Polymorphisme génétique et sélection

Auteur :
VEUILLE Michel, Directeur d'Etudes, Ecole pratique des hautes études, Paris

24-03-2018

Le polymorphisme a suscité des controverses sur son rôle dans l'évolution. Mais s'il suit essentiellement une évolution neutre, il sert de référence, par contraste, à l'étude de la sélection naturelle. Il est aussi utilisé par les écologues en biologie de la conservation pour reconstituer l'histoire passée des espèces.

1. Mutations, dérive aléatoire et évolution neutre

Le polymorphisme est constitué par les mutations qui échappent aux systèmes de réparation de l'ADN au fil des divisions cellulaires. Leur taux d'apparition est donc une variable biologique. Chez l'homme et le chimpanzé, il est de $\mu \approx 10^{-8}$ mutations par nucléotide et par génération. La quantité considérable de spermatozoïdes produits par les mammifères mâles entraine qu'il y a beaucoup plus de divisions cellulaires en lignée germinale mâle qu'en lignée germinale femelle : 380 contre 23 à 30 ans (soit 16 fois plus), et plus encore quand l'homme vieillit (840 contre 23 à 50 ans, soit 36 fois plus). Cela entraîne que dans ces espèces les mutations sont majoritairement produites en lignée mâle et dépendent de l'âge du père. Chaque naissance produit environ 100 mutations nouvelles par génome, mais du fait que seule une petite partie du génome est codante, 99% d'entre elles sont sans effet sur la survie ou la fertilité. Elles sont dites neutres. Un allèle nouveau peut être **neutre**, **délétère** ou **avantageux**. Les mutations neutres sont les plus étudiées, car elles permettent d'écrire des modèles prédictifs pour explorer l'histoire des populations. Leur distribution sert aussi d'hypothèse nulle pour interpréter, par comparaison, celle des mutations délétères ou avantageuses.

Nous pourrions penser que dans un **génome ne comprenant que des allèles neutres**, la dérive des fréquences alléliques se compenserait d'une fluctuation sur l'autre et que la diversité allélique H resterait stable sur le long terme. Mais cette impression est fausse. **Petit à petit la diversité s'érode**. Le phénomène est très analogue à la perte de diversité des noms de famille, un phénomène certes lent, mais sensible dans les isolats humains tels les villages reculés. Lorsqu'une famille n'a pas de garçon, elle ne transmet pas son patronyme. Le même patronyme peut être transmis par des familles apparentées, mais plus la population est petite, plus grande est la probabilité que des noms soient perdus. Ceci n'est évidemment pas dû à quelque propriété biologique du chromosome Y, qui accompagne les naissances mâles. Le hasard suffit à l'expliquer. Cette propriété traduit le fait que la constitution d'une génération fille à partir d'une population parentale suit le principe d'un tirage avec remise.

De même que les chromosomes Y, certains gènes de la génération parentale ne sont pas tirés, et ne se retrouvent pas dans la population fille. Si les gènes de la descendance sont tirés au hasard dans une population de taille constante, la probabilité que des gènes ne soient pas tirés est donné par une loi de Poisson de paramètre 1 comme $q_{(0)} = e^{-1} = 0{,}367$. Ces gènes non tirés (plus d'un tiers) disparaissent sans descendance. Leur absence est compensée par les gènes parentaux qui, par hasard, ont laissé plus de descendants. S'il n'en allait pas ainsi, les lignées ancestrales resteraient parallèles sans jamais se rencontrer. Le regroupement des lignées ancestrales quand on remonte vers le passé n'est pas différent d'une perte de diversité quand on descend vers le présent, et n'est pas différent non plus de ce qu'on appelle la **consanguinité**.

La mesure de la diversité de la formule (1) a une propriété utile : elle dépend de la taille de l'échantillon sur laquelle elle est estimée. Quand une population fille "t+1" est échantillonnée en tirant n gènes d'une génération parentale "t" la génération fille montre une perte de variation égale à $1/n$, suivant la formule $E(H_s) = H_t\,(1 - 1/n)$. Ceci est vrai même si la population fille est plus grande que la population mère, puisqu'il s'agit d'un tirage avec remise, mais plus la population est grande, moins la diversité est érodée. Il suffit que la population soit de **taille finie**, ce que sont toutes les populations réelles. Par convention, les généticiens désignent cette perte de variation par $1/N_e$, où N_e est appelé **effectif efficace** des chromosomes [1]. Ainsi d'une génération 1 à une génération 2 :

$$E(H_2) = H_1\,(1 - 1/N_e) \quad (3)$$

L'effectif efficace est presque toujours beaucoup plus petit que l'effectif réel des chromosomes, pour des raisons qui seront discutées plus loin. Par exemple, on estime que dans le passé de la lignée humaine, l'effectif efficace était de l'ordre de grandeur de 10.000 chromosomes. S'il n'y avait pas de mutations, on démontre que la population deviendrait monomorphe au bout d'un temps T, d'espérance :

$$E(T) = 2\,N_e \quad (4)$$

Il y a deux conséquences à cela : d'abord, le polymorphisme d'une espèce est toujours "récent" à l'échelle de la durée d'une espèce, puisqu'il dépend de mutations qui ont restauré le polymorphisme malgré l'érosion de diversité qui accompagne la dérive des fréquences alléliques. Ensuite, le niveau de polymorphisme est un compromis entre deux mécanismes opposés, créant l' **équilibre neutre mutation-dérive**.

La disparition du polymorphisme au fil du temps peut s'exprimer en sens inverse : lorsqu'on remonte le temps, il existe toujours un dernier ancêtre commun entre deux gènes d'un même locus. C'est ce que John Kingman a appelé le processus de **coalescence**. L'ancêtre n'est pas le même pour des locus différents, car la sexualité multiplie le nombre des ancêtres, donc aussi les ancêtres communs des gènes. Si la probabilité d'avoir un ancêtre commun à la génération précédente $q = 1/N_e$, reste constante au fil du temps, la distribution des ancêtres suit une loi exponentielle $t = q.e^{-qt}$. L'espérance de l'âge de ces ancêtres est égal à N_e. Deux gènes seront génétiquement similaires si aucune mutation n'est intervenue depuis ce temps. Mais il suffit qu'une mutation soit intervenue dans l'une des lignées menant de l'ancêtre à chacun des deux gènes pour que les deux gènes soient allèles. On en déduit que le nombre de différences nucléotidiques entre ces deux gènes est égal à $\theta = N_e \times 2\mu$, où μ est le **taux de mutation neutre**. Cette valeur θ, définie comme $\theta = 2N_e\mu$, est un paramètre fondamental de la génétique des populations.

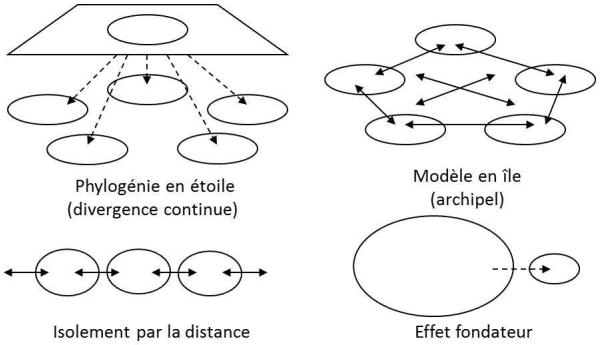

Figure 1. Quelques modèles de structuration des populations. Les flèches continues représentent des taux de migration entre les populations, figurées par des ellipses. Les flèches interrompues représentent la différenciation entre populations due à la dérive aléatoire. Dans la phylogénie en étoile et l'effet fondateur, la divergence est continue au cours du temps. Dans les autres modèles, il y a équilibre entre la dérive aléatoire au sein des populations et l'homogénéisation des populations par migration. Les profils observés de variation génétique et les FST entre couples de populations permettent de savoir si l'histoire d'une espèce donnée se rapproche plus ou moins de l'un de ces scénarios.

L'évolution neutre des populations naturelles est très importante en biologie de la conservation, car elle permet de reconstruire l'histoire des espèces. Depuis longtemps les généticiens savent que la **dérive génétique aléatoire** leur permet d'inférer les modèles de différenciation des populations et de structuration des espèces dans l'espace (figure1). Pendant la deuxième moitié du 20ème siècle, l'indicateur le plus utilisé pour étudier la structuration d'une population en sous-populations a été le $\boldsymbol{F_{ST}}$, de formule :

$$F_{ST} = 1\text{-}H_S/H_T \quad (5)$$

où H_S est la moyenne des diversités des sous-populations et H_T la diversité de la population totale [2].

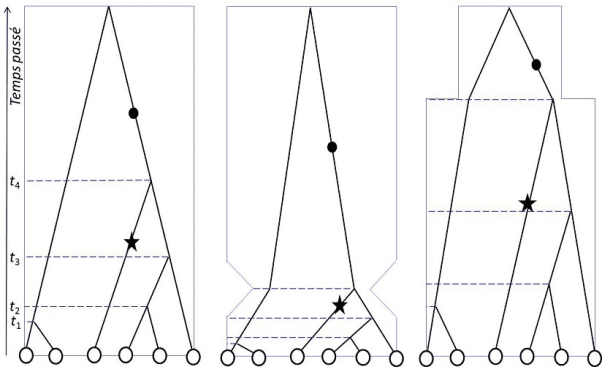

Figure 2. Coalescence et changements démographiques. La généalogie d'un échantillon de 6 gènes (cercles blancs) d'un même locus est examinée dans trois populations présentant des histoires démographiques différentes où l'encadrement représente la taille de la population : effectif constant (à gauche), en expansion récente après un goulot d'étranglement (au centre) et en expansion ancienne (à droite). Le temps (en générations) remonte vers le passé dans le haut du diagramme. Les mutations sont figurées par des cercles noirs (mutations des branches "internes", communes à plusieurs gènes de l'échantillon) ou par des étoiles noires (mutations des branches "externes", uniques dans l'échantillon car menant à un seul gène de l'échantillon). On voit que la distribution de l'âge des ancêtres communs de l'échantillon est très différente selon les conditions. Le rapport de longueur entre les branches externes (terminales) et les branches internes de la généalogie est différent. Cela entraînera, pour un taux de mutation constant en fonction du temps, une proportion relative différente des mutations externes et des mutations internes. Ce rapport des mutations rares/fréquentes dans l'échantillon est l'un des indicateurs qui permettent de reconstruire l'histoire de la population. D'autres jeux d'indicateurs du polymorphisme (dont H et π, vus dans le texte) ont aussi cette propriété.

Au XXIème siècle, âge de l'analyse numérique des génomes, la **théorie de la coalescence** [3], imaginée indépendamment par Kingman, Hudson et Tajima en 1982-83, permet, outre l'étude de la structuration, de savoir si les populations sont restées

stables, ou ont connu des changements démographiques (figure 2).

2. Modèle neutre et gestion de la biodiversité

Les figures 1 et 2 illustrent combien les profils de variation génétique sont affectés par l'histoire des populations : structuration spatiale, colonisation, migration, changement d'effectif démographique sont autant d'événements qui impriment une signature spécifique dans le polymorphisme moléculaire des espèces, et permettent aux écologues d'en retracer l'histoire. Au cours de l'ère quaternaire- la période géologique actuelle -, les climats de la planète ont changé de façon cyclique, entraînant une modification périodique du trait des côtes, un déplacement nord-sud des associations biologiques et des glaciers, déclenchant des périodes de climat humide ou sec sous toutes les latitudes. Les déplacements, diminutions, augmentations, invasions de populations qui en ont résulté, indices des **réponses des espèces aux changements de leur milieu**, sont systématiquement répertoriés par les biologistes des populations avant d'entreprendre toute initiative de **gestion des populations naturelles**. La plus grande partie des applications de la génétique des populations s'exerce de nos jours en **biologie de la conservation**.

3. Les mutations délétères

Du fait que les gènes codent pour des protéines, la plupart des mutations intervenant dans les régions codantes modifient la séquence des protéines (environ 3/4 des mutations, une proportion qui varie selon la composition de la séquence). Dans la lignée humaine environ 40% de ces changements sont délétères, c'est-à-dire qu'ils manquent quand on fait le bilan de l'évolution du génome de cette espèce depuis sa séparation d'avec la lignée du chimpanzé. Si une mutation était neutre, elle aurait une chance $1/N_e$ de remplacer un jour les autres gènes présents à ce locus (dans une population d'effectif efficace N_e, les autres gènes pris ensemble sont en proportion de $1-1/N_e$, et chacun a également une chance $1/N_e$ de remplacer tous les autres). Mais une mutation peut être délétère et affecter la santé ou la fertilité des individus qui la portent. Sa fréquence peut fluctuer pendant quelques générations par dérive aléatoire avant de disparaître par sélection (quarante générations en moyenne chez la drosophile). Tous les membres d'une espèce sont porteurs de mutations délétères. Vous et moi le sommes. Elles sont presque toujours à l'état hétérozygote, car si une mutation a une fréquence, par exemple, de 1/1000, elle aura mille fois moins de représentants à l'état homozygote qu'à l'état hétérozygote. C'est le léger désavantage des hétérozygotes qui éliminent la mutation plutôt que le désavantage souvent beaucoup plus important de l'homozygote. Les effets des mutations délétères sur plusieurs locus étant cumulatifs, le **fardeau mutationnel** devient une variable quantitative comme une autre dont les effets additifs peuvent être indétectables, et néanmoins efficaces sur le long terme pour purger le génome en permanence. Cela explique que les protéines restent fonctionnelles et que les mutations délétères restent de faible fréquence. Elles sont sans doute l'un des facteurs qui expliquent le maintien de la recombinaison génétique. Celle-ci permet à la fois de regrouper les mutations délétères entre elles pour les éliminer, et de limiter les conséquences de leur élimination sur les régions adjacentes des chromosomes.

4. Les mutations avantageuses

Que sont les 60 % de mutations affectant les protéines sans effet délétère ? Comme les mutations affectant d'autres régions du génome, elles peuvent être "neutres", c'est-à-dire sans effet sur la santé ou la fécondité dans tel environnement et dans tel système de communication d'une espèce. Leur fréquence fluctue aléatoirement dans les populations naturelles. Mais si les conditions changent, elles peuvent s'avérer avantageuses. Elles relèvent alors de la **sélection naturelle** et de la **sélection sexuelle** imaginées par Darwin, mais aussi de la **sélection** au sens premier du mot, c'est-à-dire celle opérée par l'homme sur ses espèces domestiques. Il existe deux types de polymorphisme sélectionné : le polymorphisme transitoire et le polymorphisme équilibré.

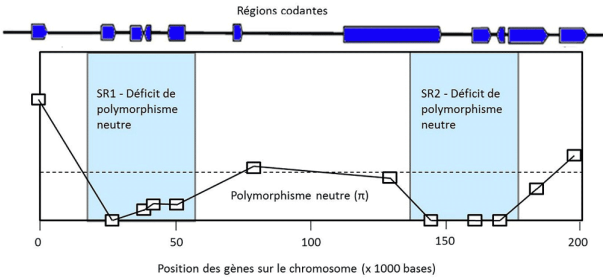

Figure 3. Balayage sélectif. La recombinaison permet de découpler l'évolution des régions adjacentes du génome. S'il n'y avait pas de sélection, le polymorphisme neutre atteindrait des valeurs comparables le long du chromosome par un processus d'équilibre très lent. Lorsqu'une mutation est avantagée dans une région et se fixe à la fréquence de 1, le processus est très rapide, et elle balaie le polymorphisme neutre de cette région, mais non celui des régions adjacentes. Le contraste du niveau de polymorphisme neutre dans les régions balayées et dans les régions neutres permet d'affirmer que c'est bien la sélection qui a agi dans les premières, et exclut le raisonnement circulaire qui admettrait que "ce qui est adapté, c'est ce que nous voyons". Cet exemple montre deux zones contiguës de balayage sélectif sur le chromosome X de Drosophila simulans. Elles permettent d'identifier deux complexes de gènes qui agissent simultanément pour modifier à leur profit les proportions mendéliennes dans la descendance des drosophiles(voir réf [6])(gènes dits "égoïstes"). Dans ces deux zones (SR1 et SR2), la sélection a fait disparaître le polymorphisme neutre.

Le **polymorphisme transitoire** correspond au cas d'une mutation avantageuse qui se "fixe" progressivement en éliminant les allèles alternatifs, ce qui peut la mener à la fréquence de 1. C'est le cas, par exemple, des gènes de résistance aux insecticides chez les moustiques, de résistance aux antibiotiques chez les bactéries et de résistance aux médicaments antipaludéens chez le parasite conférant le paludisme : ces mutations n'auraient sans doute pas eu d'avantage dans les conditions naturelles, mais dans le contexte environnemental imposé par la médecine, ces allèles augmentent de fréquence. C'est aussi le cas des trois allèles de régulation de l'expression de la lactase, enzyme qui permet aux humains de digérer le sucre du lait (lactose) non seulement à l'état de nouveau-né, comme chez les autres mammifères, mais aussi chez l'adulte. Ces mutations sont devenues bénéfiques chez des populations pratiquant l'élevage, tandis que nos ancêtres chasseurs-cueilleurs n'avaient l'occasion que de digérer le sucre des fruits (saccharose) à l'âge adulte. Dans tous ces cas de polymorphisme transitoire, le locus sur lequel porte la sélection se "trahit" par une signature dans le génome : l'expansion rapide de sa fréquence fait disparaître la variation neutre adjacente sur le chromosome. C'est un cas de **balayage sélectif**, qui permet d'affirmer que la fixation d'un allèle n'est pas due à la dérive aléatoire, mais à la sélection (figure 3, [4]).

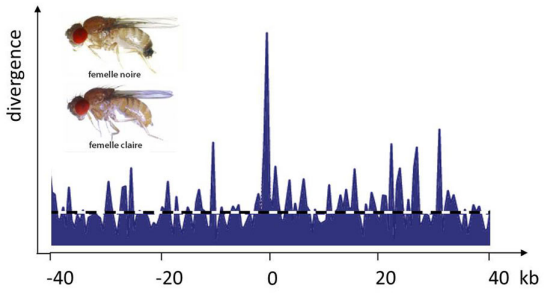

Figure 4. Polymorphisme équilibré. Lorsque la sélection naturelle maintient la coexistence de deux allèles, leurs séquences divergent de plus en plus, au point d'accumuler entre eux plus de mutations que ne le prévoit le modèle neutre pour le reste du génome. C'est le cas de deux allèles du gène tan de la drosophile (coordonnée 0), dont les deux allèles maintiennent depuis environ trois millions années la coexistence de femelles à abdomen clair ou noir. Ces motifs de coloration interviennent dans la communication entre mâles et femelles lors de l'accouplement, mais aucun ne peut éliminer l'autre, probablement parce que leur avantage sélectif diminue quand ils deviennent trop fréquents. Trait interrompu : valeur attendue de la divergence entre chromosomes ; en bleu : valeur effectivement observée (voir réf [7]).

Le **polymorphisme équilibré** correspond à des situations où deux allèles coexistent parce que chacun est avantagé dans certaines conditions, mais où aucun ne peut l'emporter sur l'autre dans toutes les circonstances du temps ou de l'espace. Un exemple en est donné par les cas où l'avantage sélectif d'un génotype croit en raison inverse de sa fréquence. On parle alors de **sélection fréquence-dépendante**. De telles situations de polymorphisme équilibré sont fréquentes dans les cas de sélection sexuelle (figure 4, [5]).

5. Le polymorphisme est-il utile ?

Dans les années 1930 à 1960, les généticiens des populations naturelles découvrirent un nombre croissant de polymorphismes dans la nature. Ils voulurent en évaluer l'étendue et en découvrir l'utilité éventuelle du point de vue de l'évolution. Des débats opposèrent des chercheurs estimant que la diversité génétique conférait un avantage en elle-même et que la sélection la maintenait à de hauts niveaux, à des chercheurs estimant que la sélection menait à un phénotype sauvage assez homogène, les variations restantes étant plutôt délétères. Aucun n'avait raison. Le français Gustave Malécot avait démontré dès les années 1950 que le polymorphisme neutre était une conséquence des lois de Mendel dans une population de taille finie [6]. C'est finalement la découverte en 1966 des niveaux extrêmement élevés du polymorphisme moléculaire, lesquels ne pouvaient s'expliquer par la seule sélection naturelle [7], qui permit aux japonais Kimura et Ohta de faire valoir la **théorie neutraliste** [8]. On réalisa que l'alternative à la théorie darwinienne de la sélection naturelle n'était pas la fixité des espèces (comme le pensaient par exemple les opposants à Darwin) mais un changement génétique continu prédit par le modèle neutre, analogue à la marche aléatoire d'un phénomène de diffusion en physique. Cette vision remporta définitivement l'adhésion au cours des années 1980. Cependant, la valeur très faible de l'effectif efficace mesurée dans toutes les espèces, comparée à l'effectif de la population reproductive, indique que des forces érodent la diversité génétique bien plus que ne le prévoient les modèles neutralistes. Cette érosion résulte pour une part, encore mal estimée, de la sélection naturelle qui élimine les mutations délétères et fixe les variations avantageuses, et de ce fait augmente les effets de dérive sur la variation neutre. Bien qu'extrêmement importants pour l'avenir des espèces, les polymorphismes sélectionnés ne constituent certes qu'une fraction minime des cas de polymorphisme.

Le polymorphisme moléculaire neutre procure la théorie de base, le modèle de référence, à partir duquel s'étudient la sélection et l'histoire des populations. Le paradoxe est que, désormais, on recherche dans le génome les signatures moléculaires de la sélection naturelle en se servant de la théorie neutraliste.

L'existence des forces sélectives qui conservent le système de recombinaison, les lois de Mendel [9], et le brassage génétique de

la sexualité est un argument pour considérer que le polymorphisme, qu'elles entretiennent ainsi, présente un avantage à court terme dans les populations naturelles.

Références et notes

[1] Jusqu'en 2000 environ, l'effectif efficace était exprimé en individus et non en chromosomes, l'effectif efficace des chromosomes était donc de $2N_e$ pour les autosomes, de $1{,}5N_e$ pour les chromosomes X, et de $0{,}5N_e$ pour les chromosomes Y et pour les mitochondries, sous réserve que le nombre de mâles et de femelles lors de la reproduction soit le même. Ces formules peuvent se rencontrer dans des manuels.

[2] Cette formule, ici très générale, prend plusieurs formes et dénominations selon le modèle génétique utilisé : F_{ST} de Wright (pour deux allèles), G_{ST} de Nei (sa généralisation, dont la formule 5 ci-dessus est une variante), Φ_{ST}, ϱ_{ST}, etc… Elle peut être remplacée par des statistiques aux propriétés similaires : D_{XY}, AMOVA. Cette redondance montre surtout le succès des " *F-statitics"* en écologie. Du fait de la dépendance de l'estimation sur la taille de l'échantillon, l'utilisation d'estimateurs sans biais doit intégrer en outre les particularités du plan d'observation. *Cf* : Weir B.S. & Cockerham C.C. (1984) *Estimating F-statistics for the analysis of population structure.* Evolution 38: 1358-1370

[3] Kingman J.F.C. (1982) *On the genealogy of large populations.* Journal of Applied Probability 19A:27–43 ; Hudson R.R. (1983) *Properties of a neutral allele model with intragenic recombination*. Theoretical Population Biology 23:183–201 ; Tajima F. (1983) *Evolutionary relationship of DNA sequences in finite populations*. Genetics 105:437–460.

[4] Derome N., K. Métayer, C. Montchamp-Moreau & M. Veuille (2004) Signature of selective sweep associated with the evolution of *sex-ratio* drive in *Drosophila simulans*. *Genetics* 166: 1357-1366 ; Derome N., E. Baudry, D. Ogereau, M. Veuille & C. Montchamp-Moreau (2008) Selective sweeps reveal a two-locus model for *sex-ratio* meiotic drive in *Drosophila simulans. Molecular Biology and Evolution*, 25: 409-416.

[5] Yassin A., Bastide H., Chung H., Veuille M., David J.R. & Pool J.E. (2016) *Ancient Balancing Selection at tan Underlies Female Colour Dimorphism in Drosophila erecta.* Nature Communications DOI: 10.1038/ncomms10400.

[6] Malécot G. (1948) *Les mathématiques de l'hérédité*. Masson et Cie ; Nagylaki T. (1989) *Gustave Malécot and the transition from classical to modern population genetics*. Genetics 122, 253–268.

[7] Lewontin R.C. & Hubby J.L. (1966) *Molecular Approach to the Study of Genic Heterozygosity in Natural Populations*. II. *Amount of Variation and Degree of Heterozygosity in Natural Populations of Drosophila pseudoobscura*, Genetics 54: 595-609 ; Lewontin R.C. (1974) *The Genetic Basis of Evolutionary Change*. Columbia Univ. Press, New York.

[8] Kimura M. (1969) *The Rate of Molecular Evolution Considered from the Standpoint of Population Genetics*. Proceedings of the National Academy of Sciences, 63, 1181-1188.

[9] http://uel.unisciel.fr/biologie/analgen/analgen_ch01/co/apprendre_ch1_01_01.html

---

L'Encyclopédie de l'environnement est publiée par l'Université Grenoble Alpes - www.univ-grenoble-alpes.fr

Pour citer cet article: **Auteur :** VEUILLE Michel (2018), Polymorphisme génétique et sélection, Encyclopédie de l'Environnement, [en ligne ISSN 2555-0950] url : http://www.encyclopedie-environnement.org/?p=1936

Les articles de l'Encyclopédie de l'environnement sont mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.