Mismatches between niches and geographic ranges of plants depend on life history traits

Jörn Pagel[1], Martina Treurnicht[1,2,3], William J. Bond[4], Tineke Kraaij[5], Henning Nottebrock[6], AnneLise Schutte-Vlok[7], Jeanne Tonnabel[8], Karen J. Esler[2] & Frank M. Schurr[1]

[1]*Institute of Landscape and Plant Ecology, University of Hohenheim, August-von-Hartmann-Str. 3, 70599 Stuttgart, Germany.*

[2]*Department of Conservation Ecology and Entomology, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa.*

[3]*South African Environmental Observation Network (SAEON), Private Bag X7, Claremont 7735, South Africa.*

[4]*Department of Biological Sciences, University of Cape Town, Rondebosch 7701, South Africa.*

[5]*School of Natural Resource Management, Nelson Mandela University, George 6529, South Africa.*

[6]*Department of Biology and Microbiology, SD AES, South Dakota State University, Brookings, South Dakota 57005, USA.*

[7]*Scientific Services, CapeNature, Private Bag X658, Oudtshoorn, 6620, South Africa.*

[8]*Department of Ecology and Evolution, Le Biophore, Quartier UNIL-Sorge, University of Lausanne, 1015 Lausanne, Switzerland.*

The ecological niche of a species describes the variation in population growth rates along environmental gradients[1,2] that drives geographic range dynamics[3]. Niches are thus central for understanding species' geographic distributions and forecasting biodiversity responses to environmental change[4]. However, theory predicts that migration limitation, source-sink dynamics and time-lagged local extinction can cause mismatches between niches and geographic distributions[3-5]. It is still unclear how relevant these niche-distribution mismatches are for biodiversity dynamics and how they depend on species life history traits. This is mainly due to a lack of the comprehensive, range-wide demographic data needed to directly infer ecological niches for multiple species[5-8]. Here we quantify niches from extensive demographic measurements along environmental gradients across the geographic ranges of 26 plant species (Proteaceae; Cape Floristic Region, South Africa). We then test whether life history explains variation in species' niches and niche-distribution mismatches. Niches are generally wider for species with high seed dispersal and/or persistence abilities. Life history traits also explain the considerable interspecific variation in niche-distribution mismatches: poorer dispersers are absent from larger parts of their potential geographic ranges, whereas species with higher persistence ability more frequently occupy environments outside their ecological niche. Our study thus identifies major demographic and functional determinants of species' niches and geographic distributions. It highlights that the inference of ecological niches from geographical distributions is most problematic for poorly dispersed and highly persistent species. We conclude that the direct quantification of ecological niches from demographic responses to environmental variation is a crucial step towards a better predictive understanding of biodiversity dynamics under environmental change.

In 1957, George Evelyn Hutchinson introduced his seminal concept of a species' ecological niche[9]. The Hutchinsonian niche is defined as the set of environmental conditions for which demographic rates result in a positive intrinsic population growth rate and thus permit a species to form self-perpetuating populations[1,2]. This niche concept has received much attention as a theoretical foundation for explaining the geographic distributions of species and forecasting range shifts under environmental change. However, this use of the Hutchinsonian niche concept has been critically revisited in recent years[4,5]. Theoretical models of range dynamics predict that the geographic distribution of a species does not necessarily match the geographic projection of the species' niche. This is because geographic distributions are structured by extinction and colonization events that arise from a dynamic interplay of spatial variation in demographic rates, local population persistence and dispersal[3]. Migration limitation, source-sink dynamics and time-lagged extinction can thus cause mismatches between niches and geographic distributions[4,5]. Strong niche-distribution mismatches can severely bias niche estimates and forecasts of species distribution models[10,11], which are widely used in global change-biodiversity assessments[12]. The extent of niche-distribution mismatches is, however, poorly known because demography-based niches were so far only quantified for a few single species[13-15], which precluded comparative analyses. More than 60 years after Hutchinson introduced his niche concept, it is thus unclear how relevant niche-distribution mismatches are in the real world and how they depend on the dispersal and persistence ability of species.

Here we quantify niches from extensive demographic data for 26 closely related plant species to analyse how niche sizes, geographic range sizes and, finally, niche-distribution mismatches depend on species' life history traits. Our study species are shrubs of the Proteaceae family endemic to the Cape Floristic Region, a global biodiversity hotspot[16]. All study species are serotinous: they store their seeds over multiple years in a canopy seedbank until fire triggers seed release, wind-driven seed dispersal and subsequent establishment of new recruits[17]. Fire is also the predominant cause of mortality of established adults[18]. This fire-linked life cycle allows the efficient measurement of key demographic rates in single visits to each population[19] and enabled us to collect data that are informative on variation in population growth rates across each species' geographic range.

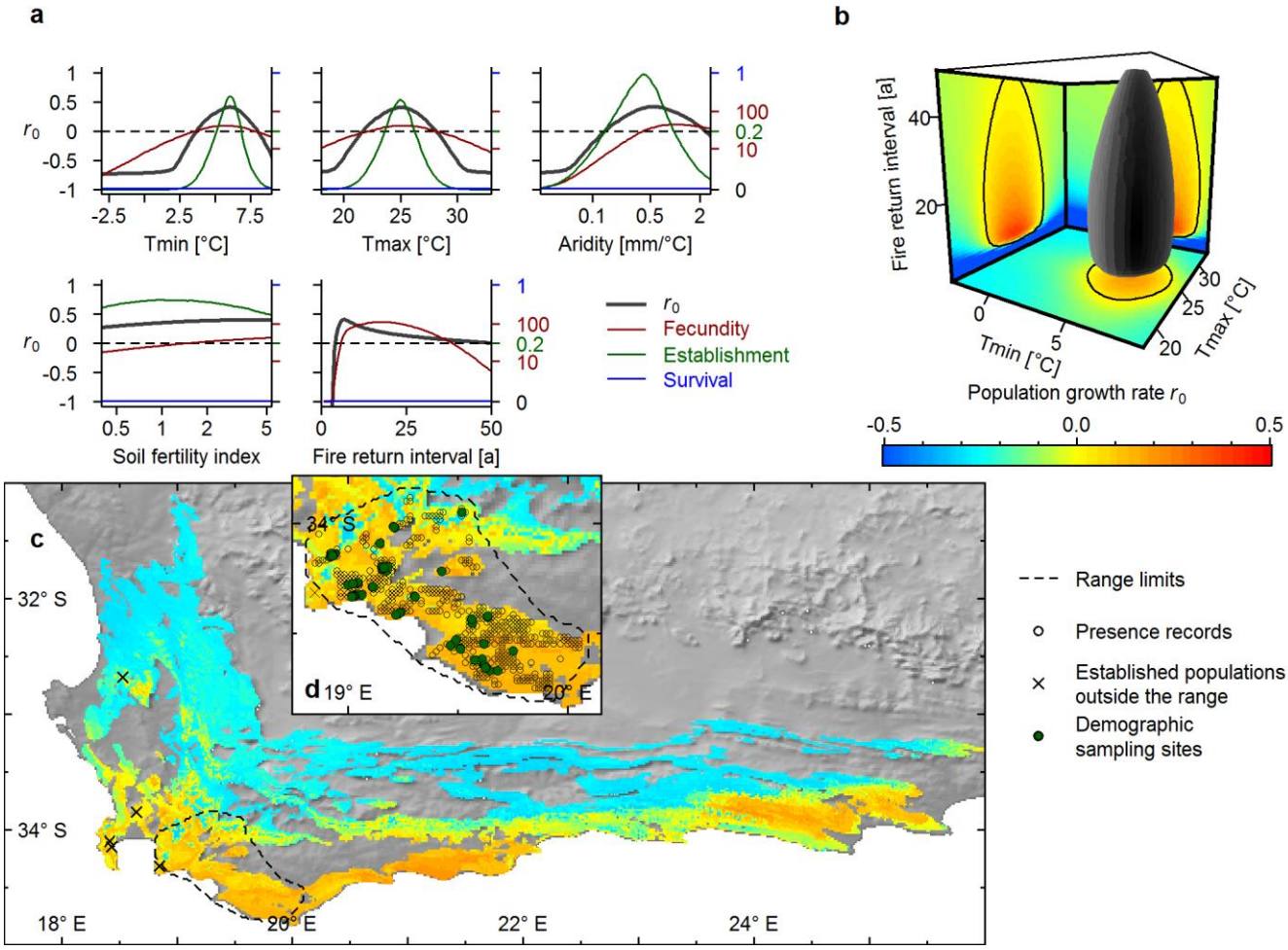

Figure 1 | The Hutchinsonian niche and geographic distribution of *Protea longifolia*. a, Responses of key demographic rates and the resulting annual intrinsic population growth rate (r_0) to variation in minimum winter temperature (T_{min}), maximum summer temperature (T_{max}), indices of summer aridity and soil fertility, and fire return interval. **b,** The niche hypervolume (grey) in a 3-d environmental sub-space delimits the conditions for which $r_0 > 0$. The marginal 2-d heat maps show the predicted r_0 when all other niche axes are set to their respective optima. **c,** Geographic projection of r_0 across the Fynbos biome (coloured areas) in comparison to the natural geographic range (dashed lines) and to populations established outside the natural range (crosses). **d,** Enlarged map showing presence records of natural populations (open circles) and demographic sampling sites (green circles). Model predictions in all subplots are the medians of the respective Bayesian posterior distributions.

We analysed a total of 3,617 population-level records of fecundity, recruitment and adult fire survival (Extended Data Table 1). The analyses used hierarchical demographic response models that describe for each key demographic rate the species-specific response curves to variation in climatic-edaphic conditions (minimum winter temperature, maximum summer temperature, summer aridity, soil fertility), fire return intervals and intraspecific density (Fig. 1a; Extended Data Fig. 1 for all study species). The fitted response curves explained much variation in long-term fecundity (Nagelkerke's R^2_N 0.29–0.91 across species, mean = 0.58; Extended Data Table 2), recruitment (R^2_N = 0.05–0.84, mean = 0.42) and adult fire survival (R^2_N = 0.41–0.83, mean = 0.62). The response curves of all three

demographic rates were then integrated in models of density-dependent local population dynamics that predict variation of intrinsic (low-density) population growth rates (r_0) along environmental gradients. This delimits each species' niche as a hypervolume in environmental space for which r_0 is positive (Fig. 1b).

We first examined how the size of the estimated niches is related to the dispersal and persistence abilities of species. Niche sizes were quantified separately for fire return intervals, which show strong temporal variation in any given location[20], and in the four-dimensional niche space spanned by gradients in long-term averages of climatic-edaphic conditions (hereafter called 'disturbance niche size' and 'environmental niche size', respectively). Dispersal ability was quantified from species-specific parameterizations of a trait-based, mechanistic model of wind-driven seed dispersal[21]. Persistence ability was characterized by resprouting as a key functional trait: Some of the study species ('resprouters', n = 7) possess fire-protected meristems from which they can resprout and are thus more likely to survive fire than species lacking this trait ('nonsprouters', n = 19)[18,19]. Since resprouter populations do not exclusively rely on successful reproduction in each fire cycle, they are expected to be less vulnerable to short fire return intervals that prevent the build-up of canopy seed banks[18]. We indeed found that resprouting ability had a clear positive effect on disturbance niche size (Fig. 2a), whereas dispersal ability had no effect (Extended Data Table 3). In contrast, environmental

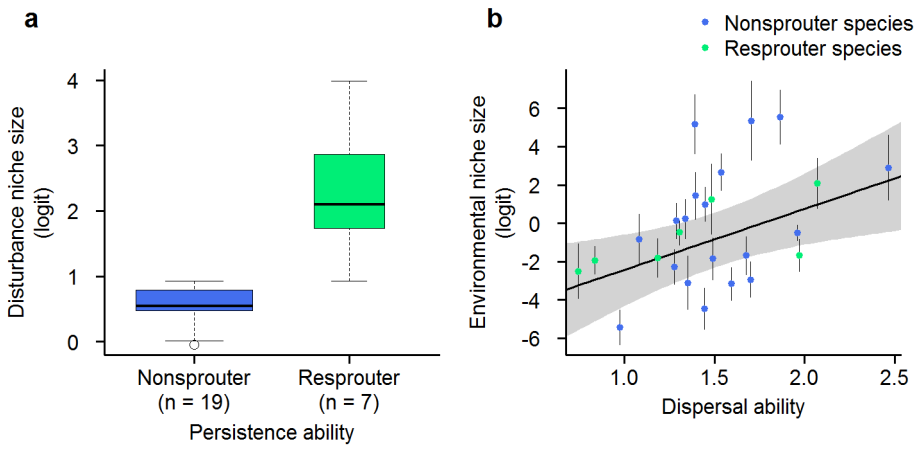

Figure 2 | Life history trait effects on niche sizes. a, Effect of persistence ability on disturbance niche size. **b,** Effect of dispersal ability on environmental niche size (points: posterior means, bars: posterior standard deviations). The line shows the estimated linear regression (posterior means, 90% credibility interval as shaded areas, slope = 1.22, p = 0.016).

niche size showed a strong positive relationship with dispersal ability (Fig. 2b). This finding is consistent with a scenario of correlational selection on niche size and dispersal, where narrower niches select for lower dispersal distances and vice versa[22,23].

To investigate mismatches between species' demographic niches and geographic distributions, we projected niches from environmental space into geographic space (Fig. 1c; Extended Data Fig. 2 for all study species). For each species, we then compared this potential geographic range (the region of demographic suitability where predicted $r_0 > 0$) to independent and extensive distribution records[17]. Several species showed a remarkably strong agreement between potential ranges and observed geographic distributions (AUC values >0.8 for ten of the 26 study species). However, there also was a high variation in this agreement (resp. mismatch) between demographic suitability and geographic distributions across species (AUC 0.55-0.97, mean = 0.77, Extended Data Table 2). Since processes that can generate these mismatches are expected to act on different spatial scales[24], we further analysed the relationship between demographic suitability and species occurrence separately at large and small spatial scales.

On large spatial scales, dispersal limitation can cause an incomplete filling of potential ranges, since species are unable to reach suitable areas[3-5]. We thus tested for a positive relationship between dispersal ability and range filling (measured as the proportion of the potential range that is covered by a species' geographic range). As expected, range filling strongly increased with dispersal ability (Fig. 3a). Hence, good dispersers not only have larger environmental niches (Fig. 2b) and thus tend to have larger potential geographic ranges but they also fill more of their potential ranges, so that both factors add up to explain the larger geographic ranges of good dispersers (Extended Data Table 3). In contrast, persistence ability had no effect on range filling. Absence from suitable areas could of course also indicate that occurrence is limited by environmental factors not considered in our analyses. Ideally, transplant experiments could be used to test model predictions of suitable areas outside the range[25]. Such large-scale transplant experiments do, however, pose substantial ethical and logistic problems[26]. Instead, we made use of the fact that most of our study species were established by humans (e.g. flower producers) in natural ecosystems outside their geographic range[17] (Fig. 1c). When evaluating our model extrapolations for these naturalized populations, we found that the predicted r_0 was positive

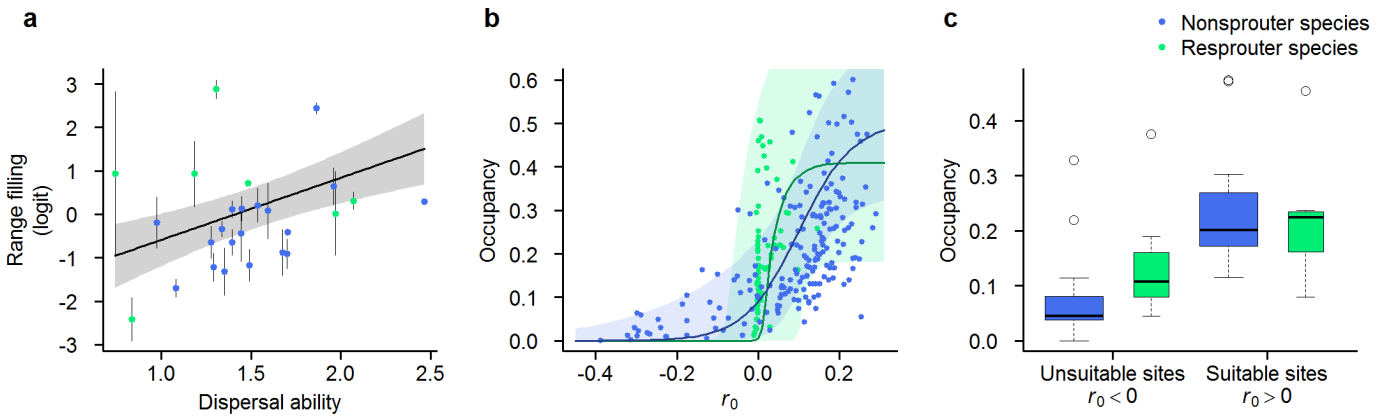

Figure 3 | Life history trait effects on the mismatch between niches and geographic distributions. a, Effect of dispersal ability on range filling (points: posterior means, bars: posterior standard deviations). The line shows the estimated linear regression (posterior means, 90% credibility interval as shaded areas, slope = 0.55, p = 0.022). **b,** Relationship between demographic suitability (predicted r_0) and occupancy within the range. Points show the mean occupancy in sites that were binned according to deciles of predicted r_0 (i.e. ten points per species). The lines show average predictions of this relationship for species with different persistence ability (posterior means, 90% credibility interval as shaded areas). **c,** Variation in species' mean occupancy of sites within their ranges that are predicted to be unsuitable ($r_0 > 0$) resp. suitable ($r_0 > 0$) among species with different persistence ability.

for an average of 80% populations per species (Extended Data Table 4). This quasi-experimental evidence suggests that the demographic niche models capture key factors limiting the geographic distributions of our study species.

On small spatial scales, niche-distribution mismatches can arise in particular from source-sink effects and from time-delayed extinction[3-5]. To assess the match between demographic suitability and a species' occurrence within its geographic range, we regressed spatial variation in occupancy (on a 1′ grid, c. 1.55 × 1.85 km²) against the locally predicted r_0. Occupancy generally increased with predicted r_0 (Fig. 3b), but the strength of this relationship varied considerably between species (R^2_N: 0.01–0.55, mean = 0.18; Extended Data Table 2) and variation in occupancy was better explained for nonsprouters than for resprouters (ANOVA, $F_{1,24}$ = 4.74, p = 0.039). For resprouters, small-scale niche-distribution mismatches are greater mainly because populations more frequently occur in unsuitable sites (Fig. 3c). This can be explained by populations of more persistent species being less vulnerable to adverse conditions in both temporally fluctuating[27] and directionally changing environments[28]. Dispersal ability had no positive effect on the occupancy of unsuitable sites. Hence, we found no indication of source-sink effects at the spatial scale of our analysis.

In summary, our comparative analysis of demography-based niches indicates that key life history traits shape the geographic distributions of plant species not only because of a relationship between those traits and niche sizes but also by trait effects on niche-distribution mismatches. Specifically, mismatches between niches and geographic distributions arise because poorly dispersed species are absent from suitable sites beyond their range limits and because species with high persistence ability are present in sites that are unsuitable under current, average environmental conditions. Importantly, this identifies poorly dispersing and highly persistent species as cases where static, correlative species distribution models are more likely to fail. For such species, range forecasts require dynamic species distribution models that incorporate demographic niche estimates[7,10,11]. From a theoretical perspective, the quantification of spatial variation in species' intrinsic population growth rates is also the first step towards understanding large-scale species coexistence and effects of biotic interactions on range dynamics[29,30]. A demographic quantification of ecological niches, particularly in well-studied model systems, thus holds great promise for better integrating ecological theory and empirical biogeography. This is urgently needed to advance our predictive understanding of biodiversity dynamics under environmental change[6,8].

Methods

Study species and demographic data. We studied 26 species of the Proteaceae family, specifically of the genera *Protea* (16 species) and *Leucadendron* (10 species), that are endemic to the Cape Floristic Region[17]. These species were chosen to represent variation in geographic distributions as well as variation in dispersal and resprouting ability. For each species we obtained data on between-population variation in key demographic rates across the entire life cycle, namely the total fecundity of adult plants since the last fire (size of individual canopy seed banks), per capita post-fire seedling recruitment (ratio between post-fire recruits and pre-fire adults) and adult fire survival. The latter two rates were measured on recently burned sites (<3 years after fire), where burned pre-fire adults were still identifiable[31,32]. Study sites for demographic sampling were selected to cover major environmental gradients across the global geographic distribution of each study species. The final data set comprised 3,617 population-level records from an average of 99 (median = 85) study sites per species (Extended Data Table 1). For details on the demographic data collection see ref[19].

Study region and environmental variables. Our study area was defined on a regular grid with a spatial resolution of 1′ × 1′ (c. 1.55 km × 1.85 km) and included all grid cells of the Cape Floristic Region in which > 5% of the area is covered by Fynbos vegetation[33]. Climatic and edaphic variables that are expected to be main determinants of the performance and survival of serotinous Proteaceae were extracted from the South African Atlas of Climatology and Agrohydrology[34]. We included *January maximum daily temperature* (T_{max}), *July minimum daily temperature* (T_{min}) and a *January aridity index* (AI) calculated as the ratio between the mean values of precipitation (P) and temperature (T): AI = P/(T + 10°C) [35]. Climatic variables are averages over the years 1950–2000. As an edaphic variable we used a *soil fertility* index[34] that combines soil texture and base status and ranges from 0 to 10. Information on the fire return interval was obtained from both observational records and model predictions. For the demographic sampling sites, information on the fire history (time since the last fire and length of the previous fire interval) was inferred from a combination of measured plant ages, historical records[36,37] and MODIS satellite observations[38] (see ref[19] for details). For predictions of

population growth rates across the study region (see below), we used probability distributions of fire return intervals predicted from a climate-driven model of post-fire ecosystem recovery[20].

Demographic response model. We used a hierarchical Bayesian modelling approach for estimating the species-specific responses of key demographic rates (fecundity, per-seed establishment and adult fire survival) to environmental covariates (Extended Data Fig. 3). The model considers effects of climatic and edaphic conditions, variable fire return intervals and intraspecific density dependence at both the adult and the seedling stage. Below, we describe the submodels for variation in each demographic rate.

Fecundity. The recorded size of the canopy seed bank ($Seed.count_{i,j}$) of plant j in population i is described by an overdispersed Poisson distribution

$$Seed.count_{i,j} \sim \text{Poisson}(Fec_i)$$

$$Fec_i \sim \text{Gamma}\left(\frac{\mu.fec_i}{k.fec}, k.fec\right)$$

where the expected value of mean fecundity $\mu.fec_i$ is determined by limiting effects of post-fire stand age (*Age*), environmental covariates (**X**) and population density(*D*):

$$\mu.fec_i = max.fec \cdot f(Age_i) \cdot g(\mathbf{X}_i) \cdot h(D_i)$$

Effects of stand age on fecundity arise from the time of maturation until the first flowering and cone production, increasing accumulation of standing cones on growing plants, cone loss and possibly senescence of aged individuals:

$$f(Age_i) = M_i \cdot \exp(\beta.fec_1 \cdot Age_i + \beta.fec_2 \cdot (Age_i)^2)$$

where M_i is a binary random variable (0, 1) indicating maturity. The probability of population-level maturity is calculated from a Weibull distribution for the age (*t.mat*) of first cone production:

$$M_i \sim \text{Bernoulli}(p.mat_i)$$

$$p.mat_i = Pr(t.mat < Age_i)$$

$$t.mat \sim \text{Weibull}(sh.mat, sc.mat)$$

The species-specific time to reproductive maturity (*t.mat*) was constrained to be at least three years for nonsprouters[17]. The effects of the environmental covariates $k = 1 \ldots K$ are described by Gaussian demographic response functions:

$$g(X_i) = exp\left(\sum_{k=1}^{K} \frac{-(X_{i,k} - opt.fec_k)^2}{2 \cdot sig.fec_k^2}\right)$$

where $opt.fec_k$ denotes the optimal conditions and $sig.fec_k$ measures the width of the response curve. Effects of population density D_i on fecundity are described as[39]:

$$h(D_i) = \exp(-\gamma.fec \cdot D_i)$$

Establishment. The establishment of new recruits from seeds is modelled as a binomial process where the number of recruits ($\#Recruits_i$) in population i depends on the total number of available seeds ($\#Seeds_i$) in the canopy seed bank at the time of the last fire and the per-seed establishment rate $\pi.est_i$:

$$\#Recruits_i \sim \text{Binomial}(Seeds_i, \pi.est_i)$$

Since $\#Seeds_i$ is unknown for recently burned sites where recruitment was recorded, it is modelled as a latent state variable

$$\#Seeds_i \sim \text{Poisson}(\#Parents_i \cdot Fec_i)$$

where $\#Parents_i$ denotes the number of pre-fire seed sources (only females for dioecious *Leucadendron* species) and Fec_i depends on environmental covariates ($\mathbf{X}_i$), the post-fire stand age (Age_i) and the adult population density (D_i) at the time of the previous fire as described in the fecundity submodel. Establishment rate $\pi.est_i$ is affected by environmental covariates ($\mathbf{X}_i$) and by the densities of seeds $SD_i = \#Seeds_i/Area_i$ and fire-surviving adults $AD_i = \#Adults_i/Area_i$.

$$\pi.est_i = max.est \cdot g(\mathbf{X}_i) \cdot h(SD_i, AD_i)$$

As for fecundity, the effects of different environmental covariates $k = 1 \ldots K$ are described by Gaussian demographic response functions:

$$g(\mathbf{X}_i) = exp\left(\sum_{k=1}^{K} \frac{-\left(X_{i,k} - opt.est_k\right)^2}{2 \cdot sig.est_k^2}\right)$$

Density effects on establishment result from the density of seeds (SD_i) as well of from the density of fire-surviving adults (AD_i), with different strengths ($\gamma.est.SD$ resp. $\gamma.est.AD$) for each of these density effects:

$$h(SD_i, AD_i) = \frac{1}{1 + c_i(\gamma.est.SD \cdot SD_i + \gamma.est.AD \cdot AD_i)}$$

Density-dependent mortality of recruits (self-thinning) is a continuous process, which for Fynbos Proteaceae generally occurs within the first three years after a fire[40]. This is described by weighting the density effects with a factor c_i that depends on the post-fire stand age ($pf.Age_i$) at the time of sampling:

$$c_i = \begin{cases} \left(\frac{pf.Age_i}{3}\right)^{\beta.est} & if\ pf.Age_i < 3 \\ 1 & if\ pf.Age_i > 3 \end{cases}$$

Thereby the model accounts for the fact that more seedlings can be observed if a site is surveyed just shortly after germination[19].

Survival. Adult fire survival is modelled as a binomial process for the proportion of survivors among all pre-fire adults ($\#All.Adults_i$):.

$$\#Survivors_i \sim \text{Binomial}(\#All.Adults_i, \pi.surv_i)$$

$$\text{logit}(\pi.surv_i) = \mu.surv + g(\mathbf{X}_i, Age_i) + h(D_i) + \varepsilon.surv_i$$

Similar as for fecundity, the effects of different environmental covariates $k = 1 \ldots K$ and of post-fire stand age (Age_i) are described by Gaussian response functions and effects of population density D_i by a negative-exponential function:

$$f(Age_i) = exp\left(\frac{-\left(Age_i - opt.surv_{Age}\right)^2}{2 \cdot sig.surv_{Age}^2}\right)$$

$$g(\mathbf{X}_i) = exp\left(\sum_{k=1}^{K} \frac{-\left(X_{i,k} - opt.surv_k\right)^2}{2 \cdot sig.surv_{\mathrm{k}}^2}\right)$$

$$h(D_i) = \exp(-\gamma.surv \cdot D_i)$$

Since adult fire survival rates of nonsprouters are generally low with little intraspecific variation[19], we modelled them as species-specific constants and considered effects of covariates only for the survival rates of resprouters.

Bayesian parameter estimation. Parameters of the model were estimated independently for each study species. All environmental variables were scaled and centred and the *aridity index* and *soil fertility index* were additionally log-transformed before the analyses. The hierarchical model was formulated in a Bayesian framework and samples from the parameter posterior distribution were generated with Markov chain Monte Carlo (MCMC) methods in the software JAGS[41] (Supplementary Information). An overview of parameter prior distributions is given in Extended Data Table 5. In three independent MCMC chains, posteriors were sampled from 100,000 iterations after a burn-in period of 500,000 iterations. Convergence of the MCMC sampler was checked by the multivariate scale reduction factor[42] being smaller than 1.1. For all further analyses, the posterior samples were regularly thinned to a sample size of 1,000 for each chain, resp. 3,000 samples in total.

Model evaluation. For each species we assessed the model fit separately for each observed demographic variable (fecundity, recruit:parent ratio, adult fire survival) by calculating Nagelkerke's[43] general R_N^2 relative to null models in which demographic rates (*π.est*, *π.surv*, *μ.fec*) are species-specific constants. The explained variance in each demographic variable for each study species is shown in Extended Data Table 2.

Prediction of niche sizes and geographic variation in r_0. The demographic response model predicts fire survival rate *π.surv*, fecundity *μ.fec* and establishment rate *π.est* as functions of environmental covariates **X**, fire interval *T* and population density at different stages. Based on these demographic rates the expected population size *N* after a fire interval of length *T* can be calculated as the sum of fire survivors and new recruits:

$$N_{t+T} = N_t \cdot \pi.surv(\mathbf{X}, T, D) + N_t \cdot p.fem \cdot \mu.fec(\mathbf{X}, T, D) \cdot \pi.est(\mathbf{X}, SD, AD)$$

For the dioecious *Leucadendron* species the parameter *p.fem* specifies the proportion of female individuals in a population and accounts for the fact that fecundity rate *μ.fec* was defined per female. The niche of a species is defined as the set of environmental conditions for which the intrinsic growth rate of small populations (r_0) is positive. To calculate r_0 we set all density variables to zero and first calculated the rate of change in population size per fire interval

$$\lambda_0(\mathbf{X}, T) = \frac{N_{t+T}}{N_t} = \pi.surv(\mathbf{X}, T) + \mu.fec(\mathbf{X}, T) \cdot \pi.est(\mathbf{X}) \cdot p.fem$$

The intrinsic growth rate r_0 was then calculated on an annual basis as

$$r_0(\mathbf{X}, T) = \frac{\log[\lambda_0(\mathbf{X}, T)]}{T}$$

To quantify species' niches, r_0 was predicted on a 5-dimensional grid spanned by niche axes according to the four climatic-edaphic covariates (T_{max}, T_{min}, AI, soil fertility) and the fire return interval (log-transformed). Since the demographic data analysed in the demographic response models were measured in natural communities and thus incorporate effects of interspecific biotic interactions, the predicted r_0 represents the post-interactive[1] (or 'realised') niche. For commensurability of the different niche axes, these were confined and scaled to the respective range of values that occur throughout the Fynbos biome for each variable (so that each axis ranges from zero to one) and each niche axes was regularly sampled with a resolution of 0.01 (10^{10} grid points). The niche size was then quantified separately for fire return interval ('disturbance niche size') and for the four climatic-edaphic variables ('environmental niche size'). The disturbance niche size was determined as the range of fire return intervals for which a positive r_0 is predicted when the climatic-edaphic variables are set to their optimal values. Likewise, the environmental niche size was determined as the hypervolume in the 4-

dimensional climatic-edaphic sub-space for which the predicted r_0 is positive when setting the fire return interval to its species-specific optimum.

To geographically project r_0 across the study region, r_0 was not predicted for a single fixed fire return interval but integrated as a weighted geometric mean over the site-specific probability distribution of fire return intervals[20]. In all cases, values of r_0 were predicted from each posterior sample of the demographic response model, yielding full posterior distributions for each predicted value.

Species distribution data and geographic ranges. The Protea Atlas Project was an extensive citizen science project that used a standardized protocol to collected complete species lists from Proteaceae communities across the Cape Floristic Region[17]. For our study region, the Protea Atlas data base contains 54,642 sampling locations with a total of 126,690 recorded presences of the 26 study species (Extended Data Table 1). We aggregated these occurrence data to the proportion of presence records (occupancy) among the sampled communities in each $1' \times 1'$ grid cell. As an overall assessment of how well species occurrence is predicted by demographic suitability, we calculated the area under the receiver operating characteristic curve (AUC)[44] for a binary classification of grid cell presence-absence by the predicted r_0. We furthermore described the large-scale geographic range of each species by an alpha-convex hull[45] over all presence records, using the *alphahull* package[46] in R with parameter $\alpha = 0.5$. The size of each species' geographic range was calculated as the overlap of this alpha-convex hull with the Fynbos biome (study region). The degree of range filling was calculated as the proportion of potentially suitable area (cells with predicted $r_0>0$) that lies within the geographic range. This calculation was conducted for each posterior sample of predicted r_0 in order to account for uncertainty in estimated range filling.

Statistical analyses of relationships between life history traits, niche characteristics and species geographic ranges. For the two measures of niche size (disturbance niche size and environmental niche size) as well as for species' geographic range size and range filling, we performed separate regression analyses to test for effects of key life-history traits. The traits considered as explanatory

variables were persistence ability (0 = nonsprouter, 1 = resprouter) and long-distance dispersal ability. The species-specific relative long-distance dispersal ability was derived from a trait-based mechanistic model of primary and secondary wind-dispersal[21] and measured as the number of neighbouring cells on a 1′ × 1′ rectangular grid that can be reached by dispersal from a source cell with a probability of at least 10^{-4}. This dispersal measure was then log-transformed and scaled.

In each regression model we accounted for both measurements errors (posterior variance of the respective response variable) and phylogenetic dependence[47]. Hence, the likelihood of the estimated species-specific posterior means (**Y.mean**) of the response variable depends on the vector of its true (but unknown) values **Y** and the respective posterior variances (**Y.var**):

$$\mathbf{Y.mean} \mid \mathbf{Y}, \mathbf{Y.var} \sim \mathrm{Normal}(\mathbf{Y}, \mathbf{Y.var})$$

For the true **Y** we then formulated a multivariate normal model

$$\mathbf{Y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma, \boldsymbol{\Sigma}, \lambda \sim \mathrm{MVN}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{\Sigma}_\lambda)$$

where $\mathbf{X}\boldsymbol{\beta}$ is the linear predictor of the regression and $\boldsymbol{\Sigma}_\lambda$ is an adjusted variance-covariance matrix to account for phylogenetic dependencies. We first calculated the variance-covariance matrix $\boldsymbol{\Sigma}$ under a Brownian motion model (using R package *ape*[48]). A molecular phylogeny of our study species (Extended Data Fig. 4) was obtained by pruning a phylogenetic tree of 291 Proteaceae species. This phylogenetic tree was constructed from a supermatrix combining molecular markers for *Leucadendron*[49], *Protea*[50] and related species of the Proteaceae family[51]. As quantitative measure of the degree of phylogenetic dependence the model furthermore includes Pagel's λ (ranging from zero to one)[52] and the adjusted $\boldsymbol{\Sigma}_\lambda$ is calculated as

$$\boldsymbol{\Sigma}_\lambda = \lambda\boldsymbol{\Sigma} + (1 - \lambda)\mathbf{I}$$

where **I** is the identity matrix.

Bayesian parameter estimations were performed in JAGS[41] using largely uninformative prior distributions (Supplementary Information). For each model we ran three independent MCMC chains with 200,000 iterations, the first half of which was discarded as burn-in and convergence was checked

by the multivariate scale reduction factor[42] being smaller than 1.1. For each response variable, we estimated a full model that included effects of persistence ability, dispersal ability and their interaction. All simplified models nested in the full model were then compared by the deviance information criteria[53] (DIC) and we report parameter estimates for the DIC-minimal models (Extended Data Table 3).

Statistical analysis of the relationship between occupancy and demographic suitability. We used a binomial non-linear regression model to analyse the relationship between each species' occupancy in the 1′ × 1′ grid cells within its geographic range and the respectively predicted intrinsic population growth rate r_0. The model describes the number of presence records ($y_{s,i}$) for species s in the grid cell i as

$$y_{s,i} \sim \text{Binomial}(\psi_{s,i}, N_i)$$

$$\psi_{s,i} = a_s / (1 + \exp(-b_s(r0_{s,i} - c_s)))$$

where N_i is the number of Protea Atlas sampling locations within the grid cell and a_s, b_s, c_s are species-specific regression parameters. Bayesian parameter estimation was performed in JAGS[41] using largely uninformative prior distributions (Supplementary Information). We ran three independent MCMC chains with 20,000 iterations, the first half of which was discarded as burn-in and convergence was checked by the multivariate scale reduction factor[42] being smaller than 1.1.

Data and code availability Phylogenetic data is available from the TreeBASE repository (http://purl.org/phylo/treebase/phylows/study/TB2:S23259). Demographic data were partly used under license from CapeNature for the current study, and so are not publicly available. These data are however available from the authors upon reasonable request and with permission of CapeNature. Documented JAGS code for the statistical analyses is provided in Supplementary information. Additionally used R code and data generated during the analyses is available from the corresponding author upon request.

References:

1. Hutchinson, G. E. *An Introduction to Population Ecology* (Yale University Press, 1978).
2. Maguire, B. Niche response structure and analytical potentials of its relationship to habitat. *Am. Nat.* **107**, 213–246 (1973).
3. Pulliam, H. R. On the relationship between niche and distribution. *Ecol. Lett.* **3**, 349–361 (2000).
4. Holt, R. D. Bringing the Hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19659–19665 (2009).
5. Schurr, F. M. *et al.* How to understand species' niches and range dynamics: a demographic research agenda for biogeography. *J. Biogeogr* **39**, 2146–2162 (2012).
6. Ehrlén, J. & Morris, W. F. Predicting changes in the distribution and abundance of species under environmental change. *Ecol. Lett.* **18**, 303–314 (2015).
7. Evans, M. E. K., Merow, C., Record, S., McMahon, S. M. & Enquist, B. J. Towards process-based range modeling of many species. *Trends Ecol. Evol.* **31**, 860–871 (2016).
8. Urban, M.C. *et al.* Improving the forecast for biodiversity under climate change. *Science* **353**, aad8466 (2016).
9. Hutchinson, G. E. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* **22**, 415–427 (1957).
10. Pagel, J. & Schurr, F. M. Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Glob. Ecol. Biogeogr.* **21**, 293–304 (2012).
11. Zurell, D. *et al.* Benchmarking novel approaches for modelling species range dynamics. *Glob. Change Biol*. **22**, 2651–2664 (2016).
12. Settele, J. *et al.* Terrestrial and inland water systems. *Climate Change 2014 Impacts, Adaptation and Vulnerability: Part A: Global and Sectoral Aspects*, 271–359 (Cambridge University Press, 2015).
13. Diez, J. M., Giladi, I., Warren, R. & Pulliam, H. R. Probabilistic and spatially variable niches inferred from demography. *J. Ecol.* **102**, 544–554 (2014).

14. Merow, C. *et al.* On using integral projection models to generate demographically driven predictions of species' distributions: development and validation using sparse data. *Ecography* **37**, 1167–1183 (2014).

15. Pironon, S. *et al.* The 'Hutchinsonian niche' as an assemblage of demographic niches: implications for species geographic ranges. *Ecography* **41**, 1103–1113 (2018).

16. Myers, N., Mittermeier, R.A., Mittermeier, C.G., Fonseca, G.A.B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–859 (2000).

17. Rebelo, A. G. *Proteas: A Field Guide to the Proteas of Southern Africa* (Fernwood Press, Vlaeberg, South Africa, 2001).

18. Bond, W. J. & Van Wilgen, B. W. *Fire and Plants*. *Population and Community Biology Series 14* (Chapman & Hall, 1996).

19. Treurnicht, M. *et al.* Environmental drivers of demographic variation across the global geographical range of 26 plant species. *J. Ecol.* **104**, 331–342 (2016).

20. Wilson, A. M., Latimer, A. M. & Silander, J. A. Climatic controls on ecosystem resilience: postfire regeneration in the Cape Floristic Region of South Africa. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 9058–9063 (2015).

21. Schurr, F. M. *et al.* Colonization and persistence ability explain the extent to which plant species fill their potential range. *Glob. Ecol. Biogeogr.* **16**, 449–459 (2007).

22. Thompson, K., Gaston, K. J. & Band, S. R. Range size, dispersal and niche breadth in the herbaceous flora of central England. *J. Ecol.* **87**, 150–155 (1999).

23. Lester, S. E., Ruttenberg, B. I., Gaines, S. D. & Kinlan, B. P. The relationship between dispersal ability and geographic range size. *Ecol. Lett.* **10**, 745–758 (2007).

24. Gaston, K. J. *The Structure and Dynamics of Geographic Ranges* (Oxford University Press, 2003).

25. Hargreaves, A. L., Karen, E. S. & Eckert, C. G. Are species' range limits simply niche limits writ large? A review of transplant experiments beyond the range. *Am. Nat.* **183**, 157–173 (2013).

26. Latimer, A.M., Silander, J.A., Rebelo, A.G. & Midgley, G.F. Experimental biogeography: the role of environmental gradients in high geographic diversity in Cape Proteaceae. *Oecologia* **160**, 151–162 (2009).

27. Higgins, S. I., Pickett, S. T. & Bond, W. J. Predicting extinction risks for plants: environmental stochasticity can save declining populations. *Trends Ecol. Evol.* **15**, 516–520 (2000).

28. Enright, N. J., Fontaine, J. B., Lamont, B. B., Miller, B. P. & Westcott, V. C. Resistance and resilience to changing climate and fire regime depend on plant functional traits. *J. Ecol.* **102**, 1572–1581 (2014).

29. Godsoe, W., Jankowski, J., Holt, R. D. & Gravel, D. Integrating biogeography with contemporary niche theory. *Trends Ecol. Evol.* **32**, 488–499 (2017).

30. Usinowicz, J. & Levine, J. M. Species persistence under climate change: a geographical scale coexistence problem. *Ecol. Lett.* **21**, 1589–1603 (2018).

31. Bond, W. J., Vlok, J. & Viviers, M. Variation in seedling recruitment of Cape Proteaceae after fire. *J. Ecol.* **72**, 209–221 (1984).

32. Bond, W. J., Maze, K. & Desmet, P. Fire life histories and the seeds of chaos. *Écoscience* **2**, 252–260 (1995).

33. South African National Biodiversity Institute. *Vegetation Map of South Africa, Lesotho and Swaziland 2012* (http://bgis.sanbi.org/SpatialDataset/Detail/18, 2012) [vector geospatial dataset, downloaded on 11 August 2017]

34. Schulze, R. E. *South African Atlas of Climatology and Agrohydrology, Technical Report 1489/1/06* (Water Research Commission, Pretoria, South Africa, 2007).

35. De Martonne, E. Aréisme et indice artidite. *C. R. Acad. Sci.* **182**, 1395–1398 (1926).

36. De Klerk, H. A pragmatic assessment of the usefulness of the MODIS (Terra and Aqua) 1-km active fire (MOD14A2 and MYD14A2) products for mapping fires in the fynbos biome. *Int. J. Wildland Fire* **17**, 166–178 (2008).

37. Kraaij, T., Cowling, R. M., Van Wilgen, B. W. & Schutte-Vlok, A. Historical fire-regimes in a poorly understood, fire-prone ecosystem: eastern coastal fynbos. *Int. J. Wildland Fire* **22**, 277–287 (2013).

38. Roy, D. P., Boschetti, L., Justice, C. O. & Ju, J. The collection 5 MODIS burned area product – Global evaluation by comparison with the MODIS active fire product. *Remote Sens. Environ.* **112**, 3690–3707 (2008).

39. Nottebrock, H. *et al.* Coexistence of plant species in a biodiversity hotspot is stabilized by competition but not by seed predation. *Oikos* **126**, 276–284 (2017).

40. Manders, P. & Smith, R. Effects of artificially established depth to water table gradients and soil type on the growth of Cape fynbos and forest plants. *S. Afr. J. Bot.* **58**, 195–201 (1992).

41. Plummer, M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing* (2003).

42. Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7**, 457–472 (1992).

43. Nagelkerke, N. J. A note on a general definition of the coefficient of determination. *Biometrika* **78**, 691–692 (1991).

44. Hanley, J. A. & McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29–36 (1982).

45. Burgman, M.A. & Fox, J.C. Bias in species range estimates from minimum convex polygons: implications for conservation and options for improved planning. *Anim. Conserv.* **6**, 19–28 (2003).

46. Pateiro-Lopez, B. & Rodriguez-Casal, A. Generalizing the convex hull of a sample: the R package alphahull. *J. Stat. Softw.* **34**, 1–28 (2010).

47. de Villemereuil, P., Wells, J. A., Edwards, R. D. & Blomberg, S. P. Bayesian models for comparative analysis integrating phylogenetic uncertainty. *BMC Evol. Biol.* **12**, 102 (2012).

48. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).

49. Tonnabel, J. *et al.* Convergent and correlated evolution of major life-history traits in the angiosperm genus *Leucadendron* (Proteaceae). *Evolution* **68**, 2775–2792 (2014).

50. Valente, L. M. *et al.* Diversification of the African genus Protea (Proteaceae) in the Cape biodiversity hotspot and beyond: equal rates in different biomes. *Evolution* **64**: 745–760 (2010).

51. Sauquet, H. *et al.* Contrasted patterns of hyperdiversification in Mediterranean hotspots. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 221–225 (2009).

52. Pagel, M. Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884 (1999).

53. Gelman, A., Hwang, J. & Vehtari, A. Understanding predictive information criteria for Bayesian models. *Stat. Comput.* **24**, 997–1016 (2014).

Acknowledgements This work was primarily funded by the German Research Foundation (DFG, grant SCHU 2259/5-1). M.T. acknowledges additional funding from Stellenbosch University and the DST-NRF's Professional Development Programme (South Africa). We are grateful to CapeNature, SANParks and R.M. Cowling for access to demographic data and to B. Olivier, all field assistants, reserve managers and private landholders who supported our field work. Data were collected under the CapeNature permit AAA0028-AAA005-00213, Eastern Cape Parks permit CRO91/12CR and SANParks permit (Agulhas National Park). Many thanks to C. Buchmann, H. Cooksley, C.S. Sheppard and J. Walter for commenting on earlier drafts. We also thank Jake Alexander for helpful discussion and comments on a previous version of the manuscript.

Author Contributions J.P. and F.M.S. conceived and designed the study. M.T. collected the data with contributions from W.J.B., T.K., H.N., AL.S.-V. and F.M.S.. J.T. provided phylogenetic analyses. J.P. analysed the data and wrote the first draft. All co-authors contributed to the writing of the final manuscript.