

# The Delta K method alone is not sufficient to determine best clustering solutions for Bayesian analysis of population genetic structure in empirical data sets

Kathryn Stankiewicz<sup>1</sup>, Kate Vasquez<sup>1</sup>, Microsatellite Group<sup>1</sup>, and Iliana Baums<sup>1</sup>

<sup>1</sup>Pennsylvania State University

April 29, 2020

## Abstract

The software program STRUCTURE relies on a Bayesian iterative clustering algorithm to group samples using multi-locus genotype data and is one of the most cited tools for determining population structure. To infer the optimal number of clusters from STRUCTURE output, the  $\Delta K$  method is often applied. However, a recent study relying on modeled microsatellite data suggested that this method has a downward bias in its estimation of K and is sensitive to uneven sampling. If this finding holds for empirical microsatellite datasets, conclusions about the scale of gene flow may have to be revised for a large number of studies. Here, we apply recently described estimators of K to re-estimate gene flow in 41 empirical microsatellite datasets; 15 from a broad range of taxa and 26 focused on a complex study system, coral. These datasets included 35 species, spanning seven continents, from diverse biological systems across the Tree of Life. After comparison of alternative estimates of K (Puechmaille statistics) with traditional ( $\Delta K$  and posterior probability) estimates, we conclude that  $\Delta K$  alone is insufficient for determining the most optimal number of clusters and sampling evenness does not necessarily predict agreement with traditional estimators. To better infer population structure, we suggest a combination of visual inspection of STRUCTURE plots and calculation of the alternative estimators at various thresholds in addition to  $\Delta K$ . Differences between estimators could reveal patterns with important biological implications, such as the potential for more population structure than previously estimated, as was the case for many studies reanalyzed here.

## Hosted file

Msat\_paper\_final\_fig\_embedded\_revcom\_allcoauths\_acceptedChanges.docx available at <https://authorea.com/users/316668/articles/446839-the-delta-k-method-alone-is-not-sufficient-to-determine-best-clustering-solutions-for-bayesian-analysis-of-population-genetic-structure-in-empirical-data-sets>