Exercises on diversity metrics

M. Humberto Reyes-Valdés

June 23, 2019

In this exercise, you will compare among four diversity metrics for allelic diversity: Shannon entropy, exponential Shannon entropy, Gini-Simpson index and allele richness. The comparison will be based on simulated data for loci with different allele numbers and frequencies. The following code will be useful:

```r
#Function to get a simulated list
sim.p<-function(n){
    a<-list()
    for (i in 1:n){
        x<-floor(runif(1)*10)+1 #Decide number of classes
        y<-runif(x) #Random probabilities
        y<-y/sum(y) #Normalize
        a[[i]]<-y
        a
    }
}
#Shannon diversity
#x log x
MyLog2p<-function(x){if(x==0) 0 else x*log(x,2)}
entropy<-function(x){-sum(sapply(x,MyLog2p),na.rm=T)}
#Gini-Simpson index
gs<-function(x){1-sum(x^2)}
#Allele richness
ar<-function(x){length(x)}
```

1. Simulate 1000 random loci with *x<-sim.p(1000)*. The variable *x* will be a list of vectors of allele frequencies for 1000 loci.

2. Make a vector for each on of the diversity metrics, by applying the functions to the list *x*. For each metrics, the vector should be of size 1000. Hint: use *lapply* and *unlist*. The way to calculate the exponential Shannon entropy is up to you.

3. The simulations were designed to have a uniform distribution of numbers of alleles and individual allele frequencies. How the distributions of the different measures of diversity look like?

4. Make a matrix of scatter plot diagrams. How the measures are related to each other? Which one is the most linear with allele richness? Can you draw some conclusions. Hint: use *pairs*.