011; Prunier et al., 2012). Further, Carlson et al. (2011) detected differential selection on leaf traits in *Protea* section *Exsertae* in response to winter rainfall differences. Both our findings and those of Lexer et al. (2014) lend support to the hypothesis that adaptive processes contribute to differentiation in the CFR. However, the nature of this relationship differs between the two species.

These differences raise the question of why *R. capensis* appears to be in the process of speciating in response to environmental differences while *P. repens* is not. The answer might lie in fundamental biological differences between the two plants. Both species are broadly distributed in the CFR, but species in Restionaceae are generally sensitive to water availability and soil types (Araya et al., 2010), whereas species in the genus *Protea* are generally less sensitive to water availability because they are deeply rooted and are less reliant on surface water (Richards et al., 1995; Watt and Evans, 1999). Thus, the different patterns observed in these two analyses could arise if *R. capensis* is inherently more likely to differentiate along environmental gradients than *P. repens*. These two species also differ in pollination syndrome; *R. capensis* is wind pollinated (Lexer et al., 2014), whereas *P. repens* is pollinated by birds and insects (Schmid et al., 2015). These differences in pollination can have a significant effect on genetic isolation between populations (Levin and Kerster, 1974), which perhaps explains why we observe isolation by distance in *P. repens* but not *R. capensis*.

Historical barriers to gene flow also contributed to genetic differentiation in *P. repens*, but only in the loci with high amounts of among-population differentiation. We recognize these barriers as the boundaries between phytogeographic zones defined by breaks in the ranges of species from many groups of plants in the CFR (Goldblatt and Manning, 2002). It is curious that we identified this effect only for highly differentiated loci. If the phytogeographic zones were acting as barriers to dispersal, they should affect all loci in a similar manner, like isolation by distance. Instead, the pattern we observed suggests that the boundaries between phytogeographic zones reflect environmental differences, not included in our PCA axes, that might differ between phytogeographic zones, such as soil type (Linder, 2001; Goldblatt and Manning, 2002) or fire return interval (Litsios et al., 2014), or perhaps combinations of environmental factors.

This work could begin to identify the genes and alleles responsible for local adaptation in *P. repens*. Carlson et al. (2015) found that in the populations studied here, leaf, stomatal, and physiological traits vary with total rainfall and winter rainfall in particular (similar to our PC3-MAP and PC2-winter rain). Genomic regions surrounding our outlier loci may well contain some of the genes related to this trait variation. Although we need more complete SNP coverage before good candidate genes can be confidently identified, some of our outlier loci have intriguing relationships to rainfall patterns. For example, among-population differences in allele frequency at locus TP26245, which occurs in gene AT5G48930 (involved in auxin transport and growth in *A. thaliana*), are correlated with changes in PC2-winter rain (Appendix S8). This coincides with a decrease in growth rate in *P. repens* populations with lower winter rainfall (Merow et al., 2014).

CONCLUSIONS

We have documented that a combination of adaptive and nonadaptive processes is associated with diversification in a widespread CFR species. Our results provide evidence consistent with postglacial colonization of the eastern part of the region. They also suggest that while physical distance reduces gene flow between distant populations, there is gene flow between nearby populations. Interestingly, historical barriers that have contributed to diversification in other groups (phytogeographic zones) also contribute to divergence in *P. repens* populations, but only in loci showing the largest amounts of among-population differentiation. We also found evidence that some level of local adaptation is occurring in response to different environmental regimes, but there is no evidence of

ecological speciation occurring. This work adds to the currently limited number of studies of the microevolutionary processes driving diversification in the CFR. Given the differences between the patterns observed in *P. repens* and *R. capensis*, additional studies of plants from the CFR are required to discover whether the patterns shown by either species hold broadly in the region.

ACKNOWLEDGEMENTS

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

DATA ACCESSIBILITY

De-multiplexed reads for each individual can be found in the GenBank Short Read Archive BioProject PRJNA380116 (accession nos. SRS2113368–SRS2113584 and SRS2114681–SRS2115182). The output of the UNEAK pipeline (allele calls, read counts per locus per individual, and fasta files for each locus) can be found at Dryad (data available from the Dryad Digital Repository: http://doi.org/10.5061/dryad.g8t8k). Custom code for the outlier analysis can be found at doi: 10.5281/zenodo.54919. A graphical depiction of the data flow, all other custom code, and all other data files can be found at both Dryad (http://doi.org/10.5061/dryad.g8t8k) and Github (http://github.com/ctkremer/P_repens).

LITERATURE CITED

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

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

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

Araya, Y. N., J. Silvertown, D. J. Gowing, K. McConway, P. Linder, and G. Midgley. 2010. Variation in d13C among species and sexes in the family Restionaceae along a fine-scale hydrological gradient. *Austral Ecology* 35: 818–824.

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

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

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

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

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

Coop, G., D. Witonsky, A. Di Rienzo, and J. K. Pritchard. 2010. Using environmental correlations to identify loci underlying local adaptation. *Genetics* 185: 1411–1423.

Cowling, R. M., B. M. Campbell, P. Mustart, D. J. McDonald, M. L. Jarman, and E. J. Moll. 1988. Vegetation classification in a floristically complex area: The Agulhas Plain. *South African Journal of Botany* 54: 290–300.

Cowling, R. M., Ş. Procheş, and T. C. Partridge. 2009. Explaining the uniqueness of the Cape flora: Incorporating geomorphic evolution as a factor for explaining its diversification. *Molecular Phylogenetics and Evolution* 51: 64–74.

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

Earl, D. A., and B. M. vonHoldt. 2012. Structure Harvester: A website and program for visualizing Structure output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359–361.

Ellis, A. G., G. A. Verboom, T. van der Niet, S. D. Johnson, and H. P. Linder. 2014. Speciation and extinction in the Greater Cape Floristic Region. *In* N. Allsopp, J. F. Colville, G. A. Verboom, and R. M. Cowling [eds.], Fynbos: Ecology, evolution and conservation of a megadiverse region, 119–141. Oxford University Press, Oxford, UK.

Ellis, A. G., and A. E. Weis. 2006. Coexistence and differentiation of "flowering stones": The role of local adaptation to soil microenvironment. *Journal of Ecology* 94: 322–335.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6: e19379.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software Structure: A simulation study. *Molecular Ecology* 14: 2611–2620.

Excoffier, L., T. Hofer, and M. Foll. 2009. Detecting loci under selection in a hierarchically structured population. *Heredity* 103: 285–298.

Excoffier, L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567.

Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587.

Fielding, A., and H. Goldstein. 2006. Cross-classified and multiple membership structures in multilevel models: An introduction and review. University of Birmingham Research Reports 1–69.

Gibby, M., S. Hinnah, E. M. Marais, and F. Albers. 1996. Cytological variation and evolution within *Pelargonium* section *Hoarea* (Geraniaceae). *Plant Systematics and Evolution* 203: 111–142.

Goldblatt, P., and J. C. Manning. 2002. Plant diversity of the Cape region of southern Africa. *Annals of the Missouri Botanical Garden* 89: 281–302.

Goldblatt, P., V. Savolainen, O. Porteous, I. Sostaric, M. Powell, G. Reeves, J. C. Manning, et al. 2002. Radiation in the Cape flora and the phylogeny of peacock irises *Moraea* (Iridaceae) based on four plastid DNA regions. *Molecular Phylogenetics and Evolution* 25: 341–360.

Gregorius, H. R., and J. H. Roberds. 1986. Measurement of genetical differentiation among sub-populations. *Theoretical and Applied Genetics* 71: 826–834.

Guo, F., D. K. Dey, and K. E. Holsinger. 2009. A Bayesian hierarchical model for analysis of single-nucleotide polymorphisms diversity in multilocus, multipopulation samples. *Journal of the American Statistical Association* 104: 142–154.

Hadfield, J. 2010. MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software* 33: 1–22.

Heelemann, S., Ş. Procheş, A. G. Rebelo, B. W. van Wilgen, S. Porembski, and R. M. Cowling. 2008. Fire season effects on the recruitment of non-sprouting serotinous Proteaceae in the eastern (bimodal rainfall) fynbos biome, South Africa. *Austral Ecology* 33: 119–127.

Hockey, P., W. R. J. Dean, and P. G. Ryan. 2005. Roberts birds of southern Africa, 7th ed. Trustees of the South African Bird Book Fund, Cape Town.

Holsinger, K. E. 2016. Genotyping by sequencing in *Protea repens* [data set]. Zenodo. http://doi.org/10.5281/zenodo.54919.

Jakobsson, M., and N. A. Rosenberg. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics (Oxford, England)* 23: 1801–1806.

Johnson, S. D. 1996. Pollination, adaptation and speciation models in the Cape flora of South Africa. *Taxon* 45: 59–66.

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.

Knowles, L. L. 2001. Did the Pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshoppers. *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.

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.

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.

Levin, D. A., and H. W. Kerster. 1974. Gene flow in seed plants. *In* T. Dobzhansky, M. K. Hecht, and W. C. Steere [eds.], Evolutionary biology, vol. 7, 139–220. Springer, Boston, Massachusetts, USA.

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.

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 Restionaceae. *Systematic Biology* 50: 892–912.

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

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

Lotterhos, K. E., and M. C. Whitlock. 2014. Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology* 23: 2178–2192.

Lu, F., A. E. Lipka, J. Glaubitz, R. Elshire, J. H. Cherney, M. D. Casler, E. S. Buckler, and D. E. Costich. 2013. Switchgrass genomic diversity, ploidy, and evolution: Novel insights from a network-based SNP discovery protocol. *PLoS Genetics* 9: e1003215.

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

Meyer, M., M. Kircher, M.-T. Gansauge, H. Li, F. Racimo, S. Mallick, J. G. Schraiber, et al. 2012. A high-coverage genome sequence from an archaic denisovan individual. *Science* 338: 222–226.

Morris, G. P., P. P. Grabowski, and J. O. Borevitz. 2011. Genomic diversity in switchgrass (*Panicum virgatum*): From the continental scale to a dune landscape. *Molecular Ecology* 20: 4938–4952.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. daFonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403: 853–858.

Nicotra, A. B., C. Chong, J. G. Bragg, C. R. Ong, N. C. Aitken, A. Chuah, B. Lepschi, and J. O. Borevitz. 2016. Population and phylogenomic decomposition via genotyping-by-sequencing in Australian *Pelargonium. Molecular Ecology* 25: 2000–2014.

Nosil, P., D. J. Funk, and D. Ortiz-Barrientos. 2009. Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* 18: 375–402.

Patterson, N., P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, et al. 2012. Ancient admixture in human history. *Genetics* 192: 1065–1093.

Patterson, N., A. L. Price, and D. Reich. 2006. Population structure and eigenanalysis. *PLoS Genetics* 2: e190.

Pickrell, J. K., and J. K. Pritchard. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics* 8: e1002967.

Pirie, M. D., E. G. H. Oliver, A. Mugrabi de Kuppler, B. Gehrke, N. C. Le Maitre, M. Kandziora, D. U. Bellstedt, et al. 2016. The biodiversity hotspot as evolutionary hot-bed: Spectacular radiation of Erica in the Cape Floristic Region. *BMC Evolutionary Biology* 16: 190.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

Prunier, R., and K. E. Holsinger. 2010. Was it an explosion? Using population genetics to explore the dynamics of a recent radiation within *Protea* (Proteaceae L.). *Molecular Ecology* 19: 3968–3980.

Prunier, R., K. E. Holsinger, and J. E. Carlson. 2012. The effect of historical legacy on adaptation: Do closely related species respond to the environment in the same way? *Journal of Evolutionary Biology* 25: 1636–1649.

Raghavan, M., P. Skoglund, K. E. Graf, M. Metspalu, A. Albrechtsen, I. Moltke, S. Rasmussen, et al. 2014. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505: 87–91.

Rebelo, T. G. 2001. Sasol proteas: A field guide to the proteas of southern Africa, 2nd ed. Fernwood Press, Vlaeberg, South Africa.

Rebelo, T. G. 2006. Protea Atlas Project website. http://www.proteaatlas.org.za/.

Richards, M. B., W. D. Stock, and R. M. Cowling. 1995. Water relations of seedings and adults of two fynbos species in relation to their distribution patterns. *Functional Ecology* 9: 575–583.

Rieseberg, L. H., and J. M. Burke. 2001. Commentary: A genic view of species integration. *Journal of Evolutionary Biology* 14: 883–886.

Rosenberg, N. A. 2003. Distruct: A program for the graphical display of population structure. *Molecular Ecology Notes* 4: 137–138.

Rymer, P. D., J. C. Manning, P. Goldblatt, M. P. Powell, and V. Savolainen. 2010. Evidence of recent and continuous speciation in a biodiversity hotspot: A population genetic approach in southern African gladioli (Gladiolus; Iridaceae). *Molecular Ecology* 19: 4765–4782.

Sauquet, H., P. H. Weston, N. P. Barker, C. L. Anderson, D. J. Cantrill, and V. Savolainen. 2009. Using fossils and molecular data to reveal the origins of the Cape proteas (subfamily Proteoideae). *Molecular Phylogenetics and Evolution* 51: 31–43.

Schmid, B., H. Nottebrock, K. J. Esler, J. Pagel, A. Pauw, and K. Böhning-Gaese. 2015. Reward quality predicts effects of bird-pollinators on the reproduction of African *Protea* shrubs. *Perspectives in Plant Ecology, Evolution and Systematics* 17: 209–217.

Schnitzler, J., T. G. Barraclough, J. S. Boatwright, P. Goldblatt, J. C. Manning, M. P. Powell, T. Rebelo, and V. Savolainen. 2011. Causes of plant diversification in the Cape biodiversity hotspot of South Africa. *Systematic Biology* 60: 343–357.

Schultze, R. E. 1997. South African atlas of climatology and agrohydrology. WRC Report TT82/96; ACRU Report 46. RSA, Pretoria, South Africa.

Schultze, R. E. 2007. South African atlas of climatology and agrohydrology. WRC Report 1489/1/06. RSA, Pretoria, South Africa.

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.

van der Niet, T., and S. D. Johnson. 2009. Patterns of plant speciation in the Cape floristic region. *Molecular Phylogenetics and Evolution* 51: 85–93.

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.

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.

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, UK.

Wang, I. J., and G. S. Bradburd. 2014. Isolation by environment. *Molecular Ecology* 23: 5649–5662.

Watt, M., and J. R. Evans. 1999. Proteoid roots: Physiology and development. *Plant Physiology* 121: 317–323.

Whittaker, R. H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs* 30: 279–338.

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

Wu, C. I. 2001. The genic view of the process of speciation. *Journal of Evolutionary Biology* 14: 851–865.

APPENDIX 1 Voucher specimens for each *Protea repens* population included in this study. All are deposited at the University of Connecticut (CONN) herbarium.

Population number – population name, CONN accession number:

1 – VanRhynsdorp, 228487; 2 – Cederberg, 228484; 3 – Banghoek, 228482; 4 – Riverlands, 228493; 5 – Ceres, 228492; 6 – Kleinmond, 228489; 7 – Montagu, 228485; 8 – Riviersonderend, 228484; 9 – Bredasdorp, 228488; 10 – DeHoop, 228496; 11 – Anysberg, 228480; 12 – Klein Swartberg, 228486; 13 – Garcia's Pass, 228490; 14 – Swartberg, 228477; 15 – Uniondale, 228481; 16 – Baviaanskloof, 228483; 17 – Kareedouw, 228479; 18 – Loerie Dam, 228491; 19 – Alicedale, 228495.