MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 4373–4386 doi: 10.1111/mec.12870

Genomics of the divergence continuum in an African plant biodiversity hotspot, I: drivers of population divergence in *Restio capensis* (Restionaceae)

C. LEXER,* R. O. WÜEST,† ‡ § S. MANGILI,* M. HEUERTZ,*¶ K. N. STÖLTING,* P. B. PEARMAN,† F. FOREST,** N. SALAMIN,† † ‡ ‡ N. E. ZIMMERMANN† and E. BOSSOLINI*§§

**Unit of Ecology & Evolution, Department of Biology, University of Fribourg, Fribourg CH-1700, Switzerland, †Landscape Dynamics Unit, Swiss Federal Research Institute WSL, Birmensdorf CH-8903, Switzerland, ‡Univ. Grenoble Alpes, LECA, Grenoble F-38000, France, §CNRS, LECA, Grenoble F-38000, France, ¶INIA Forest Research Centre, carretera de A Coruña km 7.5, E-28040 Madrid, Spain, **Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW 3DS, U.K., ††Department of Ecology & Evolution, University of Lausanne, CH-1015 Lausanne, Switzerland, ‡‡Swiss Institute of Bioinformatics, Quartier Sorge, CH-1015 Lausanne, Switzerland, §§Bayer CropScience, Technologiepark 38, 9052 Gent, Belgium*

Abstract

Understanding the drivers of population divergence, speciation and species persistence is of great interest to molecular ecology, especially for species-rich radiations inhabiting the world's biodiversity hotspots. The toolbox of population genomics holds great promise for addressing these key issues, especially if genomic data are analysed within a spatially and ecologically explicit context. We have studied the earliest stages of the divergence continuum in the Restionaceae, a species-rich and ecologically important plant family of the Cape Floristic Region (CFR) of South Africa, using the widespread CFR endemic *Restio capensis* (L.) H.P. Linder & C.R. Hardy as an example. We studied diverging populations of this morphotaxon for plastid DNA sequences and >14 400 nuclear DNA polymorphisms from Restriction site Associated DNA (RAD) sequencing and analysed the results jointly with spatial, climatic and phytogeographic data, using a Bayesian generalized linear mixed modelling (GLMM) approach. The results indicate that population divergence across the extreme environmental mosaic of the CFR is mostly driven by isolation by environment (IBE) rather than isolation by distance (IBD) for both neutral and non-neutral markers, consistent with genome hitchhiking or coupling effects during early stages of divergence. Mixed modelling of plastid DNA and single divergent outlier loci from a Bayesian genome scan confirmed the predominant role of climate and pointed to additional drivers of divergence, such as drift and ecological agents of selection captured by phytogeographic zones. Our study demonstrates the usefulness of population genomics for disentangling the effects of IBD and IBE along the divergence continuum often found in species radiations across heterogeneous ecological landscapes.

Keywords: Cape Floristic Region, isolation by adaptation, isolation by environment, population divergence, RAD sequencing, speciation

Received 26 March 2014; revision received 18 June 2014; accepted 9 July 2014

Introduction

Understanding the drivers of population divergence, speciation and species persistence is of great interest to molecular ecology, evolutionary biology and conservation

Correspondence: Christian Lexer, Fax: +41 26 300 9698; zE-mail: christian.lexer@unifr.ch

biology (Coyne & Orr 2004; Höglund 2009; Nosil 2012). Rapid recent progress in genomics is imparting fresh perspectives on studies of these topics, including the opportunity to assay many thousands of DNA sequence polymorphisms in diverging populations of nonmodel species (Feder *et al.* 2012; Gompert *et al.* 2012; Nosil 2012; Lexer *et al.* 2013). As a result of these developments, the potential to address the drivers of species diversification and persistence has never been greater, especially if genomic data are analysed within a spatially and ecologically explicit context.

Spatial patterns of divergence and gene flow for neutral genetic markers are often concordant with a model of isolation by distance (IBD), in which drift causes populations to become more different from one another at greater geographic distances (Wright 1943). This gene flow scenario is particularly frequent in plants, which often experience restricted dispersal of pollen or seeds (Vekemans & Hardy 2004). In alternative scenarios, coined 'isolation by environment' (IBE) and 'isolation by adaptation' (IBA), gene flow among populations living in different environments is limited primarily by selection against maladapted migrants (Nosil *et al.* 2009); note that IBE is related to 'isolation by adaptation' (IBA), but that the former is generally defined via environmental and the latter via adaptive phenotypic differences; Shafer & Wolf 2013). Indeed, diversity and gene flow patterns consistent with IBE are frequently re-covered in animals and plants with neutral genetic markers (reviewed by Sexton *et al.* 2013), presumably because spatially varying selection and the build-up of isolation among locally adapted populations lead to genetic associations (linkage disequilibrium, LD) between neutral markers and selected loci via genetic hitchhiking (Nosil 2012).

Disentangling the effects of geography and ecology on patterns of divergence and gene flow is greatly facilitated when many thousands of DNA markers can be studied in diverging populations (Wang *et al.* 2013), for example using recently developed genotyping tools such as Restriction site Associated DNA sequencing (RAD-seq; Baird *et al.* 2008). A population genomic approach allows discerning genome regions that diverge neutrally from those that respond to divergent selection across heterogeneous land- or seascapes (e.g. Hohenlohe *et al.* 2010; Gompert *et al.* 2012; Roesti *et al.* 2012; Stölting *et al.* 2013). Consequently, examining IBD and IBE for many genome regions or loci individually could provide a much more accurate and nuanced picture of the drivers of divergence compared with traditional neutral marker studies (Nosil 2012; Wang *et al.* 2013). Also, we now know that genomes often remain 'porous' throughout the divergence continuum ranging from population and ecotype divergence to complete speciation (Wu & Ting 2004; Feder *et al.* 2012; Nosil 2012; Stölting *et al.* 2013; The *Heliconius Genome Consortium* 2012). Thus, IBD and IBE can inform us about important mechanisms acting during early stages of this process (Nosil 2012; Shafer & Wolf 2013).

South Africa's Cape Floristic Region (CFR) represents a global plant biodiversity hotspot (>9000 vascular plant species in <90 000 km^2, including approximately 70% endemics; Goldblatt & Manning 2002). The CFR also represents a stunning ecological mosaic of different local climates (winter-rainfall seasonality and summer drought in the west, far more mesic conditions and less seasonal rainfall in the east), topographic features (rugged, mountainous terrain), soil types, plant communities and biomes (Linder 2001, 2005; Goldblatt & Manning 2002). The hypothesis has been put forward that this extreme environmental heterogeneity may have contributed to the origin and maintenance of biological diversity in the CFR (Midgley *et al.* 2003; Linder 2005; Schnitzler *et al.* 2011; Litsios *et al.* 2013). Population genomic studies of the divergence continuum in particularly diverse and ecologically dominant plant families in the CFR, such as those of the Restionaceae, Proteaecae and Ericaceae (Linder 2005), would help shed light on the roles of IBD and IBE in population divergence and speciation in the CFR. Unfortunately, few phylogeographic and population genetic studies are currently available for CFR plant taxa (reviewed by Lexer *et al.* 2013) including, to our knowledge, only one with sufficient power for population genomics (Rymer *et al.* 2010).

The South African Restionaceae (restiads) represent one of the largest (>300 species) and most intensely studied plant radiations of southern Africa (Linder 2005; Hardy *et al.* 2008; Linder & Hardy 2010). These reed-like graminoids form ecologically important elements of the hyperdiverse fynbos biome of the CFR (Goldblatt & Manning 2002). Restiad species are sensitive to soil moisture and segregate according to hydrological conditions (Araya *et al.* 2010), thus they likely respond to the steep climatic gradients present in the CFR. The well-developed phylogeny for Restionaceae (Hardy *et al.* 2008; Linder & Hardy 2010; Litsios *et al.* 2013), available knowledge on their physiological and phenological responses to water logging and drought (Araya *et al.* 2010), their 100% outcrossing (dioecious) breeding system and the known existence of morphologically recognizable subspecific forms in some species (HP Linder, IntKey available at: http://www.systbot.uzh.ch/ Bestimmungsschluessel/Restionaceae_en.html) all make restiads a highly relevant study group for addressing the drivers of population divergence and speciation in this important biodiversity hotspot.

Here, we use a combination of RAD sequencing (RAD-seq) of nuclear DNA and Sanger sequencing of

plastid DNA to address the drivers of population divergence in the CFR endemic *Restio capensis* (L.) H. P. Linder & C. R. Hardy. As this taxon is widespread within the CFR, occurs across heterogeneous and geographically extensive environmental gradients and is relatively homogenous with regard to morphology compared with other restiads (HP Linder, IntKey), we use it as an example for the earliest steps of the 'divergence continuum' in this important plant radiation. First, we characterize spatial patterns of genomic diversity for both nuclear and plastid genomes. Second, we use a Bayesian approach to examine the genomic distribution of divergence for nuclear RAD-seq polymorphisms and identify divergent outlier loci that are potentially affected by divergent selection across the environmental mosaic of the CFR. Third, we use spatial and climatic information and a generalized linear mixed modelling approach to test whether population divergence of different genomic fractions (plastid DNA, neutral RAD loci and divergent outlier RAD loci) and single nuclear RAD polymorphisms is driven by (i) isolation by distance (IBD), (ii) isolation by environment (IBE) due to the steep climatic gradients present in the CFR, (iii) effective dispersal limitation associated with membership of populations in different phytogeographic zones (areas of species endemism), that is IBE due to ecological and biogeographical determinants responsible for the formation of phytogeographic zone boundaries.

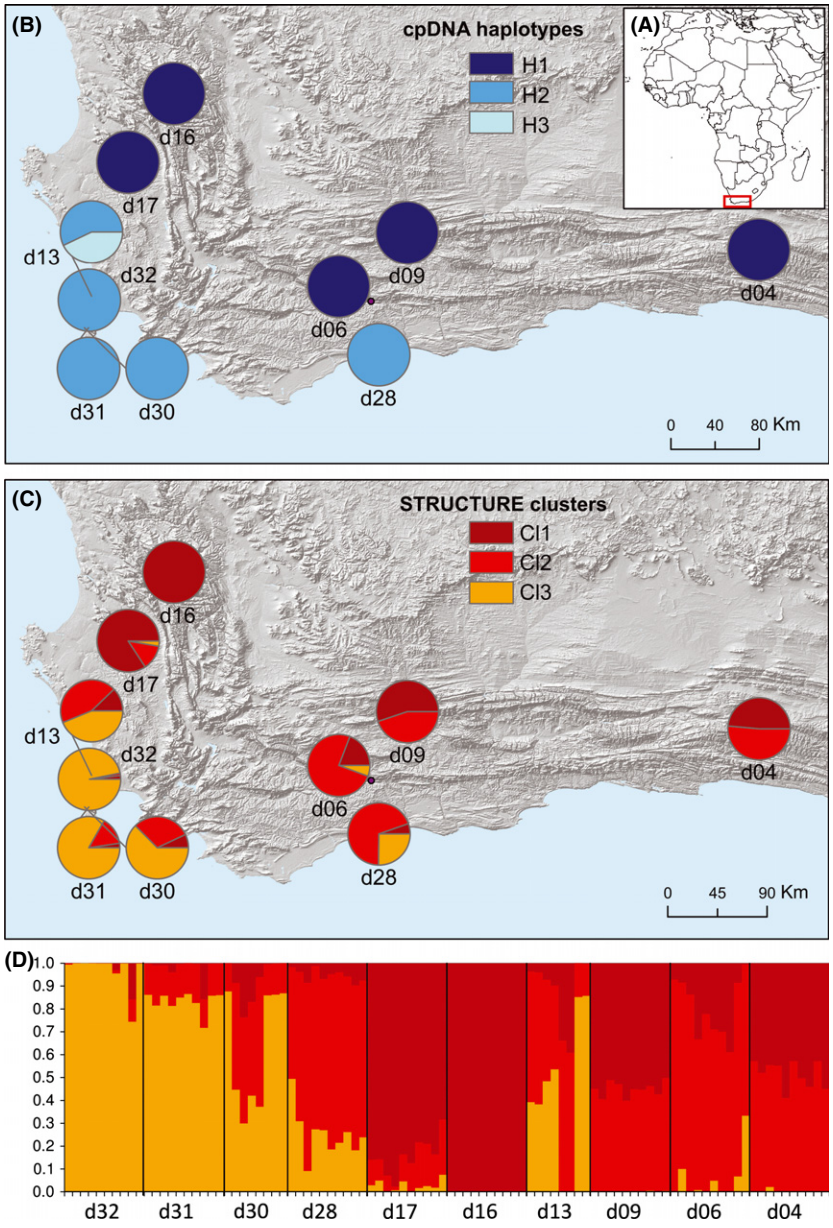

Fig. 1 Spatial organization of plastid DNA and nuclear genetic diversity in populations of *R. capensis* sampled in the CFR of South Africa. (A) Study region, the CFR is indicated by a red rectangle. (B) Spatial distribution of three plastid DNA haplotypes identified by sequencing, separated by two and three mutation steps. (C) Spatial and within-locality distribution of three 'genetic units' or clusters identified by Bayesian structure analysis of 14 434 nuclear RAD-seq polymorphisms. (D) Admixture proportions of individual specimens of *R. capensis* for the best-supported (K = 3 clusters) model from the same analysis.

Materials and methods

Plant materials

The CFR endemic *R. capensis* is an obligate outcrosser with wind-based pollen and passive seed dispersal (Dorrat-Haaksma & Linder 2000). Aboveground material from ten populations of *R. capensis* was collected in the South African CFR, covering most of the species' current geographic distribution. Populations were collected in a stratified manner with the goal of maximizing geographic area and climatic variation (Fig. 1; Table 1; Fig. S1, Supporting Information) while at the same time keeping the overall number of populations (and thus field work and DNA sequencing costs) manageable. The sample set was subjected to RAD-seq of nuclear DNA and Sanger sequencing of plastid DNA to yield information regarding the structure of nuclear and organellar genomic diversity in geographic and climate space. Up to ten individuals represented each population, thus facilitating the analysis of genetic variation present among 192 haploid nuclear genomes with codominant RAD markers. Our sample sizes for RAD-seq (16–20 allele copies per population or 192 copies in total; Table 1) are sufficient to detect rare alleles in the total data set, consistent with simulation results for biallelic markers (B-Rao 2001) and taking into account our filtering criteria implemented during polymorphism detection (below). Rare alleles may, however, remain undetected within local populations, which may lead to a slight overestimation of F_{ST} (Willing *et al.* 2012). A smaller number of individuals per population (between five and eight) were studied for plastid DNA polymorphisms (Table 1), which appeared justified based on low levels of within-population variation observed in pilot trials (see below).

RAD sequencing

Lyophilized stems of *R. capensis* were thoroughly ground with mortar and pestle in liquid nitrogen. DNA was extracted with the DNeasy Plant Mini kit (Qiagen) following the manufacturer's instructions. RAD library preparation was carried out in the laboratories of Floragenex (Eugene, Oregon, USA) following the protocol of Etter *et al.* (2011), using the high-fidelity restriction endonuclease *PstI* (New England Biolabs). This methylation-sensitive restriction enzyme effectively enriches for coding DNA among the sequenced fragments. Libraries of different individuals were barcoded with unique six-base tags that differed by at least two nucleotides. The libraries were pooled and sequenced with an Illumina HiSeq 2000 instrument. Individual barcode (=identifier) sequences were removed with Stacks (Catchen *et al.* 2011) retaining only the first 90 high-quality bases. Sequences containing only RAD library adapters or plastid DNA were filtered with a BLASTN search.

Identification and sequencing of plastid DNA polymorphisms

Sanger sequence-based genetic markers were developed in the plastome of *R. capensis*. The sequenced plastomes of two monocot species (*Thamnochortus insignis* Mast., Restionaceae; *Centrolepis monogyna* Benth., Centrolepidaceae) were used for primer selection (Givnish *et al.* 2010). Circa 150 primer pairs amplifying regions known to be highly polymorphic across a broad range of plant taxa (Shaw *et al.* 2007; Ebert & Peakall 2009; Scarcelli *et al.* 2011) were localized in these two plastomes. In addition, primer pairs were developed *de novo* from the two plastome sequences to target simple sequence repeats (SSRs) using FASTPCR version 6.1 (Kalendar *et al.*

Table 1 Spatial coordinates, sample sizes and estimates of genetic diversity for 10 populations of *Restio capensis* studied in the CFR of South Africa for polymorphic plastid DNA and nuclear RAD-seq markers

Population	Latitude (S)	Longitude (E)	N (plastid DNA/nrDNA)*	Gene diversity/ plastid DNA	Nucleotide diversity/RAD-SNPs†	% polymorphic RAD-SNPs
d04	−33.5657	23.8665	7/20	0	0.103 ± 0.051	50.2
d06	−33.9893	20.7134	8/20	0	0.060 ± 0.030	40.9
d09	−33.4259	21.0056	8/20	0	0.102 ± 0.051	53.5
d13	−33.9455	18.4364	7/16	0.571 ± 0.120	0.082 ± 0.041	42.3
d16	−32.4065	19.1072	8/20	0	0.108 ± 0.054	52.0
d17	−32.8502	18.7340	8/20	0	0.100 ± 0.050	53.8
d28	−34.4229	20.7741	6/20	0	0.060 ± 0.030	41.1
d30	−34.2035	18.3796	5/16	0	0.062 ± 0.031	36.5
d31	−34.2043	18.4123	8/20	0	0.082 ± 0.041	46.1
d32	−33.9774	18.4191	6/20	0	0.086 ± 0.043	41.9

*No. of plastid genomes and haploid nuclear genomes sampled for each population.
†Averages and standard deviations over 14 434 loci.

2011). Amplicons were designed to be 700–1500 base pairs (bp) in length. In total, 21 potential plastid DNA sequence regions were identified in this manner. All primer pairs were then tested on 12 representative samples of three different restiad species (*R. capensis* (L.) H.P. Linder & C.R. Hardy, *R. triticeus* Rottb., *Hypodiscus aristatus* (Thunb.) C. Krauss) to screen for polymorphism and robustness of amplicons within and among species. Three sequence regions (*ssrt9*, *ndhA* intron and *psbD* exon) were amplified consistently and yielded polymorphisms in *R. capensis*, and these were used in the present study (Table S1, Supporting Information). The plastid DNA regions were amplified in 50 μl reactions essentially following the protocols of Scarcelli *et al.* (2011), purified with ExoSap, concentrated by ethanol precipitation and Sanger-sequenced by Macrogen Inc. (Amsterdam, the Netherlands). Sequence chromatograms were edited and assembled with SEQUENCHER 4.8 (Gene Codes Corporation, Ann Arbor, Michigan, USA). Sequences from the three plastid DNA regions were concatenated for population genetic analysis, as the plastid genome is essentially a nonrecombining DNA molecule.

Data analyses

Pseudoreference-based RAD-SNP calling. We built a set of unique RAD-seq clusters from the sequences of the four best-covered individuals from four different populations, using the program SEED (Bao *et al.* 2011), allowing a mismatch tolerance of three nucleotides. A custom PERL script concatenated the obtained nonredundant reads into a single pseudoreference sequence. We then aligned reads of all individuals to this pseudoreference using BWA 0.5.9 (Li & Durbin 2009). We identified single nucleotide polymorphism (SNP) variants with the UNIFIED GENOTYPER of the Genome Analysis Tool Kit (GATK), using the SNP genotype likelihood model (McKenna *et al.* 2010). A custom PERL script filtered the table of variant calls, using a minimum PHRED quality threshold of 20 and a minimum and maximum variant coverage of 7 and 200 reads, respectively, and retaining only SNPs present in at least three individuals and with <40% missing data within populations. The N = 20 allele copies typically sampled in local populations translate into a detection threshold of at least 12 copies per population, that is our RAD-seq study captured primarily alleles with a local minor allele frequency (MAF) >10%. To reduce spurious variant calls from potentially paralogous loci, we subsequently eliminated all RAD stacks with more than three SNPs and all markers with a significant excess of heterozygotes relative to Hardy–Weinberg expectations (χ^2 test; $P < 0.05$). We converted the variant call format (vcf) file into input files for downstream data analyses with various software programs using custom PERL scripts. These bioinformatic analyses were conducted with the Vital-IT computing cluster maintained at the University of Lausanne and École Polytechnique Fédérale de Lausanne (EPFL), Switzerland.

Patterns of diversity and differentiation for nuclear and plastid DNA. We calculated diversity indices and analysis of molecular variance (AMOVA) for RAD-seq markers with ARLEQUIN 3.5.1 (Excoffier & Lischer 2010). We assessed patterns of genomic diversity of nuclear RAD loci using two complementary approaches, principal component analysis (PCA) in the ADE4 R package and the Bayesian approach implemented in the STRUCTURE 2.3.4 software (Falush *et al.* 2003). STRUCTURE was run under the admixture model allowing for correlated allele frequencies, using a burn-in of 50 000 followed by 100 000 Markov chain Monte Carlo (MCMC) iterations. Seven replicate runs were performed for numbers of genetic clusters from $K = 1$ to $K = 11$. We examined traces graphically to confirm chain convergence and inferred the most likely K present in the data following Evanno *et al.* (2005). Admixture proportions from STRUCTURE were averaged across runs for the best K using CLUMPP 1.1.2. (Jakobsson & Rosenberg 2007) and plotted on geographic maps using ARCMAP 10 (ESRI, Redlands, CA, USA).

We obtained estimates of plastid DNA diversity with ARLEQUIN 3.5 (Excoffier & Lischer 2010) and explored population genetic structure of plastid DNA using AMOVA in ARLEQUIN and spatial analysis of molecular variance (SAMOVA), making use of the simulated annealing procedure of Dupanloup *et al.* (2002). SAMOVA defines the number of groups of populations (K) that are maximally differentiated from each other but geographically homogeneous. Different values of K were tested (up to $K = 6$). The corresponding fixation indices (F_{SC}, F_{ST}, and F_{CT}) were calculated and their significance tested with 10 000 permutations. The best configuration of K corresponds to the maximum value of F_{CT} (divergence among groups) without any groups composed of single populations. The annealing process was repeated 100 times. Genetic barriers revealed by SAMOVA were then drawn on geographic climate maps (obtained from PCA) or phytogeographic zone maps (see below). We conducted 10 000 bootstrap replicates of a mismatch analysis in ARLEQUIN to test for sudden spatial or demographic expansion based on the sums of squared deviations (SSD) statistic and to determine significance levels.

Bayesian genome scan for potentially non-neutral outlier loci. We used the BAYESCAN 2.1 software (Foll & Gaggiotti 2008) to distinguish neutral RAD loci from those

potentially under divergent selection between populations (i.e. outlier loci with unusually large F_{ST} values). This fully Bayesian approach decomposes F_{ST} into a population-specific (beta) and a locus-specific (alpha) component, with significant alpha components indicating selection. The software implements reversible jump MCMC to estimate the posterior probabilities of models for each locus with or without the alpha component (i.e. with or without selection) and posterior odds ratios (analogous to Bayes factors) help reach decisions on evidence for or against selection. We ran the software with default priors and parameter settings explained in the user manual. We characterized the potential gene function of RAD sequences under divergent selection using BLAST2GO V.2.7.0 with a threshold of E-06 and a highest scoring pair (HSP) cut-off of 25 to compare them to known genes.

Generalized linear mixed modelling of genomic, spatial and environmental data. We used a generalized linear mixed modelling approach to test whether population divergence of different genomic fractions (plastid DNA, neutral RAD loci and divergent outlier RAD loci) and single unusually divergent outlier RAD loci (see above) was driven by (i) isolation by distance (IBD), (ii) isolation by environment (IBE) due to steep climatic gradients or (iii) IBE due to environmental determinants (related to climate, soil or biotic interactions), as captured in the delimitation of phytogeographic zones. We used the R package MCMCGLMM (Hadfield 2010) to identify geographic and ecological correlates of genetic divergence using generalized linear mixed models (GLMM; scripts and example input files in File S3). This approach accounted for nonindependence in the data resulting from the use of pairwise matrices of population-level metrics (see below). Different genetic divergence metrics were used as response variables: the number of pairwise differences estimated by ARLEQUIN for plastid DNA, F_{ST} for multilocus nuclear RAD-seq data sets (neutral vs. highly divergent outlier loci) and the allele frequency differential delta (Zhu *et al.* 2005) for single divergent outlier RAD loci (above). These distance metrics were chosen to account for the particular properties of each type of genetic data: the number of pairwise differences is well suited for relatively information-poor haplotypic plastid DNA data, F_{ST} is widely used to estimate divergence for multilocus nuclear DNA data, and delta measures divergence for single nuclear loci without confounding effects of within-population heterozygosity.

As predictor variables we used geospheric distances between populations to assess the effect of geography (IBD, 'GEO' from here onwards), Euclidean distances along the first axis of a principal component analysis (PCA) of 19 WorldClim variables (Hijmans *et al.* 2005) based on all pixels of the CFR at a resolution of 30 arc seconds ('ENV' for environment from here onwards; Table S2, Supporting Information) and a factor variable stating whether populations are situated in the same or in different phytogeographic zones (Linder 2001; Goldblatt & Manning 2002; 'PHY' from here onwards). The predictor ENV captured mainly annual precipitation and temperature seasonality (Table S2; Fig. S1, Supporting Information). Our geographic maps containing the positions of all sampled populations relative to projections of ENV (climate) and phytogeographic zones for the CFR (Fig. S1A,B, Supporting Information) illustrate the potentials and limits of our sample set to disentangle our three predictor variables. Note that for all studied populations, phytogeographic zone memberships (PHY) following Goldblatt & Manning (2002) coincided with zone memberships based on Linder (2001). Also note that IBD *sensu stricto* is defined as the relationship between $F_{ST}/(1-F_{ST})$ and log-transformed geographic distances. Nevertheless, initial analyses with these more complex metrics yielded very similar results (Mantel correlations and R^2) as regressions of untransformed geographic distances on F_{ST}, and thus, we used the latter in GLMM analyses.

Eight different models resulted from combinations of the three predictor variables: a null model without any predictor, three models with a single predictor variable (GEO, ENV or PHY), three with different combinations of two predictors and one with all three predictors. We used the deviance information criterion (DIC) and associated DIC differences and weights to compare all models for each genomic fraction and outlier RAD locus and to draw conclusions on the relative roles of different drivers of divergence (GEO, ENV or PHY) for each of them. Initial model comparisons also included GEO × ENV interaction terms, but these were never among the best-supported models (Table S3, Supporting Information), thus the more parsimonious model comparisons without interaction terms are discussed throughout this study. In the case of divergent outlier RAD loci, we also estimated 95% credible intervals (CI's) to determine the significance of each predictor in their respective models. Each outlier locus was then assigned to one of three scenarios, GEO, PHY or ENV. These were constructed by counting the number of loci for which a particular predictor was consistently significant in *all* models that included that predictor. MCMCGLMM was initiated with standard priors and run with a burn-in of 500 000 followed by 2 000 000 iterations with a thinning interval of 750. Chain convergence of MCMCGLMM was confirmed by inspecting trace plots using the CODA R package and by examining autocorrelation between successive MCMC samples. In addition to mixed modelling, the relationships between observed

genetic, geographic and environmental distances were explored graphically in the form of heat maps obtained by x-y interpolation through simple inverse distance weighting.

Results

Patterns of genomic diversity and differentiation

Sanger sequencing of plastid DNA revealed extremely low levels of variation (Table 1; Fig. 1B; three haplotypes defined by five SNPs identified among 1188 sequenced nucleotides) and clear structuring of plastid DNA haplotype diversity into a coastal and an inland lineage (Fig. 1B). This is consistent with the presence of genetic barriers along a major axis of climate differentiation across the CFR (Fig. S1, Supporting Information), although alternative explanations are possible (discussed below). Mismatch analysis of plastid DNA allowed us to reject the hypotheses of sudden demographic or spatial expansions (probabilities of SSD test statistics $P < 0.05$).

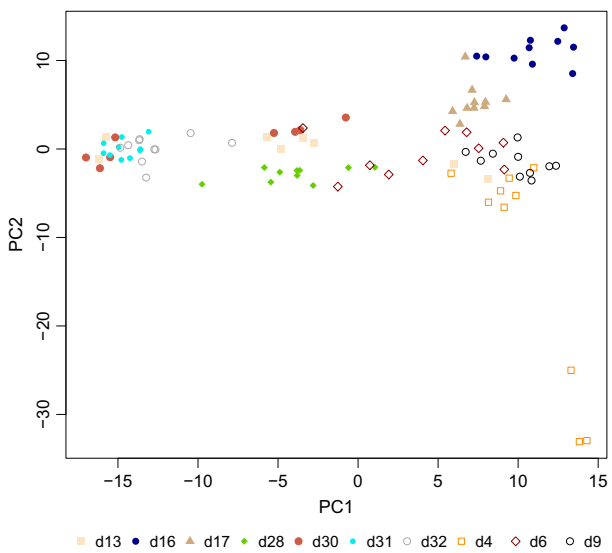

Fig. 2 First two axes from Principal Component Analysis of 14 434 high-quality RAD-seq polymorphisms studied in *R. capensis*. Populations are colour-coded following the legend at the bottom of the graph, and population IDs are consistent with Table 1 and Fig. 1. The first two axes (PC1 and PC2) explained 6.3 and 3.0% of the total variation in the data, respectively.

The pseudoreference sequence constructed from RAD-seq reads was composed of 856 799 unique clusters (Dryad entry doi:10.5061/dryad.060d2; Files S1 and S2). A total of 14 434 high-quality SNPs were called (Table S4; Table S5, Supporting Information, available from Dryad, doi:10.5061/dryad.060d2). Bayesian analysis of nuclear gene pool structure based on these SNPs suggested the presence of three genetic units or clusters in the data, consistent with gradual genetic differentiation between eastern/coastal and western/inland regions of the CFR (Fig. 1C). The spatial arrangement of these nuclear genetic clusters was largely congruent with the pattern re-covered by plastid DNA (Fig. 1). Principal component analysis (PCA) of the RAD-seq data revealed additional, more subtle patterns of population subdivision (Fig. 2) beyond those visible in the STRUCTURE analysis, with individuals from populations of each STRUCTURE cluster forming adjacent data clouds in principal component space. This is readily illustrated by individuals from populations d30, d31 and d32, which are dominated by the bright orange STRUCTURE cluster in Fig. 1C, visible on the left-hand side in the PCA in Fig. 2, and populations d16, d17 and d9, dominated by the dark red cluster in Fig. 1C, visible on the right-hand side in Fig. 2. Analysis of molecular variance (AMOVA) revealed contrasting partitionings of genetic diversity for plastid DNA and nuclear RAD-seq polymorphisms, with >93% and <3% of the genetic variance residing among local sampling localities for plastid DNA and nuclear RAD loci, respectively, considering 14 278 neutral RAD-seq markers only (Table 2).

Table 2 Analysis of molecular variance (AMOVA) in *R. capensis* estimated for three different genomic fractions: plastid DNA, neutral RAD-seq markers, highly divergent outlier RAD-seq markers

Genome fraction	Source of variation	d.f.*	Variance components	% of variation	F_{ST}†
Plastid DNA	Among	9	1.266	93.76	0.938***
	Within	61	0.084	6.24	
RAD-seq/neutral	Among	9	18.625	2.99	0.030***
	Within	182	603.775	97.01	
RAD-seq/outlier	Among	9	5.336	33.96	0.340***
	Within	182	10.375	66.04	

*Degrees of freedom.
†Three asterisks indicate significance at $P < 0.001$.

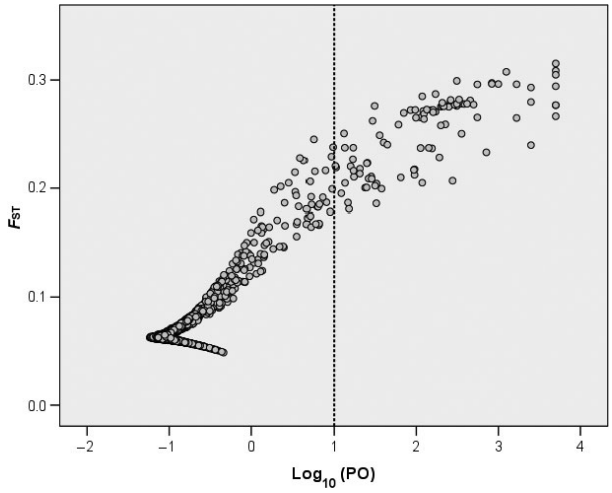

Fig. 3 Results of a Bayesian-based genome scan for genetic divergence (F_{ST}) outlier RAD-seq polymorphisms in *R. capensis*. *X*-axis, logarithm to base 10 of the posterior odds (PO) for models including selection for each locus. *Y*-axis, genetic divergence (F_{ST}) for each locus. The dashed line indicates the PO threshold of 1 used to define candidate outlier RAD loci potentially affected by divergent selection.

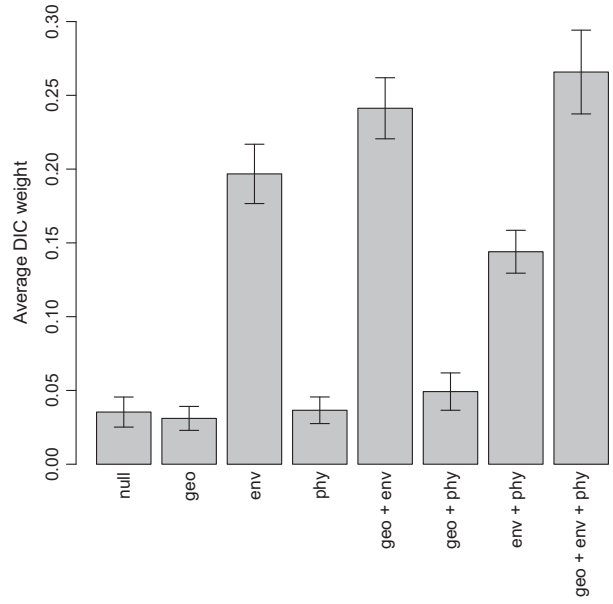

Fig. 4 Average deviance information criterion (DIC) weights (bars) and standard errors (whiskers) of mixed models predicting genetic divergence in 156 single outlier RAD-seq polymorphisms with geographic distance (GEO), an environmental distance (ENV) based on climate data and phytogeographic zone membership (PHY). On average, models that include 'ENV' as predictor variable receive significantly stronger support (=greater DIC weights) than models that do not include this predictor. See text for details.

Genome scan for divergently selected RAD-seq polymorphisms

A Bayesian-based genome scan for potentially non-neutral outlier RAD loci revealed 156 outliers in the upper tail (Fig. 3; based on a posterior odds threshold of 1), thus indicating that RAD-seq polymorphisms are potentially affected by divergent selection across the environmental mosaic of the CFR. For 16 RAD sequences, divergent outlier status was confirmed by more than one SNP, and thus, a total of 140 RAD sequences were identified as potential targets of selection (Table S6, Supporting Information). As expected, AMOVA of neutral vs. outlier RAD-seq polymorphisms revealed greatly increased among-population variance for the latter (3% vs. 34% of the total genetic variance, respectively; Table 2). BLAST and gene ontology analysis of the 140 outlier RAD sequences with BLAST2GO yielded 15 with significant hits, and for ten of these, annotations were obtained at various gene ontology (GO) levels (Table S7, Supporting Information).

Modelling of genetic, spatial and climate niche data

Generalized linear mixed modelling (GLMM) of genetic divergence for plastid DNA, neutral and unusually divergent outlier RAD-seq polymorphisms with GEO, ENV and PHY as predictor variables revealed climate (ENV) as a consistent driver of variation in the genomic data (Table 3A; Fig. 4). The environmental variable ENV (the top principal component of WorldClim climate data, capturing annual precipitation and temperature seasonality; Table S2, Supporting Information) consistently achieved the greatest DIC support, easily visible based on its zero information difference (delta DIC) to the best-supported model and the greatest DIC weights among all models for all three genomic fractions (plastid DNA, RAD/neutral, RAD/outlier; Table 3A). The ENV predictor was significant in all models of all three genomic fractions.

Modelling of genetic divergence for *each* of the 156 outlier RAD loci identified by BAYESCAN revealed the full breath of different locus-specific drivers of variation (Fig. 4; Fig. 5), including RAD loci for which divergence was clearly best explained by climate (Table 3B, locus 123), geography (Table 3B, locus 55) or phytogeographic zone membership (Table 3B, locus 62), respectively. These different drivers of variation are easily visualized in the form of heat maps of observed genetic, geographic and environmental (climate) distances (Fig. 5). Figure 5A illustrates the significant effect of geography on RAD locus 55, Fig. 5B the significant effect of climate on locus 123, and Fig. 5C the absence of effects of either geographic or climate distances on locus 62, a SNP for which divergence was best predicted by phytogeographic zoning. Fig. 5D illustrates the significant effect of climate on variation in plastid DNA, also detected by mixed modelling (Table 3).

Table 3 Exemplary results of generalized linear mixed models set up to predict genetic divergence between populations of *R. capensis* with geographic distance (GEO), an environmental (ENV) distance based on climate data and phytogeographic zone membership (PHY)

	Plastid DNA			RAD/neutral			RAD/outlier		
Model	DIC	Delta DIC	DIC weight	DIC	Delta DIC	DIC weight	DIC	Delta DIC	DIC weight
A									
Null	207.590	42.124	0.000	−223.704	28.824	0.000	−33.703	44.104	0.000
GEO	206.999	41.533	0.000	−220.718	31.810	0.000	−34.584	43.223	0.000
ENV	***165.466***	***0.000***	***0.461***	***−252.528***	***0.000***	***0.369***	***−77.807***	***0.000***	***0.598***
PHY	199.821	34.354	0.000	−236.533	15.995	0.000	−52.522	25.285	0.000
GEO + ENV	166.736	1.269	0.244	−251.584	0.944	0.230	−75.454	2.353	0.184
GEO + PHY	201.763	36.296	0.000	−240.405	12.123	0.001	−53.183	24.624	0.000
ENV + PHY	167.158	1.691	0.198	−250.640	1.888	0.143	−75.640	2.168	0.202
GEO + ENV + PHY	168.584	3.118	0.097	−251.805	0.723	0.257	−70.423	7.384	0.015
B									
Model	RAD-SNP no. 123			RAD-SNP no. 55			RAD-SNP no. 62		
	DIC	Delta DIC	DIC weight	DIC	Delta DIC	DIC weight	DIC	Delta DIC	DIC weight
Null	34.678	53.110	0.000	−62.073	3.880	0.038	−27.922	5.887	0.024
GEO	34.558	52.990	0.000	***−65.953***	***0.000***	***0.266***	−28.817	4.991	0.038
ENV	***−18.432***	***0.000***	***0.530***	−59.760	6.193	0.012	−27.394	6.414	0.019
PHY	27.182	45.614	0.000	−62.508	3.445	0.048	***−33.809***	***0.000***	***0.462***
GEO + ENV	−16.510	1.922	0.203	−65.853	0.100	0.253	−27.220	6.589	0.017
GEO + PHY	28.984	47.417	0.000	−65.121	0.832	0.176	−31.867	1.942	0.175
ENV + PHY	−16.428	2.004	0.195	−63.429	2.524	0.075	−32.047	1.762	0.191
GEO + ENV + PHY	−14.442	3.991	0.072	−64.541	1.412	0.131	−30.152	3.657	0.074

Shown are the deviance information criterion (DIC), DIC difference to the best-supported model (delta DIC) and DIC weights for each model. (A) Model comparisons for plastid DNA, 14 278 neutral RAD loci and 156 highly divergent outlier RAD loci. (B) Model comparisons for exemplary RAD loci for which divergence was best predicted by ENV (locus 123), GEO (locus 55) and PHY (locus 62). For each model comparison, the best-supported model is shown in bold italics.

Despite this diversity of results, climate (ENV) clearly was the most dominant driver of variation in multilocus and single-locus RAD-seq data. This becomes apparent from the greatly increased DIC weights of models that included ENV as opposed to models that did not include this predictor (Table 3; Fig. 4) and from the great number of RAD loci with consistently significant effects (tested via their 95% CI's) of climate (61 loci) compared with geography (13 loci) or phytogeographic zone membership (4 loci). For six of the climate-associated loci, preliminary functional classification was possible by BLAST searches and gene ontologies (Table S7, Supporting Information).

Discussion

In the present contribution, we have addressed the drivers of diversification at an early stage of the divergence continuum present in a species-rich and ecologically important plant radiation of the CFR of South Africa (Goldblatt & Manning 2002; Linder 2005; Hardy *et al.* 2008; Linder & Hardy 2010). We studied a set of diverging populations of *R. capensis*, a widespread Cape endemic, across the heterogeneous landscape mosaic of the CFR. This taxon is morphologically cohesive according to an available taxonomic key (see Introduction), but this does not preclude the presence of population subdivision or hitherto undetected ('cryptic') phenotypic variants or ecotypes. We used a population genomic approach (Feder *et al.* 2012; Nosil 2012) and analysed the genomic data within a spatially and ecologically explicit context (Nosil *et al.* 2009; Lee & Mitchell-Olds 2011; Andrew *et al.* 2012; Shafer & Wolf 2013; Wang *et al.* 2013).

Spatial patterns of genomic diversity

Sequencing of divergent populations of *R. capensis* for nuclear DNA (>14 400 markers from RAD-seq) and

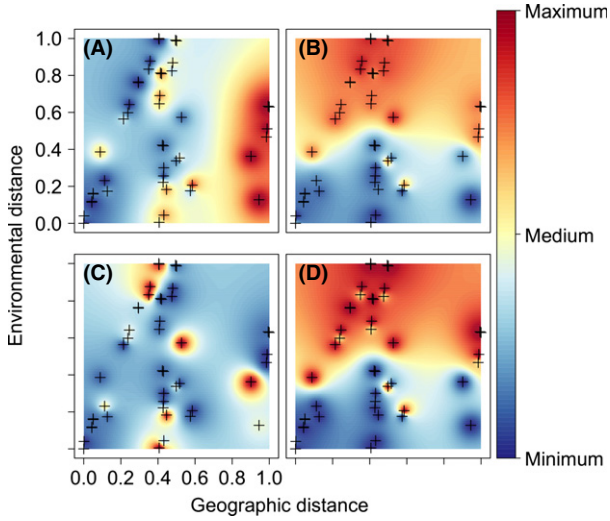

Fig. 5 Heat maps of observed data, showing single genetic loci in *R. capensis* for which divergence was driven either by geographic distance (GEO), an environmental distance (ENV) based on climate data or phytogeographic zone membership (PHY). Each heat map shows geographic distance (GEO) along the *x*-axis, climate distance (ENV) along the *y*-axis and colour-scaled genetic distance (low distances in blue to high distances in red, see legend). Examples are identical to those for which modelling results are shown in Table 3. (A) Locus 55, best predicted by geography (GEO). (B) Locus 123, best predicted by climate (ENV). (C) Locus 62, best predicted by phytogeographic zoning (PHY), thus revealing no obvious relationship with either axis. (D) Plastid DNA, best predicted by climate (ENV).

plastid DNA revealed rather homogenous levels of diversity for nuclear DNA throughout the sampling range and greatly reduced diversity for plastid DNA (Table 1). Spatial patterns of nuclear genomic differentiation inferred from Bayesian analysis of gene pool structure and from PCA of RAD-seq polymorphisms were largely concordant with those re-covered by plastid DNA (Figs. 1 and 2). Genomic differentiation into coastal and inland genetic clusters was less pronounced for nuclear DNA than for plastid DNA, as expected from the wind-pollinated, outcrossing breeding system of this species (Fig. 1C). Indeed, the highly contrasting patterns of genetic divergence seen in AMOVA, with <3% and more than 93% of the among-population variance explained by biparentally inherited neutral RAD-seq markers vs. presumably maternally transmitted plastid DNA, point to extremely efficient dispersal of pollen compared with seeds. Nevertheless, we refrained from estimating pollen to seed flow ratios (Ennos 1994) from the data because of doubts regarding the neutrality of plastid DNA (see below) and low statistical power due to low plastid DNA diversity.

Our Bayesian genome scan based on >14 400 markers from RAD-seq revealed a broad genomic distribution of divergence. In the upper tail of the genomic divergence distribution, 156 polymorphisms located on 140 different RAD sequences exhibited greater population divergence than expected under a neutral scenario (mean AMOVA $F_{ST} = 0.340$; $P < 0.001$; Table 3). Divergence of this magnitude is generally considered 'very high' in the literature and translates into <1 migrant per generation under equilibrium conditions (Conner & Hartl 2004), a value often considered the minimum for maintaining species cohesion. Our results suggest that despite the uniform morphological appearance of *R. capensis* and little divergence for neutral markers sampled from the nuclear genome (Fig. 3; Table 2), many markers or genome regions in *R. capensis* populations from across the ecological mosaic of the CFR have already diverged beyond common expectations for conspecific populations.

Genetic isolation by distance or environment?

Generalized linear mixed modelling of population divergence for different genomic fractions of DNA with GEO, ENV and PHY as predictors consistently indicated climate (ENV) as the main driver of genetic divergence among populations of *R. capensis* in the South African CFR (Table 3A, Fig. 4). The results strongly suggest IBE rather than IBD as the predominant mechanism explaining patterns of divergence and gene flow in this CFR endemic. How can consistent IBE for organellar and nuclear genomic compartments and for neutral and selected fractions of the nuclear genome, be reconciled with available knowledge of the divergence process?

In northern temperate taxa, genetic structure such as that seen in Fig. 1 is plausibly interpreted as signature of demographic history, that is colonization and population expansion from different glacial refugia (Hewitt 2000). We consider this a less likely explanation for the CFR taxa studied here, as currently available phylogeographic surveys of animals and plants from the CFR do not provide indications for consistent coastal/inland differentiation due to Pleistocene demographic shifts (reviewed by Lexer *et al.* 2013). Also, caution should be exercised when interpreting the gradual transition among genetic clusters identified by STRUCTURE (Fig. 1B) in terms of nuclear admixture, as PCA of the same nuclear genomic data indicates the presence of additional, more subtle patterns of differentiation (Fig. 2). Indeed, genetic differentiation for neutral RAD-seq polymorphisms in *R. capensis* was low (Table 2), which speaks against a dominant role for demographic processes in shaping the observed patterns of diversity. This view is supported by the absence of signatures of sudden demographic or range expansions in plastid DNA mismatch distributions (see Results). Thus, a more plausible explanation for IBE in this widespread

CFR endemic may lie in the mechanisms responsible for population divergence across an extremely heterogeneous ecological landscape (Linder 2001; Goldblatt & Manning 2002; Midgley *et al.* 2003; Schnitzler *et al.* 2011; Litsios *et al.* 2013).

According to current models of population divergence and speciation in the face of gene flow (Wu & Ting 2004; Feder *et al.* 2012; and Nosil 2012), divergence starts with direct selection on loci involved in local adaptation (a.k.a. 'speciation genes'; Wu & Ting 2004). This process may be followed by divergence hitchhiking (i.e. formation of 'differentiation islands' around selected loci because of reduced effective recombination between diverging populations), genome hitchhiking (genome-wide reduction of gene flow caused by selection) and, finally, post-speciation divergence. As pointed out by Feder *et al.* (2012), the observation of IBA (conceptually related to IBE; Shafer & Wolf 2013) for neutral markers suggests the presence of genome hitchhiking, that is selection on multiple unlinked mutations in the genome causes isolation and divergence. Thus, our observation of consistent IBE in a CFR endemic distributed across a highly heterogeneous landscape (Midgley *et al.* 2003; Schnitzler *et al.* 2011; Litsios *et al.* 2013) suggests the presence of genome hitchhiking or coupling of isolation factors more generally (Barton & de Cara 2009), during early stages of divergence. Note that this interpretation is also compatible with IBE being due to the maintenance of differentiation among locally adapted lineages in the face of gene flow (our first scenario). These scenarios are closely related, because restiads such as *R. capensis* are wind-pollinated obligate outcrossers (Dorrat-Haaksma & Linder 2000). Thus, divergence of populations or incipient species at the narrow spatial scale of the CFR likely involves opportunities for genetic contact and gene flow, despite possible episodes of allopatry.

Disentangling IBD and IBE with population genomic tools

Extending our mixed modelling of genetic, geographic, climatic and phytogeographic data to *individual* highly divergent outlier loci from RAD-seq confirmed climate (ENV) as the predominant driver of genetic divergence (Fig. 4) and provided additional insights. Perhaps most importantly, our population genomic approach of detecting locus-specific effects revealed 61 loci that had significant effects of ENV in each model that included that predictor (see Materials & Methods for all possible models), 13 loci with consistent effects of geography (GEO; the classical IBD scenario) and four loci with consistent effects of phytogeographic zone membership (PHY). Table 3B and Fig. 5 present examples for loci with clear model support (DIC weights) for each predictor. Our results corroborate the view of diverging populations as 'genomic mosaics', in which loci in the genome are differentially affected by selection, migration and drift (Wu & Ting 2004; Hohenlohe *et al.* 2010; Feder *et al.* 2012; Gompert *et al.* 2012; Nosil 2012). Three specific aspects of our modelling results for single RAD-seq loci appear particularly relevant to the genomics of the 'divergence continuum' here.

First, our finding of a non-negligible number of outlier loci with consistently significant effects of geography (GEO; 13 loci) indicates that drivers other than the ENV variable indeed contribute to population divergence for these genome regions. Potential candidate mechanisms include genetic drift [recall that IBD essentially signifies migration–drift equilibrium (Conner & Hartl 2004)] and divergent selection due to ecological agents that covary more strongly with geography than with climate predictors (Coop *et al.* 2010).

Second, our finding of several loci with consistent effects of phytogeographic zone membership (PHY; four loci) also indicates that additional ecological agents of selection contribute to adaptive population divergence. Phytogeographic zones are defined by explicit mathematical or descriptive botanical inference of areas of species endemism and are generally interpreted in terms of dispersal limitation due to a great number of ecogeographic barriers (Linder 2001; Goldblatt & Manning 2002). In the case of the CFR, these may include differences in soil types, topographic features (e.g. mountains separated by lowlands), climate zones (Linder 2001) and fire regimes (Litsios *et al.* 2013). Seen through the population geneticist's eyes, successful (i.e. effective) dispersal or its limitation will depend crucially on ecological selection during multiple life stages, especially seedling establishment. Thus, our PHY predictor represents a black box or surrogate containing numerous ecological agents of selection, which combined constitute ecogeographic barriers between populations or species (Linder 2001; Coyne & Orr 2004).

A third point of potential interest refers to our finding that genetic divergence of plastid DNA, in principle a single nonrecombining locus, was better predicted by climate than by geographic distance (Table 3A; Fig. 5D). This, together with the greatly reduced diversity observed for plastid DNA (Table 1), clear differentiation into coastal and inland lineages (Fig. 1; Fig. S1, Supporting Information), and the absence of a recent, sudden expansion signature (see mismatch distributions), raises the possibility that the patterns we found may result from a past selective sweep. Sweeps in plant plastid genomes have been reported previously (Muir & Filatov 2007) and are plausible in an environment with great variation in abiotic (drought) stress. Resequencing of complete plastomes would provide the

necessary power to test this hypothesis by searching for molecular signatures of positive selection.

Conclusions and perspectives

Analysis of population genomic data in a spatially and ecologically explicit context using appropriate analytical tools can disentangle the drivers of diversification along divergence continuums such as those seen in Restionaceae from the CFR. This approach has often been limited by the ability to type many genetic markers in the genomes of nonmodel species (unlikely to be a major limitation with available and upcoming sequencing technologies) and by the availability of information on ecological population differences that are at least partially uncoupled from geographic distances. The CFR represents a prime example for this type of spatial and environmental setting, and we suspect that studies of other biodiversity-rich regions (Myers *et al.* 2000) with strong environmental structuring will be similarly fruitful. Spatial variation in climate appears to be a major driver of diversification at early stages of the divergence continuum in these African restiads, and many genetic loci across the genome of *R. capensis* have responded to it by genome hitchhiking or coupling effects. Among highly divergent outlier RAD sequences with strong indications for climate associations, several exhibit great sequence similarity to proteins with known function during stress response (e.g. glycosyl transferases, protein kinases and stress signalling proteins; Table S7, Supporting Information). Although we caution against over-interpretation, the sequence similarities indicate a potential to identify the functional roles of adaptive alleles by cross-checking highly divergent RAD sequences with reference transcriptomes or genomes, once available. Studies that extend the spatially and ecologically explicit genomic approach employed here to multiple species at different stages of the divergence continuum in Restionaceae are forthcoming. The use of additional ecological variables, including more refined and more proximal climate and environmental variables (e.g. site water balance or soil fertility), will aid the identification of the drivers responsible for the origin and maintenance of genomic diversity. If carried out for multiple radiations in a comparative manner, this research may ultimately reveal which aspects of current and projected environmental change are most crucial in the context of protecting the biological diversity present in one of the world's richest floras.

Acknowledgements

We thank Adrian Möhl for his valuable assistance during fieldwork, Cape Nature and the Province of the Eastern Cape, South Africa, for issuing sampling permits, Alexa Oppliger and Tressa Atwood for help in the laboratory, and the team of Vital-IT/Swiss Institute of Bioinformatics for support during data analysis. This study has benefited from helpful discussions with participants and organizers of the AETFAT (Association pour l'Etude Taxonomique de la Flore d'Afrique Tropicale) 2014 conference held in Stellenbosch, South Africa. Financial support came from Swiss National Science Foundation (SNF) Sinergia grant CRSII3_125240 and a Marie Curie Intra-European Fellowship to MH.

References

Andrew RL, Ostevik KL, Ebert DP, Rieseberg LH (2012) Adaptation with gene flow across the landscape in a dune sunflower. *Molecular Ecology*, **21**, 2078–2091.

Araya YN, Silvertown J, Gowing DJ, McConway K, Linder HP, Midgley GF (2010) Variation in d13C among species and sexes in the family Restionaceae along a fine-scale hydrological gradient. *Austral Ecology*, **35**, 818–824.

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.

Bao E, Jiang T, Kaloshian I, Girke T (2011) SEED: efficient clustering of next-generation sequences. *Bioinformatics*, **27**, 2502–2509.

Barton NH, de Cara MAR (2009) The evolution of strong reproductive isolation. *Evolution*, **63**, 1171–1190.

B-Rao C (2001) Sample size considerations in genetic polymorphism studies. *Human Heredity*, **52**, 191–200.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci *de novo* from short-read sequences. *Genes, Genomes, Genetics*, **1**, 171–182.

Conner JK, Hartl DL (2004) *A Primer of Ecological Genetics*. Sinauer Associates, Sunderland, Massachusetts.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics*, **185**, 1411–1423.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, Massachusetts.

Dorrat-Haaksma E, Linder HP (2000) *Restios of the Fynbos*. Struik Nature, Cape Town.

Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, **11**, 2571–2581.

Ebert D, Peakall R (2009) A new set of universal *de novo* sequencing primers for extensive coverage of noncoding chloroplast DNA: new opportunities for phylogenetic studies and cpSSR discovery. *Molecular Ecology Resources*, **9**, 777–783.

Ennos RA (1994) Estimating the relative rates of pollen and seed migration among plant populations. *Heredity*, **72**, 250–259.

Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA (2011) SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Molecular Methods for Evolutionary Genetics*, **772**, 157–178.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology*, **10**, 564–567.

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

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28**, 342–350.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.

Givnish TJ, Ames M, McNeal JR *et al.* (2010) Assembling the tree of the Monocotyledons: plastome sequence phylogeny and evolution of Poales. *Annals of the Missouri Botanical Garden*, **97**, 584–616.

Goldblatt P, Manning JC (2002) Plant diversity of the Cape region of southern Africa. *Annals of the Missouri Botanical Gardens*, **89**, 281–302.

Gompert Z, Lucas LK, Nice CC, Fordyce JA, Forister ML, Buerkle CA (2012) Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution*, **66**, 2167–2181.

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

Hardy CR, Moline P, Linder HP (2008) A phylogeny for the African Restionaceae and new perspectives on morphology's role in generating complete species phylogenies for large clades. *International Journal of Plant Sciences*, **169**, 377–390.

Heliconius Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, **487**, 94–98.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Höglund J (2009) *Evolutionary Conservation Genetics*. Oxford University Press, Oxford.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *Plos Genetics*, **6**, e1000862.

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

Kalendar R, Lee D, Schulman AH (2011) Java web tools for PCR, *in silico* PCR, and oligonucleotide assembly and analysis. *Genomics*, **98**, 137–144.

Lee CR, Mitchell-Olds T (2011) Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Molecular Ecology*, **20**, 4631–4642.

Lexer C, Mangili S, Bossolini E *et al.* (2013) 'Next generation' biogeography: towards understanding the drivers of species diversification and persistence. *Journal of Biogeography*, **40**, 1013–1022.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Linder HP (2001) On areas of endemism, with an example from the African Restionaceae. *Systematic Biology*, **50**, 892–912.

Linder HP (2005) Evolution of diversity: the Cape flora. *Trends in Ecology and Evolution*, **10**, 536–541.

Linder HP, Hardy CR (2010) A generic classification of the Restioneae (Restionaceae), southern Africa. *Bothalia*, **40**, 1–35.

Litsios G, Wüest RO, Kostikova A *et al.* (2013) Effects of a fire response trait on diversification in replicated radiations. *Evolution*, **68**, 453–465.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.

Midgley GF, Hannah L, Millar D, Thuiller W, Booth A (2003) Developing regional and species-level assessments of climate change impacts on biodiversity in the Cape Floristic Region. *Biological Conservation*, **112**, 87–97.

Muir G, Filatov D (2007) Selective sweep in the chloroplast DNA of dioecious Silene (Section Elisanthe). *Genetics*, **177**, 1239–1247.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Nosil P (2012) *Ecological Speciation*. Oxford University Press, Oxford.

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

Roesti M, Hendry AP, Salzburger W, Berner D (2012) Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology*, **21**, 2852–2862.

Rymer PD, Manning JC, Goldblatt P, Powell MP, Savolainen V (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.

Scarcelli N, Barnaud A, Eiserhardt W *et al.* (2011) A set of 100 chloroplast DNA primer pairs to study population genetics and phylogeny in monocotyledons. *PLoS ONE*, **6**, e19954.

Schnitzler J, Barraclough TG, Boatwright JS *et al.* (2011) Causes of plant diversification in the Cape biodiversity hotspot of South Africa. *Systematic Biology*, **60**, 343–357.

Sexton JP, Hangartner SB, Hoffmann AA (2013) Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution*, **68**, 1–15.

Shafer ABA, Wolf JBW (2013) Widespread evidence for incipient ecological speciation: -a meta-analysis of isolation-by-ecology. *Ecology Letters*, **16**, 940–950.

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany*, **94**, 275–288.

Stölting KN, Nipper R, Lindtke D *et al.* (2013) Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Molecular Ecology*, **22**, 842–855.

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, **13**, 921–935.

Wang IJ, Glor RE, Losos JB (2013) Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology Letters*, **16**, 175–182.

Willing E-M, Dreyer C, van Oosterhout C (2012) Estimates of genetic differentiation measured by F_{ST} do not necessarily require large sample sizes when using many SNP markers. *PLoS ONE*, **7**, e42649.

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

Wu CI, Ting CT (2004) Genes and speciation. *Nature Reviews Genetics*, **5**, 114–122.

Zhu X, Luke A, Cooper R *et al.* (2005) Admixture mapping for hypertension loci with genome-scan markers. *Nature Genetics*, **37**, 177–181.

C.L., R.O.W., P.B.P., N.E.Z., N.S. and F.F. designed the research, S.M., E.B. and R.O.W. collected data, R.O.W., S.M., M.H., K.N.S., P.B.P., E.B. and C.L. analysed the data, and C.L. wrote the manuscript with input from all co-authors.

Data accessibility

Pseudoreference sequences and the complete RAD-seq genotype data set are available on Dryad (doi:10.5061/dryad.060d2) as Files S1 and S2 and Table S5 (Supporting Information), respectively. The RAD sequences harbouring 156 highly divergent outlier polymorphisms are available as Table S6 (Supporting Information). Sanger sequences for plastid DNA are available from GenBank, accession nos KM204769 – KM204981.

Supporting information

Additional supporting information may be found in the online version of this article.

File S1 Pseudoreference of concatenated unique RAD-seq clusters, separated by a stretch of five Ns, pseudochromosomes 1–6. Zip-compressed fasta file format. Available on Dryad, doi:10.5061/dryad.060d2.

File S2 Pseudoreference of concatenated unique RAD-seq clusters, separated by a stretch of five Ns, pseudochromosomes 7–12. Zip-compressed fasta file format. Available on Dryad, doi:10.5061/dryad.060d2.

File S3 R-script and input files (=data sets) used for generalized linear mixed modelling (GLMM) of genomic, spatial and environmental data. Details are provided in comment lines of script.

Table S1 Primer and polymorphism information for three plastid DNA sequence regions studied in *R. capensis*. Genbank accession nos KM204769–KM204981.

Table S2 Proportions of variance explained and standard deviations of the top three principal components (PC) of 19 WorldClim variables (BIO1-19) for the study region in the Cape Floristic Region of South Africa.

Table S3 Results of GLMM set up to predict genetic divergence between populations of *R. capensis* with geographic and environmental predictors as in Table 3 of main study, but including GEO × ENV interaction terms.

Table S4 RAD sequencing summary statistics for 96 individuals studied in *R. capensis*.

Table S5 RAD sequencing data set for 96 individuals from 10 populations of *R. capensis* from the Cape Floristic Region of South Africa. Available on Dryad, doi:10.5061/dryad.060d2.

Table S6 Positional and sequence information for 156 high-divergence outlier SNPs identified by BAYESCAN, located on 140 different RAD sequences.

Table S7 Gene ontology (GO) results for *R. capensis* high-divergence outlier sequences.

Fig. S1 Genetic barriers from plastid DNA SAMOVA drawn on climate and geographic maps.