Inferring Continuous and Discrete Population Structure

Gideon Bradburd
gbradburd@ucdavis.edu
@gbradburd
genescape.org
UC Davis

Patterns of population differentiation

discrete

Greenish Warblers

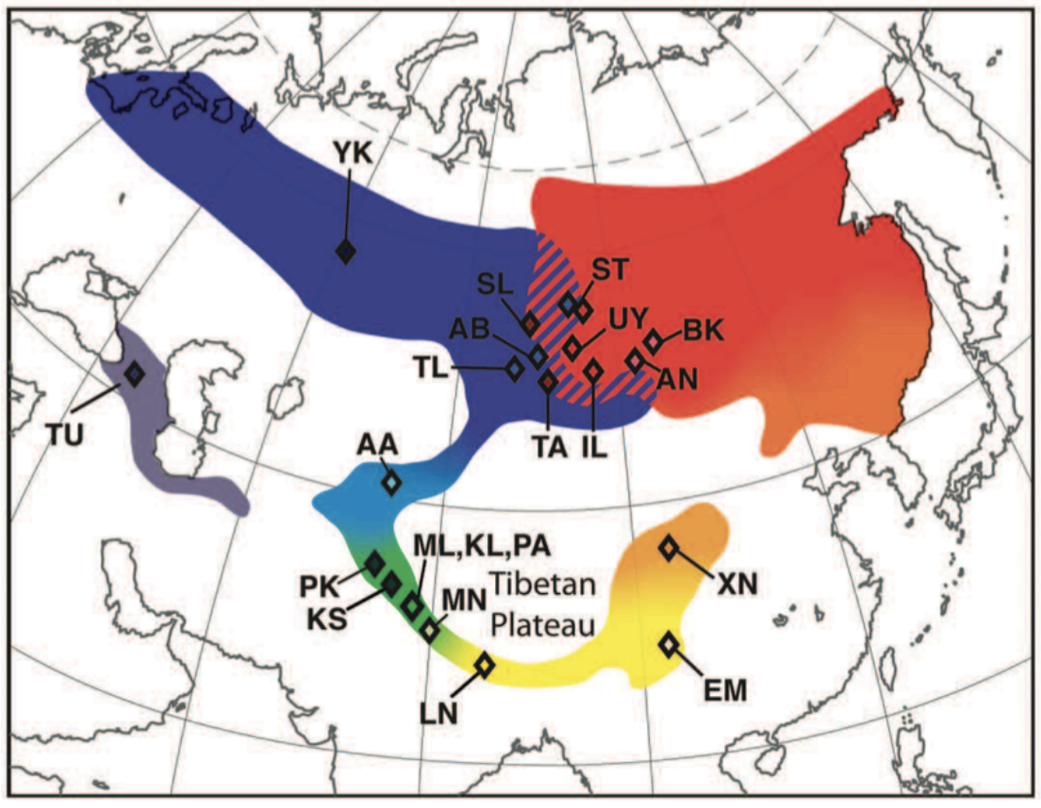
photo credit: Darren Irwin

Patterns of population differentiation

discrete

Greenish Warblers

continuous



Phylloscopus trochiloides subspecies

photo credit: Darren Irwin

Alcaide *et al.* 2014

Patterns of population differentiation

discrete & continuous

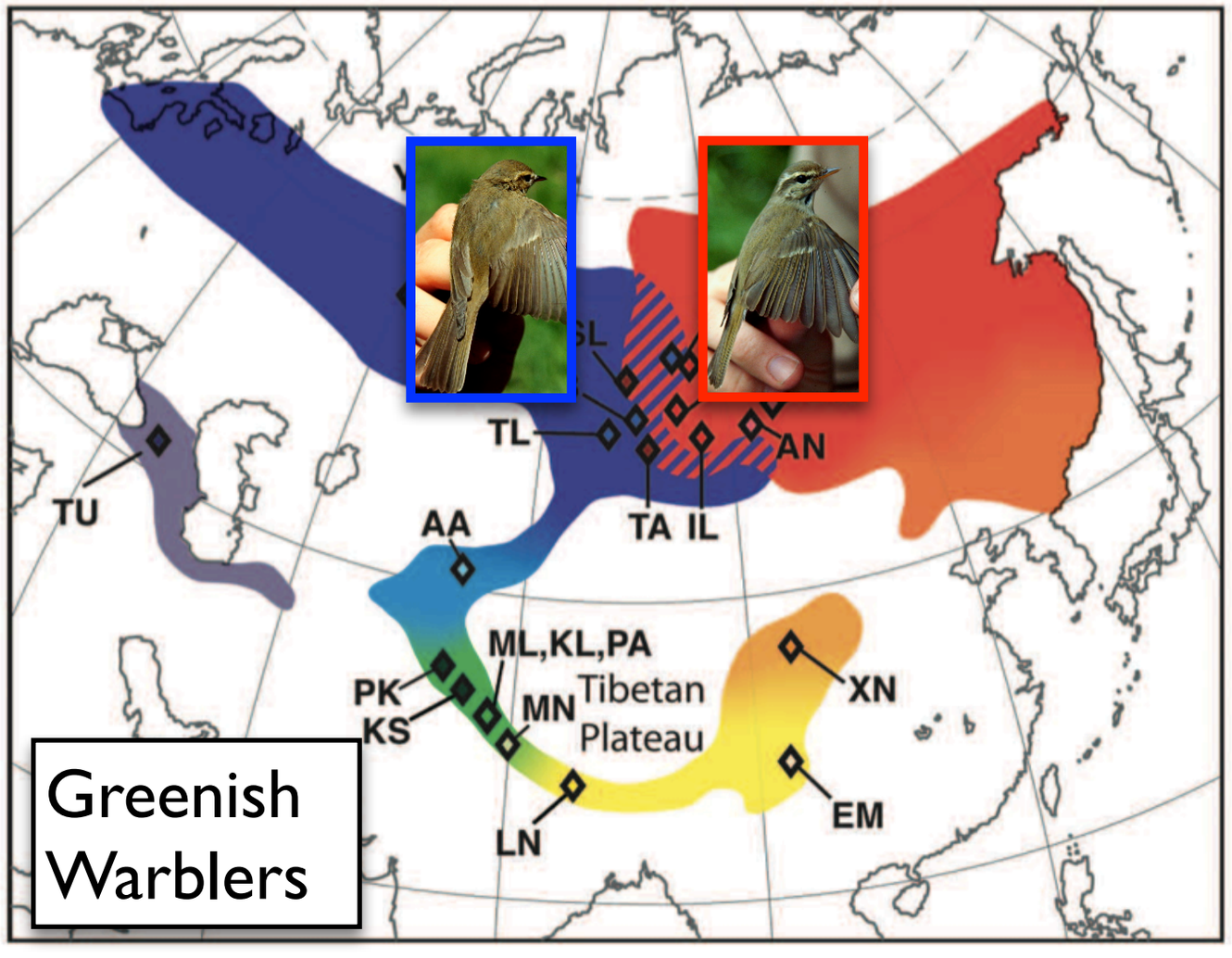
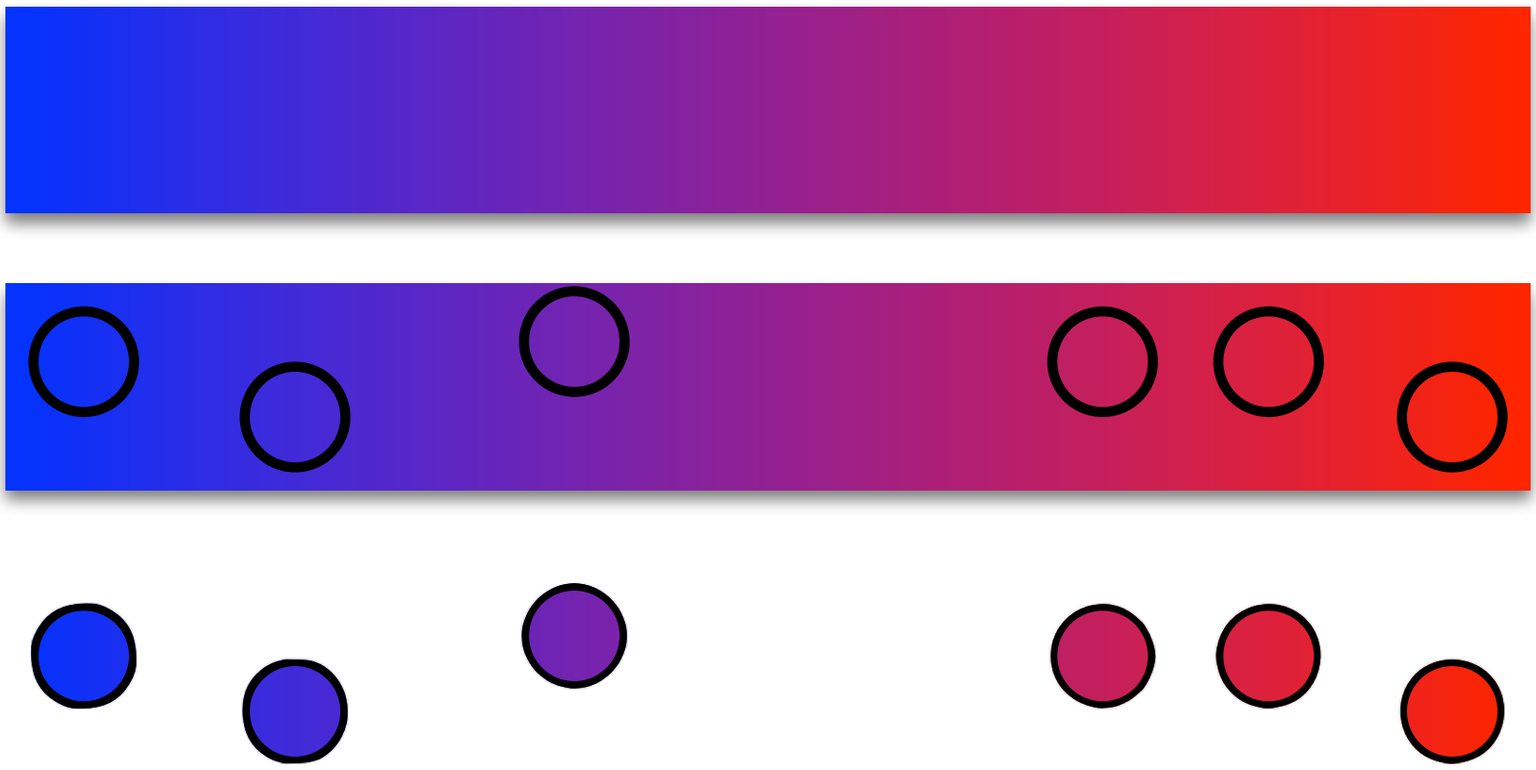


photo credit: Darren Irwin

Alcaide *et al.* 2014

Classic problem in inference and visualization of population structure:

Clines vs. Clusters



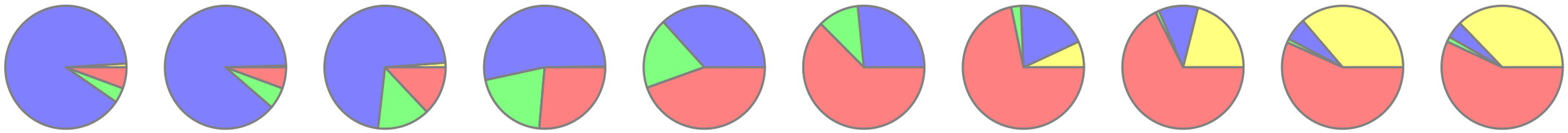
Rosenberg *et al.* 2002
Serre and Pääbo 2004
Rosenberg *et al.* 2005

geoStructure

a new method for modeling
continuous and discrete
differentiation

geoStructure

a new method for modeling
continuous and discrete
differentiation



1) Motivate with an empirical example

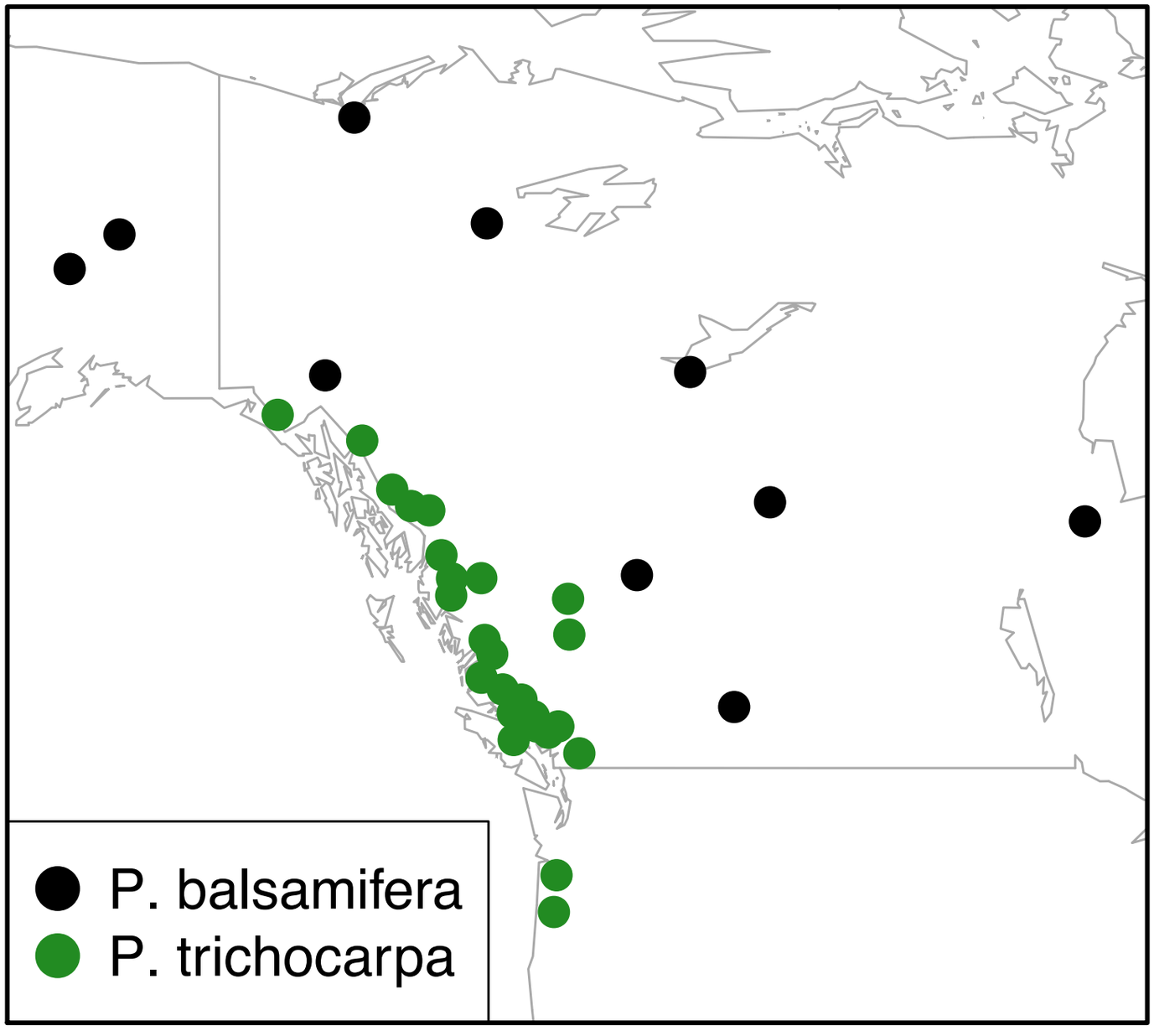
2) Method mechanics

3) Method application

Empirical Illustration:

Populus trichocarpa

Populus balsamifera



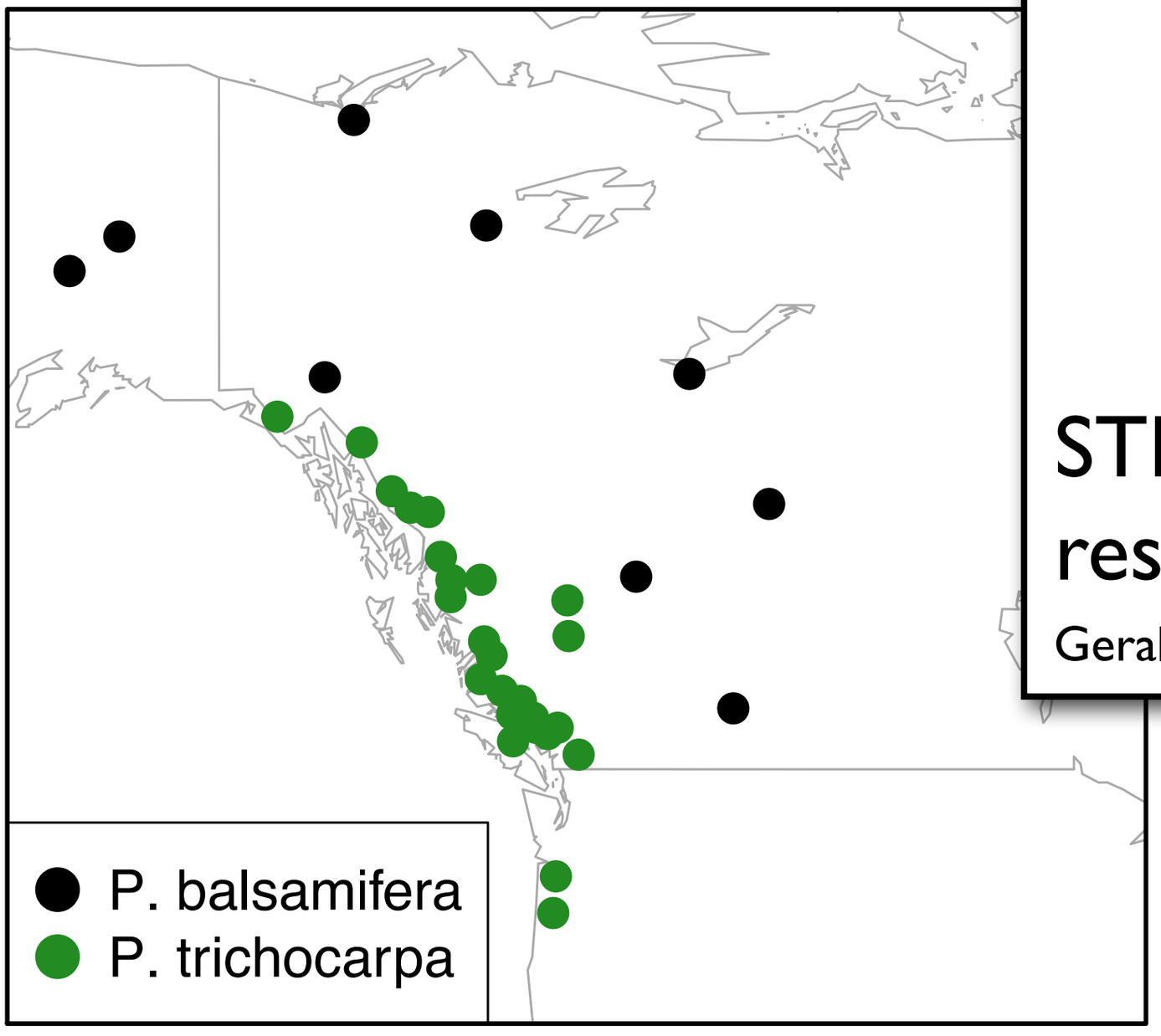
Geraldes *et al.* 2013

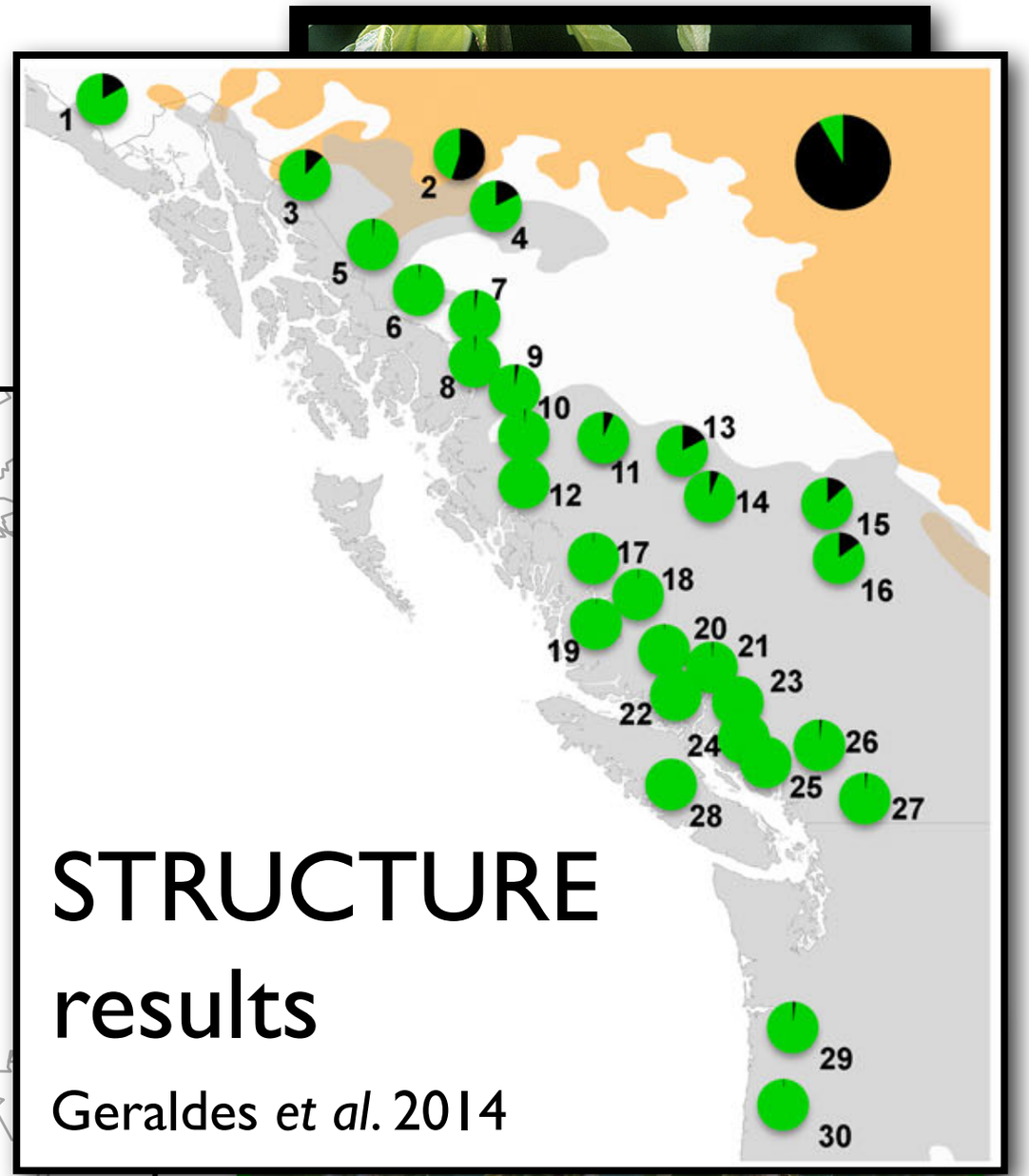
Geraldes *et al.* 2014

Empirical Illustration:

Populus trichocarpa

Populus balsamifera





STRUCTURE results

Geraldes *et al.* 2014

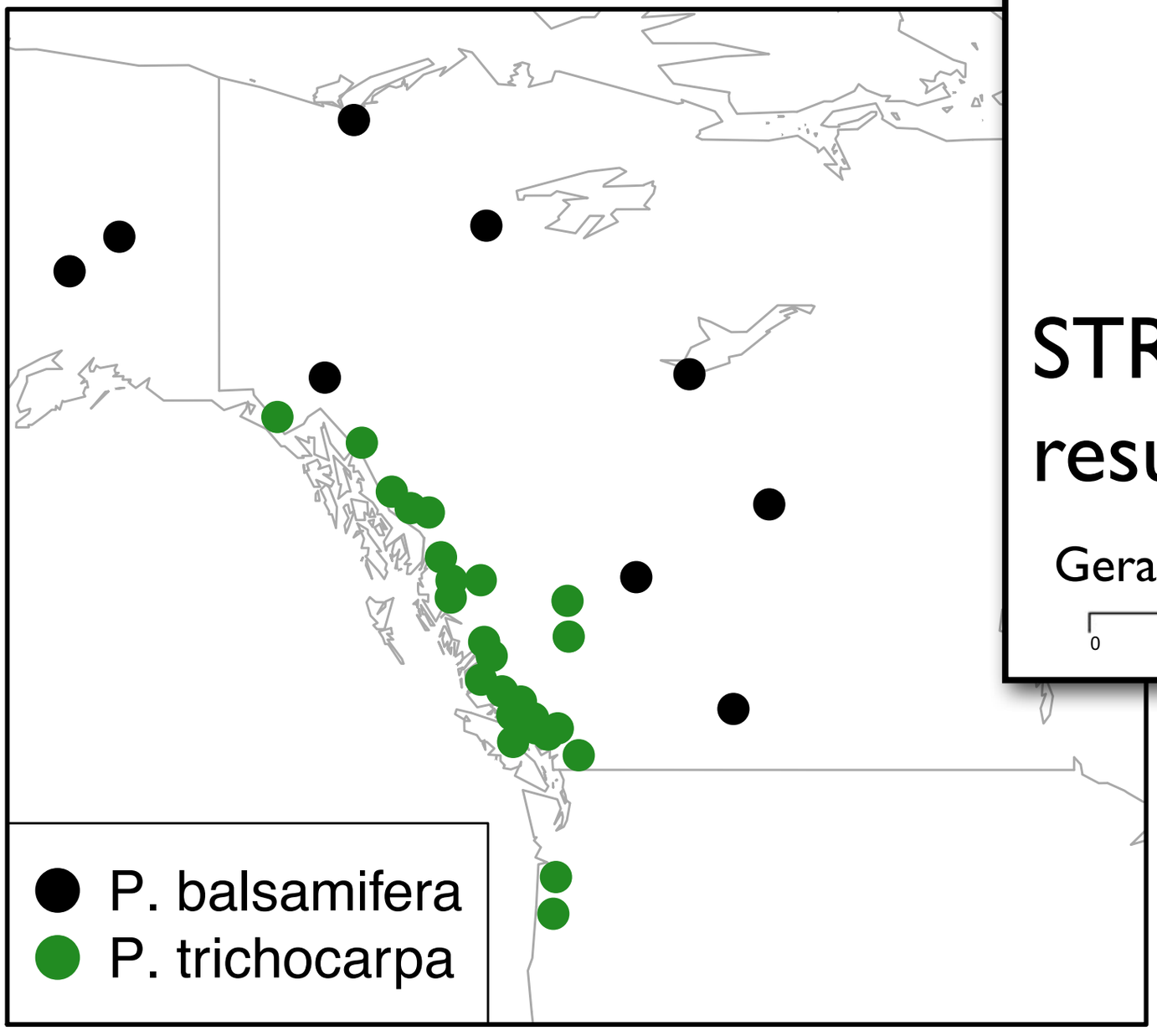
Geraldes *et al.* 2013

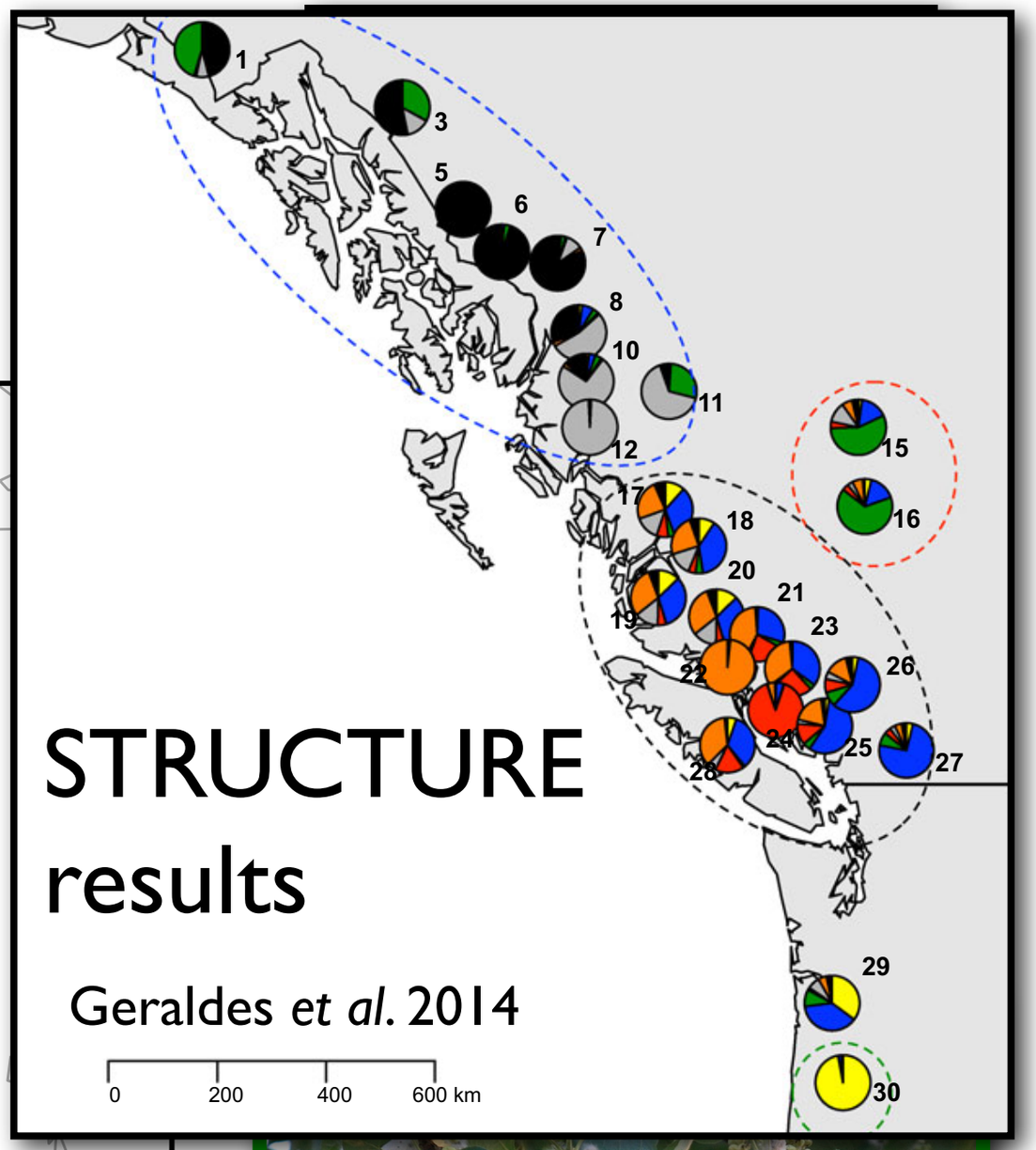
Geraldes *et al.* 2014

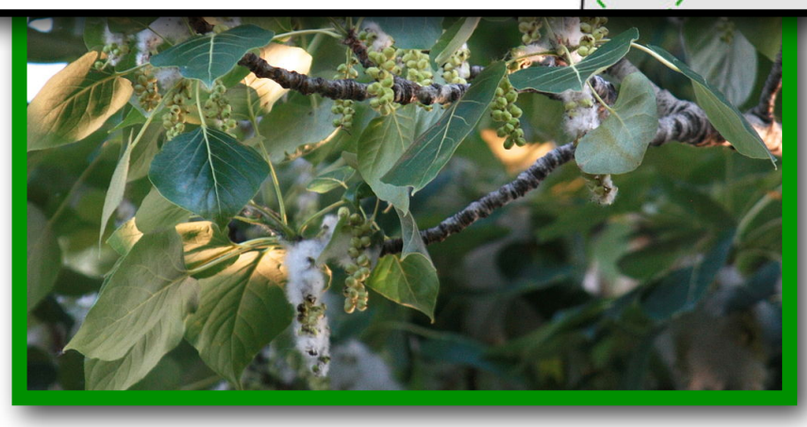
Empirical Illustration:

Populus trichocarpa

Populus balsamifera





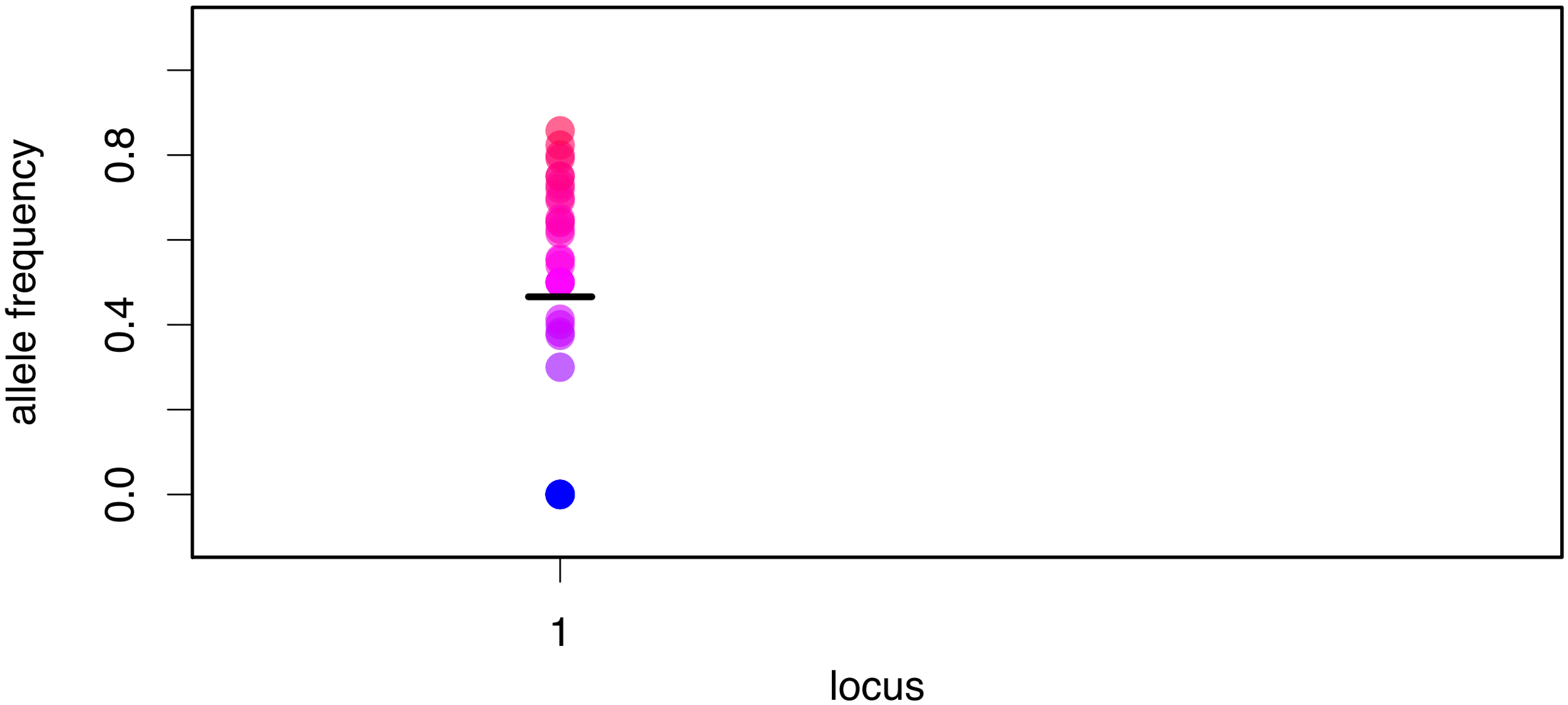


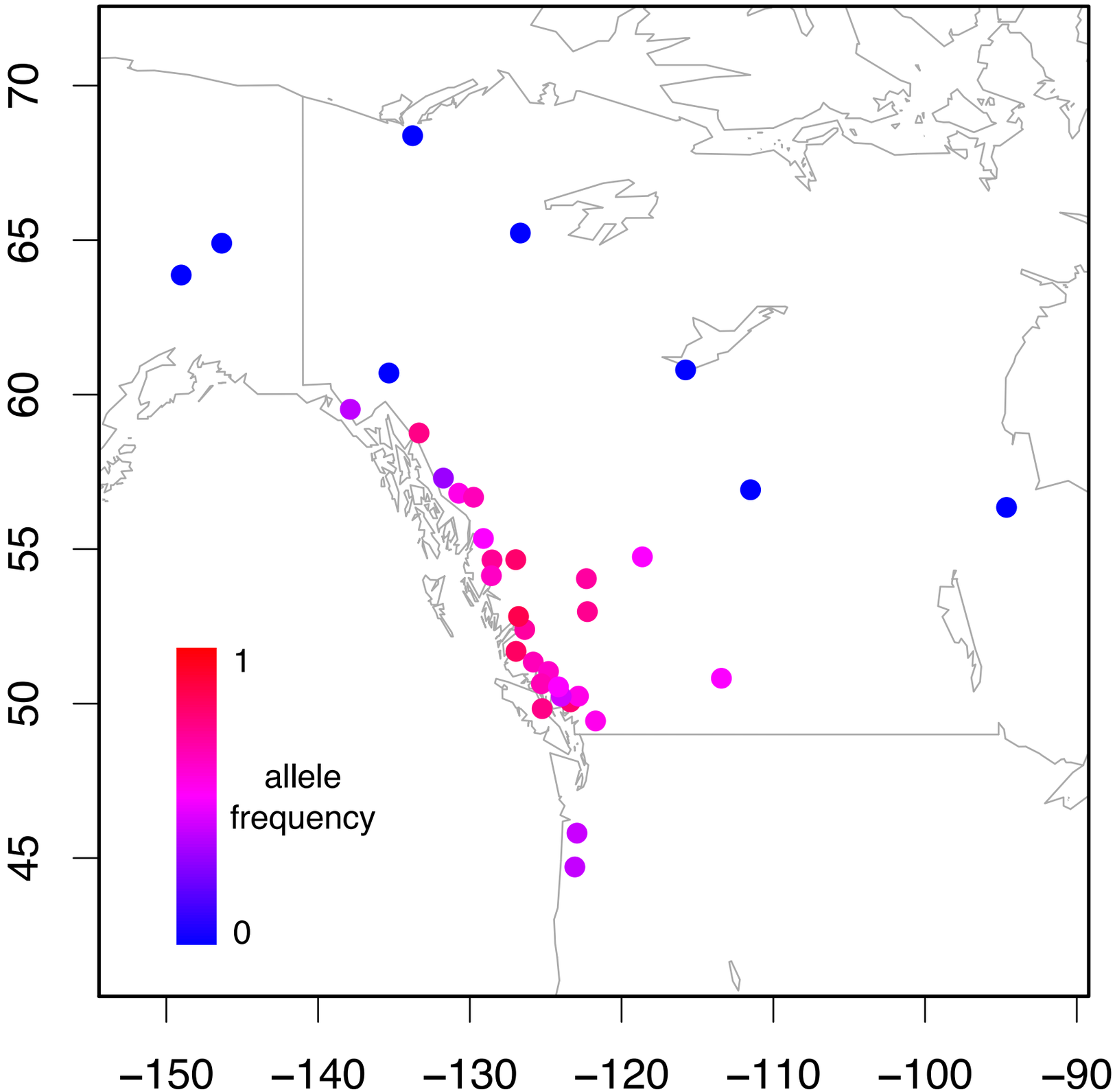
Geraldes *et al.* 2013

Geraldes *et al.* 2014

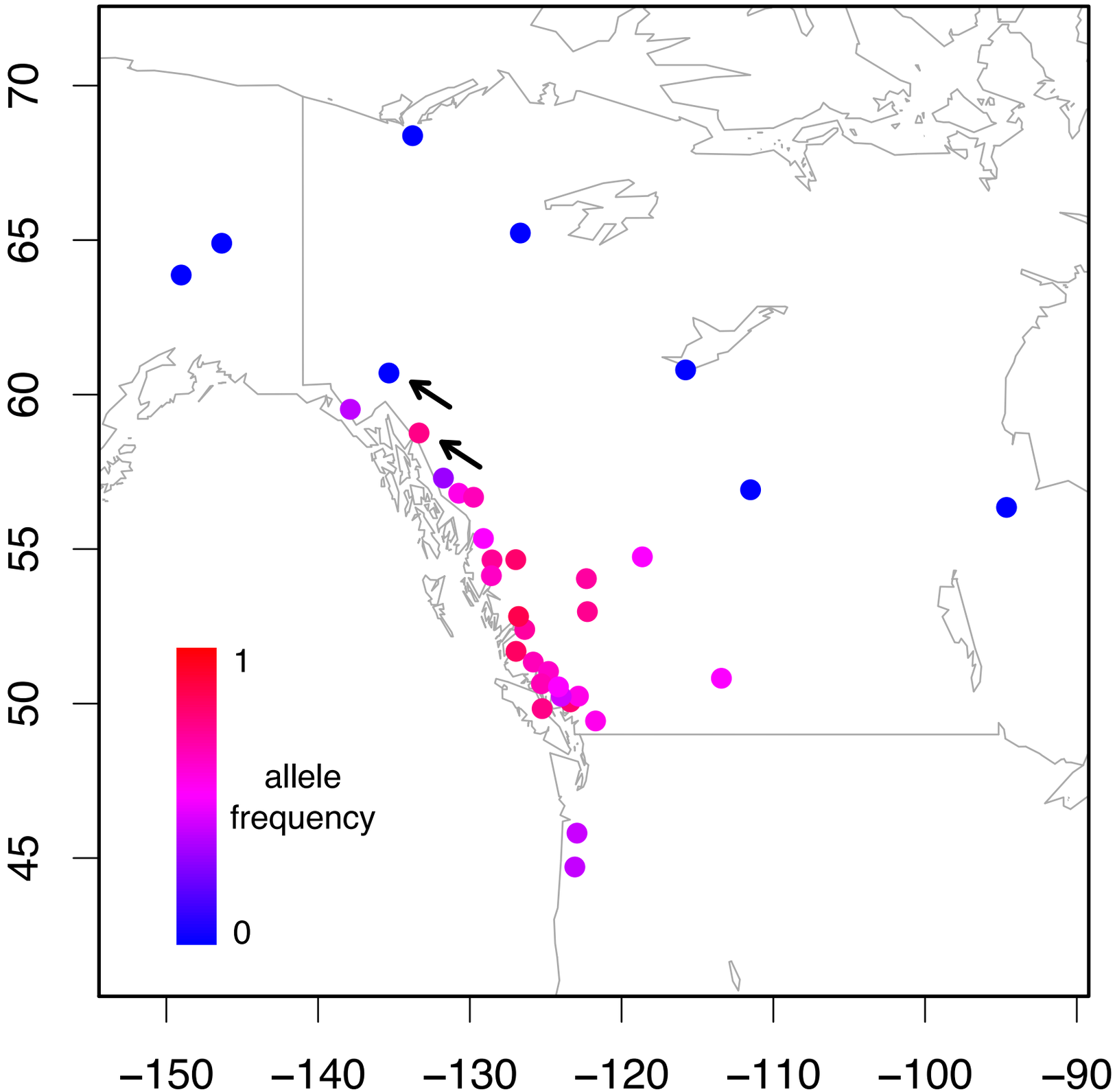
Empirical Illustration: Populus

Allele frequencies at one locus

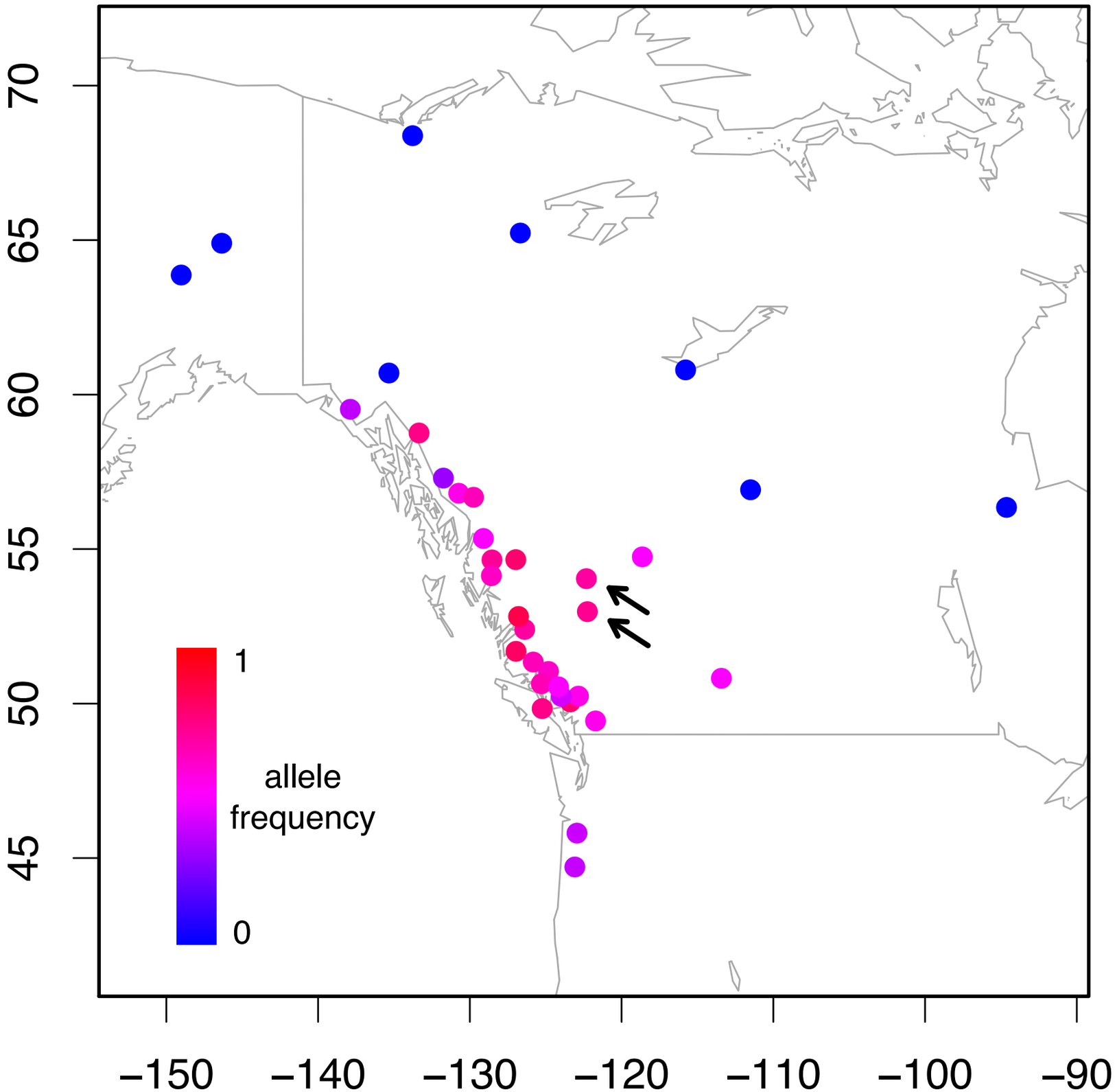


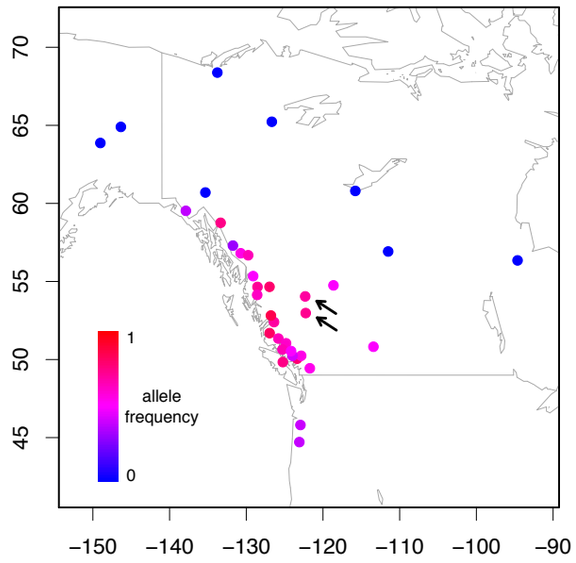

70
65
60
55
50
45
1
allele
frequency
0
−150
−140
−130
−120
−110
−100
−90

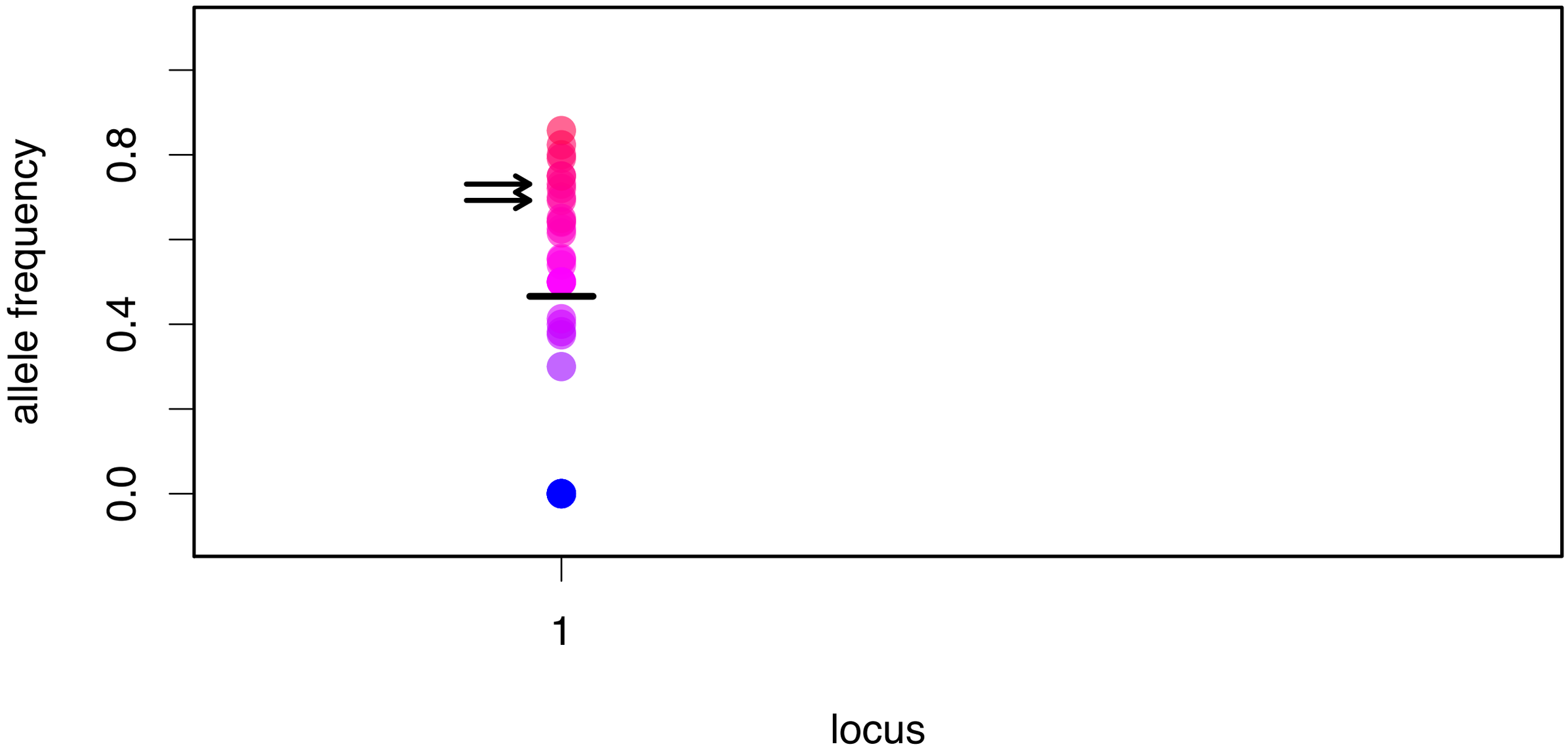
discrete



continuous

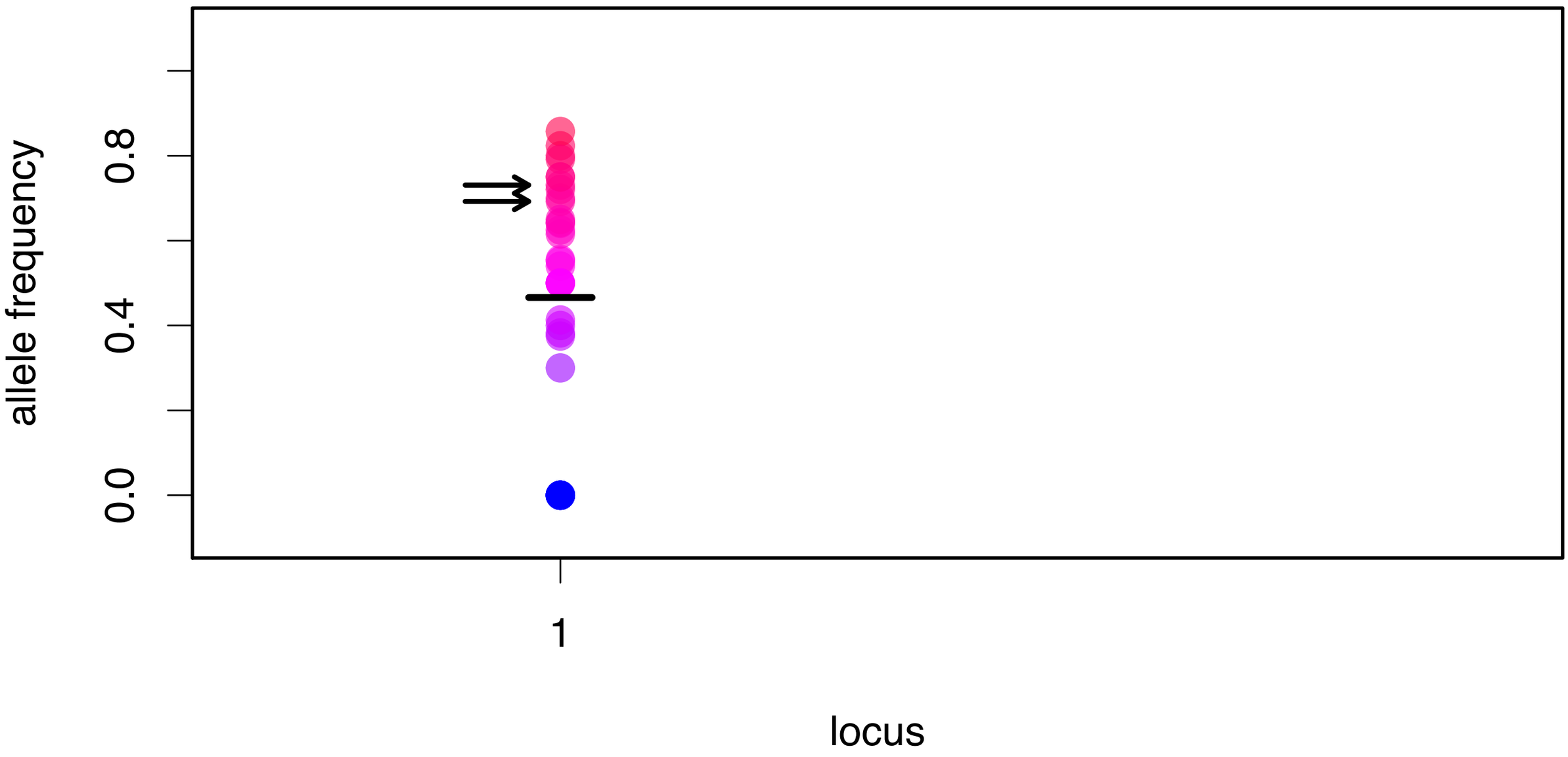



70
65
60
55
50
45
1
allele
frequency
0
−150
−140
−130
−120
−110
−100
−90


allele frequency
0.0
0.4
0.8
1
locus

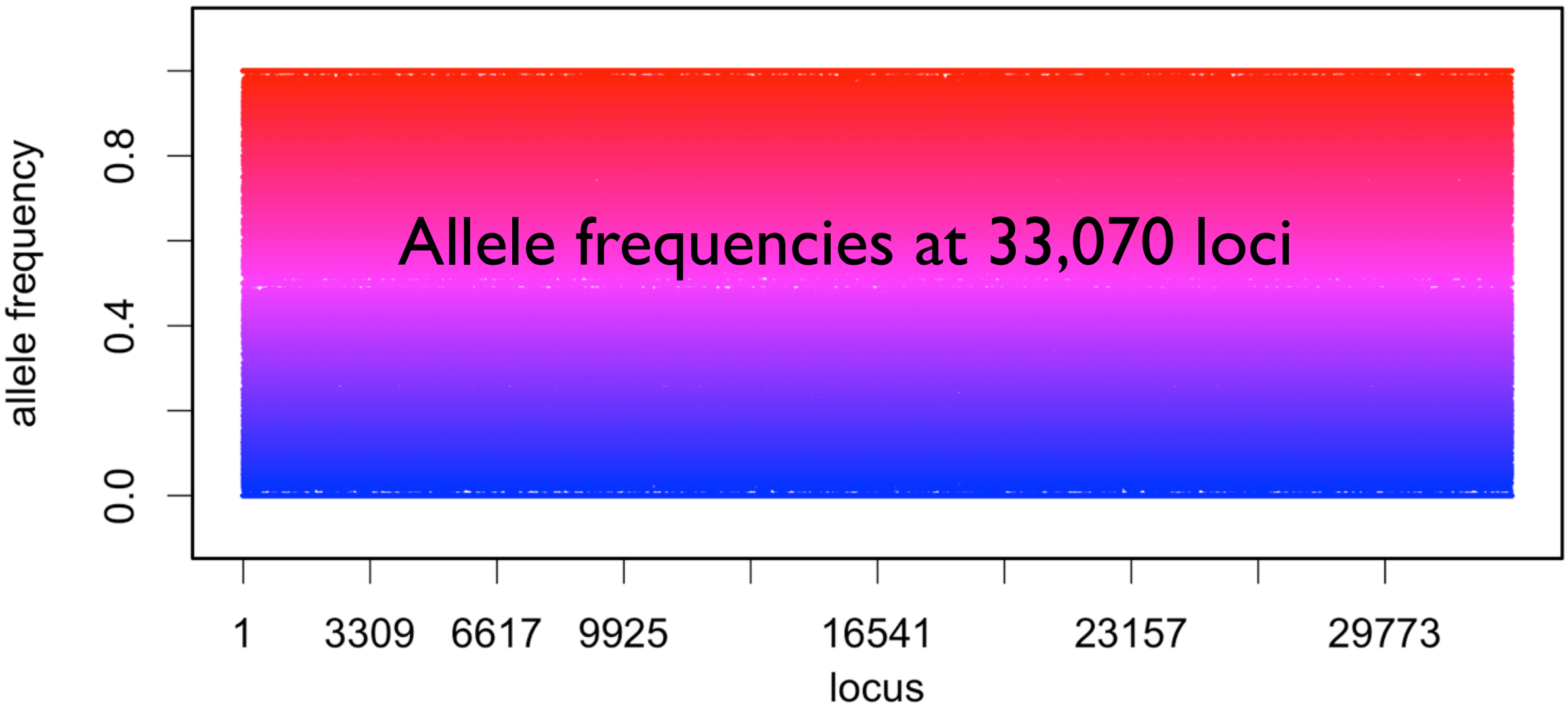
Sample covariance in allele frequencies

$$\widehat{\Omega}_{i,j} = \frac{1}{L} \sum_{\ell=1}^{L} \left(\hat{f}_{i,\ell} - \bar{f}_{\ell} \right) \left(\hat{f}_{j,\ell} - \bar{f}_{\ell} \right)$$

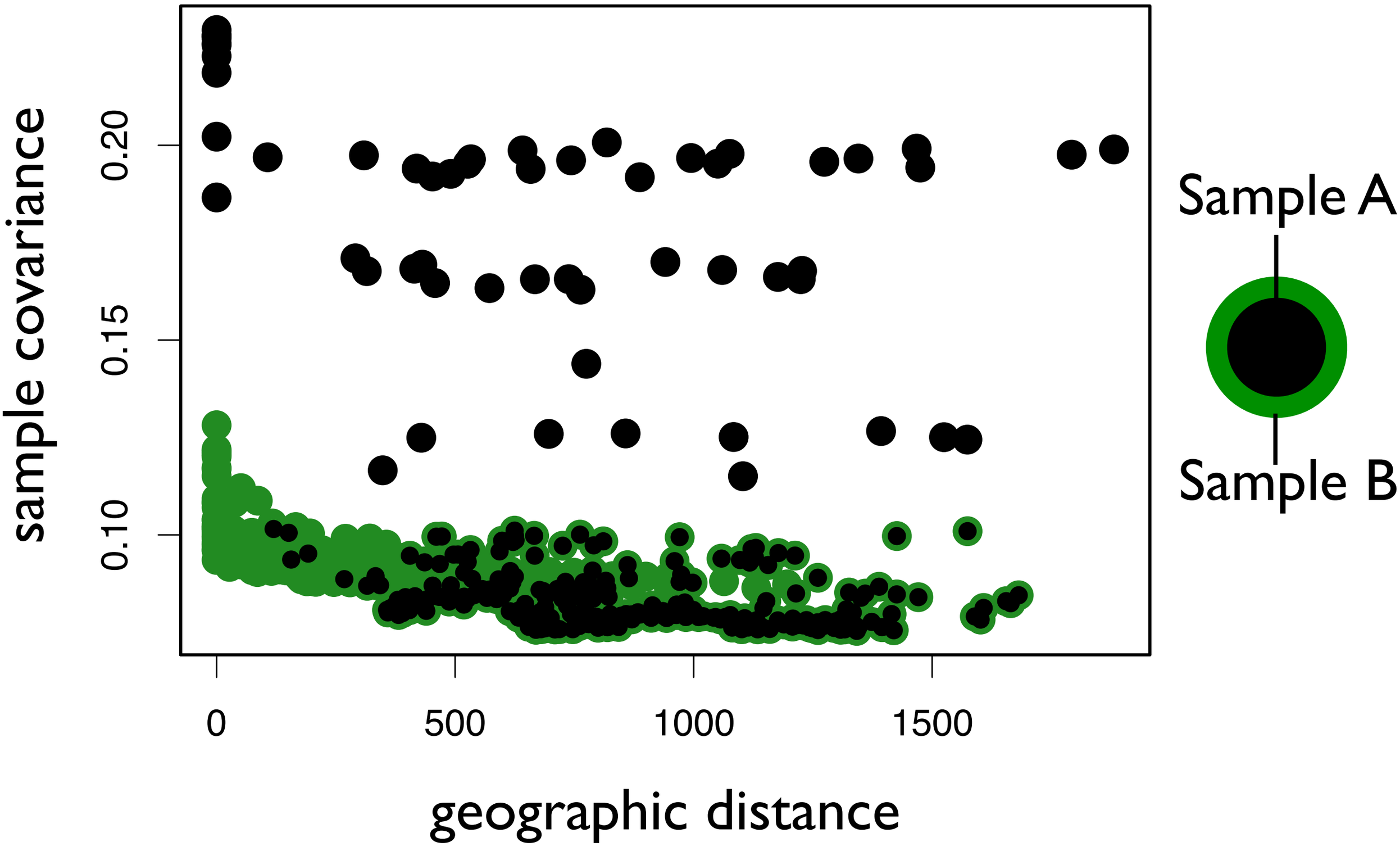


Sample covariance in allele frequencies

$$\widehat{\Omega}_{i,j} = \frac{1}{L} \sum_{\ell=1}^{L} \left(\hat{f}_{i,\ell} - \bar{f}_{\ell} \right) \left(\hat{f}_{j,\ell} - \bar{f}_{\ell} \right)$$



Populus trichocarpa and *Populus balsamifera*



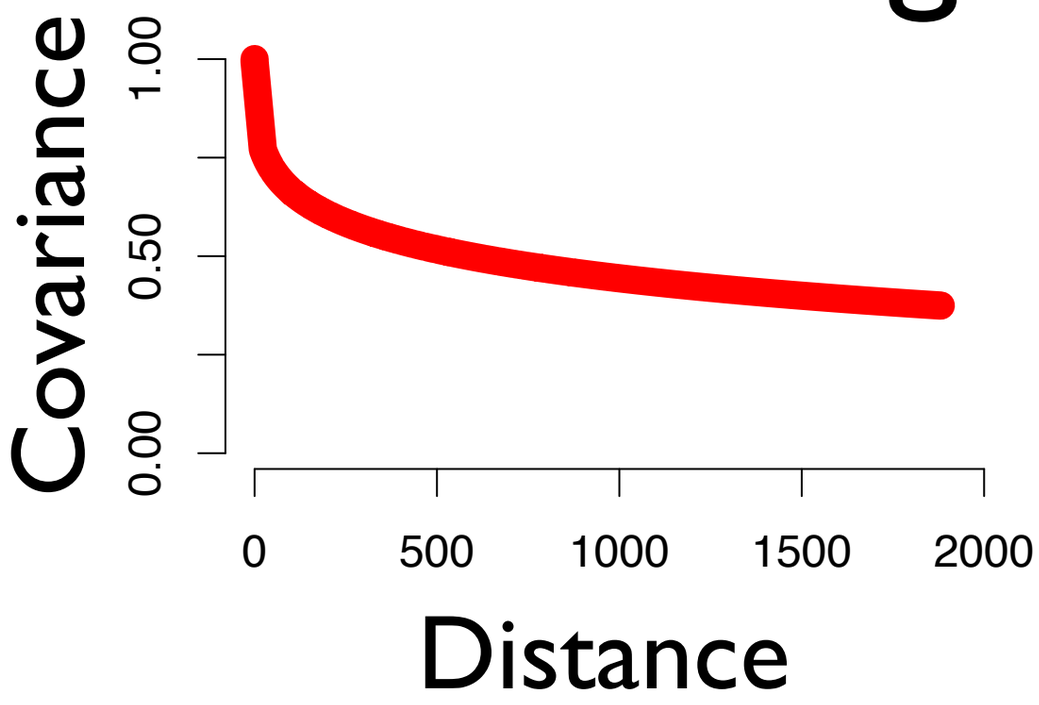
geoStructure mechanics

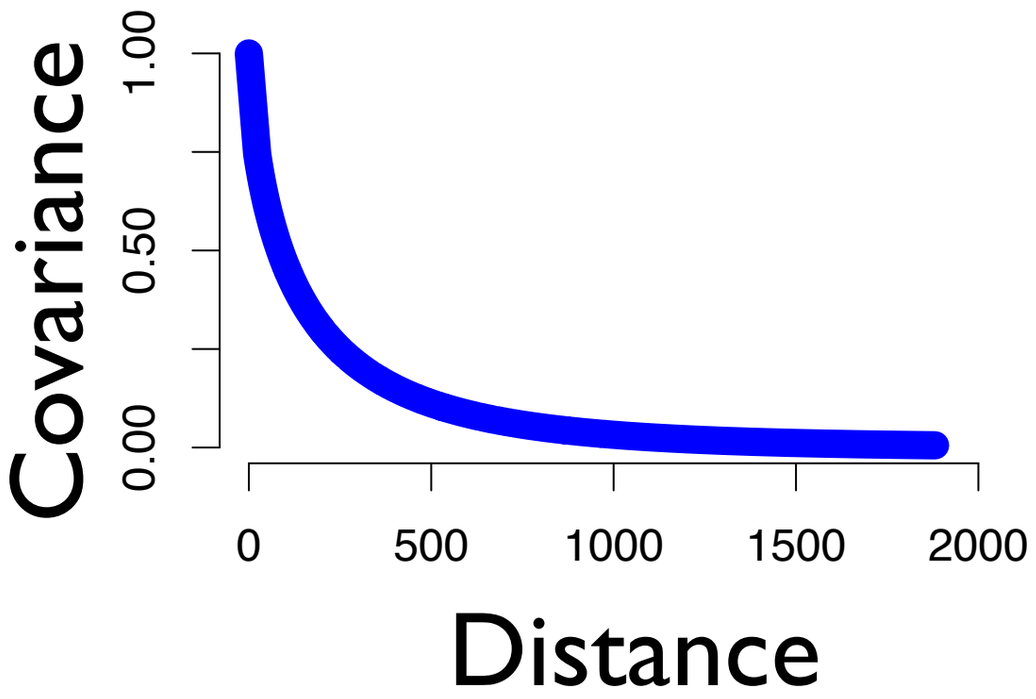
Model allele frequency covariance:

- discrete population clusters with
- continuous differentiation within clusters

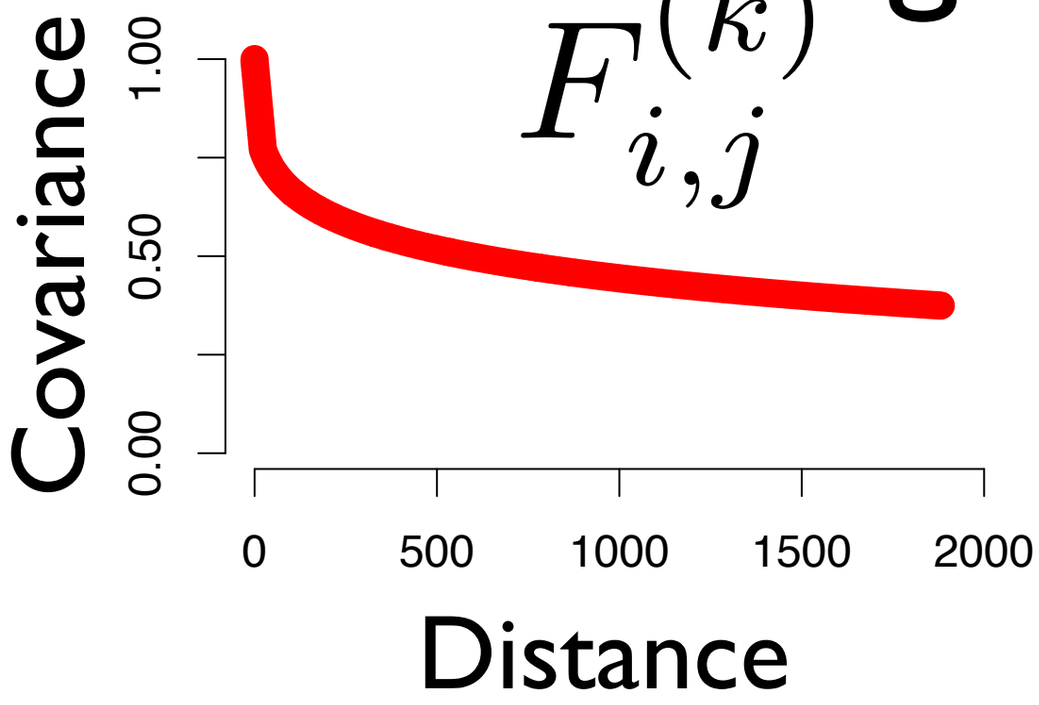
geoStructure mechanics

geoStructure mechanics

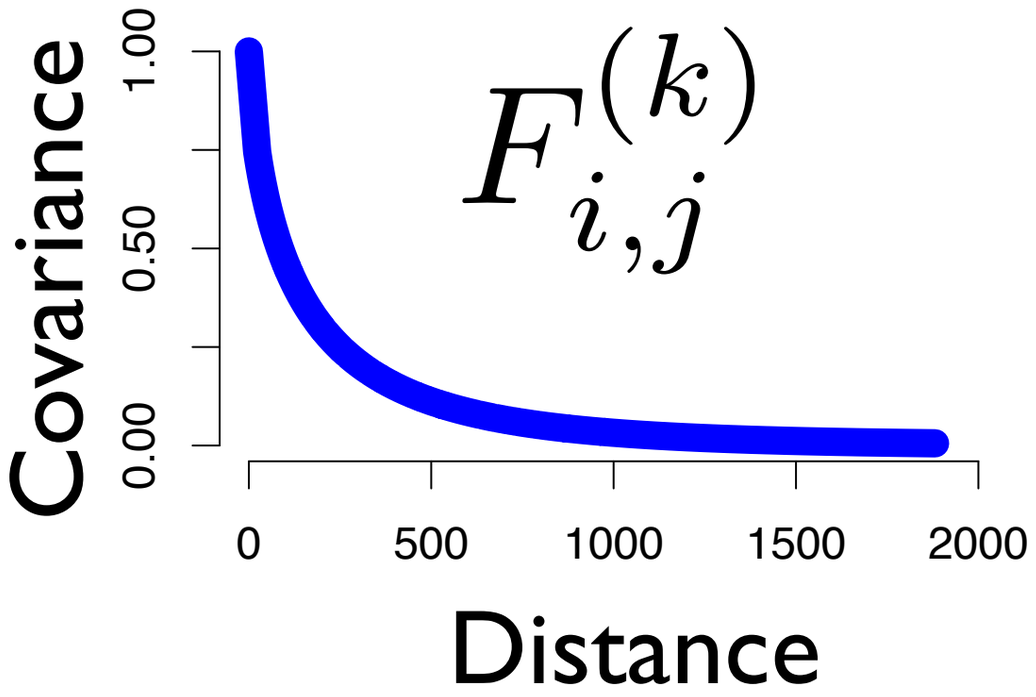




geoStructure mechanics

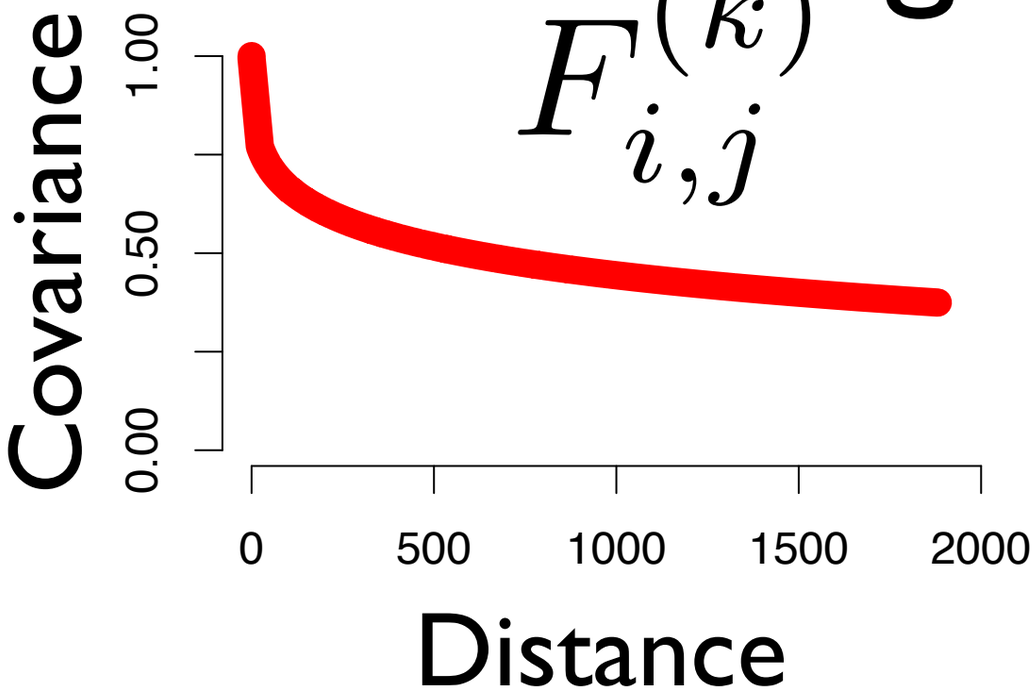


$$F_{i,j}^{(k)}$$



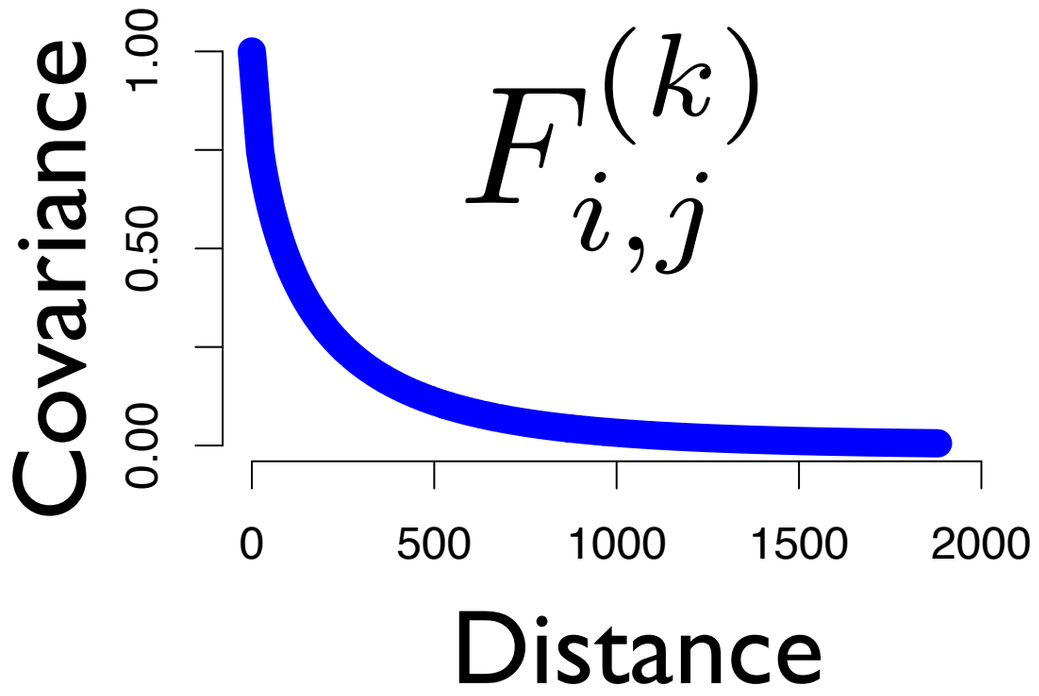
$$F_{i,j}^{(k)}$$

geoStructure mechanics



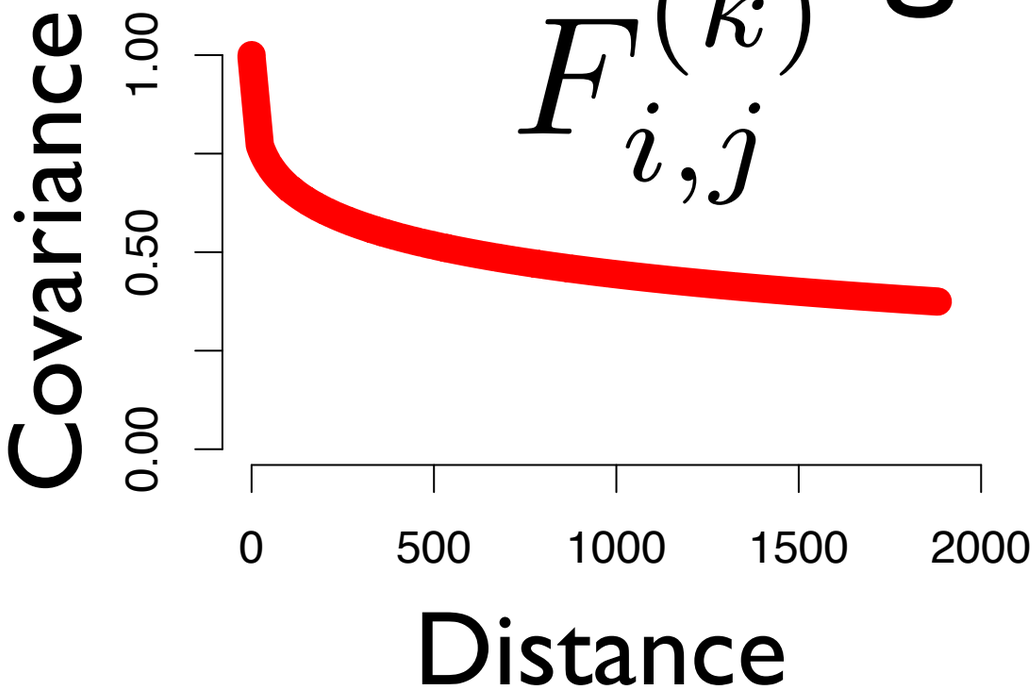
$$F_{i,j}^{(k)}$$

admixture proportion



$$F_{i,j}^{(k)}$$

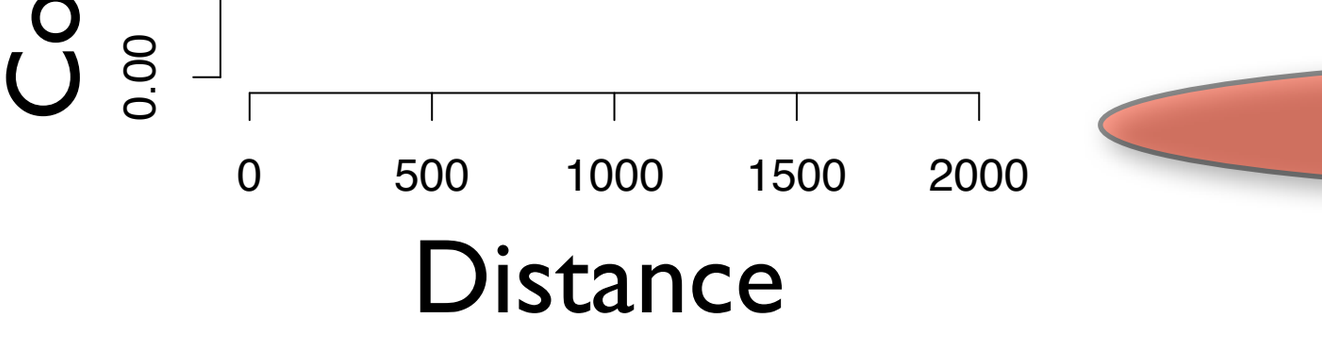
geoStructure mechanics



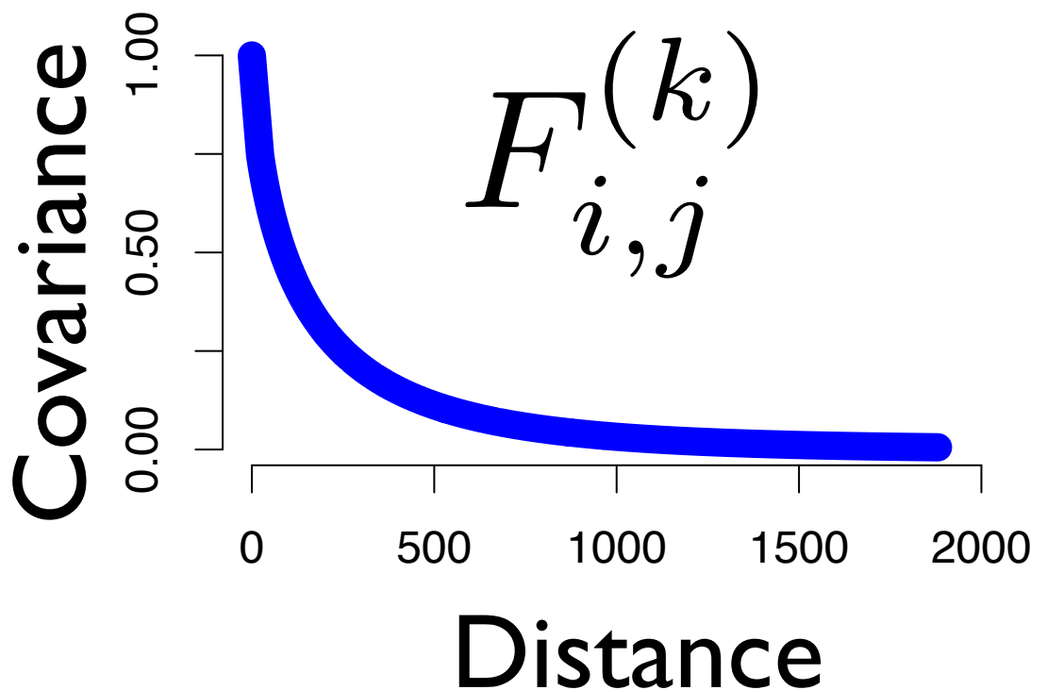
$F_{i,j}^{(k)}$

admixture proportion

$w_i^{(k)}$



$\}w_i^{(k)}$



$F_{i,j}^{(k)}$

geoStructure mechanics

parametric covariance

$$\overbrace{\Omega_{i,j} = \sum_{K} \left(w_i^{(k)} w_j^{(k)} \left(F_{i,j}^{(k)} + \mu^{(k)} \right) \right)}$$

geoStructure mechanics

parametric covariance

cluster-specific spatial covariance

$$\Omega_{i,j} = \sum_{K} \left(w_i^{(k)} w_j^{(k)} \left(F_{i,j}^{(k)} + \mu^{(k)} \right) \right)$$

geoStructure mechanics

parametric covariance

cluster-specific spatial covariance

$$\Omega_{i,j} = \sum_{K} \left(w_i^{(k)} w_j^{(k)} \left(F_{i,j}^{(k)} + \mu^{(k)} \right) \right)$$

cluster-specific shared drift

geoStructure mechanics

parametric covariance

cluster-specific spatial covariance

$$\Omega_{i,j} = \sum_{K} \left(w_i^{(k)} w_j^{(k)} \left(F_{i,j}^{(k)} + \mu^{(k)} \right) \right)$$

admixture proportions

cluster-specific shared drift

geoStructure mechanics

parametric covariance

cluster-specific spatial covariance

$$\Omega_{i,j} = \sum_{K} \left(w_i^{(k)} w_j^{(k)} \left(F_{i,j}^{(k)} + \mu^{(k)} \right) \right)$$

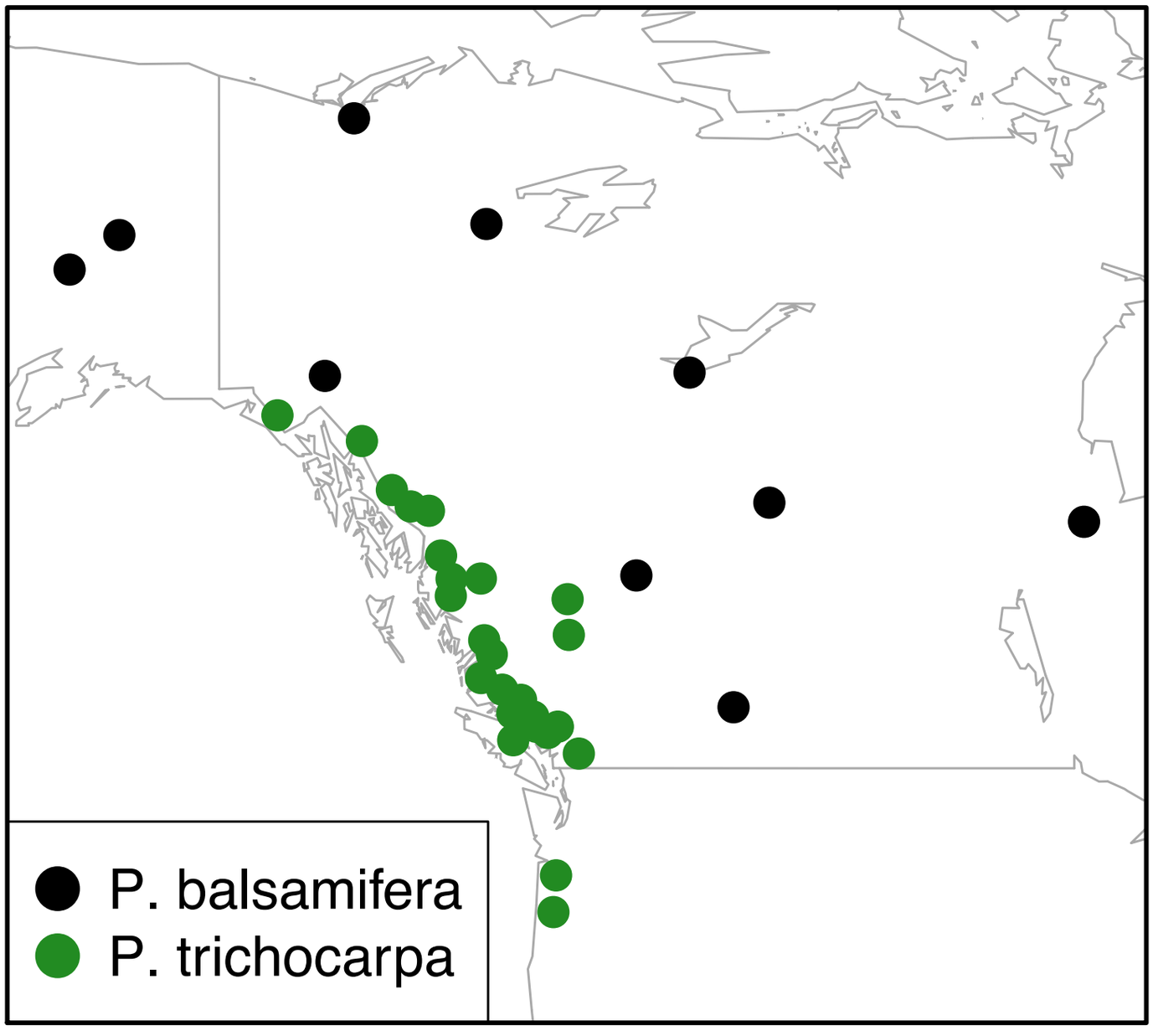
specific
d drift

$$\hat{\Omega} \sim \mathcal{W}(\Omega, L)$$

Empirical Application:

Populus trichocarpa

Populus balsamifera



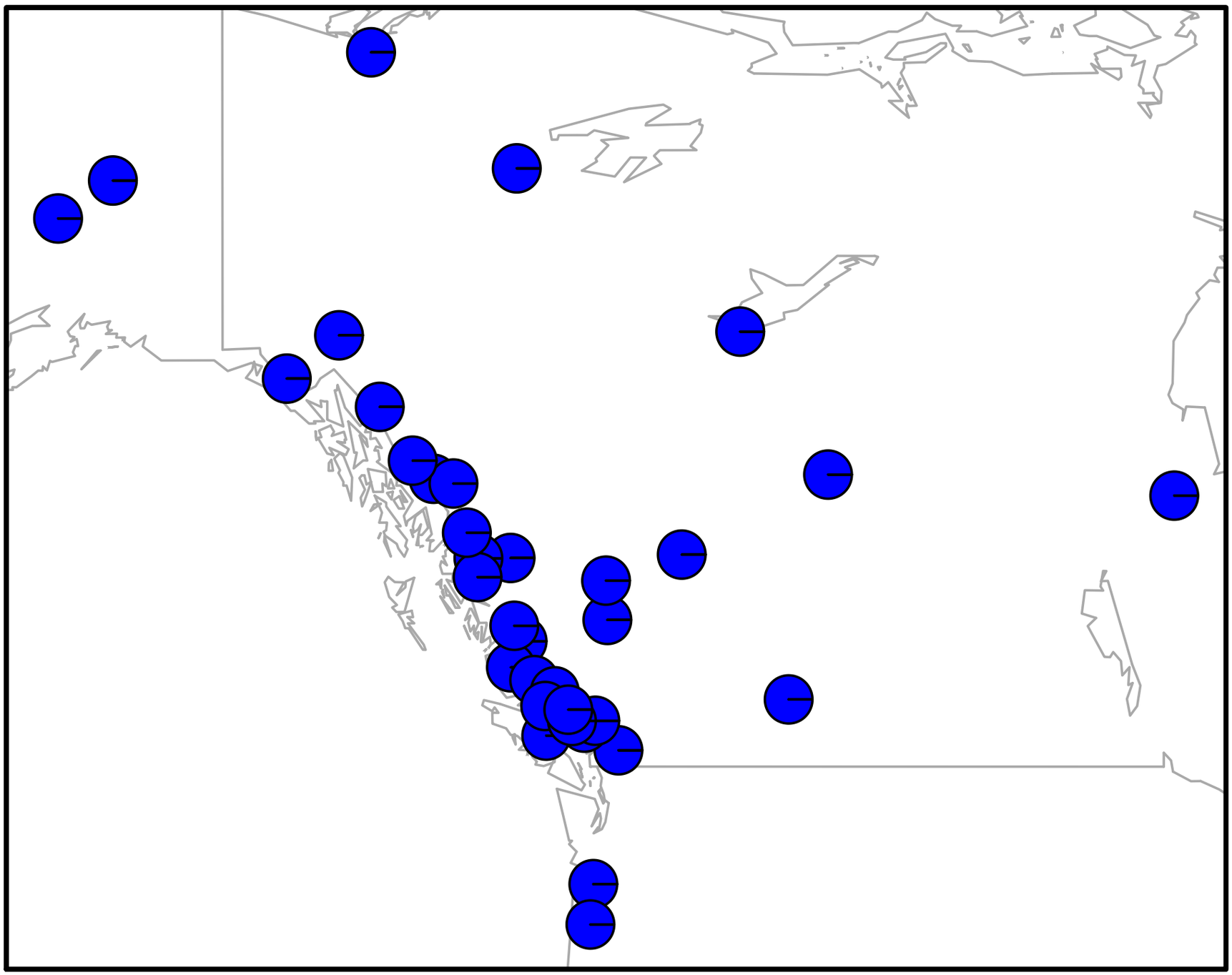
Geraldes *et al.* 2013

Geraldes *et al.* 2014

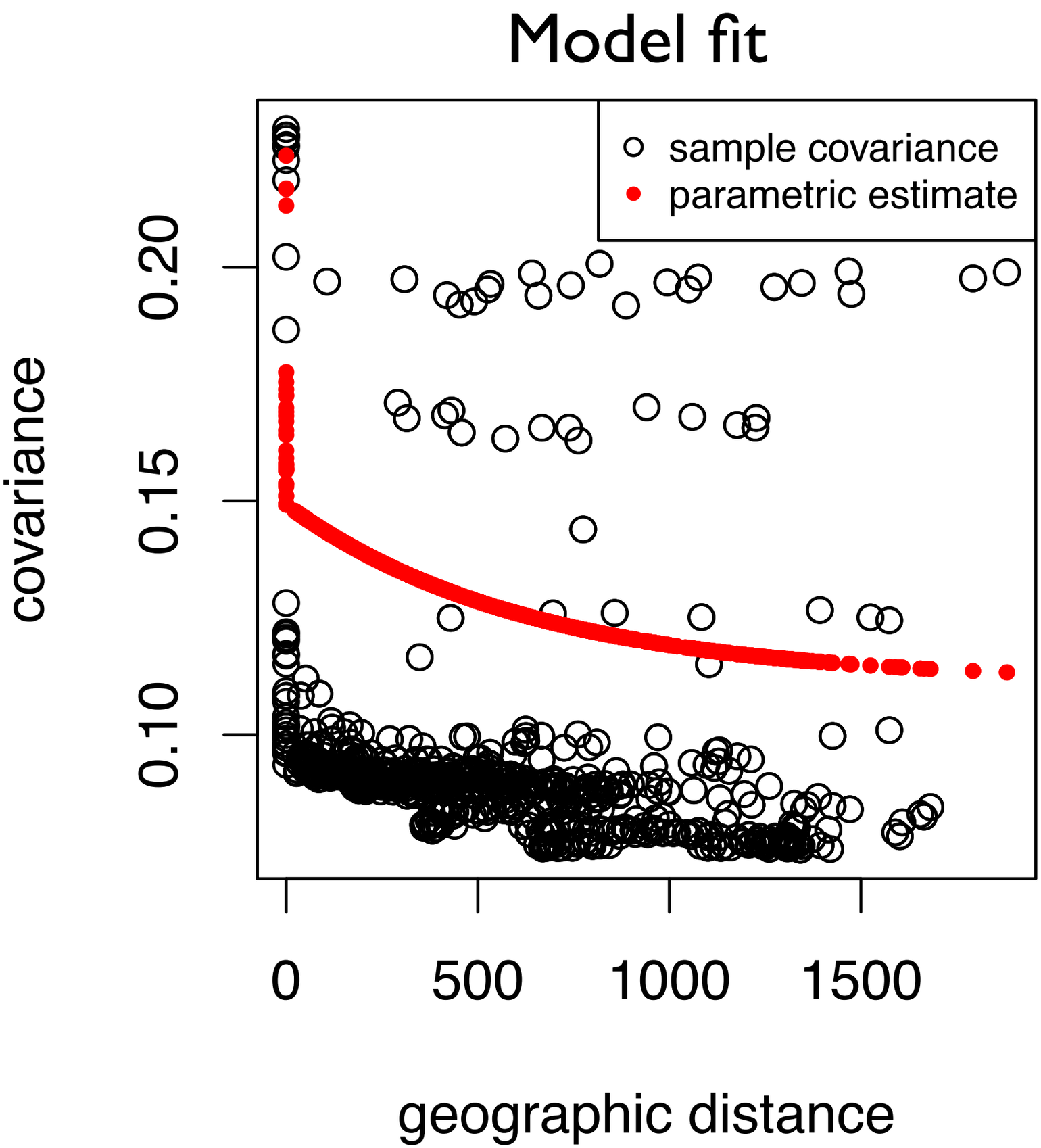
$$K = I$$

K = 1

Admixture proportion map

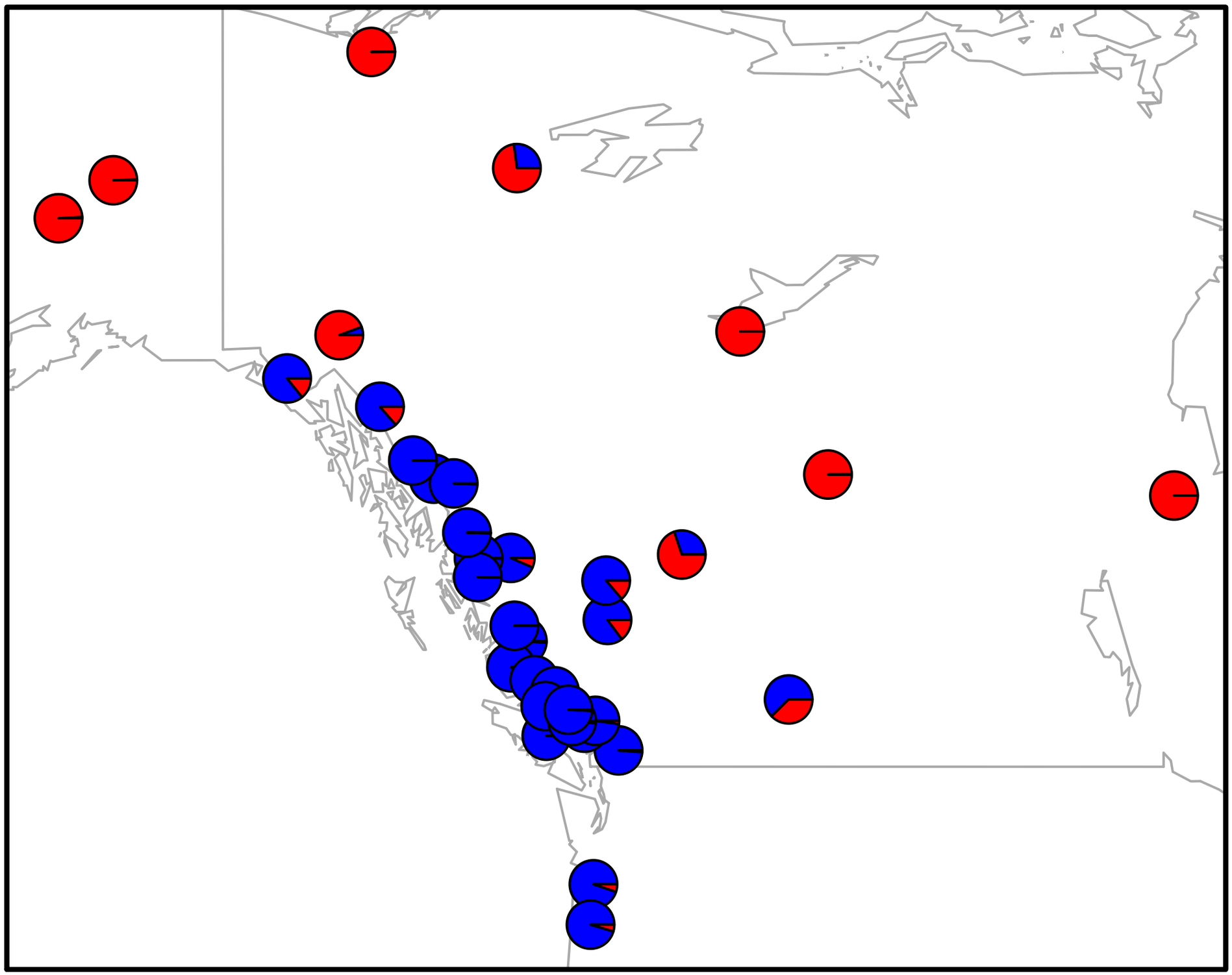


K = 1


Model fit
sample covariance
parametric estimate
covariance
0.20
0.15
0.10
0
500
1000
1500
geographic distance

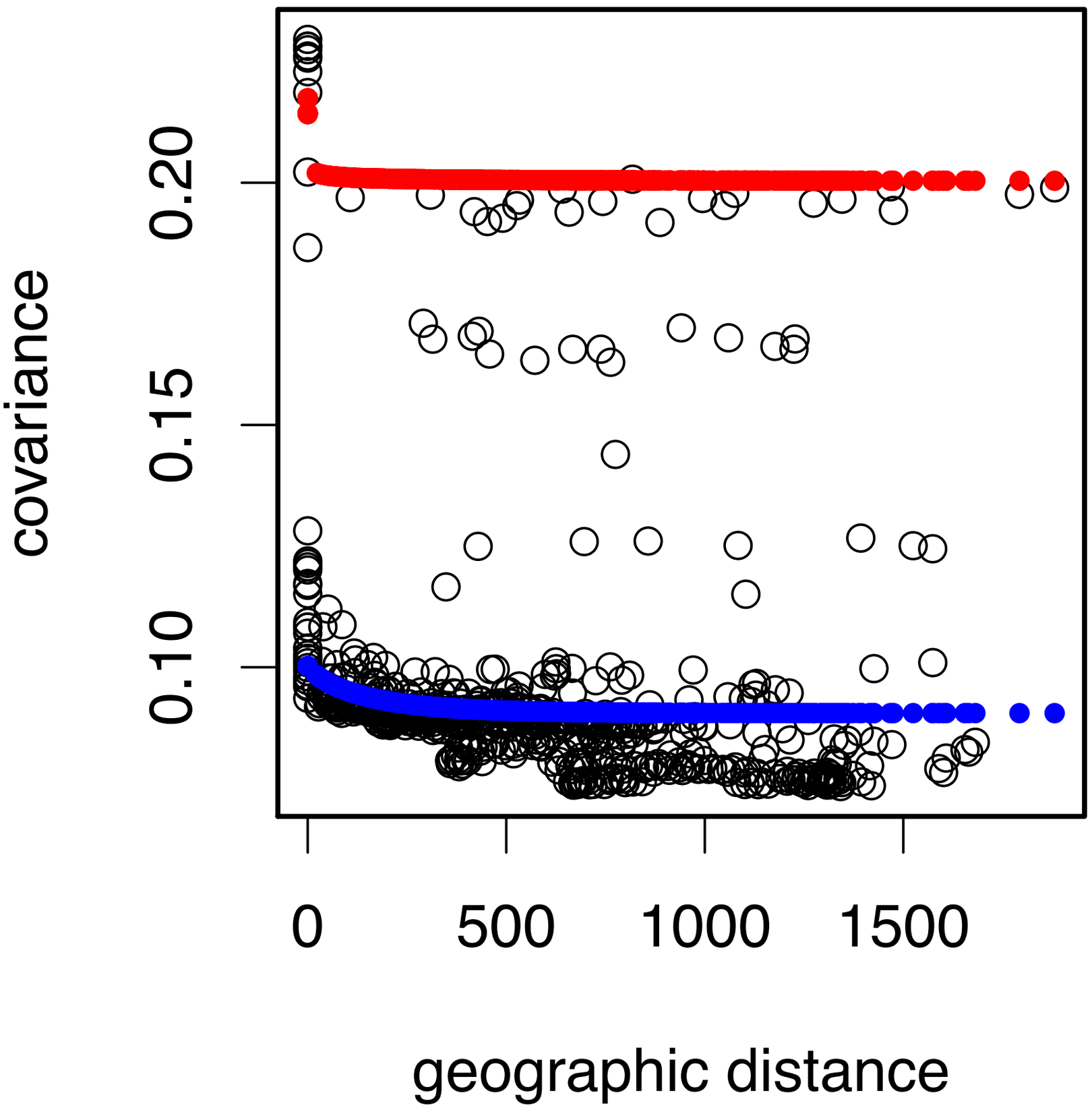
K = 2

Admixture proportion map



Spatial covariance functions

K = 2



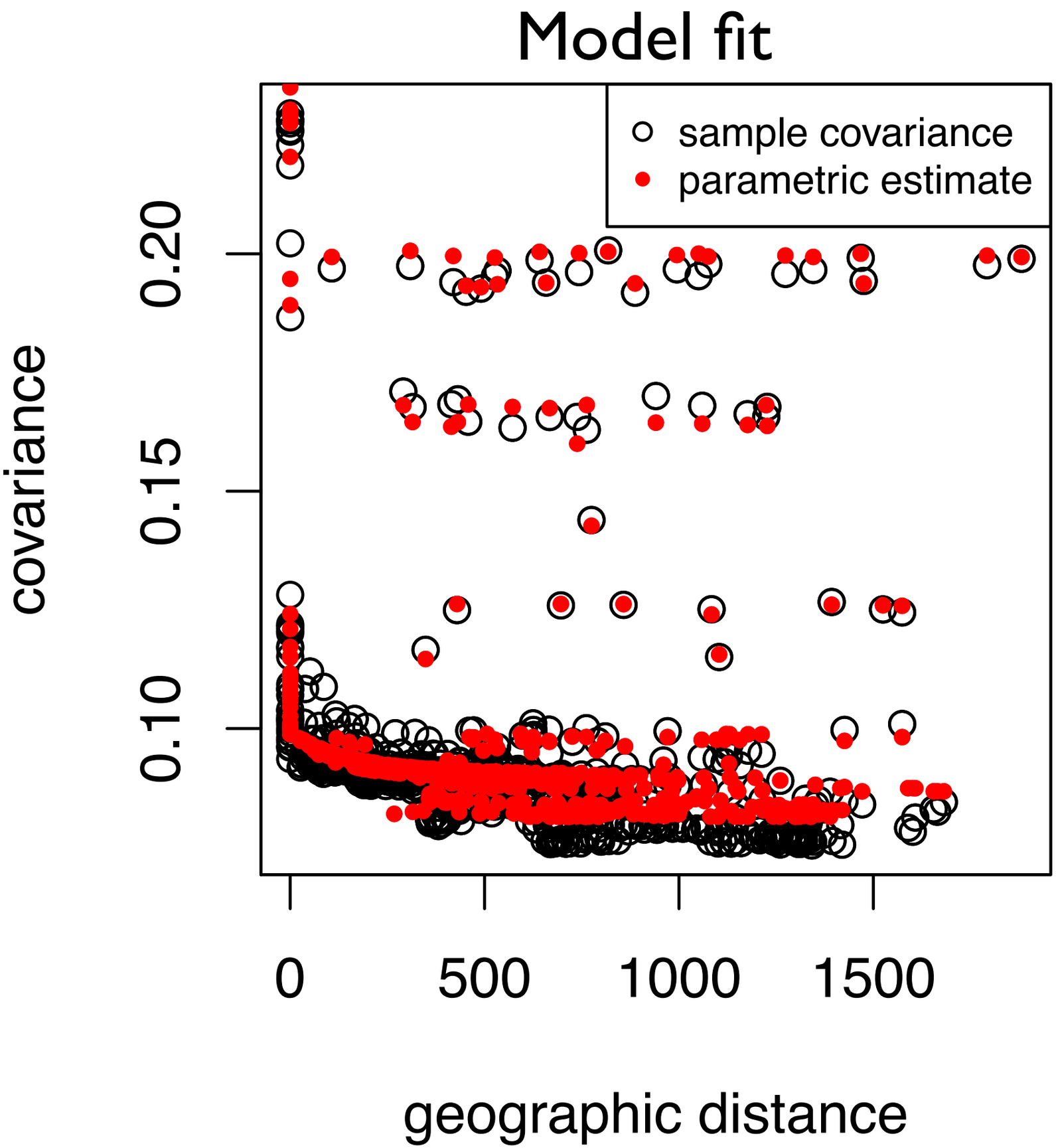
$$F_{i,j}^{(k=1)}$$

$$F_{i,j}^{(k=2)}$$

Model fit

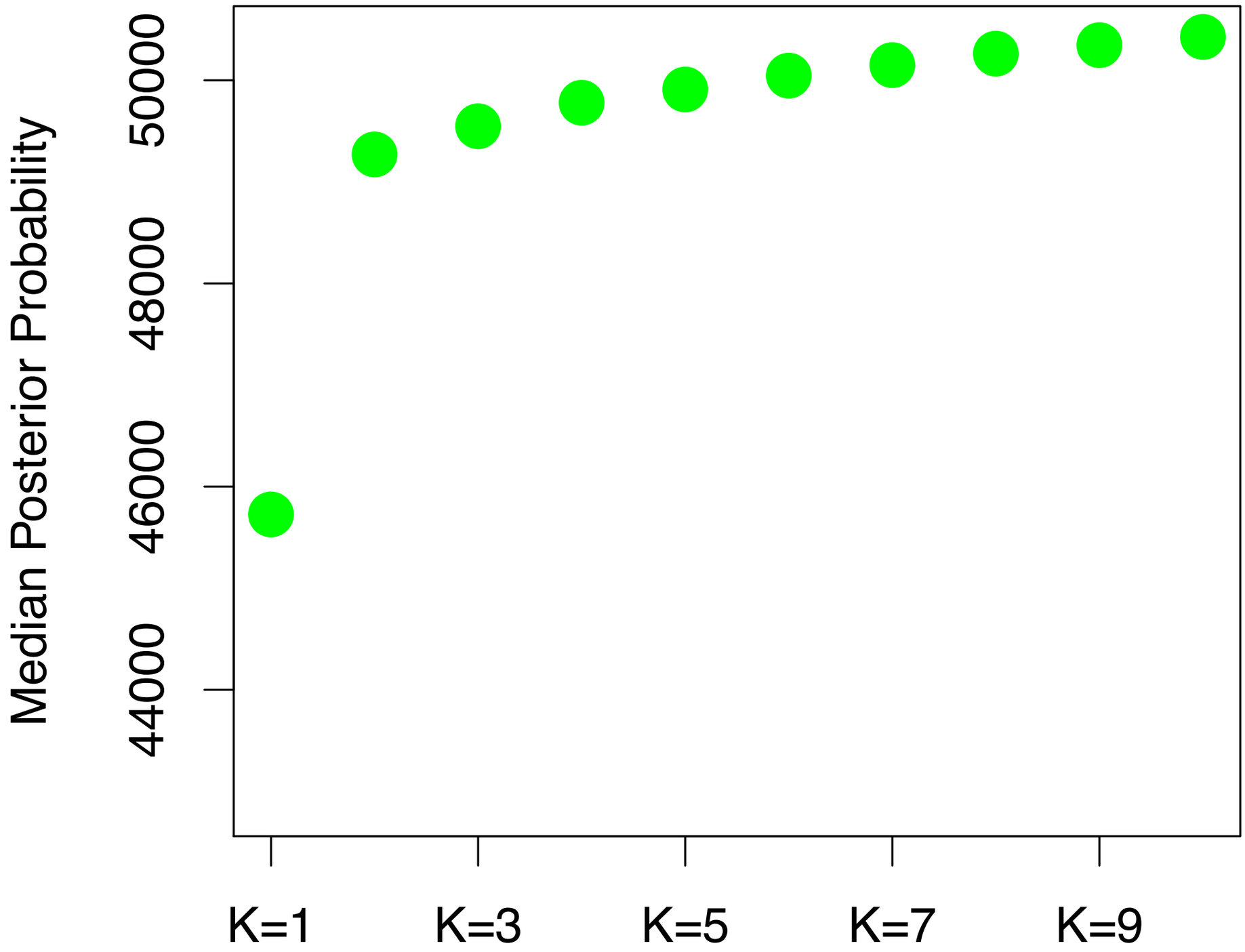
K = 2

K = 2

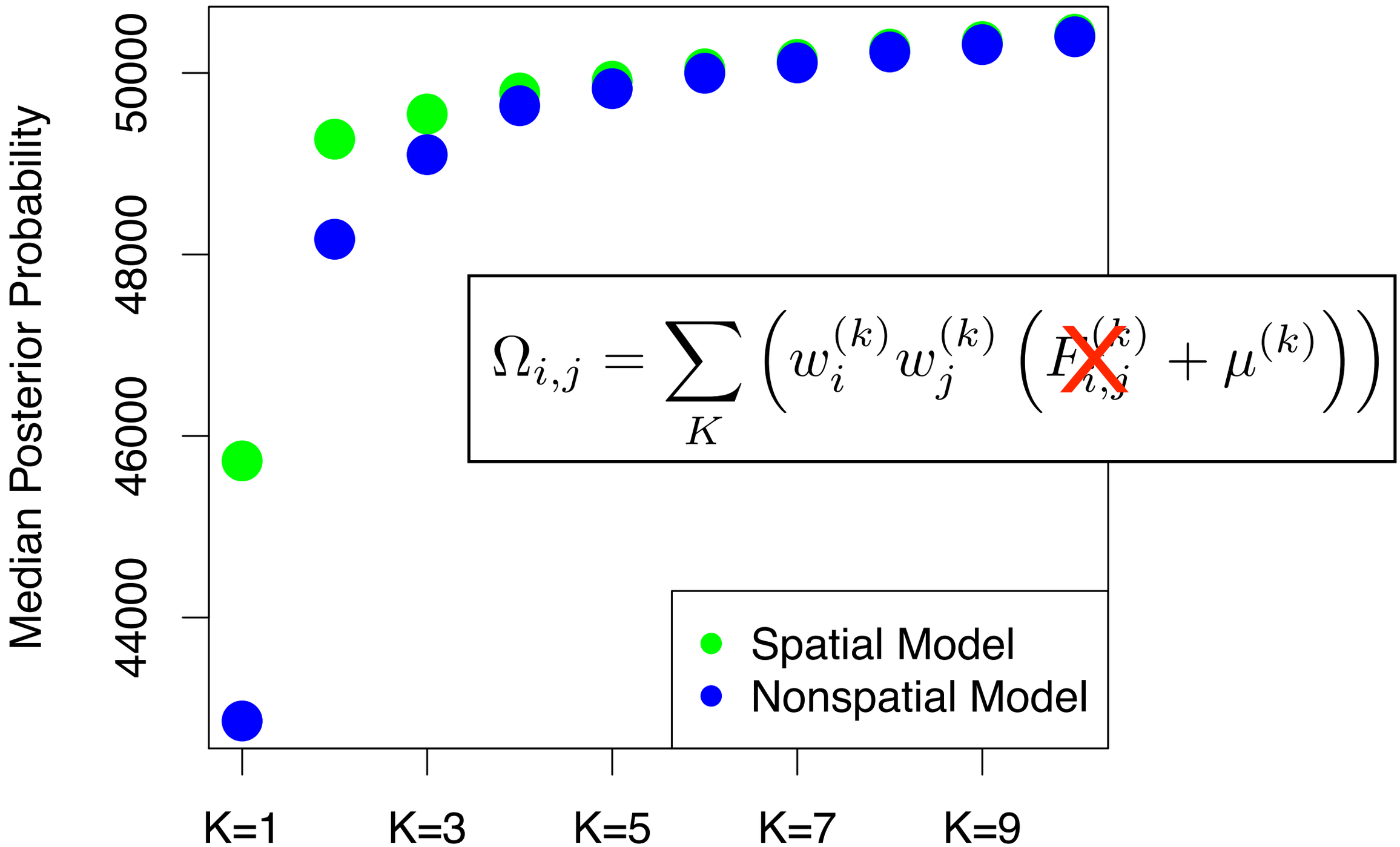

Model fit
sample covariance
parametric estimate
covariance
0.20
0.15
0.10
0
500
1000
1500
geographic distance

Model comparisons

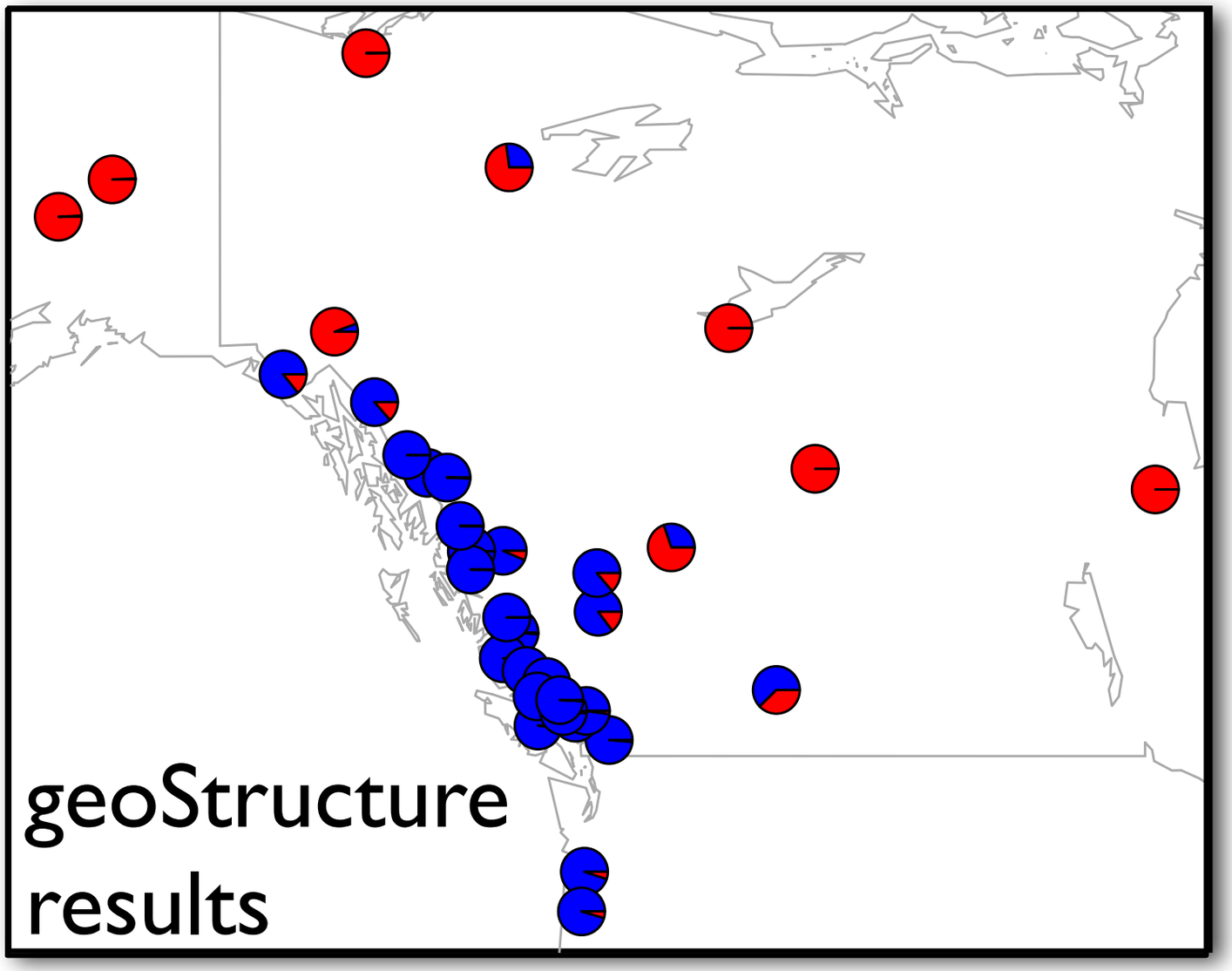
Model comparisons

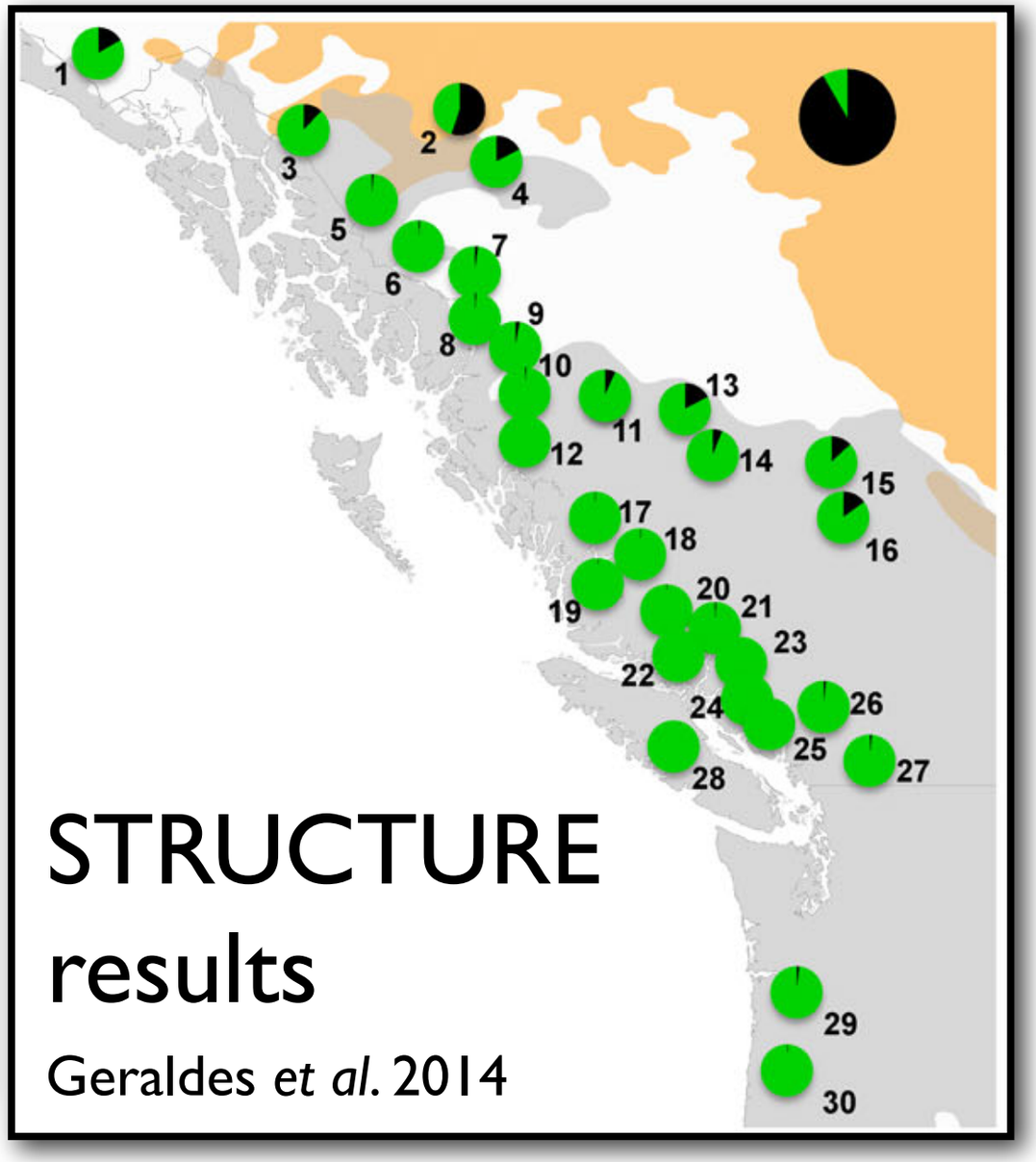


Model comparisons



Model comparisons

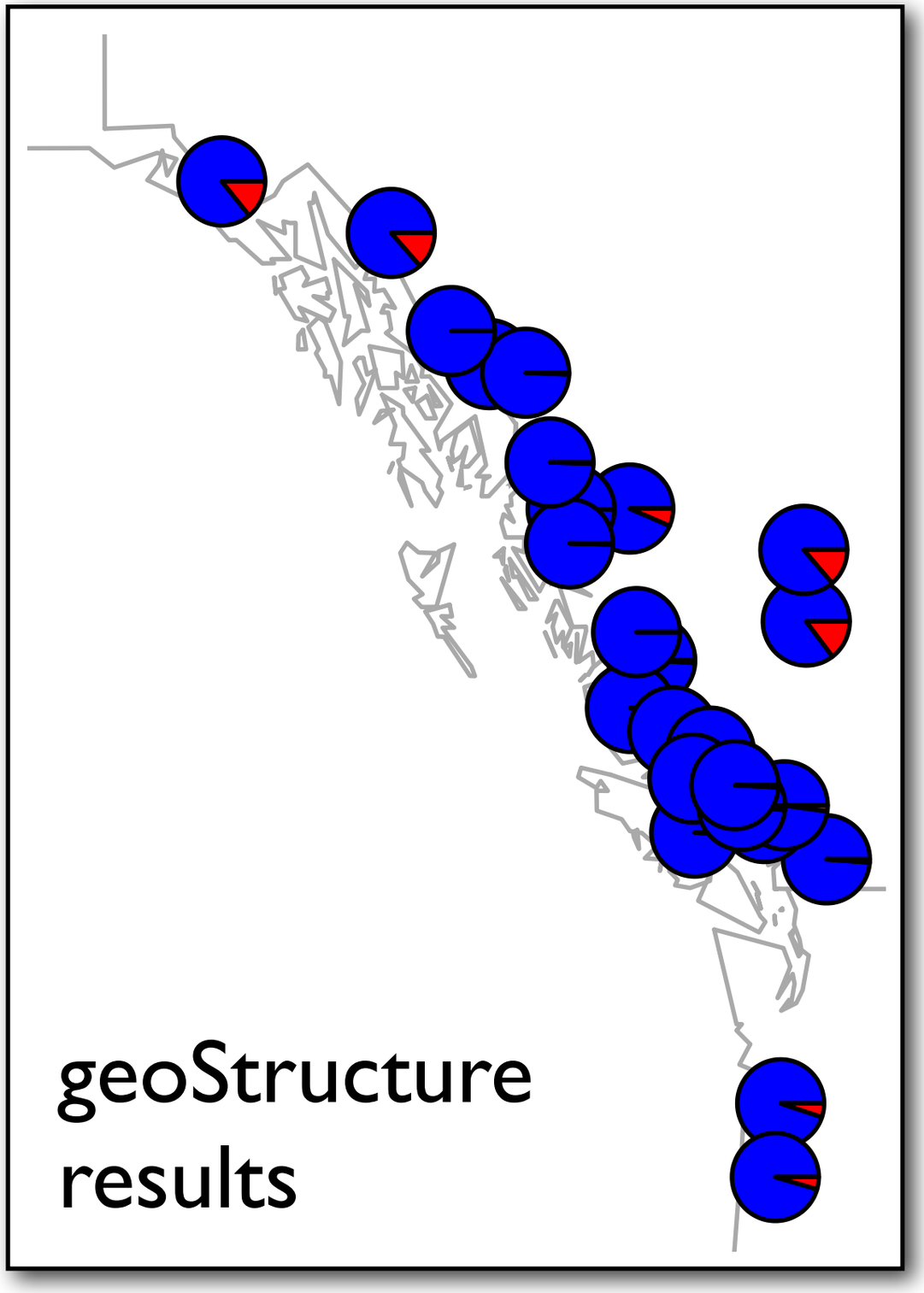


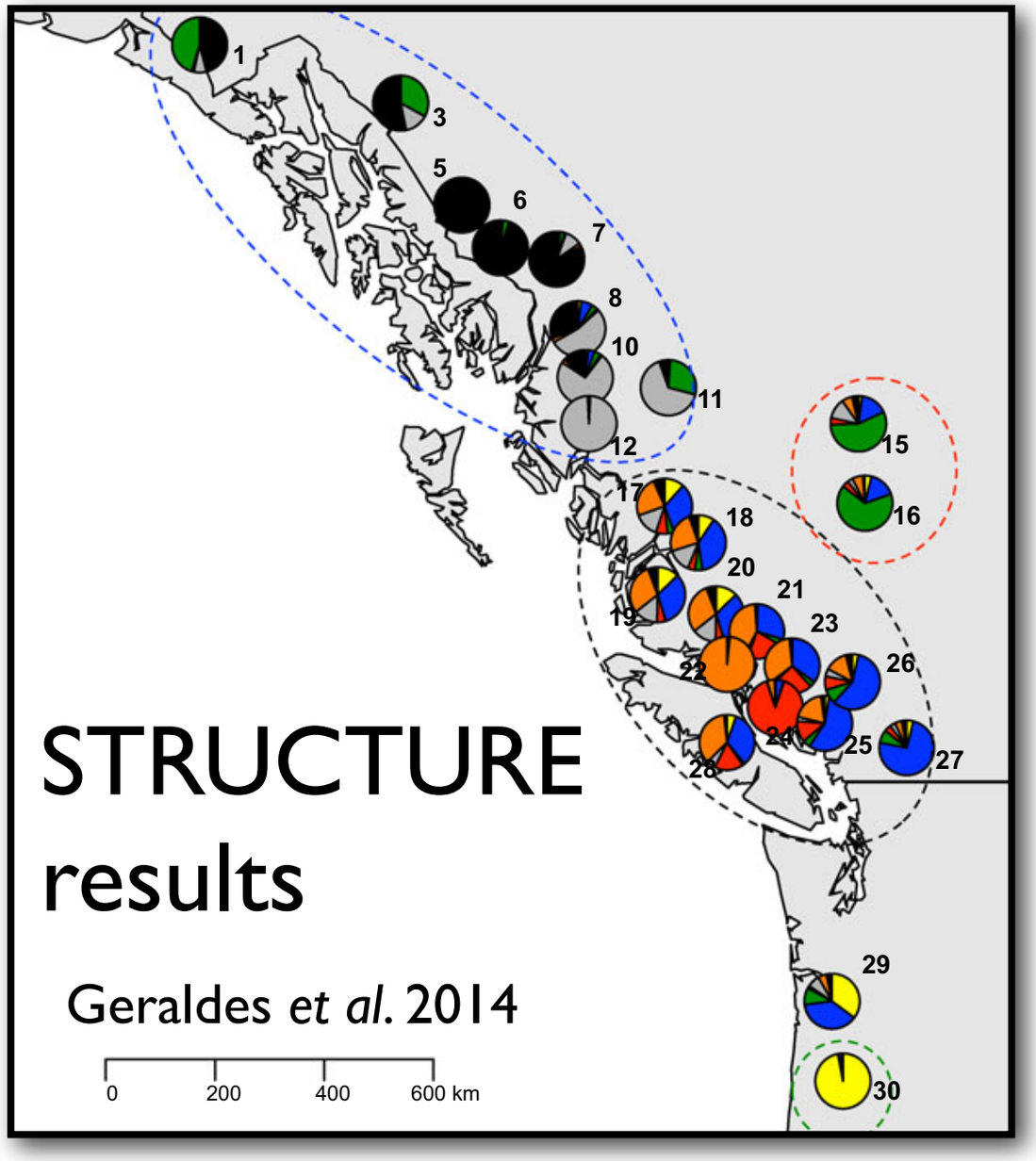


Populus trichocarpa

Populus balsamifera

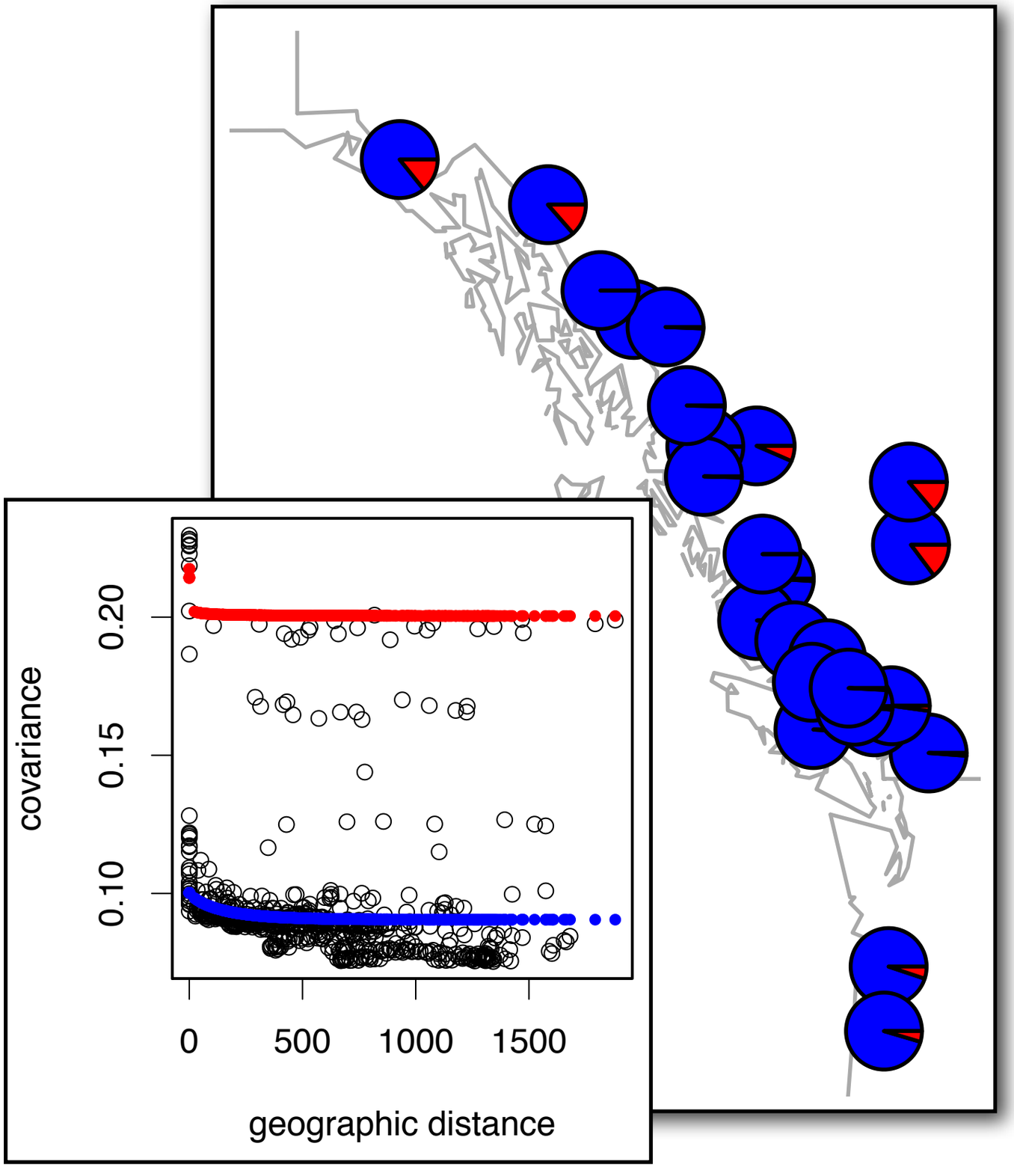
Model comparisons

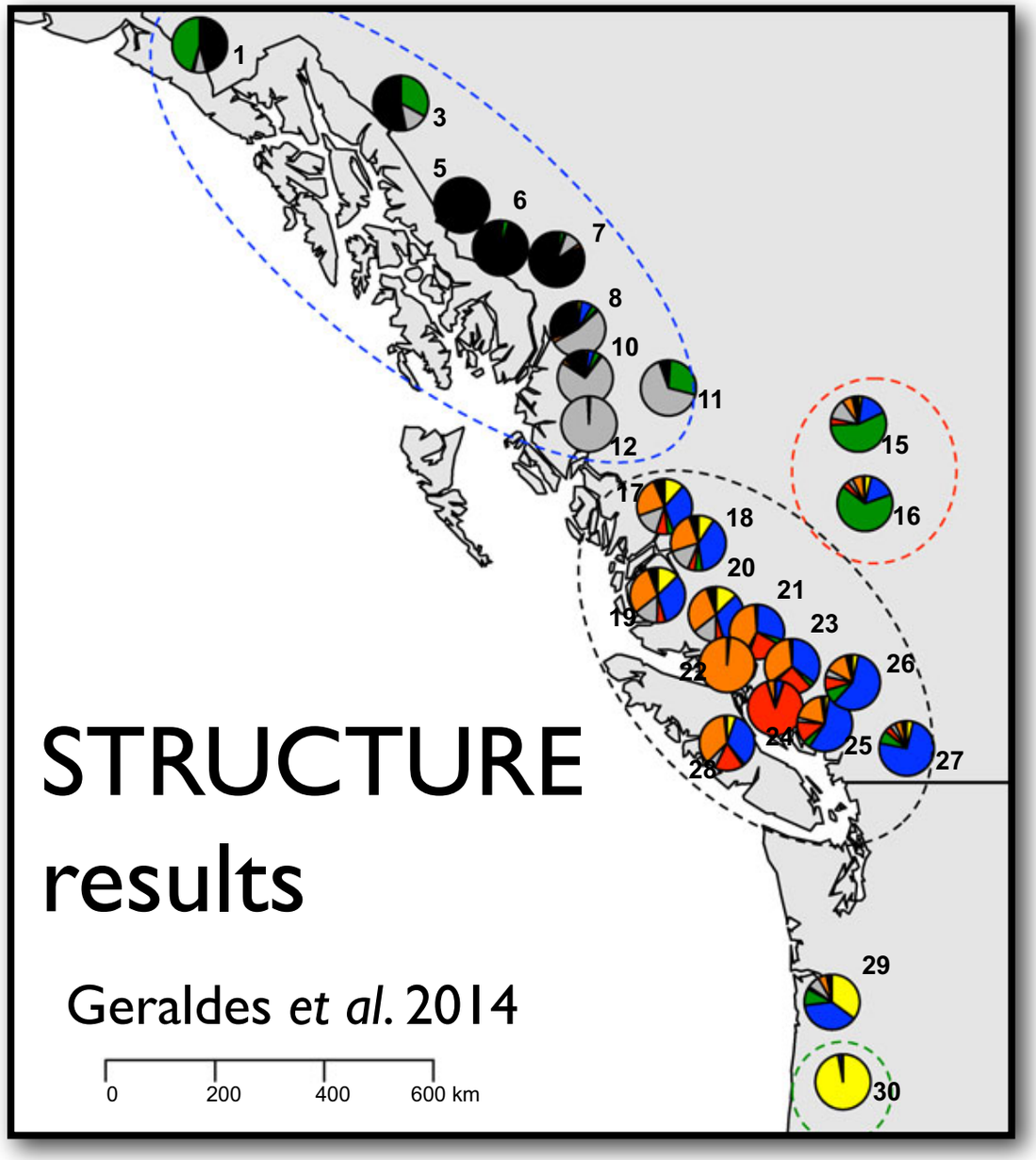




Populus trichocarpa

Model comparisons



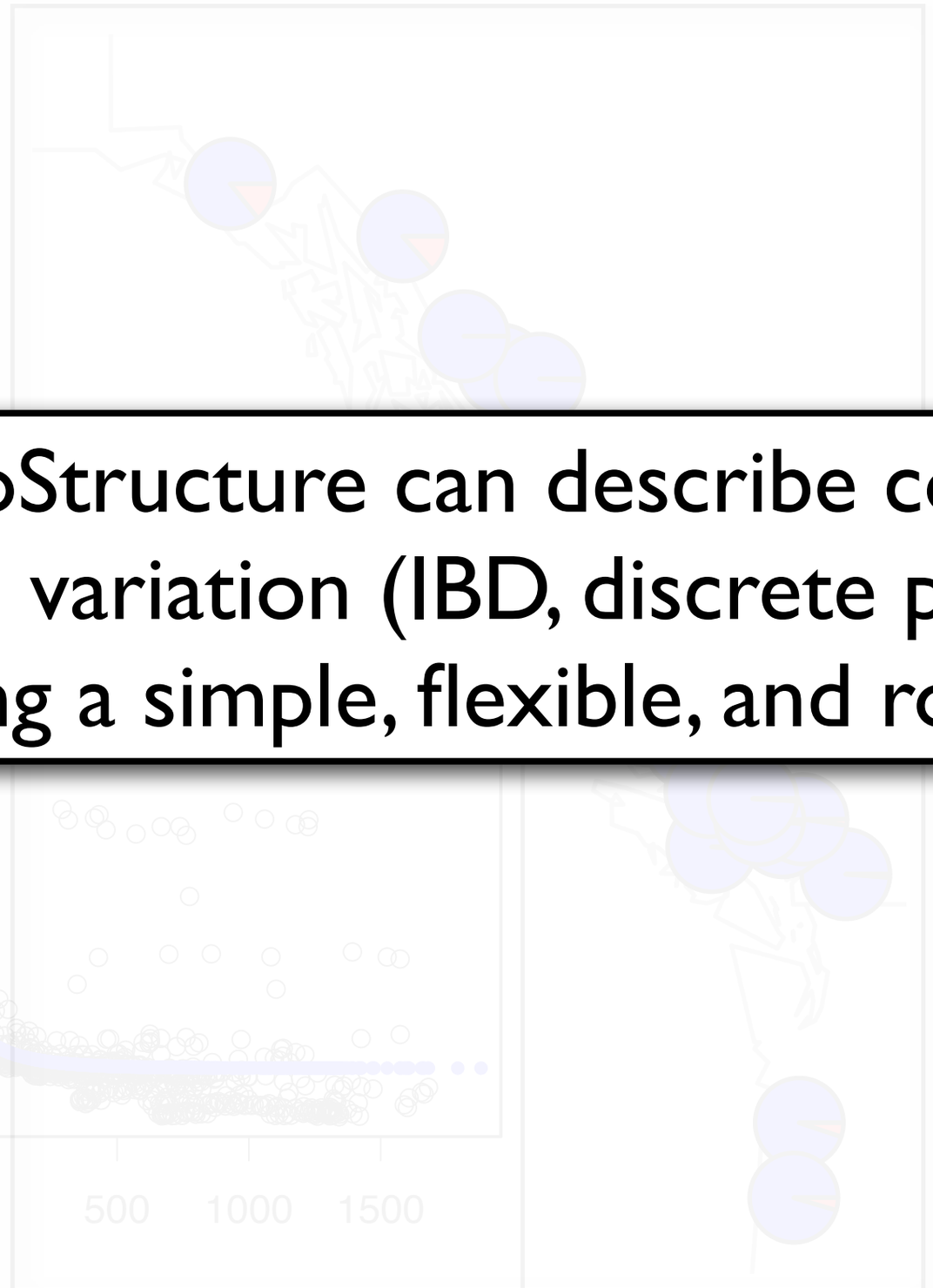


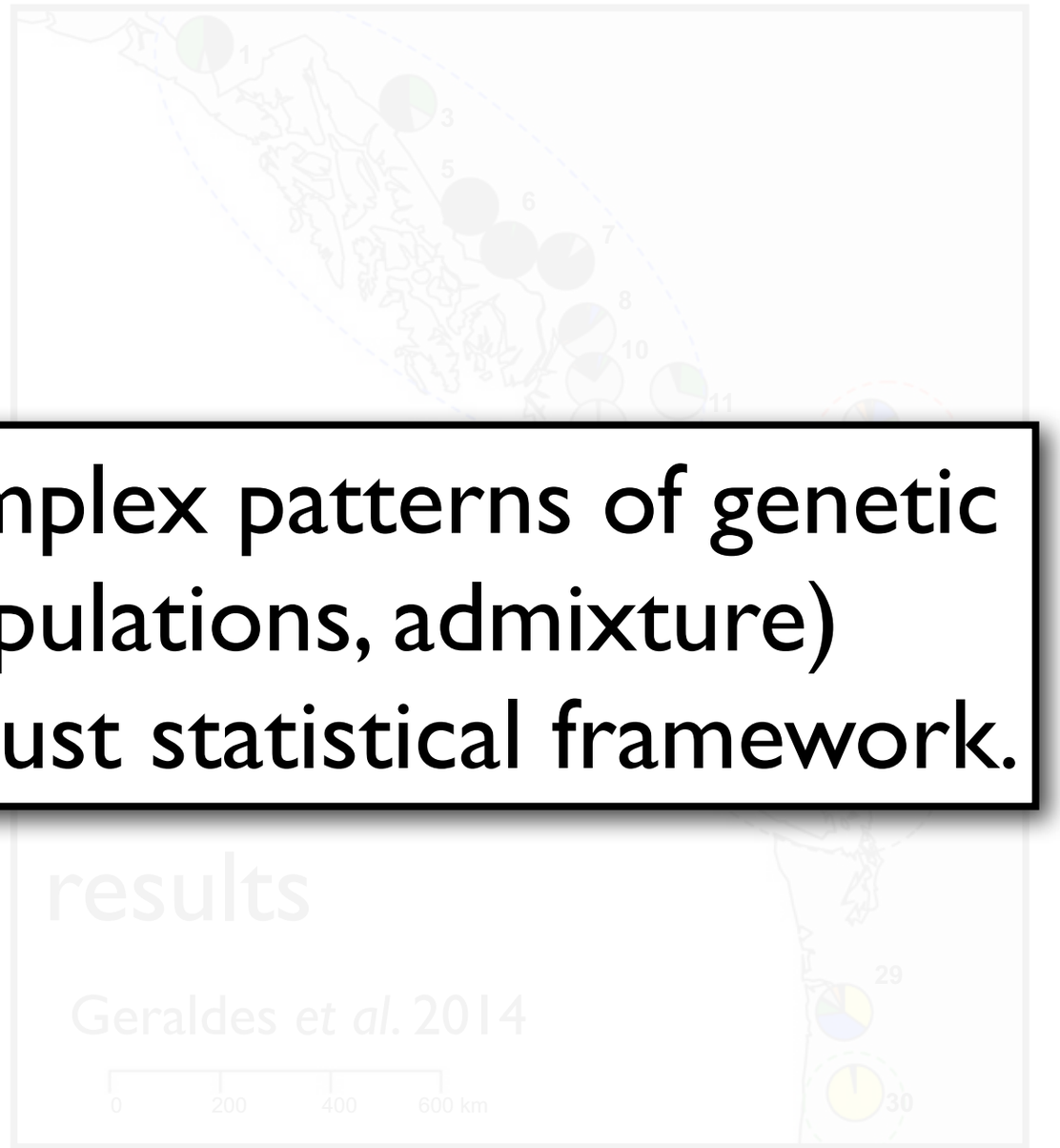
STRUCTURE results

Geraldes *et al.* 2014

Populus trichocarpa

Model comparisons





geoStructure can describe complex patterns of genetic variation (IBD, discrete populations, admixture) using a simple, flexible, and robust statistical framework.

results

Geraldes *et al.* 2014

Populus trichocarpa

Extensions

Extensions

Cluster relationships:

-treelike

-spatial

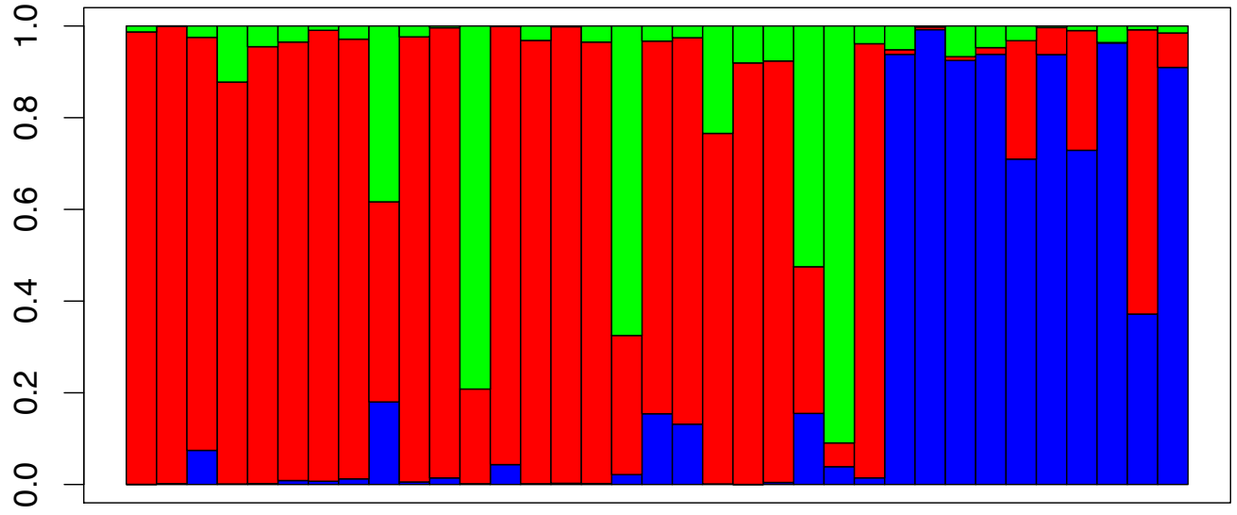
Spatiotemporal
covariance functions

Extensions

Cluster relationships:

-treelike

-spatial



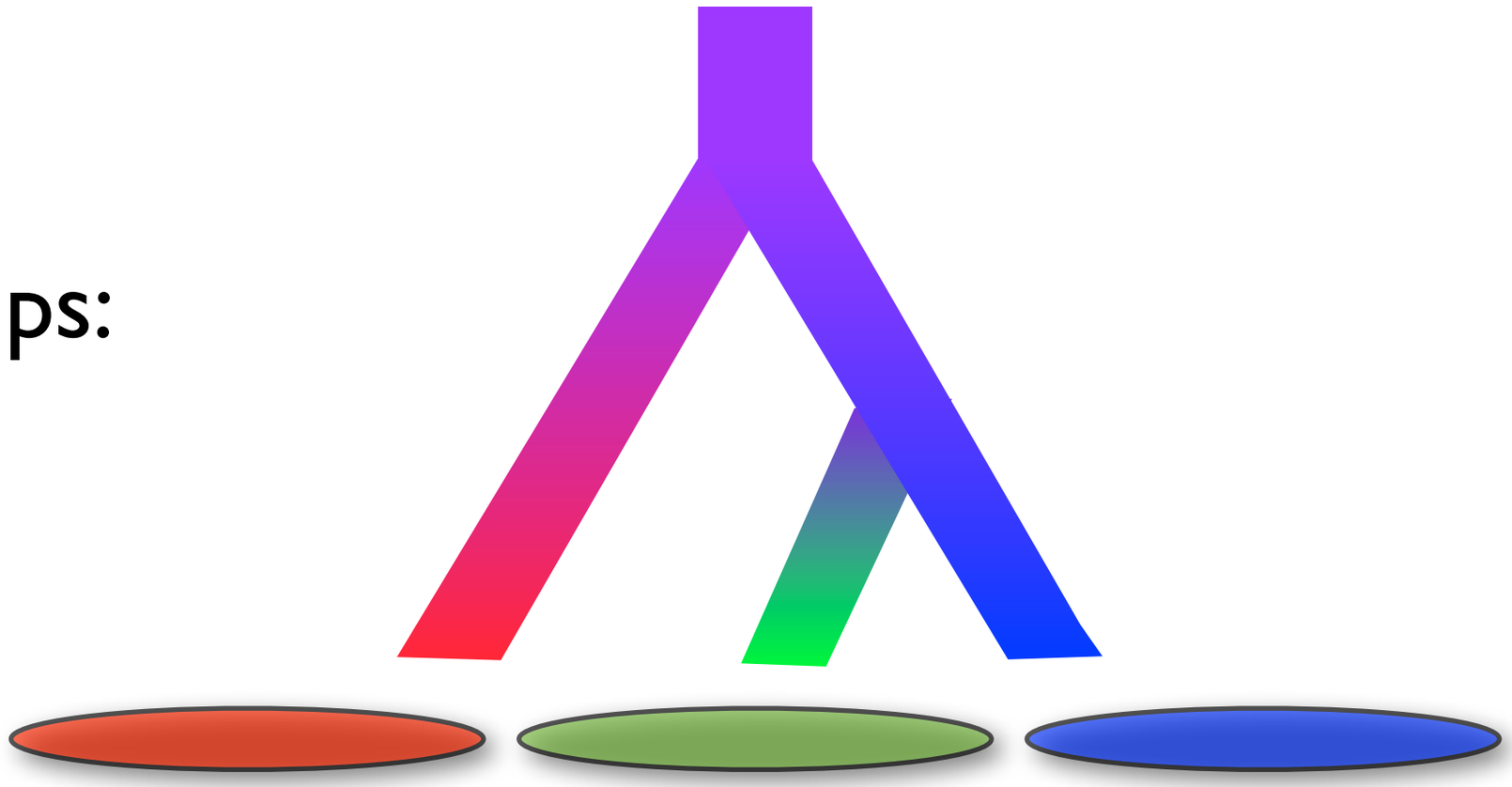
Spatiotemporal covariance functions

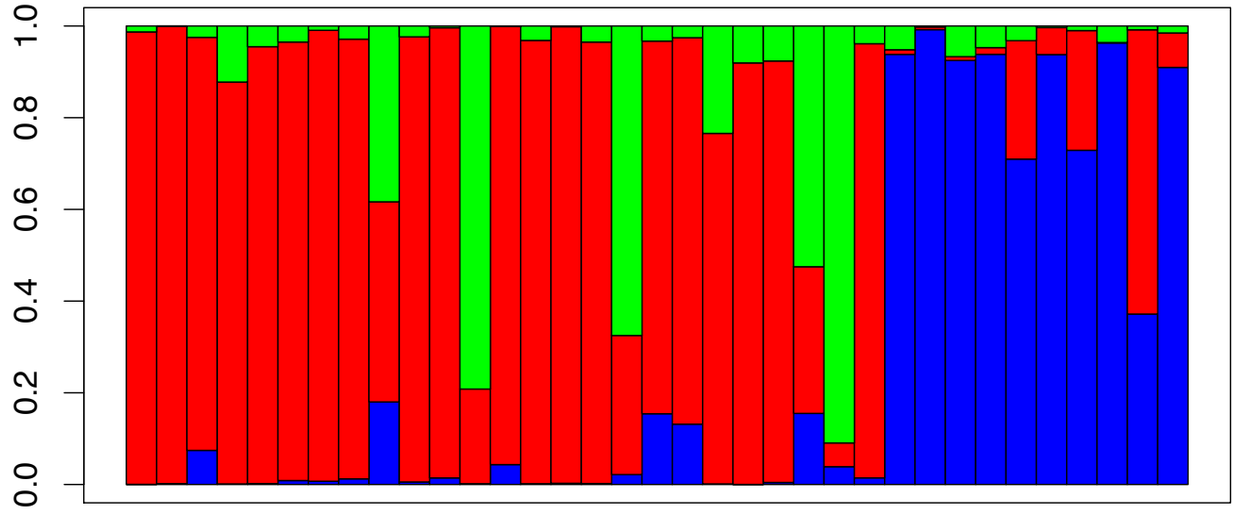
Extensions

Cluster relationships:

-treelike

-spatial





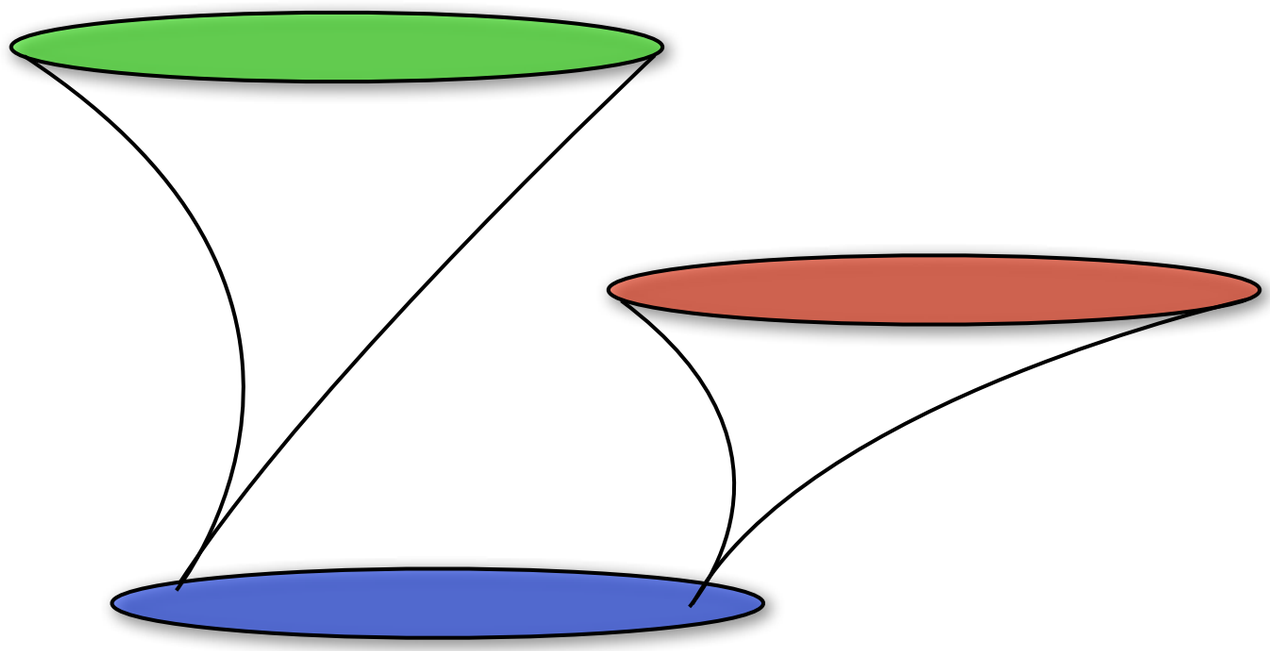
Spatiotemporal covariance functions

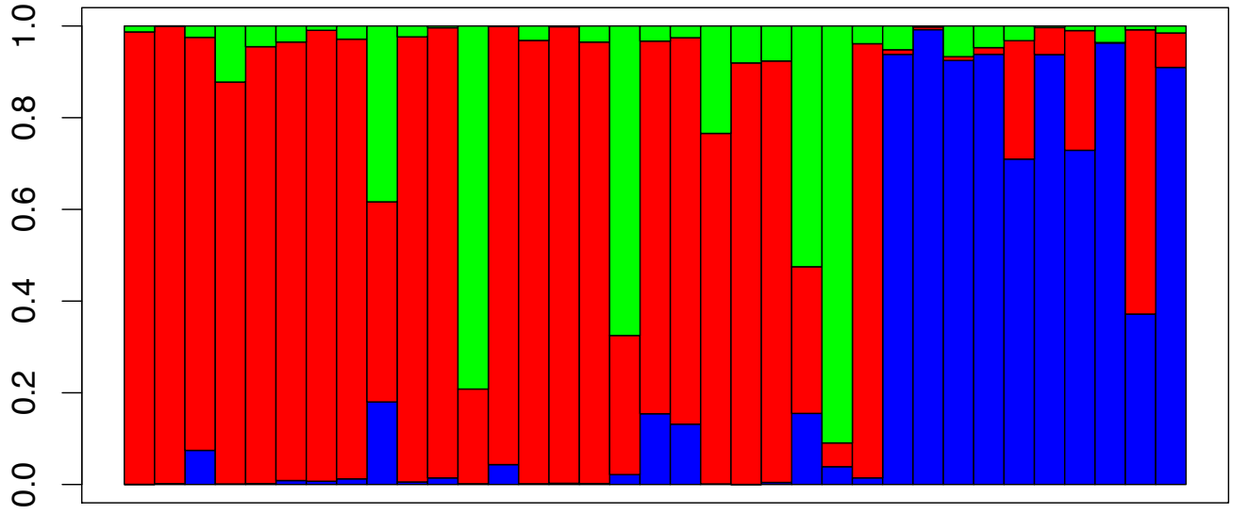
Extensions

Cluster relationships:

-treelike

-spatial





Spatiotemporal covariance functions

Extensions

Cluster relationships:

-treelike

-spatial

Spatiotemporal
covariance functions

Extensions

Cluster relationships:

-treelike

-spatial

$$F_{i,j}^{(k)} = \mathrm{f}(D_{i,j})$$

Spatiotemporal
covariance functions

Extensions

Cluster relationships:

-treelike

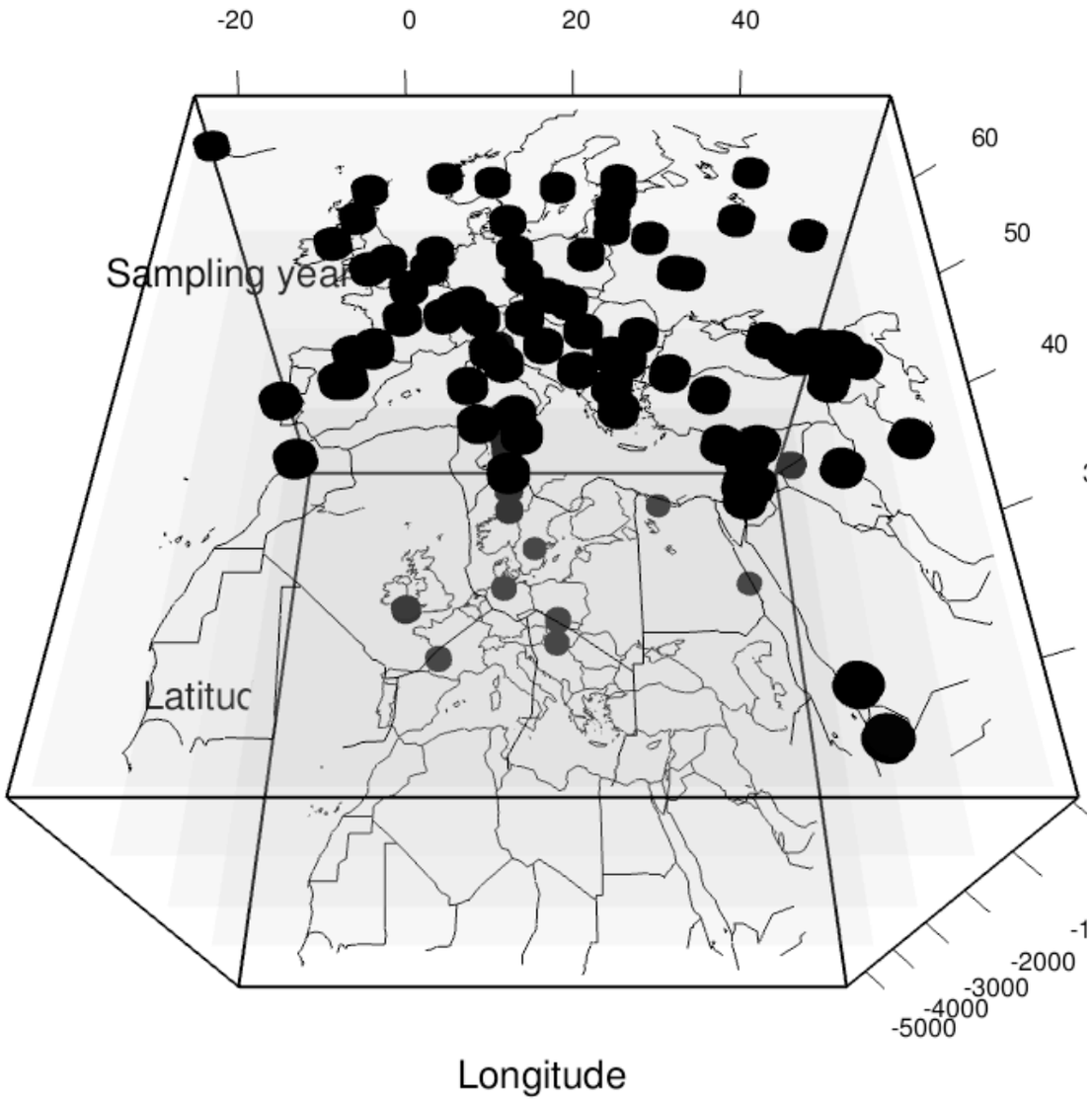
-spatial

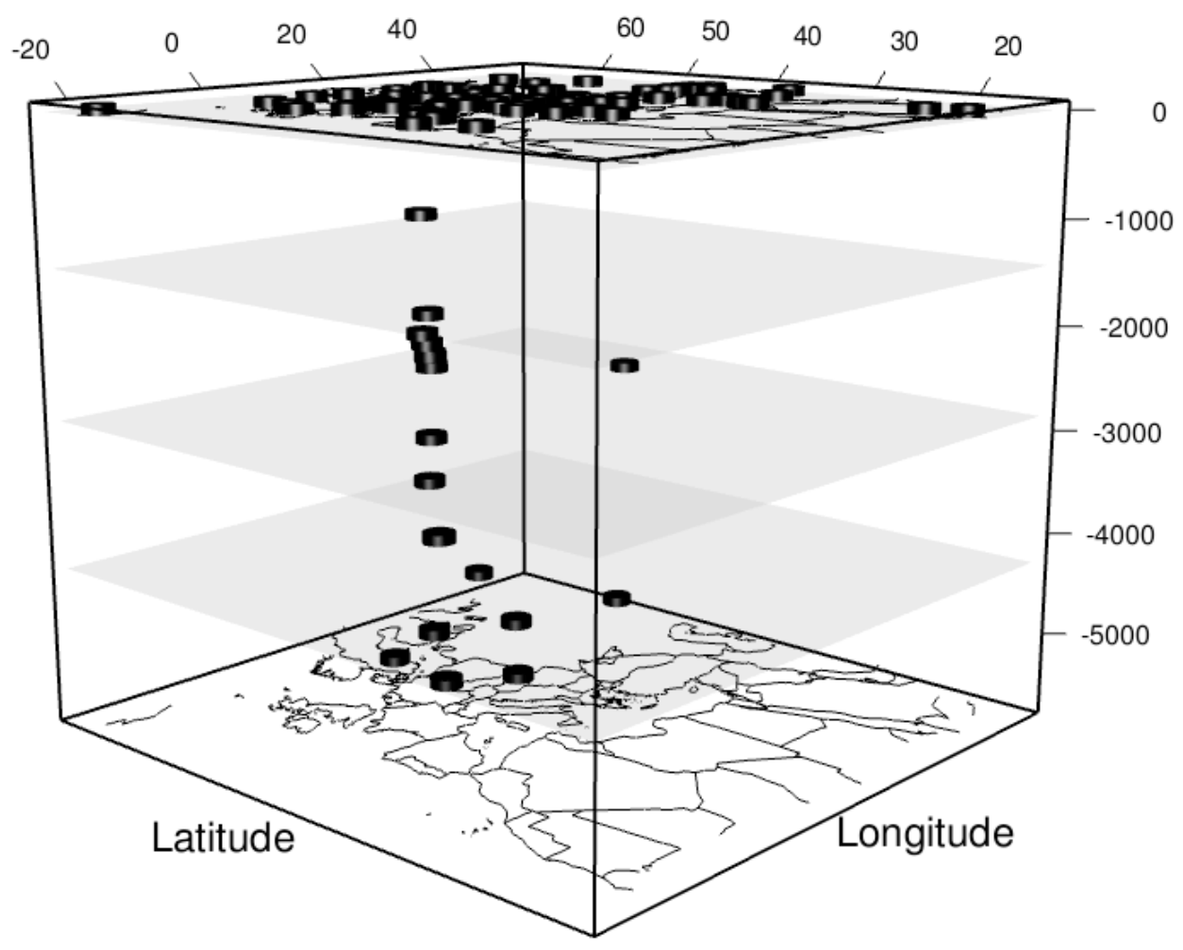
$$F_{i,j}^{(k)} = \mathrm{f}(D_{i,j})$$

$$F_{i,j}^{(k)} = \mathrm{f}(D_{i,j}, \tau_{i,j})$$

Spatiotemporal
covariance functions

Extensions

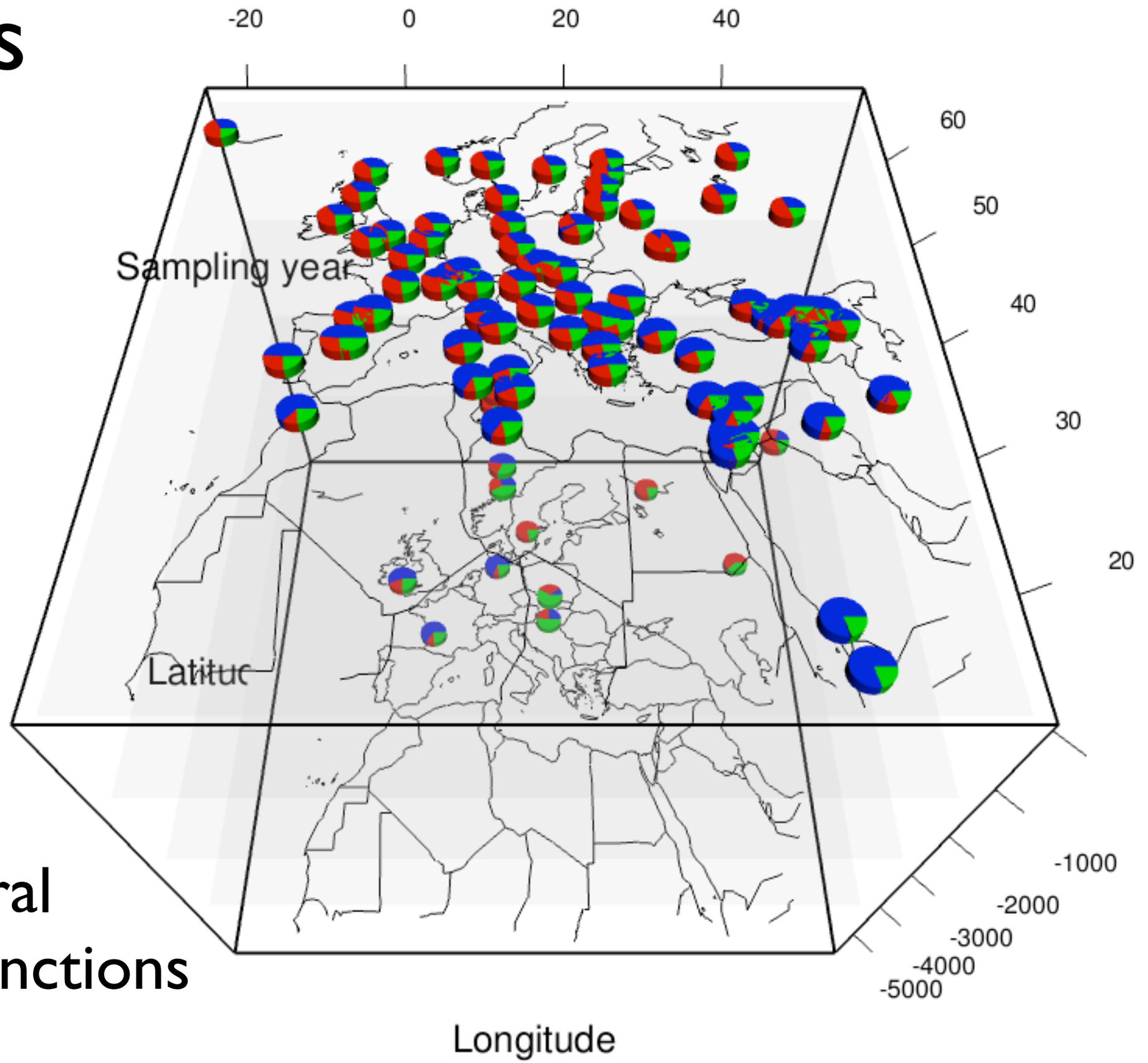




Spatiotemporal covariance functions

Haak *et al.* 2015

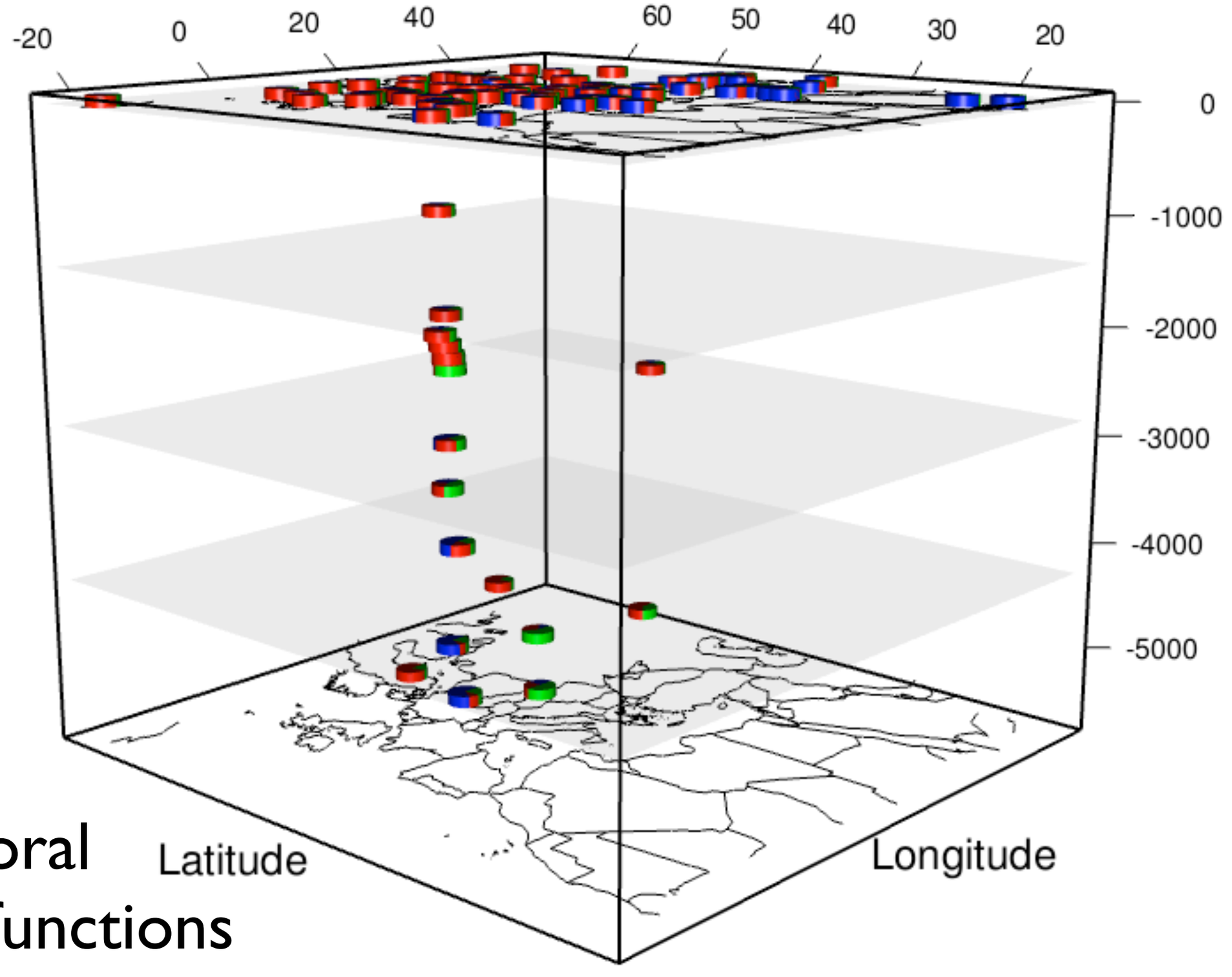
Extensions



Spatiotemporal covariance functions

Haak *et al.* 2015

Extensions



Spatiotemporal covariance functions

Haak *et al.* 2015

Graham Coop
Brad Shaffer

Peter Ralph
Yaniv Brandvain
Luke Mahler

Center for
Population Biology

Marjorie Weber
Will Wetzel
Sharon Strauss
Annie Schmitt
Michael Turelli

Thanks!

@gbradburd
genescape.org
github.com/gbradburd/

Marjorie Weber
Will Wetzel
Sharon Strauss
Annie Schmitt
Michael Turelli

BRADBURD LAB

Fall 2016!

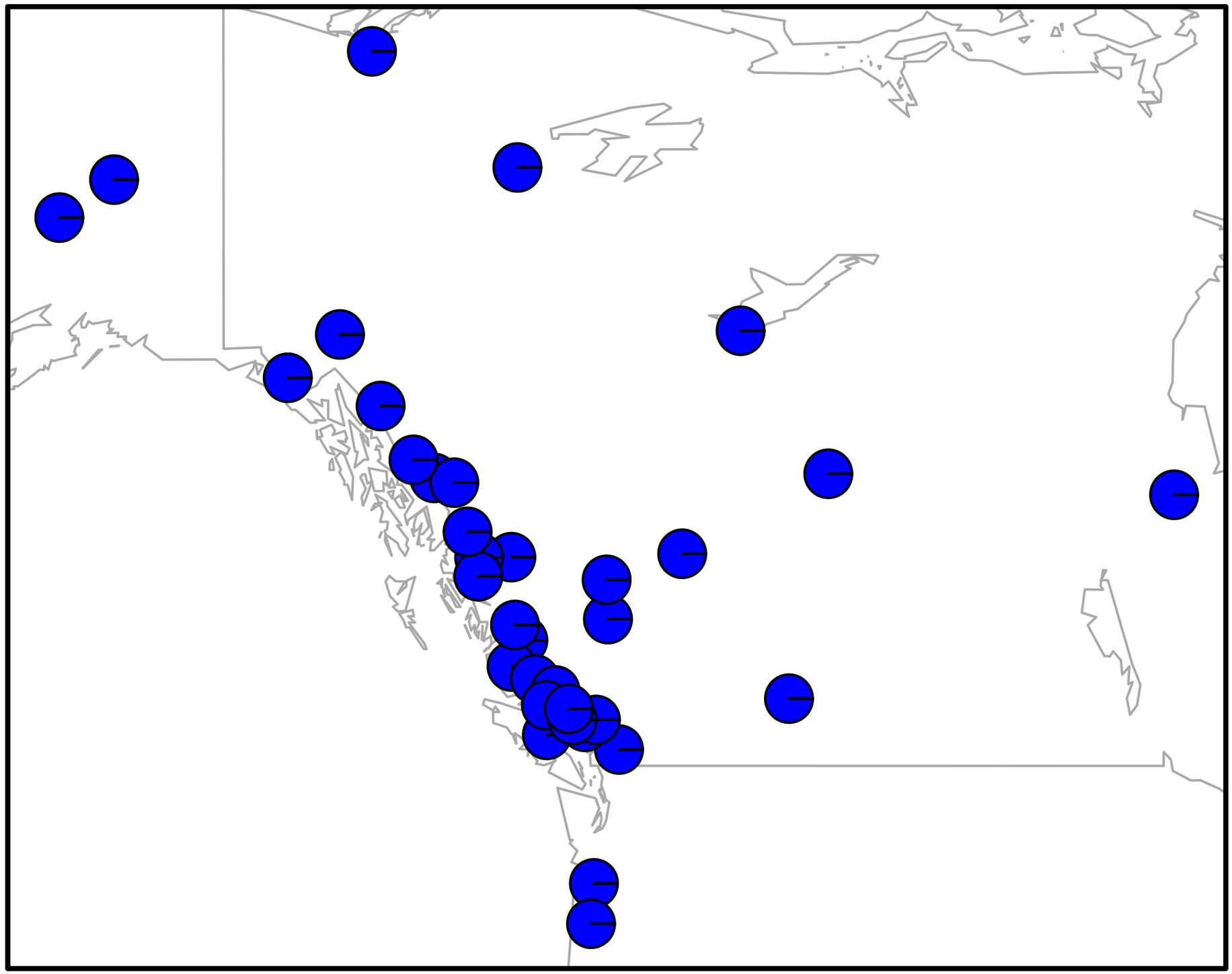
MICHIGAN STATE
UNIVERSITY

gbradburd@
ucdavis.edu

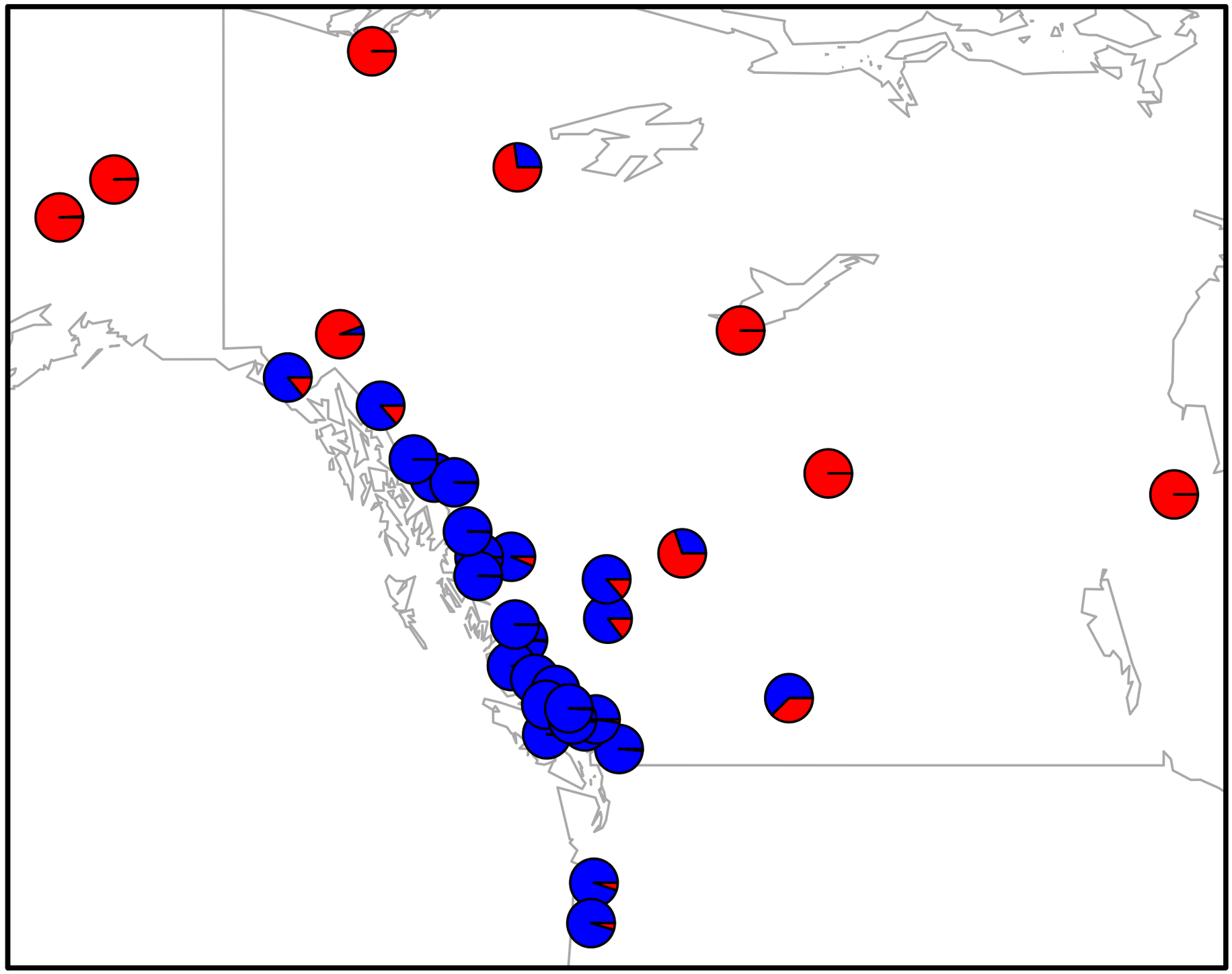
@gbradburd
genescape.org

SLIDES FOR DAYS

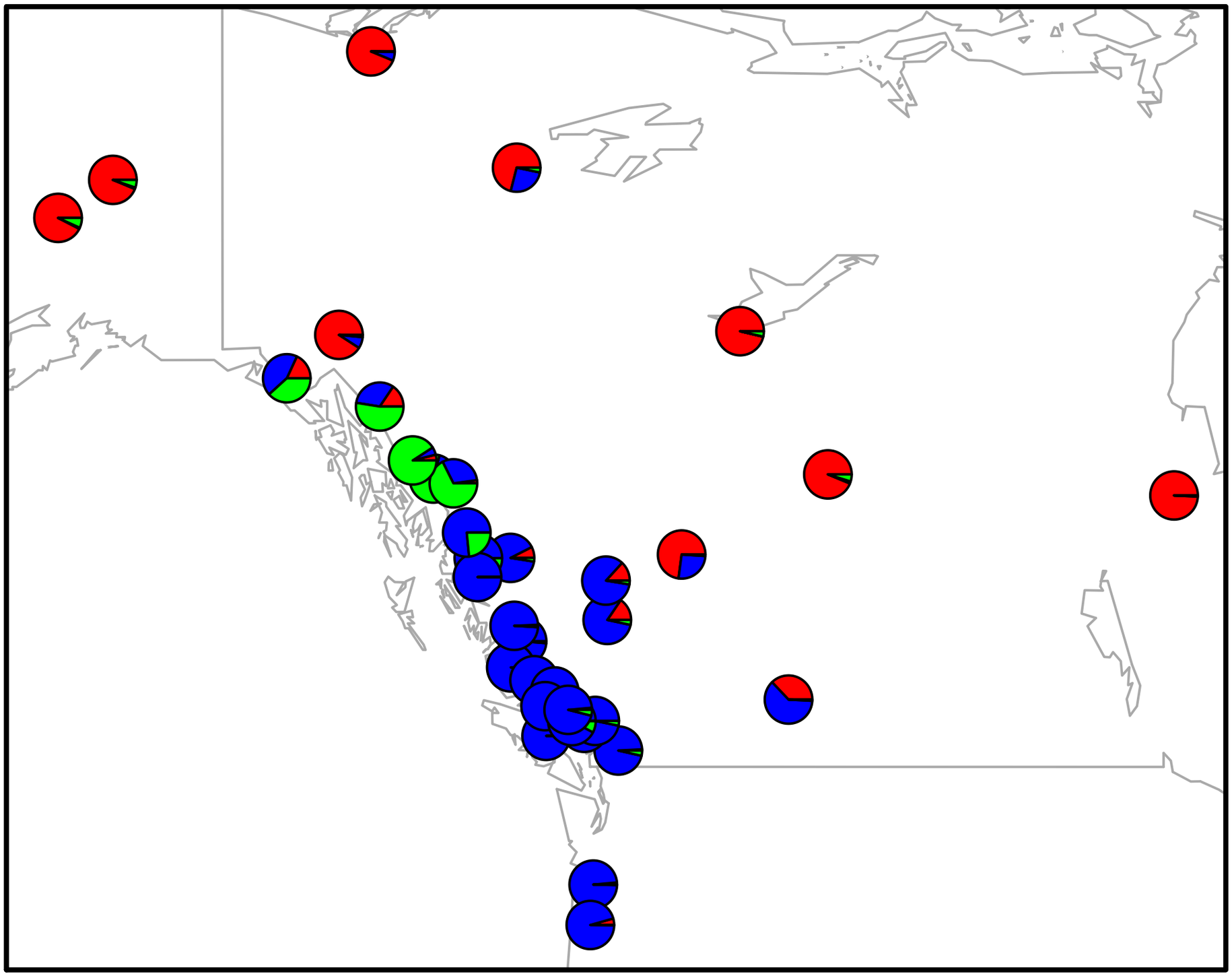
Spatial: K=1



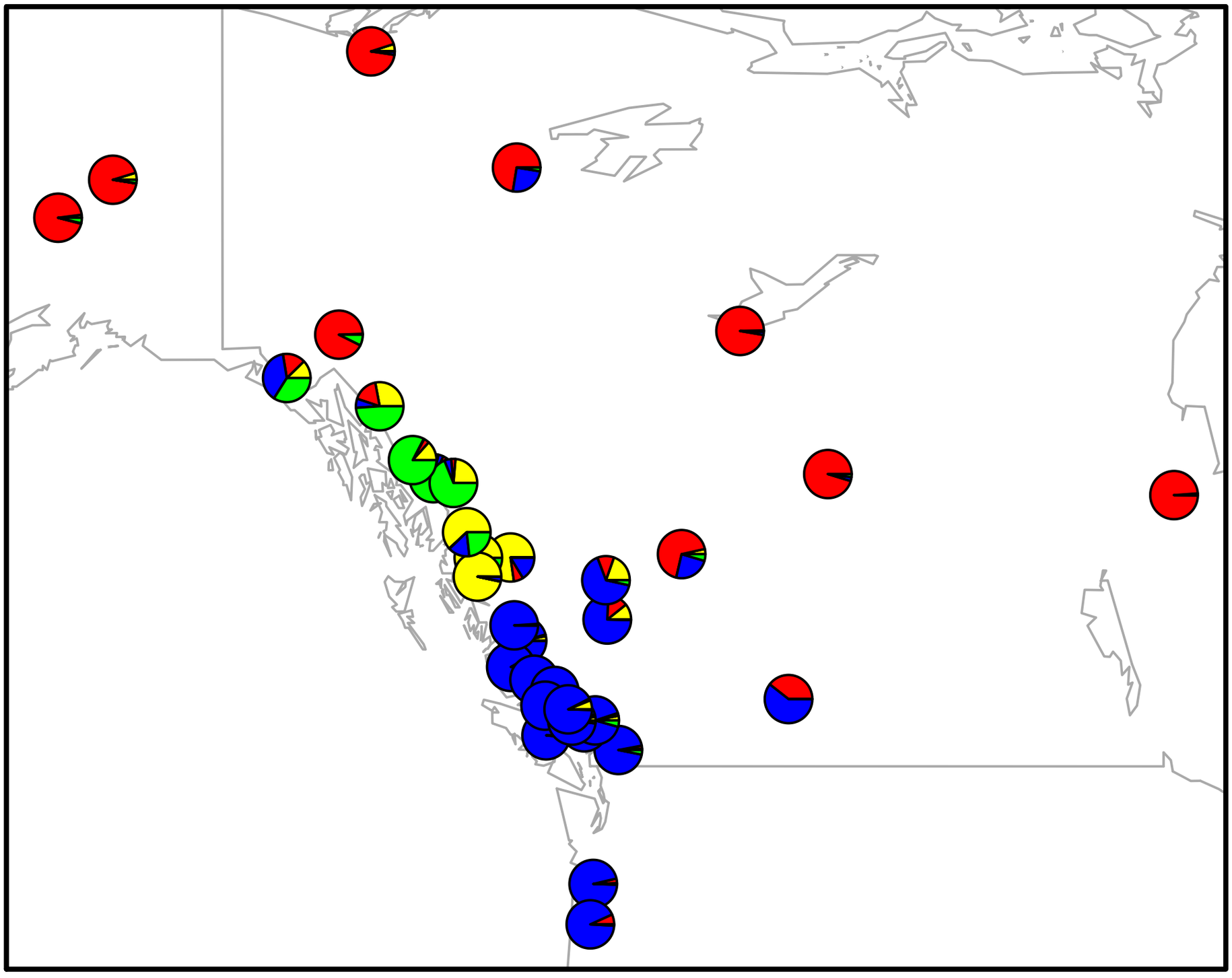
Spatial: K=2



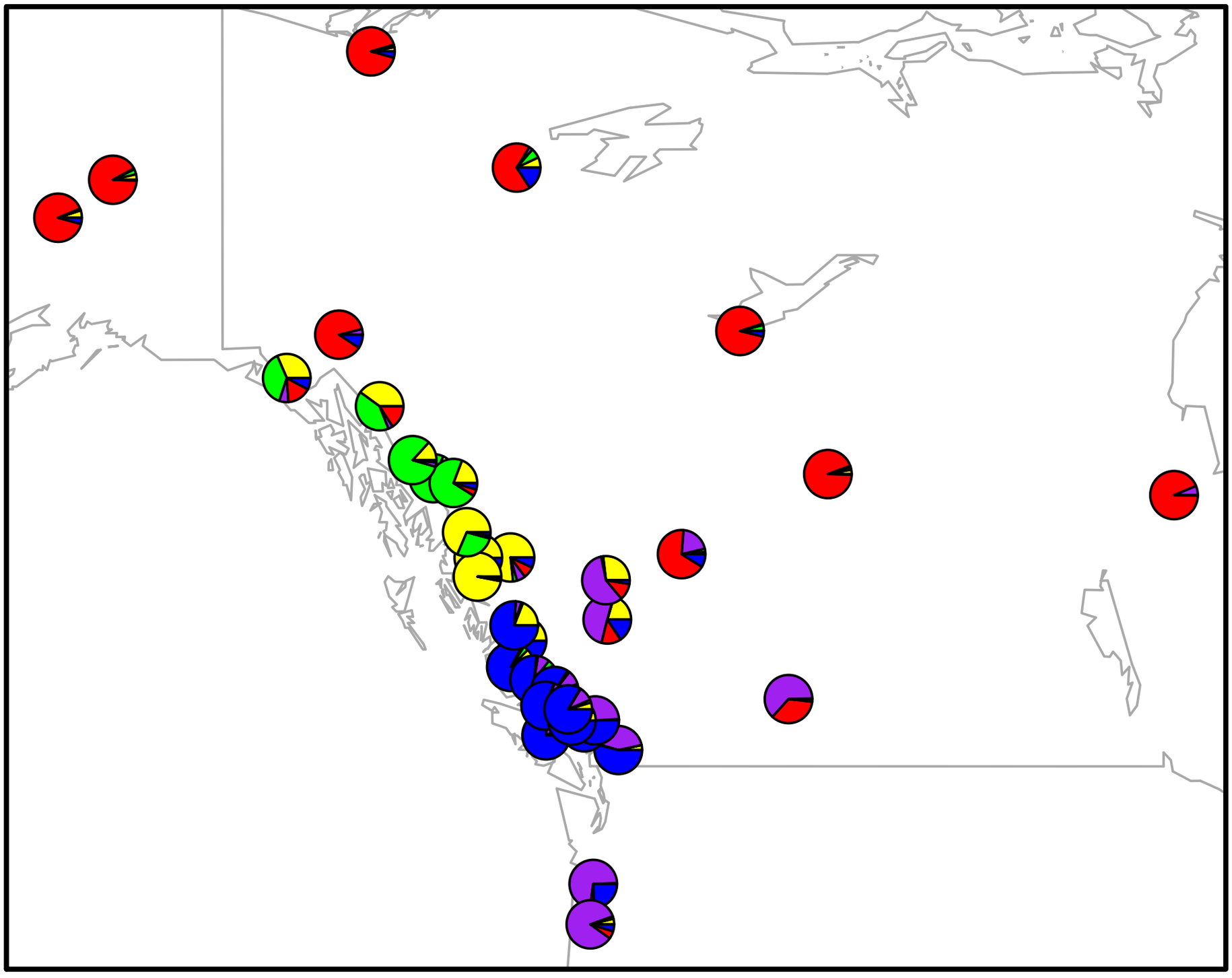
Spatial: K=3



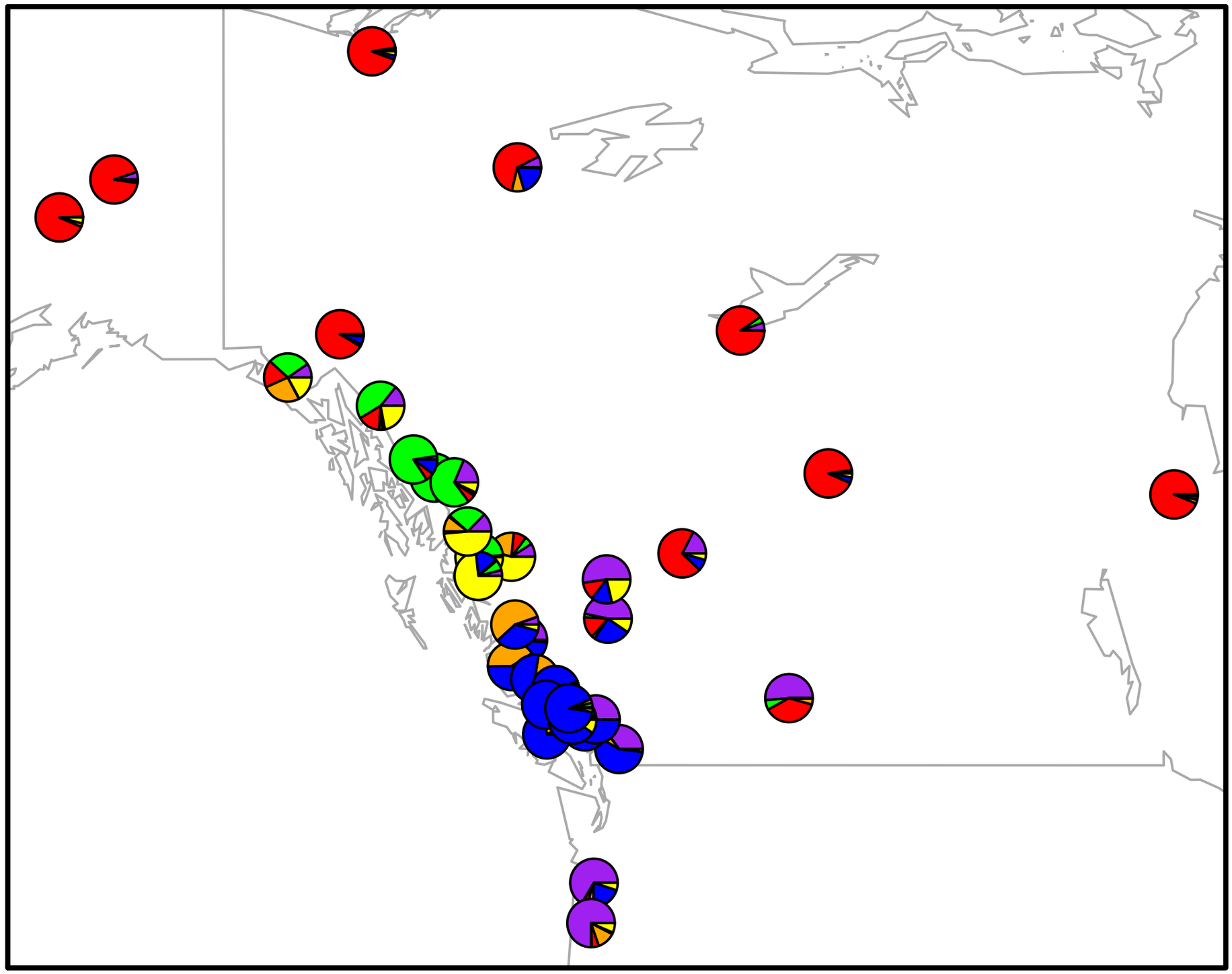
Spatial: K=4



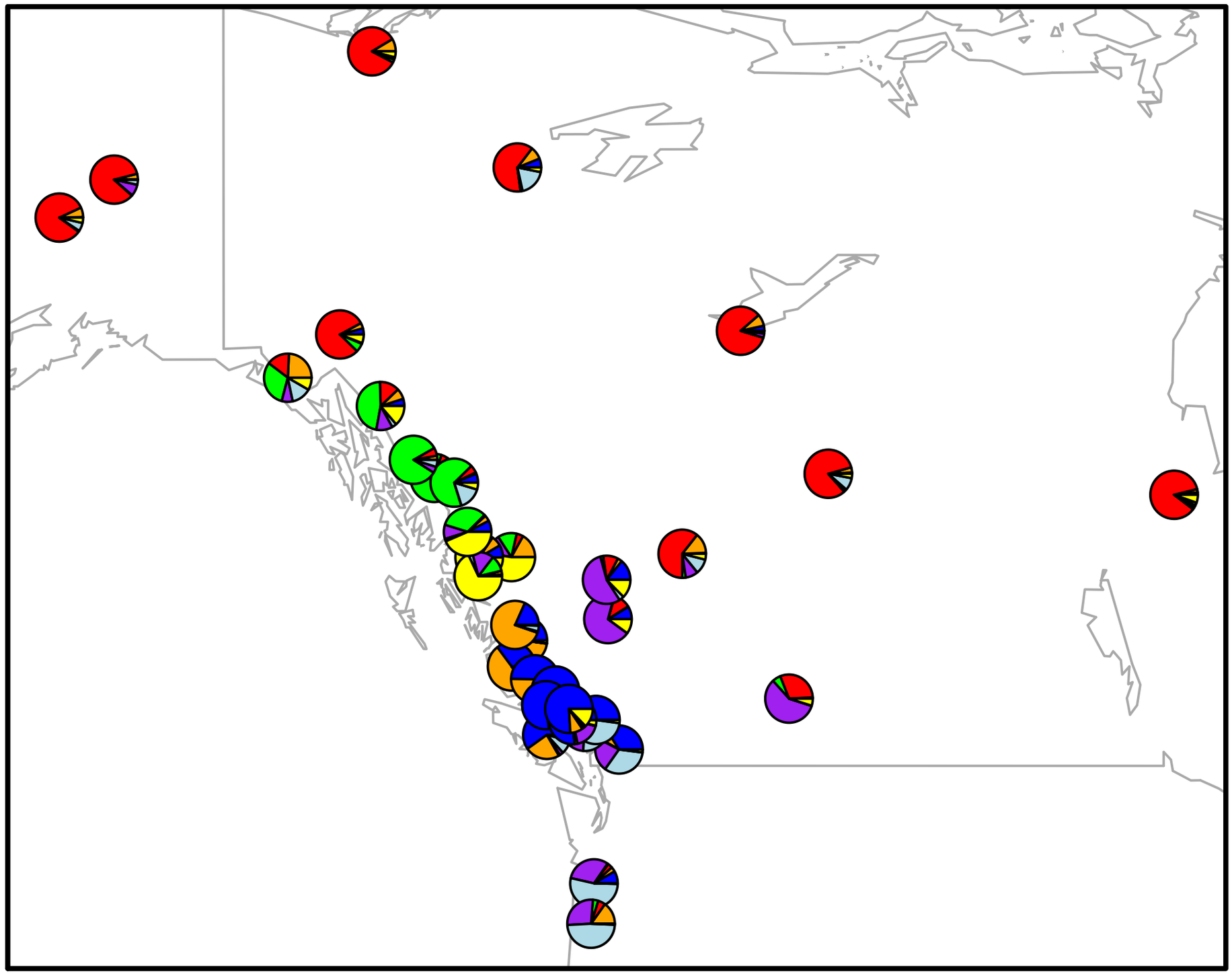
Spatial: K=5



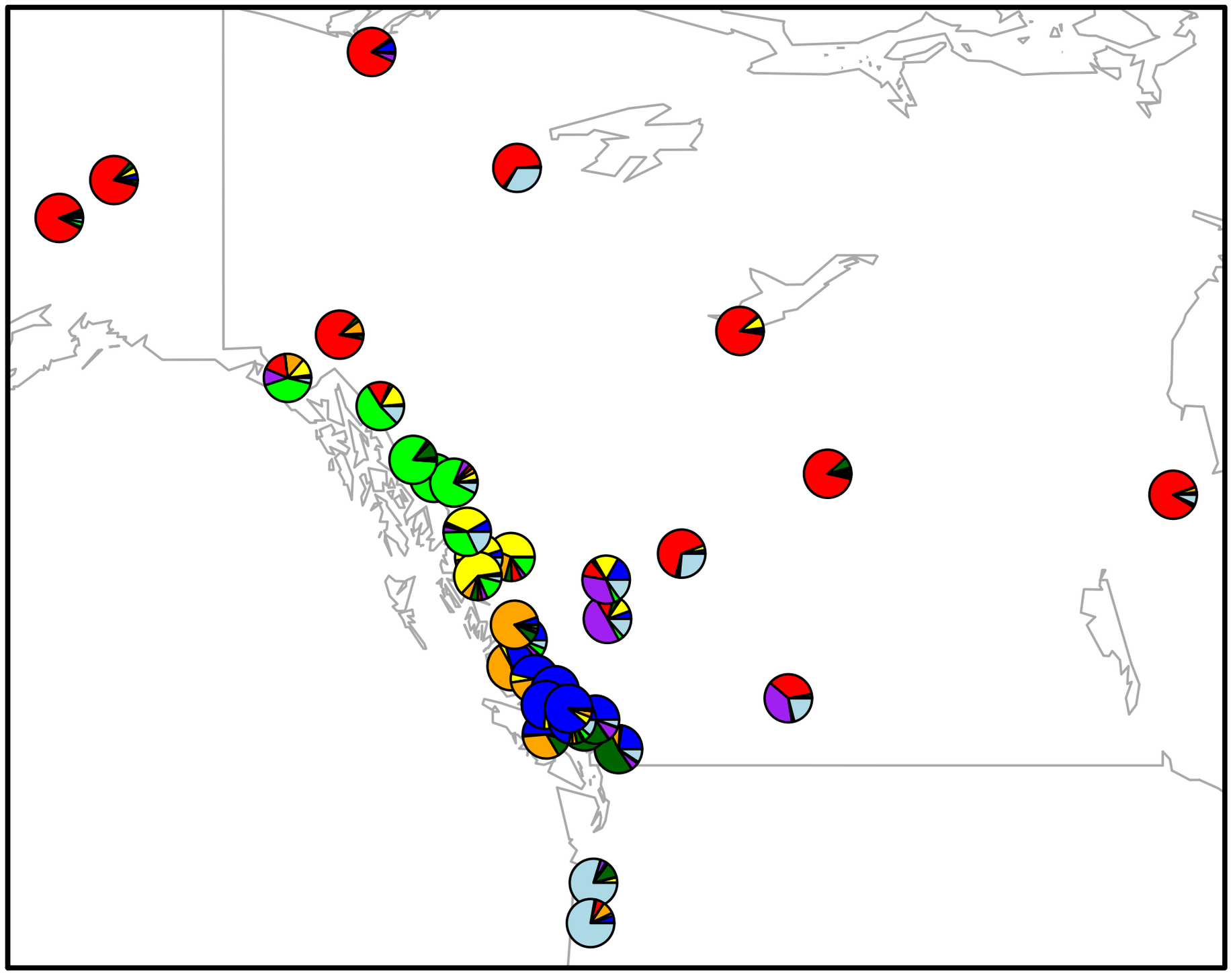
Spatial: K=6



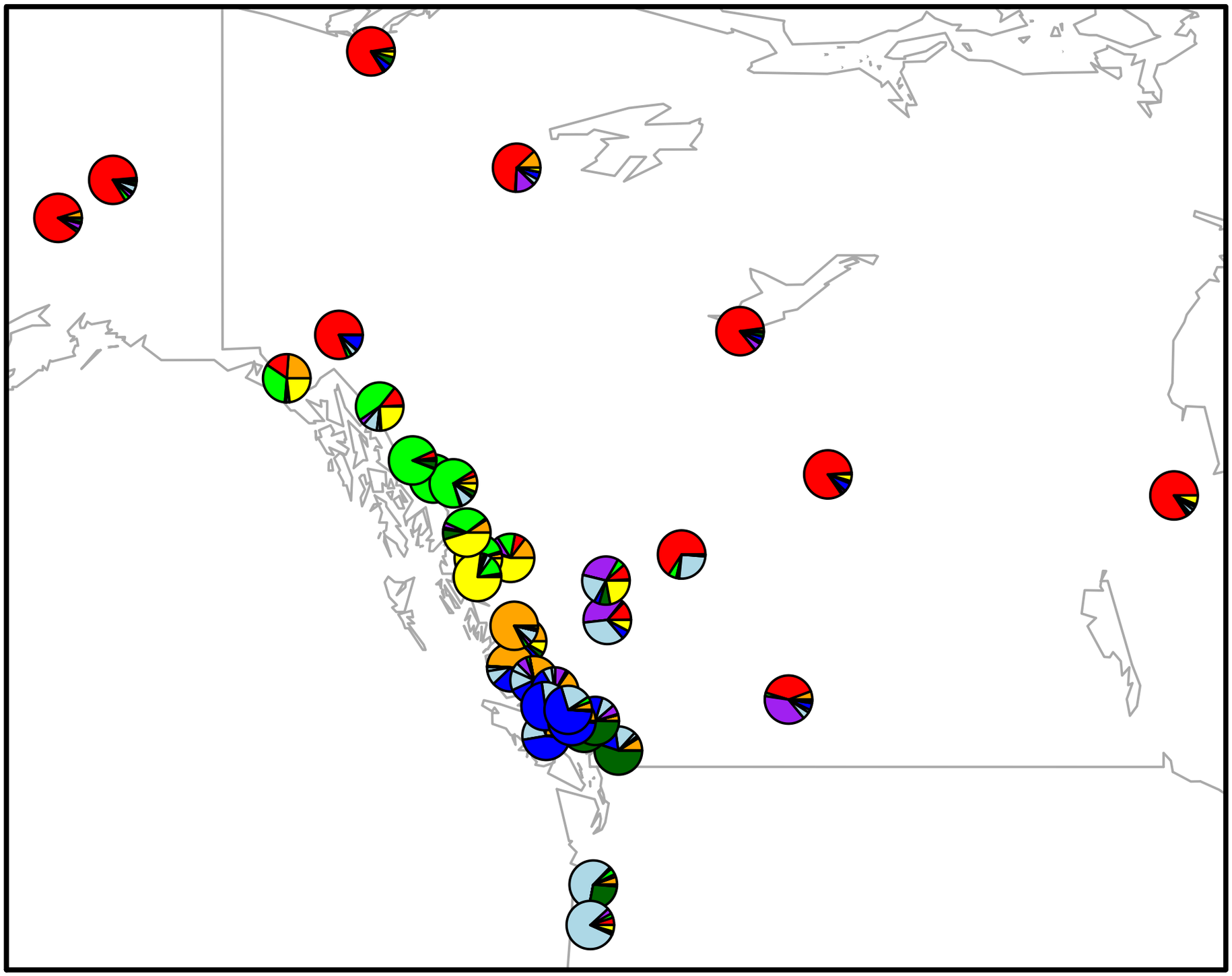
Spatial: K=7



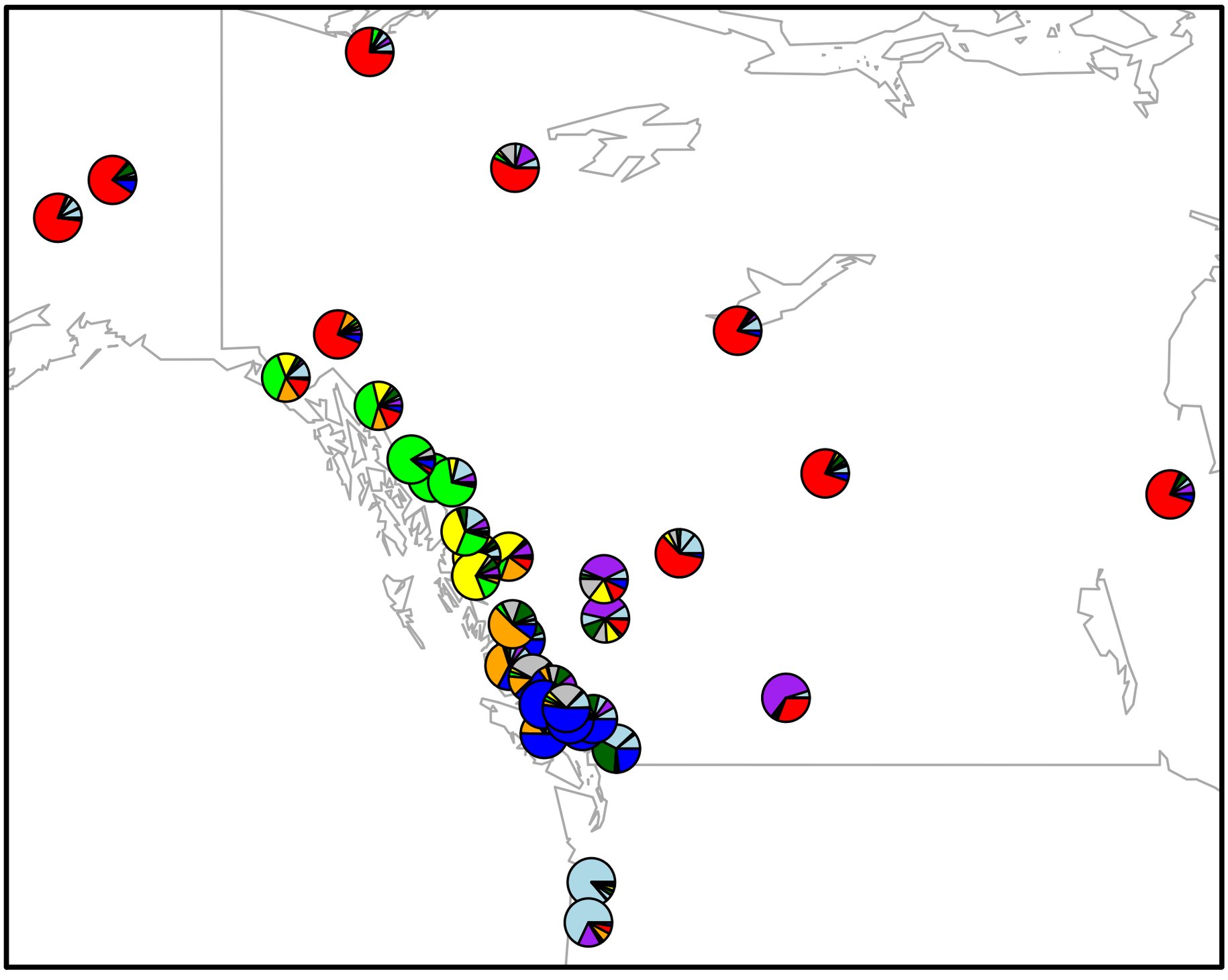
Spatial: K=8



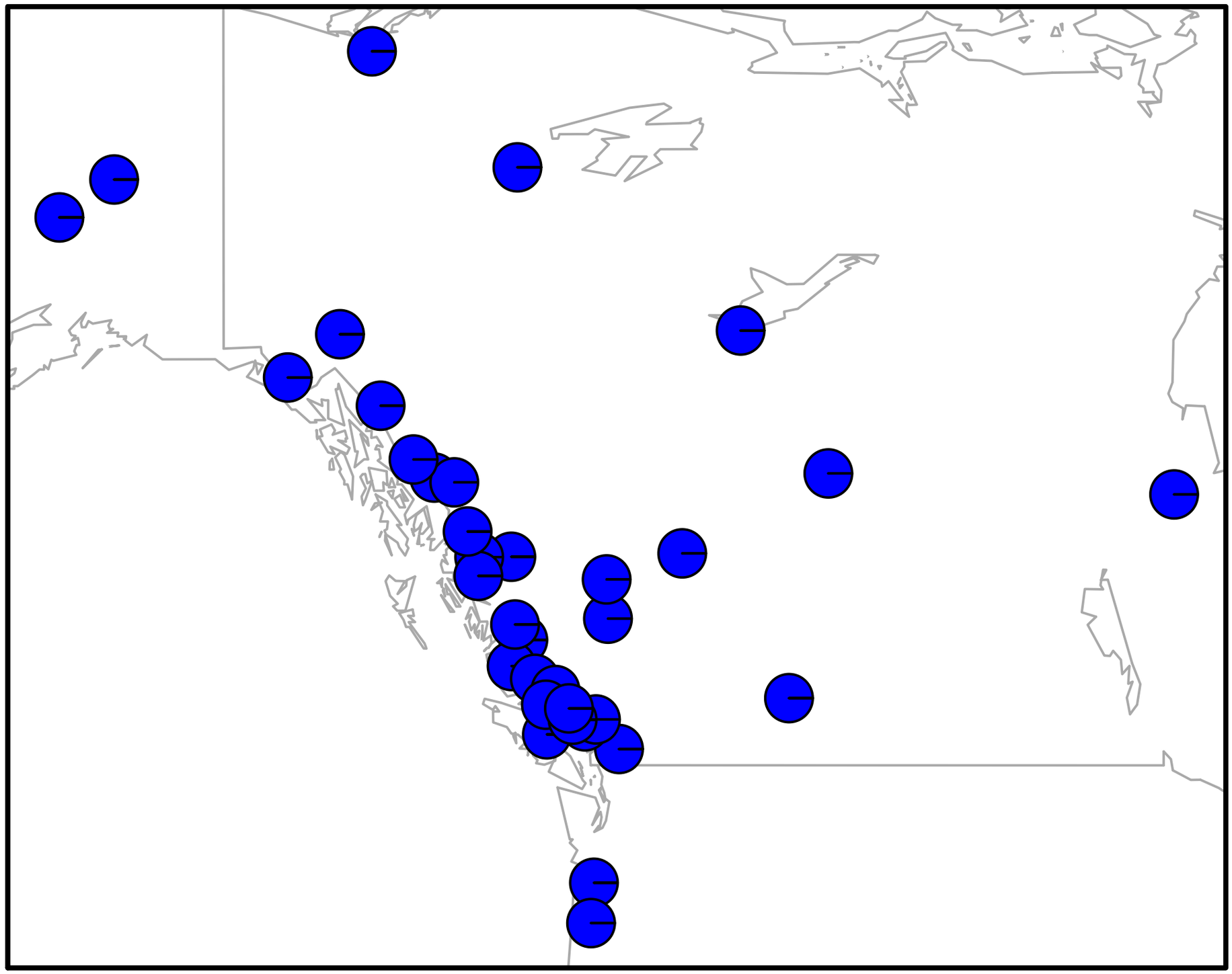
Spatial: K=9



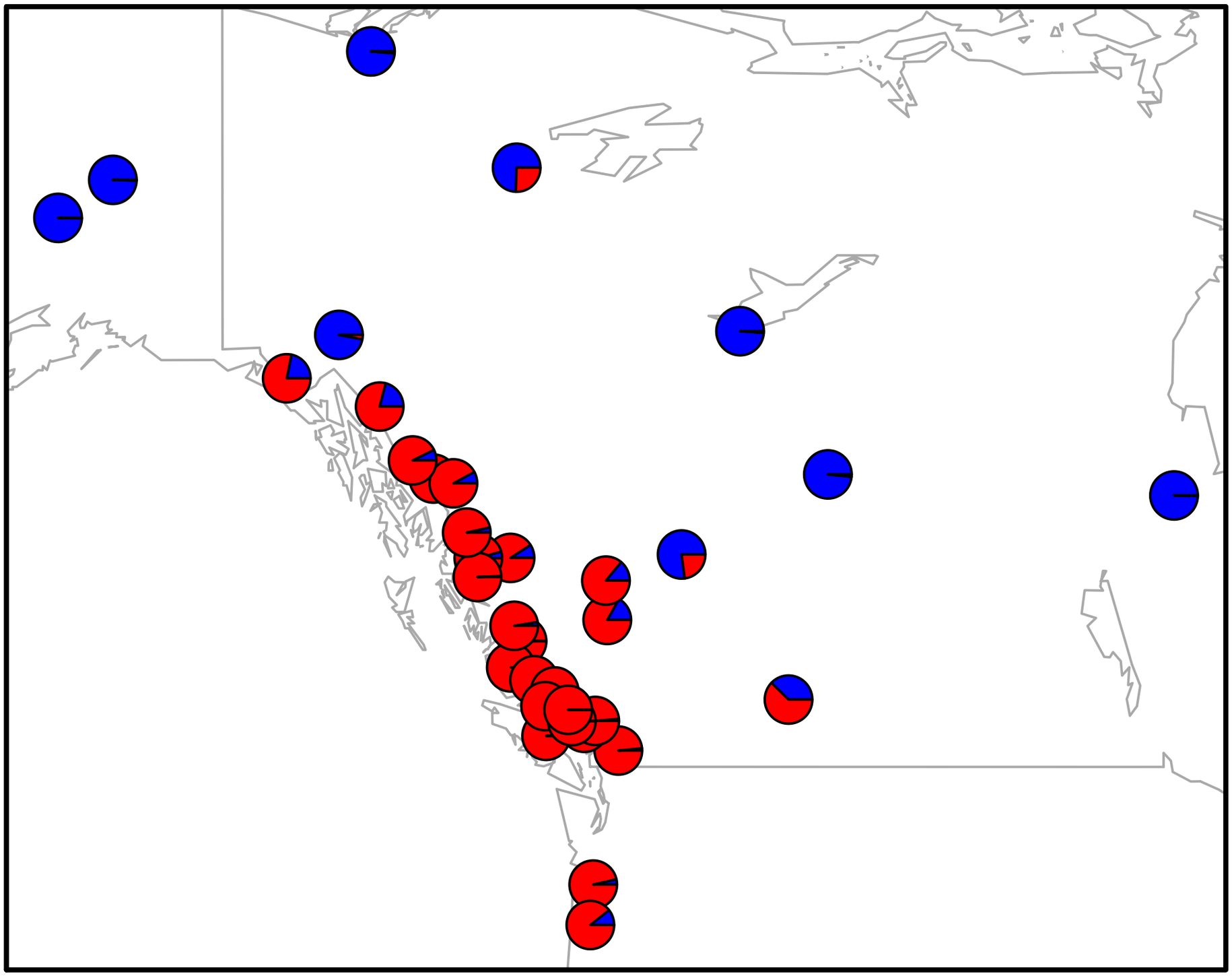
Spatial: K=10



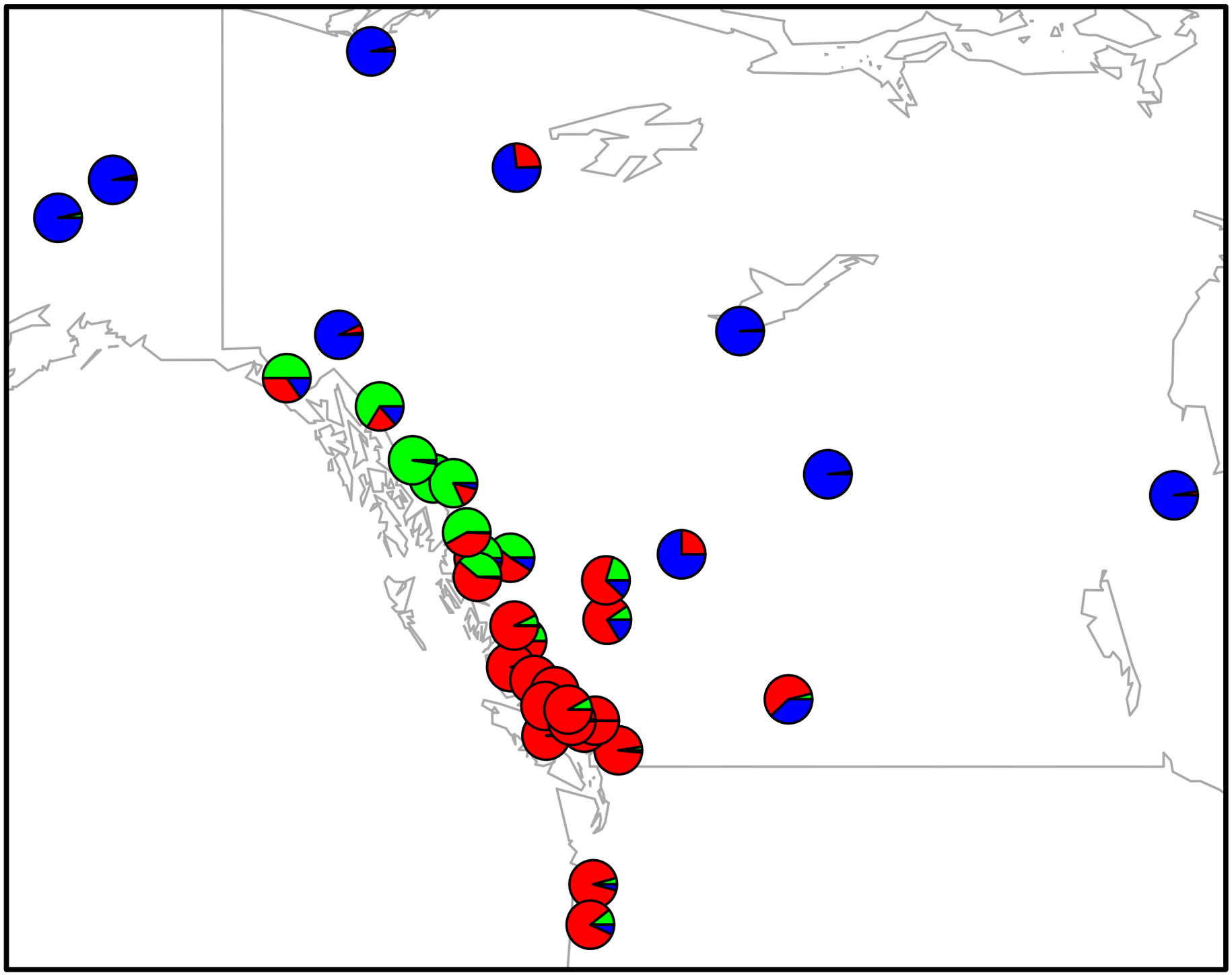
Nonspatial: K=1



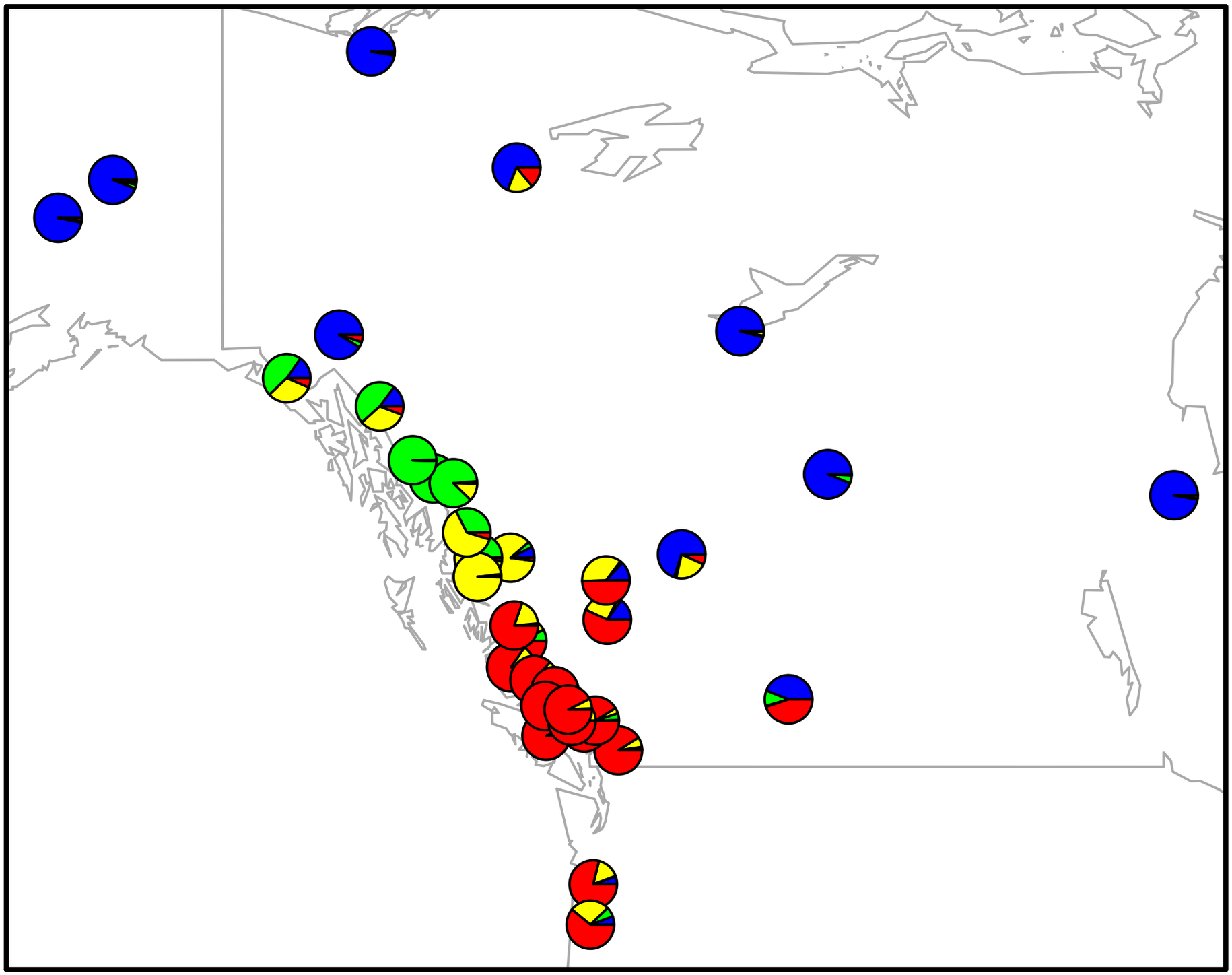
Nonspatial: K=2



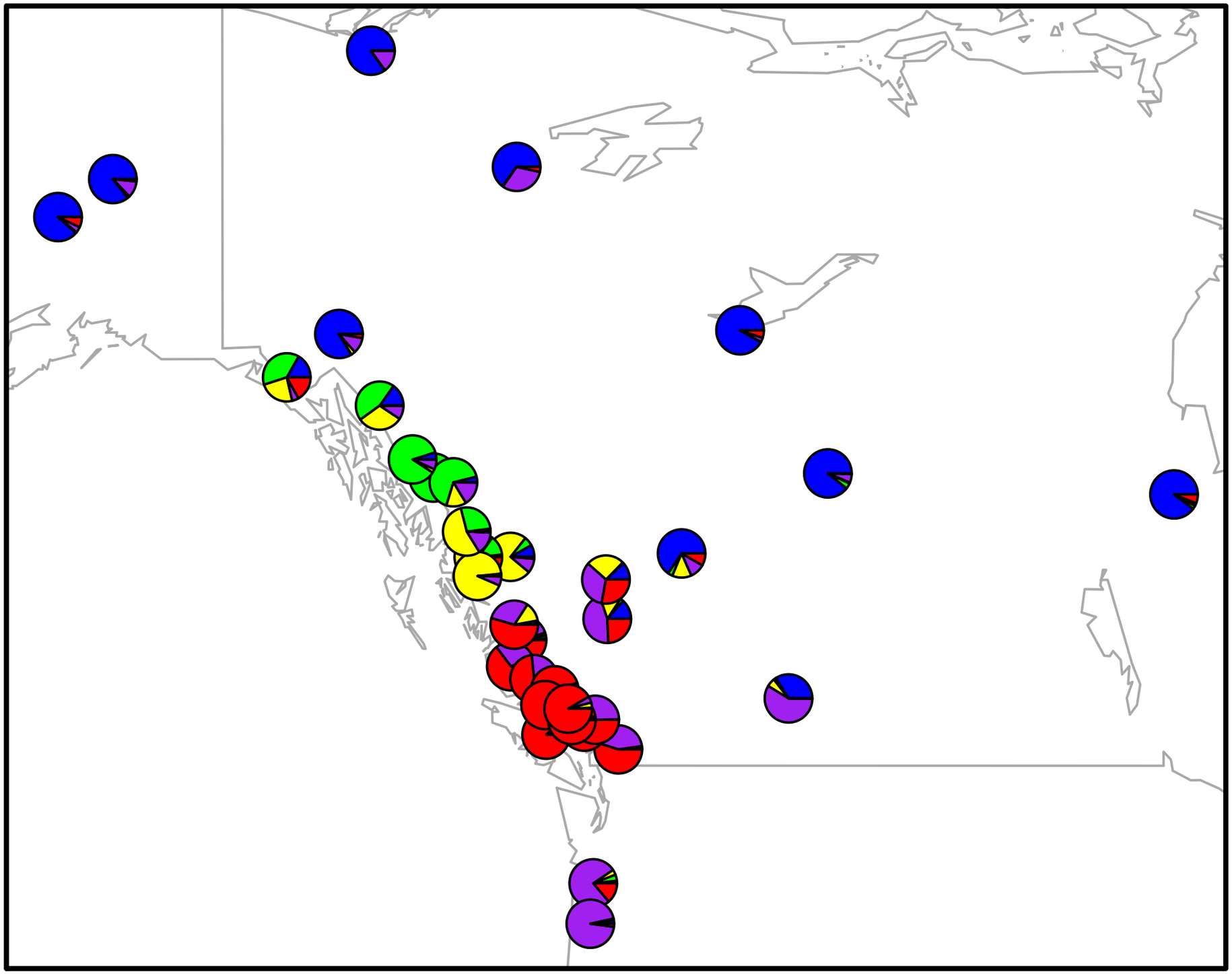
Nonspatial: K=3



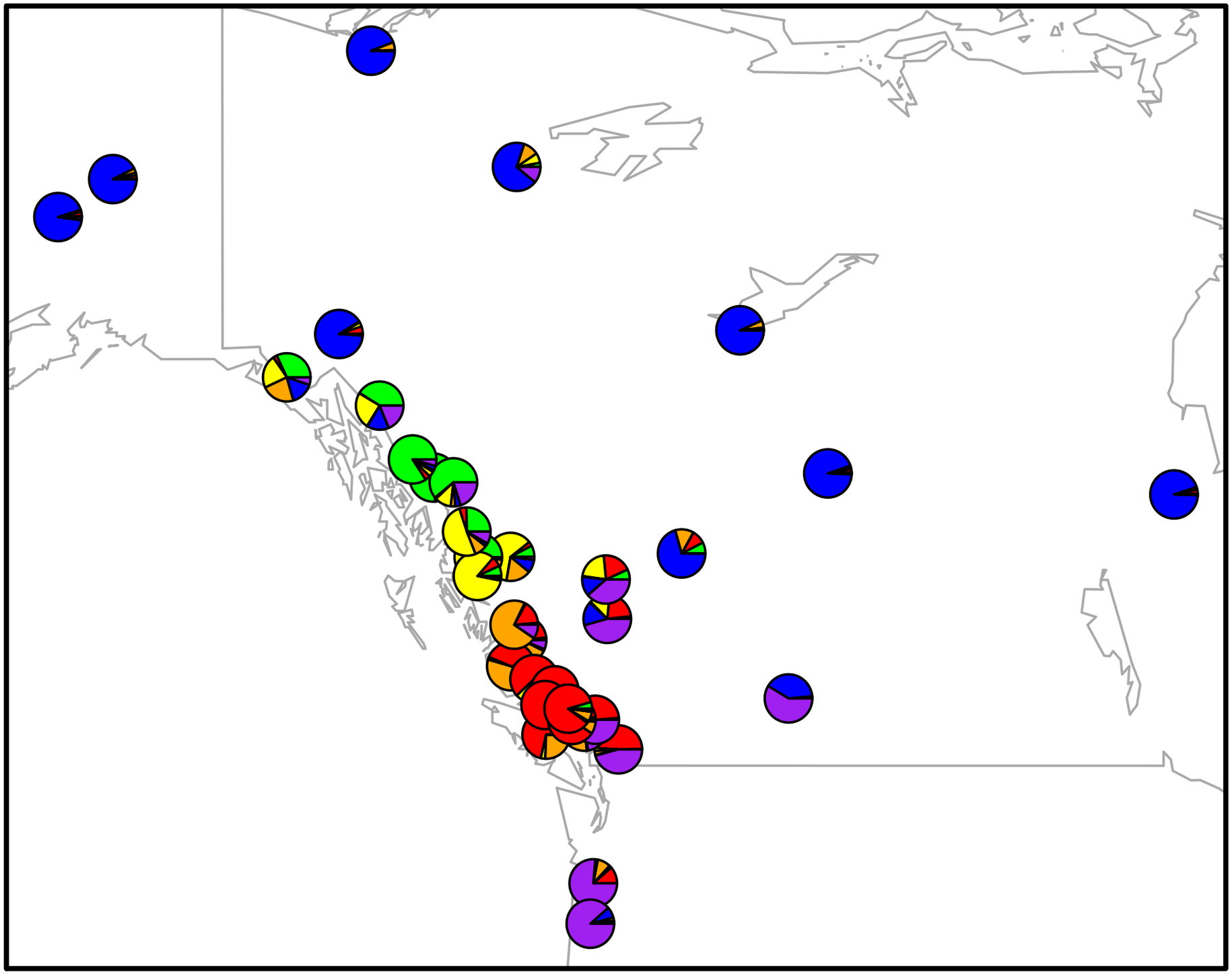
Nonspatial: K=4



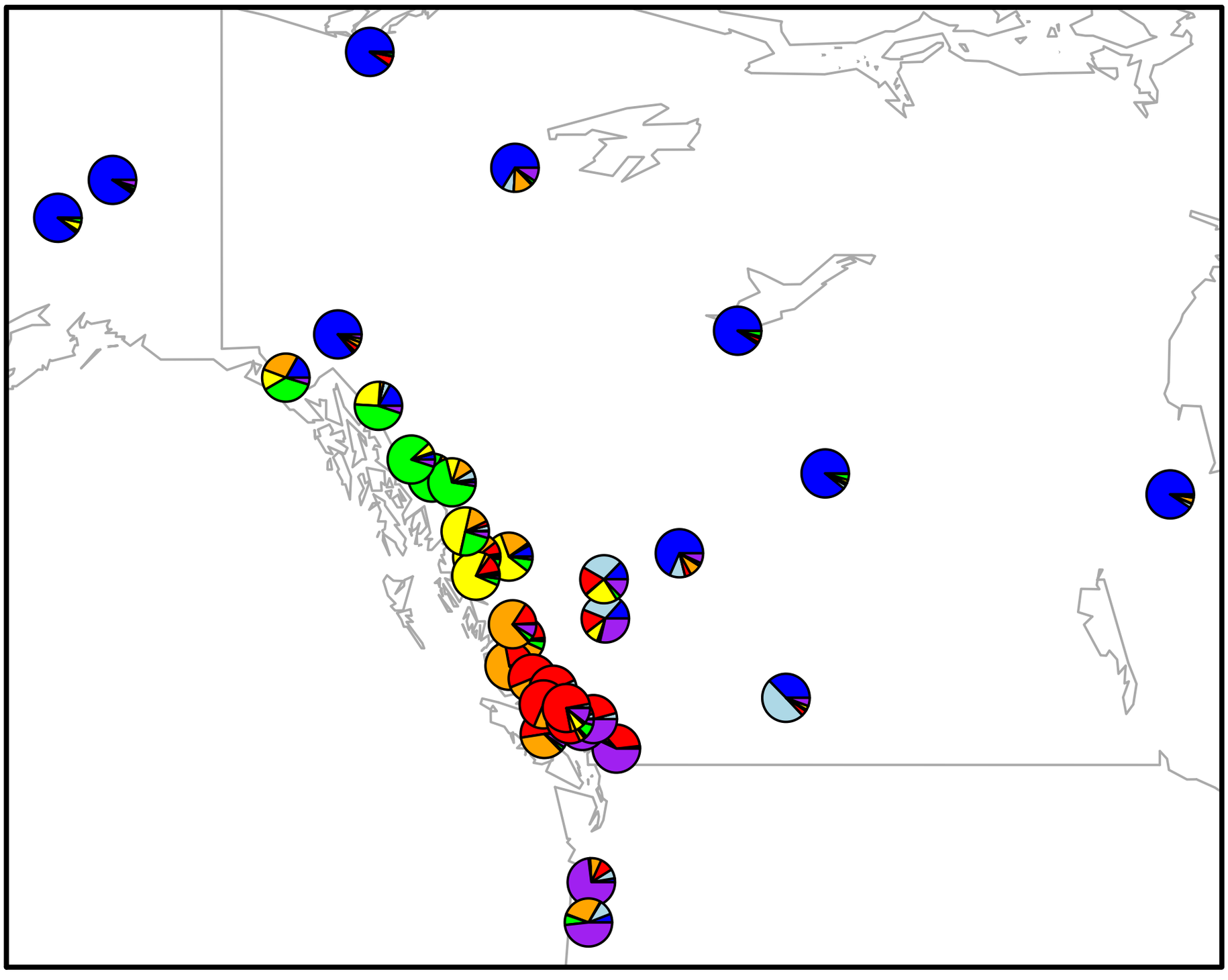
Nonspatial: K=5



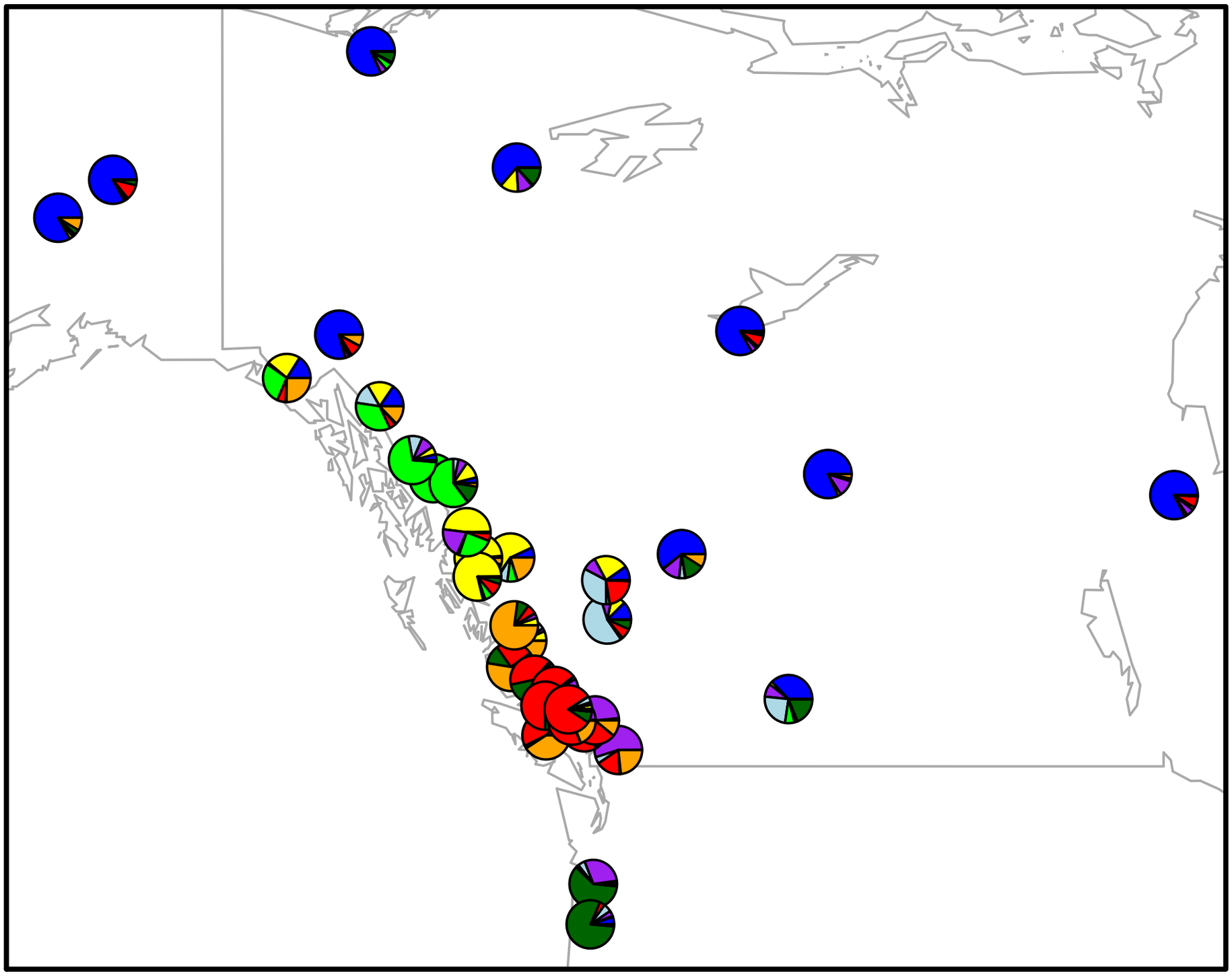
Nonspatial: K=6



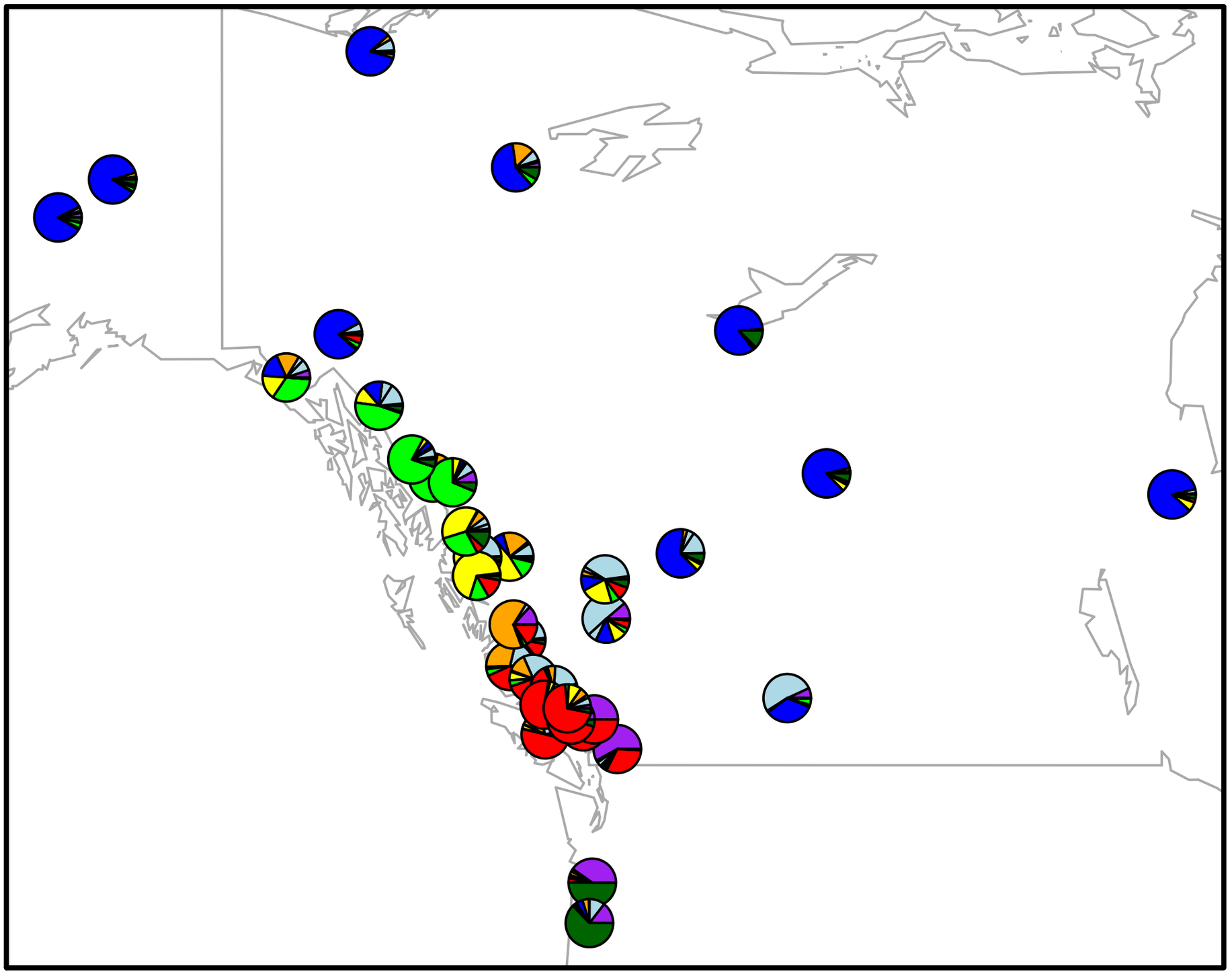
Nonspatial: K=7



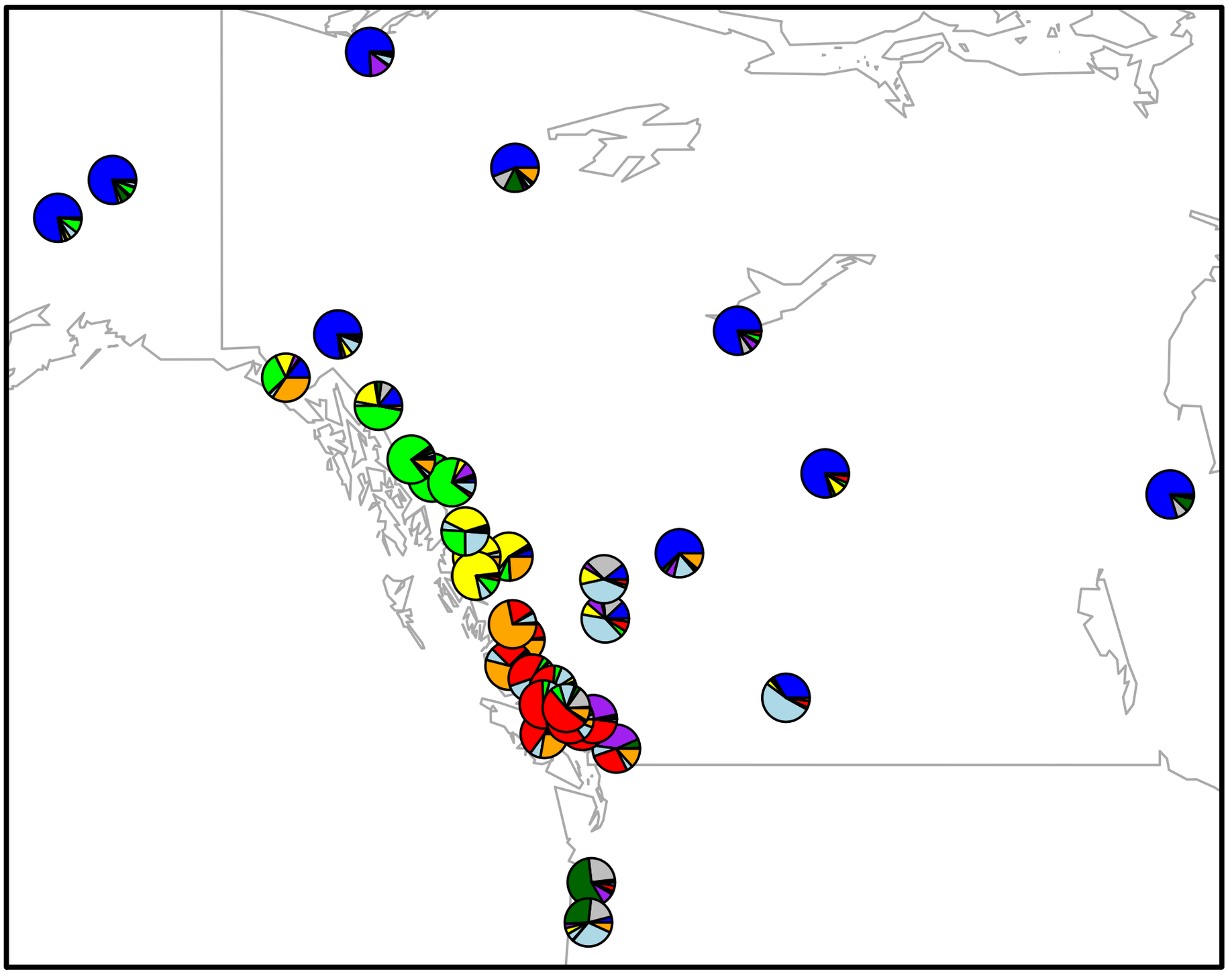
Nonspatial: K=8



Nonspatial: K=9

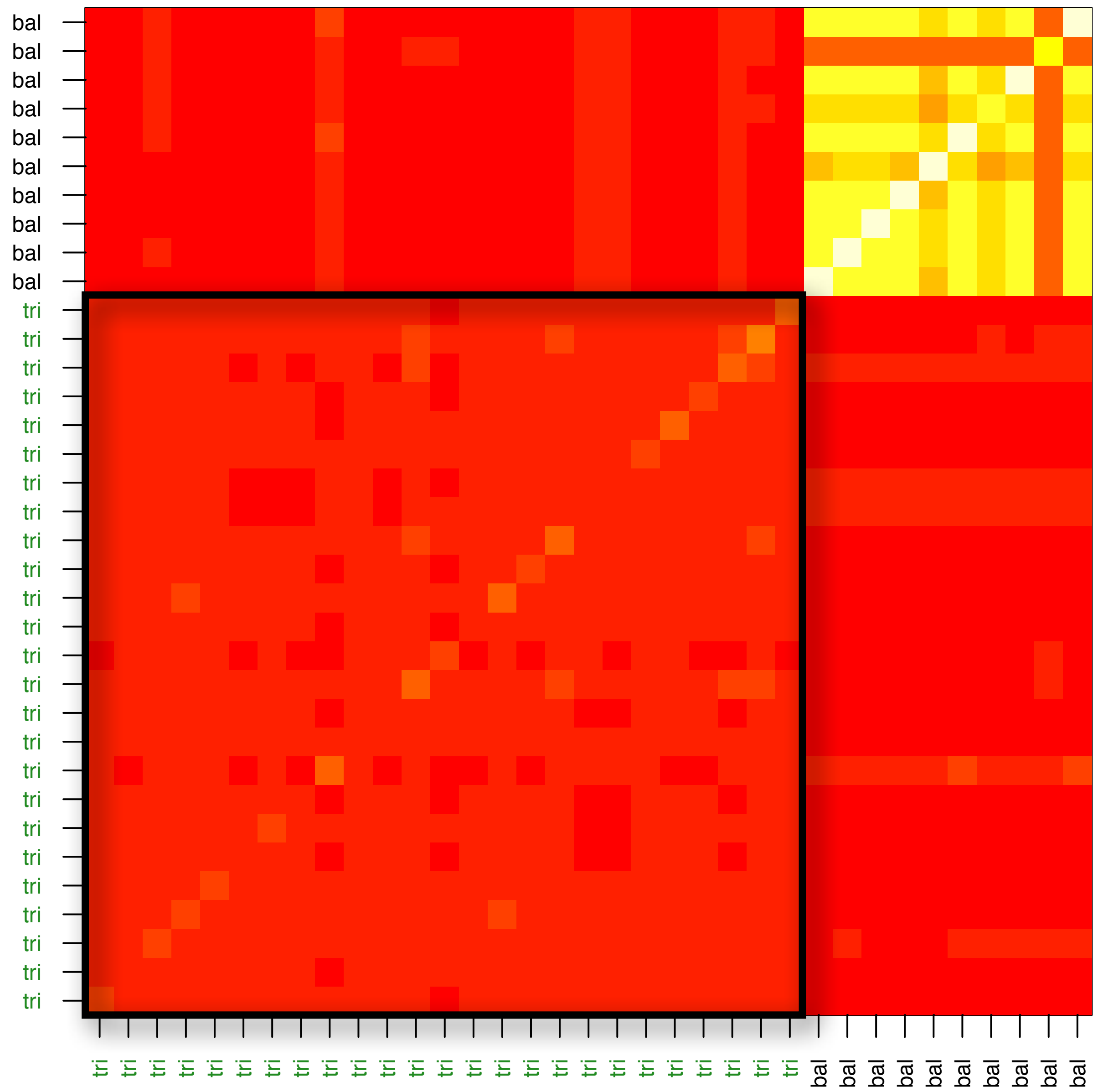


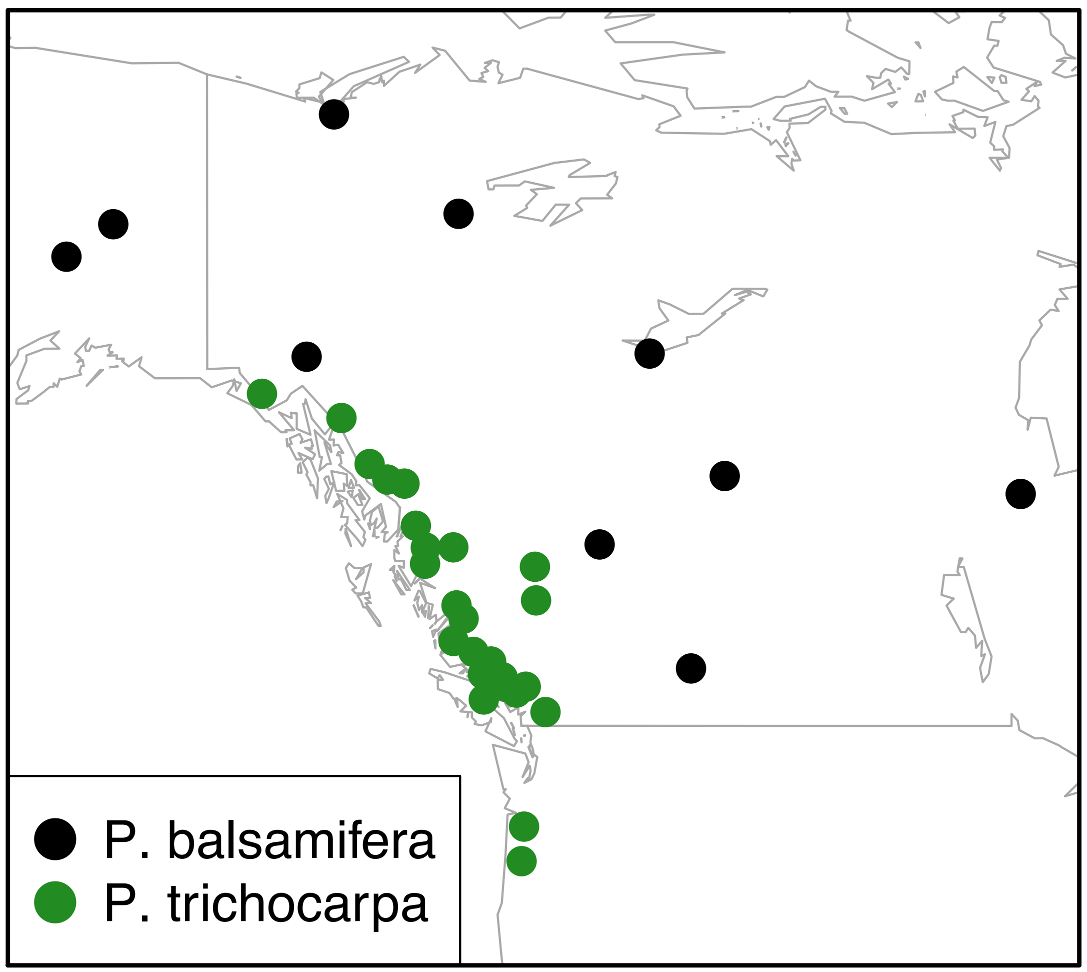
Nonspatial: K=10



Empirical Illustration:
Populus trichocarpa and *Populus balsamifera*

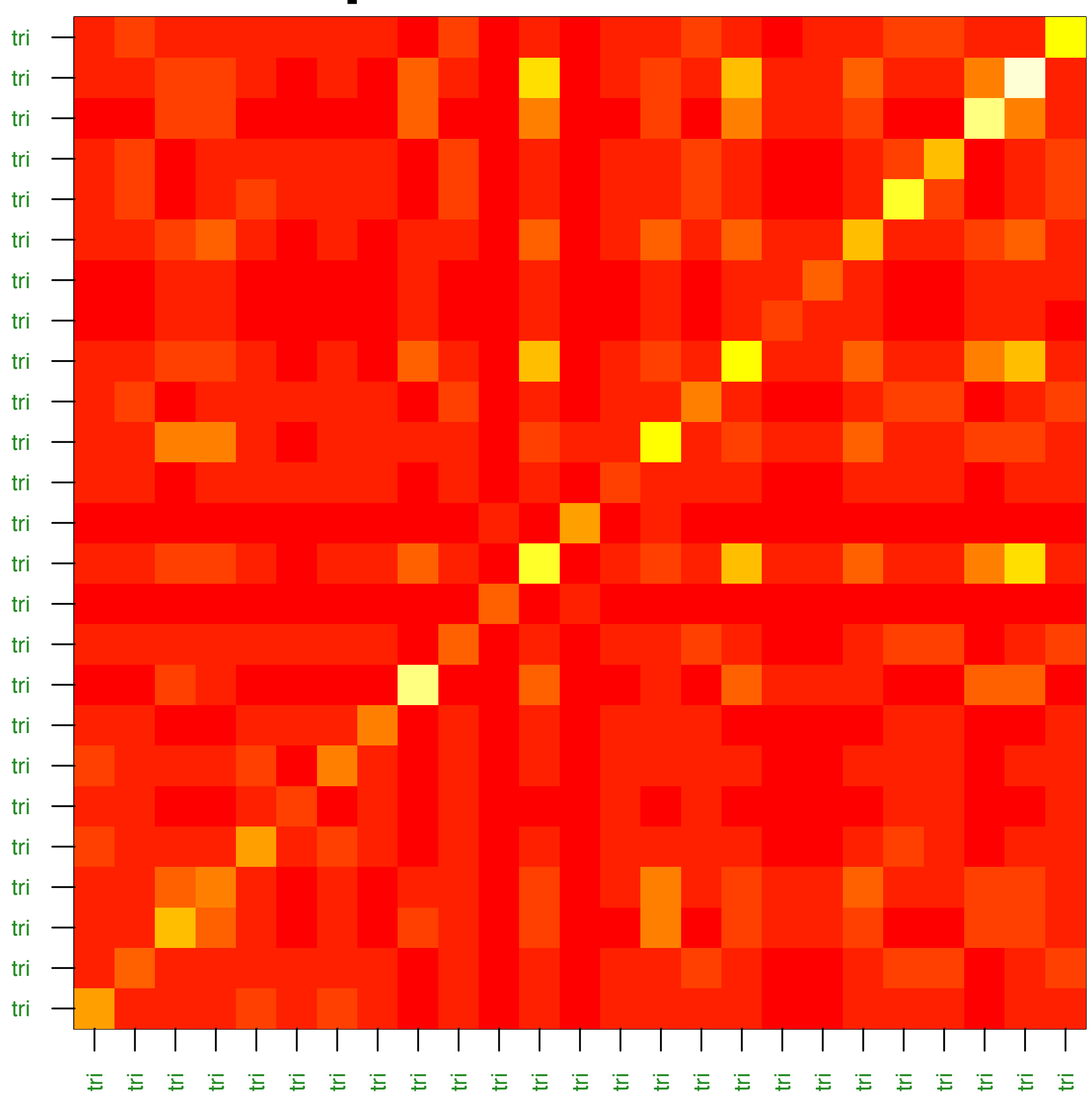
sample covariance

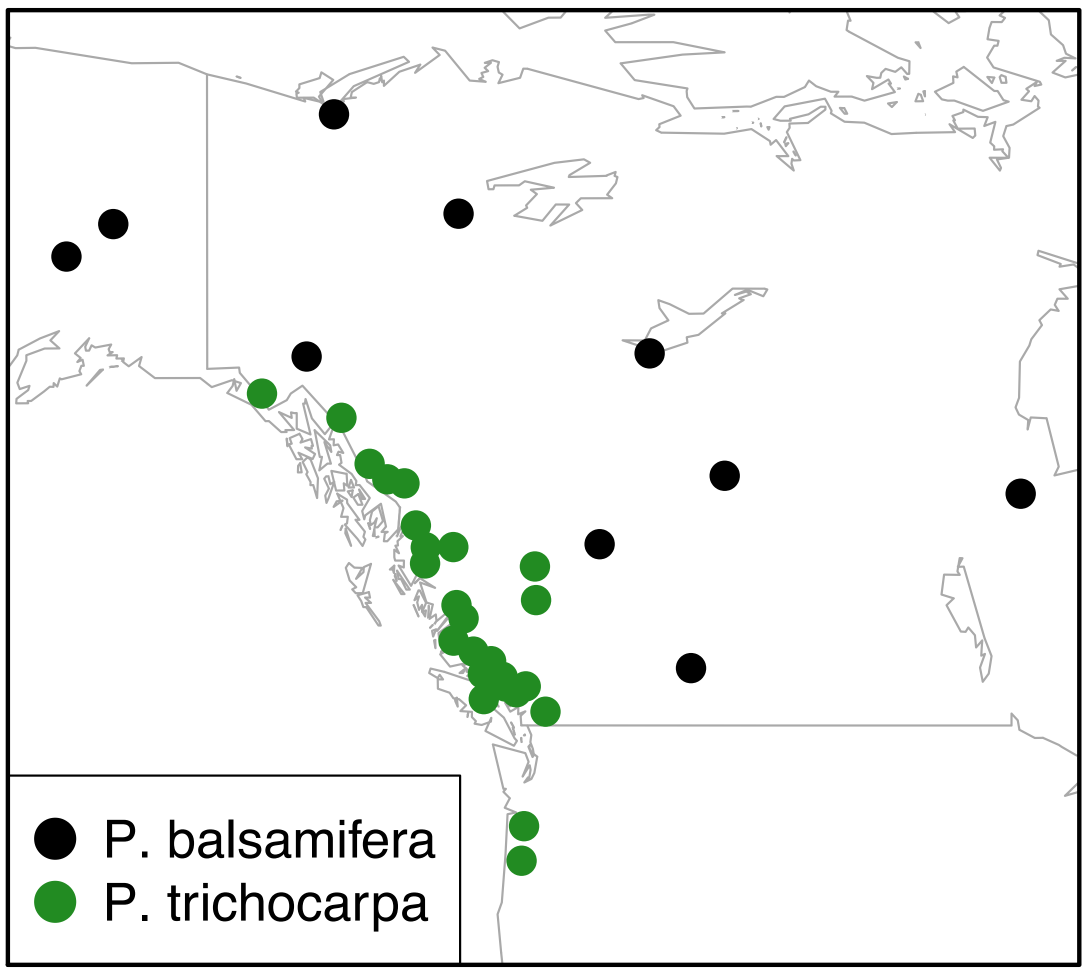




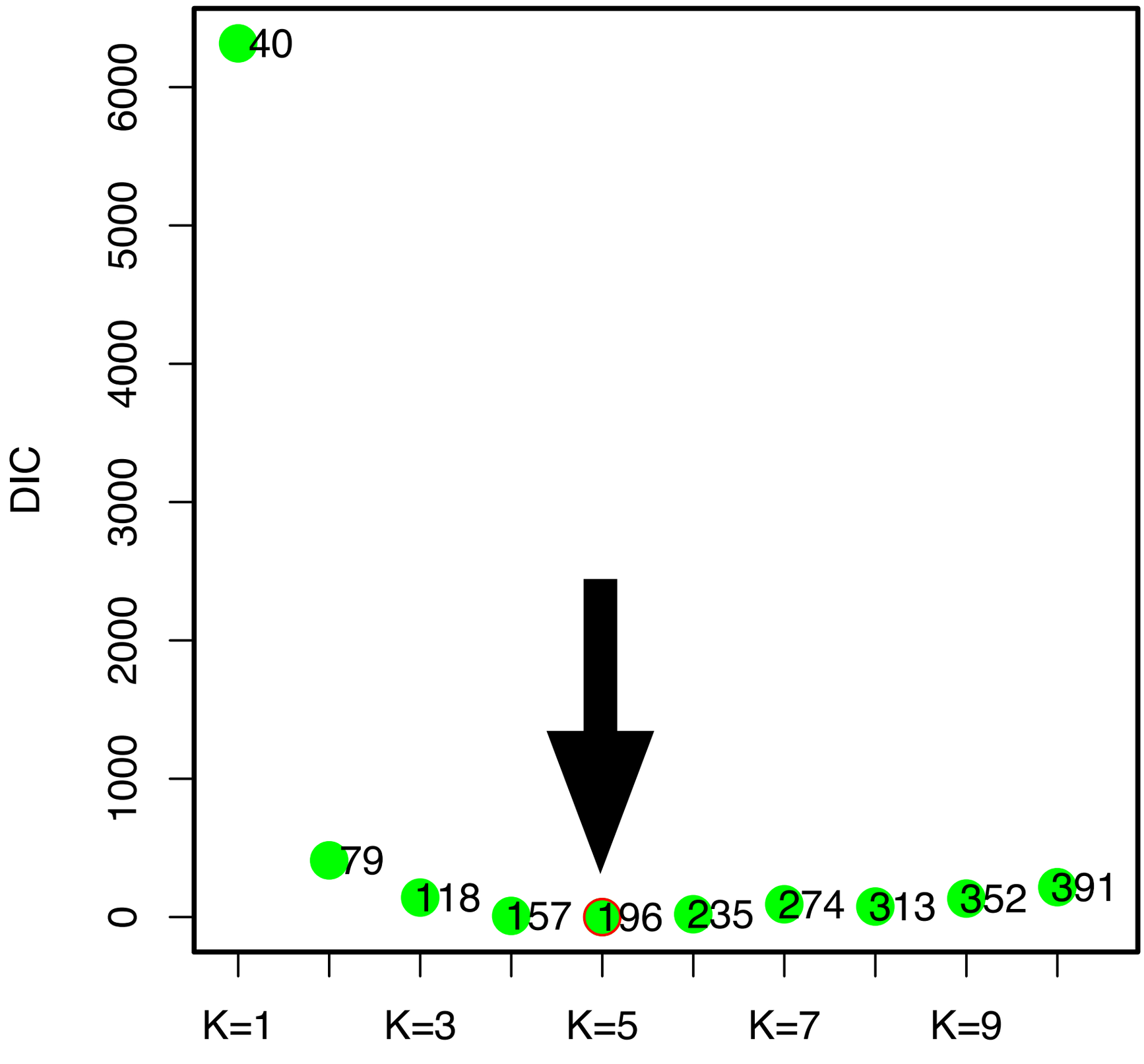
Empirical Illustration:
Populus trichocarpa and *Populus balsamifera*

sample covariance

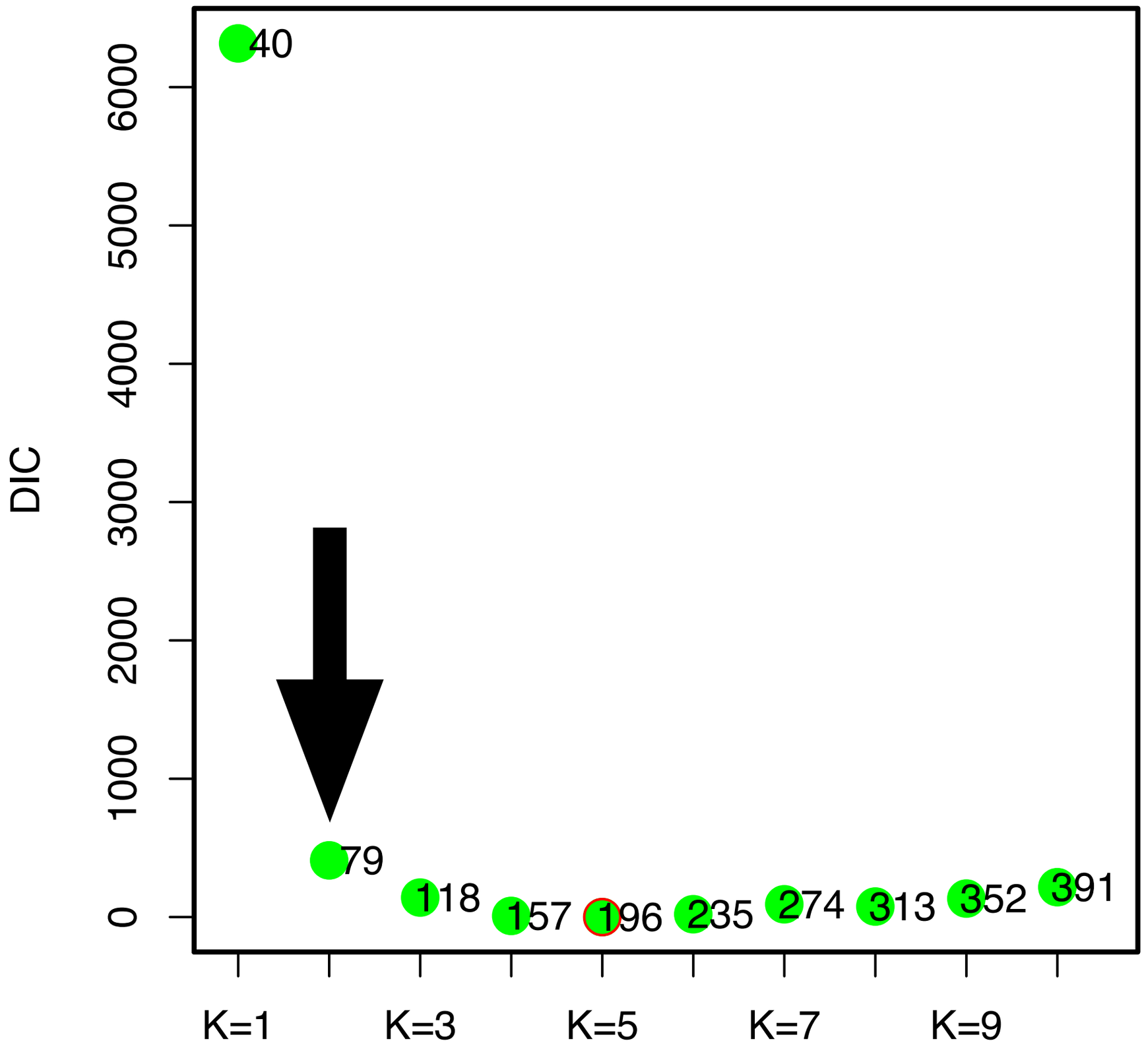


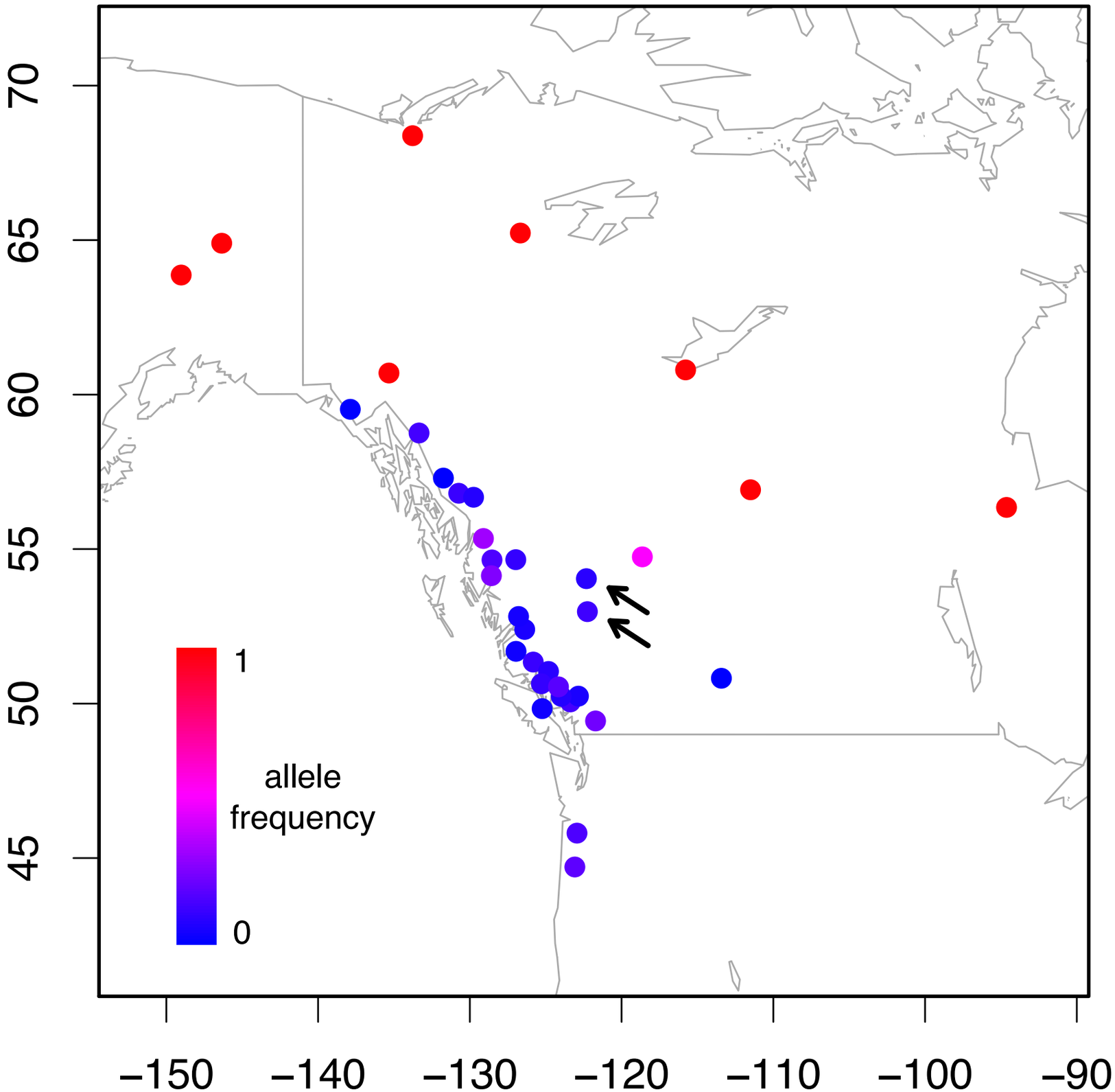


Model comparisons

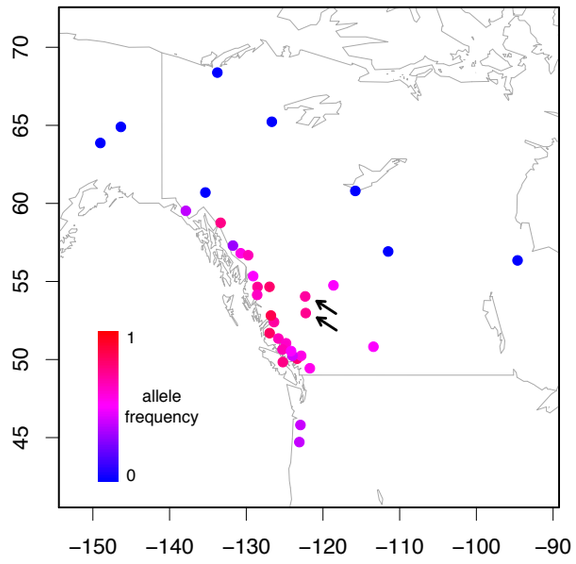


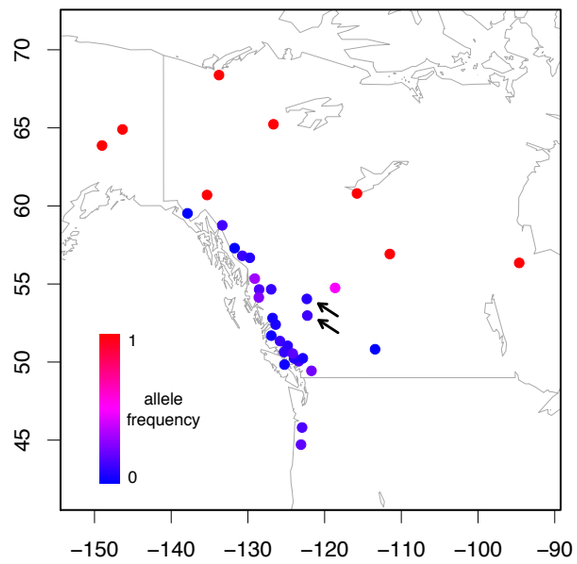
Model comparisons

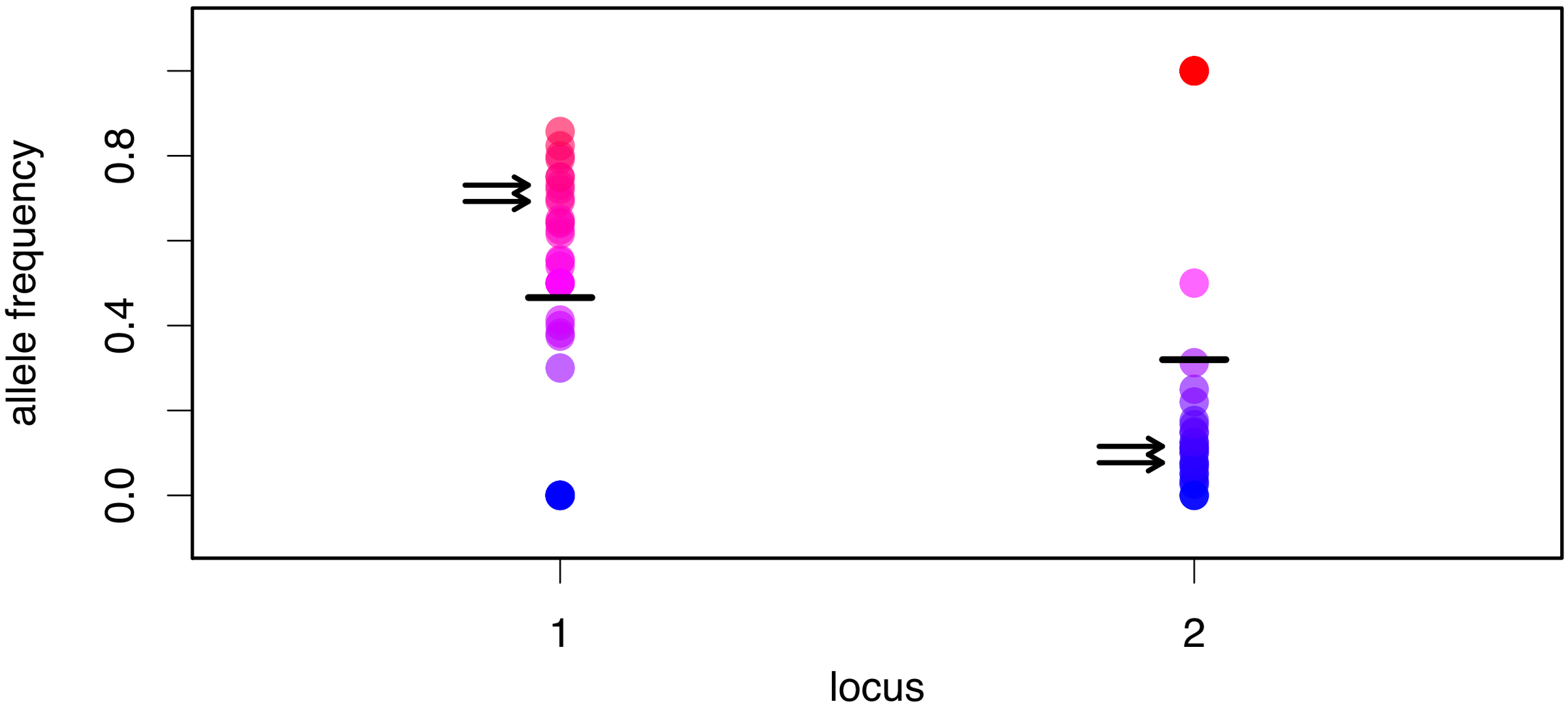



70
65
60
55
50
45
1
allele
frequency
0
−150
−140
−130
−120
−110
−100
−90

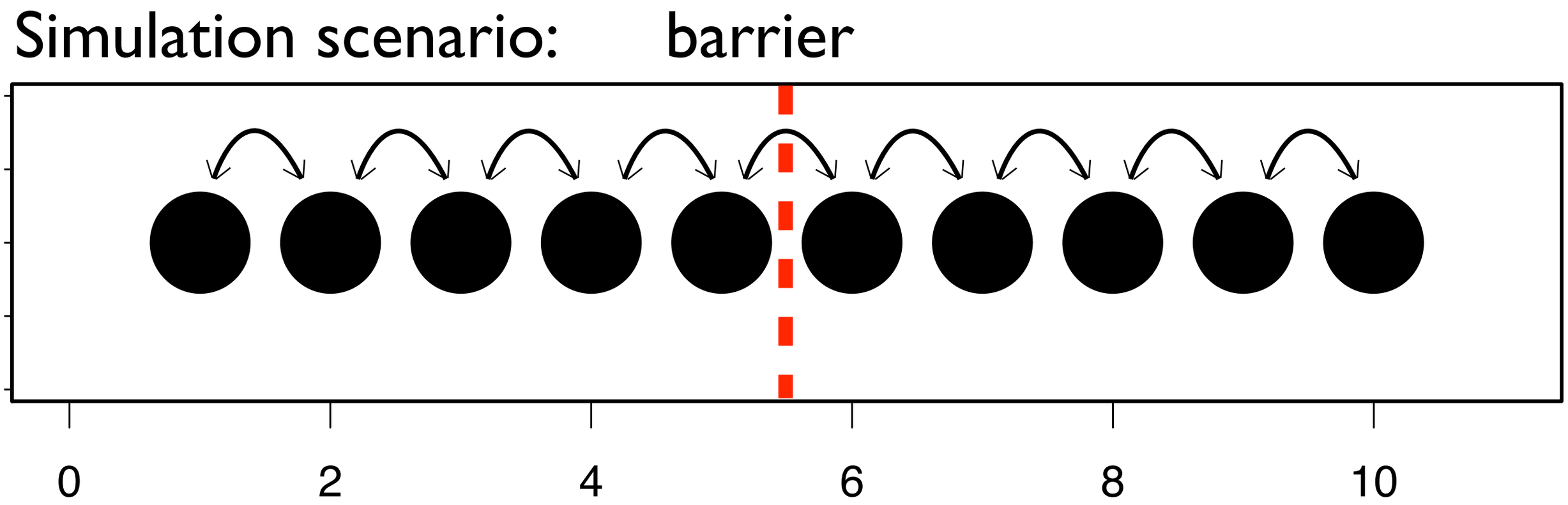
Allele frequencies at two loci





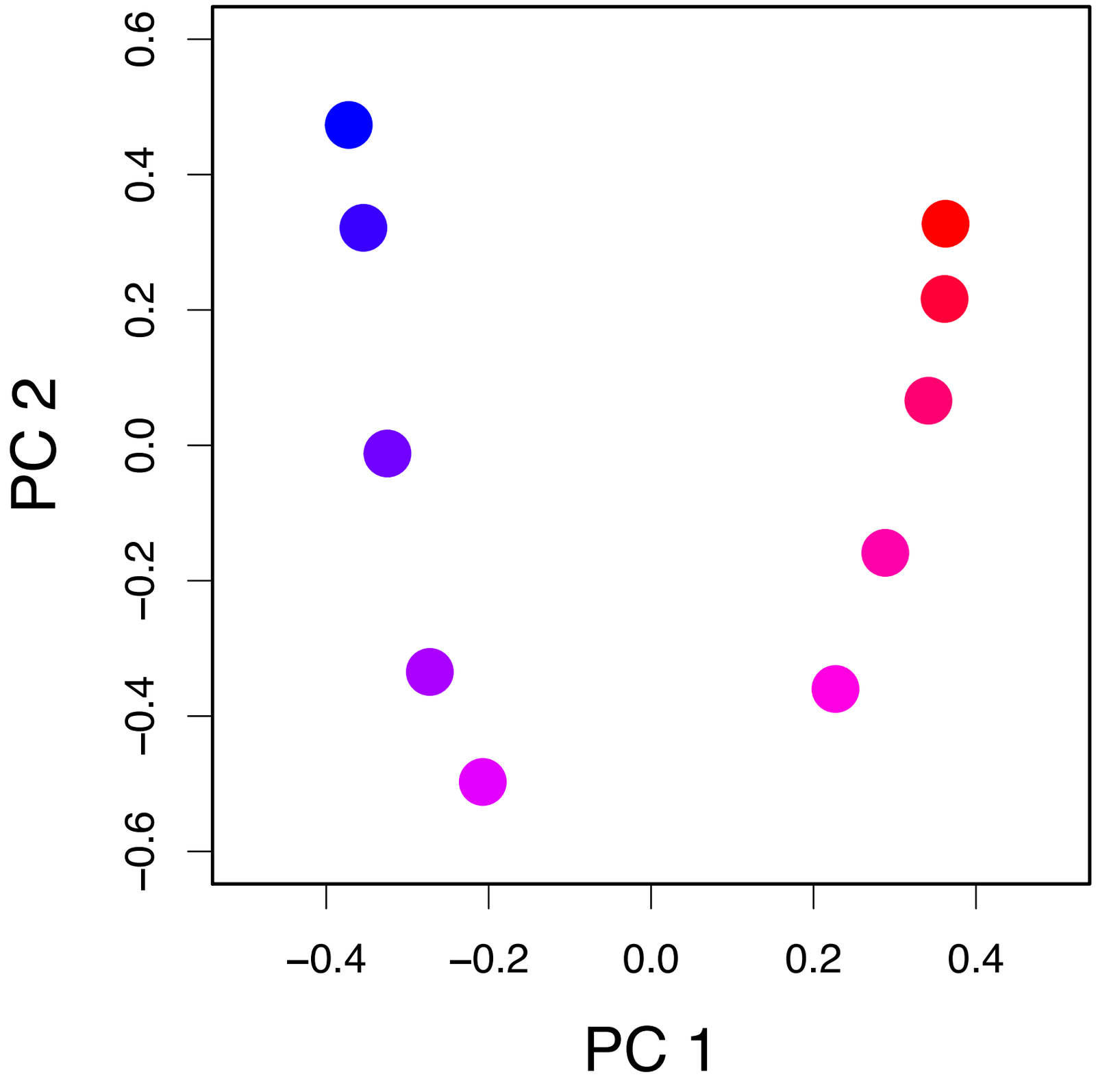


Clines vs. Clusters



Clines vs. Clusters

PCA results

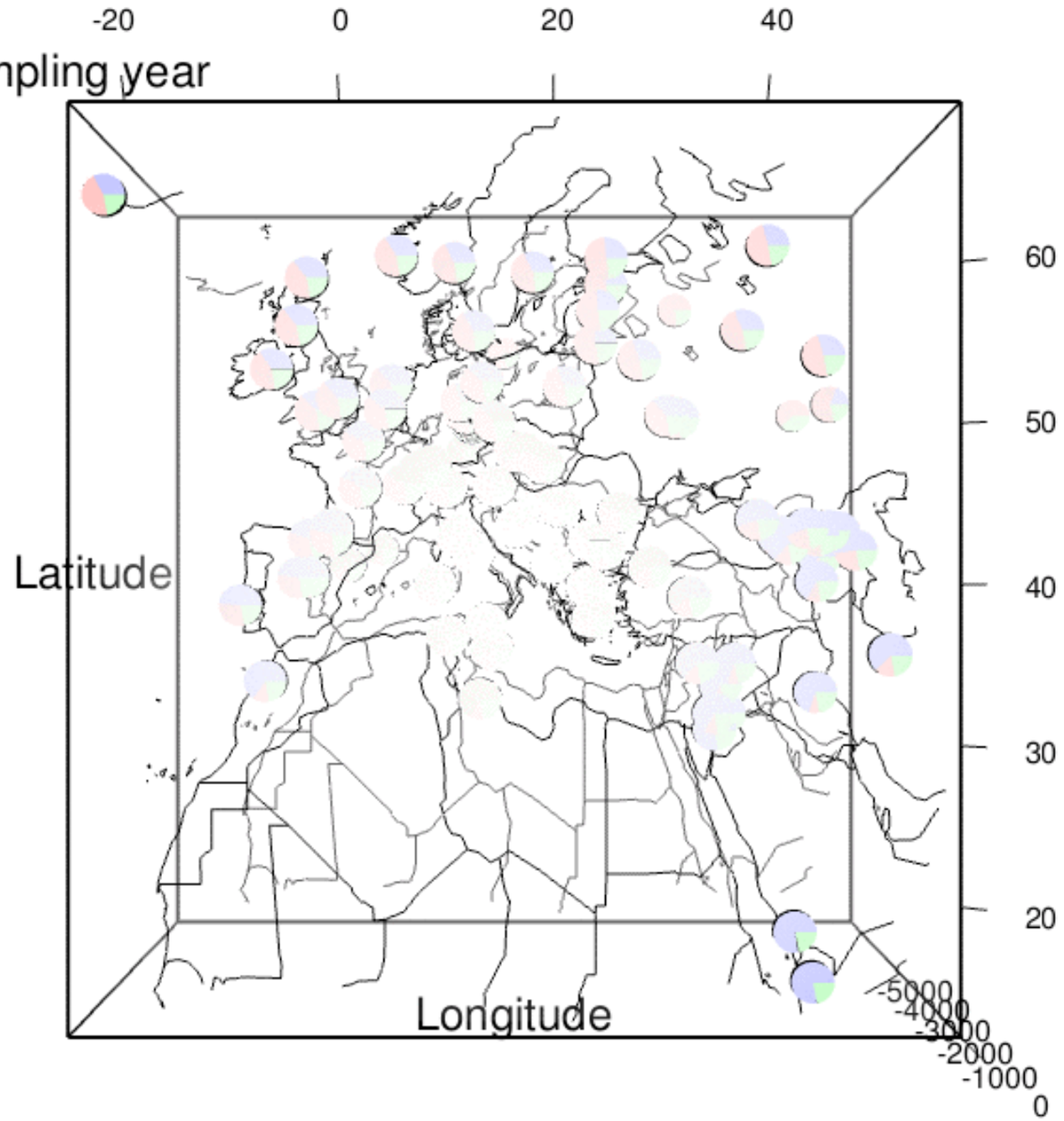


Extensions

Cluster relationships:

-treelike

-spatial

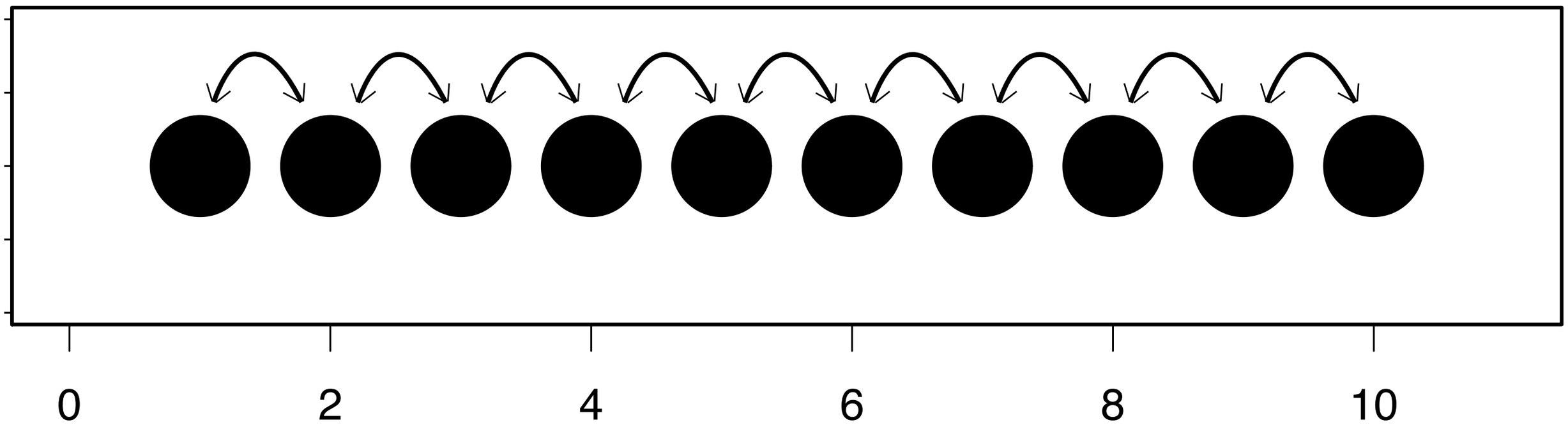


Spatiotemporal covariance functions

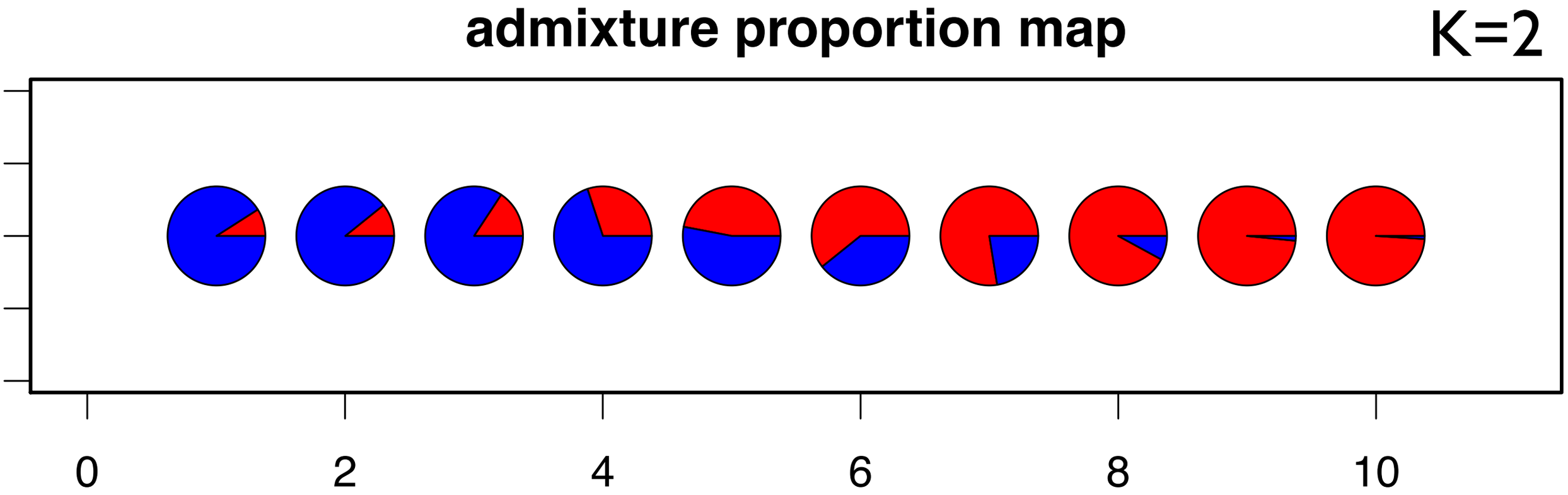
Patterns of population differentiation

Clines vs. Clusters

Simulation scenario: nearest-neighbor migration



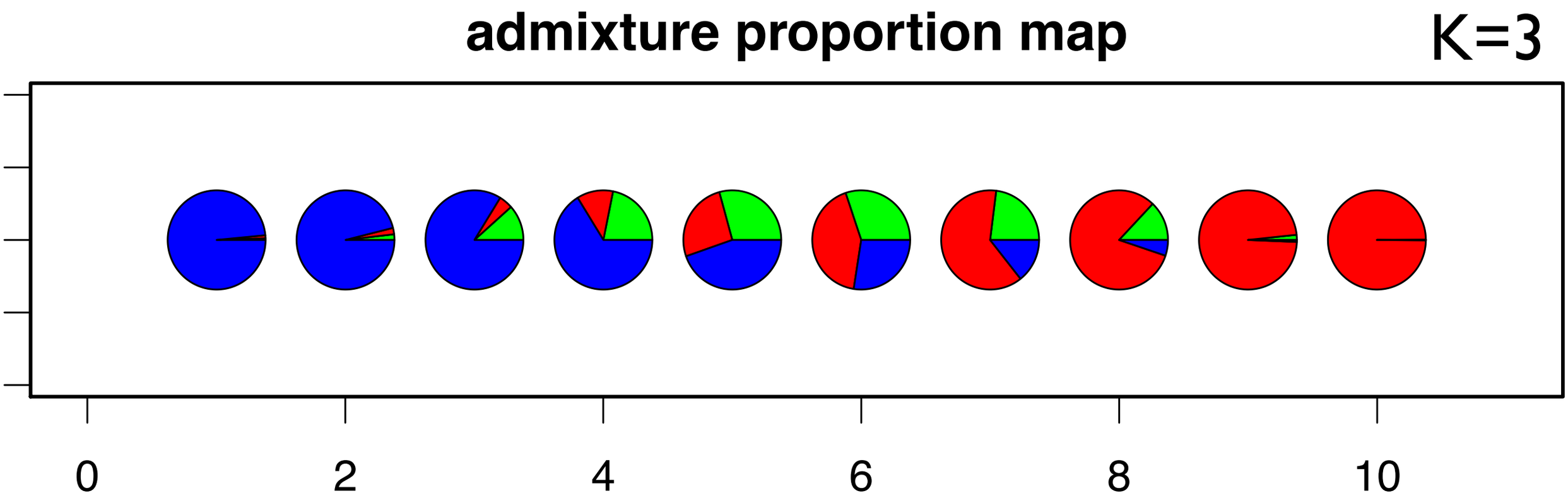
Clines vs. Clusters



STRUCTURE results

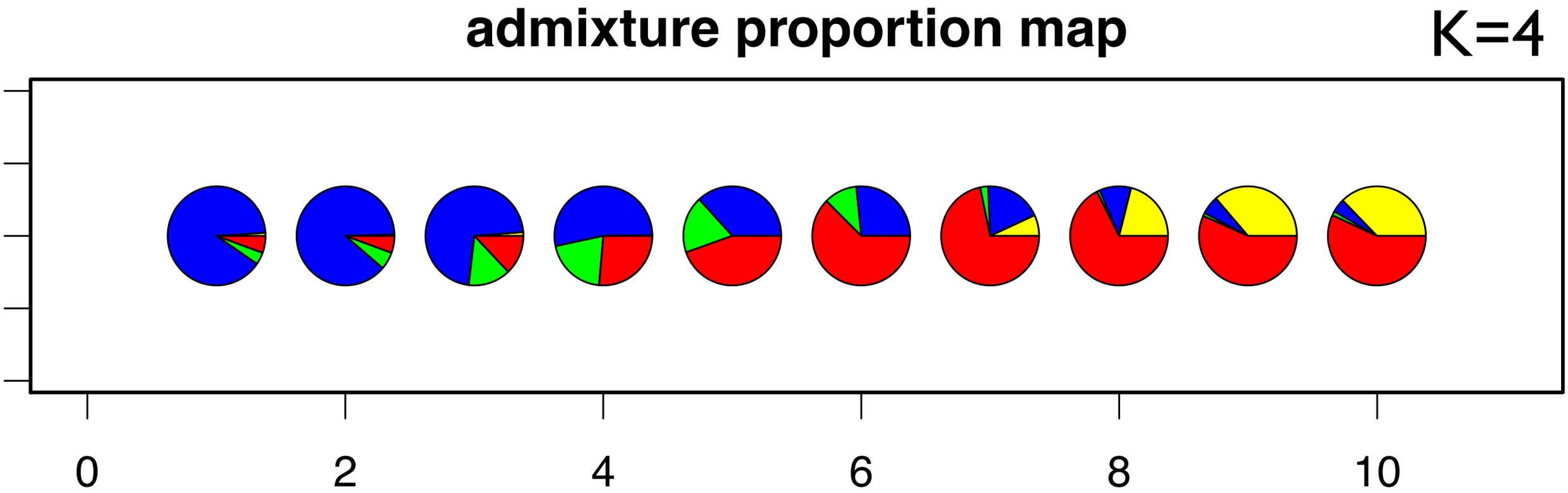
Pritchard *et al.* 2000

Clines vs. Clusters



STRUCTURE results

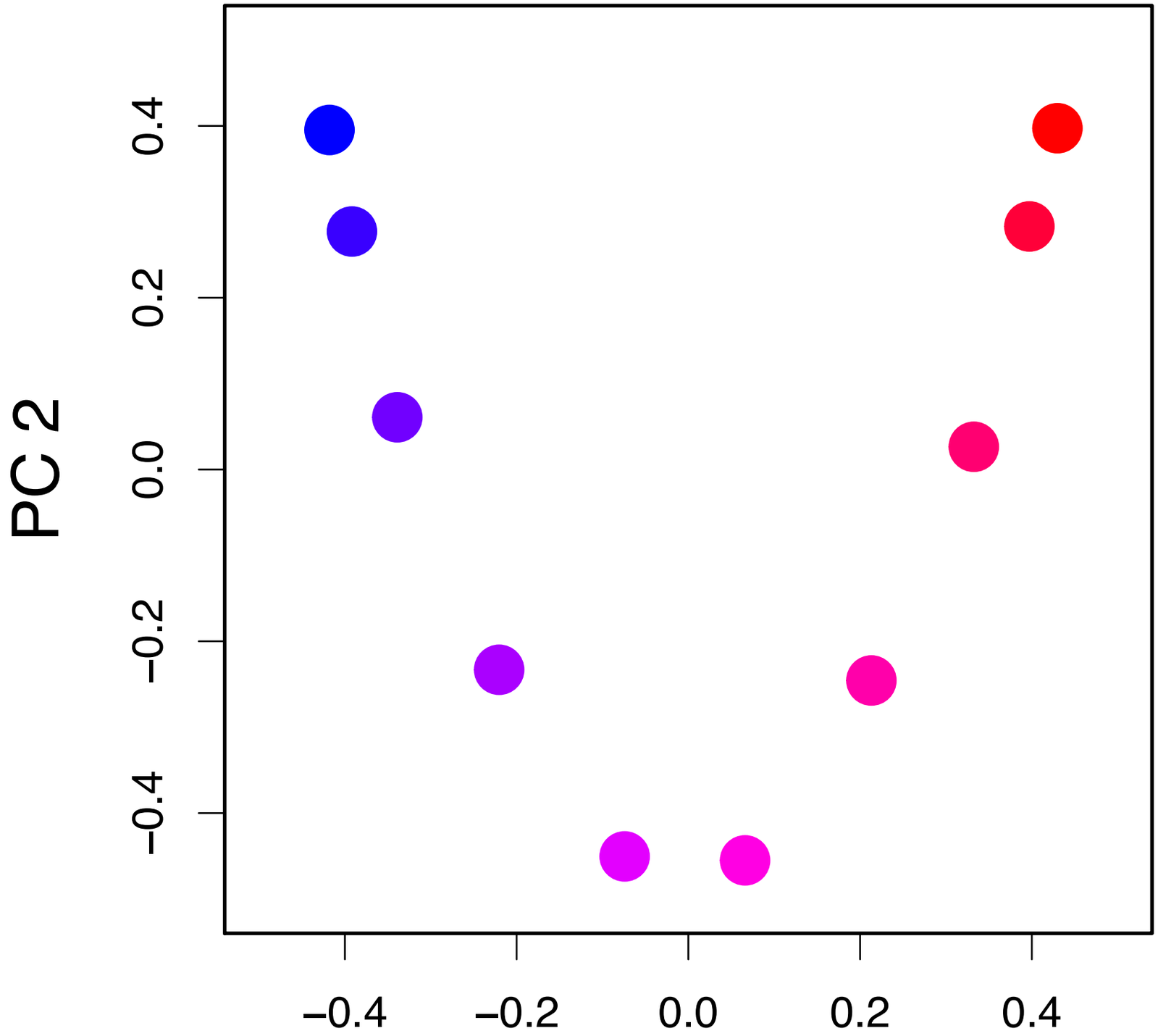
Clines vs. Clusters



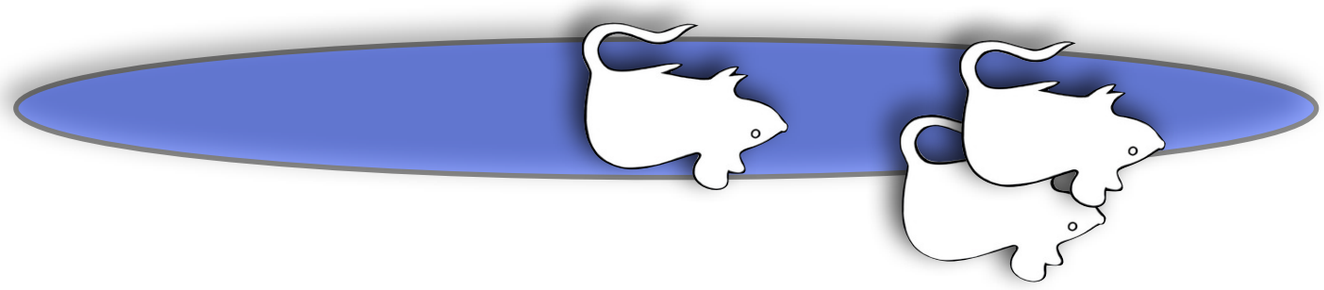
STRUCTURE results

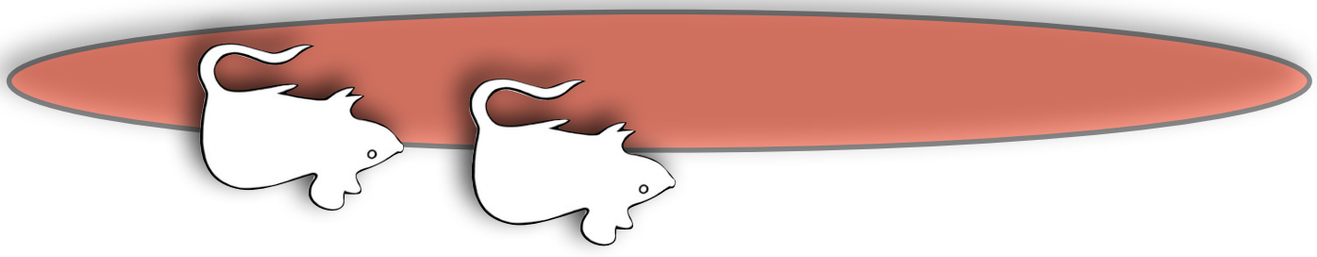
Clines vs. Clusters

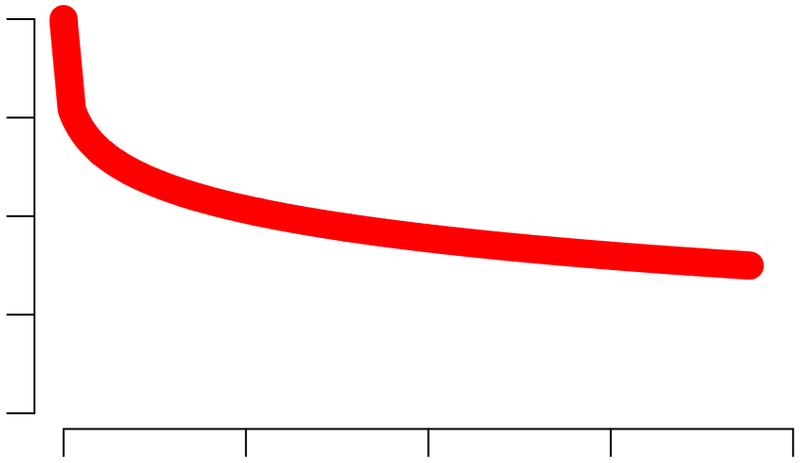
PCA results

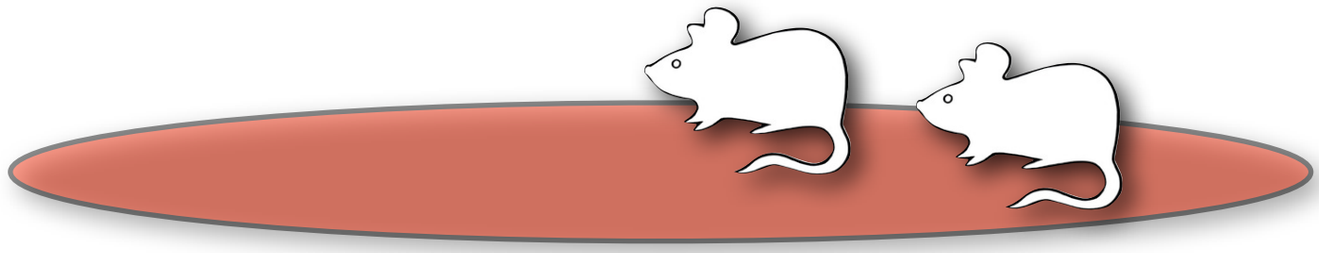


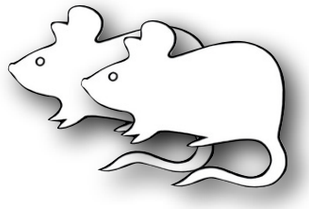
Patterson *et al.* 2006
Novembre and Stephens 2008

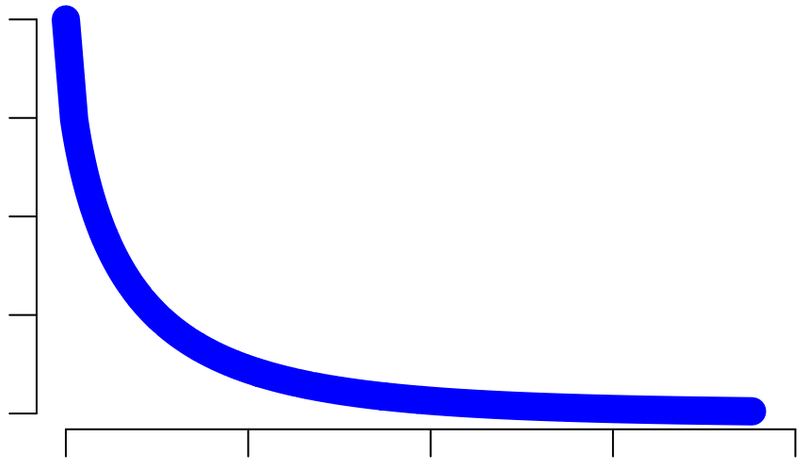










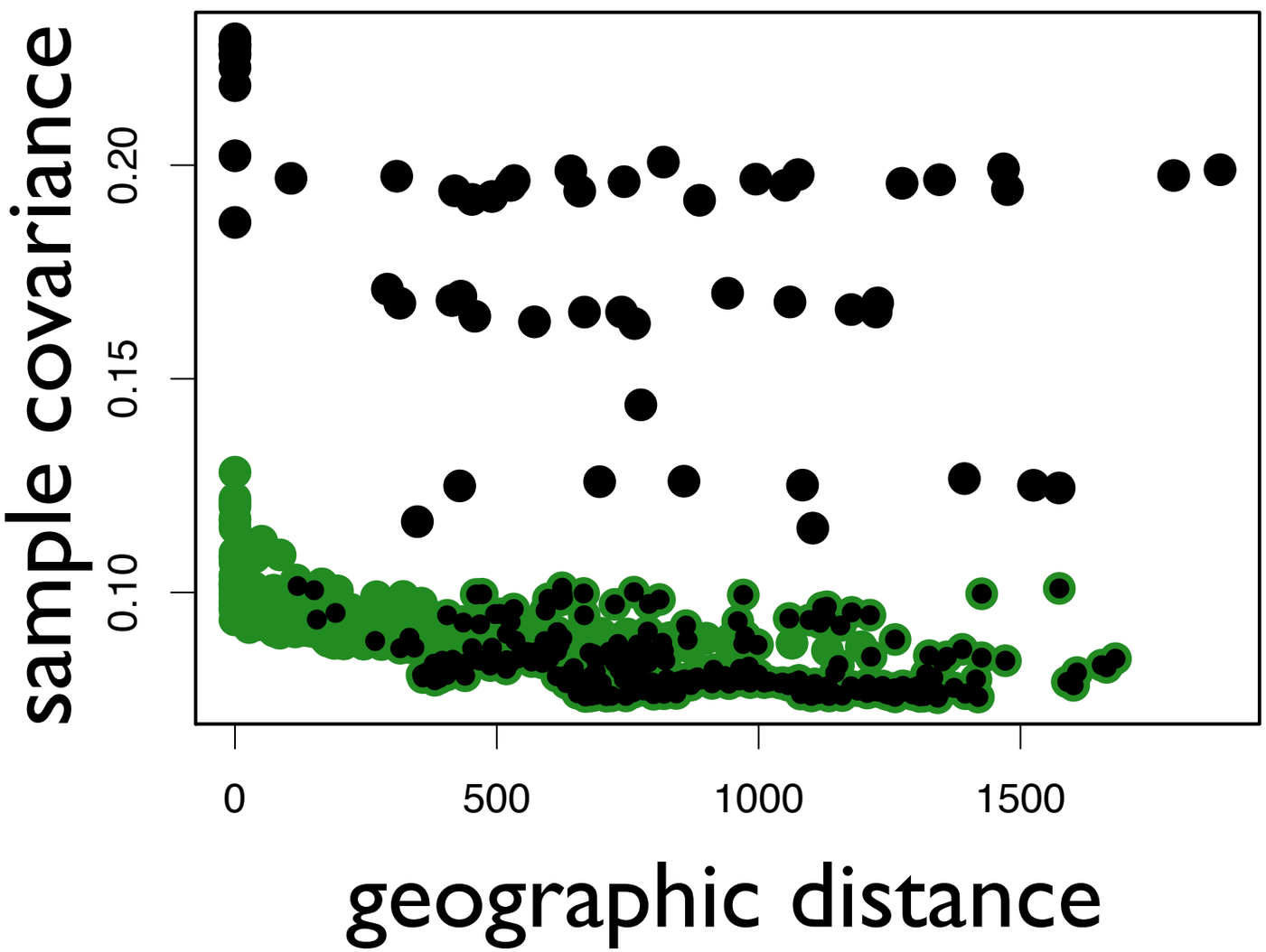


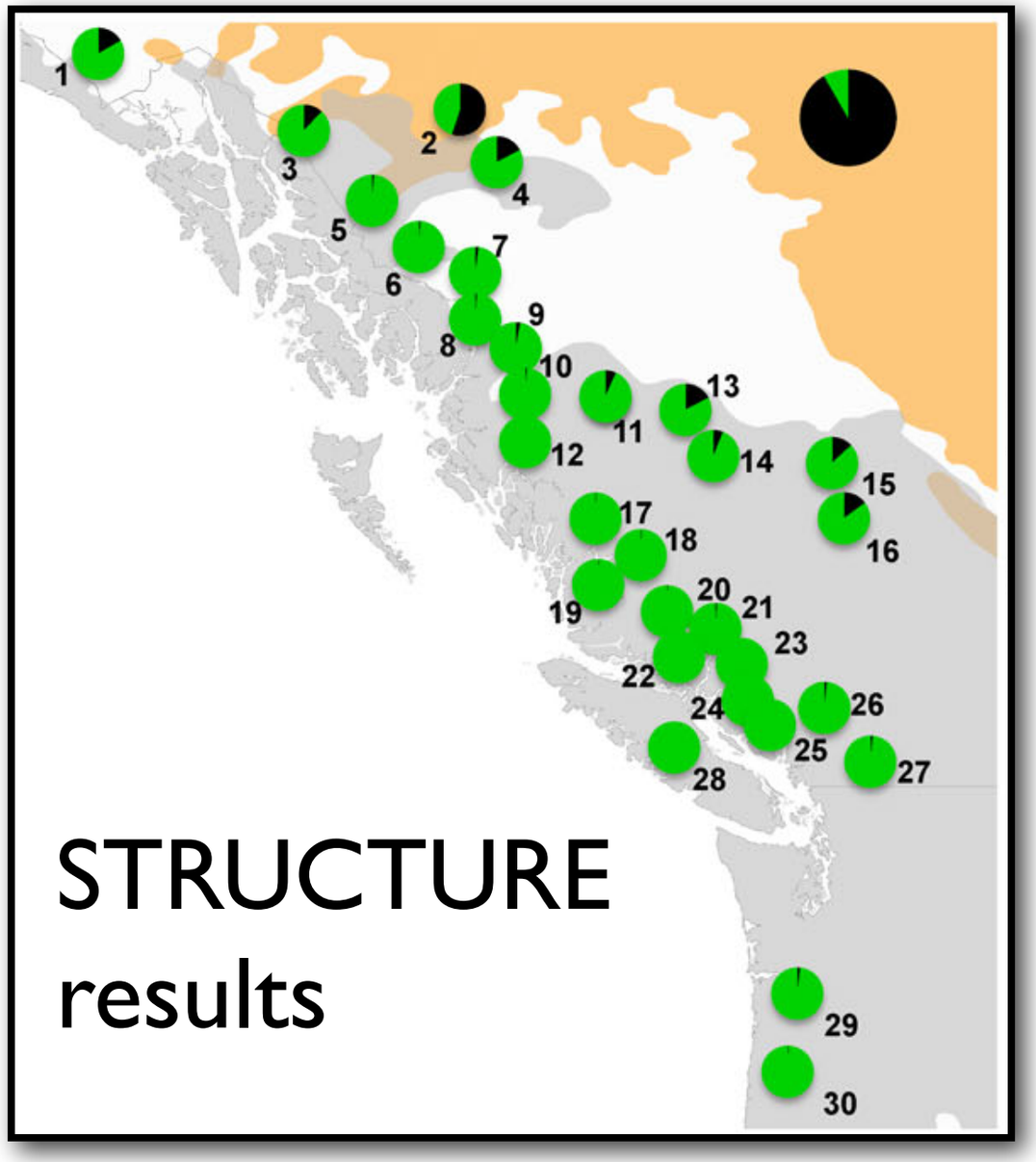
Empirical Illustration:

Populus trichocarpa and *Populus balsamifera*

Sample A

Sample B





Empirical Illustration:

Populus trichocarpa and *Populus balsamifera*