Supporting Information

Contents

Read Me

This is the pdf version of the Supporting Information. This file contains the same information as the html version, but without interactive graphics.

A. Spectral Diversity Metrics

Many spectral diversity studies are based on imagery, where the spectral reflectance of pixels are compared among one on another. Because our data are based on spectral leaf libraries, the methods we used to calculate spectral diversity were either: 1) modified from image-based studies to be based on reflectance spectra per plot, and abundance weighted when possible or 2) taken from existing methods from the functional diversity literature. These functional diversity metrics did not take into account species abundance.

Because hyperspectral reflectance data has a large number of possible wavelengths that could be used in calculating spectral diversity metrics, the same metric can be calculated in many ways. Reasons for selecting particular wavelengths and the number of wavelengths vary between studies. For instance, one of the more commonly used indices, the coefficient of variation, could be calculated with as few as five wavelengths or up to hundreds of them. As the number of wavelengths used to calculate spectral diversity increases, its underlying dimensionality grows at the trade-off of higher collinearity between wavelengths (Petchey and Gaston 2002). In the case of their trait diversity, i.e., Functional Diversity (FD) from Petchey and Gaston (2002), higher dimensionality in a diversity metric weighs species identity less, so that replacement of any one species with a set of traits (or spectra) in a community does not dramatically affect the end diversity result (in effect down-weighting redundancy). In the case of trait diversity, higher dimensionality has been shown to force relationships with species richness to become more linear instead of asymptotic– a desired feature if one wants to accurately forecast species richness across the entire range of possible richness values using linear models (Petchey and Gaston 2002). We chose to base our metric selection on predictive capacity assessed through cross-validation using all 500 wavelengths, since we felt that the heart of the surrogacy question is primarily one of prediction.

Coefficient of variation metric (SpectraCV)

The coefficient of variation (CV) spectral diversity metric was measured by taking the abundance-weighted reflectance values of all species at a plot and calculating the CV for each wavelength.

Abundance weighting was done by multiplying the percent cover, w, of species j at site k by the reflectance, r, across wavelengths, i. This can be represented as:

$$w_{j,k}r_{i,j,k}$$

The mean abundance weighted reflectance at 450 nanometers for a plot, k, with n species would be:

$$\bar{R}_{450,k} = \frac{w_{1,k}r_{1,450,k} + w_{2,k}r_{2,450,k} + ... + w_{n,k}r_{n,450,k}}{n}$$

Similarly, the standard deviation in this case would then be:

$$\hat{sd}_{450,k} = \sqrt{\frac{1}{n}\sum_{j=1}^{n} w_{j,k}(r_{j,450,k} - \bar{R}_{450,k})}$$

The coefficient of variation for wavelength 450 at site k, would be calculated as:

$$CV_{450,k} = \frac{\hat{sd}_{450,k}}{\bar{R}_{450,k}}$$

We then took the average of all CV values across all wavelengths (500 total) to arrive a single diversity value for the plot:

$$\bar{CV}_k = \frac{\sum_{i=450}^{949} CV_{i,k}}{500}$$

Spectral angle and spectral information divergence metrics (SpectraSAM and SpectraSID)

Spectral angle (Kruse et al. (1993)) and spectral information divergence (Chang (2000)) are calculated based on the difference between two spectra, a reference and a sample spectra. We compared the abundance weighted plot average spectral signal to that of the spectral signal for species each within the plot. This results in a vector of angle (or divergence) values with length equal to the number of species present at a plot. A single plot-level spectral diversity value is calculated by taking the average of the angle/divergence values within the plot. This approach is similar to the functional dispersion (FDis) measure of functional diversity, see Villéger, Mason, and Mouillot (2008).

We calculated spectral angle based on the abundance weighting (w), each individual species reflectance (r), and the average plot spectral signal ($\bar{R}$) as:

$$SAM = \cos^{-1}\frac{\sum_{i=450}^{949} w_{j,k}r_{i,j,k}\bar{R_{i,k}}}{\sqrt{\sum_{i=450}^{949} w_{j,k}r_{i,j,k}}\sqrt{\sum_{i=450}^{949}\bar{R_{i,k}}}}$$

where each wavelength i of species j at site k were weighted by the percent cover (w) of species, j, at site k. The average abundance-weighted wavelength i at site k is represented by $\bar{R_{i,k}}$. This formula gives the spectral distance between one species, j, and the average spectra at site k. This was repeated for all species within the site and then the all angles were averaged to arrive at a single value representing the average spectral angle of species within a plot. Similar approaches were used to modify the other spectral diversity calculations such as the coefficient of variation and spectral information divergence.

Convex hull area (SpectraCVA)

Similar to the prior two metrics, the convex hull area (replicated from Gholizadeh et al. (2018)) is based on the comparison of each abundance weighted species spectral signature,$r_{i,j,k}$, to the plot average spectral signature, $\bar{R}_{450}$. Using the value of each wavelength between the two of these, we can draw a convex hull and calculate the area of that hull. If a species is similar to the plot signature then it has a small hull value; a species with a more unique signature compared to the plot average would have a larger hull value. This procedure calculates n hull values, i.e., the same number of hulls as there are species in the plot, and then averages them to arrive a single value for the plot.

PCA volume ordination (SpectralCVV)

In order to calculate the convex hull volume inspired by Dahlin (2016), we ran a principal component analysis on the entire GCFR spectral reflectance matrix (where rows were species within each subregion and columns were wavelengths). Taking the first three most explanatory axes, we then calculated hull volume values based on the species membership for each site. A large volume would denote a wide spread of spectral values and thus high spectral diversity.

This convex hull volume method was calculated based on spectra that were not abundance weighted because the PCA was run directly over the entire GCFR spectral library. Our spectral measurements did not always occur alongside the plot inventories so that there is not always a one to one relationship between spectral measurements and their abundances. Species could be assigned multiple abundance values if they appear in more than one plot within a subregion; however, our spectral library was averaged to the species level within subregions.

Metrics from the trait diversity literature (SpectralRichness and SpectralDistance)

We applied existing functional diversity metrics to spectral diversity using two methods: FRic proposed by Villéger, Mason, and Mouillot (2008) and Functional Diversity (FD) proposed by Petchey and Gaston (2002). When applied to spectral diversity we refer to these as "spectral richness" and "spectral distance," respectively. The former method is similar to a convex hull volume (but the number of axes is chosen automatically) while the latter is a branch length distance method. Neither of these are abundance weighted unlike the previous custom built spectral diversity methods.

B. Correlations between spectral diversity metrics

Spectral diversity metrics that are constructed in similar manners are more likely to correlated with each other as seen in SI Figure 1. For instance, spectral convex hull volume (SpectraCVV) and Spectral Richness are both convex volume metrics without abundance weighting and are highly related (0.92); they only vary in the number of principle coordinate axes selected. In general the metrics that did not have abundance weighting were highly correlated with each other, i.e., Spectral Richness, Spectral Distance and SpectraCVV.

The names of the spectral diversity metrics are abbreviated below as such: **SpectralRichness** = Spectral richness metric based on FRic from Villeger et al., (2012), **SpectraCV** = coefficient of variation, **SpectraSAM** = spectral angle mapping, **SpectraCVA** = convex hull area, **SpectraSID** = spectral information divergence, **SpectraCVV** = convex hull volume, **SpectralDistance** = dendrogram based method based on FD from Petchey and Gaston (2002).

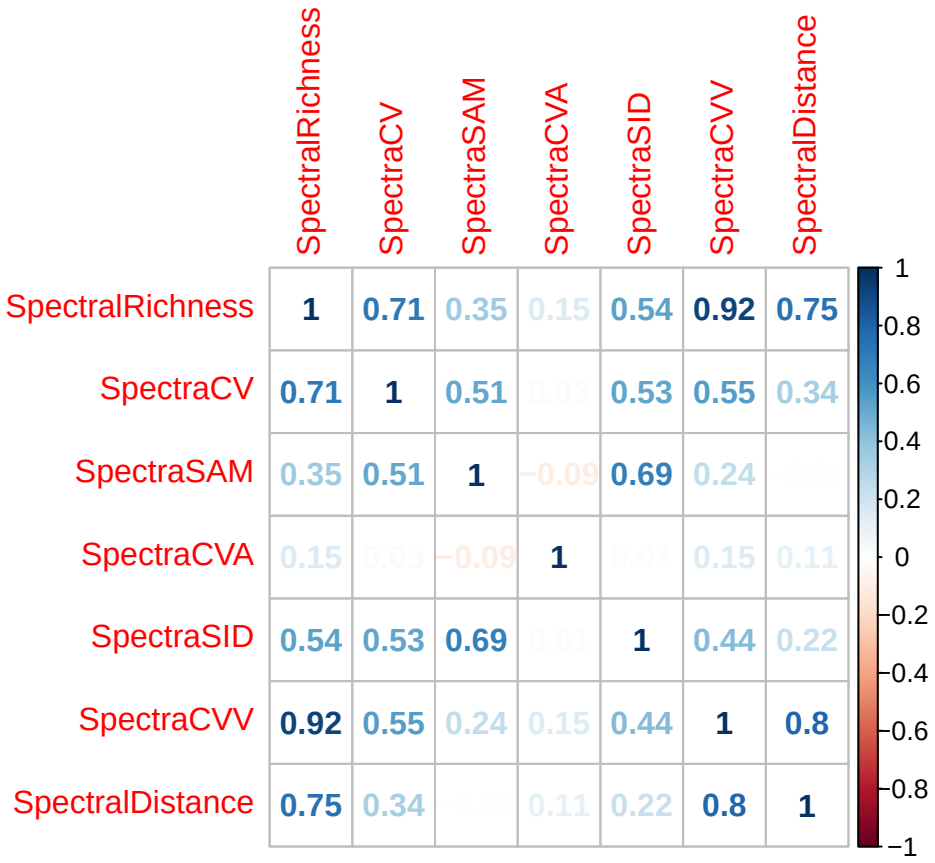

Figure 1: Correlation plots of spectral diversity metrics. Lightly colored numbers suggest low correlations while fully colored ones suggest high correlations.

```
##                  SpectralRichness  SpectraCV   SpectraSAM  SpectraCVA
## SpectralRichness        1.0000000 0.71489910  0.354108786  0.15006360
## SpectraCV               0.7148991 1.00000000  0.514896862  0.02815896
## SpectraSAM              0.3541088 0.51489686  1.000000000 -0.09397543
## SpectraCVA              0.1500636 0.02815896 -0.093975428  1.00000000
## SpectraSID              0.5361740 0.52775172  0.688011484  0.01438251
## SpectraCVV              0.9239786 0.55292765  0.240743401  0.14675687
## SpectralDistance        0.7493205 0.33941512 -0.008236795  0.10726154
##                  SpectraSID SpectraCVV SpectralDistance
## SpectralRichness 0.53617399  0.9239786      0.749320483
## SpectraCV        0.52775172  0.5529276      0.339415125
## SpectraSAM       0.68801148  0.2407434     -0.008236795
## SpectraCVA       0.01438251  0.1467569      0.107261538
## SpectraSID       1.00000000  0.4398785      0.221630497
## SpectraCVV       0.43987845  1.0000000      0.798528745
## SpectralDistance 0.22163050  0.7985287      1.000000000
```

C. Comparing spectral diversity metrics with cross-validation

We ran two forms of cross-validation to compare the ability of different spectral diversity metrics to predict biodiversity across the Greater Cape Floristic Region. In the first form, we used a ten fold cross validation where plots were randomly assigned to excluded folds. Error was assessed for each excluded fold based on root mean square error (RMSE). Using notation from James et al. (2013), this results in 10 folds that can be averaged to arrive a single cross-validated value:

$$CV_{(10)} = \frac{1}{10}\sum_{i=1}^{10} RMSE_i$$

We present the full range of RMSE values from each fold to not only show the model bias based on the average RMSE error, but all the model variance based on the range of RMSE values. This idea is captured within the boxplot comparison below.

In the second form, we use a blocked cross-validation strategy where individual subregions were excluded from the model and then RMSE was calculated based on the excluded subregion. This was done in order to investigate the influence of subregion as a dependent structure (*sensu* Roberts et al. (2017)) in the metrics.

Lastly, in order to present results such that errors can be compared between models with different scales, we present results in both RMSE and normalized RMSE values. The normalized RMSE values allow us to directly ask whether a certain spectral diversity metric was more accurate for species diversity compared phylogenetic diversity. We normalized values based on the following minimum-maximum normalization procedure:

$$\frac{RMSE_i - min(RMSE)}{max(RMSE) - min(RMSE)}$$

The names of the spectral diversity metrics are abbreviated below as such: **SpectralRichness** = Spectral richness metric based on FRic from Villéger, Mason, and Mouillot (2008), **SpectraCV** = coefficient of variation, **SpectraSAM** = spectral angle mapping, **SpectraCVA** = convex hull area, **SpectraSID** = spectral information divergence, **SpectraCVV** = convex hull volume, **SpectralDistance** = dendrogram based method based on FD from Petchey and Gaston (2002).

Random Ten Fold Cross-validation

According to SI Figures 2-3, the spectral distance metric consistently had the lowest RMSE (and NRMSE) value compared to the other metrics. In many ways, this is unsurprising given the similar construction of functional diversity, phylogenetic diversity, and spectral distance metrics, i.e., all were distance-based branch length methods. Based on the NRMSE results, spectral distance most accurately predicted species richness denoting how strong the correlation to species richness is between all of these metrics.

It is noteworthy that even the spectral diversity metrics with the high RMSE values, often spectral angle (spectraSAM) or convex hull area (spectraCVA), still had predictive accuracies around 1 standard deviation of the plots. For instance, the spectral angle measure had median RMSE values of 8.42 for species richness. Given that the plot average species richness is around ~15 with a standard deviation ~8 across the GCFR, this is still promising sign of the strength of the relationship between spectral diversity biodiversity.

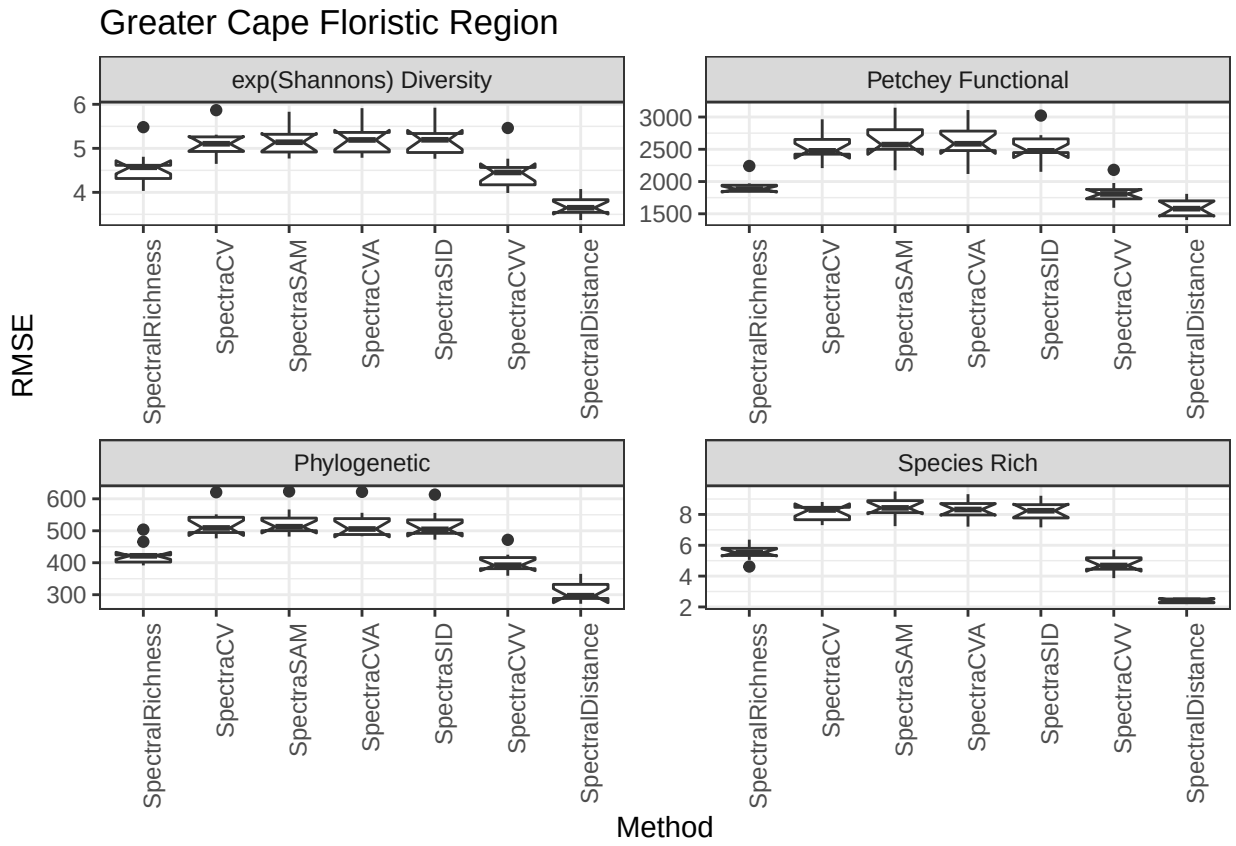

Figure 2: Root mean square error comparison based on random ten-fold cross-validation of across all 1,267 sites.

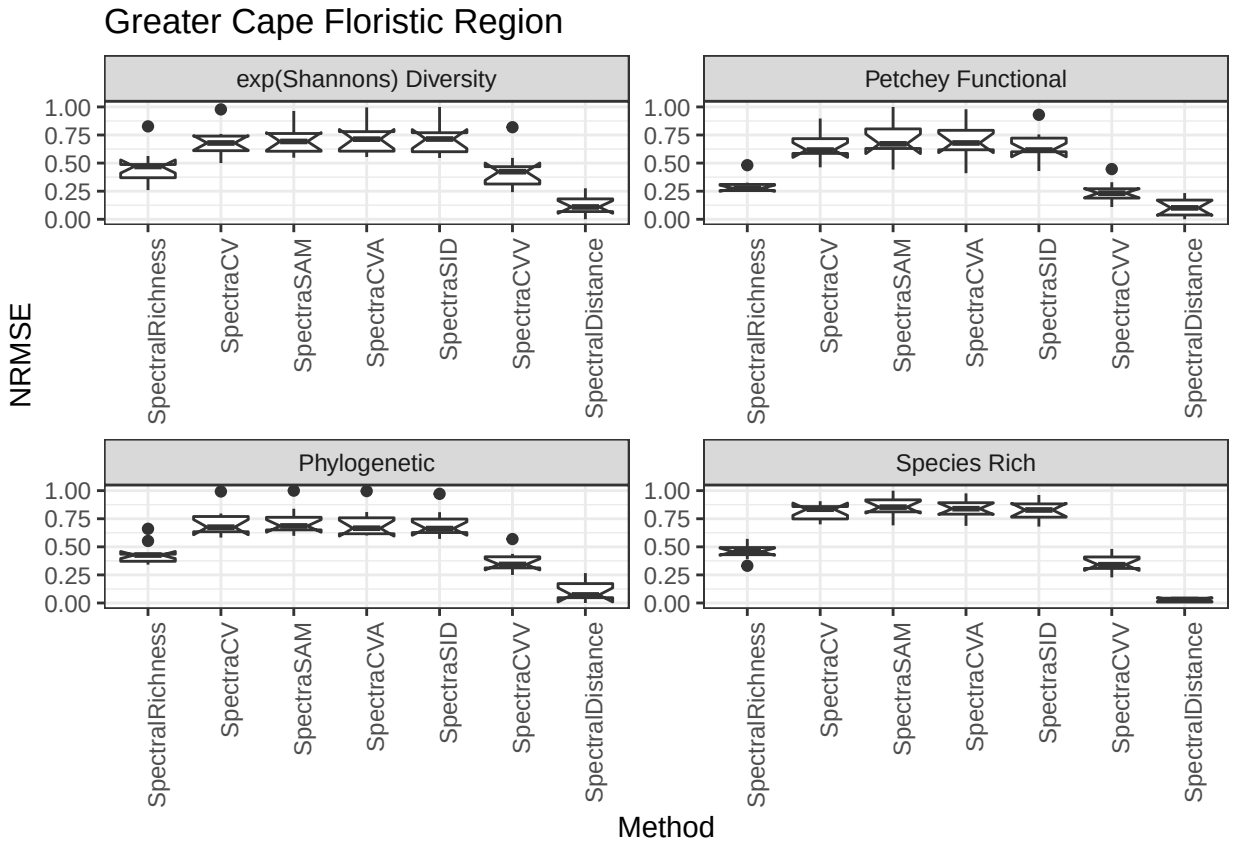

Figure 3: Normalised root mean square error comparison based on random ten-fold cross-validation of across all 1,267 sites.

Subregion Block Cross-Validation

As in the case of the random cross-validation, the blocked cross-validation results showed that spectral distance metric is the strongest predictor of other forms of biodiversity. Note that variance in predictive accuracy is higher (as evidenced by the wider boxplot intervals) compared to the random cross-validation procedure, but the models are making predictions based on information from the other subregions.

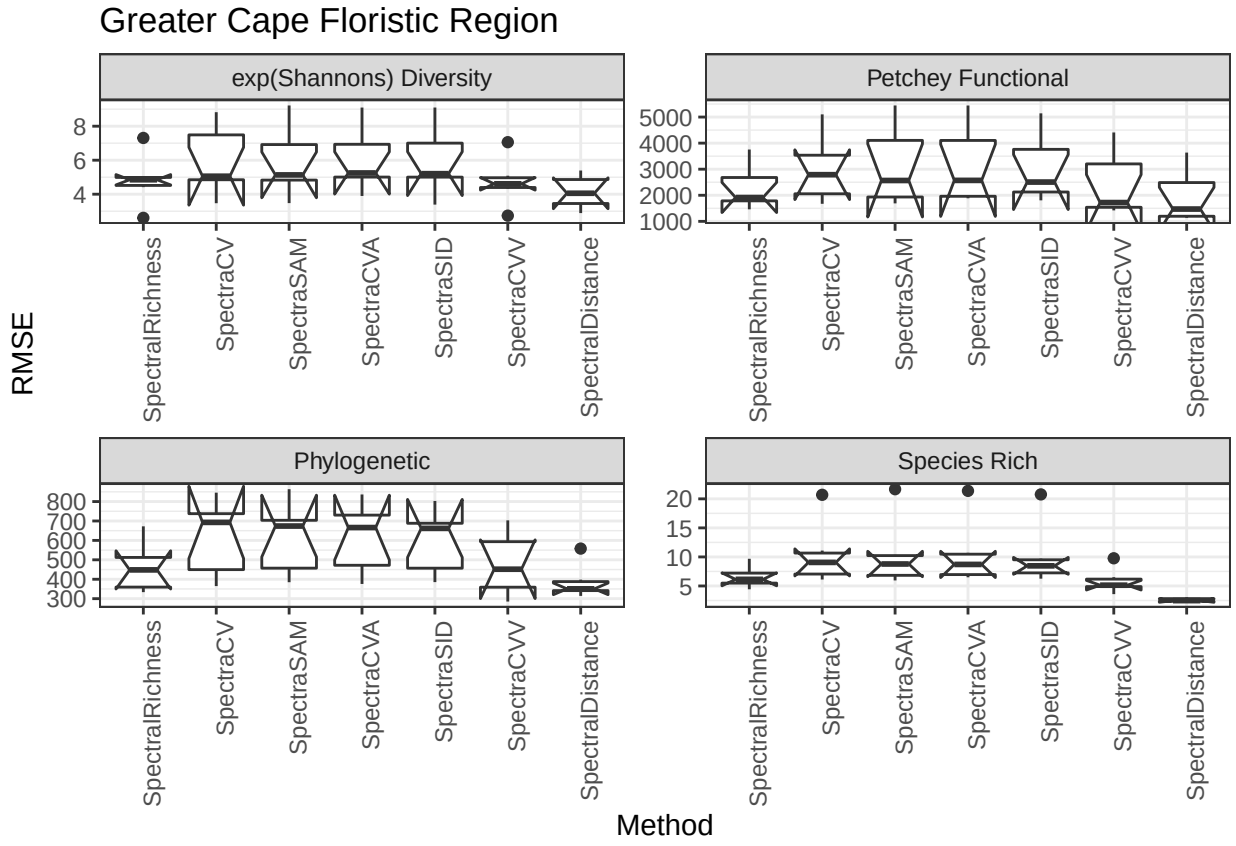

Figure 4: Root mean square error comparison based on blocked cross-validation of subregions across all 1,267 sites.

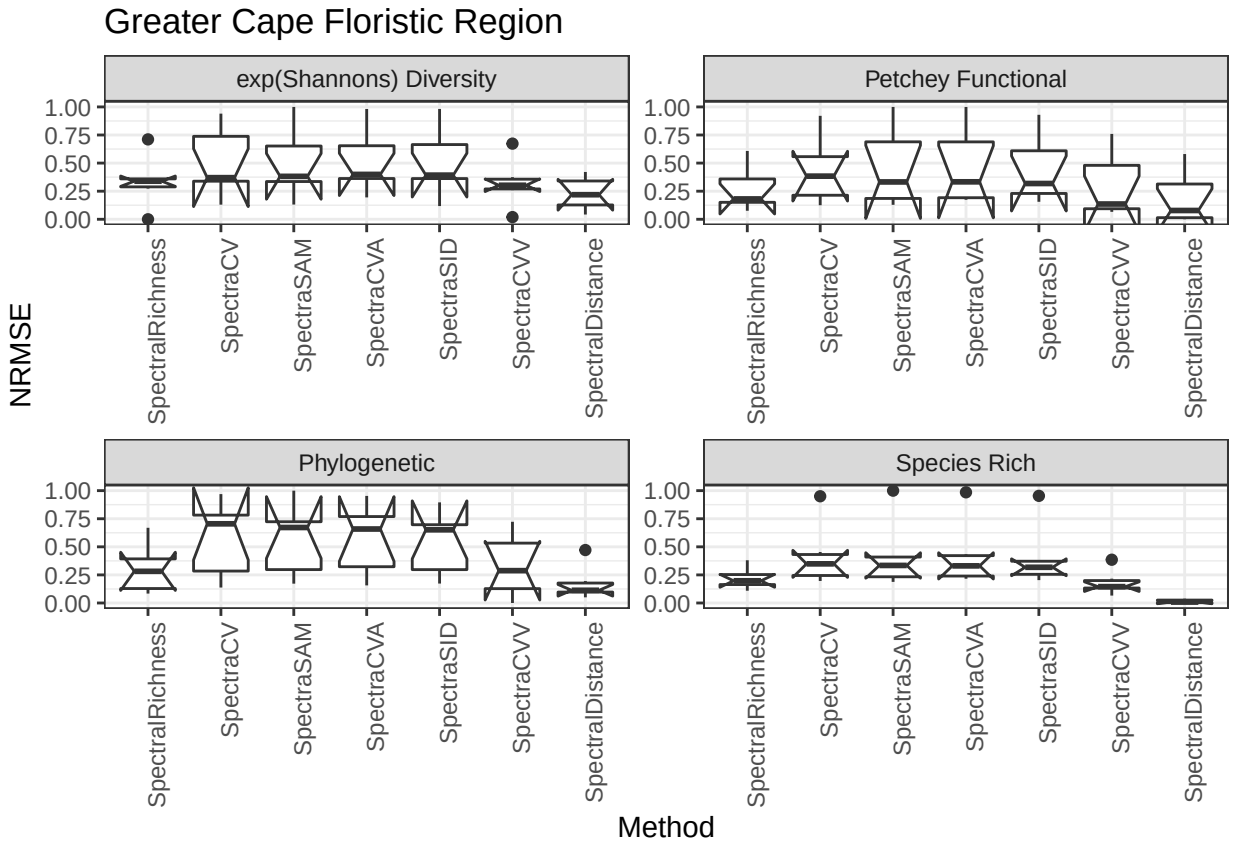

Figure 5: Normalized root mean square error comparison based on blocked cross-validation of subregions across all 1,267 sites.

D. General relationship between spectral diversity and biodiversity (Fig. 3A-D)

Given the overall predictive performance of the spectral distance metric, we proceed with it as our main metric for the following analyses. The models presented below are the same as Figure 3A-D in the manuscript and evaluate the hypothesis that spectral diversity is an accurate predictor of biodiversity across all sites. Both variables are scaled, i.e., centered and standardized. All models below are Gaussian linear models with the following model formulation:

$$biodiversity_i = \alpha_i + \beta_1 * spectraldiversity_i$$

Spectra diversity prediction of Shannon's diversity

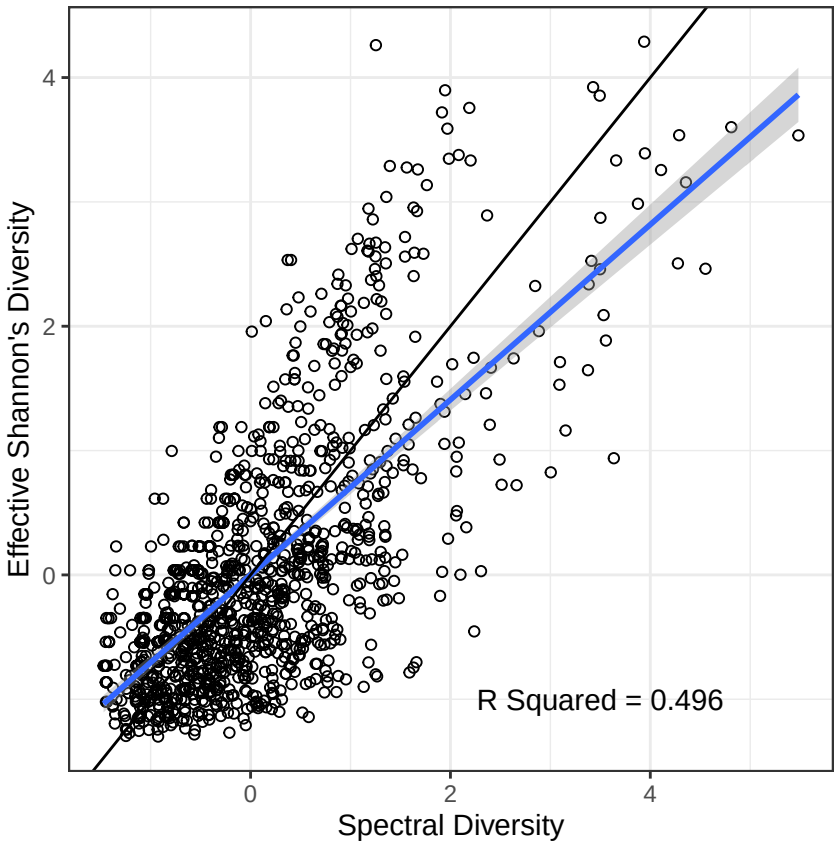

Figure 6: Relationship of spectral diversity and effective Shannon's diversity.

Spectra diversity prediction of species richness

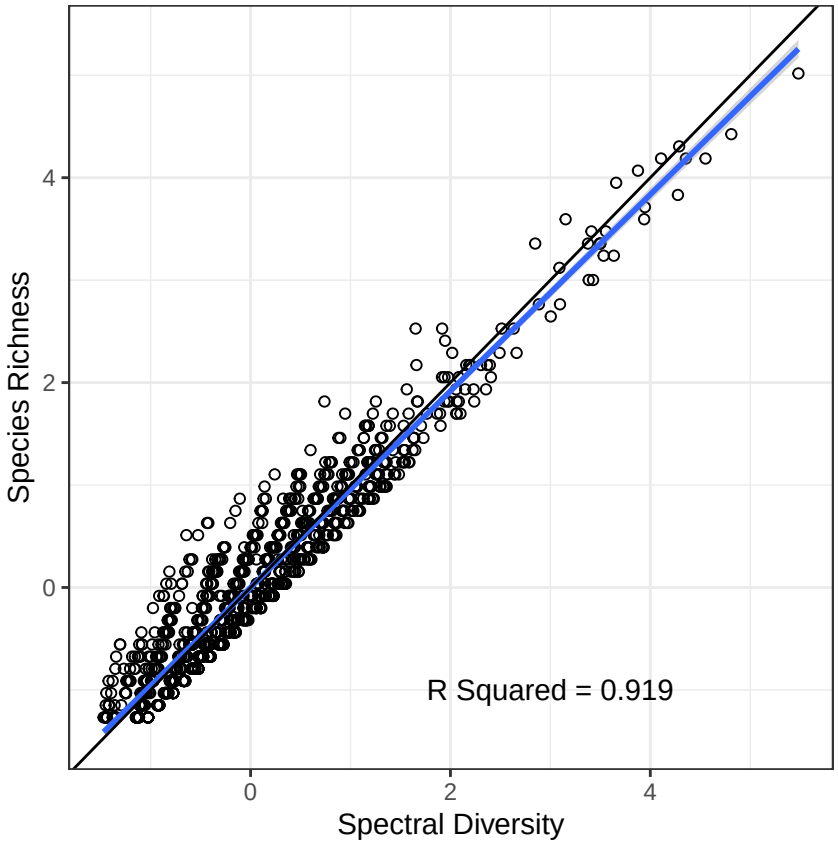

Figure 7: Relationship of spectral diversity and species richness.

Spectral diversity prediction of Functional Diversity

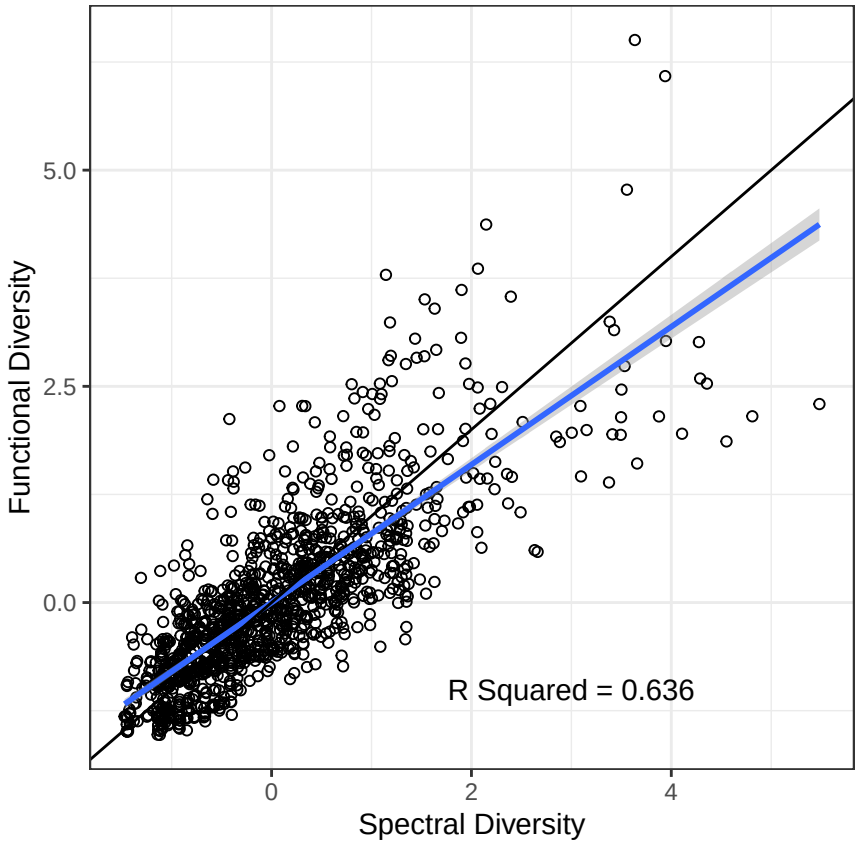

Figure 8: Relationship of spectral diversity and Functional Diversity.

Spectral diversity prediction of Phylogenetic Diversity

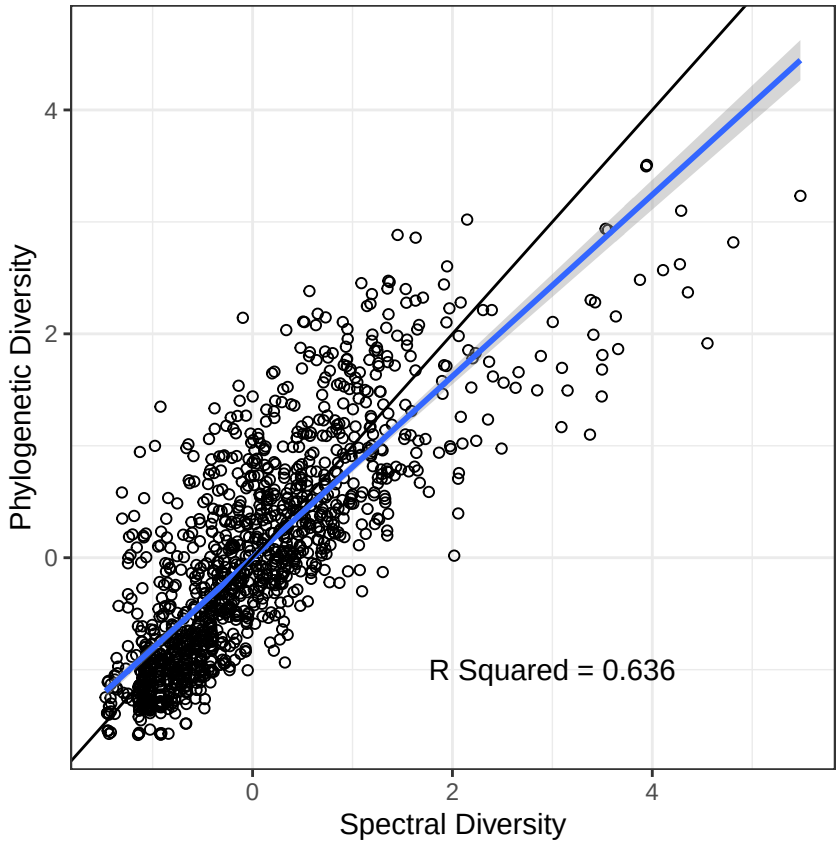

Figure 9: Relationship of spectral diversity and Phylogenetic Diversity.

E. Biome models

The following are model outputs and diagnostics for Figure 4A-C. All models are Gaussian linear models and are assessed with standard diagnostics for model assumptions (i.e., normality of errors, homogeneity of variance). Predictor variables (i.e., spectral diversity) were centered to the mean and scaled to address the mismatch in scales between effective Shannon's diversity and spectral diversity. The general model form was:

$$biodiversity_i = \alpha_i + \beta_1 * spectraldiversity_i + \beta_2 * biome_j + \beta_3 * biome_j * spectraldiversity_i$$

Note that the following models use treatment level contrasts for the biome categories. This means that biome coefficients are in reference to the first one listed, e.g., if the Fynbos intercept is the in first row then the Succulent Karoo intercept is in reference to the Fynbos intercept. Using SI Table 1 as an example, the Fynbos slope is 0.77 while the Karoo slope is in reference to the Fynbos slope and would be (0.77 + 0.19 = 0.96). Similarly, the Fynbos intercept in SI Table 1 is -0.12 while the Karoo intercept would be (-0.12+0.62 = 0.50)

Effective Shannon's diversity (Fig. 4A)

Relations of scaled variables

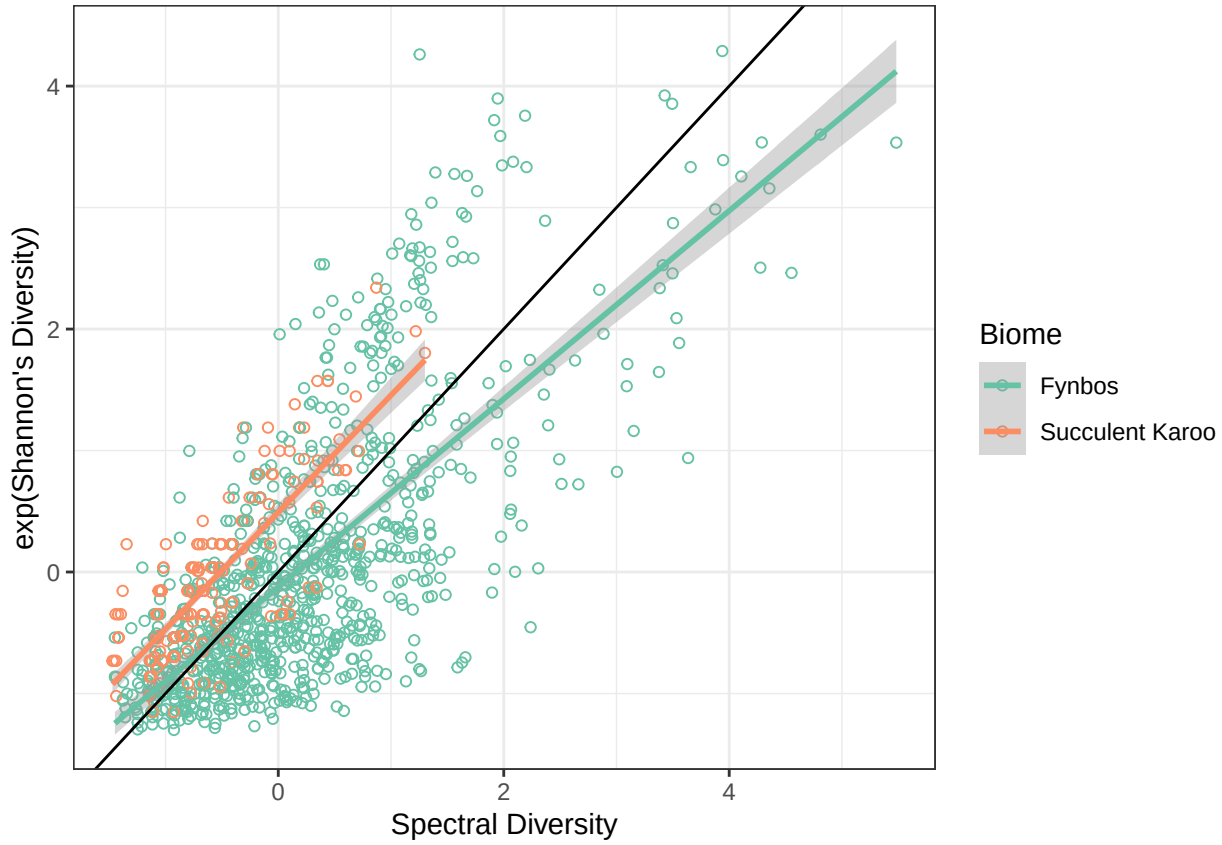

Figure 10: Spectral diversity predicting exponentiated Shannon's diversity within Fynbos and Succulent Karoo Biomes. Individual biomes can be turned on and off by clicking on them in the legend. Detailed information of each point is revealed when the mouse is hovered over the point. The black line is a 1:1 line representing maximal surrogacy.

Table 1: Model coefficients for spectral diversity predicting effective Shannon's diversity within biomes

Term	Estimate	Standard Error	Test Statistic	P value	Significant
Intercept Reference (Fynbos)	-0.123	0.023	-5.273	0.000	Yes
Spectral Diversity Slope Reference (Fynbos)	0.774	0.022	34.711	0.000	Yes
Succulent Karoo Intercept	0.616	0.074	8.366	0.000	Yes
Succulent Karoo Slope	0.188	0.086	2.181	0.029	Yes

```
## [1] "Adjusted R-squared:  0.533"
```

Relations of unscaled variables

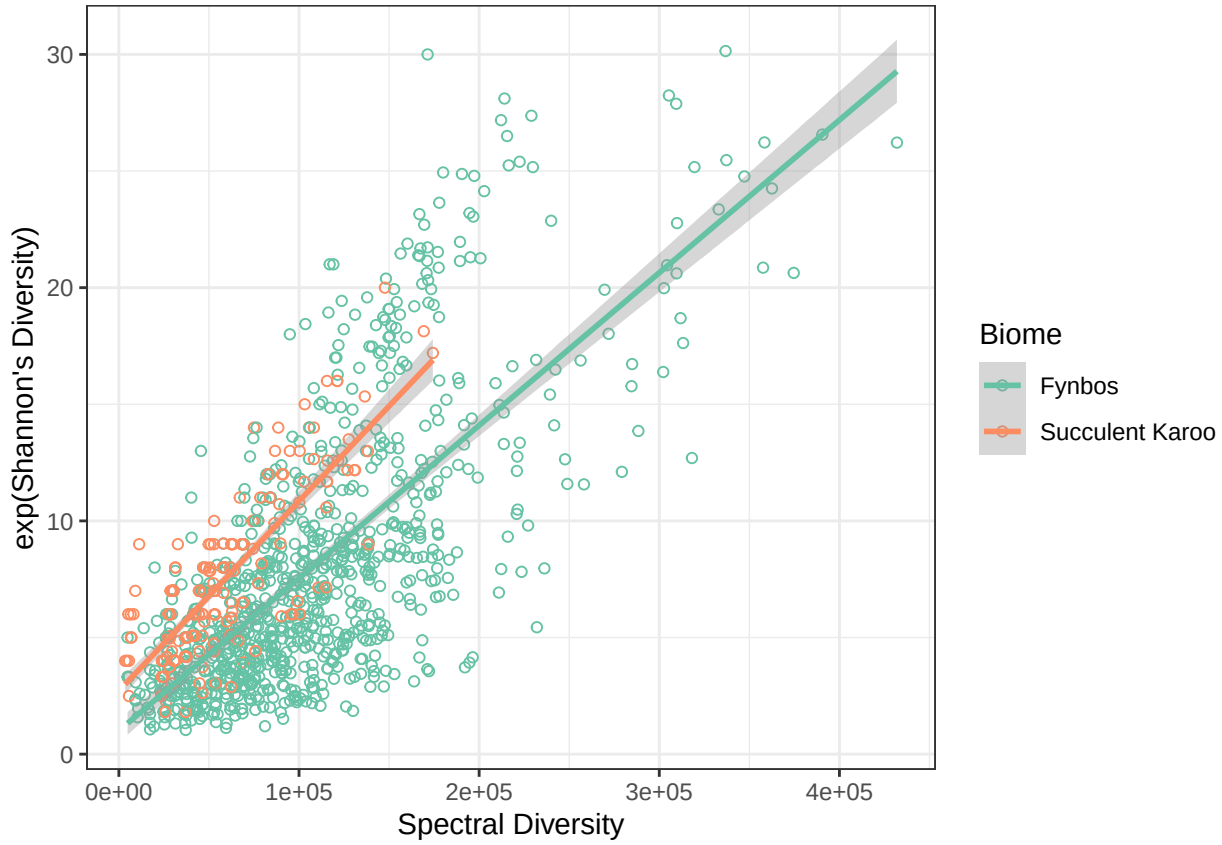

Figure 11: Unscaled version of spectral diversity predicting exponentiated Shannon's diversity within Fynbos and Succulent Karoo Biomes.

Functional diversity (Fig. 4B)

Relations of scaled variables

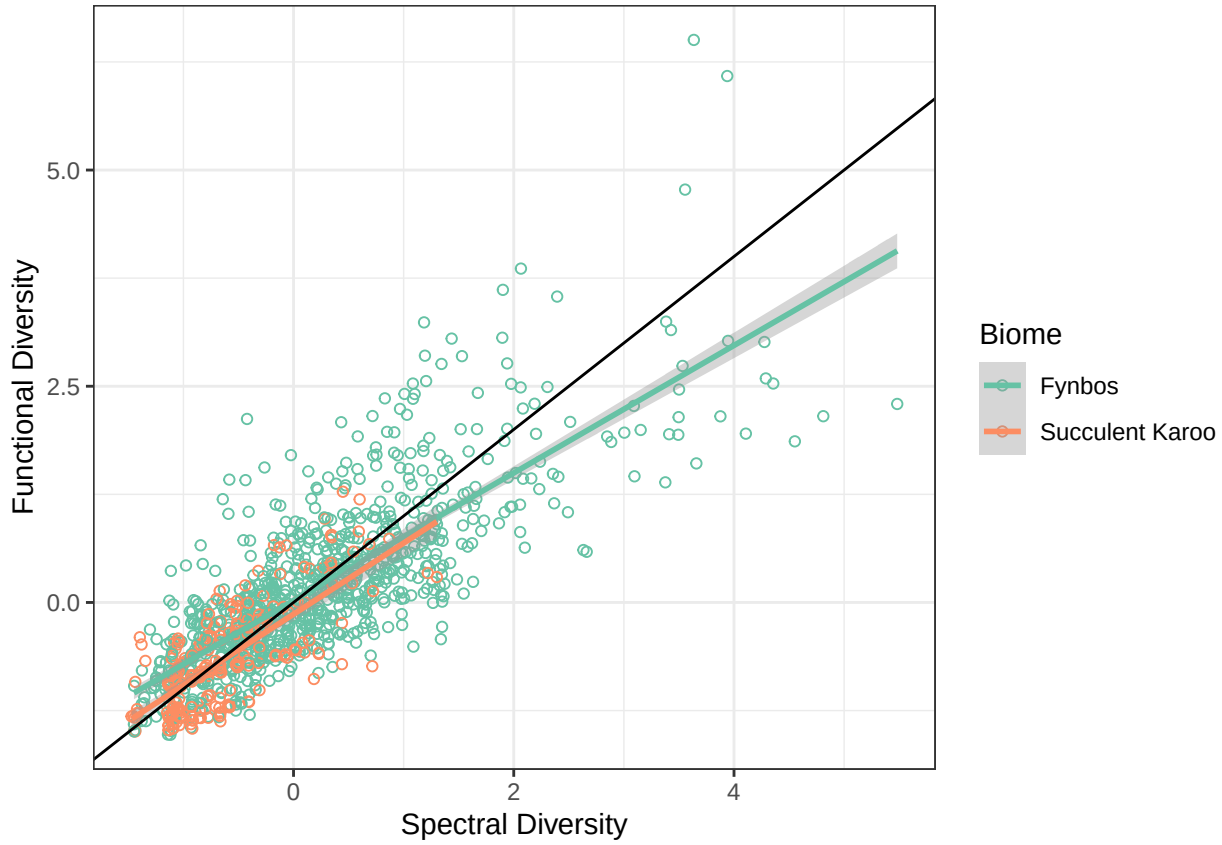

Figure 12: Spectral diversity predicting Functional Diversity within Fynbos and Succulent Karoo Biomes. Individual biomes can be turned on and off by clicking on them in the legend. Detailed information of each point is revealed when the mouse is hovered over the point. The black line is a 1:1 line representing maximal surrogacy.

Table 2: Model coefficients for spectral diversity predicting functional diversity within biomes

Term	Estimate	Standard Error	Test Statistic	P value	Significant
Intercept Reference (Fynbos)	0.023	0.019	1.204	0.229	
Spectral Diversity Slope Reference (Fynbos)	0.737	0.018	41.273	0.000	Yes
Succulent Karoo Intercept	-0.159	0.059	-2.692	0.007	Yes
Succulent Karoo Slope	0.086	0.069	1.239	0.216	

```
## [1] "Adjusted R-squared:  0.664"
```

Relations of unscaled variables

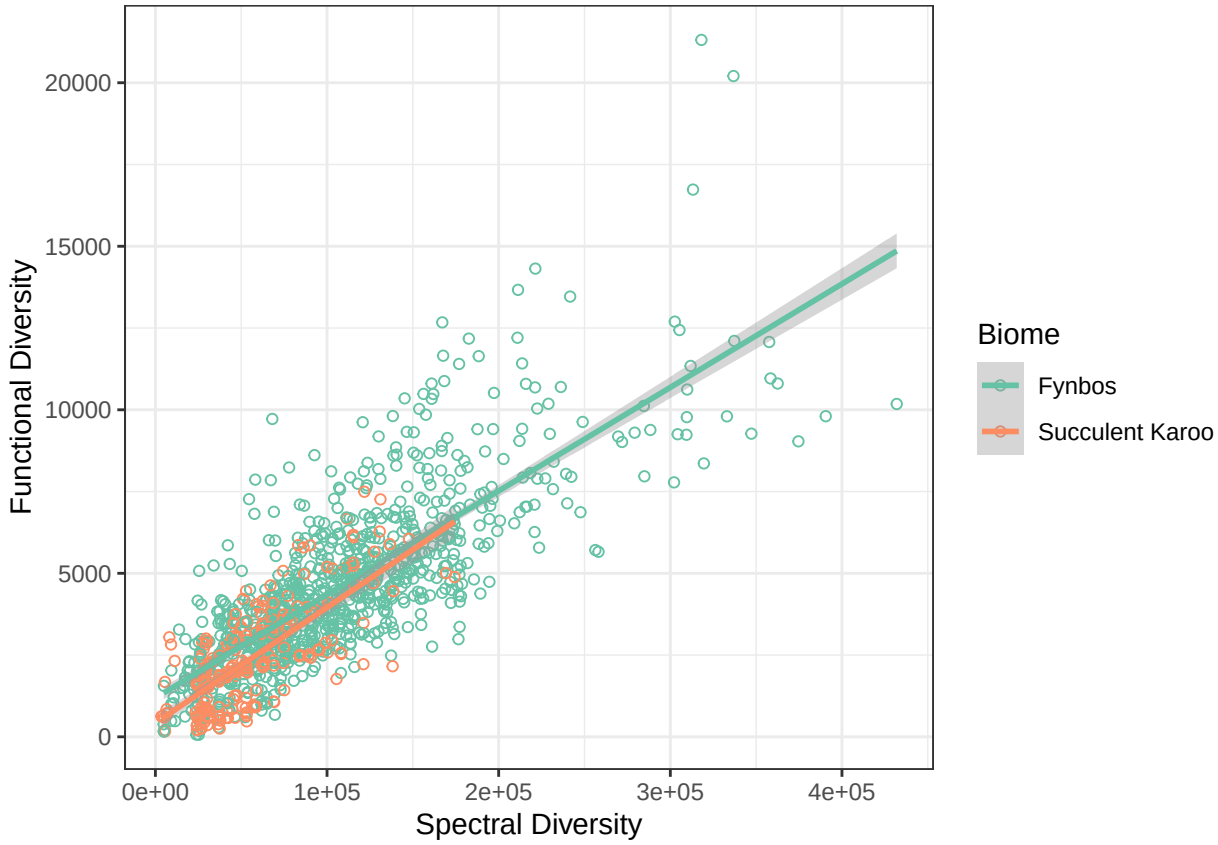

Figure 13: Unscaled version of spectral diversity predicting functional diversity within Fynbos and Succulent Karoo Biomes.

Phylogenetic diversity (Fig. 4C)

Relations of scaled variables

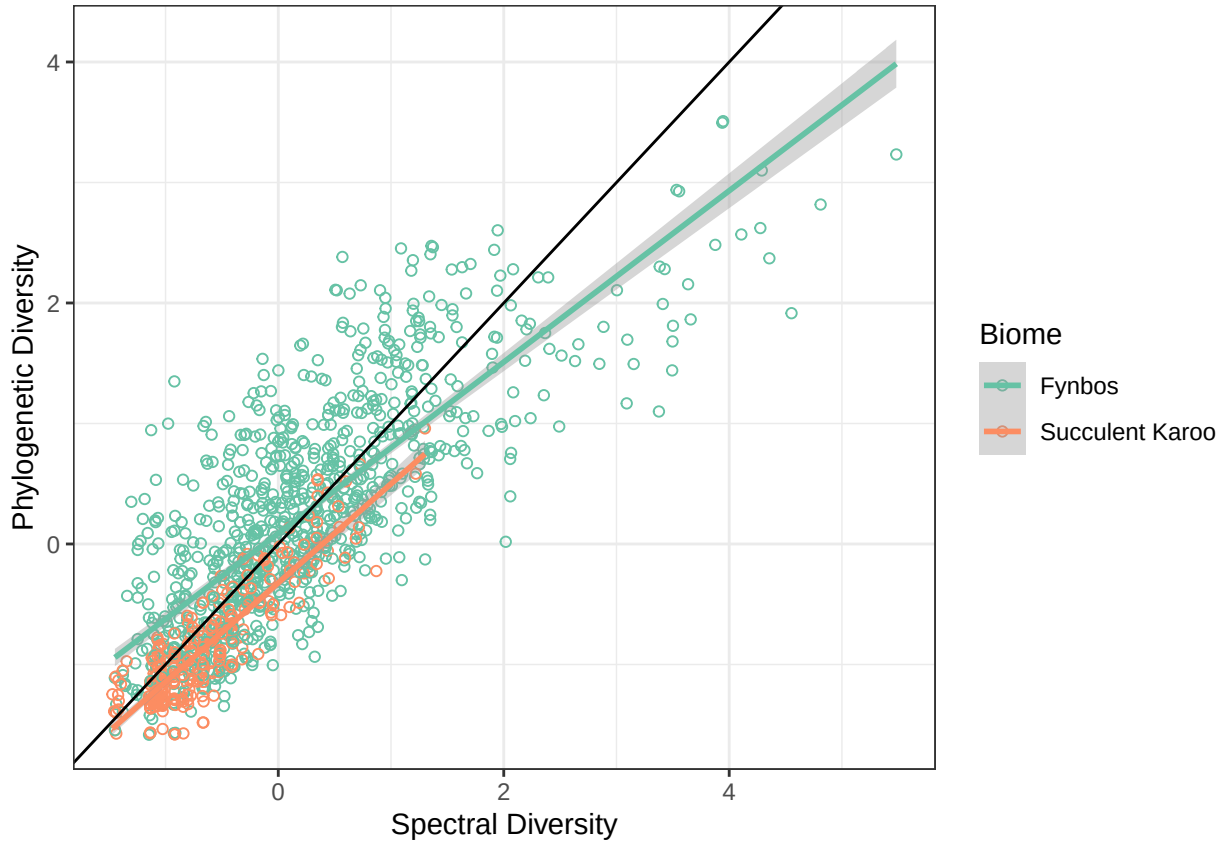

Figure 14: Spectral diversity predicting phylogenetic diversity within Fynbos and Succulent Karoo Biomes. Individual biomes can be turned on and off by clicking on them in the legend. Detailed information of each point is revealed when the mouse is hovered over the point. The black line is a 1:1 line representing maximal surrogacy.

Table 3: Model coefficients for spectral diversity predicting phylogenetic diversity within biomes

Term	Estimate	Standard Error	Test Statistic	P value	Significant
Intercept Reference (Fynbos)	0.088	0.018	4.971	0.000	Yes
Spectral Diversity Slope Reference (Fynbos)	0.711	0.017	41.993	0.000	Yes
Succulent Karoo Intercept	-0.411	0.056	-7.352	0.000	Yes
Succulent Karoo Slope	0.111	0.066	1.689	0.091	

```
## [1] "Adjusted R-squared:  0.705"
```

Relations of unscaled variables

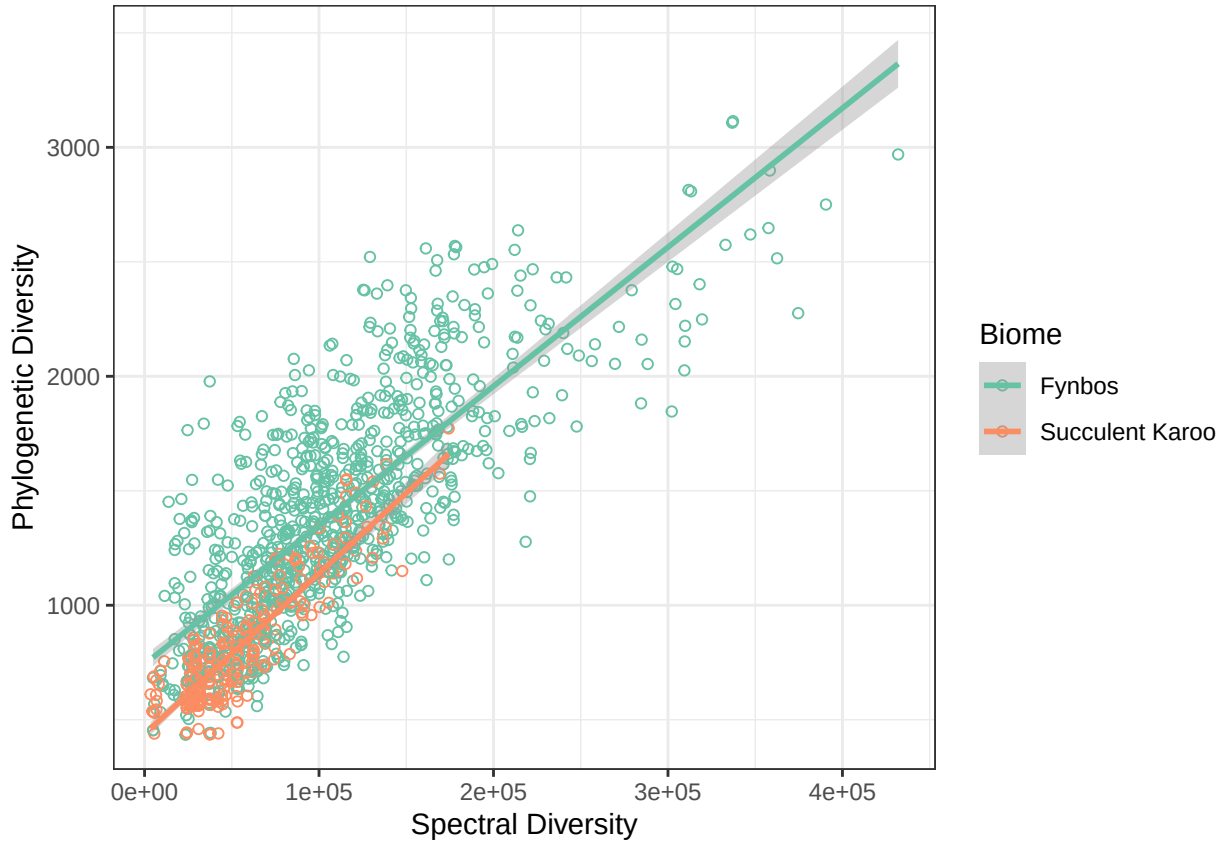

Figure 15: Unscaled version of spectral diversity predicting phylogenetic diversity within Fynbos and Succulent Karoo Biomes.

F. Subregion models

The following are model outputs and diagnostics for Figure 4D-F. Similar to the above, all models are Gaussian linear models and are assessed with standard diagnostics for model assumptions (i.e., normality of errors, homogeneity of variance). Predictor variables (i.e., spectral diversity) were centered to the mean and scaled to address the mismatch in scales between effective Shannon's diversity and spectral diversity. Because there were numerous subregion categories without a natural reference, we used sum-to-zero contrasts. In this situation, the first slope and intercept coefficients are the average slope and intercept across all groups. Individual subregions are then in reference to the average. For instance, in SI Table 4, the average intercept is -0.259 and the Cape Point intercept is -0.523 lower than the average for an absolute intercept of -0.782 (-0.259 + -0.523). Thus, a negative intercept or slope value suggests that the values are below the average intercept/slope, not that the values themselves are negative.

Effective Shannon's diversity (Fig. 4D)

Relations of scaled variables

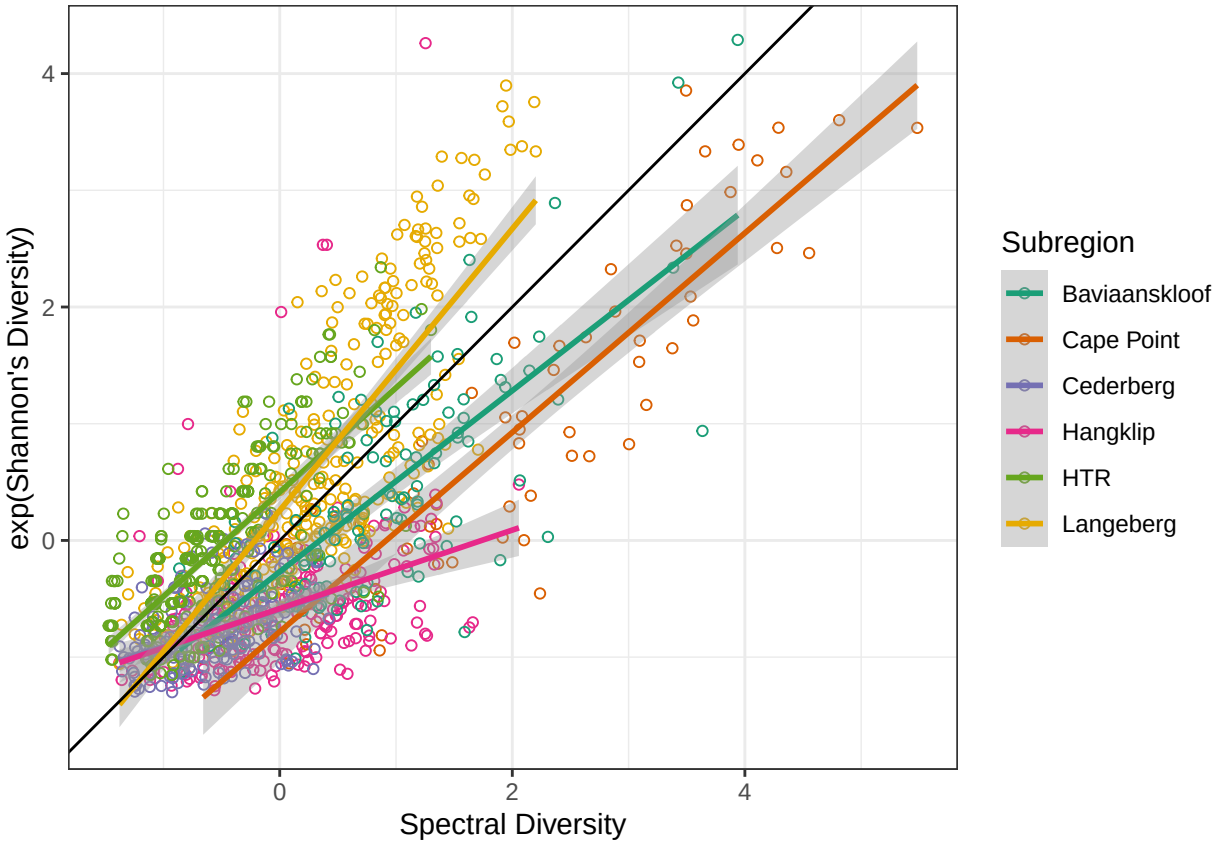

Figure 16: Spectral diversity predicting exponentiated Shannon's diversity within subregions. Individual subregions can be turned on and off by clicking on them. Detailed information of each point is revealed when the mouse is hovered over the point. The black line is a 1:1 line representing maximal surrogacy.

Table 4: Model coefficients for spectral diversity predicting exponentiated Shannon's diversity within subregions

Term	Estimate	Standard Error	Test Statistic	P value	Significant
Average Intercept for All	-0.259	0.028	-9.412	0.000	Yes
Average Slope for All	0.731	0.024	29.935	0.000	Yes
Baviaanskloof Intercept	-0.010	0.060	-0.163	0.870	
Cape Point Intercept	-0.523	0.100	-5.231	0.000	Yes
Cederberg Intercept	-0.330	0.063	-5.282	0.000	Yes
Hangklip Intercept	-0.325	0.040	-8.086	0.000	Yes
HTR Intercept	0.673	0.045	15.116	0.000	Yes
Baviaanskloof Slope	0.045	0.051	0.888	0.375	
Cape Point Slope	0.123	0.045	2.745	0.006	Yes
Cederberg Slope	-0.412	0.082	-5.012	0.000	Yes
Hangklip Slope	-0.395	0.049	-8.037	0.000	Yes
HTR Slope	0.163	0.049	3.316	0.001	Yes

```
## [1] "Adjusted R-squared:  0.715"
```

Relations of unscaled variables

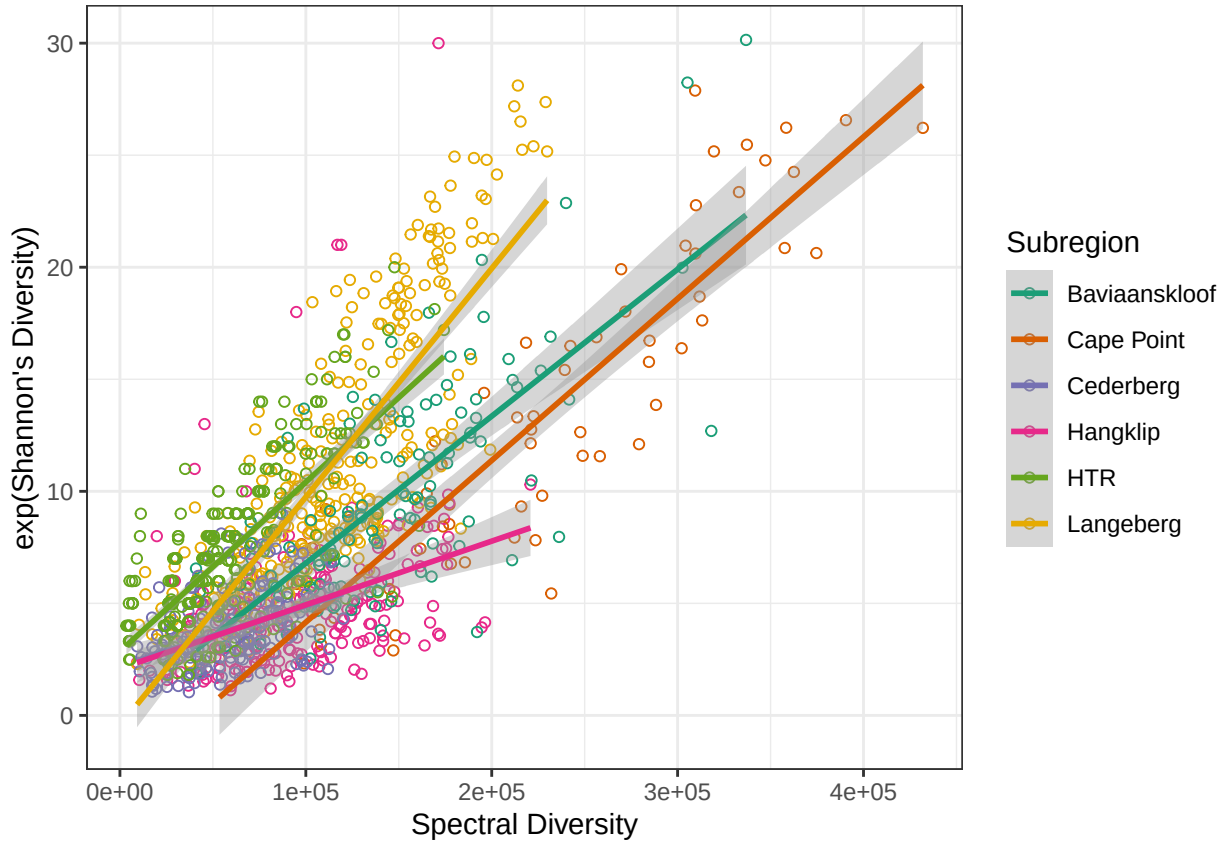

Figure 17: Unscaled version of spectral diversity predicting exponentiated Shannon's diversity within subregions.

Functional diversity (Fig. 4E)

Relations of scaled variables

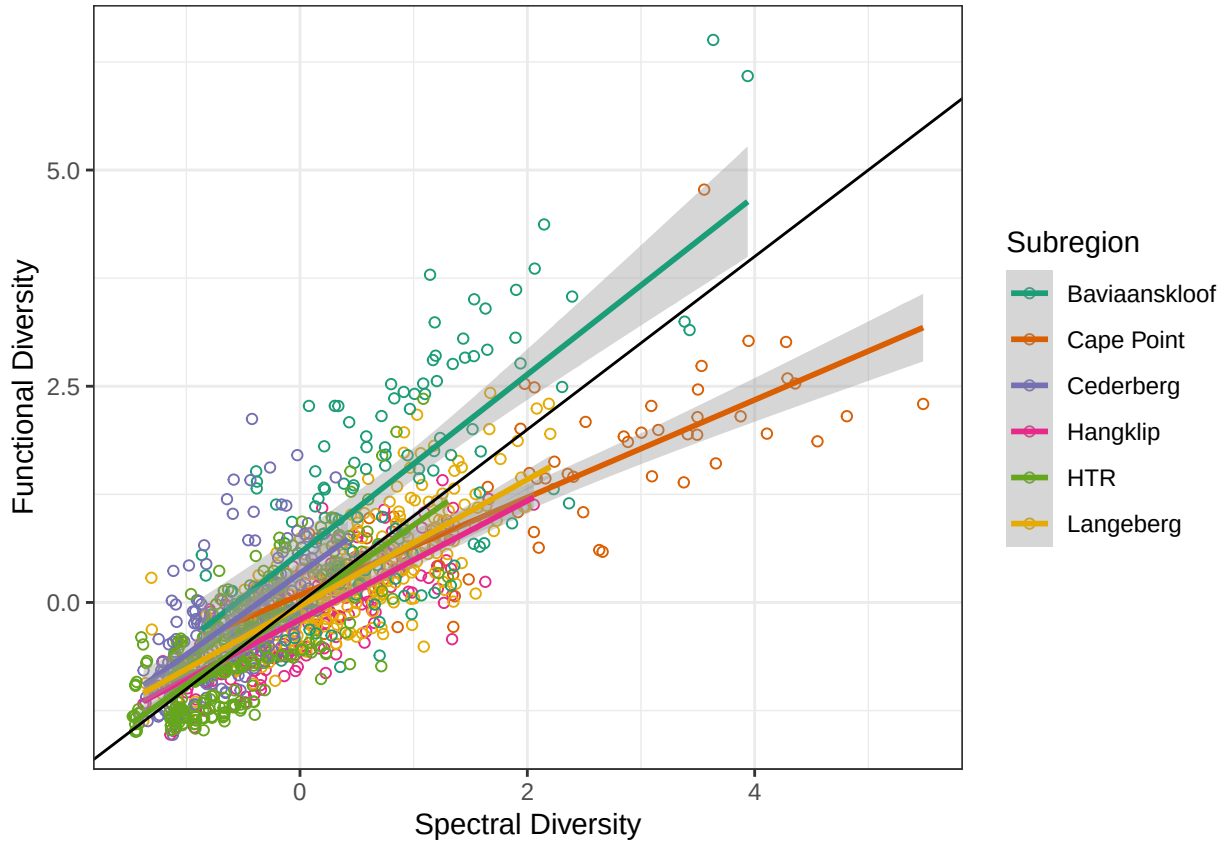

Figure 18: Spectral Diversity predicting functional diversity within subregions. Individual subregions can be turned on and off by clicking on them. Detailed information of each point is revealed when the mouse is hovered over the point. The black line is a 1:1 line representing maximal surrogacy.

Table 5: Model coefficients for spectral diversity predicting functional diversity

Term	Estimate	Standard Error	Test Statistic	P value	Significant
Average Intercept for All	0.122	0.027	4.587	0.000	Yes
Average Slope for All	0.813	0.024	34.412	0.000	Yes
Baviaanskloof Intercept	0.452	0.058	7.751	0.000	Yes
Cape PointIntercept	-0.035	0.097	-0.367	0.714	
Cederberg Intercept	0.216	0.060	3.567	0.000	Yes
Hangklip Intercept	-0.319	0.039	-8.206	0.000	Yes
HTR Intercept	-0.150	0.043	-3.494	0.000	Yes
Baviaanskloof Slope	0.218	0.049	4.464	0.000	Yes
Cape Point Slope	-0.249	0.043	-5.738	0.000	Yes
Cederberg Slope	0.128	0.079	1.609	0.108	
Hangklip Slope	-0.126	0.047	-2.653	0.008	Yes
HTR Slope	0.110	0.047	2.329	0.020	Yes

```
## [1] "Adjusted R-squared:  0.733"
```

Relations of unscaled variables

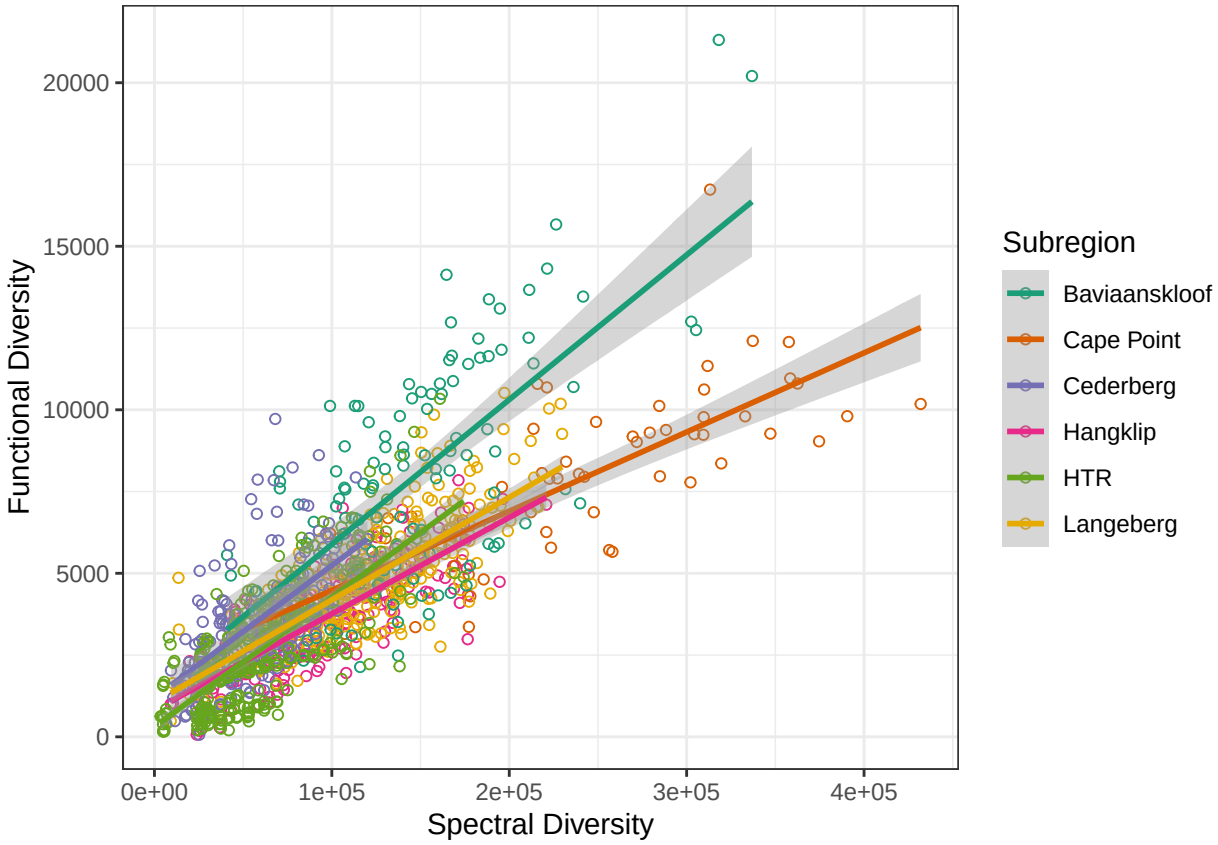

Figure 19: Unscaled version of spectral Diversity predicting functional diversity within subregions.

Phylogenetic diversity (Fig. 4F)

Relations of scaled variables

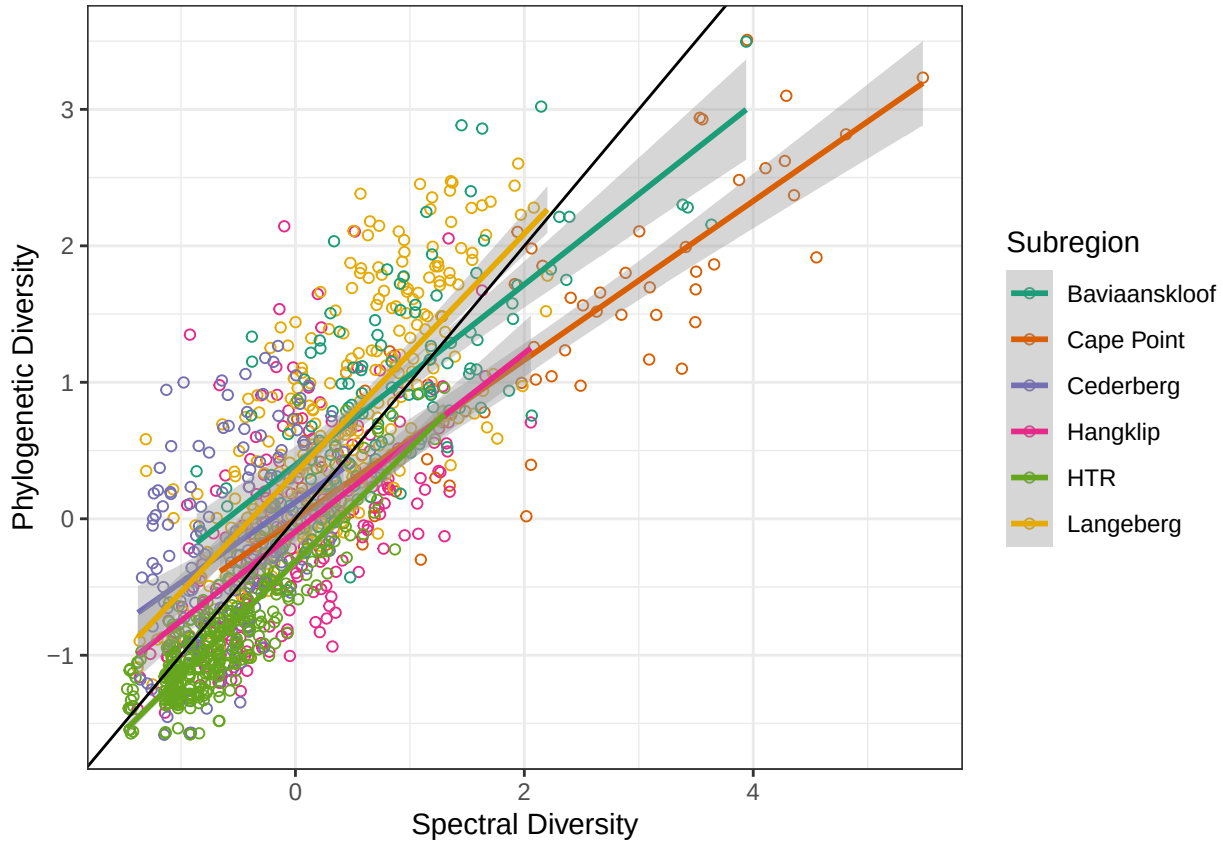

Figure 20: Spectral diversity predicting phylogenetic diversity within subregions. Individual subregions can be turned on and off by clicking on them. Detailed information of each point is revealed when the mouse is hovered over the point. The black line is a 1:1 line representing maximal surrogacy.

Table 6: Model coefficients for spectral diversity predicting phylogenetic diversity within subregions

Term	Estimate	Standard Error	Test Statistic	P value	Significant
Average Intercept for All	0.074	0.025	2.940	0.003	Yes
Average Slope for All	0.699	0.022	31.301	0.000	Yes
Baviaanskloof Intercept	0.321	0.055	5.825	0.000	Yes
Cape PointIntercept	-0.076	0.091	-0.835	0.404	
Cederberg Intercept	0.048	0.057	0.847	0.397	
Hangklip Intercept	-0.168	0.037	-4.573	0.000	Yes
HTR Intercept	-0.389	0.041	-9.555	0.000	Yes
Baviaanskloof Slope	-0.038	0.046	-0.812	0.417	
Cape Point Slope	-0.116	0.041	-2.830	0.005	Yes
Cederberg Slope	-0.110	0.075	-1.461	0.144	
Hangklip Slope	-0.045	0.045	-1.004	0.315	
HTR Slope	0.131	0.045	2.926	0.003	Yes

```
## [1] "Adjusted R-squared:  0.762"
```

Relations of unscaled variables

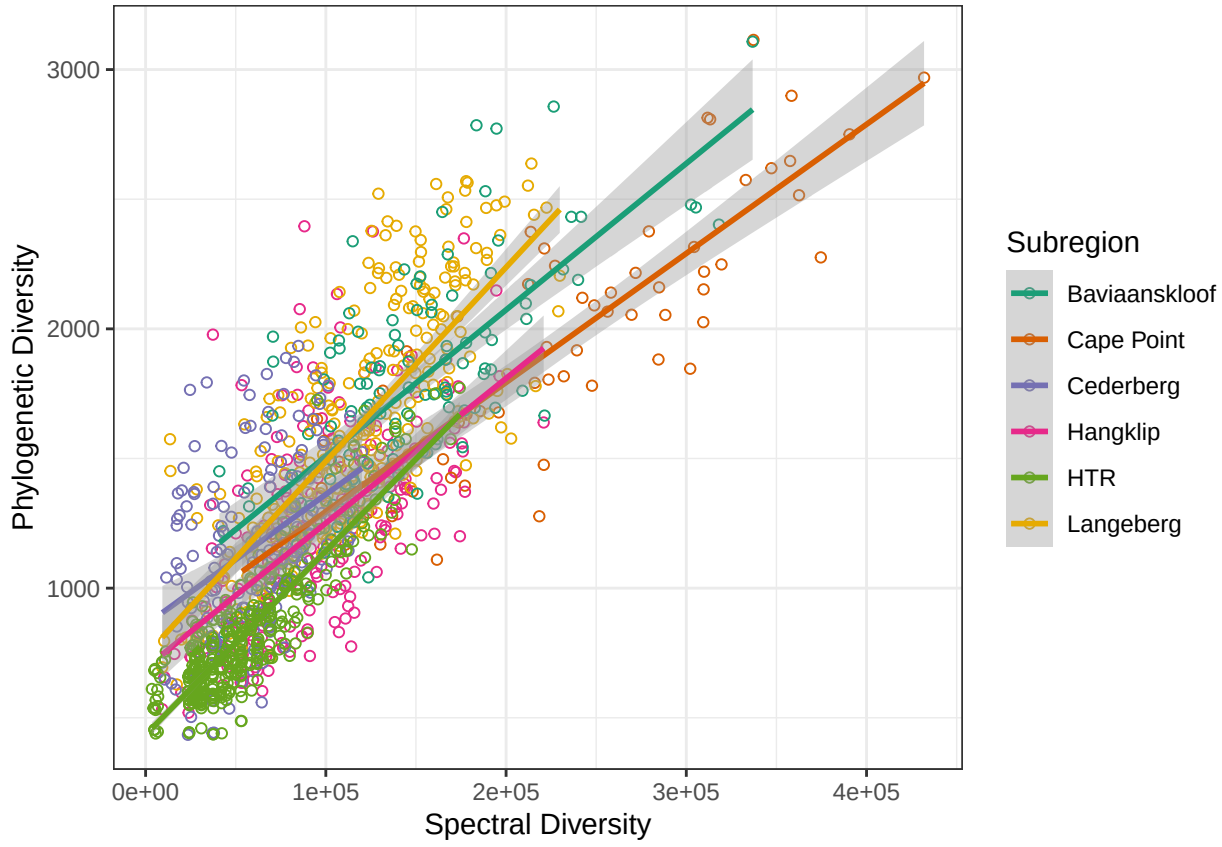

Figure 21: Unscaled version of spectral diversity predicting phylogenetic diversity within subregions.

G. Family dominance

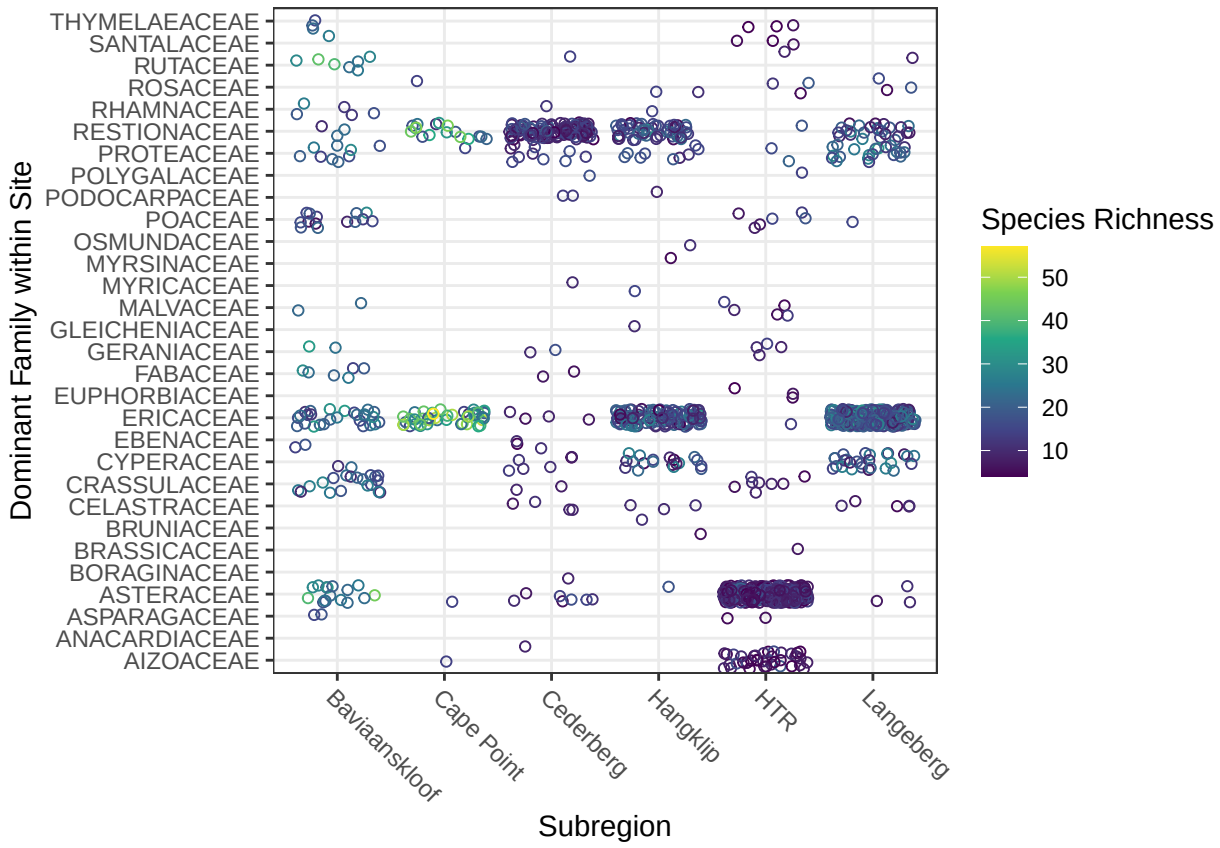

Figure 22: Dominant families within a site by each subregion and species richness. Family dominance was determined by the family with the highest abundance in a site.

Table 7: Number of plots in which a particular family was single the most dominant family, split by subregion.

Dominant Family	Baviannskloof	Cape Point	Cederberg	Hangklip	HTR	Langeberg	Total
AIZOACEAE	NA	1	NA	NA	40	NA	41
ANACARDIACEAE	NA	NA	1	NA	NA	NA	1
ASPARAGACEAE	2	NA	NA	NA	2	NA	4
ASTERACEAE	16	1	7	1	310	3	338
BORAGINACEAE	NA	NA	1	NA	NA	NA	1
BRASSICACEAE	NA	NA	NA	NA	1	NA	1
BRUNIACEAE	NA	NA	NA	2	NA	NA	2
CELASTRACEAE	NA	NA	4	3	NA	5	12
CRASSULACEAE	19	NA	2	NA	8	NA	29
CYPERACEAE	2	NA	6	19	NA	27	54
EBENACEAE	2	NA	3	NA	NA	NA	5
ERICACEAE	27	45	4	126	1	216	419
EUPHORBIACEAE	NA	NA	NA	NA	3	NA	3
FABACEAE	6	NA	2	NA	NA	NA	8
GERANIACEAE	2	NA	2	NA	4	NA	8
GLEICHENIACEAE	NA	NA	NA	1	NA	NA	1
MALVACEAE	2	NA	NA	NA	5	NA	7
MYRICACEAE	NA	NA	1	1	NA	NA	2
MYRSINACEAE	NA	NA	NA	1	NA	NA	1
OSMUNDACEAE	NA	NA	NA	1	NA	NA	1
POACEAE	14	NA	NA	NA	6	1	21
PODOCARPACEAE	NA	NA	2	1	NA	NA	3
POLYGALACEAE	NA	NA	1	NA	1	NA	2
PROTEACEAE	9	1	10	9	2	21	52
RESTIONACEAE	3	17	114	55	1	25	215
RHAMNACEAE	5	NA	1	1	NA	NA	7
ROSACEAE	NA	1	NA	2	3	3	9
RUTACEAE	7	NA	1	NA	NA	1	9
SANTALACEAE	1	NA	NA	NA	4	NA	5
THYMELAEACEAE	3	NA	NA	NA	3	NA	6

H. Biome redundancy (Fig. 5A-C)

Functional redundancy

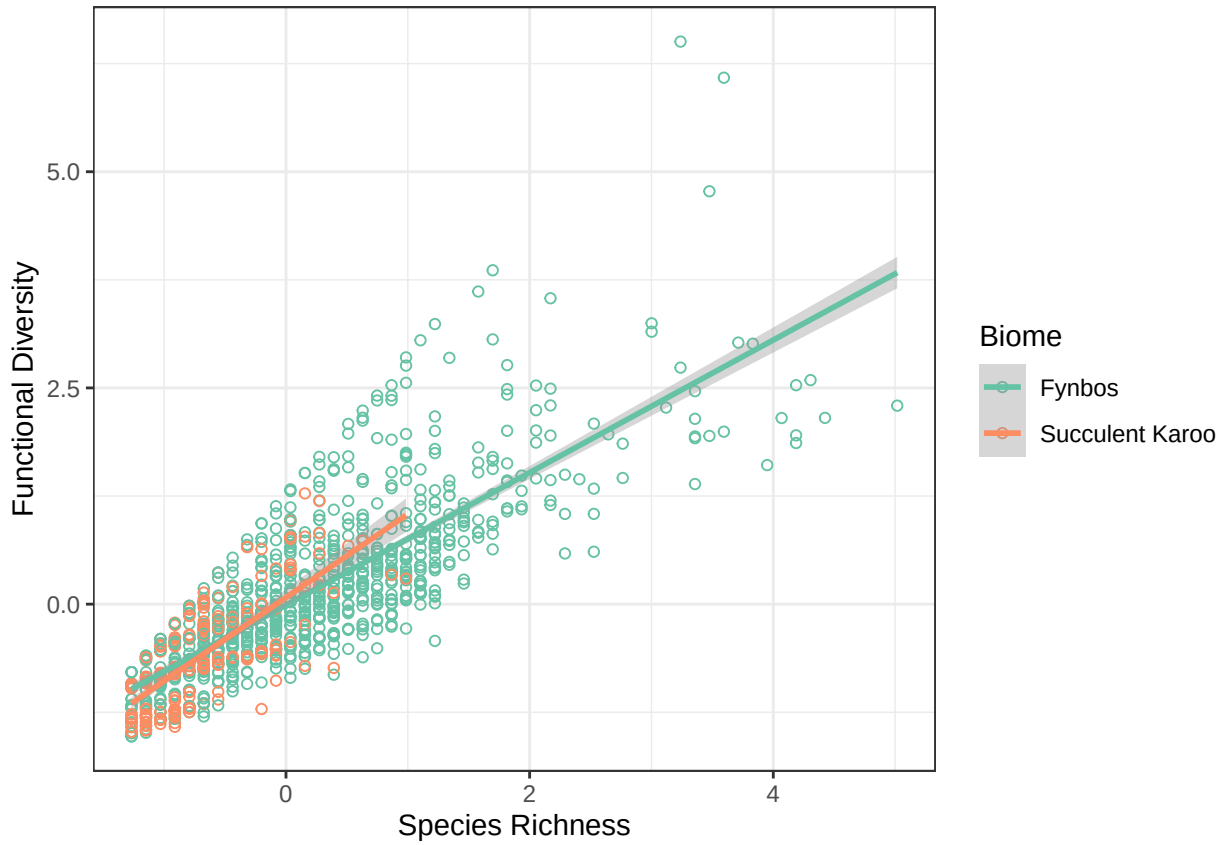

Figure 23: Species richness as a predictor of functional diversity within biomes. This graph explores the degree to which biomes differ based on functional redundancy. Individual biomes can be turned on and off by clicking on them. Detailed information of each point is revealed when the mouse is hovered over the point.

Table 8: Model coefficients for species richness predicting functional diversity within biomes

Term	Estimate	Standard Error	Test Statistic	P value	Significant
Intercept Reference (Fynbos)	-0.012	0.018	-0.629	0.529	
Slope Reference (Fynbos)	0.766	0.018	42.940	0.000	Yes
Succulent Karoo Intercept	0.088	0.071	1.248	0.212	
Succulent Karoo Slope	0.199	0.078	2.552	0.011	Yes

```
## [1] "Adjusted R-squared:  0.68"
```

Phylogenetic redundancy

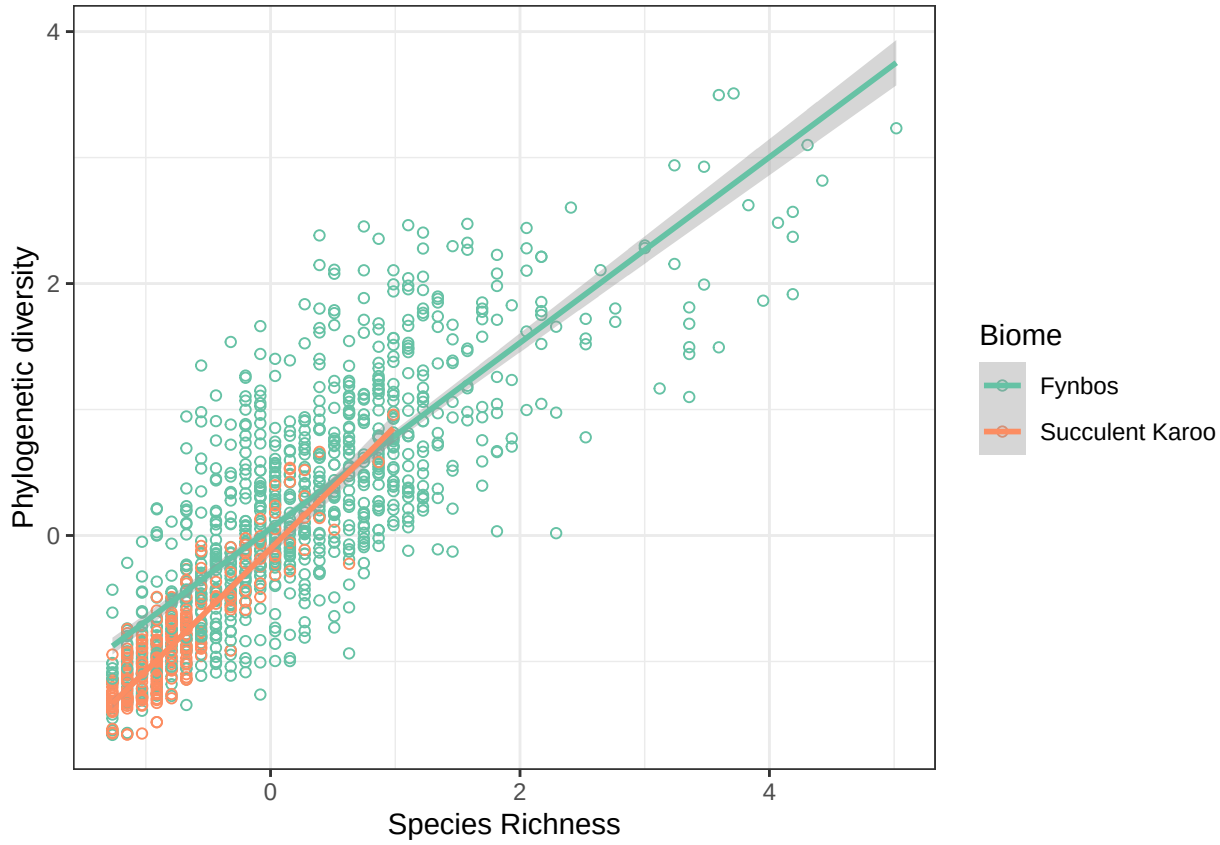

Figure 24: Species richness as a predictor of phylogenetic diversity within biomes. This graph explores the degree to which biomes differ based on phylogenetic redundancy. Individual biomes can be turned on and off by clicking on them. Detailed information of each point is revealed when the mouse is hovered over the point.

Table 9: Model coefficients for species richness predicting phylogenetic diversity within biomes

Term	Estimate	Standard Error	Test Statistic	P value	Significant
Intercept Reference (Fynbos)	0.056	0.018	3.185	0.001	Yes
Slope Reference (Fynbos)	0.737	0.017	43.479	0.000	Yes
Succulent Karoo Intercept	-0.160	0.067	-2.393	0.017	Yes
Succulent Karoo Slope	0.233	0.074	3.160	0.002	Yes

```
## [1] "Adjusted R-squared:  0.718"
```

Spectral redundancy

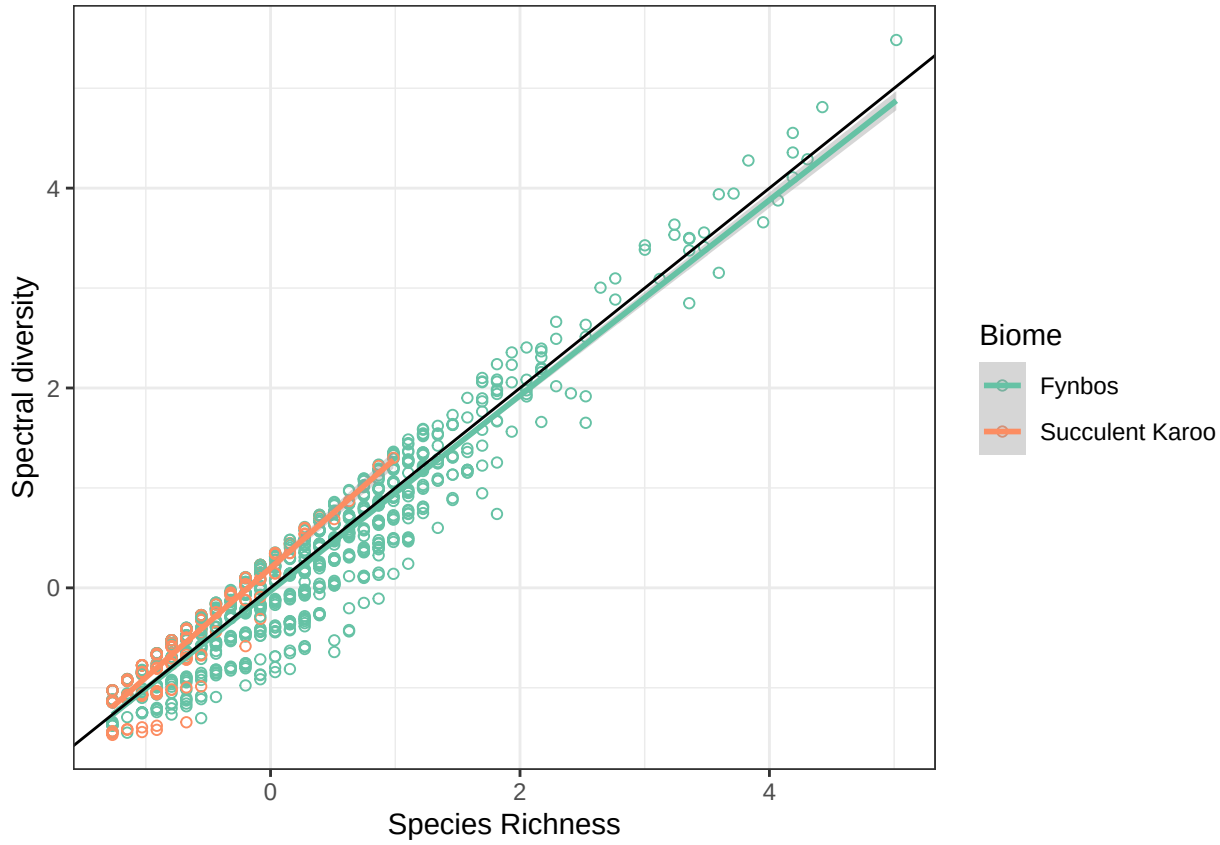

Figure 25: Species richness as a predictor of spectral diversity within biomes. This graph explores the degree to which biomes differ based on spectrl redundancy. Individual biomes can be turned on and off by clicking on them. Detailed information of each point is revealed when the mouse is hovered over the point. The black line is a 1:1 line representing maximal surrogacy.

Table 10: Model coefficients for species richness predicting spectral diversity within biomes

Term	Estimate	Standard Error	Test Statistic	P value	Significant
Intercept Reference (Fynbos)	-0.031	0.009	-3.261	0.001	Yes
Slope Reference (Fynbos)	0.978	0.009	106.839	0.000	Yes
Succulent Karoo Intercept	0.229	0.036	6.338	0.000	Yes
Succulent Karoo Slope	0.118	0.040	2.967	0.003	Yes

```
## [1] "Adjusted R-squared:  0.923"
```

I. Biome Evenness Comparison

To provide possible explanations in the differences in surrogacy patterns between biomes we compare Shannon's species evenness between the Fynbos and Succulent Karoo Biomes. As expected, the Succulent Karoo is far more even in it's species composition compared to the Fynbos.

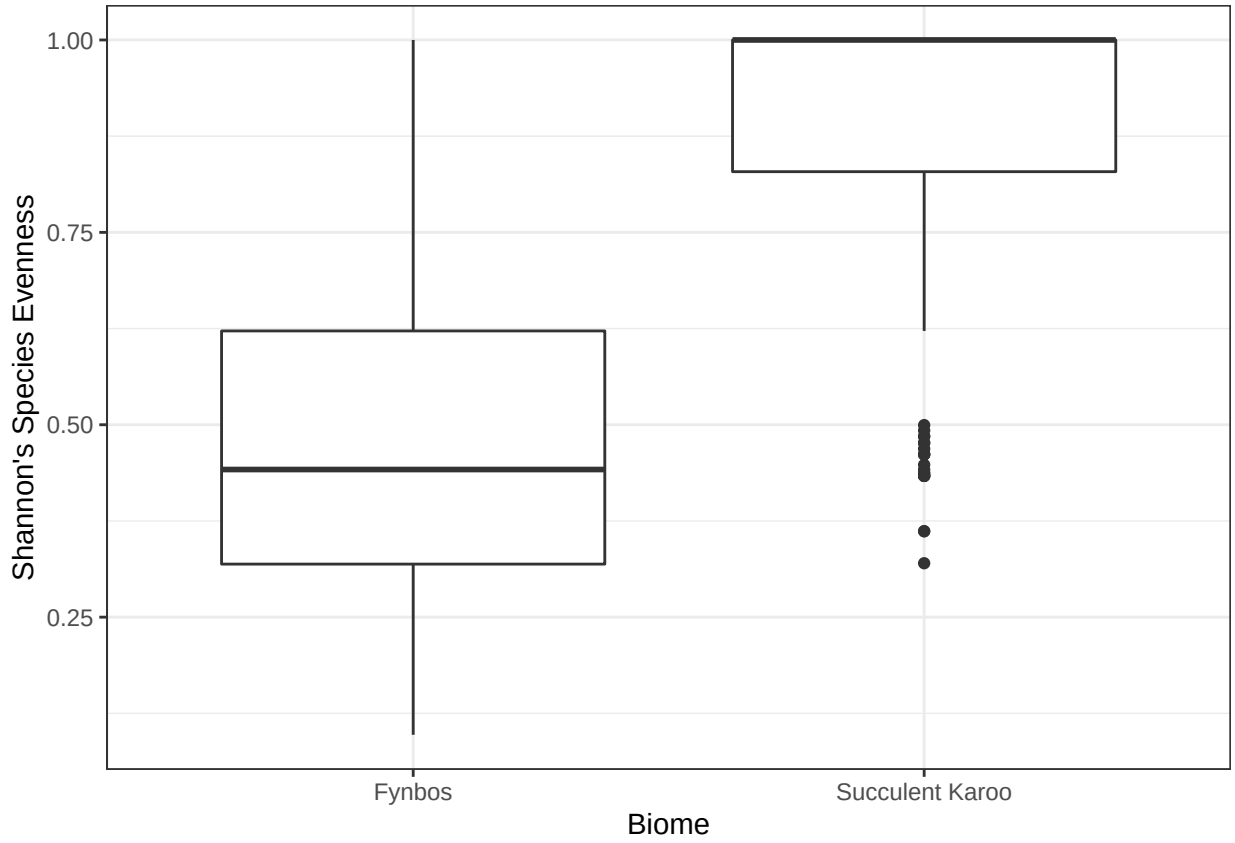

Figure 26: Comparison of Shannon's species evenness between biomes.

J. Family spectra and trait differences

Certain "iconic" families within the Greater Cape Florisitc Region are predominant in different areas and biomes. Where these families radiated is likely a factor of both environment and their ability to adapt to those environments constrained by their previous ancestry. In aggregrate, this leads to differences in average community traits and therefore spectra. Below we show the differences in spectral reflectance and traits used to calculate functional diversity between four of these families: Proteaceae, Ericaceae, Restionaceae, and Aizoaceae. The former three are characteristic of Fynbos while the latter is common and abundant in the Succulent Karoo.

Spectra of four iconic families

Families differ widely in their spectral reflectance suggesting evidence of a mechanistic underpinning between environment, community membership and ultimately biodiversity. For instance, the Aizoaceae are succulent and adapted to arid environments such as the Succulent Karoo. The Aizoaceae have a lower percent reflectance in the near infrared range (750-900 nm), a region of the spectral signature commonly associated with water content and leaf structure. On the other hand, the Proteaceae, which are typically broad leaved shrubs in the Fynbos, show more absorption (lower reflectance) of photosynthetically active radiation in the red and blue regions (and larger green hump), suggesting higher amounts of photosynthetically active pigments than other families .

Proteaceae show more absorption of photosynthetically active radiation in the red and blue (and a more pronounced green hump) indicating higher photosynthetic pigment content than other families.

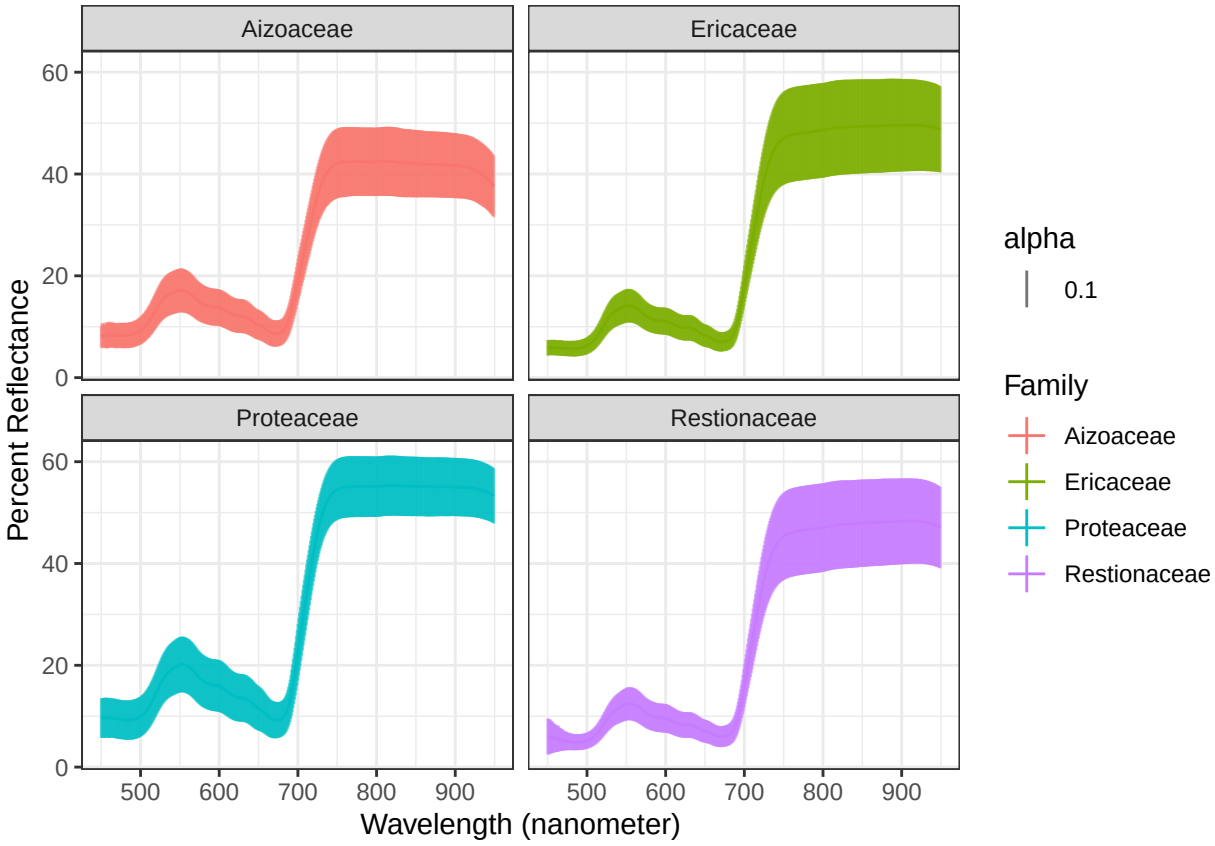

Figure 27: Comparison of mean family spectral reflectance surrounded by one standard deviatrion.

Intraspecific spectral variation of a single species

Due to the high species turnover in the GCFR, few species are found across multiple subregions in our data. We show the spectral reflectance of one species, *Leucadendron salignum*, which appeared in five of six subregions. There is substantial variation in both the visible and near infrared range suggesting that species can vary in the spectral reflectance across subregions.

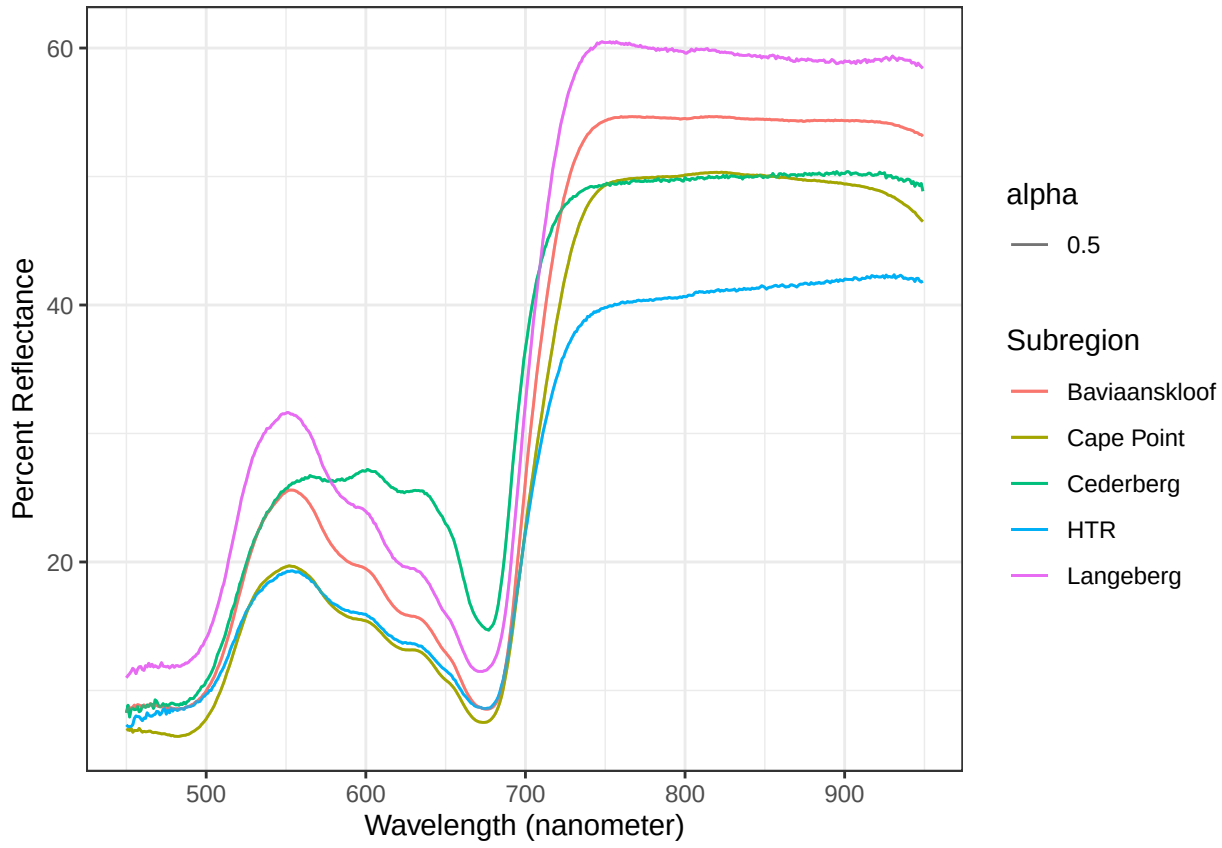

Figure 28: Spectral intra-specific variation within Leucadendron salignum across five subregions.

Trait variation of four iconic families

A key to understanding how plant spectra vary along landscapes is to understand how traits vary along landscapes. While these questions are outside of the scope of this study, we show trait differences between four iconic GCFR families, to provide motivation in how traits will provide an important context in interpreting results.

Leaf mass per area

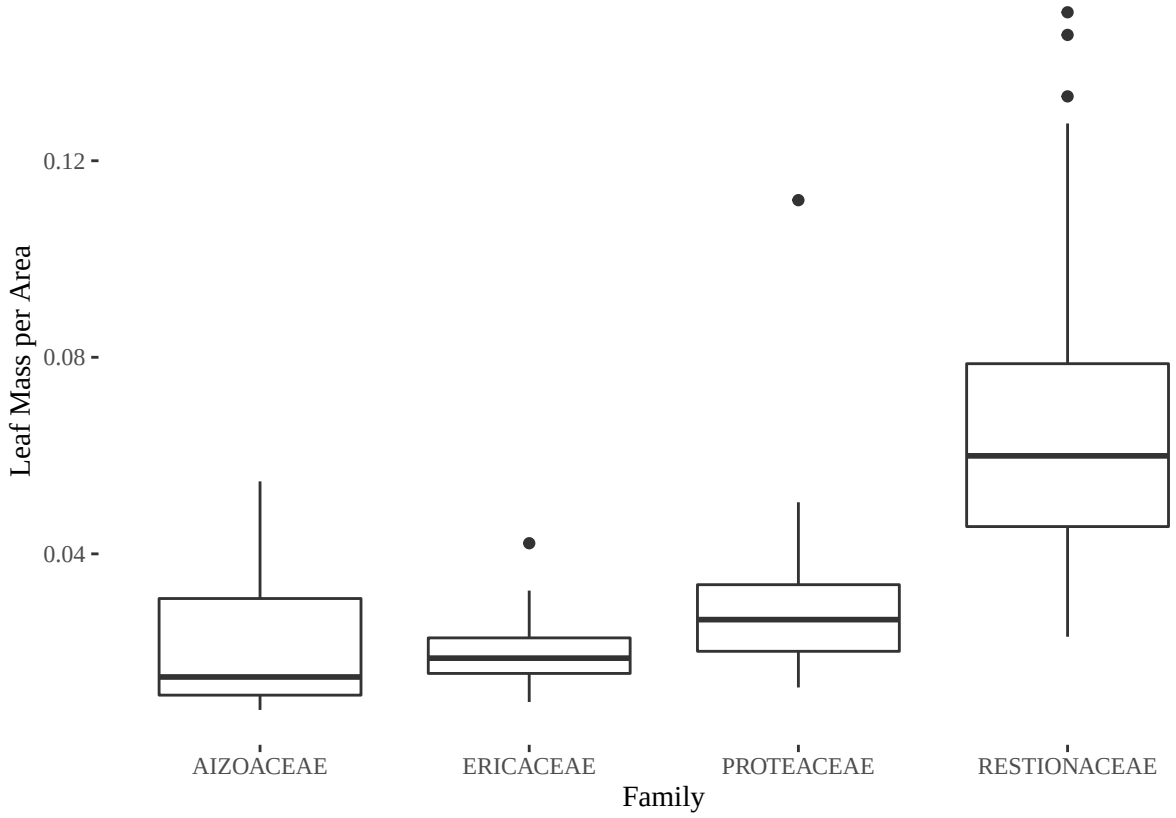

Figure 29: Comparison of leaf mass per area between iconic families.

Leaf Water Content

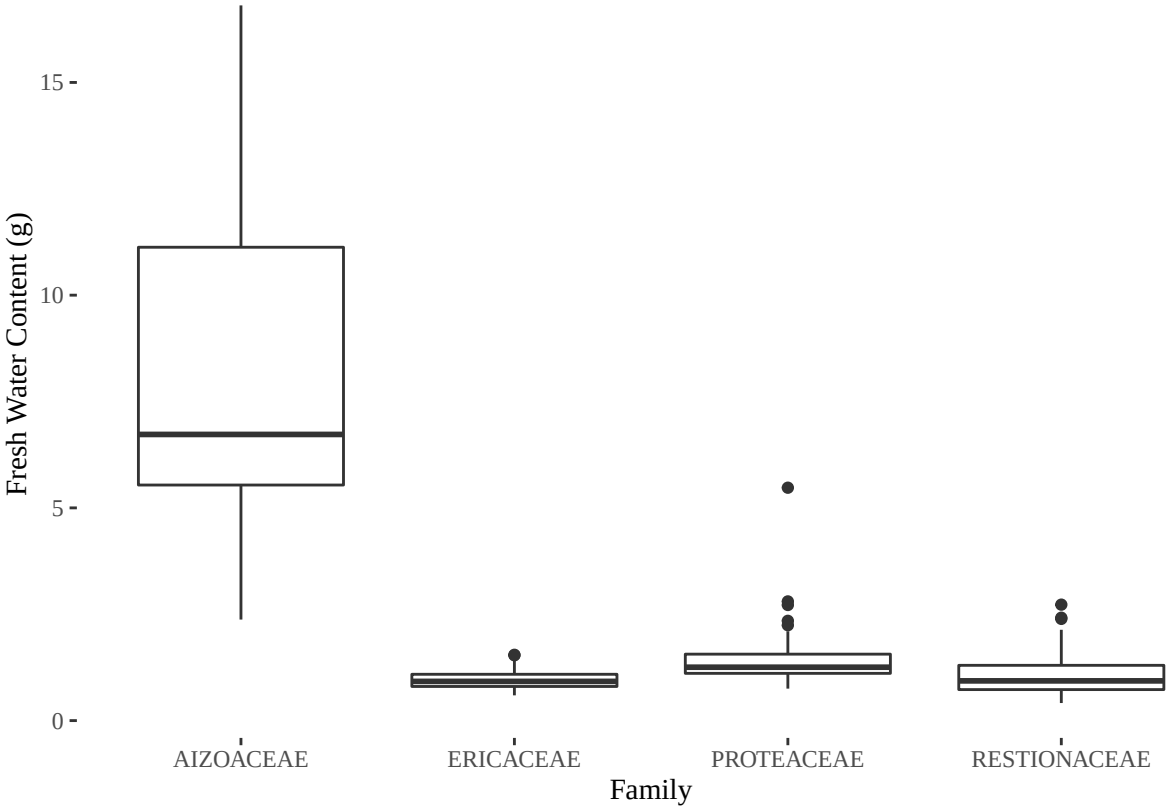

Figure 30: Comparison of leaf fresh water content between iconic families.

Percent nitrogen

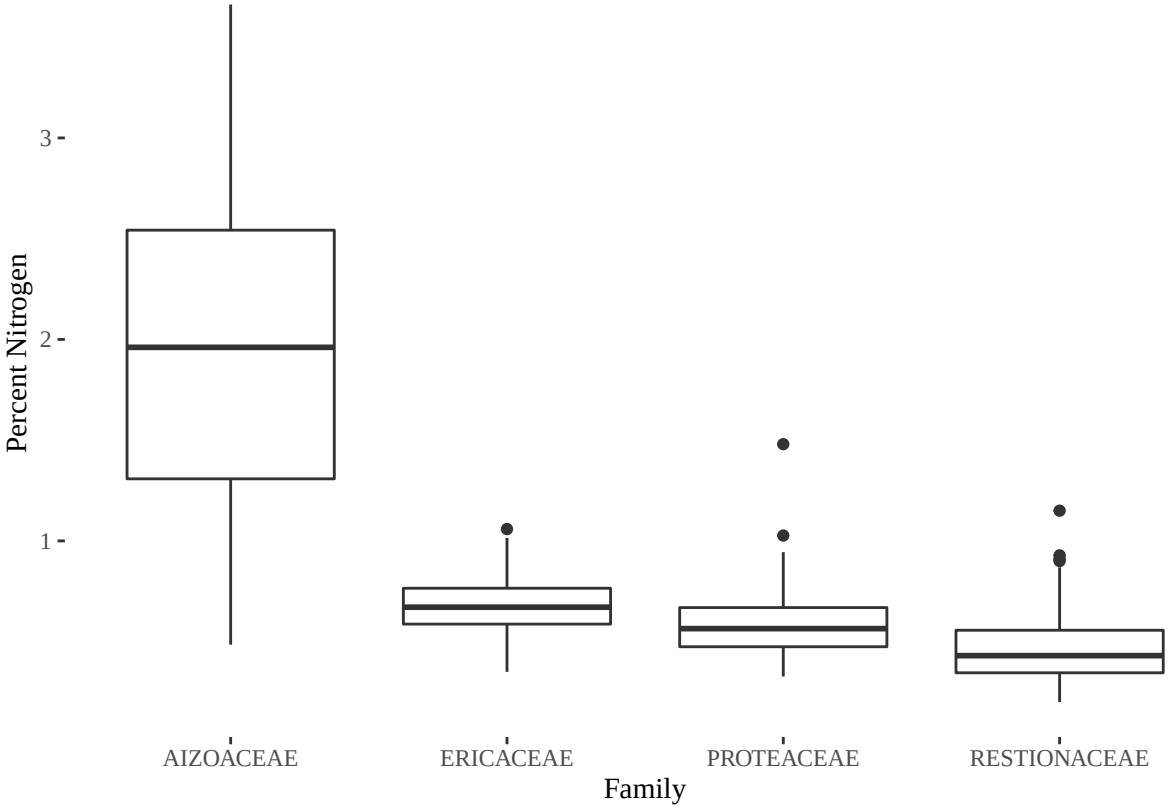

Figure 31: Comparison of percent nitrogen content between iconic families.

Trait variation between subregions

Similar to the section above, we show that traits (not abundance weighted) vary among subregions.

Leaf mass per area

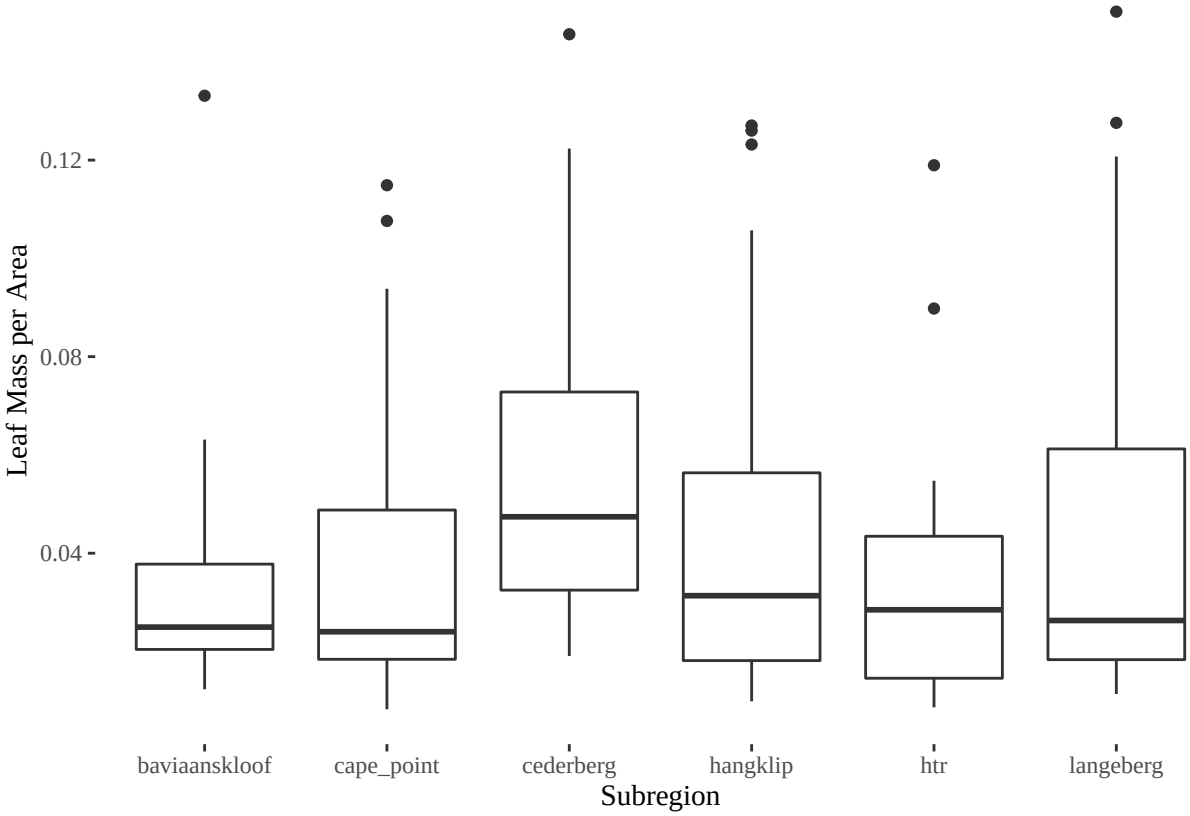

Figure 32: Comparison of leaf mass per area between subregions.

Leaf Water Content

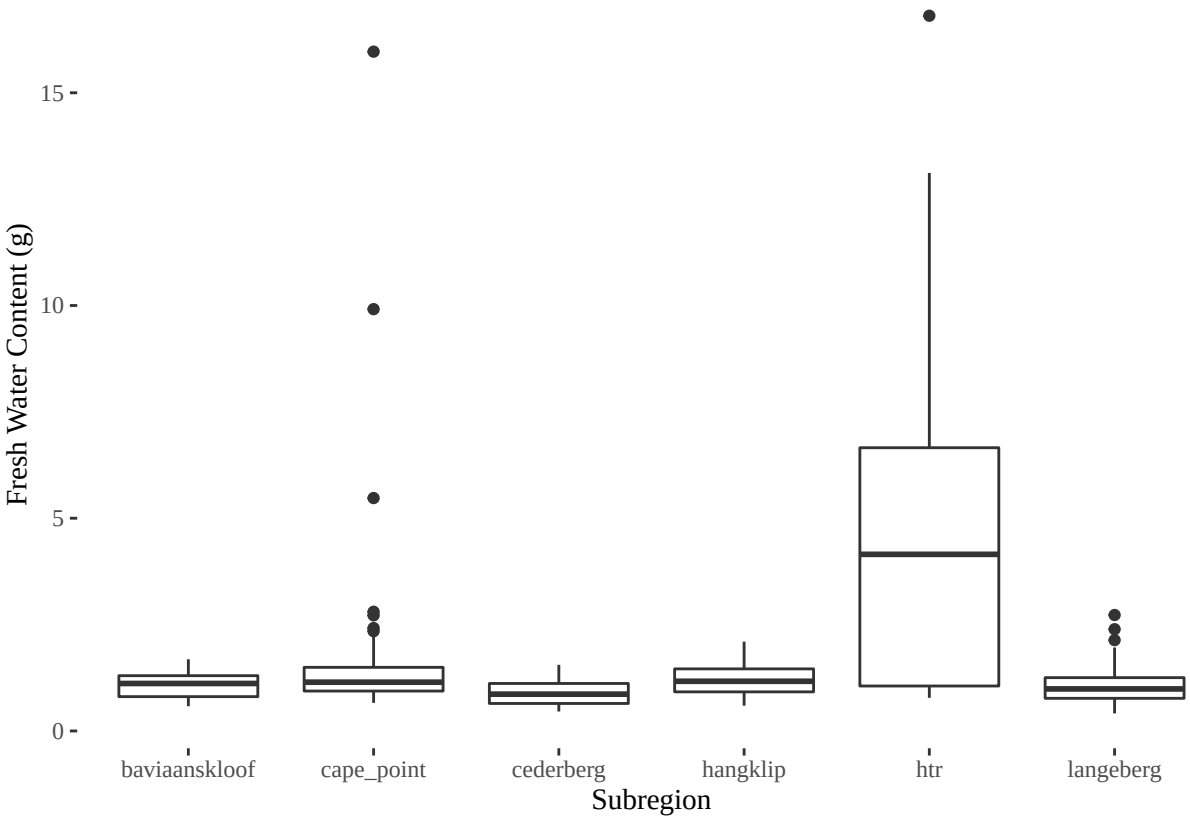

Figure 33: Comparison of leaf fresh water content between subregions.

Percent nitrogen

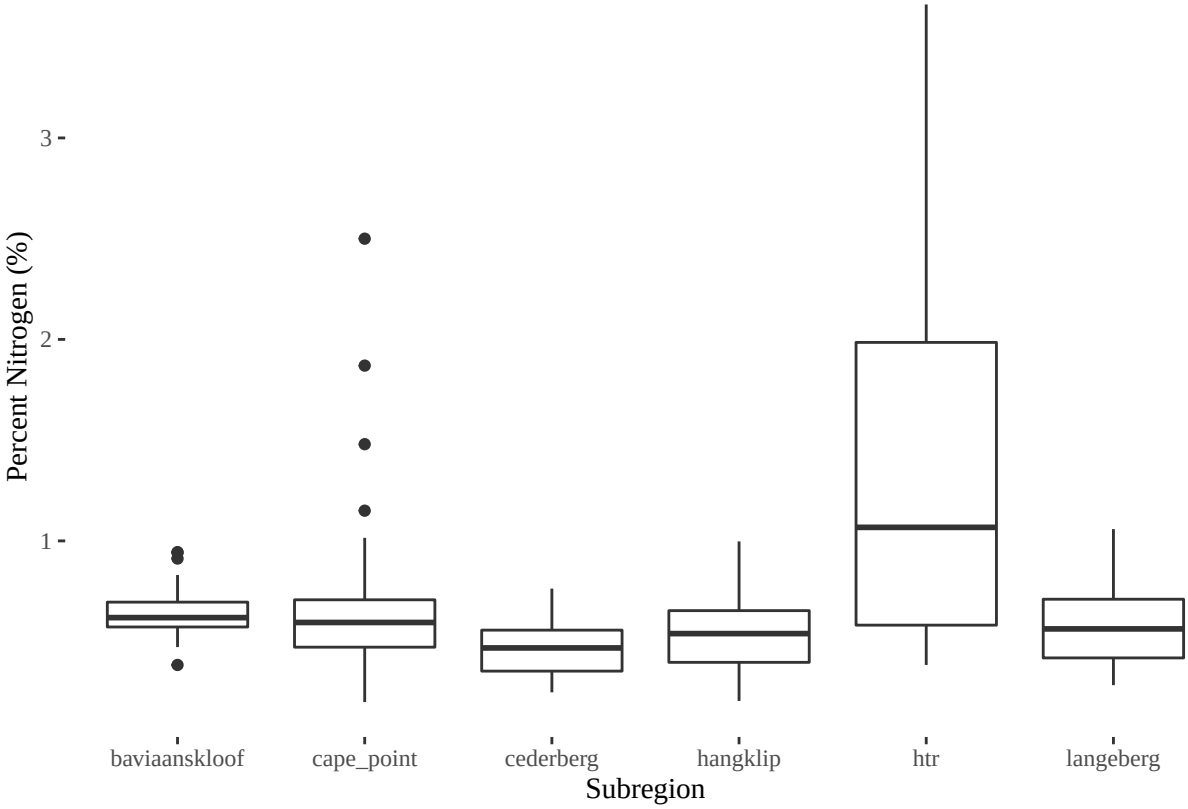

Figure 34: Comparison of percent nitrogen content between subregions.

Bibliography

Chang, Chein-I. 2000. "An Information-Theoretic Approach to Spectral Variability, Similarity, and Discrimination for Hyperspectral Image Analysis." *IEEE Transactions on Information Theory* 46 (5): 1927–32.

Dahlin, Kyla Marie. 2016. "Spectral Diversity Area Relationships for Assessing Biodiversity in a Wildland-Agriculture Matrix." *Ecological Applications* 26 (8): 2758–68. https://doi.org/10.1002/eap.1390.

Gholizadeh, Hamed, John A Gamon, Arthur I Zygielbaum, Ran Wang, Anna K Schweiger, and Jeannine Cavender-Bares. 2018. "Remote Sensing of Biodiversity: Soil Correction and Data Dimension Reduction Methods Improve Assessment of -Diversity (Species Richness) in Prairie Ecosystems." *Remote Sensing of Environment* 206: 240–53.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. "Resampling Methods." In *An Introduction to Statistical Learning*, 175–201. Springer.

Kruse, Fred A, AB Lefkoff, JW Boardman, KB Heidebrecht, AT Shapiro, PJ Barloon, and AFH Goetz. 1993. "The Spectral Image Processing System (SIPS)—Interactive Visualization and Analysis of Imaging Spectrometer Data." *Remote Sensing of Environment* 44 (2-3): 145–63.

Petchey, Owen L, and Kevin J Gaston. 2002. "Functional Diversity (FD), Species Richness and Community Composition." *Ecology Letters* 5 (3): 402–11.

Roberts, David R, Volker Bahn, Simone Ciuti, Mark S Boyce, Jane Elith, Gurutzeta Guillera-Arroita, Severin Hauenstein, et al. 2017. "Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure." *Ecography* 40 (8): 913–29.

Villéger, Sébastien, Norman WH Mason, and David Mouillot. 2008. "New Multidimensional Functional Diversity Indices for a Multifaceted Framework in Functional Ecology." *Ecology* 89 (8): 2290–2301.