

Biologie

Chapitre 8

La génétique des populations comme outil en épidémiologie

T. DE MEEÛS^(a) et K. MCCOY^(a)

1. Introduction 278
2. Qu'est-ce qu'un marqueur génétique? 278
3. Concepts de base en génétique des populations 280
4. Outils statistiques d'aide à la décision 288
5. Des exemples concrets d'application en épidémiologie 290
6. Avenir de la génétique des populations en épidémiologie 299

^a UMR 2724 Génétique et Évolution des Maladies Infectieuses, Montpellier, France

1. Introduction

Les organismes **parasites** représentent une part significative des espèces décrites (De Meeûs et Renaud 2002) et malgré la récente explosion des études moléculaires, les études populationnelles des systèmes hôtes-parasites sont encore beaucoup trop rares (Criscione et coll. 2005). Les agents pathogènes et leurs **vecteurs** sont, en effet, des organismes dont l'écologie, le mode de reproduction (reproduction sexuée ou asexuée, croisements au hasard ou **autofécondation** partielle ou totale, etc.), les déplacements ou la taille de **populations** restent difficiles (voir impossibles) d'accès par observation directe. Or, la compréhension de l'épidémiologie d'une maladie, l'évaluation des risques d'invasion ou d'épidémie, la perception du risque de diffusion de **gènes** de résistance ou de l'effet d'une stratégie de lutte sur les populations cibles, tout cela ne peut se passer d'une connaissance minimale du fonctionnement des populations concernées. L'écologie des populations de parasites et de leurs vecteurs peut, la plupart du temps, n'être inférée que par les méthodes que Slatkin (Slatkin 1985) appelle « indirectes » (Nadler 1995, De Meeûs et coll. 2002a,b, McCoy et coll. 2003, De Meeûs et coll. 2004a, 2007a). Ces méthodes se caractérisent par l'utilisation de marqueurs moléculaires (génétiques) **polymorphes** (variables) et par l'étude des variations de ces marqueurs dans et entre un certain nombre de groupes d'individus prédéfinis comme sous-populations. L'hypothèse de base sous-tendue est que la distribution de la variabilité génétique reflète les paramètres écologiques cités plus haut.

Le principal objectif de ce chapitre est d'évoquer les méthodes disponibles à ce jour, leur mérite, leur puissance ainsi que leurs limites, et les concepts et hypothèses biologiques de base qui permettent leur mise en œuvre. Ces notions ainsi que leur utilité épidémiologique seront en partie abordées au travers d'exemples concrets issus des travaux de notre équipe. Pour des revues plus techniques, le lecteur averti pourra se reporter aux revues disponibles (Tibayrenc 1999, Criscione et Blouin 2005, Criscione et coll. 2005, De Meeûs et coll. 2006, 2007a).

2. Qu'est-ce qu'un marqueur génétique ?

Un marqueur génétique correspond à une portion de l'ADN de l'organisme étudié, ou un sous-produit codé par cet ADN (comme une protéine). L'ADN est la molécule porteuse de l'hérédité chez tous les êtres vivants (à l'exception de certains virus à ARN). Pour que ce marqueur soit utilisable, il est important que cette portion d'ADN reste la même (même localisation dans le génome) d'un individu à l'autre, d'où le terme **locus**. Il importe de regarder ce qui se passe sur

cette même portion d'ADN chez plusieurs individus de différents échantillons (spatialement et/ou temporellement différents). Un locus peut correspondre à un gène codant comme c'est le cas pour les loci enzymatiques (ou iso-enzymatiques), mais il peut aussi correspondre à une zone non codante et donc *a priori* non fonctionnelle de l'ADN, comme c'est le cas de la plupart des microsatellites. Cependant, rappelons qu'un locus non codant peut se trouver dans un **intron**, c'est-à-dire dans un gène, et peut donc subir des phénomènes sélectifs par son association à ce gène. On appelle ce phénomène l'**auto-stop** (ou *hitchhiking* en anglais). Cela reste valable pour un locus situé en dehors de tout gène mais à proximité d'un locus **sélectionné** ou simplement parce que le régime de reproduction de l'organisme étudié limite ou empêche la **recombinaison** entre loci. La variabilité sélectionnée ne reflète pas uniquement les événements démographiques, mais aussi l'histoire sélective de la population. Dans ce chapitre, nous considérerons un organisme eucaryote **diploïde** (comme une tique ou une moustique), c'est-à-dire que chaque portion d'ADN (chaque locus) dispose de deux représentants (**allèles**) par individu, exception faite des marqueurs liés au sexe (sur un chromosome sexuel). Il est préférable d'analyser les populations naturelles au travers de plusieurs loci (plus il y en a et mieux c'est!) de nature identique (microsatellites, iso-enzymes, RFLP, etc.). Pour être informatif, autrement dit pour fournir des indices démographiques, un locus doit être variable (on dit qu'il est polymorphe), c'est-à-dire qu'il présente plusieurs allèles dans le groupe d'individus échantillonnés. Ce locus doit être **neutre** et **codominant** (les **hétérozygotes** diffèrent des **homozygotes**). Les mérites et différences entre les différents marqueurs disponibles ont été largement étudiés (Roderick 1996, Caterino et coll. 2000, Sunnucks 2000) et nous n'aborderons pas cette question ici. En règle générale, les loci microsatellites sont les marqueurs les plus utilisés, en particulier chez les parasites. Ces loci sont considérés comme étant en général très polymorphes, codominants, abondants dans tous les génomes et relativement aisés à manipuler (Lehmann et coll. 1996). Grâce à l'utilisation de l'outil **PCR** (*Polymerase Chain Reaction*) et des derniers progrès faits en la matière, il est possible de travailler à partir de quantités infimes de matériel biologique⁽¹⁾. Ainsi, les oocystes de *Plasmodium falciparum* (moins de 50 µm) peuvent être analysés individuellement au niveau de sept marqueurs microsatellites (Razakandrainibe et coll. 2005). Le lecteur désireux de s'informer davantage sur les propriétés des microsatellites et leurs diverses applications est invité à consulter l'ouvrage édité par Goldstein et Schlötterer (Goldstein et Schlötterer 1999).

¹ Une explication très didactique peut être consultée sur le site <http://www.ens-lyon.fr/RELIE/PCR/principe/principe.htm>

3. Concepts de base en génétique des populations

3.1 Fréquences génotypiques et alléliques

Considérons que nous travaillons sur un locus codominant chez un organisme diploïde. Admettons que nous ayons génotypé au niveau d'un locus à deux allèles (1 et 2) N individus de cet organisme dans un site particulier que nous supposons contenir une population. Parmi ces N individus, N_{11} se sont révélés être de **génotype** 11, N_{12} de génotype 12 et N_{22} de génotype 22. Notez que N est nécessairement égal à $N_{11} + N_{12} + N_{22}$. Soient p_1 et p_2 les fréquences des allèles 1 et 2 respectivement dans l'échantillon de N individus. Il y a $2N$ allèles en tout puisque nous sommes chez des diploïdes. Les fréquences des allèles 1 et 2 dans l'échantillon sont donc :

$$p_1 = \frac{2N_{11} + N_{12}}{2N} \quad (1)$$

et

$$p_2 = \frac{2N_{22} + N_{12}}{2N} = 1 - p_1 \quad (2)$$

Notez que ces valeurs sont aussi des estimations des fréquences alléliques de la population échantillonnée et que, grâce à la codominance du marqueur, nous n'avons pas eu à faire d'hypothèse pour estimer ces mêmes fréquences alléliques (en réalité, une hypothèse a ici été faite, mais laquelle?)⁽²⁾

3.2 Qu'est ce que l'équilibre de Hardy-Weinberg ?

Le modèle de population idéale imaginé par Hardy (Hardy 1908) et Weinberg (Weinberg 1908) forme la clé de voûte de toute la génétique des populations. Il s'agit du modèle nul⁽³⁾ par excellence, permettant, en relâchant chaque hypothèse une à une et en observant les conséquences, d'apprécier quelles informations seront pertinentes pour inférer l'écologie des populations étudiées.

² Cette hypothèse est que l'échantillonnage n'a pas modifié les fréquences des allèles dans la population réelle.

³ Ce qui ne signifie pas qu'il ne vaut rien mais qu'il sert de point de départ, comme l'hypothèse nulle H_0 en statistiques.

Il s'agit d'une population diploïde de taille infinie, sans **mutation**, sans migration, sans **sélection**, à reproduction sexuée obligatoire, **hermaphrodite** autocompatible, panmictique (rencontre au hasard des gamètes) à générations non-chevauchantes et non-subdivisées. Dans ce contexte, pour un locus comptant deux allèles (1 et 2) de fréquences p_1 et $p_2=1-p_1$ respectivement, on peut alors calculer la probabilité de former des individus diploïdes 11, 12 et 22 à la génération suivante. Sachant que nous puisons dans une population infinie, la probabilité d'obtenir un individu 11 est exactement égale à la probabilité de piocher un allèle 1, soit p_1 par cette même probabilité, ce qui donne p_1^2 . On aura donc p_1^2 11 et de la même façon p_2^2 22. Il existe ensuite deux manières de construire un **hétérozygote** 12, soit en piochant un allèle 1 (p_1) puis un allèle 2 (p_2), soit inversement en piochant d'abord 2 puis 1, soit une probabilité de $p_1p_2+p_2p_1$. Il en résulte que nous aurons p_1^2 11, p_2^2 22 et $2p_1p_2$ 12. Il est facile de constater que la somme de ces fréquences fait 1 (il ne manque donc aucune valeur). Il est facile également de montrer, en appliquant les équations (1) et (2), que les fréquences des allèles n'ont pas été modifiées. Comme la population est infinie, ces proportions génotypiques chez les **zygotes** représentent aussi celles des adultes de la génération précédente. Nous avons donc invariablement les mêmes fréquences d'allèles et la même structure génotypique p_1^2 , $2p_1p_2$ et p_2^2 qui correspondent à ce qu'on appelle l'équilibre de Hardy-Weinberg. Ce qui se passe quand les hypothèses du modèle sont relâchées une à une est illustré dans l'encadré 8.1.

Encadré 8.1 Déviations du modèle de Hardy-Weinberg et ses conséquences

1) Si la population est de taille finie N , les fréquences alléliques vont varier aléatoirement d'une génération à l'autre car on a peu de chance d'obtenir une représentation à l'identique des fréquences par tirage aléatoire de $2N$ gènes. Les fréquences alléliques changent donc, et ce d'autant plus brutalement que N est petit. C'est ce que l'on appelle la dérive génétique. L'observation de la variation des fréquences alléliques d'une génération à l'autre donne donc une indication sur l'**effectif efficace** (N_e) de la population suivie (Waples 1989).

2) La mutation correspond aux erreurs de copie lors de la duplication de l'ADN, durant la construction des gamètes par exemple. Il existe plusieurs modèles de mutation: le **IAM** (*infinite allele model*), le **KAM** (*k allele model*), le **SMM** (*stepwise mutation model*), ainsi que des modèles intermédiaires comme le **TPM** (*two phases model*). Quoiqu'il en soit, la mutation va bien évidemment modifier les fréquences alléliques des loci qu'elle affecte. Il faut noter cependant que les taux de mutations sont en général assez bas (sauf pour quelques microsatellites) et que la mutation seule ne peut donc pas représenter une force très puissante de l'évolution des populations. Il n'y aurait néanmoins pas d'évolution sans mutation car c'est la seule source de nouveauté héritable, et, associée à la dérive et à la sélection, la mutation représente en effet la clé de l'évolution.

3) Les populations naturelles ne sont pas isolées totalement les unes des autres. Elles reçoivent donc des propagules (adultes, larves, spores, gamètes, etc.) venant d'autres populations plus ou moins éloignées et présentant, à l'ensemble du génome, des fréquences d'allèles plus ou moins différentes. Ils peuvent être haploïdes ou diploïdes. La migration peut être forte. Elle a tendance à homogénéiser les populations entre elles (génétiquement). C'est donc une force potentiellement majeure de l'évolution des populations.

4) La sélection, parce qu'elle modifie la probabilité de transmission et/ou de survie des gènes, conduit à modifier les fréquences des allèles d'une génération à l'autre. Néanmoins, elle n'affecte qu'un ou quelques loci, alors que les phénomènes démographiques que nous allons rechercher à mettre en évidence affectent le génome en entier et les loci neutres d'une façon homogène. Des marqueurs neutres doivent donc tous donner le même signal, exception faite de problèmes techniques (allèles nuls) et de systèmes particuliers de reproduction (**clonalité** partielle). La sélection est abordée plus en détail dans d'autres chapitres de cet ouvrage (cf. Chapitres 6, 7 et 10).

5) Il existe plusieurs formes de reproduction non-panmictique. La clonalité, ou reproduction asexuée (ou **végétative**), et certaines formes de **parthénogenèse** apomictique reproduisent à l'identique les individus qui la pratiquent. Elle ne peut donc rien changer à elle seule. Cependant, en populations finies subissant la dérive génétique, avec un taux de mutation constant, des excès d'hétérozygotes sont attendus par rapport aux fréquences génotypiques de Hardy-Weinberg à l'échelle de l'ensemble du génome (Balloux et coll. 2003, De Meeûs et Balloux 2005, De Meeûs et coll. 2006). L'**autofécondation** n'est bien sûr possible que chez des organismes hermaphrodites. Elle contribue à fortement diminuer la proportion d'hétérozygotes par rapport à l'attendu de Hardy-Weinberg à tous les loci (signature génomique). C'est ce qui se passe par exemple chez le ver solitaire *Taenia solium* (De Meeûs et coll. 2003) qui est obligé de s'autoféconder pour se reproduire car il est retrouvé presque toujours seul dans l'intestin de son hôte. Les croisements systématiques entre apparentés conduisent à un résultat similaire mais moins rapide. Chez la guêpe parasitoïde *Nasonia vitripennis*, la femelle pond plusieurs œufs (frères-sœurs) dans une même chenille et les adultes qui en émergent ensuite doivent rapidement s'accoupler. Ceci a tendance à favoriser les croisements entre frères et sœurs (Shuker et coll. 2004). L'homogamie conduit les individus de même génotype, ou leurs gamètes, à s'accoupler entre eux. Les conséquences sont identiques à l'autofécondation (baisse d'hétérozygotie), sauf qu'elles ne concernent que les gènes responsables du caractère et ceux qui leur sont liés (**auto-stop**). Des caractères tels que la taille à la maturité sexuelle ou la résistance aux pathogènes ont presque toujours, ou au moins en partie, un déterminisme génétique. Or, il est prouvé que dans de nombreuses espèces, ces caractères conditionnent l'appariement (*assortative mating*) des partenaires sexuels (Thomas et coll. 1995). L'hétérogamie, ou préférence de la différence, ne peut exister sans sélection fréquence dépendante. Notons qu'elle ne concerne que les loci responsables du caractère. Dans ce cas, on s'attend à un excès d'hétérozygotes sur tous les loci impliqués. Le polymorphisme du complexe majeur d'histocompatibilité (CMH; HLA chez l'homme) est soumis à ce genre de sélection (Wedekind et Penn 2000). Enfin, chez les organismes à sexes séparés, il n'y a jamais panmixie totale car les gènes contenus dans les femelles ne peuvent s'associer qu'à ceux contenus dans les mâles. Chez les espèces **dioïques** ou chez les hermaphrodites auto-incompatibles, on s'attend à détecter des excès d'hétérozygotes si la population est de taille réduite.

6) Si les générations se chevauchent, ce phénomène seul ne va pas créer une perturbation, mais combiné à la dérive, il va créer une hétérogénéité dans la population (effet **Wahlund**, voir plus loin) qui va se traduire par un déficit en hétérozygote proportionnel à l'intensité de la dérive. Cela aura aussi tendance à minimiser certaines mesures de différenciation entre populations telles que le F_{ST} .

7) La plupart des populations naturelles sont subdivisées en sous-populations. Au niveau du polymorphisme neutre, ces sous-populations vont subir la dérive de façon d'autant plus indépendante que le taux de migration est faible et d'autant plus rapide que la taille des sous-populations est réduite. Une différenciation va s'établir entre sous-populations de telle sorte que les fréquences alléliques vont diverger. Les fréquences génotypiques totales ne seront plus conformes à Hardy-Weinberg car il y aura un déficit global en hétérozygote ou effet Wahlund (Wahlund 1928). Il peut s'établir entre migration, dérive et mutation un équilibre stable neutre. Notons que si la subdivision correspond à un **pas japonais** ou un **voisinage**, on observera un **cline** neutre. Il est important de garder ceci à l'esprit car on pourra toujours trouver en population naturelle un gradient écologique qui calquera un cline génétique. Il faudra donc toujours s'assurer que la dérive n'explique pas en totalité ce cline avant d'invoquer des causalités sélectives. Le schéma de différenciation entre sous-échantillons et son intensité offrent donc des informations précieuses sur l'importance de la dispersion et la taille des sous-populations.

3.3 Systèmes de reproduction et populations structurées

Il existe deux principaux systèmes de reproduction, la reproduction asexuée et la reproduction sexuée. La reproduction asexuée, ou clonale, correspond à la production de descendants génétiquement identiques à l'organisme parental (à la mutation **somatique** près). La reproduction **végétative** pratiquée par tous les organismes unicellulaires et certains organismes pluricellulaires en est la forme la plus emblématique. Certaines formes de **parthénogenèse** y sont également assimilables. Pour plus de détail, le lecteur pourra consulter une revue récente sur la question (De Meeûs et coll. 2007b). La reproduction sexuée peut se pratiquer sous différentes formes (cf. Encadré 8.1).

Les fréquences des génotypes et des allèles peuvent varier d'un site à l'autre en réponse au système de reproduction et à la structure des populations (c'est-à-dire la taille des sous-populations et l'intensité de la migration entre les sous-populations). On parle souvent de **métapopulation** pour caractériser cet ensemble de sous-populations connectées entre elles (cf. Chapitres 1 et 5, par exemple).

S'il n'y a qu'une seule population, c'est-à-dire que tous les individus partagent les mêmes paramètres démographiques et écologiques (en particulier la même pression de régulation), la probabilité de laisser un descendant à la génération suivante ne dépend pas de la position géographique des individus les uns par rapport aux autres.

Dans le cas où la position géographique des individus importe, la population est alors dite structurée. Plusieurs modèles peuvent rendre compte de situations plus ou moins contrastées. Ici nous ne considérerons que les deux principaux modèles de populations structurées. Le **modèle en îles** de Wright (Wright 1951) est sans doute le plus utilisé. Le modèle initial comprenait une infinité de sous-populations (encore appelées îles) de taille identique constante égale à N et à générations non-chevauchantes (les parents cessent de se reproduire quand leurs descendants commencent à le faire, et ce de telle sorte que tous les individus reproducteurs appartiennent à la même génération). À chaque génération, chaque sous-population est composée de mN migrants et de $(1-m)N$ résidents. Les migrants sont issus de parents de n'importe quelle île du modèle, alors que les résidents proviennent de parents nés sur place. Les organismes marins à long stade larvaire pélagique, tels que les crustacés ou les bivalves, ainsi que les plantes anémophiles correspondent assez bien à ce type de population. Les modèles dits en **pas japonais** ou *stepping stone* (Kimura et Weiss 1964, Slatkin 1985) stipulent que les n sous-populations ne peuvent échanger des migrants qu'entre sous-populations directement adjacentes. Ces métapopulations peuvent s'étendre sur une dimension, auquel cas chaque population n'a que deux voisines, ou bien sur deux ou sur trois dimensions. Le premier type de modèles coïncide typiquement avec les milieux côtiers, les rivières ou les écotones (milieux de bordures). Les modèles en deux dimensions rappellent les milieux continentaux, et ceux en trois dimensions les systèmes aquatiques et forestiers avec des couches ou des strates horizontales. Des modifications peuvent être apportées à ces modèles en pas japonais, en particulier les modalités de connexion entre sous-populations. À l'extrême de ce type de schémas, figure un autre modèle où l'individu lui-même devient unité. Il s'agit des **modèles en voisinage** ou *neighbourhood models* (Wright 1965, Rousset 2000, Leblois et coll. 2004) où l'apparement entre deux individus décroît progressivement avec la distance géographique qui sépare ces deux individus. Enfin, les métapopulations « vraies », au sens écologique, admettent l'extinction et/ou la recolonisation de certaines populations dans un schéma en îles ou en pas japonais.

3.4 La consanguinité, la consanguinité relative, ses différentes origines possibles et ses traductions en indices de fixation (homozygotie relative)

La **consanguinité** est une notion abstraite qui traduit le degré d'apparement qui peut exister entre deux allèles d'un même locus à l'intérieur d'un individu (si diploïde) ou entre individus d'une même sous-population ou de sous-populations distinctes. En termes formels, cela se traduit par des probabilités d'identité par descendance que nous noterons Q_i (probabilité intra-individuelle), Q_r (entre indi-

vidus d'une même sous-population) et Q_T (entre individus de sous-populations différentes) pour une structure de population à trois niveaux (individu, sous-population et total).

a) Indices de fixation ou hiérarchisation des causes de la consanguinité

Si l'on souhaite obtenir une mesure multi-allélique (plus de deux allèles), multi-échantillons et multi-locus (ce qui est souhaitable pour minimiser la variance d'estimation), on doit utiliser un estimateur de l'hétérozygotie moyenne de chaque sous-population, que l'on appelle diversité génétique notée H_s . On doit aussi définir une diversité moyenne totale, ou H_T . Autrement dit H_s mesure la diversité locale moyenne ou l'hétérozygotie moyenne (sur les allèles, loci et sous-populations) que l'on observerait si toutes les sous-populations étaient en **panmixie** locale et H_T mesure la diversité globale ou l'hétérozygotie moyenne (sur les allèles et loci) si la panmixie était globale (Nei 1987). Notons que pour deux allèles et une seule population : $H_s = H_T = 2p_1p_2$. À partir de là, on peut donner les définitions de Nei et Chesser des statistiques F de Wright (Nei et Chesser 1983) :

$$\begin{cases} F_{IS} = \frac{H_s - \overline{H}_0}{H_s} \\ F_{ST} = \frac{H_T - H_s}{H_T} \\ F_{IT} = \frac{H_T - \overline{H}_0}{H_T} \end{cases} \quad (3)$$

où \overline{H}_0 est l'hétérozygotie moyenne observée.

Idéalement, le F_{IS} représente la part relative de consanguinité (homozygotie) des individus (I) dans les sous-populations (S) due au système de reproduction. Le F_{ST} mesure la part relative de consanguinité des sous-populations (S) dans le tout (T) due à la subdivision (**effet Wahlund**), car les membres d'une même sous-population tendent à être apparentés entre eux par **dérive**. Ceci a pour effet d'entraîner la divergence entre sous-populations, raison pour laquelle le F_{ST} mesure la différenciation entre sous-populations. Enfin, le F_{IT} reflète la consanguinité des individus (I) dans le tout (T) et correspond à la résultante des deux précédents. Il est d'ailleurs aisé de démontrer à partir de (3) que les trois statistiques sont reliées par la relation :

$$(1 - F_{IT}) = (1 - F_{IS})(1 - F_{ST}) \quad (4)$$

Le F_{IS} varie de -1 (sous-population hétérozygote pour deux allèles), 0 (panmixie) à 1 (individus tous homozygotes pour différents allèles). Les valeurs négatives

(causes techniques et sélectives mises à part) reflètent un mode clonal de reproduction, une différenciation entre partenaires sexuels (zone d'hybrides F1, biais de dispersion sexe spécifique) ou une faible taille de populations d'autoincompatibles (comme les vers monogènes), **dioïques** ou de champignons **hétérothalliques**. Les valeurs positives correspondent à un mode consanguin de reproduction (autofécondation, croisements entre apparentés). Le F_{ST} varie de 0 (homogénéité des fréquences) à 1 (sous-populations toutes fixées pour l'un ou l'autre des allèles présents). Le F_{IT} varie de -1 à 1.

Maintenant, en rappelant que i) $Q_I = 1 - \overline{H}_0$ est la probabilité d'identité de deux allèles d'un même locus au sein d'un même individu, que ii) $Q_S = 1 - H_s$ est la probabilité de tirer deux allèles identiques de deux individus différents de la même sous-population, et que iii) $Q_T = 1 - H_T$ est la probabilité de tirer deux allèles identiques de deux individus de deux sous-populations différentes, alors nous pouvons donner les formules généralisées des F -statistiques pour un degré 3 de subdivision (individu, sous-population et total) (Rousset 2004):

$$\begin{cases} F_{IS} = \frac{Q_I - Q_S}{1 - Q_S} \\ F_{ST} = \frac{Q_S - Q_T}{1 - Q_T} \\ F_{IT} = \frac{Q_I - Q_T}{1 - Q_T} \end{cases} \quad (5)$$

b) Outils d'inférences

Avec l'estimation du F_{IS} on peut inférer la proportion s d'autofécondation (Hartl et Clark 1989) ou c_{fs} de croisements frère-sœur (Chevillon et coll. 2007b):

$$\begin{cases} s = \frac{2F_{IS}}{1 + F_{IS}} \\ c_{fs} = \frac{4F_{IS}}{1 + 3F_{IS}} \end{cases} \quad (6)$$

Dans un modèle en îles, en modèle de mutation **KAM** (*K Allele Model*), et si les taux de mutation et de migration sont faibles, on montre (relativement) facilement, en partant de Rousset (Rousset 1996), que:

$$F_{ST} \approx \frac{1}{1 + 4 \left[N - \frac{(sN - 1)}{2} \left(1 - 2u \frac{K}{K-1} - 2m \right) \right] \left(m \frac{n}{n-1} + u \frac{K}{K-1} \right)} \quad (7)$$

Rappelons que N est le nombre d'individus par sous-population, s le taux d'auto-fécondation, u le taux de mutation, K le nombre maximum d'allèles possibles, m le taux de migration et n le nombre total de sous-populations. En panmixie ($s = 1/N$), pour $u \ll m$ et pour K et n grands, on retrouve la formule simplifiée (Wright 1931):

$$F_{ST} \approx \frac{1}{4Nm + 1} \quad (8)$$

D'où on tire facilement que le nombre de migrants efficaces (reproducteurs) par génération est: $Nm = (1 - F_{ST})/4F_{ST}$.

De façon alternative, si nous nous trouvons dans une configuration du type isolement par la distance (comme dans un modèle en pas japonais), Rousset (1997) montre que: $F_{ST}/(1 - F_{ST}) = a + bD_G$ en modèle en une dimension, ou $F_{ST}/(1 - F_{ST}) = a + b \text{Ln}(D_G)$ en deux dimensions, où D_G est la distance géographique séparant chaque paire de sous-populations. Ceci permet d'obtenir une estimation du produit $D\sigma^2$ de la densité D d'individus reproducteurs par le carré de la distance σ séparant les adultes du lieu où étaient leurs géniteurs: $D\sigma^2 = 1/4b$ pour un modèle en une dimension et $D\sigma^2 = 1/4\pi b$ pour deux dimensions. Si on parvient à estimer séparément la densité, on peut estimer la distance de dispersion. Pour un voisinage strict, il existe une variante interindividuelle du $F_{ST}/(1 - F_{ST})$ définie par Rousset (Rousset 2000) avec les mêmes propriétés d'inférence.

Ces possibilités d'inférence expliquent que ces statistiques sont les plus souvent utilisées pour décrire la structure génétique des populations sur le terrain (Nagylaki 1998).

c) Problèmes d'estimation

Jusqu'ici nous avons essentiellement mentionné les paramètres et leurs utilités. Cependant, comme il est impossible de pouvoir échantillonner toute l'information génétique d'une population, il est nécessaire de disposer d'outils d'estimation de ces paramètres qui ne soient pas dépendants de l'effort d'échantillonnage. Par exemple, on aura systématiquement un $F_{ST} > 0$ entre sous-échantillons récoltés dans une population panmictique par simple variance d'échantillonnage. Nous ne pouvons entrer dans les détails ici, mais il existe des estimateurs non-biaisés des statistiques F de Wright développés par Weir et Cockerham (Weir et Cockerham 1984), appelés f (pour le F_{IS}), θ (pour le F_{ST}) et F (pour le F_{IT}). Une particularité concerne θ qui peut varier de -1 à 1 , des valeurs négatives exprimant une ressemblance entre sous-populations supérieure à l'attendu sous l'hypothèse d'absence de différenciation. Ici, contrairement à F_{ST} , l'espérance de son estimateur θ est centrée sur 0 en l'absence de différenciation entre sous-populations.

d) Autres paramètres utiles

Il existe des multitudes de paramètres utiles en génétique des populations dont l'énumération ici serait fastidieuse. Nous ne mentionnerons dans ce paragraphe que les principaux d'entre eux. Dans le cas de marqueurs génétiques hypervariables comme les microsatellites, la valeur maximale du F_{ST} ne sera pas 1, car il y a plus d'allèles que de sous-échantillons (voir l'équation 7 pour u très grand). Pour pallier à ce problème, Hedrick (Hedrick 1999, 2005) propose une méthode simple pour visualiser de combien le F_{ST} observé peut être éloigné de sa valeur maximale possible que l'on observerait sans migration entre sous-populations. Dans une telle situation, et si le nombre de sous-populations est assez grand, $Q_T = 0$ (probabilité d'identité entre individus de sous-populations différentes) et, d'après (5), le F_{ST} est alors égal à $Q_s = 1 - H_s = F_{STmax}$. On peut ensuite diviser la valeur observée par cette valeur maximale afin d'avoir une meilleure appréciation (et non pas une mesure exacte) du flux de gènes échangés entre les sous-populations échantillonnées.

L'indice d'assignement (AI) correspond au calcul de la probabilité qu'un individu a , compte tenu de son génotype multilocus, d'appartenir à un échantillon donné. Cette mesure peut être utile pour déterminer le statut d'immigrant ou d'autochtone des individus. Plusieurs méthodes permettent de calculer un indice d'assignement (Paetkau et coll. 1995, Baudouin et coll. 2004, Piry et coll. 2004).

4. Outils statistiques d'aide à la décision

4.1 Bases nécessaires

Nous avons donc des paramètres, leurs estimateurs et les modèles d'inférences en termes de fonctionnement des populations. Il s'agit maintenant de décider si ces valeurs estimées ont un sens, c'est-à-dire si la déviation de l'estimation par rapport à l'attendu sous une hypothèse précise (panmixie, libre circulation des gènes) dépasse ce que la variance d'échantillonnage permet d'expliquer. L'hypothèse précise comme par exemple celle qui consiste à dire qu'il existe une panmixie dans les sous-populations ou que les sous-populations ne sont pas différentes génétiquement et qu'elles proviennent d'une population homogène s'appelle l'hypothèse nulle; et elle se note H_0 . L'hypothèse alternative, notée H_1 , peut prendre différentes formes selon les circonstances. Par exemple, pour tester s'il y a panmixie locale (test sur le F_{IS}), on peut poser en H_1 que F_{IS} a une valeur plus extrême (en + ou en -) qu'attendu sous H_0 (test bilatéral), que F_{IS} est significativement plus grand qu'attendu (test unilatéral +) ou plus petit qu'attendu (test unilatéral -). Le test

statistique effectué consistera à calculer avec quelle probabilité, appelée P -value (ou risque de première espèce α), le hasard permet-il d'expliquer la déviation observée par rapport à l'attendu. Par convention, on admet qu'à partir d'une P -value $\leq 0,05$ on rejette H_0 . Dans certaines circonstances, on peut être plus ou moins sévère, et c'est au manipulateur de prendre la décision, les statistiques n'étant ici que des outils d'aide à la décision.

4.2 Estimation des P -values par randomisations : outil privilégié en génétique des populations

Le principe sera presque toujours le même car il s'agira, grâce à l'outil informatique et à l'effort de programmation de collègues méritants, de simuler l'hypothèse nulle un grand nombre de fois (randomisations), à l'aide de l'échantillon réel, afin d'obtenir une distribution possible sous cette hypothèse du paramètre estimé ou d'une statistique appropriée. La proportion de fois qu'une valeur aussi grande ou plus grande que l'observée sera obtenue dans cette distribution correspondra à la P -value du test.

Il existe plusieurs types de randomisations (De Meeûs et coll. 2007a). Elles sont basées sur le principe de Monte Carlo. Le nom provient d'un clin d'œil de Metropolis à son collègue Stan Ulam), à propos des jeux de hasard (Metropolis 1987). Il s'agit de rééchantillonnages aléatoires (ou pseudo-aléatoires) des données. Nous ne verrons ici que les cas les plus courants car pour d'autres contextes la méthode suivra le plus souvent les mêmes principes.

Pour tester si la consanguinité relative provient du régime de reproduction au sein des sous-échantillons, il suffit de réassocier aléatoirement les allèles dans chaque sous-échantillon pour chaque locus un grand nombre de fois. À chaque randomisation on mesure le F_{IS} (ou une autre statistique pertinente) à l'aide de f . Pour un test unilatéral + où H_1 est « $F_{IS} > 0$, il y a trop d'homozygotes » la P -value du test correspond à la proportion de f randomisés trouvés aussi grands ou plus grands que l'observé. Si le test est unilatéral - avec H_1 « $F_{IS} < 0$, il n'y a pas assez d'homozygotes » la P -value du test correspond à la proportion de fois où le f randomisé est trouvé être inférieur ou égal à l'observé. Pour un test bilatéral ($H_1: F_{IS} \neq 0$) il faut combiner le résultat des deux tests précédents de telle sorte que la P -value = $P_{\min} + (1 - p_{\max})$ où P_{\min} et P_{\max} représentent la plus petite et la plus grande des deux P -values obtenues lors des deux tests unilatéraux. Le test unilatéral (+) est pertinent pour tester s'il y a une reproduction consanguine (autofécondation, croisements frères-sœurs) ou s'il y a une hétérogénéité (effet Wahlund temporel ou spatial). Le test unilatéral (-) s'applique dans des cas où sont soupçonnées une **clonalité** stricte et/ou une différenciation entre femelles et mâles.

Enfin le test bilatéral s'applique si on ne sait pas dans quelle direction le signal doit aller, comme pour typiquement les populations clonales pratiquant un peu de reproduction sexuée (De Meeüs et coll. 2006).

Pour tester si la consanguinité relative provient de la subdivision ($F_{ST} > 0$, les populations sont différenciées), ce sont les individus qui sont permutés entre les différents sous-échantillons. La statistique la plus performante est ici le G , ou logarithme népérien du rapport de maximum de vraisemblance sur le tableau des fréquences d'allèles (Goudet et coll. 1996).

Si un isolement par la distance est soupçonné, on va alors chercher s'il existe une corrélation entre la matrice des distances génétiques (cf. $F_{ST}/(1-F_{ST})$, § 2.4.b) et celles des distances géographiques (une dimension) ou leur log népérien (deux dimensions) séparant les sous-échantillons deux à deux. Dans ce cas, les mesures ne sont pas toutes indépendantes : il y a auto-corrélation car chaque sous-échantillon est utilisé autant de fois qu'il y a d'autres sous-échantillons. On utilise alors un test de Mantel (Mantel 1967) de randomisation des cases d'une des deux matrices et où la statistique utilisée est en général un coefficient de corrélation.

5. Des exemples concrets d'application en épidémiologie

5.1 Le système de reproduction de la grande douve du foie et de son vecteur, la lymnée tronquée

La grande douve du foie, *Fasciola hepatica*, est un parasite communément trouvé dans le foie de nombreuses espèces de vertébrés à sang chaud, le plus souvent des mammifères, dont l'homme (Hurtrez-Boussès et coll. 2001), et chez qui elle cause une maladie actuellement réémergente, la fasciolose. Environ 17 millions de personnes sont infectées dans le monde (Meunier et coll. 2001) et 180 millions se trouvent exposées au risque d'infection (Hurtrez-Boussès et coll. 2001). Les vers adultes sont hermaphrodites et à reproduction sexuée obligatoire chez l'hôte vertébré, alors qu'une reproduction asexuée intervient chez l'hôte intermédiaire (un escargot d'eau douce). L'hôte intermédiaire, la lymnée tronquée (*Galba truncatula*) la plupart du temps, est également hermaphrodite. Alors que la douve semble panmictique chez l'hôte vertébré ($F_{IS} \approx 0$ pour la plupart des microsatellites étudiés) (Hurtrez-Boussès et coll. 2004), son hôte intermédiaire, la lymnée tronquée, montre une préférence très marquée pour l'autofécondation. En effet, l'estimation de s par la méthode du F_{IS} (voir l'Équation 6), mesuré sur des échan-

tillons de terrain à partir de loci microsatellites, est remarquablement proche ($s = 0,94$) de celle obtenue par l'étude de croisements contrôlés au laboratoire ($s = 0,98$) (Meunier et coll. 2004). Il a également été montré à l'aide de ces marqueurs que dans l'Altiplano bolivien, où les prévalences atteignent les plus hautes valeurs connues chez des populations humaines (Mas-Coma et coll. 1999), un seul génotype de *G. truncatula* peut être trouvé. Ce phénomène, résultant sans doute d'un récent et très violent **goulot d'étranglement** (Meunier et coll. 2001), pourrait expliquer le succès de la douve dans cette région, où elle ne doit s'adapter qu'à un seul génotype de lymnée, et suggère aussi une vulnérabilité potentielle importante du vecteur à des traitements avec un risque limité d'évolution de résistances.

5.2 Les tiques et les maladies qu'elles transmettent : le cas des tiques *Ixodes ricinus* et *I. uriae*, et de la borréliose de Lyme

Ixodes ricinus et *I. uriae* intéressent les chercheurs pour beaucoup de raisons mais en particulier parce que ces tiques représentent les principaux vecteurs de la borréliose de Lyme, en Europe continentale et en Afrique du Nord (en altitude) pour *I. ricinus* et sur les côtes marines boréales et australes pour *I. uriae* (De Meeûs et coll. 2007a). La borréliose de Lyme concerne en fait un complexe de nombreuses espèces de bactéries (spirochètes), *Borrelia burgdorferi sensu lato* dont certaines sont spécifiques de différents hôtes vertébrés (lézards, oiseaux, micromammifères) et dont seulement quatre se montrent pathogènes pour les humains sains (non immunodéprimés) : *Borrelia burgdorferi*, *B. afzelii*, *B. garinii* et *B. spielmanii*. Chacune de ces espèces sont de surcroît associées à des manifestations cliniques différentes. L'utilisation de marqueurs moléculaires (microsatellites) a permis d'élargir considérablement les voies possibles de recherches pour mieux comprendre l'épidémiologie de cette maladie qui reste assez mystérieuse jusqu'à présent. Il a ainsi pu être fortement suggéré qu'*I. ricinus* disperse largement (population peu structurée) mais essentiellement par les tiques mâles (De Meeûs et coll. 2002a). Ceci est suggéré par les indices d'assignement des individus à leur échantillon d'origine, très supérieurs pour les tiques femelles par rapport aux tiques mâles. Si les femelles sont mieux assignées que les mâles, cela peut vouloir dire que les tiques femelles comprennent localement moins d'individus immigrants que les tiques mâles. Ceci n'est possible, étant donné les capacités dispersives des tiques elles-mêmes, que par un biais de spécificité d'hôte sexe-spécifique, les femelles se nourrissant aux stades larvaire et nymphal préférentiellement sur des hôtes peu dispersants (micromammifères). Il en résulterait alors un biais éga-

lement au niveau des micropathogènes transmis par les tiques des deux sexes (De Meeûs et coll. 2002a). Ceci semble être le cas pour une des espèces de borréliés, spécifique de micromammifères, et plus particulièrement de l'écureuil dans la zone testée, *Borrelia burgdorferi*, qui est effectivement retrouvée plus souvent chez les tiques femelles que chez les mâles (De Meeûs et coll. 2004b). Le schéma se complexifie avec une autre espèce testée, *Borrelia afzelii*, spécifique d'écureuils, de campagnols et de mulots (au moins dans la zone étudiée). Il apparaît en effet un biais de structuration relié à la présence de cette borrélie chez les tiques. Les tiques trouvées infectées par cette borrélie de micromammifères montrent, sur des loci microsattellites, un F_{ST} supérieur aux tiques non infectées par *B. afzelii* (De Meeûs et coll. 2004b). Cette observation peut être interprétée soit par une mortalité plus grande des tiques dispersantes si infectées par cette bactérie, soit une diminution, voire une suppression, de la dispersion des tiques porteuses de *B. afzelii*. La deuxième interprétation suppose que les tiques infectées préfèrent, en tant que larve ou nymphe, se nourrir sur des hôtes peu ou pas dispersants, tels que les mulots, les campagnols ou les écureuils, trois hôtes habituels de cette borrélie. Il y aurait donc ce qu'on nomme une manipulation parasitaire. Il est intéressant de noter également que ce phénomène est bien plus prononcé pour les tiques mâles que pour les tiques femelles. Ceci est cohérent avec la conclusion précédente qui supposait une dispersion biaisée pour les tiques femelles. Dans ce cas, les femelles dispersant déjà très peu, le fait qu'elles soient infectées par *B. afzelii* ne peut produire une grande différence, alors que pour les tiques mâles, qui dispersent beaucoup, la différence sera très grande pour les quelques mâles infectés et dispersant très peu. Ici, l'utilisation des méthodes de la génétique des populations a permis d'ouvrir des champs d'investigation totalement insoupçonnés sur le rôle du sexe des tiques dans la dissémination d'un organisme pathogène, ainsi qu'également sur le rôle des pathogènes eux-mêmes dans leur propre dispersion. Une meilleure compréhension des différents mécanismes en jeu devrait permettre à l'avenir de redessiner le tableau épidémiologique complexe de ces agents pathogènes transmis par les tiques.

Au niveau du cycle marin côtier de la maladie de Lyme, le même type d'études a permis de caractériser un complexe d'espèces ou de races d'hôtes chez *I. uriae* (le vecteur), avec semble-t-il une espèce de tique par espèce d'oiseau marin (McCoy et coll. 2001, McCoy et coll. 2005, De Meeûs et coll. 2007a). Ce phénomène ressort très clairement si on effectue une ACP, ou analyse en composantes principales, présentée sur la figure 8.1. Les analyses multivariées permettent souvent une représentation didactique de l'organisation générale de la variabilité génétique globale des échantillons génotypés. Il en existe plusieurs types, de même nature, mais offrant des possibilités différentes. Le principe est de projeter chaque individu ou sous-échantillon sur les plans des axes principaux (qui traversent le mieux

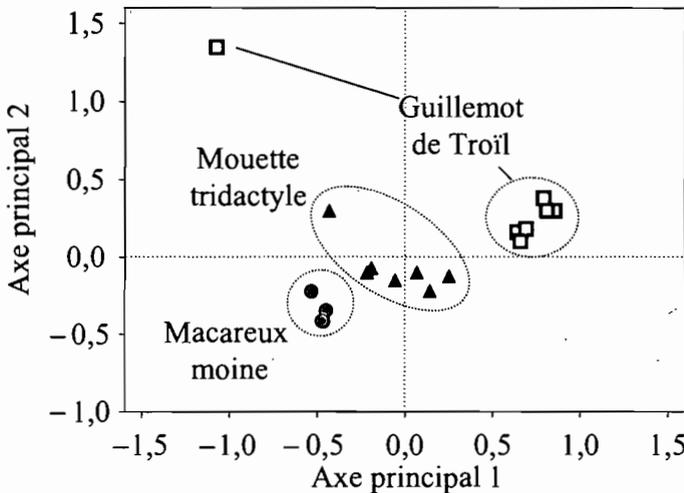
le nuage de points). L'ACP va utiliser les effectifs d'allèles dans les sous-échantillons pour transformer les données génotypiques en coordonnées multivariées.

Il y aura $\left(\sum_{i=1}^L K_i \right) - 1$ axes (L loci et K_i allèles au locus i). Notons que la distance

entre les points projetés sur les axes principaux se traduit directement en termes de différenciation relative (F_{ST}) (Guinand 1996). Ce résultat de l'ACP obtenue sur *I. uriae* suggère une forte spécificité des tiques vis-à-vis de ces différentes espèces d'oiseaux. Cette spécificité doit inévitablement entraîner des effets en cascade sur les borrélioses elles-mêmes et sur leur spécificité résultante, ainsi que sur le cloisonnement de ces différents spirochètes dont certains sont connus pour être pathogènes pour l'homme (Duneau et coll. 2008). Ces travaux posent aussi la double question de la connexion entre les cycles continentaux et marins de spirochètes de la borréliose de Lyme, d'une part ainsi que des liens possibles de transmission entre les deux hémisphères nord et sud d'autre part. À la lumière de ces résultats, les modèles épidémiologiques classiques de cette maladie sont aujourd'hui remis en question.

FIG. 8.1 | ACP ou analyse en composantes principales

ACP sur génotypes à huit loci microsatellites de différents sous-échantillons de la tique *I. uriae* prélevés sur différentes espèces d'oiseaux coloniaux marins et différents sites des côtes Atlantique-Nord. Chaque point au sein d'un type d'hôte correspond à un site différent mais certains sites contiennent plusieurs hôtes échantillonnés en sympatrie. On voit bien que c'est l'axe principal 1 (43,71% de l'inertie totale) qui discrimine le mieux les tiques d'hôtes différents alors que l'axe principal 2 (24,38% de l'inertie) semble répondre d'avantage à la position géographique. Redessiné à partir de McCoy et coll. (2005).



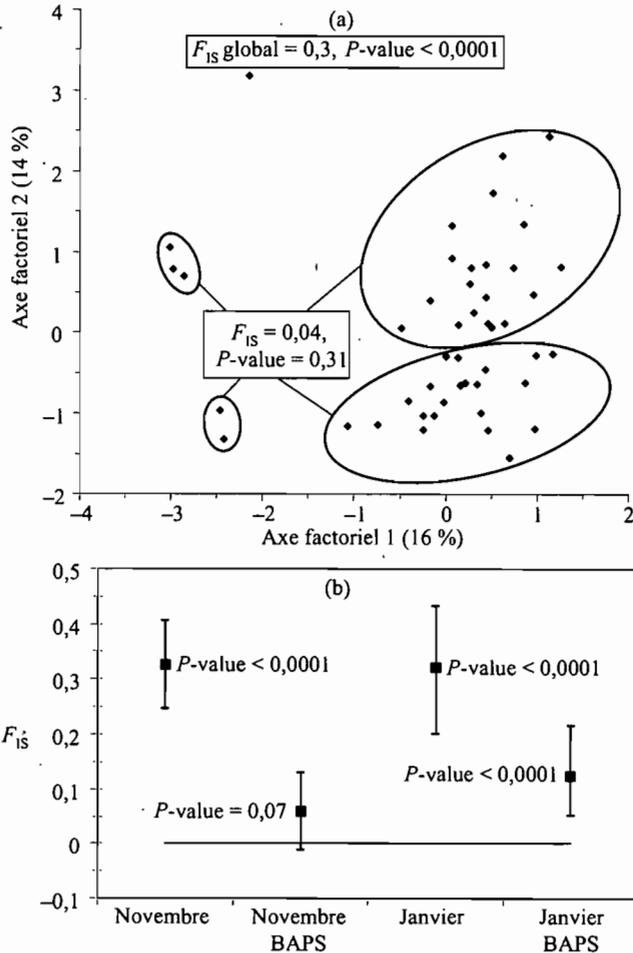
5.3 Les mouches tsétsés en Afrique

Glossina palpalis sensu lato est un complexe d'espèces de mouches tsétsés (ou glossines) d'Afrique de l'Ouest, où elles sont vectrices des trypanosomoses africaines animales et humaine, et que l'on nomme plus communément maladie du sommeil. Les premières analyses de marqueurs moléculaires (loci microsatellites) de ces glossines sur la rivière Koba au Burkina-Faso et à Bonon en Côte d'Ivoire ont tout d'abord révélé de très forts déficits en hétérozygotes que les **allèles nuls** et la dominance des allèles courts ne pouvaient entièrement expliquer (Solano et coll. 2000, Ravel et coll. 2007). L'utilisation de l'AFC (analyse factorielle des correspondances), une analyse multivariée individu-centrée, a permis de définir des sous-groupes au sein des échantillons (Solano et coll. 2000) (Figure 8.2). Une procédure de regroupement d'individus implémentée à l'aide du logiciel BAPS (comme *Bayesian Analysis of Population Structure*) (Corander et coll. 2003, Corander et coll. 2004, Corander et coll. 2006) a pu être utilisée dans une autre étude de ce système (Ravel et coll. 2007). Ici, le logiciel explore les données suivant un processus complexe afin de déterminer la structure la plus probable, partitionnée en groupes (clusters) les plus différents les uns des autres et où les individus qui les composent sont génétiquement le plus homogène possible. Entre 25 et 21 sous-groupes ont été définis selon la date de l'échantillon (novembre 2000 et janvier 2001, respectivement). Le fait que dans les deux cas les déficits en hétérozygotes calculés dans les sous-groupes ainsi définis se révèlent sensiblement être inférieurs (et non significatifs pour la plupart) au F_{IS} initial (cf. Figure 8.2) indique qu'un effet Wahlund (mélange d'individus issus de groupes ou populations génétiquement différenciés) vient s'ajouter aux autres phénomènes et que les populations de mouches tsétsés sont structurées à une échelle nettement plus fine que celle concernée par les échantillonnages. Les populations de glossines échantillonnées sont donc structurées à une échelle micro-géographique et donc probablement adaptées à des conditions locales sur une petite échelle spatiale. Ces insectes naissent et s'accouplent au même endroit. Leur dispersion due au nourrissage ne reflète donc pas ce phénomène et les mouches capturées dans les pièges proviennent vraisemblablement de zones différentes favorables à la larviposition (les tsétsés sont vivipares) et à l'accouplement, telles que les zones de présence de porcs, des points d'eau ou d'écotones (c'est-à-dire des « lisières » formées entre deux paysages, par exemple à l'interface forêt/champs cultivé) en Côte d'Ivoire (*G. palpalis palpalis*) ou le long de la rivière Koba au Burkina-Faso (*G. palpalis gambiensis*). Ces résultats suggèrent également que les pièges utilisés sont efficaces (capture de glossines sur un large territoire) mais que, même si elles se montrent très philopatrides, les capacités de recolonisation des glossines sont vraisemblablement considérables. Les campagnes d'éradication futures devront donc

tenir compte de ces éléments de connaissance et se faire sur une large échelle pour être efficaces.

FIG. 8.2 | Analyse multivariée individu-centrée

Mises en évidence d'effets Wahlund dans les échantillons de glossines par AFC au Burkina-Faso (a) et par utilisation du logiciel BAPS en Côte d'Ivoire (b). En (a) on distingue quatre groupes à l'intérieur desquels le F_{IS} n'est plus significatif alors qu'il l'est quand tous les individus sont regroupés. En (b) la baisse du F_{IS} entre les données initiales et dans les groupes définis par BAPS est flagrante tant pour l'échantillon de Novembre 2000 que pour celui de Janvier 2001. (a) redessiné d'après Solano et coll. (1999) et (b) d'après Ravel et Coll. (2007).



5.4 La tique du bétail en Nouvelle-Calédonie

La tique du bétail *Rhipicephalus (Boophilus) microplus* est une tique originaire de l'île de Java ; elle est connue pour se nourrir sur le zébu (*Bos indicus*), hôte d'origine qui contrôle immunitairement les charges de cette tique. Les races européennes de bovins (*Bos taurus*, races limousines ou charolaises par exemple) sont quant à elles incapables de résister à cette tique. Ainsi, en zone intertropicale où ces races occidentales ont souvent été introduites, les élevages souffrent partout des ravages causés par cette tique (Frisch 1999), ce qui rend indispensable l'utilisation d'acaricides contre cette peste. De plus, cette tique montre de remarquables aptitudes à développer des résistances à ces biocides (Chevillon et coll. 2007a). Cet arthropode a été accidentellement introduit en Nouvelle-Calédonie en 1942 à partir de l'Australie, après quoi l'île a été mise en quarantaine afin de prévenir une nouvelle introduction et surtout empêcher l'introduction des maladies comme la babésiose et l'anaplasmose, deux autres maladies dont cette tique est aussi vectrice. Afin de mieux comprendre la biologie des populations de cette tique, et notamment ses capacités de dispersion (dissémination de gènes de résistance ou d'éventuels pathogènes introduits), et de tester si la quarantaine a été efficace, des populations de cette tique originaires de Nouvelle-Calédonie ont été analysées à l'aide de marqueurs microsatellites. Ces recherches ont en outre montré un déficit en hétérozygotes au sein de chaque ferme analysée, mais aussi au sein de chaque individu bovin prélevé (plusieurs tiques prélevées par vache) (Koffi et coll. 2006) ; un résultat explicable par un taux de croisements frère-sœur entre spécimens de tique (cf. l'Équation 6) compris entre 0,1 et 0,22 (Chevillon et coll. 2007b). Ce résultat est en apparence contradiction avec l'appariement au hasard des mâles et des femelles de tiques trouvées accouplées lors de ces études et semble en fait mieux expliqué par une forte variance du succès de survie des pontes : certaines pontes donnent beaucoup d'adultes reproducteurs, alors que d'autres ne sont représentées que par très peu d'adultes survivants (Chevillon et coll. 2007b). Par ailleurs, l'hétérozygotie trouvée au sein de chaque ferme était compatible avec l'existence d'un goulot d'étranglement selon la méthode de Cornuet et Luikart (Cornuet et Luikart 1996) datant de 200 à 250 générations environ (Koffi et coll. 2006). Un isolement par la distance très significatif a également pu être détecté qui, par la méthode $D\sigma^2 = 1/4\pi b$ (où b est la pente de la droite $F_{ST}/(1-F_{ST}) = a+b\text{Ln}(D_G)$; cf. § 2.4.b), et en complément d'observations sur le terrain, a permis d'estimer que la dispersion de cette tique n'était que de quelques centaines de mètres par génération et que ces tiques maintiennent des tailles importantes de populations locales (plusieurs centaines à plusieurs milliers par km^2) (Koffi et coll. 2006), et ce malgré des traitements acaricides intensifs (toutes les trois semaines). Nos conclusions sont qu'ici la stratégie de tout traitement par acaricide s'avère sans issue eu égard aux formidables capacités

d'adaptation de cette tique (grandes tailles de populations, taux de mutations probablement élevés) et que très certainement une gestion par jachère avec des hôtes résistants serait plus efficace pour contrôler les populations de cette peste.

Un autre résultat intéressant de ce travail a été qu'en analysant la répartition intra-ferme de la diversité génétique de cette tique, nous avons constaté que les différents individus hôtes (vaches) abritaient chacun un échantillon représentatif des tiques présentes dans chacune des fermes. Or, le cycle « officiel » de *R. microplus* stipule que l'ensemble des stades de cette tique se succèdent, après mue, sur un seul et même individu hôte (cycle monophasique). Les femelles tombent au sol où elles laissent en mourant entre 1000 et 3000 œufs environ (Sutherst et coll. 1999). Chaque hôte se trouve ainsi massivement infecté par des larves issues d'une ou de quelques pontes. Le reste du développement se faisant sur place, les adultes qui se croiseront sont en théorie issus de ce phénomène d'infestation massif.

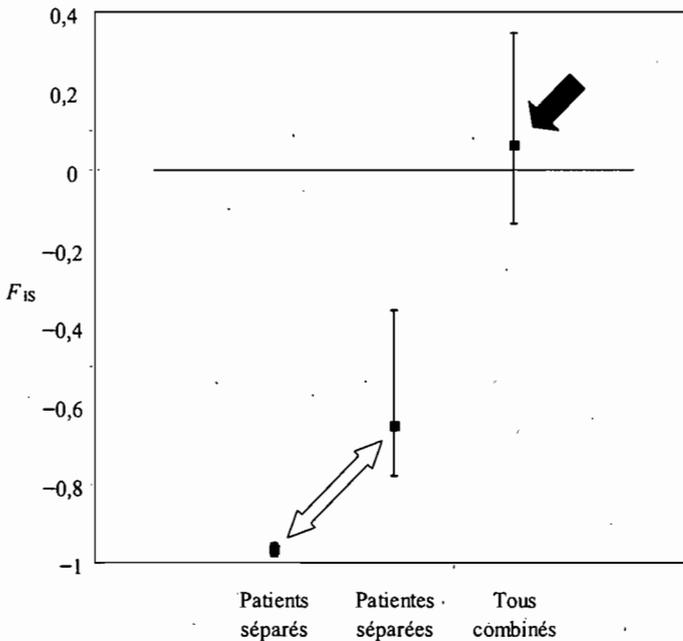
Une démonstration convaincante de l'apport de la génétique des populations à l'épidémiologie de la transmission des maladies à tique concerne l'exemple d'*Anaplasma marginale*. Il s'agit d'une bactérie appartenant aux Rickettsiales, un agent infectieux important du bétail, non contagieux (transmis par des tiques) et que l'on retrouve aussi chez *R. microplus*. Cependant, la vectorisation de cet agent pathogène par la tique a toujours été remise en cause par l'impossibilité de transmission trans-ovarienne de ce pathogène par cette tique monophasique (Uilenberg 1976). Au contraire, nos résultats montrent que *R. microplus* peut être un vecteur très efficace car les individus se redistribuent largement d'un individu hôte à l'autre dans un même pré.

5.5 Le cas spécial des levures opportunistes de patients immunodéprimés

Candida albicans est une levure diploïde. Elle fait partie de la flore commensale du tube digestif et des muqueuses génito-urinaires de la plupart des mammifères dont l'homme. Cette levure n'est pathogène que chez les sujets immunodéprimés (Hull et coll. 2000, Berman et Sudbery 2002). La biologie des populations et le régime de reproduction de cette levure restent des sujets très controversés. Dans une étude récente basée sur 14 loci enzymatiques (Nébavi et coll. 2006), une approche hiérarchique à quatre niveaux d'organisation, l'individu levure (on parle dans ce cas d'isolat), l'individu patient (cinq isolats par patients ont été prélevés pour l'étude), le sexe du patient (19 femmes et 23 hommes) et la population totale des 42 patients immunodéprimés (tous se révélant être HIV+), a été utilisée. Ces analyses, combinées avec des approches théoriques récentes sur la clonalité (De Meeûs et Balloux 2005), ont permis de déterminer un régime de reproduction presque entièrement (voire totale-

FIG. 8.3 | Indice d'hétérozygotie intra-échantillon des isolats de la levure *Candida albicans*

Résultat des analyses de l'indice d'hétérozygotie intra-échantillon (F_{IS}) des isolats de la levure *Candida albicans* prélevés chez des patients immunodéprimés de Côte d'Ivoire (Abidjan). Deux catégories d'échantillons sont présentées. Les échantillons où les patients hommes et femmes sont analysés séparément, et où les isolats de chaque patient constituent un sous-échantillon, sont indiqués par une double flèche vide. L'échantillon où tous les isolats sont combinés en un seul et même échantillon est indiqué par une flèche noire pleine. Dans le premier cas les forts excès d'hétérozygotes observés sont compatibles avec une reproduction clonale absolue ou quasi-absolue dans une population très structurée (forte différenciation entre isolats de patients différents) en petites sous-unités échangeant peu de migrants. On remarque également la claire différence entre isolats hébergés par les patientes et ceux isolés des patients. Dans le deuxième cas, en augmentant l'estimation de l'hétérozygotie, l'effet Wahlund rapproche l'échantillon de ce que l'on attend sous régime sexué panmictique et ne permet pas de voir de différence entre patients ou entre sexes. Les intervalles de confiance sont obtenus par ré-échantillonnage des loci (bootstrap).



ment) clonal, un effet majeur des individus patients dans la subdivision (chaque patient semble porter une population originale de levure), ainsi qu'une hétérogénéité très forte entre patients de sexe différent. Cette différence entre sexe est interprétable soit par le fait que les femmes présentent plus d'individus levures (individus génétiques), soit parce que les levures qu'elles abritent ont davantage connu (ou plutôt moins rarement) des événements de recombinaison. Les conclusions de ce travail ont

été rendues possibles grâce à une stratégie exceptionnelle d'échantillonnage (plusieurs isolats par patient, tailles équilibrées d'échantillons). L'individu patient n'étant pratiquement jamais pris en compte dans les études d'épidémiologie moléculaire, les auteurs ont également démontré que l'ignorance du facteur patient aurait produit une estimation totalement erronée du F_{IS} moyen, proche alors de sa valeur espérée en panmixie (Figure 8.3). Ignorer l'effet patient, qui est majeur ici, aurait également entraîné l'ignorance de l'effet du sexe du patient. Les levures du type *C. albicans* de chaque individu hôte semblent donc se comporter comme des unités clonales isolées les unes des autres. Cela signifie que les phénomènes de contagion et les infections nosocomiales par *C. albicans* sont sans doute très rares et sûrement consécutives à une élimination massive de la flore d'origine par traitements antifongiques. L'aspect clonal strict de la reproduction de ce micro-organisme indique que les souches d'intérêt peuvent facilement être marquées par n'importe quel locus (tous les loci du génome sont liés), alors que l'identification de gènes d'intérêt ne sera pas facile sur l'espèce elle-même (études d'associations très difficiles puisque tout est associé à tout). Enfin, sachant que *C. albicans* est présent chez tous les individus de l'espèce humaine, de mammifères et sans doute d'oiseaux (Hubalek 2004, Cafarchia et coll. 2006, Garcia et coll. 2007), cela donne une idée des formidables capacités qu'à ce champignon de maintenir et diffuser une diversité génétique infinie.

6. Avenir de la génétique des populations en épidémiologie

Nous espérons avoir convaincu le lecteur de l'intérêt d'utiliser les outils moléculaires et de génétique des populations. Ce type d'approche offre, en effet, l'opportunité d'accéder à des informations clés, comme nous avons essayé de l'illustrer rapidement, sur la biologie d'organismes très difficiles à étudier. Armé des connaissances de bases requises et des hypothèses attachées à l'application de chacune de ces méthodes, l'épidémiologiste de terrain devrait être aujourd'hui en mesure d'estimer, à des coûts aujourd'hui raisonnables, la taille des populations des pathogènes et/ou de leurs vecteurs, leurs niveaux de subdivision et leur capacité de dispersion dans l'espace et le temps, ainsi que leur modalité de reproduction. Cependant, nous aimerions insister sur le facteur le plus limitant dans de telles études et qui concerne la stratégie d'échantillonnage, comme nous avons essayé de l'illustrer à l'aide des exemples présentés dans ce chapitre. Les études sont encore trop rares qui tiennent compte de l'ensemble de ces paramètres. Ne pas tenir compte du sexe du patient, de la diversité du vecteur, de l'espèce hôte ou de l'individu hôte lui-même peut conduire à des images complètement erronées ou à des limitations sérieuses dans l'interpré-

tation des résultats et des observations. L'accès grandissant à des séquences génomiques complètes apporte des perspectives fascinantes dans ce domaine. Cela ouvre le champ des possibles dans l'investigation du comportement comparé de marqueurs neutres et de gènes sélectionnés par des facteurs d'intérêt (cf. Chapitre 6). De tels apports seront en effet essentiels dans notre compréhension de l'épidémiologie et de la génétique évolutive des agents infectieux.

➤ À RETENIR

- La génétique des populations permet dans certains cas des estimations précises du fonctionnement des populations de pathogènes et de leurs vecteurs pour des coûts de plus en plus raisonnables, mais requiert tout de même une bonne maîtrise des hypothèses sous-jacentes.
- Une des difficultés réside dans la multiplicité des outils et procédures statistiques, prix inévitable à payer pour atteindre une efficacité parfois très spectaculaire.
- Les précautions à prendre se multiplient, ce qui peut conduire à des analyses d'apparence fastidieuse mais néanmoins payantes.
- La diversité des cas de figures rencontrés rend parfois nécessaire la création d'outils nouveaux et donc aussi dans certains cas de faire appel à des théoriciens chevronnés, qu'il ne faut donc pas hésiter à consulter.

➤ QUESTIONS DE DISCUSSION

- Quelles sont les principales qualités requises pour un marqueur génétique ?
- Quelle est l'hypothèse faite lors de l'estimation d'une fréquence allélique d'une population à partir d'un échantillon d'individus génotypés de cette population ?
- Quelles sont les hypothèses du modèle de Hardy-Weinberg ?
- Pourquoi un fort taux de mutation, ou de sélection, brouille-t-il la perception que l'on peut avoir de la différenciation génétique ?
- Pourquoi a-t-on besoin d'estimateurs ?
- Quel est le principe d'un test par randomisation ?
- Pourquoi est-il souhaitable d'utiliser plusieurs loci dans l'analyse de la structure des populations naturelles ?
- Quels types d'inférence peut-on espérer faire à partir de données de génétique des populations ?

LECTURES POUR ALLER PLUS LOIN

Manly, B.F.J. (1985). *The Statistics of Natural Selection*. Chapman & Hall, London.

Contient énormément d'informations et de procédures diverses fort utiles, des lignes de programmes informatiques sur des domaines de la biologie empirique qui débordent largement du seul cadre de la génétique des populations.

Weir, B.S. (1996). *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*. Sinauer Associates, Sunderland, Massachusetts.

La référence standard. Contient des exercices avec leur solution et une liste de programmes informatiques disponibles gratuitement sur internet.

Hedrick, P.W. (2005). *Genetics of Populations*. Third Edition. Jones and Bartlett Publishers, Boston.

Plus accessible que les précédents. Présenté comme le plus courant et le plus complet des ouvrages d'introduction à la discipline pour les étudiants de maîtrise et de troisième cycle ainsi que pour les chercheurs intéressés par les problèmes liés à la génétique, l'évolution et la biologie de la conservation. Comprend des présentations claires de la théorie et de la génétique des populations expérimentales et de terrain. Le propos est illustré d'exemples variés, classiques et plus récents.

Glossaire

Allèle: état héréditaire dans lequel un locus se présente. Chez les diploïdes, chaque individu présente deux allèles à chaque locus. Ces allèles peuvent être identiques (**homozygote**) ou différents (hétérozygote).

Allèle nul: allèle qui reste invisible à cause de problèmes techniques et qui est par définition **récessif** et caractérisé par une absence de signal à l'état homozygote.

Autofécondation: phénomène de reproduction sexuée possible en théorie chez tous les organismes hermaphrodites et qui consiste à la fécondation d'un ovule d'un individu par un spermatozoïde produit par le même individu.

Auto-stop: *hitchhiking* en anglais. Phénomène sélectif au cours duquel la sélection à un locus entraîne des modifications de la distribution des fréquences génotypiques à un autre locus lié physiquement (proche sur le même chromosome) ou statistiquement quand le mode de reproduction est fermé (clonalité, autofécondation, etc.).

Bottleneck: en français voir goulot d'étranglement.

Chromosome: du grec *chroma*, couleur et *soma*, corps, élément. Élément microscopique constitué de molécules d'ADN. Dans les cellules eucaryotes, ils prennent la forme soit d'un bâtonnet, soit d'un échecaveau, selon qu'ils sont condensés ou non.

Chromosome sexuel: type de chromosomes existant chez certains animaux et déterminant leur sexe. On parle d'**hétérosomes** car un des deux sexes (sexe hétérogamétique) est hétérozygote pour ce chromosome (antonymique d'auto-some). Quand les femelles correspondent au sexe hétérogamétique (oiseau, schistosomes) on appelle ces chromosomes Z et W (les femelles sont ZW et les mâles ZZ) et quand ce sont les mâles (mammifères, drosophiles, tiques) on a X et Y (femelles XX et mâles XY). Il existe aussi des variantes où les mâles sont haploïdes à ce chromosome (appelé alors X0).

Cline: généralement géographique, il correspond à l'augmentation ou à la diminution graduelle des fréquences alléliques à un ou plusieurs loci le long d'un axe géographique et/ou d'un gradient écologique.

Clonalité: reproduction asexuée où la descendance est produite sans subir ni ségrégation ni recombinaison (améiotique) et se retrouve donc génétiquement strictement identique à l'individu parental, à la mutation somatique près.

Codominant: décrit un marqueur génétique pour lequel tous les hétérozygotes sont distinguables des homozygotes (ni dominant, ni récessif).

Consanguinité: condition génétique traduisant le degré de parenté des allèles entre eux, au sein des individus (diploïdes) ou entre individus. Formellement la probabilité d'identité par descendance.

Dérive génétique: décrit le processus par lequel les fréquences alléliques changent d'une génération à l'autre à cause d'un échantillonnage aléatoire des individus (gamètes, zygotes, adultes) qui doivent participer à la reproduction de la génération suivante dans une population

de taille finie. Ce processus est aléatoire et conduit toujours, en l'absence d'autres forces (migration, mutation, sélection, etc.) à la fixation d'un allèle (érosion de la diversité).

Dioïque: synonyme de **gonochorique** (terme un peu désuet aujourd'hui) et signifiant que l'espèce étudiée est séparée en deux sexes (femelles et mâles) (antonymique de **monoïque**).

Diploïde: caractérise un organisme ou une cellule possédant un matériel génétique (chromosomes) en double, à l'exception des chromosomes sexuels quand ces derniers existent.

Dominant: caractérise un marqueur génétique pour lequel un des allèles masque à l'état hétérozygote les autres allèles. Caractérise aussi un tel allèle (antonymique de récessif).

Effectif efficace: noté habituellement N_e , représente une mesure de la vitesse avec laquelle une population de taille N_c (taille de recensement ou *census size* en anglais) perd de la variabilité génétique par dérive génétique aléatoire. N_e peut différer de N_c si N_c varie au cours des générations, si le **sexe-ratio** est déséquilibré, s'il n'y a pas panmixie, etc. Par exemple $N_e=4$ pour un troupeau de 99 vaches et un seul taureau, troupeau qui dérive donc aussi vite que si constitué de deux vaches et deux taureaux.

Gène: une portion d'ADN qui code pour une fonction, c'est-à-dire transcrite en ARN de transfert et ensuite en ARN messager (ou ARNm). L'ARN messager devant lui-même être traduit en molécule active telle une enzyme.

Génotype: donne la composition allélique complète d'un individu à un locus donné ou à une série de loci spécifiques (quand précisé).

Germinal: qui provient de la lignée du même nom, cellules souches des cellules sexuelles (ou gamètes).

Gonochorique: terme un peu désuet aujourd'hui synonyme de dioïque (antonymique d'hermaphrodite).

Goulot d'étranglement: en anglais *bottleneck*. Désigne un processus démographique durant lequel une population subit une chute brutale du nombre d'individus reproducteurs.

Haploïde: caractérise un organisme ou une cellule avec un matériel génétique (chromosomes) présent en un seul exemplaire. Les cellules sexuelles (gamètes) sont typiquement haploïdes.

Hermaphrodite: se dit d'une espèce à reproduction sexuée où chaque individu peut assurer les deux fonctions femelle et male (antonymique de gonochorique).

Hétérosome: se dit d'un chromosome sexuel (antonymique d'autosome). ▫

Hétérothallique: se dit de champignons dotés d'un *mating type* du type \pm qui détermine la compatibilité sexuelle entre individus. Les $+$ ne pouvant donner des zygotes qu'avec des $-$ et vice-versa. Il y a donc autoincompatibilité (antonymique d'**homothallique**).

Hétérozygote: état d'un locus chez un individu diploïde présentant deux allèles différents (antonymique d'homozygote).

Hitchhiking: voir la définition française d'auto-stop.

Homoplasie: phénomène décrivant l'identité entre deux allèles ne résultant pas d'une parenté commune récente, qui sont alors qualifiés d'identiques par état.

Homothallique: se dit de champignons autocompatibles où chaque individu peut produire des thalles qui fusionnent pour donner des zygotes assimilables à de

l'autofécondation (antonymique d'hétérothallique).

Homozygote: état d'un locus chez un individu diploïde présentant deux fois le même allèle (antonymique d'hétérozygote).

IAM (*Infinite Allele Model*): modèle de mutation où chaque mutation génère un nouvel allèle qui n'existait pas auparavant dans la population et sera définitivement perdu s'il disparaît. Ne permet aucune **homoplasie**.

Îles (modèle en): modèle théorique de population structurée en n dèmes de tailles identiques N composés à chaque génération non-chevauchante de $(1-m)N$ individus autochtones et de mN individus migrants provenant aléatoirement de n'importe quel des n dèmes.

Intron: partie d'un gène qui ne sera pas traduite en protéine car éliminée lors du passage de l'ARN de transfert vers l'ARN messager (phénomène d'épissage) (antonymique d'exon).

KAM (*K Allele Model*): modèle de mutation en nombre fini (K) d'allèles. Modèle de mutation où chaque mutation change un allèle dans un autre allèle parmi les K existants, avec la même probabilité, y compris en lui-même. Plus K est petit, plus fréquente est l'homoplasie.

Locus: décrit une portion de l'ADN située dans une position spécifique du génome. Un locus ne correspond pas nécessairement à un gène.

Méiose: processus de production des cellules de la reproduction sexuée ou gamètes. C'est au cours de ce processus qu'ont lieu la ségrégation des allèles à chaque locus et la recombinaison entre loci, pour aboutir à la formation de cellules **haploïdes** (gamètes).

Métapopulation : modèle de population complexe où les individus se répartissent en plusieurs sous-populations qui peuvent ou non subir des extinctions et recolonisations, croissance ou déclin.

Monoïque : synonyme d'hermaphrodite (antonymique de dioïque).

Mutation : erreur héréditaire intervenant lors de la duplication de l'ADN ou durant la méiose.

Neighbourhood model : modèle en voisinage. Un modèle théorique de population structurée où la migration de chaque individu est limitée par la distance, de telle sorte que l'apparementement entre individus devient une fonction décroissante de la distance qui les sépare, même en l'absence de toute barrière ou délimitation visible.

Neutre : définit un locus ou un caractère dont le polymorphisme n'est soumis à aucune pression sélective d'aucune sorte (antonymique de sélectionné).

Panmixie : décrit un mode de reproduction sexuée où les zygotes sont formés par rencontre aléatoire de tous les gamètes de la population.

Parasite : un organisme vivant dans ou sur un individu d'une autre espèce vivante hôte de taille plus grande et qui représente une part variable des besoins écologiques du parasites (habitat, nourriture, lieu de reproduction) pour une durée plus ou moins longue mais de façon obligatoire. Le pou est typiquement un parasite, le lion non et le moustique est marginal de ce point de vue (seule la femelle se nourrit sur des animaux, et ce sur plusieurs individus hôtes différents).

Parthénogenèse : du grec *parthenos* (vierge) et *genesis* (genèse), quand une mère produit des filles à partir d'ovules non fécondés.

Pas japonais (Modèle en) : *stepping-stone model* en anglais. Modèle théorique de population subdivisée où les migrants ne s'échangent qu'entre sous-populations adjacentes.

PCR : *Polymerase Chain Reaction* (l'acronyme français ACP pour amplification en chaîne par polymérisation est très rarement utilisé). Méthode de biologie moléculaire d'amplification génique *in vitro*, qui permet de copier en grand nombre (avec un facteur de multiplication de l'ordre du milliard) une séquence d'ADN ou d'ARN connue, à partir d'une faible quantité (de l'ordre de quelques picogrammes) d'acide nucléique, à l'aide d'amorces spécifiques constituées d'oligonucléotides de synthèse de 20 à 25 nucléotides.

Polymorphe : condition qui décrit qu'un locus est variable d'un individu à l'autre, c'est-à-dire qu'il présente plus d'un allèle dans l'échantillon d'individus génotypés.

Population : groupe d'individus partageant les mêmes paramètres démographiques, en particulier la régulation de la population, et partageant une ascendance commune plus probable avec les individus de la même unité qu'avec des individus d'autres populations définies comme telles, exception faite des migrants, bien entendu.

Récessif : caractérise un allèle qui est masqué dans le cas d'hétérozygotie avec un autre allèle (antonymique de dominant).

Recombinaison : processus durant lequel les allèles de loci différents, auparavant associés, se retrouvent dissociés et réassociés à d'autres allèles. C'est ce qui se passe durant la méiose entre loci de chromosomes différents ou du même chromosome après crossing-over.

Sélection: processus durant lequel la survie et/ou le succès reproducteur d'un individu dépend de son phénotype ou de son génotype d'une manière plus ou moins directe.

Sélectionné: s'applique pour un locus ou un caractère soumis à sélection (antonymique de **neutre**).

Sexe-ratio: rapport du nombre de mâles sur le nombre de femelles dans une population. Égal à un quand il est équilibré.

SMM (*Stepwise Mutation Model*): mécanisme de mutation au cours duquel chaque mutation augmente ou diminue, avec une égale probabilité, la taille de l'allèle d'une unité (step) prédéfinie. Ce mode de mutation génère beaucoup d'homoplasie et aboutit au fait qu'une ressemblance de taille peut se traduire par une ascendance commune de deux allèles.

Somatique: ce qui vient du soma, c'est-à-dire n'impliquant pas les cellules de la lignée dite **germinale** (antonymique de germinale).

Stepping-stone model: voir l'expression française de pas japonais.

TPM (*Two Phases Model*): modèle de mutation combinant le KAM et le SMM avec une proportion variable de mutations générées par l'un ou l'autre des mécanismes correspondants.

Transition: mutation ponctuelle consistant au remplacement d'une purine par une autre purine ($A \leftrightarrow G$) ou d'une

pyrimidine par une autre pyrimidine ($C \leftrightarrow T$) (antonymique de **transversion**).

Transversion: mutation ponctuelle consistant au remplacement d'une purine par une pyrimidine ou d'une pyrimidine par une purine ($A \leftrightarrow T$, $A \leftrightarrow C$, $G \leftrightarrow C$, $G \leftrightarrow T$) (antonymique de **transition**).

Vecteur: souvent lui-même parasite, le vecteur est un être vivant faisant partie intégrante du cycle d'un parasite qu'il transporte vers l'hôte suivant. Il s'agit le plus souvent d'un arthropode piqueur tel qu'une tique qui transmet les agents de la piroplasmose, un moustique anophèle qui transmet la malaria ou un puceron qui transmet un virus pathogène pour la plante hôte. Certains hôtes intermédiaires sont aussi considérés comme vecteurs par certains, tels que les mollusques hôtes intermédiaires de distomatoses (douve du foie, bilharziose), par extension quelque peu abusive.

Végétative: mode de reproduction purement asexuée où un individu donne naissance à plusieurs autres individus par simple division (mitose ou scissiparité).

Voisinage (Modèle en): voir **Neighborhood model**.

Wahlund (effet): diminution de l'hétérozygotie observée que produit le mélange dans un même échantillon d'individus hétérogènes génétiquement.

Zygote: résultat de la fusion de deux gamètes. Le terme œuf est aussi parfois usité.

Références

- Balloux, F., Lehmann, L. et De Meeûs, T. (2003). The population genetics of clonal and partially clonal diploids. *Genetics* 164: 1635-1644.
- Baudouin, L., Piry, S. et Cornuet, J.M. (2004). Analytical bayesian approach for assigning individuals to populations. *Journal of Heredity* 95: 217-224.
- Berman, J. et Sudbery, P.E. (2002). *Candida Albicans*: a molecular revolution built on lessons from budding yeast. *Nature Reviews Genetics* 3: 918-930.
- Cafarchia, C., Camarda, A., Romito, D., Campolo, M., Quaglia, N.C., Tullio, D. et Otranto, D. (2006). Occurrence of yeasts in cloacae of migratory birds. *Mycopathologia* 161: 229-234.
- Caterino, M.S., Cho, S. et Sperling, F.A.H. (2000). The current state of insect molecular systematics: A thriving Tower of Babel. *Annual Review of Entomology* 45: 1-54.
- Chevillon, C., Ducomet, S., De Meeûs, T., Koffi, B.B., Gaia, H., Delathière, J.M. et Barré, N. (2007a). Accumulation of acaricide resistance mechanisms in *Rhipicephalus (Boophilus) microplus* (Acari: Ixodidae) populations from New Caledonia Island. *Veterinary Parasitology* 147: 276-288.
- Chevillon, C., Koffi, B.B., Barré, N., Durand, P., Amathau, C. et De Meeûs, T. (2007b). Direct and indirect inferences on parasite mating and gene transmission patterns. Pangamy in the cattle tick *Rhipicephalus (Boophilus) microplus*. *Infection, Genetics and Evolution* 7: 298-304.
- Corander, J., Waldmann, P. et Sillanpaa, M.J. (2003). Bayesian analysis of genetic differentiation between populations. *Genetics* 163: 367-374.
- Corander, J., Waldmann, P., Marttinen, P. et Sillanpaa, M.J. (2004). BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics* 20: 2363-2369.
- Corander, J., Marttinen, P. et Mantyniemi, S. (2006). A Bayesian method for identification of stock mixtures from molecular marker data. *Fishery Bulletin* 104: 550-558.
- Cornuet, J.M. et Luikart, G. (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144: 2001-2014.
- Criscione, C.D. et Blouin, M.S. (2005). Effective sizes of macroparasite populations: a conceptual model. *Trends in Parasitology* 21: 212-217.
- Criscione, C.D., Poulin, R. et Blouin, M.S. (2005). Molecular ecology of parasites: elucidating ecological and microevolutionary processes. *Molecular Ecology* 14: 2247-2257.
- De Meeûs, T., Beati, L., Delaye, C., Aeschlimann, A. et Renaud, F. (2002a). Sex-biased genetic structure in the vector of Lyme disease, *Ixodes ricinus*. *Evolution* 56: 1802-1807.
- De Meeûs, T. et Renaud, F. (2002). Parasites within the new phylogeny of eukaryotes. *Trends in Parasitology* 18: 247-251.
- De Meeûs, T., Renaud, F., Mouveroux, E., Reynes, J., Galeazzi, G., Mallié, M. et Bastide, J.M. (2002b). Genetic structure of *Candida glabrata* populations in AIDS and non-AIDS patients. *Journal of Clinical Microbiology* 40: 2199-2206.
- De Meeûs, T., Durand, P. et Renaud, F. (2003). Species concepts: what for? *Trends in Parasitology* 19: 425-427.

- De Meeûs, T., Humair, P.F., Grunau, C., Delaye, C. et Renaud, F. (2004a). Non-Mendelian transmission of alleles at microsatellite loci: an example in *Ixodes ricinus*, the vector of Lyme disease. *International Journal for Parasitology* 34: 943-950.
- De Meeûs, T., Lorimier, Y. et Renaud, F. (2004b). Lyme borreliosis agents and the genetics and sex of their vector, *Ixodes ricinus*. *Microbes and Infection* 6: 299-304.
- De Meeûs, T. et Balloux, F. (2005). F-statistics of clonal diploids structured in numerous demes. *Molecular Ecology* 14: 2695-2702.
- De Meeûs, T., Lehmann, L. et Balloux, F. (2006). Molecular epidemiology of clonal diploids: A quick overview and a short DIY (do it yourself) notice. *Infection, Genetics and Evolution* 6: 163-170.
- De Meeûs, T., McCoy, K.D., Prugnolle, F., Chevillon, C., Durand, P., Hurtrez-Boussès, S. et Renaud, F. (2007a). Population genetics and molecular epidemiology or how to «débuisquer la bête». *Infection, Genetics and Evolution* 7: 308-332.
- De Meeûs, T., Prugnolle, F. et Agnew, P. (2007b). Asexual reproduction: Genetics and evolutionary aspects. *Cellular and Molecular Life Sciences* 64: 1355-1372.
- Duneau, D., Bouludier, T., Gómez-Díaz, E., Petersen, A., Tveraa, T., Barrett, R.T. et McCoy, K.D. (2008). Prevalence and diversity of Lyme borreliosis bacteria in marine birds. *Infection, Genetics and Evolution* 8: 352-359.
- Frisch, J.E. (1999). Towards a permanent solution for controlling cattle ticks. *International Journal for Parasitology* 29: 57-71.
- Garcia, M.E., Lanzarot, P., Rodas, V.L., Costas, E. et Blanco, J.L. (2007). Fungal flora in the trachea of birds from a wildlife rehabilitation centre in Spain. *Veterinari Medicina* 52: 464-470.
- Goldstein, D.B. et Schlötterer, C. (1999). *Microsatellites, Evolution and Applications*. Oxford University Press, Oxford.
- Goudet, J., Raymond, M., De Meeûs, T. et Rousset, F. (1996). Testing differentiation in diploid populations. *Genetics* 144: 1933-1940.
- Guinand, B. (1996). Use of a multivariate model using allele frequency distributions to analyse patterns of genetic differentiation among populations. *Biological Journal of the Linnean Society* London 58: 173-195.
- Hardy, G.H. (1908). Mendelian proportions in a mixed population. *Science* 28: 49-50.
- Hartl, D.L. et Clark, A.G. (1989). *Principles of Population Genetics, Second Edition*. Sinauer Associates, Inc., Sunderland, Massachusetts.
- Hedrick, P.W. (1999). Perspective: Highly variable loci and their interpretation in evolution and conservation. *Evolution* 53: 313-318.
- Hedrick, P.W. (2005). A standardized genetic differentiation measure. *Evolution* 59: 1633-1638.
- Hubalek, Z. (2004). An annotated checklist of pathogenic microorganisms associated with migratory birds. *Journal of Wildlife Diseases* 40: 639-659.
- Hull, C.M., Raisner, R.M. et Johnson, A.D. (2000). Evidence for mating of the "asexual" yeast *Candida albicans* in a mammalian host. *Science* 289: 307-310.
- Hurtrez-Boussès, S., Meunier, C., Durand, P. et Renaud, F. (2001). Dynamics of host-parasite interactions: the example of population biology of the liver fluke (*Fasciola hepatica*). *Microbes and Infection* 3: 841-849.
- Hurtrez-Boussès, S., Durand, P., Jabbour-Zahab, R., Guégan, J.F., Meunier, C., Bargues, M.D., Mas-Coma, S. et Renaud, F. (2004). Isolation and characterization of microsatellite markers in the liver fluke

- (*Fasciola hepatica*). *Molecular Ecology Notes* 4: 689-690.
- Kimura, M. et Weiss, G.H. (1964). The Stepping Stone Model of Population Structure and the Decrease of Genetic Correlation with Distance. *Genetics* 49: 561-576.
- Koffi, B.B., De Meeûs, T., Barré, N., Durand, P., Arnathau, C. et Chevillon, C. (2006). Founder effects, inbreeding and effective sizes in the Southern cattle tick: the effect of transmission dynamics and implications for pest management. *Molecular Ecology* 15: 4603-4611.
- Leblois, R., Rousset, F. et Estoup, A. (2004). Influence of spatial and temporal heterogeneities on the estimation of demographic parameters in a continuous population using individual microsatellite data. *Genetics* 166: 1081-1092.
- Lehmann, T., Hawley, W.A., Kamau, L., Fontenille, D., Simard, F. et Collins, F.H. (1996). Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: comparison of microsatellite and allozyme loci. *Heredity* 77: 192-200.
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research* 27: 209-220.
- Mas-Coma, S., Anglés, R., Esteban, J.G., Bargues, M.D., Buchon, P., Franken, M. et Strauss, W. (1999). The Northern Bolivian Altiplano: a region highly endemic for human fascioliasis. *Tropical Medicine and International Health* 4: 454-467.
- McCoy, K.D., Boulinier, T., Tirard, C. et Michalakis, Y. (2001). Host specificity of a generalist parasite: genetic evidence of sympatric host races in the seabird tick *Ixodes uriae*. *Journal of Evolutionary Biology* 14: 395-405.
- McCoy, K.D., Boulinier, T., Tirard, C. et Michalakis, Y. (2003). Host-dependent genetic structure of parasite populations: differential dispersal of seabird tick host races. *Evolution* 57: 288-296.
- McCoy, K.D., Chapuis, E., Tirard, C., Boulinier, T., Michalakis, Y., Le Bohec, C., Le Maho, Y. et Gauthier-Clerc, M. (2005). Recurrent evolution of host-specialized races in a globally distributed parasite. *Proceedings of the Royal Society of London, Ser. B* 272: 2389-2395.
- Metropolis, N. (1987). The beginning of the Monte Carlo method. *Los Alamos Science* 15: 125-130.
- Meunier, C., Tirard, C., Hurtrez-Boussès, S., Durand, P., Bargues, M.D., Mas-Coma, S., Pointier, J.P., Jourdan, J. et Renaud, F. (2001). Lack of molluscan host diversity and the transmission of an emerging parasitic disease in Bolivia. *Molecular Ecology* 10: 1333-1340.
- Meunier, C., Hurtrez-Boussès, S., Jabbour-Zahab, R., Durand, P., Rondelaud, D. et Renaud, F. (2004). Field and experimental evidence of preferential selfing in the freshwater mollusc *Lymnaea truncatula* (Gastropoda, Pulmonata). *Heredity* 92: 316-322.
- Nadler, S.A. (1995). Microevolution and the Genetic-Structure of Parasite Populations. *Journal of Parasitology* 81: 395-403.
- Nagylaki, T. (1998). Fixation indices in subdivided populations. *Genetics* 148: 1325-1332.
- Nébavi, F., Ayala, F.J., Renaud, F., Bertout, S., Eholié, S., Moussa, K., Mallié, M. et De Meeûs, T. (2006). Clonal population structure and genetic diversity of *Candida albicans* in AIDS patients from Abidjan (Cote d'Ivoire) (vol 103, pg 3663, 2006). *Proceedings of the National Academy of Sciences of the United States of America* 103: 5632-5632.
- Nei, M. et Chesser, R.K. (1983). Estimation of fixation indices and gene diversities. *Annals of Human Genetics* 47: 253-259.

- Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press, New York
- Paetkau, D., Calvert, W., Stirling, I. et Strobeck, C. (1995). Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* 4: 347-354.
- Piry, S., Alapetite, A., Comuet, J.M., Paetkau, D., Baudouin, L. et Estoup, A. (2004). GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity* 95: 536-539.
- Ravel, S., De Meeûs, T., Dujardin, J.P., Zeze, D.G., Gooding, R.H., Dufour, I., Sane, B., Cuny, G. et Solano, P. (2007). The tsetse fly *Glossina palpalis palpalis* is composed of several genetically differentiated small populations in the sleeping sickness focus of Bonon, Côte d'Ivoire. *Infection, Genetics and Evolution* 7: 116-125.
- Razakandrainibe, F.G., Durand, P., Koella, J.C., De Meeûs, T., Rousset, F., Ayala, F.J. et Renaud, F. (2005). "Clonal" population structure of the malaria agent *Plasmodium falciparum* in high-infection regions. *Proceedings of the National Academy of Sciences of the United States of America* 102: 17388-17393.
- Roderick, G.K. (1996). Geographic structure insect populations: Gene flow, phylogeography, and their uses. *Annual Review of Entomology* 41: 325-352.
- Rousset, F. (1997). Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics* 145: 1219-1228.
- Rousset, F. (1996). Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics* 142: 1357-1362.
- Rousset, F. (2000). Genetic differentiation between individuals. *Journal of Evolutionary Biology* 13: 58-62.
- Rousset, F. (2004). *Genetic Structure and Selection in Subdivided Populations*. Princeton University Press, Princeton.
- Shuker, D.M., Reece, S.E., Whitehom, P.R. et West, S.A. (2004). Sib-mating does not lead to facultative sex ratio adjustment in the parasitoid wasp, *Nasonia vitripennis*. *Evolutionary Ecology Research* 6: 473-480.
- Slatkin, M. (1985). Gene flow in natural populations. *Annual Review of Ecology and Systematics* 16: 393-430.
- Solano, P., De La Rocque, S., De Meeûs, T., Cuny, G., Duvallet, G. et Cuisance, D. (2000). Microsatellite DNA markers reveal genetic differentiation among populations of *Glossina palpalis gambiensis* collected in the agro-pastoral zone of Sideradougou, Burkina Faso. *Insect Molecular Biology* 9: 433-439.
- Sunnucks, P. (2000). Efficient genetic markers for population biology. *Trends in Ecology & Evolution* 15: 199-203.
- Sutherst, R.W., Boume, A.S. et Sutherland, I.D. (1999). Production and survival of eggs of the cattle tick *Boophilus microplus* (Canestrini) (Acarina: Ixodidae) in the wet and dry tropics of north Queensland. *Australian Journal of Entomology* 38: 340-347.
- Thomas, F., Renaud, F., Derothe, J.M., Lambert, A., Demeëus, T. et Cezilly, F. (1995). Assortative Pairing in *Gammarus insensibilis* (Amphipoda) Infected by a Trematode Parasite. *Oecologia* 104: 259-264.
- Tibayrenc, M. (1999). Toward an integrated genetic epidemiology of parasitic protozoa and other pathogens. *Annual Review of Genetics* 33: 449-477.
- Uilenberg, G. (1976). Tick-borne livestock diseases and their vectors. 2. Epidemiology of tick-borne diseases. *World Animal Review* 17: 8-15.
- Wahlund, S. (1928). Zusammensetzung von populationen und korrelationserscheinungen von standpunkt der vererbungslehre aus betrachtet. *Hereditas* 11: 65-108.

Waples, R.S. (1989). A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* 121: 379-391.

Wedekind, C. et Penn, D. (2000). MHC genes, body odours, and odour preferences. *Nephrology Dialysis Transplantation* 15: 1269-1271.

Weinberg, W. (1908). Über den Nachweis der Verebung beim Menschen. *Jahresheft des Vereins für Vaterländische Naturkunde in Württemberg* 64: 368-382.

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

Wright, S. (1931). Evolution in mendelian populations. *Genetics* 16: 97-159.

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics* 15: 323-354.

Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to system of mating. *Evolution* 19: 395-420.

De Meeûs Thierry, McCoy Karen.

La génétique des populations comme outil en épidémiologie.

In : Guégan Jean-François (ed.), Choisy Marc (ed.). Introduction à l'épidémiologie intégrative des maladies infectieuses et parasitaires. Bruxelles : De Boeck, 2009, p. 277-310.

(LMD.Licence Maîtrise Doctorat.Cours Biologie). ISBN 978-2-8041-5948-1