

Tommasi N., Ferrari A., Labra A., Galimberti A., Biella P.*, Harnessing the power of metabarcoding in the ecological interpretation of plant-pollinator DNA data: strategies and consequences of filtering approaches. Diversity 2021, 13

Text S1 - Script for estimating filtering thresholds with an approach based on ROC curves.

The following lines can be run in the R environment, please note that implementations different from the following one are possible, and they might affect the estimated thresholds.

```
# the data of the table below the script are to be included in an object named "y"
fit1 <- glm(reads ~ label, family="quasipoisson", data=y) #estimate the false positives
predictions <- data.frame(label = y$label, pred = NA)
predictions[,"pred"] <- predict(fit1, type="response")

#estimate cutting tresholds
require(pROC)
roc1 <- roc(as.factor(predictions$label), predictions$pred, quiet=T)
coords(roc1, "best")[1] #threshold

#data: see the following table. Load the reads per taxon (rows) in a sample
```

reads	label
9	P
5	P
14	P
0	N
0	N
13	P
415	P
132	P
4	P
76	P
370	P
21	P

0	N
0	N
184	P
0	N
0	N
0	N
0	N
145	P
2	P
5	P
2	P
0	N
0	N
0	N
13	P
5968	P
29	P
12	P
412	P
1007	P
0	N
0	N
0	N
0	N
0	N
0	N
137	P
0	N
0	N
0	N
0	N
0	N

