Improved normalization of species count data in ecology by scaling with ranked subsampling (SRS): application to microbial communities.

**CONCLUSIONS**

Normalization of OTU or species counts by scaling with ranked subsampling preserves the original community structure by minimizing subsampling error. We therefore propose SRS for the normalization of biological count data.

**BACKGROUND**

Analysis of species count data in ecology often requires normalization to an identical sample size. Rarefying (random subsampling without replacement), which is the standard method for normalization, incurs sampling error, impairing the reproducibility and potentially distorting the community structure.

**RESULTS**

SRS showed greater reproducibility and preserved OTU frequencies and alpha diversity better than rarefying. The variance in Shannon diversity increased with the reduction of the library size after rarefying, but remained zero for SRS. Relative abundances of OTUs strongly varied among libraries generated by rarefying, whereas libraries normalized by SRS showed only negligible variation.

**METHODS**

SRS consists of two steps:

1. The scaling step. The counts for all species or operational taxonomic units (OTUs) are divided by a scaling factor chosen in such a way that the sum of scaled counts equals the selected total number of counts, \( C_{\min} \). The relative frequencies of all OTUs remain unchanged.

2. Ranked subsampling step. Non-integer count values are converted into integers by an algorithm that minimizes subsampling error with regard to the population structure (relative frequencies of species or OTUs) while keeping the total number of counts equal \( C_{\min} \).

**CONCLUSIONS**

Here we introduce a normalization method for species count data called scaling with ranked subsampling (SRS) and demonstrate its suitability for analyzing microbial communities.