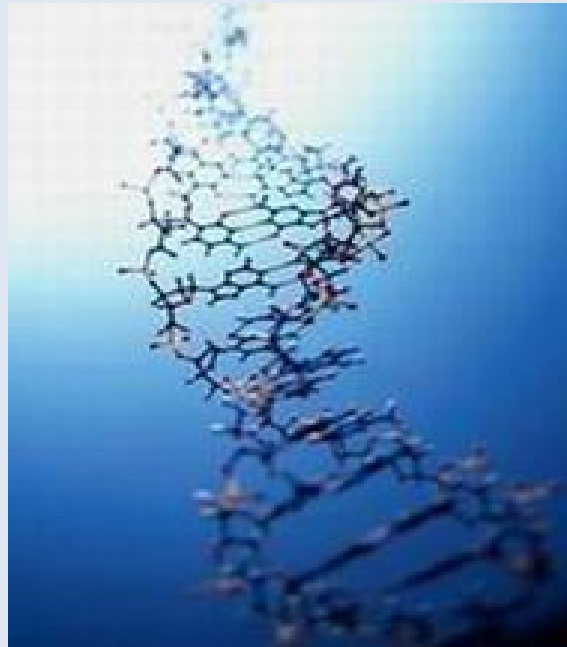


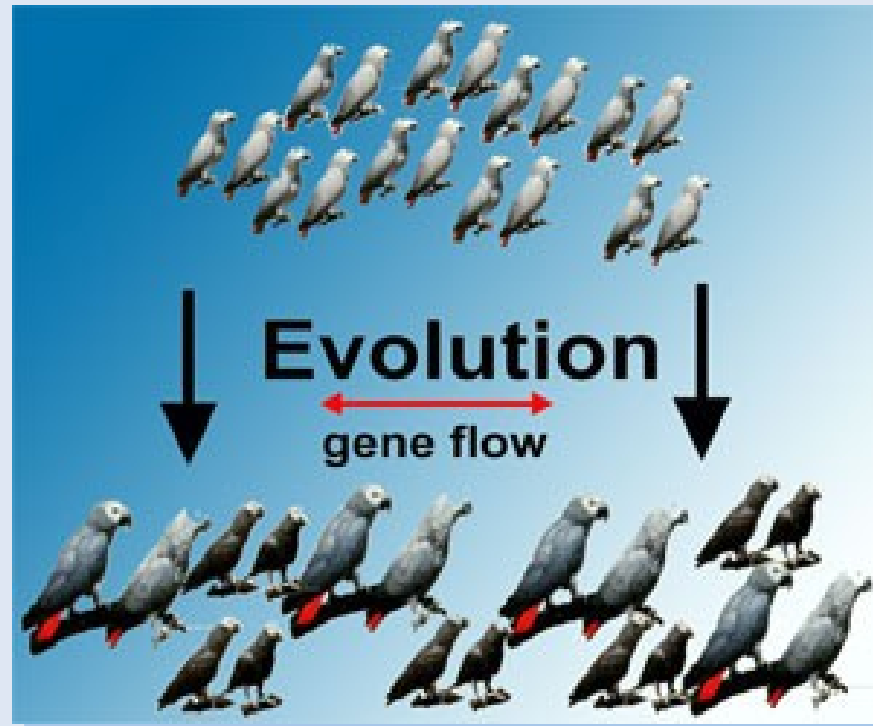
# A simulated annealing approach to define the genetic structure of populations

I. DOPANLOUP, S. SCHNEIDER and L. EXCOFFIER  
Molecular Ecology 2002



# Evoluzione, popolazione e flusso genico

- Evoluzione: processo che porta ad un cambiamento del genotipo degli individui.
- Popolazione: insieme di individui tra i quali esiste un completo incrocio casuale ed un flusso genico aperto.
- Flusso genico: scambio genetico fra popolazioni, implica migrazione e riproduzione. Aumenta il polimorfismo entro popolazione ma riduce l'assortimento genico tra popolazioni.



\*A simulated annealing approach to define the genetic structure of populations\*

# Movimenti migratori e variazione delle frequenze alleliche

- I movimenti migratori sono influenzati dalla distanza geografica ma anche dalla presenza di barriere ecologiche e culturali.
- Si è ampiamente cercato di stabilire l'impatto di queste barriere sul flusso genico e sono state sviluppate molte tecniche per evidenziare la presenza di barriere genetiche.
- Barriera genetica: area in cui il tasso di variazione delle frequenze alleliche è alto.
- Frequenza allelica: frequenza con cui i diversi alleli si manifestano all'interno di una popolazione
- Evidenziare le barriere genetiche presenti potrebbe essere interessante per poter definire la distribuzione geografica di una eventuale popolazione in esame, capirne il suo differenziamento, la sua evoluzione e la sua struttura genetica.
- L'algoritmo qui presentato consente di definire gruppi di popolazione che sono molto differenziati dagli altri (sono separati quindi da barriere genetiche) basandosi esclusivamente su dati genetici e senza l'utilizzo di interpolazione (che può risultare complessa quando i punti non sono regolarmente spazati).

# Spatial Analysis of MOlecular VAriance (SAMOVA). Partizionamento della popolazione in gruppi geneticamente e geograficamente omogenei.

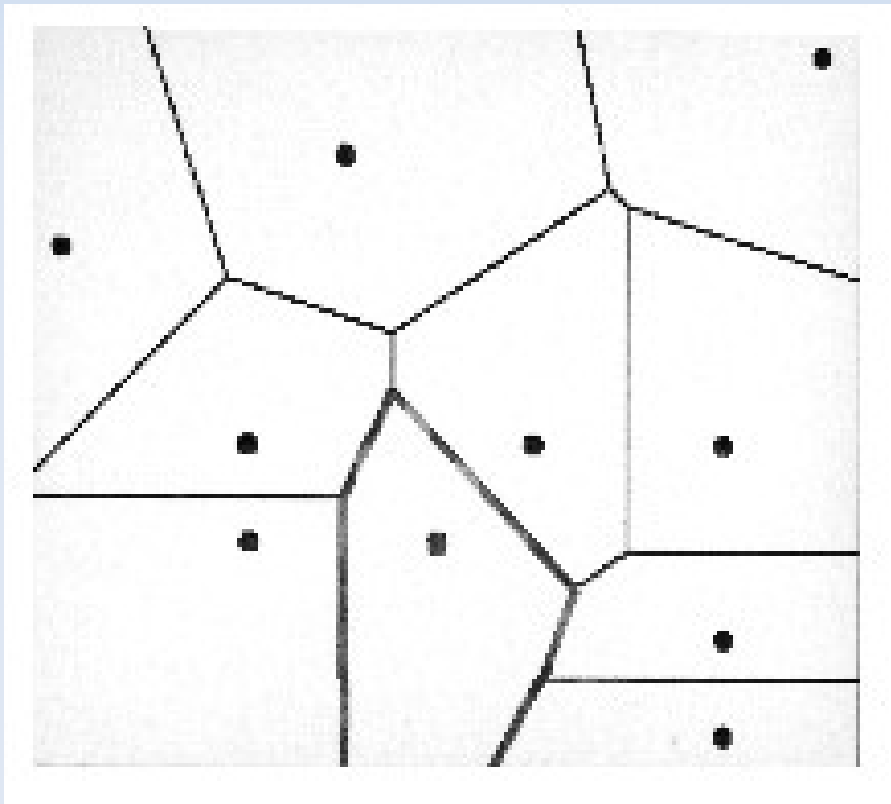
## Step preliminari:

- Costruzione di un set di *poligoni di Voronoi* partendo dalle locazioni geografiche degli  $n$  punti campione.



Dati un insieme  $Q$  di punti il diagramma di Voronoi corrisponde al partizionamento dello spazio in  $n$  poligoni convessi che corrispondono alle varie popolazioni.

- Arbitraria partizione delle  $n$  popolazioni in  $K$  gruppi.



$K$  è un numero impostato dall'utente. Ogni gruppo, eccetto uno, è composto da una singola popolazione e l'ultimo gruppo contiene le popolazioni che non sono state assegnate.

Nel caso mostrato in figura  $K$  è uguale a 2.

La barriera genetica tra i  $K$  gruppi è identificata dai lati del poligono di Voronoi.

Viene calcolato un indice  $F_{ct}$  associato ai  $K$  gruppi, corrisponde alla proporzione della varianza genetica totale tra i gruppi della popolazione.

## Simulated annealing steps:

In fisica si chiama *annealing* quel processo termico atto ad ottenere stati di bassa energia in un solido immerso in un bagno di calore.

- Si aumenta la temperatura del bagno di calore in modo tale che il solido si scioglie (le particelle sono disposte in modo casuale);
- Si decresce lentamente la temperatura del bagno di calore fino a quando le particelle del solido si sistemano in uno stato base (le particelle si dispongono in reticoli strutturati e l'energia è minima).

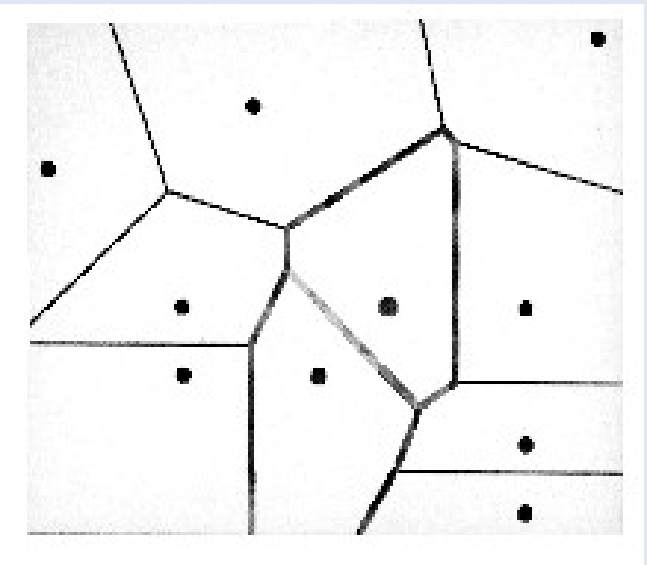
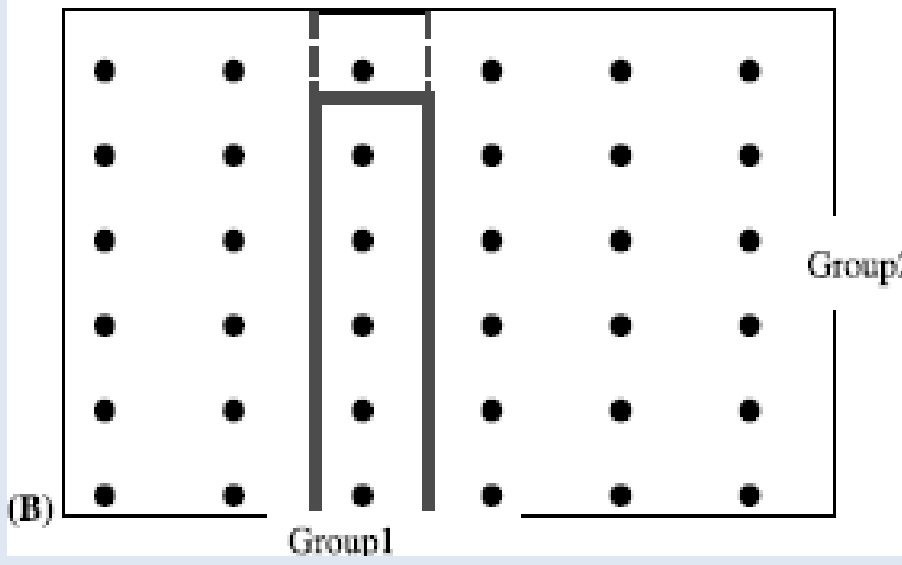
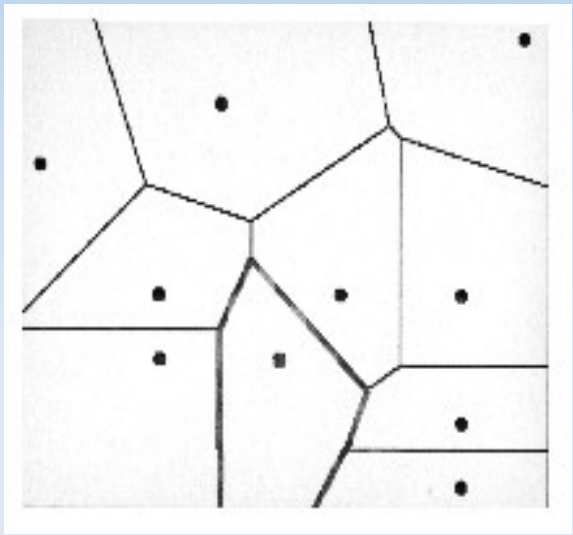
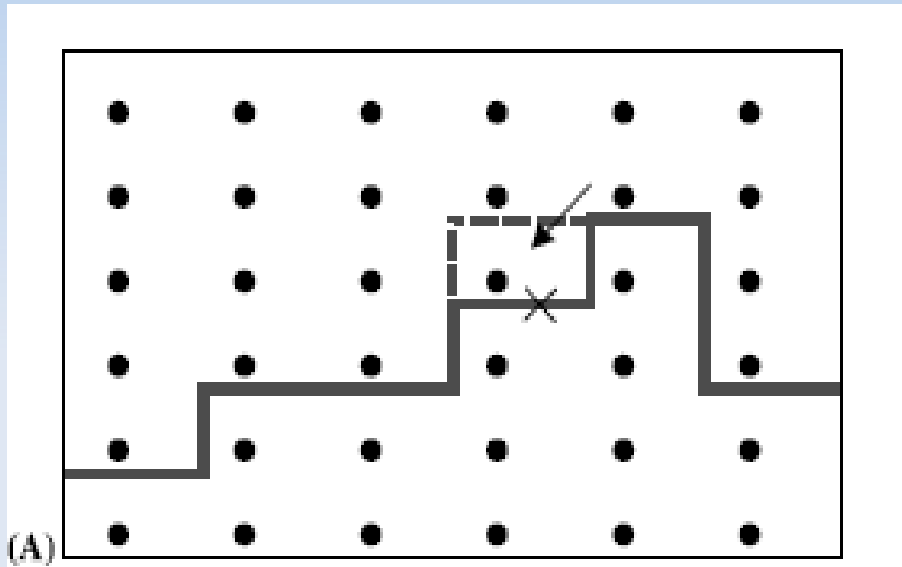
L' algoritmo simula l'evoluzione di un solido immerso in un bagno termico verso l'equilibrio. Tipicamente viene usato come un metodo di ottimizzazione globale, vengono generati punti casuali secondo una certa distribuzione e la loro accettazione o rifiuto viene fatta in funzione della temperatura.

Soluzioni del problema = Stati di un sistema fisico  
Costo di una soluzione = Energia dello stato  
Parametro di controllo = Temperatura

- Si seleziona un lato a caso della barriera.
- Le due popolazioni ai lati vengono selezionate e una andrà a far parte dell'altro gruppo.
- La barriera genetica viene modificata, si aggiornano le popolazioni che compongono i vari gruppi.
- Viene calcolato un nuovo indice  $Fct^*$  associato alla nuova partizione.
- Se  $Fct^*$  è maggiore di  $Fct$  la nuova struttura viene accettata con  $P=1$ .
- Se  $Fct^*$  è minore di  $Fct$  si accetta il cambio con una probabilità  $P = e^{-(Fct^*-Fct) \times Sa}$   
 $S$  è il numero degli step effettuati e  $a$  è una costante = 0,91.  
 Accetti un cambiamento verso un valore peggiore solo quando  $S$  è piccolo.

Questi step vengono ripetuti 10.000 volte partendo ogni volta da partizionamenti diversi.

La configurazione con il valore maggiore di  $Fct$  dopo i 10.000 processi indipendenti è considerata come il miglior raggruppamento della popolazione.



\*A simulated annealing approach to define the genetic structure of populations\*

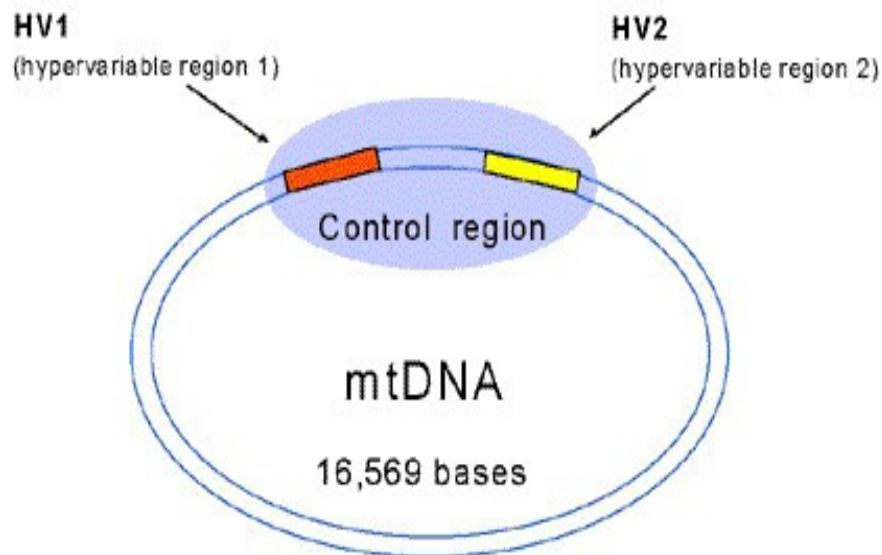


# Applicazione ad una popolazione di caprioli europei

La popolazione di caprioli europei è una razza ampiamente sparsa in tutta Europa, la loro storia evolutiva è stata fortemente influenzata dalle glaciazioni e dai più recenti confinamenti in aree protette dalla caccia.

Per definire gruppi genetici e trovare le più importanti barriere genetiche si è pensato di applicare l'algoritmo SAMOVA.

Si sono analizzate 18 popolazioni di capriolo europeo testate per il DNA mitocondriale nella regione polimorfica HVR1.



Il test sul dna mitocondriale è utile per stabilire la parentela tra 2 o più persone seguendo la linea materna.

# I risultati

(A)



- Sono stati identificati 4 gruppi contraddistinti da 4 simboli diversi.
- Come testimonianza dal corretto funzionamento dell'algoritmo è stata identificata una forte associazione tra popolazioni della Liguria e popolazioni balcaniche. La popolazione ligure è stata reintrodotta da una fonte balcanica

# La barriera genetica



•A simulated annealing approach to define the genetic structure of populations•

# Conclusioni

- Il metodo SAMOVA è molto efficace per definire gruppi le cui popolazioni non sono geograficamente adiacenti.
- Essendo basato su Simulated Annealing permette di trovare il massimo globale e evita di rimanere intrappolati in ottimi locali.
- Può essere applicato su qualsiasi tipo di popolazione, non richiede equilibrio di Hardy-Weinberg e nemmeno la presenza di linkage tra i loci analizzati.
- Si identifica una barriera genetica sulla base di soli dati sperimentali.
- E' stato sviluppato un software (SAMOVA 1.0) che implementa questo algoritmo.
- Richiede in input le coordinate geografiche delle popolazioni campione e i rispettivi dati genetici.  
(<http://web.unife.it/progetti/genetica/Isabelle/samova.html>)



\*A simulated annealing approach to define the genetic structure of populations\*