

Isolement par la distance et estimation de paramètres démographiques

Plan du cours :

1. Introduction à l'estimation de paramètres démographiques
2. Modèles démographiques structurés
3. Analyses/Méthode de la régression
 - Estimation
 - Significativité
4. Tests et comparaisons avec démographie
5. Pourquoi ça marche bien
6. Exemples d'applications

Module génétique des populations M2 MNHN

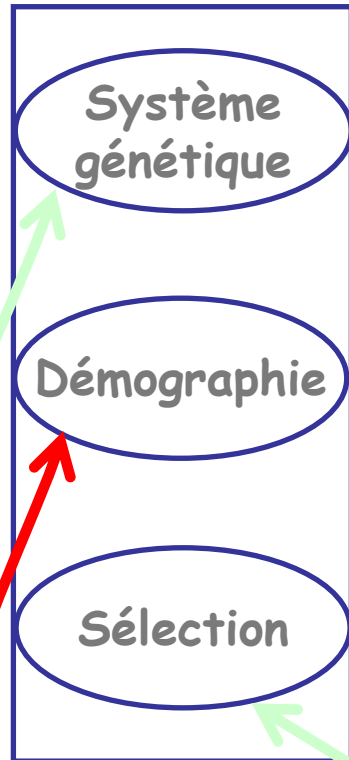
Raphaël Leblois, MC MNHN, dep^t Systématique & Evolution

Origine, Structure et Evolution de la biodiversité (UMR CNRS/MNHN/IRD 5202)

Problématique
générale

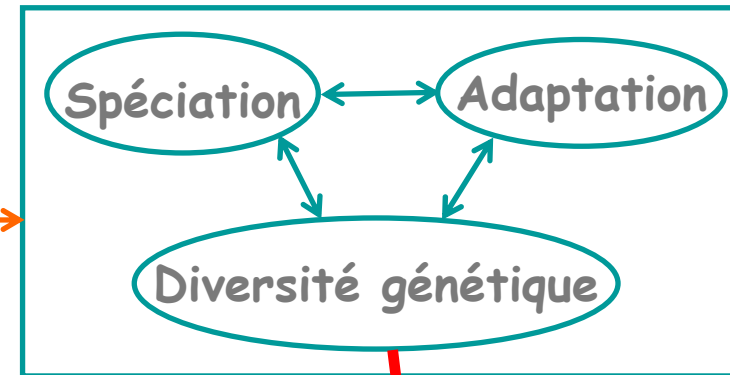
Génétique & biologie évolutive

Forces évolutives



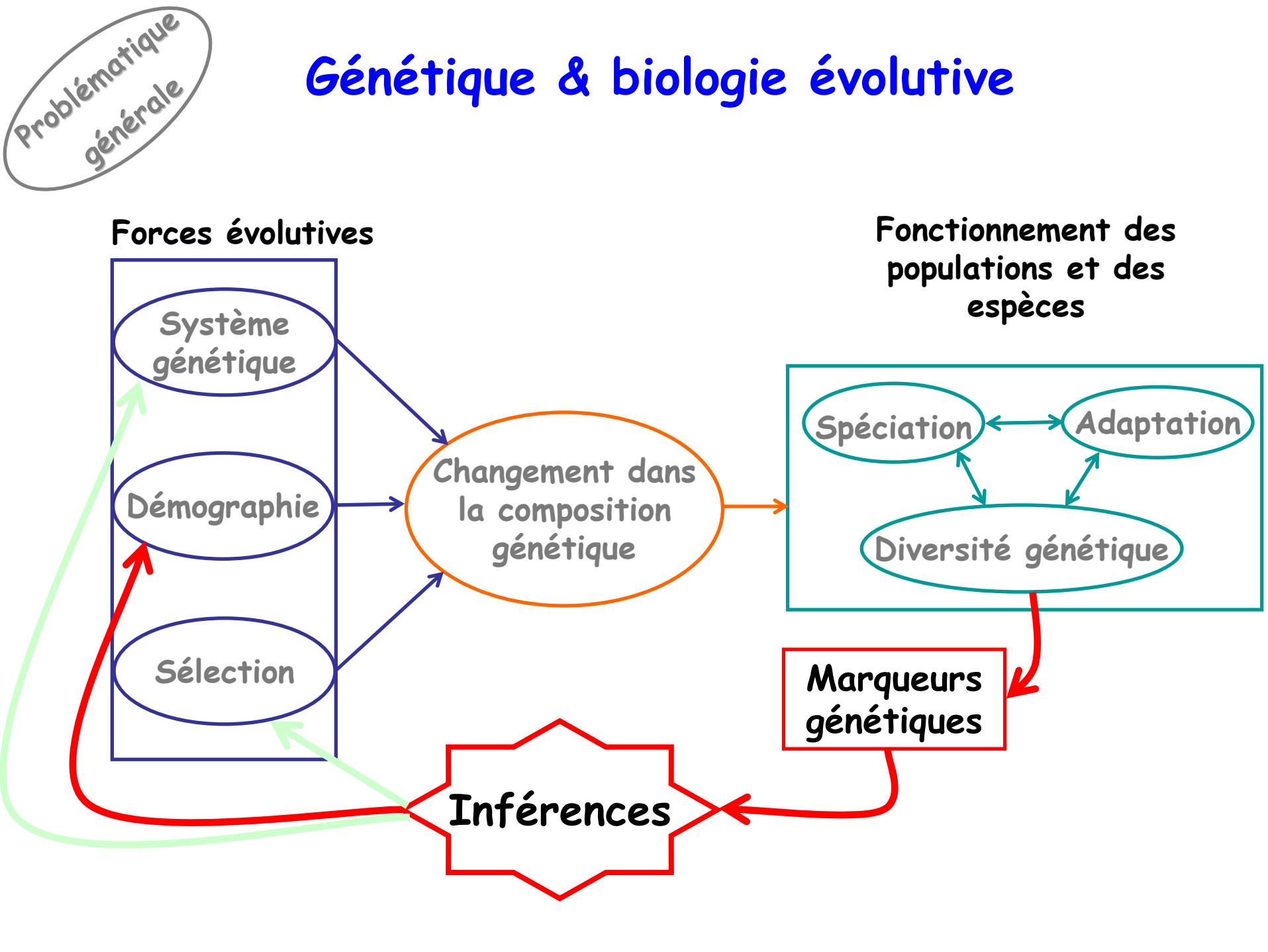
Changement dans
la composition
génétique

Fonctionnement des populations et des espèces



Marqueurs
génétiques

Inférences



1.1 Estimation de paramètres démographiques : pourquoi?

- Biologie de la conservation
 - Petites populations → risques d'extinction (dérive, stochasticité démographique et environnementale)
 - Plus de risques si populations isolées ou taux de migration faibles
- Bio-invasions
 - Suivi des tailles des populations envahissantes
 - Vitesse de colonisation
- Adaptation locale des populations naturelles à leur environnement

1.1 Estimation de paramètres démographiques : comment?

- **Méthodes directes:** capture-marquage-recapture
 - Pb : variabilité temporelle, demande beaucoup de temps
- **Méthodes indirectes:** variabilité génétique neutre
 - Améliorations récentes:
 - Marqueurs hautement polymorphes
 - Nouvelles méthodes d'analyse statistiques
- Ces approches sont-elles équivalentes?
 - ✓ approches **indirectes** → paramètres « **efficaces** »

capture-marquage-recapture → paramètres démo **actuels**

variabilité génétique \leadsto paramètres démo « **passés** »

1.1 Estimation de paramètres démographiques : comment?

- **Méthodes directes:** capture-marquage-recapture
 - Pb : variabilité temporelle, demande beaucoup de temps
- **Méthodes indirectes:** variabilité génétique neutre
 - Améliorations récentes:
 - Marqueurs hautement polymorphes
 - Nouvelles méthodes d'analyse statistiques
- Ces approches sont-elles équivalentes?
 - ✓ approches **indirectes** → paramètres « **efficaces** »

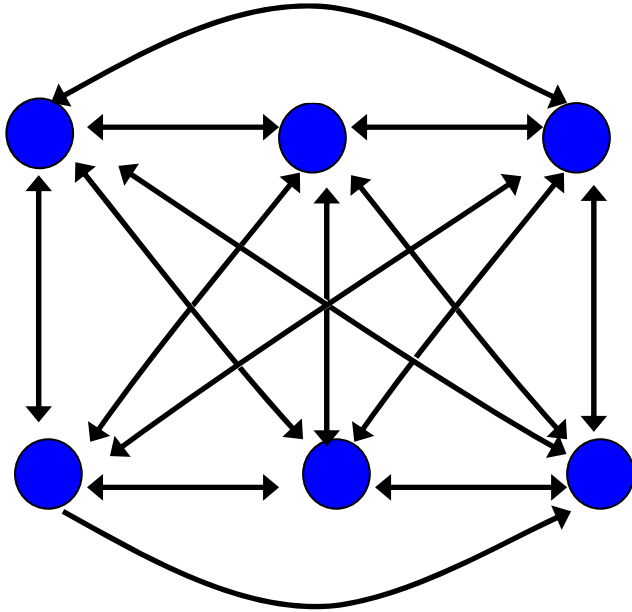
capture-marquage-recapture → paramètres démo **actuels**

variabilité génétique → paramètres démo « ~~passés~~ » **actuels**

2. Modèles démographiques de populations structurées :

2.1 Le modèle en îles

➤ Wright (1931,1937)



Simple car homogénéité réduit à 3 le nombre de paramètres :

d = nombre de sous-populations (ou ∞)

N = taille des sous-populations

m = taux de migration

On a alors la fameuse relation:

$F_{st} = 1/(1+4Nm)$ en nombre d'iles infini

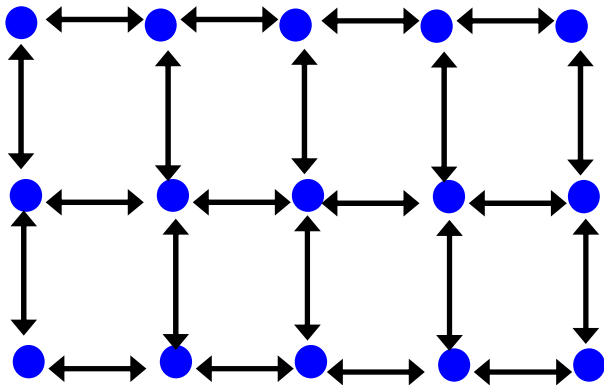
Problème principal : Migration indépendante de la distance entre sous-populations -> pas très réaliste

Bon modèle pour études théoriques mais pas pour estimation précise de paramètres démographiques

2. Modèles démographiques de populations structurées :

2.2 Le modèle de migration par pas (stepping stone)

➤ Kimura (1953), Malécot (1959), Kimura and Weiss (1964)



Aussi très simple, Peu de paramètres

d = nombre de sous-populations

N = taille des sous-populations

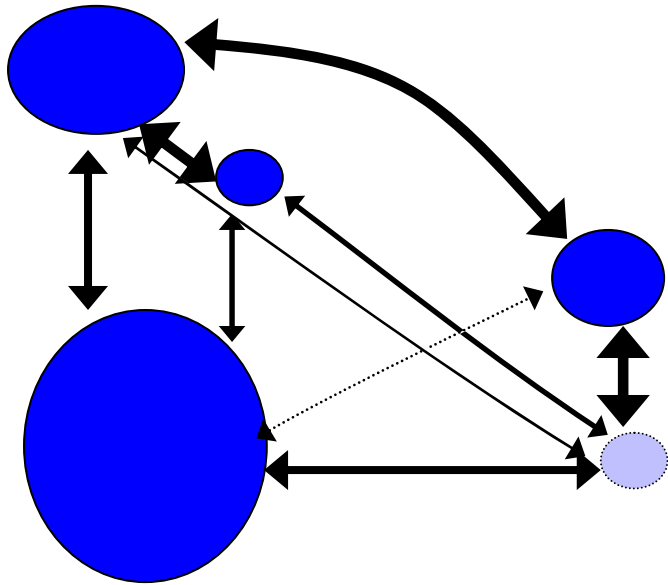
m = taux de migration

Migration seulement entre sous-populations adjacentes
→ pas très réaliste mais 1^{er} modèle analysable avec dispersion localisé dans l'espace

2. Modèles démographiques de populations structurées :

2.3 Le modèle de méta-populations

Une métapopulation est une population de populations (Levins 1969,1970) avec des extinctions possibles



Non homogène -> Beaucoup de paramètres

d = nombre de sous-populations

$\{N_1, \dots, N_d\}$ = taille des sous-populations

$\{m_{ij}\}$ = taux de migration entre paires de populations

$\{e_i\}$ = taux d'extinction des population

Très réaliste mais trop de paramètres -> **problème pour l'estimation**

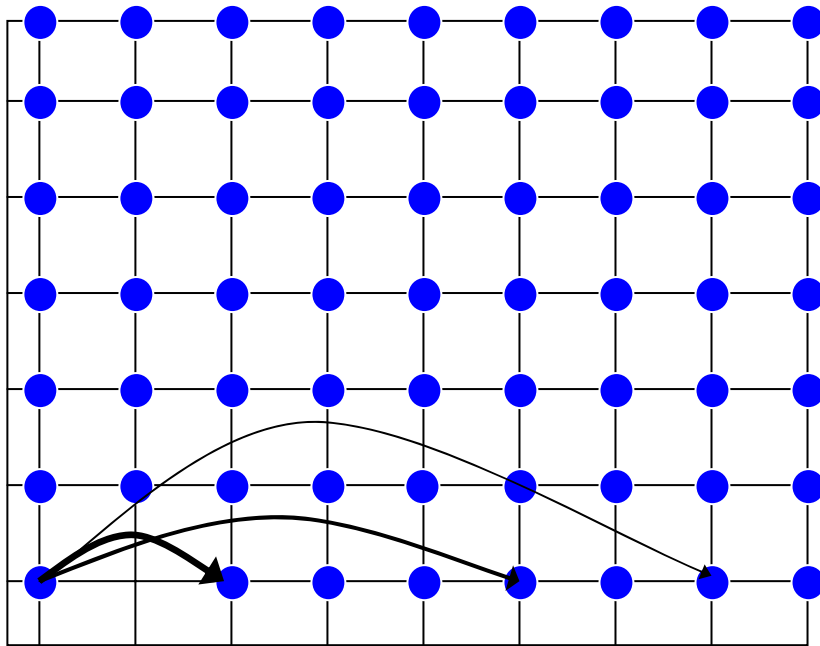
Homogénéisation -> modèle en îles avec extinction

2. Modèles démographiques de populations structurées :

2.4 Le modèle d'isolement par la distance

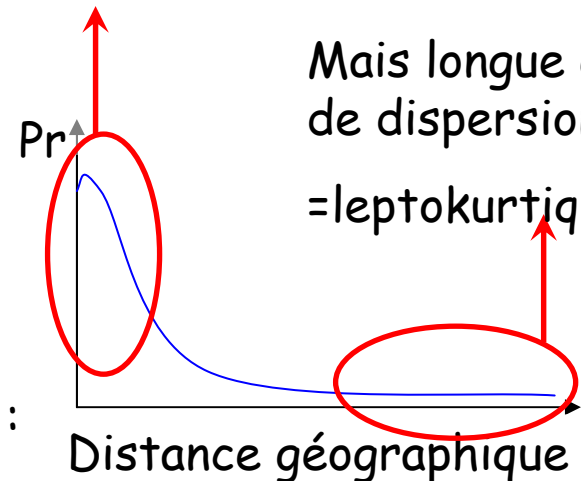
Dispersion limitée dans l'espace \leftrightarrow 2 individus ont plus de chance de se reproduire ensemble si ils sont proches géographiquement

Endler 1977 (revue biblio): la majorité des espèces ont une dispersion localisé



Majorité de la dispersion à très courte distance

Mais longue queue de dispersion
= leptokurtique

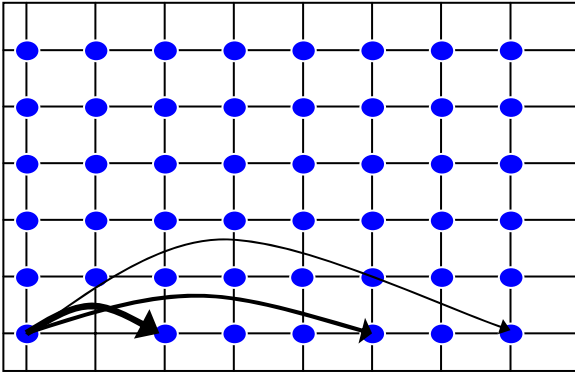


Migration fonction de la distribution de dispersion :

2. Modèles démographiques de populations structurées :

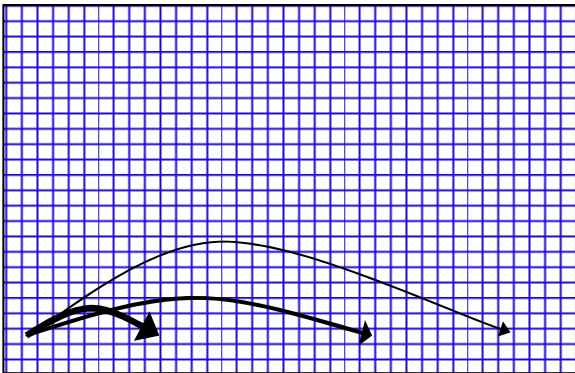
2.4 Les modèle d'isolement par la distance

2 modèles en fonction du type de distribution des organismes dans le paysage :



Population en dèmes

Chaque nœud du réseau correspond à une sous population panmictique



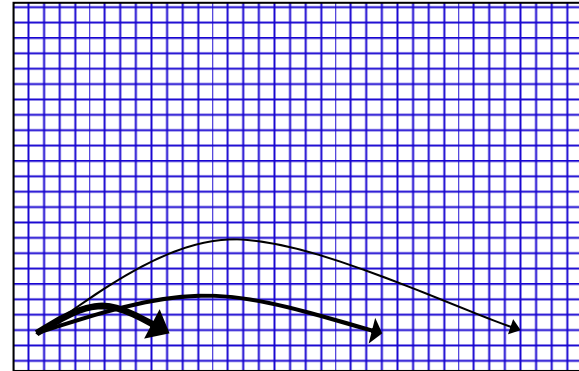
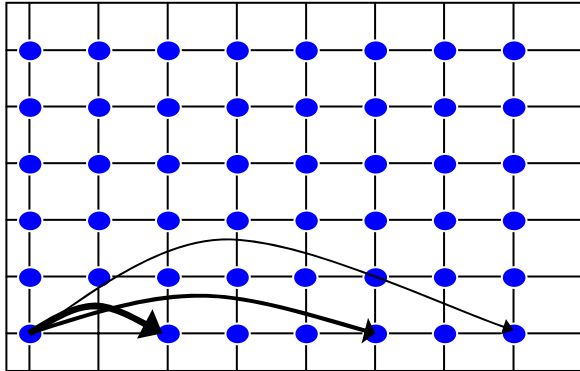
Population "continue" en réseaux

Chaque nœud du réseau correspond à 1 individu

2. Modèles démographiques de populations structurées :

2.4 Les modèle d'isolement par la distance

2 modèles en fonction du type de distribution des organismes dans le paysage :



Dans les 2 cas, homogénéité spatiale :

Dèmes de taille identique ou densité d'individus identique sur tout le réseau

Distribution de dispersion identique en tout point du réseau

-> peu de paramètres (2) :

σ^2 = Carré moyen de la distance de dispersion parent-descendant
= inverse de "force de l'isolement par la distance"

N = taille des sous-populations quand structure en dèmes
ou D = densité d'individu quand population continue

3. Analyses des modèles d'isolement par la distance :

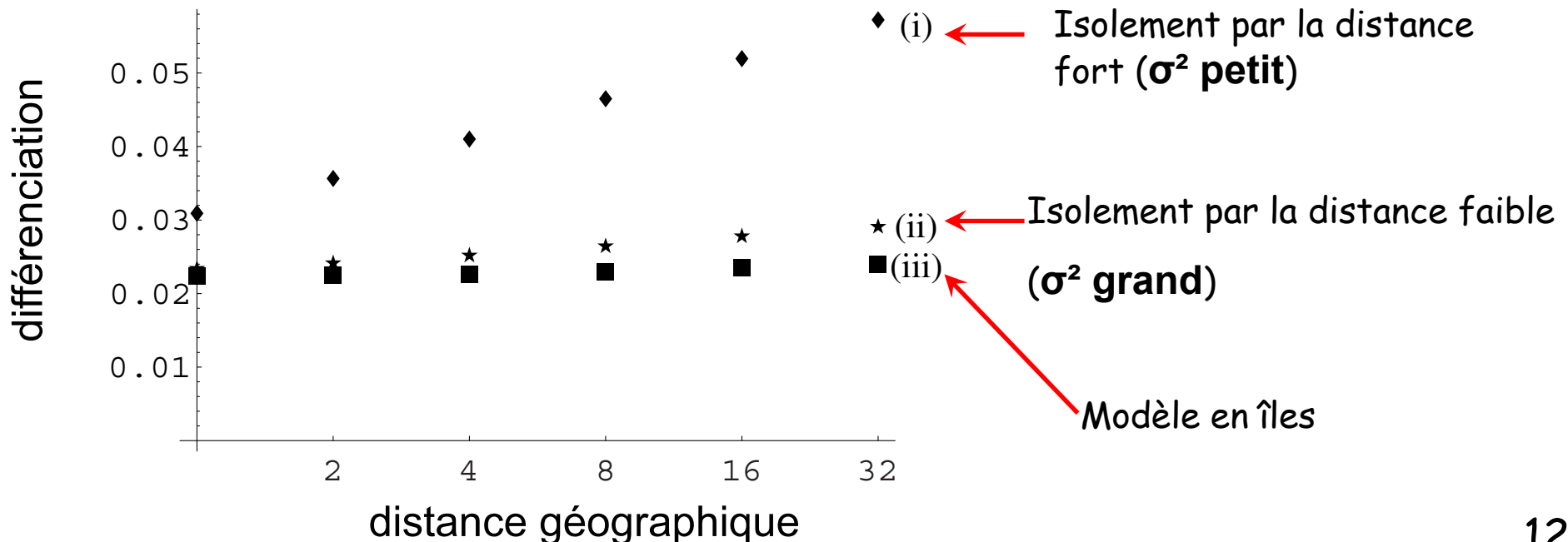
1. Résultats mathématiques

Analyse mathématique des modèles en dèmes et en populations continues :
(Wright 1943, 1946; Malécot 1948, 1950, 1967; Rousset 1997, 2000)

Ces modèles sont restés longtemps peu utilisés car leur analyse est assez complexe

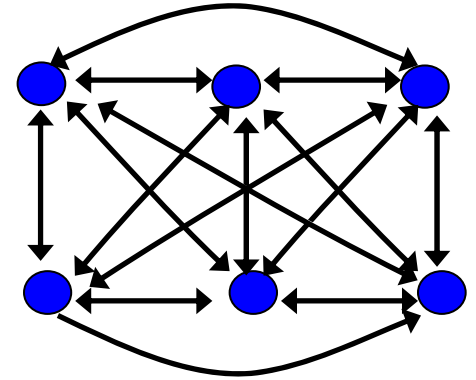
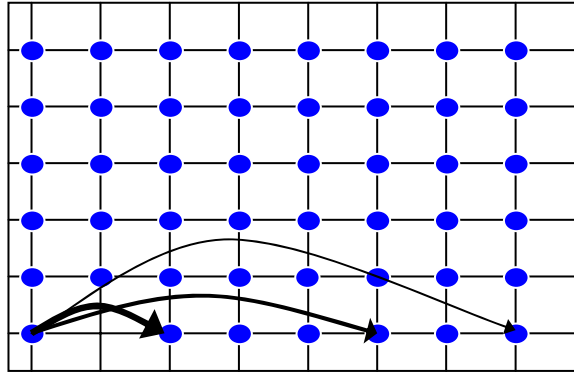
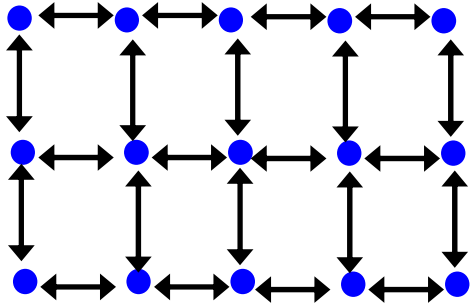
Caractéristique majeure

= la différenciation génétique augmente avec la distance géographique



3. Analyses des modèles d'isolement par la distance :

1. Résultats mathématiques



relationship between geographic and genetic distance :

Stepping stone >

IBD

> Island Model

$$\sigma^2 < 1$$

$$1 < \sigma^2 < 500$$

$$\sigma^2 = \infty$$

3. Analyses des modèles d'isolement par la distance :

1. Résultats mathématiques

Principaux résultats de l'analyse en terme de probabilité d'identité (Rousset 1997, 2000) (analyse similaire à celle présentée pour le modèle en île par R. Vitalis):

On s'intéresse à :

$$a_r \equiv \frac{Q_1 - Q_r}{1 - Q_1} \approx \frac{F_{ST}}{1 - F_{ST}} \text{ calculé à une distance géographique } r$$

Avec

$$F_{ST} \equiv \frac{Q_1 - Q_2}{1 - Q_2} \text{ et } Q_2 \Leftrightarrow Q_r$$

!! Population continue : a_r calculé entre individus et non entre populations !!

-> Q_1 =intra individu et $Q_2=Q_r$ =entre 2 individus à une distance r

3. Analyses des modèles d'isolement par la distance :

1. Résultats mathématiques

Principaux résultats de l'analyse en terme de probabilité d'identité (Rousset 1997, 2000):

En une dimension (= modèle linéaire)

$$a_r \equiv \frac{Q_1 - Q_r}{1 - Q_1} \approx \frac{1 - e^{\frac{-\sqrt{2\mu r}}{\sigma}}}{4N\sigma\sqrt{2\mu}} + \frac{A_1}{4N\sigma}$$
$$\approx^{r \text{ et } \mu \text{ petit}} \frac{r}{4N\sigma^2} + \frac{A_1}{4N\sigma} \approx^{N \rightarrow D} \frac{r}{4D\sigma^2} + \frac{A'_1}{4D\sigma}$$

-> relation linéaire entre la différenciation et la distance à petite distance et à taux de mutation faibles

3. Analyses des modèles d'isolement par la distance :

1. Résultats mathématiques

Principaux résultats de l'analyse en terme de probabilité d'identité (Rousset 1997, 2000):

En deux dimension (= modèle en réseau)

$$a_r \equiv \frac{Q_1 - Q_r}{1 - Q_1} \approx \frac{-\ln(\sqrt{2\mu}) + K_0(\sqrt{2\mu} r \sigma) + 2\pi A_2}{4\pi N \sigma^2}$$
$$\approx_{r \text{ et } \mu \text{ petit}} \frac{\ln(r)}{4\pi N \sigma^2} + \frac{\ln(\sigma) - 0.116 + 2\pi A_2}{4\pi N \sigma^2}$$
$$\approx_{N \rightarrow D} \frac{\ln(r)}{4\pi D \sigma^2} + \frac{\ln(\sigma) - 0.116 + 2\pi A'_2}{4\pi D \sigma^2}$$

-> relation linéaire entre la différenciation et $\ln(\text{distance})$ à petite distance et à taux de mutation faibles

3. Analyses des modèles d'isolement par la distance :

2. La méthode de la régression

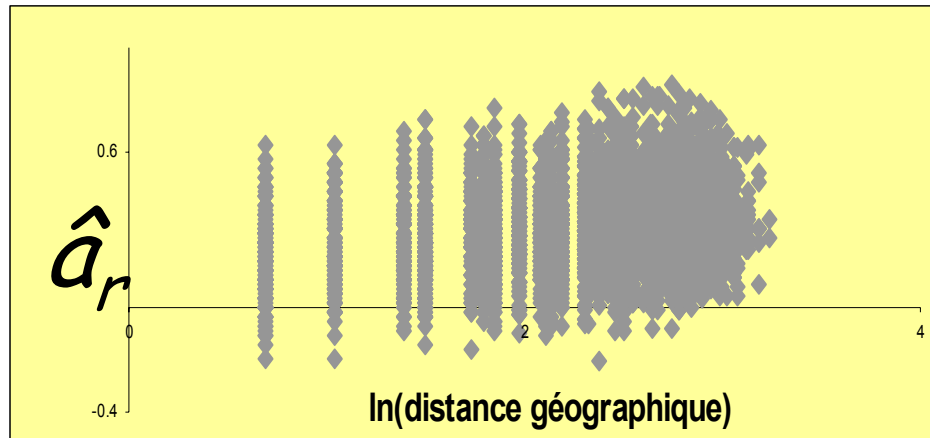
Une méthode indirecte pour estimer $D\sigma^2$ = Rousset 1997,2000

D = densité d'individus adultes

σ^2 = carré moyen de la distance de dispersion parent- descendant

- ✓ Isolement par la distance en population "continue"
- ✓ Utilise la relation linéaire entre différenciation et distance

✓ Régression



- \hat{a}_r calculé à partir des génotypes multilocus (ex: microsatellites) pour chaque paire d'individus ou de sous-populations

3. Analyses des modèles d'isolement par la distance :

2. La méthode de la régression

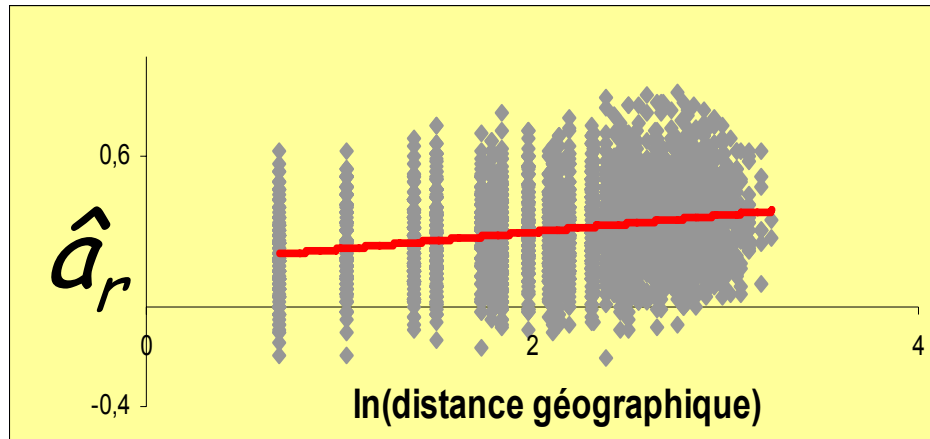
Une méthode indirecte pour estimer $D\sigma^2$ = Rousset 1997,2000

D = densité d'individus adultes

σ^2 = carré moyen de la distance de dispersion parent- descendant

- ✓ Isolement par la distance en population "continue"
- ✓ Utilise la relation linéaire entre différenciation et distance

✓ Régression



➡ 1/pente → estimateur de $4\pi D\sigma^2$

3. Analyses des modèles d'isolement par la distance :

2. La méthode de la régression

Une méthode indirecte pour estimer $D\sigma^2$ = Rousset 1997, 2000

D = densité d'individus adultes

σ^2 = carré moyen de la distance de dispersion parent- descendant

1. Estimation : $1/\text{pente} \rightarrow$ estimateur de $4\pi D\sigma^2$

2. Significativité :

a) **Test de Mantel** : test la corrélation (r^2 ou test de rang) entre 2 matrices (ici génétique et géographique) par permutation des lignes et colonnes d'une des matrices

-> Significatif si la corrélation initiale est supérieure à la corrélation sur les matrices permutées (=dans les 5% des matrices les plus corrélées)

b) **Bootstrap** : rééchantillonnage dans les locus (indépendants)

-> Intervalle de confiance (IC) pour la pente

-> si l'IC contient 0 pas significatif

**Mais bootstrap pas très puissant pour la pente...
donc utiliser plutôt le test de Mantel**

4. Tests et comparaisons avec démographie:

1. Comment tester une méthode

Deux grand principes pour tester une méthode d'estimation:

1. Tests par simulations :

a) **Données simulées sous le bon modèle** (i.e. celui de l'analyse) : on regarde si les valeurs trouvées par la méthode sont proches des valeurs spécifiées dans les simulations

-> **Précision** dans le meilleur des cas

b) **Données simulées sous modèle avec violations des hypothèses du modèle d'analyse :**

-> **Robustesse** des estimations/hypothèses du modèle d'analyse

2. Test sur données réelles pour lesquelles on connaît l'attendu :

Dans notre cas, comparaisons avec données démographiques (suivi, capture/marquage/recapture)

4. Tests et comparaisons avec démographie :

2. Tests par simulations : Méthode

(1) Choix des caractéristiques mutationnelles et démographiques



(2) Simulations 1000 répétitions



(3) Analyse des 1000 échantillons de génotypes à 10 locus
→ 1000 estimations de la pente de régression

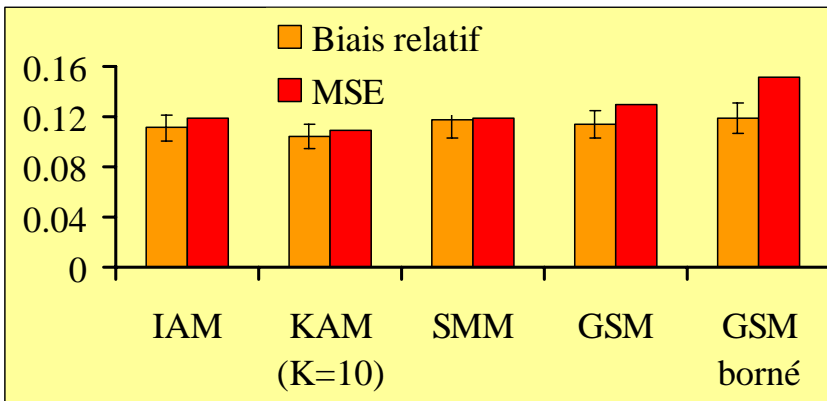


(4) Comparaisons avec la valeur « attendue » de la pente :

- Biais relatif
- Moyenne du carré de l'erreur relative (MSE)
- Estimations dans [$D\sigma^2_{att.}/2$; $2 * D\sigma^2_{att.}$] (= bonne estimation)

4. Tests et comparaisons avec démographie :

3. Tests par simulations: Résultats (Leblois et al. 2003 MBE)

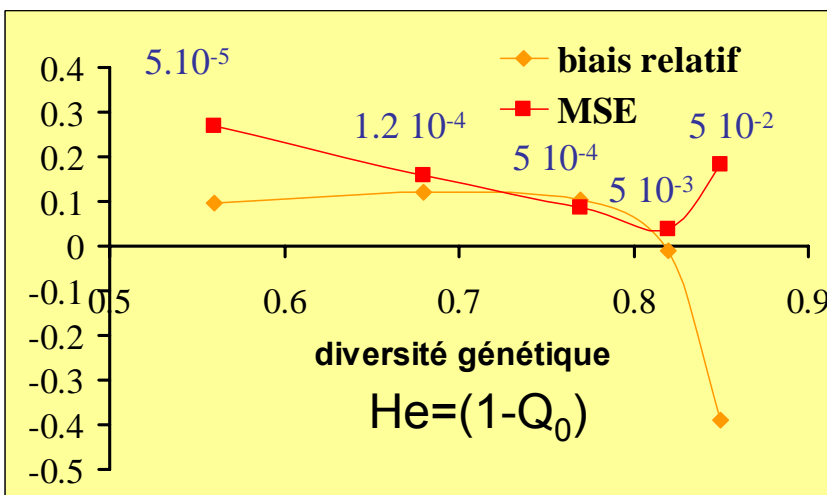


Influence des processus mutationnel :

Méthode fondés sur identité par états (IBD)

Marqueurs non IBD (ex: microsats en SMM)

Simulation -> méthode robuste car peu de changement de biais et MSE avec les différents modèles



Influence du taux de mutation (diversité) :

Méthode : hyp. μ faibles mais nécessite de la variabilité

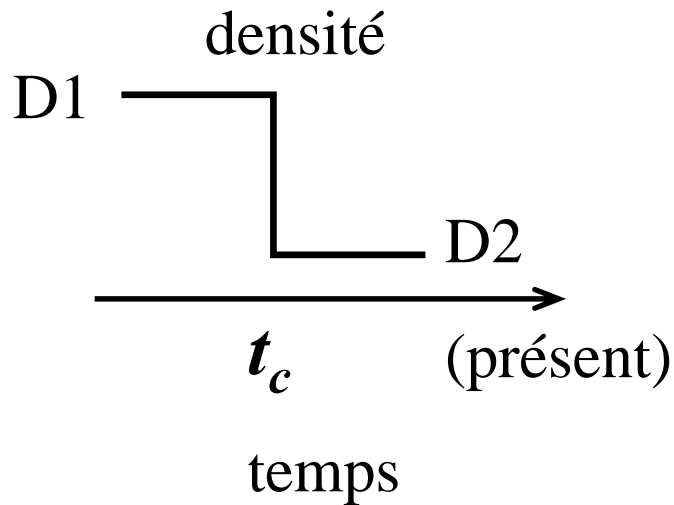
Simulation -> méthode plus précise avec diversité assez forte (0.7-0.8)

Biais fort que $div > 0.85$ (= influence des taux de mutation fort)

Microsat bon marqueur car $\mu \approx 10^{-4} \text{ } 10^{-3}$

4. Tests et comparaisons avec démographie :

3. Tests par simulations : Résultats (Leblois et al. 2004 Genetics)

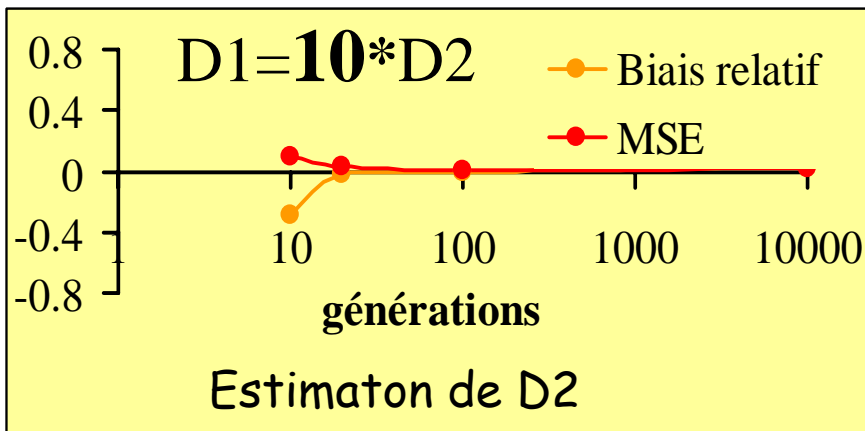


Influence des processus démographiques passés:

Ex : baisse de densité (bottleneck)

Test par simulations \rightarrow méthode robuste car peu d'influence des densités passées

Autres tests : augmentation de la densité, expansion spatiale, densité spatialement hétérogène

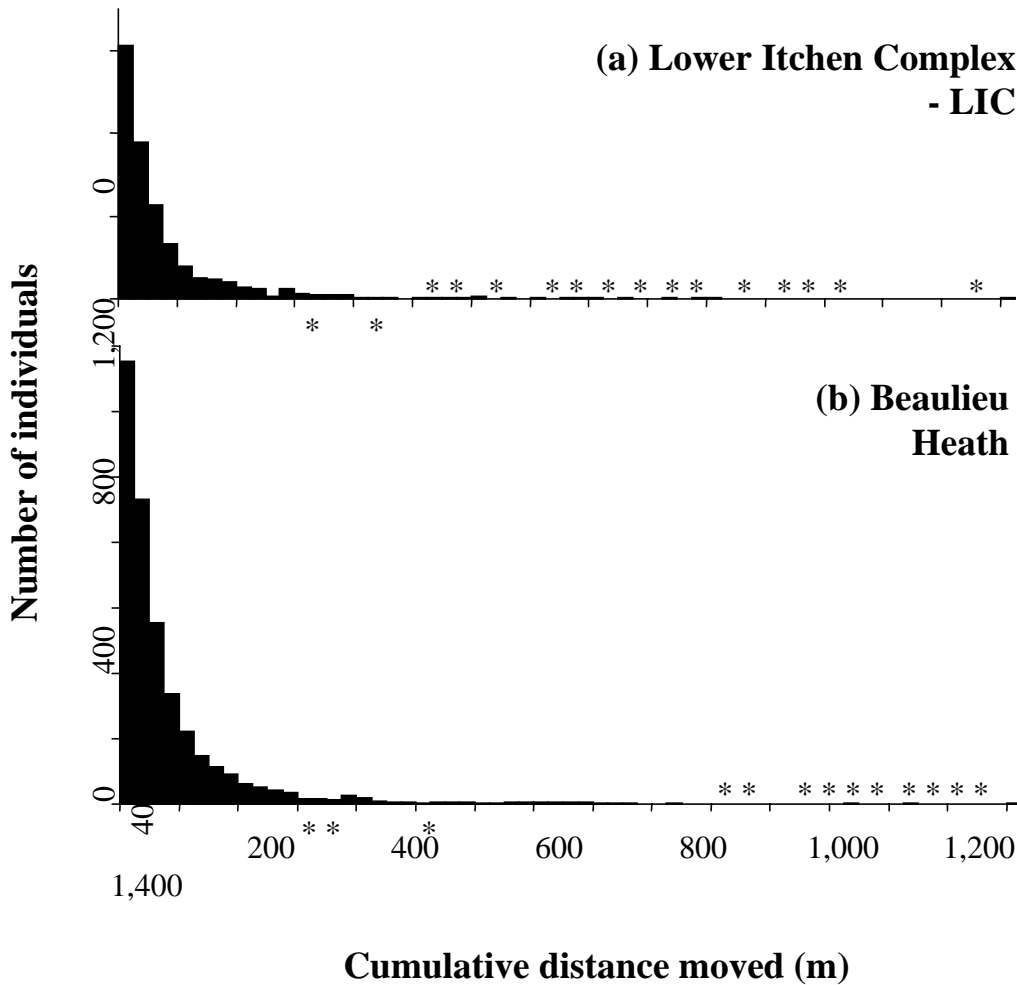


Globalement la méthode est robuste au hétérogénéités spatiales et temporelles : estime bien les paramètres actuels et locaux des populations échantillonnées

4. Tests et comparaisons avec démographie :

4. Comparaisons avec données démographiques

- Exemple des libellules (Watt et al. 2007 Mol.Ecol.)



Données démographique
(Capture marquage recapture)

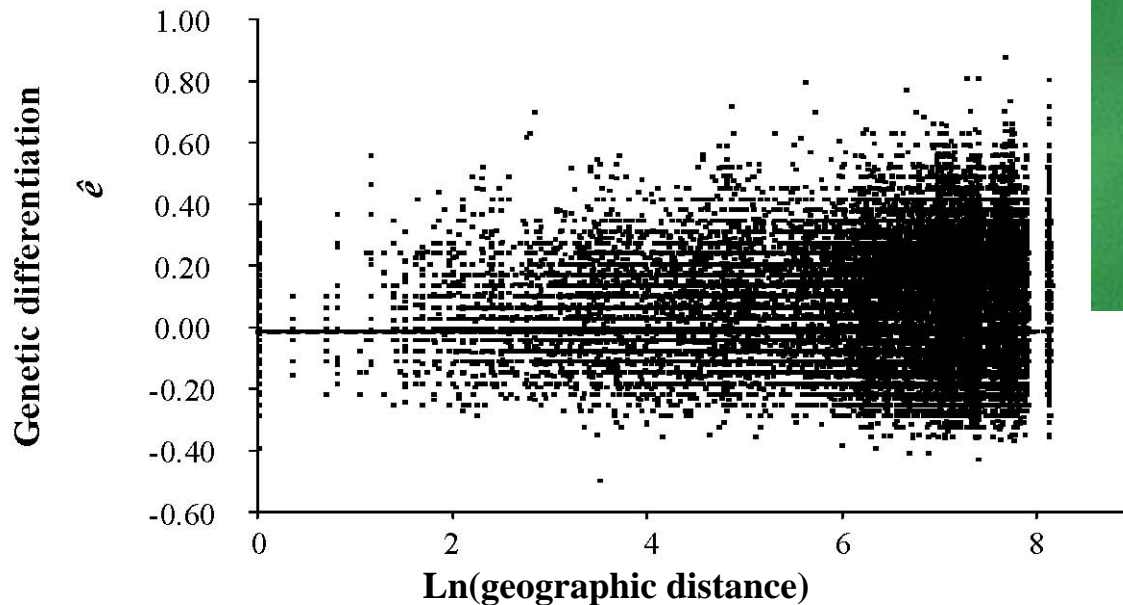
-> Estimation distribution de
dispersion et densité



4. Tests et comparaisons avec démographie :

4. Comparaisons avec données démographiques

- Exemple des libellules



Données génétiques : 13 marqueurs microsatellites

-> Estimation indirecte du produit $D\sigma^2$

4. Tests et comparaisons avec démographie :

4. Comparaisons avec données démographiques

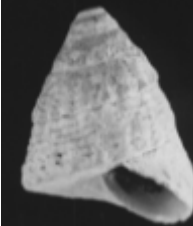
- Exemple des libellules

Très bonne concordance entre
démographie et génétique



	Estimation de $D\sigma^2$	
	Direct (démo)	Indirect (génét)
Site 1	277	222
Site 2	249	259
Site 3	555	606

4. More comparisons between genetic and demographic inference of $D\sigma^2$:



	Direct (Demography)	Indirect (genetic)
American Marten (<i>Martes americana</i>)	7.5	3.8
Kangaroo rats (<i>Dipodomys</i>)	1.43	2.58
intertidal snails (<i>Bembicium vittatum</i>)	2.4	3.6
Forest lizards (<i>Gnypetoscincus queenslandiae</i>)	11.5	5.5
Humans in the rainforest (Papous)	29.3	21.1
Legumin (<i>Chamaecrista fasciculata</i>)	9.6	13.9

4. Tests et comparaisons avec démographie :

4. Comparaisons avec données démographiques

	Direct (Demography)	Indirect (genetic)
American Marten (<i>Martes americana</i>)	7.5	3.8
Kangaroo rats (<i>Dipodomys</i>)	1.43	2.58
intertidal snails (<i>Bembicium vittatum</i>)	2.4	3.6
Forest lizards (<i>Gnypetoscincus queenslandiae</i>)	11.5	5.5
Humans in the rainforest (Papous)	29.3	21.1
Legumin (<i>Chamaecrista fasciculata</i>)	9.6	13.9

1ere fois que l'on a concordance systématique entre estimation directes et indirectes

-> validation du modèle d'isolement par la distance et de la méthode de la régression

5. Pourquoi ca marche bien

1. Les critiques des estimation indirectes

Principaux points problématiques de l'estimation de paramètres démographiques à partir de données génétiques (Hasting et Harrison 1994, Koenig et al. 1996, Slatkin 1994) :

- Modèles peu réalistes, mauvaise description de la dispersion (modèle en îles)
 - Hypothèses de stabilité démographiques dans le temps et dans l'espace
 - Hypothèses associées aux taux de mutation et processus mutationnels
 - Hypothèses de neutralité des marqueurs utilisés
-
- Ainsi Whitlock & McCauley 1999 (Heredity) :

Indirect measure of gene flow and migration : $F_{st} \approx 1/(1+4Nm)$

??? Pourquoi ca marche bien en isolement par la distance avec la méthode la régression ???

5. Pourquoi ca marche bien

2. Les différents facteurs en jeu

- Le modèle d'isolement par la distance = modèle réaliste
 - Bonne description de la dispersion (distribution avec dispersion localisé)
 - Possibilité de prendre en compte des population continue
 - > pas de nécessité de définir a priori des sous-population (svt problématique)
- La méthode d'estimation
 - Relation F-statistiques/distance plus facilement interprétable que F-stat seules
 - Pas d'hypothèses sur la forme de la distribution (OK avec leptokurtic)
 - Échelle d'étude locale (échantillonnage a petite distance) :
 - > Moins d'hétérogénéité spatiale à petite échelle (démographique et sélective)
- Les marqueurs utilisés : microsatellites
 - Bonne variabilité (donc beaucoup d'information)
 - "+ neutre" que des séquences codantes

5. Pourquoi ça marche bien

3. Interprétation à l'aide des temps de coalescence

Les probabilité d'identité de paires de gènes peuvent être exprimées en fonction des distribution des temps de coalescence pour les paires de gènes considérés selon la forme :

$$Q_i = \sum_{t=0}^{\infty} C_{i,t} (1-\mu)^{2t}$$

Des paramètres tels que

$$F_{IS} \equiv \frac{Q_0 - Q_1}{1 - Q_1}, F_{ST} \equiv \frac{Q_1 - Q_2}{1 - Q_2}, \text{ ou } a_r \equiv \frac{Q_1 - Q_r}{1 - Q_1}$$

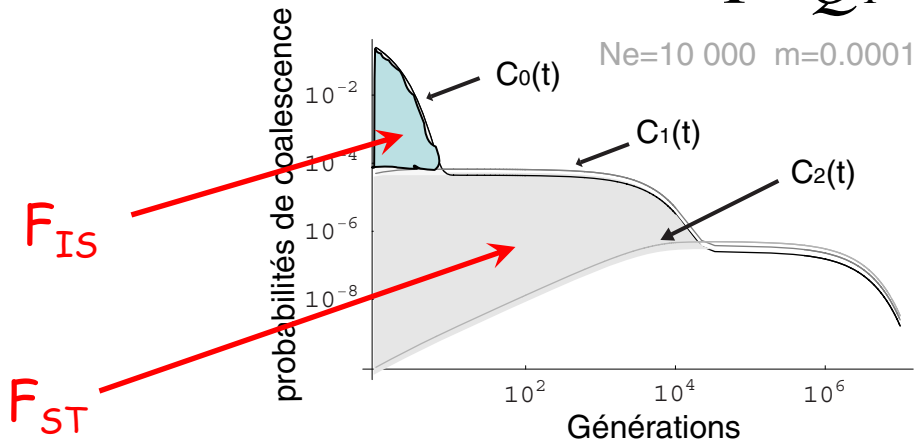
dépendent donc de la différence entre les distribution des temps de coalescence des paires de gènes considérées (ex : C_1 et C_2 pour a_r)

$$a_r \propto (C_1 - C_2)$$

5. Pourquoi ça marche bien

3. Interprétation à l'aide des temps de coalescence

$$Q_i = \sum_{t=0}^{\infty} C_{i,t} (1-\mu)^{2t}$$
$$F_{IS} \equiv \frac{Q_0 - Q_1}{1 - Q_1}, F_{ST} \equiv \frac{Q_1 - Q_2}{1 - Q_2}, \text{ ou } a_r \equiv \frac{Q_1 - Q_r}{1 - Q_1}$$



Ex : modèle en îles

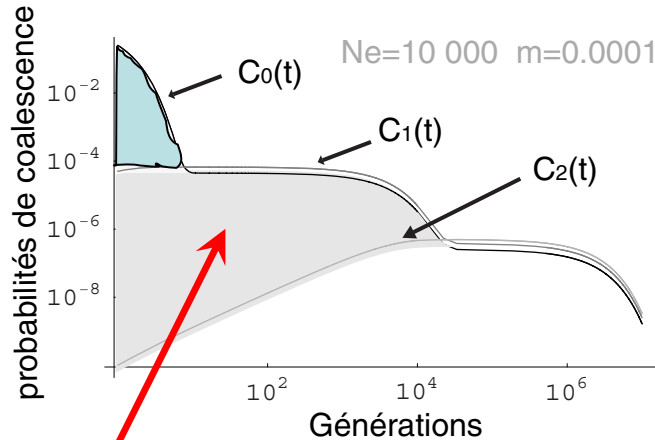
F_{IS} : temps récents donc peu influencé par mutation, etc...

F_{ST} : temps plus anciens donc plus influencé par processus anciens (mutation, hétérogénéités temporelles,...)

Ces courbes varient selon les tailles de pops et les taux de migration et m

5. Pourquoi ca marche bien

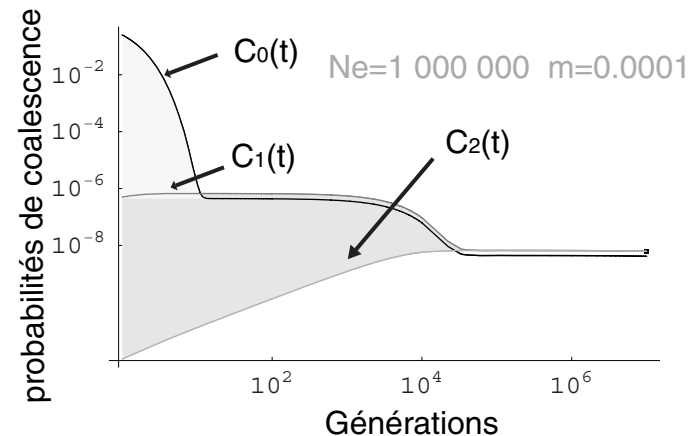
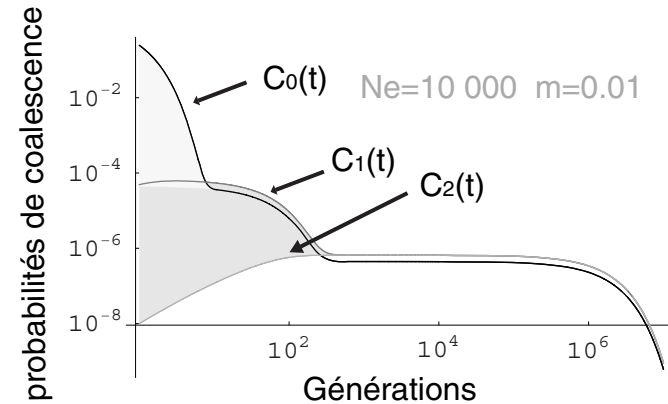
3. Interprétation à l'aide des temps de coalescence



Ex : modèle en îles

Ces courbes (surtout C_1 et C_2) varient selon les tailles de pops et les taux de migration et m

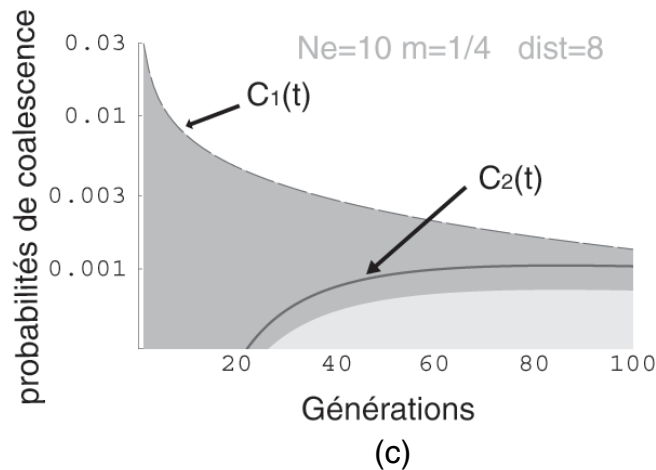
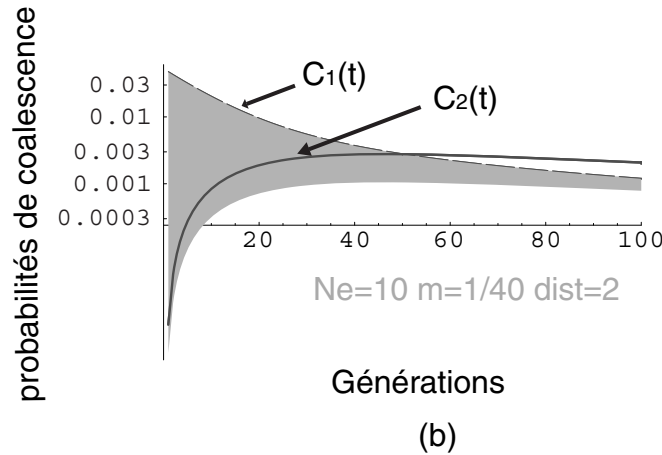
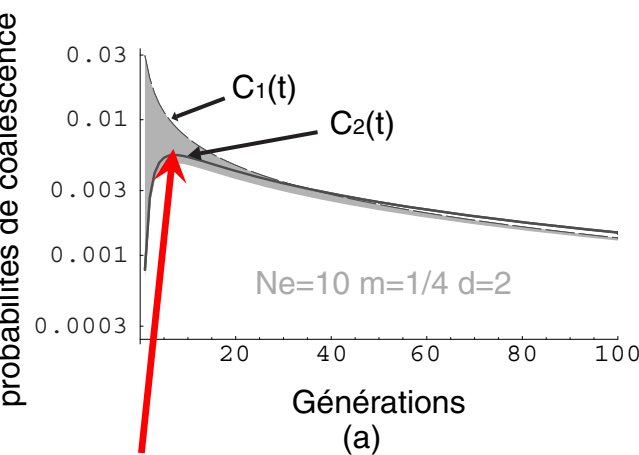
Quand $m \downarrow$ et $N \uparrow$, C_1 - C_2 va de plus en plus vers le passé, donc plus d'influence des processus anciens



5. Pourquoi ça marche bien

3. Interprétation à l'aide des temps de coalescence

En isolement par la distance :



Même influence des
tailles de pops (non
montré)

Même influence des
taux de migration

Influence
supplémentaire de
l'échelle
d'échantillonnage (dist.
max. dans échantillon)

-> plus l'échantillon est à grande
échelle plus les processus anciens
ont de l'influence

5. Pourquoi ca marche bien

3. Interprétation à l'aide des temps de coalescence

Les représentation des temps de coalescence permet de mieux comprendre pourquoi la méthode de la régression en isolement par la distance marche bien :

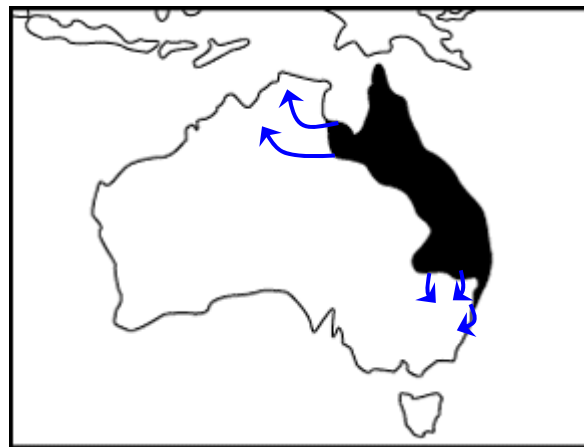
- Petites populations (modèle continu 1 pop = 1individu)
- Fort taux de migration (> 0.2)
- Échantillonnage a très petite échelle

-> faible influence des processus passés

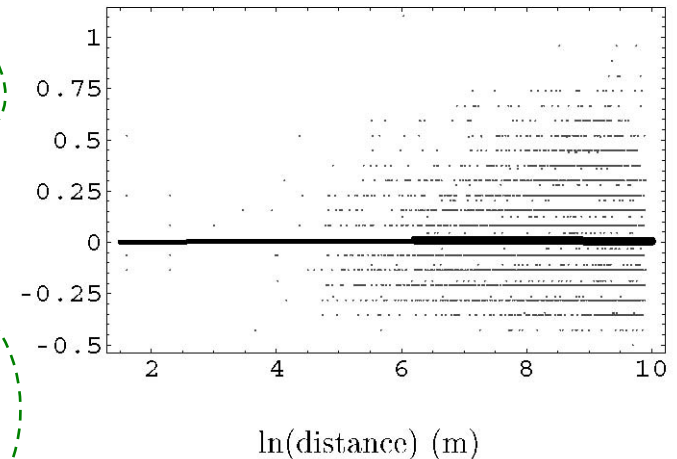
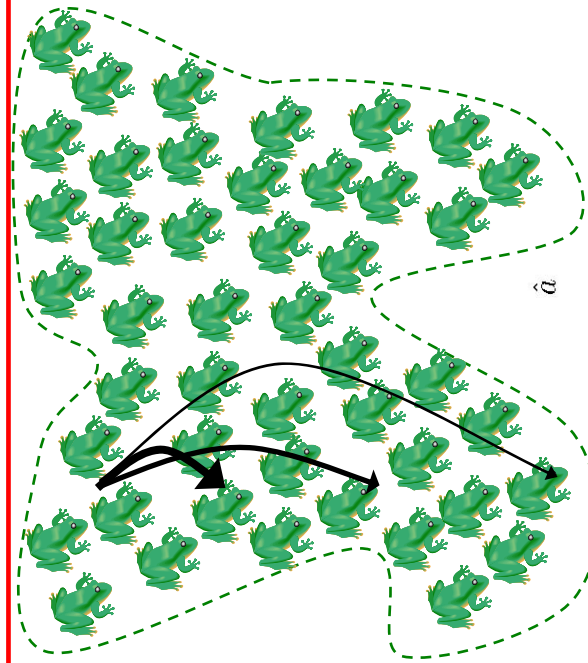
Importance du modèle, de la méthode, des marqueurs et de l'échantillonnage

6. Exemple d'applications:

1. Biologie des espèces invasives



➤ Isolement par la distance



- Pas d'isolement par la distance significatif dans les pops envahissantes -> **Forte dispersion lors de l'invasion**
- Peut expliquer la colonisation rapide de la côte Est de l'Australie (50km par an), dispersion par l'homme pas forcément en cause

6. Implications to real data set analyses:

2. Conservation genetics of forest skinks

Molecular Ecology (2004) 13, 259–269

doi: 10.1046/j.1365-294X.2003.02056.x

Limited effect of anthropogenic habitat fragmentation on molecular diversity in a rain forest skink, *Gnypetoscincus queenslandiae*

JOANNA SUMNER,*TIM JESSOP,†DAVID PAETKAU‡ and CRAIG MORITZ§

*Department of Zoology and Entomology and the Rainforest CRC, University of Queensland, St Lucia, Qld 4072, Australia, †Center for Reproduction of Endangered Species, Zoological Society of San Diego, San Diego, CA, 92112, USA, ‡Wildlife Genetics International, Box 274, Nelson, BC, V1L 5P9, Canada, §Museum of Vertebrate Zoology, University of California, Berkeley, CA, 94720, USA

Documented habitat reduction, 10 skink generations ago
→ reduced genetic diversity ?

No decrease in Na, He detected with 9 microsatellites...no sign of bottlenecks

Forest skink



6. Implications to real data set analyses:

2. Conservation genetics of forest skinks

Documented habitat reduction, 10 skink generations ago

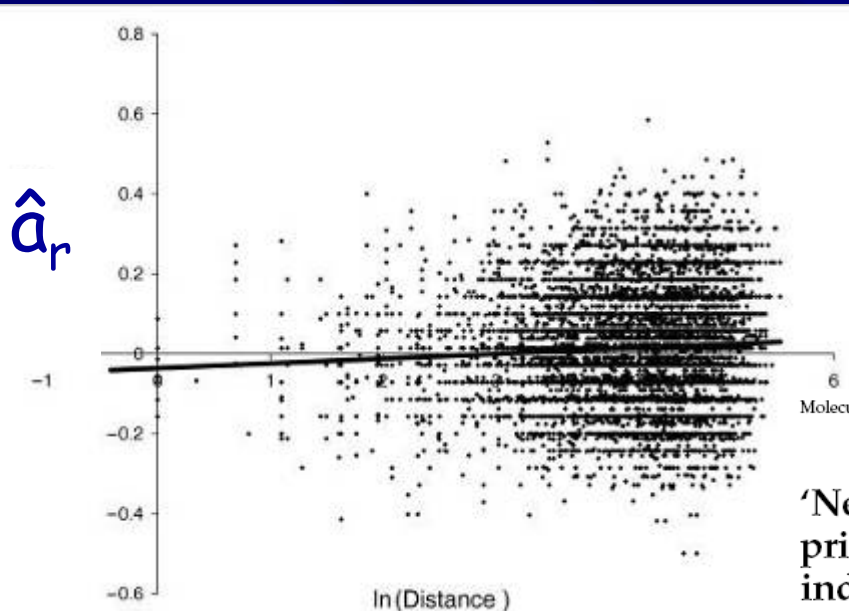
→ reduced genetic diversity ?

No decrease in N_a , H_e detected with 9 microsatellites...no signs of bottlenecks

Forest skink



But strong isolation by distance $D\sigma^2 = 5.5-11.5$



Molecular Ecology (2001) 10, 1917–1927

‘Neighbourhood’ size, dispersal and density estimates in the prickly forest skink (*Gnypetoscincus queenslandiae*) using individual genetic and demographic methods

J. SUMNER,*† F. ROUSSET,‡ A. ESTOUP*§ and C. MORITZ*

*Department of Zoology and Entomology, University of Queensland, Qld 4072, Australia, †Laboratoire Génétique et Environnement, CNRS-UMR 5554, 34095 Montpellier, France, §Centre de Biologie et de Gestion des Populations, INRA, 34980 Montferrier/Lez, France

6. Implications to real data set analyses:

2. Conservation genetics of forest skinks

Documented habitat reduction, 10 skink generations ago

→ reduced genetic diversity ?

No decrease in Na, He detected with 9 microsatellites...no signs of bottlenecks

Forest skink



Molecular Ecology (2006) 15, 3601–3615

doi: 10.1111/j.1365-294X.2006.03046.x

Genetics of recent habitat contraction and reduction in population size: does isolation by distance matter?

RAPHAEL LEBLOIS,^{*,†‡} ARNAUD ESTOUP[†] and REJANE STREIFF[†]

^{}Laboratoire Génétique et Environnement, CNRS-UMR 5554, 34095 Montpellier, France, [†]Centre de Biologie et de Gestion des Populations, INRA, Campus International de Baillarguet, CS 30016, 34988 Montferrier sur Lez cedex, France*

6. Implications to real data set analyses:

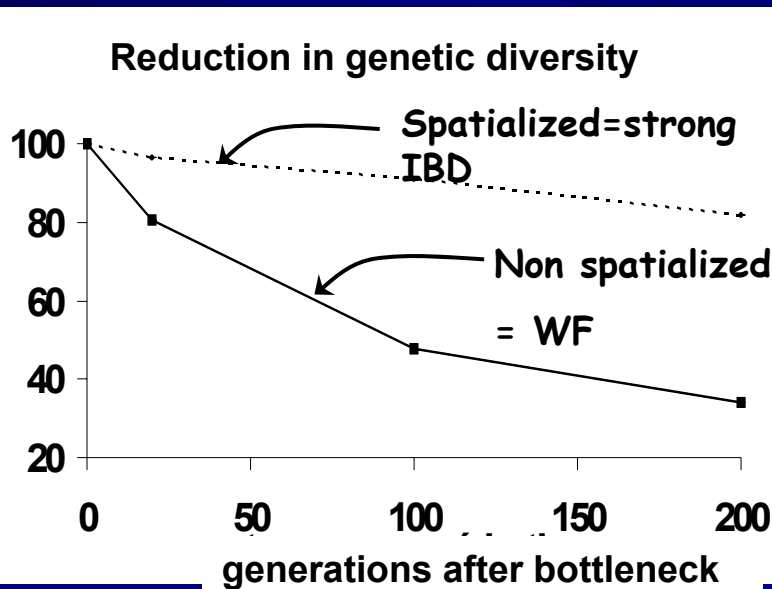
2. Conservation genetics of forest skinks

Documented habitat reduction, 10 skink generations ago

→ reduced genetic diversity ?

No decrease in N_a , H_e detected with 9 microsatellites...no signs of bottlenecks

Effect of spatial structure (IBD) :



Forest skink



➤ Genetic diversity is only weakly reduced under IBD after a bottleneck

+ No bottleneck detection under IBD (BOTTLENECK Cornuet & Luikart, M Garza & Williamson 96)

Importance of the spatial features and localized dispersal

6. Implications to real data set analyses:

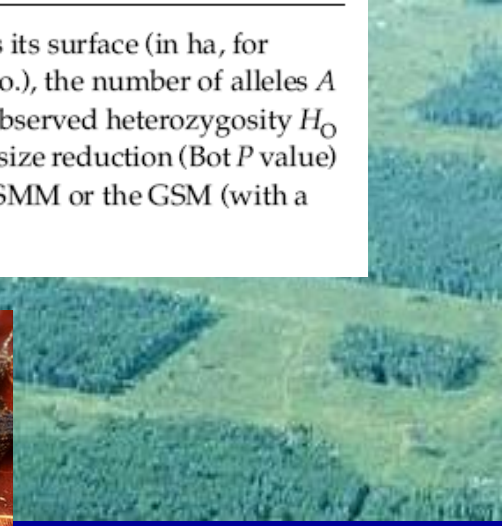
2. Conservation genetics of forest skinks

Table 2 Summary of skink population data set

Population name	Type	Area	Size	No.	A	H_E	H_O	M	Bot P value		Exp P value	
									SMM	GSM	SMM	GSM
Souita Falls (F1)	F	2.0	101–384	25	7.35	0.70	0.68	0.458	0.99	0.85	0.007**	0.18
Maalan Road (F2)	F	2.46	124–472	42	7.44	0.63	0.66	0.549	1.00	0.99	0.001**	0.01**
Waltham (F3)	F	2.64	133–507	27	7.75	0.68	0.67	0.496	0.99	0.90	0.007**	0.13
Pat Daley Park (F4)	F	5.96	300–1144	18	7.39	0.63	0.63	0.406	1.00	0.99	0.002**	0.01**
Nose Ring (F5)	F	24.19	1217–4645	29	9.31	0.70	0.68	0.524	1.00	0.99	0.005**	0.007**
Whiteing Road (F6)	F	36.31	1826–6972	30	8.77	0.71	0.69	0.522	1.00	0.99	0.005**	0.01**
Millaa Millaa Falls (F7)	F	65.06	3273–12497	27	7.88	0.66	0.65	0.465	1.00	1.00	0.001**	0.003**
Brotherton(C1)	C	NA	>> 40 000	28	8.26	0.69	0.68	0.523	0.99	0.98	0.019**	0.065
Cross-eye(C2)	C	NA	>> 40 000	30	8.00	0.69	0.70	0.526	1.00	1.00	0.001**	0.005**
Mount Father Clancy(C3)	C	NA	>> 40 000	32	8.77	0.69	0.72	0.575	1.00	1.00	0.002**	0.005**
Reynolds(C4)	C	NA	>> 40 000	28	9.18	0.66	0.71	0.566	1.00	1.00	0.002**	0.002**
Massey Creek(C5)	C	NA	>> 40 000	94	7.95	0.62	0.62	0.580	1.00	1.00	0.001**	0.002**

For each population, habitat type (F, fragmented and C, continuous, i.e. nonfragmented) is reported, as well as its surface (in ha, for fragments only), its approximate size in term of number of individuals, the number of individuals sampled (No.), the number of alleles A (adjusted for a sample size of 30 individuals using Ewens 1972's sampling formula), the gene diversity H_E , the observed heterozygosity H_O and the value of M statistics. The probability of rejecting the hypothesis of equilibrium in favour of a population size reduction (Bot P value) or expansion (Exp P value) was computed using the software BOTTLENECK (Piry *et al.* 1999) and assuming the SMM or the GSM (with a variance of 0.36) as mutation model.

Importance of the spatial features and localized dispersal



6. Implications to real data set analyses:

3. Spatialized clustering : Geneland

Copyright © 2005 by the Genetics Society of America
DOI: 10.1534/genetics.104.033803

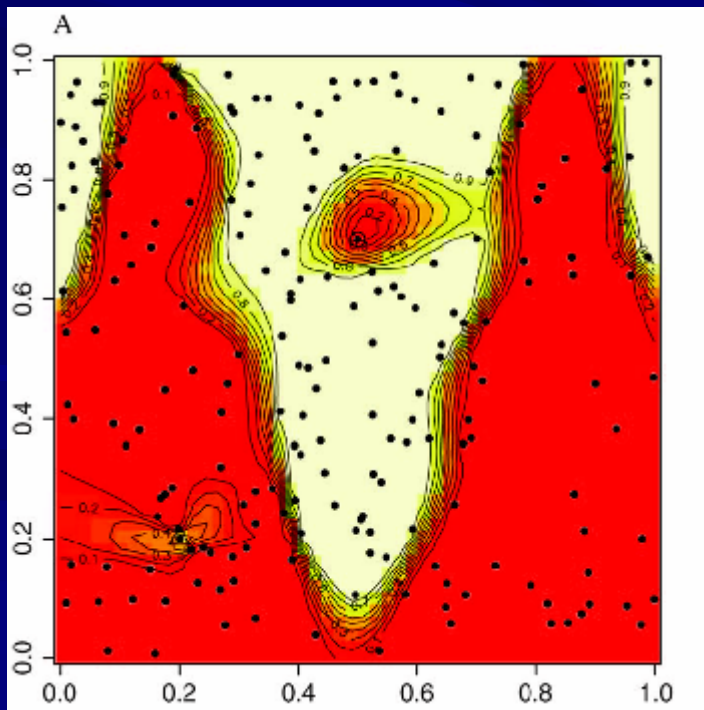
A Spatial Statistical Model for Landscape Genetics

Gilles Guillot,^{*,1} Arnaud Estoup,[†] Frédéric Mortier[‡] and Jean François Cosson[§]

**Unité de Mathématiques et Informatique Appliquées, INRA-INAPG-ENGREF, Paris, France 75231, †Centre de Biologie et de Gestion des Populations, INRA-ENSAM-IRD-CIRAD, Montpellier, France F-34198, ‡CIRAD, Département Forêt, Montpellier, France F-34398 and §Centre de Biologie et de Gestion des Populations, INRA-ENSAM-IRD-CIRAD, Montpellier, France F-34988*

Manuscript received July 21, 2004

Accepted for publication October 31, 2004



Spatial delimitation of populations from individual genotypes with geographic coordinates (cf cours assignation)
= localize genetic discontinuities in space

6. Implications to real data set analyses:

3. Spatialized clustering : Geneland

Method based on Wright-Fisher population assumptions :

- Minimize HW & linkage disequilibrium

Simulation test : Influence of isolation by distance

True population delimitation



Simulated under WF



Simulated under IBD



Wrong delimitation of populations under IBD

7. Conclusions

Isolation by distance (IBD) seems :

- to be a quite realistic model for lots of plant and animal species
 - to have good characteristics for the estimation of local and current dispersal and density
 - To have strong influence on various aspects of polymorphism data analyses
- > seems important to take into account IBD for the future development of spatial population genetics