

Abstract

Background

The genomic literature contains many examples of cluster analysis. Most of these appear to be incorrect at no fault of the authors, but rather shortcomings of common cluster and genomic distance software plus the lack of relevant mathematics background on the part of reviewers.

Results

This report illustrates the situation with examples from 4 major publishers across the past 2 decades. The natures of the errors are pinpointed and remedies are provided.

Conclusions

The findings are significant enough to merit elimination of pair-group analysis and distance dendrograms from publication in peer-reviewed journals. Authors should also be required to demonstrate by citation or mathematics that their distance measure is a metric.

Keywords

genetic distance, counting measure, dissimilarity, similarity measure, clades, clustering, nearest neighbor, distance network, pair group