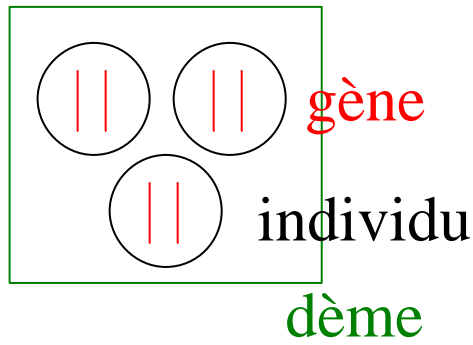


# **Structuration génétique : quelques idées utiles et leur application aux microsatellites**

F. Rousset

Génétique de l'adaptation  
Institut des Sciences de l'Evolution  
Montpellier

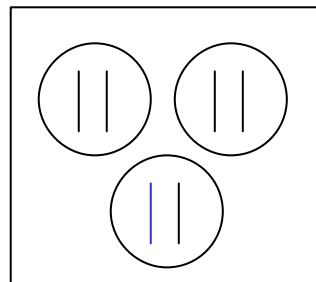
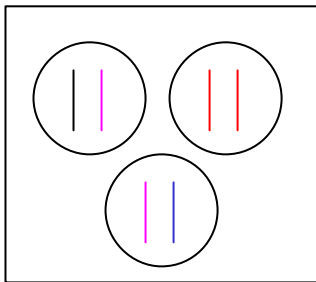
- Discussion de deux applications des microsatellites en génétique des populations:
  - estimation des « flux géniques »
  - estimation de l'« apparentement »



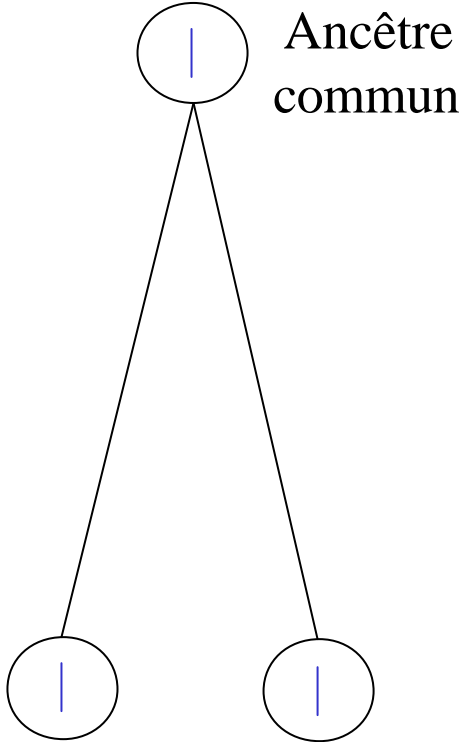
## Ces fameuses $F$ -« statistiques »

$$F \equiv \frac{Q_w - Q_b}{1 - Q_b}$$

$w$ [ithin]	$b$ [etween]	$F$
intra individu ( $Q_0$ )	individus intra dème ( $Q_1$ )	$F_{IS} \equiv \frac{Q_0 - Q_1}{1 - Q_1}$
entre individus intra demes ( $Q_1$ )	inter dème ( $Q_2$ )	$F_{ST} \equiv \frac{Q_1 - Q_2}{1 - Q_2}$

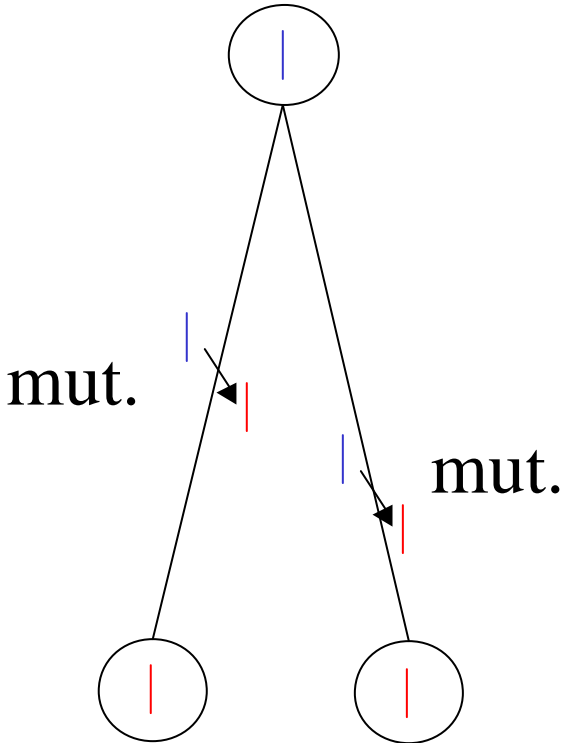


# L'identité dans les modèles avec mutation



Identité par descendance

$$\dot{Q}$$

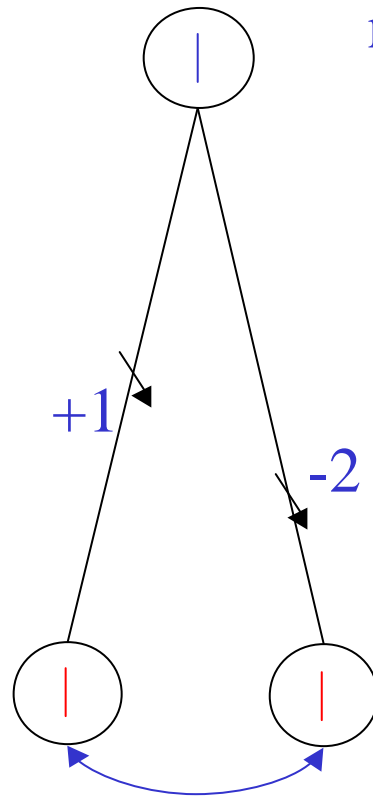


Identité par état

$$Q$$

Taille

n

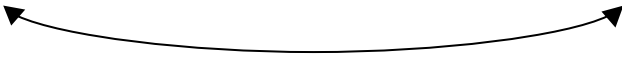


Différence<sup>2</sup>=3<sup>2</sup>

S

Identité ( $Q$ )

Taille<sup>2</sup> ( $S$ )

$$F_{ST} \equiv \frac{Q_w - Q_b}{1 - Q_b} \qquad r_{ST} \equiv \frac{S_b - S_w}{S_b}$$

$$1 - Q_j \leftrightarrow S_j$$

(variante inutilement compliquée: « Rst »)

$$\frac{F_j}{1 - F_j} = \frac{Q_w - Q_j}{1 - Q_w}$$

Fst « IAM based »... « Rst » « SMM based »?

## **Il faut raisonner autrement**

(1) définition des estimateurs classiques

(2) paramètres : faibles taux de mutation

(3) paramètres : fort taux de mutation

(4) variance des estimateurs?

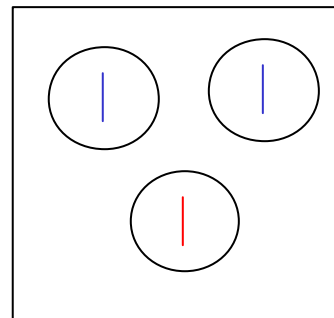
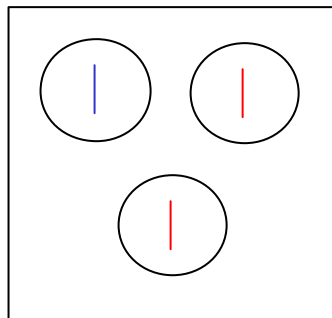
## Remarque sur l'estimation des F

Les rapports tels que  $\frac{Q_w - Q_b}{1 - Q_b}$

peuvent s'estimer par  $\frac{\hat{Q}_w - \hat{Q}_b}{1 - \hat{Q}_b}$

où les  $\hat{Q}$  sont les fréquences de paires de gènes identiques dans l'échantillon

**Exemple:**



$$\hat{Q}_w = \frac{1}{3}, \quad \hat{Q}_b = \frac{4}{9}, \quad \hat{F} = -\frac{1}{5}$$

Cockerham...

**Cette méthode ne dépend de rien**

□<sub>ST</sub> : même argument

Fst « IAM based »... « Rst » « SMM based »?

## **Il faut raisonner autrement**

(1) définition des estimateurs classiques

(2) paramètres : faibles taux de mutation

(3) paramètres : fort taux de mutation

(4) variance des estimateurs?

Fst « IAM based »... « Rst » « SMM based »?

## **Il faut raisonner autrement**

(1) définition des estimateurs classiques

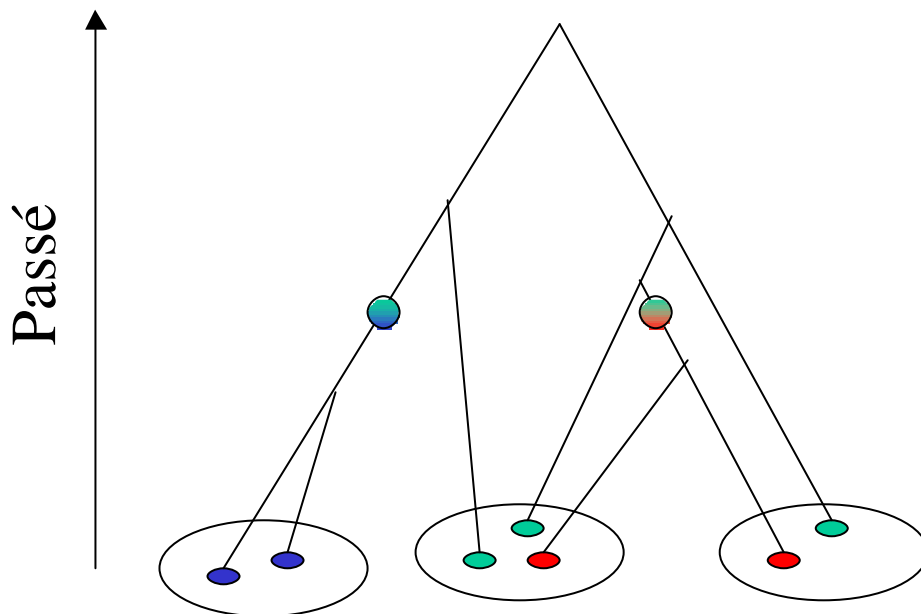
(2) paramètres : faibles taux de mutation

(3) paramètres : fort taux de mutation

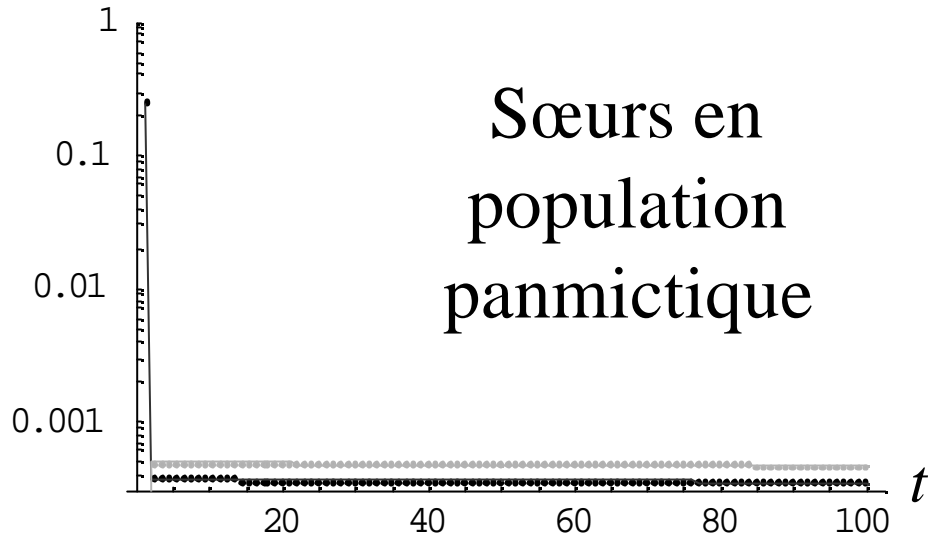
(4) variance des estimateurs?

## la coalescence

mode(s) de raisonnement pour déterminer la probabilité d'un échantillon en fonction de paramètres démographiques passés, des modèles mutationnels, etc

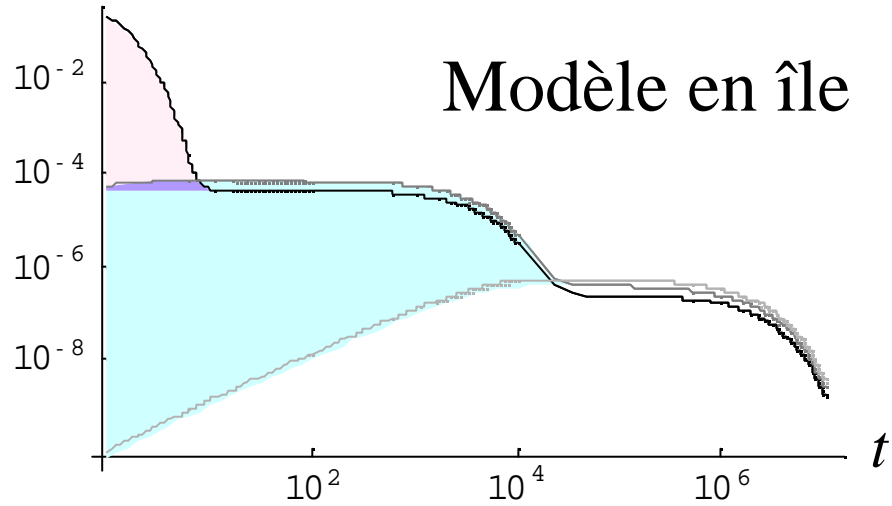


*Pr coalescence at  $t$*



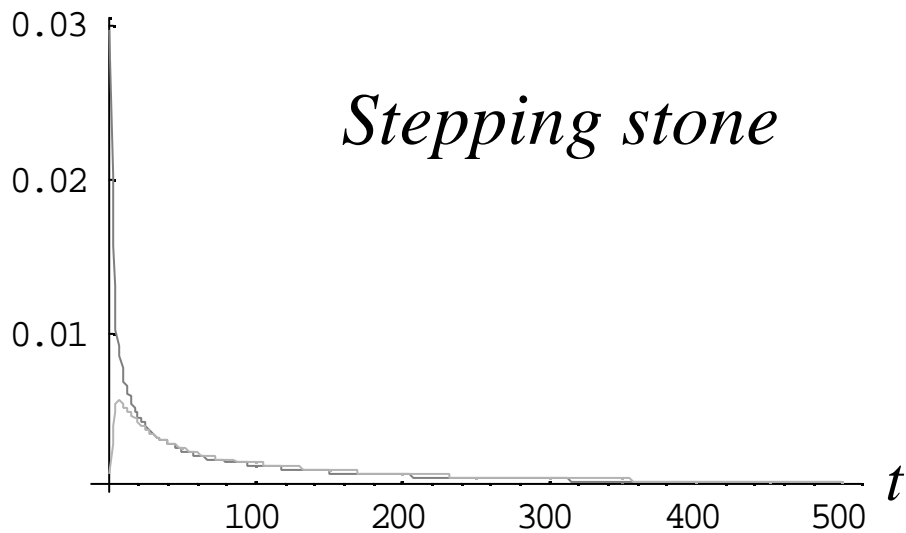
Sœurs en  
population  
panmictique

*Pr coalescence at  $t$*



Modèle en île

*Pr coalescence at  $t$*



*Stepping stone*

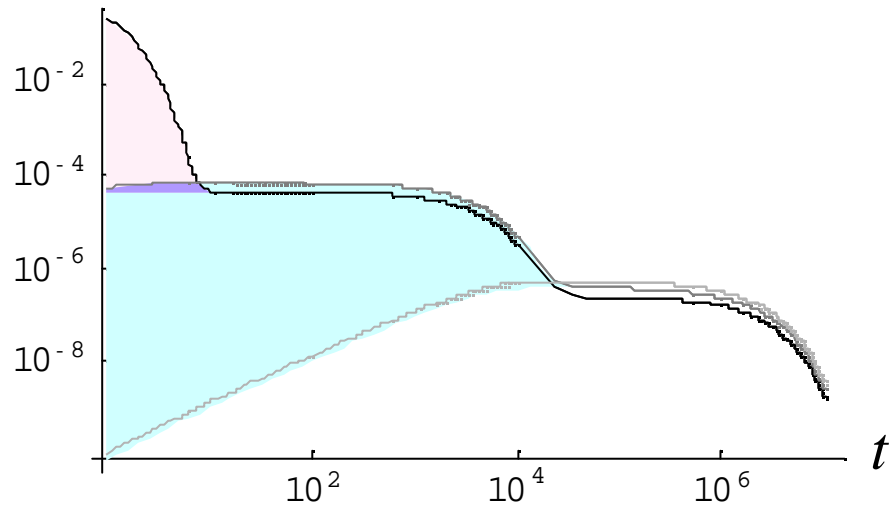
# Modèle en île

$$F_{ST} \approx 1/(1+4Nm)$$

Slatkin 1995 :

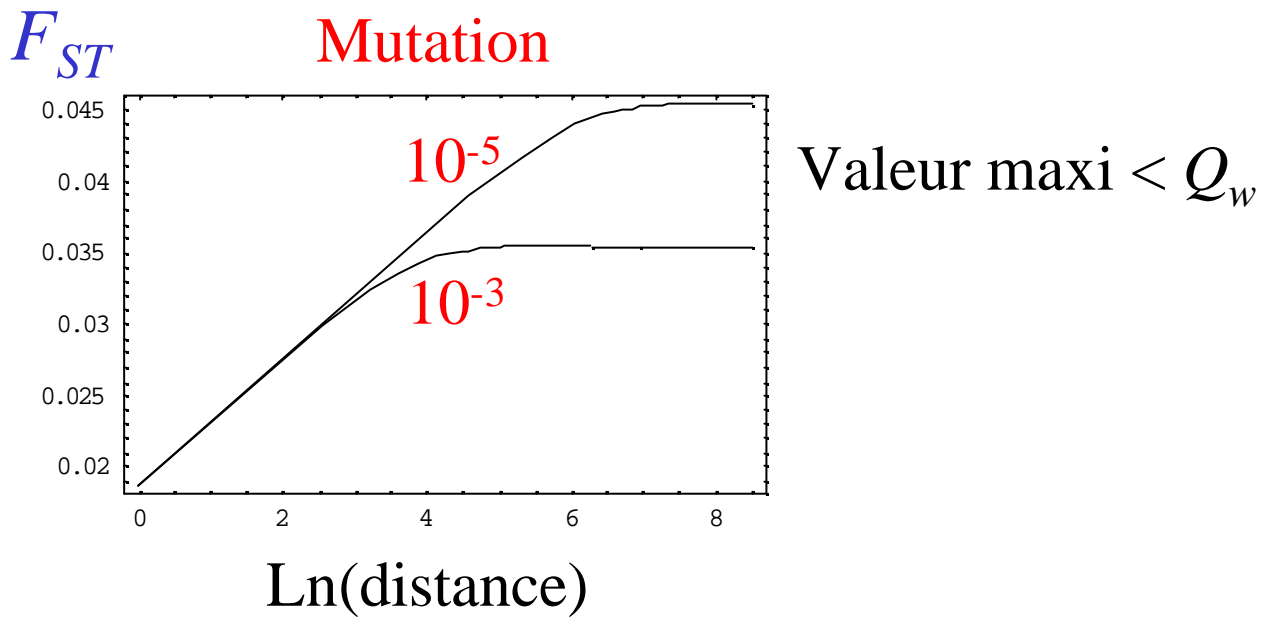
(1) ça marche pas

Pr coalescence at  $t$

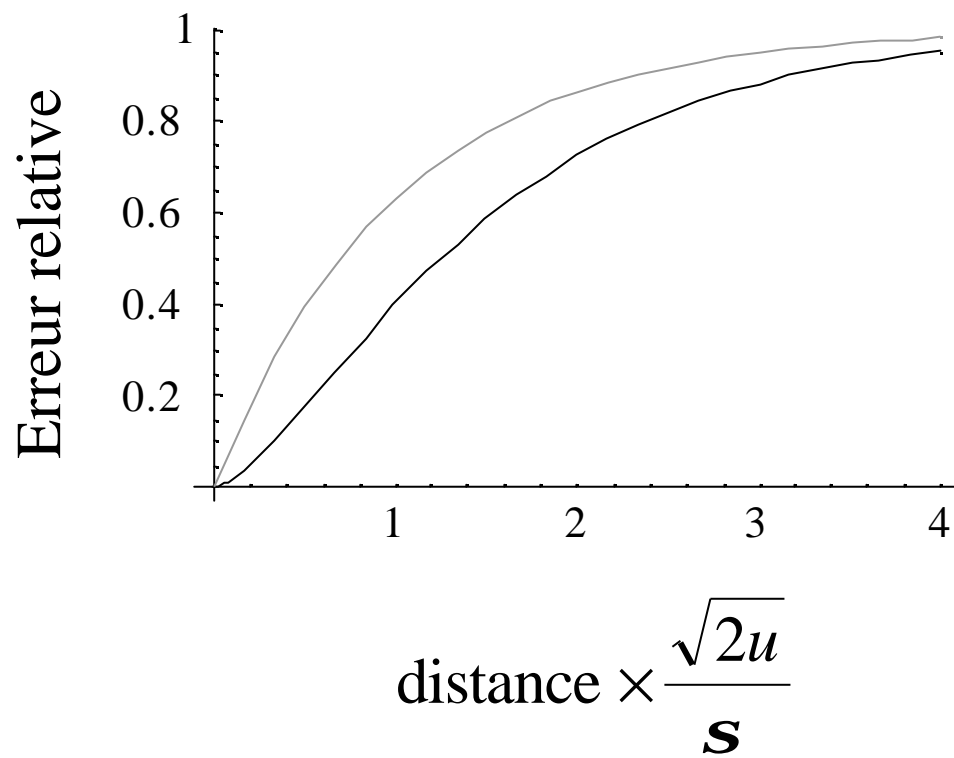


(2) « Rst » c'est mieux

Le modèle le plus approprié pour discuter ces questions...

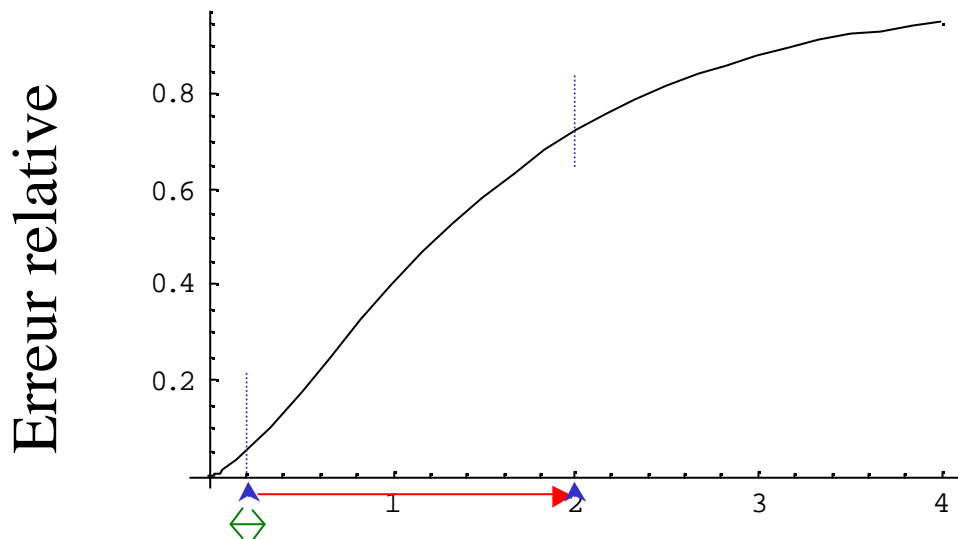


## L'erreur sur des estimations de paramètres démographiques



Autre vision du même problème:

Mutation  $\ll 0 \gg$   $\star 10^{-5}$   $\star 10^{-3}$



$$\text{distance} \times \frac{\sqrt{2u}}{\mathbf{S}}$$

Modèle mutationnel...

“estimates of **gene flow** based on **Fst** are unlikely to be very reliable. However, these estimates are **likely to be correct within a few orders of magnitude.**”  
(Whitlock & McCauley 1999).

Quelques manques dans leur biblio...

$$D_S^2 \frac{\text{génétique}}{\text{démographique}}$$

- 0.5 (populations humaines, Nouvelle Guinée);
- 1.5 (gastéropode littoral);
- 1.8 (Rongeur Nord-Américain)



Les modèles doivent préciser comment estimer des paramètres démographiques à partir de données génétiques, mais aussi **quand** cela peut être plus efficace

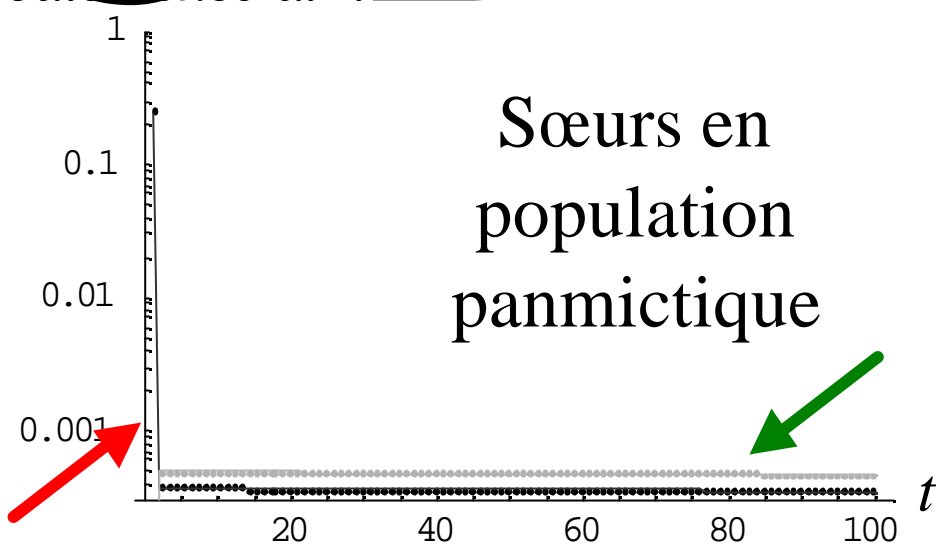
Démographie, mutation... le message :  
**plus c'est loin, plus c'est nul**

<u>Distance maxi <math>r</math></u>	<u><math>r/\diamond</math></u>
14 km	8.5
1600 m	69
1016 m	12.9

Nordborg et Donnelly 1997:

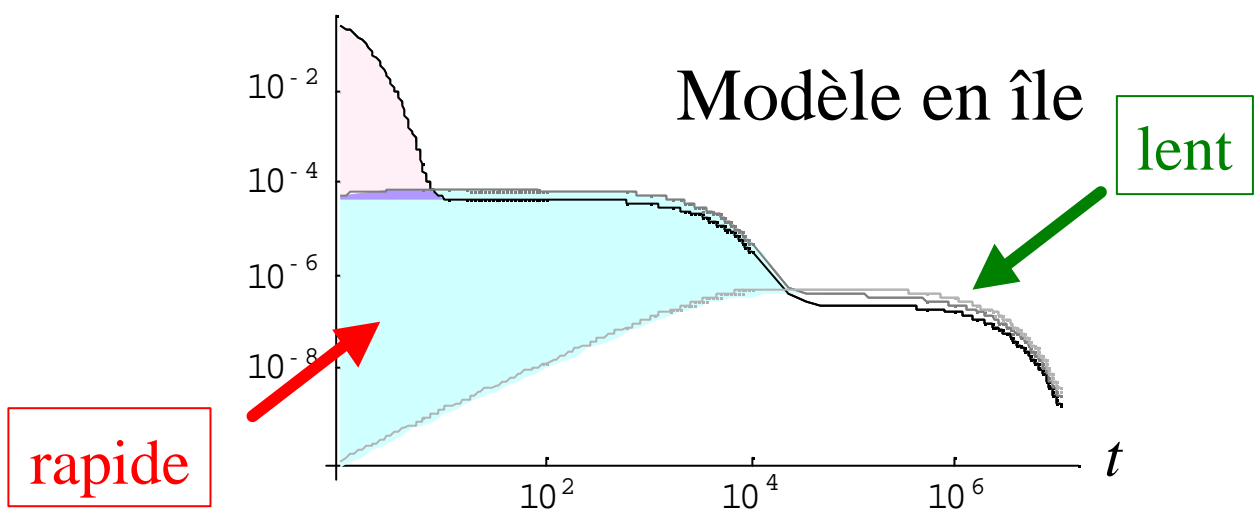
- distinction « *fast* » et « *slow* » *process*
- Calcul probabilité d'un échantillon distinguant ces deux termes
  - données de séquence
  - autofécondation, panmixie...
- analyses de structuration devraient être fondées sur processus « rapides »

*Pr coalescence at  $t$*



Sœurs en population panmictique

*Pr coalescence at  $t$*

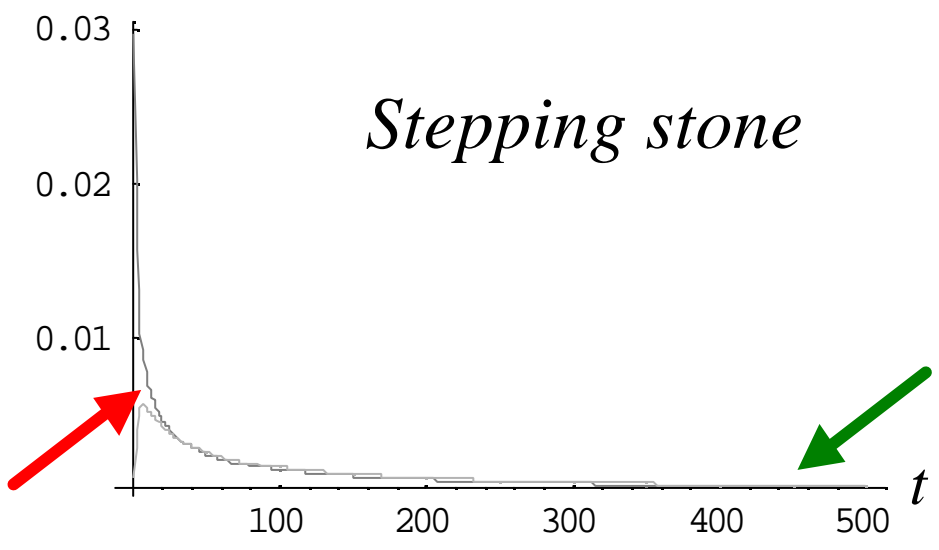


Modèle en île

rapide

lent

*Pr coalescence at  $t$*



Stepping stone

Fst « IAM based »... « Rst » « SMM based »?

## **Il faut raisonner autrement**

(1) définition des estimateurs classiques

(2) paramètres : faibles taux de mutation

(3) paramètres : fort taux de mutation

(4) variance des estimateurs?

Et pour « Rst » ( $\square_{ST}$  !)  
?

Idée générale : idem

Fst vs  $\square_{ST}$  ?

« Fst based estimates are always better than Rst when sample sizes are moderate or small ... and the number of loci scored is low ... These are the conditions under which many real investigations are carried out and therefore we conclude that in many cases the most conservative approach is to use Fst » (Gaggiotti et al. 1999)

Les forts taux de mutation ont **un avantage évident** : l'augmentation de la « puissance » de différentes analyses statistiques.

- **Analyses impossible sans polymorphisme**

- Peu d'études plus quantitatives de l'effet de la diversité sur la puissance des analyses statistiques:

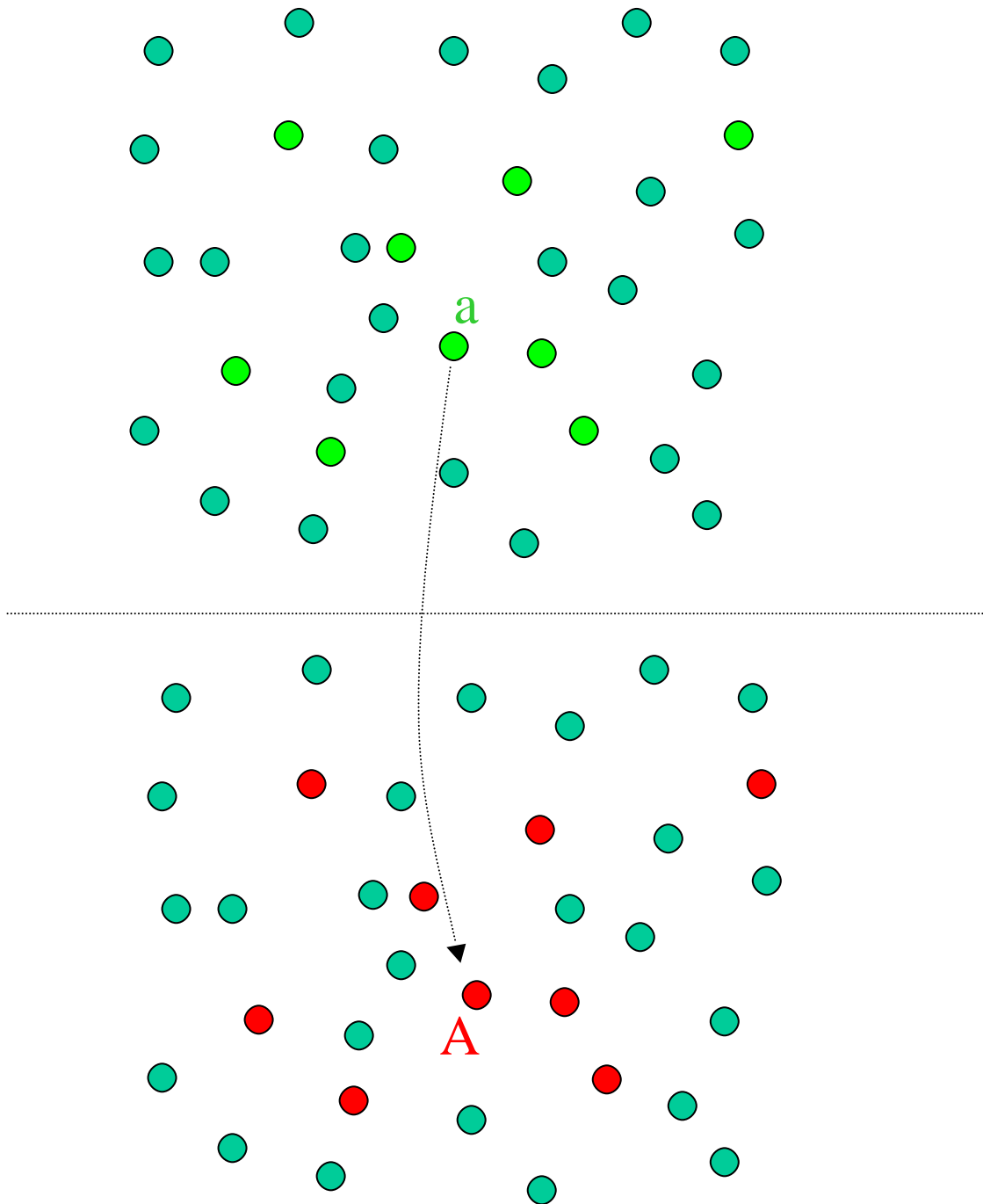
- tests (Robertson et Hill 1984, Rousset et Raymond 1995, Goudet et al. 1996)

- **Mais: peu de puissance si tous les gènes sont différents!**

Niveau **intermédiaire** de polymorphisme **optimal**, mal cerné.

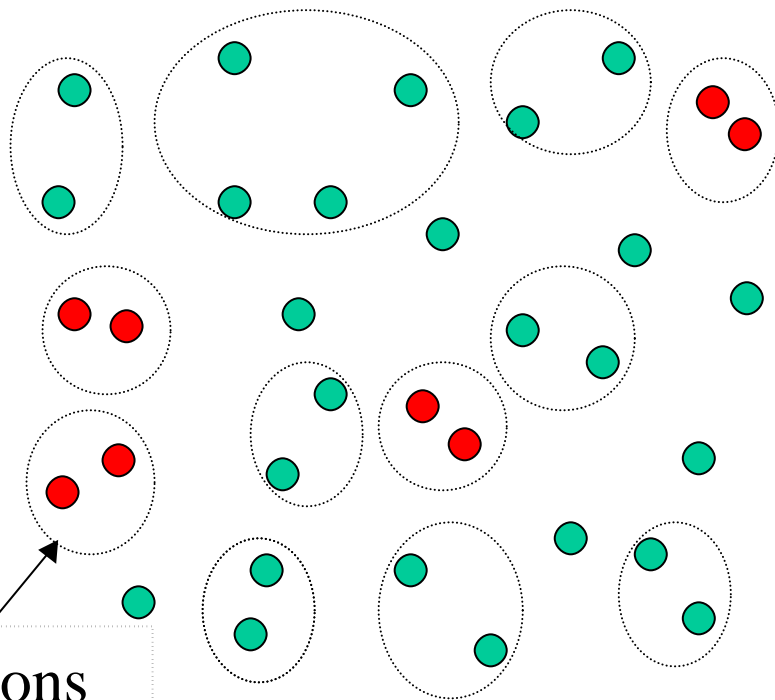
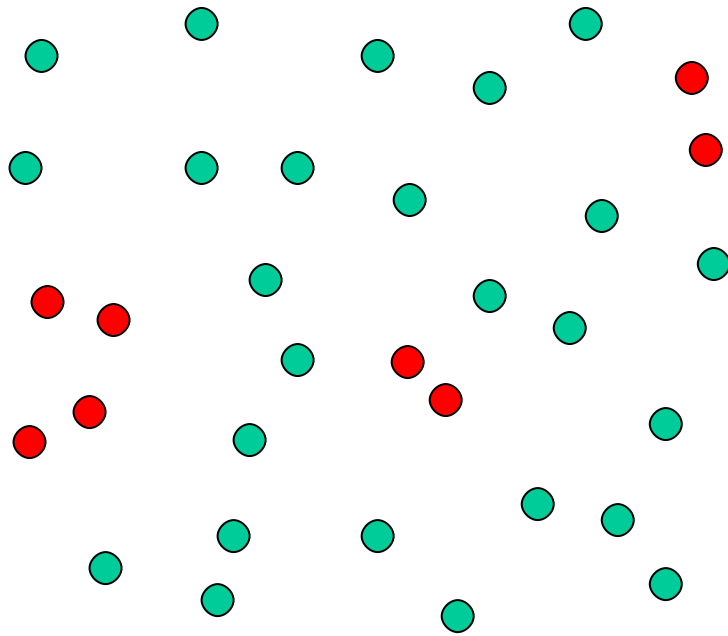
# “Apparemment”

Le problème de l’ « altruisme »

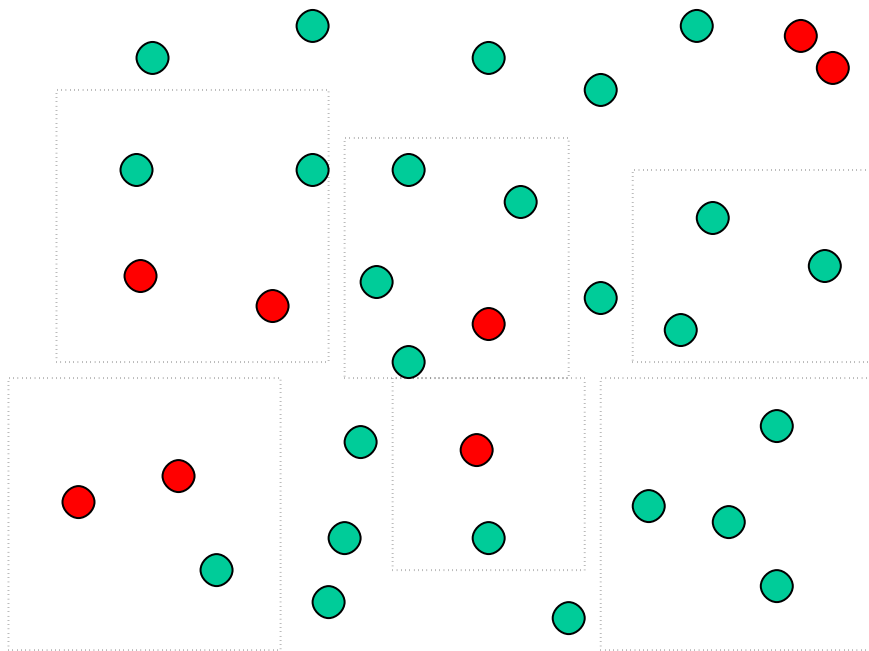


Si **A** diminue le nombre de descendants mais augmente celui de ses voisins ↴ éliminé

# Sélection de parentèle= apparemment+interactions locales



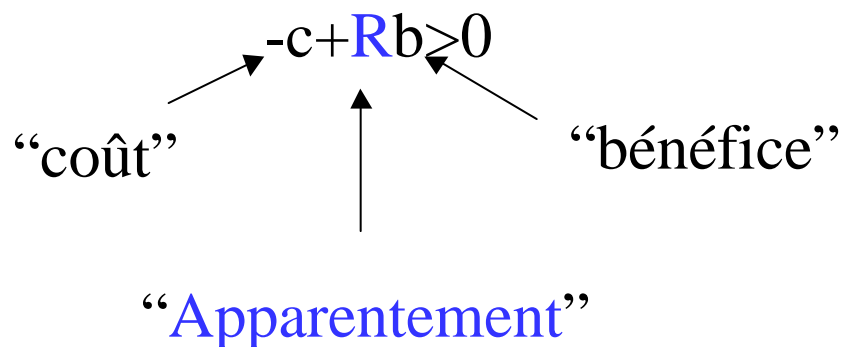
Interactions  
intra-familles



Interactions spatialement localisées

+ migration spatialement localisée

Ces questions peuvent en principe se résoudre sous la forme de la règle de Hamilton,  $-c+Rb>0$



- Estimer l' « apparement » ?...
- ... en population subdivisées ??

Un mesure de « fitness » utile est

$$- \sum_{j=0,1,\dots} \frac{f_j w_j}{f_j z_j} \left( \lim_{u \rightarrow 0} \frac{1 - Q_j}{1 - Q_0^D} \right) = 0$$

↑

Probabilité d'identité des gamètes/juvéniles qui viennent en compétition pour un dème

Mesures d'apparentement à estimer:

$$\frac{1 - Q_j}{1 - Q_0^D} = 1 + \frac{F_j}{1 - F_j}$$

## Conclusions: conseils pratiques?

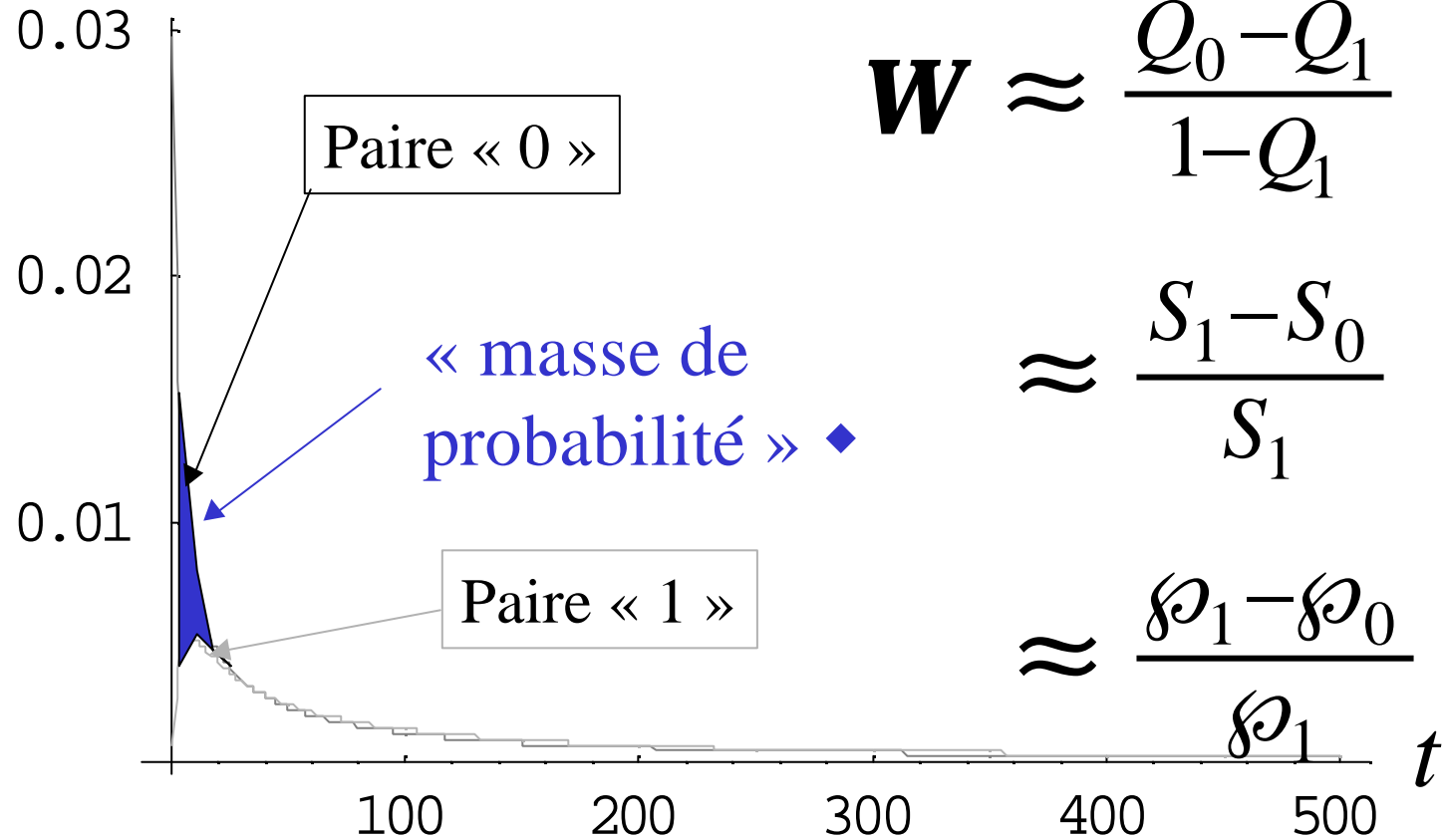
1) Utilisez des microsatellites polymorphes mais pas trop ( $Q_0=0.3$  ??)

2) Oubliez les Rst

3) Une analyse quantitative de structuration spatiale n'est raisonnable qu'à petite échelle ( $< 100$  ♦ ?), en raison du fort taux de mutation des microsatellites si ce n'est pour d'autres raisons.

... attention aux allèles nuls...

*Pr coalescence at t*



Pour  $\phi = n$  'importe quelle mesure de divergence entre paire de gènes

	Fst	« Rst »
biais <b>A</b> des paramètres génétiques par rapport aux paramètres démographiques ( <i>low-mutation limit</i> )	variable	variable
biais <b>B</b> des estimateurs	faible	large
variance <b>Var</b> des estimateurs	« moyenne »	large
<hr/>		
MSE totale <b><math>A^2 + B^2 + Var</math></b> /paramètre démographique	variable	variable+
MSE si <b>A</b> négligeable	« moyenne »	large !