

Article original

# Structure spatiale génétique et niveau de diversité intrapopulation chez les arbres forestiers

Agnès Doligez, Claire Baril, Hélène I. Joly\*

Cirad-Forêt, U. R. Diversité et amélioration génétique, Baillarguet,  
BP 5035, 34032 Montpellier cedex 01, France

**Abstract – Spatial genetic structure and diversity within forest tree populations.** An isolation by distance model is presented, which is better fitted to forest tree populations than those available to date. This model was developed in order to understand which factors might determine genetic diversity and spatial structure within populations of these species, for in situ management of their genetic resources. It allows a nonuniform spatial distribution of individuals, as well as some selfing. This model was studied by means of simulations, using a complete factorial design with 32 highly contrasted cases, which resulted from the combination of different values of the following input variables: global density, seed and pollen dispersal distances, generation overlap and overdominant selection. Our results show that much of the variation in spatial genetic structure and inbreeding at the individual level can be explained by variations in clumping and outcrossing rate. Inbreeding at the population level is mainly determined by the degree of generation overlap. As a general conclusion, all the input variables studied here do affect the genetic characteristics of simulated populations, and most interaction effects between these variables are highly significant. These results suggest that the explanation of differences in the level and/or structure of within-population genetic diversity between tree species could presumably be improved by using additional classification factors such as density, the level of clumping, generation overlapping, and/or selection, as well as by taking into account the interaction effects between classification factors. © Inra/Elsevier, Paris  
**isolation by distance / forest tree / model / genetic diversity / spatial genetic structure**

**Résumé –** Nous avons développé un modèle d'isolement par la distance adapté aux populations d'arbres forestiers, afin de comprendre les facteurs qui déterminent la structure spatiale génétique et le niveau de diversité intrapopulation chez ces espèces, en vue de la gestion in situ de leurs ressources génétiques. Il permet de prendre en compte une répartition spatiale non uniforme des individus. On a étudié par

---

\* Correspondance et tirés à part

simulation, en utilisant un plan factoriel complet, le comportement de ce modèle pour 32 combinaisons de valeurs fortement contrastées de la densité globale, des distances de dissémination des graines et de pollinisation, du chevauchement des générations, et de la sélection en faveur des hétérozygotes. Les variations d'agrégativité et du taux d'allofécondation permettent d'expliquer une grande partie des variations de la structure spatiale génétique et de la consanguinité individuelle obtenues. La consanguinité de la population est principalement déterminée par l'ampleur du chevauchement des générations. Toutes les variables d'entrée étudiées ont un effet sur l'état génétique des populations simulées, et la plupart des effets d'interaction sont très fortement significatifs. Ces résultats suggèrent qu'en utilisant comme facteurs de classification supplémentaires la densité, l'agrégativité, le chevauchement des générations, la sélection, ainsi que les interactions entre facteurs de classification, il serait possible d'améliorer l'explication des différences génétiques observées entre espèces d'arbres. © Inra/Elsevier, Paris

**isolement par la distance / arbre forestier / modèle / diversité génétique / structure spatiale génétique**

## 1. INTRODUCTION

La conservation in situ des ressources génétiques forestières nécessite une bonne compréhension des facteurs qui déterminent le niveau et la structure de la diversité génétique intraspécifique chez les espèces concernées. Or le nombre d'espèces d'arbres forestiers pour lesquelles une gestion raisonnée des ressources génétiques est souhaitable est très élevé, si l'on considère aussi bien les espèces tempérées ou boréales, que les espèces tropicales. Il n'est donc pas possible d'effectuer une étude de la diversité génétique de chacune de ces espèces avant de commencer à établir des règles de gestion de leurs ressources génétiques. L'utilisation de modèles de génétique des populations adaptés aux arbres forestiers peut aider à mieux comprendre les facteurs et les mécanismes qui influencent la diversité génétique et son organisation. De tels modèles sont particulièrement intéressants pour caractériser le lien entre la diversité génétique et certaines caractéristiques biologiques, écologiques et/ou démographiques plus facilement mesurables que les paramètres génétiques, et par conséquent pour permettre de décider de règles de gestion in situ adaptées à chaque espèce.

Nous présentons ici un modèle stochastique d'isolement par la distance (population continue structurée) mieux adapté à un large spectre d'espèces forestières que les modèles développés jusqu'à présent, notamment par Rohlf et Schnell (1971), Turner et al. (1982), Sokal et Wartenberg (1983), Bos et van der Haring (1988), Sokal et al. (1989), Ohsawa et al. (1993), et Epperson (1990, 1995a, 1995b), et nous quantifions la diversité génétique intrapopulation et sa structuration spatiale (que l'on appellera structure spatiale génétique) obtenues sous les hypothèses de ce modèle dans un grand nombre de cas contrastés. La principale caractéristique de ce modèle par rapport aux précédents est la possibilité de répartition spatiale non uniforme des individus de la population. En effet, les auteurs qui se sont attachés à quantifier la structure spatiale génétique intrapopulation (à l'aide de statistiques d'autocorrélations spatiales), ont tous supposé que les individus de la population étaient uniformément répartis dans l'espace. Par ailleurs, d'autres auteurs ont simulé des populations à répartition non uniforme (Slatkin et Barton, 1989; Ronfort et Couvet, 1995;

Kawata, 1995), mais ils n'ont pas cherché à décrire finement la structure spatiale génétique obtenue (seulement à l'aide de  $FST$  pour un seul découpage arbitraire des populations). Notre modèle présente donc l'avantage de permettre l'étude des relations entre la répartition spatiale des individus et la structure spatiale génétique. Il permet également d'étudier l'effet d'un système mixte de reproduction, du chevauchement des générations, d'une sélection de type superdominance à certains loci, et de distances de dissémination des graines supérieures aux distances de pollinisation.

## 2. MATÉRIEL ET MÉTHODES

Les hypothèses du modèle sont les suivantes : on suppose que la population est constituée de  $N = 10\,000$  individus diploïdes, monoïques, tous reproducteurs, et que sa taille est constante dans le temps. On connaît l'emplacement de chaque individu, son génotype à un locus neutre diallélique et son génotype à un locus diallélique indépendant auquel s'exerce une sélection de type superdominance symétrique, avec un coefficient de sélection  $S$  (soit une valeur sélective de 1 pour les hétérozygotes et de  $1 - S$  pour les homozygotes). Le modèle de superdominance ne semble pas être le modèle de déterminisme de la valeur sélective le mieux validé par les quelques tests effectués jusqu'à présent à l'aide de données réelles (Barret et Charlesworth, 1991). Toutefois, c'est le seul, parmi les modèles actuellement disponibles, qui permette d'obtenir des valeurs négatives de l'indice de fixation après sélection (Charlesworth et al., 1990). Or de telles valeurs sont souvent observées dans les populations naturelles d'arbres forestiers. C'est pourquoi on a utilisé ici ce type de modèle, sachant qu'il sera nécessaire par la suite de compléter la présente étude en le remplaçant par des modèles alternatifs (de type mutation-dominance) tendant au contraire à diminuer la variabilité génétique.

On suppose que la population est divisée en  $C$  classes d'âges de mêmes effectifs (en l'absence de générations chevauchantes,  $C = 1$ ). Pour créer chaque nouvel individu, on tire au hasard selon une loi uniforme une mère parmi l'ensemble des individus (non nouvellement créés) de la population. Puis on tire au hasard selon une loi uniforme un emplacement dans le voisinage de dissémination de cette mère (i.e. sur la surface du disque de rayon  $D$  centré sur cet individu). Si cet emplacement n'est pas encore occupé par un nouvel individu ou par un individu destiné à survivre, on tire ensuite au hasard selon une loi uniforme un père parmi l'ensemble des individus (non nouvellement créés) présents dans le voisinage de pollinisation de la mère (i.e. sur la surface du disque de rayon  $P$  centré sur la mère). La mère est elle-même un père potentiel, de sorte que le taux d'allofécondation est ici déterminé par la densité locale des reproducteurs (i.e. la densité au voisinage des individus). Pour chaque locus, on tire au hasard selon une loi uniforme un gène chez le père et un gène chez la mère. La valeur sélective déduite du génotype au locus sélectionné est la probabilité de survie du descendant potentiel (tous les individus de la population ayant la même probabilité de reproduction). Si le descendant potentiel ne survit pas, on recommence l'ensemble du processus depuis le tirage de la mère jusqu'à l'obtention d'un descendant qui survive. La distribution des distances de dispersion des graines et du pollen (uniforme sur la surface, ce qui implique des probabilités de dispersion plus élevées pour les grandes distances

jusqu'à  $D$  ou  $P$ , et nulles ensuite) est assez irréaliste. Nous l'avons utilisée par cohérence avec les précédentes études par simulation du modèle d'isolement par la distance, qui reposent, pour la plupart, sur une hypothèse similaire. Les conséquences de cette hypothèse sont discutées par Doligez et al., 1998.

Afin de faciliter l'étude par simulation du comportement de ce modèle, on a supposé que les individus de la population sont tous localisés aux intersections d'une grille carrée torique de taille  $L \times L$ . Le nombre d'intersections de la grille étant supérieur à  $N$ , tous les points de la grille ne sont pas occupés. On a également supposé que le temps est discret et constitué de cycles successifs. Chaque cycle est constitué : d'une phase de reproduction, pendant laquelle  $1\,000/C$  nouveaux individus sont créés selon le processus décrit ci-dessus ; d'une phase de vieillissement, durant laquelle tous les individus passent dans la classe d'âge supérieure (les individus nouvellement créés sont assignés à la première classe d'âge) ; et d'une phase de mortalité, au cours de laquelle tous les individus de la dernière classe d'âge meurent.

On a utilisé ces hypothèses pour simuler l'évolution de la population pendant 200 générations à partir d'un état initial pour lequel l'emplacement et les gènes de chaque individu sont tirés au hasard selon une loi uniforme (avec une fréquence allélique égale à 0,5 aux deux loci). Les précédentes études par simulation du modèle d'isolement par la distance ont en effet montré qu'à la génération 200 la structure spatiale génétique intrapopulation a, dans tous les cas étudiés, atteint un état qualifié par leurs auteurs de quasi-stationnaire (Sokal et Wartenberg, 1983 ; Epperson, 1990), c'est-à-dire pour lequel la vitesse d'augmentation de la moyenne des autocorrélations spatiales a considérablement diminué par rapport à la vitesse initiale dans les 50 premières générations. Nous avons simulé l'évolution de la population pour différentes combinaisons de valeurs des cinq variables d'entrée suivantes :  $L$  (taille de la grille, qui détermine la densité globale puisque la taille de la population est constante),  $D$  (distance maximale de dissémination des descendants),  $P$  (distance maximale de pollinisation),  $C$  (nombre de classes d'âge de la population, qui indique l'intensité du chevauchement des générations), et  $S$  (coefficient de sélection). Afin de caractériser les effets relatifs et d'interaction de ces variables d'entrée, on a utilisé un plan factoriel complet croisant deux valeurs extrêmes par variable (ce qui donne 32 cas différents fortement contrastés ; cf. *tableau I*), avec 50 répétitions par combinaison. Ces valeurs extrêmes ont été choisies d'après les données disponibles dans la littérature sur les arbres forestiers tropicaux (Doligez, 1996).

**Table I.** Valeurs des variables d'entrée utilisées dans le plan d'expérience.

Variable d'entrée	Borne minimale	Borne maximale
Taille de la grille $L$	1 000	2 000
Distance maximale de dissémination $D$	10	100
Distance maximale de pollinisation $P$	10	100
Nombre de classes d'âge $C$	1	10
Coefficient de sélection $S$	0,0	0,5

L'unité de  $D$  et  $P$  est le pas de la grille.

Les principales variables non génétiques mesurées en sortie sont le taux d'allofécondation  $T$  et l'agrégativité des individus, caractérisée par la statistique de Diggle  $dw$  (Upton et Fingleton, 1985). Cette statistique mesure la déviation de la répartition spatiale des individus observée par rapport à une répartition spatiale aléatoire. Le taux d'allofécondation est ici une variable non génétique, car son déterminisme est uniquement environnemental (pas de système génétique d'auto-incompatibilité). Les principales variables génétiques mesurées en sortie à chaque locus sont : la structure spatiale génétique, mesurée par des statistiques d'autocorrélation spatiale des fréquences alléliques individuelles (indice  $I$  de Moran; Upton et Fingleton, 1985; ces statistiques mesurent la similarité génétique des individus en fonction de la distance qui les sépare); la consanguinité de la population  $Fp$  (probabilité qu'au locus concerné, deux gènes tirés au hasard sans remise parmi l'ensemble des gènes de la population dérivent d'un même gène-ancêtre); la consanguinité individuelle  $Fi$  (probabilité que les deux gènes, au locus concerné, d'un individu tiré au hasard parmi l'ensemble des individus de la population dérivent d'un même gène-ancêtre). Pour calculer ces indices de consanguinité, on a affecté à chaque gène de la population initiale un numéro distinct, ce qui permet de savoir exactement à chaque génération quels gènes sont identiques par descendance. Ainsi, chaque locus (le locus neutre et le locus sélectionné) peut être considéré comme ayant initialement  $2N$  allèles. À chaque génération,  $Fi$  est alors l'homozygotie observée et

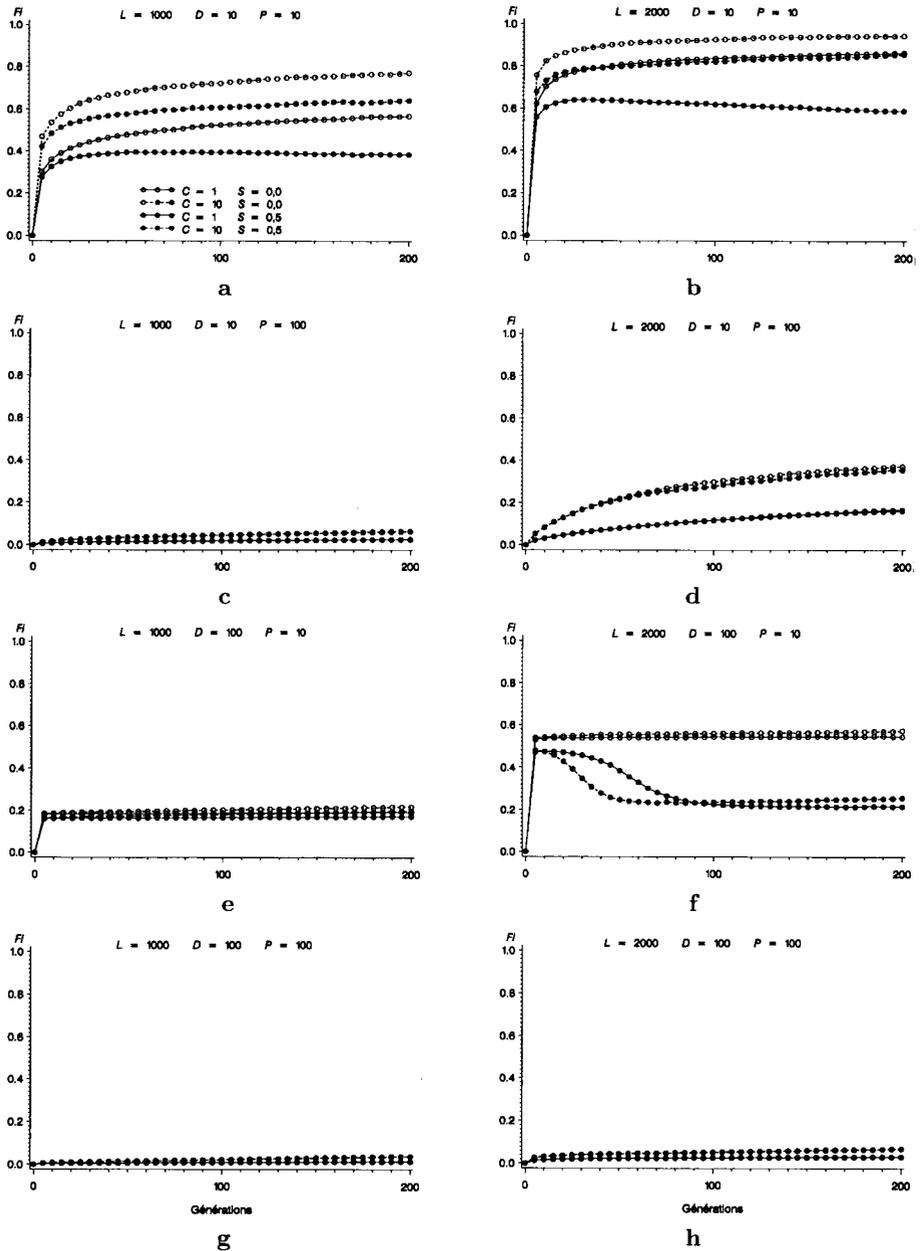
$$Fp = \frac{\sum_{i=0}^{i=2N} n_i(n_i - 1)}{2N(2N - 1)}$$

où  $n_i$  est le nombre de copies du gène numéro  $i$  dans la population ( $i = 1, \dots, 2N$ ). Les variables spatiales ( $dw$  et  $I$ ) ont été mesurées uniquement à la génération 200; les autres variables ont été mesurées toutes les cinq générations.

### 3. RÉSULTATS

L'évolution des valeurs moyennes de la consanguinité individuelle ( $Fi$ ) (*figure 1*) et du taux d'allofécondation ( $T$ ) est dans tous les cas devenue relativement lente à la génération 200 par rapport aux générations initiales. Dans quelques cas pour lesquels on a pu suivre l'évolution des statistiques spatiales dans le temps, les valeurs moyennes de l'agrégativité ( $dw$ ) et de la structure spatiale génétique ( $I$ ) avaient également une évolution fortement ralentie à la génération 200. Il faut noter qu'en l'absence de mutation, des populations finies isolées telles que celles simulées ici évoluent nécessairement vers la fixation d'un seul allèle. Mais étant donné leur grande taille, cela pourra prendre très longtemps (de l'ordre de plusieurs milliers de générations; Sokal et Wartenberg, 1983).

Le *tableau II* récapitule les effets des cinq variables d'entrée sur les valeurs des trois variables de sortie génétiques au locus neutre, à la génération 200. Toutes les variables d'entrée ont un effet sur l'état génétique des populations



**Figure 1.** Évolution dans le temps, à partir d'un état initial aléatoire, de la consanguinité individuelle moyenne (sur 50 répétitions) au locus neutre pour les 32 cas simulés. Chaque cas correspond à une combinaison extrême de valeurs des variables  $L$  (taille de la grille, proportionnelle à la densité globale),  $D$  (distance maximale de dissémination des descendants),  $P$  (distance maximale de pollinisation),  $C$  (chevauchement des générations), et  $S$  (coefficient de sélection à un locus sélectionné indépendant du locus neutre).

Table II. Effet des différentes variables d'entrée sur les variables génétiques au locus neutre.

Variable d'entrée	Effet de cette variable sur :		
	la consanguinité de la population $Fp$	la consanguinité individuelle $Fi$	la structure spatiale génétique $I^{(a)}$
$\searrow$ Densité globale de la population ( $\nearrow$ taille de la grille, $L$ )	$Fp \nearrow$ si $P$ minimal et $S$ maximal	$Fi \nearrow$ toujours	$I \nearrow$ si $D$ minimal
$\nearrow$ Distance maximale de dissémination ( $D$ )	$Fp \searrow$ si $P$ minimal et $S$ maximal	$Fi \searrow$ toujours	$I \searrow$ toujours
$\nearrow$ Distance maximale de pollinisation ( $P$ )	$Fp \searrow$ si $L$ et $S$ maximaux ou si $D$ minimal et $S$ maximal	$Fi \searrow$ toujours	$I \searrow$ si $D$ minimal
$\nearrow$ Chevauchement des générations (nombre de classes d'âge, $C$ )	$Fp \nearrow$ toujours	$Fi \nearrow$ toujours	$I \nearrow$ si $D$ minimal
$\nearrow$ Coefficient de sélection ( $S$ )	$Fp \nearrow$ si $D$ et $P$ minimaux ou si $L$ maximal et $P$ minimal	$Fi \searrow$ si $P$ minimal	$I \searrow$ si $D$ et $P$ minimaux

(a) Autocorrélations spatiales génétiques dans la 1<sup>re</sup> classe de distance, dont la largeur est égale à dix fois le pas de la grille. L'unité de  $D$  et  $P$  est le pas de la grille.

simulées (c'est-à-dire sur le niveau de diversité intrapopulation et/ou la structure de cette diversité). De plus, la plupart des effets d'interaction entre ces variables d'entrée sont très fortement significatifs. En particulier, certaines de ces variables ne peuvent avoir d'effet que pour des valeurs données des autres variables d'entrée. En revanche, les classements des résultats ne sont presque jamais modifiés (très peu d'inversions d'effets). Nous allons décrire et expliquer ici ces différents effets, notamment par les variations de l'agrégativité et du taux d'allofécondation.

La diminution de la densité globale (augmentation de  $L$ ) conduit à une augmentation de l'agrégativité due à l'éloignement des agrégats quand la distance maximale de dissémination ( $D$ ) est minimale. Dans ce cas, la densité locale (densité au voisinage des individus), et donc le taux d'allofécondation ( $T$ ), ne sont pas modifiés; l'éloignement des agrégats conduit à une plus grande fixation à l'intérieur des agrégats, et par conséquent à une augmentation de la structure spatiale génétique ( $I$ ) (*tableau III*) et de la consanguinité individuelle ( $Fi$ ) (*figure 1*). Quand  $D$  est maximal, l'agrégativité est minimale dans la plupart des cas ( $dw$  n'est alors pas significativement différent de zéro), donc la diminution de la densité globale ne modifie pas l'agrégativité et s'accompagne d'une diminution de la densité locale (et donc d'une diminution de  $T$  et d'une augmentation de  $Fi$  quand  $P$ , la distance maximale de pollinisation, est minimale), sauf dans les deux cas pour lesquels  $D$  et  $S$  sont maximaux et  $P$  est minimal (pour  $C = 1$  et  $C = 10$ ). Dans ces deux cas, la diminution de la densité globale entraîne une forte augmentation de l'agrégativité, due au désavantage relatif des individus isolés (et donc se reproduisant par autofécondation) par rapport aux individus non isolés, pour le nombre de descendants laissés à la génération suivante. Ce désavantage relatif conduit en quelques générations à un regroupement de tous les individus de la population en un unique agrégat présentant la même densité locale (et donc le même taux d'allofécondation,  $T$ ) que lorsque la densité globale est maximale. Quand  $D$  est maximal,  $I$  est toujours proche de zéro, donc un changement de la densité globale ne modifie pas sa valeur. Enfin, la diminution de la densité globale conduit à une augmentation de  $Fp$  uniquement quand  $P$  est minimal et  $S$  maximal (*figure 2*).

L'augmentation de la distance maximale de dissémination des descendants ( $D$ ) conduit toujours à une diminution de l'agrégativité ( $dw$ ) résultant d'une diminution de la densité locale. Donc quand la distance maximale de pollinisation ( $P$ ) est minimale, l'augmentation de  $D$  conduit également à une diminution du taux d'allofécondation  $T$  plus ou moins forte selon la diminution de  $dw$ . L'augmentation de  $D$  entraîne toujours une diminution de  $I$  et de  $Fi$ , malgré la diminution de  $T$ . Cette diminution est due à l'augmentation des flux de gènes par les graines. Quand  $D$  est maximal, les autocorrélations spatiales génétiques sont proches de zéro quelles que soient les valeurs des autres variables d'entrée, alors que la consanguinité individuelle n'est minimale que lorsque  $P$ , et donc le taux d'allofécondation, sont maximaux. L'augmentation de  $D$  ne conduit à une diminution de la consanguinité de la population que dans le cadre d'une interaction d'ordre 4 ( $L \times D \times P \times S$ ).

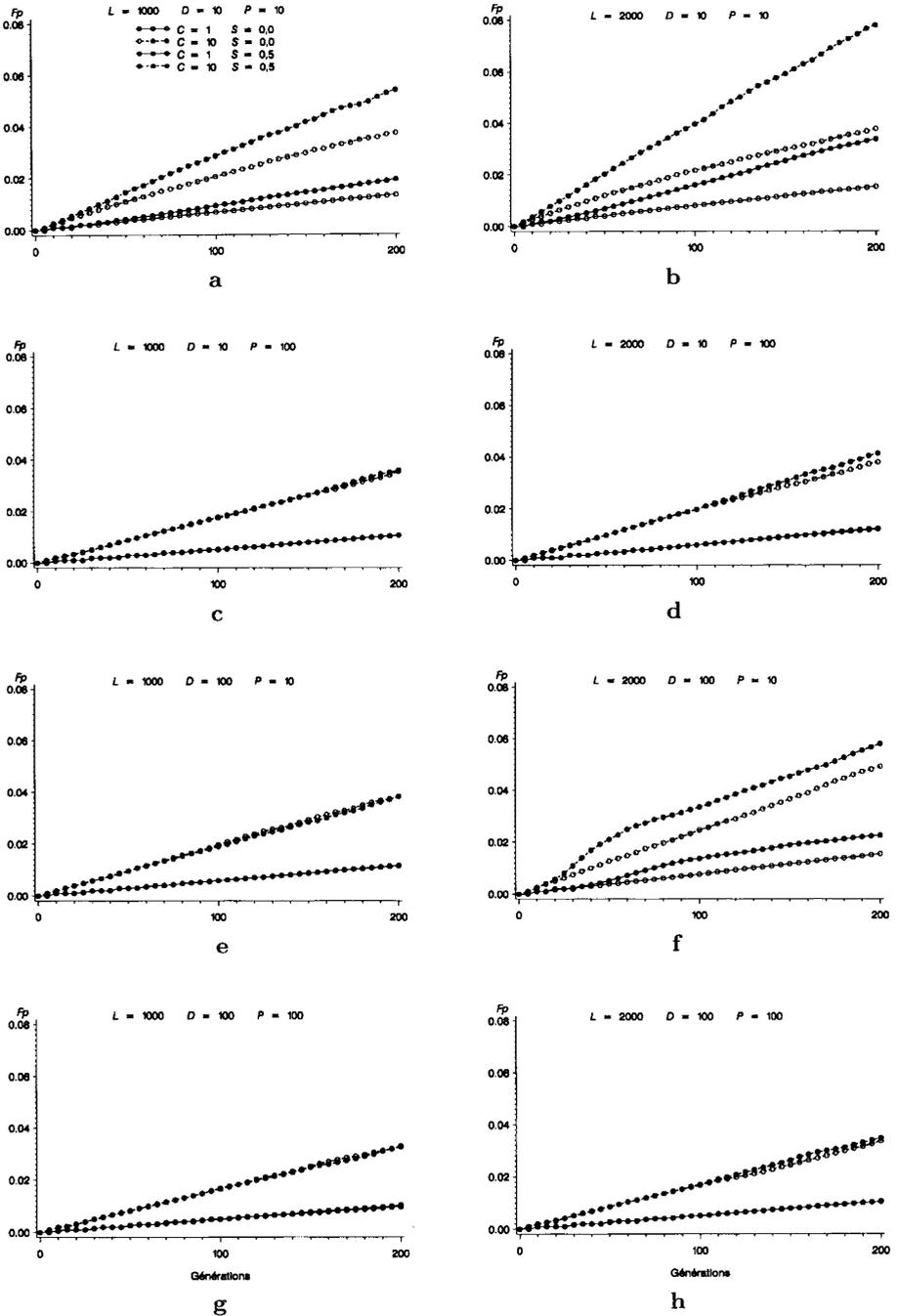
L'augmentation de la distance maximale de pollinisation ( $P$ ) ne conduit à une diminution de l'agrégativité ( $dw$ ) et de la consanguinité de la population ( $Fp$ ) que dans le cadre de cette même interaction d'ordre 4. L'augmentation de  $P$  conduit toujours à une augmentation de  $T$ , ce qui découle directement

**Table III.** Valeurs moyennes (sur 50 répétitions) des autocorrélations spatiales génétiques ( $I$ ) dans la 1<sup>re</sup> classe de distance, à la génération 200.

$D$	$P$	$C$	$S$	$L = 1\ 000$	$L = 2\ 000$
10	10	1	0,0	0,59	0,86
		10	0,0	0,81	0,95
		1	0,5	0,42	0,62
		10	0,5	0,69	0,87
10	100	1	0,0	0,05	0,28
		10	0,0	0,12	0,54
		1	0,5	0,05	0,27
		10	0,5	0,12	0,51
100	10	1	0,0	0,01	0,06
		10	0,0	0,02	0,10
		1	0,5	0,01	0,01
		10	0,5	0,02	0,02
100	100	1	0,0	0,00	0,03
		10	0,0	0,01	0,08
		1	0,5	0,01	0,03
		10	0,5	0,01	0,07

du déterminisme de  $T$  sous les hypothèses du présent modèle. L'augmentation de  $P$  entraîne systématiquement une diminution de  $I$  quand  $D$  est minimal, mais ne conduit jamais à une absence de structure spatiale génétique ( $I$  proche de zéro), contrairement à l'augmentation de  $D$ . Quand  $D$  est maximal,  $I$  est toujours proche de zéro, et  $P$  n'a pas d'effet sur  $I$ . La distance maximale de pollinisation ( $P$ ) a donc moins d'effet sur la structure spatiale génétique ( $I$ ) que la distance maximale de dissémination ( $D$ ), ce qui est dû d'une part au fait que chaque graine dispersée contribue deux fois plus aux flux de gènes que chaque grain de pollen dispersé, et d'autre part au fait que lorsque  $D$  est maximal,  $I$  est toujours proche de zéro, et sa valeur n'est pas influencée par le taux d'allofécondation ( $T$ ). L'augmentation de  $P$  conduit toujours à une diminution de la consanguinité individuelle ( $F_i$ ), même quand  $D$  est maximal, ce qui est dû non seulement à l'augmentation des flux de gènes par le pollen, mais aussi à l'augmentation de  $T$ . Ceci implique que  $P$  a plus d'effet sur la consanguinité individuelle que  $D$ , contrairement à ce qui se passe pour la structure spatiale génétique ( $I$ ).

L'augmentation du nombre de classes d'âge ( $C$ ) et donc du chevauchement des générations, conduit à une augmentation de l'agrégativité quand  $D$  est minimal, par augmentation de la densité locale (due à la présence simultanée dans la population de certains individus et de leurs descendants faiblement disséminés). Donc quand  $P$  est minimal, l'augmentation de  $C$  entraîne aussi une augmentation de  $T$ . Quand  $D$  est minimal, l'augmentation de  $C$  provoque une augmentation de  $I$  malgré l'augmentation de la densité locale et de  $T$ , car les individus proches sont encore plus apparentés à cause du chevauchement des générations. Quand  $D$  est maximal,  $C$  n'a pas d'effet sur  $I$ , qui est proche de zéro. L'augmentation de  $C$  conduit dans tous les cas à une augmentation de



**Figure 2.** Évolution dans le temps, à partir d'un état initial aléatoire, de la moyenne (sur 50 répétitions) de la consanguinité de la population au locus neutre pour les 32 cas simulés.

$Fp$ ; c'est la variable d'entrée qui a globalement le plus d'effet sur  $Fp$ . Quand  $C$  augmente,  $Fi$  augmente dans tous les cas, malgré l'augmentation de  $T$ , et de façon plus marquée dans les cas où  $I$  augmente aussi.

L'augmentation du coefficient de sélection ( $S$ ) conduit à une augmentation de l'agrégativité par une plus grande densité locale quand  $D$  et  $P$  sont minimaux. Cet effet est dû à la fois à l'avantage relatif des mères hétérozygotes au locus sélectionné pour le nombre de descendants laissés à la génération suivante (cf. Doligez et al., 1998), et à la faible dissémination de ces descendants par rapport à l'emplacement de leur mère. L'augmentation de  $S$  conduit également à une augmentation de l'agrégativité et de la densité locale dans les deux cas particuliers déjà expliqués ci-dessus (cf. effet de la densité globale). Globalement, l'influence de  $S$  sur l'agrégativité dépend donc des valeurs de  $L$ ,  $D$  et  $P$ ; c'est pourquoi nous parlons ici d'interaction d'ordre 4. L'augmentation de  $S$  provoque toujours une augmentation de  $T$  quand  $P$  est minimal, non seulement à cause de l'augmentation de la densité locale dans le cadre de l'interaction d'ordre 4, mais aussi à cause de la meilleure survie des individus issus d'allofécondation dans tous les cas. Quand  $P$  est maximal,  $T$  est toujours maximal. L'augmentation de  $S$  (au locus sélectionné) entraîne toujours une diminution de  $I$  au locus neutre (indépendant) quand  $D$  et  $P$  sont minimaux, car l'augmentation de la densité locale entraîne un plus grand mélange des génotypes aux faibles distances. L'augmentation de  $S$  conduit à une augmentation de  $Fp$  au locus neutre dans tous les cas où elle a un effet sur l'agrégativité (cf. interaction d'ordre 4), malgré l'augmentation de  $T$ . Enfin, l'augmentation de  $S$  conduit toujours à une diminution de  $Fi$  au locus neutre quand  $P$  est minimal (malgré l'augmentation de  $Fp$ ), ce qui est lié à l'augmentation de  $T$ .

#### 4. DISCUSSION

Les simulations présentées ici ont permis de montrer que les variations de la structure spatiale génétique ( $I$ ) et de la consanguinité individuelle ( $Fi$ ) peuvent être expliquées en grande partie par les variations de l'agrégativité, de la densité locale, et/ou du taux d'allofécondation. Une part importante de ces variations n'aurait donc pas pu être détectée sous l'hypothèse d'une répartition spatiale uniforme des individus. En ce qui concerne les variations de la consanguinité de la population, elles dépendent principalement du chevauchement des générations.

Aussi bien pour la consanguinité individuelle ( $Fi$ ) que pour la structure spatiale génétique, la prise en compte d'une répartition spatiale non-uniforme des individus dans notre modèle a permis de montrer que la dispersion des gènes par le pollen peut avoir un effet très différent de celui de la dispersion par les graines. Cette différence est en partie liée au fait que sous les hypothèses de notre modèle, l'effet des distances de pollinisation est systématiquement accentué par les variations du taux d'allofécondation et/ou de la densité locale, alors que l'effet des distances de dissémination est liée aux variations de l'agrégativité. Cette différence n'a jamais été décrite dans les précédentes études par simulation du modèle d'isolement par la distance (Sokal et Wartenberg, 1983; Sokal et al., 1989, Epperson, 1990, 1995a, 1995b), car ces auteurs ne font pas varier séparément ces deux modes de dispersion. L'effet sur la structure spatiale génétique d'une sélection de type superdominance à un locus

indépendant n'a jamais été décrit auparavant non plus, car il est lié à des variations de la densité locale et les précédents travaux de simulation reposent tous sur l'hypothèse d'une répartition spatiale uniforme des individus. De plus, Sokal et al. (1989) et Epperson (1990) ont décrit uniquement la structure spatiale au locus sélectionné, et non à des loci neutres indépendants.

Nos résultats pour  $F_i$  sont qualitativement en accord avec les résultats classiques concernant l'effet de différentes variables sur la taille efficace de consanguinité  $N_e$  d'une population finie : le chevauchement des générations diminue  $N_e$  et l'augmentation des distances de dispersion relatives augmente  $N_e$  (Hartl et Clark, 1989). Cependant, une comparaison quantitative nécessiterait la connaissance de variables telles que la variance du nombre effectif de gamètes produits par individu. Les valeurs de  $F_i$  dépendent de deux sources de consanguinité différentes : l'autofécondation et les croisements entre individus apparentés (ou consanguinité biparentale), dont l'importance peut être estimée par les valeurs de  $I$ . Selon les cas, l'effet prépondérant sur les valeurs de  $F_i$  sera celui de l'autofécondation ou de la consanguinité biparentale. Par exemple, quand il y a très peu de consanguinité biparentale ( $I$  proche de zéro), les variations de  $F_i$  sont fortement liées à celles du taux d'alofécondation. Cependant, pour pouvoir comparer quantitativement l'influence sur  $F_i$  de ces deux sources de consanguinité, il serait nécessaire de mesurer exactement la valeur de la consanguinité biparentale.

Hamrick et al. (1992) ont tenté d'expliquer les différences de niveau et de structure du polymorphisme enzymatique observées entre les espèces pérennes pour lesquelles des résultats sont disponibles dans la littérature. Pour cela, ils ont classé ces espèces selon les critères suivants : statut taxonomique, étendue de l'aire naturelle, distribution géographique, système de reproduction, dissémination des graines, mode de reproduction, et statut successional. Ils ont ainsi réussi à expliquer 39 % des différences de diversité intrapopulation et 16 % des différences de diversité interpopulation entre espèces. La consanguinité de la population mesurée dans nos simulations est proche de l'homozygotie espérée en panmixie (avec le concept d'identité par descendance). Or, Hamrick et al. (1992) ont utilisé l'hétérozygotie espérée en panmixie (avec le concept d'identité par état) comme mesure de la diversité. Nos résultats suggèrent donc qu'en utilisant des critères de classification supplémentaires tels que la densité globale, l'agrégativité, la densité locale, le chevauchement des générations, et la sélection, ainsi que les interactions entre les différents facteurs de classification, il serait sans doute possible d'expliquer une plus grande part des différences génétiques observées entre espèces d'arbres.

Bien sûr, le modèle présenté ici n'est que l'un des nombreux modèles possibles, et il serait nécessaire d'y apporter de nombreux compléments, notamment par l'introduction de variations démographiques (de la taille et de la structure en classes d'âge de la population), d'un déterminisme partiellement génétique du taux d'alofécondation (locus d'auto-incompatibilité), de mortalité densité-dépendante, de sélection de type mutation-dominance, de distributions plus réalistes des distances de dispersion des gènes par les graines et le pollen, de mortalité groupée rendant compte de la dynamique forestière (chablis), etc. L'étude de ce modèle simple a néanmoins permis de mettre en évidence plusieurs mécanismes possibles pour le déterminisme du niveau et de la structure de la diversité génétique chez les arbres forestiers. La comparaison de données

réelles obtenues sur plusieurs espèces ayant des caractéristiques biologiques et écologiques contrastées permettra de savoir s'il est nécessaire de remettre en cause certaines des hypothèses du modèle. Une fois les modifications nécessaires apportées, ce modèle pourra être utilisé comme outil de prédiction de la diversité intrapopulation et de son organisation, éventuellement en présence de perturbations telles que l'exploitation forestière, pour différentes espèces d'arbres en milieu naturel ou aménagé.

## REMERCIEMENTS

Ce travail a pu être réalisé grâce à une allocation de recherche du ministère de la Recherche attribuée à A. Doligez, et à l'aide financière du Cirad-Forêt. Les auteurs remercient C. Duby et un lecteur anonyme pour leurs commentaires sur une version antérieure du manuscrit, D. Doligez pour ses conseils concernant le programme de simulation, et L. Houde pour ses suggestions à propos des aspects statistiques.

## RÉFÉRENCES

Barret S.C.H., Charlesworth D., Effects of a change in the level of inbreeding on the genetic load, *Nature* 352 (1991) 522–524.

Bos M., van der Haring E., Gene flow in *Plantago*. II. Gene flow pattern and population structure. A simulation study, *Heredity* 61 (1988), 1–11.

Charlesworth D., Morgan M.T., Charlesworth B., Inbreeding depression, genetic load, and the evolution of outcrossing rates in a multilocus system with no linkage, *Evolution* 44 (1990) 1469–1489.

Doligez A., Évolution de la diversité génétique intrapopulation et de sa structure : étude d'un modèle de simulation spatialisé en vue de la gestion des ressources génétiques forestières tropicales. Thèse de l'Institut national agronomique de Paris-Grignon, 1996.

Doligez A., Baril C., Joly H.I., Fine-scale spatial genetic structure with non-uniform distribution of individuals, *Genetics* 148 (1998) 905–919.

Epperson B.K., Spatial autocorrelation of genotypes under directional selection, *Genetics* 124 (1990) 757–771.

Epperson B.K., Spatial distributions of genotypes under isolation by distance, *Genetics* 140 (1995a) 1431–1440.

Epperson B.K., Fine-scale spatial structure : correlations for individual genotypes differ from those for local gene frequencies, *Evolution* 49 (1995b) 1022–1026.

Hamrick J.L., Godt M.J.W., Sherman-Broyles S.L., Factors influencing levels of genetic diversity in woody plant species, *New Forests* 6 (1992) 95–124.

Hartl D.L., Clark A.G., Principles of population genetics. Sinauer Associates, Sunderland, M.A., 2nd ed, 1989.

Kawata M., Effective population size in a continuously distributed population, *Evolution* 49 (1995) 1046–1054.

Ohsawa R., Furuya N., Ukai Y., Effect of spatially restricted pollen flow on spatial genetic structure of an animal-pollinated allogamous plant population, *Heredity* 71 (1993) 64–73.

Rohlf F.J., Schnell G.D., An investigation of the isolation-by-distance model, *Am. Nat.* 105 (1971) 295–324.

Ronfort J., Couvet D., A stochastic model of selection on selfing rates in structured populations, *Genet. Res. Camb.* 65 (1995) 209–222.

Slatkin M., Barton N.H., A comparison of three indirect methods for estimating average levels of gene flow, *Evolution* 43 (1989) 1349–1368.

Sokal R.R., Wartenberg D.E., A test of spatial autocorrelation analysis using an isolation-by-distance model, *Genetics* 105 (1983) 219–237.

Sokal R.R., Jacquez G.M., Wooten M.C., Spatial autocorrelation analysis of migration and selection, *Genetics* 121 (1989) 845–855.

Turner M.E., Stephens J.C., Anderson W.W., Homozygosity and patch structure in plant populations as a result of nearest-neighbor pollination, *Proc. Natl. Acad. Sci. USA* 79 (1982) 203–207.

Upton G.J.G., Fingleton B., *Spatial Data Analysis by Example. Volume I. Point Pattern and Quantitative Data*, John Wiley & Sons, Chichester, 1985.