MICHIGAN STATE
U N I V E R S I T Y

Submission Cover Letter
PLoS Biology

To the Editor(s),

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

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 both simulations and empirical applications, with datasets from poplars and black bears. We also release an associated R package, called `conStruct`, for "*con*tinuous *struct*ure."

This work is motivated by two observations: first, that patterns of isolation by distance (IBD) are ubiquitous in nature; and second, that existing model-based clustering methods, such as `STRUCTURE` (Pritchard *et al.* 2000) or `ADMIXTURE` (Alexander *et al.* 2009), can only describe continuous patterns of variation using discrete clusters, and so must use a possibly large number of clusters to describe continuous population differentiation. Inference of these spurious clusters 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

In this manuscript, we use simulations and empirical data to demonstrate the ways in which nonspatial clustering approaches fail in the presence of IBD; we also present two different metrics for choosing an appropriate number of layers to describe the data, and, using both, are able to establish that our method performs quite well.

As of this submission, `STRUCTURE` alone has been cited more than 20,000 times, and the falling costs of sequencing are facilitating an explosion of next-generation sequence datasets. The first step in the analysis of these data will, for most researchers, involve a model-based clustering analysis. 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.

Sincerely, Gideon Bradburd, Graham Coop, and Peter Ralph