SAMOVA 2.0

Introduction

- SAMOVA 2.0 implements an approach to define groups of populations that are geographically homogeneous and maximally differentiated from each other.
- The method is based on a simulated annealing procedure that aims at maximizing the proportion of total genetic variance due to differences between groups of populations.

Annealing procedure

- It is inspired by the process through which a metal cools and freezes into a crystalline structure with minimum energy (the annealing process).
- The underlying assumption is that, as times passes, we should get closer to a global optimum and be less prone to accept departure from that optimum.

Preliminary steps

• Constructing Voronoi tessellation and the corresponding Delaunay triangulation.



- An arbitrary partition of the *n* populations into *K* groups is chosen at random.
- The genetic barrier(s) between the K groups are identified as edges of Voronoi polygons separating groups of populations.
- The Fct index associated to the K groups is computed.

Simulated annealing steps (1)Geographically homogeneous

(2) without constraint for the geographic composition of the groups

Geographically homogeneous



- The new *Fct* value (noted *Fct**) associated with the new partition is computed.
- The new structure is accepted with probability.

$$p = \begin{cases} 1 & \text{if } F_{\text{CT}}^* \ge F_{\text{CT}} \\ e^{(F_{\text{CT}}^* - F_{\text{CT}}) S^{\text{A}}} & \text{if } F_{\text{CT}}^* < F_{\text{CT}} \end{cases}$$

without constraint for the geographic composition of the groups



The result of the SAMOVA



Input file

- inputfile.geo
- inputfile.arp

Input file

• inputfile.geo

"
"
"
"
"
"
"
"
0"
1"
2"
3"
4"
5"

-2.300	35.100	1
-3.400	37.400	1
-2.600	38.900	1
-3.700	41.600	1
12.100	45.500	1
5.661	44.053	1
3.986	35.468	1
5.529	45.002	1
14.046	42.406	1
15.492	45.279	1
8.279	46.200	1
4.051	34.782	1
19.456	41.371	1
14.953	48.753	1
17.768	46.679	1

Input file

• inputfile.arp

#Arlequin input file written by the simulation program fastsimcoal.exe

[Profile]

Title="A series of simulated samples" NbSamples=25

- GenotypicData=0 GameticPhase=0 RecessiveData=0 DataType=DNA LocusSeparator=NONE MissingData='?'
- [Data]

[[Samples]]

#Number of independent chromosomes: 1

• Inputfile.sar

website

http://cmpg.unibe.ch/software/samova2/#Algorithm

Thanks