

# Biologie

## Chapitre 9

# Évolution adaptative des pathogènes, identification des mécanismes et conséquences épidémiologiques

F. PRUGNOLLE<sup>(a)</sup> et C. CHEVILLON<sup>(b)</sup>

1. Quelques généralités sur l'adaptation et la sélection 312
2. Évolution à court terme de la sélection positive 317
3. Évolution à long terme et empreintes de la sélection dans les séquences codantes 328
4. Dynamique adaptative et épidémiologie 335
5. Conclusion 337

<sup>a</sup> UMR 2724 Génétique et Évolution des Maladies Infectieuses, IRD-CNRS-Université de Montpellier I, Montpellier, France

<sup>b</sup> Centre interdisciplinaire sur la Mondialisation et les Maladies Infectieuses, École des Hautes Études en Santé Publique, Montpellier, France.

Tout comme pour celle de leurs hôtes et éventuels vecteurs, la diversité génétique des **pathogènes** subit les actions conjointes de la mutation, de la migration, de la dérive aléatoire, des caractéristiques du **système de reproduction** et, pour une fraction de leur génome, de l'action éventuelle de **pressions de sélection** en un moment et un lieu donnés. Le chapitre 8 a présenté une revue des actions conjointes de la mutation, de la dérive, de la migration et du système de reproduction, ainsi qu'une revue des outils dont le chercheur dispose pour retracer ces actions à partir de données de **polymorphisme** dit « neutre », pour « non-sélectionné ». Le présent chapitre a lui pour objet de décrire les formes généralement prises par la **sélection** sur les agents pathogènes, d'en dégager les conséquences sur la particularité des polymorphismes sélectionnés relativement au polymorphisme « neutre » et de présenter les outils aptes à détecter l'action d'une **pression de sélection** à un **locus** déterminé. Nous insisterons surtout sur le cas de la **sélection positive** du fait qu'il s'agit de la forme de sélection la plus à même de signer le développement d'**adaptations** de pathogènes à de nouveaux habitats (élargissement de la gamme d'espèces-hôtes, acquisition de résistances aux médicaments utilisés en lutte, etc.).

## 1. Quelques généralités sur l'adaptation et la sélection

### 1.1 Pourquoi s'intéresser à l'évolution adaptative des parasites ?

Les pathogènes, à l'instar des autres organismes vivants, sont en permanence soumis à des pressions de sélection de la part de leur environnement, c'est-à-dire à une variabilité d'habitats telle que certains réduisent fortement la survie et/ou la reproduction de la plupart des **génotypes**. Ces pressions de sélection peuvent être multiples, d'intensités différentes, varier dans l'espace et dans le temps, telles que celles définies par les défenses immunitaires de l'hôte, la lutte médicamenteuse ou vaccinale, les changements d'espèces-hôtes, voire les variations de l'environnement cellulaire de l'hôte pour un **pathogène** intracellulaire. Comment les **agents infectieux** s'adaptent-ils à ces variations environnementales et comment parviennent-ils, à terme, à coloniser des habitats peu propices à leur survie et à leur développement ?

Appréhender une évolution adaptative c'est comprendre, par exemple, l'impact à long terme des luttes thérapeutiques menées contre les virus, bactéries pathogène-

nes et/ou **parasites** eucaryotes. Les drogues utilisées pour rompre la transmission et/ou tuer les agents infectieux imposent, en effet, de fortes pressions de sélection sur ces derniers. En retour, les pathogènes sont susceptibles d'évoluer en contournant l'effet de ces produits, c'est-à-dire en développant des résistances aux produits utilisés (**chimiorésistances**). Comment se réalisent de telles évolutions, rendant les moyens thérapeutiques inefficaces et obsolètes? Est-il facile pour les pathogènes de contourner les luttes thérapeutiques? Quels gènes sont impliqués dans l'évolution de ces résistances? Le développement de résistances à certains antibiotiques chez des bactéries responsables d'infections nosocomiales telles que le staphylocoque doré (*Staphylococcus aureus*) (Bancroft 2007) ou à des drogues antipaludiques chez l'agent le plus mortel des paludismes humains, *Plasmodium falciparum* (Trape 2001), sont autant d'exemples rappelant l'importance et l'actualité de ces questions. Parallèlement, l'émergence dans les populations humaines ou domestiques de pathogènes n'ayant jusqu'alors principalement circulé que dans des écosystèmes sauvages – tels les agents du SRAS (syndrome respiratoire aigü sévère) ou de la grippe aviaire – interroge aussi sur la possibilité que les pathogènes ont à coloniser, à se développer et donc à s'adapter à de nouvelles espèces-hôtes.

## 1.2 Identifier les bases moléculaires des adaptations

Imaginons que, quelques temps après la mise sur le marché d'une nouvelle molécule capable de tuer un pathogène, plusieurs médecins diagnostiquent des échecs de traitements, suspectant alors le développement d'une nouvelle adaptation: la résistance du pathogène à l'action de ce nouveau médicament. Être capable d'identifier les bases moléculaires (gènes et mutations) de cette adaptation et de déterminer si elles sont ou non les mêmes que pour les résistances plus anciennes sont des informations précieuses pour optimiser le choix de traitements à effectuer.

Plusieurs approches sont utilisées pour identifier les bases moléculaires (gènes et mutations) des adaptations. Les études d'associations utilisent le diagnostic entre **phénotypes** « adapté » et « non-adapté », c'est-à-dire, dans notre exemple, la différence entre la résistance et la sensibilité à l'action de la dernière molécule mise sur le marché. Le principe consiste à **génotyper** des individus des deux phénotypes à différents locus et de chercher à identifier un ou plusieurs polymorphismes qui soient corrélés à la différence de phénotypes (de Bakker et coll. 2005). Il existe deux limites à cette approche. D'une part, elle suppose une discrimination facile entre les phénotypes « adapté » et « non-adapté ». Si cet objectif ne pose pas trop de problème dans notre exemple de résistance/sensibilité à une molécule connue,

il peut être plus difficile à atteindre dans d'autres cas d'adaptation. On ne sait pas toujours, en effet, quel phénotype a évolué pour permettre l'adaptation à un nouveau milieu. D'autre part, ces études d'associations sont coûteuses : elles demandent de géotyper un grand nombre d'individus de chaque phénotype pour s'assurer que les associations détectées sont fiables (c'est-à-dire qu'elles ne sont pas dues à des hasards d'échantillonnage), et d'effectuer ce géotypage à un grand nombre de locus pour obtenir des résultats précis.

Les approches dites par « gènes candidats » se basent sur des principes similaires. Là encore, les individus sont rangés selon la nature de leur phénotype et on recherche les polymorphismes qui distinguent les phénotypes l'un de l'autre. En revanche, on se limite ici à un ou quelques gènes dont l'implication dans l'adaptation étudiée est déjà suspectée. Encore faut-il que le chercheur connaisse déjà bien le pathogène et l'adaptation étudiée pour émettre une hypothèse solide quant à l'identité de ces « gènes candidats » !

En cas de difficulté, voire d'impossibilité de ranger les individus en « phénotype adapté » et « phénotype non-adapté » de manière fiable, et/ou d'absence d'indice sur l'identité des gènes impliqués dans l'adaptation, le chercheur n'est pas totalement démuné. Une approche alternative consiste à s'intéresser à l'ensemble du génome et à rechercher les régions présentant des structures de polymorphisme différentes de celles observées pour le reste du génome (Nielsen et coll. 2005, Sabeti et coll. 2007). C'est cette alternative que nous allons détailler dans les sections suivantes, notamment au travers de l'exemple des chimiorésistances développées chez *P. falciparum* (cf. § 2.6). Nous retiendrons ici que si cette approche nécessite l'acquisition d'une quantité considérable de données de polymorphisme, elle présente plusieurs avantages sur les deux précédentes. L'avantage le plus important est qu'elle est applicable à toutes les espèces, alors que les approches précédentes restent surtout limitées à celles que l'on peut croiser et élever au laboratoire. De plus, elle ne nécessite pas de connaître un trait phénotypique de l'adaptation qui soit mesurable, ni la fonction ou l'identité des gènes potentiellement impliqués. Par conséquent, elle est capable de détecter des phénotypes et des géotypes d'adaptation qui étaient insoupçonnables à travers l'état des connaissances de la biologie ou de la génétique de l'espèce pathogène étudiée.

### **1.3 Différentes formes de sélection**

De manière générale, considérons le cas où le pathogène étudié est diploïde (c'est-à-dire porteur de deux lots de chromosomes analogues) et que la sélection s'applique sur un locus  $L$  où coexistent deux allèles  $A$  et  $a$  dans la population étudiée. Les individus de la population étudiée sont alors de géotypes  $AA$ ,  $Aa$  ou  $aa$ . Dire

qu'une sélection agit sur le polymorphisme observé au locus L revient à dire que le nombre moyen de descendants par génotype (ou **valeur sélective**  $\omega_{AA}$ ,  $\omega_{Aa}$ ,  $\omega_{aa}$ ) n'est pas identique d'un génotype à l'autre. La distinction première entre les différentes formes de sélection se définit par rapport aux différences entre ces valeurs sélectives. L'encadré 1 détaille ces définitions.

**Encadré1** Les différentes formes de sélection

Pour distinguer les principales formes de sélection, nous considérerons le cas d'une population d'individus diploïdes et d'un polymorphisme au locus L défini par la coexistence des seuls allèles A et a. Les individus de la population étudiée sont alors de génotype AA, Aa ou aa.

**1. Sélections directionnelles : sélection purifiante et sélection positive**

La **sélection directionnelle** correspond à l'une ou l'autre des situations dans lesquelles un allèle est soit favorisé, soit défavorisé par rapport à tous les autres allèles dans l'environnement étudié. Dans le cas d'un polymorphisme à deux allèles, elle est donc définie par la relation  $\omega_{AA} \geq \omega_{Aa} > \omega_{aa}$  (l'allèle A est toujours favorisé relativement à l'allèle a) ou par la relation  $\omega_{AA} < \omega_{Aa} \leq \omega_{aa}$  (l'allèle A est toujours contre-sélectionné relativement à l'allèle a). Par définition, une telle sélection va toujours avoir tendance à éliminer le polymorphisme au sein de la population où elle s'applique.

Lorsque la sélection élimine en moyenne tous les allèles apparaissant par mutation (ou immigration des populations voisines) pour ne conserver qu'un seul allèle « ancestral », on la qualifie de **sélection négative** ou **purifiante** (cf. Figure E1). On parle aussi de contraintes sélectives qui sont à même d'indiquer, dans le cas de gènes, des hypothèses fonctionnelles (par exemple, site actif d'enzymes, conformation nécessaire au bon fonctionnement de la protéine, etc.).

Lorsqu'au contraire, la sélection directionnelle favorise un nouveau mutant  $\alpha$  relativement au polymorphisme préexistant (allèles A et a), on parle de sélection positive (cf. Figure E1). La sélection positive est la forme de sélection qui suscite le plus fort intérêt en biologie évolutive. Ceci parce qu'elle est associée à l'évolution de nouvelles formes, de nouvelles fonctions et de nouveaux traits chez les organismes vivants, c'est-à-dire à l'acquisition d'adaptations nouvelles à des contraintes environnementales fortes (par exemple, résistance aux antibiotiques, acquisition de la capacité à infester un nouvel hôte, etc.).

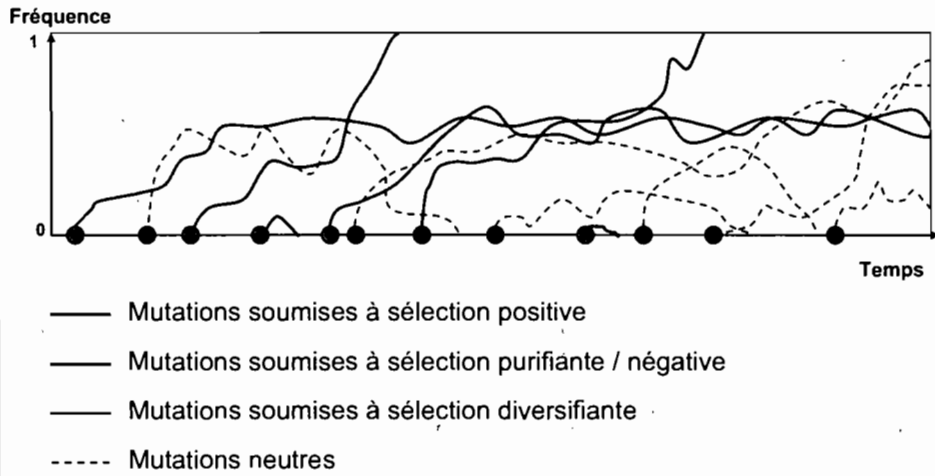
**2. Sélection diversifiante : superdominance et sélection fréquence-dépendante**

On appelle **sélection diversifiante**, tout mécanisme conduisant à conserver plus de polymorphisme et plus longtemps qu'attendu sous l'hypothèse de neutralité. La **superdominance** en est un premier mécanisme. Elle définit le cas où les individus associés à la plus forte valeur sélective sont les hétérozygotes dans un environnement ou en moyenne sur une gamme d'environnements définissant des sélections différentes (on parle alors de superdominance marginale) (cf. Figure E1). Dans l'exemple de polymorphisme diallélique, elle se caractérise par l'ensemble des deux relations  $\omega_{AA} < \omega_{Aa}$  et  $\omega_{Aa} > \omega_{aa}$ . En pratique, la superdominance est parfois assimilée à de la sélection positive en cas de détection précoce de la sélection. La raison de cette confusion provient du fait que les nouveaux mutants apparaissent

à l'état hétérozygote. On notera toutefois que dès lors que tous les allèles présentent des fréquences non négligeables dans la (les) population(s) étudiée(s), la superdominance engendre un excès d'hétérozygotes au locus sélectionné, relativement au standard défini par le polymorphisme neutre.

Un second mécanisme est la **sélection fréquence-dépendante** pour laquelle la valeur sélective associée à un génotype (ou un allèle) est inversement proportionnelle à la fréquence du génotype (ou allèle) dans la population étudiée. Un allèle rare est donc positivement sélectionné, relativement aux allèles fréquents et, au contraire, les allèles les plus fréquents sont contre-sélectionnés par rapport à tous ceux de moindre fréquence. Ce type de sélection entraîne une stabilisation de la distribution des fréquences alléliques. C'est, par exemple, ce mécanisme qui régit la sex-ratio des espèces di-sexuées et/ou la distribution des fréquences de génotype de compatibilité sexuelle chez certaines espèces de champignons et de plantes (Awadalla et Charlesworth 1999, Milgroom et Cortesi 1999, Uyenoyama 2005). Dans le cas de systèmes hôte-parasite, la sélection fréquence-dépendante est particulièrement invoquée pour les bases génétiques du système immunitaire des vertébrés (Garrigan et Hedrick 2003). On remarquera que parfois l'expression «sélection diversifiante» est appliquée au cas où l'on compare deux espèces-sœurs qui ont chacune fixé des mutations favorables, différentes au même locus, par des épisodes indépendants de sélection positive. La terminologie appliquée à la sélection, comme le phénomène lui-même, est sensible aux variations d'échelle d'observation.

**Figure E1.** Évolution au cours du temps de la fréquence de mutations sélectionnées, relativement à celle de mutations neutres.



## 2. Évolution à court terme de la sélection positive

Considérons toujours l'exemple de l'apparition d'échecs thérapeutiques avec la dernière molécule mise sur le marché pour tuer un pathogène déterminé. Dès son (leur) apparition, la (ou les) mutation(s) responsable(s) de cette résistance va (vont) augmenter en fréquence dans les populations traitées par cette dernière molécule puisqu'elles sont les seules à permettre au pathogène de survivre. Une telle augmentation de fréquence constitue une première définition de l'épisode de sélection positive que l'on cherche à détecter. À ce jour, les techniques de détection d'épisodes de sélection positive n'ont encore été que très rarement appliquées à l'étude des pathogènes. Les points 2.1 à 2.5 ci-après en détaillent les principes et le point 2.6 illustre comment l'application de certaines d'entre elles a permis de mieux comprendre l'évolution de chimiorésistances chez l'agent le plus sévère des paludismes humains, *P. falciparum*.

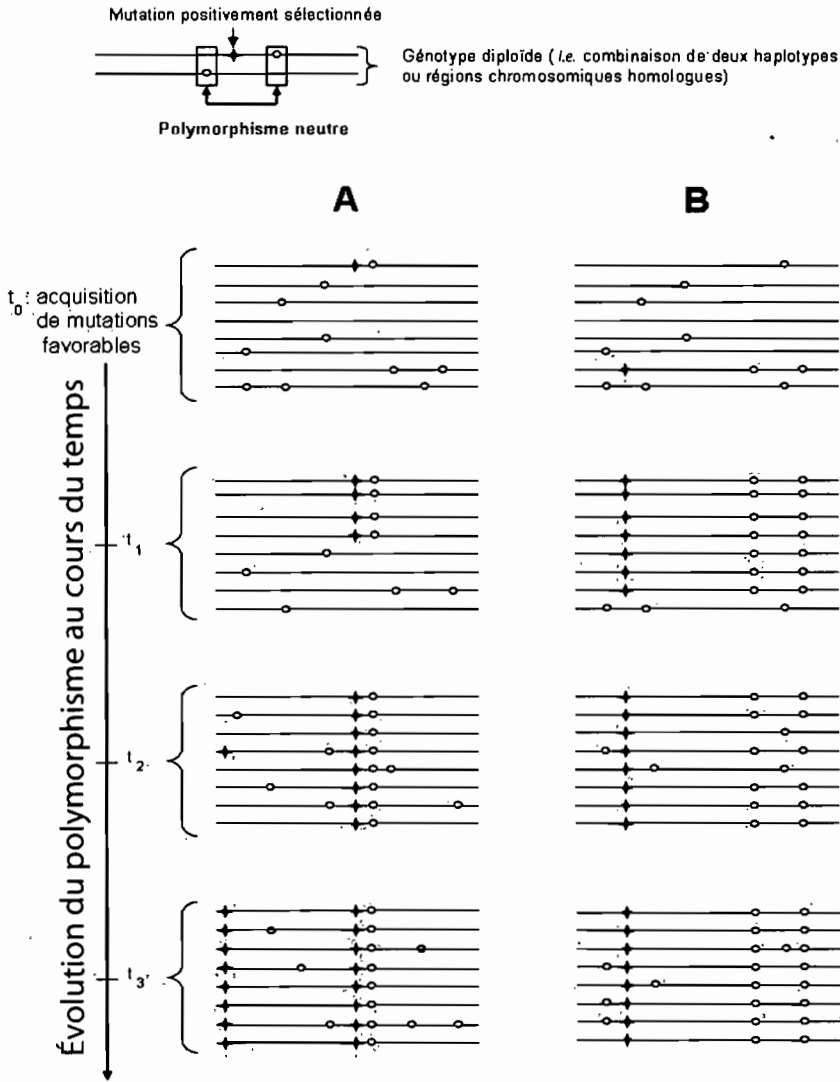
En préambule, on retiendra que quelque soit la technique de détection utilisée, un seul et même principe permettra de vérifier *in fine* que le polymorphisme d'un gène donné ait bien été modelé par la sélection : on attend en moyenne les mêmes signatures de structure du polymorphisme à tous les locus ayant évolué libres de toute action de la sélection (cf. Chapitre 8) alors qu'on attend une structure de polymorphisme singulière au locus sélectionné dans les populations où la sélection s'est exercée. Ceci définit alors la structuration (intragénotype, intra et/ou interpopulations) de polymorphisme neutre comme un « standard neutre » pour les populations étudiées ; « standard neutre » relativement auquel se définiront les signatures laissées par la sélection.

### 2.1 Que se passe-t-il au cours d'un épisode de sélection positive ?

La figure 9.1 illustre les conséquences d'un épisode de sélection positive qui démarrerait à la date  $t_0$ . Dans notre exemple de développement d'une nouvelle résistance à une drogue médicamenteuse, cette date  $t_0$  devient la date d'apparition de la première mutation permettant au pathogène de survivre en présence du médicament. La figure 9.1 suit la composition du polymorphisme présent le long de l'**haplotype** qui porte cette mutation, dans le cas où la sélection subie par cette mutation est plus (cas B) ou moins (cas A) forte. On observe que, tant en A qu'en B, l'augmentation en fréquence de la mutation sélectionnée modifie le polymorphisme neutre dans les régions voisines. Ce phénomène est appelé « **auto-stop** »

**FIG. 9.1** | Schéma illustrant un balayage sélectif et ses conséquences génétiques sur la zone soumise à sélection et les régions génomiques situées à proximité

Les barres horizontales représentent les chromosomes; les étoiles rouges des mutations favorables et les points bleus des mutations neutres. La date  $t_0$  représente la date d'apparition de la mutation qui va subir une sélection positive, c'est-à-dire qui va conférer un avantage sélectif aux individus qui la portent dans les populations suivies. La composition du polymorphisme présent le long de l'**haplotype** considéré est figurée en  $t_0$  et en trois dates plus récentes. Le cas A, à gauche, et le cas B, à droite, représentent deux trajectoires partant d'un même initial de polymorphisme neutre mais telles que la mutation favorable, apparue en A en  $t_0$ , subit une sélection positive moindre que celle apparue en B.



**génique**» ou «**balayage sélectif**» (Andolfatto 2001). Ses causes sont les suivantes. La mutation favorable apparaît par hasard sur un chromosome particulier d'un individu donné et est donc associée par hasard à un haplotype particulier. Par définition même de la sélection de la mutation favorable, l'haplotype portant cette mutation va se reproduire en un plus grand nombre de copies que les haplotypes ne portant pas la mutation. Ce phénomène d'entraînement s'altèrera dès qu'un événement de **recombinaison** emportera la mutation favorable dans un autre haplotype que celui où elle apparut, ou dès que des mutations recréeront un polymorphisme génétique neutre dans l'ensemble des descendants de l'haplotype dans lequel est apparue la mutation favorable. Ce phénomène d'auto-stop s'accompagne de différentes signatures de la sélection qui vont différencier ce que l'on observe sur les régions voisines de la mutation favorable de celles observées sur les autres régions neutres du génome. Ce sont ces signatures, décrites ci-dessous tour à tour, qui sont utilisées pour détecter les épisodes de sélection positive (Nielsen 2005).

## 2.2 Réduction de la diversité génétique locale

### a) La signature

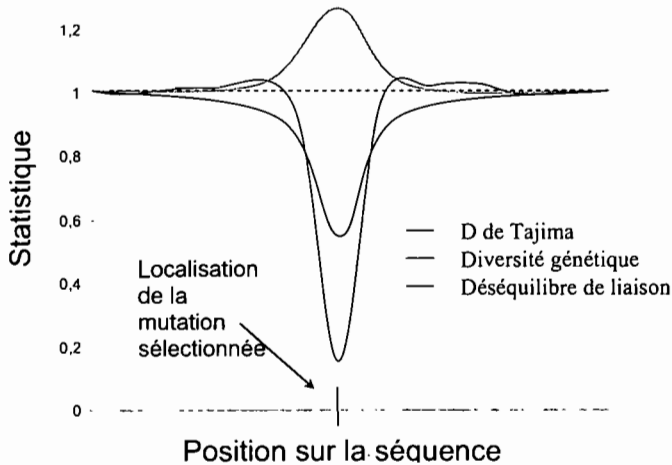
La probabilité de **recombinaison** entre le locus directement soumis à sélection et un locus voisin est d'autant plus forte que la distance entre eux est faible. Comme la recombinaison est le seul processus avec la mutation venant contrecarrer le phénomène de balayage sélectif, la diversité génétique va augmenter à mesure que l'on s'éloigne du locus où est apparue la mutation favorable. Au final, la sélection positive génère donc une signature caractéristique, présentée dans la figure 9.2, de réduction de la variabilité génétique en forme de vallée autour du locus sélectionné (Kim et Stephan 2000). Si le balayage sélectif n'est pas encore complet, la diversité génétique sera réduite principalement chez les individus porteurs de la mutation avantageuse (voir la comparaison des individus porteurs et non-porteurs de la mutation favorable dans la figure 9.1).

### b) Ses caractéristiques

Plusieurs facteurs modifient la taille de la région génomique affectée par le balayage sélectif. Le premier est le taux de recombinaison entre la mutation adaptative et la région considérée: plus ce taux est élevé, moins la réduction de diversité sera forte. Le second est l'intensité de la sélection positive et donc la vitesse à laquelle une mutation favorable augmente en fréquence dans la population: plus cette augmentation est rapide, plus la zone affectée par la diminution de diversité génétique est grande (voir le contraste entre les cas A et B de la figure 9.1). Le

**FIG. 9.2** | Schéma illustrant l'effet d'un balayage sélectif sur la variabilité génétique

Cette figure montre en particulier la réduction de variabilité génétique, l'augmentation du déséquilibre de liaison et la diminution du D de Tajima au niveau et autour de la région soumise à sélection positive. L'échelle de chaque statistique a été modifiée de telle manière à ce que les valeurs attendues sous l'hypothèse de neutralité soient égales à 1. Figure modifiée à partir de Nielsen (2005).



troisième facteur est la fréquence initiale de la mutation favorable lorsqu'elle commence à être favorisée par la sélection : plus cette fréquence initiale est élevée, moins la réduction de diversité génétique sera grande. Enfin, mais dans une moindre mesure, le taux de mutation peut jouer en recréant un polymorphisme au sein des descendants des haplotypes initialement associés à la mutation sélectionnée (Kim et Stephan 2000). Le retour au niveau de diversité génétique ayant précédé l'épisode de sélection positive est un processus très long. Par conséquent, la signature présentée sur la figure 9.2 peut perdurer très longtemps après l'arrêt d'un épisode de sélection positive.

### 2.3 Modification des relations entre allèles (fréquence et ressemblance)

#### a) Les signatures

L'épisode de sélection positive affecte aussi la distribution de fréquence des haplotypes distincts coexistant au sein de la population. En  $t_0$  – c'est-à-dire avant que l'épisode de sélection positive ait lieu – les haplotypes distincts sont tous plus ou moins présents à la même fréquence, alors que les dates postérieures voient un haplotype dominer en fréquence (cf. Figure 9.1). En d'autres termes, relativement

à l'attendu de distribution de fréquence d'allèles neutres, la sélection positive crée un excès de l'allèle fréquent (l'allèle positivement sélectionné), un excès d'allèles rares et un déficit d'allèles présentant des fréquences intermédiaires.

Par ailleurs, le jeu de la sélection positive change l'apparementement entre allèles, c'est-à-dire la distribution du nombre moyen de mutations différenciant deux haplotypes quelconques présents dans la population. Le cas B présenté sur la figure 9.1 permet de visualiser ce phénomène en comparant les échantillons collectés en  $t_0$  (avant l'épisode de sélection positive) et aux dates ultérieures. À  $t_0$ , l'échantillonnage de deux haplotypes quelconques conduit à l'observation de 1 à 6 mutations les différenciant l'un de l'autre. En  $t_1$ , ce nombre de mutations entre deux haplotypes quelconques ne prend que deux valeurs extrêmes : 0 dans les 21 occasions où on échantillonne deux fois l'haplotype sur lequel est apparue la mutation sélectionnée, et 6 dans les 7 autres cas. Le jeu de la recombinaison au sein de la région observée et des mutations neutres continuant à apparaître recrée des nombres intermédiaires à partir de  $t_2$ .

## b) « Spectre de fréquence » et test d'Ewens-Watterson

Le test d'Ewens-Watterson est basé sur l'analyse du « spectre de fréquence » des allèles dans un échantillon. Le « **spectre de fréquence** » correspond à la distribution du décompte d'allèles présentant une fréquence  $x_i = i/n$  pour  $i = 1, 2, 3, \dots, n-1$  dans un échantillon de  $n$  individus haploïdes. Un exemple d'obtention de spectre de fréquence est donné dans le tableau IX.1 pour un échantillonnage de 10 individus haploïdes ayant conduit à l'identification de 5 allèles distincts au locus étudié. Dans cet échantillon, l'allèle A est présent quatre fois, l'allèle B deux fois, l'allèle C une fois, l'allèle D deux fois et l'allèle E une fois. Le « spectre de fréquence » est donc  $\{2; 2; 0; 1; 0; 0; 0; 0; 0\}$  (cf. Tableau IX.1).

La probabilité d'obtenir un spectre de fréquence donné, sachant le nombre d'allèles observés et la taille de l'échantillon sous les hypothèses de neutralité et de **modèle en allèles infinis**, a été formalisée par Ewens (1972). Cette formule fournit donc le « standard » neutre de la distribution attendue des fréquences alléliques ainsi que de toute quantité entièrement déterminée par cette distribution sous l'hypothèse neutre, telle que l'homozygotie attendue. C'est cette homozygotie attendue sous l'hypothèse de neutralité qu'utilise le test d'Ewens-Watterson comme « standard » de polymorphisme neutre auquel comparer l'homozygotie observée dans l'échantillon (Watterson 1978). Un phénomène de sélection positive, en induisant la présence d'un allèle très fréquent et d'un grand nombre d'allèles rares, conduit à observer une valeur de l'homozygotie trop forte relativement à l'attendu neutre. Ce test est applicable à tout type de données pour lequel l'hypothèse de modèle de mutation en allèles infinis est raisonnable, c'est-à-dire tant aux données de polymorphisme enzymatique qu'aux données de **séquences**.

TAB IX.1   Détermination du « spectre de fréquence »									
Étape 1 : analyse de l'échantillonnage									
Échantillonnage	Nombre d'allèles distincts								
	Allèle A	Allèle B	Allèle C	Allèle D	Allèle E				
Individu n° 1	1	0	0	0	0				
Individu n° 2	0	1	0	0	0				
Individu n° 3	1	0	0	0	0				
Individu n° 4	0	0	0	1	0				
Individu n° 5	1	0	0	0	0				
Individu n° 6	0	0	0	1	0				
Individu n° 7	1	0	0	0	0				
Individu n° 8	0	1	0	0	0				
Individu n° 9	0	0	0	0	0			1	
Individu n° 10	0	0	1	0	0			0	
<b>Décompte total</b>	<b>4</b>	<b>2</b>	<b>1</b>	<b>2</b>	<b>1</b>				

Étape 2 : obtention du spectre de fréquence									
Nombre d'occurrence par allèle	1*	2	3	4	5	6	7	8	9
Nombre d'allèles	2	2	0	1	0	0	0	0	0

Dans une première étape (étape 1), les lignes renvoient aux allèles distincts observés dans l'échantillon et les colonnes aux individus échantillonnés; les chiffres 1 ou 0 marquent respectivement la présence ou l'absence d'un allèle dans le génotype d'un individu. Une seconde étape (étape 2) réarrange les observations précédentes pour déterminer le spectre de fréquence observé dans l'échantillon. Ici, la taille d'échantillonnage est de  $n = 10$  individus haploïdes; le spectre de fréquence est donc une suite de  $9 (= n-1)$  chiffres:

\* Correspond au nombre d'observations d'un allèle donné sur l'ensemble des 10 individus (ici, allèles observés une seule fois). Dans cet exemple, deux allèles distincts ne sont observés qu'une seule fois dans les 10 individus: l'allèle C et l'allèle E.

### c) La statistique D de Tajima (1989)

La statistique D de Tajima (1989) concerne précisément les données de séquence et les estimations de quantité de polymorphisme attendu dans une population. Dans le cas d'une population de taille constante  $N$  et à l'équilibre mutation-dérive, la quantité de polymorphisme neutre attendu  $\theta = 4N\mu$  peut être aussi bien estimé par l'estimateur  $\theta_S$ , basé sur le **nombre S de sites polymorphes** dans la population, que par l'estimateur  $\theta_\pi$ , basé sur l'utilisation du nombre moyen  $\pi$  de mutations différenciant deux séquences quelconques l'une de l'autre. Sous l'hypothèse de neutralité et d'absence de variations démographiques marquées,  $\theta = \theta_S = \theta_\pi$ . La statistique D de Tajima (1989) mesure la différence pondérée entre  $\theta_S$  et  $\theta_\pi$ , et prend donc une valeur nulle sous l'hypothèse de neutralité. En revanche, un épisode de sélection positive tendra à faire prendre à D une valeur négative (cf. Figure 9.2), et une valeur d'autant plus petite que le polymorphisme de séquence

observé est proche de la mutation sélectionnée (Braverman et coll. 1995). Des valeurs positives de  $D$  sont attendues dans le cas de phénomènes de **sélection diversifiante**.

Cette statistique  $D$  et ses variantes, comme le test de Fu et Li (1993), ont été parmi les méthodes les plus traditionnellement utilisées dans la détection d'éventuels épisodes de sélection positive dans le génome d'organismes variés. Il est important de noter cependant que cette statistique ne teste pas seulement l'écart à la neutralité sélective, mais aussi les écarts marqués à la stabilité démographique qui peuvent donc aussi faire dévier la valeur prise par  $D$ , de zéro vers les valeurs positives, dans le cas de goulots d'étranglements, et vers des valeurs négatives dans les cas opposés de forte expansion démographique. On peut toutefois différencier les cas où la cause d'un écart de  $D$  de la valeur nulle relève d'épisodes de sélection et ceux où la cause est d'ordre démographique, en utilisant des données de séquences prises dans différentes régions du génome. En cas de causes démographiques, l'ensemble du génome est affecté de la même façon et donc toutes les séquences aboutiront au même signal. En cas d'épisodes sélectifs, seule(s) la (les) région(s) sélectionnée(s) et ses (leurs) régions voisines aboutiront à un écart à l'attendu neutre.

## 2.4 Augmentation locale du déséquilibre de liaison et modification de la composition haplotypique

### a) La signature

Il y a **déséquilibre de liaison** lorsque les allèles présents à différents locus ne sont pas associés de manière aléatoire au sein des individus de la même population, mais que les polymorphismes observés à différents locus sont au contraire corrélés. Par corrélation, il faut entendre que l'observation d'un allèle particulier à un locus donné renseigne plus ou moins sur l'allèle présent à un autre locus. La sélection positive tend à augmenter le déséquilibre de liaison à proximité de la mutation favorable, du fait que la combinaison de polymorphisme neutre aux locus proches du locus sous sélection a été entraînée par l'augmentation de fréquence de l'haplotype (cf. Figures 9.1 et 9.2). La recombinaison au sein de l'haplotype est un des phénomènes (avec la mutation) qui va «casser» ces associations de polymorphisme entre le locus sélectionné et ses voisins. Par conséquent, la signature de déséquilibre de liaison associée à un épisode de sélection positive décroît à mesure que l'on s'éloigne du locus directement sous sélection.

## b) Un outil de détection de la sélection

Sous l'hypothèse de neutralité, on s'attend à ce que les allèles d'un locus qui sont présents en fréquence élevée dans une population aient eu le temps de s'associer avec différents allèles aux locus voisins par le jeu de la mutation et de la recombinaison. En revanche, après un épisode de sélection positive, l'allèle le plus fréquent au locus sélectionné n'a pas encore eu le temps de casser les déséquilibres de liaison avec les allèles présents dans le même haplotype au moment de l'apparition de la mutation favorable. Plusieurs tests utilisent cette différence de déséquilibre de liaison entre locus physiquement liés ou la composition haplotypique pour identifier les régions du génome ayant connu des événements passés de sélection positive. C'est le cas, par exemple, du test dit de « l'haplotype étendu ». Le principe de ce test consiste simplement à analyser la taille et la variation des haplotypes autour d'un allèle et de déterminer par simulations si les valeurs prises par ces deux paramètres sont attendues sous l'hypothèse de neutralité.

## 2.5 Différenciation génétique entre populations

### a) La signature

Le chapitre 8 a montré comment les actions combinées du système de reproduction de l'espèce parasite, de la dérive génétique au sein de chacune de ces populations et de la migration entre ces populations définissent un niveau de différenciation génétique « neutre » entre populations de ce parasite.

Imaginons à présent que deux populations différenciées de ce parasite soient confrontées aux mêmes conditions environnementales; c'est-à-dire qu'elles correspondent à une même espèce-hôte et plus particulièrement à des individus-hôtes sélectionnant de la même manière les individus parasites. Imaginons aussi qu'une mutation permettant une meilleure valeur sélective de ce parasite pour l'exploitation de ces individus-hôtes apparaisse dans la première population et migre rapidement dans la seconde. Dans l'une comme dans l'autre population du parasite, vont alors se répéter quasiment à l'identique la même histoire d'auto-stop pour toutes les régions voisines de la mutation sélectionnée dans cet environnement commun. En d'autres termes, les mesures de différenciation génétique vont être faibles lorsqu'elles s'appuieront sur le polymorphisme observé dans les **régions flanquantes**, relativement aux mesures de « standard neutre » s'appuyant sur le polymorphisme observé dans d'autres parties du génome du parasite.

Imaginons à l'inverse que deux populations d'une même espèce de parasites subissent chacune des pressions de sélection différentes (par exemple, deux traitements thérapeutiques différents ou deux espèces-hôtes sélectionnant le parasite

différemment). La sélection positive va alors vraisemblablement tendre à fixer certaines mutations dans l'une de ces populations, et d'autres mutations dans l'autre. On parle d'adaptation locale pour signifier que ces populations se sont adaptées aux conditions environnementales locales dans lesquelles elles évoluent. Sous l'hypothèse d'adaptation locale, les différences de sélection entre populations vont augmenter les mesures de différenciation génétique entre populations, résultant du jeu combiné de la dérive, de la migration et du régime de reproduction de l'espèce parasite étudiée uniquement pour les régions du génome voisines des locus sélectionnés. Autrement dit, les mesures de différenciations génétiques entre populations de parasites, obtenues pour les régions voisines des locus sélectionnés, vont être significativement plus fortes que celles de « standards neutres » obtenues pour d'autres régions du génome du parasite.

### b) Un outil de détection: analyses comparées entre locus

L'analyse comparée de mesure de différenciation génétique par les méthodes décrites au chapitre 8 entre locus dont le polymorphisme est soit « neutre », soit potentiellement sélectionné, a été utilisée de manière récurrente pour vérifier les hypothèses de sélection (Beaumont et Balding 2004).

## 2.6 Évolution des chimiorésistances chez un des agents du paludisme, *Plasmodium falciparum*

Les médicaments utilisés en lutte antiparasitaire ont pour effets de tuer les parasites, d'empêcher leur reproduction et/ou de bloquer leur transmission. Dès lors, toute mutation qui permet au parasite de réduire de tels effets confère un avantage sélectif important en présence de traitement. Dans le cas extrême, seul les individus porteurs d'une mutation conférant une chimiorésistance persistent dans la population d'hôtes traitée. Ces caractéristiques de fortes intensités de la lutte antiparasitaire et de la sélection pour la chimiorésistance font de ces chimiorésistances un exemple classique de sélection positive. Rechercher les empreintes de sélection positive liées à leur développement permet d'identifier les gènes responsables et de retracer l'histoire de ces adaptations, comme le montrent les exemples d'évolution de résistances à différents antipaludiques chez l'agent majeur du paludisme humain, *Plasmodium falciparum*.

Chaque année, le paludisme est responsable de 300 à 500 millions de malades et 1,5 à 2,7 millions de décès chez l'homme avec 80% des cas enregistrés en Afrique sub-saharienne, majoritairement chez les enfants de moins de cinq ans et les femmes enceintes. *P. falciparum* est l'agent responsable de sa forme la plus grave et sa répartition intertropicale concerne aussi bien l'Afrique que l'Asie ou l'Amé-

rique du Sud. Différents traitements médicamenteux ciblent ce parasite (cf. Tableau IX.2). La chloroquine est celui qui, avec la quinine, a été le plus employé tant en traitement préventif que curatif. En 1960 apparaissent les premiers cas de chloroquino-résistances en Amérique du Sud et en Asie du Sud-Est (Anderson et Roper 2005). Celles-ci s'étendent ensuite très rapidement aux autres zones impaludées africaines (Trape 2001, Anderson et Roper 2005). Sur l'ensemble de l'aire de répartition de *P. falciparum*, l'introduction de la chloroquine s'est traduite par une augmentation progressive de la fréquence des résistants jusqu'à atteindre 100% dans certaines régions géographiques. Quels gènes sont impliqués dans cette résistance? Est-ce que les mêmes gènes, voire les mêmes mutations, sont impliqués dans la résistance d'une région à l'autre? À quelle vitesse a évolué la résistance?

TAB IX.2   Exemples de traitements curatifs ou prophylactiques utilisés contre <i>P. falciparum</i>	
<b>Traitements curatifs</b>	Chloroquine
	Sulfadoxine-pyriméthamine
	Quinine
	ACT*
<b>Traitements prophylactiques</b>	Chloroquine
	Proguanil
	Association pyriméthamine-dapsone
	Méfloquine
	Doxycycline

\* Artemisin-based Combination Therapies

Pour répondre à ces questions, Wootton et coll. (2002) ont analysé le polymorphisme de marqueurs microsatellites répartis sur l'ensemble des 14 chromosomes du génome de 87 parasites échantillonnés en Afrique, Asie, Papouasie Nouvelle-Guinée et Amérique du Sud. Ils analysèrent la diversité génétique le long de chacun des chromosomes, tant chez les parasites résistants que chez les parasites sensibles, et n'observèrent de réduction forte de cette diversité chez les parasites résistants que dans une région du génome localisée sur le chromosome 7. Cette région est centrée autour du gène *Pfprt*, un gène codant pour un transporteur localisé dans les vacuoles digestives du parasite, qui était déjà suspecté jouer un rôle dans la résistance à la chloroquine (Fidock et coll. 2000). Leur observation confortant cette hypothèse, ils analysèrent plus finement le polymorphisme de *Pfprt* et de ses régions flanquantes et comparèrent le niveau de différenciation génétique, mesuré à l'aide de ces polymorphismes, aux mesures obtenues pour le polymorphisme observé dans d'autres régions du génome du parasite. L'analyse locale du polymorphisme et celle de la différenciation génétique entre les régions géographiques leur a permis d'identifier un minimum de quatre apparitions indé-

pendantes de résistances à la chloroquine. Une première mutation conférant une chloroquino-résistance s'avère être apparue au locus *Pfcr* en Asie puis avoir colonisé l'ensemble des populations asiatiques et africaines du parasite. Cette conclusion provient de l'observation d'une mesure de différenciation génétique entre échantillons asiatiques et africains, significativement plus faible lorsqu'elle s'appuie sur le polymorphisme observé dans les régions flanquantes du locus *Pfcr* que lorsqu'elle se base sur le polymorphisme observé dans d'autres parties du génome du parasite. Il s'avère aussi que, de manière indépendante à cette mutation ayant envahi l'Asie et l'Afrique, une seconde mutation résistante soit apparue au même locus *Pfcr* en Papouasie Nouvelle-Guinée et deux autres en Amérique du Sud. L'analyse de la diminution du déséquilibre de liaison dans les régions chromosomiques voisines du gène *Pfcr* leur a aussi permis de déterminer la force de la sélection et la vitesse avec laquelle ces mutations avaient envahi différentes zones géographiques. Les auteurs en conclurent, en particulier, que la résistance à la chloroquine avait évolué au cours des 20 à 80 générations suivant l'introduction de la chloroquine. En évaluant le temps de génération du parasite à 4 mois, cette estimation d'évolution de la résistance en termes de générations de parasite correspondrait à une fourchette de 6 à 20 ans, ce qui est conforme aux données épidémiologiques. Cette étude, fondée sur l'analyse des signatures de sélection (réduction de diversité génétique, déséquilibre de liaison, différenciation génétique) montre tout l'intérêt des approches sans *a priori* fonctionnel. Dans ce cas particulier, elle a permis d'identifier (i) un gène majeur de la résistance à la chloroquine chez *P. falciparum*, *Pfcr*, (ii) des allèles de résistance d'identité variable entre régions géographiques, (iii) au moins quatre événements d'apparition indépendants de la chimiorésistance et (iv) des événements de migration longue-distance d'un de ces allèles de résistance.

Lorsque la chloroquine n'a plus eu d'effet thérapeutique du fait de la présence de chloroquino-résistances, on a généralement utilisé la sulfadoxine-pyriméthamine (ou Fansidar®) comme traitement de substitution. Cette molécule agit en inhibant deux enzymes indispensables à la croissance du parasite : la dihydrofolate réductase, codée par le gène *dhfr*, et la dihydroptéroate synthétase, codée par le gène *dhps*. Malheureusement, des souches résistantes sont apparues cinq ans seulement après l'introduction de cette molécule en traitement thérapeutique (Hurwitz et coll. 1981). L'analyse du polymorphisme de *dhfr* et *dhps* a rapidement montré l'implication, au moins partielle, de mutations à ces locus dans les phénomènes de résistance cités (Plowe et coll. 1997; Basco et coll. 1998). Là encore, cette conclusion provient de l'observation de signatures caractéristiques d'un phénomène de sélection positive, telles qu'une forte réduction de la diversité génétique et une augmentation du déséquilibre de liaison aux alentours de ces deux locus situés respectivement sur les chromosomes 4 et 8 (Nair et coll. 2003,

McCollum et coll. 2007). En analysant la variabilité génétique sur l'ensemble du génome et en recherchant diverses signatures de sélection, Volkman et coll. (2007) ont montré l'implication probable d'une troisième région génomique, non encore suspectée et de fonction encore inconnue, localisée sur le chromosome 13.

Tous les pays où sévit le paludisme n'ont pas eu recours aux mêmes traitements; la nature chimique du médicament, sa dose et sa fréquence d'application varient. Comment le parasite a-t-il réagi à ces variations de pressions de sélection? Anderson et coll. (2005) ont démontré que le parasite s'adaptait aux conditions locales de traitement. Pour ce faire, ils ont étudié la différenciation génétique de parasites provenant de populations humaines différant par leur histoire thérapeutique. Ils ont mené cette étude à la fois au niveau de différents gènes du parasite potentiellement impliqués et à d'autres locus dont le polymorphisme était potentiellement neutre. Comme attendu sous l'hypothèse d'adaptations du parasite aux traitements locaux, ils observèrent une plus forte différenciation génétique au niveau des gènes impliqués dans les différentes chimiorésistances que pour les locus de polymorphisme *a priori* neutre (Anderson et coll. 2005).

Ces exemples illustrent l'importance et la puissance des recherches de signatures de sélection positive dans la compréhension de l'évolution adaptative des agents parasitaires et infectieux. Leur utilisation n'est bien sûr pas limitée aux seuls cas d'évolution de chimiorésistances mais est applicable à l'analyse d'adaptation à tout autre changement environnemental identifié (colonisation d'une nouvelle espèce-hôte, changement de traits d'histoire de vie des hôtes, etc.).

### 3. Évolution à long terme et empreintes de la sélection dans les séquences codantes

#### 3.1 Les principes

Le fonctionnement démographique en populations – c'est-à-dire en entités regroupant la majorité des descendants d'un groupe de parents (voire la majorité des couples reproducteurs dans le cas d'espèces sexuées) – induit nécessairement un signal de différenciation génétique entre populations (cf. Chapitre 8). Imaginons que les populations étudiées colonisent deux habitats, et que l'on cherche à tester l'hypothèse selon laquelle certains génotypes déterminent une meilleure adaptation du parasite à l'un ou l'autre de ces habitats. Comment différencier les signaux de différenciation génétique entre populations colonisant des habitats différents parmi des signaux ne résultant que du fonctionnement démographique

de l'espèce étudiée? Les points 2.3c et 2.5 de ce chapitre nous ont indiqué une stratégie laborieuse mais puissante dans la détection d'épisodes sélectifs récents, qui est basée sur la multiplication du nombre de marqueurs indépendants: les marqueurs « neutres » vont tous conduire au même « standard neutre » de différenciation génétique entre habitats qui sera une mesure plus faible que celle obtenue pour les marqueurs dont le polymorphisme est impliqué dans l'adaptation à l'un et l'autre des habitats.

Une stratégie complémentaire a été élaborée pour des épisodes sélectifs plus anciens, c'est-à-dire pour ceux dont le jeu de la mutation et de la recombinaison a réussi à « casser » les déséquilibres de liaison créés par les épisodes sélectifs à l'intérieur d'haplotypes de quelques centaines ou milliers de bases. Cette stratégie est applicable aux données de séquences de régions codantes et se base sur les deux propriétés suivantes de ces régions codantes. Premièrement, tous les sites nucléotidiques d'un même gène sont égaux face à la probabilité d'apparition d'une mutation et ont expérimenté la même histoire évolutive « neutre » (c'est-à-dire résultant des changements de taille de population, des événements de migration et du système de reproduction). Dès lors, aux événements de sélection près, tous les sites nucléotidiques d'un gène sont égaux face aux probabilités de conserver un polymorphisme ou de fixer des mutations différentes dans chaque groupe de populations. Plus spécifiquement, pour une population de démographie stable comptant  $N$  individus diploïdes et un taux de mutation  $\mu$  au gène considéré: la probabilité d'apparition d'une nouvelle mutation à ce gène vaut  $p_{\mu} = 2N\mu$ ; la probabilité que cette nouvelle mutation se fixe dans la population sous la seule action de la dérive vaut  $p_f = 1/2N$ ; la probabilité de fixation de nouvelles mutations neutres vaut donc  $p_{\mu} \times p_f = \mu$  pour une quantité de polymorphisme neutre attendu dans la population de  $\theta = 4N\mu$ .

Deuxièmement, du fait de la dégénérescence du code génétique, une mutation d'un **site nucléotidique** d'une région codante a un impact très variable sur le résultat protéique. Ainsi, par exemple, les mutations affectant la troisième position d'un **codon** n'affectent qu'assez rarement l'acide aminé incorporé et sont même absolument sans effet dans 32 cas sur 64. On peut alors classer les mutations affectant les régions codantes en **mutations « synonymes »**, sans effet sur la protéine, et en **mutations « non-synonymes »**, qui induisent soit un changement d'acide aminé, soit l'arrêt de la traduction *via* l'introduction d'un codon « stop ». Faisons alors l'hypothèse, faiblement contraignante, que les adaptations à de nouvelles contraintes environnementales passent par une modification de séquence(s) protéique(s), c'est-à-dire par une sélection de mutations nécessairement non-synonymes. Le « standard neutre » de différenciation génétique est alors donné par les taux de substitution de mutations synonymes (c'est-à-dire le rapport entre le nombre de sites nucléotidiques présentant des mutations synonymes sur le

nombre total de sites nucléotidiques polymorphes). Notons  $dS$  ce taux de substitution de mutations synonymes et  $dN$  le taux de substitution des mutations non-synonymes observé pour le même jeu de séquences et donc pour la même histoire évolutive. Les deux types de mutations étant créées à la même fréquence, tester l'hypothèse que le gène étudié ait expérimenté un épisode de sélection revient à comparer les taux de substitution  $dN$  (potentiellement sélectionné) et  $dS$  (standard neutre). La relation  $dN = dS$  est attendue sous l'hypothèse de neutralité. La relation  $dN < dS$  indique que les mutations non-synonymes apparues ont eu tendance à être éliminées par la sélection dans la suite de leur histoire. Cette relation repère donc les événements de **sélection purifiante**. Enfin, la relation  $dN > dS$  indique qu'au contraire les mutations non-synonymes ont été plus souvent retenues qu'attendues sous l'hypothèse de neutralité, et repère donc les événements de sélection positive (Yang et Bielawski 2000).

### a) Les méthodes utilisant la différence « synonymes/non-synonymes »

Le principe d'utilisation des mutations synonymes comme standard « neutre » a d'abord été appliqué à la détection d'épisodes sélectifs ayant accompagné des événements de **spéciation**. La méthode de MacDonald-Kreitman (McDonald et Kreitman 1991) teste ainsi l'indépendance ligne-colonne dans un tableau de contingence  $2 \times 2$  dénombrant les mutations observées en fonction, d'une part, de leur statut « synonyme » ou « non-synonyme », et d'autre part, de leur occurrence en tant que polymorphisme intraspécifique ou de fixation d'états alternatifs dans chaque espèce (voir l'exemple donné dans le tableau IX.3).

Nombre de sites	Mutation synonyme	Mutation non-synonyme
polymorphes dans au moins une espèce	42	2
fixés pour des états différents dans chaque espèce	17	7

Les chiffres donnés ici sont ceux publiés par McDonald et Kreitman (1991) dans l'article original proposant ce test. Il s'agit de valeurs obtenues dans le cas du gène codant pour l'alcool deshydrogénase chez 3 espèces du sous-groupe *Drosophila melanogaster*. Sous l'hypothèse de neutralité, le polymorphisme partagé entre espèce et celui fixé différemment dans les espèces comparées ne dépendent que du taux de mutations et doivent donc faire appel aux mêmes fréquences de « mutations synonyme » et de « mutations non-synonymes ». Ici, en revanche, on observe un excès de mutations « non-synonymes » parmi les sites fixés entre espèces, relativement à leur fréquence au sein des sites polymorphes ( $p = 0,006$ ). Cet excès suggère que ce gène a expérimenté des épisodes de sélection positive différentes entre espèces.

L'affinement suivant a consisté à utiliser le **signal phylogénétique** contenu dans les variations de séquences (que les mutations soient ou non synonymes) puis à retracer par des méthodes d'**estimation du maximum de vraisemblance**

l'histoire des mutations synonymes et non-synonymes le long de l'arbre phylogénétique obtenu (Yang et Bielawski 2000). L'hypothèse nulle première sur laquelle s'appuie ces méthodes est que tous les nucléotides d'un même gène et d'un même arbre phylogénétique évoluent avec la même valeur du rapport  $\omega = dN/dS$ , c'est-à-dire sous un même régime de sélection  $\omega_0$  qu'il s'agit d'estimer. Cette hypothèse nulle conduit presque toujours à une estimation de  $\omega_0 < 1$ , c'est-à-dire à mettre en évidence qu'en moyenne les régions codantes évoluent sous l'action d'une sélection purifiante. Un premier lot d'hypothèses alternatives laisse la possibilité aux codons d'un même gène de se distribuer dans des classes définies par différentes valeurs du rapport  $\omega = dN/dS$ ; ces rapports restent invariables le long de l'arbre phylogénétique correspondant à l'échantillonnage étudié. Dans le cas le plus simple, le nombre de classes reste restreint à deux ( $\omega_0 < 1$  signalant l'action de sélection purifiante,  $\omega_1 = 1$  signalant un polymorphisme nucléotidique neutre) ou trois ( $\omega_0 < 1$ ,  $\omega_1 = 1$  et  $\omega_2 > 1$ , cette dernière classe signalant l'action de sélection positive), mais d'autres types de distribution des rapports  $\omega$  ont été proposés et utilisés (voir pour une revue Yang et coll. 2000). Un second assouplissement de ces méthodes a consisté à permettre une variation du rapport  $\omega = dN/dS$  s'appliquant à une classe donnée de sites nucléotidiques le long de l'arbre phylogénétique (Zhang et coll. 2005). Parallèlement, les méthodes d'identification de polymorphismes associés à des valeurs  $\omega_x > 1$  se sont améliorées en diminuant le nombre d'erreurs de détection d'épisodes de sélection positive (Yang et coll. 2005). Enfin, un travail intense de simulations a été effectué dans le but de déterminer la robustesse des résultats. Il a été ainsi montré que le rapport  $\omega = dN/dS$  restait bien indépendant des variations locales du taux de mutation (Yang et Bielawski 2000) et que les biais introduits par les variations démographiques ou la recombinaison restaient globalement raisonnables (voir Anisimova et coll. 2001, 2002, Anisimova et coll. 2003, Anisimova et Yang 2007, par exemple). Au bilan, le chercheur dispose donc aujourd'hui d'un corpus de méthodes puissantes et relativement robustes pour détecter des événements de sélection de nature relativement diverse (par exemple, purifiante ou positive, constante ou variable dans le temps et/ou dans l'espace, attachée à un groupe de populations expérimentant des habitats variables, ou associée à des événements de spéciation).

## b) L'application de ces méthodes aux parasites

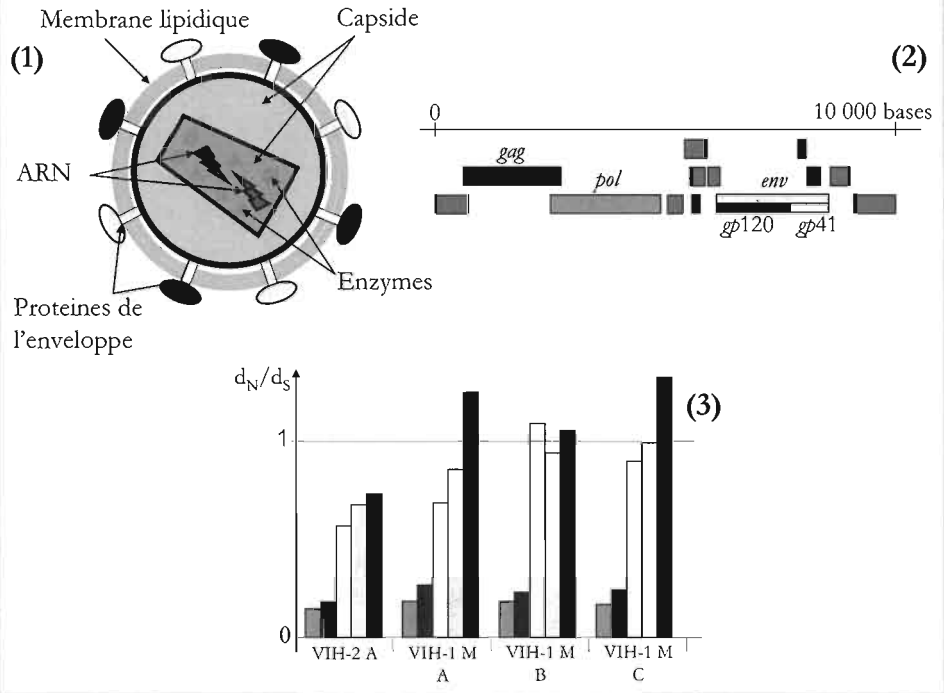
Deux des facteurs évolutifs, soupçonnés comme ayant joué (et jouant toujours) des rôles majeurs dans l'évolution des virus d'immunodéficience humaine, VIH responsables du Sida (syndrome d'immuno-déficience acquise), sont associés aux variations de l'hôte. Le premier est un changement d'espèce-hôte avec l'estimation de trois passages indépendants de grands singes à l'homme dans le cas du

virus VIH-1 et de sept passages indépendants de grands singes à l'homme dans le cas du virus VIH-2 (cf. Chapitre 7). Le second est l'évolution intra-hôte du système immunitaire humain qui tente d'éliminer le virus pendant toute la phase d'infection pré-sidaïque. Sous ces hypothèses, les régions du génome viral ayant le plus souvent et fortement expérimenté des épisodes de sélection positive devraient être associées aux sites de reconnaissance du virus par l'hôte. C'est ce qu'ont cherché à tester Choisy et ses collaborateurs (2004) au travers d'analyses du rapport  $\omega = dN/dS$  sur les gènes viraux *gag*, *pol* et *env*, ce dernier codant la protéine d'enveloppe virale, c'est-à-dire la protéine virale la plus « visible » par le système immunitaire. Ces auteurs ont mené la même analyse sur 22 jeux de données de virus isolés chez l'homme. Ces jeux de données se différenciaient soit par le gène, soit par le sous-type viral étudiés. Leurs premiers résultats confortent l'hypothèse d'un rôle majeur de la reconnaissance immunitaire dans l'évolution virale: les plus fortes signatures de sélection positive ont été identifiées sur le gène *env* (Figure 9.3). Ces auteurs ont ensuite cherché à tester si les sites nucléotidiques soumis à sélection positive le long du gène *env* étaient ou non les mêmes entre sous-types viraux. Ils ont découvert une tendance significative pour une localisation regroupée des sites portant des signatures de sélection positives au sein des mêmes régions du gène *env*: 10 des 16 positions soumises à sélection positives étaient identiques chez les virus VIH-1 de sous-type M, ainsi que 6 des 8 positions associées à  $\omega > 1$  entre virus VIH-1 de sous-type M ou O. Ces auteurs ont poursuivi leur investigation en testant si ces sites évoluant sous l'action répétée de sélection positive correspondaient à ceux qui codaient des régions protéiques reconnues par des effecteurs du système immunitaire (anticorps, lymphocytes T, etc.) ou par des enzymes de **glycosylation**. Ils observèrent une tendance significative pour que ces sites sous sélection positive récurrente diffèrent des zones de reconnaissance des effecteurs cellulaires de l'immunité humaine. En revanche, ces sites sont beaucoup plus souvent des sites de reconnaissance d'enzymes de glycolysation, autrement dit des sites où des sucres de l'hôte se lient à l'enveloppe protéique virale. Ceci suggère qu'un des avantages adaptatifs clefs de ces virus tient à leur capacité à se rendre « invisible » au regard du système immunitaire en masquant leur enveloppe protéique par une couverture de sucres.

Les mêmes méthodes de détection de sélection positive par l'analyse des rapports  $\omega = dN/dS$  ont également été appliquées sur des virus pour lesquels les relations hôtes-parasites sont beaucoup moins étudiées et comprises que pour les virus VIH-1 et VIH-2A. C'est le cas, par exemple, des quatre **sérotypes** de *Flavivirus* de la dengue (DEN-1, DEN-2, DEN-3 et DEN-4) pour lesquels les mécanismes mis en jeu au cours des infections humaines et, en particulier, dans le développement de pathologies restent du domaine de l'hypothèse. Ces virus sont distribués dans quasiment la totalité de la ceinture intertropicale où ils sont

**FIG. 9.3** | Schéma illustrant l'évolution adaptative de différents gènes du virus VIH

1. Structure du virus VIH. 2. Localisation des différents gènes le long du génome du virus. 3. Rapport  $d_N/d_S$  calculé pour chaque gène ou partie de gène (*gag*, *pol*, *env*, *env-gp120* et *env-gp41*) pour les virus VIH-1 de groupe M et de sous-types A, B et C, et les virus VIH-2 de sous-type A. Dans 1, 2 et 3, chaque gène analysé est symbolisé par une couleur donnée (par exemple, le gène *env* est associé à la couleur bleue).



généralement transmis à l'homme par deux moustiques adaptés aux habitats humains, *Aedes aegypti* et *Ae. albopictus*. Quelque soit leur sérotype, ces virus sont responsables de pathologies cliniques d'intensités très variables (allant d'infections asymptomatiques à des formes mortelles), et sont la cible d'une acquisition d'immunité à vie après guérison de l'infection, empêchant toute autre infection d'un virus de même sérotype. Là encore, une des hypothèses émises sur les facteurs décisifs de l'évolution de ces virus est la reconnaissance des virus par le système immunitaire humain. L'analyse de la variation des rapports  $\omega = d_N/d_S$  le long de la séquence génomique virale et entre régions géographiques montre un tableau légèrement plus complexe. Des événements de sélection positive ont été détectés chez certains génotypes de DEN-2 et dans certaines populations humaines, mais tous ces sites ne correspondent pas à des zones de l'enveloppe protéique reconnues par le système immunitaire (Twiddy et coll. 2002). Parallèlement, le

suivi au cours du temps de la diversité génétique des virus ayant sévi dans l'île de Porto Rico a permis d'identifier des épisodes de sélection positive sur des gènes viraux n'intervenant pas dans la reconnaissance immunitaire (Bennett et coll. 2003, Foster et coll. 2003). De telles détections sont autant de renseignements précieux dans la compréhension des relations hôtes-parasites se déroulant au cours de l'infection.

### 3.2 Avantages et inconvénients

Une des questions récurrentes, lorsque l'on s'intéresse à l'apparition d'espèces parasites, concerne les avantages et inconvénients (pour leur propre valeur sélective) de leur virulence vis-à-vis de leurs hôtes. Est-il préférable pour la survie à long terme du parasite qu'il exploite rapidement son hôte avant qu'un compétiteur ne le fasse (quitte à le tuer et donc à éteindre cette ressource)? Ou, au contraire, est-il préférable sur le long terme de ménager le plus possible son hôte quitte à voir des parasites compétiteurs piller cette ressource? Une question corollaire tient à la part relative des gènes du parasite intervenant dans la régulation de sa virulence dans l'histoire adaptative de cette espèce, relativement à ses espèces-sœurs à mode de vie libre. Tel est l'angle d'étude choisi par Anisimova et ses collaborateurs (2007) dans leur analyse des relations entre la variation du rapport  $\omega = dN/dS$  et l'évolution du parasitisme dans le genre bactérien *Streptococcus*. Ce genre comprend des espèces à mode de vie libre comme *S. thermophilus* ou *S. mutans* et des espèces pathogènes, de l'homme par exemple, telles que *S. pneumoniae* (susceptibles d'être responsables de méningites, de pneumonies, d'infections oculaires, voire de septicémies générales), *S. pyrogenes* pouvant engendrer des pharyngites, des dermatoses ou des pneumonies, ou *S. agalactiae* pouvant être responsables de méningites ou de pneumonies chez le nouveau-né ou d'infections urinaires ou de dermatoses chez l'adulte. Il émerge que sur 1 736 groupes de gènes analysés, 136 conduisent à la détection d'un épisode de sélection positive chez les espèces parasites. Ceci dit, si 29% des cas de sélection positive sont associés à des gènes intervenant dans la virulence des *Streptococcus* vis-à-vis de leurs hôtes, 19% sont associés à des protéines de même fonction chez les espèces libres et parasitaires, mais dont l'expression est connue pour être tissu-dépendante en cas de parasitisme (c'est-à-dire des protéines qui ne sont exprimées que dans certains tissus de l'hôte). Sachant que les analyses d'expression ont retenu moins d'attention que les recherches de déterminants de la virulence, ceci relativise la part de la virulence proprement dite dans les trajectoires adaptatives des espèces parasites (cf. Chapitre 10) au bénéfice de leur aptitude à la spécialisation et/ou adaptation à autant d'habitats différents que de tissus de leurs individus-hôtes. Ces résultats suggèrent aussi qu'une part de la virulence peut aussi émerger en

tant que traits complexes liés à l'aptitude à coloniser tel ou tel tissu de l'hôte (c'est-à-dire à s'adapter à des environnements précis ou, au contraire, à une gamme d'environnements différents).

## 4. Dynamique adaptative et épidémiologie

Les paragraphes précédents ont fait le point sur les méthodologies d'identification d'épisodes sélectifs. Il est temps de rediscuter brièvement de l'intérêt de s'interroger sur les adaptations des pathogènes pour la compréhension de leur dynamique épidémiologique. L'exemple précédent de détection de sélection positive sur des gènes d'expression tissu-dépendante chez les espèces parasites du genre *Streptococcus* illustre un cas où s'interroger sur les histoires adaptatives des parasites nous informe sur leur dynamique intra-hôte et la base phénotypique de leur virulence. Un exemple complémentaire est illustré par le cas des virus VIH-1 et VIH-2 responsables du Sida. Nous avons vu plus haut que la diversité virale observée entre infections indépendantes nous renseignait sur l'**avantage adaptatif** de certains génotypes viraux du type VIH-1 à limiter la probabilité d'être reconnus par les systèmes immunitaires, et ce au travers de la protection offerte par la liaison sur l'enveloppe virale. Une autre étude a recherché des signatures de sélection positive au cours de l'infection d'un individu-hôte (Lemey et coll. 2007). Cette étude a alors montré que la forte variation de durée de l'infection avant que le système immunitaire ne soit devenu inopérant (de 6 mois à 20 ans, dans le cas d'infection par VIH-1) n'était corrélé ni au rapport  $\omega = dN/dS$  ni au taux de substitutions non-synonymes ( $dN$ ), mais au taux de substitutions synonymes ( $dS$ ). En d'autres termes, à cette échelle intra-hôte, la variation de virulence n'est pas un trait sélectionné du parasite, ni même apparemment un trait d'interaction hôte-parasite, mais semble une conséquence de la qualité initiale de l'hôte : plus le système immunitaire permettra un fort taux de réplication virale en début d'infection, plus l'infection sera virulente.

L'étude de Coyne et ses collaborateurs (2007) sur un virus de chat (calicivirus) s'est adressée à la fois à l'échelle de l'infection d'un hôte et à celle d'une population d'hôtes. Là encore, les épisodes de sélection positive se sont avérés cibler en priorité les régions protéiques virales reconnues par le système immunitaire des hôtes. En combinant les analyses de polymorphismes sélectionnés et neutres, ces auteurs ont obtenu plusieurs résultats complémentaires. Ils ont confirmé l'hypothèse communément admise selon laquelle la persistance du calicivirus dans une population de chats pouvait résulter d'infections de longue durée. En revanche, ils ont aussi montré que ces infections de longue durée étaient relativement rares.

Il s'avère que l'évolution de la diversité virale au sein d'une population de chats résulte aussi d'épisodes de diversification virale intra-hôtes (tant au locus évoluant sous pression immunitaire qu'aux autres locus), associés à de fréquentes réinfections des mêmes chats.

Les virus de la grippe humaine A constituent finalement l'exemple pour lequel la tentative d'unification des recherches de dynamique adaptative et de dynamique épidémiologique est la plus poussée à ce jour (voir pour de plus amples détails le chapitre 11). L'analyse de l'évolution des virus de la grippe A s'est d'abord focalisée sur deux des trois protéines se combinant pour former l'enveloppe virale : la neuraminidase (NA) et l'hémagglutinine (HA). Ceci tient à l'hypothèse centrale que l'on cherchait à tester, selon laquelle l'évolution de parasites obligatoires comme les virus est largement dominée par la **course aux « armements »** entre hôtes et pathogènes. La vision de l'évolution de ces virus, donnée par la variabilité des gènes codant HA et NA, correspond en effet au patron attendu : l'arbre obtenu présente un glissement continu des formes protéiques reconnues par le système immunitaire tel qu'à chaque instant l'état dominant est remplacé par un et un seul variant (cf. Chapitre 11). Parallèlement, on note qu'environ tous les trois ans, le glissement des états reconnaissables par le système immunitaire correspond à la fois à un saut identitaire plus marqué (c'est-à-dire par une baisse plus importante de la reconnaissance du virus par les anticorps sélectionnés au cours des années précédentes) et à un accroissement marqué du nombre d'infections humaines (Bush et coll. 1999). Toutefois, est-ce que cette course aux armements suffit à expliquer les trajectoires évolutives de ces virus ? Pour y répondre, des études plus récentes ont analysé l'évolution temporelle du polymorphisme à d'autres locus que ceux participant à l'identité de l'enveloppe ; d'autres ont augmenté l'échantillonnage effectué au sein des mêmes populations humaines à une date donnée (voir pour revues complémentaires le chapitre 11, Nelson et Holmes 2007). Il s'avère alors d'une part que, pour un même type antigénique viral, des génotypes HANA co-circulent plus ou moins longtemps dans les mêmes populations humaines avant de se remplacer et d'autre part que la variation des descendants d'un de ces génotypes ne dépend pas uniquement de l'histoire immunitaire des populations humaines, mais également du génotype viral à d'autres locus que ceux codant HA ou NA. Notons à présent que le génome de ces virus de grippe est fragmenté en huit segments, équivalents fonctionnels des chromosomes des eucaryotes portant des gènes différents, et que cette segmentation favorise la réassociation des génotypes en nouvelles combinaisons multilocus, relativement aux événements de recombinaisons intra-haplotypes (du fait de la segmentation, on parle ici de combinaison au lieu de recombinaison). L'image qui émerge alors de cette analyse globale de la variabilité génétique virale, aux locus intervenant ou non dans la reconnaissance immunitaire par l'hôte, diminue

la place accordée à la course aux armements entre virus de grippe A et populations humaines, au bénéfice d'adaptation liée à la **coévolution** entre segments viraux. Ces résultats ouvrent aussi de nouvelles pistes de recherches concernant les causes de la saisonnalité des cas de grippe et l'importance de la circulation intertropicale dans la dynamique épidémiologique de ces virus en zones tempérées (Nelson et Holmes 2007, cf. Chapitre 11). Parallèlement, cet exemple illustre une fois de plus le danger qu'il y a à tenter de retracer la dynamique adaptative d'un parasite par la seule étude des gènes impliqués dans l'histoire d'interaction hôte-parasite d'intérêt.

## 5. Conclusion

La démocratisation du séquençage et le développement d'outils statistiques puissants que sont les méthodes basées sur l'analyse des signatures de sélection vont favoriser les cas de détection d'épisodes sélectifs chez les pathogènes les plus communément étudiés, comme chez ceux dont le génome de petite taille va rapidement permettre de travailler à l'échelle de leur génome entier. Ceci permettra de lever la contrainte de ne rechercher des épisodes de sélection qu'aux locus correspondant à des gènes de fonction connue et donc d'amener des informations complémentaires de celles produites en génétique inverse ou en analyse de **protéomique** sur la nature des déterminants moléculaires des interactions hôtes-parasites. Bien que ces approches n'en soient encore qu'à leur balbutiement dans le cas des pathogènes, il y a fort à parier qu'elles vont connaître une utilisation exponentielle chez les espèces parasites encore peu étudiées, comme c'est le cas aujourd'hui pour les organismes dont les génomes sont entièrement séquencés.

### À RETENIR

**Point 1 :** La sélection se définit relativement à un habitat, mais aussi relativement aux génotypes en compétition dans l'habitat considéré.

**Point 2 :** Pour être complète et effectivement informative sur l'histoire adaptative, une recherche de signature de l'action de la sélection doit aussi prendre en compte les informations apportées par l'analyse de polymorphisme neutre sur les actions conjointes des autres forces évolutives (migration, dérive, mutation, système de reproduction).

**Point 3 :** L'accessibilité de plus en plus grande de données moléculaires génomiques (sur l'ensemble du génome) et le développement des outils statistiques d'analyse permet d'appréhender l'évolution adaptative des agents infectieux à différentes échelles de temps et donc de mieux comprendre comment ils s'adaptent aux variations de leur environnement.

**Point 4 :** L'adaptation à des modifications environnementales passe par des épisodes de sélection positive qui laissent différentes empreintes détectables au niveau de la structure du polymorphisme dans et entre populations. Les approches de la génétique des populations permettent d'identifier les régions du génome qui, par l'effet de la sélection positive, ont joué un rôle dans l'évolution de ces adaptations.

**Point 5 :** L'adaptation des agents pathogènes à de nouvelles conditions environnementales est un facteur clé de leur dynamique. Une évolution adaptative peut déterminer de nouvelles flambées épidémiques, expliquer l'émergence ou la réémergence de nouveaux pathogènes ou de nouvelles souches (évolution des chimiorésistances, évolution de la virulence directement pour la qualité de trait ou au travers de caractéristiques d'exploitation de l'hôte, par exemple).

### QUESTIONS DE DISCUSSION

**Question 1 :** Que se passe-t-il lorsqu'une pression de sélection disparaît ?

**Question 2 :** Quel rôle jouent les zones de régulation de l'expression des gènes dans l'adaptation des pathogènes à de nouveaux environnements ?

➤ **LECTURES POUR ALLER PLUS LOIN**

**Sélection fréquence-dépendante : lire Garrigan et Hedrick (2003) et Charlesworth (2006).** Le présent chapitre s'est focalisé sur les signatures de la sélection positive de mutants au cours de l'évolution de parasites. Ces deux articles font une revue critique de l'identification d'autres formes de sélection déterminantes dans l'évolution des relations hôtes-parasites : les sélections fréquence-dépendantes. La revue de Garrigan et Hedrick fait, en particulier, la revue des recherches de ce type de sélection sur les gènes contrôlant le complexe majeur d'histocompatibilité (CMH, ou MHC en anglais) chez les vertébrés, et une revue du contournement possible des problèmes de puissance de tests pris isolément. Celui de Charlesworth complète cette revue de détail par les dernières données et avancées techniques.

**Impact épidémiologique des études de sélection adaptative ; outre le chapitre 11 lire aussi les deux articles de revue suivants :** Grenfell et coll. (2004) et Gupta et coll. (1998).

## Glossaire

**Adaptation:** processus d'évolution de fréquences génotypiques conduisant à l'amélioration des probabilités de survie et/ou de reproduction des individus présents dans un environnement particulier.

**Agent infectieux:** voir pathogène.

**Auto-stop génique (ou balayage sélectif):** augmentation en fréquence dans la population d'un allèle neutre de sa liaison physique avec une mutation soumise à sélection positive.

**Avantage adaptatif:** un génotype présente un avantage adaptatif dans un environnement donné lorsque sa valeur sélective est supérieure à la moyenne des autres génotypes de la même population. La valeur de cet avantage dépend à la fois des caractéristiques du génotype, de la composition génotypique de la population étudiée et des caractéristiques environnementales de cette population.

**Balayage sélectif:** voir auto-stop génique.

**Chimiorésistance:** capacité à contourner les effets délétères d'une molécule sur la survie, la reproduction ou la transmission.

**Codon:** unité du code génétique de l'ADN ou de l'ARN chromosomique, constitué de trois nucléotides. La plupart des codons codent des acides aminés.

**Course aux armements:** dans le cas particulier des systèmes hôtes-parasites, désigne tout spécifiquement la mise en place permanente par l'hôte de nouvelles défenses pour contrôler la démographie du parasite, et la sélection continue dans les populations du parasite de toute

mutation permettant de contourner les défenses de l'hôte.

**Coévolution:** le terme coévolution désigne les transformations qui se produisent au cours de l'évolution de deux espèces suite à leurs influences réciproques.

**Estimation de maximum de vraisemblance:** méthode statistique utilisée pour inférer les paramètres de la distribution de probabilité d'un échantillon donné.

**Évolution:** caractérise les changements phénotypiques ou génétiques qui opèrent au sein d'une population ou d'un groupe d'individus de la même espèce au cours des générations.

**Génotype:** le génotype correspond à l'ensemble des formes génétiques (allèles) que présente un individu à un locus donné. Chez les organismes diploïdes, ce génotype peut être constitué au plus de deux formes différentes.

**Génotyper:** génotyper, c'est caractériser le génotype d'un individu.

**Glycosylation:** la glycosylation est une réaction enzymatique consistant à lier un glucide à une chaîne peptidique ou une protéine.

**Haplotype:** le terme haplotype est une extension du terme allèle. Le terme «allèle» réfère aux états distincts observés à un locus donné et le terme «haplotypes» aux états distincts observés pour une combinaison de locus voisins.

**Locus:** localisation physique sur le brin chromosomique du marqueur génétique étudié. Ce terme peut donc renvoyer indifféremment à un gène (région codante), une séquence microsatellite (cf. Chapitre 8), voire même à un site nucléo-

tidique dans le cas d'analyse de séquence et d'estimation de rapport  $\omega = dN/dS$ .

**Mutations non-synonymes:** mutation d'une région codante qui en changeant un codon modifie la protéine codée par le gène.

**Mutations synonymes:** mutation d'une région codante qui change un codon sans modifier l'acide aminé incorporé dans la protéine.

**Modèle en allèles infinis:** modèle de mutation tel que toute nouvelle mutation crée un allèle différent de tous ceux existant ou ayant déjà existé.

**Nombre de sites polymorphes:** nombre de sites nucléotidiques présentant des états différents le long d'une séquence et au sein d'une population.

**Parasite:** organisme qui vit aux dépens d'un ou plusieurs autres organismes.

**Pathogène (synonyme, agent infectieux):** les pathogènes sont des micro-organismes qui induisent des maladies chez leurs hôtes.

**Phénotype:** le phénotype est un caractère observable à un trait donné chez un individu (par exemple, une caractéristique morphologique ou anatomique).

**Polymorphisme:** le polymorphisme désigne les variations de formes (allèles) observées entre individus au sein d'une population.

**Pression de sélection:** force dépendante de l'environnement et de la composition génotypique d'une population se traduisant par une différence de valeur sélective entre génotypes.

**Protéomique:** étude du protéome c'est-à-dire l'ensemble des protéines produites par un génome dans des conditions données, à un moment donné.

**Recombinaison:** événement aboutissant à l'échange de matériel génétique entre deux brins chromosomiques homologues. La probabilité de recombinaison définit la distance génétique entre deux locus (distance exprimée en centimorgan pour une probabilité exprimée en pourcentage). Distance génétique et distance physique (exprimée en nombre de paires de bases) ne sont pas proportionnelles au niveau d'un génome pris dans sa totalité mais tendent à l'être au niveau d'une région chromosomique de quelques centaines ou milliers de bases.

**Régions flanquantes d'un locus:** correspond aux deux régions du génome situées de part et d'autre du locus considéré.

**Sélection directionnelle:** type de sélection pour laquelle la valeur sélective d'un génotype est

**Sélection fréquence-dépendante:** type de sélection pour lequel la valeur sélective de chaque allèle ou génotype est d'autant plus forte que l'allèle ou le génotype est rare dans la population.

**Sélection positive:** type de sélection pour lequel la valeur sélective d'un génotype surpasse celle de tous les autres.

**Sélection purifiante:** type de sélection pour lequel la valeur sélective de tout nouvel allèle est inférieure à celle de l'allèle fixé dans la population.

**Sélection:** voir pression de sélection.

**Séquence:** portion du génome caractérisée par l'enchaînement de nucléotides.

**Sérotypes:** catégories dans lesquelles certains virus ou bactéries sont classés selon leur réaction en présence d'un sérum (partie liquidienne du sang) qui contient des anticorps spécifiques contre les bactéries ou les virus en question.

**Site nucléotidique:** portion d'ADN génomique correspondant à un nucléotide donné.

**Signal phylogénétique:** relation de descendance entre allèles ou entre copies de gènes échantillonnés dans différentes espèces. Dans ce dernier cas, par extension, le signal phylogénétique peut être interprété comme les relations de descendance entre les espèces échantillonnées.

**Spéciation:** la spéciation correspond au processus évolutif par lequel de nouvelles espèces sont créées.

**Spectre de fréquence:** correspond à la distribution du décompte du nombre d'allèles qui présentent une fréquence  $x_i = i / n$  for  $i = 1, 2, 3, \dots, n-1$  dans un échantillon de  $n$  individus haploïdes.

**Superdominance:** type de sélection pour lequel les génotypes hétérozygotes sont associés à de plus fortes valeurs sélectives que les homozygotes.

**Système de reproduction:** le système de reproduction décrit les modalités de rencontres des gamètes au sein d'une espèce.

**Traitement curatif:** le traitement curatif désigne le processus que l'on utilise pour éliminer une infection une fois que celle-ci s'est développée chez un individu.

**Traitement prophylactique (ou prophylaxie):** une prophylaxie désigne le processus actif ou passif ayant pour but de prévenir l'apparition ou la propagation d'une maladie (infectieuse ou non).

**Valeur sélective (fitness):** la valeur sélective d'un allèle (ou d'un génotype) désigne le nombre de descendants de cet allèle (ou génotype) relativement aux autres allèles (ou génotypes) présents dans la population. Le terme anglais «fitness» est parfois utilisé en français en lieu et place de «valeur sélective».

## Références

Anderson, T.J. et Roper, C. (2005). The origins and spread of antimalarial drug resistance: lessons for policy makers. *Acta Tropica* 94: 269-280.

Anderson, T.J.C., Nair, S., Sudimack, D., Williams, J.T., Mayxay, M., Newton, P.N., Guthmann, J.P., Smithuis, F.M., Hien, T.T., van den Broek, I.V.F., White, N.J. et Nosten, F. (2005). Geographical distribution of selected and putatively neutral SNPs in Southeast Asian malaria parasites. *Molecular Biology and Evolution* 22: 2362-2374.

Andolfatto, P. (2001). Adaptive hitchhiking effects on genome variability. *Current Opinion in Genetics and Development* 11: 635-641.

Anisimova, M., Bielawski, J.P. et Yang, Z.H. (2001). Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Molecular Biology and Evolution* 18: 1585-1592.

Anisimova, M., Bielawski, J.P. et Yang, Z.H. (2002). Accuracy and power of Bayes prediction of amino acid sites

- under positive selection. *Molecular Biology and Evolution* 19: 950-958.
- Anisimova, M., Nielsen, R. et Yang, Z.H. (2003). Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* 164: 1229-1236.
- Anisimova, M., Bielawski, J., Dunn, K. et Yang, Z.H. (2007). Phylogenomic analysis of natural selection pressure in *Streptococcus* genomes. *BMC Evolutionary Biology* 7: 154.
- Anisimova, M. et Yang, Z.H. (2007). Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. *Molecular Biology and Evolution* 24: 1219-1228.
- Awadalla, P. et Charlesworth, D. (1999). Recombination and selection at *Brassica* self-incompatibility loci. *Genetics* 152: 413-425.
- Bancroft, E. (2007). Antimicrobial resistance: it's not just for hospitals. *JAMA* 298: 1803-1804.
- Basco, L.K., Tahar, R. et Ringwald, P. (1998). Molecular basis of in vivo resistance to sulfadoxine-pyrimethamine in African adult patients infected with *Plasmodium falciparum* malaria parasites. *Antimicrobial Agents and Chemotherapy* 42: 1811-1814.
- Beaumont, M.A. et Balding, D.J. (2004). Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* 13: 969-980.
- Bennett, S.N., Holmes, E.C., Chirivella, M., Rodriguez, D.M., Beltran, M., Vorndam, V., Gubler, D.J. et McMillan, W.O. (2003). Selection-driven evolution of emergent dengue virus. *Molecular Biology and Evolution* 20: 1650-1658.
- Braverman, J.M., Hudson, R.R., Kaplan, N.L., Langley, C.H. et Stephan, W. (1995). The Hitchhiking Effect on the Site Frequency-Spectrum of DNA Polymorphisms. *Genetics* 140: 783-796.
- Bush, R.M., Fitch, W.M., Bender, C.A. et Cox, N.J. (1999). Positive selection on the H3 hemagglutinin gene of human influenza virus A. *Molecular Biology and Evolution* 16: 1457-1465.
- Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics* 2: 379-384.
- Choisy, M., Woelk, C.H., Guégan, J.F. et Robertson, D.L. (2004). Comparative study of adaptive molecular evolution in different human immunodeficiency virus groups and subtypes. *Journal of Virology* 78: 1962-1970.
- Coyne, K.P., Gaskell, R.M., Dawson, S., Porter, C.J. et Radford, A.D. (2007). Evolutionary mechanisms of persistence and diversification of a calicivirus within endemically infected natural host populations. *Journal of Virology* 81: 1961-1971.
- de Bakker, P.I., Yelensky, R., Pe'er, I., Gabriel, S.B., Daly, M.J. et Altshuler, D. (2005). Efficiency and power in genetic association studies. *Nature Genetics* 37, 1217-1223.
- Ewens, W.J. (1972). Sampling Theory of Selectively Neutral Alleles. *Theoretical Population Biology* 3: 87-112.
- Fidock, D.A., Nomura, T., Talley, A.K., Cooper, R.A., Dzekunov, S.M., Ferdig, M.T., Ursos, L.M.B., Sidhu, A.B.S., Naude, B., Deitsch, K.W., Su, X.Z., Wootton, J.C., Roepe, P.D. et Wellems, T.E. (2000). Mutations in the *P. falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Molecular Cell* 6: 861-871.
- Foster, J.E., Bennett, S.N., Vaughan, H., Vorndam, V., McMillan, W.O. et Carrington, C.V.F. (2003). Molecular evolution and phylogeny of dengue type 4 virus in the Caribbean. *Virology* 306: 126-134.

- Fu, Y.X. et Li, W.H. (1993). Statistical Tests of Neutrality of Mutations. *Genetics* 133: 693-709.
- Garrigan, D. et Hedrick, P.W. (2003). Detecting adaptive molecular polymorphism, lessons from the MHC. *American Journal of Human Genetics* 73: 375-375.
- Grenfell, B.T., Pybus, O.G., Gog, J.R., Wood, J.L.N., Daly, J.M., Mumford, J.A. et Holmes, E.C. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303: 327-332.
- Gupta, S., Ferguson, N. et Anderson, R. (1998). Chaos, persistence, and evolution of strain structure in antigenically diverse infectious agents. *Science* 280: 912-915.
- Hurwitz, E.S., Johnson, D. et Campbell, C.C. (1981). Resistance of *Plasmodium falciparum* Malaria to Sulfadoxine-Pyrimethamine (Fansidar) in a Refugee Camp in Thailand. *Lancet* 1: 1068-1070.
- Kim, Y. et Stephan, W. (2000). Joint effects of genetic hitchhiking and background selection on neutral variation. *Genetics* 155: 1415-1427.
- Lemey, P., Pond, S.L.K., Drummond, A.J., Pybus, O.G., Shapiro, B., Barroso, H., Taveira, N. et Rambaut, A. (2007). Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. *PLoS Computational Biology* 3: 282-292.
- McCollum, A.M., Mueller, K., Villegas, L., Udhayakumar, V. et Escalante, A.A. (2007). Common origin and fixation of *Plasmodium falciparum* dhfr and dhps mutations associated with sulfadoxine-pyrimethamine resistance in a low-transmission area in South America. *Antimicrobial Agents and Chemotherapy* 51: 2085-2091.
- McDonald, J.H. et Kreitman, M. (1991). Adaptive Protein Evolution at the *Adh* Locus in *Drosophila*. *Nature* 351: 652-654.
- Milgroom, M.G. et Cortesi, P. (1999). Analysis of population structure of the chestnut blight fungus based on vegetative incompatibility genotypes. *Proceedings of the National Academy of Sciences of the United States of America* 96: 10518-10523.
- Nair, S., Williams, J.T., Brockman, A., Phipun, L., Mayxay, M., Newton, P.N., Guthmann, J.P., Smithuis, F.M., Hien, T.T., White, N.J., Nosten, F. et Anderson, T.J.C. (2003). A selective sweep driven by pyrimethamine treatment in southeast Asian malaria parasites. *Molecular Biology and Evolution* 20: 1526-1536.
- Nelson, M.I. et Holmes, E.C. (2007). The evolution of epidemic influenza. *Nature Reviews Genetics* 8: 196-205.
- Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics* 39: 197-218.
- Nielsen, R., Williamson, S., Kim, Y., Hubisz, M.J., Clark, A.G. et Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome Research* 15: 1566-1575.
- Plowe, C.V., Cortese, J.F., Djimde, A., Nwanyanwu, O.C., Watkins, W.M., Winstanley, P.A., Estrada-Franco, J.G., Mollinedo, R.E., Avila, J.C., Cespedes, J.L., Carter, D. et Doumbo, O.K. (1997). Mutations in *Plasmodium falciparum* dihydrofolate reductase and dihydropteroate synthase and epidemiologic patterns of pyrimethamine-sulfadoxine use and resistance. *Journal of Infectious Diseases* 176: 1590-1596.
- Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X.H., Byrne, E.H., McCarroll, S.A., Gaudet, R., Schaffner, S.F., Lander, E.S. et The International HapMap Consortium (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature* 449: 913-919.
- Tajima, F. (1989). Statistical-Method for Testing the Neutral Mutation Hypothe-

- sis by DNA Polymorphism. *Genetics* 123: 585-595.
- Trape, J.F. (2001). The public health impact of chloroquine resistance in Africa. *American Journal of Tropical Medicine and Hygiene* 64: 12-17.
- Twiddy, S.S., Farrar, J.J., Chau, N.V., Wills, B., Gould, E.A., Gritsun, T., Lloyd, G. et Holmes, E.C. (2002). Phylogenetic relationships and differential selection pressures among genotypes of dengue-2 virus. *Virology* 298: 63-72.
- Uyenoyama, M.K. (2005). Evolution under tight linkage to mating type. *New Phytologist* 165: 63-70.
- Volkman, S.K., Sabeti, P.C., DeCaprio, D., Neafsey, D.E., Schaffner, S.F., Milner, D.A., Daily, J.P., Sarr, O., Ndiaye, D., Ndir, O., Mboup, S., Duraisingh, M.T., Lukens, A., Derr, A., Stange-Thomann, N., Waggoner, S., Onofrio, R., Ziaugra, L., Mauceli, E., Gnerre, S., Jaffe, D.B., Zainoun, J., Wiegand, R.C., Birren, B.W., Hartl, D.L., Galagan, J.E., Lander, E.S. et Wirth, D.F. (2007). A genome-wide map of diversity in *Plasmodium falciparum*. *Nature Genetics* 39: 113-119.
- Watterson, G.A. (1978). Homozygosity Test of Neutrality. *Genetics* 88: 405-417.
- Wootton, J.C., Feng, X., Ferdig, M.T., Cooper, R.A., Mu, J., Baruch, D.I., Magill, A.J. et Su, X.Z. (2002). Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature* 418: 320-323.
- Yang, Z.H. et Bielawski, J.P. (2000). Statistical methods for detecting molecular adaptation. *Trends in Ecology and Evolution* 15: 496-503.
- Yang, Z.H., Nielsen, R., Goldman, N. et Pedersen, A.M.K. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155: 431-449.
- Yang, Z.H., Wong, W.S.W. et Nielsen, R. (2005). Bayes empirical inference of amino acid sites under positive selection. *Molecular Biology and Evolution* 22: 1107-1118.
- Zhang, J.Z., Nielsen, R. et Yang, Z.H. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution* 22: 2472-2479.

Prugnotte Franck, Chevillon Christine.

Evolution adaptative des pathogènes,  
identification des mécanismes et  
conséquences épidémiologiques.

In : Guégan Jean-François (ed.), Choisy Marc  
(ed.). Introduction à l'épidémiologie  
intégrative des maladies infectieuses et  
parasitaires. Bruxelles : De Boeck, 2009, p.  
311-345.

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