Journal of Biogeography

Supporting Information

Environmental heterogeneity explains contrasting plant species richness between the South African Cape and southwestern Australia

Ruan van Mazijk, Michael D. Cramer and G. Anthony Verboom

Species occurrence data cleaning

For the purposes of this study, the GCFR was treated as the area occupied by the Succulent Karoo and Fynbos Biomes (Mucina & Rutherford, 2006), while the SWAFR was treated as the area occupied by Southwest Australia Savanna, Swan Coastal Plain Scrub and Woodlands, Jarrah-Karri Forest and Shrublands, Southwest Australia Woodlands, Esperance Mallee, and Coolgardie Woodlands (Olson et al., 2001) in order to match the current delimitation of the SWAFR (Gioia & Hopper, 2017; Hopper & Gioia, 2004).

To compare vascular plant species richness between the GCFR and SWAFR, geospatially explicit occurrence records of tracheophytes from within the borders of each region were obtained from the Global Biodiversity Information Facility (GBIF; Table S1). We retained only records identified to the species level and ignored intraspecific taxa. This resulted in the retention of 14,147 and 8,912 unique species names for the GCFR and SWAFR, respectively. The R package "taxize" (Chamberlain et al., 2016) was then used to query each species name against two major taxonomic databases, the Global Name Resolver (GNR) and the Taxonomic Name Resolution Service (TNRS; Boyle et al., 2013). Where either or both databases returned a match for a name, the name was retained; where not, it was excluded. Although the number of species thus excluded is high (GCFR: 692; SWAFR: 1,171), the geographically random distribution of the records associated with these names suggests that exclusion of these names will not significantly influence spatial patterns of species richness.

In order to ensure that no species were listed under multiple synonyms, the retained names were then queried against the Tropicos and Integrated Taxonomic Information System (ITIS) for known synonyms, again using "taxize". We then removed all records of species identified as non-native, using lists of invasive plants for South Africa and Australia from the IUCN's Global Invasive Species Database (http://www.iucngisd.org/gisd/).

Finally, we removed species with fewer than five total collection records in total, in order to exclude collections with potentially low-confidence identifications. This, and the exclusion of occurrence data originating from coastal pixels at the 0.05° resolution, brought the total number of species in each region down to 9,419 and 6,696 in the GCFR and SWAFR respectively.

This is the supporting information for the accepted version of the following article: Van Mazijk, R., Cramer, M.D. and Verboom, G.A. (2021), Environmental heterogeneity explains contrasting plant species richness between the South African Cape and southwestern Australia. J Biogeogr, 48: 1875-1888, which has been published in final form at https://doi.org/10.1111/jbi.14118.

Selecting environmental variables

To compare environmental heterogeneity between the GCFR and SWAFR, we acquired a suite of nine geospatially-explicit environmental variables (Table S1), selected to represent axes which are considered regionally important and independent. For example, the inclusion of PDQ in addition to MAP is justified on the basis that, where the latter captures variation in overall rainfall amount, the former measures the intensity of seasonal aridity, a key feature of mediterranean-type climates (Belda et al., 2014). Variable selection was, however, constrained by the availability of suitable raster-layers. Thus, although soil phosphorus concentration ($[P]_{soil}$) is probably an important determinant of plant distribution in both the GCFR and SWAFR (Lambers et al., 2006, 2010; Shane et al., 2008), this variable could not be included owing to a lack of comparable data layers for the two regions. Indeed, wherever possible, we made use of remote sensing derived layers.

Quantifying environmental heterogeneity

Our index of heterogeneity, based on raster data, employs nested squares at various spatial scales. We quantified the environmental heterogeneity of a given square as the variance of the environmental conditions of its four sub-squares. This was done both using the "aggregate" function in the R package "raster" (Hijmans, 2016), with variance set as the aggregation function, and "tidyverse" packages (Wickham et al., 2019). Regarding the latter, data at the QDS-, HDS- and DS-scales were analysed in data-frames labelled with grid-cell codes (sensu Larsen et al., 2009; downloaded from https://mindland.com/wp/projects/quarter-degree-grid-cells/download-qdgc/ [Accessed 27 February, 2020]), to ensure the heterogeneity and richness data were derived from the same grid-cells.

Tables

Table S1: Georeferenced environmental data[1] and vascular plant species occurrence data sources used in this study. Data were acquired for the GCFR and SWAFR, with the temporal extent of data products used described where applicable.

Dataset(s)	Source	Temporal extent	Citation(s)
Plant species occurrences	GBIF		GBIF (2017a,b)
Elevation	SRTM (v2.0)		Farr et al. (2007)
NDVI, Surface T	MODIS (v006)	Feb. 2000 to Apr. 2017	NASA (2017a,b)
MAP, PDQ	CHIRPS (v2.0)	Jan. 1981 to Feb. 2017	Funk et al. (2015)
CEC, clay, soil C, pH	SoilGrids250m		Hengl et al. (2017)

[1] Abbreviations are as follows: NDVI, normalized difference vegetation index; T, temperature; MAP, mean annual precipitation; PDQ, precipitation in the driest quarter; CEC, cation exchange capacity; C, carbon.

Table S2 (next page): Analyses of variances (ANOVAs) of the three multiple regression models of vascular plant species richness, across the GCFR and SWAFR, including species richness hotspots (Figure 5). The variables in each model are arranged in descending order according to their proportion of variance explained. The significance[1] (where applicable) of each variable's contribution to each model is also shown. Abbreviations are as in Table S1.

[1] Represented as follows: ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$; ~, $P < 0.1$; blank, NS; –, NA

Response	Term	Variance explained	
(a) S_{QDS}	(Residuals)	0.72	-
$R^2_{adj.}$ = 0.27	MAP	0.13	***
	Elevation	0.06	***
	NDVI	0.03	***
	PDQ × Region	0.02	***
	Region	0.02	***
	NDVI × Region	0.01	***
	Clay	0.01	**
	pH	0.01	**
	CEC	0.01	*
	pH × Region	0.01	*
	PDQ	< 0.01	
	MAP × Region	< 0.01	
(b) S_{HDS}	(Residuals)	0.54	-
$R^2_{adj.}$ = 0.43	MAP	0.21	***
	Elevation	0.1	***
	pH × Region	0.04	***
	NDVI	0.02	**
	MAP × Region	0.02	**
	pH	0.02	*
	Clay	0.02	*
	PDQ	0.02	*
	Surface T × Region	0.01	*
	Surface T	0.01	
	Soil C × Region	< 0.01	
	Region	< 0.01	
	Soil C	< 0.01	
(c) S_{DS}	NDVI	0.17	***
$R^2_{adj.}$ = 0.85	Elevation	0.15	***
	PDQ	0.14	***
	pH × Region	0.11	***
	Clay	0.07	***
	(Residuals)	0.07	-
	NDVI × Region	0.06	**
	Clay × Region	0.06	**
	Soil C × Region	0.04	**
	Surface T	0.04	**
	CEC	0.03	*
	Soil C	0.02	~
	Elevation × Region	0.02	~
	PDQ × Region	0.01	
	pH	0.01	
	Surface T × Region	< 0.01	
	Region	< 0.01	
	CEC × Region	< 0.01	

Table S3: Comparisons of the standard deviations (*SD*) of residuals from PC1-based and multivariate (MV) models using datasets both including excluding vascular plant species richness hotspots across the GCFR and SWAFR, across the three spatial scales. Hotspots excluded from each model were those with residuals greater than two standard deviations from the mean for that model. All pairs of GCFR and SWAFR *SD*-values differed significantly ($P < 0.01$; two-sided *F*-tests).

		SD of model residuals			
		Including hotspots		Excluding hotspots	
Scale	Region	PC1	MV	PC1	MV
(a) QDS	GCFR	343.46	312.89	234.93	217.56
	SWAFR	245.77	223.05	203.91	174.97
(b) HDS	GCFR	638.62	519.19	460.39	360.47
	SWAFR	334.15	290.90	326.06	273.31
(c) DS	GCFR	811.89	6.22	588.58	*NA*
	SWAFR	311.80	226.50	297.22	*NA*

Figures

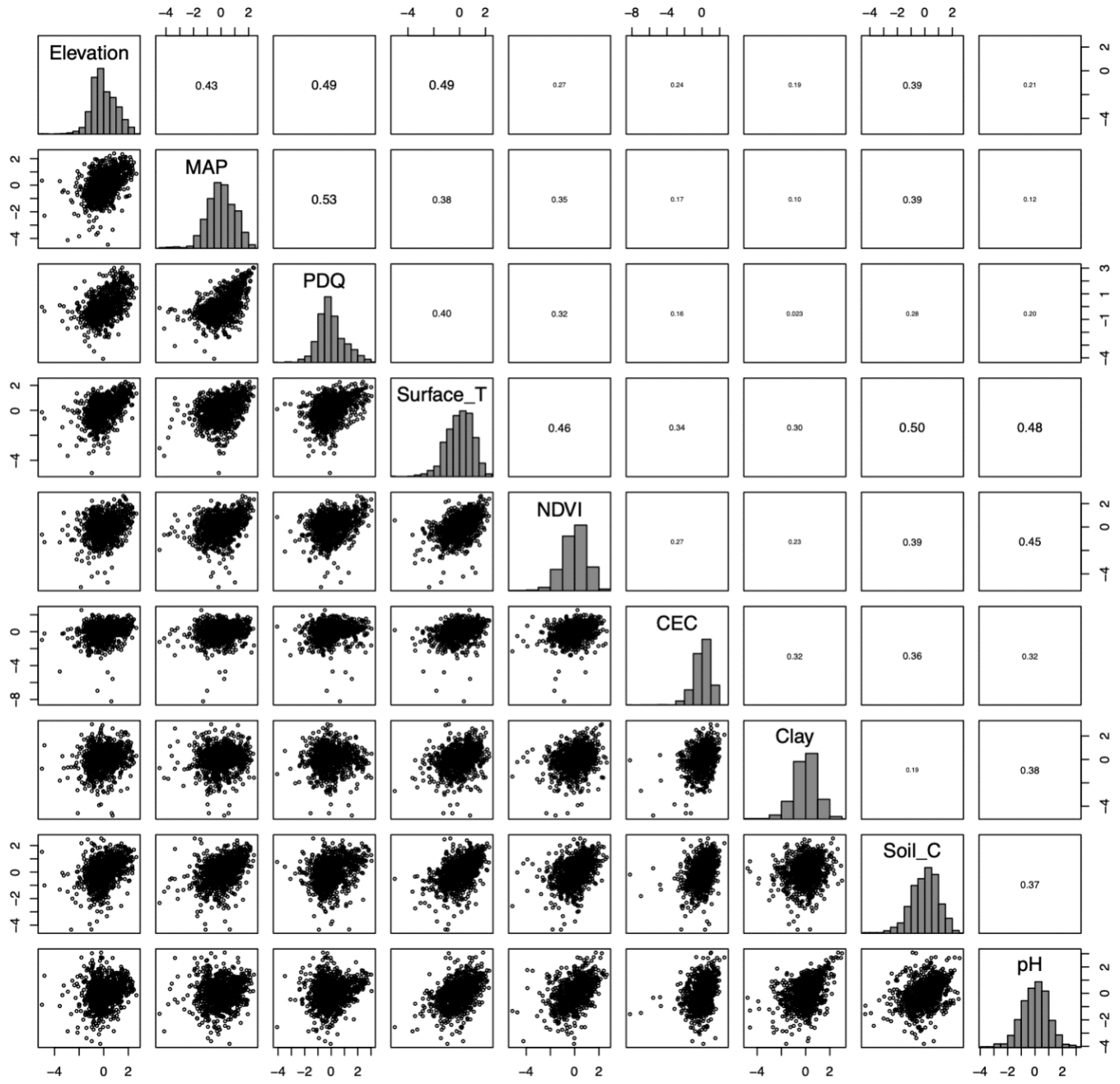
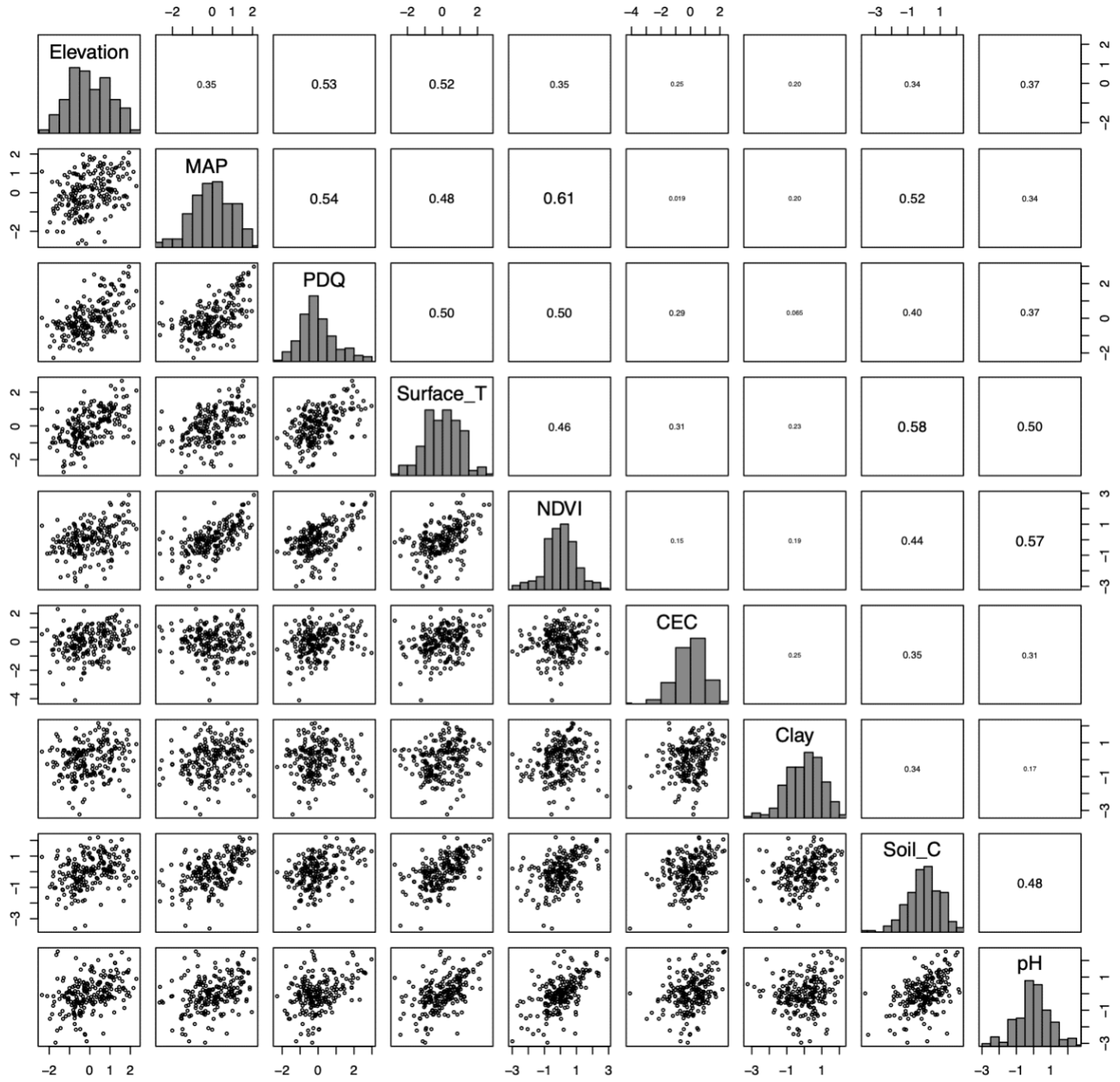


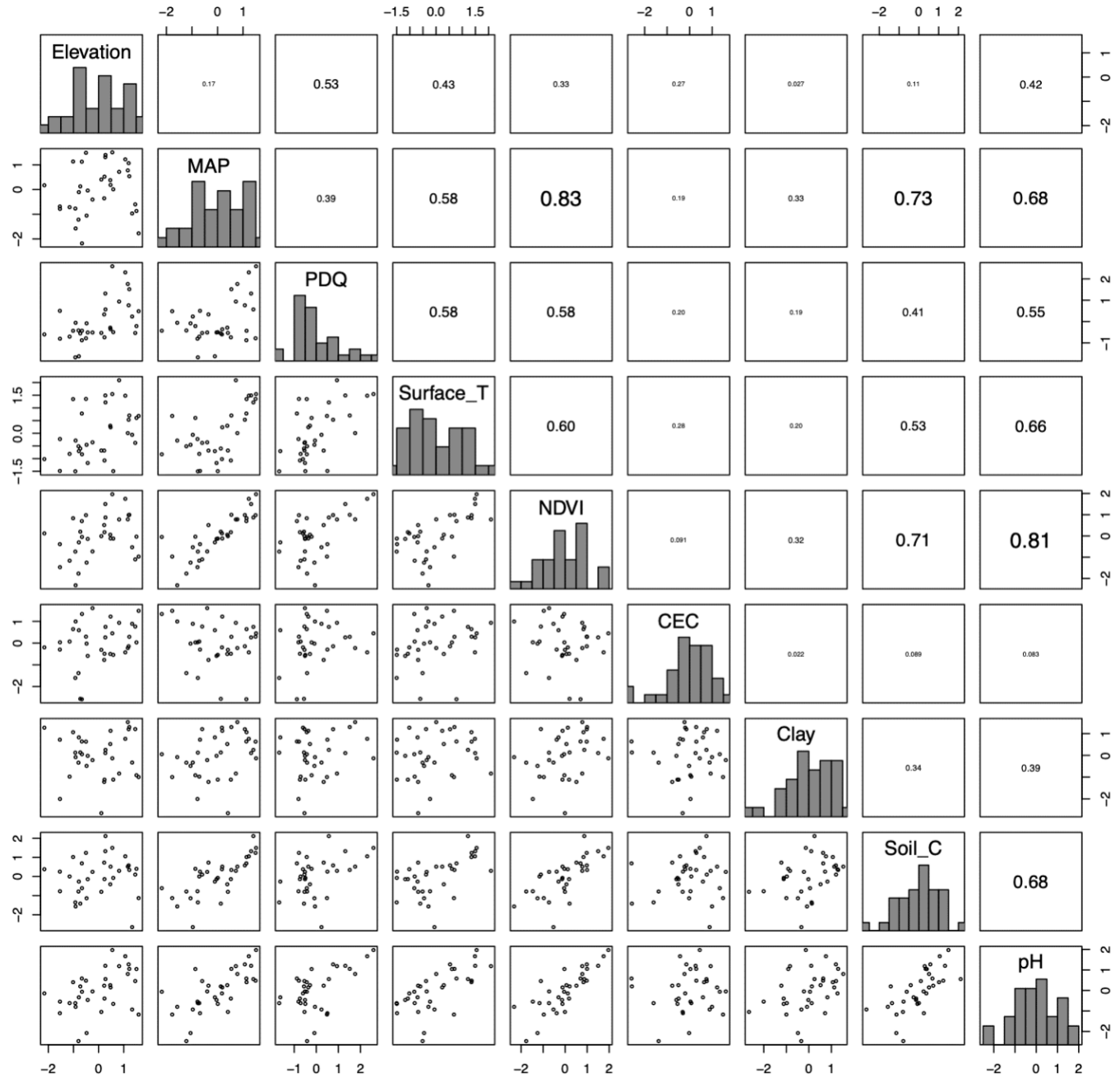
Figure S1: Pairwise correlation coefficients (upper-right panels), scatter plots (lower-left panels) and distributions (diagonal panels) of different forms of environmental heterogeneity (QDS-scale; $\log_{10}$-transformed) across the GCFR and SWAFR, demonstrating the broad independence of these variables across the study regions. Abbreviations follow that in Tables S1.



87

88 **Figure S2:** Pairwise correlation coefficients (upper-right panels), scatter plots (lower-left panels)
89 and distributions (diagonal panels) of different forms of environmental heterogeneity (HDS-
90 scale; $\log_{10}$-transformed) across the GCFR and SWAFR, demonstrating the broad independence
91 of these variables across the study regions. Abbreviations follow that in Tables S1.

92



93

94 **Figure S3:** Pairwise correlation coefficients (upper-right panels), scatter plots (lower-left panels)
95 and distributions (diagonal panels) of different forms of environmental heterogeneity (DS-scale;
96 $\log_{10}$-transformed) across the GCFR and SWAFR, demonstrating the broad independence of
97 these variables across the study regions. Abbreviations follow that in Tables S1.

98

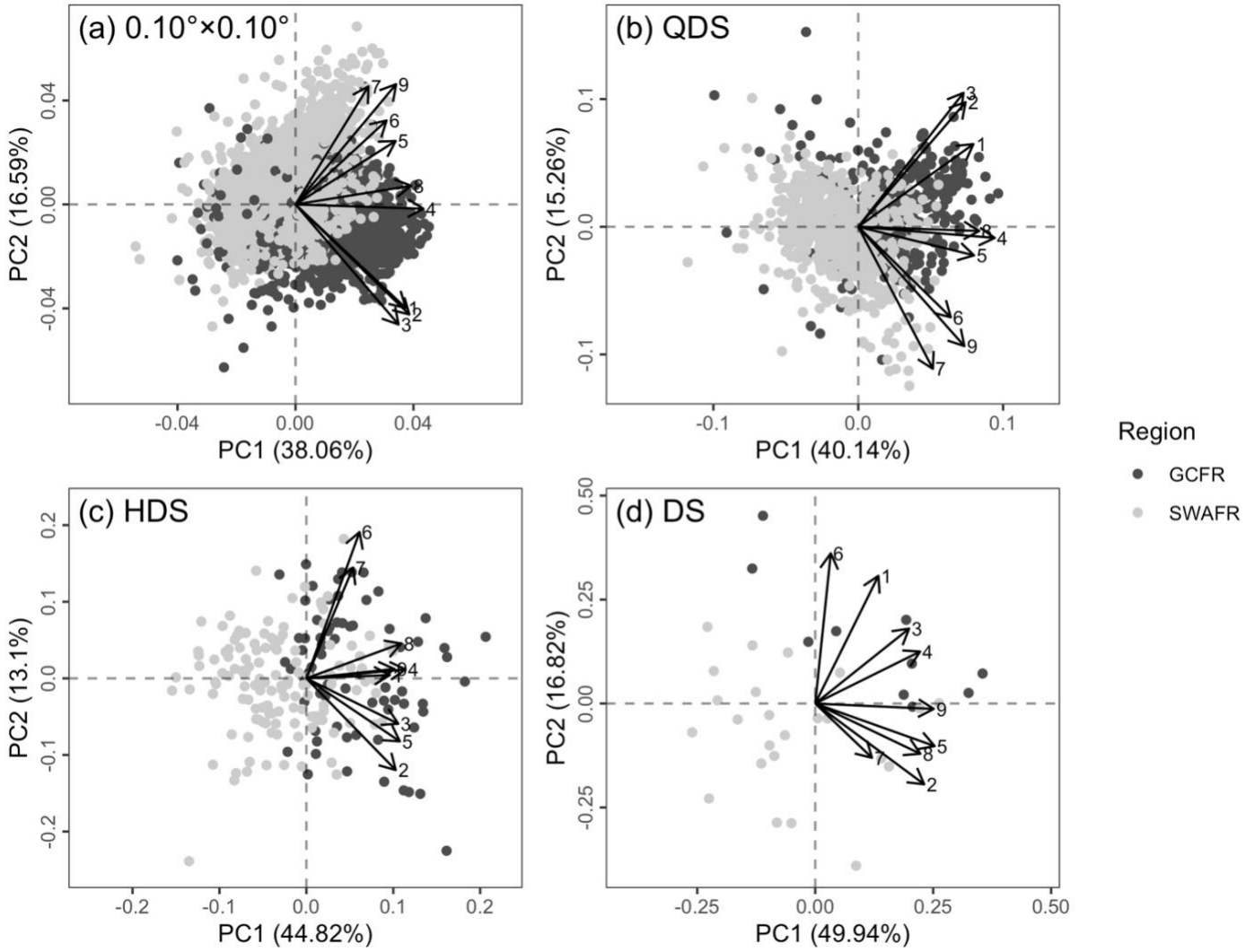
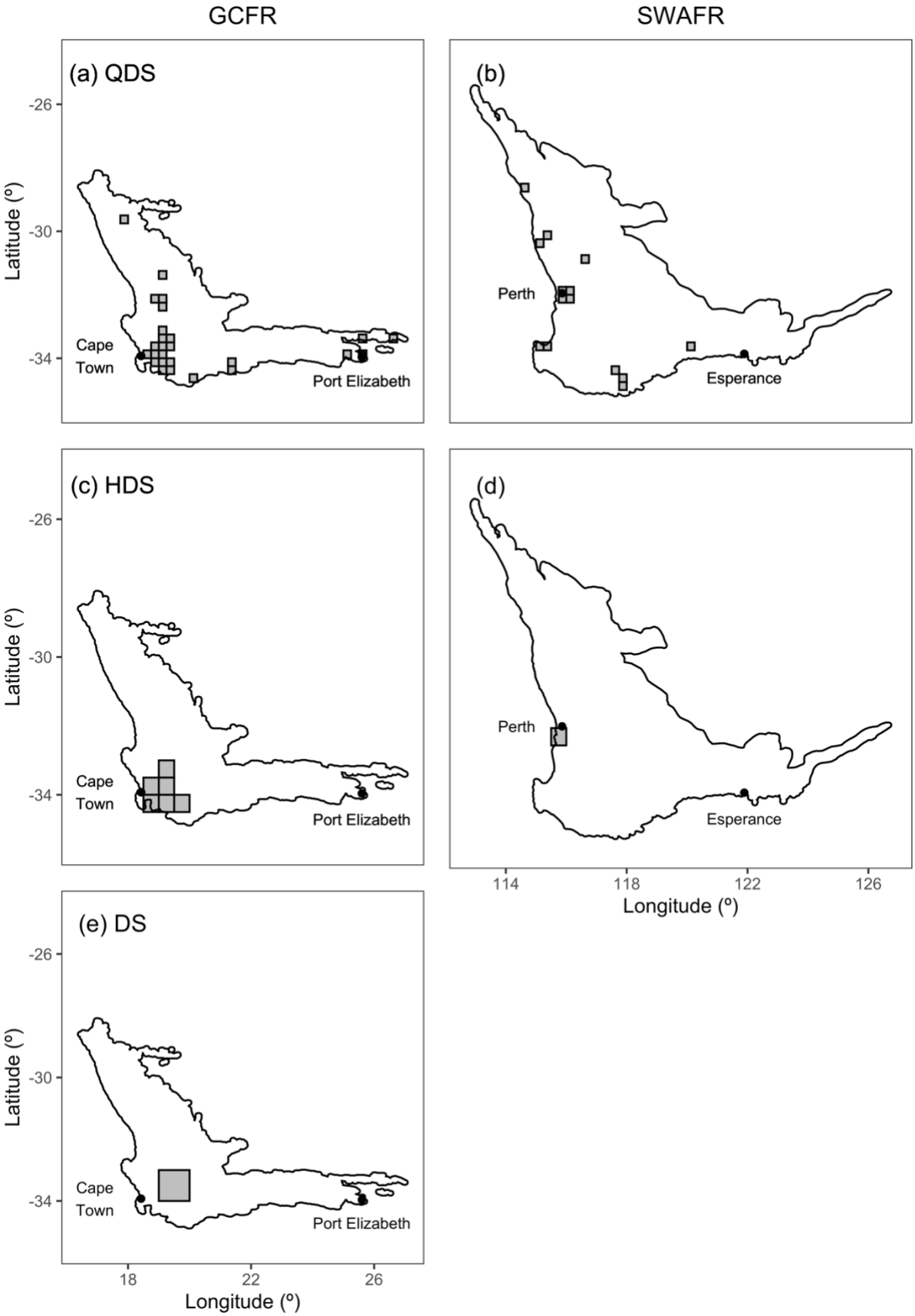


Figure S4: Scatter plots of the first (PC1) and second (PC2) axes following PCAs of the nine forms of environmental heterogeneity (log_{10}-transformed and re-scaled) across the GCFR and SWAFR, calculated at the (a) 0.10°×0.10°-, (b) QDS-, (c) HDS- and (d) DS-scales. The percentage of variance in environmental heterogeneity explained by each axis is noted in parentheses in each panel. Arrows for each heterogeneity variable show each variable's associations with PC1 and PC2, and are labelled as follows: 1, elevation; 2, MAP; 3, PDQ; 4, surface T; 5, NDVI; 6, CEC; 7, clay; 8, soil C; 9, pH.


GCFR
SWAFR
(a) QDS
(b)
(c) HDS
(d)
(e) DS
Latitude (°)
-26
-30
-34
Cape Town
Port Elizabeth
Perth
Esperance
114
118
122
126
18
22
26
Longitude (°)

108

109

Figure S5 (previous page): Grid-cells identified as having outstanding vascular plant species richness (i.e. hotspots) across the GCFR and SWAFR, following univariate regressions (Table 1, Figure 4) against the major axis of environmental heterogeneity (PC1) from the PCA (Figure S4). Hotspots were identified as those cells with residual richness greater than two standard deviations from the mean for that model. No hotspots were found for the SWAFR at the DS-scale. Map projection used: WGS84.

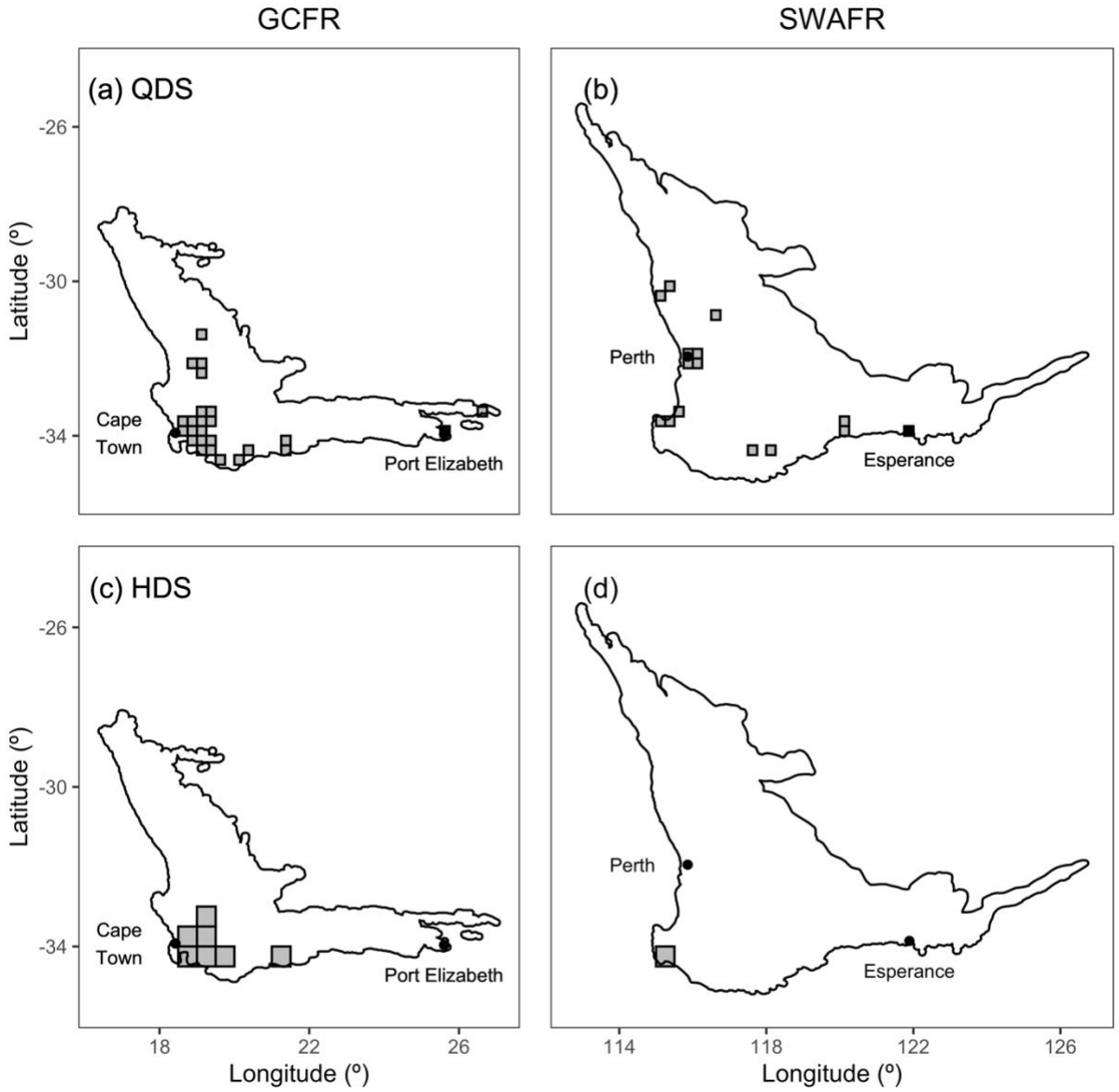


Figure S6: Grid-cells identified as having outstanding vascular plant species richness (i.e. hotspots) across the GCFR and SWAFR, following multivariate regressions (Table S2, Figure 5) against the nine axes of environmental heterogeneity. Hotspots were identified as those cells with residual richness greater than two standard deviations from the mean for that model. Note, no hotspots were found for either region at the DS-scale. Map projection used: WGS84.

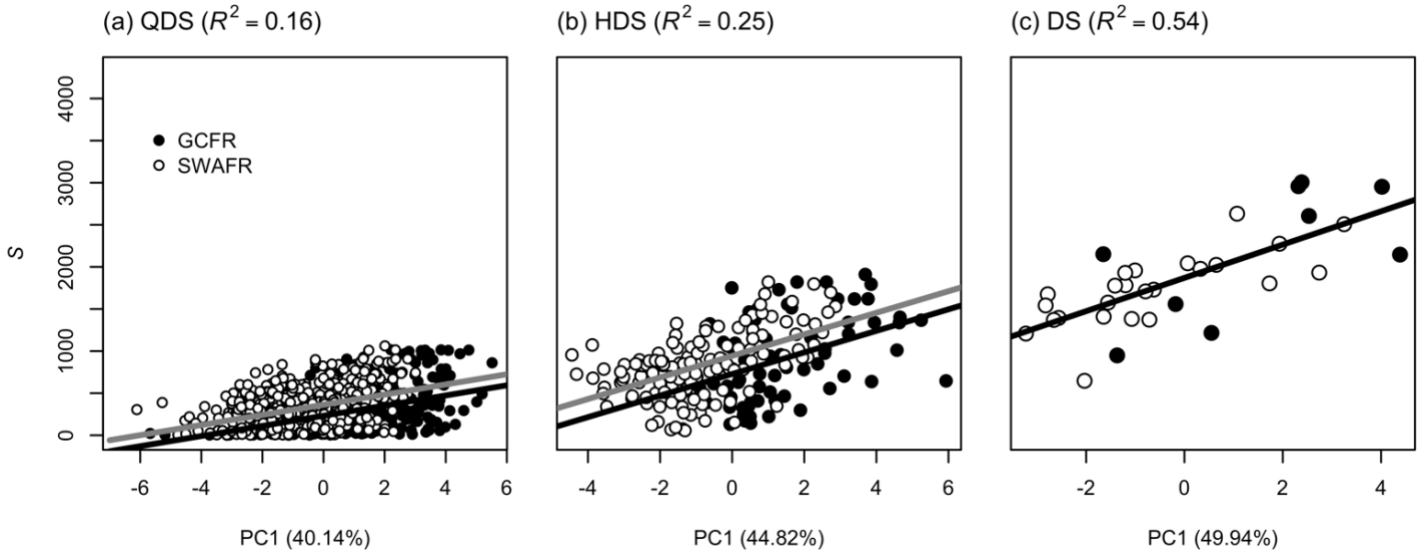
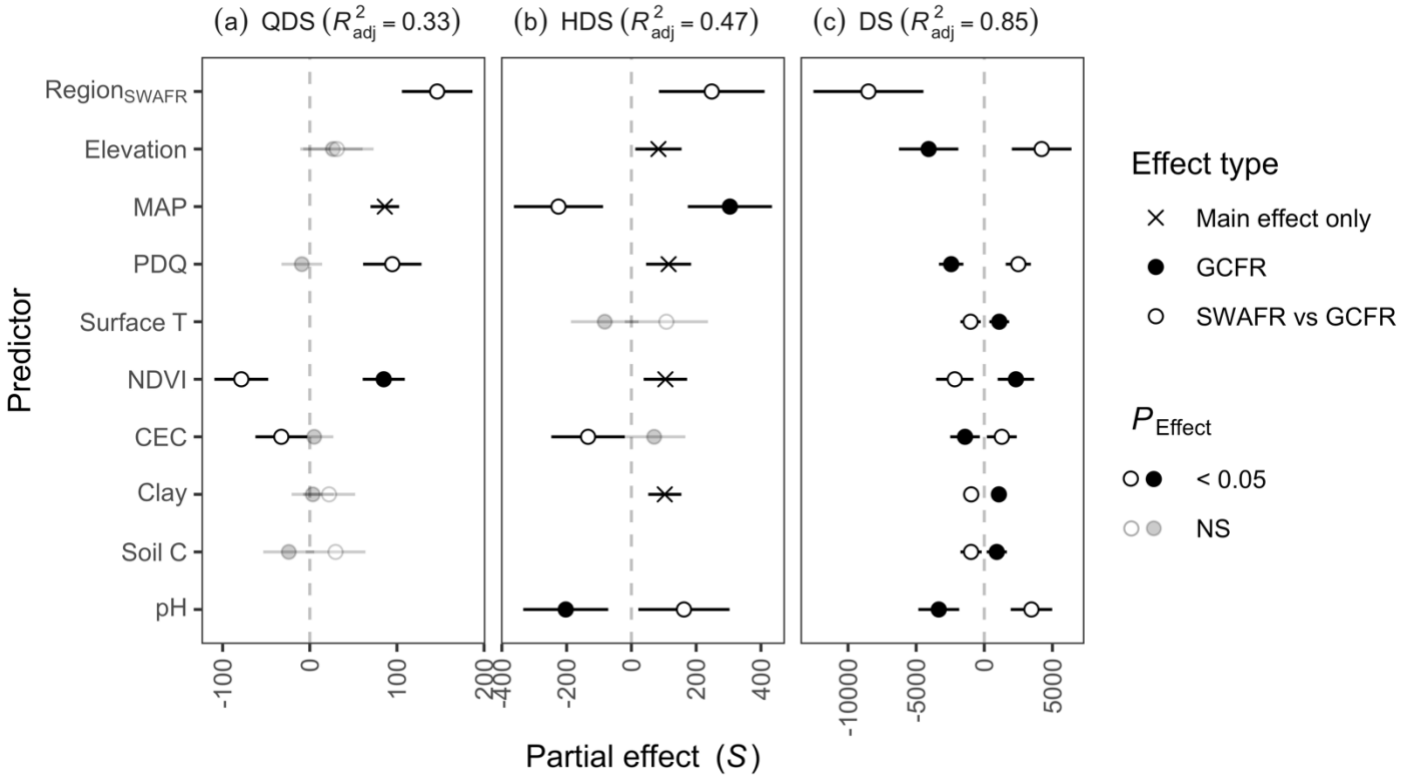


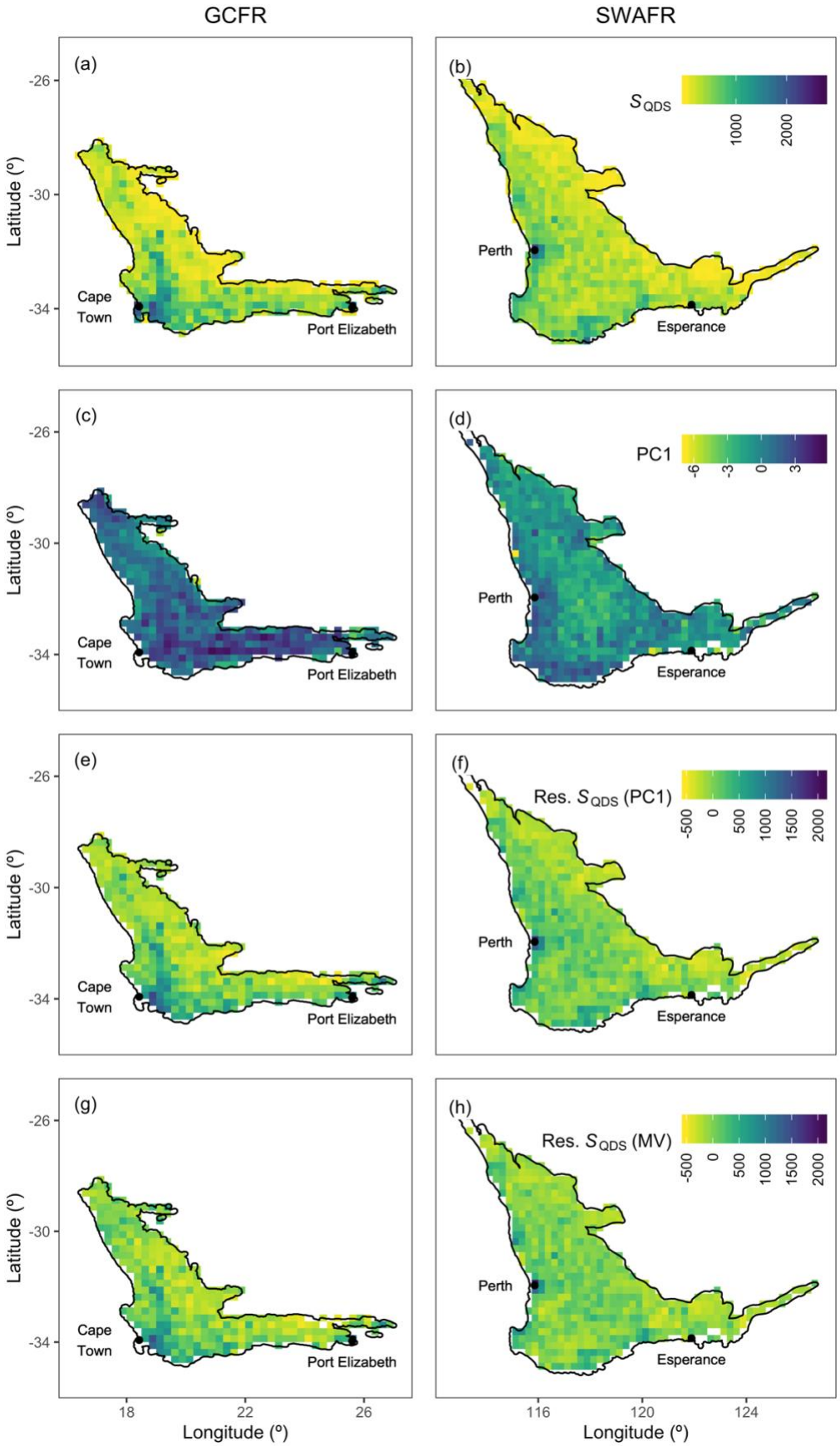
Figure S7: Simple linear regressions of vascular plant species richness, excluding species richness hotspots (defined in terms of their residuals from the full PC1-based ANCOVAs; Figure S5), as (a) S_{QDS}, (b) S_{HDS} and (c) S_{DS} against each respective scale's major axis of environmental heterogeneity (PC1) across the GCFR and SWAFR. These three linear models all have highly significant slopes ($P < 0.001$). For S_{QDS} (a) and S_{HDS} (b) the separate fits for the GCFR (black) and SWAFR (grey) are presented, following the best fitting model at those scales: a "main effect + region" model ($S \sim \beta_0 + \beta_1 X + \beta_2 Region$). The R^2-values of each model and the variation in environmental heterogeneity explained by PC1 from each of the three PCAs are noted in parentheses in the panel and horizontal axis headings respectively.



135

136 **Figure S8:** Slope estimates from multiple linear regressions of vascular plant species richness,
137 excluding species richness hotspots (defined in terms of their residuals from the original
138 multivariate regressions; Figure S6), as (a) S_{QDS}, (b) S_{HDS} and (c) S_{DS} against the various forms of
139 environmental heterogeneity (log_{10}-transformed and re-scaled) across GCFR and SWAFR. Note,
140 as no outliers were found in either region at the DS-scale, panel (c) is identical to that in 5. Points
141 with error bars denote partial effect estimates and their 95% confidence intervals. Filled and
142 empty points represent effect estimates for the GCFR and SWAFR respectively when region-
143 interaction terms were retained during stepwise model selection, while crosses represent main
144 effects (i.e. no region-interaction term retained). Estimates illustrated in black were significant (P
145 < 0.05), while those in grey were not, but still retained during stepwise model selection. The
146 multiple adjusted R^2-values of each model are noted in parentheses in the panel headings.
147 Abbreviations follow that in Tables S1–S3 and Figures S1–S4.

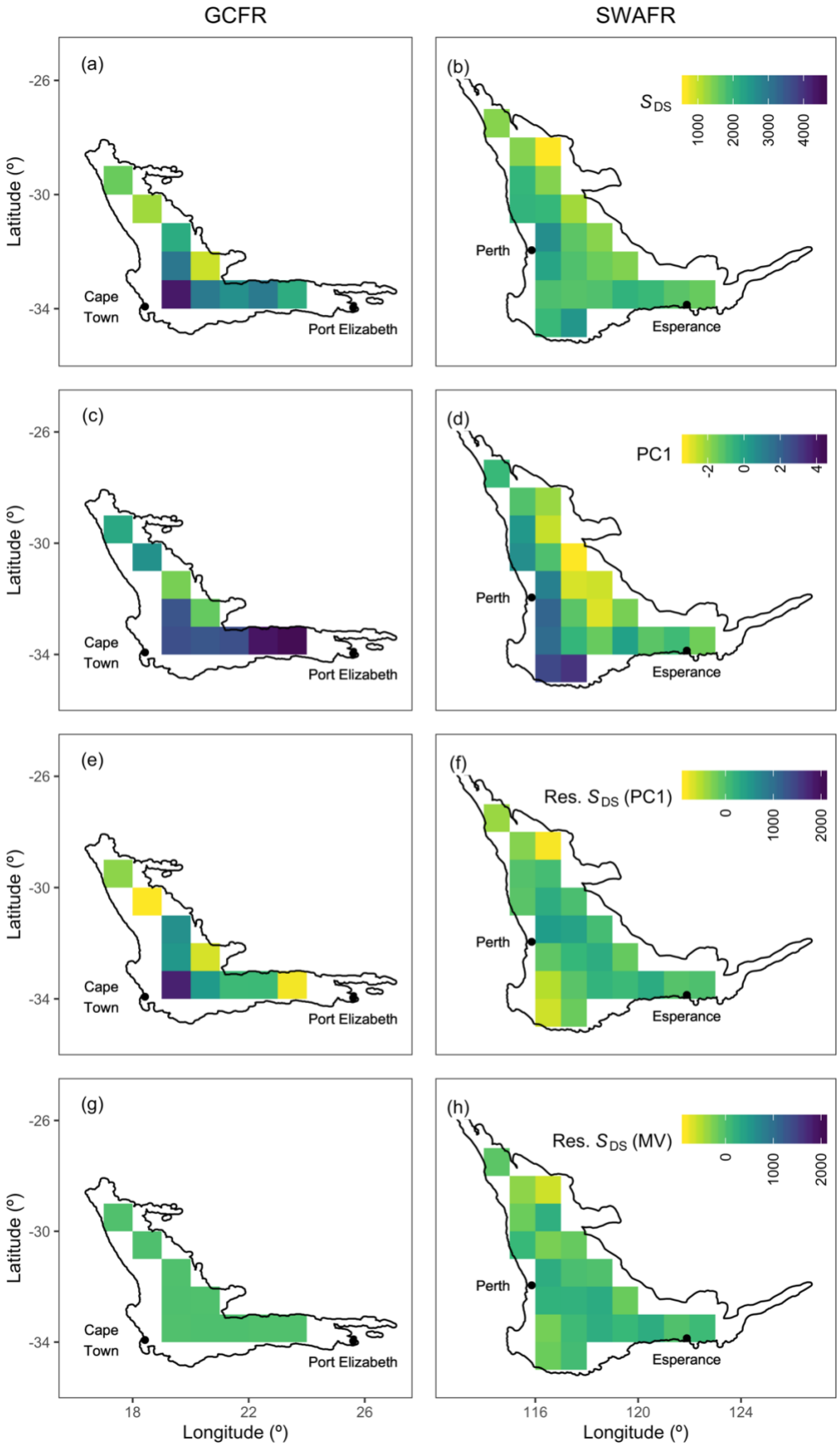
148


GCFR
SWAFR
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
Latitude (°)
-26
-30
-34
Cape Town
Port Elizabeth
Perth
Esperance
S_{QDS}
1000
2000
PC1
-6
-3
0
3
Res. S_{QDS} (PC1)
Res. S_{QDS} (MV)
-500
0
500
1000
1500
2000
18
22
26
116
120
124
Longitude (°)

149

150

Figure S9 (previous page): QDS-scale maps for the GCFR and SWAFR of (a,b) vascular plant species richness, (c,d) the major axis of environmental heterogeneity (PC1) from the PCA of nine forms of environmental heterogeneity (log_{10}-transformed), residuals from regressions of species richness against (e,f) PC1 (Figure 4b) and (g,h) the multivariate (MV) model (Figure 5b). Map projection used: WGS84.


GCFR
SWAFR
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
Latitude (°)
-26
-30
-34
Cape Town
Port Elizabeth
Perth
Esperance
S_{DS}
1000
2000
3000
4000
PC1
-2
0
2
4
Res. S_{DS} (PC1)
Res. S_{DS} (MV)
0
1000
2000
18
22
26
116
120
124
Longitude (°)

157

158

Figure S10 (previous page): DS-scale maps for the GCFR and SWAFR of (a,b) vascular plant species richness, (c,d) the major axis of environmental heterogeneity (PC1) from the PCA of nine forms of environmental heterogeneity (log_{10}-transformed), residuals from regressions of species richness against (e,f) PC1 (Figure 4b) and (g,h) the multivariate (MV) model (Figure 5b). Map projection used: WGS84.

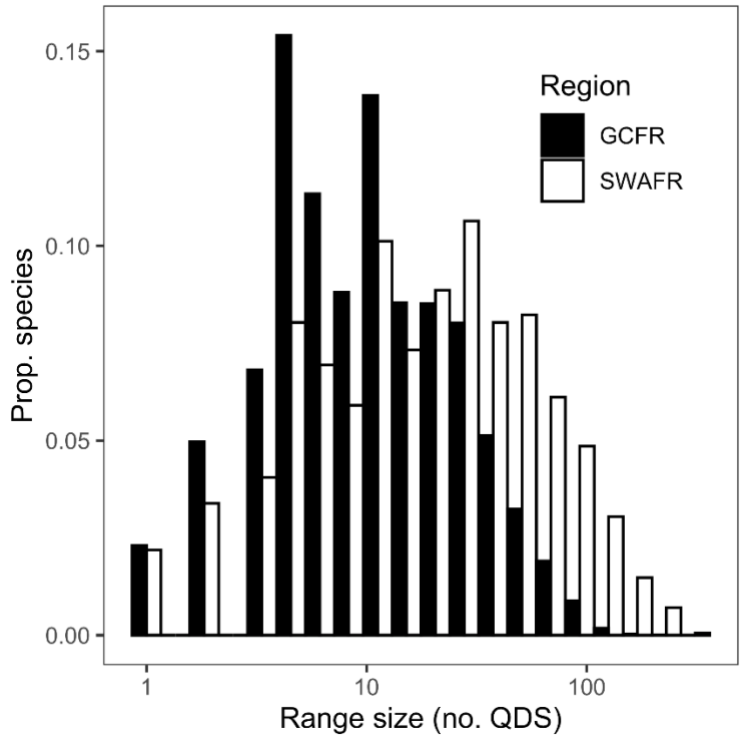
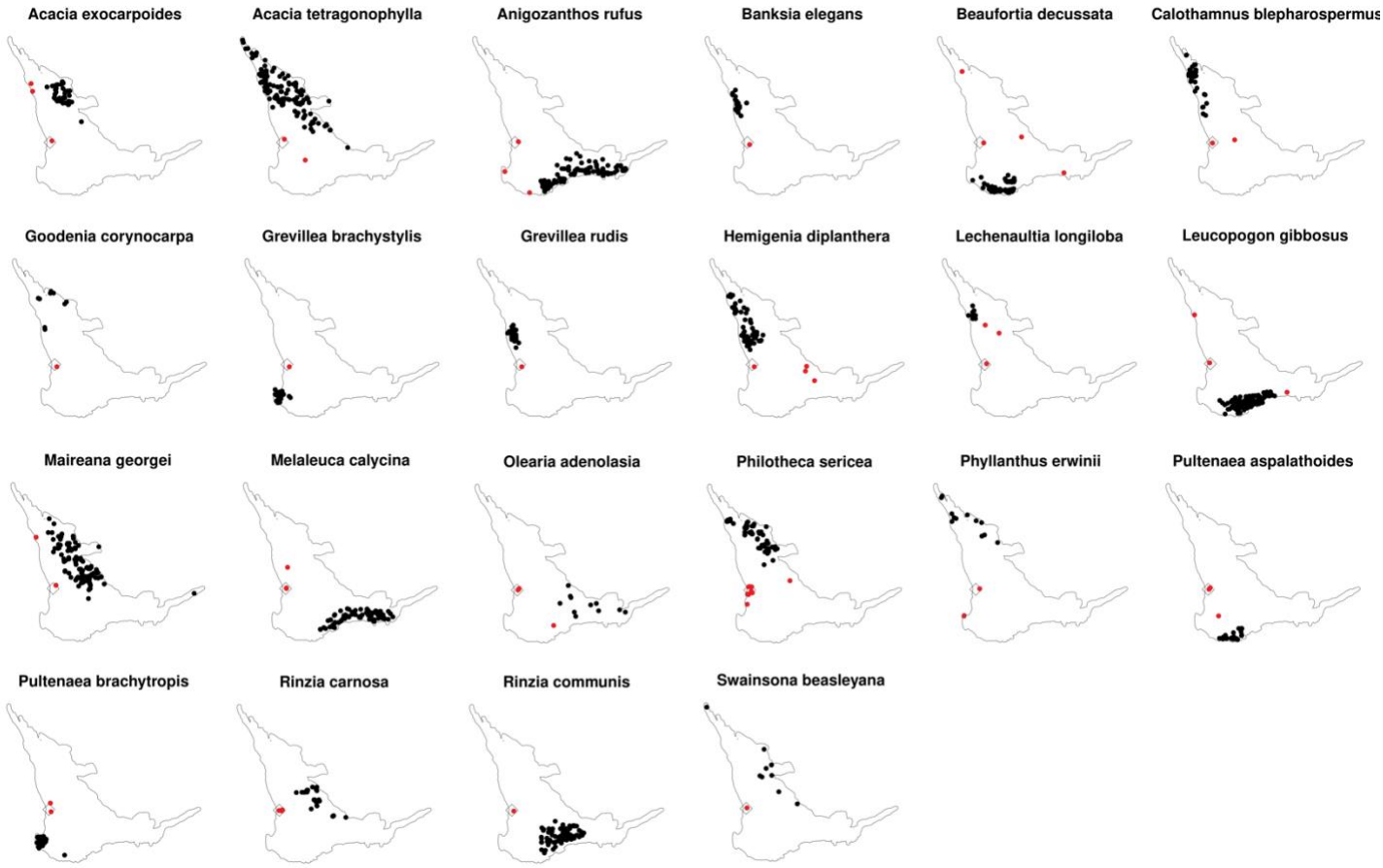


Figure S11: Frequency distribution of vascular plant species' range sizes (quantified as the number of QDS occupied; $\log_{10}$-axis) in the GCFR and SWAFR, based on the species occurrence dataset used here (see text). Frequencies are scaled as the proportions of species within each region's flora. GCFR species have, on average, smaller ranges (ca. 9.7 QDS) than SWAFR species (ca. 17.6 QDS) ($P < 0.001$; two-sided *t*-test).



172

173 **Figure S12:** Examples of vascular plant species from the SWAFR whose occurrence records
174 include outlier localities (in red) near Perth (diamond). Outlier records were determined as those
175 that fall outside the described range for that species (according to distribution maps from
176 FloraBase (https://florabase.dpaw.wa.gov.au/ [Accessed 2 March, 2020]). These 22 species were
177 found within a sample of 300 random species in our occurrence dataset that have records within
178 the four QDS-cells including Perth with exceptional richness (Figure S6b).

179

References

Belda, M., Holtanová, E., Halenka, T., & Kalvová, J. (2014). Climate classification revisited: from Köppen to Trewartha. *Climate Research*, *59*(1), 1–13.

Boyle, B., Hopkins, N., Lu, Z., Garay, J. A. R., Mozzherin, D., Rees, T., … others. (2013). The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinformatics*, *14*(1), 16.

Chamberlain, S., Szocs, E., Boettiger, C., Ram, K., Bartomeus, I., Baumgartner, J., … O'Donnell, J. (2016). taxize: Taxonomic information from around the web. R package version 0.7.8. Retrieved from https://github.com/ropensci/taxize

Farr, T., Rosen, P., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., & Alsdorf, D. (2007). The shuttle radar topography mission. *Reviews of Geophysics*, 45, 1–33. https://doi.org/10.1029/2005RG000183

Funk, C.C., Peterson, P.J., Landsfeld, M., Pedreros, D.H., Verdin, J., Shukla, S., Husak, G., Rowland, J.D., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, 2, 150066. https://doi.org/10.1038/sdata.2015.66

GBIF (24 July 2017a). Global Biodiversity Information Facility (GBIF) Occurrence Download. https://doi.org/10.15468/dl.n6u6n0

GBIF (24 July 2017b). Global Biodiversity Information Facility (GBIF) Occurrence Download. https://doi.org/10.15468/dl.46okua

Gioia, P., & Hopper, S. D. (2017). A new phytogeographic map for the Southwest Australian Floristic Region after an exceptional decade of collection and discovery. *Botanical Journal of the Linnean Society*, *184*(1), 1–15. https://doi.org/10.1093/botlinnean/box010

Global Names Resolver (GNR). http://resolver.globalnames.org/

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagoti, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017) SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE*, 12, e0169748. https://doi.org/10.1371/journal.pone.0169748

Hijmans, R. J. (2016). *raster: Geographic Data Analysis and Modeling. R package version 2.5-8*. https://cran.r-project.org/package=raster

Hopper, S. D., & Gioia, P. (2004). The Southwest Australian floristic region: Evolution and conservation of a global hot spot of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, *35*, 623–650. https://doi.org/10.1146/annurev.ecolsys.35.112202.130201

Integrated Taxonomic Information System (ITIS). http://www.itis.gov

Lambers, H., Brundrett, M. C., Raven, J. A., & Hopper, S. D. (2010). Plant mineral nutrition in ancient landscapes: high plant species diversity on infertile soils is linked to functional diversity for nutritional strategies. *Plant and Soil*, *334*(1–2), 11–31. https://doi.org/10.1007/s11104-010-0444-9

Lambers, H., Shane, M. W., Cramer, M. D., Pearse, S. J., & Veneklaas, E. J. (2006). Root structure and functioning for efficient acquisition of phosphorus: Matching morphological and physiological traits. *Annals of Botany*, *98*(4), 693–713. https://doi.org/10.1093/aob/mcl114

Larsen, R., Holmern, T., Prager, S. D., Maliti, H., & Røskaft, E. (2009). Using the extended quarter degree grid cell system to unify mapping and sharing of biodiversity data. *African Journal of Ecology*, *47*(3), 382–392. https://doi.org/10.1111/j.1365-2028.2008.00997.x

Mucina, L., & Rutherford, M. C. (2006). *The vegetation of South Africa, Lesotho and Swaziland.* South African National Biodiversity Institute.

NASA (2017a). MOD13C2 MODIS/Vegetation Indices Monthly L3 Global 0.05Deg CMG (MOD13C2) v006. NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, U.S.A. https://doi.org/10.5067/MODIS/MOD13C2.006

NASA (2017b). MOD11C3 MODIS/Land Surface Temperature Emissivity Monthly L3 Global 0.05Deg CMG (MOD11C3) v006. NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, U.S.A. https://doi.org/10.5067/MODIS/MOD1132.006

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., & others. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, *51*(11), 933–938.

Shane, M. W., Cramer, M. D., & Lambers, H. (2008). Root of edaphically controlled Proteaceae turnover on the Agulhas Plain, South Africa: phosphate uptake regulation and growth. *Plant, Cell & Environment*, *31*(12), 1825–1833.

Taxonomic Name Resolution Service (TNRS) v4.0. http://tnrs.iplantcollaborative.org/

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686