

checking BCI species selection estimates: generation normalized

preamble

This checks Ben's frequency independent selection estimates against calculations done from the the beta estimates. The selection estimates that we are given are normalized for generation time. The betas are given in terms of census years. We should be able to calculate the selection estimates from the betas given some additional information.

libraries

data

The betas estimates: estimated in terms of census years

```
beta<-readRDS("beta_freqinde.rds")
beta<-as.data.frame(beta)
dim(beta)
```

```
## [1] 268 3
```

The selection estimates: estimated in terms of generation time

```
sel<-readRDS("estimates_selection.rds")
sel<-as.data.frame(sel)
dim(sel)
```

```
## [1] 269 7
```

The immigration rates: estimated in terms of census years

```
d<-readRDS("immigration.rds")
```

get species names for the beta estimates. The last variant has to have beta=0 to allow the model to be identifiable. We will assume that the selection list, which is alphabetical, is the right order

```
beta$species<-head(sel$species,-1)
```

Add the last species in. By definition its betas are zero

```
species<-tail(sel$species, 1)
last<-as.data.frame(species)
last$beta_freqinde_2.5<-0
last$beta_freqinde_50<-0
last$beta_freqinde_97.5<-0
last<-last%>%
  select(beta_freqinde_2.5,beta_freqinde_50,beta_freqinde_97.5, species)
beta1<-rbind(beta,last)
dim(beta1)
```

```
## [1] 269 4
```

get generation times for all betas. We need to get the **original** names, without the repeat-migration suffix

```
beta1$species_original<-str_split_fixed(beta1$species, "_", 2)[,1]
```

these are the generation times

```
gen<-read.csv("generation_time_imputed.csv")
gen<-gen%>%
select(genus_species_subspecies, generation_time)
names(gen)<-c("species", "generation_time")
beta2<-merge(beta1,gen, by.x="species_original", "species")
dim(beta2)
```

```
## [1] 269 6
```

Get the average number of years between census_years census data. This should be 4.7 years since we have added in the 2005 data; but we'll use 5.5 years for the moment as, I think, was done by Ben. In fact it makes hardly any difference.

```
censusyears<-c(1982,1985,1990,1995,2000,2005,2010,2015)
m<-mean(diff(censusyears))
m
```

```
## [1] 4.714286
```

```
#m<-5.5
```

normalize beta_freqinde_50 to generation time

```
beta2$beta_freqinde_50_gennorm<-beta2$beta_freqinde_50/m*beta2$generation_time
```

Now estimate absolute fitness, W . We need an estimate of delta (or μ). We use the median estimate of the frequency independent rate.

```
delta<-d[3,1]
delta
```

```
## [1] 0.0002521865
```

```
beta2$W_gennorm=(1 - delta)*(1+beta2$beta_freqinde_50_gennorm)
beta2$w_gennorm=beta2$W_gennorm/median(beta2$W_gennorm)
beta2$s_gennorm=beta2$w_gennorm-1
```

Now, compare them our estimates of s to Ben's. This df contains all Ben's generation-normalized selection estimates

```
sel<-sel%>%
select(species, s_freqinde_50)
```

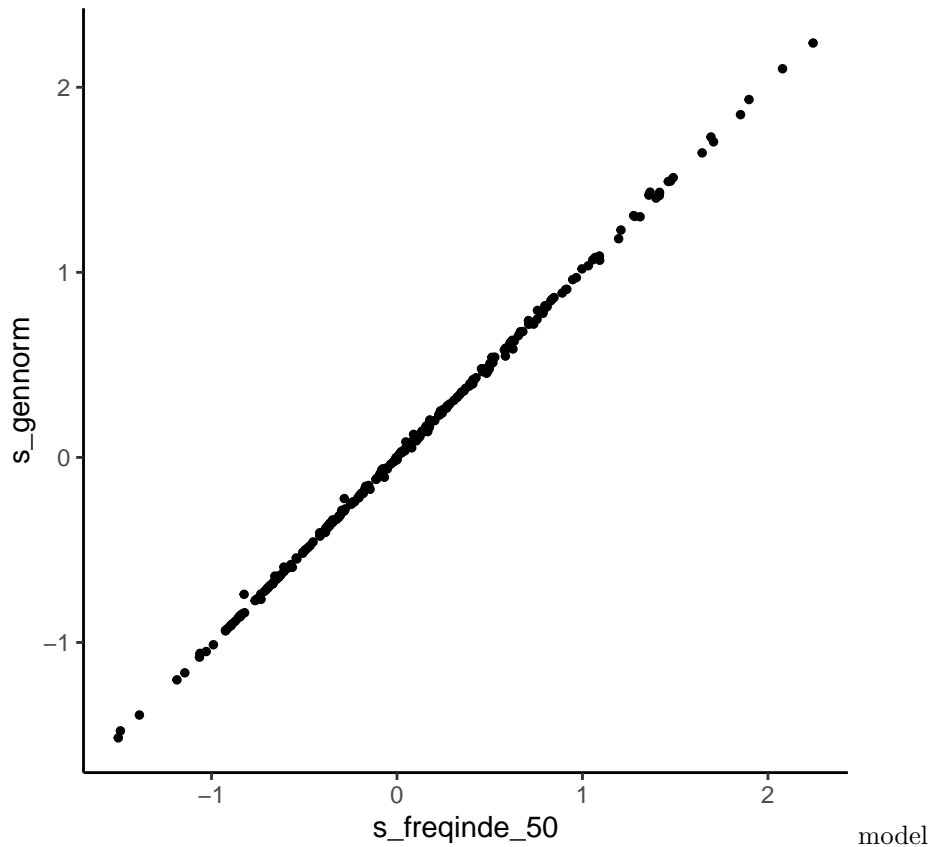
merge

```
beta3<-merge(beta2, sel, by="species")
beta3<-beta3%>%
select(species, s_freqinde_50, s_gennorm)%>%
arrange(desc(s_freqinde_50))
dim(beta3)
```

```
## [1] 269 3
```

plot

```
ggplot(beta3, aes(x=s_freqinde_50, y=s_gennorm))+
geom_point(size=1)+
theme_classic()+
theme(aspect.ratio=1)
```



```
summary(lm(s_freqinde_50~s_gennorm, data=beta3))
```

```
##
## Call:
## lm(formula = s_freqinde_50 ~ s_gennorm, data = beta3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.092105 -0.003374  0.001112  0.004546  0.043856
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0009858  0.0008469   1.164   0.245
## s_gennorm    0.9908940  0.0012320 804.270 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01384 on 267 degrees of freedom
## Multiple R-squared:  0.9996, Adjusted R-squared:  0.9996
## F-statistic: 6.468e+05 on 1 and 267 DF, p-value: < 2.2e-16
```

slope ~1 and $r^2=0.999$

check values

```
head(beta3)
```

```
##              species s_freqinde_50 s_gennorm
## 1 Aspidosperma spruceanum      2.243540  2.239103
```

```
## 2 Erythrina costaricensis      2.078749  2.099977
## 3   Prioria copaifera          1.897880  1.933693
## 4   Drypetes standleyi        1.852163  1.851992
## 5   Tabebuia guayacan         1.706671  1.705160
## 6   Sloanea terniflora        1.692857  1.731339
```

```
tail(beta3)
```

```
##                species s_freqinde_50 s_gennorm
## 264 Psychotria tenuifolia_2      -1.064669 -1.079219
## 265 Psychotria hoffmannseggiana_1 -1.144009 -1.163895
## 266   Pterocarpus rohrii         -1.186110 -1.201834
## 267   Bactris coloradonis        -1.388692 -1.391975
## 268     Inga oerstediana         -1.490963 -1.477414
## 269     Inga pezizifera_1        -1.502629 -1.515416
```

So, they're very close but not identical. Using an among-census factor of 5.5 instead of 4.7 brings the values *slightly* closer.