

CHOISIR OU ÊTRE CHOISI

APPROCHES CRITIQUES DE LA SÉLECTION



DIRECTION
Julien Gargani
Annick Jacq



14

CHOISIR OU ÊTRE CHOISI

Approches critiques de la sélection

DIRECTION

Julien Gargani

Annick Jacq

COLLECTION « ACTES »

Comité éditorial

Elsa Bansard (Université Paris-Saclay, MSH Paris-Saclay)
Marianne Blidon (Université Paris 1 Panthéon-Sorbonne, IDUP)
Stefano Bosi (Université d'Évry, CEPS)
Maryse Bresson (UVSQ, Printemps)
Sophie Chiari (IHRIM, MSH Clermont-Ferrand)
Claude Didry (CNRS, Centre Maurice Halbwachs)
Pierre Guibentif (Iscte Institut universitaire de Lisbonne, Dinâmia'CET_Iscte)
Christian Hottin (ministère de la Culture, direction générale des Patrimoines
et de l'Architecture, UMR Héritages, CTHS)
Flavie Lavallée (Université Paris-Saclay, MSH Paris-Saclay)
Sébastien Oliveau (Université Paris-Saclay, MSH Paris-Saclay)
Delphine Placidi-Frot (Université Paris-Saclay, IEDP, associée au Printemps)
André Torre (INRAE, AgroParisTech)

* * *



MSH Paris-Saclay Éditions, Université Paris-Saclay, 2025.

4, avenue des Sciences, 91190 Gif-sur-Yvette

www.msh-paris-saclay.fr

Collection « Actes »

ISSN 2800-7891



Cet ouvrage est publié en accès ouvert selon les termes de la licence Creative Commons Attribution – Utilisation non commerciale – Pas d'œuvre dérivée 4.0 International (CC BY-NC-ND 4.0), qui permet le partage de l'œuvre originale (copie, distribution, communication) par tous moyens et sous tous formats, sauf à des fins commerciales, à condition que l'œuvre originale soit correctement citée et diffusée sans modification, dans son intégralité.

Pour plus d'informations : <https://creativecommons.org/licenses/by-nc-nd/4.0/>

ISBN 978-2-9597054-1-0

La sélection naturelle en génétique des populations

Histoire, statut et perspectives

Guillaume ACHAZ

Bien que la sélection naturelle fût à l'origine de la théorie de l'évolution et qu'elle occupa une place prépondérante en biologie évolutive moléculaire jusqu'en dans les années 1960, sa place fut disputée lors du débat opposant les sélectionnistes aux neutralistes dans les années 1970-1980. Démise de sa place de modèle de référence par la dérive génétique, la sélection positive, dite darwinienne, fut reléguée au second plan des processus « importants » en génétique des populations. Seules les variations neutres étaient compatibles avec les patrons génétiques observés de divergence entre les espèces et de diversité au sein des espèces. Cependant, plusieurs observations nouvelles suggèrent qu'il est aujourd'hui nécessaire de rediscuter le dogme neutraliste établi pour appréhender sereinement les véritables causes de l'évolution moléculaire et caractériser au mieux la place de la sélection naturelle dans les processus qui façonnent la diversité génétique au sein et entre les espèces.

MOTS-CLÉS : sélection naturelle, génétique, évolution, théorie, diversité

Sélection naturelle (Darwin, 1859) : mot-valise qui porte de nombreux sens et sous-entendus, souvent assortis d'une généreuse dose d'à-peu-près. Selon l'interlocutrice ou l'interlocuteur, on entendra implicitement la survie du plus apte, une mesure de la fécondité des individus ou encore le moteur de la diversification de la vie sur Terre. Dans une branche sulfureuse des sciences humaines nommée « darwinisme social », on utilise un sens pervers du concept pour cautionner des pratiques de sélection sociales hautement questionnables.

Dans ce chapitre, je me concentrerai sur le cadre théorique de la biologie évolutive, et en particulier celui de la génétique des populations.

Je tâcherai de définir la plupart des termes utilisés en évitant autant que possible l'écueil d'une foule de détails techniques pouvant perdre le lecteur. Tout point ou définition incorrecte est à ma seule charge.

Un scientifique biologiste n'énonce aucune vérité. Il propose un cadre interprétatif aux observations qu'il fait. Avec ou sans nous, la nature *est*. Nous faisons au mieux pour construire des édifices théoriques qui donnent un sens logique aux observations que nous collectons sur le monde vivant. Sans prétention d'exhaustivité ou de vérité, je voudrais proposer un angle historico-scientifique de l'évolution moléculaire pour permettre d'attraper quelques idées simples et de prendre un peu de recul sur le sens pratique et théorique que l'on a donné à la sélection naturelle. Je me concentrerai notamment sur son rôle explicatif dans les patrons de diversité observés au sein des espèces.

La sélection naturelle en évolution moléculaire

L'objet du présent chapitre est de replacer l'usage et l'histoire de la sélection naturelle dans le domaine de l'évolution moléculaire. Pour cela, je présenterai les modèles fondamentaux, leurs évolutions au cours du dernier siècle et leurs applications à l'analyse des données. Bien que certaines circonvolutions peuvent sembler parfois un peu abstraites et sans lien direct avec le sujet, le fil conducteur restera néanmoins la place de la sélection naturelle, hier et aujourd'hui, pour les généticiens des populations.

En premier lieu, cartographions les limites du champ d'investigation du domaine nommé *évolution moléculaire*.

L'évolution moléculaire

Le domaine peut être vu comme l'étude des *composantes* moléculaires des patrons et des processus évolutifs. Il ne s'agit pas de *bases* moléculaires, car penser que l'évolution aurait ses fondements au niveau des molécules impliquerait une chaîne causale du microscopique vers le macroscopique, raisonnement réductionniste probablement erroné, sûrement discutable.

L'évolution moléculaire peut être divisée arbitrairement en deux facettes distinctes qui diffèrent par l'usage que l'on fait des données moléculaires.

Depuis la découverte des chromosomes, les chercheurs étudient l'évolution des molécules d'acide désoxyribonucléique (ADN) et des protéines qu'elles encodent. C'est l'objet par exemple de la génomique comparative ou de l'évolution des structures protéiques. On examine comment les chromosomes se fracturent et se recomposent au cours des temps longs qui séparent les espèces. La génomique comparative s'intéresse également, entre autres, à l'augmentation et à la diminution de la taille du génome ou du répertoire génique, à la dynamique de changement de l'ordre des gènes dans les chromosomes, voire à l'évolution des séquences régulatrices. Par l'approche comparative, on reconstruit la fresque temporelle de l'évolution des gènes et des génomes eux-mêmes. Pour des biochimistes structuralistes de l'évolution, on regarde les protéines codées par ces gènes, dont on peut énumérer les différentes structures, les regrouper en familles, dater l'apparition de certains repliements dans l'histoire de la vie, ou encore comprendre l'évolution corrélée des différents acides aminés d'une même protéine.

Bien que les molécules du vivant soient un objet d'étude en soi, on les utilise aussi comme des marqueurs de l'évolution des espèces. Dans cette seconde perspective, le génome n'est plus l'objet d'étude principal, il est l'outil qui permet d'explorer l'évolution du ou des individus des espèces d'intérêt. Ainsi, en génétique des populations, l'étude des variants qui ségrègent dans la population est un observable dont on se sert pour déterminer les caractéristiques de l'espèce : est-elle structurée ? A-t-elle connu des variations démographiques ? Est-elle soumise à un fort effet de sélection naturelle ? Les variants moléculaires en soi ne deviennent alors qu'un maillon de la chaîne de raisonnement.

En plus de cette distinction sur l'usage des données moléculaires, les évolutionnistes divisent généralement les échelles de temps évolutif – exprimé en générations – en deux catégories. En microévolution, on s'intéresse aux « petites » échelles temporelles, qui couvrent des périodes allant de quelques générations à plusieurs dizaines de milliers de générations. L'attention se porte principalement sur les évolutions au sein d'une espèce, que l'on imagine se produire sur des échelles de temps raisonnables, parfois même en laboratoire dans des dispositifs dits d'*évolution expérimentale*. *A contrario*, en macroévolution, on s'intéresse à des temps évolutifs beaucoup plus longs. On étudie par exemple les relations d'apparement

entre espèces ou leur formation, comme en phylogénie. Je concède sans discussion que cette distinction temporelle est caricaturale et que de nombreuses études empruntent aux deux échelles de temps.

Ce chapitre se concentre sur la place de la sélection naturelle en génétique des populations, assimilée à de la microévolution. Cependant comme implicitement les phénomènes de macroévolution agrègent des séquences d'événements de la microévolution, la place de la sélection naturelle en génétique des populations diffuse dans celle de la macroévolution, et vice-versa.

Origine de la génétique des populations

La génétique des populations a émergé il y a environ une centaine d'années. On date usuellement les articles fondateurs de la discipline aux années 1920 (Fisher, 1919 ; Wright, 1921 ; Haldane, 1927), bien qu'il y ait quelques articles antérieurs, comme ceux de Godfrey Harold Hardy (1908) et de Wilhem Weinberg (1908), ou encore ceux des biométriciens (Galton, 1889). Au cours du dernier siècle (1920-2020), la transformation du domaine a été marquée par divers événements, que j'aborderai rapidement. Ce rapide tour historique me semble crucial pour apprécier l'évolution du champ conceptuel et la posture scientifique qui lui est associée.

En premier lieu, il convient de rappeler que la naissance de la discipline s'opère à une époque où les observations de génétique sont limitées, voire absentes. Notre habitude d'analyser des séquences d'ADN pourrait nous faire oublier que ces données aujourd'hui si banales et abondantes étaient à cette époque totalement inexistantes. La question du support physique de l'hérédité génétique n'était ainsi même pas entièrement résolue, car la théorie chromosomique (Sutton, 1903 ; Boveri, 1904) débattue pendant plus d'une décennie venait d'avoir une réelle reconnaissance des « généticiens ». Les séquences d'ADN étaient quant à elle totalement inconnues.

Quelques définitions et principes

La génétique des populations des origines a donc été largement développée à travers des modèles mathématiques mêlant deux approches complémentaires, conformément aux pratiques courantes de modélisation. L'approche

« déterministe » repose sur la spécification d'un état initial et vise à prédire sans incertitude un état ultérieur. C'est une approche naturelle appliquée en physique mécanique qui permet de prédire la position d'un objet en mouvement. L'approche « stochastique » fait quant à elle l'hypothèse que le déroulement d'un processus ne peut pas être déterminé avec certitude, mais qu'il est mieux décrit par des lois de probabilités où, depuis l'état initial, certains états futurs sont plus probables que d'autres. En physique statistique, cette approche s'applique au comportement d'un ensemble de particules dont chacune se comporte complètement au hasard. En génétique évolutive, la modélisation stochastique est intimement liée au phénomène de « dérive génétique ».

En génétique des populations, nos observations sont en premier lieu les fréquences des variants qui ségrègent dans la population. En effet, à un locus (dans une région) du génome, plusieurs variants (allèles) existent pour les différents individus. En examinant la répartition de ces variants dans la population, on voudrait comprendre les processus qui ont généré ces patrons de diversité.

La définition précise d'un locus est laissée volontairement floue. On peut néanmoins considérer qu'il s'agit d'un segment génomique dont la taille et les caractéristiques ne sont pas clairement spécifiées. En génétique des populations (et en évolution moléculaire), notre intérêt se porte principalement sur les locus dits polymorphes, c'est-à-dire ceux qui abritent plusieurs allèles. Ainsi, les différentes « formes » d'un locus sont nommées « allèles ». Aujourd'hui, on sait qu'en pratique, il s'agit de séquences de nucléotides. On parle également volontiers de « variants » pour désigner les différents allèles qui ségrègent à un locus.

Contrairement à la génétique classique ou médicale, qui se concentre sur des pedigrees choisis comprenant un nombre restreint d'individus, la génétique des populations se focalise sur le « pool génétique », c'est-à-dire l'ensemble des variants portés par l'ensemble des individus d'une population. La notion de « population » elle-même n'est pas définie clairement. On l'envisage comme un ensemble d'individus capables de s'échanger du matériel génétique, allant d'un petit groupe, comme une tribu, à l'ensemble des individus d'une espèce.

Dans la Figure 1, un pool génétique est représenté. Certains individus sont coloriés en noir, pour figurer un allèle particulier, d'autres sont en jaune, et d'autres encore sont en blanc. L'objectif du domaine est de comprendre l'évolution de ce pool génétique au fil des générations et, *in fine*, d'expliquer la composition du pool. Par exemple, dans la Figure 1, les allèles blancs disparaissent progressivement au cours du temps, tandis que les allèles jaunes croissent.

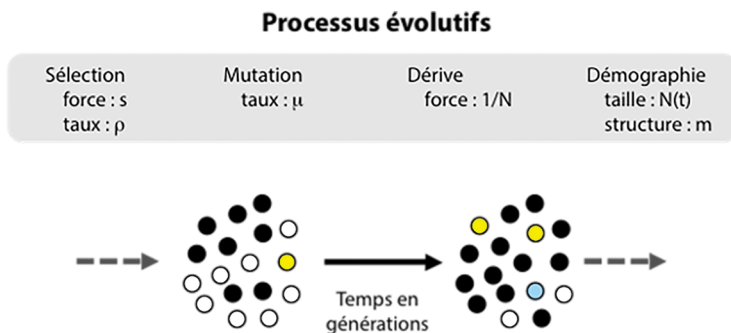


Figure 1 – Génétique des populations

Les individus des populations sont réduits à leur simple matériel génétique. On s'intéresse à la composition du pool génétique et notamment à son évolution au cours du temps, exprimé en générations. Les processus standards ayant un impact sur ce pool génétique sont la sélection (de taux p et de coefficient de sélection s), la mutation (de taux μ), la dérive génétique et les processus démographiques comme les variations de la taille de population (N ou $N(t)$ si elle varie dans le temps) et la structuration paramétrée par le taux de migration (m).

Les processus en jeu incluent notamment la mutation, qui crée de nouveaux allèles, mais aussi naturellement la sélection naturelle, qui par son action favorise ou défavorise certains variants. C'est un moteur très puissant pouvant rapidement changer la composition du pool génétique. L'autre processus majeur, pourtant méconnu du grand public, est celui de la dérive génétique. Il s'agit d'un processus stochastique causé par la finitude de la population qui entraîne un échantillonnage des allèles au cours des générations. Ainsi, par effet d'échantillonnage, lorsque peu d'individus sont présents, des allèles peuvent rapidement envahir l'ensemble de

la population, sans aucune implication de la sélection naturelle. Enfin, il convient de noter que les populations naturelles présentent une structuration en sous-populations qui ne se mélangent pas uniformément. On dit qu'elles ne sont pas panmictiques. Avant l'avènement des moyens mécaniques de locomotion, par exemple, les humains trouvaient leurs conjoints, avec lesquels ils ont eu des descendants, principalement dans leur voisinage. Ces structures démographiques spatiales interfèrent avec la variation du nombre total d'individus, qui connaît des phases alternées de croissance et de décroissance.

Quelques mots sur les fondateurs

Dans les années 1920-1930, la génétique des populations émerge sous l'impulsion de trois figures considérées *a posteriori* comme les « fondateurs » du domaine.

Ronald Aylmer Fisher (1890-1962) est un professeur de génétique (Londres puis Cambridge, Royaume-Uni), célèbre pour ses travaux en biologie et en statistiques. Il fait partie de l'école des biométriciens, dans la lignée de Francis Galton ou Karl Pearson, qui ont fondé les statistiques modernes au tournant du XIX^e et du XX^e siècle, et qui cherchent à rationaliser et à quantifier la biologie. C'est en 1930 que Fisher publie son ouvrage *The Genetical Theory of Natural Selection* (Fisher, 1930), livre qui taille une place prépondérante à la sélection naturelle pour expliquer l'évolution du monde vivant. Bien que le langage soutenu du livre puisse rendre sa lecture difficile, Fisher explore méthodiquement l'impact de la sélection naturelle sur les pools génétiques, utilisant les concepts de gènes et de locus, bien que leur nature physique soit encore floue. Le théorème fondamental de la sélection naturelle, selon Fisher, interroge le rôle de la sélection naturelle sur ces pools génétiques d'individus porteurs de chromosomes, une question abordée méthodiquement dans ses nombreux écrits rigoureux.

Sewall Wright (1889-1988), autre professeur de génétique exerçant à Columbia (New York, États-Unis) s'est consacré aux liens d'apparentement et à la consanguinité des individus. Son parcours, couvrant tout le développement de la génétique au XX^e siècle, est composé d'une collection d'articles théoriques plutôt que d'un livre (bien qu'il ait publié plus

tardivement dans sa carrière des manuels exhaustifs). Deux de ses articles sont particulièrement connus : « Evolution in Mendelian Populations » (Wright, 1931), dans lequel il décrit *in extenso* les effets des processus évolutifs, dont notamment la sélection naturelle et la dérive génétique sur les pools génétiques », et « The Roles of Mutation, Inbreeding, Crossbreeding and Selection in Evolution » (Wright, 1932), dans lequel il introduit la métaphore des paysages adaptatifs. Dans cette métaphore très populaire, les individus sont tous décrits par des combinaisons génétiques qui correspondent aux coordonnées d'un « paysage », où la hauteur est une valeur de fitness (succès évolutif). La sélection naturelle fait évoluer les populations en maximisant la fitness moyenne de la population.

John Burdon Sanderson Haldane (1892-1964) est également professeur de génétique (Londres, Royaume-Uni), auteur de l'ouvrage *The Causes of Evolution* (Haldane, 1932), qui suscite l'intérêt en explorant la variation au sein des espèces, la sélection naturelle et la notion de fitness. Haldane, caractérisé par son esprit de synthèse, a contribué très largement à la compréhension de ces concepts. À cette époque, l'attention se porte sur la manière dont la sélection naturelle explique des phénomènes, tels que le changement de couleur de blanc à noir des ailes de la phalène du bouleau suite à la pollution industrielle dans la région de Manchester.

Adaptation

Revenons au point principal de ce chapitre : la place de la sélection naturelle. Pour éclairer ce concept fondamental en génétique des populations, nous devons considérer une population d'individus. Sur le plan génétique, concentrons-nous sur un locus particulier où plusieurs variants ségrègent. Nous suivons la fréquence de ces variants au sein de cette population. Dans ce canevas typique de la génétique des populations, l'évolution est assimilée aux changements de fréquences des allèles au cours des générations.

Supposons maintenant qu'un variant avantageux apparaisse à ce locus dans un premier chromosome d'un individu de la population. La sélection naturelle favorisera la propagation de ce variant s'il confère un trait avantageux à ses porteurs. Aussi, bien qu'initialement rare, le variant pourra augmenter progressivement en fréquence jusqu'à devenir

prédominant et, *in fine*, remplacer tous les autres variants préexistants. Lorsque la fréquence atteint 100 %, on parle de fixation. Cette séquence d'événements – apparition d'un nouveau variant, augmentation en fréquence et fixation du variant – est nommée « substitution ». Un variant remplace tous les autres de la population à ce locus. Elle illustre la composante génétique (pour un unique locus) de l'adaptation évolutive qui se produit continuellement pour les organismes vivants. Ce phénomène de substitution se produit probablement en même temps à plusieurs locus. En effet, d'une part, les traits phénotypiques sont généralement gouvernés par plusieurs locus (on parle de traits polygéniques) et, d'autre part, plusieurs traits, chacun lié à des locus différents, sont probablement sélectionnés à chaque instant.

Ce schéma global dépeint donc l'évolution comme une succession de substitutions de variants avantageux dispersés dans plusieurs locus du génome, chacun améliorant un ou des traits sélectionnés. Ce processus constitue une course perpétuelle vers l'adaptation et entraîne en cascade une évolution continue de la composition génétique partout dans le génome. Cela devient d'autant plus flagrant dès lors que l'on considère que l'environnement est lui-même composé principalement d'autres organismes, eux-mêmes en perpétuelle évolution. Cette incessante course à l'adaptation a été popularisée par la métaphore de la « reine rouge » de Leigh Van Valen (1973). Dans le roman de Lewis Carroll (1865), Alice, qui se retrouve de l'autre côté du miroir, doit sans cesse courir aux côtés de la reine rouge d'échec pour essayer de se maintenir sur place. Si elle s'arrête, elle recule. Pour une espèce, reculer, c'est perdre des individus, ce qui signifie donc, au final, l'extinction. Une espèce doit sans arrêt se déplacer en avant dans le paysage évolutif (par mutation et sélection), sous peine de disparaître.

À gros traits, les variants sont souvent classés comme bénéfiques, et donc favorisés et propagés par la sélection, ou comme délétères, et donc purgés et éliminés de la population. Dans les deux cas, la variation est éliminée. Les variants dits neutres n'apportent, eux, aucun avantage ou désavantage et dérivent au gré des aléas. Il est toutefois impératif de reconnaître que cette représentation demeure une caricature, car les

processus réels sont inconnus et vraisemblablement bien plus complexes que ceux imaginés dans ces modèles simples.

Outre cette sélection « classique », d'autres formes de sélection plus alambiquées méritent d'être mentionnées, comme la sélection sexuelle. Un exemple emblématique en est la queue du paon, qui, bien que handicap majeur face aux prédateurs, séduit les femelles de cette espèce. Plus généralement, la sélection des mâles par les femelles peut conduire au développement de caractéristiques apparemment défavorables à la survie de l'individu, qui pourtant augmentent son nombre de descendants. Néanmoins, comme la sélection liée à l'adaptation à l'environnement, la sélection sexuelle élimine la variation. On parle volontiers de sélection directionnelle pour tous ces cas où la sélection entraîne les populations vers une seule « direction ».

Il existe cependant d'autres mécanismes sélectifs qui maintiennent la diversité génétique au sein des populations. En raison de leur effet d'équilibre sur les fréquences des variants, on les qualifie de sélection balancée. Contrairement à la tendance éliminatoire inhérente à la sélection directionnelle, ces processus particuliers maintiennent la variation génétique. L'exemple typique est celui de l'avantage de l'hétérozygote, où la meilleure valeur de fitness est associée aux individus qui possèdent deux variants différents sur leurs deux chromosomes homologues. On imagine que c'est le cas pour les gènes du système immunitaire en charge de reconnaître le non-soi. Bien que les détails ne soient pas abordés ici en profondeur, il faut noter l'existence de ces mécanismes de sélection capables de préserver la diversité génétique. Ils offrent une perspective plus nuancée sur la génétique des populations et ont été mis en avant lors du débat entre les neutralistes et les sélectionnistes que nous détaillerons plus loin.

Changement de paradigme

La synthèse moderne

Entre les années 1940 et 1960 eut lieu ce que les scientifiques impliqués dans ce projet dénommèrent la *synthèse moderne de l'évolution*. Le projet était d'offrir une vue intégrée des processus et des patrons évolutifs en

combinant les spécialités de tous les domaines des sciences de l'évolution tels que la paléontologie, la zoologie, la botanique, l'écologie, le développement, la systématique et la génétique. Un ensemble de biologistes de l'évolution, triés sur le volet, ont ainsi mis en commun leurs connaissances pour élaborer une synthèse des avancées de l'époque. Cette synthèse moderne de l'évolution a surtout mis l'accent sur l'implication de la sélection naturelle et a fait peu cas du processus stochastique d'échantillonnage de la dérive génétique.

Ainsi, même au cours de cette période charnière des années 1940-1960, les explications du fonctionnement du monde et les questions captivantes qui en découlent étaient orientées vers la compréhension et la caractérisation du rôle de la sélection naturelle.

La théorie neutraliste

Cependant, un événement majeur allait déplacer le centre d'intérêt de la discipline : les premières mesures de la diversité moléculaire (Lewontin & Hubby, 1966). À la lumière de l'amplitude insoupçonnée de cette diversité moléculaire, il devenait difficile d'accepter simplement que le processus prédominant en évolution moléculaire était la sélection naturelle et, en particulier, la sélection directionnelle. En effet, la sélection naturelle a une propension si forte à éliminer la diversité, qu'il faudrait des taux de mutation très élevés pour maintenir de la diversité à l'équilibre. Ces taux n'étaient pas en accord avec les mesures expérimentales des taux de mutation réels (Mukai, 1964). Combinée avec l'observation que les fixations de variants dans les protéines se produisaient à un taux constant, comme le tic-tac d'une horloge (Zuckerlandl & Pauling, 1962), et avec la démonstration que seul un petit nombre de variants pouvaient être sélectionnés à un instant donné (Haldane, 1957), émergea l'idée que la très grande majorité des variants observés n'étaient probablement soumis à aucune forme de sélection (Kimura, 1968 ; King & Jukes, 1969). Ainsi, l'évolution de ces variants génétiques n'était plus le fait d'une sélection positive ou négative, mais celui du processus d'échantillonnage neutre nommé « dérive génétique ». Si les phénotypes liés aux allèles d'un gène particulier n'ont aucune incidence significative sur la survie ou la

reproduction de l'organisme qui les porte, la diversité génétique au sein des populations peut être préservée.

Les prémices de ce changement de perspective remontent au début des années 1970. L'un des défenseurs les plus éminents de l'idée selon laquelle les variants génétiques observés sont principalement neutres est alors Motoo Kimura, un chercheur japonais ayant effectué sa thèse aux États-Unis avant de retourner établir un laboratoire d'évolution à Mishima, au Japon. Pendant les années 1950, Kimura s'est consacré à l'élaboration de modèles mathématiques complexes en génétique des populations et en évolution moléculaire. Il s'est notamment attelé à la caractérisation fine du rôle de la sélection naturelle, combinée aux autres processus (dérive, mutation, recombinaison, etc.). La seconde moitié de sa carrière, à partir des années 1970, est dédiée à défendre et à promouvoir l'idée selon laquelle les variants observés ne sont pas soumis à la sélection. Son ouvrage *The Neutral Theory of Molecular Evolution* (Kimura, 1983) marque en quelque sorte la fin de sa carrière, expliquant clairement sa position vis-à-vis de la sélection naturelle. Il ne réfute en rien son existence, mais soutient que cette sélection ne concerne pas la grande quantité de variants observés à des fréquences intermédiaires. Comme la sélection naturelle élimine rapidement les variants délétères et fixe les variants bénéfiques, seuls ceux qui sont neutres restent suffisamment longtemps pour être observés à fréquence intermédiaire.

Il faut souligner la contribution exceptionnelle de Kimura qui, d'une part, possédait une expertise en modélisation mathématique très pointue et, d'autre part, un recul remarquable sur la portée des modèles et leur pouvoir à prédire des observations vérifiables. Cette double compétence est très rare et il faut souligner la grande contribution de cette figure atypique.

En l'absence de sélection naturelle, que se passe-t-il pour ces variants neutres ? Ces derniers, tout comme leurs homologues soumis à la sélection, apparaissent aléatoirement au sein de la population par mutation ou par immigration. Puis, leurs fréquences évoluent par oscillation entre 0 et 1, sans tendance à l'augmentation ou à la diminution, jusqu'à leur perte ou leur fixation.

Un calcul simple montre que l'écrasante majorité des variants neutres qui apparaissent sont voués à disparaître en quelques générations. Plus encore, un variant atteignant une fréquence de 50 % peut finalement disparaître avec une probabilité de 0,5 (égale à sa fréquence). Parfois, très rarement, certains de ces variants parviennent à passer de fréquences négligeables à une fréquence de 1, signifiant leur fixation complète sur l'ensemble des individus de la population. Il est crucial de souligner que la durée de cette invasion est très lente. Cette lenteur laisse largement le temps d'observer des états polymorphes. En effet, à l'échelle de la dérive génétique, la fenêtre temporelle est de l'ordre de N générations, où N est la taille de la population. Cela contraste fortement avec l'effet de la sélection, où les périodes de polymorphisme sont très courtes, rendant improbable l'observation de variants sélectionnés à l'état polymorphe.

Le modèle standard neutre

Bien qu'il ne soit souvent introduit que tardivement dans la plupart des manuels de génétique des populations, le modèle de dérive génétique est aujourd'hui *LE modèle de référence* en évolution moléculaire. Il occupe la place d'hypothèse nulle au sens statistique. Pour argumenter que d'autres processus sont ou ont été à l'œuvre dans l'évolution d'un jeu de données quelconque, il faut rejeter en premier lieu le modèle standard neutre à l'aide de tests statistiques dédiés, qualifiés de *tests de neutralité*.

La Figure 2 propose un schéma du paradigme établi par le modèle standard neutre. À chaque génération, il y a un apport de nouveaux mutants, issus des lésions naturelles et mal réparées qui se produisent dans les génomes, mais aussi des erreurs de copies de l'ADN lors des divisions cellulaires. Cet apport est contrebalancé par la dérive génétique qui purge cette diversité à une vitesse très lente. Ainsi, même pour des taux de mutations faibles, l'équilibre prédit de la diversité génétique.

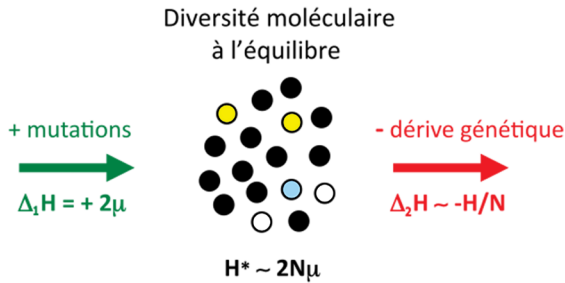


Figure 2 – Équilibre mutation-dérive

La diversité moléculaire (H) observée résulte d'un équilibre entre les mutations qui augmentent la diversité et la dérive génétique qui la réduit. Ici, $H_t = H_{t-1} + \Delta H$ est modélisée comme une suite mathématique qui, à chaque génération, d'une part, augmente par mutation ($\Delta_1 H = +2\mu$) et, d'autre part, diminue par dérive génétique ($\Delta_2 H = -H/N$). À l'équilibre ($\Delta H^* = \Delta_1 H^* + \Delta_2 H^* = 0$), la diversité prédite (H^*) est proportionnelle au produit de la taille de population (N) par le taux de mutation (μ).

Le modèle de Wright-Fisher

La description la plus commune du processus de dérive génétique repose sur un modèle de population de taille N finie et constante, appelé modèle de Wright-Fisher en référence à ses créateurs Sewall Wright Ronald Aylmer Fisher. Les générations sont non chevauchantes. Dans cette population, il n'y a pas de sélection et le régime de reproduction est celui d'une « panmixie stricte », ce qui signifie que n'importe quel individu peut être parent de n'importe quel descendant avec la même probabilité. Ainsi, il n'y a pas de structuration spatiale ou comportementale, aucune division en sous-populations. On peut assimiler une population Wright-Fisher à une population de génomes portés par des organismes qui ne sont pas structurés. Chaque génome porte un allèle et l'on suit la répartition des allèles dans la population au cours du temps. Comme la population est de taille constante, chaque individu a en moyenne un descendant. Plus précisément, par ce processus d'échantillonnage, le nombre de

descendants par individu est bien approximé par une loi de Poisson (loi de probabilité associée à tous les entiers naturels et proposée par Siméon Denis Poisson en 1837) de moyenne 1, lorsque N est au moins de plusieurs dizaines.

La Figure 3 représente une réalisation du processus de dérive dans le modèle de Wright-Fisher (Ishida & Rosales, 2020) sur 18 générations. Le temps s'écoule de gauche à droite et chaque colonne de disques représente la population à un temps donné. Dans la figure, la population comporte $N=7$ individus. À chaque génération, tous les individus meurent et sont remplacés par ceux de la génération suivante, comme pour une population de plantes annuelles. Les liens entre les disques représentent leur filiation. Dans cette figure, les individus sont haploïdes et n'ont donc qu'un unique parent. Il est aisé de généraliser le modèle pour des diploïdes sans changer grandement la teneur du propos.

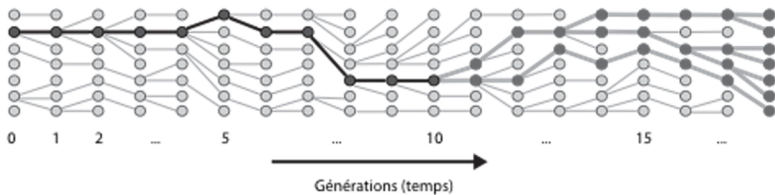


Figure 3 – Modèle de Wright-Fisher

Une réalisation du modèle stochastique de Wright-Fisher sur 18 générations. Chaque colonne représente la population (ici $N=7$ individus) pour une génération donnée. Les filiations entre les individus sont représentées par des lignes. Le trait noir épais relie tous les individus fondateurs (individus dont la descendance envahit toute la population), formant ainsi une unique lignée persistante. Il est également possible de suivre les ancêtres dans le temps rétrospectif du temps, par exemple, de la génération 18 à la génération 10. L'arbre ainsi obtenu est appelé arbre de coalescence.

Les générations se succèdent sans chevauchement, mais la transmission du matériel génétique à la nouvelle génération est parfaite. Si l'on suit la descendance de tous les individus à une génération donnée, on note que seule la descendance d'un unique individu persistera à l'infini

dans ce modèle. La descendance de tous les autres individus s'éteint. L'unique individu dont la descendance persiste peut être qualifié de *fondateur*. Comme il existe un tel individu à chaque génération, on observe une unique lignée persistante reliant tous les fondateurs de toutes les générations. Dans la Figure 3, cette lignée persistante est la ligne épaisse et noire entre les générations 0 et 10. Tous les autres individus ne laisseront aucun descendant sur le long terme. Seule la lignée persistante fonde l'intégralité de la population dans le futur. Il ne s'agit cependant pas d'un processus déterministe, car il n'est pas possible de prédire, à l'avance, quel sera l'individu fondateur, bien qu'il soit garanti qu'il en existe un et un seul pour un segment de génome non recombinaison. Lorsque plusieurs segments de génome sont assemblés dans un unique individu par recombinaison, il existe un ancêtre différent pour chacun de ces segments.

Le processus de fixation

Jusqu'aux années 1980, ce modèle a été finement caractérisé dans le sens prospectif du temps, exactement comme nous venons de le décrire, en choisissant une génération de référence et en suivant la descendance d'un individu, fondateur ou pas. On suit en général dans les générations ultérieures la fraction de la population composée par la descendance de l'individu choisi. Lorsque la descendance d'un fondateur atteint 100 % de la population, on parle de fixation. En examinant la fréquence d'un allèle focal, on s'intéresse au processus d'envahissement (éventuellement raté) de la population par une lignée particulière. La séquence des fixations successives des lignées fondatrices entraîne à rythme régulier la fixation des allèles portés par ces lignées. Ce phénomène est connu sous le nom d'« horloge moléculaire » (Zuckerkandl & Pauling, 1962). Au cours des générations, à un rythme régulier, des mutations se produisent chez certains individus (ceux de la lignée des fondateurs), qui seront propagées sur les temps plus longs à l'ensemble de l'espèce, créant ainsi la divergence entre les espèces.

Le processus de coalescence

Depuis les années 1980, le même modèle a été examiné dans le sens rétrospectif du temps, en remontant les générations dans le passé. Au temps de référence, on échantillonne certains individus d'une population dont on suit les lignées ancestrales. Il s'agit bien du même modèle, avec les mêmes règles et les mêmes distributions de descendants. Cependant, au lieu de l'écrire dans le sens naturel du temps, on le réécrit en remontant dans le passé. Ce principe est illustré dans la Figure 3, où les ancêtres des individus de la génération 18 sont marqués en trait gris épais jusqu'à l'ancêtre commun de la génération 10. Cet ancêtre commun est bien entendu de la lignée des fondateurs. L'arbre reliant tous ces individus est nommé *arbre de coalescence* et l'étude de ses arbres *théorie de la coalescence* (Kingman, 1982).

La littérature autour de ce modèle, que ce soit dans le sens prospectif ou rétrospectif du temps, est extrêmement généreuse et foisonnante. Beaucoup de caractéristiques (probabilité de fixation, temps de fixation, temps de perte, temps de coalescence, etc.) ont été calculées et forment un corpus mathématique, statistique, informatique et biologique très robuste.

Le paradigme actuel

Ainsi, le paradigme dans lequel la plupart des travaux sont réalisés aujourd'hui consiste à admettre que le modèle de Wright-Fisher représente une certaine forme de réalité, dans laquelle les populations sont de taille finie et où chaque individu, par hasard, donne ou non des descendants. Dans ce canevas, la dérive génétique est le moteur principal des variations de fréquences alléliques au cours des générations. C'est le cadre théorique fondamental (bien que parfois ignoré par les pratiquants de la discipline) dans lequel s'inscrivent la plupart des études actuelles menées en génétique des populations. Il n'y a aucune place pour la sélection naturelle. Toute idée que les allèles modifient le nombre de descendants a été bannie.

Modèle neutre et diversité génétique

Diversité génétique

Sous ce modèle neutre, il est possible de calculer la parenté génétique entre les individus. Je souligne que la parenté génétique peut être différente de la parenté des pédigrées. En effet, deux individus peuvent potentiellement avoir un ancêtre commun dans leur pédigrée, sans avoir aucun matériel génétique en commun. Par exemple, deux descendants actuels de Louis XIV, séparés par deux fois 15 générations, ont une grande chance de n'avoir rien hérité du roi soleil sur le plan génétique. En moyenne, ils ne possèdent qu'une petite fraction de son génome ($1/2^{15} \sim 1/30\,000$) et ne partagent entre eux qu'une fraction microscopique ($1/2^{30} \sim 1$ milliardième). Sur le plan génétique, on se concentre donc sur les ancêtres génétiques ayant contribué aux génomes actuels. Cette parenté génétique est stochastique, mais on peut en calculer une moyenne.

L'ancêtre de deux chromosomes homologues, échantillonnés au hasard dans une population de N individus haploïdes (ou diploïdes), vivait en moyenne il y a N générations dans le passé (ou $2N$ pour les diploïdes). Ainsi, dans une petite population, les ancêtres génétiques sont plus récents et les individus plus « consanguins ». Comme les deux chromosomes sont séparés par $2N$ générations, ils sont également séparés par $2N\mu$ événements de mutations, où μ note le taux de mutation. C'est la prédiction du modèle standard neutre (Figure 2).

La variance associée à ce nombre de mutations est cependant très grande, car elle intègre la variance du processus de généalogie, combinée avec celle du processus de mutation. Pour chaque segment non recombiné du génome, il existe un arbre généalogique particulier dont l'ancêtre commun le plus proche a existé à un temps différent du passé. À cette variabilité généalogique s'ajoute celle des mutations qui se produisent le long de l'arbre généalogique et génèrent un site polymorphe. De la conjonction de la stochasticité des arbres et des mutations découle une densité de polymorphismes très variable le long du génome (Figure 4) : les locus présentent en effet une diversité

très variable le long du génome. Certains rares locus présentent néanmoins une très forte diversité ou, comme sur la figure, une absence quasi complète de diversité. Que peut-on conclure de ce type de patrons ?

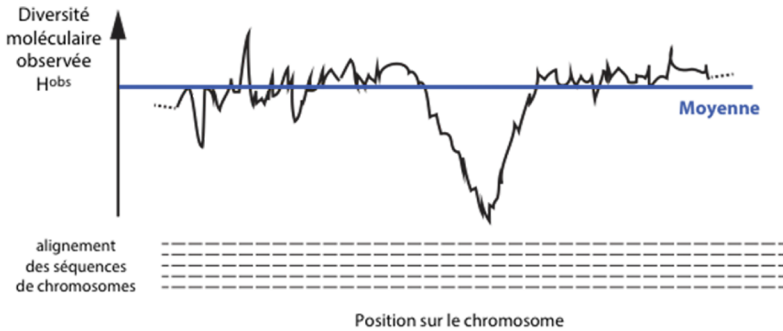


Figure 4 – Diversité observée le long des chromosomes

La diversité moléculaire, calculée pour une fenêtre glissante à partir d'un alignement de séquences chromosomiques, présente de fortes variations le long des chromosomes. Elle oscille généralement autour de la diversité moyenne, mais certains locus présentent des patrons atypiques, comme ici une forte diminution de la diversité pour les locus au centre du dessin.

La recherche de locus génomiques particuliers du point de vue de la diversité (trop grande ou trop faible) a été l'objet de nombreuses initiatives visant à cataloguer les gènes dont l'évolution ne semble pas régie par le modèle standard neutre. En arrière-plan, on imagine que ces locus ne sont justement pas neutres, mais qu'ils sont ou ont été soumis à l'action de la sélection naturelle. Typiquement, les locus ne présentant aucune diversité sont supposément ceux ayant été le théâtre d'un événement sélectif directionnel ayant purgé toute la diversité, tandis que les locus présentant une très grande diversité sont peut-être les lieux d'une sélection balancée.

Ainsi, le rôle de la sélection naturelle est d'expliquer les patrons de diversité de quelques rares locus, qui présentent des diversités trop petites ou trop grandes. Pour la grande majorité des locus, le processus d'évolution accepté est celui du modèle standard, c'est-à-dire un

équilibre entre mutation et dérive génétique. Ce processus stochastique de dérive très lent laisse une diversité notable le long du génome, bien qu'il possède une grande variance.

À certains locus rares et particuliers du génome, des variants bénéfiques apparus récemment ont envahi toute la population et ont ainsi éliminé toute la diversité avoisinante. Lorsque l'on examine un autre locus suffisamment éloigné de l'événement de sélection, il n'y a aucune réduction de diversité liée de cet événement dit de *balayage sélectif*. Par contre, aux régions voisines du locus sélectionné, on note une réduction (partielle) de la diversité génétique. Cette réduction est la conséquence de l'organisation du matériel génétique en chromosomes. Tous les allèles d'un chromosome sont physiquement liés et donc généralement cotransmis lors d'une division cellulaire. Seule la recombinaison crée des chromosomes mosaïques et donc brise l'association entre les allèles d'un même chromosome. Plus les locus sont physiquement éloignés, plus ils recombinent et plus les variations des fréquences alléliques deviennent indépendantes. Ainsi, le destin particulier d'un locus n'interfère que sur ses voisins proches.

En examinant la diversité le long des génomes de n'importe quelle espèce, on peut rechercher les locus soumis à la sélection et proposer des explications naturalistes. On cherche à raconter une histoire remarquable, où la sélection naturelle est exceptionnellement impliquée. La règle est l'absence de sélection, mais certains locus dans le génome de certaines espèces sont soumis à la sélection naturelle. Ils deviennent ainsi dignes d'une narration particulière.

Par rapport aux années 1920-1930, le canevas de travail est bien différent. On admet un cadre neutre dans lequel la sélection ne joue qu'un rôle mineur. Chez les humains seuls, certains locus présentent une diversité fortement réduite et témoignent donc d'une sélection très forte. Les exemples sont relativement peu nombreux. Le plus célèbre est probablement celui du gène de la lactase, une enzyme qui dégrade le lait chez les nourrissons et les jeunes enfants. Les adultes, comme dans la plupart des espèces de mammifères, ne se nourrissent généralement plus de lait. Mais chez les humains, dans certaines populations (notamment les populations européennes), la capacité à digérer le lait à l'âge

adulte serait apparue grâce à une mutation « avantageuse » permettant l'expression du gène même chez les adultes (Ingram *et al.*, 2009). Cet avantage a été ensuite mis en relation avec l'apparition de l'agriculture et l'apport nutritif du lait de vache en cas de famine. Cette très jolie histoire présente néanmoins plusieurs zones d'ombres (Ségurel & Bon, 2017). La tentation de construire une narration haletante est omniprésente en biologie. Cette fascination est motrice de la quête de connaissances, mais gardons-nous d'accepter toutes ces histoires comme des vérités avérées.

Pour la plupart des espèces, la liste de tous les gènes portant une trace claire de la sélection naturelle n'est pas très longue. Chez les humains, quelques gènes ayant trait à une histoire particulière (comme la lactase ou d'autres liés à l'adaptation à l'altitude), ainsi que certains gènes liés à l'immunité (Fan *et al.*, 2016), présentent des traces évidentes de sélection. Cette liste réduite conforte l'idée que les événements de sélection naturelle impactant fortement les patrons génétiques sont rares.

La taille efficace

Je voudrais maintenant aborder le concept de *taille efficace*, que les généticiens des populations (moi y compris) utilisent en routine. L'idée est brillante, voire lumineuse, mais elle se dérobe à notre compréhension lorsqu'elle est examinée à la loupe. Elle devient alors source de nombreuses confusions. Le sens de la taille efficace varie beaucoup selon les contextes, les époques, les communautés et les personnes qui le manient. Lorsque l'expression « taille efficace » est évoquée, chaque scientifique se l'approprie à sa manière, bien que très peu soient capables d'en donner un sens précis et encore moins une signification générale. Contrairement aux concepts de gènes, de locus ou de sélection, qui sont reconnus pour leurs visages multiples, la taille efficace reste souvent très mystérieuse.

La vision la plus correcte de la taille efficace est celle d'un nombre théorique, parfois estimé à partir de données, qui est profondément lié au modèle de Wright-Fisher. Il représente la taille de la population Wright-Fisher idéalisée qui aurait les mêmes variations de fréquences

alléliques que la population considérée. Cette dernière peut correspondre soit à un modèle différent du modèle standard (par exemple avec distinction entre mâles et femelles), soit à une population réelle caractérisée par des indicateurs de diversité. Selon la manière choisie pour mesurer les variations de fréquences alléliques, les valeurs de taille efficace peuvent varier. Les mesures sont-elles faites sur une ou plusieurs générations ? Sont-elles mesurées dans le sens prospectif du temps (dérive génétique) ou dans le sens rétrospectif (coalescence) ? Cherche-t-on à rendre compte du sex-ratio, des variations de taille, de la structuration ou de la sélection naturelle ? Il faut retenir que ces nombres estimés sont généralement interprétés dans le cadre du modèle standard neutre. Cependant, il faut absolument garder à l'esprit que ces tailles efficaces n'ont pas de raisons évidentes d'être liées à un nombre d'individus. La taille efficace peut être vue comme la projection, au sens mathématique, de la population étudiée dans une population standard neutre de type Wright-Fisher. À cet égard, l'interprétation d'une valeur de taille efficace doit être faite avec délicatesse.

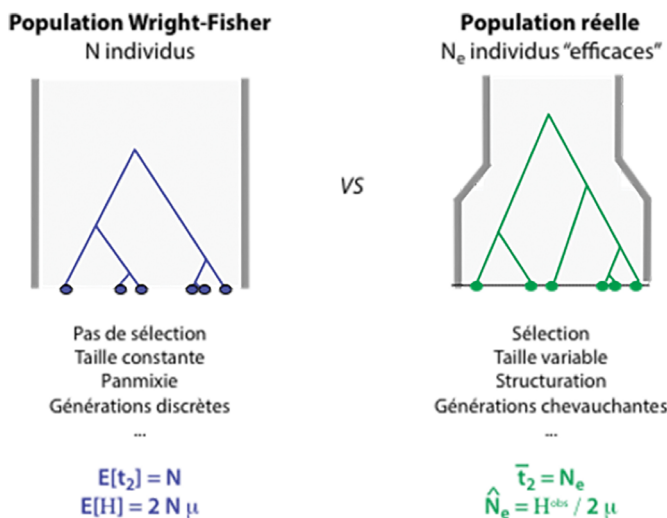
Beaucoup de scientifiques se sont attelés à évaluer la pertinence du modèle standard neutre. Notamment, au moment de l'avènement de la théorie neutraliste, de houleux débats ont opposé les « neutralistes » aux « sélectionnistes ». Ces derniers étaient foncièrement opposés à l'idée que le modèle neutre puisse expliquer les patrons d'évolution moléculaire et la diversité génétique observée. Dans cette tradition critique, aventurons-nous sur ce terrain : le modèle standard neutre explique-t-il correctement la diversité génétique observée ?

L'un des piliers soutenant la théorie neutraliste vient de l'idée, très juste, que la sélection naturelle directionnelle purge avec une redoutable efficacité la variation génétique, sur des échelles de temps évolutives courtes (de l'ordre de $1/s$ génération, où s est l'avantage sélectif relatif d'un allèle sur les autres, typiquement entre $1/10$ et $1/1000$). Pour équilibrer l'effet de la sélection naturelle et atteindre des diversités génétiques telles qu'elles sont mesurées, il faudrait des taux de mutation importants (du même ordre que s). Or, ceux-ci seraient beaucoup trop importants pour être compatibles avec les taux de mutation mesurés dans les espèces vivantes. *A contrario*, la dérive génétique purge la

variation sur des échelles de temps beaucoup plus longues, de l'ordre de N générations. Aussi, le taux de mutation nécessaire pour équilibrer l'effet de la dérive doit être de l'ordre de $1/N$ pour générer de la diversité. Ces taux de mutation sont pour le coup tout à fait compatibles avec ceux mesurés expérimentalement. À l'époque où les projets de séquençage de génomes complets fleurissent partout et concernent toutes les branches du vivant, il est temps de reposer la question : le modèle standard neutre prédit-il correctement la diversité génétique mesurée au sein des espèces ?

Du modèle à la population réelle

Revenons sur l'idée de *taille efficace* et présentons plus précisément une des méthodes d'estimation fondée sur la diversité observée. Le panneau de gauche de la Figure 5 dépeint l'abstraction Wright-Fisher censée représenter le monde. Dans cette population-modèle, deux chromosomes homologues, échantillonnés au hasard, sont séparés en moyenne par $2N\mu$ événements de mutations et donc, en négligeant les sites mutés plusieurs fois, présentent autant de différences génétiques. Sur le panneau de droite est figurée une « vraie » population. Comme les processus évolutifs sont très lents, il est en général très difficile de les observer. En conséquence, on ne connaît pas leur importance relative. On peut cependant formuler l'hypothèse raisonnable que les vraies populations ont connu des variations de taille (il n'y a donc pas de constance de la taille), sont probablement structurées en sous-populations (il n'y a donc pas de panmixie), sont soumises aux effets de la sélection naturelle (il n'y a donc pas de neutralité), présentent des générations chevauchantes et une forte variance dans le nombre de descendants (supérieure à la moyenne), etc. Tous les ingrédients indispensables aux prédictions du modèle de Wright-Fisher sont probablement faux.

**Figure 5** – Taille efficace

La taille efficace, notée N_e , d'une population est toujours estimée par comparaison avec le modèle de Wright-Fisher. Un des estimateurs de la taille efficace d'une population est obtenu en divisant la diversité de la population (ici, mesurée par H le nombre moyen de différences entre deux séquences de l'échantillon) par le taux de mutation. Si les mutations sont effectivement neutres, cet estimateur nous renseigne sur le temps auquel vivait l'ancêtre génétique moyen de la population, noté t_2 .

Il est cependant aisé, pour cette population réelle, de mesurer sa diversité génétique à partir des séquences de génomes. Il est alors trivial de comparer la prédiction du modèle de Wright-Fisher ($2N\mu$) avec la diversité mesurée. Pourvu que le taux de mutation ait été mesuré expérimentalement (même sans grande précision), il est très simple d'estimer N en divisant la diversité par le taux de mutation. Il s'agit donc bien de trouver la taille d'une population idéalisée Wright-Fisher qui produirait la même diversité à l'équilibre que celle qui est observée. C'est l'une des méthodes permettant d'estimer la taille efficace d'une population à partir de données génétiques. On peut noter que, si l'on fait l'hypothèse que la grande majorité des mutations sont neutres, cette méthode permet d'estimer le

nombre moyen de générations dans le passé où vivait l'ancêtre génétique commun de deux génomes présents. Si les prédictions du modèle standard neutre sont correctes, cette taille efficace devrait être approximativement égale à la taille de la population. Or, ce n'est pas du tout le cas.

Évaluer la diversité au sein des espèces

Cette quête de mesurer la diversité moléculaire a débuté par l'étude pionnière qui mesura, pour la première fois, la « variabilité de charge » de 9 enzymes au sein de l'espèce *Drosophila pseudoobscura* (Lewontin & Hubby, 1966). En mesurant les variants possédant une charge différente, l'ambition était d'estimer la diversité moléculaire dans une population naturelle. Mesurer le polymorphisme dans les espèces naturelles est ensuite devenu un champ d'investigation très en vogue pour les généticiens des populations. D'abord à l'aide des variations de charges dans les protéines, puis dans les séquences protéiques, dans les nucléotides des gènes, et aujourd'hui à l'échelle des génomes complets. Des centaines, puis des milliers d'espèces ont ainsi été séquencées. En compilant des données de la littérature, Michael Lynch et John S. Conery (2003) ont listé des estimations du produit $N_e\mu$ (taille efficace multipliée par le taux de mutation) pour une quarantaine d'espèces. D'autres catalogues d'estimateurs de la diversité sont publiés régulièrement (Leffler *et al.*, 2012 ; Buffalo, 2021).

Pour certaines espèces, les génomes de dizaines, voire de centaines de milliers d'individus (pour les humains) ont été séquencés puis comparés entre eux. Les projets de séquençage ont fleuri au point que l'on ambitionne de séquencer tous les animaux et plantes de la planète. Cette tâche paraît ardue, puisqu'on estime à une dizaine de millions le nombre total d'espèces de multicellulaires (Mora *et al.*, 2011), dont seule une dizaine de pour cent est décrite par les taxonomistes.

Mais l'abondance des données ne change pas les premières observations qui suggéraient déjà que la diversité moléculaire observée n'était pas proportionnelle au nombre d'individus vivant au sein de l'espèce. Peu enclins à embrasser la vision de la théorie neutraliste, les sélectionnistes cherchent ses failles et exhibent les incohérences de ses prédictions. Dans son ouvrage sur la génétique évolutive, Richard C. Lewontin (1974) met

en avant l'inadéquation entre la diversité prédite par la théorie neutraliste, qui devrait être linéaire avec le nombre d'individus, et la diversité moléculaire observée. Cette incohérence a été postérieurement nommée « paradoxe de Lewontin ».

Voici quelques nombres qui permettent de mesurer l'ampleur de l'inadéquation entre la taille efficace inférée à partir de la diversité observée et la taille de la population.

Les humains sont approximativement 7-8 milliards, que l'on arrondira à 10^{10} , mais la taille efficace estimée par la diversité est d'environ 10^4 (Takahata, 1993). Il y a donc pour les humains une réduction entre N et N_e d'un facteur 10^6 , environ. Les gorilles sont environ 10^5 , mais leur taille efficace est de 10^4 (Yu *et al.*, 2004). Un facteur 10, beaucoup plus raisonnable. La fameuse espèce de mouche à vinaigre *Drosophila melanogaster* possède une taille efficace pour l'ensemble de son espèce de 10^6 (Kreitman, 1983). Il est difficile de donner le nombre total d'individus qui peuplent la Terre pour cette espèce de petite mouche, mais c'est probablement une dizaine d'ordres de grandeur au-dessus. Si l'on regarde le monde bactérien, les nombres donnent encore plus le vertige. Pour l'espèce *Escherichia coli*, on estime la taille efficace à 10^8 (Hartl & Dykhuizen, 1984), mais le nombre total d'individus avoisine les 10^{20} (Olivier Tenaillon, communication personnelle). Un facteur 10^{12} , c'est beaucoup. Pour les virus, c'est le même patron. Par exemple, la taille efficace de *Human Immunodeficiency Virus* (HIV) au sein d'un unique patient est estimée à 10^3 (Achaz *et al.*, 2004), alors que le nombre de virions est 10^{10} , soit une différence d'un facteur de 10 millions.

Dans la plupart de ces exemples, le ratio entre le nombre d'individus réel et la taille efficace de la population est de plusieurs ordres de grandeur, allant jusqu'à des facteurs de 10^{12} pour le cas de bactérie *E. coli*. Pour faire fonctionner les logiciels modernes utilisés en routine pour les analyses populationnelles et fondés sur le modèle standard neutre, il faut remplacer les tailles réelles par les tailles efficaces. C'est gênant. Les causes de ces différences astronomiques, illustrant magnifiquement le paradoxe de Lewontin, restent encore à explorer.

Plusieurs pistes semblent *a priori* possibles. La première, invoquée pour le cas des humains, est la variation de taille des populations. À l'échelle

de N_e générations, soit quelques dizaines de milliers chez les humains, c'est-à-dire plusieurs centaines de milliers d'années, le nombre d'humains total aurait avoisiné cette petite taille : 10^4 humains. Cependant, l'omniprésence de ce patron répété de petites tailles efficaces signifierait-il que toutes les espèces actuelles sont issues de croissances fortes récentes ou qu'elles aient traversé des goulots d'étranglement avant N_e générations en remontant dans le passé ?

Bien que ce scénario soit possible, d'autres sont envisageables, et il convient maintenant de réexaminer l'influence de la sélection naturelle. Les processus sélectifs ont tendance à répandre rapidement dans les populations les allèles associés à un avantage sélectif. L'apparition par mutation d'un allèle bénéfique, suivie de sa sélection positive, entraîne ainsi une diminution drastique de la diversité au locus sélectionné. De plus, ce balayage sélectif entraîne également une diminution de la diversité sur les régions génomiques avoisinantes. C'est l'idée d'« autostop génétique ».

L'aspiration génétique (genetic draft)

L'autostop génétique est l'effet indirect d'un variant sélectionné à un locus sur ses locus avoisinants. Au locus sélectionné apparaît dans la population un variant bénéfique, soit par mutation, soit par migration. Puis, poussé par la sélection naturelle, le variant envahit rapidement la population. En l'absence de recombinaison, l'entière du chromosome dans lequel est apparue cette variation bénéfique est entraînée par cette vague invasive, remplaçant tous les autres chromosomes homologues dans la population. Ainsi, sans recombinaison, ce n'est pas uniquement le locus sélectionné qui se retrouve dépourvu de variations après la fixation, mais tout le matériel génétique physiquement lié au variant bénéfique. Cet effet disparaît cependant rapidement si la recombinaison est fréquente, car cette dernière rend indépendants les variants des différents locus. Cet effet d'entraînement, qui décroît avec la distance au locus sélectionné, est nommé « autostop génétique » (Smith & Haigh, 1974).

Après chaque adaptation génétique, une large région avoisinant le variant sélectionné positivement se retrouve appauvrie en diversité. L'effet d'un seul variant sélectionné n'est donc pas restreint au locus sélectionné,

mais affecte une région génomique bien plus large pour laquelle les événements de recombinaison restent rares.

Dans un régime d'*aspiration génétique* (*genetic draft*) (Gillespie, 2000), l'effet est magnifié par l'arrivée incessante de nouveaux variants bénéfiques. Cela se produit d'autant plus qu'une marche adaptative n'est pas causée par la fixation d'une unique variation bénéfique, mais correspond plutôt à une séquence d'événements de fixation dispersés dans le génome. De plus, il faut rappeler que l'adaptation continue d'un organisme correspond probablement à de multiples marches adaptatives affectant une grande quantité de traits permettant la survie à un environnement sans cesse renouvelé. En anglais, le terme de *genetic draft* fait un clin d'œil à celui de *genetic drift*, utilisé pour la dérive génétique, mais aussi à l'effet d'aspiration créé par le premier coureur sur ceux qui le suivent. Ici, les coureurs sont les variations neutres avoisinantes, « aspirées » à moindre effort par les variations sélectionnées positivement.

Ainsi, la plupart des régions qui ne sont pas soumises à l'effet direct de la sélection naturelle se retrouvent régulièrement purgées de leur diversité neutre par des effets indirects liés à la sélection naturelle des locus avoisinants. Pour que cet effet puisse expliquer une réduction de diversité par rapport au modèle neutre, il faut simplement qu'il se produise plus fréquemment que sur l'échelle temporelle très lente de la dérive génétique (N générations). Par exemple, chez les humains actuels, il suffit que l'adaptation soit actuellement plus fréquente que l'échelle de 10^{10} générations, soit environ 200 milliards d'années. Il est très probable que ce soit le cas dans la plupart des espèces dont les tailles de populations sont grandes, voire très grandes.

Ces modèles d'aspiration génétique sont particulièrement séduisants pour expliquer la faible diversité d'espèces ayant un très grand nombre d'individus, tels que les microorganismes. Pourquoi y a-t-il une si petite diversité dans les génomes d'*E. coli*, alors qu'il y a un si grand nombre d'individus ? L'effet de la sélection naturelle doit simplement se produire dans une échelle de temps de N_e (10^8) générations pour un site donné et surpasser l'effet de la dérive qui se produit à l'échelle de 10^{20} générations. L'hypothèse est séduisante.

À l'exception des espèces ayant un très petit nombre d'individus, une adaptation continue et régulière suffit à éroder la diversité génétique. Il n'y a nul besoin de se produire très fortement au cours des temps évolutifs. Dans ce régime d'aspiration génétique, comme dans la plupart des modèles populationnels, il n'existe à chaque génération qu'une unique lignée fondatrice de toute la population dans le futur. Toutes les mutations portées par cette lignée (quelques mutations bénéfiques et une grande quantité de mutation presque neutres) correspondent à des variations qui seront fixées au cours des temps longs. Si la proportion de mutations (presque) neutres est majoritaire, il n'y a presque aucune distorsion de l'horloge moléculaire par rapport à la dérive génétique. Ainsi, l'observation empirique d'une horloge moléculaire dans les données n'est pas incompatible avec des effets de liaison de rares variants sous sélection. Intéressant.

Conclusion

À travers ce chapitre, nous avons pu aborder la place occupée par la sélection naturelle dans le domaine de l'évolution moléculaire. Nous avons illustré comment, depuis les années 1920 où elle occupait la place de processus fondamental et causal en génétique évolutive, sa position a brusquement reculé pour presque s'effacer au profit de la dérive génétique au début des années 1970. Les trois arguments forts de la théorie neutraliste sont la constance de l'horloge moléculaire, le coût de la sélection naturelle et la grande diversité des organismes. Ces trois arguments réfutent légitimement le modèle pansélectionniste où toutes les variations sont soumises à sélection. Néanmoins, nous avons également montré que le modèle standard de dérive génétique n'est pas compatible avec la diversité génétique observée. Nous avons ensuite montré comment quelques événements sélectifs sont suffisants pour réduire la diversité sur de grandes régions génomiques. On assiste actuellement à un foisonnement d'applications de ces modèles expliquant les effets indirects de la sélection. Ils relancent les débats entre sélectionnistes et neutralistes (Kern & Hahn, 2018 ; Jensen *et al.*, 2019). L'apogée et l'omniscience de la dérive génétique comme explication causale ultime de la diversité moléculaire pourraient bien appartenir au passé. Elle fut

une étape majeure dans notre compréhension de la variation et permet d'aborder l'analyse des données génomiques sereinement. Mais elle n'explique pas pourquoi la diversité au sein des espèces est si faible au vu du nombre d'individus qui les composent. L'explication définitive reste cependant à débattre. L'histoire tranchera.

Références bibliographiques

- ACHAZ Guillaume, PALMER Sarah, KEARNEY Mary, MALDARELLI Frank, MELLORS John W., COFFIN John M., WAKELEY John, 2004. « A Robust Measure of HIV-1 Population Turnover Within Chronically Infected Individuals », *Molecular Biology and Evolution*, 21 (10), p. 1902-1912, <https://doi.org/10.1093/molbev/msh196>.
- BOVERI Theodor, 1904. *Ergebnisse über die Konstitution der chromatischen Substanz des Zellkerns*, Jena, G. Fischer, <https://doi.org/10.5962/bhl.title.28064>.
- BUFFALO Vince, 2021. « Quantifying the Relationship between Genetic Diversity and Population Size Suggests Natural Selection Cannot Explain Lewontin's Paradox », *eLife*, 10, e67509, <https://doi.org/10.7554/eLife.67509>.
- CARROLL Lewis, 1865. *Alice's Adventures in Wonderland*, London, Macmillan and Co.
- DARWIN Charles, 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, London, John Murray.
- FAN Shaohua, HANSEN Matthew E. B., LO Yancy, TISHKOFF Sarah A., 2016. « Going Global by Adapting Local: A Review of Recent Human Adaptation », *Science (New York, N.Y.)*, 354 (6308), p. 54-59, <https://doi.org/10.1126/science.aaf5098>.
- FISHER Ronald Aylmer, 1919. « XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance », *Earth and Environmental Science Transactions of The Royal Society of Edinburgh*, 52 (2), p. 399-433, <https://doi.org/10.1017/S0080456800012163>.
- FISHER Ronald Aylmer, 1930. *The Genetical Theory of Natural Selection*, Oxford, Clarendon Press.
- GALTON Francis, 1889. *Natural Inheritance*, London/New York, Macmillan and Co.

- GILLESPIE John H., 2000. « Genetic Drift in an Infinite Population: The Pseudohitchhiking Model », *Genetics*, 155 (2), p. 909-919, <https://doi.org/10.1093/genetics/155.2.909>.
- HALDANE John Burdon Sanderson, 1927. « A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation », *Mathematical Proceedings of the Cambridge Philosophical Society*, 23 (7), p. 838-844, <https://doi.org/10.1017/S0305004100015644>.
- HALDANE John Burdon Sanderson, 1932. *The Causes of Evolution*, London/ New York/Toronto, Longmans Green & Co.
- HALDANE John Burdon Sanderson, 1957. « The Cost of Natural Selection », *Journal of Genetics*, 55, p. 511-524, <https://doi.org/10.1007/BF02984069>.
- HARDY Godfrey Harold, 1908. « Mendelian Proportions in a Mixed Population », *Science (New York, N.Y.)*, 28 (706), p. 49-50, <https://doi.org/10.1126/science.28.706.49>.
- HARTL Daniel L., DYKHUIZEN Daniel E., 1984. « The Population Genetics of *Escherichia Coli* », *Annual Review of Genetics*, 18, p. 31-68, <https://doi.org/10.1146/annurev.ge.18.120184.000335>.
- INGRAM Catherine J. E., MULCARE Charlotte A., ITAN Yuval, THOMAS Mark G., SWALLOW Dallas M., 2009. « Lactose Digestion and the Evolutionary Genetics of Lactase Persistence », *Human Genetics*, 124, p. 579-591, <https://doi.org/10.1007/s00439-008-0593-6>.
- ISHIDA Yoichi, ROSALES Alirio, 2020. « The Origins of the Stochastic Theory of Population Genetics: The Wright-Fisher Model », *Studies in History and Philosophy of Science. Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 79, 101226, <https://doi.org/10.1016/j.shpsc.2019.101226>.
- JENSEN Jeffrey D., PAYSEUR Bret A., STEPHAN Wolfgang, AQUADRO Charles F., LYNCH Michael, CHARLESWORTH Deborah, CHARLESWORTH Brian, 2019. « The Importance of the Neutral Theory in 1968 and 50 Years On: A response to Kern and Hahn 2018 », *Evolution: International Journal of Organic Evolution*, 73 (1), p. 111-114, <https://doi.org/10.1111/evo.13650>.
- KERN Andrew D., HAHN Matthew W., 2018. « The Neutral Theory in Light of Natural Selection », *Molecular Biology and Evolution*, 35 (6), p. 1366-1371, <https://doi.org/10.1093/molbev/msy092>.
- KIMURA Motoo, 1968. « Evolutionary Rate at the Molecular Level », *Nature*, 217, p. 624-626, <https://doi.org/10.1038/217624a0>.

- KIMURA Motoo, 1983. *The Neutral Theory of Molecular Evolution*, Cambridge, Cambridge University Press.
- KING Jack Lester, JUKES Thomas H., 1969. « Non-Darwinian Evolution », *Science (New York, N.Y.)*, 164, p. 788-798, <https://doi.org/10.1126/science.164.3881.788>.
- KINGMAN John Frank Charles, 1982. « The Coalescent », *Stochastic Processes and their Applications*, 13 (3), p. 235-248, [https://doi.org/10.1016/0304-4149\(82\)90011-4](https://doi.org/10.1016/0304-4149(82)90011-4).
- KREITMAN Martin, 1983. « Nucleotide Polymorphism at the Alcohol Dehydrogenase Locus of *Drosophila Melanogaster* », *Nature*, 304, p. 412-417, <https://doi.org/10.1038/304412a0>.
- LEFFLER Ellen. M, BULLAUGHEY Kevin, MATUTE Daniel R., MEYER Wynn K., SÉGUREL Laure, VENKAT Aarti, ANDOLFATTO Peter, PRZEWORSKI Molly, 2012. « Revisiting an Old Riddle: What Determines Genetic Diversity Levels within Species? », *PLoS Biology*, 10 (9), e1001388, <https://doi.org/10.1371/journal.pbio.1001388>.
- LEWONTIN Richard C., 1974. *The Genetic Basis of Evolutionary Change*, New York, Columbia University Press.
- LEWONTIN Richard C., HUBBY John L., 1966. « A 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 (2), p. 595-609, <https://doi.org/10.1093/genetics/54.2.595>.
- LYNCH Michael, CONERY John S., 2003. « The Origins of Genome Complexity », *Science (New York, N.Y.)*, 302 (5649), p. 1401-1404, <https://doi.org/10.1126/science.1089370>.
- MORA Camilo, TITTENSOR Derek P., ADL Sina, SIMPSON Alastair G. B., WORM Boris, 2011. « How Many Species Are There on Earth and in the Ocean? », *PLoS Biology*, 9 (8), e1001127, <https://doi.org/10.1371/journal.pbio.1001127>.
- MUKAI Terumi, 1964. « The Genetic Structure of Natural Populations of *Drosophila Melanogaster*. I. Spontaneous Mutation Rate of Polygenes Controlling Viability », *Genetics*, 50 (1), p. 1-19, <https://doi.org/10.1093/genetics/50.1.1>

- SÉGUREL Laure, BON Céline, 2017. « On the Evolution of Lactase Persistence in Humans », *Annual Review of Genomics and Human Genetics*, 18, p. 297-319, <https://doi.org/10.1146/annurev-genom-091416-035340>.
- SMITH John Maynard, HAIGH John, 1974. « The Hitch-Hiking Effect of a Favourable Gene », *Genetical Research*, 23 (1), p. 23-35, <https://doi.org/10.1017/S0016672300014634>.
- SUTTON Walter S., 1903. « The Chromosomes in Heredity », *The Biological Bulletin*, 4 (5), p. 231–250, <https://doi.org/10.2307/1535741>.
- TAKAHATA Naoyuki, 1993. « Allelic Genealogy and Human Evolution », *Molecular Biology and Evolution*, 10 (1), p. 2-22, <https://doi.org/10.1093/oxfordjournals.molbev.a039995>.
- VAN VALEN Leigh, 1973. « A New Evolutionary Law », *Evolutionary Theory*, 1, p. 1-30.
- WEINBERG Wilhem, 1908. « Über den Nachweis der Vererbung beim Menschen », *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg*, 64, p. 369-382.
- WRIGHT Sewall, 1921. « Systems of Mating. IV. The Effects of Selection », *Genetics*, 6 (2), p. 162-166, <https://doi.org/10.1093/genetics/6.2.162>.
- WRIGHT Sewall, 1931. « Evolution in Mendelian Populations », *Genetics*, 16 (2), p. 97-159, <https://doi.org/10.1093/genetics/16.2.97>.
- WRIGHT Sewall, 1932. « The Roles of Mutation, Inbreeding, Crossbreeding, and Selection in Evolution », in Donald F. Jones (ed), *Proceedings of the Sixth International Congress on Genetics, Ithaca New York, 1932. Vol. 1: Transactions and General Addresses*, Brooklyn, Brooklyn Botanical Garden, p. 355-366.
- YU Ning, JENSEN-SEAMAN Michael I., CHEMNICK Leona, RYDER Oliver, LI Wen-Hsiung, 2004. « Nucleotide Diversity in Gorillas », *Genetics*, 166 (3), p. 1375-1383, <https://doi.org/10.1534/genetics.166.3.1375>.
- ZUCKERKANDL Emile, PAULING Linus, 1962. « Molecular Disease, Evolution, and Genic Heterogeneity », in Michael Kasha & Bernard Pullman (eds), *Horizons in Biochemistry*, New York, Academic Press, p. 189-225.

CHOISIR OU ÊTRE CHOISI

APPROCHES CRITIQUES DE LA SÉLECTION

La sélection est omniprésente : dans la nature comme dans la société. Elle joue un rôle clé dans l'interprétation de l'évolution biologique, à travers la théorie de l'évolution du vivant, mais aussi dans la compréhension du fonctionnement des activités économiques, par le biais de la concurrence et des « lois du marché ». Elle est souvent présentée comme le processus le plus pertinent pour atteindre la meilleure adéquation entre souhaits et possibilités, besoins et ressources. Elle régirait de nombreuses activités sociales et culturelles. L'excellence émergerait alors par la sélection. Le principe de la sélection apparaît ainsi comme une loi organisant à la fois la dynamique du monde vivant et celle des organisations sociales.

Cependant, le discours qui érige la sélection en « loi naturelle », rationnelle et efficace, s'appliquant indistinctement au monde vivant comme au monde social, est aujourd'hui largement questionné.

Connaître la sélection, ce n'est pas seulement la subir : c'est aussi prendre conscience des procédures explicites et implicites qui la produisent, des causes qui la motivent, des justifications qui la légitiment, des effets qu'elle exerce sur les individus et les collectifs. C'est ce chemin vers la connaissance que les diverses perspectives réunies dans cet ouvrage cherchent à faire émerger.

À travers trois grandes parties – sur la nature de la sélection, ses effets sur la production des inégalités, et les tensions entre sélection, organisation et émancipation –, les contributions croisent les regards de la biologie, de l'économie, de la sociologie et de l'informatique pour éclairer les logiques contemporaines de la sélection.