Nucleotide diversity among IUCN Categories

Elisabetta Canteri

I here aim to show how nucleotide diversity is distributed among the IUCN Red List categories. I calculated nucleotide diversity for all species with cytochrome-b sequences, georeferenced and not.

The distribution of genetic diversity among IUCN categories is presented in Figure 1 and Figure 2. The second boxplot is a "subset" of the first one, as it shows just fewer categories. As the data is not normally distributed, I performed a Kruskal-Wallis test to see if the means of the values in each category are the same. The test is statistically significant, with a p-value = 1.522e-11, meaning that the null hypothesis is rejected and so the means of the values of the categories differ from each other. I then performed a post-hoc Dunn's test to see which are the categories that differ when compared. The Dunn's test has been chosen because the number of values of each categories is not the same. As it is possible to see here below, LC is significantly different from all the "threatened" categories (VU, EN, CR), but not NT.

```
##
##   Kruskal-Wallis rank sum test
##
## data:  Nuc_div by IUCN
## Kruskal-Wallis chi-squared = 59.527, df = 5, p-value = 1.522e-11
```

```
##
##   Pairwise comparisons using Dunn's-test for multiple
##                         comparisons of independent samples
##
## data:  Nuc_div by factor(IUCN)
##
##      LC        NT        VU        EN        CR
## NT 0.88164 -         -         -         -
## VU 0.00014 0.20389   -         -         -
## EN 1.5e-06 0.00945   1.00000   -         -
## CR 0.00632 0.16341   1.00000   1.00000   -
## EW 0.64309 0.88164   1.00000   1.00000   1.00000
##
## P value adjustment method: holm
```

If LC and NT are grouped together and the test is run again, we see that the species at risk have significantly less nucleotide diversity than the species in the LC and NT categories.

```
##
##   Kruskal-Wallis rank sum test
##
## data:  grouped
## Kruskal-Wallis chi-squared = 57.103, df = 4, p-value = 1.177e-11
```

```
##
##   Pairwise comparisons using Dunn's-test for multiple
##                         comparisons of independent samples
##
## data:  grouped
##
##   1       2         3         4
## 2 0.00014 -         -         -
## 3 1.5e-06 1.00000   -         -
## 4 0.00479 1.00000   1.00000   -
```

```
## 5 0.51619 1.00000 1.00000 1.00000
##
## P value adjustment method: holm
```
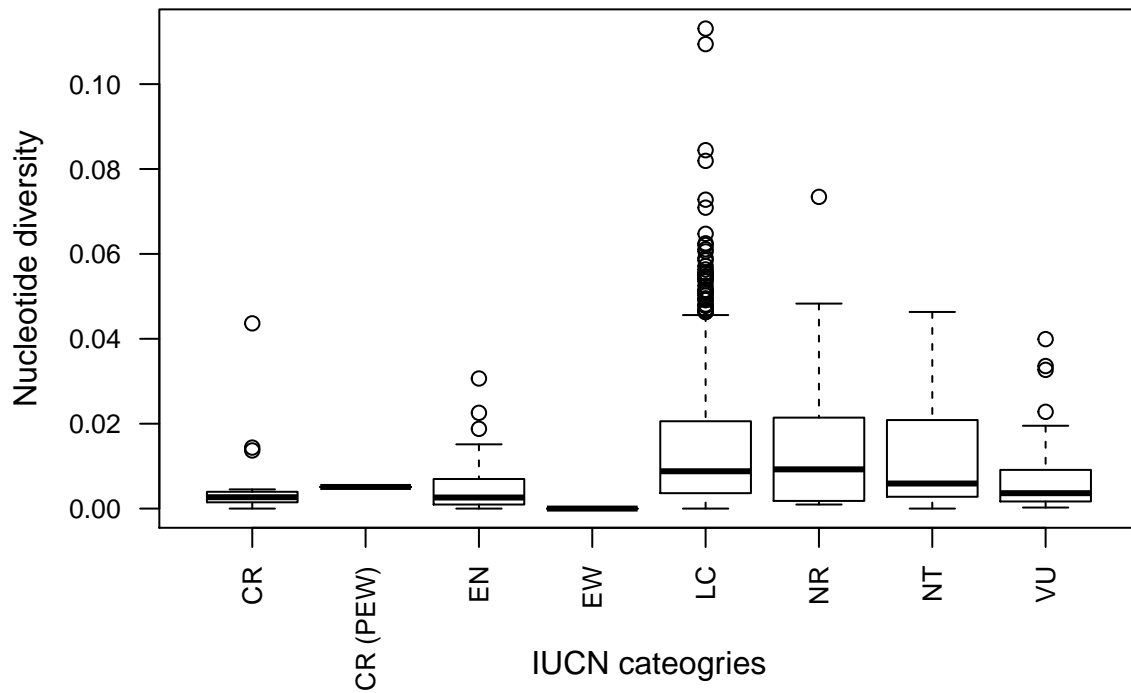
Species with more than 5 sequences

Figure 1: Nucleotide diversity among all categories present in IUCN Red List. CR = critically endangered, CR (PEW) = possibly extinct in the wild, CR (PE) = possibly extinct, DD = data deficient, EN = endangered, EW = extinct in the wild, LC = least concern, NR = not recognized, NT = near threatened, VU = vulnerable.
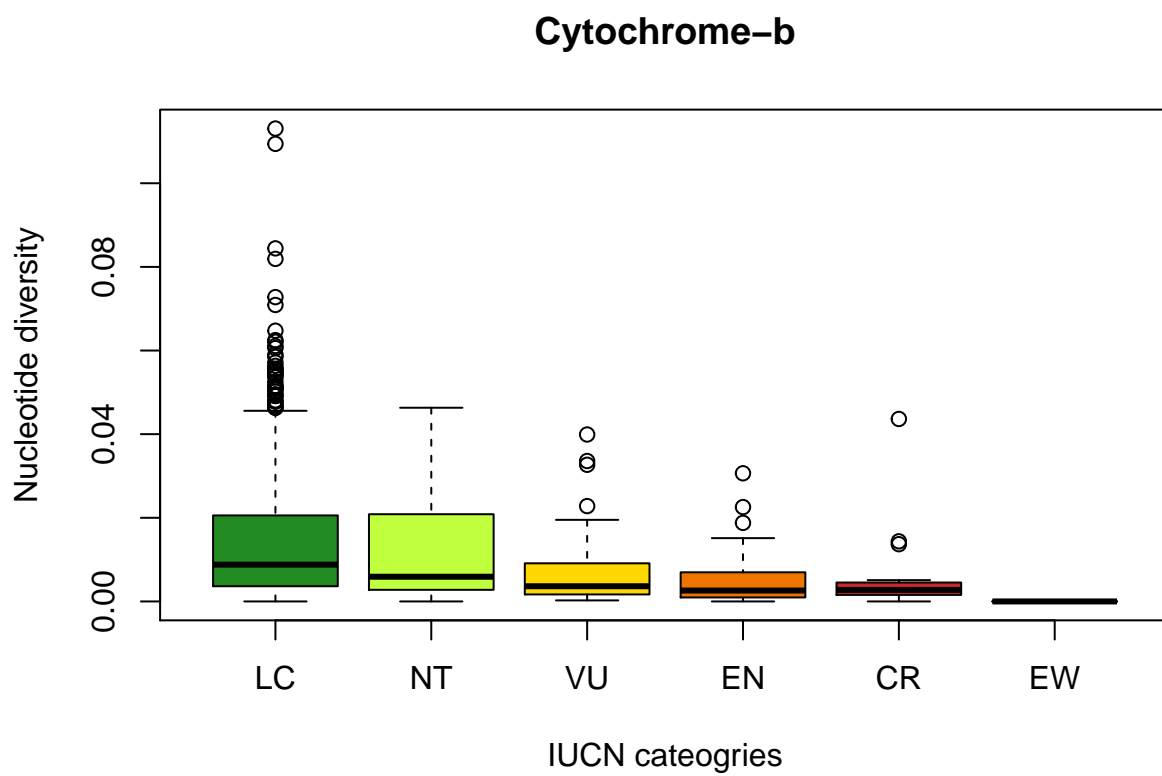
Cytochrome–b

Figure 2: Nucleotide diversity among the IUCN Red List categories. NR category is not shown.