Evolutionary Distinctiveness vs Ecological Uniqueness

Exploring the effect of tree and evolutionary model parameters

Giulio Valentino Dalla Riva

September 20, 2016

Contents

Introduction

We want to understand better if and how Evolutionary Distinctiveness is related with Ecological Uniqueness. The first is defined here [cite]; we define the second as a measure of isolation in a (continuous) trait species. We explore the effect of different phylogenetic parameters (simulating tree across a grid of such parameters), evolutionary model parameters (simulating traits from the Brownian Motion model and its punctuated and age dependent extensions), trait space dimensionality and different definitions of isolation.

Preliminaries

We need a parallel environment. Let's build it.

```
library(doMC)
doMC::registerDoMC(cores=32)
```

We simulate ultrametric tree using Liam Revell's `phytools` package. We work within the tidy data framework.

```
install.packages("pacman")
library(pacman)
p_load("devtools","plyr","tidyverse","modelr","phytools","picante","broom","viridis")
install_github("dgrtwo/widyr")
library(widyr)

source("R/trajectories_bm.R")
source("R/plot_trajectory.R")
source("R/core_functions.R")
```

We start with a simplified version of the analytic flow, keeping fixed most of the parameters.

Simulate trees

First, a list of 100(0, you name it) phylogenies, constrained to t = Time, using N_steps for the Evolutionary model. Birth rate is set to 0.5 (but can be changed) and the death rate to 0: all the trees are ultrametric.

```
Time <- 20
N_steps <- Time * 100
N_sims <- 100
Birth_rate <- 0.2
tree_list <- phytools::pbtree(t=Time,
                              b=Birth_rate,
                              scale = N_steps,
                              nsim = N_sims,
                              method = "direct",
                              type="discrete")

names(tree_list) <- seq_along(tree_list) # we give a name to the trees
```
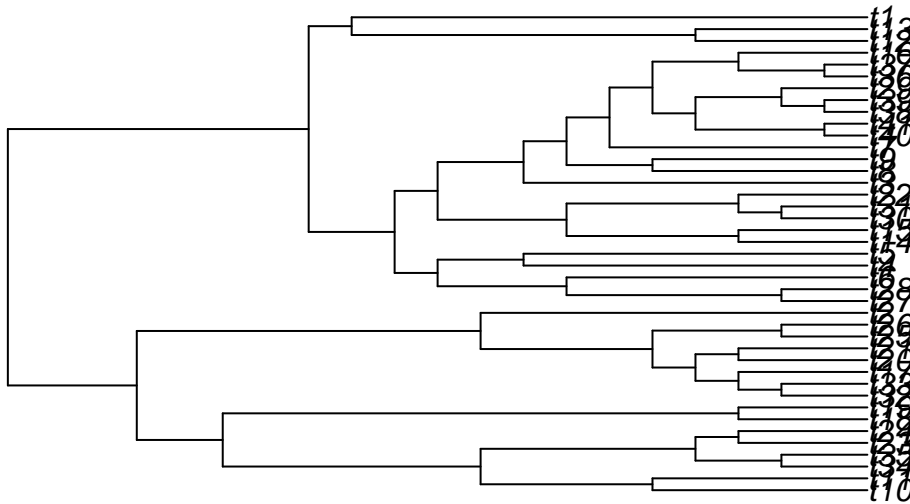
Let's observe one.

```
plot(tree_list[[1]])
```


Figure 1: A sampled phylogeny.

Simulate traits

Now we simulate the traits. Let's consider pure Brownian Motion, with 2 coordinates. We will not use the builtin simulator, but a version I wrote that allows for further generalisations.

```
sigma_simulations <- 0.2

trajectory_example <- trajectories_bm(tree_list[[1]],
                                      sigma = sigma_simulations,
                                      set_parallel = TRUE)
```

2

And let us plot the outcome.
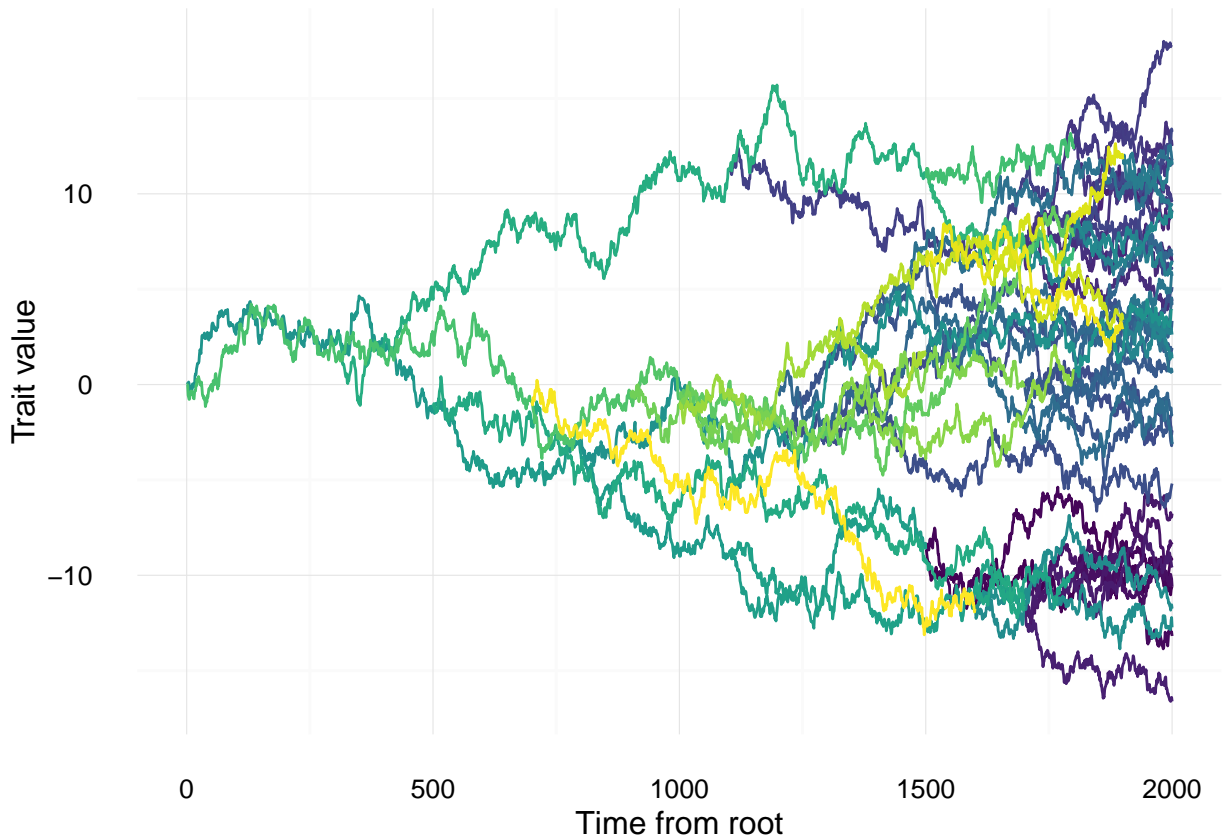
```
plot_trajectory(trajectory_example)
```


Figure 2: A sample trait trajectory.

Ok, everything seems to be working. We defined a convenient wrapping function for simulating the trajectories and preserving only the final distribution of the trait.

We us it to simulate the current trait distribution for every tree in our list. It takes a while, especially for large trees (consider using `.parallel = TRUE`)

```
traits_test <- tree_list %>%
  ldply(traits,
        sigma = sigma_simulations,
        .id = "tree",
        .parallel = TRUE) %>%
  as_tibble()
```

Compute Ecologica Uniqueness and Evolutionary Distinctiveness.

Now we are ready to compute the ecological and evolutionary indexes of uniqueness.

On one tree

We illustrate the analytic flow at first on just one tree, say the second one in the list.

```
i <- 1 # we pick a tree for example
```

We compute each tip Evolution Distinctiveness using the `picante` R package [but maybe Arne has something better in his sleeves ;-)].

```
evo_d_t1 <- picante::evol.distinct(tree_list[[i]],type="equal.splits") %>%
  mutate(Species = as.character(Species)) %>%
  as_tibble()
```

We define three different measures of Ecological Originality of a species: the first is the *average* distance of a species trait from all the other species in the community (that is, the distance between its traits and the centroid of the community's traits); the second is the *minimum* distance of a species trait from the closest trait in the community; the third is the *max* distance of a species trait from the furthest trait in the community (that is, the radius of the community centered in the focal species).

```
uniq <- traits_test %>%
  filter(tree == i) %>%
  widyr::pairwise_dist(item = Species,
                       feature = Species,
                       value = coord) %>%
  group_by(item1) %>%
  summarise(eco_ave = mean(distance), # average Ecological Originality
            eco_min = min(distance), # minimum Ecological Originality
            eco_max = max(distance)) %>% # maximum Ecological Originality
  rename(Species = item1)
```

We consolidate the ecological and evolutionary data in one data frame.

```
uniq_df <- full_join(uniq,evo_d_t1, by = "Species")
```

And let's plot this example simulation to see wether there's a correlation between ecological and any of the evolutionary distinctiveness.

```
uniq_df %>%
  gather(uniq_measure,uniq_value,eco_ave:eco_max) %>% # express it in a long format
  ggplot(aes(x = w, y = uniq_value)) + # and plot it
  geom_point()  +
  facet_wrap(~ uniq_measure, scales = "free") +
  geom_smooth(method = "lm") +
  labs(y = "Ecological Uniqueness",
       x = "Evolutionary Distinctiveness")  +
  theme_minimal()
```

Not much indeed, it seems. Let's test it.

```
models_test <- uniq_df %>%
  fit_with(lm,
           formulae(ave = ~eco_ave,
                    max = ~eco_max,
                    min = ~eco_min,
                    ~w))
```
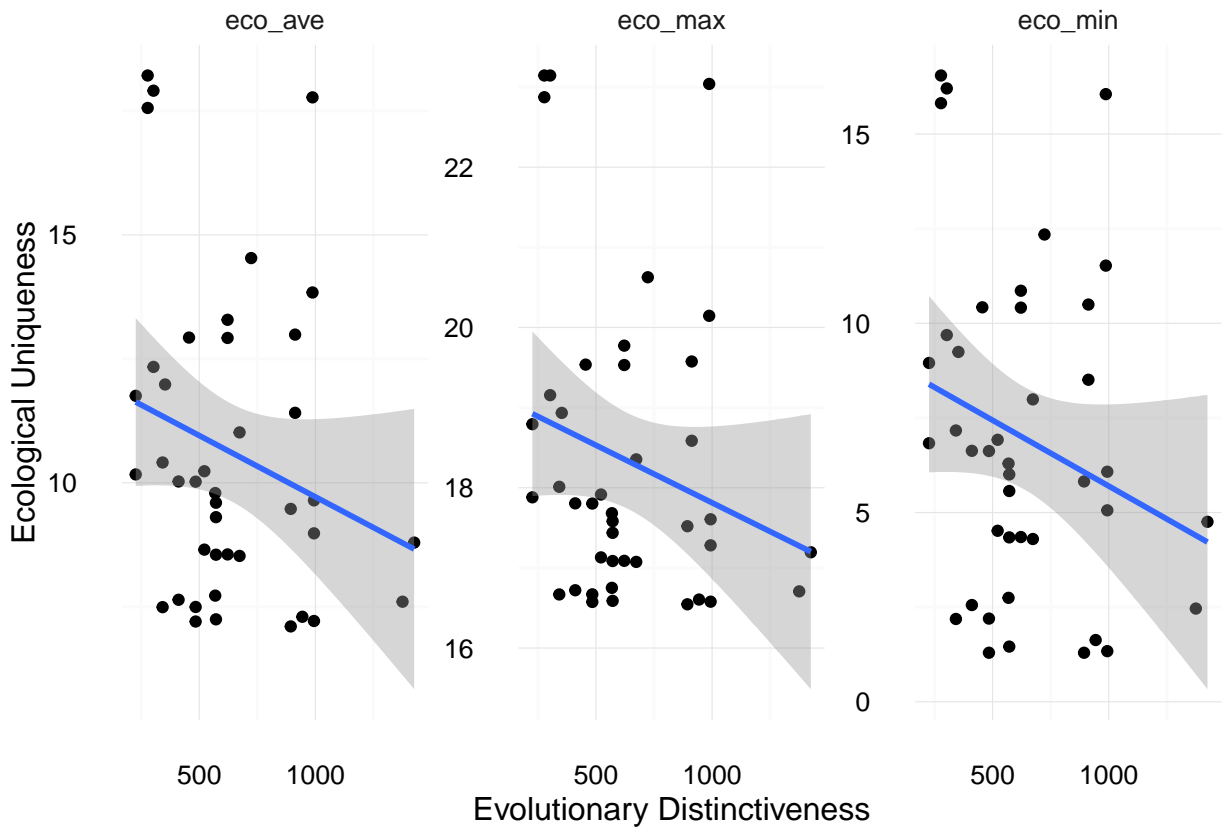
4

Figure 3: Relationship between Evolutionary Distinctiveness and Ecological Uniqueness

How do they fit? Let's take a look at the R-squared of the models.

```
models_test %>%
  laply(function(x) rsquare(x, uniq_df))
```

```
## [1] 0.05382649 0.04941186 0.05607700
```

Rather low, eh? We conveniently wrapped everything in a function. Let's see how it works:

```
uniq_df %>%
  good_of_fit()
```

```
##   ecological_uniqueness  estimate  r.squared adj.r.squared     sigma
## 1                   ave -21.75017 0.05382649    0.02956563 291.8978
## 2                   max -34.40111 0.04941186    0.02503780 292.5780
## 3                   min -16.11823 0.05607700    0.03187384 291.5504
##   statistic   p.value df    logLik      AIC      BIC deviance df.residual
## 1  2.218656 0.1443950  2 -289.8838 585.7676 590.9084  3322969          39
## 2  2.027232 0.1624538  2 -289.9792 585.9585 591.0992  3338473          39
## 3  2.316929 0.1360402  2 -289.8350 585.6700 590.8107  3315065          39
```

On all trees

We compute the Ecological Originality for each data set of traits.

```
all_eco <- traits_test %>%
  get_eco_data(variables = "tree", across = "Species")
```

And we compute the Evolutionary Distinctiviness for each tree.

```
all_evos <- ldply(tree_list,
                  evol.distinct,
                  type="equal.splits",
                  .id = "tree",
                  .parallel = TRUE) %>%
  as_tibble()
```

And we merge the two datasets.

```
all_data <- left_join(all_evos,all_eco,by=c("tree","Species"))
all_data$tree <- as.integer(all_data$tree)
```

Now we are ready to apply it to every tree in the dataset.

```
all_gof <- all_data %>%
  group_by_("tree") %>%
  do(good_of_fit(.)) %>%
  gather(key = "gof_index",
         value = "gof_value",
         estimate:df.residual)
```

We can graphically summarise the result in a histogram.

```
all_gof %>%
  filter(gof_index %in% c("r.squared","p.value")) %>%
  ggplot(aes(y = gof_value, x = ecological_uniqueness)) +
  geom_boxplot() +
  facet_grid(. ~ gof_index,scales = "free") +
  labs(y = "Goodness of fit",
       x = "Ecological Uniqueness")  +
  theme_minimal()
```
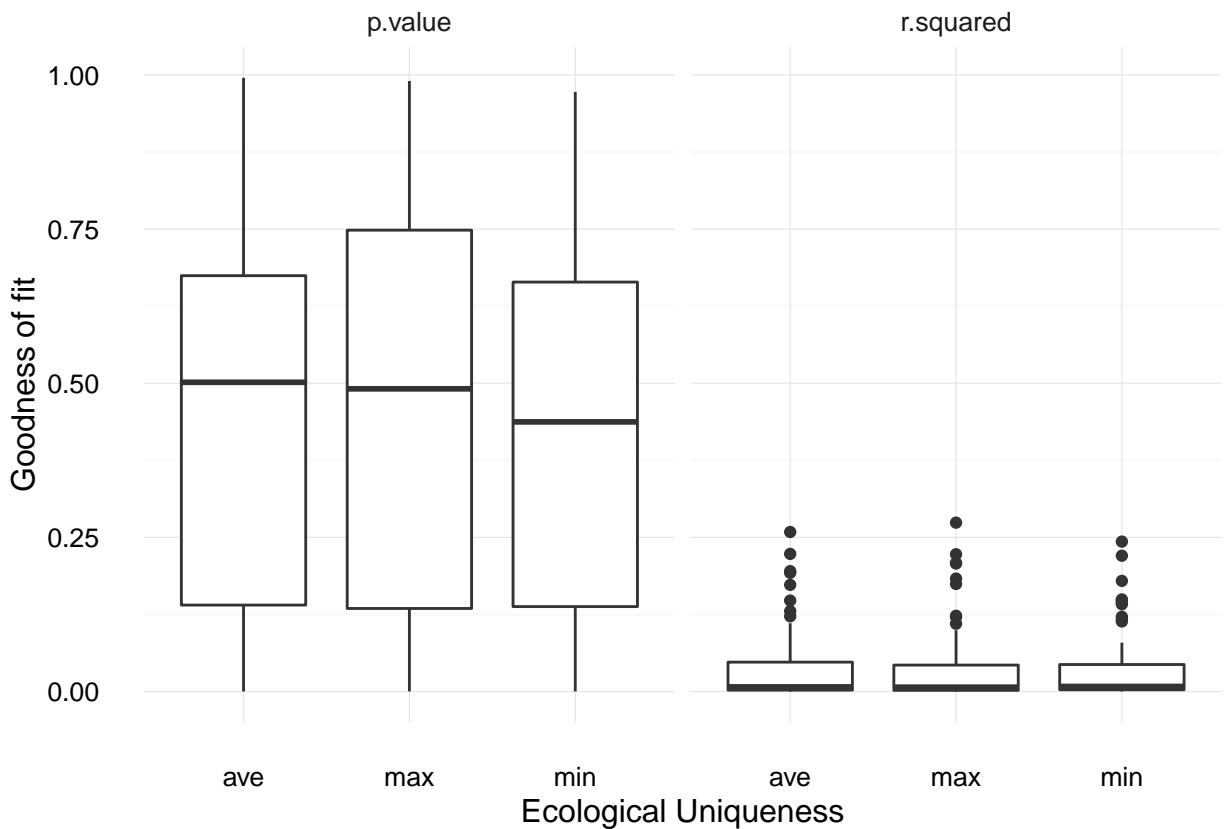

Figure 4: Are Ecological Uniqueness and Evolutionary Distinctiveness correlated?

It seems that none of the ecological uniqueness measures we implemented are significantly correlated with Evolutionary Distinctiveness: they explain less than 30% of ED and the significance is uniformly distributed over the $[0, 1]$ interval.

Next, we will explore the parameter space (for the trees, the traits and the dimensionality of the traits) to see if this result is general or not.

Tree space

Trait space

We are going to use the same trees simulated above.

The parameters of the simulation that we are interested in exploring are `speciation_jump`, determining how much of the trait drift happens at speciation event, and `early_burst`, determining how fast the drit slows

down with the age of a lineage. Here we adopt the point of view that both the lineages originating at a speciation event are *new* lineages, and restart their clock.

Let's build the parameter grid.

```
speciation_jumps <- seq(from = 10,to = 100,length.out = 5)
early_bursts <- seq(from = 0,to = 2,length.out = 5)
param_space <- expand.grid(sj = speciation_jumps,
                           eb = early_bursts)
```

And now let's replicate our analysis over the parameter space.

```
all_trjs <- param_space %>%
  adply(1,
        function(x){
          d_trj(tree_list = tree_list,
                "speciation_jump" = x[,1],
                "early_burst" = x[,2])
        },
        .parallel = TRUE
  ) %>%
  as_tibble()

my_vars <- c("tree","sj","eb")

all_gof_sj_eb <- all_trjs %>%
  get_eco_data(variables = my_vars) %>%
  consolidate(all_evos) %>%
  do_good_of_fit(variables = my_vars)
```

Let's take a look:

```
all_gof_sj_eb %>%
  group_by(sj,eb,ecological_uniqueness) %>%
  summarise(r_squared = mean(r.squared)) %>%
  ggplot(aes(x = as.factor(sj), y = as.factor(eb),fill =  r_squared)) +
  geom_tile() +
  facet_grid(. ~ ecological_uniqueness) +
  scale_fill_viridis(direction = - 1) +
  labs(y = "Early Burstness",
       x = "Jump at speciation",
       colour = "R squared")  +
  theme_minimal()
```

Nothing there.

```
all_gof_sj_eb %>%
  group_by(sj,eb,ecological_uniqueness) %>%
  summarise(estimate = mean(estimate)) %>%
  ggplot(aes(x = as.factor(sj), y = as.factor(eb),fill =  estimate)) +
  geom_tile() +
  facet_grid(. ~ ecological_uniqueness) +
  scale_fill_viridis(direction = - 1) +
```
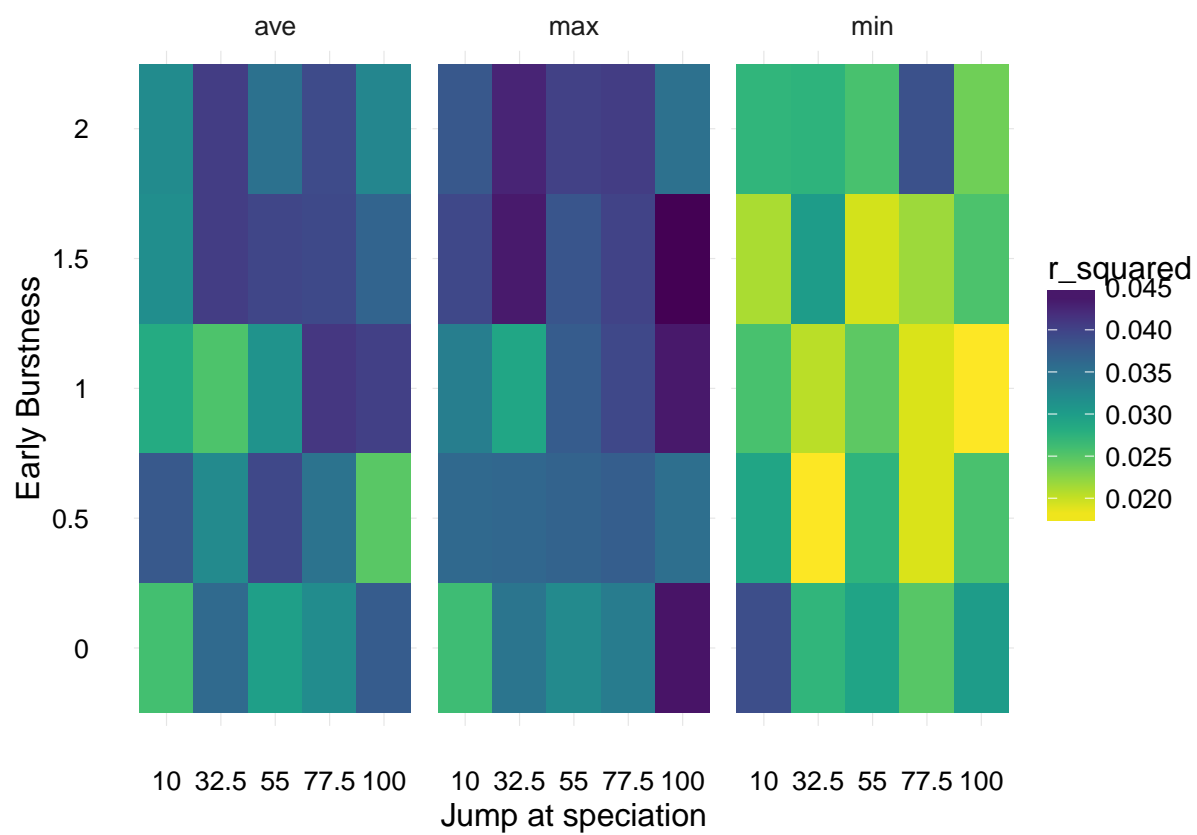
Figure 5: Goodness of fit of a linear model of Ecological Uniqueness and Evolutionary Distinctiveness.

```
    labs(y = "Early Burstness",
         x = "Jump at speciation",
         fill = "Coefficient")  +
    theme_minimal()
```
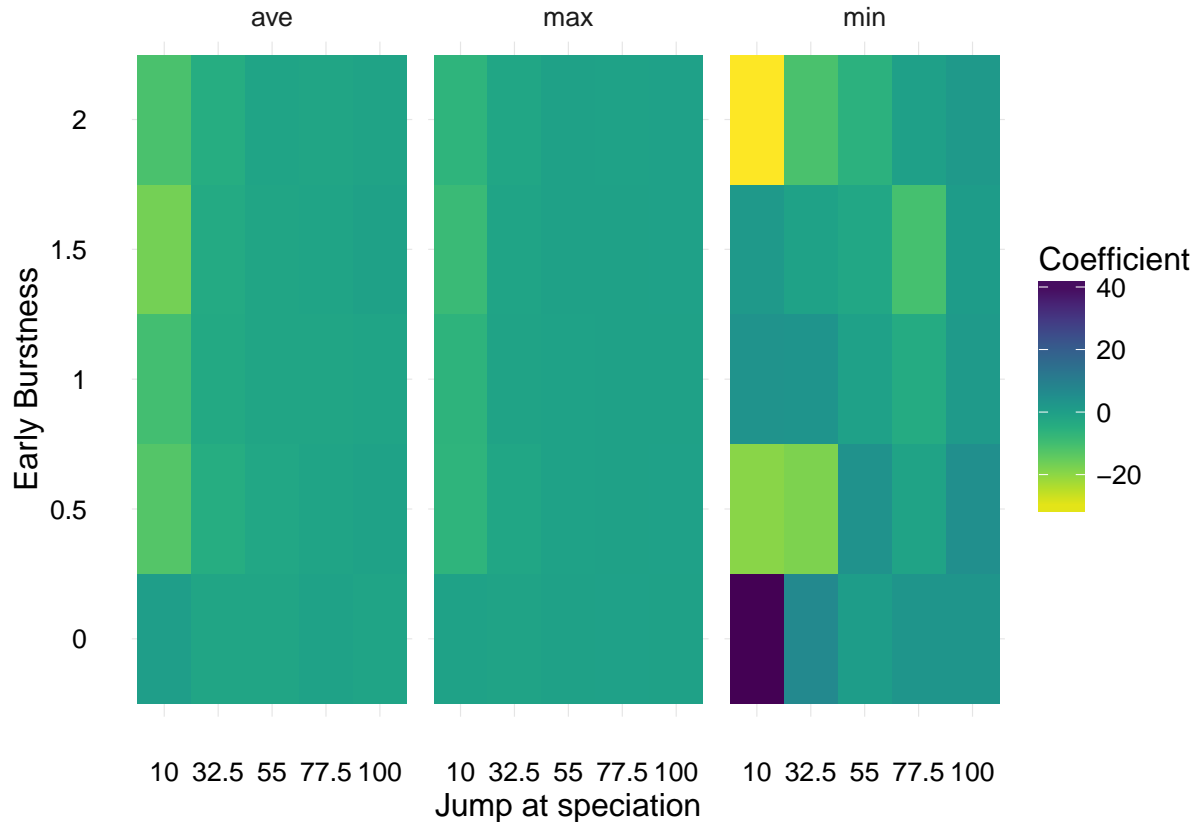

Figure 6: Estimated coefficient of a linear model of Ecological Uniqueness and Evolutionary Distinctiveness.

Traits dimensionality

One candidate explanation of that lack of correlation is saturation of the trait space. A low dimensional trait space get's saturated much quicker. Let's see how the trait dimensionality (the number of coordinates of the trait) influence the correlation.

```
speciation_jumps <- seq(from = 0,to = 100,length.out = 4)
early_bursts <- seq(from = 0,to = 2,length.out = 4)
dimensionality <- c(1,3,5,10,100)

param_space <- expand.grid(sj = speciation_jumps,
                           eb = early_bursts,
                           dims = dimensionality)

dim_ts_trjs <- param_space %>%
  adply(1,
        function(x) {
```

```
        d_trj(
          tree_list = tree_list,
          "speciation_jump" = x[, 1],
          "early_burst" = x[, 2],
          "Dim_trait" = x[, 3]
        )
      },
      .parallel = TRUE) %>%
  as_tibble()
```

We fit the regression models.

```
my_vars <- c("tree","sj","eb","dims")

stats_dim <- dim_ts_trjs %>%
  get_eco_data(variables = my_vars) %>%
  consolidate(all_evos) %>%
  do_good_of_fit(variables = my_vars)
```

And let's plot it

```
stats_dim %>%
  group_by(sj,eb,dims,ecological_uniqueness) %>%
  summarise(r_squared = mean(r.squared)) %>%
  ggplot(aes(x = as.factor(sj), y = as.factor(eb),fill =  r_squared)) +
  geom_tile() +
  facet_grid(dims ~ ecological_uniqueness) +
  scale_fill_viridis(direction = - 1) +
  labs(y = "Early Burstness",
       x = "Jump at speciation",
       fill = "R squared") +
  theme_minimal()
```

```
stats_dim %>%
  ggplot(aes(x = as.factor(dims), y = estimate, colour = ecological_uniqueness)) +
  geom_boxplot() +
  facet_grid(~ eb + sj) +
  scale_fill_viridis(direction = - 1) +
  labs(y = "Coefficient",
       x = "Trait-space dimension",
       colour = "Ecological Uniqueness") +
  theme_minimal()
```

Directional selection

The traits trajectories we simulated so far are akin to neutral drift: there is no directional component in the selection pressure. However, ecological relevant traits may be under directional selection. In certain scenario, for example during adpative radiation, the directional selection may be strong.

Here we implement directional selection as a fixed trend in the Brownian motion, constant for each branch of the phylogeny, and sampled following a centered normal distribution with standard deviation $\sigma_{trend} * \sigma$ where σ is the (initial) standard deviation of the Brownian Motion.
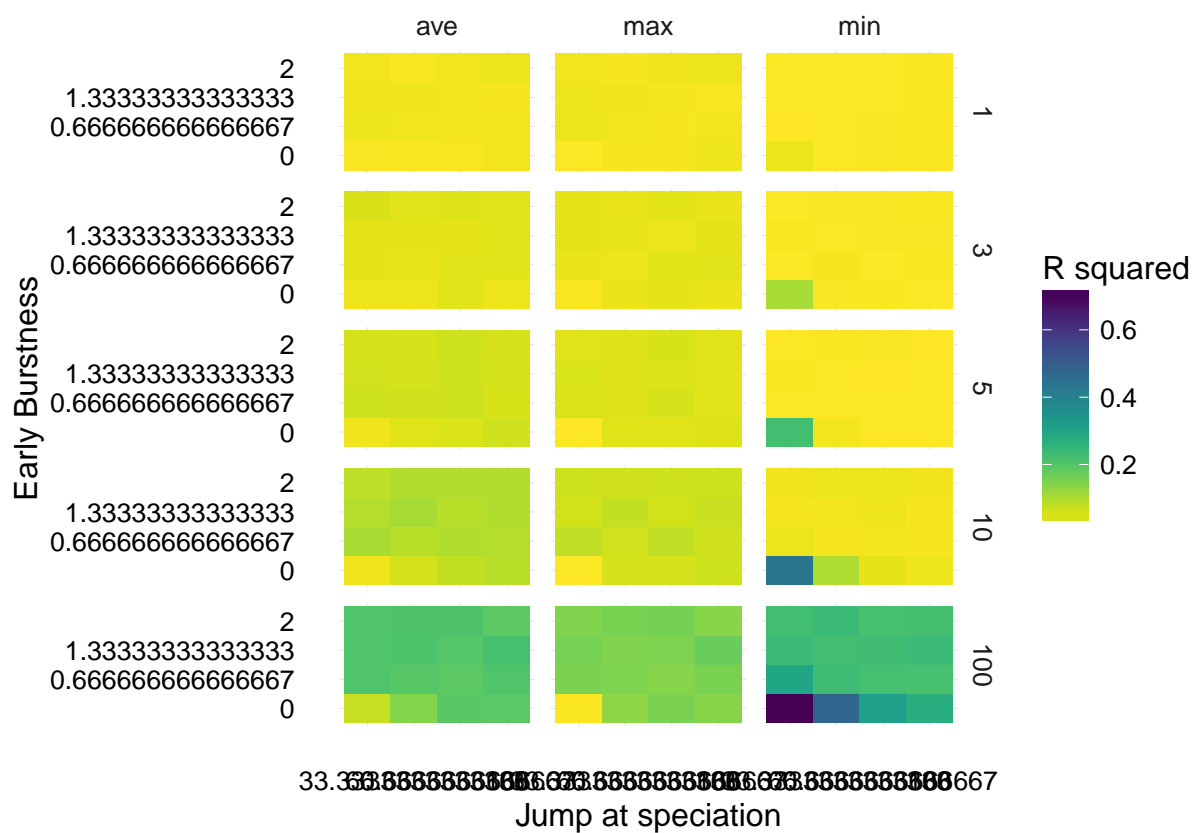
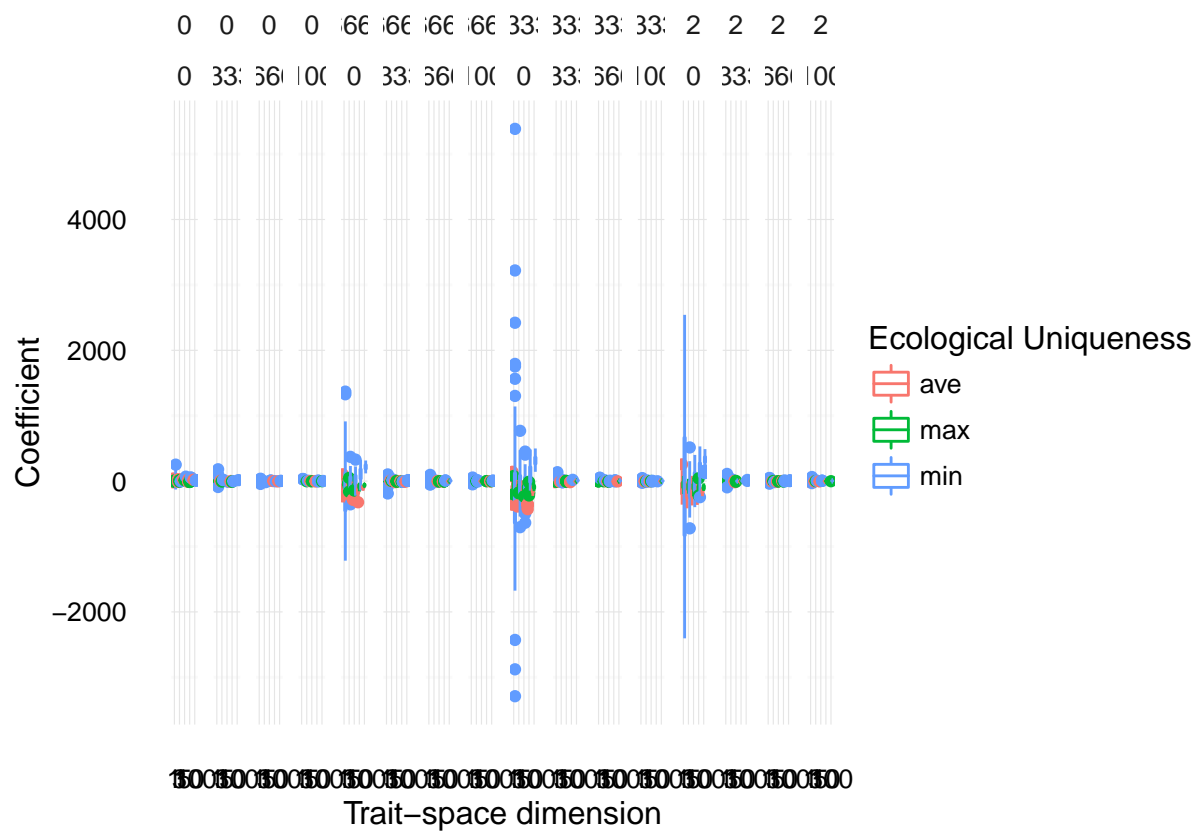Figure 7: Goodness of fit of a linear model of Ecological Uniqueness and Evolutionary Distinctiveness.

Figure 8: Estimated coefficient of a linear model of Ecological Uniqueness and Evolutionary Distinctiveness.

We consider at the beginning one simple scenario with no jump at the speciation nor early burst effect. Let's observe a sample trajectory.

Always the same tree:

```
plot(tree_list[[1]])
```
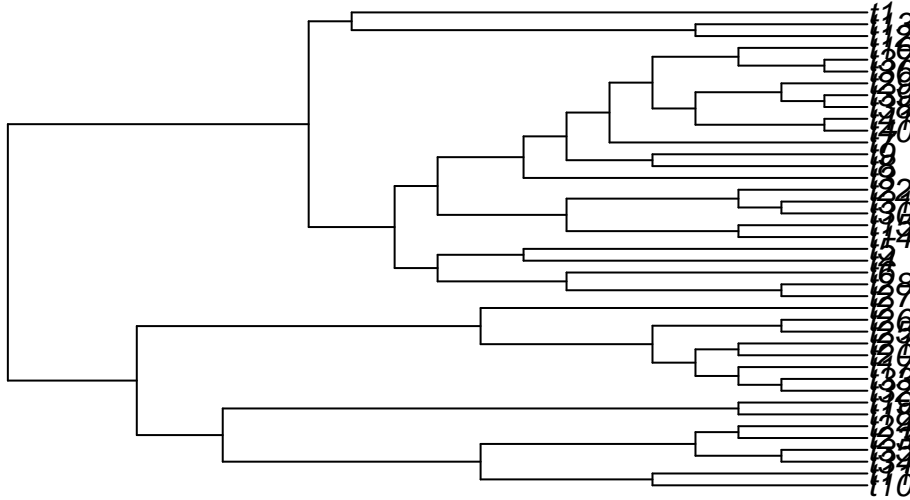

Figure 9: The same tree of before.

We sample one trajectory with a medium level of directional selection:

```
trajectories_bm(tree_list[[1]],
                sigma = 0.2,
                trend_sd = 0.1) %>%
  plot_trajectory() +
  labs(y = "Trait value",
       x = "Time from root")  +
  theme_minimal()
```

We explore the effect of different directional selection forces. Run the analysis:

```
dimensionality <- c(1,2,3,4,5,10,100)
trend_sds <- c(0.0001,0.001,0.01,0.1,1)
param_space <- expand.grid(dims = dimensionality,
                           trend_sd = trend_sds)

trajectories_trend <- param_space %>%
  adply(1,
        function(x){
          d_trj(tree_list = tree_list,
                "Dim_trait" = x[,1],
                "trend_sd" = x[,2])
        },
        .parallel = TRUE
  ) %>%
  as_tibble()
```
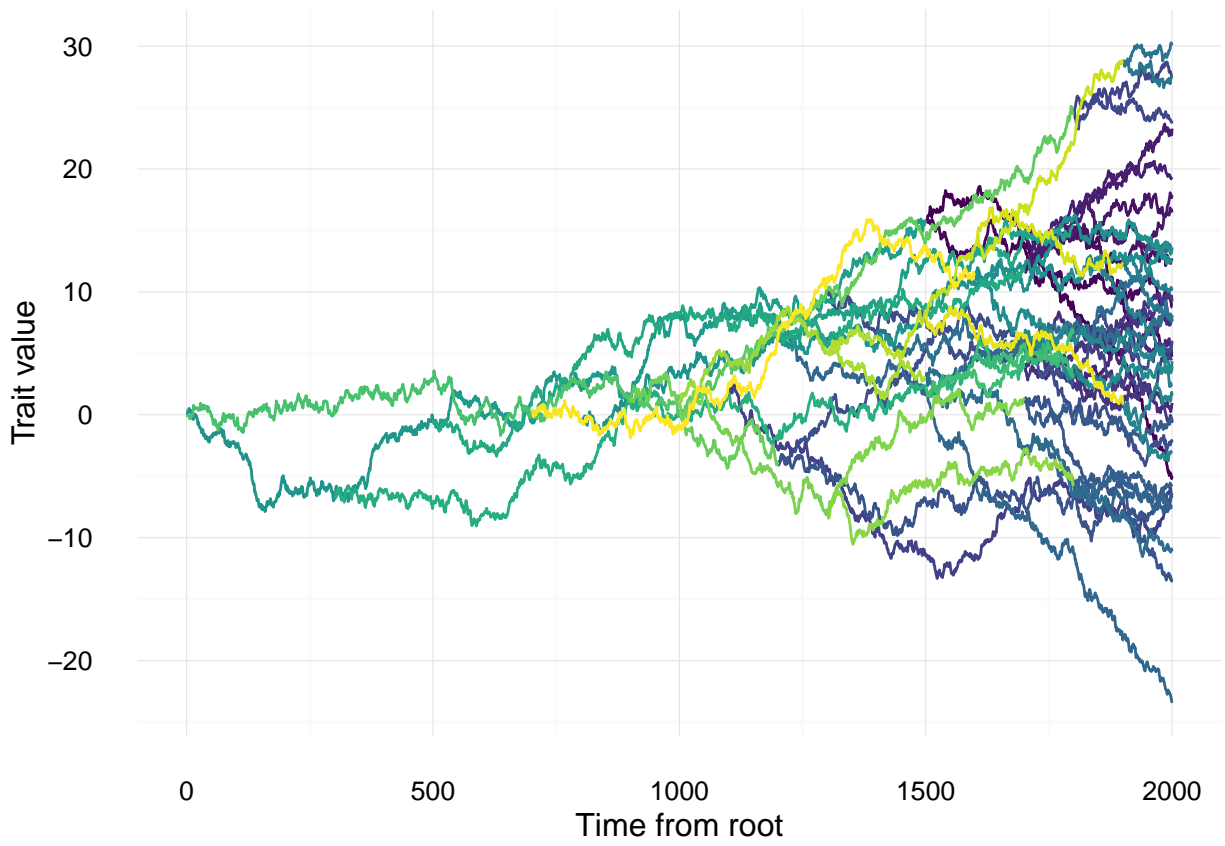
Figure 10: A sampled trajectory where selection has a directional component.

```
my_vars <- c("tree","trend_sd","dims")

stata_trend <- trajectories_trend %>%
  get_eco_data(variables = my_vars) %>%
  consolidate(all_evos) %>%
  do_good_of_fit(variables = my_vars)
```

And let's plot it:

```
stata_trend %>%
  group_by(dims, trend_sd, ecological_uniqueness) %>%
  summarise(r_squared = mean(r.squared)) %>%
  ggplot(aes(x = as.factor(dims), y = as.factor(trend_sd),fill =  r_squared)) +
  geom_tile() +
  facet_grid(. ~ ecological_uniqueness) +
  scale_fill_viridis(direction=-1) +
  labs(y = "Directional trend sd",
       x = "Trait-space dimension",
       fill = "R squared")  +
  theme_minimal()
```
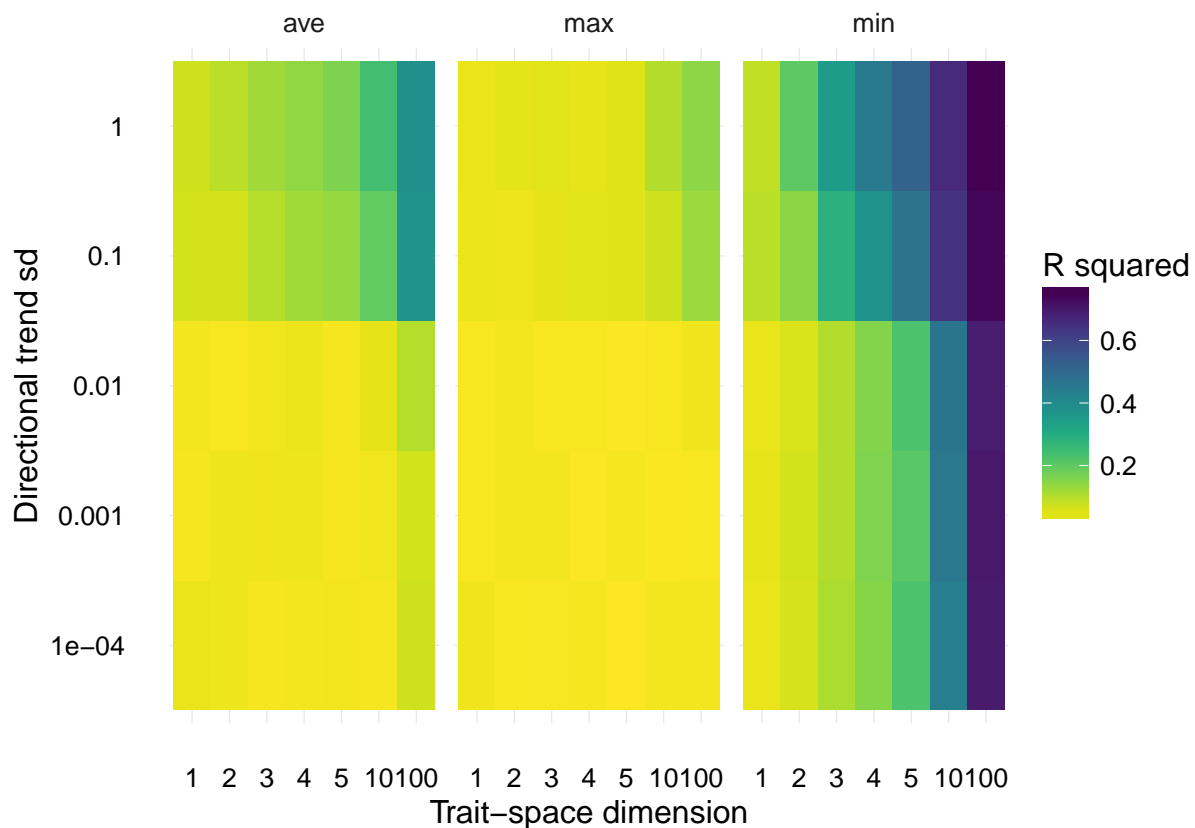

Figure 11: Goodness of fit of a linear model of Ecological Uniqueness and Evolutionary Distinctiveness.

```
stata_trend %>%
  group_by(dims, trend_sd, ecological_uniqueness) %>%
  summarise(Coefficient = mean(estimate)) %>%
  ggplot(aes(x = as.factor(dims), y = as.factor(trend_sd),fill = Coefficient)) +
  geom_tile() +
  facet_grid(. ~ ecological_uniqueness) +
  scale_fill_viridis(direction=-1) +
  labs(y = "Directional trend sd",
       x = "Trait-space dimension",
       fill = "Coefficient")  +
  theme_minimal()
```
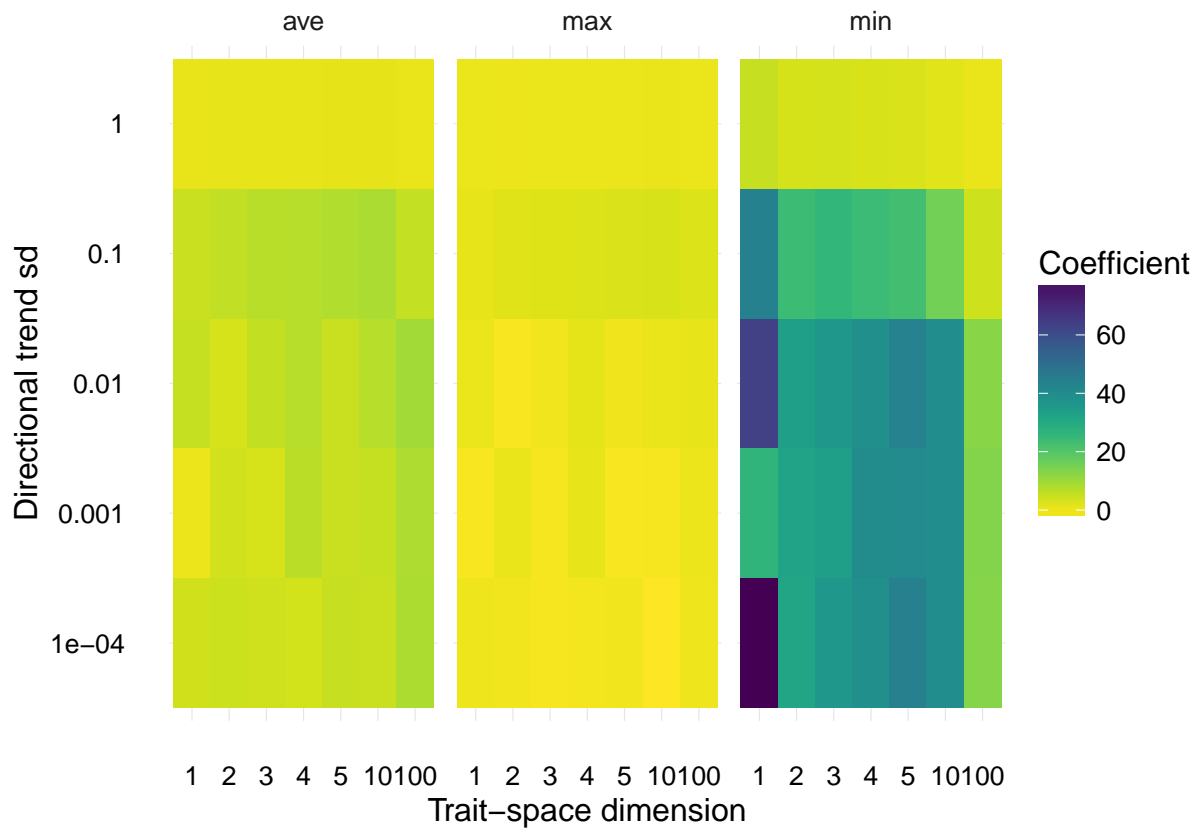

Figure 12: Estimated coefficient of a linear model of Ecological Uniqueness and Evolutionary Distinctiveness.

Discussion

From this preliminary results, it appears that may be, indeed, some relationship between Ecological and Evolutionary Distinctiveness. This is true only outside of neutral evolution scenarios, and it is easily obfuscated by the saturation of the trait space. However, in a high-dimensional trait space, under the influence of strong directional selction, Ecological Uniqueness and Evolutionary Distinctiveness are correlated.