

MICHIGAN STATE UNIVERSITY

Submission Cover Letter
PLoS Genetics

September 20, 2017

To the Editor(s),

Please find enclosed our manuscript, “Inferring Continuous and Discrete Population Genetic Structure Across Space,” for consideration for publication in PLoS Genetics.

In the manuscript, we present a novel statistical method for the simultaneous inference of continuous and discrete patterns of population structure. The method estimates ancestry proportions for each sample from a set of discrete population layers, and, within each layer, estimates a rate at which relatedness decays with distance. This model explicitly addresses the “clines vs. clusters” problem in modeling population genetic variation by jointly accommodating continuous and discrete patterns of differentiation. We demonstrate the utility of this approach using simulations and two empirical applications. We also release an associated R package, called **conStruct**, for “continuous structure.”

This work is motivated by the failure of existing model-based clustering methods, such as **STRUCTURE** (Pritchard *et al.* 2000) or **ADMIXTURE** (Alexander *et al.* 2009), to model continuous patterns of geographical variation (i.e., isolation by distance, or IBD). Such geographical patterns are ubiquitous in nature, but are described by these clustering methods using a possibly large number of discrete clusters. These clusters are spurious, and can throw a wrench in the gears of efforts to describe and categorize natural variation, one of the first steps in any analysis of genetic data, and a principal mission of biology.



**College of Natural
Sciences**

**Department of
Integrative Biology**

288 Farm Lane
Room 203
East Lansing, MI
48824-1115

Tel: 517-355-4640

Fax: 517-432-2789

integrativebiology@msu.edu
integrativebiology.natsci.msu.edu

Besides presenting a solution to this problem, we use simulations and empirical data to demonstrate the ways in which nonspatial clustering approaches fail in the presence of IBD; and provide two different metrics for choosing an appropriate number of layers to describe the data.

As of this submission, **STRUCTURE** alone has been cited more than 20,000 times. However, as the costs of sequencing continue to fall and good geospatial information becomes more readily available, the limitations we cite above will become more severe. Our method provides an important extension to **STRUCTURE** and similar approaches, one that solves a critical and widely known basic conceptual drawback of the model. We therefore believe that our work represents an important and timely contribution to the field.

John Novembre (jnovembre@uchicago.edu),
Desislava Petkova (desislavka@gmail.com),
Jonathan Pritchard (pritch@stanford.edu),
Sohini Ramachandran (sramachandran@brown.edu),
and Matthew Stephens (mstephens@uchicago.edu)
would be suitable reviewers or external editors. (as Coop is an associate editor).

Sincerely, Gideon Bradburd, Graham Coop, and Peter Ralph