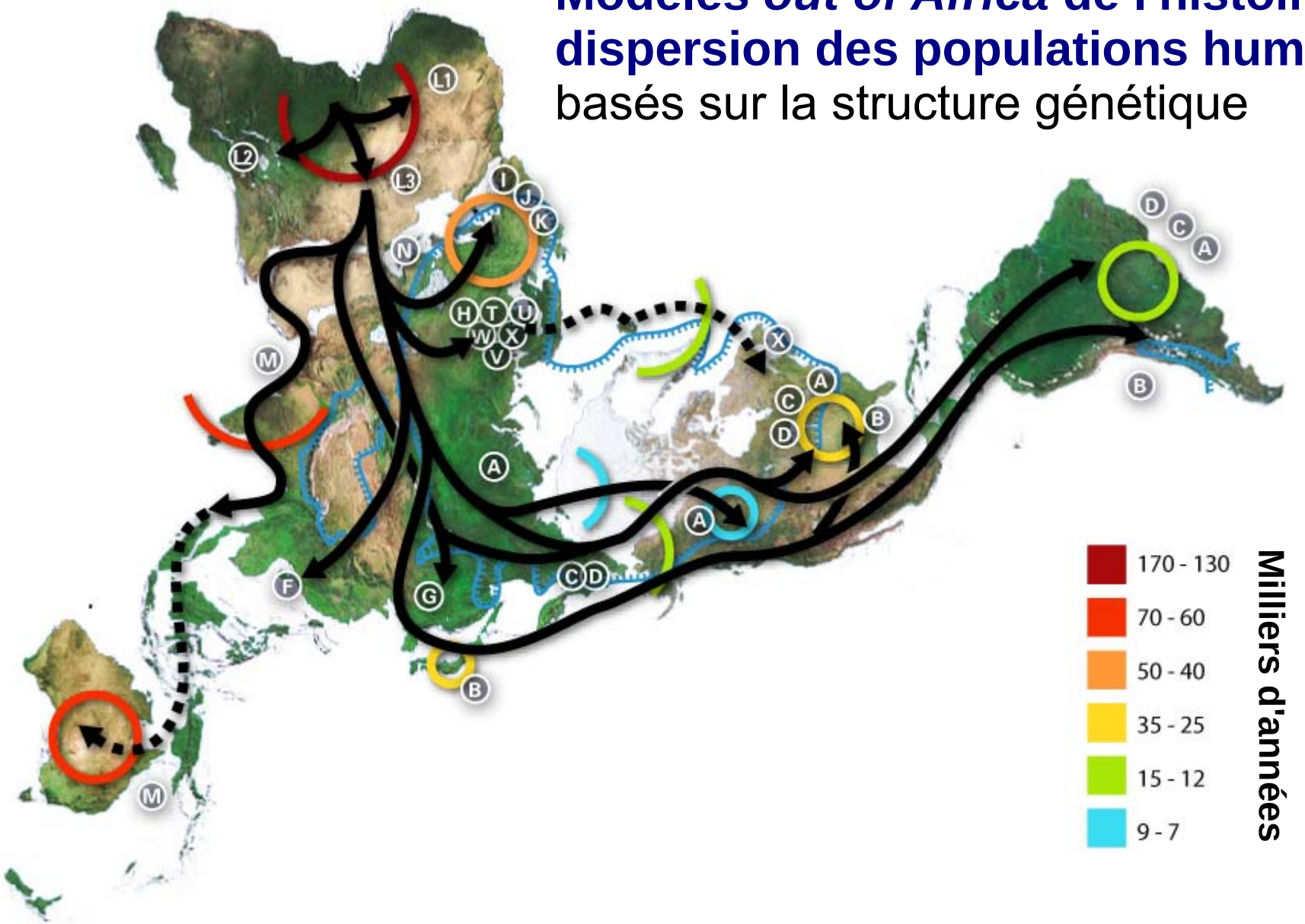
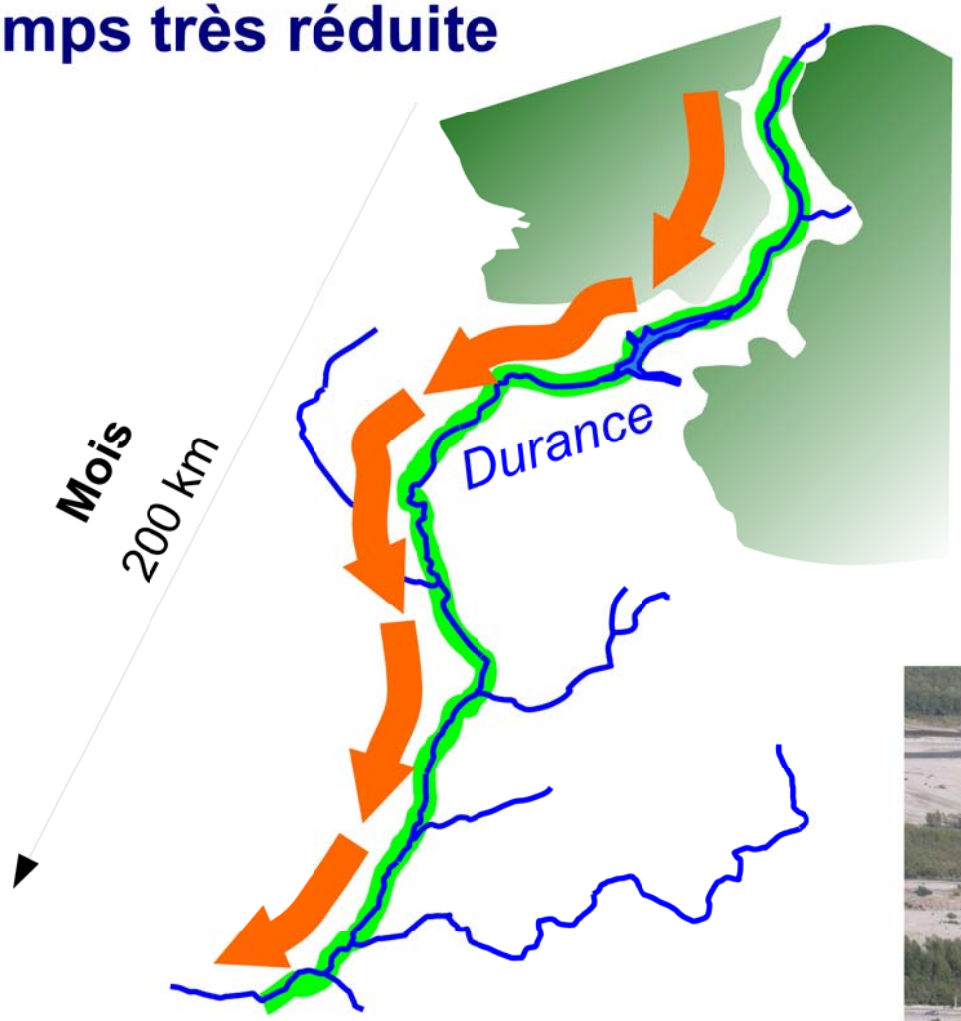
Estimation de paramètres démographiques à l'aide de la variation génétique

Modèles *out of Africa* de l'histoire de dispersion des populations humaines
basés sur la structure génétique

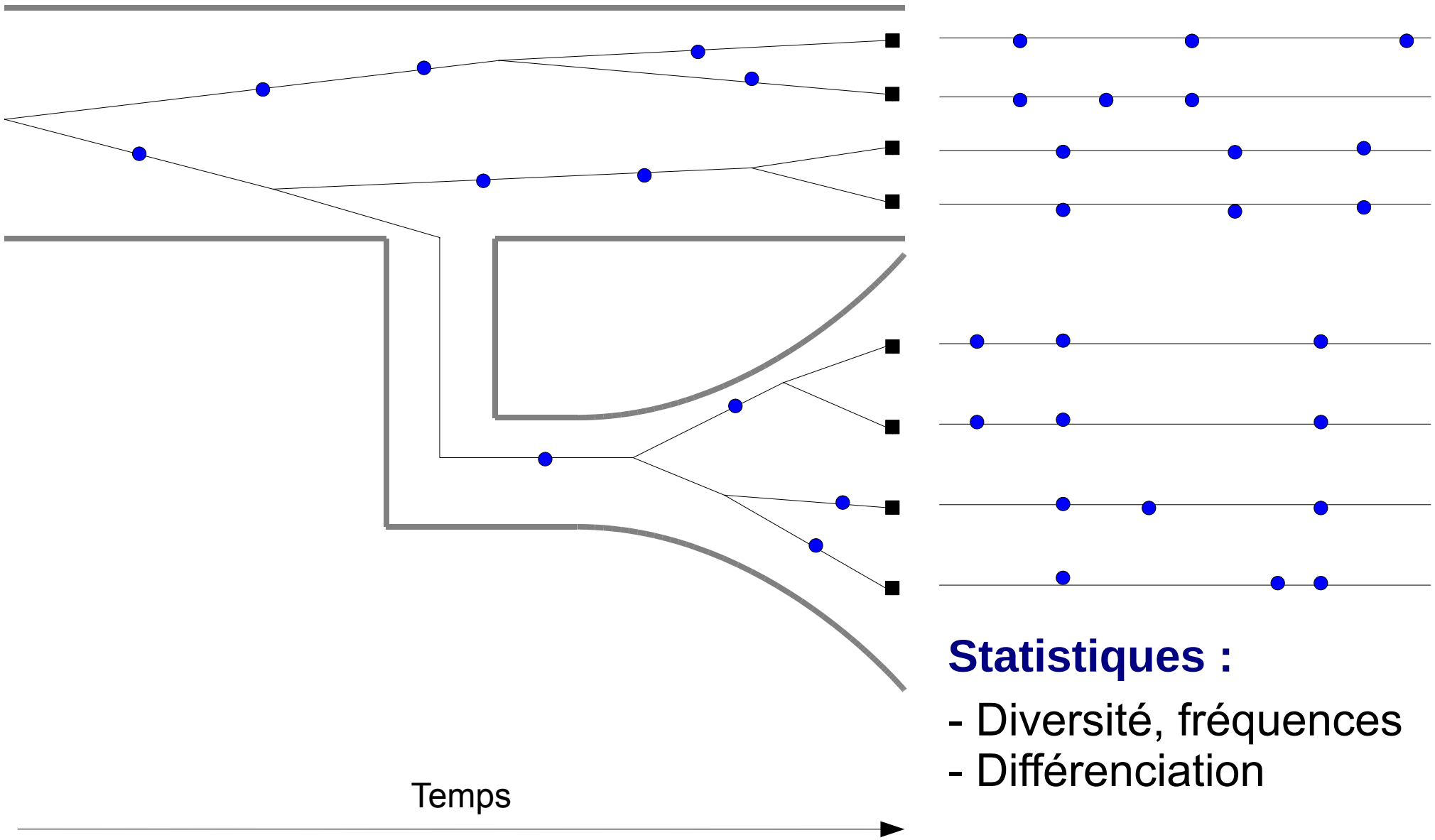


Application de la même approche au système d'invasion de la ripisylve de la Durance ?

- Échelle de temps très réduite
- Clonalité



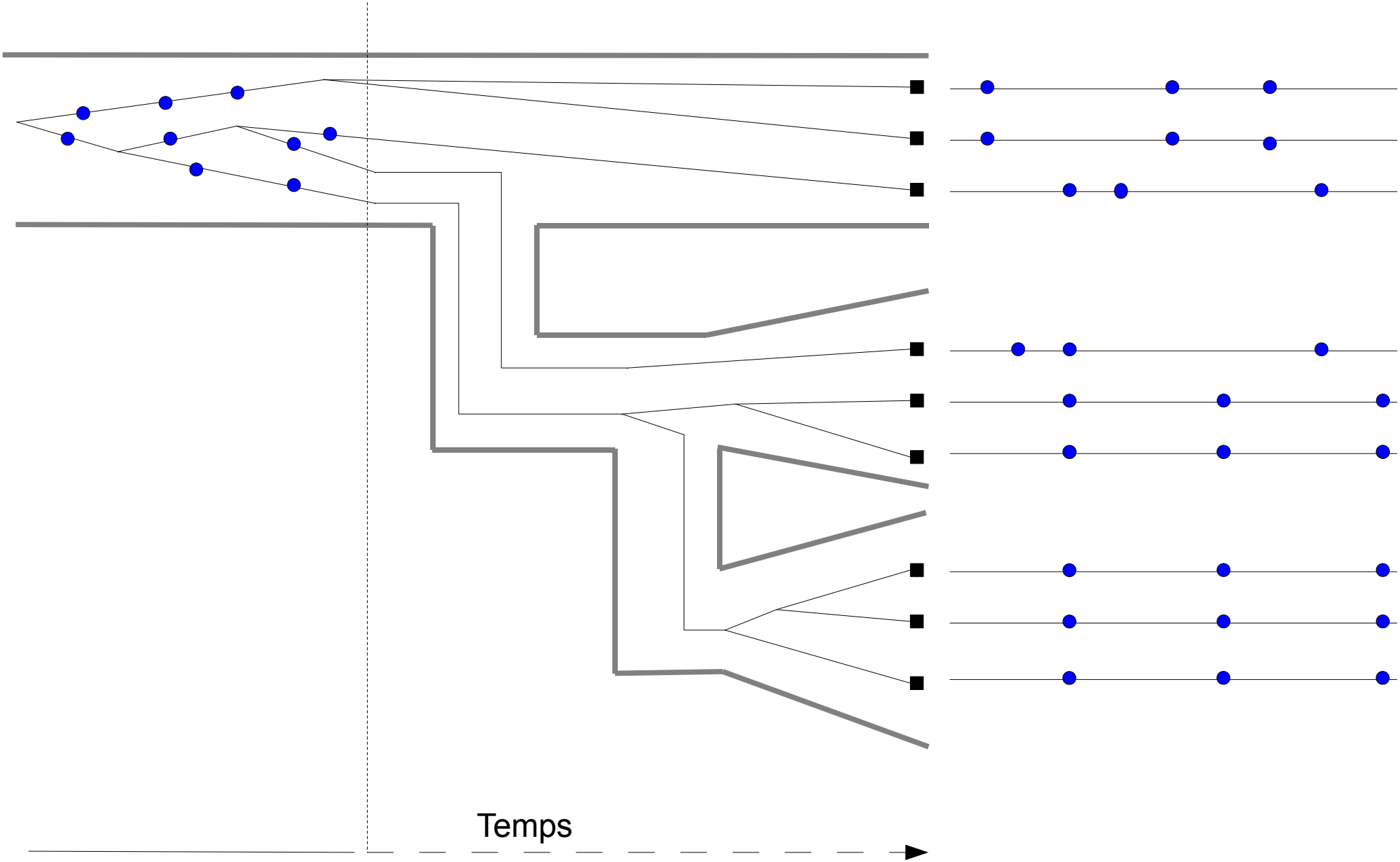
Modélisation de l'histoire des populations



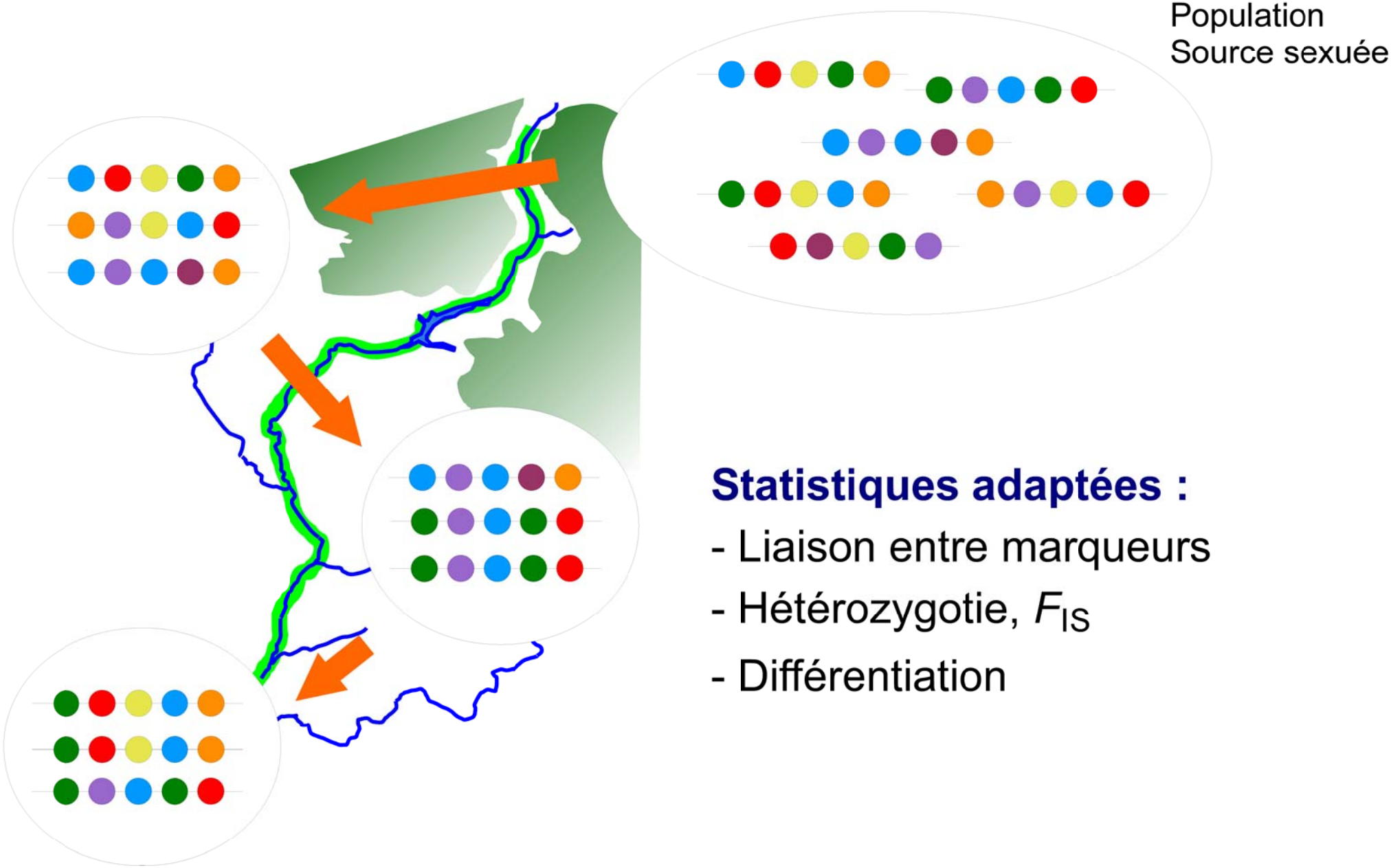
Statistiques :

- Diversité, fréquences
- Différenciation

Modélisation de l'histoire des populations à une échelle de temps réduite



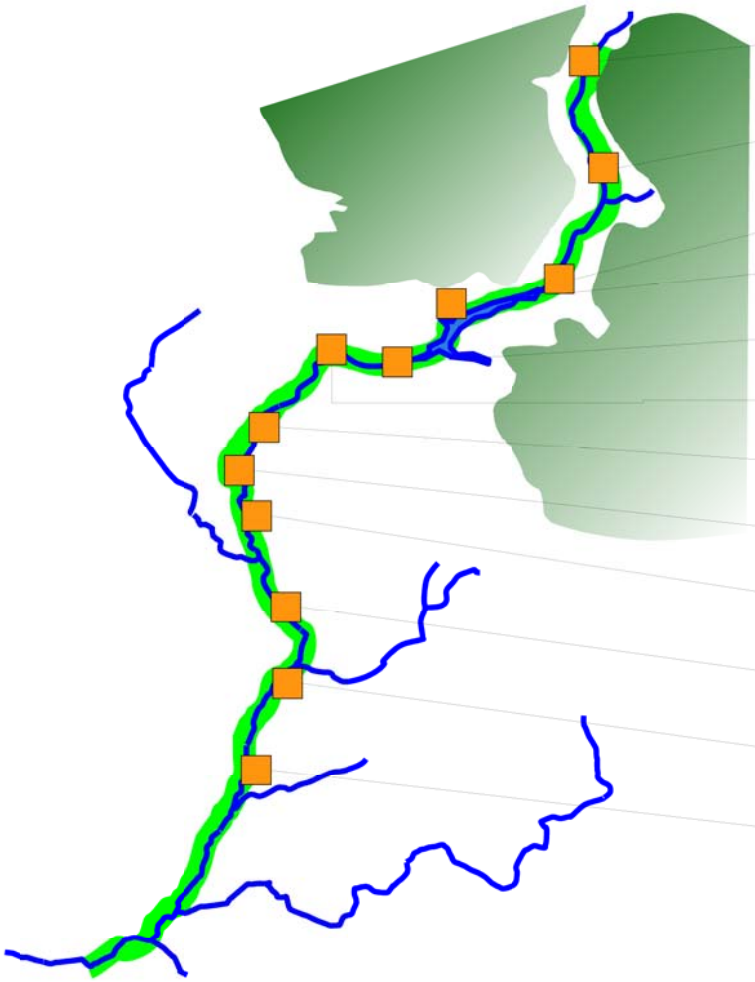
Exemple conceptuel



Statistiques adaptées :

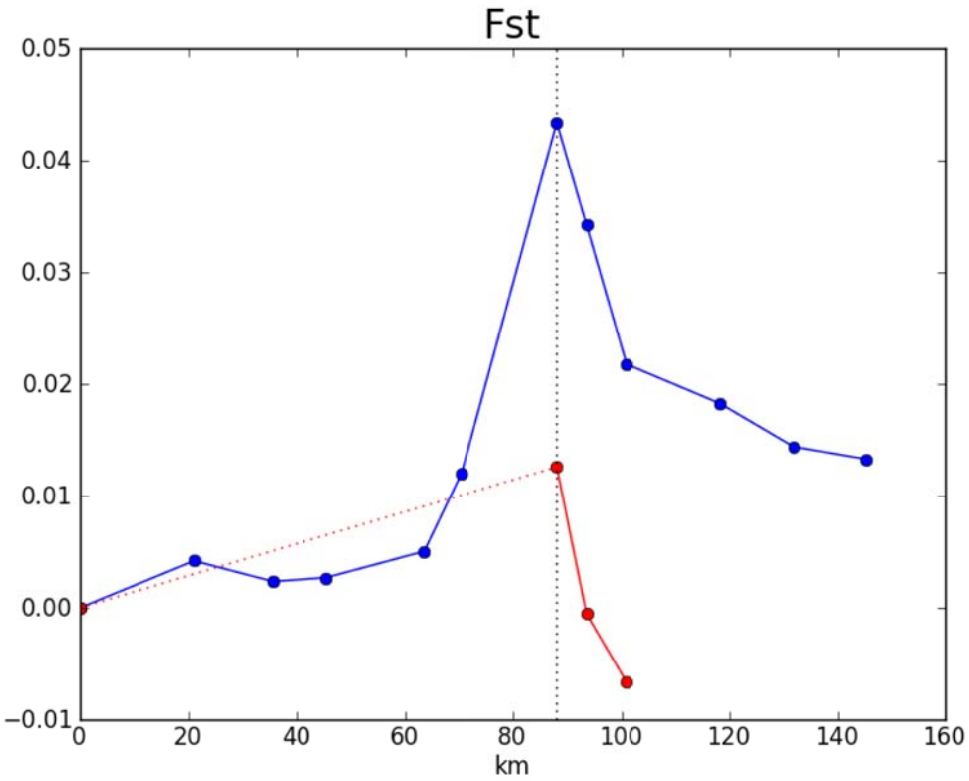
- Liaison entre marqueurs
- Hétérozygotie, F_{IS}
- Différentiation

Jeux de données



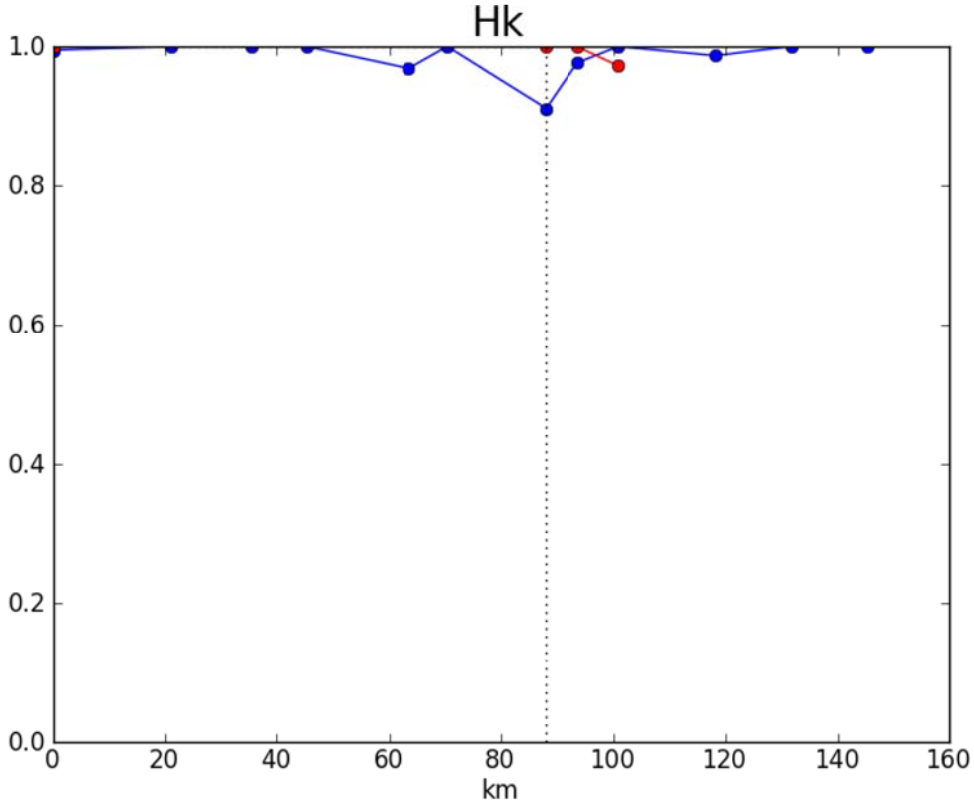
2008	2011	
20	9	
16		
23		
19		
20		
19		
10	12	Monêtier-Allemont
21	11	
9	9	
28		
19		
15		
16	**13**	**locus microsatellites**

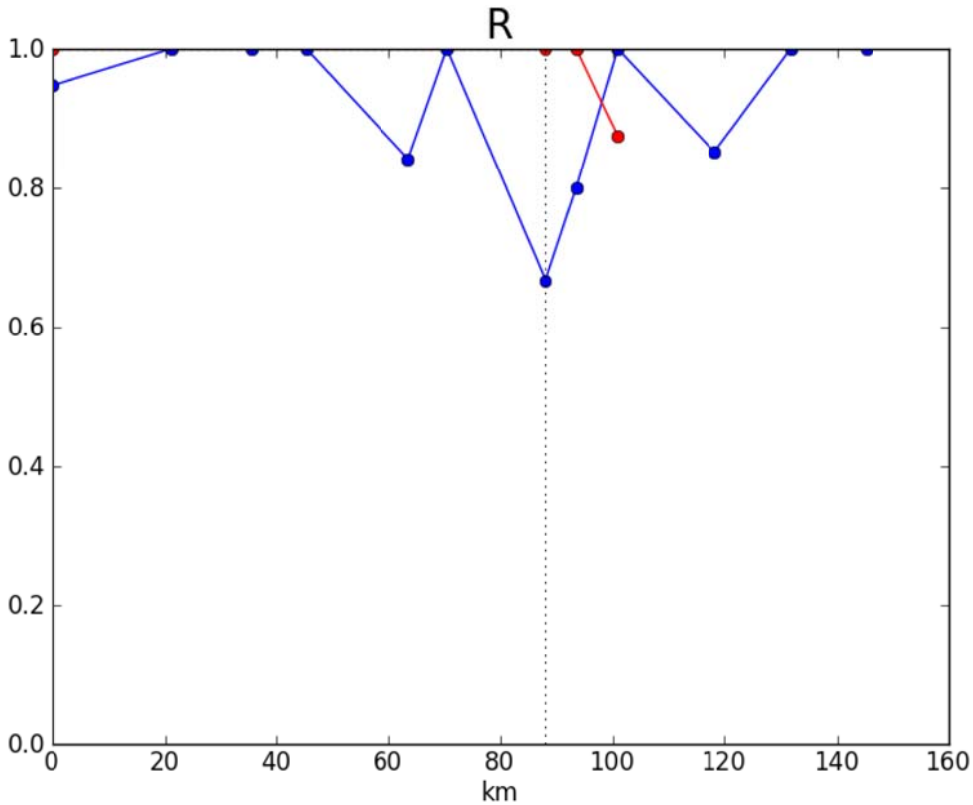
Signatures génétiques au niveau de Monêtier



La structure génétique change
(2008 et 2011)

Signatures génétiques au niveau de Monêtier



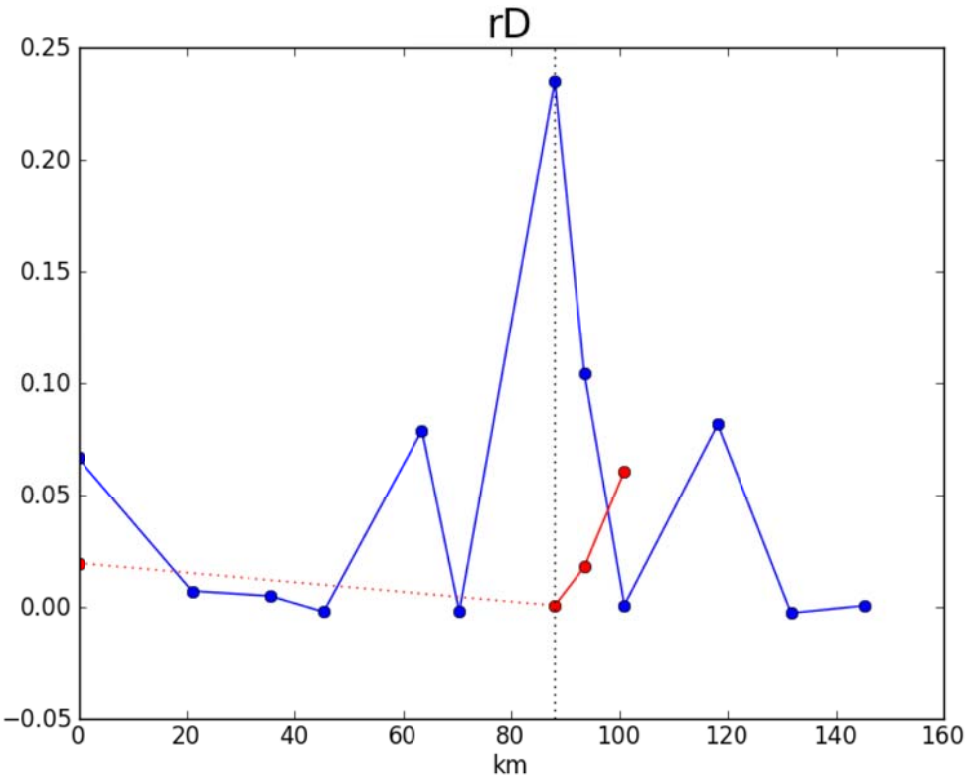


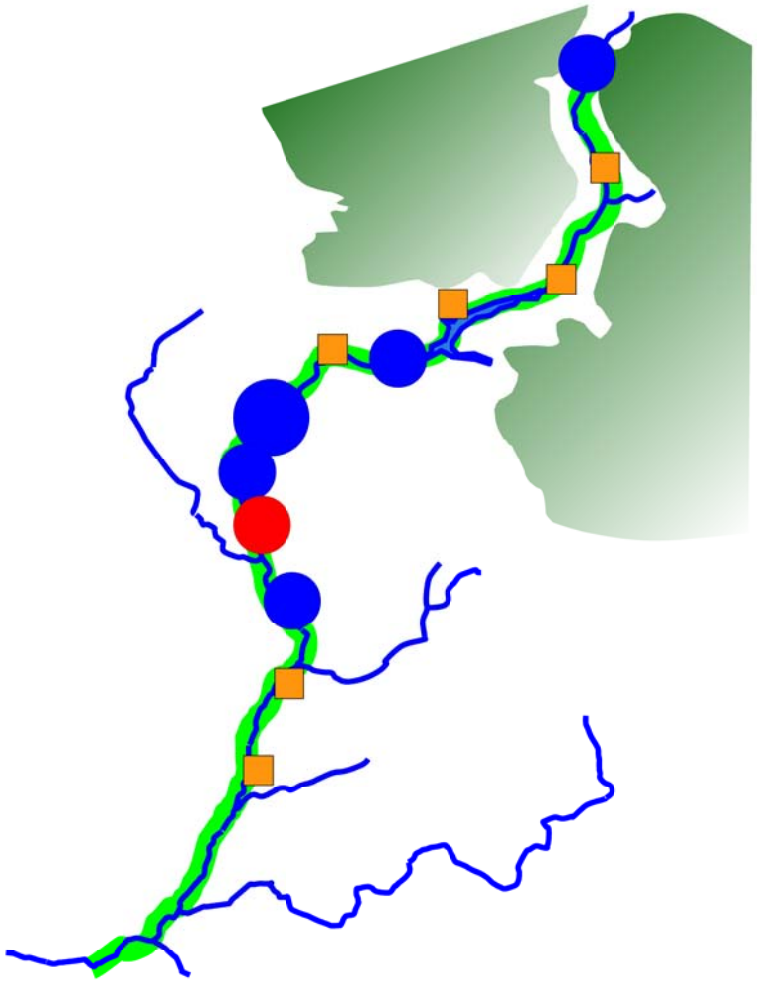
Diversité haplotypique
un peu moindre

Richesse allélique moindre :
génotypes répétés

Signatures de clonalité

Signatures génétiques au niveau de Monêtier

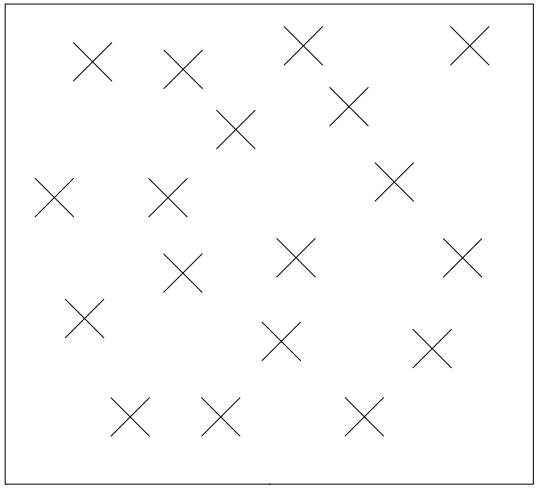


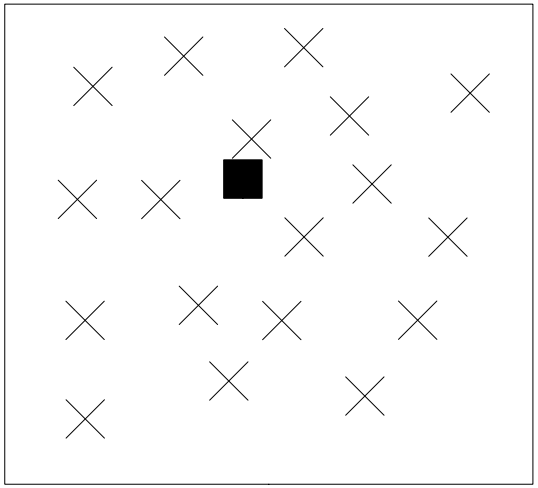


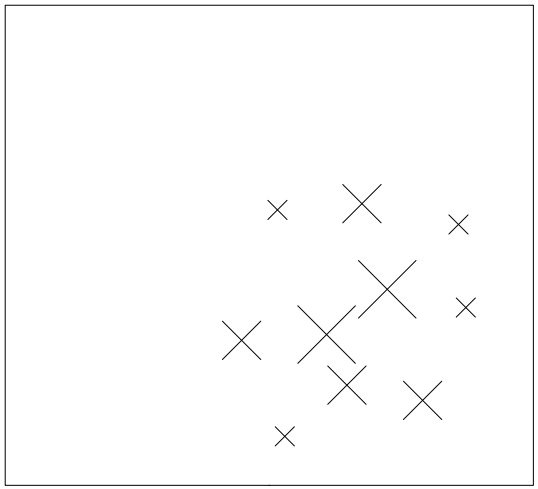
Liaisons entre marqueurs

Signatures de clonalité

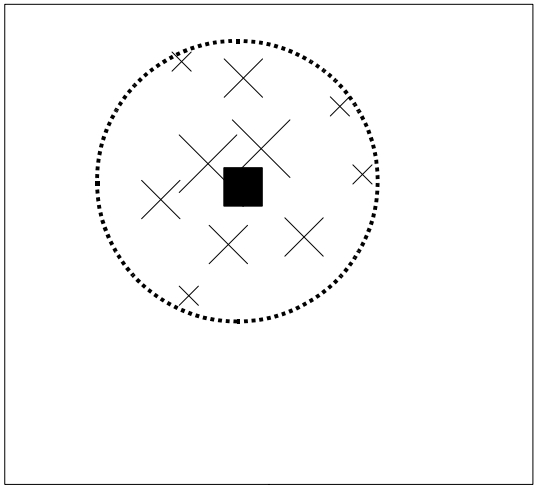
Approximate Bayesian Computation (ABC)







Paramètres



Statistiques

Pipeline ABC

1 – Observations -> statistiques descriptives

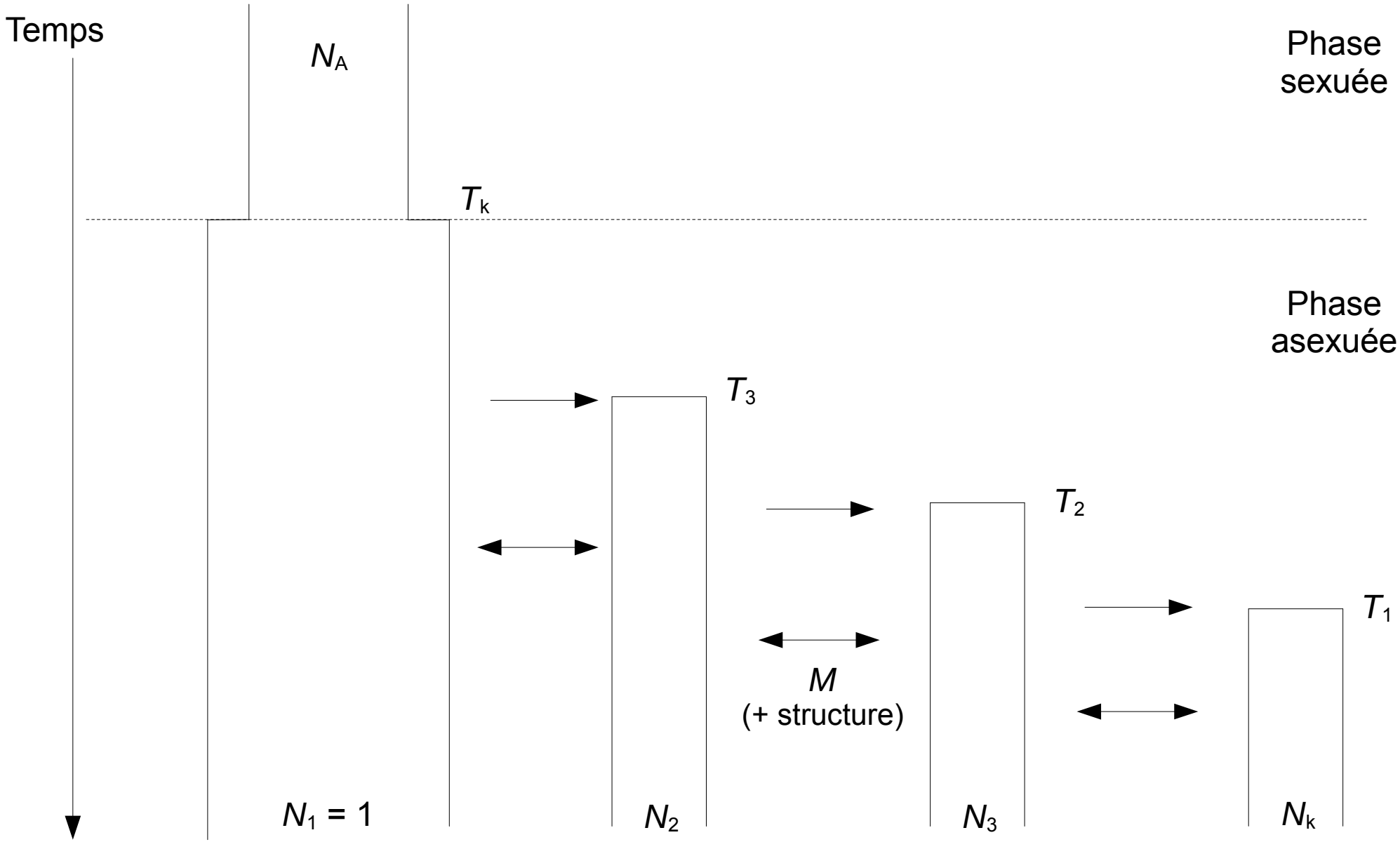
2 – Choix d'un **modèle**

3 – Distribution *a priori* des paramètres

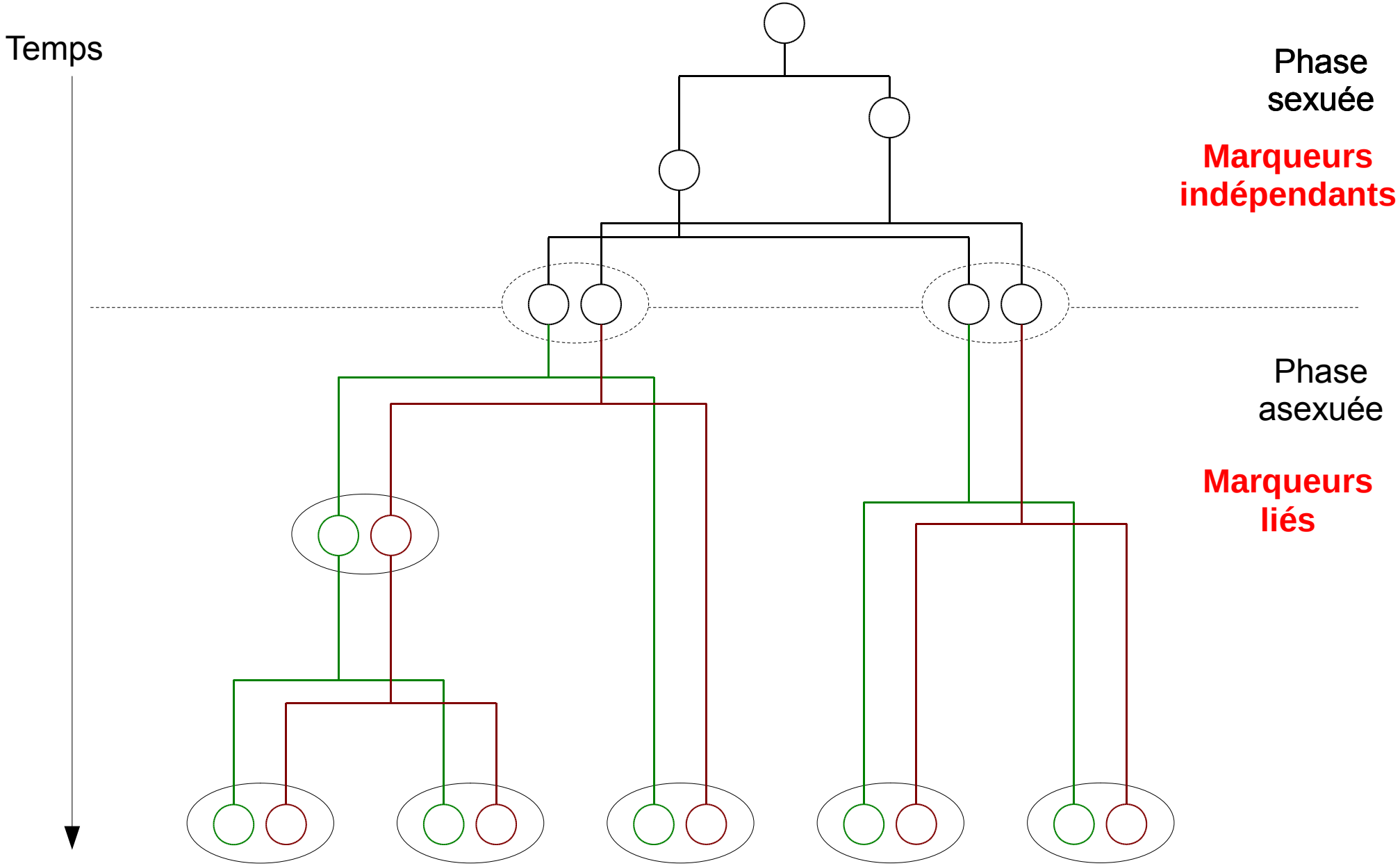
4 – Simulations selon le modèle (**régime clonal**) => coalescence

5 – Recherche des valeurs de paramètres optimales / choix de modèle

Modèle démographique



Implémentation du régiment de reproduction



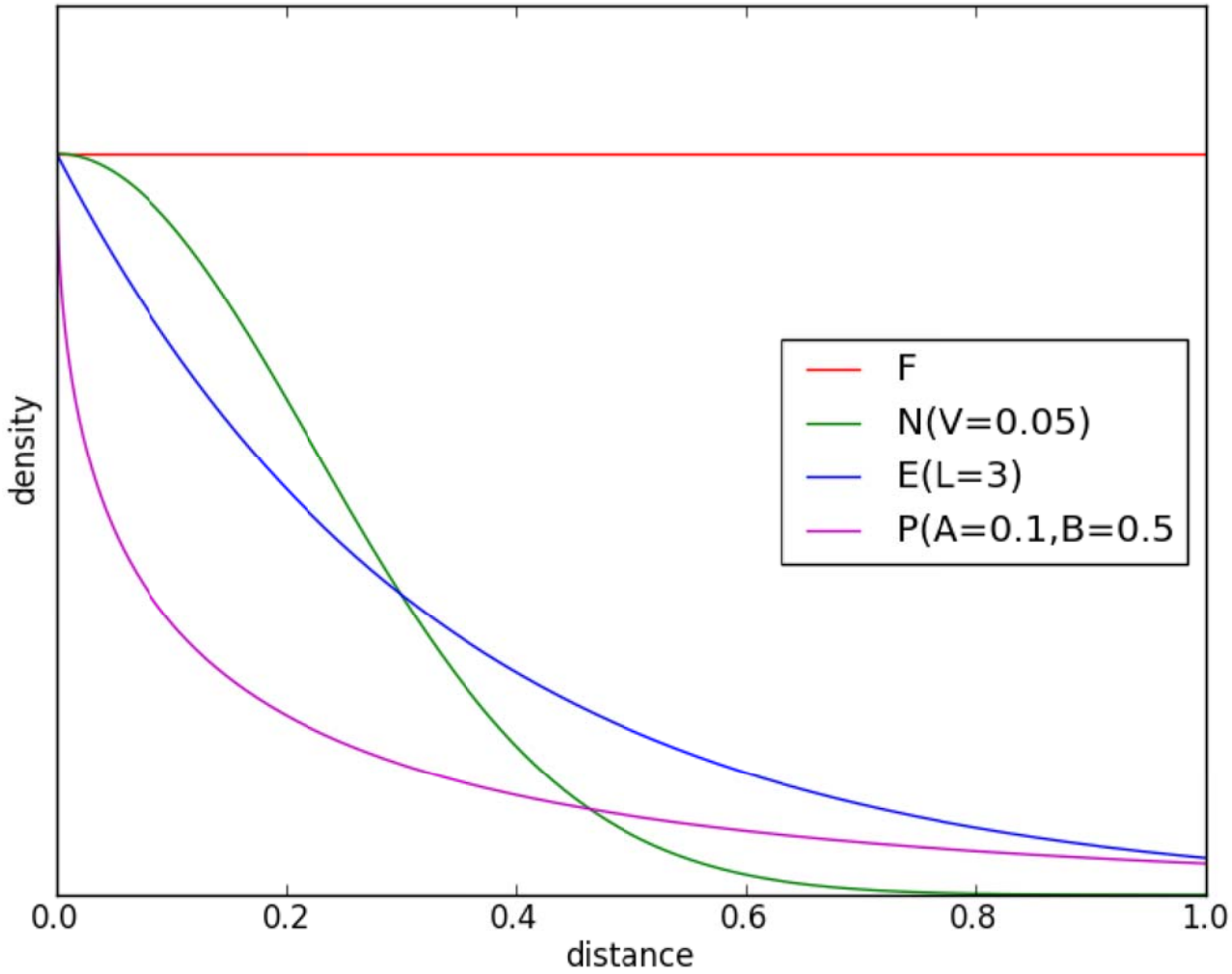
Statistiques et paramètres

Statistiques	Paramètres
$\overline{A}$	$N_2 - N_4$
$V(A)$	N_A
$\overline{H}_E$	$T_1 - T_K$
$V(H_E)$	M
$\overline{V}$	0-2 paramètres de structure de migration
$V(V)$	$\overline{\theta}$
$F_{IS1} - F_{ISK}$	$V(\theta)$
$K_1 - K_K$	p (pas de mutation)
$R_1 - R_K$	
$H_{K1} - H_{KK}$	
$R_1 - R_K$	
$r_{D1} - r_{DK}$	
$K_1 - K_k$	
$F_{ST1\text{-}2} - F_{ST1\text{-}K}$	

2008	2011
77	29

2008	2011
41-43	17-19

Modèles de migration



FLAT

Distribution uniforme
0 paramètre

NORMAL

Distribution normal $E = 0$
1 paramètre (écart-type)

EXPONENTIEL

Distribution exponentiel
1 paramètre (λ)

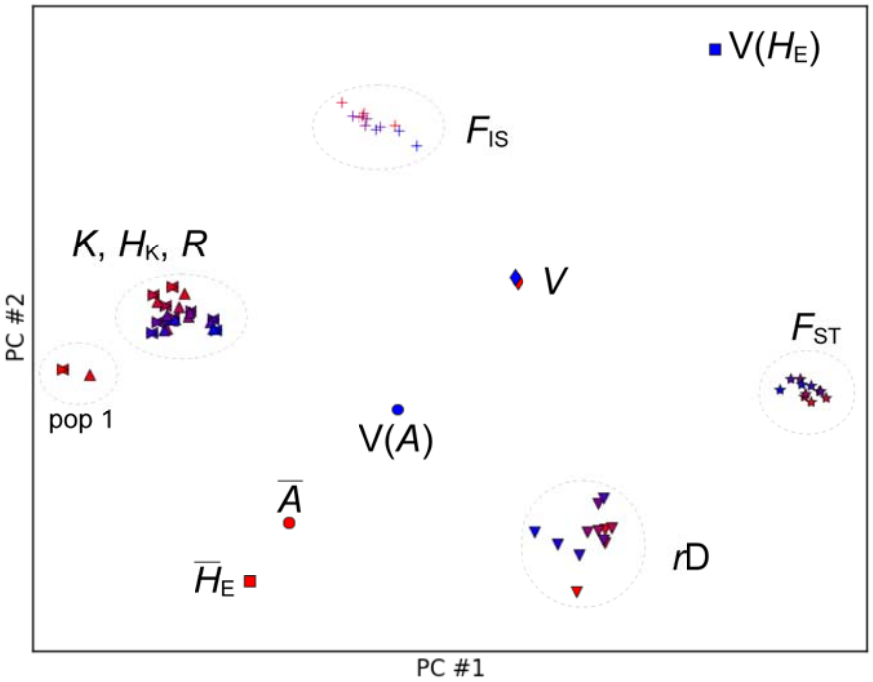
EXPONENTIEL-PUISSANCE

Distribution normale généralisée
2 paramètres :
- A : forme
- B : échelle

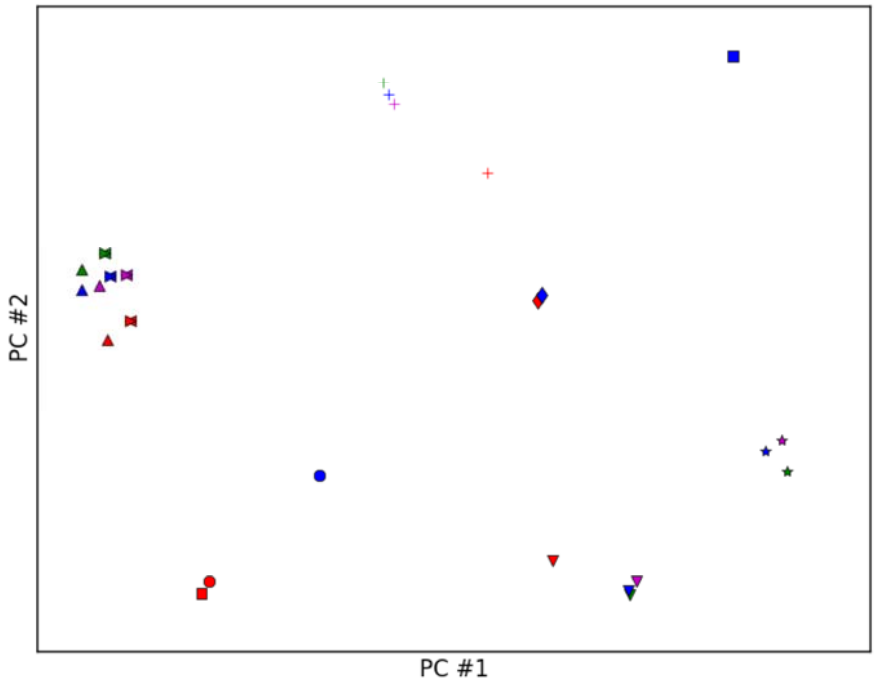
Corrélation entre statistiques – les données sont-elles bien résumées ?

ACP sur 1000 points (statistiques) pour chaque modèle

2008



2011

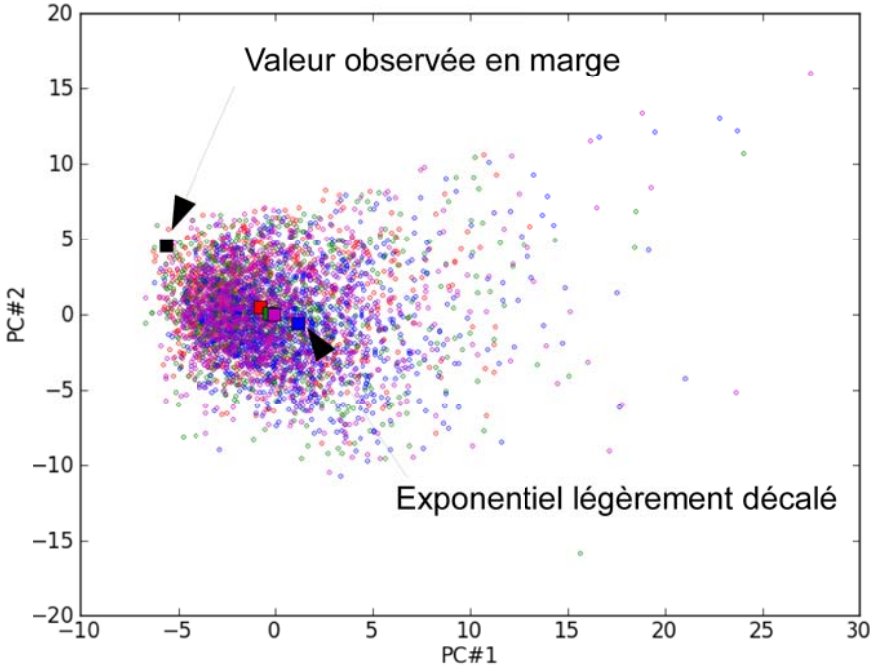


Assez de données pour estimer paramètres et choix de modèle ?

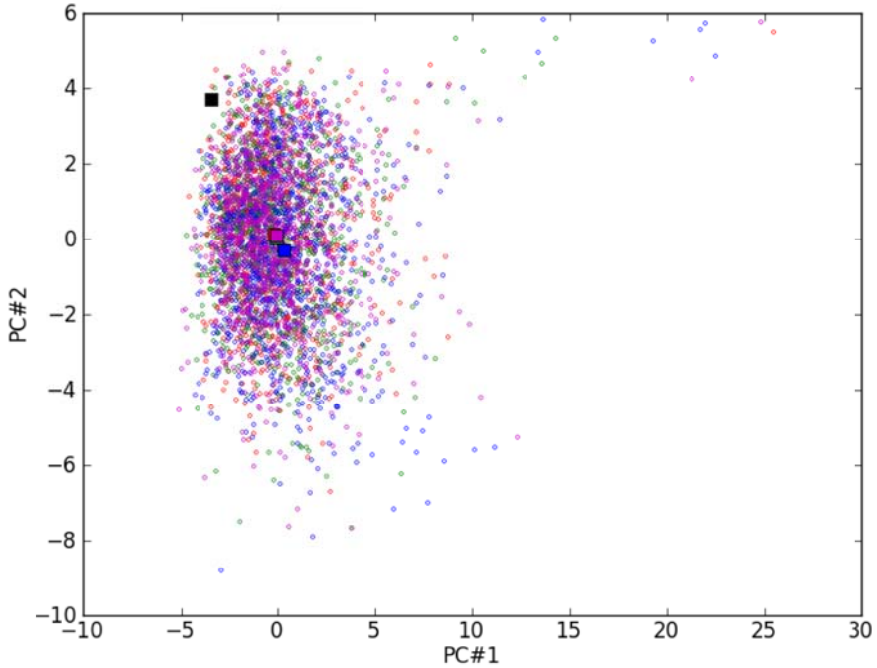
Discrimination des modèles ?

Projection de 1000 simulation par modèle

2008



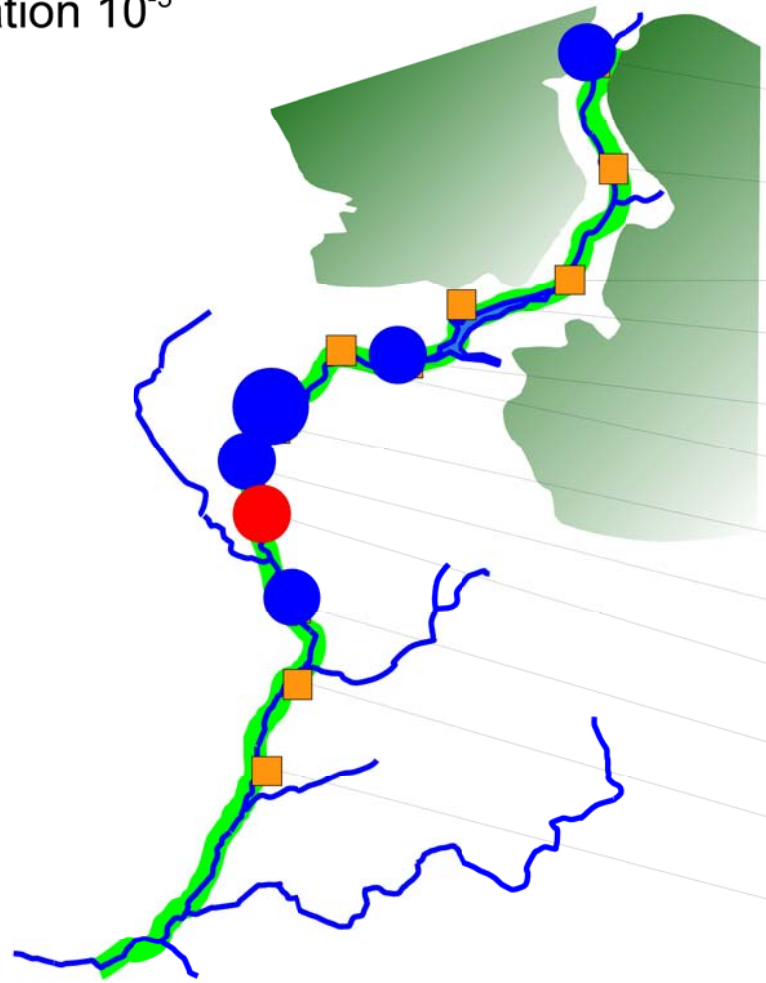
2011



Limite de l'ABC / coalescence dans ce cas ?
Modèle / paramètres inappropriés ?

Estimation des paramètres – taille des populations

> 9.5 10^6 simulations
Seuil d'acceptation 10^{-3}



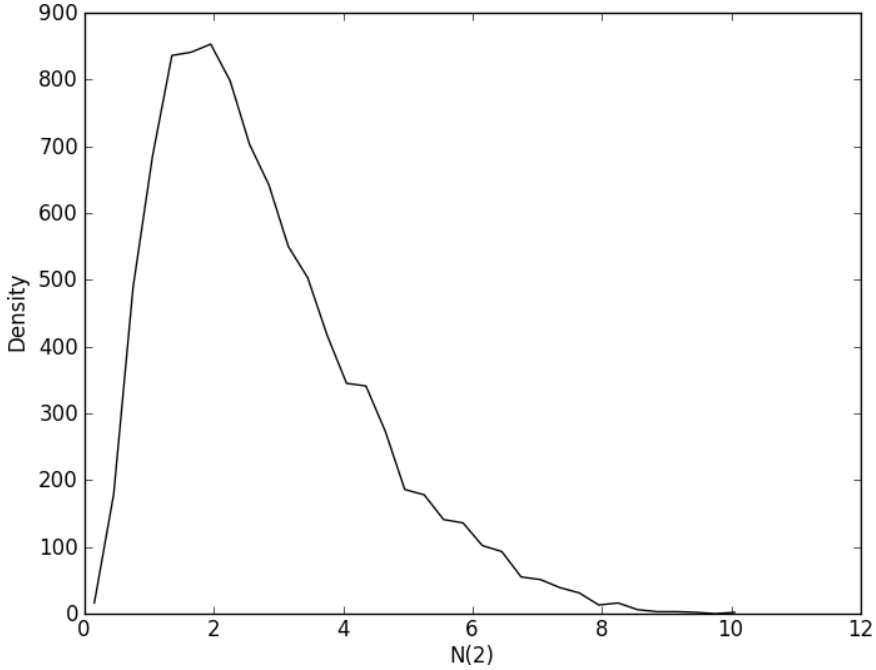
	2008	2011
N_A	0.4	0.8
	1	1
	1.7	
	1.6	
	1.4	
	0.6	
	1.5	
	0.4	1.1
	0.5	0.9
	0.9	0.6
	0.6	
	1.1	
	0.9	

Signatures de clonalité associés à faible *N*

Estimation des paramètres – dates de fondation

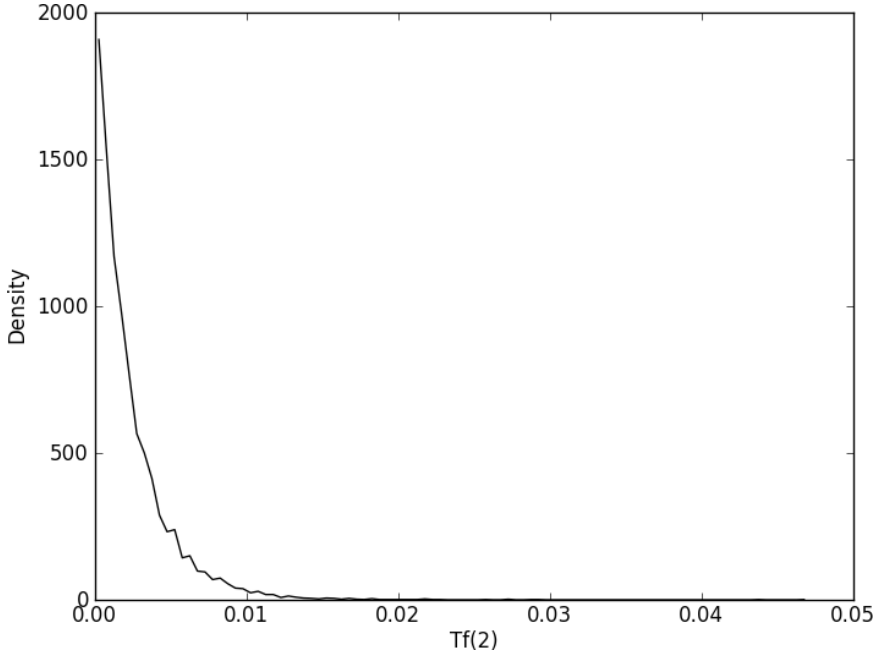
Tailles de populations :

Mode identifiable



Temps de divergence :

ne semblent pas différents de 0

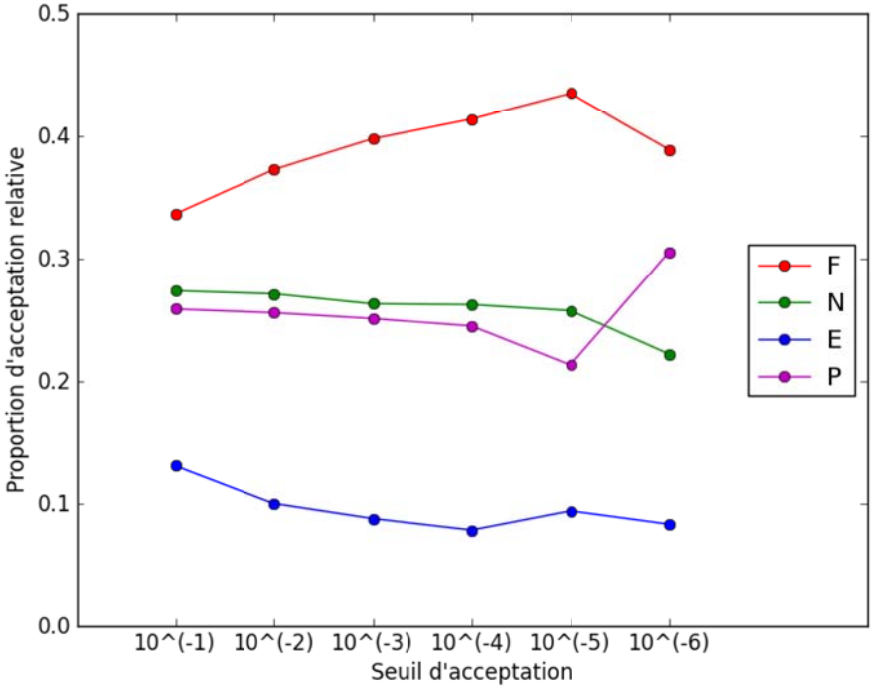


Problème d'échelle ou
problème d'échantillonnage des priors

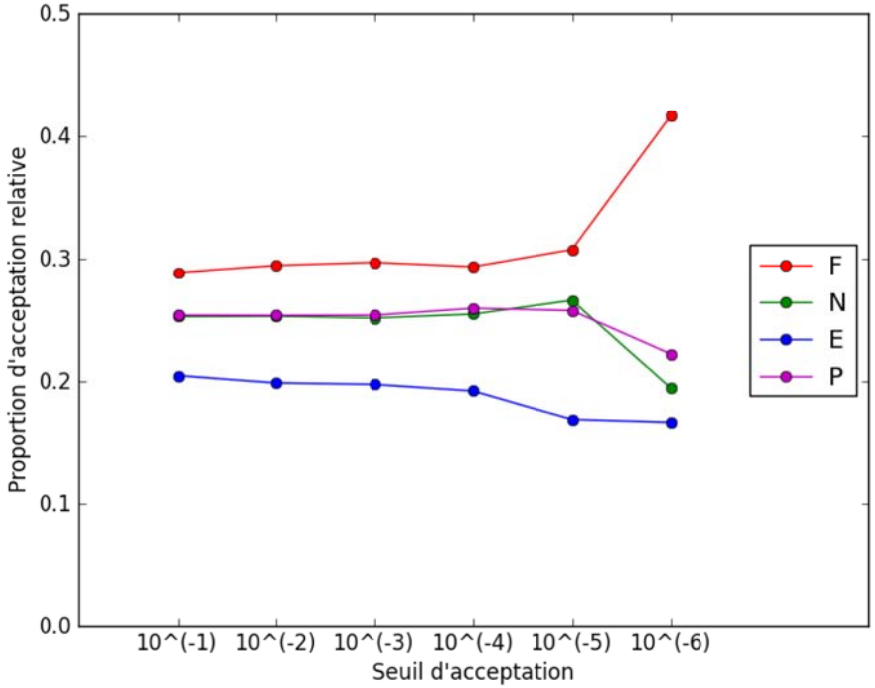
Paramètres de migration – choix de modèle

> 8 10^6 simulations par modèle
Seuil d'acceptation de 10^{-1} à 10^{-7}

2008



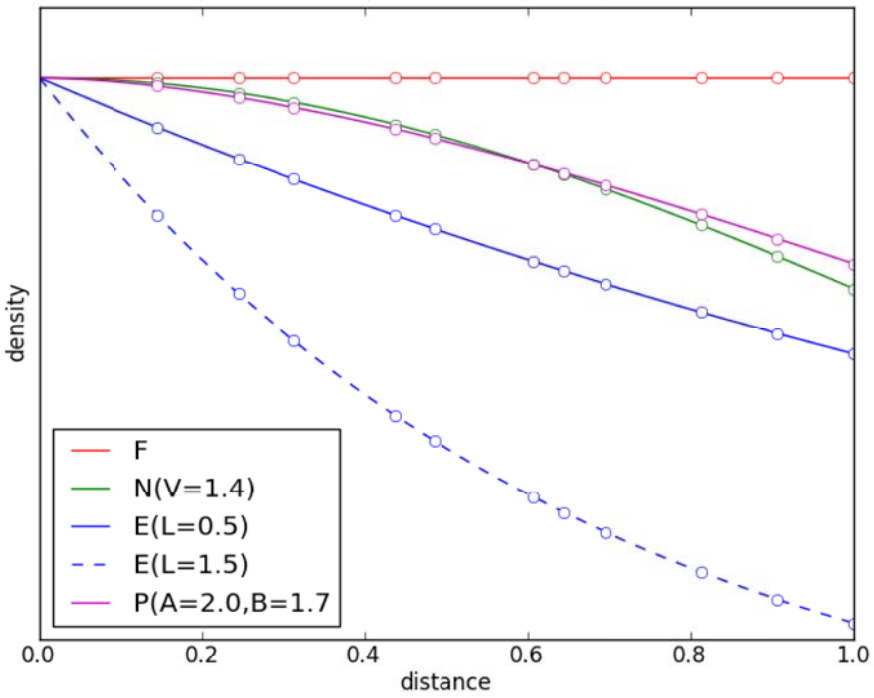
2011



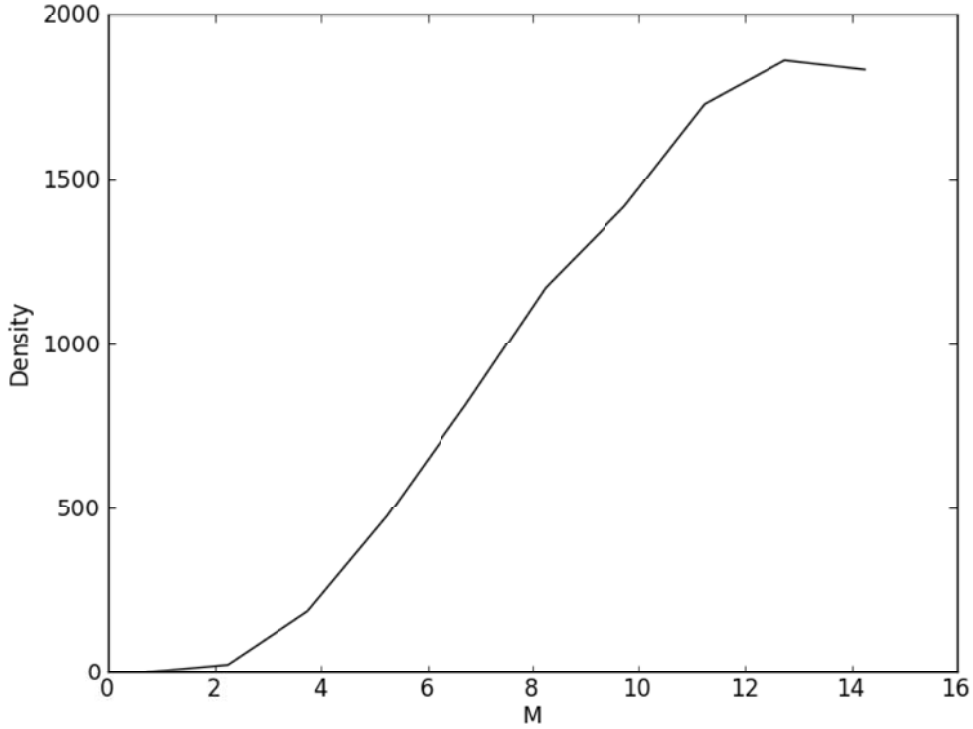
Modèle *flat* meilleur (modèle en île)

Estimation des paramètres de dispersion

Faible érosion du taux de dispersion avec la distance



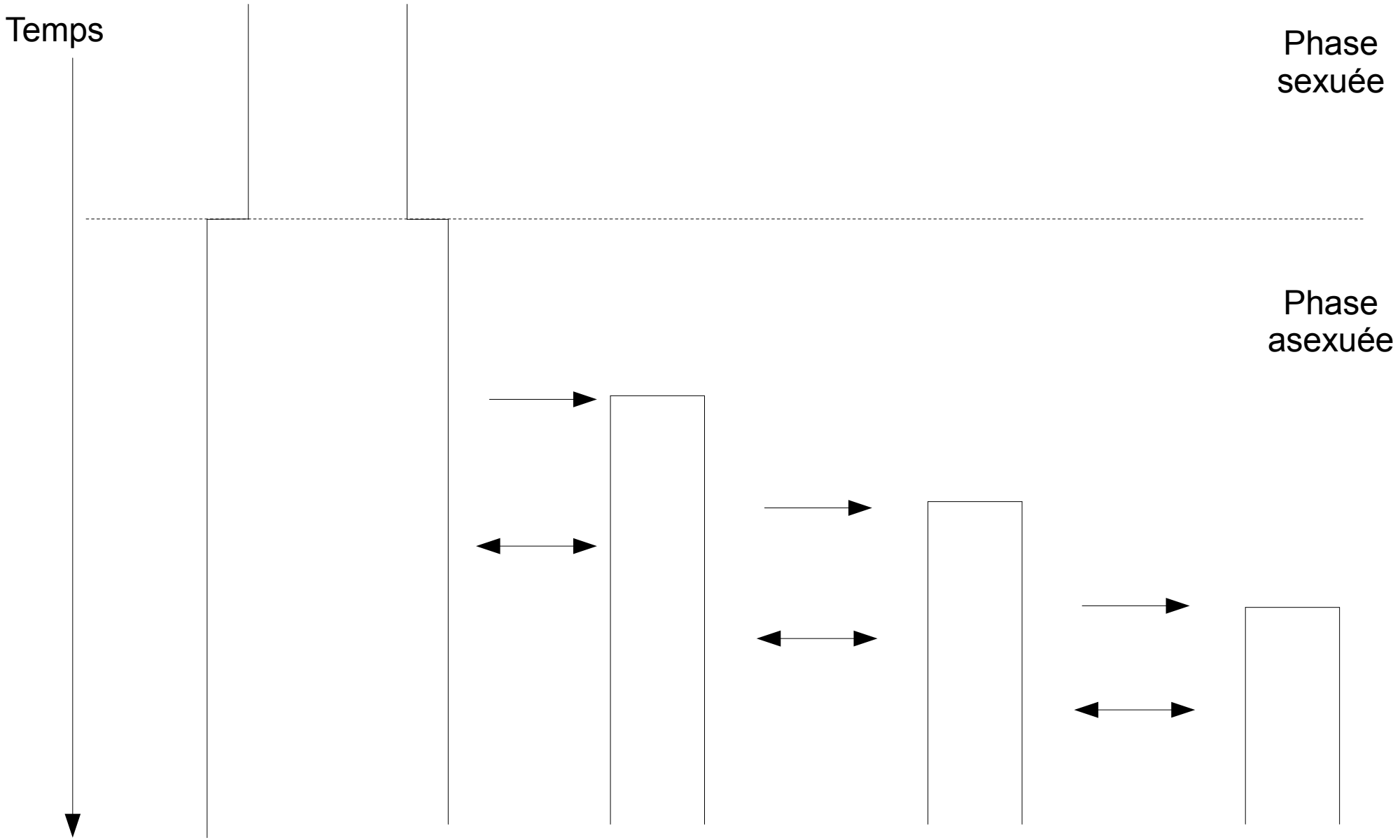
Taux de migration proche des valeurs maximales du prior



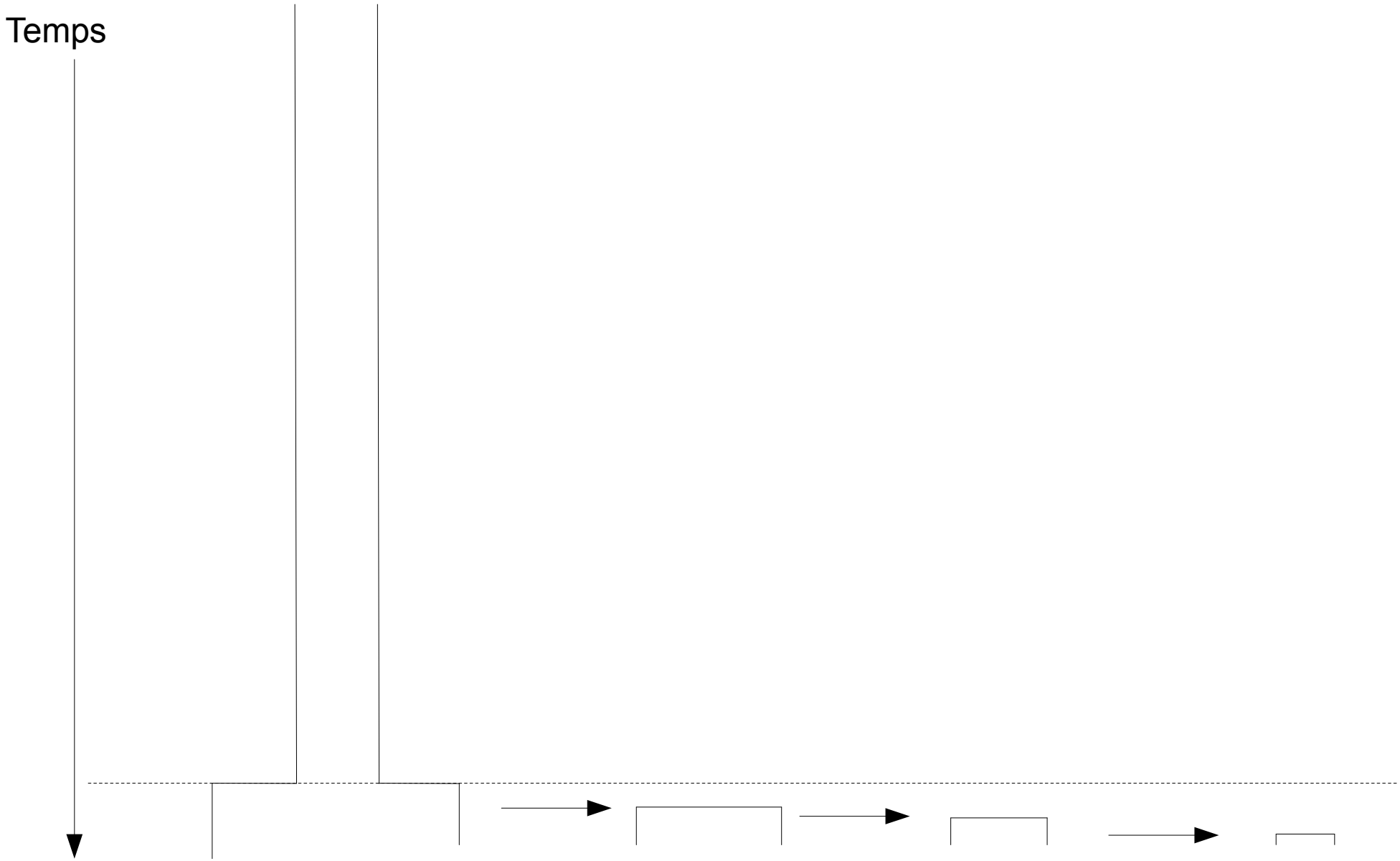
- Pas de signal de différenciation ?
(le F_{ST} peut être causé par la dérive)

- Modèle inapproprié ?

Modèle démographique



Modèle démographique



Conclusion

Certains paramètres bien estimés

- Taille de population, mais capture-t-on le bon signal ?
- Paramètres de mutation (mais différents entre années)

Temps de divergence proches de 0

- Modèle invalide ?
- Priors inappropriés ?

Structure de dispersion pas détectable

- Statistiques pas assez informatives ?
- Modèle trop compliqué ?
- Coalescent pas adapté ?

Perspectives

Affiner analyse ABC

- Priors moins naïfs
- Utilisation de données simulées
- Recherche de statistiques informatives

Simulations individu-centrées ?

- Une phase coalescent sexué + phase de dispersion *forward*

Et au-delà

- Échantillons le long de l'infection
- Intégrer différentes années (population source persistante)
- Intégration génétique + comptage ?