Les Actes du BRG, 6 (2006) 211-230
© BRG, 2006

Article original

Stratégies raisonnées d'échantillonnage pour capturer la diversité génétique et sa structuration dans les populations naturelles – Application aux mesures de gestion conservatoire

Stéphanie BRACHET[(1)*], Catherine BASTIEN[(2)], Isabelle BILGER[(3)], Corinne BURET[(2)], Jean DUFOUR[(2)], Jérôme GRANGE[(3, 5)], Vanina GUÉRIN[(2)], Véronique JORGE[(2)], Bénédicte LE GUERROUE[(4)], Marie-Claude LESAGE-DESCAUSES[(2)], Laurent LEVEQUE[(4)], Irinel MARDARE[(2)], Stéphanie MARIETTE[(3§)], Brigitte MUSCH[(4)], Leopoldo SANCHEZ[(2)], Solenn STOECKEL[(3, 5)], Marc VILLAR[(2)], Frédérique SANTI[(2)]

[(1)] ENGREF, Arboretum national des Barres, Domaine des Barres, 45290 Nogent sur Vernisson, France
[(2)] INRA, UR 588 Amélioration Génétique et Physiologie Forestières, Avenue de la Pomme de Pin, Ardon, BP 20619, 45166 Olivet Cedex, France
[(3)] Cemagref de Nogent, Domaine des Barres, 45290 Nogent sur Vernisson, France
[(4)] ONF, Conservatoire génétique des arbres forestiers USC INRA, Avenue de la Pomme de Pin, Ardon, BP 20619, 45166 Olivet Cedex, France
[(5)] Université Paris XI, UMR 8079, Laboratoire Ecologie, Systématique et Évolution, Bât. 360, 91405 Orsay, France
§ Adresse actuelle : INRA, UR 419 Espèces fruitières et Vigne, Domaine de la Grande-Ferrade, BP 81, 33883 Villenave d'Ornon, Cedex, France

Abstract: Sampling of natural genetic diversity for optimal structure and diversity capturing – applications to resource conservation policies. Both environmental and evolutionary factors determine the distribution of the genetic diversity among and within natural populations of forest tree species, resulting in genetically structured metapopulations. The existence of such distribution has major implications for resource conservation policies, as the shape and extent of *in situ* natural networks and the composition of *ex situ* collections of valuable forest tree species. A major and realistic issue we face is how to effectively sample natural variation both for optimal structure mirroring and for maximal genetic diversity capturing when designing *in situ* population networks and *ex situ* collections. Genetic differentiation descriptors like Fst and Qst represent the utmost visible consequences of underlying landscape and evolution factors acting on the metapopulation genetic diversity. The first question we address here is how to sample populations, individuals and marker loci knowing the neutral and selected genetic structure of a theoretically built metapopulation. Results

* Correspondance et tirés à part : brachet@engref.fr

of a Monte-Carlo allele-based simulation model are presented, where the sample of populations, individuals and marker loci was simultaneously modelled over ranges of Fst and Qst. Each sample strategy was not only evaluated in terms of genetic diversity potential yield relative to total diversity at the metapopulation level, but also in terms of bias and precision rendering in the estimation of Fst and Qst from the actual sample. Thus the ability of the sample to render both the diversity and its distribution over the populations was analyzed. The second question we address is whether present *ex situ* collections of two disseminated forest tree species, black poplar and wild cherry, represent a good sampling of natural populations. For that purpose, natural populations and *ex situ* collections were characterised with neutral markers and phenotypic characters. The simulations then serve to infer the degree of representativeness of existing *ex situ* collections for the two species based on the actual sample sizes and several differentiation estimates of Fst and Qst, and eventually guide the enrichment of conservation resources for these two species.

sampling/ genetic diversity/ differentiation/ *Prunus*/ *Populus*

Résumé : La diversité génétique d'une espèce se distribue entre populations et à l'intérieur des populations. Sa prise en compte à ces deux niveaux a une importance majeure pour la mise en place de modes de gestion conservatoires *in situ* et *ex situ*. La question principale qui se pose est : quel échantillonnage de la diversité faire, en terme de populations, d'individus et de gènes ? Nous abordons ici cette question, d'un point de vue théorique à travers le développement d'un modèle Monte-Carlo, puis d'un point de vue expérimental sur deux espèces forestières, le merisier et le peuplier noir, à travers une comparaison de la diversité présente en populations naturelles et en collection *ex situ*. Les deux approches sont ensuite mises en regard l'une de l'autre afin de donner des éléments sur les stratégies d'échantillonnage optimales de la diversité de ces deux espèces.

échantillonnage/ diversité génétique/ différenciation/ *Prunus*/ *Populus*

1. INTRODUCTION

La diversité génétique globale d'une espèce peut se décomposer en deux niveaux : la diversité présente à l'intérieur de chaque population et le degré de différenciation entre populations. Diversité et différenciation peuvent se mesurer au niveau de l'ADN ou à travers la quantification de la variance entre phénotypes pour différents caractères. La mise en place de collections de ressources génétiques *ex situ* ou de réseaux de conservation *in situ* devrait reposer sur la connaissance de ces deux niveaux de diversité, car ce sont les différences entre individus ayant une signification évolutive que l'on souhaite généralement conserver [13].

La question inhérente est alors : quel échantillonnage démographique et génétique faire pour constituer collections et réseaux ? Des modèles ont

d'ores et déjà examiné tout ou partie de cette question comme celui de Bataillon *et al.* [1]. Cependant, ils reposent sur la modélisation de l'ensemble des facteurs évolutifs. Dans ce projet, nous avons posé cette question en partant de la connaissance à priori de la différenciation génétique neutre (Fst) et adaptative (Qst) des populations. Nous cherchons à déterminer : i) l'échantillonnage optimal à mener pour estimer au mieux la variabilité génétique intra et inter populations d'une espèce donnée, ii) l'échantillonnage optimal des individus à faire pour constituer une collection *ex situ* d'une taille donnée représentant, à un temps donné, le maximum de diversité génétique de l'espèce, une fois cette dernière estimée.

Dans les faits, des collections *ex situ* ou des réseaux *in situ* de ressources génétiques ont jusqu'à présent été constitués en se basant principalement sur des critères écologiques. C'est particulièrement le cas des arboreta ou de certaines collections d'arbres forestiers, mis en place sans connaissance à priori de la variabilité génétique. Dans ce projet, nous avons alors voulu examiner si des collections *ex situ* de deux arbres forestiers, le merisier et le peuplier noir, constituées uniquement sur la connaissance de l'origine géographique des individus représentaient un échantillonnage adéquat de la diversité des populations naturelles et s'il fallait, au besoin, les revoir.

Nous présentons ici des éléments de réponse apportés par le modèle quant aux modalités d'échantillonnage théoriquement optimal pour estimer variabilité et différenciation. Nous les complétons par les premières estimations de la diversité et de la différenciation en populations naturelles chez le merisier et en collection chez le peuplier noir. Nous confrontons enfin les recommandations théoriques aux échantillonnages concrets qui ont été faits sur les deux espèces forestières.

2. STRATÉGIES D'ÉCHANTILLONNAGE : APPROCHE THÉORIQUE

2.1. Description du modèle

La figure 1 représente schématiquement le modèle développé et précise les paramètres fixés à priori. Un pool de gènes est d'abord créé à partir d'un certain nombre de paramètres fixés initialement. Trois cas distincts sont présentés ici : le cas a suppose que tous les loci sont identiques (*i.e.* fréquences allèliques, nombre d'allèles et valeurs de Fst identiques) ; le cas b, que les loci ont des fréquences alléliques et des nombre d'allèles différents ; le cas c, que les loci ont des fréquences alléliques identiques mais des valeurs de Fst variables. Il est par ailleurs supposé que la répartition des loci dans le génome est homogène et qu'ils sont indépendants. Il n'y a pas de déséquilibre de liaison, ni de

dominance. Le pool de gènes créé est ensuite réparti aléatoirement dans une métapopulation initiale (avec des valeurs de Fst et Qst proches de 0), constituée d'un nombre de populations défini, chaque population ayant un nombre d'individus défini. Pour obtenir une métapopulation correspondant à un couple (Fst, Qst) donné, une méthode d'optimisation appelée « recuit simulé » (ou Simulated Annealing) [12] est alors appliquée. Cette méthode consiste en une grande série d'itérations (> 10^6), qui chacune est une légère modification de la composition génotypique de la métapopulation par échanges d'allèles entre populations, d'une part aux loci neutres et d'autre part aux loci sélectionnés. Selon les valeurs de Fst et Qst calculées sur la nouvelle métapopulation obtenue à la fin de chaque itération et un critère de convergence donné par la fonction probabiliste de Boltzmann, la modification est retenue ou rejetée. Lorsque aucune modification ne peut améliorer la convergence vers les valeurs de Fst et Qst souhaitées, le processus d'itération s'arrête. La métapopulation ainsi obtenue est alors considérée comme la « vraie » métapopulation, caractérisée par un couple (Fst_M, Qst_M) donné et un nombre total et efficace d'allèles.

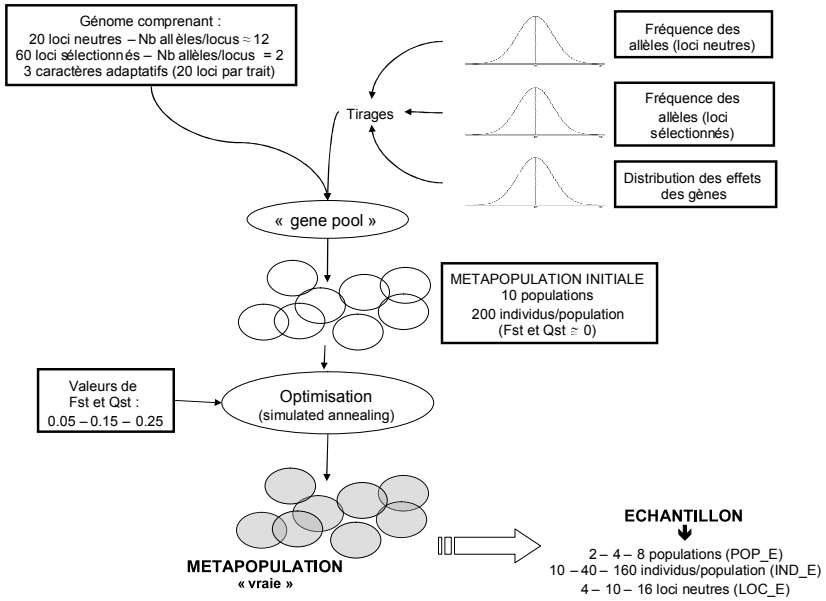


Figure 1 : Modèle d'échantillonnage. Représentation schématique du modèle d'échantillonnage. Dans les encadrés accentués, les paramètres fixés sont indiqués avec leurs valeurs. La distribution des fréquences alléliques aux loci neutres est supposée uniforme ou asymptotique et peut varier entre loci. Les valeurs de Fst sont fixées pour le génome entier (cas a) ou par locus (cas c). Les valeurs de Qst sont fixées par trait. A : nombre total d'allèles, Ae : nombre efficace d'allèles.

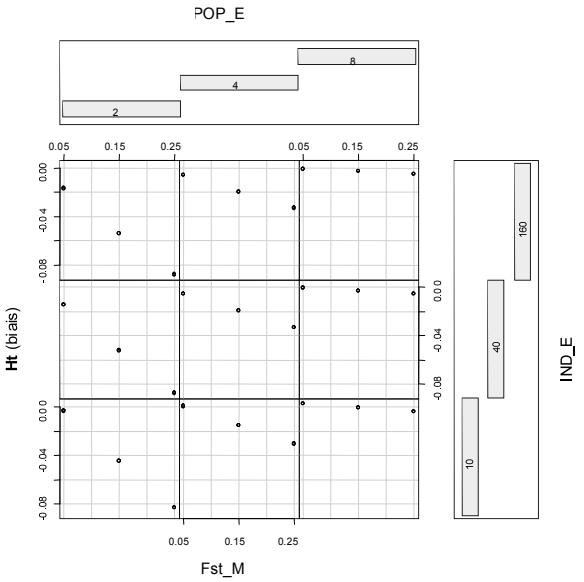
Une série d'échantillonnages est alors simulée afin de déterminer les meilleures stratégies d'échantillonnage. Chaque échantillonnage consiste en un tirage sans remise des populations, puis des individus dans les populations choisies et des loci. Le tirage des loci neutres détermine le nombre de marqueurs neutres utilisés. Pour les loci sélectionnés, il n'y a pas de tirage aléatoire, la valeur génotypique des individus étant considérée comme connue. Pour chaque couple (Fst_M, Qst_M) donné, 30 métapopulations sont construites et pour chaque métapopulation, la stratégie d'échantillonnage est répétée 10 000 fois.

Afin d'évaluer la stratégie d'échantillonnage, les moyennes et variances des variables génétiques suivantes : Hs (hétérozygotie intra-population), Ht (hétérozygotie totale), Fst, Qst, Ae (nombre efficace d'allèles) et A (nombre total d'allèles) sont calculées sur l'ensemble des répétitions et un échantillon de 150 d'entre elles est retenu pour analyser les courbes de distribution de chaque variable génétique ainsi que leurs intervalles de confiance. La comparaison avec la « vraie » métapopulation sur ces variables génétiques est faite en calculant la différence entre les valeurs moyennes estimées sur les échantillons pour chaque stratégie d'échantillonnage et les valeurs calculées sur la métapopulation définie par un couple (Fst_M, Qst_M) donné. Cette différence est appelée biais. La précision du biais est mesurée à partir des variances calculées sur l'ensemble des répétitions (la précision est forte quand les variances sont faibles). Biais et précision sont alors analysés en fonction des facteurs suivants : nombre de populations échantillonnées, nombre d'individus échantillonnés par population, nombre de loci ou marqueurs neutres échantillonnés et valeur du Fst de la « vraie » métapopulation (Fst_M).

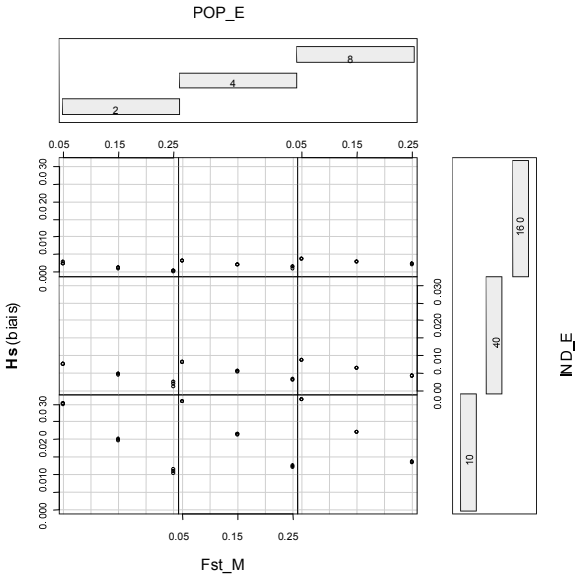
3. RÉSULTATS

Les résultats obtenus pour le cas a (loci identiques) sont illustrés par les figures 2 à 5. La figure 2 montre les biais obtenus sur les variables Ht, Hs et Fst en fonction du nombre de populations, du nombre d'individus et des valeurs de Fst_M. Un biais négatif indique une sous-estimation liée à l'échantillonnage. Concernant la valeur de Ht, on observe qu'il est toujours sous-estimé par l'échantillonnage (fig. 2a). Une diminution du nombre de populations échantillonnées entraîne une augmentation de la sous-estimation de Ht. Une réduction du nombre d'individus a, quant à elle, moins d'effet. Enfin, plus la valeur du Fst de la métapopulation est grande, plus la valeur de Ht est sous-estimée. A contrario, la valeur de Hs est toujours surestimée (fig. 2b), particulièrement lorsque le nombre d'individus échantillonnés est faible et que la valeur du Fst_M est faible. La valeur du Fst est toujours sous-estimée par l'échantillonnage (fig. 2c). Cette sous-estimation est d'autant plus forte que le nombre de populations échantillon-

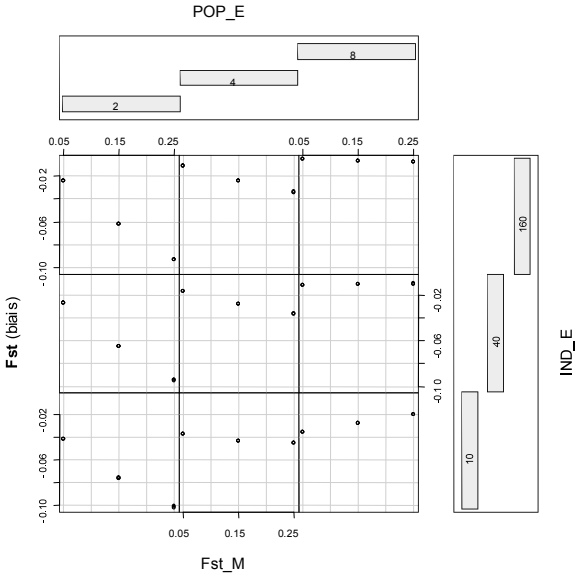
nées ou que le nombre d'individus diminue, en particulier pour des valeurs fortes du Fst_M.



2a



2b



2c

Figure 2 : Effet de l'échantillonnage sur Ht, Hs et Fst – cas a.
Valeurs des biais pour Ht (2a), Hs (2b) et Fst (2c) en fonction du nombre de populations (POP_E), du nombre d'individus (IND_E) échantillonnés et du Fst de la « vraie » métapopulation (Fst_M).
(biais = valeur calculée sur l'échantillon – valeur calculée sur la « vraie » métapopulation).

Les figures 3 à 5 présentent les variations du biais avec sa précision en fonction du nombre de loci neutres, du nombre de populations et du nombre d'individus échantillonnés pour le Fst, Qst, A et Ae. Pour chacune des valeurs estimées, la précision du biais augmente avec l'effort d'échantillonnage, notamment avec le nombre de populations échantillonnées.

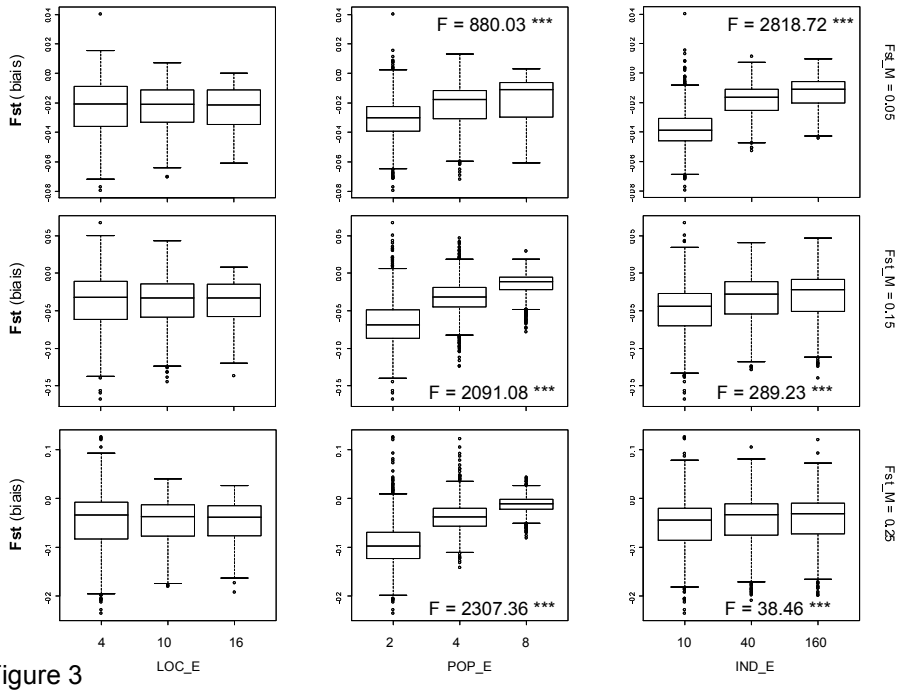


Figure 3

Figure 3 : Effet de l'échantillonnage sur Fst – cas a.
Diagramme des biais du Fst avec la précision pour différents niveaux d'échantillonnage des loci (LOC_E), des populations (POP_E), des individus (IND_E) et du Fst de la « vraie » métapopulation (Fst_M). Les valeurs minimales et maximales de chaque carré correspondent aux quartiles inférieurs et supérieurs, la moitié correspondant à la valeur médiane. Pour les facteurs significatifs, le calcul du F provenant d'une ANOVA à 3 facteurs (POP_E x IND_E x LOC_E) est donné.

Dans chaque cas, le degré de significativité du facteur est évalué à travers une analyse de variance. L'échantillonnage des populations et des individus a un effet significatif sur le Fst, l'effet individu étant plus important pour de faibles valeurs du Fst_M, alors que pour de fortes valeurs, c'est l'effet population qui est plus important. Le Qst, quant à lui, est généralement sous-estimé et le nombre de populations échantillonnées est le seul facteur jouant sur son estimation (fig. 4).

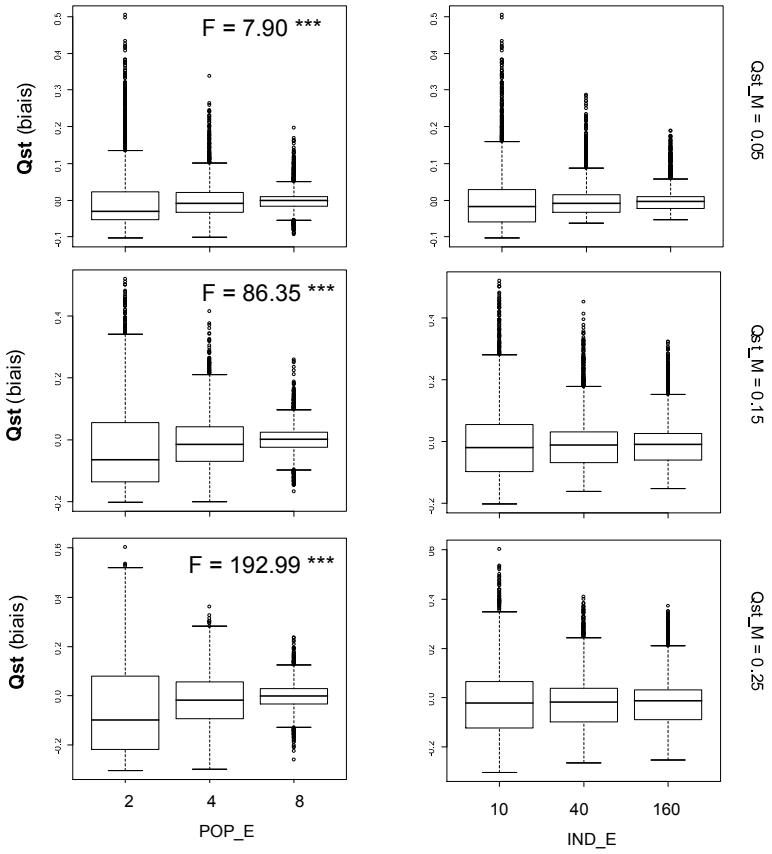
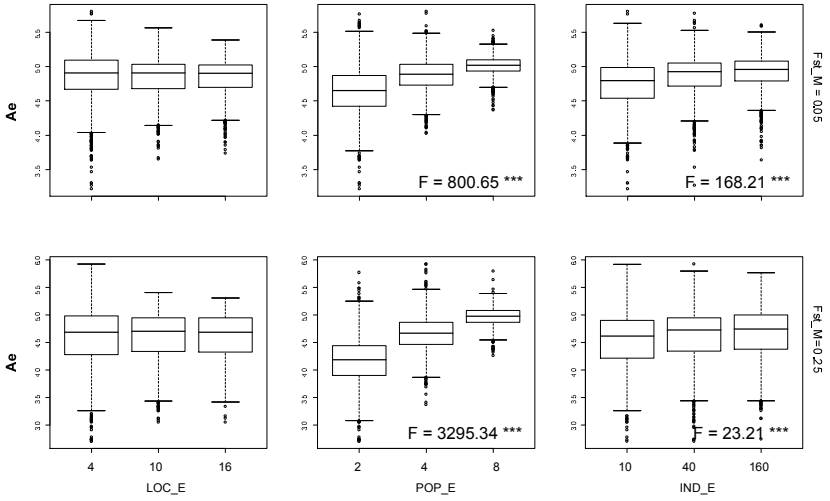
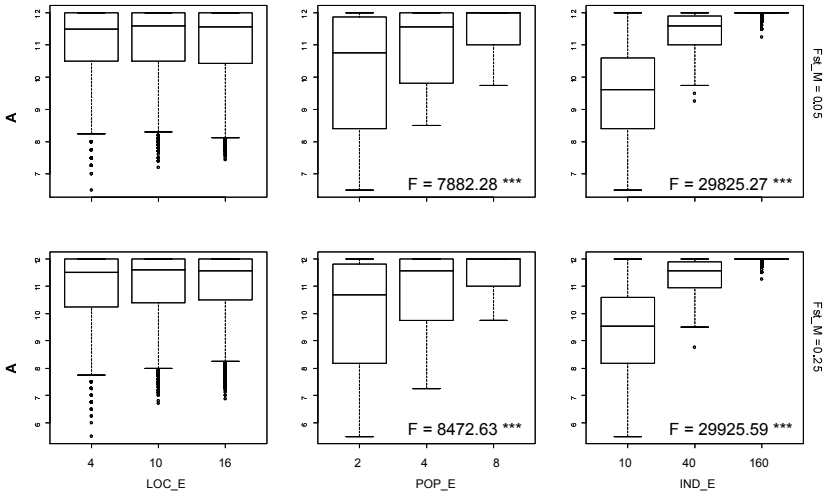


Figure 4 : Effet de l'échantillonnage sur Qst – cas a.
Idem Figure 3 pour Qst avec Qst_M = valeur du Qst de la « vraie » métapopulation.

Enfin, la figure 5 (a, b) présente l'effet de l'échantillonnage sur les valeurs de Ae et A : le nombre de populations et d'individus échantillonnés ont un effet très significatif sur ces deux valeurs, la valeur du Fst_M jouant uniquement sur la valeur de Ae. L'analyse des résultats dans le cas b où les loci sont différents entre eux quant à la distribution des fréquences alléliques et au nombre d'allèles montre un effet significatif du nombre de loci neutres échantillonnés sur Fst, A et Ae (fig. 6) mais l'effet du nombre de populations et d'individus échantillonnés reste prépondérant. Dans le cas c (non illustré), l'effet du nombre de loci échantillonnés est significatif pour A et Ae mais pas pour Fst. Le coefficient de variation du Fst fixé par locus au départ n'a pas d'effet significatif sur A, Ae et Fst.



5a



5b

Figure 5 : Effet de l'échantillonnage sur Ae et A – cas a.
Idem Figure 3 pour A et Ae.

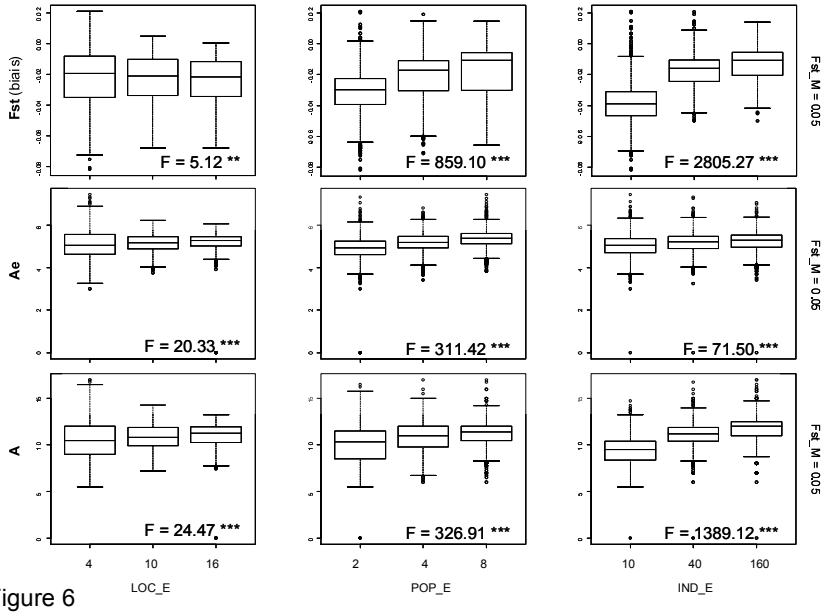


Figure 6

Figure 6 : Effet de l'échantillonnage sur A, Ae et Fst dans le cas b.
Diagramme des biais du Fst et des valeurs moyennes de A et Ae (avec la précision) pour différents niveaux d'échantillonnage des loci (LOC_E), des populations (POP_E), des individus (IND_E) [Fst de la « vraie » métapopulation, Fst_M, fixé à 0,05]. Les valeurs minimales et maximales de chaque carré correspondent aux quartiles inférieurs et supérieurs, la moitié correspondant à la valeur médiane. Pour les facteurs significatifs, le calcul du F provenant d'une ANOVA à 3 facteurs (POP_E x IND_E x LOC_E) est donné.

4. DONNÉES EXPÉRIMENTALES EN POPULATIONS NATURELLES ET *EX SITU*

Deux espèces forestières combinant reproduction sexuée et asexuée ont été étudiées : le merisier (*Prunus avium* L., Rosaceae, 2n = 16) espèce entomophile disséminée dans toute la France ; le peuplier noir (*Populus nigra* L., Salicaceae, 2n = 38) espèce dioïque, anémophile et pionnière, inféodée aux ripisylves peu anthropisées. Chez le merisier, l'allogamie est favorisée par un système d'auto incompatibilité gamétophytique. Ces espèces peuvent avoir des densités très variables et représenter de 0 à presque 100 % de la composition d'un peuplement. Elles possèdent toutes les deux un compartiment sauvage et cultivé, ce dernier comportant de nombreuses variétés, fruitières pour le merisier, ornementales et à finalité de production de bois pour le peuplier, introduites à grande échelle par l'homme.

4.1. Échantillonnage des populations et des collections

Pour chaque espèce, six populations naturelles et une collection ont été échantillonnées (voir fig. 7 et 8). Les résultats présentés dans cet article portent sur quatre des six populations (avec un échantillonnage de 50 individus par population) pour le merisier et sur la collection de 308 individus pour le peuplier (fig. 7 et 8).

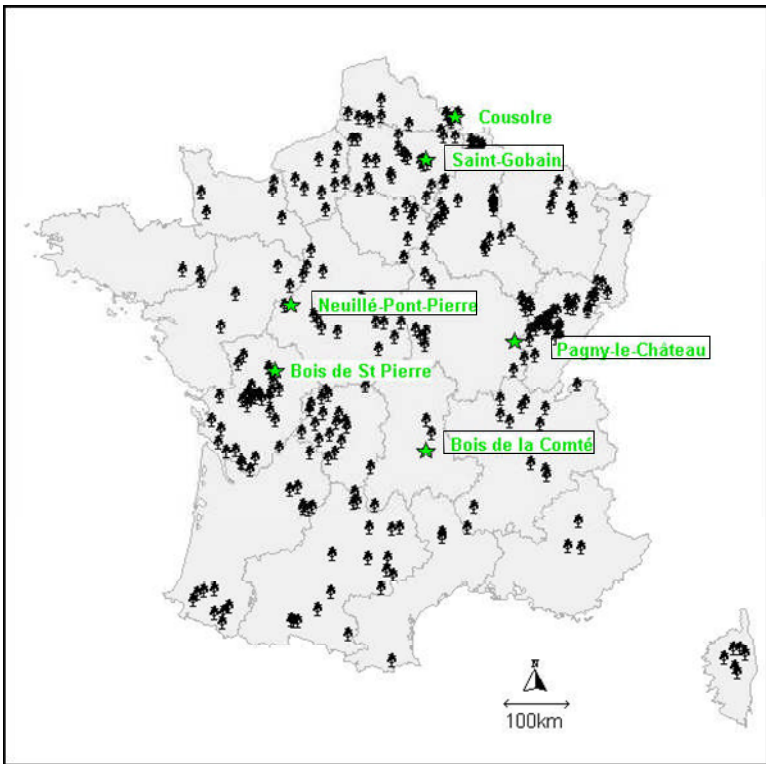


Figure 7 : Carte de répartition des échantillonnages pour le merisier.
- Localisation géographique (symbolisée par des arbres) des provenances géographiques représentées dans la collection française de merisier.
- Localisation des 6 populations naturelles échantillonnées (symbolisées par des étoiles) ; les populations analysées (50 individus par population) sont encadrées.

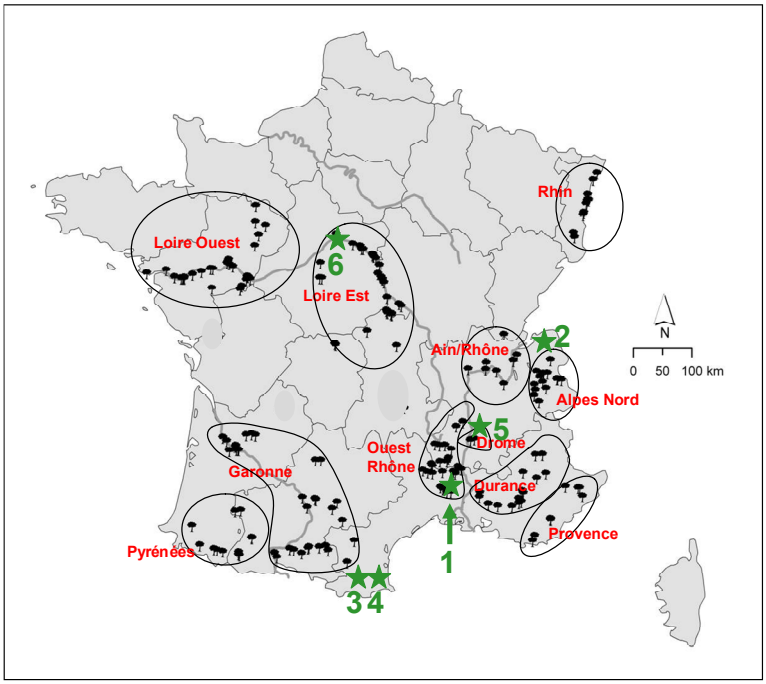


Figure 8 : Carte de répartition des échantillonnages pour le peuplier.
- Localisation géographique des 11 provenances géographiques (symbolisées par des arbres) représentées dans la collection française de peuplier noir par 308 clones.
- Localisation des 6 populations naturelles échantillonnées (symbolisées par des étoiles, 1 : Ardèche ; 2 : Dranse ; 3 : Eyne ; 4 : Nohèdes : 5 : Ramières ; 6 : St-Pryvé – St-Mesmin).

4.2. Échantillonnage et amplification des marqueurs

Pour le merisier, 8 marqueurs microsatellites ont été utilisés et les allèles du système d'incompatibilité du locus S, ont été caractérisés. Pour le peuplier noir, 12 marqueurs microsatellites, dont deux marqueurs associés à des QTL de résistance à la rouille foliaire à *Melampsora larici-populina* : PMGC667 chez *P. trichocarpa* [11] et PMGC433 chez *P. nigra* (Storme, comm. pers.) ont été utilisés (voir tabl. I).

Pour les deux espèces, l'ADN génomique total a été extrait des feuilles en utilisant le kit d'extraction « Dneasy® Qiagen 96 plant kit ». Les marqueurs microsatellites ont été amplifiés par PCR en utilisant des amorces marquées par des fluorophores (HEX, FAM et NED) selon les protocoles décrits par Stoeckel *et al.* [17] pour le merisier et Jorge *et al.* [11] pour le peuplier. Les allèles ont été séparés par migration dans un séquenceur capillaire

(ABI PRISM ® 3100 Genetic Analyzer). Les génotypes ont été visualisés et lus avec deux logiciels : GENOTYPER ® et/ou GENESCAN ®.

Tableau I : Marqueurs utilisés et localisation sur les groupes de liaison pour chaque espèce.

Marqueurs peuplier	Référence	GL	**A**	**Ae**	**Fst**	Marqueurs merisier	Référence	GL	**A**	**Ae**	**Fst**
PMGC61	[2]	VIII	13	6,99	0,0199	UDP96-005	[3]	I	9	4,3	0,044
PMGC2578	[2]	VI	27	9,55	0,0201	UDP98-411	[18]	II	8	3,1	0,049
PMGC2838	[2]	V	13	1,70	0,0435	PCEGA34	[8]	II	20	6,0	0,067
PMGC2852	[2]	I	21	5,51	0,0284	BPPCT034	[5]	II	14	4,0	0,07
PMGC333	[2]	XI	10	3,16	0,0272	BPPCT040	[5]	IV	8	3,7	0,088
PMGC486	[2]	III	18	7,014	0,0248	PS12A02	[16] [10]	IV	11	2,5	0,084
PMGC2156	[2]	V	14	1,709	0,0356	UDP98-021	[18]	VI	4	2,6	0,124
WPMS05	[2]	XII	24	8,82	0,0265	UDP98-412	[18]	VI	8	4,0	0,047
PMGC433	[2]	XVI	12	2,21	0,0281	S	[15]	VI	19	9,1	0,035
PMGC2235	[2]	IV	19	6,31	0,0351						
PMGC14	[2]	XIII	13	4,54	0,0228						
PMGC667	[2]	II	48	10,97	0,0331						
Moyenne			**19,33**	**5,71**	**0,0277**	**[sans S]**			**10,25**	**4,63**	**0,071**
Écart-type			10,43	3,13	0,002				4,86	1,61	0,009

Valeurs de A, Ae et Fst estimées sur les 308 clones de peuplier noir structurés en 11 unités géographiques, pour 12 marqueurs microsatellites, et sur les 4 populations naturelles de merisier (50 individus échantillonnés par population), pour 8 marqueurs microsatellites et le marqueur S.

4.3. Analyse phénotypique de la résistance à la rouille foliaire du peuplier dans la collection

Pour le peuplier, une étude de la structuration géographique de la variabilité génétique de la résistance partielle à la rouille foliaire a été conduite pour deux isolats de *M. larici-populina* inoculés artificiellement en laboratoire [7]. Trois composantes de la résistance partielle ont été évaluées : la latence infectieuse (LP) correspondant au nombre de jours depuis l'inoculation jusqu'à l'apparition du premier urédosore sporulant, le nombre d'urédosores sporulants (UN) au 13e jour après inoculation et la taille relative des urédosores (US) au 14e jour après inoculation, notée à l'aide d'une échelle à 5 classes puis calibrée par analyse d'images de 15 disques par classe.

4.4. Analyse des données

Les valeurs de A et Ae par locus sur l'ensemble des populations, leur moyenne et leur écart-type sur l'ensemble des loci microsatellites ont été calculés. Le Fst de Weir et Cockerham [20] pour chaque marqueur, sa valeur pour l'ensemble des microsatellites et l'écart-type de cette valeur par jackknifing sur les loci ont été estimés (Fstat v2.9.3.2 [9]). En ce qui concerne le merisier, les calculs ont été effectués en ne prenant qu'un seul représentant par tache de drageonnage supposée. Pour les données phénotypiques, la statistique Qst a été estimée de la façon suivante : $Qst = \sigma^2_b / (\sigma^2_b + \sigma^2_w)$ où σ^2_b et σ^2_w représentent respectivement la variance entre populations et la variance clonale [14].

4.5. Résultats

Pour le merisier, les résultats obtenus sur les quatre populations naturelles avec 8 marqueurs microsatellites et le marqueur du locus S donnent un A observé variant de 4 à 20 et une moyenne de 10,25 allèles par locus. Pour le peuplier, les résultats obtenus sur l'ensemble de la collection de référence et sur les 12 marqueurs microsatellites donnent un A observé qui varie de 12 à 48, avec une moyenne de 19,33 allèles par locus. La valeur du Fst aux marqueurs microsatellites est de 0,071 (± 0,009) pour le merisier et de 0,028 (± 0,002) pour le peuplier (tabl. I). Pour les données phénotypiques obtenues sur la collection nationale, une héritabilité individuelle h^2_L (σ^2_G/σ^2_P) élevée reflète l'existence d'un contrôle génétique important pour l'ensemble des composantes de la résistance partielle à la rouille foliaire. Comme chez d'autres espèces de peuplier [7], c'est pour le niveau moyen de sporulation (taille des urédosores) que la variabilité génétique individuelle est la plus élevée chez le peuplier noir. Cependant, aucune structuration géographique significative ne peut être mise en évidence pour aucune des composantes étudiées (tabl. II).

Tableau II : Variabilité génétique et différenciation géographique pour différentes composantes de la résistance à la rouille foliaire (*M. larici-populina*) au sein de la collection de référence de peuplier noir.

Résistance à E1-98ID6	h^2_L	Test F effet population	**Qst**	**Résistance à E4-98AG69**	h^2_L	Test F effet population	**Qst**
Latence infectieuse	0,42	1,16 ns	0,001	Latence infectieuse	0,29	1,45 ns	0,001
Nb urédosores	0,37	1,56 ns	0,002	Nb urédosores	0,33	1,25 ns	0,001
Taille des urédosores	0,55	0,59 ns	0,0001	Taille des urédosores	0,44	0,30 ns	0,0001

5. DISCUSSION

5.1. Résultats théoriques

L'analyse des stratégies d'échantillonnage a montré que les facteurs prépondérants à prendre en compte pour capturer au mieux la diversité génétique d'une espèce et sa structuration entre populations était le nombre de populations et le nombre d'individus échantillonnés. Le nombre de loci, lui, n'a un effet que lorsqu'il existe de la variance entre loci mais ce dernier reste moins important que les facteurs d'échantillonnage démographique. Le nombre de populations échantillonnées a un effet d'autant plus fort sur le Fst et le Ae que la structuration de la métapopulation considérée est plus forte. Le nombre d'individus échantillonnés est important lorsque la structuration de la diversité n'est pas très marquée entre populations, comme c'est le cas généralement chez les espèces forestières. Il n'a pas d'effet sur le Qst, du fait de l'hypothèse forte du modèle sur la connaissance de la valeur génotypique individuelle.

5.2. Stratégie d'échantillonnage expérimentale et résultats théoriques

Le tableau III compare les valeurs des variables génétiques, Fst, Qst, Ae et A obtenues à partir de notre modèle (dans les cas a et b) pour une stratégie d'échantillonnage similaire à celle utilisée pour l'estimation de la diversité génétique des populations naturelles de peuplier et de merisier. Les deux cas donnent des résultats proches, ce qui indique que le choix des marqueurs n'est pas capital à condition bien sûr qu'ils aient un minimum de polymorphisme. Ainsi, dans le cas b, lorsque l'écart-type du nombre d'allèles de chaque locus est comparable à celui estimé chez le peuplier et le merisier, l'estimation du Fst reste assez proche du cas a. Globalement, la distribution des valeurs de Fst montre qu'au moins 95 % des valeurs sont supérieures à 0, celle des valeurs de Qst a elle une amplitude plus large, en particulier pour des valeurs de Qst de la métapopulation égale à 0,05, plus difficiles à détecter. La stratégie d'échantillonnage choisie permet de capturer un nombre efficace d'allèles inférieur de 5,2 à 8,7 % à celui calculé sur la « vraie » métapopulation et un nombre total d'allèles de l'ordre de 94 %. Par conséquent, la stratégie d'échantillonnage adoptée pour les deux espèces à faible différenciation génétique semble permettre simultanément une bonne estimation de la différenciation inter - populations et de la richesse allélique intra - population. Toutefois, il reste encore à les examiner plus en détail, en particulier au niveau des populations échantillonnées relativement aux populations réelles dont le nombre et la délimitation sont très difficiles à connaître.

Tableau III : Statistiques des courbes de distribution des valeurs de Fst, Qst, A et Ae pour une stratégie d'échantillonnage correspondant à 4 populations sur 10, 40 individus par population sur 200 et 10 loci neutres sur 20, en fonction de plusieurs « vraies » métapopulations considérées et caractérisées par un couple (Fst_M, Qst_M).
(1) : Fst_M = 0,05, (2) : Fst_M = 0,15.

	(1)	(2)				(1)	(2)	(1)	(2)
	Fst		**Qst**			**Ae**		**A**	
Cas a (loci identiques)									
percentile 2,5%	0,023	0,093	-0,006	0,04	0,042	4,543	4,131	11,1	11,2
moyenne	0,033	0,121	0,051	0,137	0,237	4,922	4,725	11,577	11,589
Valeur « vraie »	***0,05***	***0,15***	***0,05***	***0,15***	***0,25***	***5,07***	***5,07***	***12***	***12***
percentile 97,5%	0,046	0,15	0,125	0,267	0,388	5,268	5,167	11,8	11,9
Cas b (loci différents)									
percentile 2,5%	0,021	0,089	-0,002	0,013	0,066	4,494	4,342	9,3	9,7
moyenne	0,034	0,122	0,049	0,145	0,244	5,228	5,037	11,277	11,412
Valeur « vraie »	***0,05***	***0,15***	***0,05***	***0,15***	***0,25***	***3,84 à 7,52***	***3,84 à 7,52***	***6 à 18***	***6 à 18***
percentile 97,5%	0,046	0,147	0,115	0,277	0,419	5,858	5,858	13	13,3

6. PERSPECTIVES

Concernant l'approche théorique, plusieurs points restent à préciser. Une analyse complémentaire non paramétrique des effets des différents facteurs permettra de valider les premiers résultats obtenus par analyse de variance. L'étude des stratégies d'échantillonnage à ressources constantes permettra d'évaluer séparément l'effet du nombre de populations échantillonnées et l'effet du nombre d'individus, ces deux facteurs étant confondus dans les analyses présentées. Côté expérimental, les analyses complémentaires de diversité génétique chez le peuplier et le merisier permettront une comparaison entre les populations naturelles et les collections. Il s'agira alors de confronter l'échantillonnage existant des ressources génétiques dans les collections avec un échantillonnage théorique permettant à la fois de capturer la diversité et la structuration de cette diversité.

RÉFÉRENCES

[1] Bataillon T.M., David J.L., Schoen D.J., Neutral genetic markers and conservation genetics: simulated germplasm collections, Genetics 144 (1996) 109-117.

[2] Cervera M.T., Storme V., Ivens B. *et al.* Dense genetic linkage maps of three populus species (*Populus deltoides*, *P, nigra* and *P, trichocarpa*) based on AFLP and microsatellite markers, Genetics 158 (2001) 787-809.

[3] Cipriani G., Lot G., Huang W. *et al.* AC/GT and AG/CT microsatellite repeats in peach (*Prunus persica* L, Batsch) isolation, characterisation and cross-species amplification in *Prunus*, Theor. Appl. Genet. 99 (1999) 65-72.

[4] Clarke J.B., Tobutt K.R., Development and characterization of polymorphic microsatellites from *Prunus avium* "Napoleon", Mol. Ecol. Notes 3 (2003) 578-580.

[5] Dirlewanger E., Cosson P., Tavaud M. *et al.* Development of microsatellite markers in peach *Prunus persica* (L,) Batsch and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L,), Theor. Appl. Genet. 105 (2002) 127-138.

[6] Dirlewanger E., Graziano E., Joobeur T. *et al.* Comparative mapping and marker-assisted selection in Rosaceae fruit crops, Proc. Nat. Ac. Sci. 101 (2004) 9891-9896.

[7] Dowkiw A., Husson C., Frey P. *et al.*, Partial resistance to *Melampsora larici-populina* leaf rust in hybrid poplars: genetic variability in inoculated excised leaf disk bioassay and relationship with complete resistance, Phytopath. 93 (2003) 421-427.

[8] Downey S.L., Iezzoni A.F., Polymorphic DNA markers in black cherry (*Prunus serotina*) are identified using sequences from sweet cherry, peach, and sour cherry, J. Am. Soc. Horti. Sci. 125 (2000) 76-80.

[9] Goudet J., FSTAT (Version 1,2): A computer program to calculate F-statistics, J. Hered. 86 (1995) 485-486.

[10] Joobeur T., Periam N., Vicente M.C., King G.J., Arus P., Development of a second generation linkage map for almond using RAPD and SSR markers, Genome 43 (2000) 649-655.

[11] Jorge V., Dowkiw A., Faivre-Rampant P. *et al.*, Genetic architecture of qualitative and quantitative *Melampsora larici-populina* leaf rust resistance in hybrid poplar: genetic mapping and QTL detection, New Phytol. 167 (2005) 113-127.

[12] Kirkpatrick S., Gellat C.D.J., Vecchi M.P., Optimization by simulated annealing, Science 220-4598 (1983) 671-680.

[13] Mc Kay J.K., Latta R.G. Adaptive population divergence: markers, QTL and traits, Trends Ecol. Evol. 17 (2002) 285-291.

[14] Spitze K., Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation, Genetics 135 (1993) 367-374.

[15] Sonneveld T., Robbins T.P., Tobutt K.R., Improved discrimination of self-incompatibility S-RNase alleles in cherry and high throughput genotyping by automated sizing of first intron PCR products, Plant Breeding (sous presse).

[16] Sosinski B., Gannavarapu M., Hager L.D. *et al.* Characterization of microsatellite markers in peach *Prunus persica* (L,) Batsch, Theor. Appl. Genet. 101 (2000) 421-428.

[17] Stoeckel S., Grange J., Fernandez-Manjarres J.F., Bilger I., Frascaria-Lacoste N., Mariette S., Heterozygote excess in a self-incompatible and partially forest tree species – *Prunus avium* L, Mol. Ecol. (sous presse).

[18] Testolin R., Marrazzo T., Cipriani G. *et al.* Microsatellite DNA in peach (*Prunus persica* L, Batsch) and its use in fingerprinting and testing the genetic origin of cultivars, Genome 43 (2000) 512-520.

[19] Vaughan S.P., Russell K., Characterization of novel microsatellites and development of multiplex PCR for large-scale population studies in wild cherry, *Prunus avium*, Mol. Ecol. Notes 4 (2004) 429-431.

[20] Weir B.S., Cockerham C.C., Estimating F-statistics for the analysis of population structure, Evolution 38 (1984) 1358-1370.