Supplementary information on methods

a) Implications of primer specificity and gene copy number variation

Taxon coverage in community analyses is always affected by certain biases which specifically affect the presence/absence and the absolute abundance of taxa. This holds true for microscopic as well as for molecular surveys. In molecular community analyses, primer selectivity (Stoeck et al. 2006) and gene copy number variation (Medinger et al. 2010) are among the dominant factors affecting actual sequence coverage and sequence abundance. Effects of gene copy number variation particularly result in an overestimation of alveolate taxa as these groups have a comparatively high copy number of the ribosomal operon. The generally high proportion of alveolate sequences, specifically of Ciliophora, is a well-known artefact in molecular data sets (Dyal et al. 1995, Zhu et al. 2005, Medinger et al. 2010). Similarly, the effect of primer selectivity has also been addressed in several studies and it is well known that different primers select for different sequences (Stoeck et al. 2006, Jeon et al. 2008). Both aspects introduce mainly systematic errors, i.e. distinct sequences are over- or underrepresented in a similar pattern across samples. In turn, the partitioning of rare and abundant taxa is similarly affected to some extent. However, as both rare and abundant taxa should be affected, we do not expect a significant impact on the analysis of patterns across samples. This is supported by the fact that taxonomic groups which are generally overestimated (such as ciliates) as well as habitat types of different diversity and taxon composition show similar trends with respect to the pattern shift between rare and abundant sequences. Although the absolute abundance of taxa cannot be measured and the absence or presence of sequence types cannot be categorically stated (Jeon et al. 2008), relative abundance shifts and patterns across samples can be analysed robustly.

Dyal PL, Hope S, Roberts DM, Embley TM. (1995). Use of the PCR and fluorescent-probes to recover SSU ribosomal-RNA sequences from single cells of the ciliate protozoan Sphatitium. *Molecular Ecology* 4:499–503.

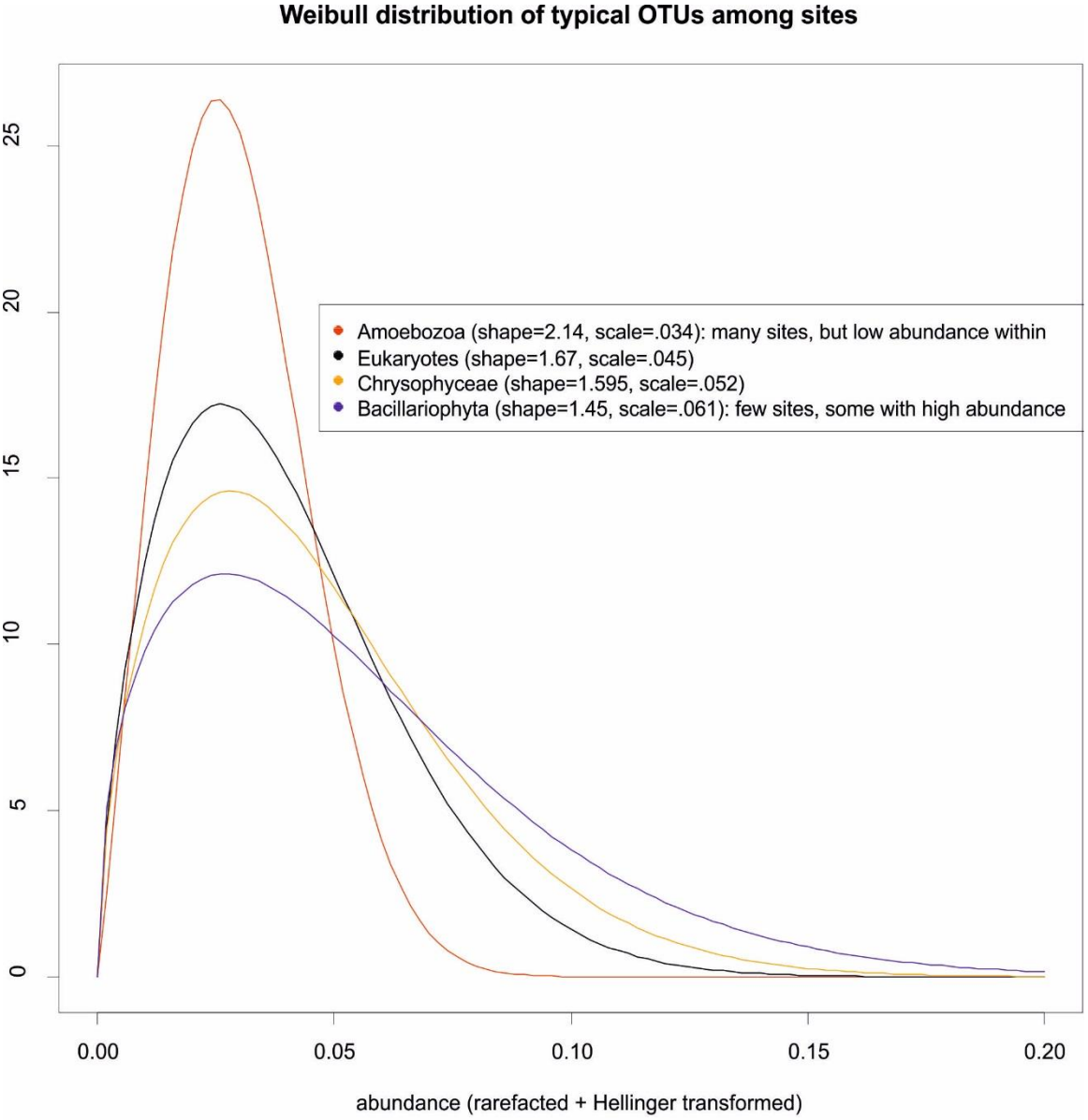
Joen S, Bunge J, Leslin C, et al. (2008). Environmental rRNA inventories miss over half of protistan diversity. *BMC Microbiology* 8:222.

Medinger R, Nolte V, Panday RV, Jost S, Ottenwälder B, Schlötterer C, Boenigk J. (2010). Diversity in a hidden world: potential and limitation of next generation sequencing for surveys of molecular diversity of Eukaryotic microorganisms. *Mol Ecol* 19:32-40.

Stoeck T, Hayward B, Taylor GT, et al. (2006). A multiple PCR-primer approach to access the microeukaryotic diversity in environmental samples. *Protist* 157:31-43.

Zhu F, Massana R, Not F, Marie D, Vaulot D. (2005). Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiology Ecology* 52:79-92.

b) Weibull distribution



Weibull distribution curves for 4 typical OTUs (based on median values of shape parameter *k* and scale parameter l [lambda]) from different taxonomic groups. Note that the skewness of the curves, which is the crucial feature concerning generalists/specialists here, is dependent only on the shape, but not on the scale parameter. This is one of the attractive properties of the 2-parameter Weibull distribution and the reason why Fig. 3 only shows the fitted shape parameters and not the fitted scale parameters. In Fig. 3, the fitting of the parameters plus corresponding KS-tests were conducted not only for 4, but for every single of about 1250 OTUs according to Ricci 2005. For values out of the range of our OTUs, the Weibull distribution would switch to an exponential distribution at k<=1, and to a Gaussian formed distribution with skewness=0 at k ~3.6.