A fair comparison

To the Editor: Recently, Paulson *et al.*[1] introduced a normalization method, reporting that it improves clustering of meta-genomic abundance data, which is very important for many applications in the fast-growing area of microbiome research. However, in our view, the perceived improvement is due to a postprocessing procedure that is preferentially combined with some, but not all, normalizations included in their method comparison, rather than to the proposed normalization itself.

Paulson *et al.*[1] compared their normalization method to three existing ones using a data set from a study of microbial communities in the mouse gut and concluded that their method, called cumulative-sum scaling (CSS), "substantially improved" the separation between two known clusters present in the data[1]. As the authors kindly provided us with the source code, we were able to reproduce their first figure (**Supplementary Fig. 1**). However, this was possible only when we applied a logarithm transformation to the data normalized with their CSS method but not to the data normalized by the other methods. Combining the log transformation with each of the normalizations shows that differences in cluster separation are due mainly to this additional transformation and not to the normalization itself (**Fig. 1**). Thus, conceptually simpler methods, such as relative-abundance normalization (also called total-sum scaling (TSS)), should not be dismissed on these grounds.

To understand the large effect of the log transformation on this comparison, it is important to note that it is nonlinear, a feature that can fundamentally change the distribution of the data (skewing reduction, for example). Because the transformation is undefined for input values ≤0, one typically adds a small value (pseudocount) to non-negative input data to avoid log(0). However, owing to the nonlinearity of the log, this value also affects the transformation result (**Supplementary Fig. 2**). Paulson *et al.*[1] set the pseudocount to 1 as a way to preserve zero counts. However, as the four normalizations compared produce output values whose ranges differ by several orders of magnitude, the same pseudocount may not be optimal for all of them. It should instead be chosen to ensure a consistent treatment: for instance, by setting it to a value smaller than the minimum abundance value before transformation (**Supplementary Fig. 2** and **Supplementary Note**).

Methodological improvements are crucial in highly complex fields such as metagenomics. We feel, however, that in a comparison of different approaches, it is important to minimize the potential confounding sources by ensuring equal treatment of all methods under study.

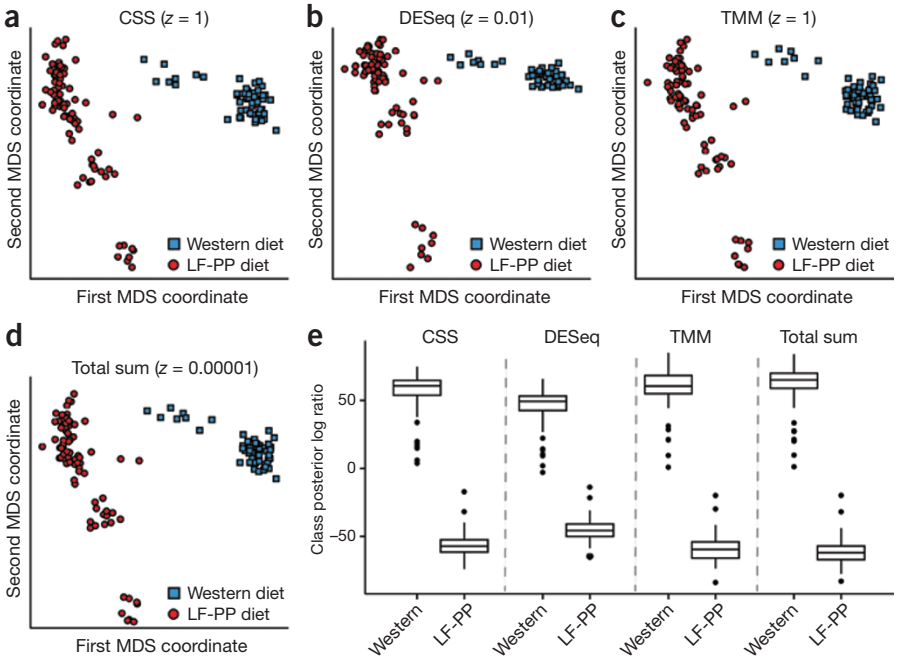

Figure 1 | Clustering analysis of different normalization methods. (**a**–**d**) First two principal coordinates of multidimensional-scaling (MDS) analysis of mouse stool data normalized by CSS (**a**), DESeq size factors (**b**), trimmed mean of *M*-values (TMM) (**c**) and total-sum scaling (**d**). The pseudocount (*z*) used with the log transformation is indicated in parentheses (**Supplementary Note**). Colors indicate clinical phenotype (diet). LF-PP, low-fat, plant polysaccharide–rich diet. All normalizations separate samples by diet. (**e**) Class posterior probability log ratio for Western diet obtained from linear discriminant analysis. Each box corresponds to the distribution of leave-one-out posterior probability of assignment to the 'Western' cluster across normalization methods. Samples were optimally distinguished by phenotypic similarity regardless of the method of normalization used. This figure corresponds to Figure 1 in Paulson *et al.*[1] (see also **Supplementary Fig. 1**).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper (doi:10.1038/nmeth.2897).

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Paul I Costea, Georg Zeller, Shinichi Sunagawa & Peer Bork

European Molecular Biology Laboratory, Heidelberg, Germany.
e-mail: bork@embl.de

1. Paulson, J.N., Stine, O.C., Bravo, H.C. & Pop, M. *Nat. Methods* **10**, 1200–1202 (2013).

Paulson *et al.* reply: Costea *et al.*[1] challenge the fairness of the results presented in the first figure of our paper[2], which explored the effect of normalization and transformation procedures on clustering analysis of marker-gene survey