

THÈSE

présentée devant l'Université Claude Bernard - LYON I

pour l'obtention du
Diplôme de Doctorat
(arrêté du 30 mars 1992)

Soutenue publiquement le 18 juin 2002

par

Nathaniel VALIÈRE

**Amélioration et optimisation des méthodes
non-invasives et des marqueurs microsatellites en
Biologie des Populations et de la Conservation**

VOLUME 1 : TEXTE PRINCIPAL

Jury

Rapporteurs : Stéphane AULAGNIER
Michael BRUFORD

Président : Dominique MOUCHIROUD (co-directrice)

Examineurs : Catherine HÄNNI
Dominique PONTIER (co-directrice)
Roger PRADEL
François SARRAZIN

UMR CNRS 5558
Laboratoire de Biométrie et Biologie Evolutive
Université Claude Bernard LYON I - Bât 711
43, boulevard du 11 novembre 1918
69622 Villeurbanne

Mis en page avec la classe thloria.

REMERCIEMENTS

Je voudrais tout d'abord remercier tous les membres du jury qui ont accepté de juger ce travail, même si parfois leur domaine de compétences ne correspondait pas complètement au sujet de la thèse. Cela n'a été que plus enrichissant.

Je voudrais ensuite remercier Dominique PONTIER et Dominique MOUCHIROUD pour m'avoir accueilli en cours de route afin de finir ma thèse dans de bonnes conditions.

Que le laboratoire de Biologie de Populations d'Altitude (Grenoble) soit aussi remercié pour m'avoir permis de commencer ce travail sur les marqueurs microsatellites et la population de loup française et pour les deux premières années de ma thèse. J'y ai beaucoup appris !

Je voudrais aussi remercier les personnes impliquées dans la partie lupine de mon travail :

- le Parc National du Mercantour, plus particulièrement Benoît LEQUETTE et les gardes-moniteurs Vésubie-Tinée.
- Gérard MILLISCHER pour son amitié et sa gentillesse
- Marie-Lazarine POULLE pour les discussions et ses encouragements

Je voudrais remercier Jacques, Véro et Philippe GAUDIN pour la relecture et la correction de certaines parties du manuscrit.

Je voudrais enfin remercier tous mes amis (je ne peux les citer tous ici mais ils se reconnaîtront) pour tout ce que l'on a partagé et que l'on partagera, et les « Zaffreux de la 108 » pour les délires et autres décrochages !

Enfin, je remercie ma famille pour m'avoir soutenu tout au long de mes études et m'avoir permis d'arriver jusque là.

à l....

*All that is gold does not glither,
Not all those who wander are lost;
The old that is strong does not wither,
Deep roots are not reached by the frost.
From the ashes a fire shall be woken,
A light from the shadows shall spring;
Renewed shall be blade that was broken,
The crownless again shall be king.*

*Bilbo, Son of Bungo, BAGGINS
In The Lord of The Rings - Book One
J.R.R. TOLKIEN*

RÉSUMÉ

Les récentes avancées en Biologie Moléculaire ont entraîné également des avancées en Biologie des Populations et de la Conservation. La mise au point de la *Polymerase Chain Reaction* (PCR) a ainsi permis d'utiliser des échantillons ne contenant qu'une très faible quantité d'ADN. On peut désormais récolter des fèces ou des poils pour en extraire de l'ADN et pouvoir typer génétiquement les échantillons grâce, par exemple, à des marqueurs microsatellites. Cet échantillonnage non-invasif constitue une stratégie intéressante pour étudier des populations ou des espèces discrètes, rares ou en danger. Néanmoins, les faibles quantité et qualité de l'ADN associées aux inhibiteurs de PCR peuvent entraîner des erreurs de génotypage et ainsi biaiser l'identification des individus. Aussi, des mises au point et des optimisations doivent être menées avant de commencer une étude à grande échelle.

Nous allons voir dans ce mémoire qu'il est possible d'optimiser l'utilisation des échantillons non-invasifs et des microsatellites. Ceci sera réalisé autour des trois grandes étapes d'une étude : l'échantillonnage, le génotypage et l'analyse des données. A chacune de ces étapes, des articles seront présentés. Nous aborderons ainsi :

- l'utilité de l'échantillonnage non-invasif pour étudier la colonisation du loup (*Canis lupus*) en France et en Suisse,
- la mise au point de l'extraction d'ADN à partir d'urine récoltée dans la neige,
- le développement d'un logiciel permettant de déterminer la stratégie d'une étude de population basée sur l'identification individuelle,
- l'utilité de la répétition des amplifications PCR pour chaque échantillon et locus (approche multitubes) pour confirmer les génotypes (application avec le logiciel sus-cité),
- un exemple de l'utilisation du logiciel sus-cité sur un jeu de données réelles,
- une comparaison (basée sur des simulations avec le logiciel sus-cité) des méthodes d'estimation de la taille de population,
- le développement d'un logiciel d'analyse des données d'identification individuelle.

Mots-clés: Echantillonnage non-invasif, microsatellites, génotypage, erreurs, approche multitubes, étude pilote.

ABSTRACT

Recent advances in molecular biology has made possible advances in population and conservation biology. The developpment of the *Polymerase Chain Reaction* (PCR) has allowed the use of samples containing small amount of (degraded) DNA. It is now possible to collect faeces or hairs, extract DNA and then to genetically type samples using, for example, microsatellite markers. Such noninvasive genetic sampling becomes an interesting strategy to study populations for elusive, discrete or endangered species. Nevertheless, low DNA quantity and quality, associated with PCR inhibitors can cause genotyping errors and thus can lead to bias in the individual identification. Optimization must be conducted before proceeding to a large scale study.

In this thesis, we will explore the possibility of developing of noninvasive sampling and microsatellites markers. The ideas will be discussed around the three main stages of a study : sampling, genotyping and analysis of data. We will see :

- the usefullness of the noninvasive sampling for studying the colonization of France and Switzerland by wolf (*Canis lupus*),
- the developpement of DNA extraction from urine samples collected in the field,
- the developpement of a software for determining the optimal strategy for a population study,
- the usefullness of the PCR replications to confirm genotype for each sample and locus,
- an example of the use of the software with a real dataset,
- a simulation-based comparison of methods for estimating population size,
- the developpement of a software to analyse individual identification data.

Keywords: noninvasive sampling, microsatellites, genotyping, errors, multitubes approach, pilot study.

Sommaire

Résumé	v
Abstract	vii
Table des figures	xv
Liste des tableaux	xix
Liste des articles présentés au cours de ce mémoire	xxi

Volume 1 : Texte principal

Introduction Générale	3
-----------------------	---

Partie I Introduction et Contexte Général

Chapitre 1 La biologie et la génétique de la conservation	7
1.1 La biologie de la conservation	8

1.2	La génétique de la conservation	10
1.3	Les marqueurs génétiques : description et applications	11
1.3.1	Présentation des outils	13
1.3.1.1	Les allozymes	13
1.3.1.2	La PCR (Polymerase Chain Reaction)	13
1.3.1.3	RFLP (Restriction Fragment Length Polymorphism)	14
1.3.1.4	Séquençage d'ADN	14
1.3.1.5	SSCP (Single-Strand Confirmational Polymorphism) et MHC (Major Histocompatibility complex)	15
1.3.1.6	RAPD (Random Amplification of Polymorphic DNA)	16
1.3.1.6.1	AFLP (Amplified Fragments Length Polymorphism)	17
1.3.1.7	VNTR (Variable Number of Tandem Repeats)	18
1.3.2	Applications et applicabilité des méthodes génétiques	20
1.3.2.1	Niveau populationnel : structure des populations	22
1.3.2.2	Niveau familial et du groupe social: la détermination des parentés	23
1.3.2.3	Niveau individuel : identification et suivi des individus	24
1.3.3	Avantages et inconvénients des microsatellites	25
1.3.3.1	Mise au point et développement	25
1.3.3.2	Révélation des marqueurs	25
1.3.3.3	Amplification PCR	26
1.3.3.4	Interprétation et utilisation des résultats	26
Chapitre 2 L'échantillonnage non-invasif		29
2.1	Les différentes formes de méthodes d'échantillonnage non-invasif	30
2.2	Les méthodes non-invasives et la génétique	31
2.2.1	Rendement d'amplification	31
2.2.2	Erreurs de génotypage liées à la quantité et la qualité de l'ADN	32
2.2.2.1	Révélation de fragments surnuméraires	33
2.2.2.2	Non révélation de fragments	34
2.2.2.3	Autres erreurs	36
2.2.2.4	Taux d'erreurs	36
Chapitre 3 Perspectives de travail et Plan du mémoire		39
3.1	Au niveau de l'échantillonnage	40
3.2	Au niveau du génotypage des échantillons	41
3.3	Au niveau du traitement des données issues du génotypage microsatellites	43

Partie II Optimisation des méthodes non-invasives et du génotypage des microsatellites

Chapitre 4 Mise en place, optimisation et amélioration de l'échantillonnage non-invasif	47
4.1 Utilisation des échantillons non-invasifs dans l'étude préliminaire de l'expansion du loup en France et en Suisse par séquençage de l'ADN mitochondrial (Article I)	48
4.2 Mise au point et utilisation d'échantillons d'urine récoltés sur le terrain (Article II)	49
4.3 Discussion	49
4.3.1 Méthodologie et logistique	49
4.3.2 Contraintes de structuration et de spatialisation	50
Chapitre 5 Optimisation de l'identification individuelle par génotypage microsatellite	53
5.1 Le logiciel GEMINI (Article III)	54
5.1.1 Modèle utilisé dans GEMINI et paramètres	54
5.1.2 Analyser les effets des erreurs et optimiser l'approche multi-tubes	56
5.1.3 Analyse des données	58
5.2 Applications de GEMINI	59
5.2.1 Nombre de répétitions et nombre de loci à utiliser dans une étude : détermination et illustration par la simulation (Article IV)	59
5.2.2 Détermination du nombre minimum de répétitions du génotypage : applications aux données de GARNIER <i>et al.</i> (2001)	61
5.3 L'optimisation du génotypage à d'autres étapes de l'analyse	62
5.3.1 Conditions de récolte et de préservation des échantillons	62
5.3.2 Extraction d'ADN	64
5.3.3 Amplification	65
5.3.4 Précautions lors de l'extraction et de l'amplification	67
5.3.5 Choix des marqueurs	68
5.3.6 Analyse et interprétation des gels	69

Chapitre 6 Analyse des données des génotypes microsatellites	71
6.1 L'estimation de l'effectif d'une population et stratégies d'échantillonnage	72
6.2 Traitement des données microsatellites : le logiciel GIMLET	74
6.2.1 Description des tâches et comparaison avec d'autres logiciels d'analyse	74
6.2.2 Description de modules	76
6.2.2.1 Le module « <i>Identification</i> »	76
6.2.2.1.1 L'identification et la comparaison des génotypes	76
6.2.2.1.2 la détermination des parentés potentielles entre individus	77
6.2.2.2 Le module « <i>Calculator</i> »	77
6.2.2.2.1 La construction des génotypes consensus et l'estimation des taux d'erreurs à partir des génotypes multilocus répétés	78
6.2.2.2.2 L'estimation des fréquences alléliques, des taux d'hétérozygotie et des probabilités d'identité à partir des génotypes	82
6.2.2.2.3 L'estimation de l'effectif de population	84

Partie III Discussion et Conclusion

Chapitre 7 L'étude pilote: une étape indispensable avant une étude à grande échelle	87
7.1 Tester les méthodes sur le terrain	88
7.2 Tester les analyses génétiques à partir des échantillons non-invasifs	88
7.2.1 Test de l'amplification d'ADN à partir des échantillons	89
7.2.2 Estimation des taux d'erreurs, des fréquences alléliques et de la probabilité d'identité	89
7.2.2.1 Nombre de répétitions par échantillon et par locus dans l'étude pilote	91
7.2.2.2 Combien d'échantillons récolter pour l'étude pilote ?	91
7.3 Détermination de la stratégie optimale	91
Chapitre 8 Conclusion Générale	93
Références	101

Volume 2 : Annexes et Articles

Annexes	121
Annexe A Identification individuelle chez le loup	123
A.1 Etude du loup en France	124
A.1.1 Situation	124
A.1.2 Pourquoi une étude génétique ?	124
A.2 Identification individuelle dans la population de loup	125
A.2.1 Choix du système de loci microsatellites	125
A.2.2 Sélection des loci « fiables »	126
A.2.3 Probabilité d'identité	127
A.2.3.1 Calcul de la probabilité d'identité pour les loups	127
A.2.3.2 Application des loci aux échantillons de fèces	128
A.2.4 Sexage des individus	130
Annexe B La biologie de la conservation: quelques précisions	133
B.1 Extinctions et émergence de la biologie de la conservation	134
B.1.1 Diversité et extinctions	134
B.1.2 Déclin et extinctions	136
B.1.2.1 Facteurs naturels influençant les déclin et les extinctions	137
B.1.2.2 Les facteurs d'origine humaine	137
B.1.2.3 Le rôle prépondérant de la taille de population	138
B.1.2.3.1 Stochasticité démographique	139
B.1.2.3.2 Stochasticité environnementale	139
B.1.2.3.3 Stochasticité génétique	139
B.1.2.3.3.1 La dérive génétique	139
B.1.2.3.3.2 Consanguinité et dépression de consanguinité	140
B.1.2.3.3.3 Notion de taille efficace génétique	140
B.2 L'émergence de la Conservation	141
B.2.1 Pourquoi et comment conserver ?	143
B.2.1.1 Pourquoi conserver ?	143
B.2.1.2 Les moyens de conservation	143
B.2.1.2.1 Les aires de protection	144

B.2.1.2.1.1 La conservation <i>ex situ</i> des ressources	144
B.2.1.2.1.2 Les mesures légales	146
B.2.1.2.1.3 Les programmes de restauration	147
B.3 Le rôle de la biologie en conservation	147
B.3.1 Diagnostic et prescription	148
B.3.1.1 L'apport de la biogéographie insulaire et du concept de métapopulation	149
B.3.1.2 Minimum Viable Population et Population Viability (Vulnerability)	
Analysis	150
B.3.2 Informations biologiques	151
B.4 La génétique de la conservation	152
B.4.1 Définition de la taxonomie et des unités de conservation	152
B.4.2 Diagnostiquer les problèmes génétiques et donner des priorités de conservation	155
B.4.2.1 La variabilité génétique	155
B.4.2.2 La consanguinité	156
B.4.2.3 L'effectif et la taille efficace d'une population	156
B.4.2.4 Le problème des hybrides	158
B.4.3 Obtention d'informations biologiques	159
Références	161
Article I	171
Article II	173
Article III	175
Article IV	177
Article V	179
Article VI	181
Article VII	183
Article VIII	185
Index	187

Table des figures

1.1	Intégration des facteurs provoquant et accélérant l'extinction de populations. Certains facteurs peuvent entraîner des diminutions d'effectif, la fragmentation du milieu, des changements de régimes de prédation, la diminution des ressources. Ces facteurs peuvent provoquer la disparition directe de la population, ou entraîner cette population dans une spirale (ou vortex GILPIN et SOULÉ 1986). L'effectif de la population se réduit alors progressivement jusqu'à (éventuellement) l'extinction de la population. Modifié d'après BARBAULT (1997, p.77) et FRANKHAM <i>et al.</i> (2002, p.32)	8
1.2	Diagrammes des problématiques posées en génétique de la conservation. D'après FRANKHAM <i>et al.</i> (2002), p.17.	12
1.3	Schéma d'un cycle de PCR illustrant les trois phases : dénaturation, hybridation et élongation. A partir d'une molécule d'ADN en début de cycle, on obtient, en fin de cycle, deux molécules d'ADN, identiques entre elles et avec la molécule de départ.	14
1.4	Principe des RFLP. Les sites de restriction de l'enzyme utilisés sont différents entre les deux brins d'ADN. La migration des brins générés par le découpage de l'ADN par l'enzyme donne, sur un gel, les résultats présentés à droite.	15
1.5	Principe du séquençage basé sur la réaction de Sanger. Ici, l'utilisation de quatre marquages fluorescents permet de séquencer un fragment en une seule piste. En bas et à gauche, un image d'un gel de séquence (logiciel GENESCAN, APPLIED BIOSYSTEMS PERKIN ELMER) où une piste (ligne verticale de couleur) correspond à une séquence.	16
1.6	Schéma général de la méthode de détection du polymorphisme de conformation simple brin (SSCP) avec en comparaison la migration des doubles brins pour les mêmes fragments d'ADN.	17
1.7	Principe de l'AFLP montrant les quatre étapes de la méthode : (1) Digestion (deux enzymes de restriction), (2) Ligation (ADN ligase) des adaptateurs spécifiques aux sites de restriction, (3) PCR pré-sélective et (4) PCR sélective.	18
1.8	Principe des marqueurs minisatellites. (a) représentation schématique de la procédure de la révélation du polymorphisme des minisatellites. (b) autoradiogramme montrant les bandes révélées pour une sonde appliquée chez 11 hommes (d'après JEFFREYS <i>et al.</i> 1985).	19

1.9	Principe des marqueurs microsatellites ; En haut : représentation de la structure des microsatellites pour un locus et deux individus et résultats sur gel d'agarose après amplification ; En bas : représentation des pics de fluorescence (une couleur par locus microsatellite) sur un gel d'acrylamide après analyse de ce gel. Chaque pic correspond à un fragment amplifié d'un locus (allèle dont la longueur est indiquée). A droite, sont indiqués les génotypes (ensemble des longueurs des fragments amplifiés pour chaque locus) des deux individus avec en gras les allèles différenciant ces individus.	20
2.1	Dispositifs de « capture » de poils pour deux types de mammifères : (a) et (b) deux dispositifs adaptés aux grizzlys (<i>Ursus arctos horribilis</i>). Modifié d'après WOODS <i>et al.</i> (1999) ; (c) dispositif adapté à la martre américaine (<i>Martes americana</i>). En bas : positionnement du dispositif sur un arbre. En haut : vue interne du dispositif. Dessin réalisé à partir des photographies dans (FORAN <i>et al.</i> 1997b).	32
2.2	Exemple de faux allèle (5 ^e piste) lors du typage multiple d'un locus pour un échantillon d'urine de canidé. Dans ce cas, le faux allèle correspond à la production d'un « allèle » de 2 paires de bases plus grand que celui attendu (+ le fragment associé, décalé d'une paire de bases). Dans ce cas précis, le faux allèle peut être très vite repéré car il apparaît chez un individu hétérozygote pour ce locus.	34
2.3	Exemple de perte d'allèle (six premières pistes) lors du typage multiple d'un locus pour un échantillon d'urine de canidé. Une seule répétition sur les huit réalisées donne le bon génotype de l'échantillon pour ce locus.	35
2.4	Exemple de modification de la taille d'un allèle pour une faible amplification (deuxième piste) par rapport à la taille lors d'une « bonne » amplification (première et troisième piste).	37
3.1	Diagramme de la procédure à tubes multiples proposée par TABERLET <i>et al.</i> (1996) pour obtenir un niveau de confiance supérieur à 99% pour chaque locus étudié (Modifiée d'après TABERLET <i>et al.</i> 1996).	42
3.2	Illustration des tâches d'analyse génétique d'une population: échantillonnage, génotypage multiple, construction des génotypes consensus, tri et identification des génotypes uniques, calcul des fréquences alléliques (Freq. all.), calcul des taux d'hétérozygotie (H), calcul des probabilités d'identité (PI), détermination des parentés possibles (parenté), estimation de la taille de population (effectif).	43
5.1	Menu principal du logiciel GEMINI avec les trois modules de travail.	55
5.2	Modèle de l'étude d'une population, utilisé dans le logiciel GEMINI	55
5.3	Menu du Module Batch de GEMINI montrant les trois tâches de ce module: Simulations complète, Simulation pour plusieurs modalités de nombre de répétitions PCR et Test du seuil d'acceptation pour la construction des génotypes consensus.	56
5.4	Résultats donnés par GEMINI après une simulation pour plusieurs modalités du nombre de répétitions de génotypage par échantillon et par locus. On voit l'augmentation du pourcentage d'identification correcte (ordonnées) en fonction du nombre de génotypes par échantillon (abscisses).	57
5.5	Exemple de construction d'un génotype consensus en fonction du seuil d'acceptabilité des allèles. A gauche, sont représentés huit électrophorégrammes pour un échantillon et un locus. A droite, sont expliquées les règles d'acceptabilité des allèles.	58

5.6	Principe de la démarche pour tester le génotypage multiple pour un échantillon et un locus. Elle consiste à génotyper l'échantillon plusieurs fois, puis de ré-échantillonner ces génotypes pour construire des génotypes consensus. En comparant ces génotypes consensus avec le génotype consensus issu de toutes les répétitions, on peut évaluer l'approche multi-tubes.	60
5.7	Photo de gel d'agarose 1,6% montrant les résultats d'amplification de l'ADN mitochondrial à partir d'échantillons de fèces de loup. Les conditions de PCR sont identiques entre les différentes modalités d'extraction: DNA STOOL KIT (à gauche) et le DNA TISSUE KIT (à droite).	65
5.8	Photo d'un gel d'agarose 1,6% comparant les intensités des bandes entre des réactions d'amplification (conditions identiques) d'ADN d'un extrait de fèces sur deux marqueurs microsatellites avec l'AMPLITAQ GOLD (en haut) et la HOTSTAR TAQ (en bas) . N : utilisation du buffer « standard » ; G : utilisation du BUFFER GOLD ; Q : utilisation du Q BUFFER.	66
5.9	Exemple de configuration des salles de manipulations et des échanges autorisés ou non entre ces salles. Une salle supplémentaire pour les échantillons « anciens » (échantillons fossiles ou de musée) peut être ajoutée séparément des autres	68
6.1	Exemple d'une relation observée entre le nombre de génotypes uniques et le nombre d'échantillons typés (o) et des courbes d'ajustement aux deux équations citées dans le texte (Equation de Kohn=... ; Equation de Chessel=-.-)	73
6.2	Exemple du format de cycle de vie utilisé dans GIMLET qui permet de prendre en compte plusieurs saisons de reproduction au cours d'une année et de prendre en compte les mois dans les âges et les dates.	75
6.3	Illustration de la tâche d'identification effectuée dans le logiciel GIMLET	76
6.4	Illustration de la tâche de regroupement et de tri effectuée dans GIMLET	77
6.5	Illustration de la détermination de la parenté potentielle entre individus sur la base des génotypes et des données démographiques (date de naissance et de mort, maturité sexuelle, durée de gestation, sexe)	77
6.6	Trois types d'évènements (et leur probabilité d'occurrence suivant les différents cas) pouvant expliquer les différences entre génotypes. τ = taux de perte d'allèles; ρ = taux de faux allèle; μ_1 = taux de changement d'allèle pour un homozygote; N_{all} est le nombre d'allèles possibles dans la population considérée.	79
6.7	Différents types d'évènements (et leur probabilité d'occurrence) pouvant expliquer le passage d'un génotype à l'autre. ν_1 = taux de changement d'un allèle pour un hétérozygote; ν_2 = taux de changement des deux allèles pour un hétérozygote; ν_3 = taux de changement d'un hétérozygote vers un homozygote.	81
6.8	Effet de l'ordre d'analyse des échantillons (30 ordres d'échantillons différents) sur la relation du nombre de génotypes uniques en fonction du nombre d'échantillons analysés.	84
7.1	<i>a.</i> Approche séquentielle proposée par TABERLET <i>et al.</i> (1999) permettant de tester l'utilisation des échantillons non-invasifs pour l'identification individuelle par génotypage microsatellite. <i>b.</i> Modification de l'approche séquentielle sur la base de TABERLET <i>et al.</i> (1999) et TABERLET et LUIKART (1999).	90
A.1	Les différentes étapes de mises au point de l'identification individuelle chez le loup à partir d'échantillons de fèces et des marqueurs microsatellites.	125

A.2	Dosage de la concentration de l'ADN dans un extrait de tissu. Le graphe (Densité Optique DO en fonction de la concentration d'ADN en ng/ μ l) montre les points de la gamme étalon et de l'extrait dilué 25 fois et 125 fois.	129
A.3	Résultats des simulations concernant le typage génétique d'un hétérozygote (modifié d'après TABERLET <i>et al.</i> 1996) avec la position (flèche) de l'échantillon dont la concentration a été déterminée (voir texte).	130
B.1	Illustration de l'histoire résumée de l'apparition de la vie (à gauche) et des extinctions (à droite) au cours des temps géologiques. L'échelle de temps est en millions d'années avant notre ère. Diagramme des extinctions, d'après BARBAULT (1997).	135
B.2	Pedigree de la population captive du cheval de Przewalski (<i>Equus przewalskii</i>) montrant que les croisements entre apparentés augmentent dans le temps dans une population restreinte. Issu de FRANKHAM <i>et al.</i> (2002).	141
B.3	L'entrée Nord du parc national du Yellowstone (Photo N.Valière). L'inscription est typique de ce que deviendra le parc: « Au profit et pour le plaisir du peuple ».	142
B.4	Principe de formes et tailles de refuges d'après la biogéographie insulaire. Pour chaque cas, la forme et la taille sur la gauche sont de meilleure qualité que celles de droite. (d'après SIMBERLOFF 1988 et JANZEN 1986).	150
B.5	Distribution hypothétique d'allèles (haplotypes) parmi quatre populations montrant les fréquences alléliques et la phylogénie en relation avec les ESU et les MU (d'après MORITZ 1994). Les populations 1 et 2 ont des fréquences alléliques similaires et font partie de la même unité de gestion (MU). Les populations 3 et 4 ont des allèles fixés et seule la population 4 montre une séparation phylogénétique. La population 3 fait donc partie de la même unité « évolutivement » significative (ESU) que les populations 1 et 2 alors que la population 4 constitue une ESU à part.	154

Liste des tableaux

1.1	Récapitulation des différentes caractéristiques des méthodes décrites dans le texte. Séq. = séquençage (ADNmt principalement); Alloz.= allozymes; msat= minisatellites multilocus; μ sat = microsattellites. Modifié d'après SCHLÖTTERER et PEMBERTON (1994), PARKER <i>et al.</i> (1998), MUELLER et WOLFENBARGER (1999), SUNNUCK (2000), FRANKHAM <i>et al.</i> (2002).	21
2.1	Valeur de succès d'amplification dans différentes études publiées. Pour chaque étude, le type d'échantillons (Type), l'espèce, le nombre de loci microsattellites utilisés (N loci) et les valeurs du succès d'amplification (Succès PCR) sont données. Les annotations entre parenthèses précisent le mode de calcul du succès : P = sur le nombre d'amplifications (PCR), E = sur le nombre d'extraits, S = sur le nombre d'échantillons, L = sur le nombre de couples locus/échantillons, # = succès d'amplification pour tous les loci, \$ = amplification positive pour au moins un locus. *= après lyophilisation, **= après étuvage.	33
2.2	Valeurs de taux d'erreurs (perte d'allèle, ADO et faux allèles, FA) pour diverses études publiées. Voir TABLEAU 2.1 pour la signification des annotations entre parenthèses.	37
3.1	Coûts approximatifs (en euros) de génotypage microsattellites (estimation du coût des consommables, hors fonctionnement) sans multiplex d'amplification mais avec 3 multiplex de charge.	42
5.1	Comparaison d'amplification de microsattellites en utilisant différentes polymé- rases: la HOTSTAR TAQ de QIAGEN (55 cycles d'amplification) et l'AMPLITAQ GOLD de PERKIN ELMER (60 cycles d'amplification). Le tableau donne les taux de succès d'amplification (P+%) et de perte d'allèle (ADO%) pour un échantillon de fèces et 10 loci. "-" l'échantillon est homozygote au locus en question. "*" = données pour 45 cycles d'amplification	67
6.1	Cinq exemples de la méthode de sélection du génotype consensus à partir du calcul des probabilités d'apparition des génotypes répliqués (huit répliqués). Les cas différents par la répartition des génotypes, les taux d'erreurs et le nombre d'allèles dans la population.	80

A.1	Nombre d'allèles (N_a), nombre d'allèles ayant une fréquence supérieure à 0.1 ($N_a > 0.1$), nombre d'individus testés ($N_{bindiv.}$), taux d'hétérozygotie observée (H_{obs}), taux attendu d'hétérozygotie à l'équilibre de Hardy-Weinberg (H_{exp}), probabilité d'identité calculée avec l'équation corrigeant la taille de population ($PI_{unbiased}$) et probabilité d'identité calculée pour les frère-soeurs (PI_{sibs}).	126
A.2	Informations concernant les 11 loci sélectionnés : noms des loci, groupes de liaison (Gr. Liais. ; valeur pour NEFF <i>et al.</i> 1999/MELLERSH <i>et al.</i> 1997), séquences des amorces (Seq. Amorces) 5' → 3', motifs des répétitions (si disponible dans Francisco <i>et al.</i> 1996), références (Réf. ; 1 = FRANCISCO <i>et al.</i> 1996 ; 2 = MELLERSH <i>et al.</i> 1997 ; 3 = NEFF <i>et al.</i> 1999), température d'hybridation utilisée en °C (THyb), étendue des tailles de fragments amplifiés déterminées à partir des 13 lous testés (Taille), marquage fluorescent (Coul. ; 6'FAM=Bleu ; HEX=Vert ; NED =Jaune).	127
A.3	Résultats d'amplification multiple (8 par échantillon et par locus) pour 4 échantillons (3 fèces et 1 tissu) et 7 loci donnant des résultats exploitables. %P+= pourcentage de PCR positives ; %ADO = pourcentage de perte d'allèle. Hmz = échantillon homozygote pour le locus.	128
B.1	Nombre d'espèces éteintes et menacées de 1600 à nos jours dans différents taxons et représentation (en %) des espèces éteintes sur les îles.	136
B.2	Potentiel relatif des systèmes de conservation pour la protection et la maintenance de la diversité biologique (modifié d'après Soulé, 1991). '0' = pas de rôle ; '+', '++', '++++' indiquent un rôle faible, modéré et significatif. Voir aussi SPELLENBERG (1992, p. 174-175) pour les aires de protection.	145
B.3	Relation entre la signification statistique et la signification biologique d'une divergence phylogénétique.	155

Liste des articles présentés au cours de ce mémoire

Article I

VALIERE N., Fumagalli L., Gielly L., Miquel C., Lequette B., Poulle M.-L., Weber J.-M., Arlettaz R. & Taberlet P.

Long distance wolf recolonization of France and Switzerland inferred from non-invasive genetic sampling over a period of 10 years.

Accepté dans *Animal Conservation*.

Article II

VALIERE N. & Taberlet P.

Urine collected in the field as a source of DNA for species and individual identification.

Molecular Ecology (2000) 9 : 2150-2152.

Article III

VALIERE N., Berthier P., Mouchiroud D. & Pontier D.

GEMINI : software for testing the effects of genotyping errors and multitubes approach for the individual identification.

Molecular Ecology Notes (2002) 2 : 83-86

Article IV

VALIERE N.

The use of multitubes approach for microsatellite identification : number of genotyping replications and number of loci.

En préparation

Article V

VALIERE N., Pontier D., Garnier J.N. & Goossens B.

Computer simulations to predict identification consistency using noninvasive sampling and the multitubes approach.

Soumis à *Conservation Genetics*

Article VI

VALIERE N.

Population size estimation and biases associated with genotyping errors.

En préparation

Article VII

VALIERE N.

GIMLET : a computer program for analysing genetic individual identification data.

Molecular Ecology Notes (2002) 2 : 377-379

Article VIII (Non décrit dans ce mémoire)

VALIERE N.

L'étude de l'expansion du loup (*Canis lupus*) en France : apport des méthodes non-invasives et des marqueurs moléculaires

In *L'étude et la Conservation des Carnivores* (2002). Chapron, G. & Moutou, F. (éditeurs). SFEPM. 167p.

Volume 1 : Texte principal

Introduction Générale

La conservation des espèces et des espaces sauvages a connu un regain d'intérêt avec le développement de la biologie de la conservation dans les années 80. La biologie de la conservation (SOULÉ 1985) a permis de cadrer les théories et les techniques disponibles pour le maintien et la réhabilitation des sites naturels ou des espèces et de les intégrer aux activités humaines.

De grands débats se sont élevés sur la taille des réserves (voir SIMBERLOFF 1988), sur les effectifs de populations à maintenir (*e.g.* SOULÉ 1980, FRANKLIN 1980), sur l'unité biologique sur laquelle faire agir les moyens de conservation (voir par exemple CROZIER 1992, AVISE 1989, BOWEN 1999, DIMMICK *et al.* 1999, ROJAS 1992, WOODRUFF 1989) et de nombreuses disciplines ont participé aux avancées de la biologie de la conservation afin de déterminer le fonctionnement de certaines populations pour mieux les protéger.

Certains de ces points n'auraient pu être possibles sans l'apport des outils moléculaires et de la génétique (HAIG 1998). La génétique met à la disposition de la biologie de la conservation de nombreuses techniques qui permettent ainsi de pouvoir répondre à de multiples questions comme celle du fonctionnement des populations (HUGHES 1998). Pour pouvoir appréhender ces questions, un outil moléculaire doit être utilisable à plusieurs échelles biologiques, être suffisamment variable, facile à mettre en oeuvre et applicable à partir d'échantillons divers (MCDONALD et POTTS 1997, POWELL 1994). C'est le cas notamment des microsatellites (BEAUMONT et BRUFORD 1998, BALDING 1998). Ces marqueurs, largement utilisés dans la littérature, ont la particularité d'être amplifiés par PCR (*Polymerase Chain Reaction*). Cette caractéristique leur permet d'être utilisés à partir d'échantillons tels que les fèces, les poils ou autres échantillons non-invasifs (*e.g.* MORIN et WOODRUFF 1996). Ces échantillons sont récoltés sur le terrain avec un dérangement minimal des animaux ce qui autorise l'étude d'espèces ou de populations qu'il n'est pas possible d'étudier par d'autres méthodes nécessitant la capture ou l'observation des individus. Ainsi, cette opportunité peut permettre d'avoir accès à des populations protégées, voire en danger pour lesquelles le dérangement est proscrit (TABERLET et LUIKART 1999).

Néanmoins, si l'association des marqueurs microsatellites et des méthodes d'échantillonnage non-invasif semblait prometteuse et révolutionnaire, il faut bien garder à l'esprit que les échantillons en question ne contiennent qu'une quantité limitée d'ADN souvent dégradé, associée à la présence d'inhibiteurs de PCR, si bien que des erreurs de génotypage¹ (souvent stochastiques) sont attendues (TABERLET et WAITS 1998). Dans ce cadre d'étude, il est nécessaire de définir des points à optimiser et à améliorer afin de réduire et corriger ces erreurs (*e.g.* TABERLET et LUIKART 1999).

Nous verrons dans la première partie de ce mémoire comment a émergé la biologie de la conservation et ce que la génétique peut apporter à cette nouvelle discipline. Après une description

¹caractérisation des loci microsatellites pour un échantillon

des différentes méthodes à disposition, nous verrons aussi pourquoi les microsatellites semblent être si importants en biologie des populations et de la conservation par rapport aux autres marqueurs génétiques.

Nous aborderons enfin l'échantillonnage non-invasif en décrivant les procédés et leur utilisation et comment ce type d'échantillonnage peut être utilisé en association avec la génétique. Notamment, nous verrons les avantages mais surtout les inconvénients au niveau de l'amplification d'ADN et du génotypage des microsatellites.

Dans une deuxième partie, nous verrons les points sur lesquels se concentrer pour pouvoir améliorer l'utilisation des méthodes d'échantillonnage non-invasif et des microsatellites à différents niveaux d'une étude génétique d'une population. C'est tout d'abord l'échantillonnage sur le terrain, puis le génotypage et la correction des erreurs, et enfin l'analyse des données. L'ensemble de ces points s'articuleront autour de sept articles publiés, soumis ou en préparation. Nous nous attarderons enfin sur l'utilité d'une étude pilote permettant de tester l'approche dans les conditions propres de l'étude.

Première partie

Introduction et Contexte Général

Chapitre 1

La biologie et la génétique de la conservation

Sommaire

1.1 La biologie de la conservation	8
1.2 La génétique de la conservation	10
1.3 Les marqueurs génétiques : description et applications	11
1.3.1 Présentation des outils	13
1.3.1.1 Les allozymes	13
1.3.1.2 La PCR (Polymerase Chain Reaction)	13
1.3.1.3 RFLP (Restriction Fragment Length Polymorphism)	14
1.3.1.4 Séquençage d'ADN	14
1.3.1.5 SSCP (Single-Strand Confirmational Polymorphism) et MHC (Major Histocompatibility complex)	15
1.3.1.6 RAPD (Random Amplification of Polymorphic DNA)	16
1.3.1.7 VNTR (Variable Number of Tandem Repeats)	18
1.3.2 Applications et applicabilité des méthodes génétiques	20
1.3.2.1 Niveau populationnel : structure des populations	22
1.3.2.2 Niveau familial et du groupe social: la détermination des parentés	23
1.3.2.3 Niveau individuel : identification et suivi des individus	24
1.3.3 Avantages et inconvénients des microsatellites	25
1.3.3.1 Mise au point et développement	25
1.3.3.2 Révélation des marqueurs	25
1.3.3.3 Amplification PCR	26
1.3.3.4 Interprétation et utilisation des résultats	26

1.1 LA BIOLOGIE DE LA CONSERVATION

Les enregistrements fossiles montrent que les diversités spécifique et écologique n'ont cessé d'augmenter globalement depuis l'apparition de la vie il y a 3,8 milliards d'années. Il existe en fait une dynamique plus complexe dans l'évolution : des phases de spéciation et d'extinction se succèdent, si bien que l'on peut compter plusieurs extinctions de masse suivies chacune d'une augmentation du nombre d'espèces.

Les extinctions et les déclin des espèces ou des populations peuvent être le fruit de nombreux facteurs. On reconnaît aujourd'hui que les risques d'extinction sont dus à des interactions entre des facteurs démographiques, environnementaux et génétiques particulièrement liés à la taille de population (FIGURE 1.1 ; voir Annexe B).

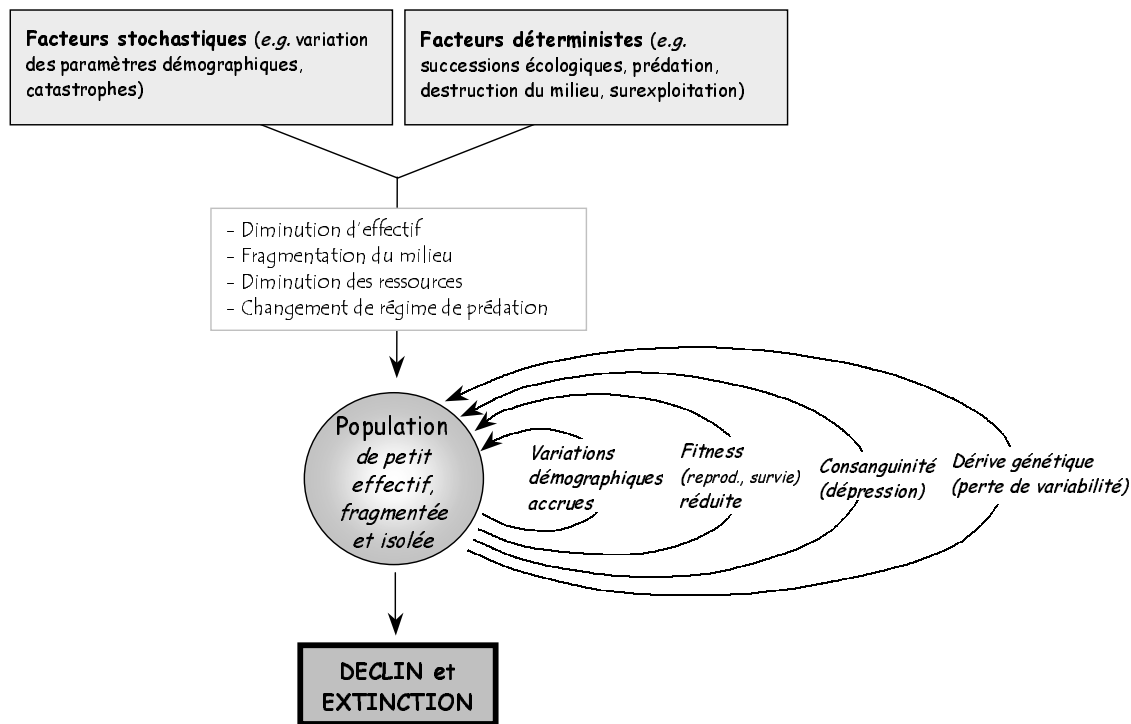


FIGURE 1.1 – Intégration des facteurs provoquant et accélérant l'extinction de populations. Certains facteurs peuvent entraîner des diminutions d'effectif, la fragmentation du milieu, des changements de régimes de prédation, la diminution des ressources. Ces facteurs peuvent provoquer la disparition directe de la population, ou entraîner cette population dans une spirale (ou vortex GILPIN et SOULÉ 1986). L'effectif de la population se réduit alors progressivement jusqu'à (éventuellement) l'extinction de la population. Modifié d'après BARBAULT (1997, p.77) et FRANKHAM et al. (2002, p.32)

Si les causes naturelles semblent être à l'origine du déclin et de l'extinction de certaines populations ou espèces, les pertes d'espèces, au cours des derniers siècles, semblent être provoquées par l'apparition de la civilisation humaine et de l'agriculture (MOROWITZ 1991) et être accélérées par l'expansion démographique humaine (MOROWITZ 1991, SPELLENBERG 1992, p.3). Les causes humaines sont la destruction et la dégradation directes de zones naturelles provoquant la diminution de la biodiversité et la fragmentation des habitats, la surexploitation de populations animales ou végétales, la dégradation indirecte ou enfin l'introduction d'espèces exotiques

qui entrent en compétition avec les espèces endémiques, ou s'en nourrissent (DIAMOND 1989, WESTERN *et al.* 1989, SOULÉ 1991, EHRLICH et WILSON 1991, SPELLENBERG 1992, pp.3-13, BARBAULT 1997). Tous ces facteurs ont provoqué et provoquent actuellement une « sixième extinction » (LEAKEY et LEWIN 1995) dont le rythme et l'amplitude dépassent l'apparition de nouvelles espèces par spéciation, ce qui entraîne une réelle crise de la biodiversité .

Bien que l'homme puisse avoir des implications néfastes sur son environnement, il peut aussi prendre conscience de son impact et agir contre la dégradation de la vie sauvage, aussi bien pour des raisons éthiques (toute forme de vie a le droit d'exister et l'Homme n'a aucun droit de mort sur une autre espèce, MYERS 1979), des raisons « esthétiques » (EHRLICH et WILSON 1991, la nature est belle et il faut la conserver), des raisons matérialistes (voir par exemple SPELLENBERG 1992, pp.13-20) car de nombreuses espèces végétales et animales fournissent des composés ou des produits qui peuvent être utilisés en médecine (*e.g.* médicaments, analgésiques, sédatifs, antibiotiques, hormones, enzymes, stimulants), en agriculture (*e.g.* pour la nourriture) ou dans l'industrie (*e.g.* latex, résines, huiles essentielles).

Au Etats Unis, la conservation prend une forme cadrée et écologique au cours du 20^e siècle grâce au développement de l'écologie évolutive sous l'influence des écologistes et « conservationnistes » Paul B. SEARS (SEARS 1935) et Aldo LEOPOLD (LEOPOLD 1949). Leur vision, imprégnée par l'idée d'équilibre, décrit la nature comme un ensemble de parties dynamiquement reliées entre elles dans un système intégré (base de l'Ecologie²) qu'il faut protéger en conservant les caractères propres des écosystèmes et de leur régulation naturelle. Mais ce n'est que récemment (dans les années 1960-70) qu'une confluence sociale et scientifique incite les professionnels (biologistes, sociologues, gestionnaires, administrateurs) à développer une nouvelle discipline de la biologie : la biologie de la conservation . Cette confluence (SOULÉ 1986, 1987, BARBAULT 1997) était due à (*i*) l'explosion des théories et des données sur l'écologie des communautés et de la biogéographie insulaire, (*ii*) la volonté de séparer la biologie « traditionnelle » (*e.g.* zoologie, botanique) de la biologie moderne des populations (foresterie, pêche, ou gestion de la vie sauvage), (*iii*) une tendance sociale des mouvements environnementalistes et enfin (*iv*) une prise de conscience de la « crise de la biodiversité » (et notamment la responsabilité de l'homme dans cette crise).

SOULÉ (1987) identifie six grands rôles des membres de la *Society for Conservation Biology* (créée en 1986) et donc de la biologie de la conservation en général :

- La modélisation et l'analyse des processus biologiques à l'échelle des populations, des communautés, des écosystèmes et de la planète.
- L'inventaire et la détermination de la systématique des formes vivantes.
- Les tests d'hypothèses (expérimentations, travaux de terrain).
- Le développement et l'évaluation des interventions technologiques et de gestion pour maintenir et restaurer la diversité et les fonctions.
- La communication des résultats (notamment par le biais du journal *Conservation Biology*).
- L'intégration des connaissances aux activités humaines.

Pour tous ces rôles, de nombreuses disciplines doivent coopérer, avec la participation de nombreux acteurs : écologie, biogéographie, systématique, génétique, science de l'évolution, épidémiologie, sociobiologie, foresterie, pêche, gestion, agronomie, médecine vétérinaire, écono-

²Le terme d'Ecologie fut introduit clairement par SEARS : « un écologiste est celui qui entre dans une forêt ou une prairie et qui ne voit pas simplement ce qu'ils sont, mais ce qu'il se passe à l'intérieur » (SEARS 1935, cité dans STEBBINS 1992).

mie, politique, ethnobiologie, anthropologie, éthique environnementale, paléontologie, climatologie, philosophie, océanographie, agences gouvernementales, organisations et associations non gouvernementales (SOULÉ 1986, EHRENFELD 1987; SPELLENBERG 1992, p.24).

Le rôle principal de la biologie dans la conservation est de faire le lien entre les gestionnaires et les scientifiques afin que cette coopération puisse identifier les problèmes et fournir des principes scientifiques à la conservation et des procédures de correction (WESTERN 1989). En réalité, et de façon plus pragmatique, on peut définir deux grandes implications : (i) identifier les unités biologiques qu'il faut conserver et déterminer les priorités et (ii) apporter des informations primordiales sur les populations, les espèces et/ou les écosystèmes pour choisir la meilleure stratégie. Dans les dernières années c'est le domaine de la génétique qui a connu le plus de développements. Nous allons voir comment la génétique peut servir à la conservation, les différentes techniques et leurs applications.

1.2 LA GÉNÉTIQUE DE LA CONSERVATION

La génétique de la conservation est l'application de la génétique pour préserver les espèces comme des entités dynamiques capables d'adaptation aux changements d'environnement. Cela regroupe la gestion génétique des petites populations, la résolution des incertitudes taxonomiques, la définition d'unités de gestion à l'intérieur des espèces et l'utilisation des analyses génétiques pour déceler le commerce illégal et le braconnage ou pour comprendre la biologie de l'espèce (FRANKHAM *et al.* 2002, p.1).

La génétique de la conservation a permis d'apporter une contribution non négligeable à la compréhension des mécanismes de déclin et d'extinction (voir Annexe B) et doit traiter 11 problèmes principaux (FRANKHAM *et al.* 2002, p.9-10) :

- les effets délétères dus à la consanguinité, sur la reproduction et la survie des individus des populations,
- la perte de diversité génétique et de la capacité à évoluer dans un environnement changeant,
- la fragmentation des populations et la réduction du flux génique,
- les processus aléatoires (dérive génétique) surpassant la sélection comme principal processus évolutif,
- l'accumulation et la perte (purge) des mutations délétères,
- l'adaptation génétique à la captivité et ses effets sur le succès de réintroduction,
- la résolution d'incertitudes taxonomiques,
- la définition d'unités de gestion à l'intérieur des espèces,
- l'utilisation des analyses génétiques moléculaires dans les cas de braconnage et de commerce illégal d'espèces,
- l'utilisation des analyses génétiques moléculaires pour comprendre les aspects importants de la biologie de l'espèce
- les effets délétères sur la *fitness* qui apparaissent parfois (*outbreeding depression*).

Dans le présent mémoire, nous allons nous intéresser à certaines applications de la génétique de la conservation, principalement grâce à l'utilisation des microsatellites ou du séquençage de l'ADN mitochondrial et des méthodes d'échantillonnage non-invasif. En effet, la mise au point de l'amplification de l'ADN par PCR (*Polymerase Chain Reaction*) et le développement parallèle de nouveaux marqueurs moléculaires a permis des avancées non négligeables dans ces domaines et la résolution de plusieurs points cités ci-dessus (voir FIGURE 1.2 et Annexe B) dont principalement :

- *La détection de l'exploitation illégale d'espèces protégées* : le braconnage et le commerce d'individus d'espèces protégées représentent l'une des menaces les plus sérieuses pour la

survie de ces espèces. L'utilisation des marqueurs moléculaires peut permettre de comparer des échantillons prélevés lorsqu'il y a soupçon de braconnage ou de commerce illégal à partir d'échantillons de référence (RUTH et FAIN 1993, MANEL *et al.* 2002). Cela a été le cas pour le braconnage sur l'oryx (MARSHALL *et al.* 1999), le cas de détection de la présence de produits dérivé du tigre *Panthera tigris* dans des médicaments chinois (HOLDEN 2000) ou le cas de détermination de l'origine du caviar dans plusieurs lots commercialisés en Amérique du Nord (BIRSTEIN *et al.* 1998). Une fraude dans une compétition de pêche a pu aussi être détectée (PRIMMER *et al.* 2000).

- *L'analyse de la diversité génétique au sein des espèces et des populations et l'analyse de la fragmentation génétique des populations et des flux géniques* peuvent être réalisées grâce à l'ADN mitochondrial (HARRISON 1989, PALUMBI *et al.* 1991, TABERLET et BOUVET 1994) ou à des marqueurs nucléaires (GOLDSTEIN *et al.* 1999 ; PAETKAU *et al.* 1995, 1998a, 1999 chez l'ours polaire *Ursus maritimus* ou l'ours brun).
- *L'étude de la biologie de certaines populations* : en biologie de la conservation de nombreux paramètres de populations doivent être déterminés pour pouvoir gérer et réagir au mieux sur les populations menacées. Les paramètres démographiques (taux de survie, de fécondité, taille de population) sont des paramètres de premier ordre pour pouvoir ajuster les moyens de conservation (par exemple, d'où vient le déclin d'une population : d'une augmentation de mortalité ou d'une baisse de fécondité, ou les deux). Le comportement apporte aussi une contribution importante à la biologie de la conservation (voir SUTHERLAND 1998). En effet les comportements sociaux, la territorialité, les patrons de dispersion, les mouvements d'individus, les liens de parenté et les patrons de reproduction sont des informations nécessaires afin d'étudier les processus d'adaptation dans différentes conditions écologiques et donc d'adapter les moyens de conservation à chaque populations. La description des modes de reproduction peut être appréhendée par la génétique (voir HUGHES 1998). La génétique a permis aussi d'étudier certains aspects des populations qui ne pouvaient être étudiés avec les méthodes classiques d'observation. On a pu se rendre compte, par exemple, en utilisant les marqueurs génétiques, que certaines espèces « socialement » monogames ne l'étaient pas génétiquement du fait de l'existence de fertilisations hors couple (WESTNEAT et WEBSTER 1994). La détermination des parentés et des paternités ainsi que du sexe des individus sont autant d'informations capitales en conservation, pour l'établissement de pedigres de populations captives ou pour les programmes de réintroduction (voir par exemple MARSHALL *et al.* 1999 chez l'oryx *Oryx leucoryx* ; TABERLET *et al.* 1993 chez l'ours brun *Ursus arctos*). La démographie, notamment la taille de population, passée et présente constitue également une information précieuse (voir SCHWARTZ *et al.* 1998, GOLDSTEIN *et al.* 1999, WAITS et LEBERG 1999) pour connaître le statut précis d'une population avant ou après une intervention de conservation.

1.3 LES MARQUEURS GÉNÉTIQUES : DESCRIPTION ET APPLICATIONS

La génétique a connu ces dernières années de nouveaux développements. Ces développements, notamment liés à la mise au point de la réaction d'amplification de l'ADN (la *Polymerase Chain Reaction*) a permis de nombreuses applications. Dans cette partie, nous allons d'abord décrire succinctement les différentes techniques disponibles et plus particulièrement les marqueurs microsatellites. Ensuite nous verrons comment ces outils génétiques peuvent être utilisés en biologie

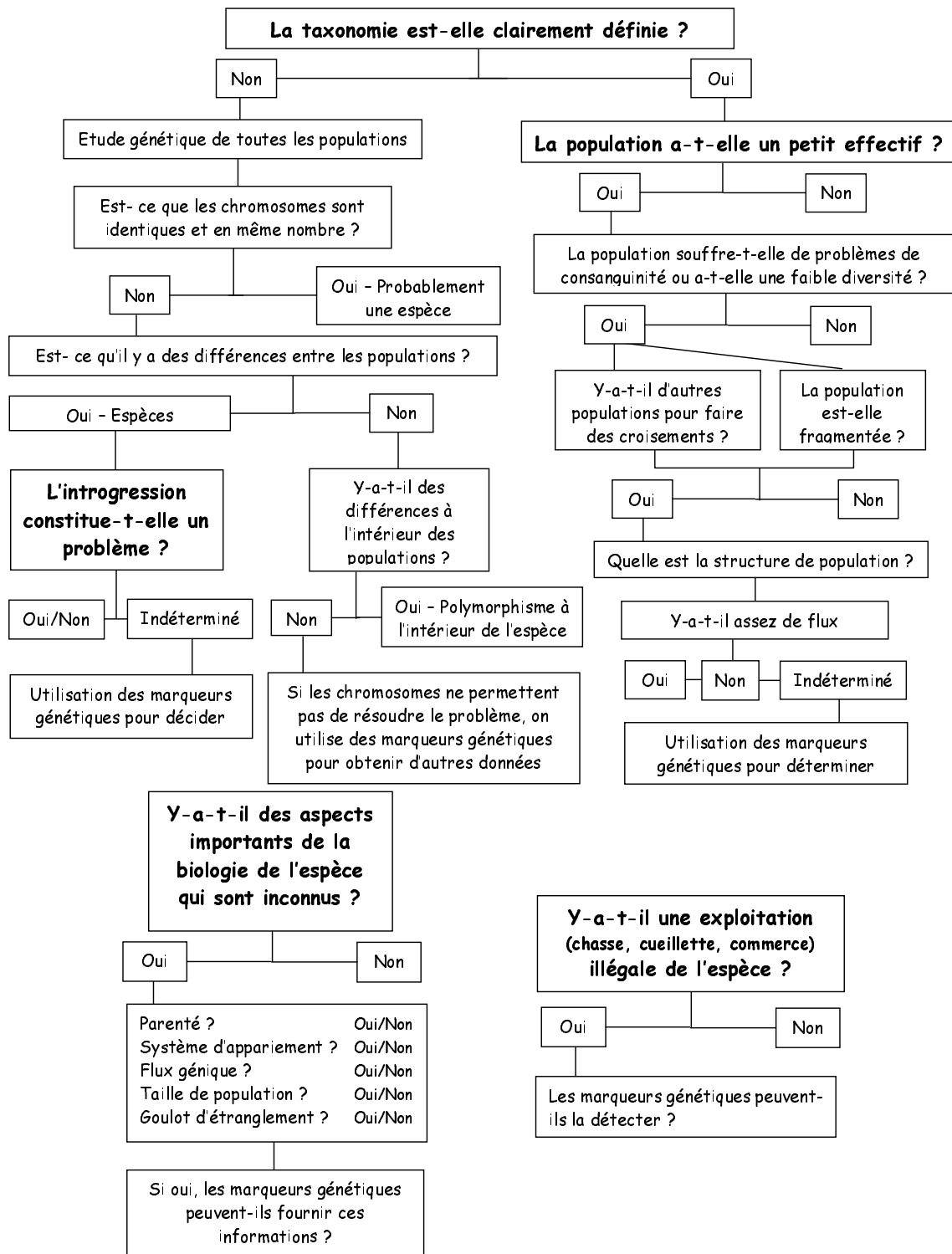


FIGURE 1.2 – Diagrammes des problématiques posées en génétique de la conservation. D'après FRANKHAM et al. (2002), p.17.

des populations et de la conservation et les avantages fournis par les microsatellites.

1.3.1 Présentation des outils

Les principaux outils moléculaires (voir tableau récapitulatif 1.1, p.21) utilisés en biologie des populations et de la conservation sont : les allozymes, le séquençage, les SSCP, les RFLP (et PCR-RFLP), les RAPD, les AFLP, et les VNTR. Une technique supplémentaire sera décrite car elle a considérablement fait évoluer et développer les autres techniques : la *Polymerase Chain Reaction* (ou PCR).

1.3.1.1 Les allozymes

Le terme d'**allozyme** réfère aux différentes **formes alléliques de protéines**. Ces différents allozymes peuvent être révélés sur des gels d'électrophorèse en faisant migrer les protéines dans un champ électrique (voir par exemple HOELZEL et DOVER 1991, MAY 1992). La séparation des variants protéiques (dépendant de leur charge électrique et de leur poids moléculaire) pour plusieurs individus conduit à classer ceux-ci en homozygotes et hétérozygotes à chaque locus (protéines) étudié. Cette méthode est considérée comme **simple, rapide et peu chère**. La technique requiert la préparation des échantillons (passage dans un tampon) afin d'extraire les protéines sans perdre leur activité. Il suffit ensuite d'appliquer les protéines extraites sur un gel d'électrophorèse pour en séparer les variants. Les principales limitations des allozymes est qu'ils ne présentent souvent **pas assez de variabilité** pour différencier des populations ou des espèces à cause des contraintes sélectives sur l'évolution des protéines. Dans certains cas, même des espèces largement séparées et différentes morphologiquement peuvent être monomorphes à plusieurs loci allozymiques (voir PARKER *et al.* 1998). Les contraintes sélectives sur les protéines étudiées peuvent ainsi limiter l'utilisation de certaines méthodes statistiques mais aussi créer un patron de variation reflétant les conditions écologiques de l'espèce ne reflétant pas les relations phylogénétiques existant entre les populations (par exemple une protéine qui influence la tolérance à la température peut avoir un patron de variation géographique par clines).

1.3.1.2 La PCR (Polymerase Chain Reaction)

La **PCR** (*Polymerase Chain Reaction*) est une réaction in vitro (SAIKI *et al.* 1988) basée sur le principe de la duplication de l'ADN dans les cellules. Elle permet d'**amplifier le nombre de copies d'ADN** grâce à l'utilisation d'une enzyme, l'ADN polymérase. La PCR a constitué une petite révolution en biologie moléculaire en permettant de travailler sur des quantités d'ADN importantes et en favorisant le développement de certaines méthodes d'analyse de l'ADN (voir plus loin). La PCR se décompose en **plusieurs cycles de trois phases** correspondant à des températures de travail différentes (voir FIGURE 1.3). A chaque cycle, chaque molécule d'ADN est dupliquée en deux molécules identiques entre elles et identiques à la molécule de départ (en théorie...). L'amplification lors d'une réaction PCR est exponentielle mais le rendement de la réaction diminue au cours des cycles et la courbe d'amplification est plutôt sigmoïde. Néanmoins, à partir d'une molécule d'ADN de départ, on arrive facilement à des millions de copies à la fin de la réaction.

Au cours de ces trois phases, l'ADN est dénaturé, des oligonucléotides (spécifiques de la zone à amplifier) sont hybridés à l'ADN cible afin d'amorcer la réaction, ces amorces sont allongées pour permettre la synthèse à l'identique des fragments parents par la complémentarité des bases. La PCR nécessite donc la présence de dinucléotides (dNTP), de différents sels et de $MgCl_2$ (pour catalyser la réaction) et de l'ADN polymérase. Jusque dans les années 80, l'enzyme utilisée était

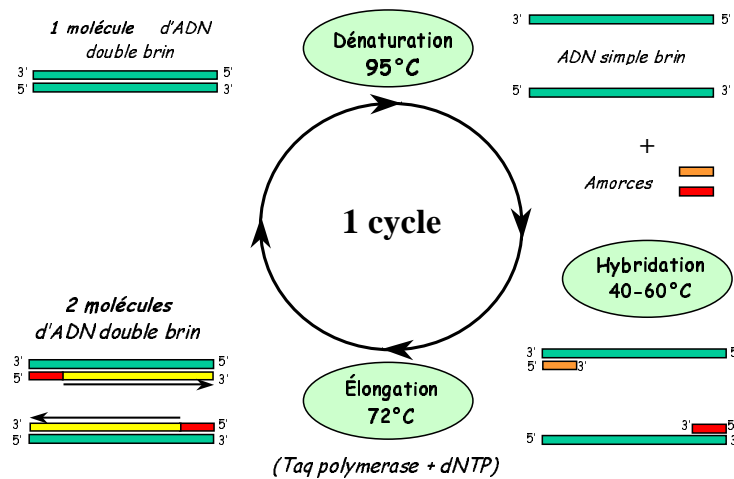


FIGURE 1.3 – Schéma d'un cycle de PCR illustrant les trois phases : dénaturation, hybridation et élongation. A partir d'une molécule d'ADN en début de cycle, on obtient, en fin de cycle, deux molécules d'ADN, identiques entre elles et avec la molécule de départ.

le fragment de Klenow de *Escherischia coli*, qui présentait le désavantage d'être inactivé par la chaleur ce qui imposait de rajouter de l'enzyme à chaque cycle de la réaction. La découverte d'une polymérase thermostable, la *Taq polymerase*, isolée à partir de *Thermus aquaticus*, a nettement amélioré les conditions et les performances de la réaction, car cette polymérase pouvait être injectée en début de réaction et restait active jusqu'à la fin de la PCR (MULLIS et FALOONA 1987).

1.3.1.3 RFLP (Restriction Fragment Length Polymorphism)

La technique **RFLP** (*Restriction Fragment Length Polymorphism*) a été la première basée sur l'ADN, utilisée en biologie des populations. La technique RFLP repose sur l'utilisation d'**enzymes de restriction** capables de couper l'ADN en des sites précis (appelés sites de restriction, déterminés par une séquence de quelques paires de bases) produisant ainsi des fragments de tailles différentes quand des mutations ont créé ou modifié ces sites de restriction (voir FIGURE 1.4). Les fragments sont visualisés par hybridation *Southern Blot* ou directement sur un gel d'électrophorèse (marquage par bromure d'éthidium), cette dernière étant plus avantageuse car l'hybridation requiert le développement de sondes qui sont souvent particulières à une espèce (WESTNEAT et WEBSTER 1994).

Une version plus efficace de cette technique est la PCR-RFLP qui utilise la PCR pour amplifier des fragments spécifiques du génome. La PCR-RFLP permet l'application de la méthode RFLP sur des plus petites quantités d'ADN initial et de sélectionner l'information génétique (par exemple l'ADNmt) sur lequel s'appliquera la méthode.

1.3.1.4 Séquençage d'ADN

Le **séquençage d'ADN** consiste en la **détermination de la succession des bases** constituant l'ADN d'une partie du génome. Le séquençage est l'outil de prédilection pour la compréhension de l'évolution moléculaire et il est utilisé de façon importante pour la **construction de**

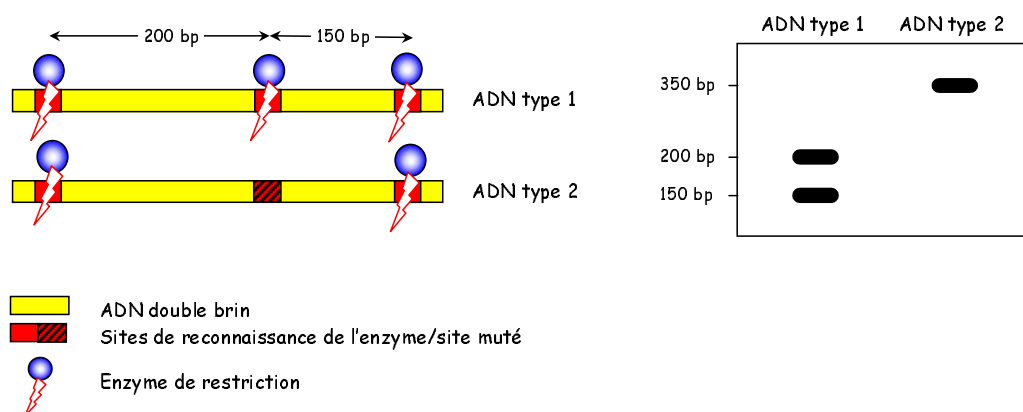


FIGURE 1.4 – Principe des RFLP. Les sites de restriction de l'enzyme utilisés sont différents entre les deux brins d'ADN. La migration des brins générés par le découpage de l'ADN par l'enzyme donne, sur un gel, les résultats présentés à droite.

phylogénies entre espèces. La détermination des séquences d'ADN est réalisée maintenant « en routine » grâce à des réactions de PCR. Pour cela, on utilise un **type de PCR particulier** au cours de laquelle des didéoxynucléotides marqués (ddNTP) sont utilisés en plus des dinucléotides (dNTP). Ces ddNTP sont des dNTP modifiés pour arrêter l'élongation des fragments synthétisés (SANGER *et al.* 1977). Ainsi au cours de la réaction, seuls les fragments se terminant par le nucléotide associé au ddNTP (ddATP, ddTTP, ddCTP, ddGTP) présent dans la réaction seront révélés (voir FIGURE 1.5). Jusqu'à récemment, le séquençage nécessitait l'emploi de composés radioactifs et de plusieurs réactions afin d'obtenir la séquence d'ADN en quatre pistes sur un gel (une piste par base : A, T, G, C ; voir par exemple HOELZEL et DOVER 1991, p.27). Aujourd'hui les ddNTP sont marqués par fluorescence (plusieurs couleurs) et on peut associer un ddNTP (donc une base) à une couleur. Une seule réaction de PCR est donc nécessaire et une séquence peut être lue sur une seule piste de séquenceur automatique (avec une couleur par base, la succession des couleurs correspondant à la succession des bases sur le brin d'ADN).

1.3.1.5 SSCP (Single-Strand Confirmational Polymorphism) et MHC (Major Histocompatibility complex)

La méthode **SSCP** (*Single-Strand Confirmational Polymorphism*) est une méthode de **détection des mutations ponctuelles ou des insertions/délétions**, qui peut (parmi d'autres techniques ; voir par exemple SCOTT et STRAUSS 2000) remplacer le séquençage complet d'une région du génome. Cette technique est basée sur la **conformation tridimensionnelle des fragments d'ADN simple-brin**. Le principe général (voir FIGURE 1.6 ; ORITA *et al.* 1989) est de faire migrer des fragments d'ADN, chimiquement dénaturés (séparation des brins complémentaires), sur un gel non dénaturant (pour ne pas altérer la configuration tridimensionnelle des brins). La séquence des fragments simples brins déterminent la charge et la structure secondaire (conformation tridimensionnelle) et ces fragments peuvent être séparés par migration sur gel. Une PCR peut être réalisée avant cette migration pour marquer les brins et pour amplifier une seule zone du génome. Cette technique peut détecter au-delà de 80% des mutations ponctuelles (SCOTT et STRAUSS 2000) et de nombreuses modifications et variantes ont été utilisées pour accroître la sensibilité des SSCP.

Dans le domaine de la conservation, la méthode SSCP est souvent associée à l'**analyse du**

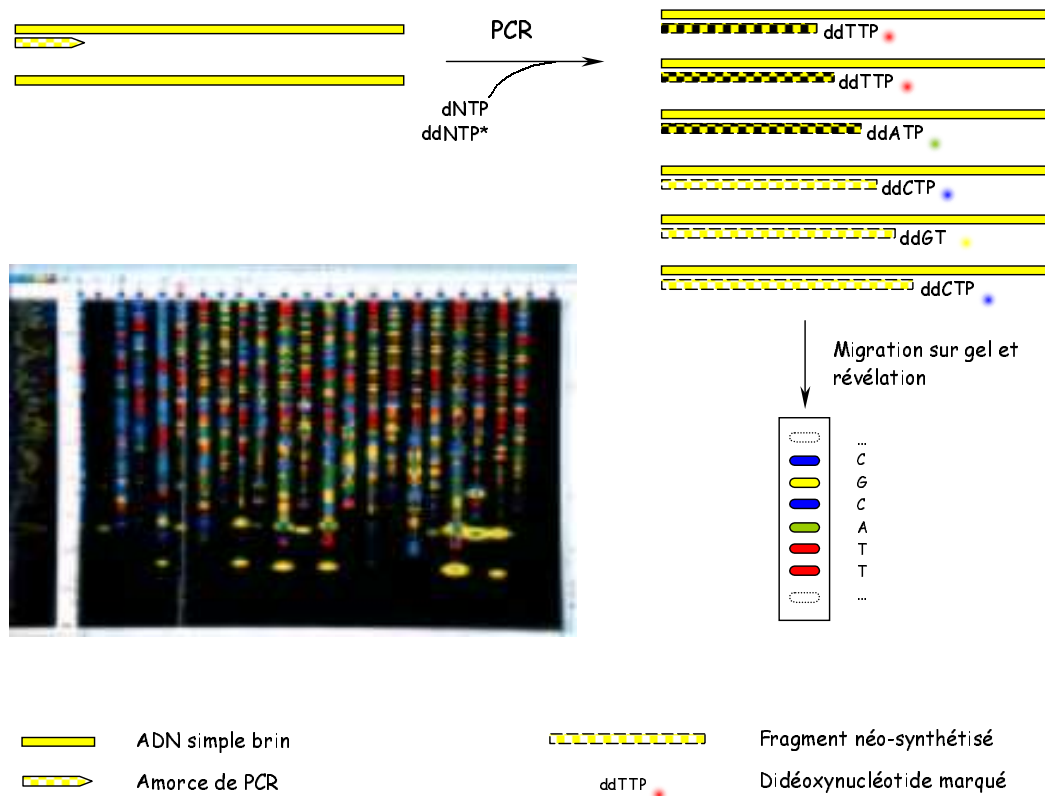


FIGURE 1.5 – Principe du séquençage basé sur la réaction de Sanger. Ici, l'utilisation de quatre marquages fluorescents permet de séquencer un fragment en une seule piste. En bas et à gauche, un image d'un gel de séquence (logiciel GENESCAN, APPLIED BIOSYSTEMS PERKIN ELMER) où une piste (ligne verticale de couleur) correspond à une séquence.

MHC (*Major histocompatibility Complex*; voir par exemple GUTIÉRREZ-ESPELETA *et al.* 1999, HEDRICK *et al.* 2000a, b, GUTIÉRREZ-ESPELETA *et al.* 2001, HEDRICK *et al.* 2001a, b). Le MHC est un des systèmes génétiques les plus importants pour la résistance aux maladies infectieuses chez les vertébrés (HEDRICK *et al.* 2000a). O'BRIEN et EVERMANN (1988) ont suggéré que les populations avec une faible diversité du MHC peuvent être particulièrement vulnérables aux maladies infectieuses (mais cette relation n'est pas forcément universelle, voir par exemple GUTIÉRREZ-ESPELETA *et al.* 2001). Les maladies peuvent être en outre responsables de l'extinction de certaines populations ou espèces (voir par exemple WOODROFFE 1999), surtout en tant que facteur terminal (CAUGHLEY et GUNN 1996, p.266). Il est donc important d'analyser la variabilité génétique du MHC pour appréhender la vulnérabilité des populations.

1.3.1.6 RAPD (Random Amplification of Polymorphic DNA)

Les marqueurs **RAPD** (*Random Amplification of Polymorphic DNA*) sont basés sur l'**amplification de fragments d'ADN par PCR en utilisant des amorces aléatoires**. Le polymorphisme provient de la différence entre les sites d'hybridation des amorces entre les échantillons analysés. Il en résulte un patron de bandes séparées par électrophorèse, différent entre individus par la présence/absence de produits d'amplification. Cette technique est très simple à mettre en oeuvre puisque aucune connaissance préalable sur le génome n'est requise. En revanche, la **diffi-**

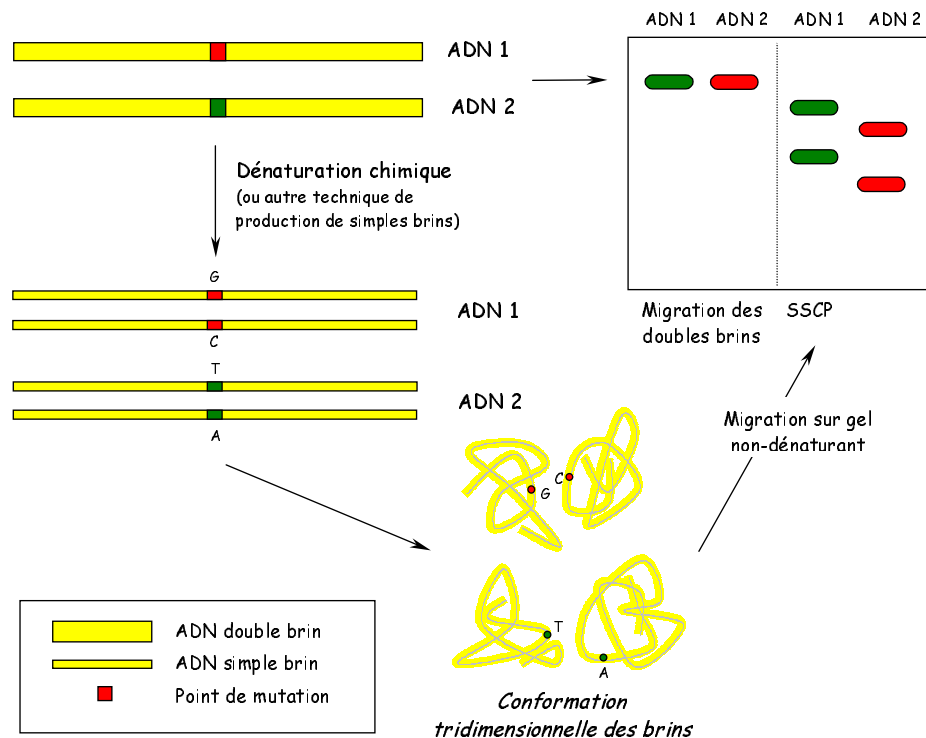


FIGURE 1.6 – Schéma général de la méthode de détection du polymorphisme de conformation simple brin (SSCP) avec en comparaison la migration des doubles brins pour les mêmes fragments d'ADN.

culté de reproductibilité des résultats (PÉREZ *et al.* 1998) et les possibles artefacts (dus à la faible longueur des fragments produits, PARKER *et al.* 1998) semblent être de sérieuses limites aux marqueurs RAPD. Cette méthode a été beaucoup utilisée chez les végétaux et semble être d'application limitée surtout pour les assignations et les exclusions de paternité (PARKER *et al.* 1998).

1.3.1.6.1 AFLP (Amplified Fragments Length Polymorphism)

Le principe des marqueurs **AFLP** (*Amplified Fragments Length Polymorphism*, acronyme voulant se rapprocher de RFLP ; VOS *et al.* 1995) est d'**amplifier des fragments d'ADN de restriction**. L'ADN génomique est découpé par deux enzymes de restriction (un enzyme à site fréquent, l'autre à site rare). Après ligation d'adaptateurs complémentaires des césures des sites de restriction, les fragments sont amplifiés grâce à l'utilisation d'amorces complémentaires des adaptateurs (voir FIGURE 1.7). Les amplifications sont sélectives dans le sens où, en ajoutant des nucléotides au bout de l'amorce complémentaire des sites de restriction, seuls certains fragments sont amplifiés ce qui réduit la complexité des patrons de polymorphisme révélés sur gel d'acrylamide. La première amplification est appelée pré-sélective et réduit le nombre de fragments par 16 (un nucléotide est ajouté aux deux amorces). La seconde amplification est dite sélective et réduit le nombre de fragments par 256 (trois nucléotides sont ajoutés aux deux amorces). Les principaux avantages des AFLP est que leur mise au point ne nécessite **aucune connaissance préalable sur le génome**, qu'elle s'applique à tout le génome (et non une seule région), et qu'il est possible de produire un très grand nombre de loci (jusqu'à 50 à 200 loci par expérience ; VOS *et al.* 1995, BLEARS *et al.* 1998). Si on peut appliquer les AFLP sur de petites quantité

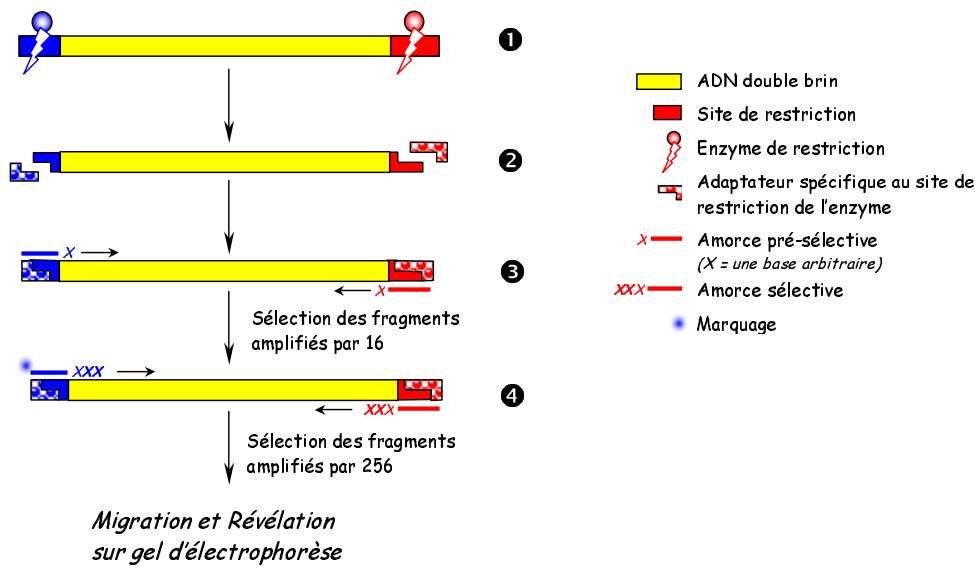


FIGURE 1.7 – Principe de l'AFLP montrant les quatre étapes de la méthode : (1) Digestion (deux enzymes de restriction), (2) Ligation (ADN ligase) des adaptateurs spécifiques aux sites de restriction, (3) PCR pré-sélective et (4) PCR sélective.

d'ADN (BLEARS *et al.* 1998, MUELLER et WOLFENBARGER 1999), il est néanmoins nécessaire d'obtenir de l'ADN de très bonne qualité (BLEARS *et al.* 1998, mais voir MUELLER et WOLFENBARGER 1999). Il faut aussi savoir que contrairement aux RAPD, la technique AFLP est **très reproductible** (JONES *et al.* 1998a), mais cette reproductibilité reste limitée par l'interprétation des bandes et l'influence de la préparation de l'ADN génomique (ROBINSON et HARRIS 1999).

1.3.1.7 VNTR (Variable Number of Tandem Repeats)

Les **VNTR** (*Variable Number of Tandem Repeats*) sont des **éléments répétés**, de tailles variables, **distribués aléatoirement** au sein du génome de nombreux organismes et qui n'ont apparemment pas de fonction biologique connue³ La variabilité de tels marqueurs provient du nombre de répétitions de chaque motif. La taille des motifs répétés est aussi variable ce qui permet de différencier arbitrairement trois types de classes : les minisatellites et les microsatellites⁴. On définit les généralement **minisatellites** par des motifs supérieurs à 10 paires de bases (JEFFREYS *et al.* 1985, QUELLER *et al.* 1993, ASHLEY et DOW 1994, CHAMBERS et MACAVOY 2000) voire 15 paires de bases (BRUFORD *et al.* 1992), et dont le nombre de répétitions est variable entre les individus. Les minisatellites semblent évoluer à des taux de mutations de 10^{-3} - 10^{-2} par génération. Cette variabilité serait principalement due à la recombinaison inégale (*crossing over* inégal ; PARKER *et al.* 1998). Le haut taux d'évolution (donc le grand nombre potentiel de répétitions) et les motifs plus longs des minisatellites impliquent que ceux-ci ne peuvent être amplifiés par PCR (fragments à amplifier trop longs). Les minisatellites doivent être transférés sur membrane (d'où une perte potentielle d'ADN) et hybridés avec une ou plusieurs sondes marquées avant de migrer sur gel (voir FIGURE 1.8).

³Il existe des cas où l'insertion d'un certain nombre de motifs dans des zones particulières peut provoquer des maladies (LUBJUHNS *et al.* 1994) et où l'instabilité du nombre de motifs est liée à l'apparition de cancer (voir MAEHARA *et al.* 2001).

⁴le troisième type de VNTR sont les satellites, voir CHAMBERS et MACAVOY 2000

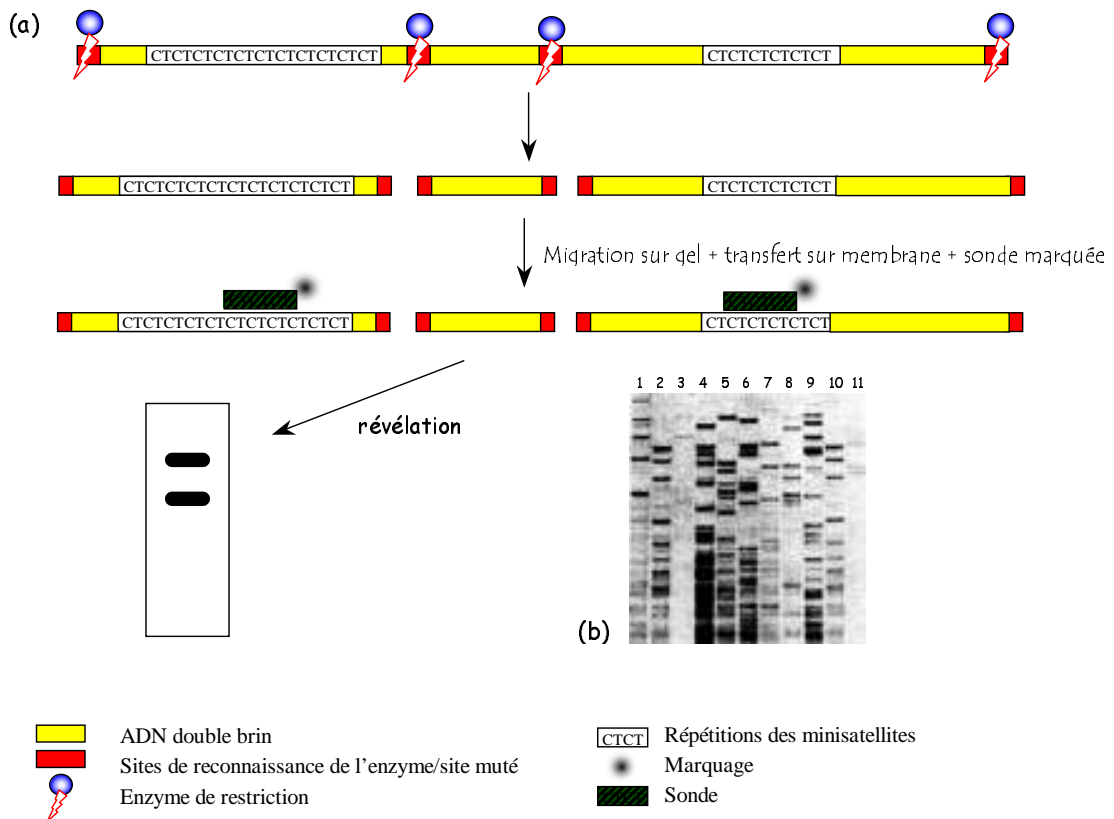


FIGURE 1.8 – Principe des marqueurs minisatellites. (a) représentation schématique de la procédure de la révélation du polymorphisme des minisatellites. (b) autoradiogramme montrant les bandes révélées pour une sonde appliquée chez 11 hommes (d'après JEFFREYS *et al.* 1985).

Les **microsatellites** (LITT et LUTY 1989) sont définis comme des répétitions de motifs de moins de six paires de bases (BECKMANN et WEBER 1992, HUGHES et QUELLER 1993, QUELLER *et al.* 1993, TAUTZ 1993, ASHLEY et DOW 1994, MCDONALD et POTTS 1997, PARKER *et al.* 1998, CHAMBERS et MACAVOY 2000), mais certains auteurs se basent sur quatre (WESTNEAT et WEBSTER 1994) ou cinq (BRUFORD et WAYNE 1993, JARNE et LAGODA 1996, MAEHARA *et al.* 2001) nucléotides. Les taux d'évolution du nombre de répétitions sont inférieurs à ceux des minisatellites (10^{-5} à 10^{-2} par génération, JARNE et LAGODA 1996, MCDONALD et POTTS 1997). Les mécanismes de mutation sont d'ailleurs différents entre les deux types de VNTR. Pour les microsatellites, la variabilité serait principalement due au décalage de la polymérase (*slippage*) et à un mauvais alignement des brins d'ADN (*misalignment*) durant la réplication (LEVINSON et GUTMAN 1987, EISEN 1998, ZHU *et al.* 2000) conduisant à l'insertion ou la délétion d'un (ou plusieurs) motif (LEVINSON et GUTMAN 1987, ASHLEY et DOW 1994, SCHLÖTTERER et PEMBERTON 1994, ZHU *et al.* 2000). Les *crossing over* inégaux (LEVINSON et GUTMAN 1987) peuvent aussi être impliqués et les mutations peuvent aussi être liées aux mécanismes de réparation de l'ADN (JARNE et LAGODA 1996). Certains auteurs (voir CHAMBERS et MACAVOY 2000) attribuent aux microsatellites une dynamique d'évolution selon un cycle de vie avec une naissance (accentuation des mutations dans une région déjà riches en motifs), une croissance (pendant laquelle le nombre d'allèle est contrôlé par les processus classiques de mutation, sélection, dérive et migration) et une mort (*e.g.* des délétions et des substitutions

faisant disparaître les motifs et les répétitions).

Les **microsatellites peuvent être amplifiés par PCR** (WEBER et MAY 1989, HUBERT *et al.* 1992) avec des amorces marquées puisque les longueurs des fragments répétés ne sont pas aussi importantes que pour les minisatellites. Les fragments amplifiés sont ensuite visualisés sur un gel d'électrophorèse (gel de polyacrylamide) après migration (voir FIGURE 1.9).

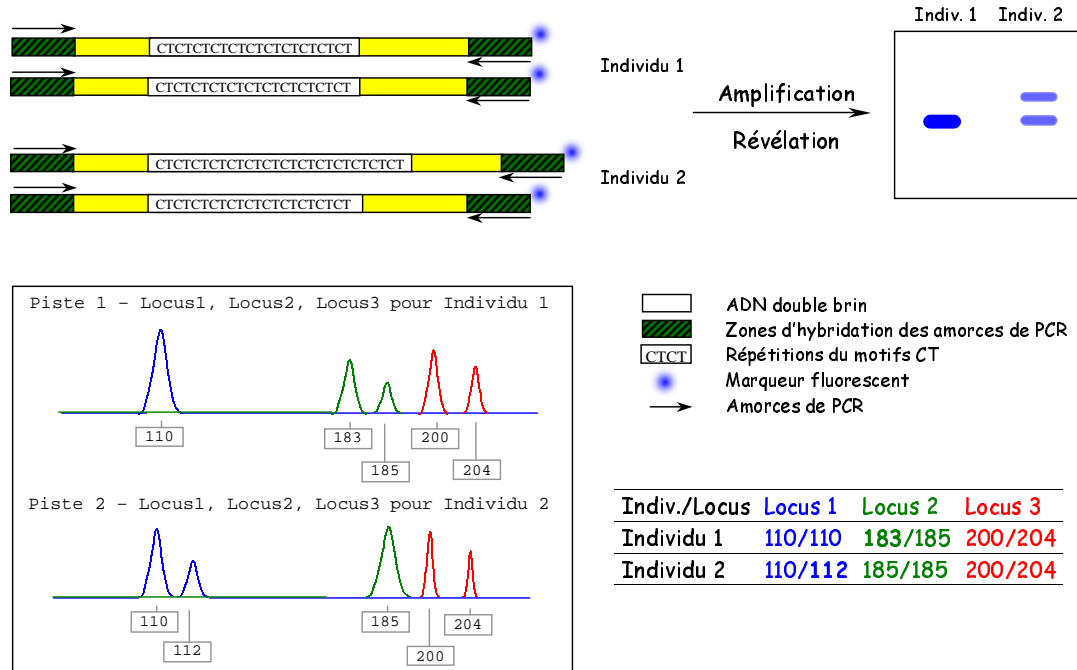


FIGURE 1.9 – Principe des marqueurs microsatellites ; En haut : représentation de la structure des microsatellites pour un locus et deux individus et résultats sur gel d'agarose après amplification ; En bas : représentation des pics de fluorescence (une couleur par locus microsatellite) sur un gel d'acrylamide après analyse de ce gel. Chaque pic correspond à un fragment amplifié d'un locus (allèle dont la longueur est indiquée). A droite, sont indiqués les génotypes (ensemble des longueurs des fragments amplifiés pour chaque locus) des deux individus avec en gras les allèles différenciant ces individus.

1.3.2 Applications et applicabilité des méthodes génétiques

Les outils génétiques peuvent être utilisés pour résoudre plusieurs questions en biologie des populations et de la conservation (voir TABLEAU 1.1) comme par exemple :

- déterminer les diversités génétiques des espèces ou des populations et les comparer entre elles,
- définir la structure des populations,
- clarifier les stratégies de reproduction,
- estimer la taille des populations,
- identifier les individus pour pouvoir les suivre au cours du temps,
- inférer les processus démographiques passés.

L'application des différentes méthodes à ces différentes questions biologiques dépend essentiellement du niveau de résolution (le polymorphisme) de chaque marqueur (PARKER *et al.* 1998,

TABLEAU 1.1 – Récapitulation des différentes caractéristiques des méthodes décrites dans le texte. Séq. = séquençage (ADNmt principalement); Alloz.= allozymes; msat= minisatellites multilocus; μ sat = microsatellites. Modifié d'après SCHLÖTTERER et PEMBERTON (1994), PARKER et al. (1998), MUELLER et WOLFENBARGER (1999), SUNNUCK (2000), FRANKHAM et al. (2002).

Caractéristiques	Séq.	SSCP	Alloz.	RFLP	RAPD	AFLP	msat.	μ sat.
<i>Application à de nouveaux taxons</i>	+	+/-	+	+	+	+	+/-	+/-
<i>Facilité</i>	+	+	++	-	+	+	-	+
<i>Niveau de variabilité</i>	- à ++	-/+	-	- à +	++	++	++	+
<i>Reproductibilité</i>	+	+	+	+	-	+/-	+/-	+
<i>ADN dégradé</i>	+	?	-	+	+	-	-	+
<i>Coût</i>	+	+	-	+	+	++	++	+
<i>Amplification</i>	+	+	-	+	+	+	-	+
<i>Soumis à sélection</i>	-/+	-/+	+	-/+	-/+	-/+	-/+	- (?)
<i>Interprétation des gels</i>	+	+	++	-	-	-	-	+
<i>Difficulté de mise au point</i>	+	?	-	++	-	+	+	++
<i>Codominance</i>	+	+	+	+	-	-/+	-	+
<i>Identification d'individus</i>	-/+	-	-	-	+	++	+++	+++
<i>Taxonomie</i>	++	+	+	++	++	?	++	+/-
<i>Structure de population</i>	++	++	++	+	+	+	+	++
<i>Parenté</i>	-	-	-	+	-	+	++	++
<i>Echantillons non-invasifs</i>	+	?	-	-/+	+	+/-	-	+
<i>Criminologie (Forensics)</i>	++	+	+	?	++	+	++	+++
<i>Taille de population</i>	+++	?	-	?	+	?	+/-	+++
<i>Taille efficace</i>	+	?	++	?	-	?	+/-	+++
<i>Détection de goulots d'étranglement</i>	++	?	++	?	++	?	+/-	+++
<i>Phylogéographie</i>	+++	?	-	+	-	?	-	+
<i>Détermination de source pour ré-introd.</i>	+	?	++	+	++	?	+++	+++

SUNNUCK 2000) mais aussi du niveau d'organisation étudié : espèce, population, groupe familial, individu (voir TABLEAU 1.1). Dans cette partie nous allons voir quelles sont les méthodes pouvant être utilisées pour certains objectifs, sans toutefois réaliser une revue exhaustive des utilisations des méthodes.

1.3.2.1 Niveau populationnel : structure des populations

La **diversité génétique** d'une population est un facteur très important pour donner des priorités ou préciser les moyens de conservation (CROZIER 1992) car elle peut renseigner sur l'**isolement**, les **flux génétiques** (dispersion génétique) existants entre les populations (MILLIGAN *et al.* 1994).

Les techniques citées peuvent, a priori, toutes servir à décrire la **structure de populations**. Les allozymes ont été utilisés à partir de 1966, avec une série de publications sur la quantification de la variation génétique de populations humaines et de drosophiles (POWELL 1994). Elles ont été longtemps utilisées pour déterminer la structure à l'intérieur des populations, la variation à l'intérieur des espèces (voir BAVERSTOCK 1988, PARKER *et al.* 1998) mais les allozymes peuvent ne pas contenir assez de variabilité pour déterminer avec précision ces structures. En revanche, le principal avantage des allozymes est leur facilité et leur rapidité de mise en oeuvre. On préférera néanmoins les marqueurs RFLP qui peuvent être facilement utilisés pour déterminer la sous-structuration de population. L'application des RFLP sur l'ADN mitochondrial (PCR-RFLP) ou le séquençage de cet ADN, peuvent être des outils très utiles pour étudier la structuration aussi bien entre populations que dans les populations (*e.g.* PALUMBI *et al.* 1991, TABERLET et BOUVET 1992, RANDI *et al.* 1995, GIRMAN *et al.* 1997, BARRATT *et al.* 1999, VILÀ *et al.* 1999, RANDI *et al.* 2000). Les séquences d'ADN mitochondrial servent en outre à quantifier la diversité génétique au sein d'une population mais aussi à déterminer la structure géographique et génétique de population, des événements de goulots d'étranglement ou des patrons de colonisation (HARRISON 1989, TABERLET 1996). En revanche, il y a quelques limites à l'utilisation de l'ADNmt, comme par exemple la transmission maternelle (une différenciation peut être détectée à cause d'un déficit de dispersion des femelles modifiant la signature génétique, TABERLET et BOUVET 1994) ou le fait que le génome mitochondrial se comporte comme une seule unité génétique (la dérive génétique peut conduire à des phylogénies incorrectes, NEI 1996).

Le principal problème de l'utilisation des marqueurs RAPD et AFLP au niveau de la population est qu'ils sont dominants (il existe un débat sur la co-dominance des marqueurs AFLP, mais la reconnaissance de la co-dominance des AFLP reste rare, M. Gaudel, comm. pers.). Dans ce cas, on ne peut avoir accès aux hétérozygotes et donc cela limite l'estimation de la diversité génétique. De plus, l'apparition de bandes artificielles et le manque de reproductibilité pour les RAPD (JONES *et al.* 1998b) limitent aussi ces applications. Néanmoins, ces deux marqueurs ont été utilisés dans quelques cas pour comparer la diversité génétique entre populations en danger et non menacées (voir FRANKHAM *et al.* 2002, p.66 pour les références)

Les minisatellites peuvent être également utilisés pour étudier les populations mais ils présentent un fort polymorphisme (BRUFORD *et al.* 1992) et ils peuvent être trop variables pour les études à ce niveau (BRUFORD *et al.* 1992, PARKER *et al.* 1998). De plus, il existe une variabilité dans la distribution des proportions des bandes partagées due à la variation du nombre de bandes qui peuvent être détectées (PARKER *et al.* 1998).

Les microsattellites sont aisément utilisables pour plusieurs applications au niveau de la population (voir plus loin) car on peut étudier un seul locus à la fois et les fréquences alléliques

peuvent être calculées. Ceci permet l'utilisation des statistiques F , initialement développées pour les allozymes. Néanmoins, la structuration de populations assez éloignées est limitée par le même problème que la détermination des distances génétiques entre taxons éloignés, c'est-à-dire la saturation du nombre de répétitions des motifs (FORBES et HOGG 1999).

1.3.2.2 Niveau familial et du groupe social : la détermination des parentés

Le **degré de consanguinité** (liée à l'apparentement des individus reproducteurs) et les **structures de reproduction** peuvent avoir d'importantes conséquences dans le contexte de la conservation de populations (MILLIGAN *et al.* 1994). Il est donc nécessaire de connaître le fonctionnement de la population au niveau de la reproduction pour pouvoir agir convenablement, surtout chez des espèces difficiles à étudier (voir le cas de la détermination de la structure de population de dauphin pilote ou globicéphale grinde, *Globicephala melas* à partir d'échantillons de produits de la chasse de ce cétacé, AMOS *et al.* 1993). Ce fonctionnement peut être appréhendé par la **détermination des liens de parenté** existant entre les individus formant les groupes sociaux.

Grâce à la détermination des parentés entre individus, l'étude de nombreux types de groupes sociaux a pu être menée (voir HUGHES 1998). Par exemple, la plupart des techniques (allozymes, RFLP, minisatellites multilocus, microsatellites) ont permis de montrer des taux de près de 20% de fertilisation hors couple chez les oiseaux (voir Appendix 1 de WESTNEAT et WEBSTER 1994).

Les liens de parenté peuvent être étudiés avec les différents outils décrits plus haut mais le polymorphisme de certains marqueurs peut limiter leur application. C'est le cas des allozymes qui présentent une variabilité assez faible (voir HUGHES et QUELLER 1993, WESTNEAT et WEBSTER 1994).

En revanche, les marqueurs RFLP sont potentiellement utilisables dans les applications de détermination de la parenté (grâce à la détection des hétérozygotes) mais l'utilisation de l'ADN mitochondrial n'est pas concluante pour ces applications fines (HUGHES et QUELLER 1993). D'autre part, l'analyse de l'ADN mitochondrial se limite à la détermination des lignées maternelles (sa transmission est uniquement maternelle) et l'exclusion de maternité n'a pas été concluante dans certaines études (voir PARKER *et al.* 1998).

En revanche, les AFLP peuvent être utilisés pour les études de parenté car le nombre de loci générés par une expérience peut être très important pour différencier avec précision les individus d'une population (MUELLER et WOLFENBARGER 1999, ROBINSON et HARRIS 1999).

Enfin les VNTR sont les marqueurs les plus adaptés à ce type d'étude. Les liens de parenté peuvent être déterminés grâce aux minisatellites multilocus, en estimant la proportion de bandes partagées par deux individus (BRUFORD *et al.* 1992). Néanmoins, les problèmes d'assignation des bandes (phénomènes d'allélisme où une bande peut provenir de plusieurs loci) associés à l'impossibilité de différencier avec précision les hétérozygotes et les homozygotes et le fait que des individus non-apparentés puissent avoir des bandes communes, diminuent grandement les possibilités de détection des liens de parenté grâce aux minisatellites (LYNCH 1988).

Les **microsatellites présentent des avantages** non négligeables pour la détermination des liens de parenté et de la structure de populations : les locus peuvent être analysés séparément, ils présentent un polymorphisme suffisant en fournissant des génotypes facilement interprétés et enfin les hétérozygotes peuvent être détectés.

1.3.2.3 Niveau individuel : identification et suivi des individus

La possibilité d'**identifier les individus** d'une population peut être un moyen extrêmement puissant pour déterminer le **fonctionnement de cette population** (voir la détermination de la parenté plus haut) et **estimer certains paramètres** comme la taille de population. L'identification visuelle par des caractéristiques morphologiques d'un individu ou le marquage est un bon moyen pour suivre les individus afin, par exemple, d'étudier leur **comportement**, connaître la **distribution géographique** précise d'un groupe, estimer la **taille du groupe**. Dans cette optique, il est nécessaire que ce type de « marque » (naturelle ou artificielle) réponde à certains critères (PALSBØLL 1999). Une marque doit ainsi :

- pouvoir être appliquée sur tous les animaux de la population quels que soient leur sexe ou leur âge,
- pouvoir être reconnue à distance et de préférence de façon non-invasive,
- ne pas être perdue au cours du temps (notamment avec l'âge de l'animal),
- pouvoir identifier les individus de la population sans ambiguïté possible,
- permettre une comparaison des marques efficace et rapide.

Toutes les marques « classiques » (utilisation de photographies pour reconnaître les individus, marques auriculaires, collier émetteur, transpondeur) ne satisfont pas toujours toutes ces conditions. Par exemple, une caractéristique morphologique d'un individu peut disparaître au cours de sa vie. L'utilisation de photographies peut être délicate à cause de l'angle de prise de vue. Une marque auriculaire ou un collier émetteur peuvent être perdus. Pour lire un transpondeur, il est souvent nécessaire de capturer les individus.

L'apparition des outils moléculaires et notamment des marqueurs microsatellites a considérablement révolutionné cette pratique. Les **marqueurs moléculaires** (comme les microsatellites par exemple) **répondent à la plupart des critères** cités :

- ils sont applicables universellement, notamment chez la majorité des animaux.
- ils peuvent être appliqués à tout individu de la population, quels que soient son âge ou son sexe, pour peu que des échantillons soient disponibles.
- ils ne sont pas perdus au cours de la vie d'un individu et restent inchangés durant cette période.
- ils peuvent être déterminés à partir d'échantillons récoltés de façon non-invasive (voir plus loin).
- ils peuvent simplement être comparés entre eux pour une identification rapide.

Néanmoins, une des difficultés de l'identification individuelle par les marqueurs génétiques est de trouver un système de marqueurs assez polymorphes⁵ pour pouvoir identifier sans ambiguïté chaque individu, c'est-à-dire d'avoir une probabilité d'identité la plus faible possible (voir Section 6.2.2.2.3, p.84). Avec les marques non génétiques, comme l'étiquetage auriculaire, l'expérimentateur choisit son système de telle façon que les marques soient facilement reconnaissables et qu'elles ne puissent pas être confondues entre elles. Or, avec les marqueurs génétiques, la marque n'est pas choisie par l'expérimentateur mais déterminée par les processus d'évolution moléculaire (*e.g.* mutations) et/ou dictée par les mécanismes mendéliens de la transmission des gènes. L'identification individuelle requiert donc un fort polymorphisme. Seuls les AFLP et les

⁵Un locus est dit polymorphe dès qu'il a au moins deux allèles et que la fréquence d'un de ces allèles ne dépasse pas 95% ou 99%. Le niveau de polymorphisme est déterminé par le nombre d'allèles (formes alléliques) et leur fréquence respective.

VNTR peuvent servir à l'identification d'individus car les autres marqueurs ne sont pas assez polymorphes (comme les allozymes), ou les techniques ne sont pas reproductibles (comme les RAPD). Il semble que l'utilisation des **marqueurs microsatellites soit la technique la plus prometteuse** pour cette application. En effet, il apparaît que les minisatellites (surtout les marqueurs multilocus) présentent des patrons génétiques souvent compliqués pour une interprétation aisée, et que les AFLP présentent une interprétation des gels assez subjective lors de la sélection des pics informatifs (ROBINSON et HARRIS 1999). Les minisatellites ont néanmoins été des marqueurs très utilisés pour l'identification des individus (FRANKHAM *et al.* 2002, p.64). En revanche, l'interprétation des gels de microsatellites est relativement aisée, et dans certains cas sans aucune ambiguïté. L'utilisation indépendante de locus permet également d'augmenter la précision de l'identification puisqu'on pourra choisir des systèmes de loci qui donneront une puissance d'identification et de discrimination des individus suffisante.

1.3.3 Avantages et inconvénients des microsatellites

Il semble que les microsatellites soient des marqueurs qui peuvent être facilement utilisée à différentes échelles biologiques (MCDONALD et POTTS 1997). Elle est la technique la plus applicable et la plus informative pour mesurer la variation de l'ADN à différent niveaux. Les microsatellites présentent donc certaines caractéristiques qui leur permettent d'être des marqueurs de choix dans beaucoup d'applications (voir revues dans QUELLER *et al.* 1993, ASHLEY et DOW 1994, SCHLÖTTERER et PEMBERTON 1994, MCDONALD et POTTS 1997, BALDING 1998, BEAUMONT et BRUFORD 1998, LUIKART et ENGLAND 1999).

1.3.3.1 Mise au point et développement

Les microsatellites sont présents chez de nombreuses espèces de vertébrés (mammifères, oiseaux, poissons, amphibiens ; voir TAUTZ et RENTZ 1984, ASHLEY et DOW 1994), mais aussi chez les invertébrés (voir par exemple HUGHES et QUELLER 1993) et les plantes en moins grande abondance. Chez les mammifères, la variabilité des microsatellites semble être un phénomène général. L'obtention de loci microsatellites (voir revue de ZANE *et al.* 2002) qui montrent une certaine variabilité dans le nombre de répétitions peut être longue. Cela nécessite la construction d'une banque génomique pour sélectionner les régions d'ADN présentant des répétitions suivie du test des loci repérés. L'amplification par PCR nécessite des oligonucléotides afin d'amorcer la réaction, ces amorces sont souvent particulières à chaque espèce ou groupe d'espèces. Il est donc nécessaire de déterminer des couples d'amorces pour chacun des taxons analysés, un travail souvent long et fastidieux (BRUFORD et WAYNE 1993). Une stratégie alternative consiste à tester les amorces déjà développées chez une espèce voisine sur l'espèce d'étude (par exemple l'homme *Homo sapiens* pour les bonobos *Pan paniscus* GERLOFF *et al.* 1995 ou les entelles *Presbytis entellus* LAUNHARDT *et al.* 1998 ; le chien *Canis familiaris* pour le loup *Canis lupus*, voir Annexe A).

1.3.3.2 Révélation des marqueurs

L'utilisation des marqueurs VNTRs multilocus ou des RFLP conduit bien souvent à des difficultés d'interprétation des gels qui sont dues :

- aux patrons compliqués des bandes (d'où l'appellation d'empreinte ADN en comparaison avec le code barre d'un produit de supermarché),
- à l'interprétation de chaque bande par rapport aux allèles et loci et à la difficulté de reproductibilité d'un individu à l'autre et surtout d'un gel à l'autre (PARKER *et al.* 1998).

Bien que l'utilisation de marqueurs unilocus (c'est-à-dire ceux pour lesquels on peut analyser un seul locus à la fois) corrige ce problème et permette une identification fiable des individus, les minisatellites et la technique *Southern blot* associée restent des méthodes coûteuses et longues car il est nécessaire de développer des sondes qui ne révèlent qu'un seul locus à la fois (PARKER *et al.* 1998 ; mais voir EPPLEN *et al.* 1991). En revanche, les marqueurs microsatellites reposent sur une application moins coûteuse (amplification par PCR ; sept fois moins coûteuse par échantillon, PARKER *et al.* 1998) et plus rapide, une fois les mises au point réalisées et le matériel acheté (séquenceur automatique par exemple). De plus, les patrons de révélation des marqueurs microsatellites peuvent être très simples à interpréter car :

- les microsatellites sont des marqueurs unilocus,
- il n'existe que deux allèles au maximum par individu,
- l'utilisation de gel de polyacrylamide permet de déterminer la taille des allèles avec une précision de 1 paire de bases.

1.3.3.3 Amplification PCR

L'un des principaux avantages des microsatellites est de pouvoir être amplifiés par PCR (WEBER et MAY 1989). Les marqueurs microsatellites, grâce aux motifs de plus petites tailles et des répétitions moins nombreuses que pour les minisatellites, représentent des fragments de 100 à 300 paires de bases, ce qui les rend aisément amplifiables par PCR. La possibilité d'amplifier de l'ADN est un avantage inestimable en biologie de la conservation (BEAUMONT et BRUFORD 1998) ou en criminologie (BALDING 1998), car des échantillons ne contenant qu'une faible quantité d'ADN (féces, urine, salive, sperme, poils), quelque fois dégradé, peuvent être analysés (e.g. application sur une seule cellule de sperme HUBERT *et al.* 1992).

1.3.3.4 Interprétation et utilisation des résultats

Il est possible de calculer les fréquences alléliques aux différents loci et il est ainsi possible d'avoir accès à de nombreuses méthodes et tests pour étudier la diversité génétique (équilibre de HARDY-WEINBERG, déséquilibre de liaison, déficit d'hétérozygote). Chaque locus peut être étudié séparément et indépendamment des autres. On peut aussi associer ces loci pour peu qu'ils soient sur des groupes chromosomiques différents et ainsi par exemple augmenter la puissance de discrimination des individus (voir WAITS *et al.* 2001).

Les marqueurs microsatellites permettent l'utilisation directe de méthodes statistiques développées initialement pour les allozymes, notamment les statistiques F (ou G) pour la différenciation de populations (voir WEIR et COCKERHAM 1984) et des applications de ces statistiques (voir par exemple la détection de la dispersion et du biais de cette dispersion entre sexe, WASER et STROBECK 1998, GOUDET *et al.* 2002, PRUGNOLLE et DE MEEUS 2002). De nouvelles approches sont développées comme par exemple les méthodes de maximum de vraisemblance, les méthodes de coalescence, les méthodes bayésiennes et les microsatellites sont des marqueurs de choix pour ce type de développement statistique (voir LUIKART et ENGLAND 1999).

Les microsatellites peuvent donc être utilisés pour de nombreuses applications dont⁶ :

- Estimation de la diversité génétique et des structures génétiques de populations
 - ALLEN *et al.* (1995) chez le phoque gris *Halichoerus grypus*
 - CIOFI et BRUFORD (1999) chez le dragon de Komodo *Varanus komodoensis*

⁶Cette liste n'est certainement pas exhaustive ; elle est principalement centrée sur l'utilisation des échantillons non-invasifs.

- FORBES et BOYD (1996) chez le loup *Canis lupus*
- FORBES et HOGG (1999) chez le bighorn *Ovis canadensis*
- GOOSSENS *et al.* (2001) chez la marmotte alpine *Marmota marmota*
- GOTTELLI *et al.* (1994) chez le loup éthiopien *Canis simiensis*
- KYLE et STROBECK (2001) chez le glouton *Gulo gulo*
- PAETKAU et STROBECK (1994), PAETKAU *et al.* (1998a, b) chez l'ours brun, *Ursus arctos*
- PAETKAU *et al.* (1995, 1999) chez l'ours polaire *Ursus maritimus*
- TAYLOR *et al.* (1994) chez le wombat à museau velu du nord *Lasiiorhinus kreftii*
- Détermination de la structure sociale, des succès de reproduction et du système de reproduction de populations
 - GAGNEUX *et al.* (1999), MORIN *et al.* (1994b) chez le chimpanzé *Pan troglodytes*
 - GARNIER *et al.* (2001) chez le rhinoceros noir *Diceros bicornis*
 - GERLOFF *et al.* (1999) chez le bonobo *Pan paniscus*
 - GIRMAN *et al.* (1997) chez le lycaon *Lycaon pictus*
 - KAYS *et al.* (2000) chez le kinkajou *Potos flavus*
 - SAY *et al.* (1999, 2001) chez le chat domestique *Felis catus*
- Estimation de taille de population
 - AMOS *et al.* (1993), PALSBOELL *et al.* (1997) chez la baleine à bosse *Megaptera novaeangliae*
 - KOHN *et al.* (1999) chez le coyote *Canis latrans*
 - MOWAT et STROBECK (2001) chez l'ours brun
 - MOWAT et PAETKAU (2002) chez la martre américaine *Martes americana*
 - NICHOLS *et al.* (2001) chez le faucon crécerelle mauricien *Falco punctatus* (taille efficace)
- et les variations de cette taille (voir les méthodes dans BEAUMONT 1999, LUIKART *et al.* 1999, PIRY *et al.* 1999, GARZA et WILLIAMSON 2001)
- Identification individuelle
 - ELLEGREN (1991) pour des échantillons d'oiseaux de musée
 - ERNEST *et al.* (2000) chez le puma *Puma concolor*
 - REED *et al.* (1997) chez le phoque gris et la veau marin *Phoca vitulina*
 - SLOANE *et al.* (2000) chez le wombat à museau velu du nord
 - TABERLET *et al.* (1997) chez l'ours brun
- Exclusion et assignation de parenté
 - GARNIER *et al.* (2001) chez le rhinoceros noir
 - MORIN et WOODRUFF (1992), MORIN *et al.* (1994b), GAGNEUX *et al.* (1999), CONSTABLE *et al.* (2001) chez le chimpanzé
 - SAY *et al.* (1999) chez le chat domestique
 - VALSECCHI *et al.* (2002) chez la baleine à bosse

Chapitre 2

L'échantillonnage non-invasif

Sommaire

2.1	Les différentes formes de méthodes d'échantillonnage non-invasif	30
2.2	Les méthodes non-invasives et la génétique	31
2.2.1	Rendement d'amplification	31
2.2.2	Erreurs de géotypage liées à la quantité et la qualité de l'ADN	32
2.2.2.1	Révélation de fragments surnuméraires	33
2.2.2.2	Non révélation de fragments	34
2.2.2.3	Autres erreurs	36
2.2.2.4	Taux d'erreurs	36

Auparavant, pour étudier la génétique des populations, l'échantillonnage nécessitait le prélèvement direct à partir des individus, de grandes quantités de tissu (muscle, peau) destinées à des méthodes moléculaires du type électrophorèse de protéines. Cette démarche « destructive » pouvait nécessiter de sacrifier l'animal pour « la cause scientifique » (TABERLET et LUIKART 1999). Pour certaines espèces qui le permettent, on peut capturer l'individu et lui prélever (méthodes non-destructives mais toutefois invasives) un petit échantillon de tissu musculaire ou de peau (*e.g.* biopsie au niveau de l'oreille) ou lui prélever du sang. Ces techniques restent très difficiles à mettre en oeuvre dans certains cas comme :

- des animaux de grande taille, à cause de la difficulté de capture et de manipulation,
- des animaux discrets (*e.g.* comportements nocturnes) et/ou farouches,
- des populations en danger car les captures peuvent engendrer une mortalité chez les individus ou les observations/captures peuvent déranger la structure sociale des populations et compromettre la viabilité de la population.

Ces considérations ont amené les scientifiques à développer des méthodes permettant d'obtenir des échantillons sans déranger ou capturer les animaux (**méthodes non-invasives**) afin d'en extraire de l'ADN pour des analyses génétiques ultérieures. Ce développement est largement lié à la mise au point de la PCR (voir Section 1.3.1.2, page 13) qui permet de multiplier le nombre de copies d'ADN d'un extrait.

2.1 LES DIFFÉRENTES FORMES DE MÉTHODES D'ÉCHANTILLONNAGE NON-INVASIF

Il existe plusieurs formes de méthodes non-invasives. Les méthodes non-invasives strictes constituent la forme la plus simple et la plus pratique car elle consiste à récolter sur le terrain des indices de présence. Cet échantillonnage consiste par exemple à récolter des fèces directement sur le sol (voir Article I ou KOHN *et al.* 1999), récolter de l'urine de Canidés dans la neige (voir Article II), de prélever des poils d'ours sur les écorces d'arbres (TABERLET et BOUVET 1992), des restes de feuilles mastiquées par des Chimpanzés (GAGNEUX *et al.* 1999) ou enfin des plumes d'oiseaux (SEGELBACHER 2002). L'échantillonnage non-invasif le plus singulier est peut être celui de fèces de mammifères marins (TIKEL *et al.* 1996 chez le dugong *Dugong dugong*, REED *et al.* 1997 chez le veau marin *Phoca vitulina* et le phoque gris *Halichoerus grypus*, PARSONS 2001 chez le dauphin souffleur *Tursiops truncatus*). Avec les méthodes strictes, l'animal n'est ni observé, ni manipulé. Cette forme d'échantillonnage ne nécessite donc pas d'infrastructure ou de logistique particulière.

Une deuxième forme de méthodes non-invasives, nécessite l'observation des individus avant le prélèvement des échantillons. Par exemple, dans les communautés de chimpanzés (*Pan troglodytes*; voir par exemple MORIN et WOODRUFF 1992, MORIN *et al.* 1994b, CONSTABLE *et al.* 2001, GAGNEUX *et al.* 1997, 1999), les poils sont récoltés dans les nids que construisent chaque soir les chimpanzés pour la nuit. Dès le matin, les poils sont prélevés uniquement dans les nids pour lesquels la construction par un chimpanzé ayant dormi seul à l'intérieur a été observée. L'observation des animaux est donc une nécessité dans ce type d'échantillonnage si l'on veut réaliser une identification correcte (pas de mélange d'échantillons) des individus.

Enfin, d'autres méthodes non-invasives nécessitent la mise en place d'un dispositif permettant de prélever des poils. WOODS *et al.* (1999), MOWAT et STROBECK (2001) et POOLE *et al.* (2002) ont développé et mis en application un dispositif permettant de récolter des poils sur des ours grizzlys (*Ursus arctos*). Les grizzlys sont de grands mammifères qui peuvent être agressifs et leur

capture nécessite de gros moyens logistiques. Le dispositif utilisé (voir FIGURE 2.1a) consiste en un appât placé en hauteur au centre d'un cercle de fils barbelés disposé à 50 cm du sol de sorte que l'ours, par l'odeur alléché, passe au-dessous des fils barbelés. Ses poils sont alors retenus par les barbelés. Un autre dispositif consiste en deux rondins de bois parcourus par des peignes ou des brosses retenant les poils des ours attirés par l'appât placé au sommet (voir FIGURE 2.1b). FORAN *et al.* (1997b) et MOWAT et PAETKAU (2002) ont développé et appliqué un dispositif permettant de prélever des poils sur des petits carnivores discrets des forêts américaines (notamment la martre d'Amérique *Martes americana*). Il s'agit de deux planches de bois placées en V le long d'un arbre. Sur les parois internes, des adhésifs collent et « capturent » les poils du dos des animaux passant à l'intérieur (voir FIGURE 2.1c). Enfin un dernier dispositif peut être cité, il s'agit simplement de rubans adhésifs placés à l'entrée de terriers afin de prélever les poils de wombats (*Lasiorhinus kreftii*) lorsqu'ils passent par l'entrée du terrier (SLOANE *et al.* 2000).

2.2 LES MÉTHODES NON-INVASIVES ET LA GÉNÉTIQUE

Les **échantillons non-invasifs** peuvent constituer une **source de matériel génétique** en extrayant l'ADN contenu dans les cellules. Les échantillons récoltés de façon non-invasive contiennent dans la plupart des cas une **quantité limitée d'ADN**. Par exemple, dans les poils, les cellules exploitables se limitent aux cellules de la racine (bien que l'on puisse extraire de l'ADN à partir de la partie kératinée mais les pigments et probablement la kératine inhibent les réactions d'amplification, MORIN et WOODRUFF 1996). Il en est de même pour les plumes, pour lesquelles, on utilise un bout de la racine (SEGELBACHER 2002). Pour les fèces, les rares cellules exploitables sont les cellules du tube digestif de l'hôte, desquamées lors du transit intestinal (FORAN *et al.* 1997a, KOHN et WAYNE 1997). Les cellules du tractus digestif sont les seules à résister aux conditions du transit et c'est donc l'ADN de ces cellules qui sera extrait des fèces. Les cellules récupérées de l'urine pourraient provenir aussi d'une desquamation du système urinaire en plus des leucocytes présents chez certains sujets (YOKOTA *et al.* 1998). Ces échantillons non-invasifs n'ont pu être utilisés en génétique des populations que grâce à la découverte de la PCR (voir Section 1.3.1.2, p.13) puisque cette réaction permet de multiplier les copies d'ADN et d'obtenir une quantité suffisante (dans la majorité des cas) d'ADN pour les analyses. En revanche, l'ADN de ces échantillons est **souvent dégradé** à cause des conditions rencontrées sur le terrain (humidité, pluie, gel/dégel; GOOSSENS *et al.* 2000, MURPHY *et al.* 2000; voir LUCCHINI *et al.* 2002) ou dans le tractus digestif dans le cas de fèces (enzymes et sucs digestifs). La qualité de l'ADN et de son extraction est par conséquent aussi liée au régime alimentaire (herbivore, carnivore) de l'animal ainsi qu'à sa variation pendant l'année (GOOSSENS *et al.* 2000, MURPHY *et al.* 2000, MAUDET *et al.* 0000). La faible quantité et qualité d'ADN, associée à la **présence d'inhibiteurs de PCR** pour certains échantillons (*e.g.* les fèces, GERLOFF *et al.* 1995 ou l'urine, KAHN *et al.* 1991 cité dans MARKLUND *et al.* 1996) limitent toutefois les analyses. D'une part, le **succès d'amplification peut être restreint**, et d'autre part des **erreurs peuvent être réalisées** lors de la détermination de génotypes microsatellites.

2.2.1 Rendement d'amplification

Les conditions citées ci-dessus ne permettent pas un bon rendement d'amplification. En effet, le pourcentage de PCR positives (c'est-à-dire les PCR donnant un produit d'amplification dont la taille est compatible avec celle attendue) est toujours inférieur à celui obtenu à partir de tissu ou de sang (*e.g.* BAYES *et al.* 2000, SLOANE *et al.* 2000). Pour les fèces ou les poils, ce pourcentage est très variable (compris en général entre 50% et 90%). Une revue des estimations

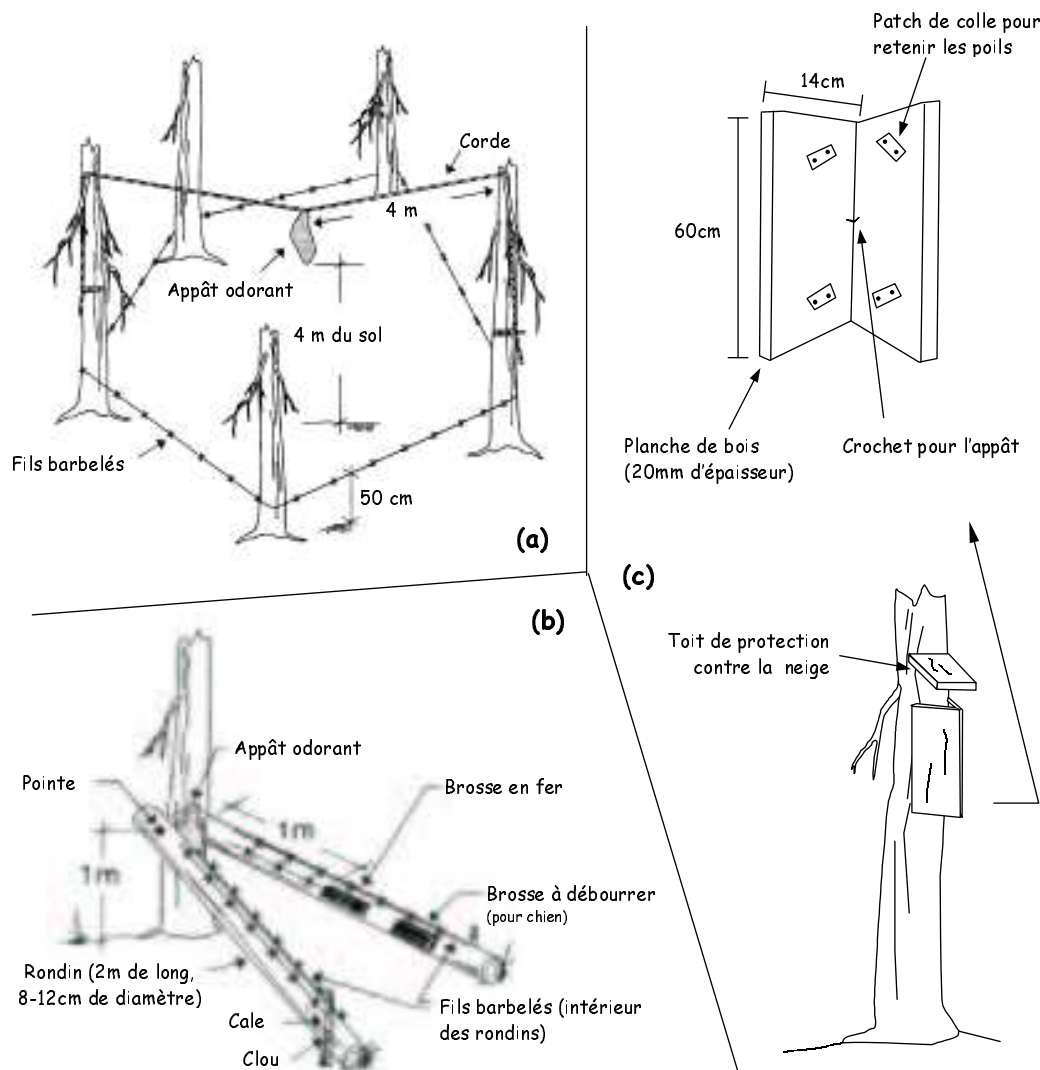


FIGURE 2.1 – Dispositifs de « capture » de poils pour deux types de mammifères : (a) et (b) deux dispositifs adaptés aux grizzlys (*Ursus arctos horribilis*). Modifié d'après WOODS *et al.* (1999) ; (c) dispositif adapté à la martre américaine (*Martes americana*). En bas : positionnement du dispositif sur un arbre. En haut : vue interne du dispositif. Dessin réalisé à partir des photographies dans (FORAN *et al.* 1997b).

de succès d'amplification est donnée dans le TABLEAU 2.1. Les chiffres diffèrent grandement en fonction du nombre de loci microsatellites pris en compte, du mode de calcul réalisé soit sur les échantillons soit sur les réactions PCR, de la différence des méthodes d'extraction utilisées, du type d'échantillon et de l'espèce étudiée.

2.2.2 Erreurs de géotypage liées à la quantité et la qualité de l'ADN

On sait aussi que le manque de quantité et de qualité de l'ADN peuvent provoquer des erreurs lors d'analyses génétiques utilisant la PCR (*e.g.* GERLOFF *et al.* 1995, TABERLET *et al.* 1996, DOUGLAS *et al.* 2002). Deux grands types d'erreurs d'amplification peuvent être causés : la production de fragments d'ADN non présents dans l'échantillon de départ et la non-révélation de fragments d'ADN.

TABLEAU 2.1 – Valeur de succès d’amplification dans différentes études publiées. Pour chaque étude, le type d’échantillons (Type), l’espèce, le nombre de loci microsatellites utilisés (N loci) et les valeurs du succès d’amplification (Succès PCR) sont données. Les annotations entre parenthèses précisent le mode de calcul du succès : P = sur le nombre d’amplifications (PCR), E = sur le nombre d’extraits, S = sur le nombre d’échantillons, L = sur le nombre de couples locus/échantillons, # = succès d’amplification pour tous les loci, \$ = amplification positive pour au moins un locus. * = après lyophilisation, ** = après étuvage.

Référence	Espèce	Type	N loci	Succès PCR
BAYES <i>et al.</i> (2000)	<i>Papio sp</i>	Fèces	8	70% (P)
CONSTABLE <i>et al.</i> (2001)	<i>Pan troglodytes</i>	Fèces	16	71% (E)
		Poils		90% (E)
MORIN <i>et al.</i> (2001)	<i>Pan troglodytes</i>	Fèces	9	84,1% (E, #) 79% (P)
GERLOFF <i>et al.</i> (1995)	<i>Pan paniscus</i>	Fèces	6	66% (P)
GOOSSENS <i>et al.</i> (2000)	<i>Pongo pygmaeus</i>	Fèces	1	65% (P)
LATHUILLIÈRE <i>et al.</i> (2001)	<i>Macaca sylvanus</i>	Fèces	3	70% (P)
LAUNHARDT <i>et al.</i> (1998)	<i>Presbytis entellus</i>	Fèces	5	70% (E)
KOHN <i>et al.</i> (1999)	<i>Canis latrans</i>	Fèces	3	96,2% (P)
				48% (S, #)
LUCCHINI <i>et al.</i> (2002)	<i>Canis lupus</i>	Fèces	6	54% (S)
MIQUEL ET TABERLET (non pub.)	<i>Canis lupus</i>	Fèces	6	94% (P)
MURPHY <i>et al.</i> (2000)	<i>Ursus arctos</i>	Fèces	1	89% (P)*
				59% (P)**
TABERLET <i>et al.</i> (1996)	<i>Ursus arctos</i>	Fèces	1	88% (P)
TABERLET <i>et al.</i> (1997)	<i>Ursus arctos</i>	Fèces	6	14,6% (S, #)
		Poils		20% (S, #)
ERNEST <i>et al.</i> (2000)	<i>Puma concolor</i>	Fèces	12	34,4% (S, #) >75% (E)
REED <i>et al.</i> (1997)	<i>Phoca vitulina</i>	Fèces	5	20% (S, #)
				93% (S, \$)
GARNIER <i>et al.</i> (2001)	<i>Diceros bicornis</i>	Fèces	10	93% (P)
PARSONS (2001)	<i>Tursiops truncatus</i>	Fèces	3	100% (P)
GAGNEUX <i>et al.</i> (1997)	<i>Pan troglodytes</i>	Poils	11	50% (E, \$)
GAGNEUX <i>et al.</i> (1999)	<i>Pan troglodytes</i>	Poils	11	63,3% (S, #)
				3
MORIN <i>et al.</i> (1994b)	<i>Pan troglodytes</i>	Poils	8	82,2% (S, \$)
MOWAT ET PAETKAU (2002)	<i>Martes Americana</i>	Poils	6	77,2% (S, #)
SLOANE <i>et al.</i> (2000)	<i>Lasiornhinus krefftii</i>	Poils	12	88% (P)
VALIÈRE ET TABERLET (2000)	<i>Canis lupus</i>	Urine	3	90% (P)
SEGELBACHER (2002)	<i>Tetrao urogallus</i>	Plumes	10	98,3% (P, #)

2.2.2.1 Révélation de fragments surnuméraires

Le premier type d’erreur peut être causé par la présence d’ADN **contaminant** ou par des fragments d’ADN issus d’**erreurs de duplication** de la polymérase (provoquée ou non par la dégradation de l’ADN). Ces copies d’ADN peuvent se trouver en nombre comparable à celui des copies d’ADN à amplifier quand ces derniers sont en très faible quantité. Si de telles

contaminations ou erreurs se produisent dans les premiers cycles de la PCR, les nombres de copies d'ADN exogène/modifié et d'ADN cible seront proches en fin de réaction. Les deux types d'ADN seront révélés produisant des résultats erronés. Pour les microsatellites, ces erreurs sont connues sous le terme de **faux allèles** (FA, *false allele* ; voir FIGURE 2.2) car les fragments d'ADN surnuméraires peuvent être interprétés comme des allèles réels, qui ne sont pas présents en réalité dans l'échantillon. Les faux allèles sont considérés comme provenant d'événements stochastiques tant qu'ils ne sont pas systématiques comme par exemple la contamination d'un lot de réactif influençant les résultats de toute la série de PCR réalisées à partir de ce lot (NAVIDI *et al.* 1992).

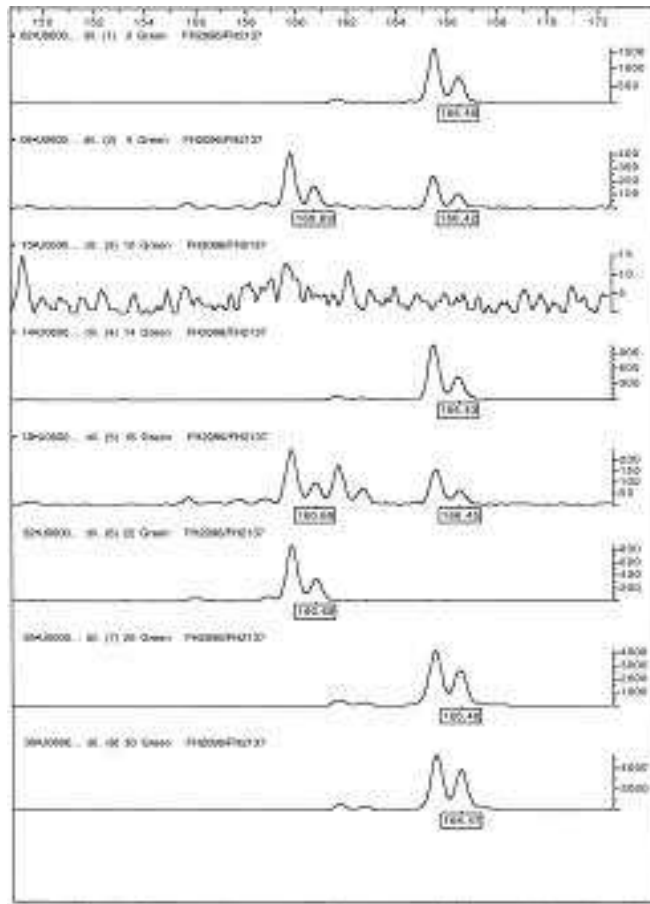


FIGURE 2.2 – Exemple de faux allèle (5^e piste) lors du typage multiple d'un locus pour un échantillon d'urine de canidé. Dans ce cas, le faux allèle correspond à la production d'un « allèle » de 2 paires de bases plus grand que celui attendu (+ le fragment associé, décalé d'une paire de bases). Dans ce cas précis, le faux allèle peut être très vite repéré car il apparaît chez un individu hétérozygote pour ce locus.

2.2.2.2 Non révélation de fragments

Le deuxième type d'erreur consiste à la non-révélation de fragments d'ADN contenu dans l'échantillon. Ces erreurs peuvent être dues à la dégradation de l'ADN (limitant l'hybridation des amorces sur l'ADN) et principalement à la **faible quantité d'ADN** dans l'extrait. En

effet, une explication de la **perte d'allèle** (ADO, *allelic dropout*; voir FIGURE 2.3) pour les microsatellites est que pendant le pipetage dans l'extrait et ensuite lors de l'amplification, un seul allèle est échantillonné, amplifié et révélé sur les deux présents chez un hétérozygote. Plus le nombre total initial de copies d'ADN sera faible, plus le risque d'une grande différence de fragments prélevés entre les deux allèles sera grand, et donc plus les pertes d'un allèle seront probables du fait de l'**échantillonnage aléatoire** qui sélectionnera l'allèle le plus abondant.

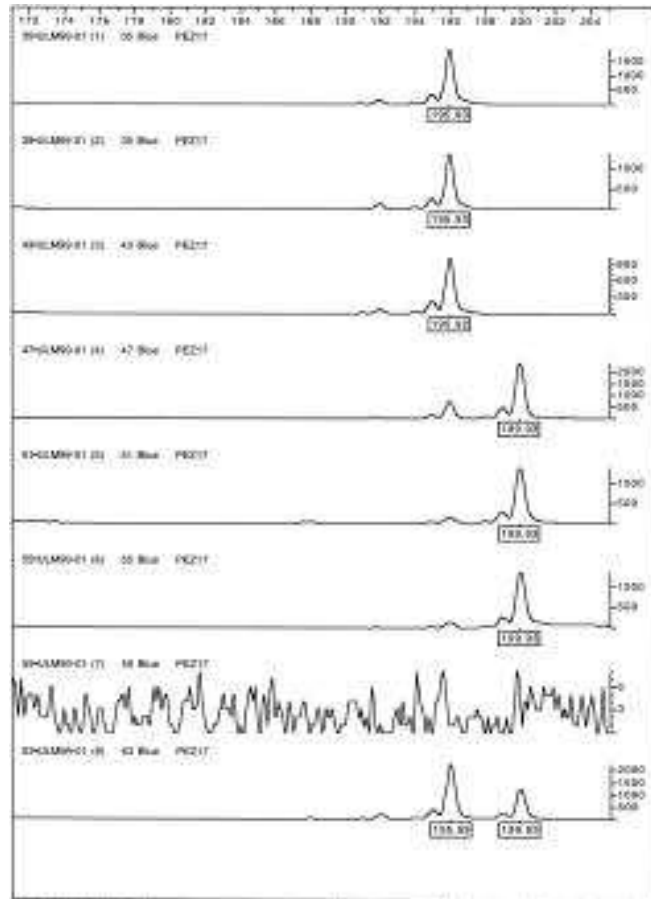


FIGURE 2.3 – Exemple de perte d'allèle (six premières pistes) lors du typage multiple d'un locus pour un échantillon d'urine de canidé. Une seule répétition sur les huit réalisées donne le bon génotype de l'échantillon pour ce locus.

D'autres explications ont été avancées comme l'amplification préférentielle des petits allèles (GERLOFF *et al.* 1995). En effet, il est observé que les petits allèles chez les hétérozygotes ont des pics d'aires plus importantes que les allèles de grandes taille (GILL *et al.* 1998, 2000a) et ceci peut être directement lié à un meilleur rendement d'amplification pour les petits fragments. Il semble en fait que les pertes d'allèle observées quand on utilise les microsatellites sur des extraits à faible quantité d'ADN soient un phénomène stochastique. En effet, aucun article n'a montré une différence significative entre la perte des grands ou des petits allèles. Un test de Chi2 sur les données de GAGNEUX *et al.* (1997) ou TABERLET *et al.* (1996) ne montre aucune différence de taux de perte d'allèles entre les différents allèles. Seules de faibles différences sont détectées

mais ne sont jamais statistiquement significatives (voir GERLOFF *et al.* 1995, CONSTABLE *et al.* 2001). Toutefois, il n'est pas complètement exclu que cette amplification préférentielle puisse contribuer à la perte d'un allèle. HU *et al.* (1992) ont montré qu'un allèle pouvait être préférentiellement amplifié en fonction des conditions de concentration en $MgCl_2$: des allèles peuvent être amplifiés à des concentrations de plus de 1mM de $MgCl_2$ et ne sont pas amplifiés en dessous de cette concentration. JEFFERY *et al.* (2001) ont aussi détecté des pertes d'allèles pour deux loci amplifiés à partir d'échantillons de bonne qualité (sang) chez *Melospiza melodia*. Ces deux loci ont été amplifiés avec un tampon NH_4 et non un tampon KCl comme les autres loci de l'étude ne montrant pas de perte d'allèle. Il est probable qu'il y ait une association des facteurs amenant à la perte d'allèle, le facteur majeur étant vraisemblablement la faible quantité d'ADN.

Notons enfin un autre type d'erreur de non-révélation d'allèle qui n'est pas dû à un faible quantité d'ADN dans les extrait. C'est le phénomène d'**allèles nuls**. Ce phénomène est dû à des mutations sur le site d'hybridation d'une des amorces de PCR ce qui prévient l'amplification (CALLEN *et al.* 1993). Ce type d'erreur est important à signaler car tout système de loci est susceptible de présenter ce phénomène (CALLEN *et al.* 1993, PAETKAU et STROBECK 1995, PEMBERTON *et al.* 1995). Contrairement au perte d'allèle, ce phénomène n'est pas stochastique et est donc systématique dans toutes les réactions réalisées pour l'extrait et le locus en question.

2.2.2.3 Autres erreurs

D'autres erreurs peuvent être commises comme une mauvaise attribution de la longueur d'un allèle. Ce type d'erreur peut être lié directement à l'analyse (automatique ou manuelle) des gels d'électrophorèse (WAITS et LEBERG 2000) ou à la faible amplification de l'ADN dans la mesure où les allèles sont souvent décalés sur l'électrophorégramme par rapport à leur véritable taille pour des extraits dont l'amplification est faible (voir FIGURE 2.4)⁷. Ce type d'erreur est très difficile à prendre en compte car difficilement quantifiable et modélisable.

Des artéfacts d'électrophorèse peuvent aussi être détectés, s'illustrant soit par une fausse identification d'un homozygote en hétérozygote soit par une assignation incorrecte d'un hétérozygote (FERNANDO *et al.* 2001). Ces erreurs se produisent quand une grande quantité de produits de PCR est chargée sur un séquenceur ABI PRISM 377. Elles peuvent être évitées en faisant passer dans le gel un courant de 3kV pendant une heure avant de charger les échantillons (cette opération modifiera les propriétés physiques du gel et éliminera ou réduira les artéfacts, FERNANDO *et al.* 2001).

2.2.2.4 Taux d'erreurs

Les taux d'erreurs (pertes d'allèle et faux allèles) observés dans les différentes études publiées varient en fonction de l'espèce, du type d'échantillon (voir TABLEAU 2.2). En moyenne, les valeurs de taux de perte d'allèle sont d'environ 5-10%, avec des valeurs d'environ 30% dans 3 cas et deux valeurs extrêmes de 50% et 75%. Les faux allèles apparaissent avec une fréquence généralement inférieure à 5% car ils sont facilement identifiable si les allèles présentent des tailles très différentes de ceux attendus.

⁷ces erreurs peuvent être évitées en se fixant des règles sur la sélection des pics, en fonction par exemple de la hauteur des pics sur l'électrophorégramme

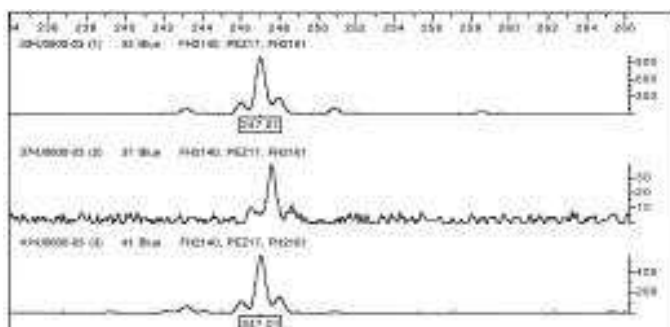


FIGURE 2.4 – Exemple de modification de la taille d'un allèle pour une faible amplification (deuxième piste) par rapport à la taille lors d'une « bonne » amplification (première et troisième piste).

TABLEAU 2.2 – Valeurs de taux d'erreurs (perte d'allèle, ADO et faux allèles, FA) pour diverses études publiées. Voir TABLEAU 2.1 pour la signification des annotations entre parenthèses.

Référence	Espèce	Type	N loci	ADO	FA
BAYES <i>et al.</i> (2000)	<i>Papio sp</i>	Fèces	8	8% (P)	1,2% (P)
CONSTABLE <i>et al.</i> (2001)	<i>Pan troglodytes</i>	Fèces	16	30% (S)	
MORIN <i>et al.</i> (2001)	<i>Pan troglodytes</i>	Fèces	9	24% (P)	
LAUNHARDT <i>et al.</i> (1998)	<i>Presbytis entellus</i>	Fèces	5	5% (S)	0% (S)
LATHUILLIÈRE <i>et al.</i> (2001)	<i>Macaca sylvanus</i>	Fèces	3	0-6% (P)	13-20% (P)
GOOSSENS <i>et al.</i> (2000)	<i>Pongo pygmaeus</i>	Fèces	1	4,2% (S)	3% (S)
ERNEST <i>et al.</i> (2000)	<i>Puma concolor</i>	Fèces	12	8% (P)	
KOHN <i>et al.</i> (1999)	<i>Canis latrans</i>	Fèces	3	2,6% (P)	2,6% (P)
LUCCHINI <i>et al.</i> (2002)	<i>Canis lupus</i>	Fèces	6	18% (P)	≈ 0% (P)
			1	0-33% (P)	
MIQUEL ET TABERLET (non pub.)	<i>Canis lupus</i>	Fèces	6	16,8% (P)	1,4% (P)
			1	1,6-37,0% (P)	0-7,1% (P)
TABERLET <i>et al.</i> (1996)	<i>Ursus arctos</i>	Fèces	1	53% (P)	<5% (P)
GARNIER <i>et al.</i> (2001)	<i>Diceros bicornis</i>	Fèces	10	30% (P)	≈ 0% (P)
PARSONS (2001)	<i>Tursiops truncatus</i>	Fèces	3	0% (P)	0,97% (P)
GAGNEUX <i>et al.</i> (1997)	<i>Pan troglodytes</i>	Poils	11	31,3% (P)	
GOOSSENS <i>et al.</i> (1998)	<i>Marmota marmota</i>	1 Poil	1	14,3% (P)	4% (P)
		3 Poils		4,9% (P)	1,4% (P)
		10 Poils		0,4% (P)	0% (P)
MOWAT et PAETKAU (2002)	<i>Martes Americana</i>	Poils	6	1,4% (S)	1,5% (S)
SLOANE <i>et al.</i> (2000)	<i>Lasiorhinus krefftii</i>	Poils	12	0% (P)	0,3% (P)
VALIÈRE et TABERLET (2000)	<i>Canis lupus</i>	Urine	3	4,2% (P)	0% (P)
			2 ech.	74,6% (P)	0% (P)
SEGELBACHER (2002)	<i>Tetrao urogallus</i>	Plumes	10	0,77% (P, #)	0% (P)

Chapitre 3

Perspectives de travail et Plan du mémoire

Sommaire

3.1	Au niveau de l'échantillonnage	40
3.2	Au niveau du génotypage des échantillons	41
3.3	Au niveau du traitement des données issues du génotypage mi- crosatellites	43

Nous avons donc vu que l'utilisation des méthodes non-invasives et des outils génétiques, notamment les microsatellites, peut être d'un apport considérable en biologie des populations et de la conservation, mais peuvent aussi présenter des limitations quant au taux d'erreur observé pour le génotypage. Il semble évident que de tels taux d'erreurs peuvent engendrer des biais énormes dans les analyses. WAITS et LEBERG (2000) ont montré que la taille de population estimée à partir des méthodes non-invasives pouvait être deux fois supérieure à la taille réelle si le taux d'erreurs est uniquement de 5%. Il est alors indispensable de prendre en compte et de corriger ces problèmes et des améliorations sont indispensables à tous les niveaux d'une analyse.

3.1 AU NIVEAU DE L'ÉCHANTILLONNAGE

L'échantillonnage sur le terrain est la première étape que l'on peut optimiser afin de maximiser les chances d'obtenir le plus d'information sur une population et pour diminuer les risques d'erreurs lors des analyses génétiques. Dans une première partie nous allons montrer combien l'échantillonnage non-invasif peut être important dans la description de la biologie d'une population. Les fèces et les poils sont déjà connus pour être des sources d'ADN et utilisés dans des études pour diverses espèces (voir par exemple GERLOFF *et al.* 1995, GAGNEUX *et al.* 1997). En France, la population de loup est actuellement étudiée depuis son retour naturel dans les Alpes. Les études consistent entre autres à récolter des fèces en vue de déterminer le régime alimentaire des loups (voir POULLE *et al.* 1997) où à repérer les zones de présence du loup grâce aux indices de présence (voir POULLE *et al.* 1999). Les fèces peuvent tout aussi bien servir à une étude génétique afin de décrire cette population et déterminer les mécanismes de colonisation de l'espèce dans un territoire vacant. Ce type d'échantillonnage est d'autant plus intéressant que les outils classiques d'étude de la répartition spatiale et temporelle des individus (*e.g.* capture, radiopistage) ne peuvent être utilisés en ce qui concerne le loup (décision politique et difficulté de capture notamment). Les distributions spatiale et temporelle des loups peuvent être analysées grâce à l'identification individuelle (par les microsatellites) mais les grandes lignes des processus de colonisation peuvent aussi être déduits de l'analyse de l'ADN mitochondrial permettant l'identification spécifique des échantillons. Un premier article (Article I) présenté fera référence à l'étude préliminaire de la colonisation des Alpes française et suisse par les loups *via* l'analyse de l'ADN mitochondrial à partir d'échantillons non-invasifs, principalement les fèces.

Dans le cas du loup en France, si les fèces peuvent être récoltées assez facilement, les poils sont extrêmement difficiles à trouver sur le terrain et aucun dispositif n'a été mis en place pour prélever des poils sur les individus comme cela a été fait pour les ours ou les mustélidés (voir précédemment). Il était ainsi intéressant de compléter le panel d'échantillons déjà disponibles en ne se limitant pas uniquement aux fèces. L'urine peut contenir de l'ADN qui peut être utilisé pour l'identification individuelle (*e.g.* MARKLUND *et al.* 1996, YOKOTA *et al.* 1998). En revanche, les études déjà menées l'ont été sur des prélèvements d'urine de chevaux après une course hippique ou des échantillons humains dans des études cliniques. Aucun essai ne semble avoir été mené sur des échantillons d'urine prélevés sur le terrain chez des animaux sauvages. Dans un second article (Article II), je présenterai donc la mise au point de l'extraction et de l'amplification d'ADN mitochondrial (ADNmt, région de contrôle) et nucléaire (microsatellites) à partir d'échantillons d'urine récoltés sur le terrain (principalement dans la neige mais aussi sur un tronc d'arbre), ses avantages et ses inconvénients.

L'utilisation d'échantillons non-invasifs nécessite certaines précautions. Les problèmes de

contamination et les précautions à prendre pour réduire ce risque, la préservation des échantillons sur le terrain et jusqu'aux analyses génétiques, mais aussi les problèmes de logistique et de mise en place des méthodes non-invasives seront abordés et discutés.

3.2 AU NIVEAU DU GÉNOTYPAGE DES ÉCHANTILLONS

Il existe de nombreuses solutions afin d'éviter ou limiter les erreurs de génotypage à tous les niveaux d'une analyse. Même en optimisant les manipulations (voir Section 5.3, p.62), on obtient des taux d'erreurs non négligeables. Les solutions proposées pour corriger « complètement » ces erreurs sont l'extraction multiple (répétition des extractions pour un échantillon ; *e.g.* GOOSSENS *et al.* 2000) et l'amplification multiple (répétition des amplifications PCR pour un extrait, NAVIDI *et al.* 1992). Ces deux solutions permettent de faire face à la variabilité (aléatoire) de la quantité d'ADN dans les échantillons (*e.g.* KOHN *et al.* 1995), à la variabilité du rendement d'extraction d'ADN pour un même échantillon, et à l'apparition d'erreurs stochastiques de génotypage lors de l'amplification PCR. D'autre part, il existe une certaine variabilité dans la quantité et/ou qualité de l'ADN entre individus (*e.g.* GERLOFF *et al.* 1995, GAGNEUX *et al.* 1997, FRANTZEN *et al.* 1998). Ainsi, il est souvent nécessaire de récolter plusieurs échantillons par individu quand cela est possible (voir par exemple GOOSSENS *et al.* 2000, GARNIER *et al.* 2001) afin de confirmer les résultats.

L'**approche multi-tubes** (ou approche à tubes multiples) , qui consiste à diviser un extrait en plusieurs parties et à analyser séparément ces aliquotes (NAVIDI *et al.* 1992), semble être la solution la plus utilisée pour faire face aux erreurs stochastiques. Dans cette approche, l'ensemble des résultats partiels forment un résultat consensus sur la base de l'apparition des allèles. Les erreurs stochastiques n'apparaîtront que partiellement dans les répétitions et seront ainsi éliminées.

Cette approche multi-tubes proposée pour diverses méthodes mettant en jeu un typage par PCR (NAVIDI *et al.* 1992) a été reprise et affinée par TABERLET *et al.* (1996) pour les microsattellites. Ces auteurs préconisent une approche basée sur deux règles :

- Règle 1 : on ne retient un allèle uniquement s'il apparaît deux fois au cours des amplifications.
- Règle 2 : pour identifier un homozygote, il faut que l'allèle soit identifié au moins 7 fois.

La première étape de la procédure consiste à réaliser trois amplifications. Si les résultats de ces amplifications sont ambigus, c'est-à-dire si les résultats ne répondent pas aux règles précédemment citées, quatre expériences supplémentaires devront être menées. Si aucune des deux règles n'est satisfaite sur la totalité des amplifications, des amplifications supplémentaires devront être réalisées (voir FIGURE 3.1).

En revanche, l'approche multi-tubes, bien que très efficace, peut s'avérer très coûteuse. En effet, dans la plupart des études, un nombre important d'échantillons sont récoltés (plusieurs centaines par exemple) et doivent être analysés. Ainsi le coût de génotypage microsatellite peut être multiplié par plusieurs facteurs et le coût de l'ensemble de l'analyse génétique peut devenir très vite exorbitant (voir TABLEAU 3.1). De plus, l'étude requerra plus de temps et plus de volume d'extrait, ce dernier pouvant être utilisé pour typer un plus grand nombre de loci (WAITS et LEBERG 2000).

Ainsi, il est absolument nécessaire de pouvoir tester la fiabilité des identifications individuelles et de prévoir le nombre minimum de répétitions d'amplification par échantillon pour obtenir des résultats d'un bon niveau de confiance avant d'entamer une étude génétique à grande échelle .

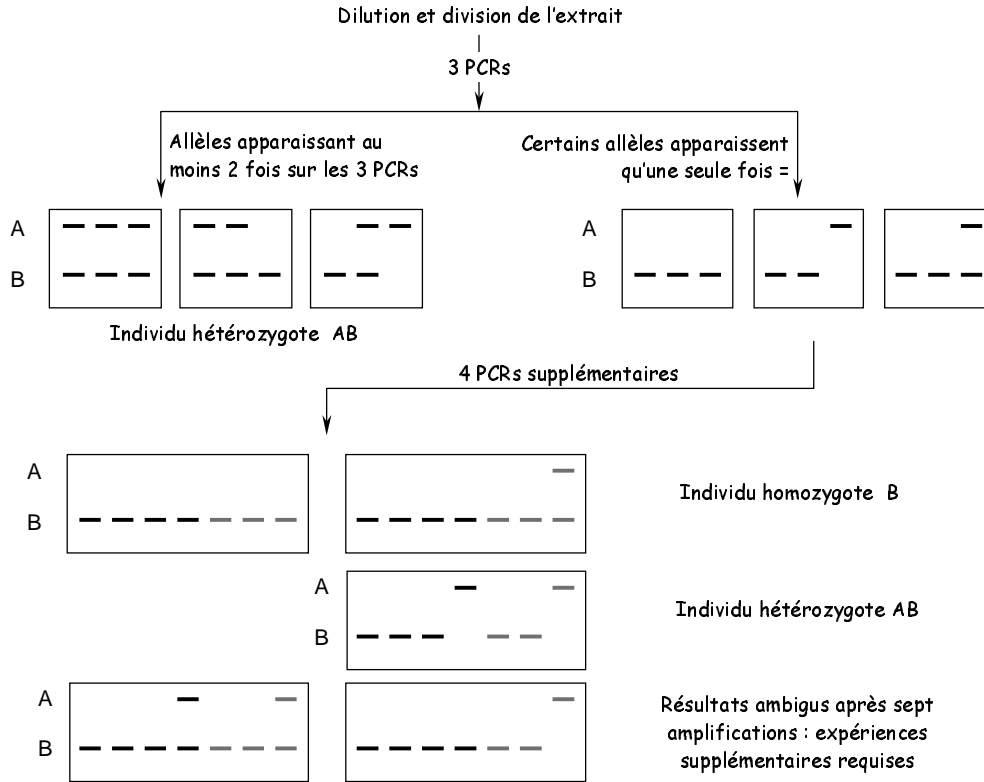


FIGURE 3.1 – Diagramme de la procédure à tubes multiples proposée par TABERLET *et al.* (1996) pour obtenir un niveau de confiance supérieur à 99% pour chaque locus étudié (Modifiée d'après TABERLET *et al.* 1996).

TABEAU 3.1 – Coûts approximatifs (en euros) de génotypage microsatellites (estimation du coût des consommables, hors fonctionnement) sans multiplex d'amplification mais avec 3 multiplex de charge.

	Simple tube		Tube multiple (8 répétitions)					
	1 échantillon	500 échantillons	1 locus	10 locus	1 échantillon	500 échantillons		
	1 locus	10 locus	1 locus	10 locus	1 locus	10 locus	1 locus	10 locus
Amplification	1,52	15,2	760	7600	12,16	121,6	6080	60800
Dépôt sur gel	1,52	4,56	760	2280	12,16	36,48	6080	18240
TOTAL	3,04	19,76	1520	9880	24,32	158,08	12160	79040

Le nombre de loci utilisés est aussi un facteur à prendre en compte pour la fiabilité des résultats. En effet, plus le nombre de loci est important plus la puissance d'identification est importante (WAITS *et al.* 2001) mais en revanche plus la probabilité d'erreur dans les génotypes multilocus est grande (WAITS et LEBERG 2000). Il faut donc trouver le nombre optimum de loci nécessaire à une étude.

Le troisième article (Article III) présentera ainsi le développement d'un logiciel informatique (GEMINI) qui réalise ces prédictions (nombre d'amplification et nombre de loci). Dans un quatrième article (Article IV), nous verrons que l'approche multitubes est très efficace pour corriger les erreurs de génotypage et je proposerai aussi une autre stratégie pour déterminer le nombre de répétitions de génotypage microsatellite que celle donnée par le logiciel. Enfin, dans un cinquième article (Article V), l'approche utilisant le logiciel sera appliquée à un jeu de données réelles pour vérifier la pertinence de notre approche et du logiciel.

3.3 AU NIVEAU DU TRAITEMENT DES DONNÉES ISSUES DU GÉNOTYPAGE MICROSATELLITES

L'association du génotypage de loci microsatellites, devenu une technique de routine dans de nombreux laboratoires, et de l'utilisation de nouvelles approches d'échantillonnage non-invasif vont permettre d'obtenir des données en grande quantité. De plus, l'utilisation de l'approche à tubes multiples produit de nombreux génotypes pour une étude de population.

La production accrue de données nécessite l'analyse et le traitement de ces dernières (LUIKART et ENGLAND 1999). Plusieurs tâches d'analyses sont nécessaires (FIGURE 3.2).

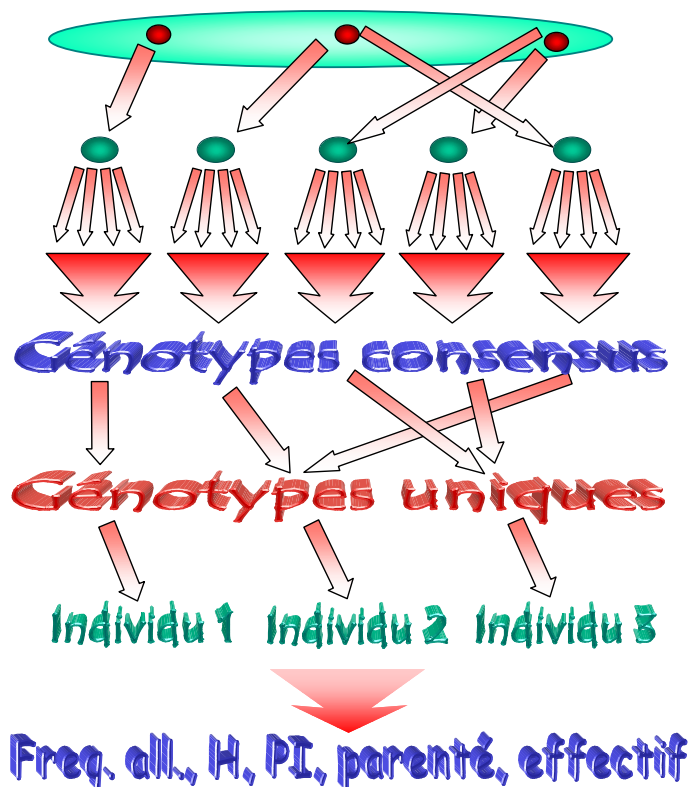


FIGURE 3.2 – Illustration des tâches d'analyse génétique d'une population : échantillonnage, génotypage multiple, construction des génotypes consensus, tri et identification des génotypes uniques, calcul des fréquences alléliques (Freq. all.), calcul des taux d'hétérozygotie (H), calcul des probabilités d'identité (PI), détermination des parentés possibles (parenté), estimation de la taille de population (effectif).

Il faut ainsi pouvoir construire les génotypes consensus à partir des répétitions pour chaque échantillon et chaque locus, estimer les erreurs de génotypage en comparant ces génotypes consensus avec les résultats de chaque répétition, trier et classer les génotypes pour regrouper ceux qui sont identiques, identifier des échantillons préalablement récoltés. La détermination des parentés potentielles entre les individus à partir de leur génotype et d'autres informations démographiques, le calcul et l'estimation de certains paramètres de populations (fréquences alléliques, hétérozygotie, probabilité d'identité, taille de population) peuvent aussi être réalisés à partir de ces données.

Dans ce contexte, j'ai développé un logiciel (GIMLET, Article VI) qui regroupe ces tâches principales d'analyse de génotypes microsatellites à partir d'un format de fichier « standard » en biologie des populations (format GENEPOP, RAYMOND et ROUSSET 1995).

Dans le cadre de l'estimation de la taille de populations à partir des échantillons non-invasifs, je présenterai une application de l'utilisation du logiciel GEMINI. En effet, plusieurs méthodes mathématiques peuvent être utilisées (modèles de capture-marquage-recapture ; courbe de raréfaction ; voir section 6.2.2.2.3, p.84) pour estimer la taille de population en identifiant chaque individu par leur génotype au cours d'une période de temps où une ou plusieurs occasions de capture sont réalisées. Dans le septième article (Article VII), j'étudierai donc les biais relatifs de chaque méthode en fonction de certains paramètres comme l'effort d'échantillonnage, de la structure de la population (influençant la probabilité de capture des individus) ou du taux d'erreurs de la détermination des génotypes.



Deuxième partie

Optimisation des méthodes non-invasives et du génotypage des microsatellites

Chapitre 4

Mise en place, optimisation et amélioration de l'échantillonnage non-invasif

Sommaire

4.1	Utilisation des échantillons non-invasifs dans l'étude préliminaire de l'expansion du loup en France et en Suisse par séquençage de l'ADN mitochondrial (Article I)	48
4.2	Mise au point et utilisation d'échantillons d'urine récoltés sur le terrain (Article II)	49
4.3	Discussion	49
4.3.1	Méthodologie et logistique	49
4.3.2	Contraintes de structuration et de spatialisation	50

L'échantillonnage est la première étape d'une étude de population. Elle est donc la première étape à prendre en compte lorsque l'on veut améliorer ou optimiser une étude complète.

Dans ce chapitre, nous verrons à partir d'un exemple, comment l'échantillonnage non-invasif peut être utilisé pour étudier certains aspects de la biologie d'une population. L'étude du loup en France constitue un bon exemple car, en France, l'échantillonnage ne peut se faire qu'indirectement (voir Section 3.1). De plus, la présence d'individus *Canis lupus* dans une région peut être certifiée par l'analyse génétique *via* l'ADN mitochondrial à partir d'échantillons non-invasifs. Le séquençage de régions de l'ADN mitochondrial permet d'identifier l'espèce et la population d'origine de l'individu associé à l'échantillon analysé.

Nous verrons aussi la mise au point et le test d'une nouvelle source non-invasive d'ADN (l'urine récoltée sur le terrain) qui peut venir compléter les autres sources déjà disponibles. Nous discuterons enfin de la mise en place concrète de l'échantillonnage non-invasif sur le terrain.

4.1 UTILISATION DES ÉCHANTILLONS NON-INVASIFS DANS L'ÉTUDE PRÉLIMINAIRE DE L'EXPANSION DU LOUP EN FRANCE ET EN SUISSE PAR SÉQUENÇAGE DE L'ADN MITOCHONDRIAL (ARTICLE I)

Article I

VALIERE N., FUMAGALLI L., GIELLY L., MIQUEL C., LEQUETTE B.,
POULLE M.-L., WEBER J.-M., ARLETTAZ R., TABERLET P.

Long distance wolf recolonization of France and Switzerland inferred
from non-invasive genetic sampling over a period of 10 years

Accepté dans Animal conservation.

Cet article présente les résultats d'une étude préliminaire de la colonisation des loups en France et en Suisse par l'identification spécifique en extrayant et séquençant l'ADN mitochondrial à partir d'échantillons de fèces, poils, tissus (individus morts) ou régurgitât. Même si l'étude de l'ADN mitochondrial peut être limitée (transmission maternelle ; voir Section 1.3.2.1, p.22), elle peut permettre de recueillir certaines informations, comme ici de certifier la présence d'une espèce dans une région et de suivre l'évolution de sa distribution spatiale et temporelle. De plus l'ADN mitochondrial est présent en plus grande quantité dans les cellules (environ 1000 copies par cellules par rapport à 2 copies par cellules). Les erreurs de génotypages sont susceptibles d'être moins nombreuses pour l'ADN mitochondrial que pour l'ADN nucléaire.

Notre étude a montré :

- Qu'il est possible d'identifier l'espèce (notamment de distinguer entre Renard roux *Vulpes vulpes*, Chien *Canis familiaris*, et Loup et la lignée génétique⁸ pour la grande majorité des échantillons non-invasifs.
- Que les échantillons de fèces sont ceux qui donnent les meilleurs résultats d'amplification.
- Que les échantillons français et suisses appartiennent à la même lignée génétique que les échantillons italiens. Cet haplotype commun est facilement discernable parmi les autres haplotypes de chiens de différentes races et de loups du monde entier.
- Que l'identification de l'espèce à partir de tels échantillons permet d'attester la présence du loup dans une région donnée et ainsi fournir des informations préliminaires sur l'expansion

⁸Une lignée génétique peut être définie sur la base d'un haplotype commun aux individus appartenant à cette lignée.

de la population de loup dans les Alpes. Nos résultats sont compatibles avec l'hypothèse d'un retour naturel des loups en France et en Suisse à partir de l'Italie. Nos résultats peuvent aussi montrer que les loups auraient colonisé les Alpes en deux étapes : une migration à longue distance de quelques loups avant l'installation de meutes.

4.2 MISE AU POINT ET UTILISATION D'ÉCHANTILLONS D'URINE RÉCOLTÉS SUR LE TERRAIN (ARTICLE II)

Article II

VALIERE N., TABERLET P.

Urine collected in the field as a source of DNA for species and individual identification.

Molecular Ecology (2000) 9 : 2150-2152.

Cet article présente la mise au point et le test de l'extraction et de l'amplification d'ADN mitochondrial et nucléaire à partir d'urine de canidés récoltée sur le terrain. Les échantillons proviennent du Parc National du Mercantour et sont principalement de la neige imprégnée d'urine de canidé (identification visuelle par les traces à proximité). Quelques morceaux d'écorce d'arbre sur lequel un Canidé est suspecté d'avoir uriné ont aussi été récoltés. L'extraction est réalisée grâce à un kit commercial (QIAAMP DNA STOOL, QIAGEN) après précipitation de l'ADN à l'acétate de sodium et à l'éthanol. L'amplification de l'ADN mitochondrial (D-Loop; amorces L15995 et H16498) et nucléaire (microsatellites; loci FH2140, FH2161, FH2050, FH2137, FH2096 et PEZ17) ont été réalisés afin de tester l'identification spécifique et individuelle à partir de ces échantillons.

Les principaux résultats sont :

- Qu'il est possible d'extraire et d'amplifier de l'ADN mitochondrial et de le séquencer afin d'identifier l'espèce (dans notre cas Chien ou Loup), à partir d'échantillons d'urine récoltés sur le terrain,
- Qu'il est possible d'extraire et d'amplifier de l'ADN nucléaire et d'identifier des individus à partir d'un génotypage microsatellite, à partir d'échantillons d'urine récoltés sur le terrain,
- Qu'il pouvait y avoir plusieurs individus identifiés dans le même échantillon d'urine,
- Que les taux d'erreurs dues à la faible quantité d'ADN (dégradé?) et/ou aux inhibiteurs de réactions enzymatiques (comme l'urée) sont comparables à ceux trouvés pour les fèces et semblent être très variables d'un échantillon à l'autre.

4.3 DISCUSSION

Lorsque l'on veut utiliser des méthodes non-invasives, il est primordial de prendre en considération plusieurs points avant d'engager un échantillonnage à grande échelle.

4.3.1 Méthodologie et logistique

Si l'un des avantages des méthodes non-invasives est la possibilité de ne pas nécessiter l'utilisation de dispositifs particuliers, il est des cas où des équipements spéciaux sont exigés. En effet, des dispositifs particuliers pour prélever les poils sont nécessaires (voir FIGURE 2.1, p.32). Ces dispositifs montrent une grande efficacité de prélèvement de poils mais exigent de pouvoir les placer en tous lieux de l'aire d'étude avec un minimum de temps et d'infrastructure. Dans

certains cas cités ci-dessus, l'utilisation d'un hélicoptère, de système GPS (*Global Positioning System*) et de techniciens expérimentés ont été nécessaires. Ces exigences sont de même nature que la logistique requise pour les captures et le piégeage d'animaux et les méthodes non-invasives peuvent devenir aussi coûteuses que ces méthodes, même si le coût propre des dispositifs peut être modique.

Dans ce type de protocole, il faut éviter l'utilisation d'appât afin d'attirer les individus vers le « piège ». En effet, l'appât en tant que « récompense » peut entraîner une réponse comportementale positive des individus, ce qui peut biaiser les estimations de taille de population par des modèles de capture-recapture (OTIS *et al.* 1978). Il faut noter que cette réponse comportementale est assez faible dans le cas de l'échantillonnage non-invasif de la martre d'Amérique (MOWAT et PAETKAU 2002). Il est préférable de réduire la réponse comportementale en n'utilisant que des appâts odorants pour lesquels les individus n'obtiennent pas de récompense lors du passage dans le dispositif (MOWAT et STROBECK 2001 ; L.Waits, comm. pers.) et de placer ces appâts hors d'atteinte des individus.

L'échantillonnage des poils fait apparaître une limitation liée à l'origine exacte de chaque poil. En effet, quand on prélève une touffe de poils sur les fils barbelés on prend le risque que ces poils proviennent de plusieurs individus. Ainsi il est bien souvent nécessaire de réduire au maximum le temps séparant les vérifications des pièges ou d'analyser un poil à la fois. Cette dernière démarche n'est pas souhaitable car cela peut engendrer de faible taux de succès de PCR et des erreurs de génotypages (voir GOOSSENS *et al.* 1998 ; Tableaux 2.1 et 2.2, pp.33,37). Les échantillons d'urine présentent également le même problème de mélange d'individus. Il faut donc récolter des échantillons pour lesquels on est sûr qu'un seul individu est représenté en fonction des traces laissés à proximité.

Certaines techniques nécessitent d'ailleurs l'observation des individus avant la récolte des échantillons pour s'assurer que les échantillons appartiennent bien à un seul individu (voir par exemple MORIN *et al.* 1993, 1994b, GAGNEUX *et al.* 1997, 1999 sur le chimpanzé ; GARNIER *et al.* 2001 chez le rhinoceros noir). Ceci est possible dans des populations qui tolèrent la présence de l'homme par exemple les deux communautés de chimpanzé suivies à Gombe en Tanzanie et à Taï en Côte d'Ivoire. Les suivis comportementaux de ces deux communautés ont débuté il y a 25 à 30 ans (BOESCH et BOESCH-ACHERMANN 2000) et les individus sont donc habitués à la présence de l'homme.

4.3.2 Contraintes de structuration et de spatialisation

Certaines espèces sont difficiles à observer et/ou capturer car le milieu dans lequel elles vivent est difficile à prospecter. S'il a été montré que les mammifères marins peuvent être échantillonnés de façon non-invasive (TIKEL *et al.* 1996, REED *et al.* 1997), ce milieu est de toute façon très difficile à échantillonner, notamment à cause des trois dimensions à appréhender mais aussi à la surface importante des territoires, par exemple, des mammifères marins (AMOS *et al.* 1993). Dans le cas particulier des mammifères marins, il peut être quelques fois nécessaire d'avoir recours à des méthodes non-invasives partielles pouvant consister au prélèvement à distance de morceaux de peaux ou de tissus (*e.g.* VALSECCHI *et al.* 2002).

Il semble donc évident que l'intensité et le type (aléatoire, systématique) de l'échantillonnage doivent être ajustés au type de terrain et à l'espèce étudiée. Dans le cadre de l'estimation de taille de population par capture-marquage-recapture (CMR), les contraintes sont identiques à celles que l'on peut trouver dans les techniques classiques de CMR (voir OTIS *et al.* 1978). MOWAT et STROBECK (2001), MOWAT et PAETKAU (2002) et POOLE *et al.* (2002) utilisent pour leur

échantillonnage de poils un quadrillage en cellules de l'aire d'étude et échantillonnent à l'intérieur de chaque cellule. La surface de chaque cellule doit être au moins aussi grande que le territoire moyen des individus pendant la période d'étude, ceci afin d'optimiser les chances de « capturer » un individu. Chaque cellule est systématiquement échantillonnée plusieurs fois, en espaçant les sites successifs d'une même cellule (1 km dans le cas des grizzlys) afin de réduire le biais par rapport à l'attraction des animaux par les appâts. Dans certains cas, la topologie du terrain oblige les expérimentateurs à adopter un échantillonnage curvi-linéaire car l'échantillonnage ne peut se faire que le long des voies praticables (voir par exemple KOHN *et al.* 1999 ou MOWAT et PAETKAU 2002). Cette contrainte est commune pour tout échantillonnage surtout dans des zones montagneuses et vallonnées.

Une des grandes contraintes vis à vis de l'échantillonnage est la présence d'une structuration spatiale forte des échantillons associée à une structuration sociale. Le dépôt de fèces ou d'urine peut ne pas être aléatoire (voir LUCCHINI *et al.* 2002) et l'optimisation de l'échantillonnage passe par une connaissance minimale des comportements de l'espèce à la fois pour optimiser la récolte d'échantillons mais également pour pouvoir interpréter convenablement la distribution spatiale et temporelle des indices. Les comportements de marquage (par exemple, défécation et urination) liés à la structure sociale peuvent influencer fortement la distribution des échantillons par rapport aux territoires des animaux et par la fréquence de dépôt des échantillons. Par exemple, GESE et RUFF (1997) remarquent une intensification des marquages par les dominants lors de la période de reproduction et à la périphérie du territoire chez le coyote. VILÀ *et al.* (1994) montrent que les loups marquent davantage le long des chemins, de surcroît peu fréquentés par l'homme, ou au niveau de monticules. Les dispositifs de capture de poils (FIGURE 2.1, p.32) sont aussi sujet à de tels biais en fonction, par exemple, de la hauteur des fils barbelés. Certains individus, dont la taille est inférieure à la hauteur des fils barbelés (donc en majorité des jeunes) passeront sous les fils sans être échantillonnés (WOODS *et al.* 1999, BOULANGER et MCLELLAN 2001). Une hétérogénéité apparaîtra ainsi au niveau de la probabilité de capture car certains individus seront moins représentés dans les prélèvements (LUCCHINI *et al.* 2002).

D'autre part, les régimes alimentaires ou le temps entre le « dépôt » de l'échantillon par l'animal et son prélèvement par les biologistes peuvent influencer le taux de réussite des analyses génétiques (GOOSSENS *et al.* 2000, MURPHY *et al.* 2000) et devra aussi être pris en compte car il peut varier en fonction des saisons (MAUDET *et al.* 0000, LUCCHINI *et al.* 2002).

Il convient donc de prendre en compte ces structurations et contraintes lors des analyses des données car elle pourront avoir une influence sur les résultats (voir Article VI sur la taille de population). Si aucune information n'est disponible sur l'espèce ou la population en question, il faudra garder à l'esprit que ces facteurs peuvent influencer les analyses.



Chapitre 5

Optimisation de l'identification individuelle par génotypage microsatellite

Sommaire

5.1 Le logiciel GEMINI (Article III)	54
5.1.1 Modèle utilisé dans GEMINI et paramètres	54
5.1.2 Analyser les effets des erreurs et optimiser l'approche multi-tubes	56
5.1.3 Analyse des données	58
5.2 Applications de GEMINI	59
5.2.1 Nombre de répétitions et nombre de loci à utiliser dans une étude : détermination et illustration par la simulation (Article IV)	59
5.2.2 Détermination du nombre minimum de répétitions du génotypage : applications aux données de GARNIER <i>et al.</i> (2001)	61
5.3 L'optimisation du génotypage à d'autres étapes de l'analyse	62
5.3.1 Conditions de récolte et de préservation des échantillons	62
5.3.2 Extraction d'ADN	64
5.3.3 Amplification	65
5.3.4 Précautions lors de l'extraction et de l'amplification	67
5.3.5 Choix des marqueurs	68
5.3.6 Analyse et interprétation des gels	69

Nous allons aborder dans ce chapitre un nouvel outil informatique dédié à l'optimisation de l'approche multi-tubes afin de sélectionner au mieux le nombre de répétitions d'amplification de PCR par échantillon et locus. Cet outil basé sur la simulation sera ensuite testé sur un jeu de données réelles afin de mettre à l'épreuve la pertinence du modèle. Enfin, nous discuterons des autres méthodes consacrées à la réduction et la limitation des erreurs d'identification individuelle par génotypage microsatellite.

5.1 LE LOGICIEL GEMINI (ARTICLE III)

Article III

VALIERE N., BERTHIER P., MOUCHIROUD D., PONTIER D.
GEMINI : software for testing the effects of genotyping errors and multitubes approach for the individual identification.
Molecular Ecology Notes (2002) 2 : 83-86

Lors du génotypage des échantillons, on peut décider de réaliser une approche à tubes multiples pour pouvoir corriger les erreurs stochastiques. Néanmoins, cette approche peut augmenter de façon considérable le coût de l'étude car il faut dupliquer chaque amplification pour chaque échantillon. Dans ce contexte, il ne s'agit pas de commencer une étude à grande échelle (plusieurs centaines d'échantillons à analyser) sans être sûr que les résultats obtenus seront assez fiables par rapport à la question que l'on se pose (TABERLET *et al.* 1999) ou de réaliser des opérations inutiles. Il est donc nécessaire d'optimiser le nombre de répétitions par échantillon et par locus, en trouvant un compromis entre coût et fiabilité des résultats.

La modélisation et la simulation peuvent être utilisées pour appréhender ce nombre optimal en faisant varier plusieurs paramètres (tels que les taux d'erreurs de génotypage, l'effort d'échantillonnage, le nombre de loci, *etc.*) dans un modèle. Elles peuvent aussi être utilisées pour tester l'effet des erreurs de génotypage, de l'effort d'échantillonnage et d'autres paramètres sur les résultats d'identification individuelle. L'approche utilisée dans le logiciel GEMINI (*Genotyping Errors and Multitubes approach for INdividual Identification*) est de simuler une analyse génétique d'une population (échantillonnage, génotypage, analyse) et de faire varier certains paramètres pour connaître l'impact de ces variations sur les résultats.

GEMINI est articulé en trois modules (voir FIGURE 5.1) : un module de simulation de chaque étape du modèle, un module de simulation de tout le modèle (*Batch*) et un module d'analyse des fichiers produits par GEMINI ou au format GENEPOP (RAYMOND et ROUSSET 1995).

5.1.1 Modèle utilisé dans GEMINI et paramètres

Le modèle utilisé dans GEMINI reprend la démarche d'une étude de population (FIGURE 5.2). La population est dite fermée : il n'y a pas de migration, pas de décès, ni de naissance au cours de la période d'étude de la population.

Les étapes du modèle sont les suivantes :

1. La création (aléatoire) des génotypes multilocus des individus d'une population à partir de fréquences alléliques fixées par l'utilisateur. Chaque allèle d'un individu est tiré aléatoirement parmi les allèles existant dans la population. La probabilité du choix d'un allèle est déterminée par les fréquences alléliques de cet allèle⁹.

⁹les fréquences alléliques dans la population simulée ne seront égales aux fréquences alléliques fixées que si la taille de population est infinie

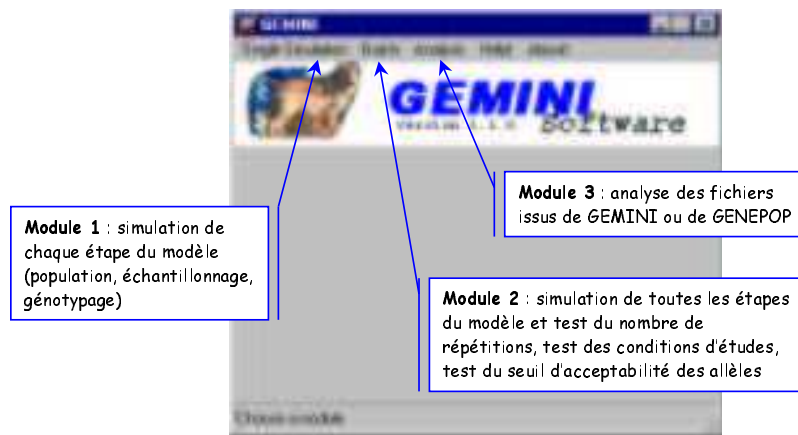


FIGURE 5.1 – Menu principal du logiciel GEMINI avec les trois modules de travail.

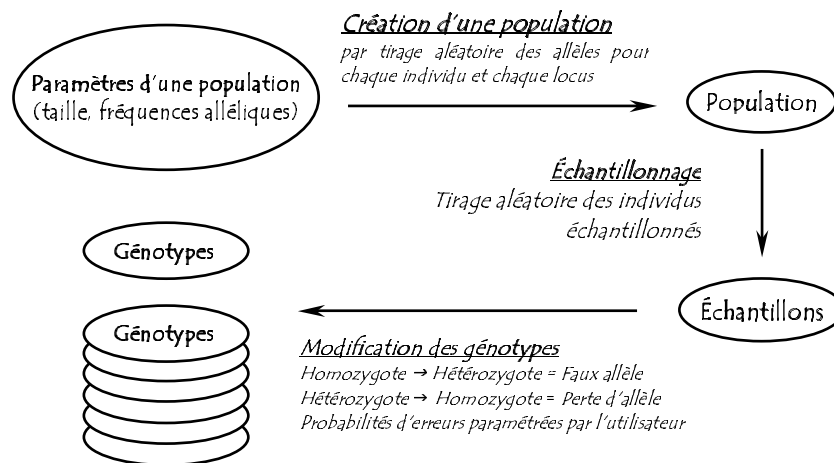


FIGURE 5.2 – Modèle de l'étude d'une population, utilisé dans le logiciel GEMINI .

- Le tirage d'échantillons à partir de la population par tirage aléatoire des individus. L'échantillonnage peut être réalisé avec remise (modèle non-invasif strict où un échantillon peut être récolté plusieurs fois au cours d'une occasion de capture) ou sans remise (un seul échantillon récolté par individu et par occasion). L'échantillonnage peut être composé de plusieurs occasions de capture.
- Le génotypage des échantillons avec l'introduction d'erreurs dans les génotypes de chaque échantillon suivant les taux d'erreurs fixés par l'utilisateur. Un hétérozygote sur un locus aura une certaine probabilité (définie par l'utilisateur) de devenir un homozygote par la perte d'un de ces allèles. Chacun des allèles d'un hétérozygote a une probabilité de $\frac{1}{2}$ d'être perdu s'il y a une perte d'allèle du fait qu'aucune preuve ne soutienne l'hypothèse d'une perte ou d'une amplification préférentielle. Un homozygote sur un locus aura une certaine probabilité de devenir un hétérozygote par la révélation d'un allèle surnuméraire. Pour simplifier le modèle, le faux allèle sera choisi parmi les allèles qui existent dans la population. Les allèles ont la même probabilité d'être choisis. Le génotypage peut être

simple (une amplification par échantillon) ou multiple (plusieurs répétitions par échantillon pour l'approche multi-tubes).

4. Le calcul de différents paramètres : fréquences alléliques, probabilité d'identité, hétérozygotie, pourcentage d'identification correcte, taille de population.

Afin de simuler une analyse fidèle à celle que l'on souhaite réaliser, il est nécessaire de pouvoir paramétrer certaines variables qui influenceront les résultats des simulations. Dans GEMINI, les paramètres suivants peuvent être fixés par l'utilisateur :

- nombre d'individus dans la population
- hétérozygotie observée dans la population
- fréquences alléliques
- nombre d'échantillons récoltés
- nombre d'occasions de capture
- type d'échantillonnage
- probabilité de capture des individus
- taux d'erreurs (perte d'allèle et faux allèles) pouvant être différent entre locus
- nombre de répétitions par échantillon pouvant être différent entre locus.

5.1.2 Analyser les effets des erreurs et optimiser l'approche multi-tubes

Le module *Batch* (voir FIGURE 5.3) est le plus important pour l'utilisateur.

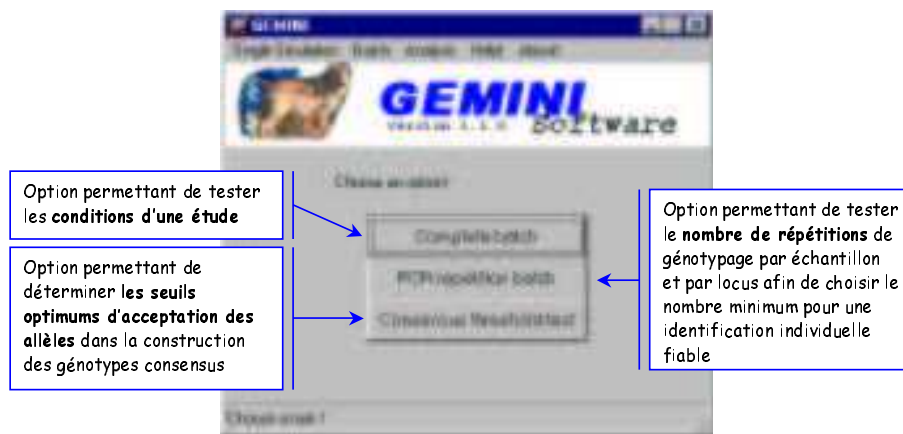


FIGURE 5.3 – Menu du Module Batch de GEMINI montrant les trois tâches de ce module : Simulations complète, Simulation pour plusieurs modalités de nombre de répétitions PCR et Test du seuil d'acceptation pour la construction des génotypes consensus.

En effet, il permet de :

- Réaliser des simulations afin de tester les conditions d'une étude. L'utilisateur fixe tous les paramètres cités ci-dessus et examine les résultats (*e.g.* pourcentage d'identifications correctes¹⁰, taille de population après génotypage comparée à celle simulée, fréquences alléliques après génotypage). Il peut ainsi décider de la stratégie à adopter en comparant plusieurs combinaisons de ces paramètres.

¹⁰c'est-à-dire les cas où le génotype déterminé est identique au génotype réel de l'échantillon

- *Tester l'approche multi-tubes et le nombre de géotypages par échantillon.* Dans ce cas, l'utilisateur fixe tous les paramètres comme ci-dessus à la différence que le nombre de répétitions de géotypage peut varier entre deux bornes. GEMINI ne calcule dans ce cas que le pourcentage d'identification correcte et affiche la relation de ce pourcentage en fonction du nombre de répétitions (voir FIGURE 5.4). L'utilisateur peut donc choisir le nombre minimum de répétitions à utiliser pour obtenir le maximum de fiabilité dans l'identification individuelle.

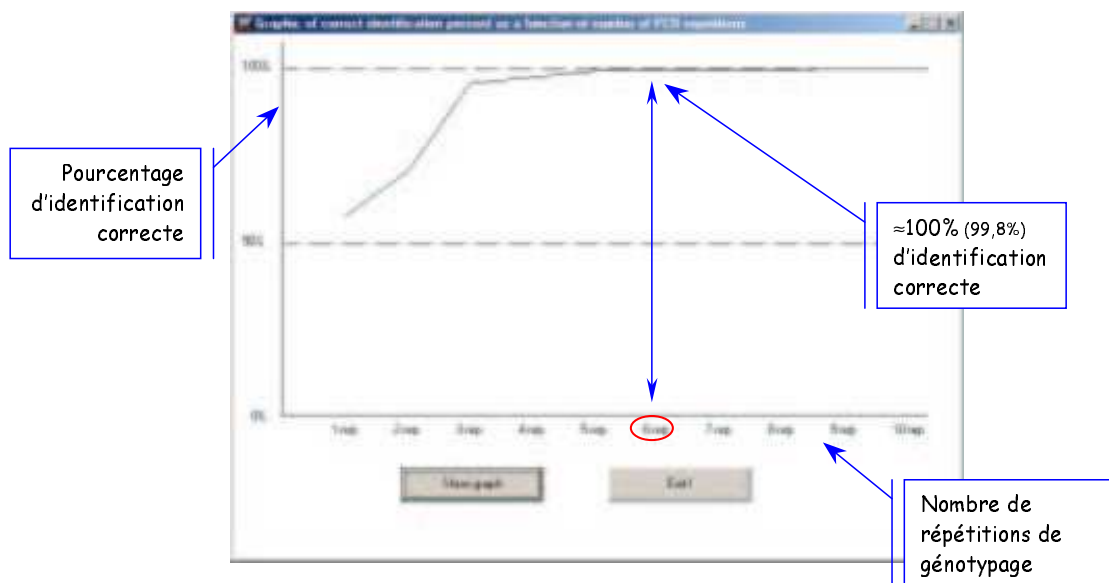


FIGURE 5.4 – Résultats donnés par GEMINI après une simulation pour plusieurs modalités du nombre de répétitions de géotypage par échantillon et par locus. On voit l'augmentation du pourcentage d'identification correcte (ordonnées) en fonction du nombre de géotypages par échantillon (abscisses).

- *Déterminer le(s) seuil(s) d'acceptabilité des allèles dans la construction des géotypes consensus.* Les géotypes consensus sont les géotypes construits à partir de plusieurs répétitions de géotypage d'un échantillon. Pour chaque locus, le géotype consensus est constitué par 1(es) allèle(s) (au plus deux) qui apparaissent un nombre de fois supérieur à un seuil d'acceptabilité (voir FIGURE 5.5).

Ce seuil doit être choisi consciencieusement car il déterminera l'exactitude du géotype consensus par rapport au vrai géotype de l'échantillon. Il est lié aux taux d'erreurs fixés dans les paramètres de géotypage. En effet, plus le taux de perte d'allèle sera élevé, plus le seuil devra être bas pour pouvoir retenir les allèles qui ont le plus souffert de la perte d'allèle. De même, plus le taux de faux allèles sera élevé, plus le seuil d'acceptabilité devra être élevé pour ne pas retenir d'éventuels faux allèles. Le programme teste chaque seuil possible et détermine le seuil optimum sur la base du pourcentage d'identification correcte. Le fichier de sortie contenant les seuils optimaux pour chaque condition sera alors utilisé dans la tâche décrite au point précédent afin que pour chaque modalité du nombre de répétitions de géotypage, on obtienne des résultats optimaux.

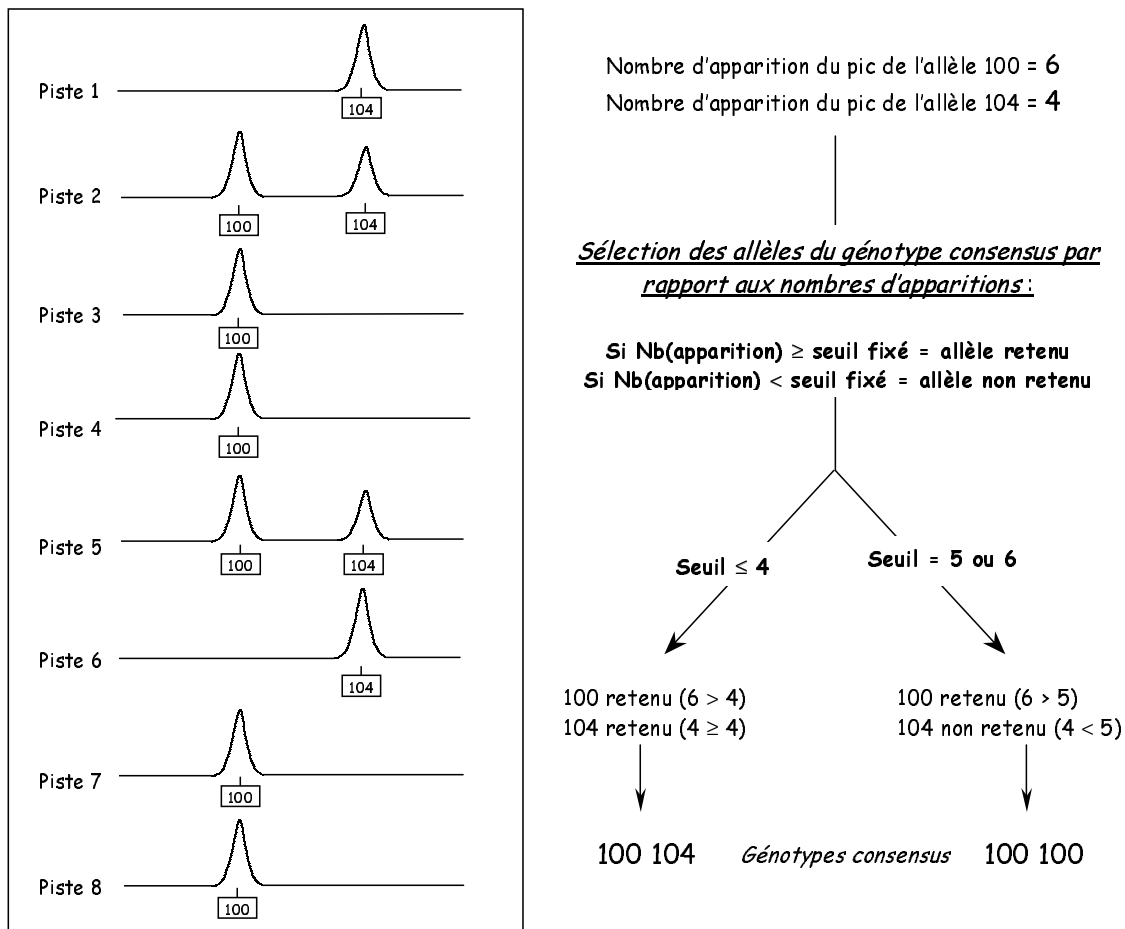


FIGURE 5.5 – Exemple de construction d'un génotype consensus en fonction du seuil d'acceptabilité des allèles. A gauche, sont représentés huit électrophorogrammes pour un échantillon et un locus. A droite, sont expliquées les règles d'acceptabilité des allèles.

5.1.3 Analyse des données

Pour étudier les effets des erreurs et/ou de la correction de ces erreurs par l'approche multitubes, sur l'analyse d'une population, GEMINI propose plusieurs analyses. Ces analyses sont réalisées d'office dans le module *Batch* et peuvent être réalisées à partir de fichiers de sortie GEMINI ou GENEPOP.

Les différentes analyses sont :

- le calcul des fréquences alléliques, des hétérozygoties et des probabilités d'identité
- la taille de population
- les pourcentages d'identifications correctes, erronées et nouvelles.

Les deux premiers points seront décrits dans la section consacrée au logiciel GIMLET (Section 6.2.2.2, p.77), car le module d'analyse n'est pas le point central de GEMINI. Cependant, le troisième point n'est pas traité par GIMLET et sera donc expliqué ci-après.

Pour quantifier l'impact des erreurs et de leur correction sur la fiabilité d'identification d'individu, il faut s'intéresser à la proportion d'échantillons qui ont été correctement attribués aux individus. Etant donné les génotypes des individus de la population, de ceux qui sont échan-

tillonnés et enfin le résultat du génotypage, il est très facile de calculer l'impact des erreurs sur l'identification des individus.

Les identifications correctes correspondent aux cas où le génotype après un ou plusieurs génotypage est identique au vrai génotype de l'individu échantillonné. Les identifications erronées correspondent aux cas où ces génotypes sont différents et où le génotype erroné peut être identifié comme un autre individu de la population. Les nouvelles identifications correspondent aussi à ce cas à la différence que le génotype ne correspond à aucun individu connu de la population. Ce génotype sera donc attribué à un individu « virtuel » n'existant pas dans la population.

5.2 APPLICATIONS DE GEMINI

5.2.1 Nombre de répétitions et nombre de loci à utiliser dans une étude : détermination et illustration par la simulation (Article IV)

Article IV

VALIERE N.

The use of multitubes approach for microsatellite identification :
number of genotyping repetitions and number of loci

Animal Conservation (PREVISION DE SOUMISSION)

Dans cet article, je présente des résultats de simulations effectuées avec GEMINI pour illustrer l'importance de l'approche multi-tubes afin de corriger les erreurs stochastiques de génotypage. La correction totale des erreurs de génotypage nécessitent plus ou moins de répétitions de PCR suivant le taux d'erreurs. Les résultats issus de GEMINI permettent facilement de choisir le nombre minimum de répétitions à réaliser pour avoir des identifications fiables en construisant le graphique entre le nombre de répétitions et le pourcentage d'identification correcte.

En plus du nombre de répétitions nécessaires à une identification fiable des échantillons, le nombre de loci étudiés est aussi très important. En effet, les simulations montrent que quand on utilise moins de loci, on diminue aussi le taux d'erreur dans les génotypes multilocus. Ceci est expliqué par le fait que la probabilité d'erreur sur un génotype multilocus est diminuée si ce génotype est construit avec un nombre restreint de locus. Une approche pour l'identification des échantillons est donc d'identifier un individu locus après locus et non sur tous les loci. Un échantillon est attribué à un individu si le nombre de loci communs entre l'échantillon et l'individu est plus important que pour les autres individus. On augmente ainsi la proportion d'identification correcte mais ces identifications se font aussi sur un nombre de loci moins important, car on retire par cette procédure les loci où une erreur s'est produite. Il faut donc prendre des précautions si on utilise cette approche.

Je propose également dans cet article une autre démarche qui peut être utilisée pour tester le nombre de répétitions. Cette démarche (FIGURE 5.6) consiste à réaliser le génotypage multiple de plusieurs échantillons grâce au système de loci à notre disposition. Le nombre de répétitions par échantillon et locus doit être suffisamment grand (par exemple 12) pour assurer la validité de l'approche. A partir des résultats de génotypage, il suffit de ré-échantillonner, pour chaque échantillon testé, les génotypes multiples en ne retenant que deux, trois, quatre, *etc.* génotypes à partir desquels nous allons construire un génotype consensus. En comparant ce génotype avec le génotype consensus construit à partir de toutes les répétitions, nous allons pouvoir évaluer le nombre minimum de répétitions par échantillon où la proportion de génotypes consensus identique aux génotypes issus de toutes les répétitions est maximum.

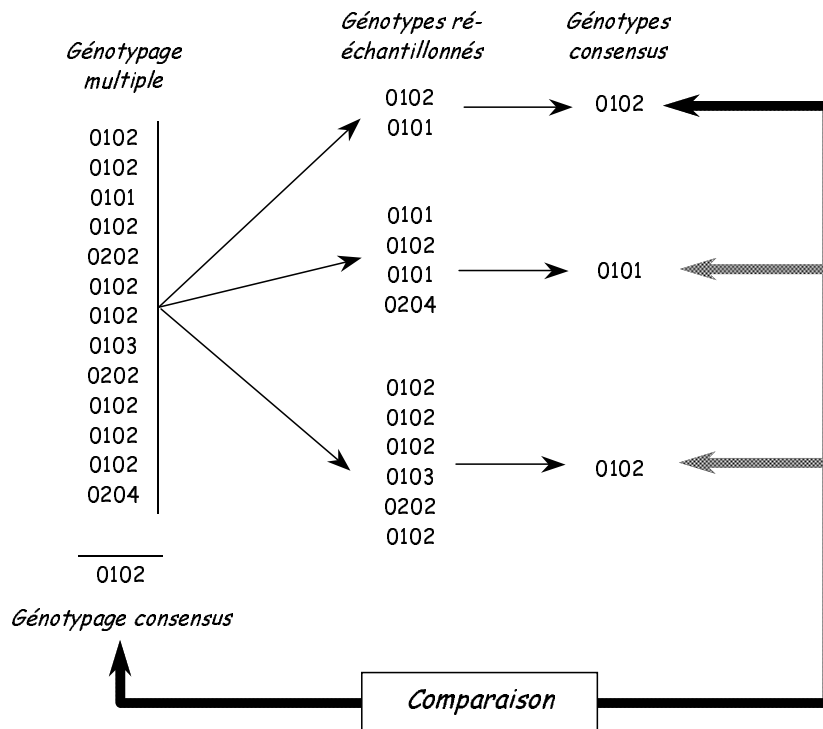


FIGURE 5.6 – Principe de la démarche pour tester le génotypage multiple pour un échantillon et un locus. Elle consiste à génotyper l'échantillon plusieurs fois, puis de ré-échantillonner ces génotypes pour construire des génotypes consensus. En comparant ces génotypes consensus avec le génotype consensus issu de toutes les répétitions, on peut évaluer l'approche multi-tubes.

Cette approche diffère de l'approche utilisant GEMINI dans le sens où elle est basée uniquement sur des résultats réels. L'approche de GEMINI n'utilise elle qu'une partie des résultats de l'étude préliminaire (fréquences alléliques, taux d'erreurs) pour réaliser des simulations. Néanmoins, l'approche décrite dans l'Article IV ne permet pas de tester les effets combinés des erreurs de génotypages, de l'échantillonnage sur les résultats (ce que peut faire GEMINI), mais permet uniquement de déterminer le nombre de génotypages requis pour chaque échantillon.

Ma méthode diffère également des méthodes proposées par MORIN *et al.* (2001) ou MILLER *et al.* (2002).

La méthode de MILLER *et al.* (2002) consiste à estimer la vraisemblance du génotype d'un échantillon après une première série de répétitions. Cette vraisemblance est calculée en décrivant un modèle pour les erreurs et en introduisant les fréquences alléliques, le taux d'hétérozygotie et le taux d'erreurs. Ensuite, en fonction de la vraisemblance du génotype, on pourra déterminer le nombre de répétitions qu'il sera nécessaire de refaire pour cet échantillon.

MORIN *et al.* (2001) proposent aussi une approche complètement différente sur l'estimation de la fiabilité de génotypage et la prévision du nombre de répétitions à effectuer. Cette approche est basée sur le dosage de l'ADN contenu dans les échantillons. Pour doser l'ADN dans des échantillons de fèces, il est nécessaire d'utiliser une technique particulière. En effet, l'ADN exploitable contenu dans un échantillon de fèces peut être de deux origines principales : l'ADN de l'hôte (celui que l'on veut doser) et celui provenant de bactéries du tractus digestif. Les techniques de dosage classiques (par une coloration dont l'intensité est proportionnelle à la quantité d'ADN contenu dans l'extrait) permettent seulement de doser l'ADN total. Il faut en fait pouvoir doser

uniquement l'ADN de l'hôte. Pour cela, on peut utiliser la réaction de PCR afin de cibler le dosage grâce à des amorces spécifiques de l'hôte. La méthode utilisée dans MORIN *et al.* (2001) est basée sur la PCR quantitative qui permet de quantifier la concentration de l'ADN amplifié tout au long d'une réaction PCR (voir JANSSEN 2001). Une fois les échantillons dosés, ces derniers sont classés en fonction de leur concentration d'ADN et un nombre de répétitions de PCR leur est attribué.

Les différences entre ces méthodes et mon approche et les points critiques sont les suivants :

- mon approche est basée sur l'ensemble des échantillons et non sur chacun d'entre eux. Les méthodes de MORIN *et al.* (2001) et MILLER *et al.* (2002) doivent être vues plus comme des méthodes de « suivi » de chaque échantillon (respectivement par dosage de l'ADN hôte et par le calcul de la vraisemblance du génotype obtenu pour l'échantillon) pour savoir combien de répétitions il est nécessaire de réaliser. Mon approche analyse un lot d'échantillons pour pouvoir déterminer le nombre moyen de répétitions nécessaires sur cet ensemble afin de l'appliquer à tous les échantillons à analyser. Cela peut être un inconvénient surtout si le taux d'erreurs est très variable entre les échantillons.
- mon approche n'est pas sensible à l'estimation des fréquences alléliques ou aux écarts de ces fréquences par rapport à l'équilibre de Hardy-Weinberg, et serait moins sensible à la variation du taux d'erreurs entre loci, par rapport à la méthode de MILLER *et al.* (2002) dans l'état actuel de la méthode (C.MILLER, comm. pers).
- la méthode de MILLER *et al.* (2002) nécessite de manipuler à plusieurs reprises les échantillons jusqu'à ce que la vraisemblance soit significative. Le fait de multiplier les manipulations (au même titre que l'approche de TABERLET *et al.* 1996, p.3.1), peut faire augmenter le risque de contamination ou d'échanges entre les échantillons.
- mon approche ne nécessite pas de matériel particulier comme thermocycleur particulier pour la PCR quantitative (voir JANSSEN 2001) comme dans l'approche de MORIN *et al.* (2001).
- les résultats de MORIN *et al.* (2001) ont été présentés sur la base de l'amplification d'une portion de 81 bp d'un gène nucléaire. Or cette taille de fragment n'est pas forcément celle des microsatellites (taille entre 100 et 250 bp par exemple). Il faudra faire très attention quant à la translation des règles présentées par MORIN *et al.* (2001) et il est préférable de refaire des tests sur chaque système de loci et d'échantillon à analyser.

5.2.2 Détermination du nombre minimum de répétitions du génotypage : applications aux données de GARNIER *et al.* (2001)

Article V

VALIERE N., PONTIER D., GARNIER J.N., GOOSSENS B.

Computer simulations to predict identification consistency using noninvasive sampling and the multitubes approach

Conservation Genetics (SOUMIS)

Les résultats des simulations conduites dans GEMINI sont-ils probants quand il s'agit de déterminer le nombre minimal de répétitions (NMR) à réaliser ? Ce point peut être résolu en illustrant l'utilisation de GEMINI sur un jeu de données réelles. Le jeu de données doit autoriser le test des résultats issus de GEMINI en permettant la détermination de l'effet réel des erreurs et de la correction de ces erreurs. J'ai utilisé les résultats de génotypage d'une communauté de 35 rhinocéros noirs (*Diceros bicornis*) au Zimbabwe (GARNIER *et al.* 2001). Ces auteurs ont réalisé un échantillonnage non-invasif et l'approche multi-tubes pour identifier génétiquement les

individus de la communauté. Ce jeu de données est composé de tous les résultats des génotypes pour chaque répétition et pour chaque échantillon (soit de 6 à 15 amplifications par échantillon et par locus) de crottes de rhinocéros sur 6 loci microsatellites.

A partir de ces données, les deux points suivants sont traités :

- La détermination de la relation réelle entre le nombre de répétitions de génotypage et la fiabilité de l'identification en appliquant un ré-échantillonnage des données brutes pour chaque modalité du nombre de répétitions,
- La réalisation des simulations avec GEMINI en utilisant des estimations de taux d'erreurs et des fréquences alléliques issus des données brutes, puis la comparaison avec les résultats du point précédent.

L'étude a permis de montrer :

- que la réalisation d'au moins quatre répétitions de génotypage par échantillon et par locus sont nécessaires pour pouvoir avoir un niveau de fiabilité d'identification de 100% dans cette étude. GARNIER *et al.* (2001) ont réalisé au moins six répétitions par échantillon et locus ce qui était donc largement suffisant pour identifier les individus et les paternités dans la communauté de rhinocéros.
- que le programme GEMINI permet de déterminer le NMR avec une bonne adéquation avec les résultats des données réelles. En effet, la relation simulée entre le nombre de réplifications de PCR et le pourcentage d'identifications correctes correspond à la relation issue des données réelles. De petites différences apparaissent quand on utilise uniquement une seule réaction de PCR. Ces résultats montrent que le modèle d'apparition des erreurs dans les génotypes multilocus utilisé dans GEMINI est assez proche de ce qu'il se passe en réalité. La pertinence du modèle utilisé dans GEMINI permet de prédire le nombre minimum de répétitions de PCR à réaliser afin d'obtenir des résultats fiables. Les écarts entre le modèle et les données réelles produisent toujours des prédictions donnant des NMR supérieurs ou égaux aux NMR déterminés à partir des données réelles.

Une approche est donc proposée afin de déterminer ce NMR. Une étude pilote doit être réalisée afin d'estimer les taux d'erreurs et les fréquences alléliques. Puis, à partir de ces données, des simulations sont réalisées afin de déterminer le NMR (*e.g.* plus petit nombre de répétitions correspondant à un pourcentage d'identification correcte supérieur à 99%) pour chaque locus et chaque échantillons.

5.3 L'OPTIMISATION DU GÉNOTYPAGE À D'AUTRES ÉTAPES DE L'ANALYSE

Les limites de l'utilisation d'échantillons non-invasifs liées à la quantité, la qualité et la présence d'inhibiteurs dans les extraits ne sont pas des obstacles infranchissables. S'il est possible de corriger les erreurs de génotypage, il est aussi nécessaire de prévenir les erreurs lors de certaines étapes d'une étude génétique.

5.3.1 Conditions de récolte et de préservation des échantillons

Le plus grand risque encouru lorsque l'on récolte des échantillons non-invasifs sont les contaminations. Ces contaminations peuvent provenir d'autres échantillons ou provenir de l'expérimentateur. Les conditions de récolte des échantillons non-invasifs sont donc extrêmement importantes et il est nécessaire de prendre le maximum de précautions (port de gants ou utilisation

d'« outils », changement des gants ou des outils entre chaque récolte). Pour la récolte de fèces, de simples morceaux de bois peuvent suffire à récolter l'échantillon sans le toucher. Néanmoins, il est possible d'utiliser un sac plastique comme un gant afin de prélever un morceau sans risque de contamination. Les fèces peuvent être transférées ensuite dans des pots en plastique pour faciliter la manipulation lors des analyses génétiques. Pour les poils, de petites pinces (type pinces à épiler ou pinces fines) peuvent être nécessaires afin de limiter les contaminations par l'expérimentateur ou par d'autres échantillons et simplifier la manipulation car les poils peuvent coller au gants en latex. Les pinces devront être brûlées entre chaque utilisation afin d'éviter les contaminations. Chaque poil devra être récolté et conservé séparément sauf si l'on est sûr que ces poils proviennent du même individu.

Le délai entre le dépôt de l'indice et la récolte est un point important pouvant influencer la qualité de l'ADN contenu dans les échantillons (MURPHY *et al.* 2000, LUCCHINI *et al.* 2002). Ce délai devra être le plus court possible pour limiter l'exposition de l'échantillon aux intempéries et aux autres facteurs pouvant influencer la qualité de l'ADN (*e.g.* humidité). Pour les fèces, on peut estimer le temps passé par l'échantillon grâce à son état (sec, entier ou morcelé ; LUCCHINI *et al.* 2002). On pourra envisager de ne pas traiter les échantillons qui seront considérés comme trop dégradés.

Une fois les échantillons récoltés, ils doivent être conservés et préservés. Le mode de préservation est très important car il va avoir une influence sur la qualité et la quantité d'ADN présent dans l'échantillon et donc sur les résultats des analyses génétiques (MURPHY *et al.* 2000). Différentes méthodes de préservation sont disponibles. Il a été montré que les billes de SILICA GEL (dessicateur ; BLAUGEL, Roth, GmbH ; environ 7 volumes de SILICA GEL pour un volume d'échantillon) sont très efficaces et très pratiques sur le terrain (WASSER *et al.* 1997, MURPHY *et al.* 2000), même si des solutions de tampons peuvent être aussi très efficaces (comme le DET : DMSO/EDTA/Tris/Salt ; FRANTZEN *et al.* 1998). En revanche, la manipulation d'une solution liquide peut s'avérer moins pratique sur le terrain. D'autre part, en utilisant un tampon, on augmente le risque de contaminations survenues lors de la préparation du tampon ou dues à la dégradation des poils (libération d'ADN) contenus dans les fèces (WASSER *et al.* 1997). Les échantillons peuvent être traités, dans le cas des fèces, à l'étuve ou au lyophilisateur afin d'éliminer toute trace d'humidité et donc de risque de moisissure et de présence d'ADN bactérien dans les échantillons (MURPHY *et al.* 2000)¹¹. En revanche, cette manipulation augmente les risques de contamination et peut éventuellement entraîner la dégradation de l'ADN dans le cas de l'étuvage (voir TABLEAU 2.1, p.33). Il est alors conseillé de traiter les échantillons pour une conservation à long terme qu'une fois ceux-ci analysés.

En ce qui concerne les poils, ils peuvent être simplement préservés à long terme dans une enveloppe de papier, également très pratique sur le terrain. Les enveloppes peuvent être conservées à température ambiante, éventuellement dans une boîte avec du SILICA GEL afin de limiter l'humidité de l'air. En revanche, dans certains cas, l'extraction d'ADN doit être effectuée le plus rapidement possible car les cellules des bulbes sèchent et peuvent se décrocher du poil ce qui limite l'application des analyses génétiques (TABERLET P., comm. pers.). Néanmoins, des poils peuvent aussi être conservés très longtemps (10 ans) en donnant de bon résultats lors de l'amplification de microsatellites (D.PONTIER et L.SAY, comm. pers.). Dans ce dernier cas, il est

¹¹Notons qu'il peut être envisagé, s'il est possible, de laisser les échantillons de crottes au congélateur -80°C pendant une semaine afin d'éliminer les risques de contamination par les larves d'*Echinococcus multilocularis* contenues dans certaines fèces de carnivores sauvages ou domestiques (F.RAOU, comm. pers.)

vraisemblablement nécessaire d'avoir obtenu une assez grande quantité de poils.

Dans le cadre de la récolte d'échantillons, il est impératif qu'une collaboration entre les généticiens et les personnes sur le terrain s'installe. Cette coopération est utile au niveau de trois points : (i) aider les généticiens à analyser le maximum d'échantillons, (ii) faire connaître les contraintes de terrain aux généticiens pour qu'ils n'exigent pas l'impossible, et (iii) faire connaître aux gens de terrain (souvent déconnectés du monde de la génétique) les exigences et les aboutissants des méthodes génétiques appliquées aux échantillons récoltés afin de les sensibiliser aux précautions qui doivent être prises lors de l'échantillonnage. Par exemple, il faudra choisir la meilleure période pour échantillonner. En effet, les conditions de terrain peuvent être particulières pour certaines saisons (hauteur de neige par exemple) et il est nécessaire de connaître ces conditions pour pouvoir échantillonner au mieux. D'autre part, le rendement d'extraction d'ADN peut varier suivant le régime alimentaire de l'animal étudié (M. MURPHY, comm. pers.) et donc de la saison (MAUDET *et al.* 0000). Ainsi, le géotypage microsatellite effectué sur des fèces d'hiver est significativement plus fiable que pour des fèces de printemps dans le cas d'ongulés de montagne. Il est donc très important de comparer et de tester ce qu'il est possible de faire sur le terrain et au laboratoire avant d'entamer une étude.

5.3.2 Extraction d'ADN

Le principe de l'extraction est de briser les membranes cellulaires, libérant l'ADN qui est ensuite récupéré. L'utilisation de kits développés à cet effet est de plus en plus répandue (DNA TISSUE KIT ou DNA STOOL KIT de QIAGEN, GmbH, Germany) car d'une part ils possèdent des protocoles bien définis et optimisés et d'autre part, ils ne nécessitent pas la préparation de solutions, minimisant notamment le risque de contamination (CONSTABLE *et al.* 2001). Il faut noter en revanche que les tubes du fournisseur peuvent contenir de l'ADN contaminant d'origine humaine (SCHMIDT *et al.* 1995) et il faudra faire des tests si l'on analyse des échantillons contenant une quantité extrêmement faible d'ADN (os fossiles ou tissus de musée). Le choix d'une méthode ou d'un kit adaptés au type d'échantillon est très important. Le kit dédié à l'extraction de l'ADN contenu dans les fèces fourni par QIAGEN donne des résultats bien meilleurs que l'utilisation du kit Tissu. Avec ce dernier on pouvait extraire de l'ADN à partir des fèces en modifiant légèrement le protocole (notamment en multipliant quatre fois les volumes des tampons dans les premières étapes du protocole). Le DNA STOOL KIT de QIAGEN permet de limiter la présence d'inhibiteurs des réactions enzymatiques dans l'extrait avec l'utilisation de pastilles INHIBITEXTM (résine) qui permettent de fixer les inhibiteurs. Les résultats obtenus sont assez convaincants (voir FIGURE 5.7), permettant l'amplification d'ADN à partir d'échantillons dont on ne pouvait amplifier de l'ADN avec le protocole « Tissu » et permettent l'amélioration importante de la qualité d'amplification pour d'autres échantillons.

Même si la méthode utilisant le CHELEX (résine chélateur, WALSH *et al.* 1991) est le protocole le plus utilisé pour les poils, les kits commerciaux DNA TISSUE KIT (QIAGEN) ou DNEASYTM BLOOD KIT (QIAGEN) peuvent être utilisés (MOWAT et PAETKAU 2002, POOLE *et al.* 2002; C.MAUDET, comm. pers.). On dit souvent que les techniques les plus simples sont parfois les meilleures. VIGILANT (1999) a montré que l'extraction d'ADN à partir de poils en utilisant simplement une digestion avec de la protéinase K dans un tampon de PCR (ALLEN *et al.* 1998) a un succès d'amplification (nombre de PCR positives) légèrement supérieur (mais non significatif) que pour le protocole du Chelex. Le protocole utilisé par VIGILANT (1999) permet l'extraction simple de l'ADN d'un poil et l'utilisation de 1/200ème de l'extrait pour une réaction PCR. Ceci autorise donc de nombreuses réactions PCR (géotypage par exemple) à partir d'un seul poil.



FIGURE 5.7 – Photo de gel d'agarose 1,6% montrant les résultats d'amplification de l'ADN mitochondrial à partir d'échantillons de fèces de loup. Les conditions de PCR sont identiques entre les différentes modalités d'extraction : DNA STOOL KIT (à gauche) et le DNA TISSUE KIT (à droite).

Une solution simple pour améliorer la quantité d'ADN extrait serait d'augmenter la quantité d'échantillon à partir duquel on va extraire l'ADN. En revanche, cette démarche comporte des limitations. Pour les fèces, augmenter la quantité de matière fécale dans les extractions signifie aussi augmenter la quantité d'inhibiteur de PCR dans l'extrait et ceci n'est pas intéressant même en utilisant le DNA STOOL KIT (QIAGEN). D'autre part, dans le cas des poils, augmenter leur nombre améliore beaucoup les résultats d'amplification (voir GOOSSENS *et al.* 1998) mais on augmente parallèlement les risques d'avoir des poils d'individus différents dans le cas où l'on a aucun renseignement sur l'origine des échantillons. Ainsi, l'augmentation du nombre de poils par extraction est faisable quand on sait si les poils proviennent du même individu, ce qui n'est pas toujours le cas pour les techniques d'échantillonnage non-invasif strict.

5.3.3 Amplification

Les conditions des PCR (températures d'hybridation, concentrations des réactifs) doivent être impérativement testées et optimisées afin d'obtenir le meilleur rendement d'amplification et de meilleurs profils microsatellites. Par exemple, la température d'hybridation doit être convenablement déterminée en partant de la température donnée par les auteurs ayant développé les amorces ou en la calculant à partir de la composition en bases des amorces. Dans le cas d'extrait avec de l'ADN dégradé, on peut diminuer légèrement la température d'hybridation afin de favoriser l'ancrage des amorces à cet ADN. En revanche, si la température est trop basse, cela peut aussi augmenter les risques d'amplification d'ADN qui ne serait pas la cible voulue puisque les amorces pourront s'hybrider avec de l'ADN dont la séquence ne sera pas totalement complémentaire des amorces (BIRCH *et al.* 1996).

Le nombre de cycles de PCR est aussi un paramètre extrêmement important. En effet, il faut généralement augmenter le nombre de cycles de PCR pour des échantillons de fèces ou de poils par rapport à des échantillons de tissu. En revanche, un trop grand nombre de cycles PCR ne donne pas toujours de meilleurs résultats et peut aussi augmenter le risque d'amplification de contaminants présents en très faible quantité. Si le nombre de cycles est trop important, la réaction de PCR n'est plus productive car la polymérase n'est plus en mesure de polymériser les copies d'ADN à cause de la perte de son activité. Il est donc impératif d'utiliser une polymérase qui garde sa fonction le plus longtemps possible. L'utilisation de l'AMPLITAQ GOLDTM (PERKIN ELMER)

peut être une solution. Cette polymérase est inactive à température ambiante et n'est active que quand la température atteint 95°C. Il a été montré (BIRCH *et al.* 1996, KEBELMANN-BETZING *et al.* 1998) que cette polymérase peut avoir une activité fonctionnelle durable, et qu'elle améliore ainsi la sensibilité (meilleure amplification même dans le cas de petite quantité d'ADN) et la spécificité (amplification de l'ADN cible), tout en étant simple à utiliser (on ajoute un cycle supplémentaire d'activation de la polymérase). L'utilisation du BUFFER GOLD (PERKIN ELMER) ne semble pas améliorer la sensibilité de cette polymérase : j'ai effectué des tests qui montrent une intensité moins importante des bandes amplifiées sur des échantillons de fèces avec deux loci microsatellites (voir FIGURE 5.8). Ceci peut être dû au fait que le BUFFER GOLD améliore la spécificité de l'hybridation ce qui ne constituerait pas une amélioration pour des extraits d'ADN dégradé déjà en faible quantité.

QIAGEN a développé une autre polymérase (HOTSTAR TAQ) associée avec un tampon (Q BUFFER) qui devait permettre de mieux amplifier des petites quantités d'ADN. Néanmoins, les résultats de tests que j'ai effectués montreraient qu'effectivement cette polymérase et ce tampon donnent de meilleurs résultats de sensibilité (voir FIGURE 5.8 et TABLEAU 5.1), mais que cela n'apporte que très peu d'amélioration au niveau des erreurs de génotypage sur plusieurs loci (voir TABLEAU 5.1). Ces résultats sont néanmoins à confirmer sur un nombre plus grand de replications.

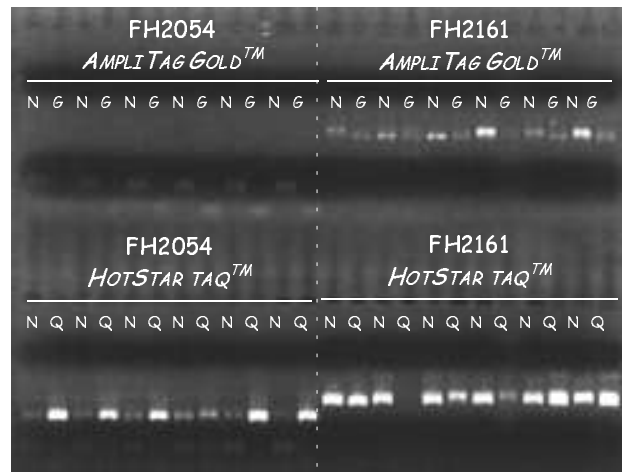


FIGURE 5.8 – Photo d'un gel d'agarose 1,6% comparant les intensités des bandes entre des réactions d'amplification (conditions identiques) d'ADN d'un extrait de fèces sur deux marqueurs microsatellites avec l'AMPLI TAG GOLD (en haut) et la HOTSTAR TAQ (en bas) . N : utilisation du buffer « standard » ; G : utilisation du BUFFER GOLD ; Q : utilisation du Q BUFFER.

Notons que l'utilisation de la polymérase T4 après la PCR (*e.g.* GINOT *et al.* 1996) peut permettre d'éliminer les ajouts d'une base à la fin des fragments par la *Taq* Polymerase. Cet ajout d'une base en fin de fragment dépend de la composition des amorces (BROWSTEIN *et al.* 1996) et peut engendrer des problèmes d'interprétation des gels et des profils microsatellites (créant des pics supplémentaires dans les profils). Ainsi, en éliminant ces « erreurs » on peut clarifier les profils ce qui peut être très utile quand on a des électrophorogrammes très peu intenses et où les

TABLEAU 5.1 – Comparaison d'amplification de microsatellites en utilisant différentes polymerases : la HOTSTAR TAQ de QIAGEN (55 cycles d'amplification) et l'AMPLITAQ GOLD de PERKIN ELMER (60 cycles d'amplification). Le tableau donne les taux de succès d'amplification (P+%) et de perte d'allèle (ADO%) pour un échantillon de fèces et 10 loci. "-" l'échantillon est homozygote au locus en question. "*" = données pour 45 cycles d'amplification

	HOTSTAR		AMPLITAQ	
	P+%	ADO%	P+%	ADO%
FH2140	100%*	100%*	100%	0%
PEZ17	88%*	0%*	100%	25%
FH2050	75%	33%	13%	100%
FH2137	100%	50%	100%	63%
FH2096	100%	0%	100%	0%
FH2199	0%	-	0%	-
FH2054	100%	-	38%	-
FH2247	100%	13%	0%	-
FH2289	88%	0%	0%	-
FH2356	50%	-	0%	-

profils sont brouillés.

5.3.4 Précautions lors de l'extraction et de l'amplification

Parce que les échantillons non-invasifs ne contiennent qu'une très faible quantité d'ADN, il faut faire très attention aux problèmes de contamination. Ces contaminations peuvent notamment être réalisées lors des étapes d'extraction et d'amplification. Plusieurs règles de conduite sont nécessaires pour éviter ces contaminations (KWOK et HIGUSHI 1989). Les plus simples sont de s'assurer de la stérilité et de la propreté de tous les ustensiles et matériel (pipettes, pinces, tubes *etc.*) en les nettoyant à la javel ou en les passant à l'étuve, mais aussi les paillasses sur lesquelles sont réalisées les manipulations (lavage à la javel ou exposition à la lumière UV, voir SARKAR et SOMMER 1990). Il faudra aussi utiliser des pointes à filtre pour les pipettes, des lames de scalpel à usage unique (une par échantillon). La précaution la plus contraignante est la séparation des pièces d'extraction d'ADN concentré (où seront traités les échantillons de tissus, de sang) et les pièces d'extraction d'ADN non concentré (où seront traités les échantillons de poils, de fèces). Cette contrainte doit être mise en place pour éviter toute contamination d'ADN concentré vers des échantillons d'ADN non concentré. De plus, pour la même raison, il est nécessaire de séparer les pièces de manipulations pré-PCR (ADN non concentré) et des pièces de manipulations post-PCR (ADN concentré car amplifié). Dans la salle post-PCR, peuvent être traités tout type d'échantillons d'ADN puisque tous seront « concentrés en ADN ». Pour minimiser les risques de contamination, des habitudes de comportement devront ainsi être prises avec une certaine discipline, par exemple ne pas travailler dans une pièce à ADN non concentré après avoir travaillé dans une pièce pour ADN concentré (voir FIGURE 5.9). Dans les cas extrême d'échantillons difficiles (os et tissus anciens), les pièces peuvent être équipées d'un système de suppression et de lampe à UV pour décontaminer l'ensemble de la pièce.

Enfin, les risques de contaminations ne peuvent être abaissés à zéro et il est nécessaire de contrôler les possibles contaminations au cours des manipulations. Ainsi, des « contrôles négatifs »

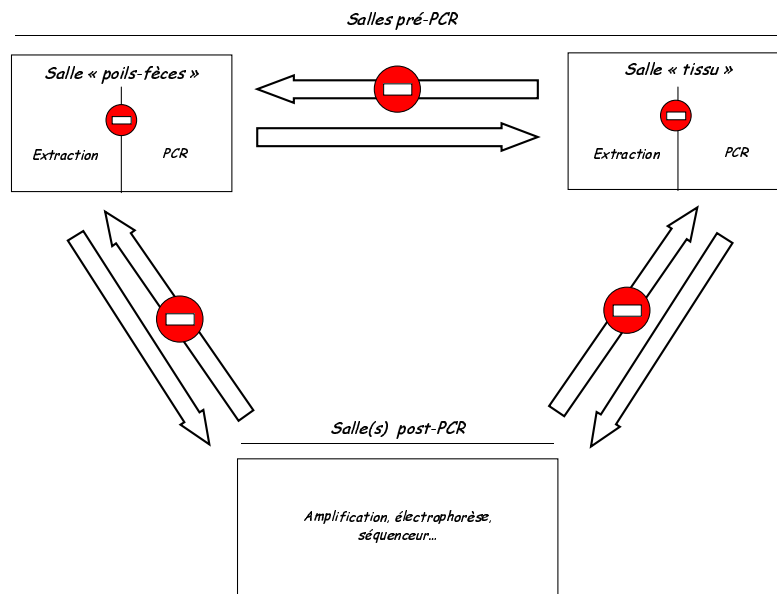


FIGURE 5.9 – Exemple de configuration des salles de manipulations et des échanges autorisés ou non entre ces salles. Une salle supplémentaire pour les échantillons « anciens » (échantillons fossiles ou de musée) peut être ajoutée séparément des autres

doivent être faits en réalisant les manipulations sans ajouter d'ADN dans les tubes. Ces tubes doivent être placés en dernier dans la série pour qu'il reflètent l'ensemble de la manipulation. Dans ces tubes, seuls les produits contaminants seront éventuellement amplifiés (preuve d'une contamination) et identifiés. Des « contrôles positifs » (pour les PCR) peuvent être réalisés en ajoutant, au lieu de l'ADN étudié, un volume d'extrait d'ADN dont on sait qu'il a déjà été amplifié auparavant afin de s'assurer que les tubes contiennent les réactifs adéquats.

5.3.5 Choix des marqueurs

Le nombre et le choix des locus microsatellites utilisés pour l'identification individuelle vont avoir une influence très forte sur les analyses. Tout d'abord, le choix de tri- ou tétra-nucléotides est préconisé car ce type de marqueurs microsatellites semblent présenter une meilleure fidélité d'amplification en ce qui concerne les bandes artéfacts (EDWARDS *et al.* 1991) et parce qu'ils permettent une meilleure définition des allèles (BRADLEY *et al.* 2000, ZHANG *et al.* 2001).

La sélection des marqueurs peut ensuite se baser sur la longueur des fragments amplifiés. Pour des échantillons dont l'ADN peut être dégradé et en faible quantité, il vaut mieux que le fragment à amplifier soit le plus court possible (de l'ordre de 200 paires de base). A partir d'amorces déjà développées il sera donc quelque fois nécessaire de redéfinir des amorces pour que les fragments amplifiés soient plus petits (voir BRADLEY *et al.* 2000).

De plus, on préconise souvent d'utiliser les marqueurs les plus informatifs, c'est-à-dire donnant les meilleures chances de discerner les individus. Ce choix peut se faire sur la base du polymorphisme et du calcul des probabilités d'identité (voir Section 6.2.2.2.2, p.82) en choisissant les loci qui présentent les plus faibles probabilités (c'est-à-dire qu'il pourront discerner au mieux les individus entre eux dans la population).

Enfin, il faudra aussi prendre en compte les taux d'erreurs associés à chacun des loci pour essayer de privilégier ceux qui sont les plus fiables. La réduction du nombre de loci peut permettre

de diminuer la charge et le coût de travail, de diminuer la quantité d'ADN requise permettant de faire plus de répétitions d'amplification par locus (TABERLET *et al.* 1999), et enfin de diminuer aussi le risque de faire une erreur sur un géotype multilocus (WAITS et LEBERG 2000 et Article VII).

5.3.6 Analyse et interprétation des gels

La détermination des allèles pour chaque locus de chacun des échantillons est aussi une étape importante où des erreurs peuvent être commises. L'utilisation de logiciels de traitement de telles tâches (par exemple GENESCAN et GENOTYPER de PE BIOSYSTEMS) facilite l'attribution des allèles. En revanche, il vaut mieux garder un certain contrôle sur l'automatisation de cette tâche. D'autre part des règles simples d'attribution et de détection des allèles par rapport à la forme des pics doivent être déterminées au début de chaque étude lorsque l'on teste les marqueurs. Ces règles devront bien évidemment être conservées tout au long des analyses. Il faut noter que certaines règles de décision ont été décrites, notamment quand les échantillons proviennent de mélange d'individus (GILL *et al.* 1997, CLAYTON *et al.* 1998, GILL *et al.* 1998, 2000b). Ces règles sont basées sur les aires ou la taille des pics des allèles, mais semblent, *a priori* difficiles à appliquer si la variation d'amplification entre les échantillons est grande. Il faut noter qu'une certaine souplesse peut être adoptée pour certains échantillons particuliers (par exemple quand ils ne contiennent que très peu d'ADN et que les pics n'ont pas la forme attendu).

L'adjonction, sur une piste d'un gel, d'un échantillon que l'on a déjà géotypé correctement (échantillon de tissu typé plusieurs fois par exemple) peut permettre de caler et d'éliminer les écarts de longueur des fragments analysés entre les gels.

Chapitre 6

Analyse des données des génotypes microsatellites

Sommaire

6.1	L'estimation de l'effectif d'une population et stratégies d'échantillonnage	72
6.2	Traitement des données microsatellites : le logiciel GIMLET . . .	74
6.2.1	Description des tâches et comparaison avec d'autres logiciels d'analyse . .	74
6.2.2	Description de modules	76
6.2.2.1	Le module « <i>Identification</i> »	76
6.2.2.2	Le module « <i>Calculator</i> »	77

6.1 L'ESTIMATION DE L'EFFECTIF D'UNE POPULATION ET STRATÉGIES D'ÉCHANTILLONNAGE

Article VI

VALIERE N.

Population size estimation and biases associated with genotyping errors
EN PREPARATION

En génétique de la conservation, l'un des paramètres primordiaux et peut-être l'un des plus délicats à acquérir est l'effectif d'une population. L'effectif d'une population est un indice très intéressant car il permet d'appréhender la productivité et la viabilité des populations (WAITS et LEBERG 1999). Il existe classiquement plusieurs solutions pour estimer la taille de population (SCHWARTZ et SEBER 1999) et les méthodes de capture-marquage-recapture (CMR) en sont une parmi d'autres. Il existe une multitude de modèles de CMR et d'applications se différenciant notamment par la façon dont la variabilité de probabilité de capture est modélisée ou par le type de population (*ouverte* ou *fermée* ; voir POLLOCK *et al.* 1990).

Les méthodes génétiques et les méthodes d'échantillonnage non-invasif peuvent être appliquées à ce type de modèles. En effet, la capture des individus peut être remplacée par la récolte d'échantillons et l'identification visuelle (marques) des individus par la caractérisation génétique (*e.g.* microsatellites) des échantillons (voir par exemple FORAN *et al.* 1997a, PALSBOELL *et al.* 1997, KOHN *et al.* 1999, WOODS *et al.* 1999, MOWAT et STROBECK 2001). En revanche, ces méthodes peuvent avoir des limites, les modèles pouvant ne pas pas s'ajuster aux données. Néanmoins, on peut essayer d'utiliser les modèles déjà existants sur les données basées sur l'identification génétique car les sources de variabilité peuvent être considérées comme équivalentes. Par exemple, la variabilité de probabilité de capture avec le temps ou entre individus (mâle/femelle ; dominant/dominé) sont tout aussi bien présents dans les deux types d'approche. Néanmoins, la variabilité associée aux classes d'âge ne pourra être corrigée avec l'approche génétique (puisque aucune information sur l'âge de l'animal à partir d'un échantillon non-invasif ne peut être obtenue, sauf si une technique spécifique est disponible, par exemple la présence d'hormones et leur dosage dans les fèces).

Si pour les méthodes classiques de CMR l'identification individuelle est sans erreur (les marques sont choisis dans ce sens), dans l'approche génétique, il est possible que plusieurs individus puissent avoir un même génotype et être donc considérés comme le même individu ; ceci peut alors entraîner une sous-estimation de la taille de population (*shadow effect*, MILLS *et al.* 2000). WAITS et LEBERG (2000) ont aussi décrit les biais provoqués par les erreurs de génotypage dans l'estimation de la taille de population. Leurs simulations montrent que le biais d'estimation peut atteindre +200% dans le cas où le taux d'erreurs par locus est de 5% pour 7 à 10 loci.

Outre les modèles de CMR, il est nécessaire de pouvoir tester d'autres méthodes d'estimation de taille de population. KOHN *et al.* (1999) proposent une méthode basée sur la courbe de raréfaction de génotypes uniques en fonction du nombre d'échantillons typés (voir FIGURE 6.1).

Les données observées sont ajustées à la courbe dont l'équation est :

$$y = \frac{ax}{b + x} \quad (6.1)$$

où y est le nombre de génotypes uniques, x le nombre d'échantillons typés.

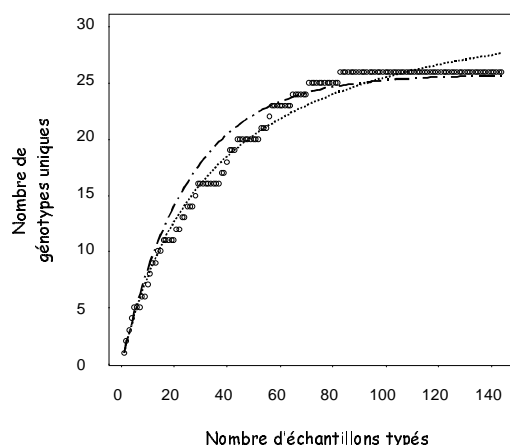


FIGURE 6.1 – Exemple d'une relation observée entre le nombre de génotypes uniques et le nombre d'échantillons typés (\circ) et des courbes d'ajustement aux deux équations citées dans le texte (Equation de Kohn=... ; Equation de Chessel=-.-).

L'estimation du paramètre a fournit une estimation de la taille de population car a est la valeur asymptotique de y .

Une autre équation proposée par D.Chessel (non publiée) peut être utilisée :

$$y = a - a \left(1 - \frac{1}{a}\right)^x \quad (6.2)$$

Elle correspond à l'espérance du nombre y de boîtes pleines (génotypes uniques) quand on distribue x balles (échantillons) dans a boîtes (individus). Contrairement à la première relation, cette deuxième relation ne comporte pas d'asymptote mais un plateau réel et serait apparemment plus proche du modèle d'estimation du nombre d'individus de la population¹².

J'ai utilisé GEMINI afin de réaliser des simulations pour compléter l'étude de WAITS et LEBERG (2000) et déterminer dans quelles conditions différentes méthodes d'estimation de la taille de population pouvaient être utilisées. Deux modèles pour chacune des approches ont été testés : les modèles Mo (CAPT-Mo) et Mh (CAPT-Mh) pour l'approche de CMR, et les modèles utilisant les équations de KOHN *et al.* (1999) (RAREF-Kohn) et de D.Chessel (RAREF-Chessel) pour la méthode de raréfaction.

Les principaux résultats des simulations montrent que :

- Les différentes méthodes d'estimation utilisées réagissent très différemment à la fois à l'introduction d'une hétérogénéité des probabilités de capture et d'erreurs de génotypage.
- en absence d'hétérogénéité de probabilité de capture et d'erreurs de génotypage, les méthodes CAPT-Mo et RAREF-Chessel montrent le moins de biais (<20%).
- en présence d'hétérogénéité des probabilités de capture, les méthodes se comportent différemment suivant l'intensité d'échantillonnage. Ainsi, RAREF-Kohn montre le plus faible biais des quatre méthodes pour une intensité d'échantillonnage faible. En revanche, pour

¹²il se peut aussi que dans certains cas, la taille estimée par ajustement soit inférieure au nombre total de génotypes

une intensité d'échantillonnage forte, les biais sont relativement équivalents et compris entre -20% et +25% (biais relatif).

- en présence d'erreurs de géotypage, la méthode RAREF-Kohn se comporte très mal en atteignant des biais relatifs jusqu'à plus de 5000% pour des taux d'erreurs de ADO=30% et FA=5% (taux d'erreurs réalistes, voir TABLEAU 2.2, p.37). La méthode qui produit le moins de biais est la méthode CAPT-Mh, tout en montrant des biais relatifs de plus de 280% pour le même couple de taux d'erreurs. Les autres méthodes se comportent sensiblement de la même façon que CAPT-Mh pour les taux d'erreur intermédiaires (entre 1% et 10%).

6.2 TRAITEMENT DES DONNÉES MICROSATELLITES : LE LOGICIEL GIMLET

Article VII

VALIERE N.

GIMLET : a computer program for analysing genetic individual identification data.

Molecular Ecology Notes 2 : 377-379

Les données microsatellites sont de plus en plus utilisées en biologie des populations et de la conservation. Quand les microsatellites sont associés aux méthodes d'échantillonnage non-invasif, l'analyse de données est particulière. Il faut en effet, avant tout, construire les géotypes consensus, puis trier et regrouper ces géotypes pour éviter les redondances des individus (voir FIGURE 3.2, p.43). Dans ce contexte, j'ai développé un logiciel, GIMLET. La philosophie générale de ce logiciel est différente de celle de GEMINI car il permet d'analyser des données réelles (géotypes microsatellites issus d'échantillons non-invasifs), et non de simuler ces données pour déterminer une stratégie.

6.2.1 Description des tâches et comparaison avec d'autres logiciels d'analyse

L'originalité de GIMLET par rapport à d'autres logiciels déjà publiés est qu'il permet de réaliser toutes les tâches avec le même outil et à partir d'un format de fichier (format GENEPOP, RAYMOND et ROUSSET 1995) unique, largement utilisé en génétique des populations (G.LUIKART, comm. pers.).

Le logiciel permet de trier et/ou d'identifier les géotypes des échantillons, ce que, à ma connaissance, seuls les logiciels IDENTITY (W.AMOS, non publié) et CERVUS (MARSHALL *et al.* 1998) réalisent, à la différence que GIMLET permet la sauvegarde des géotypes uniques dans des fichiers au format GENEPOP afin de réaliser facilement des analyses ultérieures. GIMLET reprend également les tâches classiques de calculs de fréquences alléliques, du nombre d'allèles par locus et d'hétérozygotie pouvant être réalisées dans GENEPOP (RAYMOND et ROUSSET 1995) en y ajoutant néanmoins le calcul des probabilités d'identité.

Ensuite, GIMLET permet de déterminer les parentés en se basant sur la transmission mendélienne des allèles des parents vers leur progéniture, mais aussi sur des données de traits d'histoire de vie (*e.g.* âge, sexe, maturité sexuelle). Cette démarche n'est pas développée dans des logiciels car pour un nombre limité de loci, la compatibilité des géotypes n'est pas suffisante pour rendre compte de toutes les parentés ou paternités (CHAKRABORTY *et al.* 1988) ou d'exclure tous les individus sauf un (MARSHALL *et al.* 1998). Les méthodes traditionnellement utilisées

(afin de remplacer les méthodes d'exclusion de paternité) sont des méthodes de calculs de vraisemblance (MARSHALL *et al.* 1998) ou d'une probabilité *a posteriori* de paternité (DANZMANN 1997, NIELSEN *et al.* 2001) suivant des modèles mathématiques.

Le développement du module *Kinship* de GIMLET a été motivé par l'absence d'algorithme simple basé sur la ségrégation mendélienne des allèles, ainsi que par l'absence de programme permettant de discriminer les parentés sur la base de données démographiques et de traits d'histoire de vie. Pour la partie génétique, le module *Kinship* de GIMLET doit être utilisé avec précaution et est plutôt conseillé dans les cas où un grand nombre de loci polymorphes sont testés sur un nombre limité d'individus issus d'une population largement échantillonnée. Ces conditions sont certes limitées par rapport à l'utilisation d'autres logiciels, comme CERVUS (MARSHALL *et al.* 1998), qui peuvent être appliqués à des jeux de données importants. Ce dernier logiciel ne permet pas, par contre, de prendre en compte directement les données démographiques et de traits d'histoire de vie comme peut le faire le module *Kinship* de GIMLET. Ainsi, GIMLET traite ces données (date de naissance, de mort, temps de gestation, âge de première reproduction, sexe) afin de confirmer ou d'infirmer les parentés génétiques (par exemple un parent doit être sexuellement mature et vivant à la date de conception de la progéniture). Pour cela, GIMLET adopte un format particulier du cycle de vie (voir FIGURE 6.2) permettant de prendre en compte plusieurs saisons de reproduction au cours d'une année et la possibilité de prendre en compte les mois dans les âges des individus. Ce format diffère de celui utilisé par PARENTÉ (CERCUEIL *et al.* Soumis) qui réalise aussi l'assignation génétique par compatibilité mendélienne mais qui ne considère (dans l'avancement actuel du programme) que des années entières.

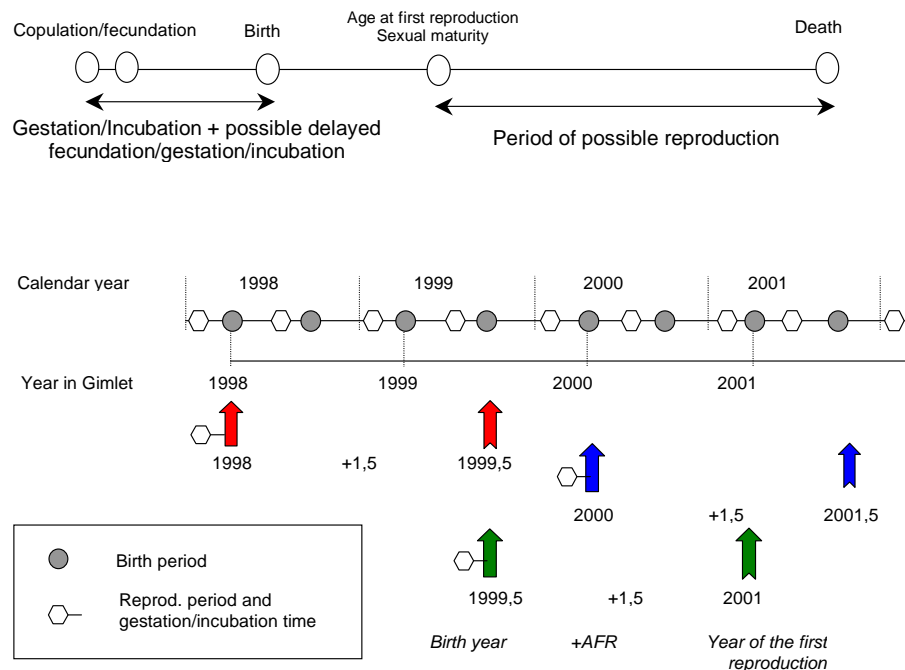


FIGURE 6.2 – Exemple du format de cycle de vie utilisé dans GIMLET qui permet de prendre en compte plusieurs saisons de reproduction au cours d'une année et de prendre en compte les mois dans les âges et les dates.

Enfin, le point le plus original du programme est la possibilité d'estimation de la taille de

population à partir d'un ensemble de génotypes prélevés en une seule ou plusieurs occasions de capture. Ceci peut être réalisé en employant les modèles classiques de capture-marquage-recapture ou la méthode de courbe de raréfaction, ce qui n'est réalisé par aucun logiciel.

6.2.2 Description de modules

GIMLET s'articule autour de deux modules généraux (« *Identification* » et « *Calculator* ») qui permettent d'effectuer respectivement les tâches d'identification et de comparaison de génotypes, et de calculs de différents paramètres de population.

6.2.2.1 Le module « *Identification* »

Ce module permet de réaliser les tâches d'identification et de comparaison de génotypes multilocus afin d'identifier des individus ou regrouper ces génotypes. Il permet aussi de déterminer les parentés par comparaison des génotypes des parents et des enfants.

6.2.2.1.1 L'identification et la comparaison des génotypes

Des comparaisons deux à deux entre génotypes peuvent être réalisées sur l'ensemble des loci afin d'identifier un (ou plusieurs) génotype(s) parmi un lot d'autres génotypes (FIGURE 6.3). Cette identification de génotypes peut aussi être conduite locus après locus. Dans ce cas, le génotype de référence identifié est celui qui aura le plus grand nombre de locus en commun avec le génotype analysé.

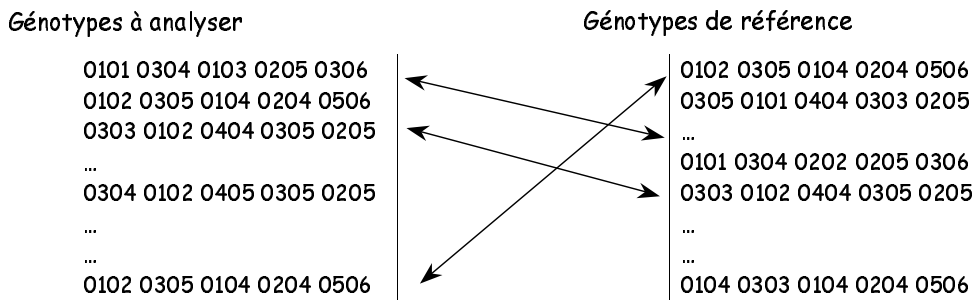


FIGURE 6.3 – Illustration de la tâche d'identification effectuée dans le logiciel GIMLET

Le regroupement de génotypes qui sont identiques entre eux (FIGURE 6.4) est aussi une tâche très importante. Cette tâche permet donc de trier les génotypes en paquets qui peuvent être considérés comme des individus différents si la puissance de discrimination des loci est suffisante (voir section 6.2.2.2.3). Ceci est très utile quand on combine l'échantillonnage non-invasif et l'identification individuelle génétique où, au cours d'une même session de capture, on peut collecter plusieurs échantillons du même individu. Cette tâche de regroupement est similaire à la tâche décrite ci-dessus : comparaisons sur tous les loci en même temps puis locus après locus. Le programme précise les cas où un seul allèle à un ou deux locus diffère(nt) entre les génotypes ou les cas où deux allèles diffèrent à deux loci.

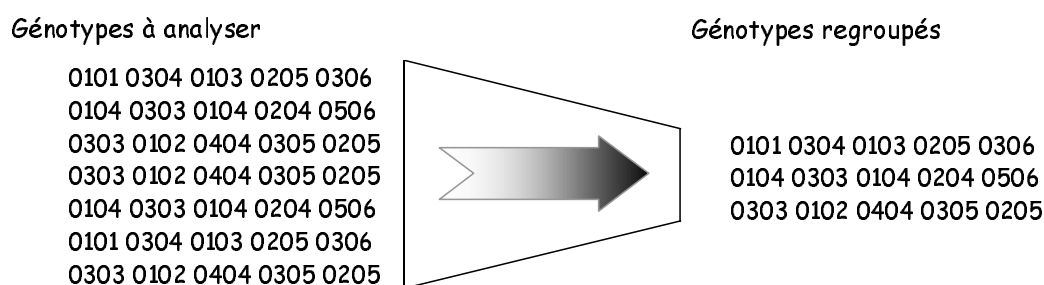


FIGURE 6.4 – Illustration de la tâche de regroupement et de tri effectuée dans GIMLET

6.2.2.1.2 la détermination des parentés potentielles entre individus

Cette détermination est réalisée en analysant la compatibilité des génotypes entre les individus sur la base de la transmission mendélienne des allèles à chaque locus, avec la possibilité de prendre en compte certains paramètres comme l'âge, le sexe des individus, le temps de gestation/incubation, l'âge moyen de première reproduction (FIGURE 6.5).

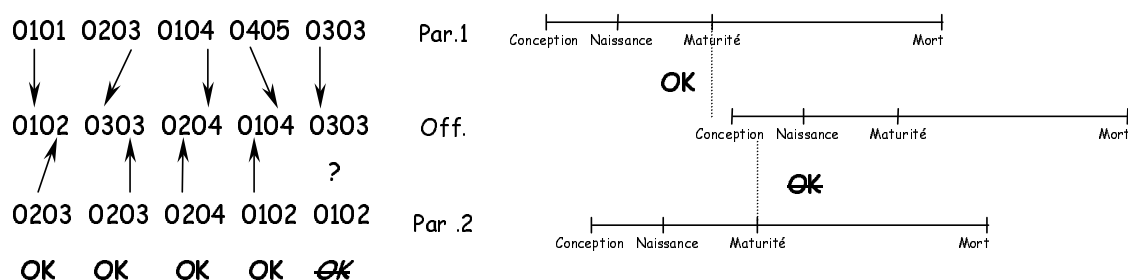


FIGURE 6.5 – Illustration de la détermination de la parenté potentielle entre individus sur la base des génotypes et des données démographiques (date de naissance et de mort, maturité sexuelle, durée de gestation, sexe)

Des individus sont potentiellement les parents d'un individu si les conditions suivantes sont respectées :

1. les génotypes des parents et de l'enfant aux différents locus doivent être compatibles avec une transmission mendélienne des allèles. Une certaine « marge d'erreur » peut être admise en précisant le nombre maximum de locus pour lesquels des incompatibilités (erreurs de génotypage, mutations, ...) sont tolérées.
2. le parent doit être vivant et en âge de se reproduire au moment de la conception de l'enfant (*i.e.* la date de naissance du jeune à laquelle on retire le temps de gestation et le temps de fécondation/incubation/gestation différée).
3. un couple de parents est validé si les individus ont des sexes différents.

6.2.2.2 Le module « Calculator »

Le module *Calculator* permet trois grandes tâches : la construction de génotypes consensus et l'estimation des taux d'erreurs, l'estimation des fréquences alléliques, des hétérozygoties et des probabilités d'identité, et enfin l'estimation de la taille de population.

6.2.2.2.1 La construction des génotypes consensus et l'estimation des taux d'erreurs à partir des génotypes multilocus répétés

La construction de consensus est réalisée sur le même principe que dans le logiciel GEMINI (voir FIGURE 5.5, p.58) avec la prise en compte d'un seuil d'acceptabilité des allèles dans le génotype consensus.

Cette dernière méthode de seuil d'acceptabilité des allèles a été développée et intégrée dans GIMLET afin de s'ajuster au mieux avec la détermination des génotypes consensus par les utilisateurs. En effet, implicitement, on déterminera un génotype consensus si les allèles qui apparaissent sont présents en nombre suffisant (et de surcroît de façon indépendante). Une autre méthode peut être adoptée afin de déterminer le génotype consensus. Cette approche, adaptée sur celle développée dans NAVIDI *et al.* (1992), consiste à calculer la probabilité d'apparition des génotypes répliqués en fonction du génotype consensus possible. NAVIDI *et al.* (1992) calculent la probabilité de tirer les allèles apparus parmi un nombre total de réplifications selon une loi binomiale. Or le modèle sur lequel sont réalisées les simulations de GEMINI nous permet de développer un autre type de calcul. A partir de ce modèle, on peut facilement calculer les probabilités d'apparition de chaque génotype suivant les taux d'erreurs paramétrés, sachant le génotype vrai.

Pour un échantillon (diploïde) donné, posons G le génotype vrai sur un locus et E l'ensemble des génotypes répétés pour le même locus. Nous considérerons qu'un seul évènement peut être à l'origine d'une différence entre deux génotypes quelconques. Prenons le cas simple où au plus deux allèles sont présents sur l'ensemble des réplifications (dans la série E , seuls les allèles A_1 et A_2 sont apparus), ces évènements sont (voir FIGURE 6.6) :

- une **perte d'allèle** (probabilité d'occurrence = τ) quand un hétérozygote au locus considéré est typé comme un homozygote, tel que cet évènement explique le passage de A_1A_2 vers A_1A_1 ou vers A_2A_2 .
- un **faux allèle** (probabilité d'occurrence = ρ) quand un homozygote au locus considéré est typé comme un hétérozygote, tel que cet évènement explique le passage de A_1A_1 ou A_2A_2 vers A_1A_2 .
- **autre cas d'erreur** (probabilité d'occurrence = μ_1) quand l'allèle typé pour un homozygote est « remplacé » par un allèle différent, tel que cet évènement explique le passage de A_1A_1 vers A_2A_2 (ou *vice versa*).

On peut calculer la probabilité $P(E|G)$ d'obtenir E sachant que l'on est en présence du génotype G en considérant les erreurs possibles.

Supposons que A_1 et A_2 sont apparus avec la répartition en génotypes suivante : k_{12} fois le génotype A_1A_2 , k_1 fois le génotype A_1A_1 et k_2 fois le génotype A_2A_2 . D'une façon générale, on a :

$$Pr(E|G) = Pr([N_{A_1A_2} = k_{12}] \cap [N_{A_1A_1} = k_1] \cap [N_{A_2A_2} = k_2] | G) \quad (6.3)$$

où $N_{A_1A_2}$ est le nombre de génotypes A_1A_2 apparus dans E (respectivement pour A_1A_1 et A_2A_2).

Si on pose $G = A_1A_2$, on a :

$$Pr(E|A_1A_2) = \left(1 - \frac{\tau}{2} - \frac{\tau}{2}\right)^{k_{12}} \times \left(\frac{\tau}{2}\right)^{k_1} \times \left(\frac{\tau}{2}\right)^{k_2} \times \frac{(k_{12} + k_1 + k_2)!}{k_{12}!k_1!k_2!}$$

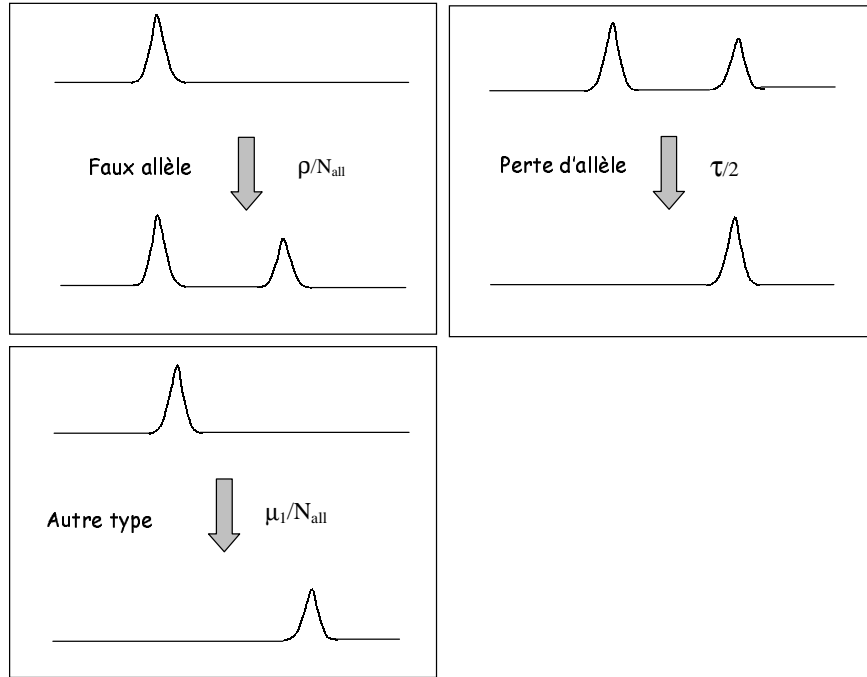


FIGURE 6.6 – Trois types d'évènements (et leur probabilité d'occurrence suivant les différents cas) pouvant expliquer les différences entre génotypes. τ = taux de perte d'allèles ; ρ = taux de faux allèle ; μ_1 = taux de changement d'allèle pour un homozygote ; N_{all} est le nombre d'allèles possibles dans la population considérée.

Les trois premiers termes sont les probabilités conditionnelles des trois évènements constituant E . Le dernier terme est le nombre de combinaisons de k_{12} , k_1 et k_2 éléments dans un ensemble de $k_{12} + k_1 + k_2$ éléments.

Soit :

$$Pr(E|A_1A_2) = (1 - \tau)^{k_{12}} \times \left(\frac{\tau}{2}\right)^{k_1+k_2} \times \frac{(k_{12} + k_1 + k_2)!}{k_{12}!k_1!k_2!}$$

Avec le même raisonnement, si $G = A_1A_1$ et si N_{all} est le nombre d'allèles différents présents dans la population, on a :

$$Pr(E|A_1A_1) = \left(\frac{\rho}{N_{all}}\right)^{k_{12}} \times (1 - \rho - \mu_1)^{k_1} \times \left(\frac{\mu_1}{N_{all}}\right)^{k_2} \times \frac{(k_{12} + k_1 + k_2)!}{k_{12}!k_1!k_2!}$$

et si $G = A_2A_2$:

$$Pr(E|A_2A_2) = \left(\frac{\rho}{N_{all}}\right)^{k_{12}} \times \left(\frac{\mu_1}{N_{all}}\right)^{k_1} \times (1 - \rho - \mu_1)^{k_2} \times \frac{(k_{12} + k_1 + k_2)!}{k_{12}!k_1!k_2!}$$

Ces probabilités sont donc calculées suivant les différents paramètres τ , ρ , μ_1 et N_{all} . On retiendra l'hypothèse nulle qui aura la plus grande valeur Φ définie par :

$$\Phi(A_1A_2) = \frac{Pr(E|A_1A_2)}{Pr(E|A_1A_2) + Pr(E|A_1A_1) + Pr(E|A_2A_2)} \quad (6.4)$$

$$\Phi(A_1A_1) = \frac{Pr(E|A_1A_1)}{Pr(E|A_1A_2) + Pr(E|A_1A_1) + Pr(E|A_2A_2)} \quad (6.5)$$

$$\Phi(A_2A_2) = \frac{Pr(E|A_2A_2)}{Pr(E|A_1A_2) + Pr(E|A_1A_1) + Pr(E|A_2A_2)} \quad (6.6)$$

Il faut bien préciser que les expressions (6.4), (6.5) et (6.6) ne représentent pas les vraisemblances des différents génotypes sachant les données mais les vraisemblances relatives des données sachant les différents génotypes possibles.

Cinq exemples sont donnés dans le TABLEAU 6.1. On peut remarquer que la méthode peut très facilement distinguer le génotype consensus de façon cohérente. La détermination du génotype consensus est évidemment sensible aux taux d'erreurs paramétrés (Cas 2 et Cas 3). Cette sensibilité semble toutefois moins importante pour le taux de changements d'allèles (μ_1) et le nombre d'allèles dans la population (Cas 2 et Cas 4; Cas 2 et Cas 5).

TABLEAU 6.1 – Cinq exemples de la méthode de sélection du génotype consensus à partir du calcul des probabilités d'apparition des génotypes répliqués (huit réplifications). Les cas différents par la répartition des génotypes, les taux d'erreurs et le nombre d'allèles dans la population.

	Cas 1	Cas 2	Cas 3	Cas 4	Cas 5
τ (ADO)	0,1	0,1	0,5	0,1	0,1
ρ (FA)	0,1	0,1	0,1	0,1	0,1
μ_1	0,005	0,005	0,005	0,00001	0,005
k_{12}	4	0	0	0	0
k_1	2	6	6	6	6
k_2	2	2	2	2	2
Nb.allele	4	4	4	4	16
$Pr(A_1A_2)$	0,0017222625	0,0000000011	0,0004272461	0,0000000011	0,0000000011
$Pr(A_1A_1)$	0,0000000002	0,0000372961	0,0000372961	0,0000000001	0,0000026285
$Pr(A_2A_2)$	0,0000000002	0,0000000000	0,0000000000	0,0000000000	0,0000000000
$\Phi(A_1A_2)$	0,9999997177	0,0000293253	0,9197143814	0,319994316	0,000415945
$\Phi(A_1A_1)$	0,0000001411	0,9999706747	0,0802856186	0,680005684	0,9995840549
$\Phi(A_2A_2)$	0,0000001411	0,0000000000	0,0000000000	0,0000000000	0,0000000000

On peut dériver les probabilités $P(E|G)$ avec plus de deux allèles apparaissant au cours des réplifications en prenant en compte les autres cas d'évènements pouvant expliquer les différences entre génotypes (voir FIGURE 6.7).

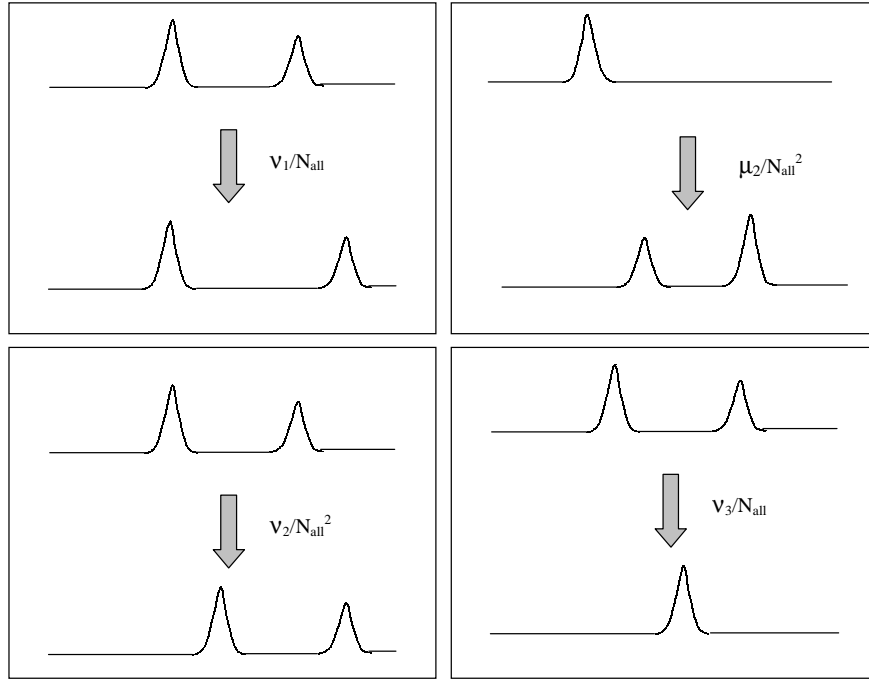


FIGURE 6.7 – Différents types d'évènements (et leur probabilité d'occurrence) pouvant expliquer le passage d'un génotype à l'autre. ν_1 = taux de changement d'un allèle pour un hétérozygote ; ν_2 = taux de changement des deux allèles pour un hétérozygote ; ν_3 = taux de changement d'un hétérozygote vers un homozygote.

Dans ce cas, on peut raisonner de la même façon que pour l'équation (6.3). Nous aurons pour tout allèle i et tout allèle j ($i \neq j$) :

$$\begin{aligned} Pr(E|A_i A_j) = & Pr \left([N_{A_i A_j} = k_{ij}] \cap [N_{A_i A_i} = k_i] \cap [N_{A_j A_j} = k_j] \cap \right. \\ & \left. [N_{A_i A_o} = k_{i_o}] \cap [N_{A_j A_o} = k_{j_o}] \cap [N_{A_o A_o} = k_o] \cap [N_{A_o A_\bullet} = k_{o_\bullet}] | G \right) \end{aligned}$$

où o et \bullet désignent tout allèle différent de i et de j .

Ainsi :

$$\begin{aligned} Pr(E|A_i A_j) = & (1 - \tau - \nu_1 - \nu_2 - \nu_3)^{k_{ij}} \times \left(\frac{\tau}{2}\right)^{k_i + k_j} \times \left(\frac{\nu_1}{N_{all}}\right)^{k_{i_o} + k_{j_o}} \times \left(\frac{\nu_2}{N_{all}^2}\right)^{k_{o_\bullet}} \times \left(\frac{\nu_3}{N_{all}}\right)^{k_o} \\ & \times \frac{(k_{ij} + k_i + k_j + k_{i_o} + k_{j_o} + k_o + k_{o_\bullet})!}{k_{ij}! k_i! k_j! k_{i_o}! k_{j_o}! k_o! k_{o_\bullet}!} \end{aligned}$$

Pour tout allèle A_i , on a :

$$Pr(E|A_i A_i) = Pr([N_{A_i A_i} = k_i] \cap [N_{A_i A_o} = k_{i_o}] \cap [N_{A_o A_o} = k_o] \cap [N_{A_o A_\bullet} = k_{o_\bullet}] | G)$$

d'où

$$Pr(E|A_i A_i) = (1 - \rho - \mu_1 - \mu_2)^{k_i} \times \left(\frac{\rho}{N_{all}}\right)^{k_{i_o}} \times \left(\frac{\mu_1}{N_{all}}\right)^{k_o} \times \left(\frac{\mu_2}{N_{all}^2}\right)^{k_{o\bullet}} \\ \times \frac{(k_i + k_{i_o} + k_o + k_{o\bullet})!}{k_i! k_{i_o}! k_o! k_{o\bullet}!}$$

Malgré la nécessité de connaître les probabilités d'occurrence de chaque type d'évènements possibles, ainsi que le nombre d'allèles présents dans la population, cette méthode peut être adoptée pour sélectionner le génotype consensus. Elle a été testée et semble donner de bons résultats. Cependant cette méthode n'est pas intégrée dans la version actuelle de GEMINI ou GIMLET car d'autres tests doivent être menés.

L'estimation des différents taux d'erreurs (et éventuellement les différents évènements cités ci-dessus) à partir des données de génotypage multiple est également possible. La proportion de pertes d'allèle et de faux allèles est calculée en comparant chaque génotype issu d'une réplication PCR avec le génotype consensus. Ce dernier peut être soit construit par GIMLET soit être fourni par l'utilisateur.

6.2.2.2 L'estimation des fréquences alléliques, des taux d'hétérozygotie et des probabilités d'identité à partir des génotypes

Les fréquences alléliques calculées sont les fréquences observées de chaque allèle à chaque locus. Le taux d'hétérozygotie observé est calculé en comptant le nombre d'hétérozygotes dans la population. L'hétérozygotie attendue (selon les proportions de Hardy-Weinberg) est calculée grâce à l'équation :

$$H_{exp} = 1 - \sum_i p_i^2 \quad (6.7)$$

où p_i est la fréquences de l'allèle i .

L'hétérozygotie (observée ou attendue) sur l'ensemble des loci est calculée par :

$$H = \frac{\sum_j H_j}{n_j} \quad (6.8)$$

où H_j est l'hétérozygotie au locus j et n_j est le nombre total de locus.

La probabilité d'identité est la probabilité que deux individus pris au hasard dans une population aient le même génotype sur un ou plusieurs loci. C'est la statistique la plus utilisée pour rendre compte de la puissance et la capacité d'un marqueur à différencier deux individus d'une population (WAITS et LEBERG 2000). La probabilité d'identité théorique peut être calculée de plusieurs façons (WAITS *et al.* 2001) à partir des fréquences alléliques des loci étudiés. Trois types de calcul sont généralement utilisés pour les marqueurs codominants. Le premier est celui utilisé pour une population dont les individus s'appartiennent au hasard. L'équation (JAMIESON 1965, PAETKAU et STROBECK 1994) pour un locus est :

$$PI = \sum_i p_i^4 + \sum_i \sum_{j>i} (2p_i p_j)^2 \quad (6.9)$$

Le deuxième calcul est dérivé de l'équation ci-dessus mais corrigée pour la taille de l'échantillon utilisé pour estimer les fréquences alléliques¹³ (KENDALL et STEWART 1977, PAETKAU *et al.* 1998b) :

$$PI_{unbiased} = \frac{n^3(2a_2^2 - a_4) - 2n^2(a_3 + 2a_2) + n(9a_2 + 2) - 6}{(n-1)(n-2)(n-3)} \quad (6.10)$$

où n est la taille d'échantillon, a_i est égal à $\sum_j p_j^i$, et p_j la fréquence de l'allèle j .

Ces équations ne prennent pas en compte l'éventuelle structure sociale d'une population. En effet, s'il existe une forte structure sociale, les individus vont avoir tendance à être apparentés et à avoir des génotypes similaires. Ainsi, une troisième équation (EVETT et WEIR 1998, TABERLET et LUIKART 1999) peut être proposée pour une population dont les individus sont uniquement des frères-soeurs (*sibs*) :

$$PI_{sibs} = \frac{1}{4} + \frac{1}{2} \sum_i p_i^2 + \frac{1}{2} \left(\sum_i p_i^2 \right)^2 - \frac{1}{4} \sum_i p_i^4 \quad (6.11)$$

WAITS *et al.* (2001) ont montré que la probabilité observée dans une population est généralement comprise entre les probabilités calculées à partir des équations (6.10) et (6.11). L'estimation de la probabilité d'identité pour un ensemble de loci peut être faite en multipliant les probabilités pour chaque locus entre elles en considérant que les loci sont indépendants entre eux :

$$PI_{allloci} = \prod_j PI_j \quad (6.12)$$

où PI_j est la probabilité d'identité calculée pour le locus j .

Les calculs de ces probabilités permettent de savoir si le système de loci que l'on va utiliser sera apte à différencier les individus d'une population. Les seuils de probabilité d'identité recommandés sont de 0.01 pour l'estimation de la taille d'une population (MILLS *et al.* 2000) et de 0.001-0.0001 (voir WAITS *et al.* 2001) pour l'identification individuelle dans les cas de braconnage par exemple (*forensics*). La différence entre ces seuils provient simplement du fait de la tolérance aux confusions d'identification dans les différentes applications. Pour des cas de braconnage, le seuil est inférieur à celui pour l'estimation de la taille de population ou le suivi d'individus, car dans le premier cas, les erreurs ne peuvent être tolérées (inculpation d'une personne) alors que dans le second, elles le sont davantage.

¹³Attention avec cette équation car la taille de l'échantillon admet un minimum d'échantillons utilisés. Par exemple pour un locus à deux allèles (fréquences : 0,75 et 0,25), si la taille de l'échantillon est en dessous de 10, la valeur de PI est négative...

6.2.2.2.3 L'estimation de l'effectif de population

Cette tâche ne permet pas directement d'estimer l'effectif d'une population à partir des données de génotypage. Elle permet de créer des fichiers de données qui seront par la suite traités pour déterminer l'effectif de population. GIMLET propose deux méthodes d'estimation à partir du génotypage des échantillons. L'approche de capture-recapture utilise les génotypes des individus comme des marques et GIMLET construit des fichiers d'histoire de capture des génotypes uniques. Ces fichiers peuvent être ultérieurement traités dans le programme CAPTURE (OTIS *et al.* 1978) pour estimer la taille de population de la même façon que pour des histoires de capture d'individus.

Pour la méthode basée sur la courbe de raréfaction (voir Section 6.1, p.72), l'ordre de typage et d'analyse des échantillons peut influencer l'estimation de la taille de population puisque cet ordre va déterminer la forme de la relation observée (voir par exemple FIGURE 6.8). L'ajustement sera donc différent entre les ordres adoptés (KOHN *et al.* 1999). Il est donc nécessaire de simuler plusieurs ordres possibles et d'estimer la taille de population à partir de l'ensemble de ces estimations.

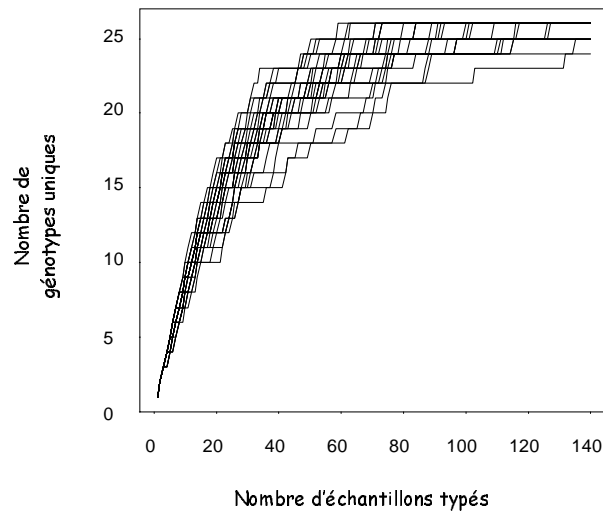


FIGURE 6.8 – Effet de l'ordre d'analyse des échantillons (30 ordres d'échantillons différents) sur la relation du nombre de génotypes uniques en fonction du nombre d'échantillons analysés.

GIMLET produit des fichiers qui regroupent le nombre d'apparition de chaque génotype unique au cours de l'analyse. Ces fichiers peuvent être analysés dans le programme R (IHAKA et GENTLEMAN 1996) en utilisant un fichier script écrit également par GIMLET. La procédure dans R consiste, à partir de ces fichiers, à construire plusieurs séries d'ordre d'analyse des échantillons, de construire la courbe moyenne de l'ensemble des observations et d'ajuster (ajustement non linéaire) cette courbe moyenne aux équations (6.1) et (6.2).

De plus, GIMLET permet aussi de construire des fichiers au format de SURGE (format BIOMECO pouvant être converti au format de SURGE avec CR, PRADEL et LEBRETON 1993, PRADEL *et al.* 1995) et de MARK (WHITE et BURNHAM 1999) afin de conduire des estimations de survie par exemple.

Troisième partie

Discussion et Conclusion

Chapitre 7

L'étude pilote : une étape indispensable avant une étude à grande échelle

Sommaire

7.1	Tester les méthodes sur le terrain	88
7.2	Tester les analyses génétiques à partir des échantillons non-invasifs	88
7.2.1	Test de l'amplification d'ADN à partir des échantillons	89
7.2.2	Estimation des taux d'erreurs, des fréquences alléliques et de la probabilité d'identité	89
7.2.2.1	Nombre de répétitions par échantillon et par locus dans l'étude pilote	91
7.2.2.2	Combien d'échantillons récolter pour l'étude pilote ?	91
7.3	Détermination de la stratégie optimale	91

Avant d'utiliser toute approche, il est impératif de tester, « en grandeur nature », son application à la fois sur le terrain et au laboratoire. Une étude pilote est alors indispensable (BAVERSTOCK et MORITZ 1990, TABERLET et LUIKART 1999, TABERLET *et al.* 2001). L'étude pilote doit permettre (i) de tester la possibilité de récolter sur le terrain les échantillons en assez grand nombre (dans les conditions discutées précédemment, voir Section 4.3, p.49) pour le besoin de l'étude, (ii) de tester les analyses génétiques à partir des échantillons récoltés et enfin (iii) de pouvoir prédire l'effort qu'il faudra fournir pour pouvoir appliquer à grande échelle les méthodes choisies.

La contribution de mon travail dans cette optique se situe essentiellement, mais pas uniquement, dans la troisième partie de cette démarche. Nous allons voir successivement les trois étapes de cette étude pilote.

7.1 TESTER LES MÉTHODES SUR LE TERRAIN

Même si certaines méthodes non-invasives ne nécessitent pas nécessairement de grands moyens, il est utile de tester préalablement l'échantillonnage sur le terrain. Trouve-t-on assez de fèces pour pouvoir analyser la biologie de l'animal en question ? Est-il possible de récolter d'autres échantillons (urine) et dans quelles conditions ? Certains dispositifs demandent des mises au point afin de maximiser les chances d'obtenir des résultats probants. Par exemple, les dispositifs proposés par WOODS *et al.* (1999) ou MOWAT *et al.* (1999) doivent être expérimentés afin de les adapter à l'espèce et au milieu écologique échantillonné. Il n'est pas évident que des dispositifs donnant de bons résultats pour les ours ou les mustélidés dans les Rocheuses américaines donnent aussi de bons résultats pour des populations de la même espèce en Europe. Comme précisé précédemment, une collaboration entre les personnes de terrain et les généticiens doit être initiée et entretenue. Par exemple, MOWAT *et al.* (1999) ou POOLE *et al.* (2002) ont confié le placement et le relevé des dispositifs de « piégeage » des poils de grizzlys à des personnes ayant une expérience considérable dans la chasse et l'observation des ours dans les secteurs concernés. A ce niveau, toutes les précautions doivent être prises en compte et les contraintes (récolte, préservation) testées, ceci d'autant plus que le type d'échantillon est nouveau ou n'a jamais été testé. Ainsi l'utilisation de gants, de pinces mais aussi les modes de préservation (tampon, Silica Gel, enveloppe) requis pour chaque échantillon prélevé devront être testés ou utilisés en gardant à l'esprit que chaque cas d'échantillonnage peut être particulier. Notons enfin que certaines techniques de récolte et de préservation des échantillons sont plus efficaces que d'autres pour certains échantillons et certaines conditions (voir discussion dans la section 5.3.1, p.62).

7.2 TESTER LES ANALYSES GÉNÉTIQUES À PARTIR DES ÉCHANTILLONS NON-INVASIFS

Comme nous l'avons déjà vu, le rendement de l'amplification PCR et la fiabilité des résultats de génotypage ne sont pas toujours assurés quand on utilise des échantillons non-invasifs. Il est donc primordial de tester la possibilité d'amplifier les marqueurs microsatellites à partir des échantillons mais aussi de tester la fiabilité des résultats obtenus à partir de ces échantillons. De nombreux articles traitent d'ailleurs du test d'extraction, d'amplification et de génotypage d'ADN à partir d'échantillons non-invasifs (BAYES *et al.* 2000, LATHUILLIÈRE *et al.* 2001, par exemple).

Je présenterai ici une démarche pour l'étude pilote génétique inspirée des approches proposées par TABERLET et LUIKART (1999) et TABERLET *et al.* (1999). TABERLET *et al.* (1999) proposent

une approche séquentielle qui permet d'appréhender au mieux la partie génétique d'une analyse. Cette approche consiste en plusieurs étapes (voir FIGURE 7.1a) : amplification d'ADN mitochondrial, amplification d'ADN nucléaire, estimation d'un seuil de tolérance pour les taux d'erreurs compatibles avec la question biologique, et enfin le test expérimental du taux d'erreur. A chaque étape, si les résultats ne sont pas satisfaisants des optimisations et améliorations sont nécessaires (sauf pour la troisième) si besoin est. Sinon, l'approche non-invasive devra être abandonnée.

TABERLET et LUIKART (1999) ont proposé, sur une idée similaire, de simuler une population virtuelle ainsi que le génotypage de ses individus en introduisant des erreurs suivant différents taux. Ceci a pour but de permettre d'estimer le taux d'erreur compatible avec la question posée suivant l'impact des erreurs. En fait, pourquoi ne pas réaliser les simulations directement avec les paramètres et les taux d'erreurs de l'étude en question. On connaîtra ainsi directement l'impact des erreurs sur les analyses. Dans cette optique, si l'on simule la correction des erreurs par l'approche multitubes il est possible de déterminer le nombre de répétitions minimum pour corriger toutes les erreurs stochastiques (ou presque) et obtenir des résultats d'identification fiables, compatibles avec la question posée.

Cette démarche (voir FIGURE 7.1b) est basée sur trois étapes principales : le test d'amplification de l'ADN, l'estimation des taux d'erreurs et des fréquences alléliques et enfin le choix du nombre de loci et de répétitions de génotypage par simulations informatiques.

7.2.1 Test de l'amplification d'ADN à partir des échantillons

La première étape de l'étude pilote génétique est de s'assurer que l'on peut amplifier l'ADN à partir des échantillons récoltés. Ce test peut tout d'abord se réaliser avec l'amplification de l'ADN mitochondrial (ADNmt). En effet, dans les cellules, l'ADNmt est en plus grande quantité (environ mille copies par cellule) que l'ADN nucléaire (ADNn; une copie par allèle et par cellule). Il est donc a priori plus facile d'amplifier de l'ADNmt que de l'ADNn à partir d'échantillons ne contenant qu'un très faible nombre de cellules exploitables. Si le test de l'ADNmt n'est pas concluant, il sera difficile d'amplifier l'ADNn à partir des mêmes échantillons. Dans ce cas là, mieux vaut par exemple essayer d'autres techniques d'extraction d'ADN ou de préservation des échantillons.

Si le test de l'ADNmt est concluant, l'étape suivante sera de tester l'amplification de l'ADNn (microsatellites). Cette étape est assez délicate, car même s'il est possible d'amplifier de l'ADNmt, il n'est pas toujours possible de pouvoir amplifier de l'ADNn. De plus, cette étape nécessite parfois de lourdes mises au point afin de déterminer les conditions optimales de PCR pour chaque locus étudié. Les conditions PCR déterminées à partir d'échantillons de tissus, par exemple, ne sont pas forcément les mêmes que pour les échantillons de fèces ou de poils. En ce qui concerne le loup, j'avais déterminé les conditions PCR optimales à partir d'échantillons de tissu dilué 20x et 100x (afin de se rapprocher des quantités d'ADN présentes dans les échantillons de fèces). Quand j'ai utilisé les échantillons de fèces, j'ai rencontré des difficultés pour amplifier l'ADNn à partir de ces échantillons et j'ai dû optimiser de nouveau les conditions PCR pour ces échantillons.

7.2.2 Estimation des taux d'erreurs, des fréquences alléliques et de la probabilité d'identité

Les échantillons non-invasifs sont connus pour la difficulté à pouvoir amplifier de l'ADN et à obtenir des résultats fiables de génotypage microsatellites. Une revue de la littérature montre que les taux de réussite sont très variables suivant l'espèce, le type d'échantillon ou la méthode d'extraction utilisée (voir TABLEAU 2.1, p.33 et TABLEAU 2.2, p.37). Il est donc assez difficile

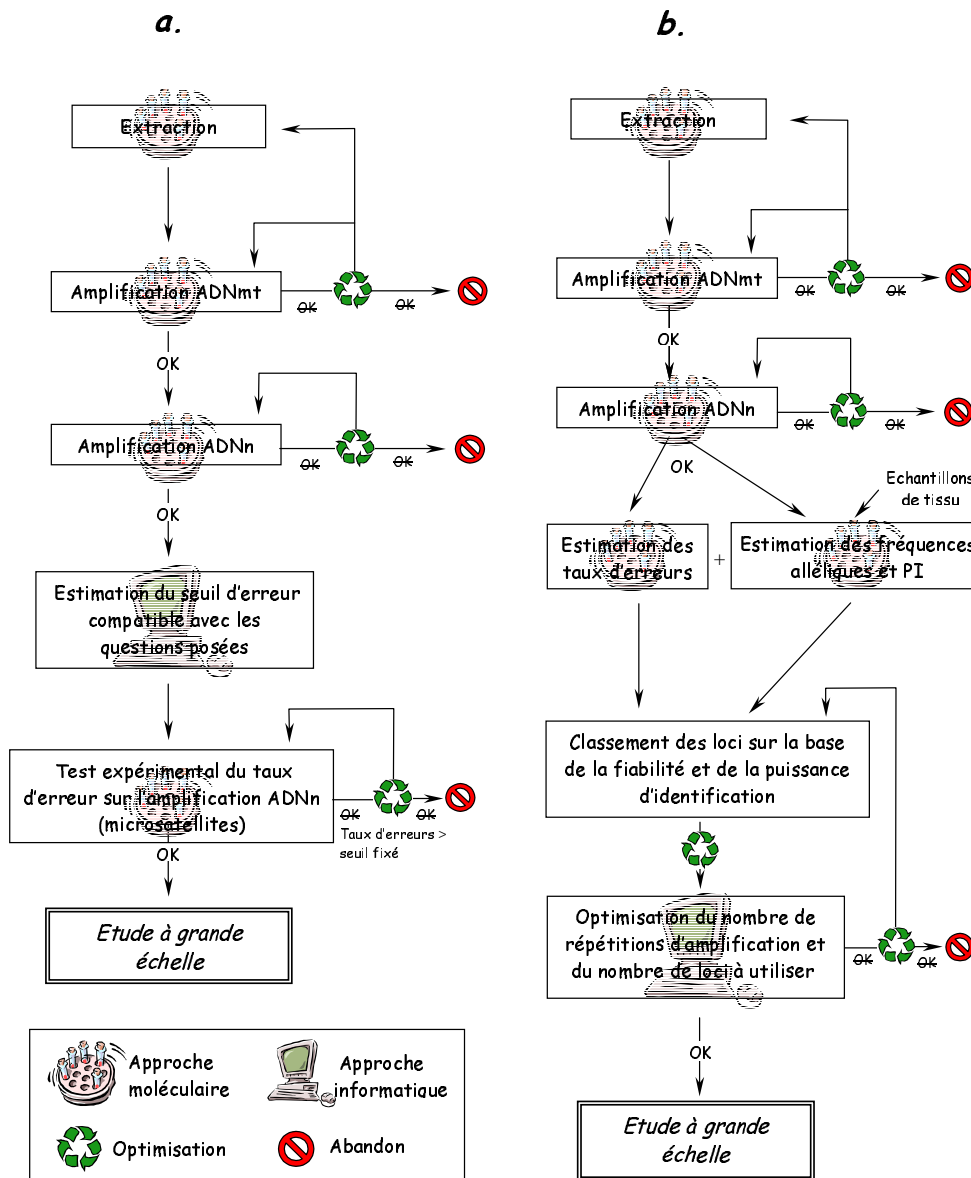


FIGURE 7.1 – a. Approche séquentielle proposée par TABERLET *et al.* (1999) permettant de tester l'utilisation des échantillons non-invasifs pour l'identification individuelle par génotypage microsatellite. b. Modification de l'approche séquentielle sur la base de TABERLET *et al.* (1999) et TABERLET et LUIKART (1999).

de se baser sur l'étude d'une espèce ou d'une population voisine pour estimer le taux d'erreurs dans un cas particulier (TABERLET *et al.* 1999). De plus, pour une même population les taux d'erreurs peuvent varier en fonction de la saison (MAUDET *et al.* 0000). Dans ces conditions il est nécessaire d'estimer les taux d'erreurs mais aussi les fréquences alléliques et la probabilité d'identité du système de locus utilisé (ces données seront utilisées dans la suite de l'étude pilote). L'approche pour évaluer le taux d'erreurs de génotypage est d'utiliser quelques échantillons et de répéter plusieurs fois l'amplification PCR pour chaque locus.

Le nombre de répétitions réalisées pour chaque locus et chaque échantillon doit être suffisant

pour pouvoir détecter les faibles taux d'erreurs (TABERLET et LUIKART 1999) et pour pouvoir obtenir des génotypes consensus fiables. Afin d'estimer les taux d'erreurs, il suffit de comparer ce génotype consensus avec les résultats de chacune des répétitions et de compter le nombre de différences. Cette procédure est réalisée par le logiciel GIMLET. De plus, le nombre d'échantillons doit être suffisant pour avoir des estimations assez précises dans la population étudiée.

7.2.2.1 Nombre de répétitions par échantillon et par locus dans l'étude pilote

Je conseille de réaliser au minimum 8 répétitions par locus et échantillon (mais voir plus loin) pour trois raisons :

- TABERLET *et al.* (1996) préconisent la réalisation de 4 à 7 répétitions par échantillon et locus afin d'obtenir un niveau de fiabilité de 99%.
- Des simulations que j'ai réalisées avec GEMINI dans le cadre de l'étude du loup français montrent que 8 répétitions sont nécessaires pour corriger les erreurs de génotypage à un niveau de fiabilité supérieur à 99% (les taux de perte d'allèles sont compris entre 0,01 et 0,4 et de faux allèles entre 0 et 0,1 selon les loci [N=6]). Il en est de même pour des simulations utilisant des fréquences alléliques provenant de FORBES et BOYD (1996) sur 10 loci avec des taux d'erreurs de 0,3 (ADO) et 0,05 (FA). Chez les rhinocéros (voir Article V), le nombre conseillé de répétitions est d'au moins 4-5 répétitions avec des valeurs de taux d'erreurs de 0,3 (ADO) et 0 (FA).
- Huit répétitions par échantillon et par locus sont « facilement » réalisables avec l'utilisation des barrettes de tubes PCR attachés et une multi-pipette.

7.2.2.2 Combien d'échantillons récolter pour l'étude pilote ?

Une question difficile à résoudre est celle du nombre d'échantillons que l'on va devoir utiliser pour l'étude pilote. En effet, ce nombre ne doit pas être trop faible afin que l'on puisse obtenir des estimations de taux d'erreurs fiables, pour rendre compte de la variabilité des taux d'erreurs entre échantillons, mais aussi pour pouvoir convenablement estimer les fréquences alléliques. D'autre part, il ne doit pas être trop grand car une étude pilote doit rester une étude pilote et ne doit pas échantillonner toute la population si les méthodes utilisées s'avèrent peu fiables. Il existe donc un compromis entre une borne inférieure de fiabilité et une borne supérieure de dépenses superflues. TABERLET et LUIKART (1999) préconisent d'échantillonner entre 20 et 50 individus mais ce chiffre pourrait dépendre de la taille de population étudiée. L'effort d'échantillonnage lors de l'étude pilote ne peut être le même dans une population de 30 individus ou dans une population de 200. De plus, il est important pour l'estimation des fréquences alléliques que les échantillons récoltés proviennent d'individus différents. Or, pour compliquer la question, il faut être sûr que la probabilité d'identité soit assez faible pour pouvoir s'assurer que l'on puisse différencier les individus.

7.3 DÉTERMINATION DE LA STRATÉGIE OPTIMALE

Cette problématique a été le point central de ma thèse avec l'élaboration d'un logiciel (GEMINI) permettant de déterminer la stratégie optimale d'une étude génétique utilisant l'identification individuelle par génotypage (de type microsatellite). Cette détermination peut se faire par des simulations suivant un modèle (voir Section 5.1.1, p.54). Pour être le plus proche possible de la situation d'étude, on doit fournir au logiciel des paramètres comme par exemple les fréquences des allèles pour chaque locus utilisé. L'utilisateur pourra alors simuler son étude et faire varier

certaines paramètres pour pouvoir déterminer la meilleure stratégie. Nous avons vu l'utilisation de GEMINI pour déterminer le nombre minimal de répétitions PCR afin d'obtenir un niveau de fiabilité d'identification maximal. Mais ce n'est pas le seul paramètre à déterminer. Par exemple le nombre de loci à utiliser est aussi un point crucial (Article IV). GEMINI permet ainsi de déterminer le nombre minimum de loci pour pouvoir avoir une grande puissance de discrimination des individus couplée à une fiabilité maximale (supérieure à 99% voire équivalente à 100%). Un autre point pour lequel GEMINI peut être utile est la détermination de l'effort d'échantillonnage qui peut donner des résultats fiable et précis, notamment en ce qui concerne la taille de population. En effet, même si des grandes directives peuvent être dégagées sur l'intensité d'échantillonnage et la méthode d'estimation de la taille de population (voir Article V), il peut être intéressant de tester sa propre approche pour savoir si elle peut donner des résultats non biaisés et précis.

Si le but de l'étude pilote est uniquement de tester la fiabilité de génotypage, d'autres stratégies sont possibles, comme les méthodes de MILLER *et al.* (2002) ou de MORIN *et al.* (2001) ou encore celle de ré-échantillonnage des répétitions de PCR (méthode présentée dans l'Article IV). Cette dernière s'intègre parfaitement bien dans le cadre d'une étude pilote. Les résultats de génotypage réalisé au cours de l'étude pilote seront évalués pour déterminer le nombre moyen de répétitions de génotypage à réaliser pour chaque échantillon. Ainsi, au lieu de huit répétitions conseillées dans la section 7.2.2.1, on réalisera 12 répétitions (afin d'augmenter la précision de la méthode) et on appliquera la méthode de ré-échantillonnage des répétitions (voir FIGURE 5.6, p.60). Par ailleurs, si l'on veut utiliser uniquement GEMINI, seules huit répétitions sont nécessaires afin d'être sûr des génotypes consensus. L'utilisation de GEMINI n'est pas remise en cause par l'approche proposée dans l'Article IV car GEMINI peut servir à tester les variations des taux d'erreurs ou d'autres paramètres comme l'effort d'échantillonnage. Ainsi, l'approche de ré-échantillonnage des répétitions sera idéale pour tester le génotypage multiple et GEMINI pour tester le génotypage multiple associé à un échantillonnage et des conditions d'étude particulières pouvant être modifiées.



Chapitre 8

Conclusion Générale

LES RÉSULTATS PRINCIPAUX

Les objectifs principaux du présent mémoire étaient d'aborder et de discuter les différentes améliorations et optimisations que l'on pouvait apporter afin d'utiliser au mieux le typage des microsattellites combiné avec l'échantillonnage non-invasif. Cette approche est nécessaire pour l'étude de nombreuses populations qui ne peuvent être étudiées que grâce aux méthodes non-invasives (populations difficiles à étudier par le biais d'observation ou de capture ; populations en danger). L'optimisation d'une telle approche est nécessaire, en partie parce que des erreurs peuvent se produire quand on travaille avec des échantillons ne contenant qu'une quantité limitée d'ADN, souvent dégradé et en présence d'inhibiteurs de réactions enzymatiques. Mais d'autres points doivent être abordés comme l'échantillonnage lui-même et l'analyse de données.

L'idée générale qui se dégage est que l'échantillonnage non-invasif est d'une utilité considérable dans certains cas et qu'il peut apporter des informations précieuses sur certains points de la biologie d'une population (voir Article I). Ce mode d'échantillonnage reste néanmoins une alternative à la capture et autres modes invasifs, notamment si on considère l'effort pour optimiser les analyses génétique réalisées à partir des échantillons. Cet effort doit se positionner sur trois niveaux : l'échantillonnage, le génotypage et l'analyse des données.

Au niveau de l'échantillonnage. Premièrement, il faudra veiller à la situation précise du site d'étude et de la biologie de l'espèce. Il a été discuté de certains problèmes liés avec la structuration de l'habitat et la structuration sociale des groupes pouvant influencer l'échantillonnage (variabilité des probabilités de capture en fonction des individus). Cette hétérogénéité peut avoir des conséquences dans l'interprétation des données, d'autant plus dans le cas de l'échantillonnage non-invasif strict où la modulation de l'effort d'échantillonnage est limité à la seule prospection d'une partie accessible par l'homme des territoires des animaux. La structure spatiale des échantillons qui sera détectée sera donc influencée par l'intensité d'échantillonnage, la topologie du terrain, la mobilité des individus, la structure sociale (représentation respective des individus dominants/dominés, mâles/femelles, jeunes/adultes) et les conditions climatiques (*e.g.* les précipitations neigeuses pour le suivi des traces afin de maximiser la récolte d'indices). Dans le contexte de l'estimation de taille de population, l'efficacité et la sensibilité des méthodes utilisées à différents facteurs dont l'hétérogénéité des probabilités de capture ou les erreurs de génotypage ont été testées (Article VI) et il a été montré que cette hétérogénéité peut avoir une grande influence sur les estimations. La prise en compte de ces influences rendra possible la correction des analyses mais aussi apportera beaucoup sur l'interprétation des résultats. D'autre part, il faudra se concentrer sur les conditions de récolte et de préservation des échantillons afin de limiter la dégradation ou la contamination de l'ADN. Dans tous les cas de figure, une concertation entre les généticiens et les biologistes de terrain (voire l'implication des mêmes personnes dans les différents stades de l'étude) doit être un point central de la phase d'échantillonnage afin d'associer les connaissances et les compétences des deux parties.

Au niveau du génotypage. Le génotypage est évidemment une étape essentielle pour de l'analyse génétique. Outre les moyens pour diminuer les risques de contamination (KWOK et HIGUSHI 1989) et pour augmenter la potentialité d'amplifier l'ADN à partir des échantillons, l'un des meilleurs moyens pour corriger les erreurs stochastiques (dues à la faible quantité de l'ADN souvent dégradé) est de répéter le processus de génotypage afin de déterminer le génotype consensus sur l'ensemble des réplifications (approche multi-tubes, NAVIDI *et al.* 1992). L'approche multi-tubes permet, en revanche, de corriger uniquement les erreurs stochastiques, c'est-à-dire les erreurs qui se produisent aléatoirement au cours des manipulations. Les erreurs systématiques

apparaissent quant à elles dans toutes les réplifications d'une série de manipulations et ne seront pas corrigées. Ces erreurs systématiques peuvent provenir, par exemple, de la contamination d'un lot de réactifs utilisé pour l'extraction d'ADN ou la préparation d'une PCR. Dans ce cas, quasiment toutes les réactions préparées à partir de ce lot présenteront des produits contaminants et il ne servira à rien de répéter les PCR pour corriger ces « erreurs ». Pour éviter ce type de problème il est nécessaire de diviser les lots de réactifs en lots plus petits pour que les contaminations ne soient pas présentes dans une grande quantité de réactif. Les allèles nuls peuvent aussi constituer des erreurs systématiques dans le sens où, pour un échantillon et quel que soit le nombre de réplifications réalisées, l'erreur sera toujours commise. Les allèles nuls doivent être considérés avec beaucoup d'intérêt car ils peuvent être présents pour la plupart des loci étudiés (CALLEN *et al.* 1993, PAETKAU et STROBECK 1995, PEMBERTON *et al.* 1995).

L'approche multi-tubes peut accroître considérablement les coûts des analyses génétique et il est nécessaire de réduire au maximum le nombre de réplifications de PCR par échantillon et par locus. Ainsi, un programme informatique a été développé (GEMINI, Article III) pour déterminer ce nombre de réplifications *via* la simulation des génotypes. En revanche, plusieurs autres méthodes de sélection ont été proposées (MORIN *et al.* 2001, MILLER *et al.* 2002, Article IV). L'utilisateur pourra choisir une méthode en fonction de (i) s'il peut investir ou avoir accès à du matériel encore peu répandu aujourd'hui (PCR quantitative; MORIN *et al.* 2001), (ii) s'il veut utiliser une méthode basée sur un modèle mathématique qui peut être sensible aux variations du taux d'erreurs entre loci ou à une mauvaise estimation des fréquences alléliques (méthode de MILLER *et al.* 2002), et enfin (iii) s'il désire faire une étude pilote génétique sur quelques échantillons et déterminer un nombre moyen de réplifications PCR pour tous les échantillons, risquant ainsi d'en faire trop ou pas assez suivant la variations du taux d'erreurs entre échantillons, mais permettant de simplifier les manipulations de l'analyse génétique (méthode présentée dans l'Article IV).

Le logiciel GEMINI, quant à lui, permet de tester, par exemple, l'influence du type d'échantillonnage, de l'effort d'échantillonnage, et du taux d'erreurs sur les résultats. Différentes modalités de ces paramètres peuvent être simulées et leurs effets peuvent être testés sur des analyses (estimations de fréquences alléliques, structuration de population, estimation de taille de population) effectuées à partir des génotypes (Article VI) afin de déterminer la combinaison des différents paramètres minimisant les biais tout en minimisant les coûts.

Au niveau de l'analyse. L'analyse des données est la dernière étape avant l'interprétation. Dans ce mémoire est présenté un nouveau logiciel d'analyse des données microsatellites qui permet de réaliser plusieurs tâches. Ce logiciel n'est pas destiné à remplacer d'autres logiciels comme GENEPOP (RAYMOND et ROUSSET 1995), GENETIX (BELKHIR *et al.* 1996), ARLEQUIN (SCHNEIDER *et al.* 2000) ou CERVUS (MARSHALL *et al.* 1998). GIMLET donne la possibilité de réaliser des analyses, comme les calculs des fréquences alléliques, des hétérozygoties ou la parenté entre individus qui sont effectués par les autres programmes. Néanmoins, il s'agissait de les intégrer pour que l'utilisateur puisse associer les différentes tâches offertes par GIMLET. Ce dernier a notamment été développé pour traiter les données issues de l'échantillonnage non-invasif en fournissant la possibilité de construire automatiquement les génotypes consensus, de trier et identifier les génotypes identiques afin de les sauvegarder dans des fichiers au format GENEPOP et de pouvoir les utiliser dans d'autres analyses. Ces tâches n'ont jamais été implémentées auparavant dans d'autres programmes. GIMLET permet également de calculer les probabilités d'identité ce que ne permettait aucun programme publié (seul PROBID pouvait les réaliser, G.LUIKART, non publié). Une autre originalité de GIMLET est de permettre la construction automatique de fichiers (fichiers d'histoire de capture des génotypes, fichiers script) qui seront utilisés pour l'estimation de la taille de population en utilisant deux types de méthodes (modèles de capture-marquage-

recapture et courbe de raréfaction).

L'étude pilote. Avant toute étude à grande échelle, il faudra donc se concentrer sur la faisabilité de la démarche choisie en testant au préalable les techniques et approches envisagées. Cette étude pilote doit se focaliser sur les trois étapes décrites précédemment dans un ensemble cohérent. Dans la littérature, certains articles traitent d'un seul des aspects de l'étude pilote. Le plus souvent il s'agit d'une étude pilote limitée à la tentative et au test d'amplification d'ADN à partir de fèces ou de poils (par exemple BAYES *et al.* 2000, LATHUILLIÈRE *et al.* 2001). Néanmoins, les trois étapes devront être testées (voir Chapitre 7, p.87). A l'issue de cette étude pilote complète, on pourra décider si l'approche non-invasive est adéquate, si elle peut être conduite à grande échelle ou si des points doivent être révisés. Ceci doit être fait en gardant à l'esprit que l'on n'aura vraisemblablement pas autant de renseignements sur les individus qu'en les capturant. En effet, la capture permet de relever des mesures (*e.g.* taille, poids) que l'on ne peut obtenir en prélevant des échantillons non-invasifs.

Par ailleurs, le statut reproducteur ou le niveau de stress peuvent être mesurés à partir des hormones contenus dans les fèces (FOLEY *et al.* 2001) ou de poils (KOREN *et al.* 2002). Il n'en reste pas moins que dans certains cas, l'échantillonnage non-invasif constitue un procédé souvent obligatoire en biologie de la conservation. Il doit donc être pris comme un échantillonnage alternatif quand les autres méthodes d'échantillonnage ne sont pas souhaitées ou réalisables.

DÉVELOPPEMENT FUTUR DES MÉTHODES NON-INVASIVES ET IMPLICATION EN BIOLOGIE DE LA CONSERVATION

Tout d'abord à court terme, les logiciels développés ici devront évoluer et être complétés par de nouvelles possibilités et avancées méthodologiques. Par exemple, dans le programme GEMINI, l'approche pour déterminer le nombre de répétitions de génotypage est différente de celle proposée par TABERLET *et al.* (1996) (voir FIGURE 3.1 p.42). L'approche de TABERLET *et al.* (1996) consiste en une détermination en deux temps du génotype d'un échantillon pour certains échantillons (principalement ceux qui seront homozygotes, pour lesquels sept réplifications doivent être réalisées). Ceci peut augmenter le temps nécessaire à l'analyse des échantillons et augmenter les risques de contamination. L'approche utilisée dans GEMINI nécessite une seule étape car tous les échantillons sont génotypés avec le même nombre de répétitions de PCR. Une perspective est d'intégrer l'approche de TABERLET *et al.* (1996) dans GEMINI et ainsi de pouvoir tester et comparer les deux approches pour savoir quelle stratégie donne les résultats de génotypage les moins biaisés tout en minimisant le nombre de réplifications d'amplification.

Les données de GARNIER *et al.* (2001) ont permis d'illustrer et de montrer que GEMINI permettait de prédire, avec une assez bonne adéquation, l'évolution du pourcentage d'identification correcte en fonction du nombre de répétitions de génotypage (Article V). En revanche, cette illustration ne constitue qu'un exemple particulier d'étude de population. Le groupe ne comptait que 35 individus et l'échantillonnage a consisté en un prélèvement de un à trois échantillons par individu. Il serait intéressant de renouveler l'étude réalisée dans l'Article V sur un cas plus général, c'est-à-dire un cas où la population est plus importante et où l'échantillonnage est également plus important de façon à prélever plusieurs échantillons d'origine inconnue. L'ultime but de cette orientation est de mener une étude multidisciplinaire, alliant le travail de terrain (échantillonnage), le travail génétique (génotypage fiable des échantillons), et une approche orientée vers la biologie informatique (analyse des données avec le développement de nouveaux outils). C'est en

développant un tel projet que l'on pourra démontrer les potentialités des approches présentées dans ce manuscrit.

GIMLET propose différentes méthodes pour estimer la taille de population notamment grâce à l'approche par capture-recapture. Un élargissement des modèles utilisés doit être une priorité dans le futur. Des applications ont déjà été menées (voir par exemple MOWAT et PAETKAU 2002) mais ils faut faire attention aux modèles utilisés. En effet, les études utilisant les dispositifs de « capture de poils » n'ont utilisé que des modèles en populations fermées (c'est-à-dire sans migration, sans décès ni naissance). Or, la plupart des populations, sur une échelle de temps plus ou moins importante, sont en fait ouvertes puisque des échanges entre cette population et « l'extérieur » sont possibles durant la période d'étude et des morts et naissances peuvent intervenir si la période d'étude est assez longue. BOULANGER et MCLELLAN (2001), par exemple, ont montré que l'hypothèse de fermeture des populations est très importante dans l'estimation de la taille de population et ont développé une méthode basée sur l'identification d'une population « noyau » à l'intérieur de la grille d'étude (les probabilités de capture des individus échantillonnés à l'intérieur de ce noyau sont moins soumises à la violation de fermeture de la population que pour l'extérieur du noyau) puis sur l'estimation de la taille de population à partir de ce noyau. Il est donc nécessaire, comme pour de nombreuses applications, de développer des modèles spécifiques de populations ouvertes (voir POLLOCK *et al.* 1990, SCHWARTZ et SEBER 1999) en prenant en compte le non-respect des hypothèses de fermeture de populations, la quasi-absence d'influence du dispositif (non-invasif) sur la probabilité de capture (pas ou peu de réponse comportementale), les variabilités de probabilité de capture âge-dépendant (hauteur des fils barbelés, comportement de marquage). En dehors de l'estimation de la taille de population, l'orientation peut être aussi prise vers l'estimation des taux de survie, de fécondité et de dispersion (LEBRETON *et al.* 1992, SCHWARTZ et SEBER 1999). En parallèle, les approches utilisant les courbes de raréfaction peuvent aussi être développées car elles présentent un potentiel non négligeable pour l'estimation de taille de population (voir Article VI).

L'avenir des approches discutées dans ce mémoire est d'autant plus prometteur que de nouvelles méthodes d'analyse des génotypes microsatellites et d'estimation de paramètres biologiques (*e.g.* taille de population, taille efficace, migration) sont maintenant développées. Les méthodes de maximum de vraisemblance (déterminant les paramètres de telle façon qu'ils maximisent la vraisemblance des données observées suivant un certain modèle), les méthodes bayésiennes (déterminant une distribution de probabilité pour le paramètre en utilisant les données et des connaissances *a priori* sur la distribution) et les méthodes de coalescence (modélisant le patron des relations ancestrales entre les allèles, qui peut apporter des informations sur les paramètres) permettent d'obtenir des informations détaillées sur des paramètres évolutifs (comme le taux de mutation), historiques ou actuels (comme la taille efficace de population et ses variations ou le taux de migration). Ces approches permettent en général d'estimer les paramètres avec plus de précision que des méthodes basées sur des statistiques (F_{ST} , H) car elles utilisent la totalité des informations sur les fréquences des allèles et non une fraction de celles-ci (BEAUMONT et BRUFORD 1998, LUIKART et ENGLAND 1999). Néanmoins, les méthodes basées sur les statistiques F_{ST} afin de déterminer la dispersion et les flux de gènes restent des outils de choix (voir par exemple GOUDET *et al.* 2002, NEIGEL 2002).

Les développements des tests d'assignation (WASER et STROBECK 1998, DAVIES *et al.* 1999, WASER *et al.* 2001 ; *e.g.* PAETKAU *et al.* 1995, CORNUET *et al.* 1999, PRITCHARD *et al.* 2000) sont aussi d'une grande importance en biologie de la conservation pour détecter les migrations et les flux géniques (WASER *et al.* 2001), mais aussi détecter les cas de braconnage ou de commerce

illégal d'animaux (voir par exemple MARSHALL *et al.* 1999, MANEL *et al.* 2002). La dispersion et les flux géniques sont des paramètres cruciaux en conservation car ils sont la clé de la colonisation de nouveaux habitats, de la restauration naturelle de populations (démographiquement par le recrutement et génétiquement par l'apport de « nouveaux gènes »). Il est donc primordial de poursuivre les efforts de développement des méthodes d'assignation et de tester celles-ci (voir MAUDET *et al.* 2002).

Les microsatellites, associés à l'échantillonnage non-invasif, permettent l'individualisation des animaux à partir d'un grand nombre d'échantillons récoltés. Ceci permet le suivi de chaque individu dans l'espace et le temps. Le suivi individuel et ses apports au niveau de la détermination et la compréhension des patrons de traits d'histoire de vie des populations sont d'autant plus importants que le comportement s'inscrit maintenant comme un volet essentiel de la biologie de la conservation (SUTHERLAND 1998). Quelles sont les structures sociales et quelles sont les relations avec les structures génétiques (BEAUMONT et BRUFORD 1998)? Quels est le comportement de dispersion dans des habitats fragmentés? Quelles sont les capacités de dispersion d'un individu? Que deviennent les individus que l'on a ré-introduits dans une population? se reproduisent-ils? se sont-ils intégrés dans un groupe? Autant de questions qui doivent être résolues et qui le peuvent grâce aux microsatellites et à l'échantillonnage non-invasif (*e.g.* MORIN *et al.* 1994a, CONSTABLE *et al.* 2001).

L'acquisition de données individuelles peut être complexe car elle est triple :

- acquisition des données génotypiques intégrant un maximum de loci informatifs (afin de diminuer la probabilité d'identité),
- acquisition des données spatiales en se servant des nouvelles technologies de positionnement GPS (*Global Positioning System*),
- acquisition des données temporelles.

Ceci implique le développement d'outils puissants de gestion des données spatialisées (*Système d'Information Géographique*), temporelles et génétiques, mais surtout de traitements et d'interprétation de celles-ci.

Bien évidemment, l'utilisation des méthodes d'échantillonnage non-invasif ne s'applique pas, en génétique, uniquement aux marqueurs microsatellites. L'ADN mitochondrial, évoluant (en terme de taux de mutation ponctuelle) généralement plus vite que l'ADN nucléaire, est ainsi un outil très important en génétique de la conservation à l'échelle de la population ou de l'espèce. Son séquençage ou la détermination de son polymorphisme (*e.g.* RFLP) permet, par exemple, d'identifier les espèces avant d'appliquer les microsatellites sur les échantillons ou des aliments contenus dans les fèces (*e.g.* Article I, KOHN *et al.* 1995, LUCCHINI *et al.* 2002), d'identifier l'hybridation entre espèces (*e.g.* PILGRIM *et al.* 1998 sur le loup et le coyote), de déterminer les structures et la diversité génétique des populations (*e.g.* RANDI *et al.* 1995). D'autre part, l'analyse des séquences d'ADN mitochondrial permet d'identifier les unités de conservation, les populations sources pour les translocations et de résoudre des incertitudes taxonomiques, qui sont d'importance en génétique de la conservation (*e.g.* GOLDBERG 1997, OAKENFULL *et al.* 2000; voir Annexe B). Néanmoins, dans le cadre des méthodes non-invasives l'utilisation de l'ADN mitochondrial est essentielle pour tester l'amplification d'ADN (voir étude pilote, Chapitre 7, p.87) à partir d'échantillons non-invasifs.

Il faut bien être conscient que l'apport de nouvelles techniques et méthodes pour étudier les populations (structure, diversité) et les individus (comportement) doit s'inscrire dans un contexte

plus large de la biologie de la conservation. Si en biologie des populations, le temps n'est pas un facteur limitant, ceci est différent pour la biologie de la conservation. En effet, la biologie de la conservation est une discipline de crise (SOULÉ 1985), et pour faire face à la diminution rapide (principalement d'origine humaine) de la biodiversité et des milieux naturels, il est impératif d'agir rapidement. Les méthodes d'échantillonnage non-invasif strict pourraient permettre dans certains cas d'obtenir un grand nombre d'échantillons dans un temps limité, en tout cas qui sera moins important que celui requis pour la mise en place d'autres stratégies d'échantillonnage. L'apport des nouvelles technologies, comme les puces à ADN (voir GERHOLD *et al.* 1999), pourront peut être permettre une étude rapide avec un très grand nombre de marqueurs pouvant être étudiés en même temps. Cependant, cette technologie est encore au stade exploratoire au niveau de l'analyse des grands jeux de données et on peut s'interroger sur la sensibilité de ces méthodes vis à vis de la possible application des puces à ADN sur des échantillons non-invasifs (ADN dégradé et en petite quantité ; voir GERHOLD *et al.* 1999).

La génétique a de nombreuses implications et applications en biologie de la conservation (voir Annexe B et FRANKHAM *et al.* 2002) à plusieurs échelles et plusieurs niveaux (populations sauvages ou captives, niveau spécifique ou populationnel, compréhension des mécanismes d'extinction et prédiction d'extinction). La génétique et la biologie de la conservation s'inscrivent aussi dans un contexte disciplinaire plus large, où les processus évolutifs (spéciation, sélection, adaptation), écologiques (succession et diversité écologiques, structure de populations, diversité génétique, taille des populations) et systématiques (particularité des espèces, phylogénie) doivent être intégrés (BOWEN 1999). Ces trois points de vues mènent souvent à une trichomie des priorités de conservation : diversité écologique, diversité génétique et particularité phylogénétique. Deux orientations sont privilégiées dans ce contexte : une conservation orientée vers l'écosystème et une conservation orientée vers l'espèce ou la population (SOULÉ 1991). Mais la ségrégation de ces deux orientations est exagérée (SOULÉ 1991) car elle sont intimement liées. L'exemple le plus frappant est dans le développement des techniques génétiques de clonage.

En effet, certains auteurs pensent que le clonage doit avoir une place importante dans la conservation d'espèces (LANZA *et al.* 2000). Mais pourquoi conserver une espèce pour pouvoir la cloner et ré-introduire des populations locales dans 100 ans, si au cours de cette période tout son habitat a disparu (VRIJENHOEK 1989) ? De même, si protéger un milieu naturel c'est aussi protéger les populations qui en font partie (WESTERN *et al.* 1989), la préservation de petits fragments peut être inefficace si la plupart des populations sont en très faible effectif, souffrent de l'isolation, d'effets délétères dus à la consanguinité, d'effets de la stochasticité démographique et environnementale. Ces facteurs, s'ils ne sont pas détectés à temps, seront fatals pour ces populations. On voit donc l'implication des deux niveaux mais il est souvent difficile d'appréhender, en même temps et entièrement, les deux dans un contexte de conservation (WESTERN *et al.* 1989). En revanche, WILSON (2000) précise que la compréhension de l'écologie des communautés passe par la connaissance des espèces, de leur systématique (phylogénie et « cartographie des biotopes ») et de leur histoire naturelle.

Cette vision dichotomique et centrée sur la biologie est trop simple pour être réaliste car, pour des raisons évidentes, d'autres facteurs clés doivent être pris en compte (WESTERN *et al.* 1989, MOROWITZ 1991, SOULÉ 1991). Les sociétés humaines, leurs traditions, leur économie en relation avec les populations animales et végétales et les milieux doivent tenir une place de choix dans la définition des programmes de conservation par les conservationnistes (non par les biologistes) spécialement dans les pays en voie de développement où la conservation doit être économiquement rentable. La génétique et la biologie en général peuvent apporter beaucoup à la conservation des milieux naturels et des espèces les habitants. Ces considérations amènent

à souligner l'incohérence d'une séparation des parties scientifique (fondamentale) et appliquée. L'interaction entre les deux est indispensable pour une bonne dynamique de la biologie de la conservation. En effet, le fait de réduire le rôle de la biologie à une production de conseils pour des cas concrets de conservation (WESTERN 1989) est néfaste. Il est évident que ce rôle doit être rempli par la biologie mais les biologistes doivent être intégrés dans les projets de conservation pour éviter certains écueils. On voit trop souvent des stratégies de conservation élaborées à partir de quelques phrases de rapports scientifiques, ce qui peut mener à un échec des opérations de conservation. D'autre part, la biologie « fondamentale » doit aussi se servir des projets de conservation (par exemple les réintroductions) comme des expériences en grandeur nature pour tester la pertinence de certaines hypothèses sur des cas réels. Cette synergie entre l'aspect scientifique fondamental et l'application des résultats de ces études doit être le point central des nouvelles avancées en biologie de la conservation.

Enfin pour finir, ne tombons pas dans la frénésie qui a mené l'Archipel des Galàpagos (GRENIER 2001) à une détérioration importante de son milieu : un mélange d'avidité de connaissances biologiques et de cupidité mercantiliste sans borne. Quand l'homme aura compris que l'on peut comprendre et jouir de la nature sans la dégrader, ce sera peut être un petit pas pour l'homme mais un grand pas pour la planète.

Références

- ALLEN, M., ENGSTRÖM, A., MEYERS, S., HANDT, O., SOLDEEN, T., VON HAESSELER, A., PÄÄBO, S., et GYLLENSTEN, U. (1998). Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps : sensivity and matching probabilities. *Journal of Forensic Science*, 43 : 453–364.
- ALLEN, P., AMOS, W., POMEROY, P., et TWISS, S. (1995). Microsatellite variation in grey seals (*Halichoerus grypus*) shows evidence of genetic differentiation between two british colonies. *Molecular Ecology*, 4 : 653–662.
- AMOS, W., SCHLÖTTERER, C., et TAUTZ, D. (1993). Social structure of pilot whales revealed by analytical DNA profiling. *Science*, 260 : 670–672.
- ASHLEY, M. et DOW, B. (1994). The use of microsatellite analysis in population biology : Background, methods and potential applications. In SCHIERWATER, B., STREIT, B., WAGNER, G., et DESALLE, R., editors, *Molecular Ecology and Evolution : Approaches and Applications*, pages 185–202. Birkhäuser Verlag, Basel.
- AVISE, J. (1989). A role for molecular genetics in the recognition and conservation of endangered species. *Trends in Ecology and Evolution*, 4(9) : 279–281.
- BALDING, D. (1998). Forensic applications of microsatellites. In GOLDSTEIN, D. et SCHLÖTTERER, C., editors, *Microsatellites : Evolution and Applications*, pages 198–210. Oxford University Press, Oxford.
- BARBAULT, R. (1997). *Biodiversité, Introduction à la Biologie de la Conservation*. Les Fondamentaux. Hachette, Paris.
- BARRATT, E., GURNELL, J., MALRKY, G., DEAVILLE, R., et BRUFORD, M. (1999). Genetic structure of fragmented populations of red squirrels (*Sciurus vulgaris*) in the UK. *Molecular Ecology*, 8(S55-S63).
- BAVERSTOCK, P. (1988). Applications of molecular genetic techniques in zoology. *Australian Journal of Zoology*, 1 : 1–13.
- BAVERSTOCK, P. et MORITZ, C. (1990). Sampling design. In HILLIS, D. et MORITZ, C., editors, *Molecular Systematics*, pages 13–24. Sinauer Associates, Sunderland.
- BAYES, M., SMITH, K., ALBERTS, S., ALTMANN, J., et BRUFORD, M. (2000). Testing the reliability of microsatellites typing from faecal DNA in the savannah baboon. *Conservation Genetics*, 1 : 173–176.
- BEAUMONT, M. (1999). Detecting population expansion and decline using microsatellites. *Ge-*

- netics*, 153 : 2013–2029.
- BEAUMONT, M. et BRUFORD, M. (1998). Microsatellites in conservation genetics. In GOLDSTEIN, D. et SCHLÖTTERER, C., editors, *Microsatellites, Evolution and Applications*, pages 165–182. Oxford University Press, Oxford.
- BECKMANN, J. et WEBER, J. (1992). Survey of human and rat microsatellites. *Genomics*, 12 : 627–631.
- BELKHIR, K., BORSA, P., CHIKHI, L., GOUDET, J., et BONHOMME, F. (1996). GENETIX 3.07, windows software for population genetics.
- BIRCH, D., KOLMODIN, L., LAIRD, W., MCKINNEY, N., WONG, J., YOUNG, K., ZANGENBERG, G., et ZOCCOLI, M. (1996). Simplified hot start PCR. *Nature*, 381 : 445–446.
- BIRSTEIN, V., DOUKARIS, P., SORKIN, B., et DESALLE, R. (1998). Population aggregation analysis of three caviar-producing species of sturgeons and implications for the species identification of black caviar. *Conservation Biology*, 12(4) : 766–775.
- BLEARS, M., DE GRANDIS, S., et TREVORS, J. (1998). Amplified fragment length polymorphism (AFLP) : a review of the procedure and its application. *Journal of Industrial Microbiology and Biotechnology*, 21 : 99–114.
- BOESCH, C. et BOESCH-ACHERMANN, H. (2000). *The Chimpanzees of the Tai Forest*. Oxford University Press, New York.
- BOULANGER, J. et MCLELLAN, B. (2001). Closure violation in DNA-based mark-recapture estimation of grizzly bear populations. *Canadian Journal of Zoology*, 79 : 642–651.
- BOWEN, B. (1999). Preserving genes, species, or ecosystems? healing the fractured foundations of conservation policy. *Molecular Ecology*, 8 : S5–S10.
- BRADLEY, B., BOESCH, C., et VIGILANT, L. (2000). Identification and redesign of human microsatellite markers for genotyping wild chimpanzee (*Pan troglodytes verus*) and gorilla (*Gorilla gorilla*) DNA from faeces. *Conservation Genetics*, 1 : 289–292.
- BROWSTEIN, M., CARPTEN, J., et SMITH, J. (1996). Modulation of non-templated nucleotide addition by Taq DNA polymerase : primer modifications that facilitate genotyping. *BioTechniques*, 20(6) : 1004–1010.
- BRUFORD, M., HANOTTE, O., BROOKFIELD, J., et BURKE, T. (1992). Single locus and multilocus DNA fingerprinting. In HOELZEL, A., editor, *Molecular genetic analysis of populations : a practical approach*, pages 225–269. IRL Press/University of Oxford, Oxford.
- BRUFORD, M. et WAYNE, R. (1993). Microsatellites and their application to population genetics studies. *Current Opinions in Genetics and Developpement*, 3 : 939–943.
- CALLEN, D., THOMPSON, A., SHEN, Y., PHILLIPS, H., RICHARDS, R., MULLEY, J., et SUTHERLAND, G. (1993). Incidence and origin of 'null' alleles in the (AC)_n microsatellite markers. *American Journal of Human Genetics*, 52 : 922–927.
- CAUGHLEY, G. et GUNN, A. (1996). *Conservation Biology in Theory and Practice*. Blackwell Science, Cambridge.
- CERCUEIL, A., BELLEMAIN, E., et MANEL, S. (Soumis). PARENTE : software package for parentage analysis. *Journal of Heredity*.

- CHAKRABORTY, R., MEAGHER, T., et SMOUSE, P. (1988). Parentage analysis with genetic markers in natural populations. i. the expected proportion of offspring with unambiguous paternity. *Genetics*, 118 : 527–536.
- CHAMBERS, G. et MACAVOY, E. (2000). Microsatellites : consensus and controversy. *Comparative Biochemistry and Physiology, Part B.*, 126 : 455–476.
- CIOFI, C. et BRUFORD, M. (1999). Genetic structure and gene flow among komodo dragon populations inferred by microsatellite loci analysis. *Molecular Ecology*, 8 : S17–S30.
- CLAYTON, T., WHITAKER, J., SPARKES, R., et GILL, P. (1998). Analysis and interpretation of mixed forensic stains using DNA STR profiling. *Forensic Science International*, 91 : 55–70.
- CONSTABLE, J., ASHLEY, M., GOODALL, J., et PUSEY, A. (2001). Noninvasive paternity assignment in Gombe chimpanzees. *Molecular Ecology*, 10 : 1279–1300.
- CORNUET, J.-M., PIRY, S., LUIKART, G., ESTOUP, A., et SOLIGNAC, M. (1999). New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics*, 153 : 1989–2000.
- CROZIER, R. (1992). Genetic diversity and the agony of choice. *Biological Conservation*, 61 : 11–15.
- DANZMANN, R. (1997). Probmax : a computer program for assigning unknown parentage in pedigree analysis from known genotypic pools of parents and progeny. *Journal of Heredity*, 88(4) : 333.
- DAVIES, N., VILLABLANCA, F., et RODERICK, G. (1999). Determining the source of individuals : multilocus genotyping in nonequilibrium population genetics. *Trends in Ecology and Evolution*, 14(1) : 17–21.
- DIAMOND, J. (1989). Overview of recent extinctions. In WESTERN, D. et PEARL, M., editors, *Conservation for the Twenty-first Century*, pages 37–41. Oxford University Press, New York.
- DIMMICK, W., GHEDOTTI, M., GROSE, M., MAGLIA, A., MEINHARDT, D., et PENNOCK, D. (1999). The importance of systematic biology in defining units of conservation. *Conservation Biology*, 13(3) : 653–660.
- DOUGLAS, J., SKOL, A., et BOEHNKE, M. (2002). Probability of detection of genotyping errors and mutations as inheritance inconsistencies in nuclear-family data. *American Journal of Human Genetics*, 70 : 487–495.
- EDWARDS, A., CIVITELLO, A., HAMMOND, H., et CASKEY, C. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics*, 49 : 746–756.
- EHRENFELD, D. (1987). Editorial. *Conservation Biology*, 1(1) : 6–7.
- EHRlich, P. et WILSON, E. (1991). Biodiversity studies : science and policy. *Science*, 253 : 758–761.
- EISEN, J. (1998). Mechanistic basis for microsatellite instability. In GOLDSTEIN, D. et SCHLÖTTERER, C., editors, *Microsatellites, Evolution and Applications*, pages 34–48. Oxford University Press, Oxford.

- ELLEGREN, H. (1991). DNA typing of museum birds. *Nature*, 354 : 113.
- EPPLEN, J., AMMER, H., EPPLEN, C., KAMMERBAUER, C., MITREITER, R., ROEWER, L., SCHWAIGER, W., STEIMLE, V., ZISCHLER, H., ALBERT, E., ANDREAS, A., BEYERMANN, B., MEYER, W., BUITKAMP, J., NANDA, I., SCHMID, M., NÜRNBERG, P., PENA, S., PÖCHE, H., SPRECHER, W., SCHARL, M., WEISING, K., et YASSOURIDIS, A. (1991). Oligonucleotide fingerprinting using simple repeat motifs : a convenient, ubiquitously applicable method to detect hypervariability for multiple purposes. In BURKE, T., DOLF, G., JEFFREYS, A., et WOLFF, R., editors, *DNA Fingerprinting : Approaches and Applications*, pages 50–69. Birkhäuser Verlag, Basel.
- ERNEST, H., PENEDO, M., MAY, B., SYVANEN, M., et BOYCE, W. (2000). Molecular tracking of mountain lions in the Yosemite valley in California : genetic analysis using microsatellites and faecal DNA. *Molecular Ecology*, 9 : 433–441.
- EVETT, I. et WEIR, B. (1998). *Interpreting DNA evidence : Statistical Genetics for Forensic Scientists*. Sinauer Associates, Sunderland.
- FERNANDO, P., EVANS, B., MORALES, J., et MELNICK, D. (2001). Electrophoresis artefacts - a previously unrecognized cause of error in microsatellite analysis. *Molecular Ecology Notes*, 1 : 325–328.
- FOLEY, C., PAPAGEORGE, S., et WASSER, S. (2001). Noninvasive stress and reproductive measures of social and ecological pressures in free ranging African elephants. *Conservation Biology*, 15(4) : 1134–1142.
- FORAN, R., CROOKS, K., et MINTA, S. (1997a). Species identification from scats : an unambiguous genetic method. *Wildlife Society Bulletin*, 25(4) : 835–839.
- FORAN, R., MINTA, S., et HEINEMEYER, K. (1997b). DNA-based analysis of hair to identify species and individuals for population research and monitoring. *Wildlife Society Bulletin*, 25(4) : 840–847.
- FORBES, S. et BOYD, D. (1996). Genetic variation of naturally colonizing wolves in the central Rocky Mountains. *Conservation Biology*, 10(4) : 1082–1090.
- FORBES, S. et HOGG, J. (1999). Assessing population structure at high levels of differentiation : microsatellites comparisons of bighorn sheep and large carnivores. *Animal Conservation*, 2 : 223–233.
- FRANCISCO, L., LANGSTON, A., MELLERSH, C., NEAL, C., et OSTRANDER, E. (1996). A class of highly polymorphic tetranucleotide repeats for canine genetic mapping. *Mammalian Genome*, 7 : 359–362.
- FRANKHAM, R., BALLOU, J., et BRISCOE, D. (2002). *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge.
- FRANKLIN, I. (1980). Evolutionary change in small populations. In SOULÉ, M. et WILCOX, B., editors, *Conservation Biology : an Evolutionary-Ecological Perspective*, pages 135–150. Sinauer, Sunderland.
- FRANTZEN, M., SILK, J., FERGUSON, J., WAYNE, R., et KONH, M. (1998). Empirical evaluation of preservation methods for faecal DNA. *Molecular Ecology*, 7 : 1423–1428.
- GAGNEUX, P., BOESCH, C., et WOODRUFF, D. (1999). Female reproductive strategies, pa-

- ternity and community structure in wild west African chimpanzees. *Animal Behaviour*, 57 : 19–32.
- GAGNEUX, P., BOESCH, C., et WOODRUFF, S. (1997). Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Molecular Ecology*, 6 : 861–868.
- GARNIER, J., BRUFORD, M., et GOOSSENS, B. (2001). Mating system and reproductive skew in the black rhinoceros. *Molecular Ecology*, 10 : 2031–2041.
- GARZA, J. et WILLIAMSON, E. (2001). Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, 10 : 305–318.
- GERHOLD, D., RUSHMORE, T., et CASKEY, C. (1999). DNA chips : promising toys have become powerful tools. *Trends in Biochemical Science*, 24 : 168–173.
- GERLOFF, U., HARTUNG, B., FRUTH, B., HOHMANN, G., et TAUTZ, D. (1999). Intra-community relationships, dispersal pattern and paternity success in a wild living community of bonobos (*Pan paniscus*) determined from DNA analysis of faecal samples. *Proceedings of the Royal Society of London, Serie B*, 266 : 1189–1195.
- GERLOFF, U., SCHLÖTTERER, C., RASSMANN, K., RAMBOLD, I., HOHMANN, G., FRUTH, B., et TAUTZ, D. (1995). Amplification of hypervariable simple sequence repeats (microsatellites) from excremental DNA of wild living bonobos (*Pan paniscus*). *Molecular Ecology*, 4 : 515–518.
- GESE, E. et RUFF, R. (1997). Scent-marking by coyotes, *Canis latrans*, the influence of social and ecological factors. *Animal Behaviour*, 54 : 1155–1166.
- GILL, P., SPARKES, R., FEREDAY, L., et WERRETT, D. (2000a). Report of the european network of forensic science institutes (ensfi) : formulation and testing of principles to evaluate STR multiplexes. *Forensic Science International*, 108 : 1–29.
- GILL, P., SPARKES, R., et KIMPTON, C. (1997). Development of guidelines to designate alleles using an STR multiplex system. *Forensic Science International*, 89 : 185–197.
- GILL, P., SPARKES, R., PINCHIN, R., CLAYTON, T., WHITAKER, J., et BUCKLETON, J. (1998). Interpreting simple STR mixtures using allele peak areas. *Forensic Science International*, 91 : 41–53.
- GILL, P., WHITAKER, J., FLAXMAN, C., BROWN, N., et BUCKLETON, J. (2000b). An investigation of the rigor of interpretation rules for STRs derived from less than 100pg of DNA. *Forensic Science International*, 112 : 17–40.
- GILPIN, M. et SOULÉ, M. (1986). Minimum viable populations : the processes of population extinction. In SOULÉ, M., editor, *Conservation Biology : the Science of Scarcity and diversity*, pages 13–34. Sinauer Associates, Sunderland.
- GINOT, F., BORDELAIS, I., NGUYEN, S., et GYAPAY, G. (1996). Correction of some genotyping errors in automated fluorescent microsatellite analysis by enzymatic removal of one base overhangs. *Nucleic Acids Research*, 24(3) : 540–541.
- GIRMAN, D., MILLS, M., GEFFEN, E., et WAYNE, R. (1997). A molecular genetic analysis of social structure, dispersal, and interpack relationships of the African wild dog (*Lycaon pictus*). *Behavioral Ecology and Sociobiology*, 40 : 187–198.

- GOLDBERG, T. (1997). Inferring the geographic origins of 'refugee' chimpanzees in Uganda from mitochondrial DNA sequences. *Conservation Biology*, 11(6) : 1441–1446.
- GOLDSTEIN, D., ROEMER, G., SMITH, D., REICH, D., BERGMAN, A., et WAYNE, R. (1999). The use of microsatellite variation to infer population structure and demographic history in natural model system. *Genetics*, 151 : 797–801.
- GOOSSENS, B., CHIKHI, L., TABERLET, P., WAITS, L., et ALLAINÉ, D. (2001). Microsatellite analysis of genetic variation among and within alpine marmot populations in the French Alps. *Molecular Ecology*, 10.
- GOOSSENS, B., CHIKHI, L., UTAMI, S., DE RUITER, J., et BRUFORD, M. (2000). A multi-samples, multi-extracts approach for microsatellite analysis of faecal samples in an arboreal ape. *Conservation Genetics*, 1 : 157–162.
- GOOSSENS, B., WAITS, L., et TABERLET, P. (1998). Plucked hair samples as a source of DNA : reliability of dinucleotide microsatellite genotyping. *Molecular Ecology*, 7 : 1237–1241.
- GOTTELLI, D., SILLERO-ZUBIRI, C., APPLEBAUM, G., ROY, M., GIRMAN, D., GARCIA-MORENO, J., OSTRANDER, E., et WAYNE, R. (1994). Molecular genetics of the most endangered canid : the Ethiopian wolf *Canis simensis*. *Molecular Ecology*, 3 : 301–312.
- GOUDET, J., PERRIN, N., et WASER, P. (2002). Tests for sex-biased dispersal using biparentally inherited genetic markers. *Molecular Ecology*, 11 : 1103–1114.
- GRENIER, C. (2001). Les Galàpagos, la conservation en échec. *Biofutur*, 211 : 38–41.
- GUTIÉRREZ-ESPELETA, G., HEDRICK, P., KALINOWSKI, S., GARRIGAN, D., et BOYCE, W. (2001). Is the decline of desert bighorn sheep from infectious disease the result of low MHC variation. *Heredity*, 86 : 439–450.
- GUTIÉRREZ-ESPELETA, G., KALINOWSKI, S., BOYCE, W., et HEDRICK, P. (1999). Genetic variation and population structure in desert bighorn sheep : implications for conservation. *Conservation Genetics*, 1 : 3–15.
- HAIG, S. (1998). Molecular contributions to conservation. *Ecology*, 79(2) : 413–425.
- HARRISON, R. (1989). Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends in Ecology and Evolution*, 4(1) : 6–11.
- HEDRICK, P., GUTIÉRREZ-ESPELETA, G., et LEE, R. (2001a). Founder effect in an island population of bighorn sheep. *Molecular Ecology*, 10 : 851–857.
- HEDRICK, P., LEE, R., et PARKER, K. (2000a). Major histocompatibility complex (MHC) variation in the endangered Mexican wolf and related canids. *Heredity*, 85 : 617–624.
- HEDRICK, P., PARKER, K., GUTIÉRREZ-ESPELETA, G., RATTINK, A., et LIEVERS, K. (2000b). Major histocompatibility complex variation in the Arabian oryx. *Evolution*, 54(6) : 2145–2151.
- HEDRICK, P., PARKER, K., et LEE, R. (2001b). Using microsatellite and MHC variation to identify species, ESUs, and MUs in the endangered sonoran topminnow. *Molecular Ecology*, 10 : 1399–1412.
- HOELZEL, A. et DOVER, G. (1991). *Molecular Genetic Ecology*. In Focus. Oxford University Press, New York.

- HOLDEN, C. (2000). DNA test to thwart tiger trade? *Science*, 287 : 963.
- HU, C.-Y., ALLEN, M., et GYLLENSTEN, U. (1992). Effect of freezing of the PCR buffer on the amplification specificity : allelic exclusion and preferential amplification of contaminating molecules. *PCR Methods and Applications*, 2 : 182–183.
- HUBERT, R., WEBER, J., SCHMITT, K., ZHANG, L., et ARNHEIM, N. (1992). A new source of polymorphic DNA markers for sperm typing : analysis of microsatellite repeats in single cells. *American Journal of Human Genetics*, 51 : 985–991.
- HUGHES, C. (1998). Integrating molecular techniques with field methods in studies of social behavior : a revolution results. *Ecology*, 79(2) : 383–399.
- HUGHES, C. et QUELLER, D. (1993). Detection of highly polymorphic microsatellite loci in a species with little allozyme polymorphism. *Molecular Ecology*, 2 : 131–137.
- IHAKA, R. et GENTLEMAN, R. (1996). R : a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5 : 299–314.
- JAMIESON, A. (1965). The genetics of transferrins in cattle. *Heredity*, 20 : 419–440.
- JANSSEN, D. (2001). Real-time PCR makes quantification really easy. *Genomics & Proteomics*, July/August : 57–59.
- JANZEN, D. (1986). The external threat. In SOULÉ, M., editor, *Conservation Biology : The Science of Scarcity and Diversity*, pages 286–303. Sinauer Associates, Sunderland.
- JARNE, P. et LAGODA, P. (1996). Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution*, 11(10) : 424–429.
- JEFFERY, K., KELLER, L., ARCESE, P., et BRUFORD, M. (2001). The development of microsatellite loci in the song sparrow, *Melospiza melodia* (Aves) and genotyping errors associated with good quality DNA. *Molecular Ecology Notes*, 1 : 11–13.
- JEFFREYS, A., WILSON, V., et THEIN, S. (1985). Individual-specific 'fingerprints' of human DNA. *Nature*, 316 : 76–77.
- JONES, C., EDWARDS, K., CASTAGLIONE, S., WINFIELD, M., SALA, F., VAN DER WIEL, C., VOSMAN, B., MATTHES, M., DALY, A., BRETTSCHEIDER, R., MAESTRI, E., MARMIROLI, N., AERT, R., VOLCKAERT, G., et KARP, A. (1998a). Reproducibility testing of AFLPs by a network of European laboratories. In KARP, A., ISAAC, P., et INGRAM, D., editors, *Molecular Tools for Screening Biodiversity*, pages 191–192. Chapman & Hall, London.
- JONES, C., EDWARDS, K., CASTAGLIONE, S., WINFIELD, M., SALA, F., VAN DER WIEL, C., VOSMAN, B., MATTHES, M., DALY, A., BRETTSCHEIDER, R., BETTINI, P., BUIATTI, P., MAESTRI, E., MARMIROLI, N., AERT, R., VOLCKAERT, G., RUEDA, J., VAZQUEZ, A., et KARP, A. (1998b). Reproducibility testing of RAPDs by a network of European laboratories. In KARP, A., ISAAC, P., et INGRAM, D., editors, *Molecular Tools for Screening Biodiversity*, pages 176–179. Chapman & Hall, London.
- KAHN, G., KANGRO, H., COATES, P., et HEATH, R. (1991). Inhibitory effects of urine on the polymerase chain reaction for cytomegalovirus DNA. *J. Clin. Pathol.*, 44 : 360–365.
- KAYS, R., GITTLEMAN, J., et WAYNE, R. (2000). Microsatellite analysis of kinkajou social organisation. *Molecular Ecology*, 9 : 743–751.

- KEBELMANN-BETZING, C., SEEGER, K., DRAGON, S., SCHMITT, G., MÖRICKE, A., SCHILD, T., HENZE, G., et BEYERMANN, B. (1998). Advantages of a new Taq DNA polymerase in multiplex PCR and time-release PCR. *BioTechniques*, 24(1) : 154–158.
- KENDALL, M. et STEWART, A. (1977). *The advanced Theory of statistics*, volume 1. Macmillan, New York.
- KOHN, M., KNAUER, F., STOFFELA, A., SCHRÖDER, W., et PÄÄBO, S. (1995). Conservation genetics of the european brown bear - a study using excremental PCR of nuclear and mitochondrial sequences. *Molecular Ecology*, 4 : 95–103.
- KOHN, M. et WAYNE, R. (1997). Facts from feces revisited. *Trends in Ecology and Evolution*, 12(6) : 223–227.
- KOHN, M., YORK, E., KAMRADT, D., HAUGHT, G., SAUVAJOT, R., et WAYNE, R. (1999). Estimating population size by genotyping faeces. *Proceedings of the Royal Society of London, Serie B*, 266 : 657–663.
- KOREN, L., MOKADY, O., KARASKOV, T., KLEIN, J., KOREN, G., et GEFFEN, E. (2002). A novel method using hair for determining hormonal levels in wildlife. *Animal Behaviour*, 63 : 403–406.
- KWOK, S. et HIGUSHI, R. (1989). Avoiding false positives with PCR. *Nature*, 339 : 237–238.
- KYLE, C. et STROBECK, C. (2001). Genetic structure of north American wolverine (*Gulo gulo*) populations. *Molecular Ecology*, 10 : 337–347.
- LANZA, R., DRESSER, B., et DAMIANI, P. (2000). Cloning Noah’s ark. *American Scientist*, 87(5) : 84–89.
- LATHUILLIÈRE, M., MÉNARD, N., GAUTIER-HION, A., et CROUAU-ROY, B. (2001). Testing the reliability of noninvasive genetic sampling by comparing analyses of blood and fecal samples in barbary macaques (*Macaca sylvanus*). *American Journal of Primatology*, 55 : 151–158.
- LAUNHARDT, K., EPPLER, C., EPPLER, J., et WINKLER, P. (1998). Amplification of microsatellites adapted from human systems in faecal DNA of wild Hanuman langurs (*Presbytis entellus*). *Electrophoresis*, 19 : 1356–1361.
- LEAKEY, R. et LEWIN, R. (1995). *The Sixth Extinction : Biodiversity and its Survival*. Phoenix, London.
- LEBRETON, J., BURNHAM, K., CLOBERT, J., et ANDERSON, D. (1992). Modeling survival and testing biological hypotheses using marked animals : a unified approach with case studies. *Ecological Monograph*, 62 : 67–118.
- LEOPOLD, A. (1949). *A Sand County Almanac*. Oxford University Press, New York.
- LEVINSON, G. et GUTMAN, G. (1987). Slipped-strand mispairing : a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution*, 4(3) : 203–221.
- LITT, M. et LUTY, J. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics*, 44 : 397–401.
- LUBJUHN, T., SCHWAIGER, F.-W., et EPPLER, J. (1994). The analysis of simple repeat loci as applied in evolutionary and behavioral science. In SCHIERWATER, B., STREIT, B.,

- WAGNER, G., et DESALLE, R., editors, *Molecular Ecology and Evolution : Approaches and Applications*, pages 33–44. Birkhäuser Verlag, Basel.
- LUCCHINI, V., FABBRI, E., MARUCCO, F., RICCI, S., BOITANI, L., et RANDI, E. (2002). Noninvasive molecular tracking of colonizing wolf (*Canis lupus*) packs in the Western Italian Alps. *Molecular Ecology*, 11(5) : 857–868.
- LUIKART, G., CORNUET, J., et ALLENDORF, F. (1999). Temporal changes in allele frequencies provide estimates for population bottleneck size. *Conservation Biology*, 13(3) : 523–530.
- LUIKART, G. et ENGLAND, P. (1999). Statistical analysis of microsatellite DNA data. *Trends in Ecology and Evolution*, 14(7) : 253–256.
- LYNCH, M. (1988). Estimation of relatedness by DNA fingerprinting. *Molecular Biology and Evolution*, 5(5) : 584–599.
- MAEHARA, Y., ODA, S., et SUGIMACHI, K. (2001). The instability within : problems in current analyses of microsatellite instability. *Mutation Research*, 461 : 249–263.
- MANEL, S., BERTHIER, P., et LUIKART, G. (2002). Detecting poaching : identifying the origin of individuals using bayesian assignment tests and multi-locus genotyping. *Conservation Biology*, in press.
- MARKLUND, S., SANDBERG, K., et ANDERSSON, L. (1996). Forensic tracing of horse identities using urine samples and DNA markers. *Animal Biotechnology*, 7(2) : 145–153.
- MARSHALL, T., SLATE, J., KRUK, L., et PEMBERTON, J. (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, 7 : 639–655.
- MARSHALL, T., SUNNUCK, P., SPALTON, J.-A., GRETH, A., et PEMBERTON, J. (1999). Use of genetic data for conservation management : the case of Arabian oryx. *Animal Conservation*, 2 : 269–278.
- MAUDET, C., LUIKART, G., DUBRAY, D., VON HARDENBERG, A., et TABERLET, P. (0000). Strong seasonal effects and low winter genotyping error rates when analysing microsatellite DNA from wild ungulates faeces. *Molecular Ecology*, 00(00) : 000.
- MAUDET, C., MILLER, C., BASSANO, B., BREITENMOSER-WÜRSTEN, C., GAUTHIER, D., OBEXER-RUFF, G., MICHALLET, J., TABERLET, P., et LUIKART, G. (2002). Microsatellite DNA and recent statistical methods in wildlife conservation management : applications in alpine ibex [*Capra ibex (ibex)*]. *Molecular Ecology*, 11 : 421–436.
- MAY, B. (1992). Starch gel electrophoresis of allozymes. In HOELZEL, A., editor, *Molecular Genetic Analysis of Populations : A Practical Approach*, The Practical Approach Series, pages 1–27. Oxford University Press, Oxford.
- MCDONALD, D. et POTTS, W. (1997). DNA microsatellites as genetic markers at several scales. In *Avian Molecular Evolution and Systematics*, pages 29–49. Academic Press.
- MELLERSH, C., LANGSTON, A., ACLAND, G., FLEMING, M., RAY, K., WIEGAND, N., FRANCISCO, L., GIBBS, M., AGUIRRE, G., et OSTRANDER, E. (1997). A linkage map of the canine genome. *Genomics*, 46 : 326–336.
- MILLER, C., JOYCE, P., et WAITS, L. (2002). Assessing allelic dropout and genotype reliability using maximum likelihood. *Genetics*, 160 : 357–366.

- MILLIGAN, B., LEEBENS-MACK, J., et STRAND, A. (1994). Conservation genetics : beyond the maintenance of marker diversity. *Molecular Ecology*, 3 : 423–435.
- MILLS, L., CITTA, J., LAIR, K., SCHWARTZ, M., et TALLMON, D. (2000). Estimating animal abundance using non-invasive DNA sampling : promise and pitfalls. *Ecological Applications*, 10(1) : 283–294.
- MORIN, P., CHAMBERS, K., BOESCH, C., et VIGILANT, L. (2001). Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Molecular Ecology*, 10 : 1835–1844.
- MORIN, P., MOORE, J., CHAKRABORTY, R., JIN, L., GOODALL, J., et WOODRUFF, D. (1994a). Kin selection, social structure, gene flow, and evolution of chimpanzees. *Science*, 265 : 1193–1201.
- MORIN, P., WALLIS, J., MOORE, J., CHAKRABORTY, R., et WOODRUFF, D. (1993). Non-invasive sampling and DNA amplification for paternity exclusion, community structure, and phylogeography in wild chimpanzees. *Primates*, 34(3) : 347–356.
- MORIN, P., WALLIS, J., MOORE, J., et WOODRUFF, D. (1994b). Paternity exclusion in a community of wild chimpanzees using hypervariable simple sequence repeats. *Molecular Ecology*, 3 : 469–478.
- MORIN, P. et WOODRUFF, D. (1992). Paternity exclusion using multiple hypervariable microsatellite loci amplified from nuclear DNA of hair cells. In MARTIN, R., DIXSON, A., et WICKINGS, E., editors, *Paternity in Primates : Genetic Test and Theories*, pages 63–81. Karger, S., Basel.
- MORIN, P. et WOODRUFF, D. (1996). Noninvasive genotyping for vertebrate conservation. In SMITH, T. et WAYNE, R., editors, *Molecular Genetic Approaches in Conservation*, pages 298–313. Oxford University Press, New York.
- MORITZ, C. (1994). Applications of mitochondrial DNA analysis in conservation : a critical review. *Molecular Ecology*, 3 : 401–411.
- MOROWITZ, H. (1991). Balancing species preservation and economic considerations. *Science*, 253 : 752–754.
- MOWAT, G. et PAETKAU, D. (2002). Estimating marten population size using hair capture and genetic tagging in southeastern British Columbia. *Journal of Wildlife Management*, 8(3).
- MOWAT, G., PAETKAU, D., PROCTOR, M., et STROBECK, C. (1999). Using genetic tagging to estimate population parameters. *Cat News (IUCN Cat specialist Group Newsletter)*, 31 : 24–27.
- MOWAT, W. et STROBECK, C. (2001). Estimating population size of grizzly bears using hair capture, DNA profiling, and mark-recapture analysis. *Journal of Wildlife Management*, 64(1) : 183–193.
- MUELLER, U. et WOLFENBARGER, L. (1999). AFLP genotyping and fingerprinting. *Trends in Ecology and Evolution*, 14(10) : 389–394.
- MULLIS, K. et FALOONA, F. (1987). Specific synthesis of DNA in vitro via a polymerase-catalysed chain reaction. *Methods in Enzymology*, 155 : 335–350.
- MURPHY, M., WAITS, L., et KENDALL, K. (2000). Quantitative evaluation of fecal drying

- methods for brown bear DNA analysis. *Wildlife Society Bulletin*, 28(4) : 951–957.
- MYERS, N. (1979). *The Sinking Ark, A New Look at the Problem of Disappearing Species*. Pergamon Press, Oxford.
- NAVIDI, W., ARNHEIM, N., et WATERMAN, M. (1992). A multitubes approach for accurate genotyping of very small DNA sample using PCR : statistical considerations. *American Journal of Human Genetics*, 50 : 347–359.
- NEFF, M., BROMAN, K., MELLERSH, C., RAY, K., ACLAND, G., AGUIRRE, G., ZIEGLE, J., OSTRANDER, E., et RINE, J. (1999). A second-generation genetic linkage map of the domestic dog, *Canis familiaris*. *Genetics*, 151 : 803–820.
- NEI, M. (1996). Phylogenetic analysis in molecular evolutionary genetics. *Annual Review of Genetics*, 30 : 371–403.
- NEIGEL, J. (2002). Is F_{st} obsolete? *Conservation Genetics*, 3(2) : 167–173.
- NICHOLS, R., BRUFORD, M., et GROOMBRIDGE, J. (2001). Sustaining genetic variation in a small population : evidence from the mauritius krestel. *Molecular Ecology*, 10 : 593–602.
- NIELSEN, R., MATTILA, D., CLAPHAM, P., et PALSBOÛLL, P. (2001). Statistical approaches to paternity analysis in natural populations and applications to the north atlantic humpback whale. *Genetics*, 157 : 1673–1682.
- OAKENFULL, E., LIM, H., et RYDER, O. (2000). A survey of equid mitochondrial DNA : implications for the evolution, genetic diversity and conservation of *Equus*. *Conservation Genetics*, 1 : 341–355.
- O'BRIEN, S. et EVERMANN, J. (1988). Interactive influence of infectious disease and genetic diversity in natural populations. *Trends in Ecology and Evolution*, 3 : 254–259.
- ORITA, M., IWAHANA, H., KANAZE, H., HAYASKI, K., et SEKIYA, T. (1989). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformational polymorphisms. *Proceedings of the National Academy of Science, USA*, 86 : 2766–2770.
- OTIS, D., BURNHAM, K., WHITE, G., et ANDERSON, D. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs*, 62.
- PAETKAU, D., AMSTRUP, S., BORN, E., CALVERT, W., DEROCHER, A., GARNER, G., MESSIER, F., STIRLING, I., TAYLOR, M., et WHIG, . (1999). Genetic structure of the world's polar bear populations. *Molecular Ecology*, 8 : 1571–1584.
- PAETKAU, D., CALVERT, W., STIRLING, I., et C., S. (1995). Microsatellite analysis of population structure in Canadian polar bear. *Molecular Ecology*, 4 : 347–354.
- PAETKAU, D., SHIELDS, G., et STROBECK, C. (1998a). Gene flow between insular, coastal and interior populations of brown bears in Alaska. *Molecular Ecology*, 7 : 1283–1292.
- PAETKAU, D. et STROBECK, C. (1994). Microsatellite analysis of genetic variation in black bear populations. *Molecular Ecology*, 3 : 489–495.
- PAETKAU, D. et STROBECK, C. (1995). The molecular basis and evolutionary history of a microsatellite null allele in bears. *Molecular Ecology*, 4 : 519–520.
- PAETKAU, D., WAITS, L., CLARKSON, P., CRAIGHEAD, L., VYSE, E., WARD, R., et STROBECK, C. (1998b). Variation in genetic diversity across the range of North American brown

- bears. *Conservation Biology*, 12(2) : 418–429.
- PALSBØLL, P. (1999). Genetic tagging : contemporary molecular ecology. *Biological Journal of the Linnean Society*, 68 : 3–22.
- PALSBØLL, P., ALLEN, J., BÉRUBÉ, M., CLAPHAM, P., FEDDERSEN, T., HAMMOND, P., HUDSON, R., JØRGENSEN, H., KATONA, S., LARSEN, A., LARSEN, F., LIEN, J., MATTILA, D., SIGURJÓNSSON, J., SEARS, R., SMITH, T., SPONER, R., STEVICK, P., et ØIEN, N. (1997). Genetic tagging of humpback whales. *Nature*, 388 : 767–769.
- PALUMBI, S., MARTIN, A., KESSING, B., et MCMILLAN, W. (1991). Detecting population structure using mitochondrial DNA. In HOELZEL, A., editor, *Genetic Ecology of Whales and Dolphins (Rep. Int. Whal. Commn., Special Issue 13)*, pages 271–278. Cambridge University Press, Cambridge, UK.
- PARKER, P., SNOW, A., SCHUG, M., BOOTON, G., et FUERST, P. (1998). What molecules can tell us about populations : choosing and using a molecular marker. *Ecology*, 79(2) : 361–382.
- PARSONS, K. (2001). Reliable microsatellite genotyping of dolphin DNA from faeces. *Molecular Ecology Notes*, 1 : 341–344.
- PEMBERTON, J., SLATE, J., BANCROFT, D., et BARRETT, J. (1995). Nonamplifying alleles at microsatellite loci : a caution for parentage and population studies. *Molecular Ecology*, 4 : 249–252.
- PILGRIM, K., BOYD, D., et FORBES, S. (1998). Testing for wolf-coyote hybridization in the Rocky Mountains using mitochondrial DNA. *Journal of Wildlife Management*, 62(2) : 683–689.
- PIRY, S., LUIKART, G., et CORNUET, J.-M. (1999). BOTTLENECK : a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, 90(4) : 502–503.
- POLLOCK, K., NICHOLS, J., BROWNIE, C., et HINES, J. (1990). *Statistical inference for capture-recapture experiments*, volume 107 de *Wildlife Monographs*. The Wildlife Society.
- POOLE, K., MOWAT, G., et FEAR, D. (2002). DNA-based population estimate for grizzly bears in northeastern British Columbia, Canada. *Wildlife Biology*, 7(2) : 105–115.
- POULLE, M., CARLES, L., et LEQUETTE, B. (1997). Significance of ungulates in the diet of recently settled wolves in the Mercantour Mountains (southern France). *Revue d'Ecologie (Terre et Vie)*, 52 : 357–367.
- POULLE, M.-L., LEQUETTE, B., et DAHIER, T. (1999). La recolonisation des Alpes par le loup de 1992 à 1998. *Bulletin mensuel de l'Office National de la Chasse*, 242 : 4–13.
- POWELL, J. (1994). Molecular techniques in population genetics : a brief history. In SCHIERWATER, B., STREIT, B., WAGNER, G., et DESALLE, R., editors, *Molecular Ecology and Evolution : Approaches and Applications*, pages 131–156. Birkhäuser Verlag, Basel.
- PRADEL, R. et LEBRETON, J.-D. (1993). *User Manual for Program SURGE, version 4.3*. CEFÉ/CNRS, Montpellier.
- PRADEL, R., REBOULET, A.-M., et VIALLEFONT, A. (1995). Testing hypotheses and estimating survival from capture histories with cr. *Journal of Applied Statistics*, 22 : 775–784.

- PÉREZ, T., ALBORNOZ, J., et DOMINGUEZ, A. (1998). An evaluation of RAPD fragment reproductability and nature. *Molecular Ecology*, 7 : 1347–1357.
- PRIMMER, C., KOSKINEN, M., et PIIRONEN, J. (2000). The one that did not get away : individual assignment using microsatellite data detects a case of fishing competition fraud. *Proceedings of the Royal Society of London, Serie B*, 267 : 1699–1704.
- PRITCHARD, J., STEPHENS, M., et DONNELLY, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155 : 945–959.
- PRUGNOLLE, F. et DE MEEUS, T. (2002). Inferring sex-biased dispersal from population genetic tools : review. *Heredity*, 88 : 161–165.
- QUELLER, D., STRASSMAN, J., et HUGHES, C. (1993). Microsatellites and kinship. *Trends in Ecology and Evolution*, 8(8) : 285–288.
- RANDI, E., FRANCISCI, F., et LUCCHINI, V. (1995). Mitochondrial DNA restriction-fragment-length monomorphism in the Italian wolf (*Canis lupus*) population. *J. Zool. Syst. Evol. Res.*, 33 : 97–100.
- RANDI, E., LUCCHINI, V., CHRISTENSEN, M., MUCCI, N., FUNK, S., DOLF, G., et LOESCHCKE, V. (2000). Mitochondrial DNA variability in Italian and East European wolves : detecting the consequences of small population size and hybridization. *Conservation Biology*, 14(2) : 464–473.
- RAYMOND, M. et ROUSSET, F. (1995). GENEPOP (version 1.2) : population genetics software for exact tests and ecumenicism. *Journal of Heredity*, 86 : 248–249.
- REED, J., TOLLIT, D., THOMPSON, P., et AMOS, W. (1997). Molecular scatology : the use of molecular genetic analysis to assign species, sex and individual identity to seal faeces. *Molecular Ecology*, 6 : 225–234.
- ROBINSON, J. et HARRIS, S. (1999). Amplified fragment length polymorphisms and microsatellites : a phylogenetic perspective. In GILLET, E., editor, *Which DNA Marker for Which Purpose ? Final Comendium of the Research Project Development, optimisation and validation of molecular tools for assessment of biodiversity in forest trees*. URL <http://web-doc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm>.
- ROJAS, M. (1992). The species problem and the conservation : what are we protecting. *Conservation Biology*, 6 : 170–178.
- RUTH, J. et FAIN, S. (1993). The 'individualization' of large North American mammals. In PENA, S., CHAKRABORTY, R., EPPLER, J., et JEFFREYS, A., editors, *DNA Fingerprinting : State of the Science*, pages 429–436. Birkhäuser Verlag, Basel.
- SAIKI, R., GELFAND, D., STOFFEL, S., SCHARF, S., HIGUCHI, R., HORN, G., MULLIS, K., et ERLICH, H. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239 : 487–491.
- SANGER, F., NICKLEN, S., et COULSON, A. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Science, USA*, 74 : 5463–5467.
- SARKAR, G. et SOMMER, S. (1990). More light on PCR contamination. *Nature*, 347 : 340–341.
- SAY, L., PONTIER, D., et NATOLI, E. (1999). High variation in multiple paternity of domestic cat (*Felis catus* L.) in relation to environmental conditions. *Proceedings of the Royal*

- Society of London, Serie B*, 266 : 2071–2074.
- SAY, L., PONTIER, D., et NATOLI, E. (2001). Influence of oestrus synchronization on male reproductive success in the domestic cat (*Felis catus* L.). *Proceedings of the Royal Society of London, Serie B*, 268 : 1049–1053.
- SCHLÖTTERER, C. et PEMBERTON, J. (1994). The use of microsatellites for genetic analysis of natural populations. In SCHIERWATER, B., STREIT, B., WAGNER, G., et DESALLE, R., editors, *Molecular Ecology and Evolution : Approaches and Applications*, pages 203–214. Birkhäuser Verlag, Basel.
- SCHMIDT, T., HUMMEL, S., et HERRMANN, B. (1995). Evidence of contamination in PCR laboratory disposables. *Naturwissenschaften*, 82 : 423–431.
- SCHNEIDER, S., ROESSLI, D., et EXCOFFIER, L. (2000). ARLEQUIN ver. 2.000 : a software for population genetics data analysis.
- SCHWARTZ, C. et SEBER, A. (1999). A review of estimating animal abundance. *Statistical Science*, 14(4) : 427–456.
- SCHWARTZ, M., TALLMON, D., et LUIKART, G. (1998). Review of DNA-based census and effective population size estimators. *Animal Conservation*, 1 : 293–299.
- SCOTT, J. et STRAUSS, N. (2000). A review of current methods in DNA fingerprinting. In SAMSON, R. et PITT, J., editors, *Integration of Modern Taxonomic Methods for Penicillium and Aspergillus Classification*, pages 209–224. Harwood Academic Publishers, Amsterdam.
- SEARS, P. (1935). *Deserts on the March*. University of Oklahoma Press, Norman, 4th (1980) edition.
- SEGELBACHER, G. (2002). Noninvasive genetic analysis in birds : testing reliability of feather samples. *Molecular Ecology Notes*, 2(3) : 367–369.
- SIMBERLOFF, D. (1988). The contribution of population and community biology to conservation science. *Annual Review of Ecology and Systematics*, 19 : 473–511.
- SLOANE, M., SUNNUCKS, P., ALPERS, D., BEHEREGARAY, L., et TAYLOR, A. (2000). Highly reliable genetic identification of individual northern hairy-nosed wombats from single remotely collected hairs : a feasible censuring method. *Molecular Ecology*, 9 : 1233–1240.
- SOULÉ, M. (1980). Thresholds for survival : maintaining fitness and evolutionary potential. In SOULÉ, M. et WILCOX, B., editors, *Conservation Biology : an Evolutionary-Ecological Perspective*, pages 111–124. Sinauer, Sunderland.
- SOULÉ, M. (1985). What is conservation biology? *Bioscience*, 35 : 727–734.
- SOULÉ, M. (1986). Conservation biology and the 'real world'. In SOULÉ, M., editor, *Conservation Biology, the Science of Scarcity and Diversity*, pages 1–12. Sinauer Associates, Sunderland.
- SOULÉ, M. (1987). History of the society for conservation biology : how and why we got here. *Conservation Biology*, 1(1) : 4–5.
- SOULÉ, M. (1991). Conservation : tactics for constant crisis. *Science*, 253 : 744–750.
- SPELLENBERG, I. (1992). *Evaluation and Assessment for Conservation*. Chapman & Hall, London.
- STEBBINS, G. (1992). Why should we conserve species and wildlands? In FIEDLER, P. et

- JAIN, S., editors, *Conservation Biology : the Theory and Practice of Nature Conservation, Preservation and Management*, pages 393–410. Chapman & Hall, London/New York.
- SUNNUCK, P. (2000). Efficient genetic markers for population biology. *Trends in Ecology and Evolution*, 15(5) : 199–203.
- SUTHERLAND, W. (1998). The importance of behavioral studies in conservation biology. *Animal Behaviour*, 56 : 801–809.
- TABERLET, P. (1996). The use of mitochondrial DNA control region sequencing in conservation genetics. In SMITH, T. et WAYNE, R., editors, *Molecular Genetic Approaches in Conservation*, pages 125–142. Oxford University Press, New York.
- TABERLET, P. et BOUVET, J. (1992). Bear conservation genetics. *Nature*, 358 : 197.
- TABERLET, P. et BOUVET, J. (1994). Mitochondrial DNA polymorphism, phylogeography, and conservation genetics of the brown bear *Ursus arctos* in Europe. *Proceedings of the Royal Society of London, Serie B*, 255 : 195–200.
- TABERLET, P., CAMARRA, J., GRIFFIN, S., UHRÈS, E., HANOTTE, O., WAITS, L., DUBOIS-PAGANON, C., BURKE, T., et BOUVET, J. (1997). Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Molecular Ecology*, 6 : 869–876.
- TABERLET, P., GRIFFIN, S., GOOSSENS, B., QUESTIAU, S., MANCEAU, V., ESCARAVAGE, N., WAITS, L., et BOUVET, J. (1996). Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, 24(16) : 3189–3194.
- TABERLET, P. et LUIKART, G. (1999). Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society*, 68 : 41–55.
- TABERLET, P., LUIKART, G., et GEFFEN, E. (2001). New methods for obtaining and analyzing genetic data from free-ranging carnivores. In GITTELMAN, J., FUNK, S., McDONALD, D., et WAYNE, R., editors, *Carnivore Conservation*, pages 313–334. Cambridge University Press, Cambridge.
- TABERLET, P., MATTOCK, H., DUBOIS-PAGANON, C., et BOUVET, J. (1993). Sexing free-ranging brown bears *Ursus arctos* using hairs found in the field. *Molecular Ecology*, 2 : 399–403.
- TABERLET, P. et WAITS, L. (1998). Non-invasive genetic samplig. *Trends in Ecology and Evolution*, 13(1) : 26–27.
- TABERLET, P., WAITS, L., et LUIKART, G. (1999). Non-invasive genetic sampling : look before you leap. *Trends in Ecology and Evolution*, 14(8) : 323–327.
- TAUTZ, D. (1993). Notes on definition and nomenclature of tandemly repetitive DNA sequences. In PENA, D., CHAKRABORTY, R., EPPLEN, J., et JEFFREYS, A., editors, *DNA Fingerprinting : State of the Science*, pages 21–28. Birkhäuser Verlag, Basel.
- TAUTZ, D. et RENTZ, M. (1984). Simple sequence are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research*, 12(10) : 4127–4138.
- TAYLOR, A., SHERWIN, W., et WAYNE, R. (1994). Genetic variation of microsatellites loci in a bottlenecked species : the nothern hairy-nosed wombat *lasiorhinus kreftii*. *Molecular Ecology*, 3 : 277–290.

- TIKEL, D., BLAIR, D., et MARSH, D. (1996). Marine mammals faeces as a source of DNA. *Molecular Ecology*, 5 : 456–457.
- VALIÈRE, N. (en preparation). Population size estimation and biases associated with genotyping errors.
- VALIÈRE, N. (en preparation). The use of multitubes approach for microsatellite genotyping : number of PCR replications and number of loci.
- VALIÈRE, N. (2002a). GIMLET : a computer program for analysing genetic individual identification data. *Molecular Ecology Notes*, 2(3) : 377–379.
- VALIÈRE, N. (2002b). L'étude de l'expansion du loup (*Canis lupus*) en France : apport des méthodes non-invasives et des marqueurs moléculaires. In CHAPRON, G. et MOUTOU, F., editors, *XXIIIème Colloque Francophone de Mammalogie*, pages 97–101, Ecole National Vétérinaire d'Alfort. SFPEM.
- VALIÈRE, N., BERTHIER, P., MOUCHIROUD, D., et PONTIER, D. (2002). GEMINI : software for testing the effects of genotyping errors and multitubes approach for individual identification. *Molecular Ecology Notes*, 2 : 83–86.
- VALIÈRE, N., FUMAGALLI, L., GIELLY, L., MIQUEL, C., LEQUETTE, B., POULLE, M.-L., WEBER, J., ARLETTAZ, R., et TABERLET, P. (Accepté). Wolf recolonisation of france and switzerland inferred from non-invasive genetic sampling over a period of 10 years. *Animal Conservation*.
- VALIÈRE, N., PONTIER, D., GARNIER, J., et GOOSSENS, B. (Soumis). Computer simulations to predict identification consistency using noninvasive sampling and the multitubes approach. *Conservation Genetics*.
- VALIÈRE, N. et TABERLET, P. (2000). Urine collected in the field as a source of DNA for species and individual identification. *Molecular Ecology*, 9 : 2150–2152.
- VALSECCHI, H., HALE, P., CORKERON, P., et AMOS, W. (2002). Social structure in migrating humpback whales (*Megaptera novaeangliae*). *Molecular Ecology*, 11 : 507–518.
- VIGILANT, L. (1999). An evaluation of techniques for the extraction and amplification of DNA from naturally shed hairs. *Biological Chemistry*, 380 : 1329–1331.
- VILÀ, C., AMORIM, I., LEONARD, J., POSADA, D., CASTROVIEJO, S., PETRUCCI-FONSECA, F., CRANDALL, K., ELLEGREN, S., et WAYNE, R. (1999). Mitochondrial DNA phylogeography and population history of the grey wolf *Canis lupus*. *Molecular Ecology*, 8 : 2089–2103.
- VILÀ, C., URIOS, V., et CATROVIELO, J. (1994). Use of faeces for scent marking in Iberian wolves (*Canis lupus*). *Canadian Journal of Zoology*, 72 : 374–377.
- VOS, P., HOGERS, R., BLEEKER, M., REIJANS, M., VAN DE LEE, T., HORNES, M., FRIJTERS, A., POT, J., PELEMAN, J., KUIPER, M., et ZABEAU, M. (1995). AFLP : a new technique for DNA fingerprinting. *Nucleic Acid Research*, 23(21) : 4407–4414.
- VRIJENHOEK, R. (1989). Population genetics and conservation. In WESTERN, D. et PEARL, M., editors, *Conservation for the Twenty-first Century*, pages 89–98. Oxford University Press, New York.
- WAITS, J. et LEBERG, P. (1999). Advances in the use of molecular markers for studies of

- population size and movement. *Transactions of the North American Wildlife Society*, pages 191–201.
- WAITS, J. et LEBERG, P. (2000). Biases associated with population estimation using molecular tagging. *Animal Conservation*, 3 : 191–199.
- WAITS, L., LUIKART, G., et TABERLET, P. (2001). Estimating the probability of identity among genotypes in natural populations : cautions and guidelines. *Molecular Ecology*, 10 : 249–256.
- WALSH, P., METZGER, D., et HIGUSHI, R. (1991). Chelex100 as a medium for simple extraction of DNA for PCR-based typing forensic material. *BioTechniques*, 10 : 506–513.
- WASER, P. et STROBECK, C. (1998). Genetic signatures of interpopulation dispersal. *Trends in Ecology and Evolution*, 13 : 43–44.
- WASER, P., STROBECK, C., et PAETKAU, D. (2001). Estimating interpopulation dispersal rates. In GITTELMAN, J., FUNK, S., MCDONALD, D., et WAYNE, R., editors, *Carnivore Conservation*, pages 484–497. Cambridge University Press, Cambridge.
- WASSER, S., HOUSTON, C., KOEHLER, G., CADD, G., et FAIN, S. (1997). Techniques for application of faecal DNA methods to field studies of ursids. *Molecular Ecology*, 6 : 1091–1097.
- WEBER, J. et MAY, P. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics*, 44 : 388–396.
- WEIR, B. et COCKERHAM, C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6) : 1358–1370.
- WESTERN, D. (1989). Conservation biology. In WESTERN, D. et PEARL, M., editors, *Conservation for the Twenty-first Century*, pages 31–36. Oxford University Press, New York.
- WESTERN, D., PEARL, M., PIMM, S., WALKER, B., ATKINSON, I., et WOODRUFF, D. (1989). An agenda for conservation action. In WESTERN, D. et PEARL, M., editors, *Conservation for the Twenty-first Century*, pages 304–323. Oxford University Press, New York.
- WESTNEAT, D. et WEBSTER, M. (1994). Molecular analysis of kinship in birds : Interesting questions and useful techniques. In SCHIERWATER, B., STREIT, B., WAGNER, G., et DESALLE, R., editors, *Molecular Ecology and Evolution : Approaches and Applications*, pages 91–128. Birkhäuser Verlag, Basel.
- WHITE, G. et BURNHAM, K. (1999). Program MARK : Survival estimation from populations of marked animals. *Bird Study*, 46(Supplement) : 120–138.
- WILSON, E. (2000). On the future of conservation biology. *Conservation Biology*, 14(1) : 1–3.
- WOODROFFE, R. (1999). Managing disease threats to wild mammals. *Animal Conservation*, 2 : 185–193.
- WOODRUFF, D. (1989). The problems of conserving genes and species. In WESTERN, D. et PEARL, M., editors, *Conservation for the Twenty-first Century*, pages 76–88. Oxford University Press, New York.
- WOODS, J., PAETKAU, D., LEWIS, D., MCLELLAN, B., PROCTOR, M., et STROBECK, C.

- (1999). Genetic tagging of free-ranging black and brown bears. *Wildlife Society Bulletin*, 27(3) : 616–627.
- YOKOTA, M., TATSUMI, N., TSUDA, I., TAKUBO, T., et HIYOSHI, M. (1998). DNA extraction from human urinary sediments. *Journal of Clinical Laboratory Analysis*, 12 : 88–91.
- ZANE, L., BARGELLONI, L., et PATARNELLO, T. (2002). Strategies for microsatellite isolation : a review. *Molecular Ecology*, 11 : 1–16.
- ZHANG, Y.-W., MORIN, P., RYDER, O., et ZHANG, Y.-P. (2001). A set of human tri- and tetra-nucleotide microsatellite loci useful for population analyses in gorillas (*Gorilla gorilla gorilla*) and orangutans (*Pongo pygmaeus*). *Conservation Genetics*, 2 : 391–395.
- ZHU, Y., STRASSMANN, J., et QUELLER, D. (2000). Insertions, substitutions, and the origin of microsatellites. *Genet. Res. Camb.*, 76 : 227–236.

THÈSE

présentée devant l'Université Claude Bernard - LYON I

pour l'obtention du
Diplôme de Doctorat
(arrêté du 30 mars 1992)

Soutenue publiquement le 18 juin 2002

par

Nathaniel VALIÈRE

**Amélioration et optimisation des méthodes
non-invasives et des marqueurs microsatellites en
Biologie des Populations et de la Conservation**

VOLUME 2 : ANNEXES ET ARTICLES

Jury

Rapporteurs : Stéphane AULAGNIER
Michael BRUFORD

Président : Dominique MOUCHIROUD (co-directrice)

Examineurs : Catherine HÄNNI
Dominique PONTIER (co-directrice)
Roger PRADEL
François SARRAZIN

UMR CNRS 5558
Laboratoire de Biométrie et Biologie Evolutive
Université Claude Bernard LYON I - Bât 711
43, boulevard du 11 novembre 1918
69622 Villeurbanne

Mis en page avec la classe thloria.

Volume 2 : Annexes et Articles

Annexes

Annexe A

Identification individuelle chez le loup

Sommaire

A.1 Etude du loup en France	124
A.1.1 Situation	124
A.1.2 Pourquoi une étude génétique ?	124
A.2 Identification individuelle dans la population de loup	125
A.2.1 Choix du système de loci microsatellites	125
A.2.2 Sélection des loci « fiables »	126
A.2.3 Probabilité d'identité	127
A.2.3.1 Calcul de la probabilité d'identité pour les loups	127
A.2.3.2 Application des loci aux échantillons de fèces	128
A.2.4 Sexage des individus	130

A.1 ETUDE DU LOUP EN FRANCE

A.1.1 Situation

Jusqu'au 19^e siècle, le loup (*Canis lupus*) peuplait la plupart des territoires européens et était présent sur 90% du territoire français (POULLE *et al.* 1995). Diabolisé dans les contes pour enfants, compétiteur du chasseur et de l'éleveur, ce grand prédateur a été chassé et a été l'objet de persécutions multiples : battues, pièges, empoisonnement, autant de moyens incités par un soutien financier et des privilèges pour les chasseurs de loups. C'est ainsi que le loup a été exterminé par l'homme en Europe durant plusieurs siècles et surtout à partir du 19^e siècle (voir BREITENMOSER 1998). L'effet conjugué de cette extermination, du déclin des populations d'ongulés, de la déforestation et du développement de l'agriculture (POULLE et LONGCHAMP 1997, BREITENMOSER 1998), a fait diminuer les effectifs de loup en Europe, menant à sa complète disparition dans certaines régions au 19^e et 20^e siècles. En France, le loup a définitivement disparu entre 1930 et 1939 (DE BEAUFORT 1987, POULLE *et al.* 1995).

En revanche, le loup n'a pas totalement été exterminé en Italie où une population (devenant isolée des autres populations européennes) subsiste dans les Apennins. L'effectif de cette population est descendu jusqu'à 100 individus (ZIMEN et BOITANI 1975) et depuis les années 1970, grâce à une protection de l'espèce, cette population connaît une expansion géographique importante vers le nord (BOITANI 1984, 1992). Le retour naturel du loup en France est vraisemblablement due à cette expansion. La colonisation des Alpes françaises par les loups est en cours, situation unique chez les Grands Mammifères. Ce modèle biologique peut nous permettre de comprendre davantage les processus d'expansion de tels prédateurs sur des territoires « vacants ». Les questions abordées peuvent concerner deux aspects de la dynamique d'expansion :

- la description démographique et génétique de la population : composition et taille des meutes, identification, sexe et parenté des individus, niveau de consanguinité, diversité génétique, dispersion, flux génique ;
- la répartition et l'évolution spatiale et temporelle de la population de loups : dynamique de colonisation de la population (axes de propagation, vitesse et type de progression, fondation de la population, évolution de la diversité génétique).

A.1.2 Pourquoi une étude génétique ?

Si les outils classiques d'études de la répartition spatiale et temporelle (*e.g.* capture, radiopistage) ne peuvent être utilisés en ce qui concerne le loup (décision politique et difficulté de capture notamment), l'outil génétique paraît alors très important tant pour la résolution de l'étude spatiale que pour l'étude génétique des populations. En effet, grâce au développement de nouvelles techniques en écologie moléculaire (méthodes non-invasives, amplification d'ADN par PCR), il est désormais possible d'extraire assez d'ADN à partir d'échantillons récoltés sur le terrain (par exemple des crottes) pour réaliser des analyses génétiques (voir VALIÈRE 2002) : identification spécifique (ADN mitochondrial) et identification individuelle (microsatellites). L'identification spécifique peut permettre d'attribuer une espèce à un échantillon par exemple lors d'une analyse de régime alimentaire pour ne retenir que les excréments de l'espèce étudiée (FORAN *et al.* 1997) ou pour prouver la présence d'une espèce (en l'occurrence le loup) dans une région donnée. Les génotypes multilocus (identification individuelle) sont d'une meilleure résolution et on peut repérer les déplacements d'un individu ou estimer les degrés de parenté entre les membres de la population. La taille de population peut par ailleurs être estimée (voir KOHN *et al.* 1999). Cette estimation est impossible avec le système actuel de suivi, basé seulement sur les indices de présence du fait de l'absence d'identification individuelle.

A.2 IDENTIFICATION INDIVIDUELLE DANS LA POPULATION DE LOUP

L'identification individuelle chez le loup à partir d'échantillons de fèces et aux marqueurs microsatellites a nécessité de nombreuses mises au point et essais. Les différentes étapes décrites dans les sections suivantes sont résumées dans la FIGURE A.1.

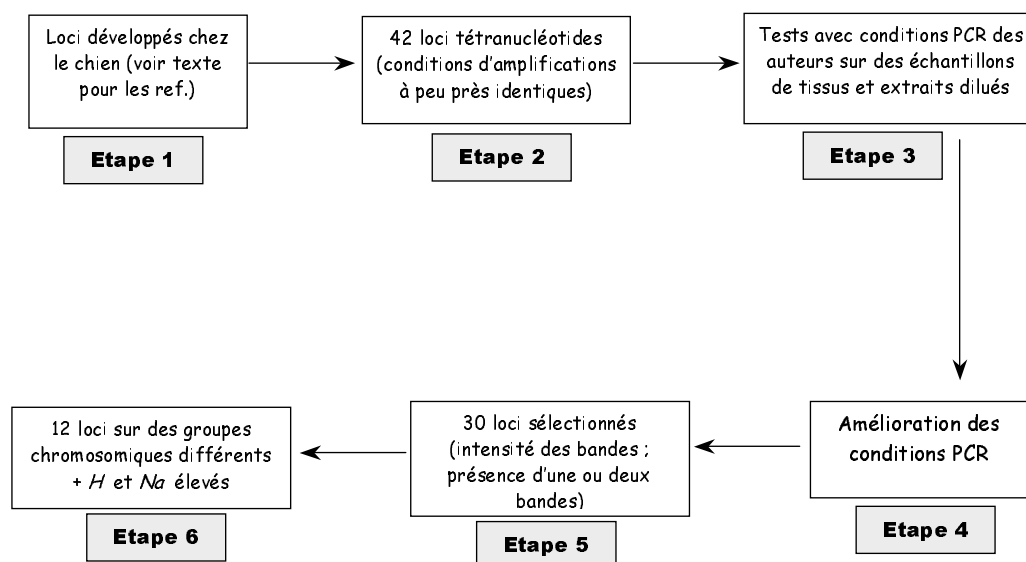


FIGURE A.1 – Les différentes étapes de mises au point de l'identification individuelle chez le loup à partir d'échantillons de fèces et de marqueurs microsatellites.

A.2.1 Choix du système de loci microsatellites

La première étape pour développer un système de loci microsatellites à des fins d'identification individuelle est la sélection des loci. Classiquement, la sélection de loci microsatellites se réalise par clonage de fragments à partir d'une banque génomique, puis par sélection des clones qui sont susceptibles de contenir des microsatellites, et enfin par l'extraction d'ADN et le test des loci (BRUFORD et WAYNE 1993, ASHLEY et DOW 1994). Ces opérations peuvent être longues requérant plusieurs semaines voir plusieurs mois avant d'avoir un nombre suffisant de loci polymorphes. Cette démarche est nécessaire pour les espèces où aucun locus microsatellite n'a encore été déterminé où si l'on désire cloner de nouveaux loci. En revanche, si des microsatellites ont été développés pour des espèces proches, il est possible de tester directement ces loci sur l'espèce étudiée. C'est le cas pour le loup où de nombreux microsatellites ont été développé chez le chien. J'ai sélectionné, parmi les microsatellites développés chez le chien (FRANCISCO *et al.* 1996, MELLERSH *et al.* 1997, NEFF *et al.* 1999), uniquement les tétra-nucléotides (Etapes 1 et 2 dans la FIGURE A.1). En effet, ce type de microsatellites seraient moins soumis aux erreurs d'amplification (meilleure fidélité d'amplification) que les di-nucléotides par exemple (EDWARDS *et al.* 1991). Quarante deux microsatellites ont été sélectionné dont l'indice de facilité d'amplification était le plus fort (quand il était disponible) et la température d'hybridation pour les 2 amorces était d'environ 55°C. Ces loci ont été testés en utilisant les conditions PCR préconisées par les auteurs avec une température d'hybridation de 55°C (Etape 3 dans la FIGURE A.1). Ces

tests ont été menés en utilisant un échantillon de tissu musculaire de loup mort. Un échantillon de tissu a été utilisé car le tissu musculaire est susceptible de contenir assez d'ADN de bonne qualité (pour peu que l'échantillon ait été convenablement préservé) pour obtenir des résultats de génotypage fiables, à la différence des fèces où la quantité et la qualité d'ADN n'est pas assurée. Afin de s'approcher des conditions de quantité d'ADN des fèces, l'extrait d'ADN de tissu été dilué 20 à 100 fois. Suivant les résultats de génotypage pour ces échantillons, les conditions PCR ont été améliorées (augmentation ou diminution de la température d'hybridation, augmentation de la concentration de $MgCl_2$ ou de dNTP) pour chaque locus (Etape 4 dans la FIGURE A.1).

A.2.2 Sélection des loci « fiables »

Sur les 42 loci retenus, 30 ont été sélectionnés sur la base des résultats d'une part du point de vue de l'intensité des bandes obtenues et d'autre part de la netteté des bandes et l'absence de bandes multiples (Etape 5 dans la FIGURE A.1). En utilisant 13 échantillons de tissus de loups morts (7 loups italiens, 4 loups français, 1 loup suisse) provenant tous de la lignée génétique (ADN mitochondrial) italienne, le nombre d'allèles pour les loci sélectionnés a été déterminé en utilisant une amorce fluorescente pour chaque couple d'amorces d'un locus. Onze loci microsatellites ayant les nombres d'allèles et les taux d'hétérozygotie les plus importants (TABLEAU A.1) et localisés sur des groupes de liaison différents (NEFF *et al.* 1999 ; voir TABLEAU A.2) ont été retenus pour le système de loci polymorphes (Etape 6 dans la FIGURE A.1).

TABLEAU A.1 – Nombre d'allèles (N_a), nombre d'allèles ayant une fréquence supérieure à 0.1 ($N_a > 0.1$), nombre d'individus testés ($N_{bindiv.}$), taux d'hétérozygotie observée (H_{obs}), taux attendu d'hétérozygotie à l'équilibre de Hardy-Weinberg (H_{exp}), probabilité d'identité calculée avec l'équation corrigeant la taille de population ($PI_{unbiased}$) et probabilité d'identité calculée pour les frère-soeurs (PI_{sibs}).

Loci	N_a	$N_a > 0.1$	Nb indiv.	H_{obs}	H_{exp}	$PI_{unbiased}$	PI_{sibs}
FH2199	8	3	13	0,92	0,79	$3,18.10^{-02}$	$3,71.10^{-01}$
FH2161	5	4	13	0,85	0,74	$7,37.10^{-02}$	$4,08.10^{-01}$
FH2054	6	4	13	0,85	0,76	$5,88.10^{-02}$	$3,96.10^{-01}$
FH2137	9	3	13	0,85	0,78	$3,28.10^{-02}$	$3,79.10^{-01}$
FH2140	3	3	13	0,85	0,58	$2,17.10^{-01}$	$5,26.10^{-01}$
FH2096	3	3	13	0,85	0,64	$1,65.10^{-01}$	$4,81.10^{-01}$
FH2356	6	2	13	0,77	0,57	$1,36.10^{-01}$	$5,18.10^{-01}$
FH2247	6	4	13	0,69	0,69	$6,93.10^{-02}$	$4,40.10^{-01}$
PEZ17	3	3	13	0,77	0,64	$1,74.10^{-01}$	$4,84.10^{-01}$
FH2289	6	3	13	0,54	0,65	$7,75.10^{-02}$	$4,62.10^{-01}$
FH2050	5	4	13	0,46	0,7	$9,43.10^{-02}$	$4,33.10^{-01}$

Ces 11 loci porteront des fluorescences de couleur différente afin qu'ils puissent être utilisés dans un ou plusieurs multiplexes de charge sur séquenceur automatique. Le TABLEAU A.2 montre les séquences des amorces pour les 11 loci sélectionnés ainsi que le groupe chromosomique, la température d'hybridation et les couleurs de fluorescence.

TABLEAU A.2 – Informations concernant les 11 loci sélectionnés : noms des loci, groupes de liaison (Gr. Liais. ; valeur pour NEFF *et al.* 1999/MELLERSH *et al.* 1997), séquences des amorces (Seq. Amorces) 5' → 3', motifs des répétitions (si disponible dans Francisco *et al.* 1996), références (Réf. ; 1 = FRANCISCO *et al.* 1996 ; 2 = MELLERSH *et al.* 1997 ; 3 = NEFF *et al.* 1999), température d'hybridation utilisée en °C (THyb), étendue des tailles de fragments amplifiés déterminées à partir des 13 loups testés (Taille), marquage fluorescent (Coul. ; 6'FAM=Bleu ; HEX=Vert ; NED =Jaune).

Locus	Gr. Liais.	Seq. Amorces (5' → 3')	Motif	Réf.	THyb	Taille	Coul.
FH2140	CFA5/L19	GGGGAAGCCATTTTAAAGC TGACCCTCTGGCATCTAGGA	(GAAA)	1	50	132-158	6'FAM
PEZ17	L21/-	CTAAGGGACTGAACTTCTCC GTGGAACCTGCTTAAGATTC	?	3	50	196-204	6'FAM
FH2161	L32/L25	TCAGCAAGAAACCCTCCAGT CATTCCCAACGGAGACTCT	(GAAA)	1	55	239-271	6'FAM
FH2199	L13/L28	GCTGAGCACTGGGTATTGTATG TGTTACAAATTAATGTGAAATGGC	?	2	55	307-343	6'FAM
FH2054	L9/L6	GCCTTATTCATTGCAGTTAGGG ATGCTGAGTTTTGAACTTTCCC	(GATA)	1	50	145-165	NED
FH2247	non lié	TTCCCCTTACTTTTATCATAGCAT CAAATGCAGATTAGGGACACA	?	2	55	188-234	NED
FH2050	L24/L27	GTATCGAGTCCC GCGTTG CTGGGAGGTGCAAGGTTTAA	(TAAA) /(AAAC)	1	55	249-261	NED
FH2289	L11/L26	CATGGTCTCAGGATCCTAGGA CTAAGCATTCCTCTGATGGTC	?	2	55	290-310	NED
FH2096	L10/L17	CCGTCTAAGAGCCTCCAG GACAAGGTTTCCTGGTTCCA	(GAAT)	1	55	96-104	HEX
FH2137	L4/L3	GCAGTCCCTTATTCCAACATG CCCCAAGTTTTGCATCTGTT	(GAAA)	1	50	155-185	HEX
FH2356	L14/non lié	CTTGCATTCCCGCTCTCACT TCCTGAAATAGCTCCAGCGC	?	2	55	208-248	HEX

A.2.3 Probabilité d'identité

La probabilité d'identité est la probabilité que deux individus pris au hasard dans une population aient le même génotype sur un ou plusieurs loci¹⁴. Voir section 6.2.2.3 (p.84) pour des détails.

A.2.3.1 Calcul de la probabilité d'identité pour les loups

En combinant les 11 loci sélectionnés et en utilisant les fréquences alléliques calculées à partir des 13 individus analysés, les probabilités théoriques d'identité sont $PI = 4,029.10^{-10}$, $PI_{unbiased} = 1,941.10^{-12}$ et $PI_{sibs} = 1,267.10^{-04}$. L'utilisation de ses 11 loci permettra donc d'identifier sans ambiguïté les individus de la population de loup. Dans ce cas, nous serons en dessous des seuils recommandés de 0.01 afin d'estimer correctement la taille d'une population (MILLS *et al.* 2000) et dans la fourchette 0.001-0.0001 (voir WAITS *et al.* 2001) pour l'identification individuelle dans le cas de braconnage. En revanche, il faut noter, que les calculs sont réalisés à partir d'un échantillon d'individus (N=13) qui n'est pas complètement représentatif de

¹⁴Il ne faut pas confondre avec la probabilité d'appariement (*match probability*) qui est la probabilité de tirer un individu dont le génotype est identique à celui d'un individu que l'on a préalablement tiré (voir par exemple WOODS *et al.* 1999)

la population française étudiée (j'ai utilisé ces échantillons car il appartiennent tous à la même lignée génétique et qu'ils étaient les seuls échantillons de tissus disponibles). On ne peut donc prendre les valeurs ci-dessus comme celles de la population française étudiée. En revanche, vu les faibles valeurs rencontrées et le fait que la population française provient de la population italienne, on peut tout de même penser que la probabilité d'identité est assez faible pour pouvoir identifier les individus de loups en France.

A.2.3.2 Application des loci aux échantillons de fèces

Les conditions PCR développées pour les échantillons de tissus même en utilisant des extraits dilués n'étaient pas optimales pour les échantillons de fèces. En effet, en utilisant les mêmes conditions PCR sur les fèces, le faible pourcentage de réussite des amplifications pour l'ensemble des loci sélectionnés a conduit à de nouvelles mises au point, en recherchant les conditions PCR donnant de meilleurs résultats (plus forte intensité des bandes et plus fort pourcentage d'amplifications positives). En définitive, seulement sept couples d'amorces amplifiaient l'ADN de façon convenable (taux de PCR positives non nul, profils lisibles sur séquenceur) à partir d'échantillons de fèces. Pour les autres amorces, soit l'optimisation n'a pas été convaincante, soit il s'agissait d'un défaut des amorces elles-mêmes reconnu par le fournisseur. Les probabilités d'identité en utilisant ces 7 amorces deviennent alors $PI = 1,291.10^{-06}$, $PI_{unbiased} = 8,336.10^{-08}$, $PI_{sibs} = 3,248.10^{-03}$. Même si les valeurs sont plus fortes que précédemment, nous sommes encore proches des valeurs recommandées (voir plus haut).

Avec une si faible efficacité des amplifications, il était raisonnable de penser que la fiabilité des génotypes obtenus à partir des fèces devait être limitée. En effet, en répétant huit fois les amplifications par échantillon et par locus, le pourcentage de PCR positives (*i.e.* donnant un produit d'amplification compatible avec les tailles des fragments que l'on doit obtenir) est extrêmement variable selon les loci et les échantillons (TABLEAU A.3). Le taux d'erreurs par perte d'allèles est de l'ordre de 25-60% et les erreurs par faux allèles négligeables.

TABLEAU A.3 – Résultats d'amplification multiple (8 par échantillon et par locus) pour 4 échantillons (3 fèces et 1 tissu) et 7 loci donnant des résultats exploitables. %P+= pourcentage de PCR positives ; %ADO = pourcentage de perte d'allèle. Hmz = échantillon homozygote pour le locus.

Echantillons	9606-277		9605-227		9605-230		T0699-02	
Loci	%P+	%ADO	%P+	%ADO	%P+	%ADO	%P+	%ADO
FH2140	100	0	50	?	100	Hmz	100	0
PEZ17	100	25	0	-	100	62,5	100	0
FH2161	50	Hmz	12,5	Hmz?	0	-	100	0
FH2050	12,5	100	0	-	37,5	?	100	Hmz
FH2137	100	62,5	0	-	75	50	100	0
FH2096	100	0	62,5	?	100	Hmz	62,5	0
FH2054	37,5	-	50	Hmz	25	Hmz?	100	12,5

Cette faible efficacité dans le génotypage peut provenir de la dégradation de l'ADN dans les fèces, des inhibiteurs de PCR contenus dans les échantillons et enfin de la faible quantité d'ADN. Afin de quantifier l'ADN de loups dans un extrait, j'ai comparé l'intensité des bandes obtenues pour l'échantillon de fèces avec une gamme de dilution à partir d'un échantillon de tissu. En effet, le dosage direct de l'ADN dans un extrait d'échantillon de fèces est impossible (mais voir MORIN

et al. 2001 pour une nouvelle technique de dosage basée sur la PCR quantitative) car il existe dans l'extrait de l'ADN hôte (de l'individu qui a déposé la fèces) et de l'ADN bactérien de la flore intestinale. Pour estimer la quantité d'ADN dans un extrait de fèces, j'ai me suis donc basé sur la concentration d'ADN dans un extrait de tissus. J'ai comparé l'amplification de cet extrait de référence dilué plusieurs fois avec l'extrait de fèces grâce à plusieurs loci microsatellites.

Le dosage de l'extrait de tissu a été effectué au fluorimètre avec la méthode du picogreen. La gamme étalon a été réalisée avec de l'ADN de phage λ (100ng/ml) dans du tampon TE (Tris-EDTA). La gamme comporte les concentrations de 0, 0.05 ng/ μ l, 0.1 ng/ μ l, 0.2 ng/ μ l, 0.4 ng/ μ l, 0.6 ng/ μ l, 0.8 ng/ μ l et 1 ng/ μ l. Le dosage de l'ADN dans l'extrait de tissu de référence est réalisé en diluant cet extrait 25 et 125 fois et en mesurant la densité optique de ces dilutions. La FIGURE A.2 montre les résultats du dosage de ces dilutions. On peut estimer la concentration d'ADN dans l'extrait (non dilué) de tissu à environ 25 ng/ μ l.

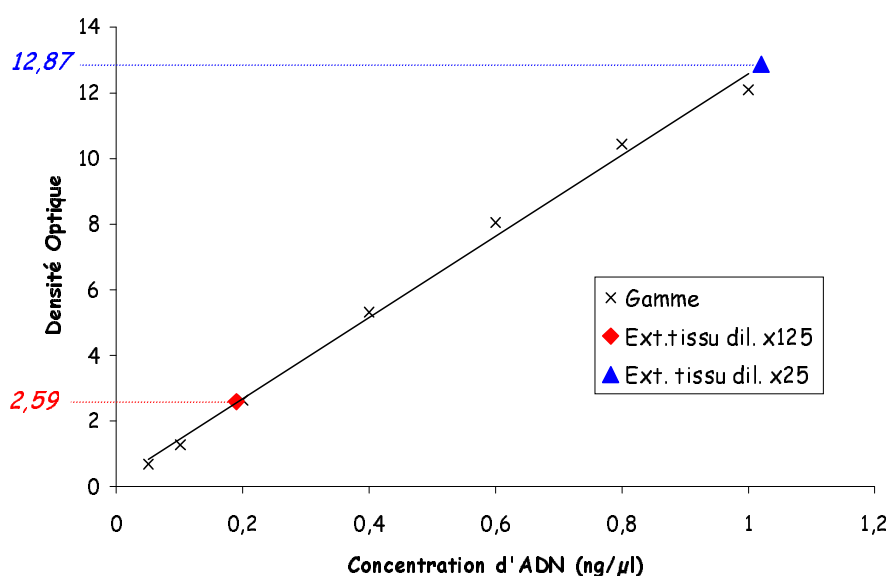


FIGURE A.2 – Dosage de la concentration de l'ADN dans un extrait de tissu. Le graphe (Densité Optique DO en fonction de la concentration d'ADN en ng/ μ l) montre les points de la gamme étalon et de l'extrait dilué 25 fois et 125 fois.

La deuxième étape consiste en l'amplification de plusieurs amorces microsatellites pour calibrer et estimer la quantité d'ADN dans un extrait de fèces par rapport à l'amplification de l'extrait de tissu. Dix paires d'amorces permettent de situer la concentration d'ADN de l'extrait de fèces en dessous de celle de l'extrait de tissu dilué 15625 fois.

On peut donc en déduire que la concentration d'ADN hôte contenue dans l'extrait de fèces de loup est d'environ 1.6 pg/ μ l soit l'équivalent d'une cellule pour 5 μ l¹⁵. Cette valeur est parmi les plus faibles par rapport à celles déterminées par MORIN *et al.* (2001) (0-2550 pg/ μ l; moyenne de 192 pg/ μ l). Ceci signifie que dans les conditions de PCR utilisées (notamment 5 μ l d'extrait par tube), nous sommes dans des conditions très défavorables en ce qui concerne les erreurs potentielles (voir TABERLET *et al.* 1996). D'après les courbes de la FIGURE II. 5, nous pouvons

¹⁵une cellule contient environ 7 pg d'ADN (TABERLET *et al.* 1996).

envisager obtenir un taux de PCR positives de 80% et un taux de perte d'allèle d'environ 60-65%, ce qui est environ les résultats obtenus pour cet échantillon (voir plus haut).

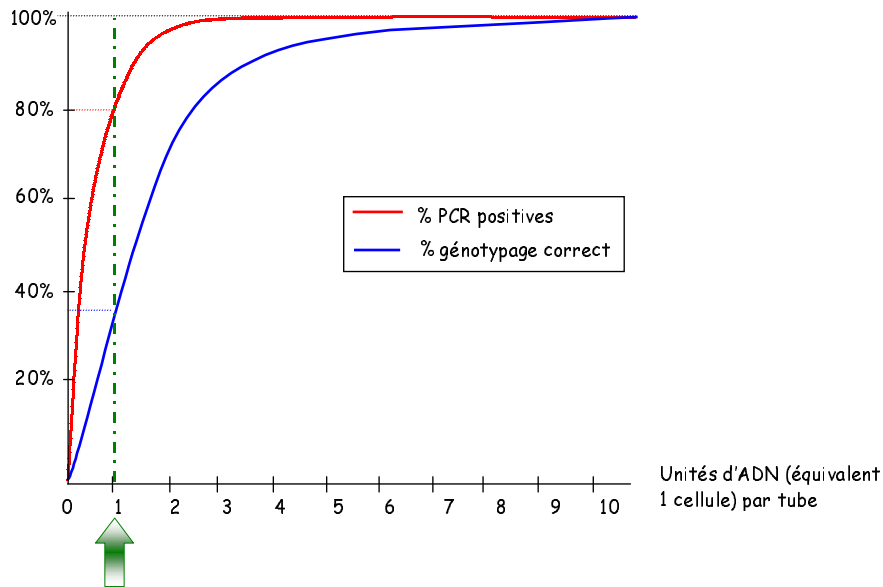


FIGURE A.3 – Résultats des simulations concernant le typage génétique d'un hétérozygote (modifié d'après TABERLET *et al.* 1996) avec la position (flèche) de l'échantillon dont la concentration a été déterminée (voir texte).

A.2.4 Sexage des individus

Dans l'étude de la population française de loup, il était nécessaire de pouvoir identifier le sexe de l'individu. Pour simplifier les expériences (c'est-à-dire en ne faisant pas de manipulations - séquençage du marqueur SRY ou du gène de l'améloglobine - différentes de l'amplification des microsatellites), j'ai décidé d'utiliser un locus microsatellite porté par le chromosome Y. A partir d'une séquence d'un microsatellite spécifique au chromosome Y donnée dans GenBank (*Accession Number* AF032441 ; Olivier, M. et Lust, G.), j'ai développé deux amorces pour pouvoir amplifier ce microsatellite. L'amplification de ce microsatellite doit donner un produit d'amplification (environ 115 paires de base) uniquement chez les mâles et aucun produit chez les femelles. Après optimisation des conditions PCR pour ces amorces et des tests d'amplification sur des loups italiens dont je connaissais le sexe, l'amplification de ce locus s'est révélée plus ou moins efficace. En effet, d'une part les échantillons de fèces contiennent peu d'ADN et souvent dégradé (voir plus haut), mais il faut préciser aussi que le microsatellite étudié ici n'est porté que par le chromosome Y. Donc que le nombre de copies de ce locus est deux fois moins important que pour les locus autosomaux utilisés précédemment. L'absence de produit peut donc être interprété soit comme l'absence du marqueur (individu femelle), soit d'un fragment présent mais non amplifié (individu mâle) car en trop petite quantité. Comme il n'existe pas de vérification pour le sexage des femelles il est impératif de faire très attention à la détermination du sexe comme par exemple de réaliser plusieurs amplifications (bien que cela ne puisse résoudre complètement le problème).

Récemment, SUNDQVIST *et al.* (2001) ont développé quatre loci microsatellites portés par le chromosome Y. Contrairement à celui que j'ai développé, ces loci sont polymorphes (de 2 à 9

allèles). Ils peuvent donc être utilisés à la fois pour la détermination du sexe, mais aussi pour l'identification individuelle des mâles chez qui apparait ces loci, où même pour étudier la structure génétiques chez les mâles (voir SUNDQVIST *et al.* 2001).

Annexe B

La biologie de la conservation : quelques précisions

Cet annexe est une version plus complète de l'introduction générale sur la biologie de la conservation

Sommaire

B.1 Extinctions et émergence de la biologie de la conservation	134
B.1.1 Diversité et extinctions	134
B.1.2 Déclin et extinctions	136
B.1.2.1 Facteurs naturels influençant les déclin et les extinctions	137
B.1.2.2 Les facteurs d'origine humaine	137
B.1.2.3 Le rôle prépondérant de la taille de population	138
B.2 L'émergence de la Conservation	141
B.2.1 Pourquoi et comment conserver ?	143
B.2.1.1 Pourquoi conserver ?	143
B.2.1.2 Les moyens de conservation	143
B.3 Le rôle de la biologie en conservation	147
B.3.1 Diagnostic et prescription	148
B.3.1.1 L'apport de la biogéographie insulaire et du concept de métapopulation .	149
B.3.1.2 Minimum Viable Population et Population Viability (Vulnerability) Analysis	150
B.3.2 Informations biologiques	151
B.4 La génétique de la conservation	152
B.4.1 Définition de la taxonomie et des unités de conservation	152
B.4.2 Diagnostiquer les problèmes génétiques et donner des priorités de conservation	155
B.4.2.1 La variabilité génétique	155
B.4.2.2 La consanguinité	156
B.4.2.3 L'effectif et la taille efficace d'une population	156
B.4.2.4 Le problème des hybrides	158
B.4.3 Obtention d'informations biologiques	159

B.1 EXTINCTIONS ET ÉMERGENCE DE LA BIOLOGIE DE LA CONSERVATION

B.1.1 Diversité et extinctions

Pour un observateur, le nombre et les formes des espèces sont toujours étonnantes, allant de microscopiques cellules procaryotes au plus grand organisme actuel, le séquoia géant (plus de 6000 tonnes). En une seule journée de prospection dans une région tempérée, une personne peut comptabiliser des centaines d'espèces différentes, des Insectes aux Mammifères en passant par les Oiseaux, et encore plus en comptabilisant les espèces végétales. Le nombre d'espèces sur une surface donnée est encore plus étonnante dans les régions tropicales, spécialement en ce qui concerne la classe des Insectes. Outre le nombre d'espèces animales et végétales, la diversité écologique est aussi incroyable. Tous les milieux ont été colonisés, même aux endroits où l'on ne pouvait suspecter la présence de formes vivantes : sources chaudes, déserts, fosses marines...

L'inventaire complet des formes biologiques est vraisemblablement loin d'être terminé et les estimations du nombre d'espèces sont très variables. Seulement 1,5 millions d'espèces seraient connues et répertoriées (MYERS 1979, WILSON 1988). La plupart des espèces vivantes ne sont pas encore découvertes et décrites et on estime le nombre total d'espèces à 5 à 10 millions (MYERS 1979) voire 5 à 30 millions (WILSON 1988)¹⁶.

Les enregistrements fossiles montrent que la diversité spécifique et la diversité écologique n'ont cessé d'augmenter globalement depuis l'apparition de la vie, il y a 3,8 milliards d'années. Le nombre de familles des groupes vivants a globalement augmenté mais de façon irrégulière car il existe une dynamique dans l'évolution, la succession de **spéciation** et d'**extinction**. La terre a ainsi connu de nombreuses phases d'**extinction de masse** plus ou moins importantes permettant une dynamique dans les radiations spécifiques au cours des temps géologiques. Par exemple, les Mammifères n'ont pu évoluer et donner toutes les formes actuelles (allant de la souris à l'éléphant en passant par l'homme) que parce que les dinosaures ont disparu à la fin du Crétacé (il y a 65 millions d'années), laissant des niches écologiques vacantes pour l'évolution des petits Mammifères (RAUP 1988). La composition spécifique du monde vivant n'a donc cessé d'être modifiée au cours des 3,8 milliards d'années depuis l'apparition de la vie sur Terre. Les causes des **cinq extinctions de masse** (voir FIGURE B.1) sont pour la plupart inconnues mais les archives fossiles montrent que ces extinctions ont eu pour conséquence la disparition de plus de 50% de familles biologiques, jusqu'à plus de 90% de certains groupes marins pour l'extinction du Permien.

Au cours des 50 000 dernières années (Pléistocène), il y eut plusieurs épisodes d'extinctions, touchant en particulier les Mammifères et les Oiseaux (CAUGHLEY et GUNN 1996) mais de façon hétérogène suivant les continents (l'Afrique étant le continent le moins touché pour les extinctions de Mammifères).

Il est très difficile de trouver un consensus sur les causes des extinctions continentales du Pléistocène (DIAMOND 1989, CAUGHLEY et GUNN 1996), mais il est possible que la conjonction de deux interprétations (bouleversement climatique et effet de la chasse et de la modification du milieu par le feu) puisse expliquer ces extinctions (MYERS 1979, p.29). En revanche, en ce qui concerne les extinctions de l'ère Holocène (après 5000 ans BP), les effets (directs et indirects) de la colonisation des Hommes sur certaines îles est fortement suspectée (DIAMOND 1989, OLSON

¹⁶La définition d'espèce utilisée dans la plupart des estimations est la définition basée sur la taxonomie ou sur le concept biologique d'espèce, bien que d'autres concepts existent.

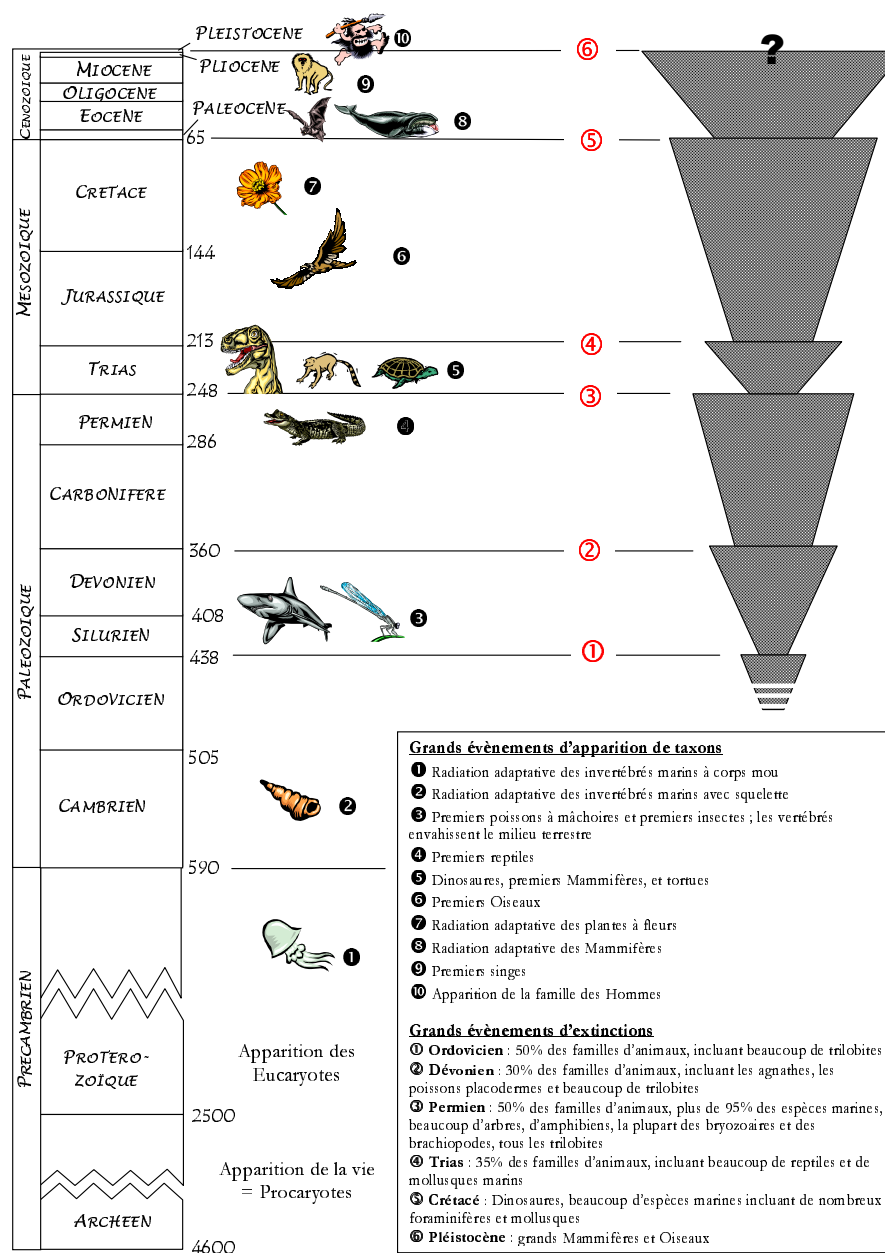


FIGURE B.1 – Illustration de l'histoire résumée de l'apparition de la vie (à gauche) et des extinctions (à droite) au cours des temps géologiques. L'échelle de temps est en millions d'années avant notre ère. Diagramme des extinctions, d'après BARBAULT (1997).

1989). Cette influence a conduit, dans certains cas, à l'extinction de la majorité des espèces de grande taille (OLSON 1989). En Nouvelle Zélande, dans les îles Hawaiï, les îles Mascareignes, les îles de la Société, les îles Marquises, la diversité des oiseaux a diminué de 30 à 50% (parfois jusqu'à 80%), pendant la période d'occupation des hommes (OLSON 1989). Pour l'époque récente, les pertes d'espèces sur les continents semblent être provoquées par l'apparition de la civilisation humaine et de l'agriculture (MOROWITZ 1991) et être accélérées par l'expansion démographique humaine (MOROWITZ 1991, SOULÉ 1991). Plus particulièrement, l'augmentation des extinctions

depuis le 17^e siècle correspondrait à la grande expansion des européens possédant des armes de plus en plus destructrices (DIAMOND 1989) et à l'industrialisation (SMITH *et al.* 1993). Il paraît donc évident que l'origine humaine de la majorité des extinctions de ces derniers siècles ne fait aucun doute (EHRlich et WILSON 1991). Les 20^e et 21^e siècles pourraient être les siècles les plus néfastes en ce qui concerne les extinctions d'espèces animales et végétales. Certains auteurs affirment qu'après cette période de 200 ans, l'extinction massive des espèces sera comparable à celle intervenue il y a 65 millions d'années et qui a vu s'éteindre la plupart des espèces qui vivaient sur terre (LANDE 1988) et qui a duré environ un million d'années. L'amplitude et le rythme de cette « **sixième extinction** » depuis l'apparition de l'homme (LEAKEY et LEWIN 1995 cité dans FRANKHAM *et al.* 2002) sont extrêmement inquiétantes (voir TABLEAU B.1). Certaines projections soutiennent que plus d'un quart des espèces sur Terre disparaîtraient dans les 50 ans à venir et ce serait aussi la première fois que les plantes subiraient une extinction massive (EHRlich et WILSON 1991)¹⁷. De plus, le processus évolutif inverse de l'extinction (la spéciation) qui doit créer des espèces, est aussi perturbé par des facteurs humains (ERWIN 1991) et le rythme d'extinction surpasse largement l'apparition de nouvelles espèces (FRANKHAM *et al.* 2002, p.2) ce qui limite l'apparition naturelle des espèces.

TABLEAU B.1 – Nombre d'espèces éteintes et menacées de 1600 à nos jours dans différents taxons et représentation (en %) des espèces éteintes sur les îles.

	Nb d'espèces éteintes ^{a/b}	% du taxon ^{a/b}	% sur îles ^a	Nb d'espèces menacées ^b
Mammifères	85/59	2,1/1,3	60	505
Oiseaux	113/116	1,3/1,2	85	1029
Reptiles	21/23	0,3/0,4	95	167
Amphibiens	2/2	0,05/0,07	0	59
Poissons	23/29	0,1/0,1	4	452
Invertébrés	98/65	0,01/0,015	49	999
Mollusques	191/191	-/0,2	79	354
TOTAL	533/485			3564

^a FRANKHAM *et al.* (2002); ^b CAUGHLEY et GUNN (1996)

B.1.2 Déclin et extinctions

Le problème actuel au niveau de la biodiversité¹⁸ est donc cette sixième extinction. On peut identifier plusieurs facteurs conduisant au déclin et à l'extinction de populations¹⁹. Ces facteurs ont des effets qui interagissent entre eux (GILPIN et SOULÉ 1986, SHAFFER 1987) si

¹⁷Certains auteurs relativisent les estimations et les prévisions des extinctions récentes et futures en mentionnant que les données sur les causes d'extinction sont difficilement disponibles (voir par exemple CAUGHLEY et GUNN 1996) et qu'il existe des biais (notamment un biais d'échantillonnage, SMITH *et al.* 1993; ou une mauvaise utilisation des théories de la biogéographie insulaire, voir MANN 1991).

¹⁸Le concept de biodiversité intègre trois composantes et leurs interactions : la diversité spécifique (richesse et abondance relative en espèces), la variabilité génétique (variabilité au niveau de l'ADN) et la diversité écologique (formes et modes de vie).

¹⁹ou d'une espèce si toutes les populations de cette espèce disparaissent

bien qu'une population peut entrer dans un spirale (ou **vortex d'extinction**, GILPIN et SOULÉ 1986) conduisant à son extinction, sinon à son déclin (voir FIGURE 1.1, p.8).

B.1.2.1 Facteurs naturels influençant les déclin et les extinctions

Les facteurs « naturels » provoquant le déclin et éventuellement les extinctions de populations et d'espèces sont de plusieurs natures :

- Les causes **déterministes** : les déclin peuvent être provoqués par des phénomènes non-aléatoires *i.e.* dont les effets ont une certaine direction et une certaine amplitude. Par exemple, lors de successions végétales, les différents stades sont amenés à disparaître et laisser la place aux stades suivants. Les facteurs humains (voir plus loin) peuvent être considérés comme des facteurs déterministes.
- Les causes **stochastiques** : ce sont des facteurs aléatoires dont les variations de direction et d'amplitude sont imprévisibles.
 - **Stochasticité démographique** : elle résulte des hasards des paramètres de natalité et de mortalité dans une population du fait de la réalisation d'un processus aléatoire dans une population à petit effectif. Par exemple, les individus d'une petite population peuvent être majoritairement des mâles (indépendamment de la mortalité) même si la sex-ratio est de 0,5 à la naissance.
 - **Stochasticité environnementale** : elle est liée aux variations des facteurs biotiques (ressources alimentaires, compétition, prédation, *etc.*) et abiotiques (climat). Les **événements catastrophiques** (*e.g.* inondations, incendies, sécheresses, météorites) sont des cas extrêmes de facteurs stochastiques environnementaux qui peuvent conduire à des extinctions soudaines de populations.
 - **Stochasticité génétique** : cette stochasticité est due aux variations aléatoires de la composition génétique des populations. La dérive génétique (perte aléatoire d'allèles) et la consanguinité (reproduction entre apparentés) peuvent influencer la viabilité d'une population et sont à la base de la génétique de la conservation dans les petites populations.

B.1.2.2 Les facteurs d'origine humaine

Les impacts de l'Homme sur la nature ont des effets différents suivant les pays (distinction Nord-Sud, SOULÉ 1991) car sont favorisés par la croissance de la population humaine, la pauvreté, la volonté de profits à court terme, l'anthropocentrisme (par exemple, les donations en Amérique, sont principalement orientées vers la religion et ne vont que minoritairement vers les associations pour la protection de la nature), les transitions culturelles et économiques (passage d'un mode de subsistance à une demande de profit croissant), et enfin aux facteurs politiques (reliés aux conditions économiques du pays).

On reconnaît généralement que la diminution de la biodiversité et la détérioration de la nature peuvent être causées par plusieurs facteurs humains directs ou indirects (MYERS 1979, DIAMOND 1989, WESTERN *et al.* 1989, EHRLICH et WILSON 1991, SOULÉ 1991, SPELLENBERG 1992, BARBAULT 1997) :

- *Destruction directe et dégradation de zones naturelles* : urbanisation, déforestation, assèchement de zones humides, défrichement, aménagement de cours d'eau, barrages. Ces dégradations provoquent souvent la fragmentation des habitats, par exemple le morcellement des forêts. La **fragmentation** des habitats entraîne tout d'abord des effets de bordure

(SOULÉ 1980a) : le centre des fragments sera plus proche de la lisière et les fragments posséderont une plus grande quantité de lisière comparée à leur surface. Les lisières sont des zones qui ont des caractéristiques particulières car elles subissent les influences des deux milieux. Même si elles peuvent contenir une plus grande diversité spécifique (car au niveau de la lisière, certaines espèces des deux milieux peuvent coexister), les espèces propres aux deux milieux peuvent disparaître. De plus, la fragmentation peut provoquer la diminution de la taille des habitats donc des isolats populationnels, la diminution de la connectivité entre populations et la restriction de la dispersion et de la colonisation. Les effets de la taille et des effets de lisières sont intimement liés entre eux (SOULÉ 1980a).

- *Dégradation indirecte* par la pollution de l'air, de l'eau et du sol et autres facteurs de l'influence humaine comme l'augmentation de l'effet de serre, pouvant induire une élévation de la température et du niveau de la mer (SOULÉ 1989). Ces facteurs peuvent affecter directement les organismes ou des cycles biologiques entiers.
- *Surexploitation de populations animales ou végétales* : chasse, pêche, cueillette. C'est le facteur le plus direct affectant la diversité spécifique. Le développement des armes et des technologies de chasse ont eu un impact sur la faune, spécialement sur les grands mammifères (voir BREITENMOSER 1998). Dorénavant, l'impact relatif de la surexploitation des espèces est peut être moins important, parce que beaucoup de candidats à l'extermination sont maintenant éteints ou vont l'être (DIAMOND 1989), mais il n'en reste pas moins que certaines espèces sont aujourd'hui menacées par le **braconnage** (*e.g.* rhinocéros, éléphant pour ne citer que les plus populaires).
- *Introduction d'espèces exotiques* prédatant ou entrant en compétition avec les espèces endémiques (l'exemple phare étant l'introduction de singes *Macaca fascicularis* et de cochons domestiques qui auraient conduit à l'extinction du dodo *Raphus cucullatus* sur les Iles Mascareignes à cause de la prédation des oeufs et des poussins, de la compétition pour la nourriture). Ceci doit être associé à l'introduction de maladies dans des régions où les espèces ne sont pas préparées et immunisées contre la nouvelle maladie (voir revue dans DASZAK *et al.* 2000).

B.1.2.3 Le rôle prépondérant de la taille de population

La taille de population est primordiale pour la gestion des populations en déclin, des populations des réserves, de zoos ou des populations insulaires (CAUGHLEY et GUNN 1996, p.165). La taille de population est d'ailleurs un des critères fondamentaux dans l'établissement des listes IUCN²⁰ (J. O'Grady, non publié, cité dans FRANKHAM *et al.* 2002). En effet, en plus des impacts directs de certains facteurs sur l'effectif des populations, les populations souffrent de conséquences diverses directement dues à leur faible effectif. Il existe des seuils en dessous desquels les forces aléatoires non-adaptatives prévalent sur les forces adaptatives et déterministes (SOULÉ 1985, 1986b). Quand l'effectif de la population est faible, plusieurs effets stochastiques sont attendus au niveau démographique (augmentation des effets dus à la stochasticité démographique et environnementale) et au niveau génétique (dérive génétique, consanguinité et baisse de la diversité génétique) qui augmentent la variabilité du taux d'accroissement de la population et la valeur de la probabilité d'extinction de la population (par exemple en diminuant la capacité à répondre aux changements d'environnement, voir FRANKHAM *et al.* 1999).

²⁰International Union for Conservation of Nature and Natural Resources

B.1.2.3.1 Stochasticité démographique

Dans une population à fort effectif, l'absence de reproduction ou la mortalité chez certains individus seront « tamponnées » s'il y a un assez grand nombre d'autres individus assurant le renouvellement de la population. En revanche, dans une petite population, dans les mêmes conditions, le taux d'accroissement devient très variable et la population peut alors s'éteindre très rapidement, car elle n'a pas un assez grand nombre de reproducteurs pour compenser la perte de reproduction et remplacer les individus qui meurent (CAUGHLEY et GUNN 1996, p.167). La modification de la distribution en âge ou en sexe a aussi une grande influence sur le déclin de la population. Par exemple, pour une population qui a perdu la grande majorité des femelles matures, la sex-ratio sera biaisée en faveur des mâles et la reproduction sera moins performante l'année suivante puisque limitée par le nombre de femelles. Dans ce contexte, on définit une taille efficace (démographique) de population, qui correspond à la taille d'une population ayant une sex-ratio équilibrée et une distribution en âge stable (population idéale) et ayant un accroissement net identique à la population étudiée. La taille efficace correspond au nombre d'individus participant réellement à la reproduction.

B.1.2.3.2 Stochasticité environnementale

Les petites populations peuvent être menacées par la stochasticité environnementale. Les conséquences de la stochasticité environnementale sont liés (comme pour la stochasticité démographique) aux effets de tirage aléatoire dans une population avec très peu d'individus. Dans le cas extrême des événements climatiques soudains de type catastrophique, le taux de mortalité peut être plus élevé chez les reproducteurs que chez les juvéniles, réduisant le recrutement à la prochaine génération. La structure en âge est modifiée et cela peut faire diminuer encore plus la taille de population.

En revanche, d'après CAUGHLEY et GUNN (1996), les effets de la stochasticité environnementale sur le taux d'accroissement d'une population seront différents de ceux de la stochasticité démographique. La variance du taux d'accroissement générée par les fluctuations environnementales serait en fait indépendante de la taille de population.

B.1.2.3.3 Stochasticité génétique

La taille de population et sa réduction sont des points cruciaux en génétique de la conservation. L'évolution génétique des populations est régie par les mutations, la migration, la sélection et l'aléatoire. Dans les petites populations, c'est le dernier facteur qui est le plus important car les taux de mutation sont faibles et leurs effets sont inexistant à court terme. De même, les effets de la sélection sont réduits ou inexistant (FRANKHAM *et al.* 2002).

La génétique des petites populations peut être articulée autour de deux points majeurs : la **dérive génétique** et la **consanguinité**.

B.1.2.3.3.1 La dérive génétique

En absence de sélection, de mutation, d'immigration, les effets combinés de la ségrégation des allèles dans les gamètes et de facteurs tels que la sex-ratio, la variance de la production de progéniture entre individus, la reproduction non aléatoire entre individus, produisent des changements aléatoires dans les fréquences des allèles provoquant une diminution de variation génétique par dérive aléatoire (LACY 1987, LANDE et BARROWCLOUGH 1987). La dérive génétique correspond donc à la perte de certains allèles et à la fixation d'autres. La perte d'allèles dans une

petite population peut être simplement expliquée par le fait que les allèles sont ségrégués parmi les gamètes de la génération parentale pour être représentés dans la génération suivante. Cette « loterie » répond à trois règles (CAUGHLEY et GUNN 1996) :

- plus la population est petite, plus les différences entre fréquences alléliques vont augmenter,
- plus la fréquence d'un allèle est faible, plus cet allèle a de chance d'être perdu à la génération suivante,
- plus la fréquence d'un allèle est élevée, plus cet allèle a de chance d'être fixé à la génération suivante.

Il en résulte une augmentation du nombre d'homozygotes et une diminution du nombre d'hétérozygotes (fixation et perte d'allèles ; *e.g.* ENGLAND 1997) ce qui peut entraîner :

- la perte de variabilité génétique qui peut diminuer le potentiel évolutif en réduisant la capacité d'adaptation à l'environnement et à ses changements (FRANKHAM *et al.* 1999).
- l'expression d'allèles délétères récessifs qui seront fixés s'ils ne sont pas létaux.
- l'augmentation de la sensibilité des lignées à des pathogènes dans le sens où les pathogènes auront beaucoup plus de facilité à se propager dans une population où les individus sont génétiquement identiques (ayant le même système d'histocompatibilité) que dans une population où les individus sont génétiquement différents (O'BRIEN 1994).

B.1.2.3.3.2 Consanguinité et dépression de consanguinité

Le faible effectif d'une population entraîne une autre conséquence : la consanguinité. La consanguinité correspond à la reproduction entre apparentés. Cette consanguinité peut être présente dans de grandes populations mais dans une petite population, la probabilité d'appariement entre apparentés augmente au fil des générations puisque les types d'appariement seront limités par le nombre d'individus présents (FIGURE B.2).

La consanguinité entraîne, sans changement des fréquences alléliques, une augmentation du taux d'homozygotie, et ainsi une exposition des allèles récessifs délétères, ce qui va entraîner une « dépression de consanguinité » se traduisant par une diminution de la moyenne des valeurs de fitness pour les caractères de reproduction (FRANKEL et SOULÉ 1981). Ces conséquences peuvent souvent amener à l'extinction de la population (voir MILLS et SMOUSE 1994, FRANKHAM et RALLS 1998, SACCHERI *et al.* 1998) surtout dans les cas de populations isolées et fermées.

Il semble que les différences de susceptibilité à la dépression de consanguinité parmi les mammifères, les oiseaux et les invertébrés sont faibles (FRANKHAM, expériences en captivité non publiées, cité dans FRANKHAM *et al.* 2002). Par contre à l'intérieur de chaque espèce, la dépression de consanguinité peut varier de façon importante entre populations. En effet, la dérive génétique peut faire disparaître les allèles délétères susceptibles d'être exposés par la consanguinité.

B.1.2.3.3.3 Notion de taille efficace génétique

La taille efficace est la taille d'une population idéale²¹ qui aurait le même taux de changement génétique que la population étudiée (KIMURA et CROW 1963). Les conséquences génétiques dans les petites populations dépendent de la taille efficace de la population et non directement de sa taille réelle. Il est donc impératif d'estimer la taille efficace d'une population (voir plus loin) pour appréhender les risques de dérive génétique et de consanguinité.

²¹ au sens de WRIGHT 1931 ; pas de migration, générations non chevauchantes, union aléatoire des gamètes, pas de sélection, les mutations sont ignorées, nombre de reproducteurs constant pour toutes les générations, taille des familles homogène, sex-ratio équilibrée.)

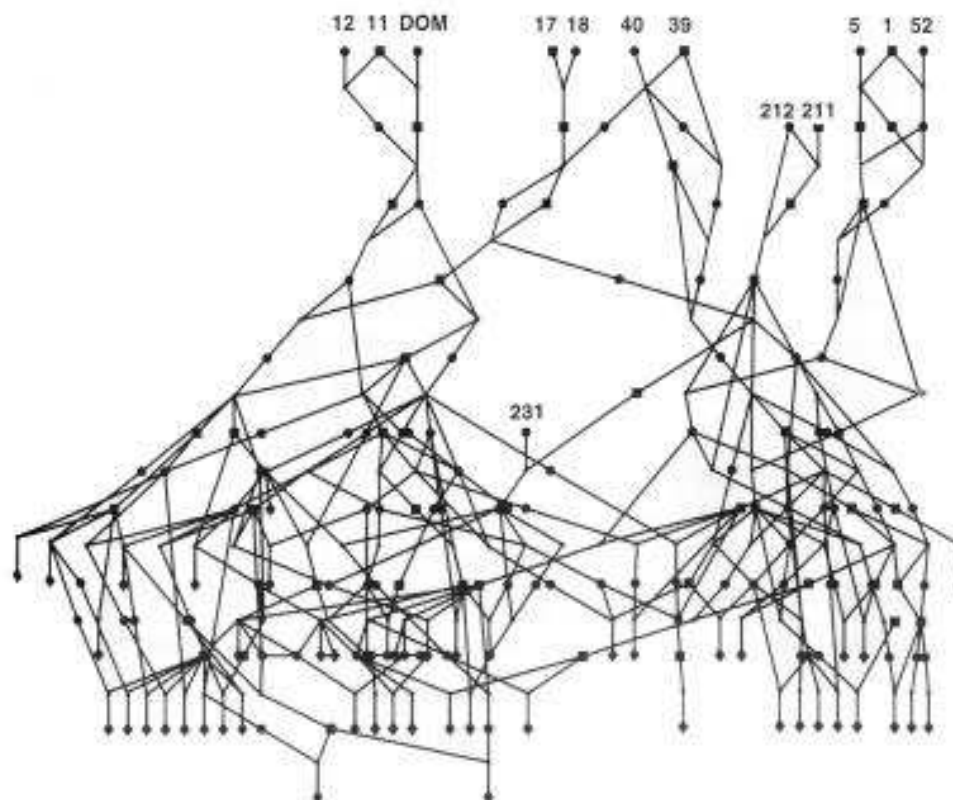


FIGURE B.2 – Pedigree de la population captive du cheval de Przewalski (*Equus przewalskii*) montrant que les croisements entre apparentés augmentent dans le temps dans une population restreinte. Issu de FRANKHAM *et al.* (2002).

Dans les populations naturelles, le rapport N_e/N est en général inférieur à 1 avec une moyenne d'environ 0.11 (méta-analyse, FRANKHAM 1995) ce qui signifie que les populations se comportent génétiquement comme si elle ne contenait que 10% de leurs individus.

B.2 L'ÉMERGENCE DE LA CONSERVATION

Le passé a donc montré de nombreuses périodes d'extinction plus ou moins importantes et on s'aperçoit que le rôle de l'Homme est largement reconnu dans la récente phase d'extinction.

L'origine de la conservation des milieux naturels remonte très loin dans l'histoire de l'Homme avec les croyances religieuses et philosophiques, notamment orientales (BARBAULT 1997, p.101). En occident, on ne sait pas trop quand cette idée a émergé (peut être avec ST THOMAS D'ACQUIN, ST FRANÇOIS D'ASSISE ou J.J.ROUSSEAU ?) mais l'idée concrète de la conservation naît au cours du 19^e siècle. L'exemple de l'Amérique est particulièrement intéressant car elle a eu un rôle non négligeable dans le développement de l'écologie et de la conservation (BARBAULT 1997). Longtemps vue comme un ensemble de territoires hostiles et sauvages, l'Amérique a été colonisée et le sentiment de dominance était réel. En 1620, le gouverneur William BRADFORD décrivait l'Amérique vierge comme une terre « hideuse et désolée » qu'il fallait conquérir. Pourtant, c'est en Amérique qu'apparaît l'idée de la responsabilité de l'homme dans la dégradation de la nature

et son devoir de corriger cette dégradation est pleinement exprimé par George P. MARSH (*Man and Nature/Earth as Modified by Human Action*, cité dans EHRENFELD 1970 p.6). Au cours du 19^e siècle, plusieurs expéditions rendent compte de la splendeur du territoire nord américain. En 1832, George CATLIN, propose de créer « un parc de la nation avec ses bêtes et ses hommes dans la fraîcheur de leur beauté sauvage ». En 1864, la première mesure concrète de protection d'un paysage naturel est prise par le gouvernement américain en donnant la vallée du Yosemite et une forêt de Séquoïa à l'état de Californie. Cela deviendra plus tard le *Yosemite National Park*. En 1872, c'est la création du premier parc national du monde, le *Yellowstone National Park*, qui deviendra très vite une sorte de parc de récréation plutôt qu'un sanctuaire pour la faune et la flore (voir FIGURE B.3).



FIGURE B.3 – L'entrée Nord du parc national du Yellowstone (Photo N. Valière). L'inscription est typique de ce que deviendra le parc : « Au profit et pour le plaisir du peuple ».

A la fin du 19^e siècle, un courant romantique naît sous l'impulsion des écrivains EMERSON et David H. THOREAU : la nature a une autre raison d'être qui ne se limite pas à des dimensions économiques et matérialistes (BARBAULT 1997). Au début du 19^e siècle, John MUIR et Gifford PINCHOT opposent leur point de vue (voir EHRENFELD 1970). D'un côté la conservation des sites et des ressources est primordiale pour stimuler les sentiments religieux et favoriser les expériences spirituelles mais aussi pour conserver les communautés d'espèces vivant dépendamment les unes des autres. De l'autre côté une vision « hyper-matérialiste »²² de la conservation des ressources naturelles (notamment des forêts).

Un autre courant apparaît avec le développement de l'écologie évolutive sous l'influence du premier conservateur américain Aldo LEOPOLD. Cette vision de la nature, encore imprégnée par l'idée d'équilibre, décrit la nature comme un ensemble de parties dynamiquement reliées entre elles dans un système intégré qu'il faut protéger en conservant les caractères propres des écosystèmes et de leur régulation naturelle. Dans les années 1960-70, une confluence sociale et scientifique (explosion des théories et des données sur l'écologie des communautés et de la

²² « L'usage adéquat des ressources naturelles est le plus grand bien pour le plus grand nombre et pour le plus longtemps ».

biogéographie insulaire ; volonté de séparer le biologie « traditionnelle » de la biologie moderne des populations incluant les domaines appliqués de la foresterie - *forestry* -, de la pêche - *fisheries* - ou de la gestion de la vie sauvage - *wildlife management* ; tendance sociale des mouvements environnementalistes et prise de conscience de la « crise de la biodiversité » ; SOULÉ 1986a, 1987a, BARBAULT 1997) incite les professionnels (biologistes, sociologues, gestionnaires, administrateurs) à développer une nouvelle discipline de la biologie : la biologie de la conservation.

Les différents mouvements décrits ici montrent les nombreux points de vue que l'on peut avoir dans l'idée de conservation. Les différents concepts s'articulent notamment dans la place de l'homme dans la nature et de son intervention dans la conservation de la vie sauvage et des espaces naturels. L'homme peut être écarté de la nature sauvage et ne doit pas influencer l'évolution naturelle (vision plutôt méditerranéenne). L'homme peut aussi être au centre de la nature et l'influencer largement (vision plutôt anglo-saxonne). Pour la vision utilitariste de la conservation, l'intérêt de l'homme doit être le but de la conservation. Cela est d'ailleurs privilégié dans les pays en développement où la conservation des espèces et des milieux ne peut en aucun cas se faire au détriment des populations locales, mais au contraire à leur avantage.

B.2.1 Pourquoi et comment conserver ?

B.2.1.1 Pourquoi conserver ?

De nombreux arguments incitent à conserver et protéger les espèces et/ou les écosystèmes. Il y a tout d'abord une dimension éthique à la conservation et la protection de la nature : toute forme de vie a le droit d'exister et l'Homme n'a aucun droit sur une espèce (MYERS 1979, p.46). Il a une responsabilité morale de protéger « ses compagnons vivants dans l'univers » (EHRlich et WILSON 1991). Un argument plus populaire est que la perte de la diversité spécifique est un problème évident (argument dit « esthétique », EHRlich et WILSON 1991). L'augmentation continue de l'intérêt du public pour la nature (*e.g.* éco-tourisme, films sur la vie sauvage) démontre le besoin de conservation de la biodiversité spécifique. C'est aussi le potentiel évolutif qu'il faut conserver car il ne s'agit pas de conserver les espèces et les écosystèmes tels qu'ils sont à l'heure actuelle, mais il faut conserver les processus écologiques et évolutifs pour permettre des adaptations futures. D'où l'intérêt de conserver la diversité génétique qui est la base du potentiel évolutif car elle gouverne la diversité phénotypique des organismes (voir WOODRUFF 1989). D'autres arguments mettent en avant les intérêts « matériels » de conserver et protéger la biodiversité (*e.g.* MYERS 1979, pp.57-81, SPELLENBERG 1992, pp.13-20). En effet, de nombreuses espèces végétales et animales fournissent des composés ou des produits qui peuvent être utilisés en médecine (*e.g.* médicament, antalgiques, antibiotiques, hormones, enzymes), en agriculture (*e.g.* pour la nourriture) ou dans l'industrie (*e.g.* latex, résines, huile essentielle).

B.2.1.2 Les moyens de conservation

Les différents arguments en faveur de la protection et de la conservation de la biodiversité ont nourri les débats sur la cible des actions : faut-il conserver les écosystèmes (processus et fonctions), les assemblages biogéographiques, les espèces, la diversité génétique... (SOULÉ 1991, WESTERN 1992). La protection d'aires naturelles permet de conserver les espèces qui y vivent et la mise en place de ces aires est souvent liée à la biodiversité existante (voir MÉDAIL 2001). En revanche, la conservation d'une espèce est souvent privilégiée surtout parce que cette espèce est « charismatique » (par exemple les grands vertébrés McNAUGHTON 1989, WOODRUFF 1989). Au-delà de la conservation des espèces, il s'agit de conserver les processus mettant en jeu ces

espèces. La diversité écologique est donc aussi un critère important dans la volonté de conservation. Comme le rappellent WESTERN *et al.* (1989), protéger les espèces, les communautés et les processus écologiques sont des actions complémentaires. Comment, en effet, exclure la conservation des espèces quand on conserve les communautés et les processus ? Parallèlement, lorsque l'on protège et conserve les espèces, c'est aussi certains processus et communautés que l'on préserve.

Il est possible de classer les modes de conservation et de protection en plusieurs niveaux (BARBAULT 1997), qui n'ont pas les mêmes buts et les mêmes conséquences sur la conservation des différents cibles biologiques (voir TABLEAU B.2).

B.2.1.2.1 Les aires de protection

C'est un mode de conservation *in situ* (c'est-à-dire sur place). Une aire protégée est selon l'IUCN, « une zone de terre ou de mer consacrée particulièrement à la protection de la biodiversité et des ressources naturelles et culturelles qui lui sont associées et gérée selon des lois ou d'autres moyens efficaces » (cité dans BARBAULT 1997). Il existe plusieurs types d'aires protégées, différenciés, en particulier, suivant les restrictions de l'intervention de l'homme. Si les parcs allient la protection d'une aire tout en permettant l'accès au public en ayant un rôle « d'éducation » (voir par exemple le Parc du Yellowstone), certaines réserves sont fermées à toute implication de l'homme. D'autres systèmes doivent prendre en compte simultanément les objectifs de conservation et de développement des sociétés humaines (par exemple les réserves de biosphères créées par l'UNESCO grâce au programme MAB, *Man and Biosphere* en 1974). Aujourd'hui, il existe plus de 4500 parcs et zones protégées répartis sur les 5 continents et comptabilisant une superficie totale de 4,5 Millions de km².

B.2.1.2.1.1 La conservation *ex situ* des ressources

Il existe plusieurs possibilités de conservation *ex situ* (c'est-à-dire hors site). Les plus traditionnelles sont les parcs zoologiques et les jardins botaniques. Ce sont deux systèmes artificiels qui ont pour but de conserver et de reproduire des organismes vivants dans des conditions contrôlées et semi-artificielles afin de conserver les espèces et les ressources génétiques.

Les zoos et les jardins botaniques sont aujourd'hui plus que des « musées du vivant » car les gestionnaires de ces établissements ont des rôles non négligeables en conservation en participant activement aux programmes de reproduction des espèces en danger avec une coopération dans des programmes de conservation *in situ* (voir FRANKHAM *et al.* 2002, pp.421-422). Les espèces qui sont gardées en captivité sont susceptibles de recouvrer la liberté dans des programmes de relâché. Néanmoins, une gestion appropriée de ces populations restreintes doit être menée afin de limiter les effets de la consanguinité (voir Section B.4.2.2, p.156), de l'augmentation d'allèles délétères dans des conditions de *sélection relâchée* (LYNCH et O'HELY 2001). Certaines espèces (dont l'oryx d'Arabie *Oryx leucoryx*, l'oryx algazelle *Oryx dammah*, l'addax *Addax nasomaculatus*, le cheval de Przewalski *Equus przewalskii*, le furet à pattes noires *Mustela nigripes*, le cerf du Père David *Elaphurus davidiensis*, le condor de Californie *Vultur californicus*, 11 espèces d'escargot du genre *Partula* et quelques végétaux sont vivantes uniquement en captivité (FRANKHAM *et al.* 2002).

D'autres systèmes consistent à préserver directement les ressources biologiques en conservant artificiellement des graines, gamètes ou embryons, des tissus, des cellules des espèces en danger. Les gamètes ou les embryons d'espèces en danger peuvent être inséminés chez des animaux domestiques, constituant ainsi des mères de remplacement (voir trois exemples chez les animaux

TABLEAU B.2 – Potentiel relatif des systèmes de conservation pour la protection et la maintenance de la diversité biologique (modifié d'après Soulé, 1991). '0' = pas de rôle; '+', '++', '+++' indiquent un rôle faible, modéré et significatif. Voir aussi SPELLENBERG (1992, p. 47175) pour les aires de protection.

Cibles de conservation	Systèmes de conservation					
	Aires de protection stricte ¹	Aires de protection ²	Ré-introduction ⁴	Aires artificielles de protection ³	Parcs et jardins ⁵	Banques (tissus, GraïneADN) ⁶
Ecosystème						
<i>Processus et fonctions</i>	+++	++	++	++	0	0
<i>Utilisation humaine traditionnelle</i>	0	++	++	+	0	0
Assemblages biogéographiques	+++	++	+	++	0	0
Espèces endémiques ou indigènes	+++	++	+	++	++	+
Populations locales	+++	++	++	++	+	+
Variation génétique des espèces						
<i>Equivalents sauvages d'esp. domest.</i>	+++	++	+	+	+++	++
<i>Variétés domestiques traditionnelles</i>	0	+	0	+	+	+++
<i>Variation génétique non-économique</i>	+++	++	++	+	+	+

dans LANZA *et al.* 2000). Cette stratégie de cryopréservation est bien plus applicable pour les plantes en utilisant les graines qui entrent en dormance. Récemment, un appel à une coordination internationale a été lancé pour unifier les efforts de conservation de l'ADN dans l'espoir de sauvegarder une trace du patrimoine génétique des espèces en danger afin, dans le futur, de recréer ce qu'étaient les organismes (MORIN 2000, POINAR et EGLINTON 2000, RYDER *et al.* 2000). Mais quel intérêt de reconstruire des organismes si les milieux dans lesquels ils vivaient ont été détruits et dégradés. La conservation d'ADN ou d'échantillons contenant l'ADN peut être par contre très utile pour des buts scientifiques comme la comparaison de la diversité génétique entre espèces actuelles et anciennes.

B.2.1.2.1.2 Les mesures légales

L'évolution des mesures légales s'est réalisée en trois phases (BEURIER 2001) : la protection des espèces utiles à l'homme (avec une contrepartie : la destruction de « nuisibles »), la protection des milieux (*e.g.* Convention de Ramsar, 1971) et l'interdiction de certains commerces d'animaux, puis enfin une approche prenant en compte les milieux, les espèces et leurs interactions (*e.g.* Convention de Berne, 1979).

Les mesures légales consistent à donner un statut législatif (ou assimilé) particulier pour certaines espèces et habitats. Ces mesures peuvent dépendre des Etats et être de la forme de listes d'espèces interdites de chasse ou d'exploitation (par exemple les ESA ou *Endangered Species Act* américain ou australien), ou de législation visant à protéger un habitat et délimiter des parcs et réserves naturels, ou encore à fixer les procédures d'étude et de réintroduction d'espèces (comme par exemple la directive « Habitats-Faune-Flore » ou « Natura 2000 » de la communauté européenne).

D'autres mesures peuvent être prises au niveau international, par exemple :

- la « liste rouge » de l'IUCN : l'IUCN (*International Union for Conservation of Nature and Natural Resources* ; 1948) est une organisation non gouvernementale internationale regroupant plusieurs Etats mais aussi une grande majorité d'organisation de protection de la nature. L'IUCN fournit une liste d'espèces classées selon leur statut²³ : éteint, éteint à l'état sauvage, en danger critique d'extinction, en danger, vulnérable, quasi menacée, préoccupation mineure, données insuffisantes, non évalué. Le statut de chaque espèce est déterminé selon plusieurs critères comme le taux d'extinction projeté, l'effectif total d'individus matures, l'aire de répartition ou la réduction de la taille de population. Cette liste n'a pas de réelle valeur législative mais encourage la coopération entre les gouvernements, les organisations nationales et internationales concernées par la protection de ces espèces.
- La convention sur le commerce international des espèces en danger de la faune et la flore sauvage (CITES, *Convention on International Trade in Endangered Species of Wild Fauna and Flora*) : cette convention internationale publie tous les deux ans trois listes. Les espèces de l'Annexe I sont des espèces en danger d'extinction et ne sont pas autorisées à être importées ou exportées par les nations signataires excepté sous certaines conditions non commerciales (par exemple scientifique) ; l'Annexe II est une liste des espèces non menacées pour lesquelles on a besoin d'un permis assurant que l'animal ne sera pas blessé lors du transport ; l'Annexe III est réservé aux espèces dont les nations pensent qu'une coopération internationale est nécessaire pour la survie et la conservation de cette espèce.

²³ Catégories et Critères de l'IUCN pour la Liste Rouge Version 3.1 (2000), catégories préparées par la Commission de la sauvegarde des espèces de l'IUCN, approuvées lors de la 51^e réunion du Conseil de l'IUCN Gland, Suisse ; disponible sur le site de l'IUCN <http://www.iucn.org/themes/ssc/redlists/RLcats2001booklet.html>

B.2.1.2.1.3 Les programmes de restauration

Les programmes de restauration consistent en des activités intenses de gestion d'une espèce menées pour augmenter ou retenir la richesse (diversité génétique par exemple), sa viabilité démographique ou sa productivité (effectif) dans des habitats dégradés (SOULÉ 1991). Elle consiste à ré-introduire ou renforcer²⁴ une population en amenant des individus de la même espèce (ou un taxon proche et interfertile). Ce type de système de préservation est extrêmement coûteux car il nécessite des études intensives sur les deux sites ou populations (captives ou non) concernés (site ou population d'origine des individus transférés et site ou population qui recevra ces individus). De lourds moyens techniques pour le prélèvement et le transport des individus d'un site à l'autre devront être mis en place. Enfin un suivi post-translocation est nécessaire pour s'assurer du succès de l'opération. Malgré le nombre croissant de programmes de réintroduction, cette pratique souffre de l'absence de cadre théorique. Néanmoins, de nombreux domaines doivent être étudiés afin de mener au mieux ces opérations : études de terrains (écologie, démographie pour connaître les causes du déclin de la population), études génétiques, simulations informatiques.

B.3 LE RÔLE DE LA BIOLOGIE EN CONSERVATION

SOULÉ (1987a) identifie six grands rôles des membres de la *Society for Conservation Biology* (créée en 1986) et donc de la biologie de la conservation :

- Modélisation et analyse des processus biologiques au niveau des populations, des communautés, des écosystèmes et de la planète,
- Inventaire et détermination de la systématique des espèces
- Test d'hypothèses (expérimentations)
- Développement et évaluation des interventions technologiques et de gestion pour maintenir et restaurer la diversité et les fonctions,
- Communication des résultats (notamment par le biais du journal *Conservation Biology*),
- Intégration des connaissances aux activités humaines.

Plus généralement, le rôle principal de la biologie de la conservation est de faire le lien entre les gestionnaires et les scientifiques afin que cette coopération puisse identifier les problèmes, fournir des principes scientifiques à la conservation et des procédures de correction (WESTERN 1989). C'est d'ailleurs grâce à la biologie (en général) que les conservationnistes vont être alertés sur les risques biologiques encourus par les populations de petites tailles et consanguines, et sur la dimension évolutive des principes de conservation (WESTERN 1989).

Pour tous ces rôles, de nombreuses disciplines doivent coopérer, avec la participation de nombreux acteurs : écologie, biogéographie, systématique, génétique, science de l'évolution, épidémiologie, sociobiologie, foresterie, pêche, gestion, agronomie, médecine vétérinaire, économie, politique, ethnobiologie, anthropologie, éthique environnementale, paléontologie, climatologie, philosophie, océanographie, agences gouvernementales, organisations et associations non gouvernementales (SOULÉ 1986a, EHRENFELD 1987).

En ce qui concerne la biologie, on peut définir deux grandes implications : *(i)* identifier les unités biologiques qu'il faut conserver et déterminer les priorités et *(ii)* apporter des informations primordiales sur les populations, les espèces et/ou les écosystèmes pour choisir la meilleure

²⁴la ré-introduction consiste à relâcher des animaux qui ont disparus de l'aire géographique ; le renforcement de population consiste à relâcher des animaux dans une zone d'où ils n'ont jamais disparus.

stratégie.

B.3.1 Diagnostic et prescription

Une approche rationnelle de la conservation veut que la première étape de cette démarche soit d'identifier les **unités de conservation** (DIMMICK *et al.* 1999). De nombreux débats ont lieu sur la sélection des aires de protection ou des espèces à protéger (voir SPELLENBERG 1992).

Les unités biologiques de conservation correspondent aux unités biologiques (espèces, sous-espèces, populations,...) qui partagent une aire géographique et une histoire commune, et qui montrent une différenciation des autres unités, pouvant ainsi être gérées indépendamment²⁵. Il existe de grandes controverses concernant l'identification et le choix des unités de conservation. On peut, par exemple, se demander s'il vaut mieux conserver les processus écologiques et évolutifs plutôt que des espèces particulières entrant dans ces processus (WESTERN 1992). De même, faut-il concentrer ses efforts sur la diversité écologique (*i.e. les processus*), sur la diversité spécifique (maintien du nombre d'espèces) ou la diversité génétique (importante dans les évolutions et adaptations futures) (WOODRUFF 1989, CROZIER 1992). La trichotomie de ce débat concorde avec les disciplines de l'écologie, de la systématique et de l'évolution (BOWEN 1999). Le choix de prendre l'espèce comme unité de conservation est alors pragmatique (CAUGHLEY et GUNN 1996) car l'espèce est une entité biologique et évolutive que tout le monde connaît. Toutefois cette entité est difficile à définir de façon consensuelle. Faut-il considérer l'espèce de façon taxonomique (typologique; SOLTIS et GITZENDANNER 1999) ou de façon évolutive. La première approche définit une espèce grâce à la nomenclature (*Genre espèce*) alors que la seconde définit une espèce comme unité d'évolution en prenant en compte les processus évolutifs (ROJAS 1992). Il existe plusieurs autres concepts alternatifs : le concept biologique de MAYR (1942) prenant l'isolement reproductif des espèces comme critère principal, le concept évolutif (SIMPSON 1961, WILEY 1978) ou divers concepts phylogénétiques (voir liste dans SOLTIS et GITZENDANNER 1999). Il est facile de comprendre que s'il existe plusieurs définitions pour l'espèce, il existera presque autant de stratégies de conservation puisque les efforts s'appliqueront sur des unités différentes (MORITZ 1994). Par exemple le nombre d'espèces de *Drimys* (Rongeurs) en Malaisie peut aller de 1 à 30 suivant le concept utilisé (voir référence dans ROJAS 1992).

Le choix de l'unité de conservation n'est pas une tâche suffisante pour commencer un plan de conservation. Encore faut-il savoir si une espèce doit être conservée ou protégée et choisir quelle espèce ou quelles populations utiliser? La conservation de sites naturels peut être définie soit de façon intuitive soit en utilisant des critères plus objectifs basés sur la diversité (écologique, en espèce) du milieu (SPELLENBERG 1992, SANTINI et ANGULO 2001). Un des arguments en faveur du choix d'une espèce, à défaut de la conservation d'un assemblage écologique, est que certaines d'entre elles peuvent représenter des « espèces-parapluies », c'est-à-dire des espèces dont la conservation permet la conservation d'autres espèces. Par exemple, les grands carnivores nécessitent de grandes aires de protection qui incluent de nombreuses communautés et espèces, autres que les grands carnivores (NOSS *et al.* 1996). Les choix d'une espèce ou d'une population doit donc prendre en compte le statut de cette espèce ou population (déclin, risque d'extinction, influence de l'homme, etc.), les coûts requis par sa conservation, ainsi que les retombées écologiques de sa protection.

²⁵Les unités de conservation peuvent être différentes des unités de gestion sur lesquelles vont agir directement les moyens de conservation.

B.3.1.1 L'apport de la biogéographie insulaire et du concept de métapopulation

La prise en compte de la structuration spatiale des populations (voir GILPIN 1987) a incité à se pencher sur l'intérêt de cette structuration dans les problèmes de conservation. Le concept de métapopulation est un exemple de cette influence. Une **métapopulation** est un ensemble de populations locales qui possèdent une dynamique propre à chacune, mais aussi une dynamique globale à l'échelle de la métapopulation (ELMHAGEN et ANGERBJÖRN 2001). Le concept de métapopulation a servi de base à la formalisation de la dynamique d'extinction/recolonisation affectant les populations locales. Dans une métapopulation, quand une population s'éteint, elle peut être remplacée par une autre population. Un des facteurs clé de la dynamique de la métapopulation est donc l'inter-connectivité existant entre les populations locales permettant la migration entre populations et la recolonisation de sites vacants. Le concept de métapopulation a, entre autres, servi de base pour l'élaboration de modèles nécessaires à la conservation de certaines espèces dont l'habitat a été fragmenté et dont les populations ont été plus ou moins dispersées en plus petits groupes. La **biogéographie insulaire** (voir MCARTHUR et WILSON 1967) était une approche de la structure spatiale des populations dans des îles tenant compte du taux d'immigration et des taux d'extinction, la distance au continent et la superficie des îles. On a assimilé ensuite les réserves et refuges à des îles et cette théorie a permis de développer des réflexions sur la détermination de la forme, la taille et la connectivité de refuges afin de maximiser notamment le nombre d'espèces dans l'ensemble des refuges. Les concepts de métapopulation et la biogéographie insulaire ont galvanisé la biologie de la conservation pendant des années, car ils devaient donner des règles sur la taille et la forme des motifs des refuges (ou réserves) et leur connectivité (SIMBERLOFF 1988).

Le débat s'est notamment orienté sur la meilleure stratégie à adopter : un seul grand refuge vaut-il mieux que plusieurs petits refuges (**Single Large Or Several Small**, SLOSS; voir FIGURE B.4) et la réponse n'est pas évidente (voir par exemple WILCOX et MURPHY 1985, SIMBERLOFF 1988) surtout si on prend aussi en compte la forme des refuges (SIMBERLOFF 1988). Comme éléments de réponse, on peut avancer l'idée que la diversité spécifique augmente moins rapidement dans les refuges au fur et à mesure de l'augmentation de leur taille. On peut ainsi penser que les petits refuges contiennent plus d'espèces que des grands proportionnellement à leur taille. En revanche, les grands refuges peuvent accueillir certaines espèces qui seront absentes dans les petits refuges, par exemple les espèces de grande taille qui représentent souvent des espèces-parapluie (WESTERN *et al.* 1989). De plus, les grands refuges sont souvent moins vulnérables (au niveau de la surface touchée) aux événements catastrophiques tels que les incendies (SOULÉ 1989). Il faut aussi prendre en considération le coût financier de la conservation. Ce coût sera beaucoup plus important pour les grands refuges que pour de petits refuges, à moins que l'on laisse les écosystèmes d'un grand refuge atteindre des cycles biologiques sans intervenir (WALKER 1989). D'autre part, une multitude de petites réserves aurait une meilleure persistance à long terme, si on prend en compte la dynamique d'extinction/recolonisation dans la métapopulation : des refuges s'éteindraient et seraient remplacés par d'autres si l'isolement des fragments n'est pas complète. Toutefois, la redistribution des individus d'une population vers des sites vacants peut entraîner des goulots d'étranglement dans les populations sources et des effets de fondation dans les nouvelles populations (d'où une possible perte de variabilité génétique, de la consanguinité, une possible stochasticité démographique).

Ce dernier point souligne l'importance de la connexion entre les refuges. La mise en place de **corridors** augmente la connexion entre les refuges permettant de redresser des extinctions locales grâce à la migration d'individus. En revanche, il faut savoir que ces corridors permettent aussi la dispersion et la transmission de maladies, d'incendies, de prédateurs. Dans l'archipel des

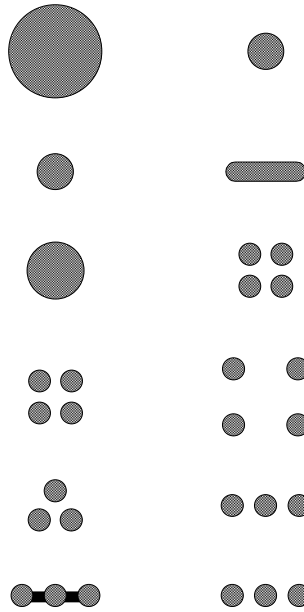


FIGURE B.4 – Principe de formes et tailles de refuges d’après la biogéographie insulaire. Pour chaque cas, la forme et la taille sur la gauche sont de meilleure qualité que celles de droite. (d’après SIMBERLOFF 1988 et JANZEN 1986).

Seychelles, l’impact de l’introduction de prédateurs a été limitée grâce à la séparation naturelle des îles et seulement deux espèces d’oiseaux ont disparu sur l’ensemble de l’archipel (SIMBERLOFF 1988). Au niveau génétique, les corridors réduisent la consanguinité et la dérive génétique en favorisant les échanges. Ils limitent en revanche l’évolution de génotypes localement adaptés tout en diminuant la diversité génétique entre les fragments (homogénéisation). En absence de connexion, les populations subdivisées perdent rapidement de la variabilité dans chaque sous-population, mais la variation (diversité entre populations) est retenue sur toutes les populations tant qu’il n’y a pas d’extinction locale (VARVIO *et al.* 1986, LACY 1987). Il y a donc, dans un système de réserves, un compromis à faire entre la dynamique d’extinction/recolonisation, la diversité spécifique, les évolutions possibles des tailles de population et la variabilité génétique.

B.3.1.2 Minimum Viable Population et Population Viability (Vulnerability) Analysis

Si le concept de métapopulation et la biogéographie insulaire se sont penchés sur la dynamique au niveau des écosystèmes, les approches qui ont pris leur suite se sont elles focalisées sur les causes des extinctions des petites populations (SIMBERLOFF 1988). En effet, la détection des priorités dans la conservation d’espèces est principalement réalisée grâce à l’estimation des probabilités d’extinction (MAGUIRE *et al.* 1987) et plus particulièrement la détermination des effets de certains événements sur l’extinction de populations. La prise en compte des facteurs influençant les populations en danger et à petits effectifs a donc fait émerger l’idée d’une approche d’analyse de viabilité (ou vulnérabilité) des populations (PVA, *Population Viability Analysis*; GILPIN et SOULÉ 1986, SOULÉ 1987b) basée sur l’appréhension de la persistance des populations au cours du temps (viabilité). La PVA tente « d’intégrer les interactions complexes entre les facteurs génétiques, démographiques, environnementaux et spatio-temporels pour déterminer les probabilités d’extinction des populations » (SOULÉ 1989). L’approche de PVA repose en partie sur le

concept de Population Minimale Viable (MVP, *Minimum Viable Population*) (en association avec le concept de métapopulation en considérant la dynamiques d'extinction/recolonisation). Ce concept doit répondre à la question : « quelles sont les conditions minimales pour la viabilité des systèmes naturels sur une certaine durée ? ». Trois causes potentielles d'extinction des populations sont considérées pour le concept de MVP : les stochasticités démographique, environnementale et génétique (uniquement la consanguinité, FRANKHAM *et al.* 2002, p.511). SHAFFER (1981) définit une MVP comme la plus petite population isolée ayant 99% de chance de rester telle quelle pendant 1000 ans en dépit des effets des stochasticités démographique, environnementale et génétique. L'approche consiste à estimer les probabilités d'extinction de l'espèce ou de la population étudiée et sa viabilité grâce à des études empiriques, des modèles analytiques, des déduction subjective et surtout des simulations informatiques à partir de données préliminaires (logiciel VORTEX, MILLER et LACY 1999 ou d'autres logiciels voir BROOK *et al.* 2000, LEGENDRE 2000). L'utilisation de modèles de viabilité (*e.g.* courbes d'isorisques entre le pourcentage de déclin et la probabilité d'extinction) ou de certains critères d'évaluation (*e.g.* déclin observé sur une période de temps, taux de croissance, taille de population, isolation et flux entre populations, la destruction de l'habitat) permet de comparer les espèces pour les classer dans des catégories de vulnérabilité (voir BARBAULT 1997, MAGUIRE *et al.* 1987) mais aussi d'estimer l'impact relatif de chaque facteur (analyse de sensibilité).

B.3.2 Informations biologiques

En biologie de la conservation de nombreux traits d'histoire de vie doivent être déterminés pour pouvoir gérer et réagir au mieux sur les populations menacées. Les paramètres démographiques (taux de survie, de fécondité, taille de population) sont bien évidemment des paramètres de première ordre pour pouvoir ajuster les moyens de conservation (par exemple, d'où vient le déclin d'une population : d'une augmentation de mortalité ou d'une baisse de fécondité, ou les deux). Le comportement a apporté aussi une contribution importante à la biologie de la conservation (voir SUTHERLAND 1998). En effet la connaissance des comportements sociaux, des patrons de dispersion, des mouvements d'individus, des liens de parenté et des patrons de reproduction sont nécessaires afin d'étudier les processus d'adaptation dans différentes conditions écologiques et donc d'adapter les moyens de conservation à chaque population.

Pour obtenir de telles informations, des études de terrain, principalement par l'observation directe, sont souvent nécessaires, mais elles ne permettent d'obtenir qu'un nombre limité d'informations. Les techniques de radiopistage ou de marquage d'individus servent aussi à suivre des individus (*e.g.* transpondeur, balise ARGOS, bague d'identification, radio-émetteur), de connaître les domaines vitaux et les habitudes de ces individus, mais aussi la dispersion et les détails de la migration de certaines espèces (qui migre, où et quand ?). Les modèles de capture-marquage-recapture peuvent être utilisés pour estimer les effectifs des populations, les taux de survie et de fécondité, ou même les taux de dispersion (SCHWARTZ et SEBER 1999). L'effectif des populations est d'ailleurs une information primordiale pour le biologiste de la conservation puisqu'une population peut s'éteindre rapidement si elle passe en dessous de certains seuils (mais voir section sur la taille efficace). Le suivi des tailles de populations peut donc mettre en évidence le déclin de certaines populations ou espèces, mais il faut noter que la détection du déclin de populations est d'autant plus difficile que la taille de population est faible (CAUGHLEY et GUNN 1996). L'importance de la taille réelle de population est à relier aussi à la l'importance de la taille efficace de population dans le diagnostic des risques d'extinction.

Les caractéristiques biologiques des populations, en particulier la taille de population, doivent être étudiées pendant et après l'application des moyens de conservation pour vérifier le bon

déroulement des processus de conservation (par exemple un suivi d'une populations réintroduite est une étape obligatoire d'un programme de réintroduction).

B.4 LA GÉNÉTIQUE DE LA CONSERVATION

La génétique de la conservation est l'application de la génétique pour préserver les espèces comme des entités dynamiques capables d'adaptation aux changements d'environnement. Cela regroupe la résolution des incertitudes taxonomiques, la définition d'unités de gestion à l'intérieur des espèces, la gestion génétique des petites populations et l'utilisation des analyses génétiques dans les cas de commerce illégal et de braconnage et pour comprendre la biologie de l'espèce (FRANKHAM *et al.* 2002, p.1).

La génétique de la conservation a permis d'apporter une contribution non négligeable à la compréhension des mécanismes de déclin et d'extinction et doit, plus précisément, traiter de 11 problèmes principaux (FRANKHAM *et al.* 2002, p.9-10) :

- les effets délétères sur la reproduction et la survie, dus à la consanguinité
- la perte de diversité génétique et la capacité à évoluer dans un environnement changeant
- la fragmentation des populations et la réduction du flux génique
- les processus aléatoires (dérive génétique) surpassant la sélection comme principal processus évolutif
- l'accumulation et la perte (purge) des mutations délétères
- l'adaptation génétique à la captivité et ces effets sur le succès de réintroduction
- la résolution d'incertitudes taxonomiques
- la définition d'unités de gestion à l'intérieur des espèces
- l'utilisation des analyses génétiques moléculaires dans les cas de braconnage et de commerce illégal d'espèces
- l'utilisation des analyses génétiques moléculaires pour comprendre les aspects de la biologie de l'espèce important pour la conservation
- les effets délétères sur la *fitness* (valeur sélective) qui apparaissent parfois (outbreeding depression).

Nous allons aborder une partie de ces points et l'exposé sera articulé autour de trois thèmes principaux déjà décrits au niveau non-génétique (Section B.3.1, p.148 et Section B.3.2, p.151) :

1. la définition de la taxonomie et des unités de conservation
2. diagnostiquer les problèmes génétiques et donner des priorités (détection de la consanguinité, de la diversité génétique, calcul de la taille efficace)
3. obtention d'informations biologiques (compréhension de la biologie de l'espèce, identification des populations et des individus pour la détection de l'exploitation illégales des espèces, études des populations pour des fins de restauration de populations).

B.4.1 Définition de la taxonomie et des unités de conservation

La génétique peut avoir un rôle important dans l'identification des espèces et des unités de conservation comme le montrent les trois exemples suivants.

L'étude de marqueurs mitochondriaux a montré que non seulement le Renard de Darwin (*Pseudolopex fulvipes*) était unique, mais qu'il était une espèce ancestrale en comparaison du Renard gris (*Pseudolopex griseus*). Le Renard de Darwin devait donc être considéré comme une unité biologique particulière qu'il fallait protéger (exemple cité dans JOHNSON *et al.* 2001).

Chez les félidés, les séquences mitochondriales ont été utilisées pour déterminer les relations phylogénétiques existant entre les petits félidés. Ainsi, on a pu montrer que les taxons comme le kodkod (*Oncifelis guigna*) et le chat de Geoffroy (*O. geoffroyi*) divergeaient relativement récemment en comparaison avec les ocelots (*Felis pardalis*) et les margays (*F. wiedi*) ou avec d'autres lignées de chat-ocelot (*Leopardus tigrina*). Cette dernière lignée particulière, avec son statut d'espèce en danger, pouvait alors retenir une attention propre dans le futur (exemple cité dans JOHNSON *et al.* 2001).

Le statut de la tortue de Kemp (*Lepidochelys kempi*) n'était pas bien défini avant les analyses génétiques. Ce taxon était considéré comme une espèce à part entière, ou comme une sous-espèce de la tortue de Ridley (*L. olivacea*), ou enfin comme un hybride entre la caouanne (*Caretta caretta*) et la tortue verte (*Chelonia mydas*) ou la tortue à écailles (*Eretmochelys imbricata*). La divergence génétique sur une partie de l'ADN mitochondrial a montré que les tortues de Ridley et de Kemp étaient des formes distinctes des autres tortues et qu'il existait une séparation suffisante entre les deux formes pour appliquer des programmes de protection séparés (exemple cité dans HAIG 1998).

Une approche rationnelle de la conservation veut donc que la première étape de cette démarche soit d'identifier les unités de conservation (DIMMICK *et al.* 1999). Sans se baser simplement sur une divergence de séquences du génome, on peut définir des unités de conservation avec des règles plus ou moins arbitraires. Les **unités « évolutivement » significatives** (*Evolutionary Significant Unit*, ESU) sont des taxons pour lesquelles on suppose un potentiel évolutif conduisant à des adaptations et qui méritent une conservation séparée (RYDER 1986). La définition des ESU a évolué au cours du temps (voir CRANDALL *et al.* 2000). Elles peuvent, par exemple, être définies sur la base d'une variation adaptative significative basée sur la concordance de différents jeux de données provenant de différentes techniques (RYDER 1986). Pour WAPLES (1991), une ESU est un ensemble de populations qui sont reproductivement séparées d'autres populations et qui ont des adaptations uniques et différentes.

MORITZ (1994) définit une ESU, au niveau génétique, comme un ensemble de populations ayant une divergence phylogénétique significative des allèles (haplotypes) sur l'ADN mitochondrial par rapport aux autres populations, avec une divergence semblable pour les fréquences alléliques de loci nucléaires (voir FIGURE B.5). Il différencie les ESU des **unités de gestion** (*Management Units*, MU). Ces dernières sont des populations démographiquement distinctes ayant des fréquences alléliques distinctes quels que soient la structure phylogénétique et le niveau de divergence génétique (voir FIGURE B.5).

Si le diagnostic et la pertinence des unités de conservation sur la base d'informations génétiques est maintenant discuté (voir par exemple TAYLOR et DIZON 1999, GOLDSTEIN *et al.* 2000), il s'avère en fait que la définition des ESU et des MU par MORITZ (1994) n'a pas suscité assez de débat et de discussion dans la communauté scientifique (PAETKAU 1999). Elle a été mise en application directement alors que ces approches comportaient des limites dans certains cas (voir PAETKAU 1999). De plus, on peut penser que les données démographiques détaillées peuvent être beaucoup plus utiles pour la biologie de la conservation car ce sont des facteurs répondant à court terme au changement d'environnement, alors que les facteurs génétiques sont des facteurs répondant à long terme (LANDE 1988, MILLIGAN *et al.* 1994). Ainsi, si on étudie la divergence entre deux populations ou sous-espèces, l'absence de divergence génétique ne signifie pas forcément que les formes étudiées sont conspécifiques mais peut simplement signifier que les données sont insuffisantes pour pouvoir rejeter l'hypothèse nulle « il n'y a pas de divergence entre les populations » (DOWLING *et al.* 1992a). HEDRICK (1999) pose aussi le problème de la relation entre la signification statistique d'une divergence et la signification biologique. En effet,

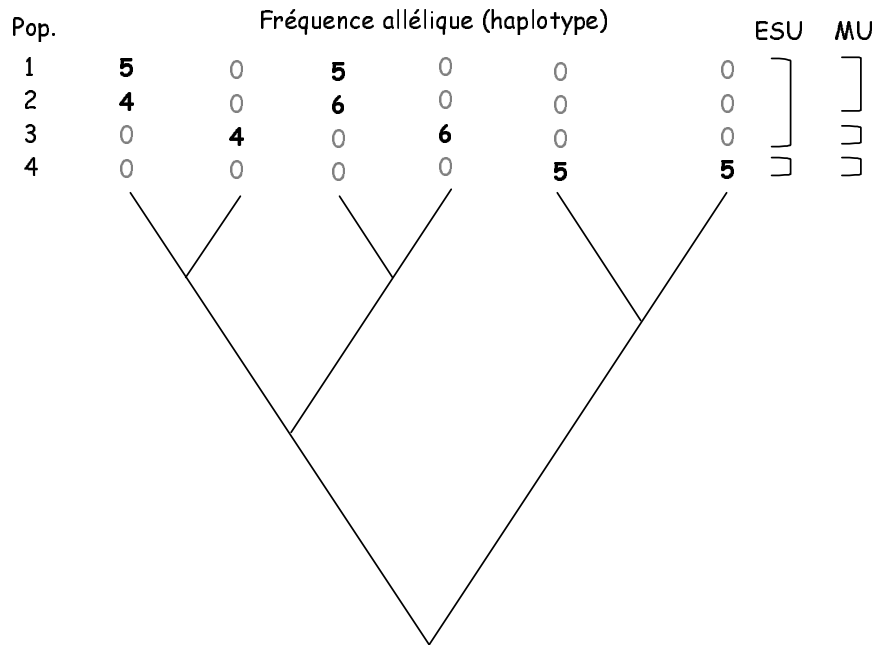


FIGURE B.5 – Distribution hypothétique d'allèles (haplotypes) parmi quatre populations montrant les fréquences alléliques et la phylogénie en relation avec les ESU et les MU (d'après MORITZ 1994). Les populations 1 et 2 ont des fréquences alléliques similaires et font partie de la même unité de gestion (MU). Les populations 3 et 4 ont des allèles fixés et seule la population 4 montre une séparation phylogénétique. La population 3 fait donc partie de la même unité « évolutivement » significative (ESU) que les populations 1 et 2 alors que la population 4 constitue une ESU à part.

il existe des cas où une signification statistique ne correspond pas forcément à une signification biologique et *vice versa* (HEDRICK 1999, CRANDALL *et al.* 2000, voir TABLEAU B.3). Il est alors impératif de prendre en compte des caractères écologiques car il peut exister une monophylie génétique (imposée par la définition de MORITZ qui ignore les différences adaptatives) pour des populations distinctes (CRANDALL *et al.* 2000).

CRANDALL *et al.* (2000) recommandent de ne plus utiliser la terminologie de ESU mais de revenir à un concept d'espèce, consistant en des populations avec des niveaux variés de flux génique et évoluant à travers la dérive génétique et la sélection naturelle. Ils placent dans cette nouvelle vision, l'idée d'« échangeabilité » tant au niveau génétique (effet du flux génique) qu'au niveau écologique (effet de la sélection naturelle ou de la dérive génétique sur la différenciation écologique des populations) dans des perspectives récentes et historiques. Suivant la combinaison de ces critères, les espèces seront considérées comme distinctes ou non (voir FIGURE 1 dans CRANDALL *et al.* 2000). Enfin, l'utilisation du seul critère génétique pour la définition d'unité de gestion (population) peut être mal appliqué. TAYLOR et DIZON (1999) préconisent de définir ces unités de gestion au cas par cas avec, en amont, des questions des gestionnaires qui doivent trouver des réponses précises dans les analyses des scientifiques.

TABLEAU B.3 – Relation entre la signification statistique et la signification biologique d'une divergence phylogénétique.

	<i>Divergence statistiquement significative</i>	<i>Divergence non statistiquement significative</i>
<i>Signification biologique</i>	OK	Pas assez de puissance (ex : allozymes) Fort flux génique réduisant la détection d'une divergence au niveau génétique
<i>Pas de signification biologique</i>	Utilisation de nombreux marqueurs très polymorphes montrant des différences significatives mais sans réelle signification biologique.	OK

B.4.2 Diagnostiquer les problèmes génétiques et donner des priorités de conservation

Le choix de l'unité de conservation n'est pas une tâche suffisante pour commencer un plan de conservation (génétique). Il est nécessaire de diagnostiquer quels problèmes génétiques sont présents et ensuite de donner des priorités de conservation sur telles ou telles unités. Il faut alors porter son attention sur plusieurs critères (notamment des facteurs influençant le déclin et la vulnérabilité des populations).

B.4.2.1 La variabilité génétique

Le maintien de la **diversité génétique** est un objectif majeur en conservation car cette diversité représente le potentiel évolutif d'une espèce ou d'une population. La variation entre individus de traits quantitatifs d'histoire de vie, principalement liés à la reproduction, est le déterminant principal du potentiel évolutif. Cette variation est en partie due à la variation génétique (au niveau de plusieurs loci : QTL, *Quantitative Traits Loci*) en interaction avec les facteurs d'environnement. La baisse de la variation génétique peut alors influencer la variation au niveau des caractères quantitatifs d'autant plus si ces traits sont très héréditaires.

La réduction de la **variation génétique quantitative** peut être due à la stochasticité génétique influençant les petites populations (consanguinité, dérive génétique), mais aussi à l'impact de croisements entre populations de fitness différentes (hétérosis et *outbreeding depression*) et à la translocation d'individus d'un environnement à un autre (FRANKHAM *et al.* 2002, p.97).

La variation au niveau des QTL est très difficile à mesurer car il est nécessaire d'avoir à la fois des marqueurs génétiques liés à des caractères quantitatifs et des mesures de traits d'histoire de vie. Les techniques moléculaires mesurant la diversité génétique au niveau de l'ADN ou des protéines doivent présenter un certain niveau de variabilité dont on attend qu'il reflète le niveau de variation qui influence l'adaptation future ou la fitness individuelle (mais voir PFENDER *et al.* 2000). Le taux d'hétérozygotie, le nombre d'allèles par locus et la fréquence des marqueurs tels que les microsatellites, les allozymes, *etc.* sont proposés pour rendre compte de la variabilité génétique d'une population (voir par exemple HARTL et PUCEK 1992, TAYLOR *et al.* 1994, PAETKAU *et al.* 1998, GUTIÉRREZ-ESPELETA *et al.* 1999).

La variabilité génétique au sein d'une espèce ou d'une population peut alors servir de critère

pour donner une priorité à la conservation d'une espèce ou d'une population car elle est liée à une augmentation du risque de déclin et d'extinction (*e.g.* LACY 1987). Il est ainsi nécessaire de conserver et de mettre en oeuvre des moyens pour protéger des populations qui perdent de la diversité génétique (en introduisant par exemple des individus d'autres populations afin de diversifier le patrimoine génétique), mais il faut aussi concentrer les efforts sur des populations qui ont assez de variabilité génétique pour assurer une adaptation future et une expansion réussie des populations (HEDRICK et MILLER 1992).

B.4.2.2 La consanguinité

La consanguinité est très importante à prendre en compte et à mesurer car elle peut avoir des conséquences néfastes sur le devenir des petites populations (effets sur la fitness ou des traits liés à la fitness, HEDRICK et KALINOWSKI 2000). Plusieurs méthodes expérimentales ont été développées pour examiner les effets de la consanguinité, et plusieurs approches statistiques ont été utilisées pour quantifier l'étendue de la dépression de consanguinité (voir HEDRICK et KALINOWSKI 2000). Afin de quantifier le niveau de consanguinité, on peut calculer le coefficient de consanguinité F correspondant à la probabilité que l'individu possède deux allèles identiques par descendance (suivant une ségrégation mendélienne des allèles) et mesure donc le degré d'apparentement existant entre ses parents. Par exemple, si les parents sont non liés, $F=0$; si les parents sont frère/soeur, mère/fils ou père/fille, $F=0.25$; si les parents sont des cousins de premier degré, $F=0.0625$. Pour déterminer l'étendue de la consanguinité dans une population entière, on calcule le coefficient de consanguinité moyen de tous les individus de cette population. Cette approche nécessite la détermination du pedigree des individus concernés (BALLOU 1983). Dans les cas de populations captives, les pedigrees peuvent être connus avec précision, et les coefficients de consanguinité calculés. Dans le cas de populations sauvages, il est souvent très difficile d'obtenir les relations exactes entre les individus. Néanmoins, les marqueurs très polymorphes permettent de déterminer certaines relations, ou de calculer des coefficients d'apparentement qui peuvent rendre compte de la consanguinité (*e.g.* LYNCH 1988, QUELLER et GOODNIGHT 1989, GOODNIGHT et QUELLER 1999).

Le niveau de consanguinité peut aussi être évalué par le déficit en hétérozygotes comparé aux proportions à l'équilibre de Hardy-Weinberg ou grâce au rapport entre l'hétérozygotie observée et attendue.

Les effets de la consanguinité (dépression de consanguinité) ne peuvent être détectés qu'en comparant les fitness de reproduction de population consanguine et non-consanguine, ce qui, la plupart du temps, est très difficile à réaliser dans des populations sauvages en danger.

B.4.2.3 L'effectif et la taille efficace d'une population

Les effets de la consanguinité et de la dérive génétique dans une population sont liés à la taille de cette population. L'effectif d'une population, et plus particulièrement l'effectif efficace génétique de la population (LANDE et BARROWCLOUGH 1987) peut être ainsi un indicateur de la vulnérabilité de cette population. Pour SOULÉ (1980b), le rôle majeur de la génétique de la conservation est de développer des critères pour déterminer la taille de population (ou l'aire minimum) qui doit être fournie pour maintenir le potentiel adaptatif et la fitness.

L'effectif réel d'une population (*census population size*) peut être utilisée pour appréhender les priorités de conservation, mais elle sera beaucoup moins informative que la taille efficace sur le statut de vulnérabilité de la population car des facteurs (*e.g.* sex ratio, systèmes d'apparie-

ment) peuvent influencer ce statut. Il existe plusieurs approches pour appréhender la taille d'une population (NUNNEY et ELAM 1994) : comptage de tous les individus, fluctuations du nombre d'individus ou du nombre de juvéniles au cours des ans, comptage des adultes reproductifs uniquement. Néanmoins, de nouvelles approches sont développées afin d'estimer la taille réelle d'une population (nombre total d'individus) grâce aux marqueurs moléculaires et aux méthodes non-invasives (KOHN *et al.* 1999, WAITS et LEBERG 1999).

L'estimation de la **taille efficace** (N_e) de population n'est pas aisée car (*i*) on définit généralement plusieurs types de taille efficace génétique (par exemple le changement génétique, exprimé dans la définition de la Section B.1.2.3.3.3, p.140, peut être le taux de consanguinité ou la perte de diversité génétique conduisant au calcul de la taille efficace de consanguinité ou propre *-eigenvalue*), et (*ii*) il existe plusieurs approches pour le calcul de la taille efficace. Ce sont les approches génétiques qui sont les plus appropriées pour des buts de conservation. Ces méthodes sont nombreuses et s'appuient sur des données génétique (généralement les fréquences alléliques) à partir d'un ou deux échantillons (voir par exemple LANDE et BARROWCLOUGH 1987, NUNNEY et ELAM 1994, SCHWARTZ *et al.* 1998, CRANDALL *et al.* 1999, SCHWARTZ *et al.* 1999, BERTHIER *et al.* 2001). Il existe aussi des approches démographiques du calcul de la taille efficace. Néanmoins, ces dernières nécessitent certains paramètres²⁶ difficiles à estimer pour des populations naturelles. Les approches démographiques permettent, en outre, de prédire les changements de N_e avec des conditions changeantes (NUNNEY et ELAM 1994) ce que ne peuvent faire les approches génétiques.

Dans l'optique de développer des critères de conservation, certains auteurs ont fourni des seuils de taille efficace pour des populations génétiquement viables. Par exemple, SOULÉ (1980b) et FRANKLIN (1980) préconisent une taille de population de plus de 50 individus pour que la population puissent faire fasse, à court terme, à la consanguinité. FRANKLIN (1980) et LANDE et BARROWCLOUGH (1987) préconisent un effectif efficace d'au moins 500 pour que la population soit viable à long terme (100-1000 ans) en faisant face à la dérive génétique. D'autres auteurs ont donné d'autres seuils : 1000-5000 (LANDE 1995) ou 500-1000 (FRANKLIN et FRANKHAM 1998) afin de retenir le potentiel évolutif; 100000-1000000 pour LANDE et BARROWCLOUGH (1987) pour retenir la diversité génétique sur un locus pendant 100000-1000000 ans.

Néanmoins, ces nombres ne doivent pas être considérés comme des nombre magiques (WILCOX 1986, SOULÉ 1987a, WOODRUFF 1989) car, d'une part, les seuils sont donnés suivant le but et la durée de viabilité exigée pour la population et d'autre part ces nombres ont été déterminés à partir de bases empiriques (notamment à partir de constats d'éleveurs sur la tolérance à la consanguinité pour les premiers seuils cités) ou théoriques (avec différents paramètres, valeurs de paramètres et modèles) et ils ne sont donc pas applicables à toute population sauvage (SIMBERLOFF 1988, LYNCH et LANDE 1998).

Afin de fixer des seuils de taille efficace mais surtout les probabilités d'extinction des populations permettant une viabilité à long terme des population sauvages, des analyses de viabilité (PVA) peuvent être réalisées en modélisant et simulant sur un grand nombre de réplifications le devenir de la population²⁷. Ainsi, on peut connaître avec une certaine confiance (voir BROOK *et al.* 2000; voir les restrictions de ELLNER *et al.* 2002) la viabilité d'une population et si elle

²⁶la taille efficace est sensible à plusieurs facteurs comme la sex-ratio inégale (la taille efficace sera limitée par le sexe le moins représenté), la variation de la taille des familles, la fluctuation de la taille de population, la consanguinité.

²⁷la détermination de taille minimale par PVA semble sensible aux erreurs dans les données démographiques et les modèles ne sont pas assez précis pour faire de telles prédictions (REED *et al.* 2002).

conservera ses chances de viabilité à long terme, et quels sont les facteurs les plus importants dans son histoire. Il faut savoir qu'en l'état actuel des modèles, seule la consanguinité peut être intégrée en tant que stochasticité génétique. Des règles de décision telles que celles présentées ci-dessus peuvent être utilisées dans les programmes de reproduction. Dans ce type de programme, il est généralement admis de conserver 90% de la diversité génétique pendant 100 ans (FRANKHAM *et al.* 2002, pp.352-353). La taille efficace pour atteindre ce but dépendra notamment du temps de génération ou du taux de consanguinité de la population en question. L'efficacité dépendra aussi de la vitesse à laquelle les populations atteindront cette taille (le taux de croissance devra être le plus grand possible pour ne pas prolonger les effets de fondation).

B.4.2.4 Le problème des hybrides

Un des avantages des méthodes génétiques en biologie de la conservation est la reconnaissance des hybrides. L'**hybridation** est le croisement d'individus provenant de populations génétiquement différentes sans prendre en compte le statut taxonomique des populations (voir RHYMER et SIMBERLOFF 1996, ALLENDORF *et al.* 2001). Le problème de l'hybridation dans les populations naturelles est lié aux effets de cette hybridation sur les populations mises en jeu, menant quelques fois à l'extinction d'une ou des deux populations (RHYMER et SIMBERLOFF 1996). En effet, l'**introgression** (le flux génique existant entre les populations dont les individus s'hybrident ; RHYMER et SIMBERLOFF 1996) peut mener au mélange complet des pools géniques et ainsi à la perturbation de complexes de gènes adaptés et à l'élimination des populations parentes (RHYMER et SIMBERLOFF 1996). Les populations parentes peuvent aussi être en danger quand la progéniture hybride a une fitness plus importante que les parents. Ceci peut mener, à long terme, à la disparition d'une ou des deux populations parentes. L'inverse peut aussi se produire quand les hybrides sont stériles et que la proportion d'accouplements hybrides est très importante par rapport aux accouplements non-hybrides (notamment dans le cas où une des deux populations parentes a un faible effectif) menant à un « gaspillage » des accouplements.

Les hybrides n'ont pas toujours été pris en compte dans les programmes de conservation, qui pouvaient préconiser simplement la destruction de ces hybrides ou bien ne préconisaient aucune mesure. Néanmoins, la reconnaissance du rôle de l'hybridation dans l'évolution (*e.g.* spéciation) a incité à prendre en compte l'hybridation en biologie de la conservation (ALLENDORF *et al.* 2001). En effet, si on se place dans un concept de conservation où l'homme doit être à l'extérieur de la nature, un taxon hybride peut être protégé s'il constitue un taxon historiquement stable et s'il ne provient pas d'un phénomène anthropogénique (DOWLING *et al.* 1992b). Bien que, pour l'identification des hybrides, l'utilisation de caractères morphologiques est importante (*e.g.* NOWAK 1992, PHILLIPS et HENRY 1992) les méthodes génétiques peuvent constituer des outils très utiles pour discuter du statut d'hybride (voir par exemple WAYNE 1992). Outre l'utilisation de l'ADN mitochondrial (séquençage ou PCR-RFLP) pour différencier les populations et montrer le statut d'hybride d'une population ou d'une espèce (*e.g.* LEHMAN *et al.* 1991, WAYNE et JENKS 1991, PILGRIM *et al.* 1998, RANDI *et al.* 2000), de nouvelles approches statistiques ont été développées et permettent d'assigner des individus à certaines populations à partir des données de génotypage multilocus (voir par exemple PAETKAU *et al.* 1995, BERTORELLE et EXCOFFIER 1998, CORNUET *et al.* 1999, PRITCHARD *et al.* 2000). Ces approches statistiques et génétiques sont d'autant plus importantes qu'elles peuvent être réalisées à partir d'échantillons non-invasifs (ALLENDORF *et al.* 2001).

B.4.3 Obtention d'informations biologiques

Les données génétiques ont émergé comme une composante essentielle de la biologie de la conservation pour l'interprétation de l'histoire passée, de l'état présent et futur des populations (O'BRIEN 1994). En dehors de la définition des priorités et des unités de conservation, la génétique a aussi un rôle dans la description des espèces et des populations. La génétique peut alors servir à obtenir des informations sur la démographie (à long terme) des populations (MILLIGAN *et al.* 1994) mais aussi sur les comportements et autres traits d'histoire de vie. Par exemple, la variabilité génétique et sa répartition géographique permettent d'appréhender la structuration des populations surtout grâce à l'ADN mitochondrial (HARRISON 1989, PALUMBI *et al.* 1991, TABERLET 1996) mais aussi avec l'ADN nucléaire (*e.g.* PAETKAU *et al.* 1995, GOLDSTEIN *et al.* 1999, PAETKAU *et al.* 1999). La génétique a permis aussi d'étudier des aspects de certaines populations qui ne pouvaient pas être étudiés avec les méthodes classiques d'observation. Par exemple, les patrons de reproduction de plusieurs espèces ont été étudiés *via* les outils génétiques. La plupart des systèmes de reproduction (monogamie, polygynie, polyandrie ; HUGHES 1998) ont été étudiés. L'apport le plus connu de la génétique dans l'étude des modes de reproduction est celui de la démonstration que certaines espèces « socialement » monogames ne le sont pas génétiquement car il existe des fertilisations et des paternités hors couple. De plus, le parasitisme intraspécifique de nid (c'est-à-dire le fait que certaines femelles pondent leurs oeufs dans le nids de femelles voisines) a aussi été mis en évidence chez certains oiseaux (HUGHES 1998). Ces fertilisations hors couple et ce parasitisme ne peuvent, dans certaines cas, être déterminés par l'observation, et seules des analyses génétiques peuvent rendre compte des parentés réelles (*i.e.* génétiques). Au-delà de cette application, les relations génétiques entre individus d'une population sont une information essentielle et primordiale dans la compréhension du fonctionnement des populations. Le suivi des individus par leur identification génétique peut permettre de connaître la composition des groupes, les relations entre les membres de ces groupes, la dispersion entre différents groupes ou populations, les territoires et aires d'habitats, *etc.* Ces informations peuvent être particulièrement utiles pour suivre les individus après une réintroduction ou un renforcement de population. Par exemple, la population d'ours français a été renforcée dans les Pyrénées par la translocation d'individus slovènes (TABERLET et BOUVET 1992, TABERLET *et al.* 1993, 1997). Les ours des Pyrénées sont régulièrement échantillonnés (poils) pour suivre l'évolution de la petite population ainsi reconstituée. Les individus sont ainsi suivis pour connaître les territoires et les relations de parenté sont déterminées pour connaître les nouvelles naissances et les couples à l'origine de ces naissances, *etc.* (TABERLET *et al.* 1997).

D'autre part, l'identification individuelle peut être très utile pour réguler et combattre le braconnage et le commerce illicite d'espèces. En effet, un individu peut être identifié génétiquement et être comparé avec des indices de terrain pour connaître l'origine (population protégée ou non) de l'individu (RUTH et FAIN 1993, MANEL *et al.* 2002) grâce à la comparaison des haplotypes mitochondriaux ou à des tests d'assignation (si les fréquences alléliques de loci nucléaires de plusieurs populations de référence sont connues).

Références

- ALLENDORF, F., LEARY, R., SPRUELL, P., et WENBURG, J. (2001). The problems with hybrids : setting conservation guidelines. *Trends in Ecology and Evolution*, 16(11) : 613–622.
- ASHLEY, M. et DOW, B. (1994). The use of microsatellite analysis in population biology : Background, methods and potential applications. In SCHIERWATER, B., STREIT, B., WAGNER, G., et DESALLE, R., editors, *Molecular Ecology and Evolution : Approaches and Applications*, pages 185–202. Birkhäuser Verlag, Basel.
- BALLOU, J. (1983). Calculating inbreeding coefficients from pedigrees. In SCHONEWALD-COX, C., CHAMBERS, S., MACBRYDE, B., et THOMAS, L., editors, *Genetics and Conservation*, pages 509–520. Benjamin/Cummings, Menlo Park.
- BARBAULT, R. (1997). *Biodiversité, Introduction à la Biologie de la Conservation*. Les Fondamentaux. Hachette, Paris.
- BERTHIER, P., BEAUMONT, M., CORNUET, J., et LUIKART, G. (2001). Likelihood-based estimation of the effective population size using temporal changes in allele frequencies : a genealogical approach. *Genetics*, 160(2) : 741–751.
- BERTORELLE, G. et EXCOFFIER, L. (1998). Inferring admixture proportions from molecular data. *Molecular Biology and Evolution*, 15(10) : 1298–1311.
- BEURIER, J.-P. (2001). Le vivant sauvage et le droit international. *Biofutur*, 211 : 42–45.
- BOITANI, L. (1984). Genetic considerations on wolf conservation in Italy. *Boll. Zool.*, 51 : 367–373.
- BOITANI, L. (1992). Wolf research and conservation in Italy. *Biological Conservation*, 61 : 125–132.
- BOWEN, B. (1999). Preserving genes, species, or ecosystems ? Healing the fractured foundations of conservation policy. *Molecular Ecology*, 8 : S5–S10.
- BREITENMOSER, U. (1998). Large predators in the alps : the fall and rise of man’s competitors. *Biological Conservation*, 83(3) : 279–289.
- BROOK, B., O’GRADY, J., CHAPMAN, A., BURGMAN, M., AKÇAKAYA, H., et FRANKHAM, R. (2000). Predictive accuracy of population viability analysis in conservation biology. *Nature*, 404 : 385–387.
- BRUFORD, M. et WAYNE, R. (1993). Microsatellites and their application to population genetics

- studies. *Current Opinions in Genetics and Development*, 3 : 939–943.
- CAUGHLEY, G. et GUNN, A. (1996). *Conservation Biology in Theory and Practice*. Blackwell Science, Cambridge.
- CORNUET, J.-M., PIRY, S., LUIKART, G., ESTOUP, A., et SOLIGNAC, M. (1999). New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics*, 153 : 1989–2000.
- CRANDALL, K., BININDA-EMONDS, O., MACE, G., et WAYNE, R. (2000). Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution*, 15(7) : 290–295.
- CRANDALL, K., POSADA, D., et VASCO, D. (1999). Effective population sizes : missing measures and missing concepts. *Animal Conservation*, 2 : 317–319.
- CROZIER, R. (1992). Genetic diversity and the agony of choice. *Biological Conservation*, 61 : 11–15.
- DASZAK, P., CUNNINGHAM, A., et HYATT, A. (2000). Emerging infectious diseases of wildlife - threats to biodiversity and human health. *Science*, 287 : 443–449.
- DE BEAUFORT, F. (1987). Le loup en France : éléments d'écologie historique. In SFPEPM , editor, *Encyclopédie des Carnivores de France*, volume 32. SFPEPM.
- DIAMOND, J. (1989). Overview of recent extinctions. In WESTERN, D. et PEARL, M., editors, *Conservation for the Twenty-first Century*, pages 37–41. Oxford University Press, New York.
- DIMMICK, W., GHEDOTTI, M., GROSE, M., MAGLIA, A., MEINHARDT, D., et PENNOCK, D. (1999). The importance of systematic biology in defining units of conservation. *Conservation Biology*, 13(3) : 653–660.
- DOWLING, T., DEMARAIS, B., MINCKLEY, W., DOUGLAS, M., et MARSCH, P. (1992a). Use of genetic characters in conservation biology. *Conservation Biology*, 6(1) : 7–8.
- DOWLING, T., MINCKLEY, W., DOUGLAS, M., MARSH, P., et DEMARAIS, B. (1992b). Response to Wayne, Nowak, and Phillips and Henry : use of molecular characters in conservation biology. *Conservation Biology*, 6(4) : 600–603.
- EDWARDS, A., CIVITELLO, A., HAMMOND, H., et CASKEY, C. (1991). Dna typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics*, 49 : 746–756.
- EHRENFELD, D. (1970). *Biological Conservation*. Modern Biology Series. Holt, Rinehart, and Winston, Inc., New York.
- EHRENFELD, D. (1987). Editorial. *Conservation Biology*, 1(1) : 6–7.
- EHRlich, P. et WILSON, E. (1991). Biodiversity studies : science and policy. *Science*, 253 : 758–761.
- ELLNER, S., FIEBERG, J., LUDWIG, D., et WILCOX, C. (2002). Precision of population viability analysis. *Conservation Biology*, 16(1) : 258–261.
- ELMHAGEN, B. et ANGERBJÖRN, A. (2001). The applicability of metapopulation theory to large mammals. *Oikos*, 94 : 89–100.
- ENGLAND, P. (1997). *Conservation genetics of population bottlenecks*. Phd thesis, Macquarie

University.

- ERWIN, T. (1991). An evolutionary basis for conservation strategy. *Science*, 253 : 750–752.
- FORAN, R., CROOKS, K., et MINTA, S. (1997). Species identification from scats : an unambiguous genetic method. *Wildlife Society Bulletin*, 25(4) : 835–839.
- FRANCISCO, L., LANGSTON, A., MELLERSH, C., NEAL, C., et OSTRANDER, E. (1996). A class of highly polymorphic tetranucleotide repeats for canine genetic mapping. *Mammalian Genome*, 7 : 359–362.
- FRANKEL, O. et SOULÉ, M. (1981). *Conservation and Evolution*. Cambridge University Press, Cambridge.
- FRANKHAM, R. (1995). Effective population size/adult population size ratios in wildlife : a review. *Genetical Research*, 66 : 95–107.
- FRANKHAM, R., BALLOU, J., et BRISCOE, D. (2002). *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge.
- FRANKHAM, R., LEES, K., MONTGOMERY, M., ENGLAND, P., LOWE, E., et BRISCOE, D. (1999). Do population size bottlenecks reduce evolutionary potential? *Animal Conservation*, 2 : 255–260.
- FRANKHAM, R. et RALLS, K. (1998). Inbreeding leads to extinction. *Nature*, 392 : 441–442.
- FRANKLIN, I. (1980). Evolutionary change in small populations. In SOULÉ, M. et WILCOX, B., editors, *Conservation Biology : an Evolutionary-Ecological Perspective*, pages 135–150. Sinauer, Sunderland.
- FRANKLIN, I. et FRANKHAM, R. (1998). How large must populations be to retain evolutionary potential? *Animal Conservation*, 1 : 69–73.
- GILPIN, M. (1987). Spatial structure and population vulnerability. In SOULÉ, M., editor, *Viable Populations for Conservation*, pages 125–139. Cambridge University Press, Cambridge.
- GILPIN, M. et SOULÉ, M. (1986). Minimum viable populations : the processes of population extinction. In SOULÉ, M., editor, *Conservation Biology : the Science of Scarcity and diversity*, pages 13–34. Sinauer Associates, Sunderland.
- GOLDSTEIN, D., ROEMER, G., SMITH, D., REICH, D., BERGMAN, A., et WAYNE, R. (1999). The use of microsatellite variation to infer population structure and demographic history in natural model system. *Genetics*, 151 : 797–801.
- GOLDSTEIN, P., DESALLE, R., AMATO, G., et VOGLER, A. (2000). Conservation genetics at the species boundary. *Conservation Biology*, 14(1) : 120–131.
- GOODNIGHT, K. et QUELLER, D. (1999). Computer software for performing likelihood tests of pedigree relationship using genetic markers. *Molecular Ecology*, 8 : 1231–1234.
- GUTIÉRREZ-ESPELETA, G., KALINOWSKI, S., BOYCE, W., et HEDRICK, P. (1999). Genetic variation and population structure in desert bighorn sheep : implications for conservation. *Conservation Genetics*, 1 : 3–15.
- HAIG, S. (1998). Molecular contributions to conservation. *Ecology*, 79(2) : 413–425.
- HARRISON, R. (1989). Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends in Ecology and Evolution*, 4(1) : 6–11.

- HARTL, G. et PUCEK, Z. (1992). Genetic depletion in the European bison (*Bison bonasus*) and the significance of electrophoretic heterozygosity for conservation. *Conservation Biology*, 8(1) : 167–174.
- HEDRICK, P. (1999). Perspective : highly variable loci and their interpretation in evolution and conservation. *Evolution*, 53(2) : 313–318.
- HEDRICK, P. et KALINOWSKI, S. (2000). Inbreeding in conservation biology. *Annual Review of Ecology and Systematics*, 31 : 139–162.
- HEDRICK, P. et MILLER, P. (1992). Conservation genetics : techniques and fundamentals. *Ecological Applications*, 2 : 30–46.
- HUGHES, C. (1998). Integrating molecular techniques with field methods in studies of social behavior : a revolution results. *Ecology*, 79(2) : 383–399.
- JANZEN, D. (1986). The external threat. In SOULÉ, M., editor, *Conservation Biology : The Science of Scarcity and Diversity*, pages 286–303. Sinauer Associates, Sunderland.
- JOHNSON, W., EIZIRIK, E., ROELKE-PARKER, M., et O'BRIEN, S. (2001). Applications of genetic concepts and molecular methods to carnivore conservation. In GITTELMAN, J., FUNK, S., McDONALD, D., et WAYNE, R., editors, *Carnivore Conservation*, pages 335–358. Cambridge University Press, Cambridge.
- KIMURA, M. et CROW, J. (1963). The measurement of effective population number. *Evolution*, 17 : 279–288.
- KOHN, M., YORK, E., KAMRADT, D., HAUGHT, G., SAUVAJOT, R., et WAYNE, R. (1999). Estimating population size by genotyping faeces. *Proceedings of the Royal Society of London, Serie B*, 266 : 657–663.
- LACY, R. (1987). Loss of genetic diversity from managed populations : interacting effects of drift, mutation, immigration, selection, and population subdivision. *Conservation Biology*, 1(2) : 143–158.
- LANDE, R. (1988). Genetics and demography in biological conservation. *Science*, 241 : 1455–1460.
- LANDE, R. (1995). Mutation and conservation. *Conservation Biology*, 9 : 782–791.
- LANDE, R. et BARROWCLOUGH, G. (1987). Effective population size, genetic variation, and their use in population management. In SOULÉ, M., editor, *Viable Populations for Conservation*, pages 87–123. Cambridge University Press, Cambridge.
- LANZA, R., DRESSER, B., et DAMIANI, P. (2000). Cloning Noah's ark. *American Scientist*, 87(5) : 84–89.
- LEAKEY, R. et LEWIN, R. (1995). *The Sixth Extinction : Biodiversity and its Survival*. Phoenix, London.
- LEGENDRE, S. (2000). The use of software in conservation biology. *Revue d'Ecologie (Terre et Vie)*, Suppl.7 : 123–125.
- LEHMAN, N., EISENHAWER, A., HANSEN, K., MECH, L., PETERSON, R., GOGNAN, P., et WAYNE, R. (1991). Introgression of coyote mitochondrial DNA into sympatric North American gray wolf populations. *Evolution*, 45(1) : 104–119.

- LYNCH, M. (1988). Estimation of relatedness by DNA fingerprinting. *Molecular Biology and Evolution*, 5(5) : 584–599.
- LYNCH, M. et LANDE, R. (1998). The critical effective size for a genetically secure population. *Animal Conservation*, 1 : 70–72.
- LYNCH, M. et O'HELY, M. (2001). Captive breeding and the genetic fitness of natural populations. *Conservation Genetics*, 2 : 363–378.
- MAGUIRE, L., SEAL, U., et BRUSSARD, P. (1987). Managing critically endangered species : the Sumatran rhino as a case study. In SOULÉ, M., editor, *Viable Populations for Conservation*, pages 141–158. Cambridge University Press, Cambridge.
- MANEL, S., BERTHIER, P., et LUIKART, G. (2002). Detecting poaching : identifying the origin of individuals using bayesian assignment tests and multi-locus genotyping. *Conservation Biology*, in press.
- MANN, C. (1991). Extinction : are ecologists crying wolf? *Science*, 253 : 736–738.
- MAYR, E. (1942). *Systematics and the origin of species*. Columbia University Press, New York.
- MCARTHUR, R. et WILSON, E. (1967). *The Theory of Island Biogeography*. Princeton University Press, London.
- MCNAUGHTON, S. (1989). Ecosystems and conservation in the twenty-first century. In WESTERN, D. et PEARL, M., editors, *Conservation for the Twenty-First Century*, pages 109–120. Oxford University Press, New York.
- MÉDAIL, F. (2001). Les *hot spots* de biodiversité : un outil pour la conservation? *Biofutur*, 211 : 37.
- MELLERSH, C., LANGSTON, A., ACLAND, G., FLEMING, M., RAY, K., WIEGAND, N., FRANCISCO, L., GIBBS, M., AGUIRRE, G., et OSTRANDER, E. (1997). A linkage map of the canine genome. *Genomics*, 46 : 326–336.
- MILLER, P. et LACY, R. (1999). *VORTEX : a stochastic simulation of the extinction process. Version 8 user's manual*. Conservation Breeding Specialist Group (SSC/IUCN), Apple Valley, MN.
- MILLIGAN, B., LEEBENS-MACK, J., et STRAND, A. (1994). Conservation genetics : beyond the maintenance of marker diversity. *Molecular Ecology*, 3 : 423–435.
- MILLS, L., CITTA, J., LAIR, K., SCHWARTZ, M., et TALLMON, D. (2000). Estimating animal abundance using non-invasive DNA sampling : promise and pitfalls. *Ecological Applications*, 10(1) : 283–294.
- MILLS, L. et SMOUSE, P. (1994). Demographic consequences of inbreeding in remnant populations. *American Naturalist*, 144 : 412–431.
- MORIN, P. (2000). Preservation of DNA from endangered species. *Science*, 289 : 725–726.
- MORIN, P., CHAMBERS, K., BOESCH, C., et VIGILANT, L. (2001). Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Molecular Ecology*, 10 : 1835–1844.
- MORITZ, C. (1994). Applications of mitochondrial DNA analysis in conservation : a critical review. *Molecular Ecology*, 3 : 401–411.

- MOROWITZ, H. (1991). Balancing species preservation and economic considerations. *Science*, 253 : 752–754.
- MYERS, N. (1979). *The Sinking Ark, A New Look at the Problem of Disappearing Species*. Pergamon Press, Oxford.
- NEFF, M., BROMAN, K., MELLERSH, C., RAY, K., ACLAND, G., AGUIRRE, G., ZIEGLE, J., OSTRANDER, E., et RINE, J. (1999). A second-generation genetic linkage map of the domestic dog, *canis familiaris*. *Genetics*, 151 : 803–820.
- NOSS, R., QUICKLEY, H., HORNOCKER, M., MERRIL, T., et PACQUET, P. (1996). Conservation biology and carnivore conservation in the Rocky Mountains. *Conservation Biology*, 10(4) : 949–963.
- NOWAK, R. (1992). The red wolf is not a hybrid. *Conservation Biology*, 6(4) : 593–595.
- NUNNEY, L. et ELAM, D. (1994). Estimating the effective population size of conserved populations. *Conservation Biology*, 8(1) : 175–184.
- O'BRIEN, S. (1994). Perspective on conservation genetics. In SCHIERWATER, B., STREIT, B., WAGNER, G., et DESALLE, R., editors, *Molecular Ecology and Evolution : Approaches and Applications*, pages 275–280. Birkhäuser Verlag, Basel.
- OLSON, S. (1989). Extinction on islands : man as a catastrophe. In WESTERN, D. et PEARL, M., editors, *Conservation for the Twenty-first Century*, pages 50–53. Oxford University Press, New York.
- PAETKAU, D. (1999). Using genetics to identify intraspecific conservation units : a critique of current methods. *Conservation Biology*, 13(6) : 1507–1509.
- PAETKAU, D., AMSTRUP, S., BORN, E., CALVERT, W., DEROCHE, A., GARNER, G., MESSIER, F., STIRLING, I., TAYLOR, M., et WIIG, . (1999). Genetic structure of the world's polar bear populations. *Molecular Ecology*, 8 : 1571–1584.
- PAETKAU, D., CALVERT, W., STIRLING, I., et C., S. (1995). Microsatellite analysis of population structure in Canadian polar bear. *Molecular Ecology*, 4 : 347–354.
- PAETKAU, D., WAITS, L., CLARKSON, P., CRAIGHEAD, L., VYSE, E., WARD, R., et STROBECK, C. (1998). Variation in genetic diversity across the range of North American brown bears. *Conservation Biology*, 12(2) : 418–429.
- PALUMBI, S., MARTIN, A., KESSING, B., et McMILLAN, W. (1991). Detecting population structure using mitochondrial DNA. In HOELZEL, A., editor, *Genetic Ecology of Whales and Dolphins (Rep. Int. Whal. Commn., Special Issue 13)*, pages 271–278. Cambridge University Press, Cambridge, UK.
- PFENDER, M., SPITZE, K., HICKS, J., MORGAN, K., LATTA, L., et LYNCH, M. (2000). Lack of concordance between genetic diversity estimates at the molecular and quantitative-trait levels. *Conservation Genetics*, 1 : 263–269.
- PHILLIPS, M. et HENRY, V. (1992). Comments on red wolf taxonomy. *Conservation Biology*, 6(4) : 596–599.
- PILGRIM, K., BOYD, D., et FORBES, S. (1998). Testing for wolf-coyote hybridization in the Rocky Mountains using mitochondrial DNA. *Journal of Wildlife Management*, 62(2) : 683–689.

- POINAR, H. et EGLINTON, G. (2000). Preservation of DNA from endangered species. *Science*, 289 : 726.
- POULLE, M., LEQUETTE, B., HOUARD, T., et DAHIER, T. (1995). Le retour du loup en France, dans le Mercantour. *Faune de Provence*, 16 : 96–99.
- POULLE, M. et LONGCHAMP, F. (1997). Régime alimentaire des loups dans le Mercantour. *Faune de Provence*, 18 : 33–40.
- PRITCHARD, J., STEPHENS, M., et DONNELLY, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155 : 945–959.
- QUELLER, D. et GOODNIGHT, K. (1989). Estimating relatedness using genetic markers. *Evolution*, 43 : 237–275.
- RANDI, E., LUCCHINI, V., CHRISTENSEN, M., MUCCI, N., FUNK, S., DOLF, G., et LOESCHCKE, V. (2000). Mitochondrial DNA variability in Italian and East European wolves : detecting the consequences of small population size and hybridization. *Conservation Biology*, 14(2) : 464–473.
- RAUP, D. (1988). Diversity crises in the geological past. In WILSON, E. et PETER, F., editors, *Biodiversity*, pages 51–57. National Academy Press, Washington DC.
- REED, J., MILLS, L., DUNNING JR, J., MENGES, E., MCKELVEY, K., FRYE, R., BEISSINGER, S., ANSTETT, M.-C., et MILLER, P. (2002). Emerging issues in population viable analysis. *Conservation Biology*, 16(1) : 7–19.
- RHYMER, J. et SIMBERLOFF, D. (1996). Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*, 27 : 83–109.
- ROJAS, M. (1992). The species problem and the conservation : what are we protecting. *Conservation Biology*, 6 : 170–178.
- RUTH, J. et FAIN, S. (1993). The 'individualization' of large North American mammals. In PENA, S., CHAKRABORTY, R., EPPLER, J., et JEFFREYS, A., editors, *DNA Fingerprinting : State of the Science*, pages 429–436. Birkhäuser Verlag, Basel.
- RYDER, O. (1986). Species conservation and systematics : the dilemma of subspecies. *Trends in Ecology and Evolution*, 1 : 9–10.
- RYDER, O., MCLAREN, A., BRENNER, S., ZHANG, Y.-P., et BENIRSCHKE, K. (2000). DNA banks for endangered animal species. *Science*, 288 : 275–276.
- SACCHERI, I., KUUSSAARI, M., KANKARE, M., VIKMAN, P., FORTELIUS, W., et HANSKI, I. (1998). Inbreeding and extinction in a butterfly metapopulation. *Nature*, 392 : 491–494.
- SANTINI, F. et ANGULO, A. (2001). Assessing conservation biology priorities through the development of biodiversity indicators. *Rivista di Biologica/Biology Forum*, 94 : 259–275.
- SCHWARTZ, C. et SEBER, A. (1999). A review of estimating animal abundance. *Statistical Science*, 14(4) : 427–456.
- SCHWARTZ, M., TALLMON, D., et LUIKART, G. (1998). Review of DNA-based census and effective population size estimators. *Animal Conservation*, 1 : 293–299.
- SCHWARTZ, M., TALLMON, D., et LUIKART, G. (1999). Using genetic to estimate the size of wild populations : many methods, much potential, uncertain utility. *Animal Conservation*,

2 : 321–323.

- SHAFFER, M. (1981). Minimum population sizes for species conservation. *Bioscience*, 31 : 131–134.
- SHAFFER, M. (1987). Minimum viable populations : coping with uncertainty. In SOULÉ, M., editor, *Viable Populations for Conservation*, pages 69–86. Cambridge University Press, Cambridge.
- SIMBERLOFF, D. (1988). The contribution of population and community biology to conservation science. *Annual Review of Ecology and Systematics*, 19 : 473–511.
- SIMPSON, G. (1961). *Principles of animal taxonomy*. Columbia University Press, New York.
- SMITH, F., MAY, R., PELLEW, R., JOHNSON, T., et WALTER, K. (1993). How much do we know about the current extinction rate? *Trends in Ecology and Evolution*, 8(10) : 375–378.
- SOLTIS, P. et GITZENDANNER, M. (1999). Molecular systematics and the conservation of rare species. *Conservation Biology*, 13(3) : 471–483.
- SOULÉ, M. (1980a). The effects of fragmentation. In SOULÉ, M., editor, *Conservation Biology, the Science of Scarcity and Diversity*, pages 233–236. Sinauer Associates, Sunderland.
- SOULÉ, M. (1980b). Thresholds for survival : maintaining fitness and evolutionary potential. In SOULÉ, M. et WILCOX, B., editors, *Conservation Biology : an Evolutionary-Ecological Perspective*, pages 111–124. Sinauer, Sunderland.
- SOULÉ, M. (1985). What is conservation biology? *Bioscience*, 35 : 727–734.
- SOULÉ, M. (1986a). Conservation biology and the 'real world'. In SOULÉ, M., editor, *Conservation Biology, the Science of Scarcity and Diversity*, pages 1–12. Sinauer Associates, Sunderland.
- SOULÉ, M. (1986b). The fitness and viability of populations. In SOULÉ, M., editor, *Conservation Biology, the Science of Scarcity and Diversity*, pages 13–18. Sinauer Associates, Sunderland.
- SOULÉ, M. (1987a). History of the society for conservation biology : how and why we got here. *Conservation Biology*, 1(1) : 4–5.
- SOULÉ, M. (1987b). Introduction. In SOULÉ, M., editor, *Viable Populations for Conservation*. Cambridge University Press, Cambridge.
- SOULÉ, M. (1989). Conservation biology in the twenty-first century : summary and outlook. In WESTERN, D. et PEARL, M., editors, *Conservation for the Twenty-first Century*, pages 297–303. Oxford University Press, New York.
- SOULÉ, M. (1991). Conservation : tactics for constant crisis. *Science*, 253 : 744–750.
- SPELLENBERG, I. (1992). *Evaluation and Assessment for Conservation*. Chapman & Hall, London.
- SUNDQVIST, A., ELLEGREN, H., OLIVIER, M., et VILÀ, C. (2001). Y chromosom haplotyping in Scandinavian wolves (*Canis lupus*) based on microsatellite markers. *Molecular Ecology*, 10 : 1959–1966.
- SUTHERLAND, W. (1998). The importance of behavioral studies in conservation biology. *Animal Behaviour*, 56 : 801–809.

- TABERLET, P. (1996). The use of mitochondrial DNA control region sequencing in conservation genetics. In SMITH, T. et WAYNE, R., editors, *Molecular Genetic Approaches in Conservation*, pages 125–142. Oxford University Press, New York.
- TABERLET, P. et BOUVET, J. (1992). Bear conservation genetics. *Nature*, 358 : 197.
- TABERLET, P., CAMARRA, J., GRIFFIN, S., UHRÈS, E., HANOTTE, O., WAITS, L., DUBOIS-PAGANON, C., BURKE, T., et BOUVET, J. (1997). Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Molecular Ecology*, 6 : 869–876.
- TABERLET, P., GRIFFIN, S., GOOSSENS, B., QUESTIAU, S., MANCEAU, V., ESCARAVAGE, N., WAITS, L., et BOUVET, J. (1996). Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, 24(16) : 3189–3194.
- TABERLET, P., MATTOCK, H., DUBOIS-PAGANON, C., et BOUVET, J. (1993). Sexing free-ranging brown bears *Ursus arctos* using hairs found in the field. *Molecular Ecology*, 2 : 399–403.
- TAYLOR, A., SHERWIN, W., et WAYNE, R. (1994). Genetic variation of microsatellites loci in a bottlenecked species : the northern hairy-nosed wombat *Lasiornhinus kreftii*. *Molecular Ecology*, 3 : 277–290.
- TAYLOR, E. et DIZON, A. (1999). First policy then science : why a management unit based solely on genetic criteria cannot work. *Molecular Ecology*, 8 : S11–S16.
- VALIÈRE, N. (2002). L'étude de l'expansion du loup (*Canis lupus*) en France : apport des méthodes non-invasives et des marqueurs moléculaires. In CHAPRON, G. et MOUTOU, F., editors, *XXIIIème Colloque Francophone de Mammalogie*, pages 97–101, Ecole National Vétérinaire d'Alfort. SFPEM.
- VARVIO, S., CHAKRABORTY, R., et NEI, M. (1986). Genetic variation in subdivided populations and conservation genetics. *Heredity*, 57 : 189–198.
- WAITS, J. et LEBERG, P. (1999). Advances in the use of molecular markers for studies of population size and movement. *Transactions of the North American Wildlife Society*, pages 191–201.
- WAITS, L., LUIKART, G., et TABERLET, P. (2001). Estimating the probability of identity among genotypes in natural populations : cautions and guidelines. *Molecular Ecology*, 10 : 249–256.
- WALKER, B. (1989). Diversity and stability in ecosystem conservation. In WESTERN, D. et PEARL, M., editors, *Conservation for the Twenty-first Century*, pages 121–132. Oxford University Press, New York.
- WAPLES, R. (1991). Pacific salmon, *Oncorhynchus spp.*, and the definition of 'species' under the endangered species act. *Mar. Fish. Rev.*, 53 : 11–22.
- WAYNE, R. (1992). On the use of morphologic and molecular genetic characters to investigate species status. *Conservation Biology*, 6(4) : 590–592.
- WAYNE, R. et JENKS, S. (1991). Mitochondrial DNA analysis implying extensive hybridization of the endangered red wolf *Canis rufus*. *Nature*, 351 : 565–568.
- WESTERN, D. (1989). Conservation biology. In WESTERN, D. et PEARL, M., editors, *Conservation for the Twenty-first Century*, pages 31–36. Oxford University Press, New York.

- WESTERN, D. (1992). The biodiversity crisis : a challenge for biology. *Oikos*, 63 : 29–38.
- WESTERN, D., PEARL, M., PIMM, S., WALKER, B., ATKINSON, I., et WOODRUFF, D. (1989). An agenda for conservation action. In WESTERN, D. et PEARL, M., editors, *Conservation for the Twenty-first Century*, pages 304–323. Oxford University Press, New York.
- WILCOX, B. (1986). Extinction models and conservation. *Trends in Ecology and Evolution*, 1(2) : 46–48.
- WILCOX, B. et MURPHY, D. (1985). Conservation strategy : the effects of fragmentation on extinction. *American Naturalist*, 125 : 879–887.
- WILEY, E. (1978). The evolutionary species concept reconsidered. *Systematic Zoology*, 27 : 17–26.
- WILSON, E. (1988). The current state of biological diversity. In WILSON, E. et PETER, F., editors, *Biodiversity*, pages 3–18. National Academy Press, Washington DC.
- WOODRUFF, D. (1989). The problems of conserving genes and species. In WESTERN, D. et PEARL, M., editors, *Conservation for the Twenty-first Century*, pages 76–88. Oxford University Press, New York.
- WOODS, J., PAETKAU, D., LEWIS, D., MCLELLAN, B., PROCTOR, M., et STROBECK, C. (1999). Genetic tagging of free-ranging black and brown bears. *Wildlife Society Bulletin*, 27(3) : 616–627.
- WRIGHT, S. (1931). Evolution in Mendelian populations. *Genetics*, 16 : 97–159.
- ZIMEN, E. et BOITANI, L. (1975). Number and distribution of wolves in Italy. *Zeitschrift für Säugetierkunde*, 40 : 102–112.

Article I

Article II

Article III

Article IV

Article V

Article VI

Article VII

Article VIII

Index

- ADN mitochondrial, 22, 48, 49, 89, 98
- Allèles nuls, 36, 95
- Approche multi-tubes, 41, 56, 94
 - Optimisation, 54, 61, 92
- Biodiversité, 8, 134, 136
 - Crise, 9
- Comportement, 11, 50, 98, 151
- Consanguinité, 10, 23, 99, 140, 156
- Conservation, 8, 9, 141
 - Biologie, 9, 143, 147
 - Génétique, 10, 152
- Contaminations, 34, 62, 63, 67, 95
- Diversité génétique, 11, 20, 99, 155
- Echantillonnage, 30, 40, 49
 - Conditions d', 62, 88, 94
 - Effort d', 44, 91, 92
 - non-invasif, 30, 40, 49
- Effectif (taille) de population, 20, 27, 50, 58, 72, 84, 97, 138, 156
 - efficace, 140, 157
- Erreurs de géotypage, 32
 - Correction, 54, 78
 - Taux, 36, 37, 49, 55, 56, 59, 62, 68
 - Estimation, 31, 82, 90
- Etude pilote, 4, 62, 88–92, 96, 98
- Fèces, 26, 30, 31, 33, 37, 40, 48, 60, 63–65, 72, 89, 96, 98, 124
- Faux allèles, 34, 37, 56, 57
- Fragmentation, 8, 137
- GEMINI, 54, 91
 - Utilisation, 59, 61, 73
- GIMLET, 44, 74–76, 95, 97
- Identification, 24, 40
 - individuelle, 20, 21, 49, 54, 76, 91, 98
 - spécifique, 21, 48, 49
- Loup, 40, 48, 124
 - Fèces, 125, 128
 - Loci étudiés, 126, 127
 - Poils, 40
 - Probabilité d'identité, 127
 - Sélection loci, 125
 - Taux d'erreurs, 128, 130
- Microsatellites, 19, 21–25
- Parenté, 11, 23, 77, 151
- Perte d'allèle, 35, 37, 55–57, 67, 78
- Poils, 26, 30–33, 37, 48–50, 63, 65, 89, 96
- Polymorphisme, 20, 24, 59, 68
- Probabilité d'identité, 24, 58, 68, 82
- Structure
 - génétique, 22, 26
 - sociale, 20, 22, 27, 44, 83, 94, 99
 - spatiale, 94