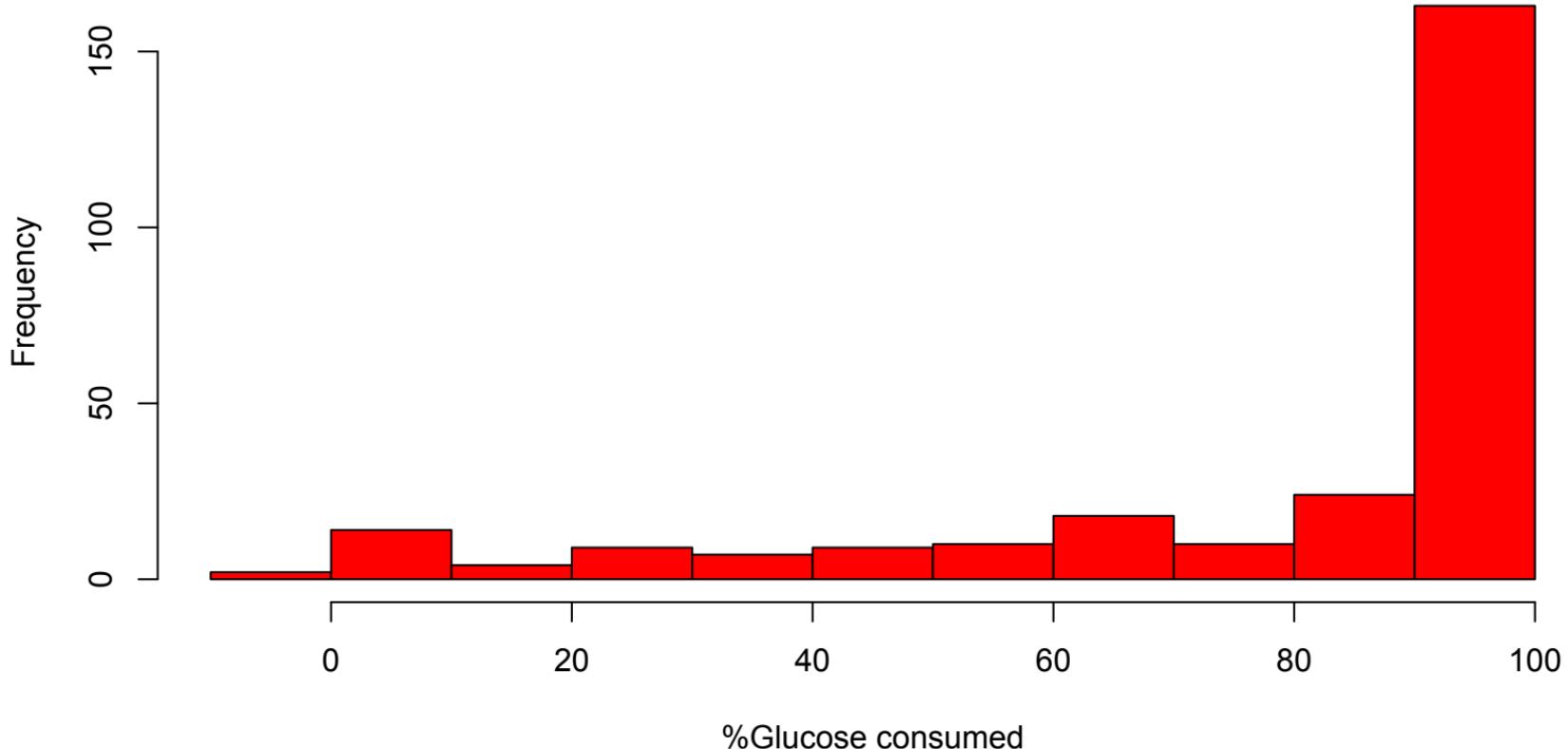
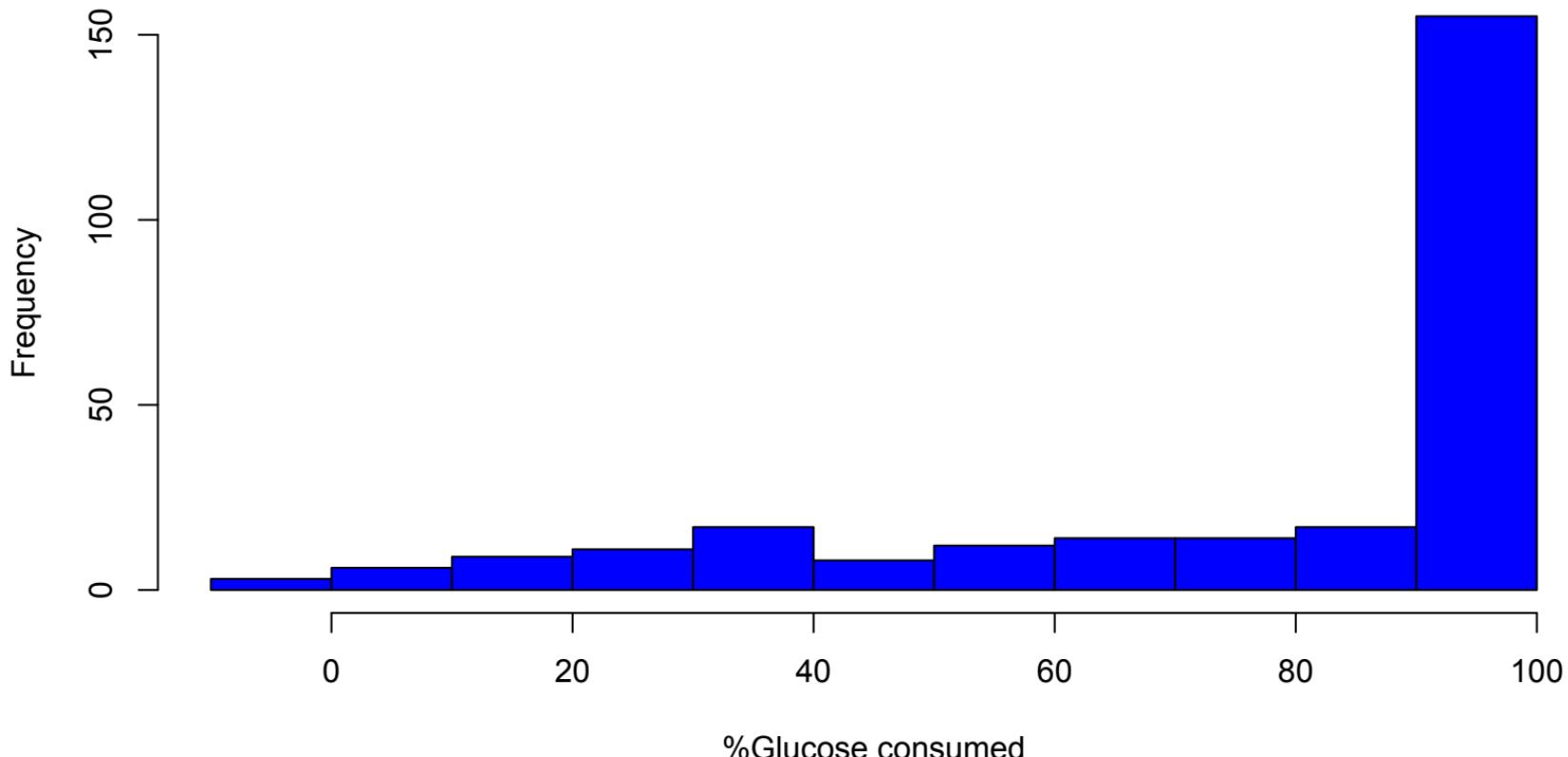


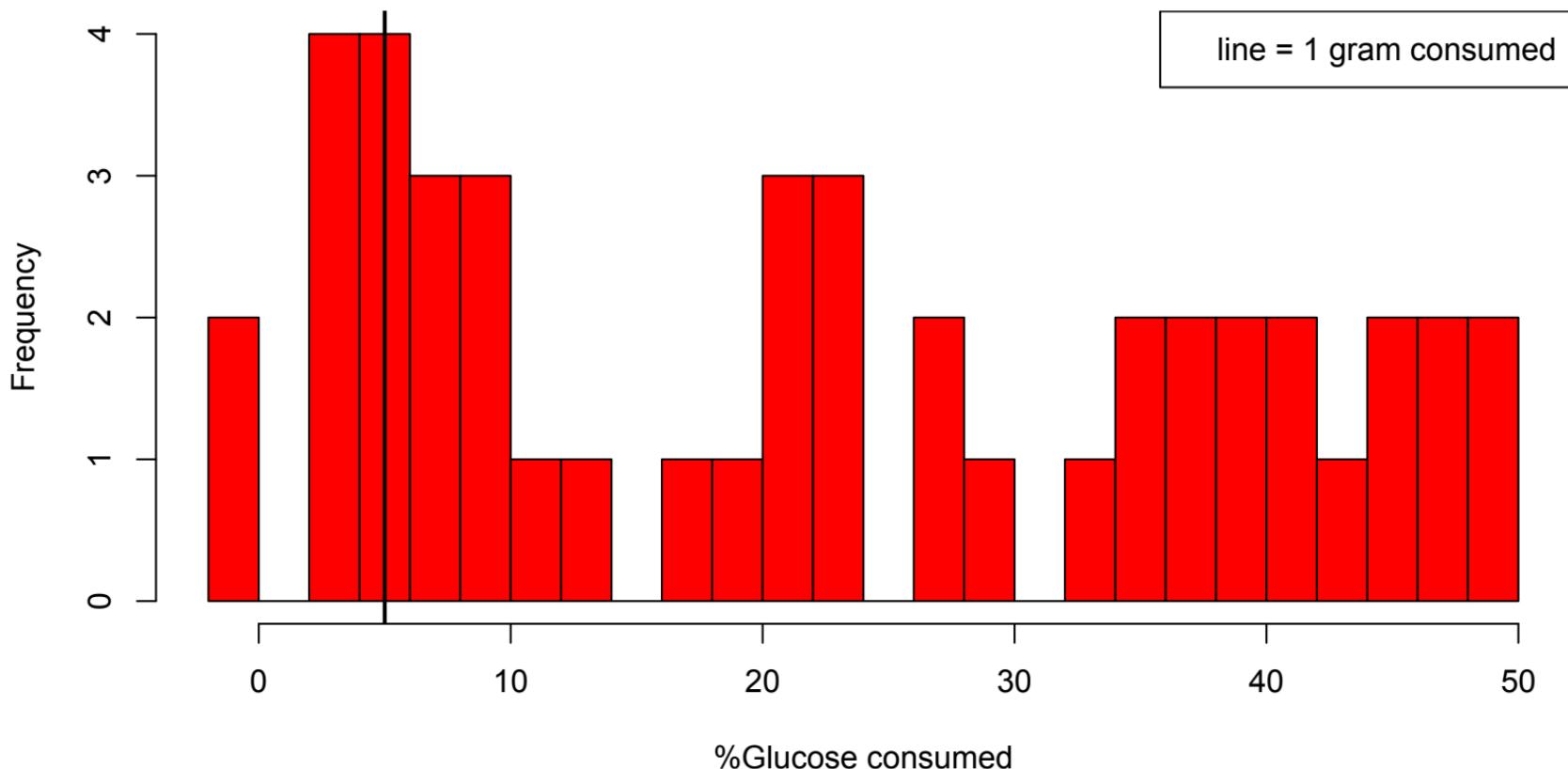
% Glucose consumed in flasks



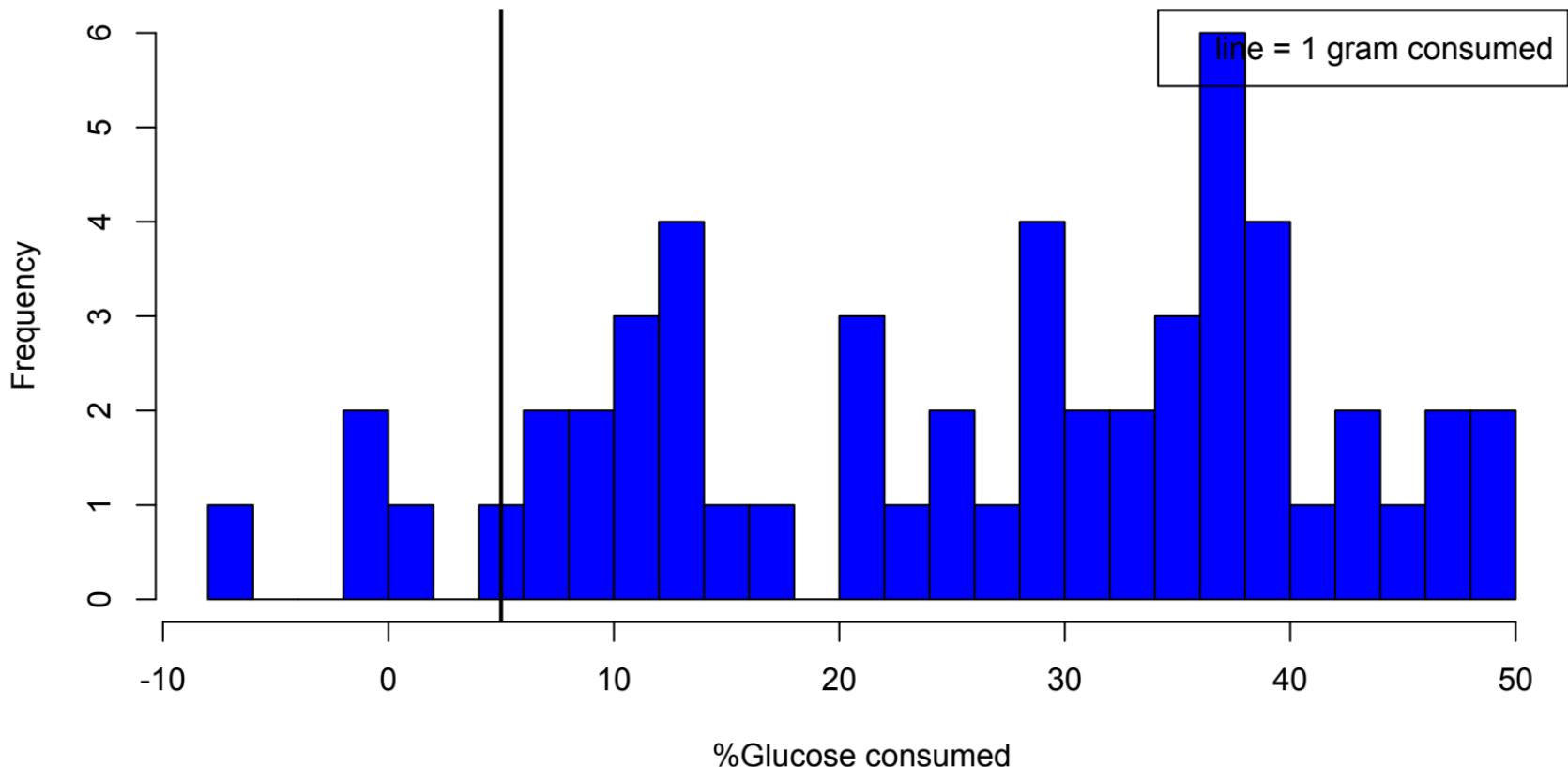
% Glucose consumed in tubes



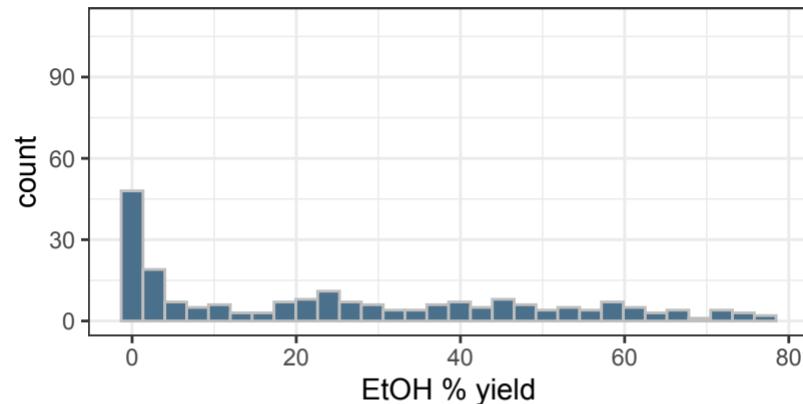
% Glucose consumed in flasks
spp. that consumed less than 50% of blank (n= 45 spp.)



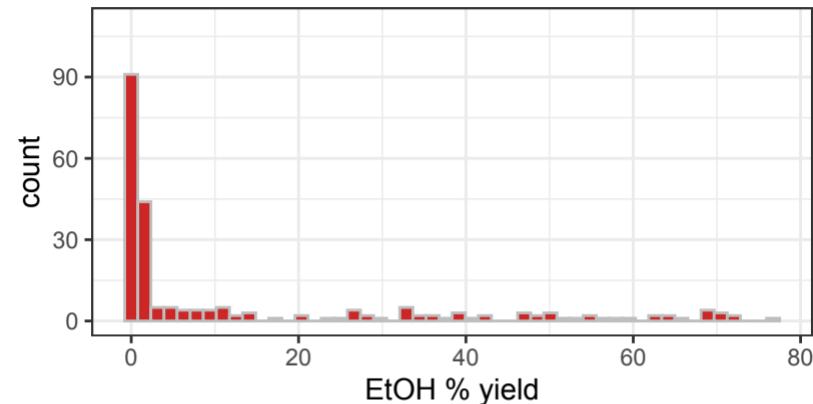
% Glucose consumed in tubes
spp. that consumed less than 50% of blank (n= 54 spp.)



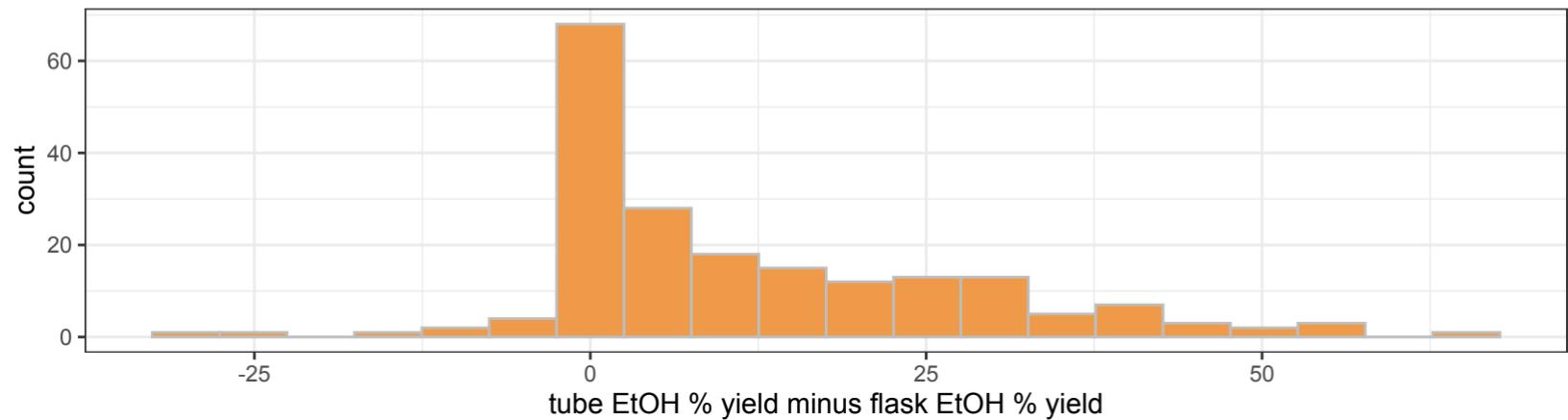
culture tube
n= 212 spp.



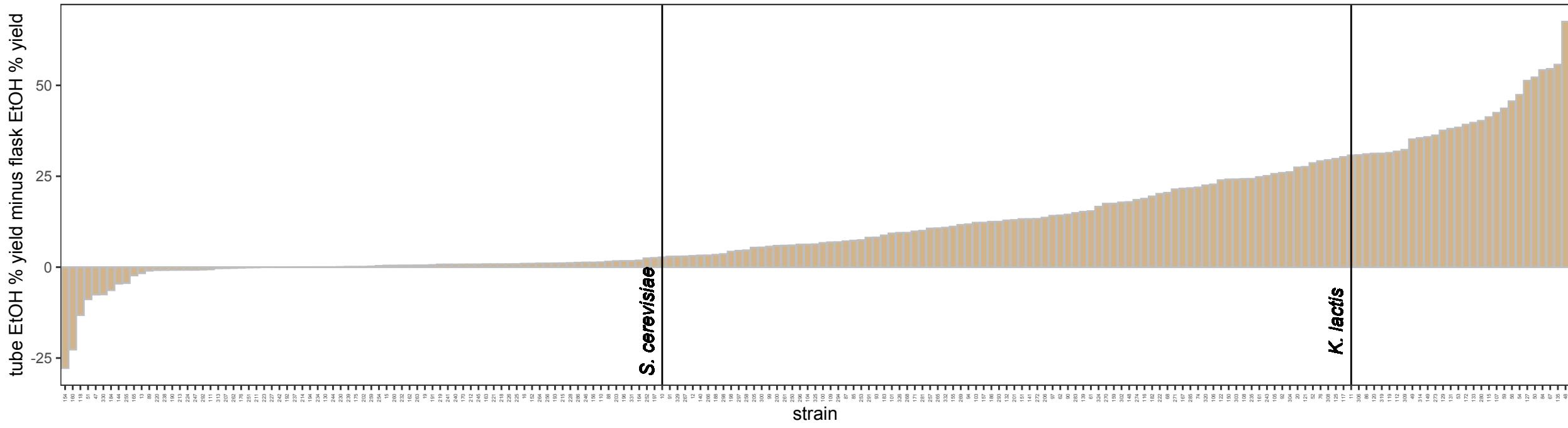
baffled flask
n=225 spp.



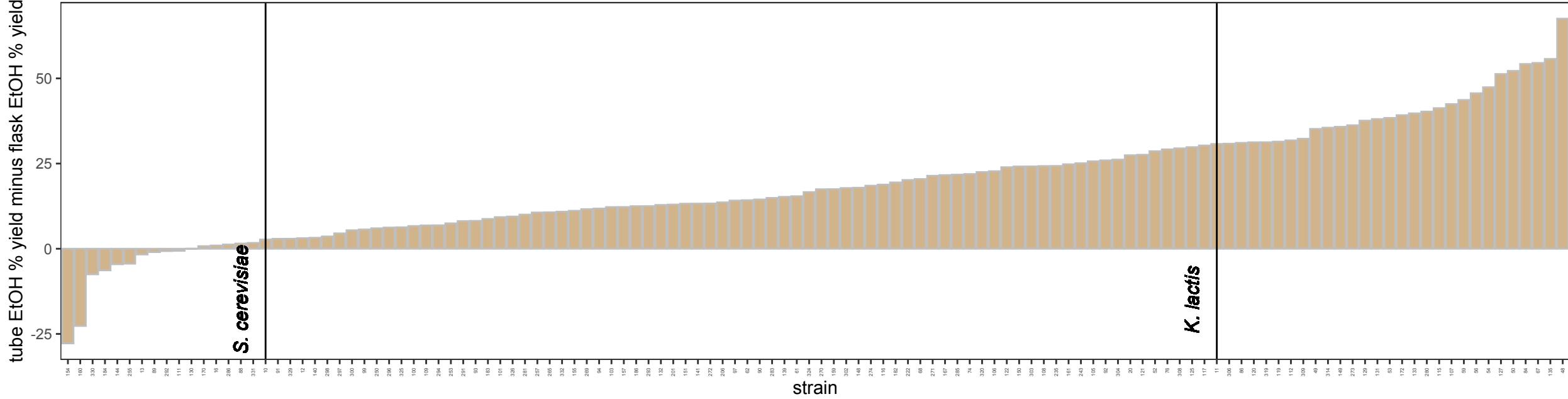
difference in EtOH % yield in low vs high O₂ n= 197 spp.



Difference in EtOH production between conditions
all 197 spp. that consumed >50% glucose



Difference in EtOH production between conditions
all 122 spp. that consumed >50% glucose in both conditions
and produced >10% yield EtOH in tubes



Goal: to categorize into biologically relevant groupings based on difference between tube and flask fermentation.

Data input: The %EtOH yield from tubes minus the same data from flasks for each species. Filtered to include only those spp. that consumed >50% of glucose in both conditions and produced >10% EtOH in tubes (evidence of fermentation in tubes).

Strategy: Jenk's natural breaks algorithm – algorithmically identifies natural breaks in 1-dimensional data distributions. Produces Goodness of Fit value.

- The best fit to a distribution of n data points is n groups – each data point as its own group would me the maximum goodness of fit.
- We want to pull out categories without artificially overfitting the data.
- Right now I'm using the lowest number of groupings with a goodness of fit >.90.

Jenks natural breaks optimization; number of classes: 5

