



Séminaire du Centre d'Alembert

Centre Interdisciplinaire d'Étude de l'Évolution des Idées, des Sciences et des Techniques

13 novembre 2018

Université Paris-Sud/Paris-Saclay, Orsay

La sélection dans tous ses états : fonctions, processus, conséquences

La sélection semble partout, dans la nature comme dans la société. Elle est souvent présentée comme l'outil idéal pour obtenir la meilleure adéquation entre souhaits et possibilités, besoins et ressources. C'est par la sélection qu'émergerait l'excellence. Est-ce le mode de fonctionnement optimisé de toutes les organisations ou un mode de gestion en situation de pénurie ? S'agit-il d'un processus rationnel pour obtenir des résultats interprétables ou d'une contingence ayant modelé l'évolution des espèces ? Comment fonctionne la sélection et existe-t-il des alternatives ? Nous réfléchirons sur les critères et les méthodes, qu'ils soient automatisés ou non, et au-delà nous interrogerons l'impact de la sélection sur le fonctionnement de nos disciplines scientifiques, sur l'établissement des normes, et sur l'organisation de nos sociétés. Pour cela, le Centre d'Alembert fera intervenir des collègues de différents domaines dans le cadre de ce séminaire : science de l'éducation, sciences de la vie, sciences et techniques des activités physiques et sportives, économie, informatique, physique...

La sélection naturelle et évolution moléculaire : histoire, statut et perspectives

Auillaume ACHAZ

Enseignant-chercheur UPMC, Muséum national d'Histoire

Résumé

Bien que la sélection naturelle fut à l'origine de la théorie de l'évolution, sa place fut disputée lors du débat opposant les sélectionnistes aux neutralistes dans les années 70-80. Démise de sa place de modèle de référence par la dérive génétique, elle fut reléguée au second plan à cause de son manque à expliquer la diversité observée au sein des espèces. Cependant, plusieurs observations nouvelles suggèrent qu'il est aujourd'hui nécessaire de rediscuter le dogme établi pour appréhender sereinement les véritables causes de l'évolution moléculaire et comprendre la place de la sélection naturelle face à la diversité génétique au sein des espèces.

TABLE DES MATIÈRES

1	La sélection naturelle en évolution moléculaire.....	4
1.1	L'évolution moléculaire	4
1.2	Petit aperçu historique concernant la génétique des populations.....	5
1.2.1	Quelques définitions et principes	5
1.2.2	Balayage historique.....	6
2	Adaptation	8
3	La synthèse moderne	9
3.1	Déplacement de paradigme - A paradigm shift	10
3.2	Motoo Kimura	11
3.3	Le modèle standard Neutre - Standard Neutral Model.....	11
3.4	Le modèle de Wright-Fischer	12
3.5	Le processus de fixation - The fixation process.....	13
3.6	Le processus de coalescence - The coalescent process.....	13
3.7	Le paradigme actuel - The current paradigm.....	14
3.8	Diversité génétique - Genome-wide = expected diversity	14
4	Est-ce que ce modèle neutre rend compte de la diversité entre les individus ?	16
4.1	Du modèle à la population réelle.....	17
4.2	Evaluation de la diversité des espèces - Assessing species diversity	17
4.3	The genetic draft	19

Je vais commencer par vous parler de sélection naturelle. Je vais essayer de parler plutôt du cadre théorique dans lequel on travaille en biologie évolutive, particulièrement dans le domaine de ce qu'on appelle « génétique des populations ». Je vais essayer de redéfinir tous les termes [sans] trop me perdre dans les détails techniques. Si jamais, un point n'était pas clair, bien sûr [vous pouvez] lever la main et on peut rediscuter, il n'y a pas de problème, on avancera tranquillement.

[Temps = 00 minutes 53 secondes]

1 La sélection naturelle en évolution moléculaire

[Diapo 1]

Je ne suis pas extrêmement légitime pour vous parler de la vie des animaux ou des bêtes puisqu'en fait, je ne travaille pas directement [à leur contact]. Je fais surtout de l'analyse de données ou de la modélisation, donc ce sera un peu plus abstrait. Mais cela dit, on peut parler du cadre théorique sans aller dans les détails techniques de la théorie, juste pour avoir une idée de la façon dont on approche cette question de sélection naturelle en biologie et en particulier en évolution moléculaire.

Le fil rouge, c'est cette place de la sélection naturelle dans l'évolution moléculaire avec une partie un peu historique, une partie du statut actuel et [la manière dont] on peut se projeter sur ce qui est en train de se passer et comment va évoluer notre façon de voir la sélection naturelle.

D'abord on va [voir] ce qu'est l'évolution moléculaire.

1.1 L'évolution moléculaire

[Diapo 2]

On peut distinguer deux facettes :

D'abord l'étude de l'évolution des molécules en elles-mêmes, par exemple comment évolue le génome au cours de la spéciation de certaines espèces (devient-il plus grand ? Plus petit ? Se casse-t-il en chromosomes ?).

Puis un autre aspect qui est peut-être plus proche de ce dont je vais parler aujourd'hui [à savoir comment] on va utiliser ce matériel génétique comme marqueur de l'évolution des espèces.

[Temps = 2 minutes et 22 secondes]

En fait, l'objet d'étude n'est pas le génome, ce n'est que l'outil qui nous permet d'accéder à notre centre d'intérêt. Classiquement, les [chercheurs] coupent les échelles évolutives en deux : les petites échelles et les grandes échelles. [Pour les] petites microévolutions, ce sont des échelles de temps de l'ordre de quelques

milliers, dizaines de milliers de générations, et l'on s'intéresse surtout à ce qui se passe au sein d'une espèce, [à la façon dont] une espèce évolue sur des échelles de temps raisonnables [que l'on peut observer en laboratoire].

En évolution moléculaire, les [chercheurs] s'intéressent également aux relations qui existent entre les espèces, ce qui nous amène dans des gestions de temps beaucoup plus lointaines dont je ne vais pas du tout parler aujourd'hui.

Je vais plutôt parler de la génétique des populations.
[Diapo 3]

1.2 Petit aperçu historique concernant la génétique des populations

On peut dire qu'elle a à peu près 100 ans puisqu'on peut en dater le début autour de 1920. Aujourd'hui, on est presque en 2020, donc ça fait à peu près 100 ans. Il y a eu plusieurs rebondissements que je vais essayer d'aborder. De façon intéressante, cette discipline est née à une époque où il y avait très peu d'observation et, au niveau moléculaire, il [n'y en] avait [pas] puisqu'on ne savait même pas quel était le support de l'information -en fait, [on] savait déjà [qu'il s'agissait] des chromosomes, mais [on] n'avait aucune idée des séquences d'ADN-.

[Temps = 4 minutes et 9 secondes]

1.2.1 Quelques définitions et principes

Donc [la génétique des populations] a été beaucoup [étudiée à partir] des modèles mathématiques avec deux approches, comme classiquement dans la modélisation.

Une approche dite « déterministe », c'est à dire que si l'on spécifie un état de départ, on peut prédire l'état d'arrivée, et une approche plus « stochastique », -on reviendra sur ce qu'on appelle la « dérive génétique »-. Les données qu'on traite dans cette discipline, typiquement, sont des morceaux de génome qu'on va appeler des « locus » en français.

Ce qu'est un *locus* ne [peut pas être précisément] défini ; [il s'agit] d'un morceau de génome dont on ne connaît pas vraiment la taille, dont on n'a pas envie de spécifier les détails. Ce qui nous intéresse en génétique des populations, en évolution moléculaire, c'est surtout quand le *locus* peut prendre plusieurs formes. C'est à dire, chez différents individus, on étudie un gène qui va avoir différentes formes chez différents individus de la même espèce. Ce sont ce qu'on appelle des *locus homologues* chez différents individus. Le cadre de pensée générale est représenté dans ce petit schéma.

[Diapo 4]

[Temps = 5 minutes et 32 secondes]

On ne s'intéresse pas, comme en génétique standard ou médicale, à un pedigree [spécifique] à quelques individus, on s'intéresse ici au pool génétique qui est porté par l'ensemble des individus d'une population, et là encore, on se garde bien de définir ce qu'est une population ; c'est un ensemble d'individus qui peuvent échanger du matériel génétique entre eux. C'est une approche très réductionniste, on simplifie les organismes, on ne regarde que le matériel génétique qu'ils portent et il peut avoir plusieurs formes.

Dans ce schéma, on a un pool génétique dans lequel il y a certains individus coloriés en noir donc qui porteraient un allèle de couleur noire, certains qui porteraient un jaune et d'autres un blanc. On va essayer de s'intéresser à [la façon dont] va évoluer ce pool génétique au cours du temps. Par exemple ici, dans ce pool génétique, les blancs sont en train de disparaître et les jaunes sont en train de se multiplier. Les forces que l'on met en présence pour étudier cela, typiquement, sont la sélection naturelle dont on va beaucoup reparler, et la mutation. On peut imaginer qu'à certaines étapes de temps apparaissent de nouveaux individus, par exemple un peu plus tard pourrait apparaître un individu bleu, et on va s'intéresser, dans le pool génétique, à ce que va devenir cet individu bleu.

La dérive génétique est le processus stochastique qui est un processus d'échantillonnage -je reviendrai dessus de façon un peu plus détaillée plus tard- et puis bien sûr, on fait une liste, mais ce n'est jamais exhaustif. Les populations naturelles sont structurées, cela veut dire que tout le monde ne se mélange pas avec tout le monde.

En général, on a tendance, avant l'invention de la bicyclette, à se croiser avec les gens qui n'habitent pas trop loin. Alors, les organismes vivants qui se déplacent très peu, ne se dispersent pas beaucoup ; ils se croisent avec ceux qui sont à côté. Cela crée de la structure, de la démographie, les populations peuvent croître, décroître, avoir toute une histoire démographique.

[Temps = 7 minutes et 43 secondes]

1.2.2 Balayage historique

[Diapo 5]

Je vais juste faire un petit balayage historique. Vous m'excuserez d'avance, je vais faire des sauts temporels. L'idée est de broser rapidement la place de la sélection naturelle dans ce cadre de travail.

Ronald Fischer

[Diapo 6]

La génétique des populations est née dans les années 20-30, [avec] trois personnes qu'on [considère] classiquement [comme les] fondateurs de la génétique des populations. L'un d'entre eux est Ronald Fischer que vous connaissez probablement pour ses travaux, soit en biologie, soit en statistiques. [Il fait partie de] l'école des biométriciens : Pierson, [Wright, Fischer, Haldane], etc. Ronald Fischer, lui, a écrit ce livre assez difficile à lire -en anglais très soutenu- en 1930 : *La théorie génétique de la sélection naturelle*. On voit bien la place de la sélection naturelle à cette époque dans la pensée des gens. Cela veut dire très clairement [que] le centre d'intérêt à cette époque-là (1920-1930) est d'expliquer ce que va faire la sélection naturelle dans ces populations. On parle déjà de gènes, de locus, [mais] on ne sait pas vraiment ce que [sont] gènes et locus.

En parallèle, quand on regarde les travaux de Morgan, [on sait] déjà ce que [sont] les chromosomes qui portent l'hérédité mais on ne sait pas ce qu'est une séquence d'ADN.

Donc [diapo 6], on s'intéresse à ces populations d'allèles. Ce que je voulais vous montrer est que le théorème fondamental de la sélection naturelle est : « Que va faire la sélection naturelle sur ces pools génétiques d'individus qui portent des chromosomes ? ».

Il va avoir une approche très méthodique de la question ; les écrits de Fischer sont extrêmement rigoureux. Il a une petite tendance à ne pas donner ses démonstrations donc les gens vont mettre beaucoup de temps, des décennies parfois, à démontrer [à nouveau] que tout ce qu'il a fait était vrai, qu'il n'y a absolument aucune erreur.

C'est parfois un peu trop « trivial » [pour Ronald Fischer de] donner la démonstration. Le problème, c'est que les autres ne sont pas capables de la retrouver [facilement], mais finalement on les a toutes retrouvées et tout va bien. Il ne s'est trompé, nulle part. Fischer est un Anglais de l'école de biométriciens.

[Temps = 10 minutes et 25 secondes]

Sewall Wright

[Diapo 7]

Un autre personnage très important [parmi] les fondateurs de la génétique des populations est un Américain qui [s'est] intéressé [aux] questions de *breeding*, de croisements agronomes : Sewall Wright. Il a une vie intéressante car [elle] couvre quasiment tout l'avènement de la génétique, depuis 1900 jusqu'à une époque récente. Il a été témoin de tout ce qui s'est passé.

Il a [écrit] des livres, mais postérieurs, qui sont plutôt des compilations de travaux théoriques à destination des chercheurs, mais il n'y a pas vraiment de livre, comme

[celui] de Fisher qui veut mettre sur la table un concept en particulier. [Je vais vous] [présenter] [diapo 7] deux papiers très célèbres [qu'il a écrits] :

Le premier [article célèbre est] « *Evolution in Mendelian populations* ». Là encore, la plupart des lignes s'intéressent à l'effet de la sélection naturelle sur ces pools génétiques. Il est aussi le père de ce qu'on appelle les paysages adaptatifs, l'équivalent des paysages énergétiques en physique, c'est à dire qu'un individu est vu comme une configuration de plusieurs gènes qui portent plusieurs formes, par exemple le gène A est de couleur noire, le gène B [correspond à la couleur] bleue, le gène C au jaune et on va s'intéresser à [leur] assemblage, [à] la valeur d'adaptation que ça va pouvoir porter.

Cet [article] a [inspiré de] nombreuses [études] dans la littérature [scientifique] et encore aujourd'hui, les gens travaillent sur les paysages adaptatifs et la sélection naturelle.

Il y a une sous-section de ces chapitres qui parle des processus stochastiques, dans cet article, mais c'est vu comme une curiosité plutôt qu'un moteur principal, [comme] de la nuisance par rapport à la sélection naturelle qui veut optimiser l'adaptation des organismes à leur avènement.

[Temps = 12 minutes et 40 secondes]

John B.S. Haldane

[Diapo 8]

Enfin, le dernier personnage, John B. S. Haldane a écrit un livre : « *Variation Within a Species, Natural Selection, What is Fitness ?* », c'est exactement ce qui intéresse les gens à ce moment-là. Haldane a une carrière intéressante ; il était en Angleterre puis à [mi]-carrière, il est parti [enseigner] en Inde et cette seconde partie de sa carrière a laissé moins de traces. Il fait partie des gens [possédant] un esprit de synthèse très fort, il a amené beaucoup de choses. Encore une fois, quand on regarde tous les écrits de cette époque-là, [on voit que] les gens s'intéressent à [la façon dont] la sélection naturelle, va permettre d'expliquer, par exemple [diapo 6] comment les papillons se « miment » les uns, les autres.

[Temps = 13 minutes et 31 secondes]

2 Adaptation

[Diapo 9]

Qu'est-ce que la sélection naturelle ? Je me suis dit que j'allais faire un petit dessin [diapo 9]. On imagine qu'on s'intéresse à une population d'individus et qu'on regarde la fréquence de certains variants dans ces populations. Dans mon petit dessin du début, [on observe] quelle fraction de la population représentent les noirs, les blancs

ou les jaunes, et ce à quoi on pourrait s'attendre est que si un variant avantageux arrive dans une population, la sélection naturelle va avoir tendance à propager ce variant dans la population. Il va être de très basse fréquence, il va monter à la fréquence de 1 [ce qui veut dire] qu'on a soit un nouveau mutant qui arrive et semble être plus en adéquation avec son environnement et, au fur et à mesure des générations, c'est lui qui va remplacer toute la population résidente.

Donc on [obtient] ces vagues. On voit un peu l'évolution comme des successions de remplacements de variants avantageux par un nouveau variant avantageux, une sorte de course éternelle et une adaptation infinie. [Voici] un dessin très simple [diapo 9] : les variants sont soit positifs, soit négatifs. S'ils sont négatifs, ils sont perdus dans la population. S'ils sont positifs, ils sont propagés dans la population. Mais bien sûr c'est beaucoup plus compliqué, ils [l']ont tout de suite compris.

Il y a d'autres patterns de sélection. Un exemple qu'on aime bien est celui de la sélection dite sexuelle [avec] l'exemple emblématique : la queue du paon. On imagine que la queue du paon n'est pas très pratique pour se cacher des prédateurs, mais elle a la préférence des femelles ; de plus en plus, les femelles préfèrent des queues de paon toujours plus grandes, donc il y a un coût un peu absurde à développer des caractères sexuels qui permettent la parade, mais comme ils sont choisis par les femelles en fonction de leur queue, c'est ce caractère qui envahit petit à petit l'espèce.

Il y a d'autres formes de sélection qu'on dit balancée ; ce sont des formes de sélection qui vont maintenir de la diversité. Puisque ce que vous voyez ici, c'est que [pour] chaque mutation qui arrive, si elle est désavantageuse, elle est purgée hors de la population, si elle est avantageuse, elle est fixée dans la population et à la fin du processus, on a à nouveau perdu la diversité. Donc ces processus de sélection ont tendance à purger la diversité au sein des populations. C'est important.

[Temps = 16 minutes et 00 seconde]

Il existe des formes de sélection qui maintiennent de la diversité. Je n'ai pas trop envie de rentrer dans les détails, mais je veux juste dire que ça a été déjà noté [dès le départ] qu'il existait de la sélection qui permettait de maintenir la diversité.

3 La synthèse moderne

[Diapo 10]

Dans les années 40-60, ce qu'on a appelé la synthèse moderne de l'évolution, -je pense que le terme vient du livre de Huxley, de la même famille que tous les Huxley que vous devez connaître-, a été promue par un certain nombre de gens [aux] backgrounds en évolution très différents : il y a des paléontologues, des botanistes, des généticiens. Ils se sont mis d'accord pour faire une sorte de synthèse des

connaissances à cette époque et, à nouveau, j'ai été regarder cela sur Wikipédia : il n'y a pas de trace de processus stochastique, il n'y a que la sélection naturelle.

[Temps = 17 minutes et 6 secondes]

Donc même dans les années 40-60, ce qui explique le monde, les questions intéressantes, ce sont les questions qui ont trait à cette place de la sélection naturelle : Pourquoi les papillons se ressemblent,...

3.1 Déplacement de paradigme - A paradigm shift

[Diapo 11]

Puis, il s'est passé quelque chose qu'on doit absolument noter. Dans les années 70, on a commencé à réaliser que [lorsqu']on regardait à l'échelle moléculaire, il y avait une diversité très forte entre les individus, une diversité tellement forte qu'on avait du mal à simplement admettre que le processus important était la sélection naturelle. Pourquoi ? Parce que la sélection naturelle a tendance à purger la diversité. Si on observe de la diversité, cela veut dire qu'il y aurait des mutations qui viendraient tout le temps remettre de la diversité dans ce pool génétique, puisque dès que la sélection naturelle agit, on purge la diversité. Les estimations de taux de mutation qu'il fallait [insérer dans les modèles] pour pouvoir expliquer cette diversité devenaient intenables. Donc on s'est dit que probablement les variants qu'on observe ne sont sélectionnés ni positivement, ni négativement, ils sont là, ils sont neutres. D'avoir la couleur verte, jaune, noire ou blanche pour le gène x ne change rien, en fait. Cela permettrait d'avoir de la diversité résidente dans ces populations.

Les tournants, ces articles en général que l'on cite, [datent du] tout début des années 70 et l'artisan le plus fort de cette idée que les variants génétiques qu'on observe sont majoritairement neutres, est un Japonais qui a fait sa thèse aux Etats-Unis, puis qui est revenu au Japon, Motoo Kimura. Il a beaucoup travaillé sur les modèles mathématiques de génétique des populations et sur la place de la sélection en génétique des populations, donc il sait très bien de quoi il parle.

[Temps = 19 minutes et 34 secondes]

Il a passé la deuxième partie de sa carrière, à partir de 70, à essayer de promouvoir cette idée que les variants qu'on observe ne sont pas soumis à sélection. Il a écrit un livre en 1983 qui est un petit peu le chant du cygne, qui est vraiment la fin de sa carrière, où il va [commencer] ses chapitres en disant « Oui je sais bien que la sélection existe, je l'ai bien compris, mais ce n'est pas ce que j'essaie de dire. Ce que j'essaie de montrer, c'est que les variants qu'on observe dans les espèces sont neutres ». Ce n'est pas que la sélection naturelle n'existe pas. Les variants qui sont délétères sont tout de suite purgés de la population, les variants qui sont positifs sont tout de suite fixés dans la population. Donc ce qu'on observe, les différences entre les individus, ne sont ni les délétères, ni les avantageux, ce sont les autres.

3.2 Motoo Kimura

[Diapo 12]

[Temps = 20 minutes et 30 secondes]

Voici son argumentaire et des chapitres qui disent qu'il a très bien compris, que la sélection existe. Il a beaucoup travaillé sur cette question et sait tout à fait de quoi il parle. C'est assez remarquable d'avoir des gens qui ont une carrière mathématiques/développement extrêmement pointue et qui, à moitié de carrière, vont basculer et faire une synthèse plus générale du cadre conceptuel dans lequel on travaille. C'est très rare et c'est pour cela que je pense que c'est important de mentionner l'existence de ce personnage.

Et s'il n'y a pas de sélection naturelle, que se passe-t-il avec ces variants neutres ? Ces variants neutres, comme les variants soumis à sélection, apparaissent par hasard, un jour, dans la population, par mutation. La fréquence de ces variants va dériver dans l'espace des possibles, entre 0 et 1, et va avoir tendance à monter ou descendre.

Ici [Diapo 13], il y a un variant qui est apparu, est monté à presque 20 % et a disparu après. Là un autre variant monte, puis certains de ces variants, par chance, – je vais revenir en détails [sur la manière] dont ça marche -vont passer de fréquences très faibles [jusqu'à] envahir l'ensemble de la population. Il est très important de noter que même quand ils envahissent, l'échelle de temps à laquelle se produit cet envahissement est très lent, ce qui fait qu'on a le temps de les observer à l'état polymorphe, contrairement à ce schéma [diapo 8] où ces temps durant lesquels il y a un polymorphisme sont très courts. Pour voir un variant sélectionné à l'état polymorphe, on a une toute petite fenêtre de tir donc en général on ne les voit pas.

Ici, dans cet autre modèle [Diapo 13], cette fenêtre de temps est très longue, de l'ordre de deux fois la taille de la population qui est de 100000 individus, [donc cela représente un temps de] 2×1000000 individus \times le temps de génération, ce n'est pas du tout la même échelle de temps, donc on peut observer des organismes.

[Temps = 22 minutes 55 secondes]

3.3 Le modèle standard Neutre - Standard Neutral Model

[Diapo 14]

En sélection, ce qui est devenu aujourd'hui le modèle de référence est le modèle de dérive génétique.

Voici [diapo 15] le paradigme dans lequel on travaille. On a des organismes, on [les] réduit dans une approche affreusement réductionniste, où l'on oublie tous les organismes pour ne s'intéresser qu'au matériel génétique.

Il y a une certaine diversité de ce matériel génétique. Qu'est ce qui augmente cette diversité ? Ce sont les nouveaux mutants qui apparaissent, par des erreurs de copies de l'ADN -il y a beaucoup de discussions [sur la provenance] des mutations qui peuvent venir d'ailleurs-. Il y a des variants qui apparaissent. Ce qui purge la diversité, c'est la dérive génétique qui, elle, se situe à une échelle très lente, donc on a le temps de voir les variants.

Qu'est-ce qu'est cette dérive génétique ? C'est important de le préciser.

[Diapo 16]

Dans la description la plus simple de ce modèle, c'est une population de taille fixe, finie, qu'on nomme souvent N , dans laquelle il n'y a pas de sélection, qui a une taille constante et dans laquelle on va faire l'hypothèse qu'il y a une « panmixie stricte » (*strict panmixia*). Cela veut dire que n'importe qui peut descendre de n'importe qui. Il n'y a pas de structuration géographique. On imagine que c'est un grand sac où tout le monde est mélangé et on suit simplement les variants au cours du temps. Comme la population est de taille constante et qu'il n'y a pas de sélection, tous les individus ont la même chance. Donc, en moyenne, chaque individu a un descendant puisqu'il faut que la taille soit constante au cours du temps. Puis, c'est un processus d'échantillonnage qu'on va décrire à l'aide de lois de probabilité.

[Temps = 24 minutes et 54 secondes]

3.4 Le modèle de Wright-Fischer

Je ne vais pas trop rentrer dans les lois de probabilités, par contre je vais [présenter] un petit dessin [Diapo 17] d'une réalisation de ces processus de dérives qui est assez éclairant. On imagine que le temps va de la gauche [vers] la droite et que chaque barre [représente] la population à un temps donné. Là ce sont 7 individus à une génération. A chaque génération, tous les individus meurent et sont remplacés par la génération suivante. On veut avoir comme règle que chaque individu a, en moyenne, un descendant et que tout le monde ait la même chance. Une des façons de spécifier cela [consisterait à dire que c'est « comme si »] chacun des descendants choisissait aléatoirement son parent à la génération d'avant. Or cela ne se passe pas comme cela dans la nature. Ce ne sont pas les enfants qui choisissent leurs parents, mais si on veut appliquer cette règle d'égalité et d'avoir une descendance en moyenne, c'est l'une des façons de les spécifier.

Par exemple, cet individu a pour parent celui-là. A chaque fois que je mets un petit trait, cela [signifie] : c'était mon parent. C'est un modèle où chaque génération succède à une autre génération, donc ce sont des générations non chevauchantes. Tout le monde meurt, est remplacé par une nouvelle génération et on transmet son matériel génétique à la nouvelle génération avec ces règles-là. C'est une vraie

réalisation du processus que j'ai faite et que j'ai redessinée et on peut voir que certains descendants ne laissent pas de descendants.

Si par exemple celui-ci a été colorié en bleu et qu'à la génération d'après il n'a pas de descendants, le bleu [disparaît]. Celui-là, par exemple, va avoir une descendance mais, sur le long terme, a disparu. Ce qu'on peut noter, c'est que finalement il n'y a qu'une seule ligne de descendance qui va persister à l'infini dans ce modèle.

Par exemple, si on s'intéresse à cet individu et qu'on suit toute sa descendance, on va voir qu'au bout d'un moment, tous les individus de ce trait vert descendent de celui-là. C'est un modèle où une seule lignée persiste au cours du temps et cette lignée va fonder l'intégralité de la population au cours du temps. Ce n'est pas déterminé parce qu'on ne peut pas savoir, à l'avance, quel est l'individu qui va fonder l'intégralité de la population, par contre il est garanti qu'il y en a un qui sera le père de tout le monde plus tard.

[Temps = 27 minutes et 36 secondes]

3.5 Le processus de fixation - The fixation process

[Diapo 18]

Pendant des années, on a étudié ce modèle vers l'avant, exactement comme je viens de le faire, c'est à dire on choisit un individu et on suit sa descendance qui envahit la population. Comme dans les schémas que je vous ai montrés : par exemple, ici, il n'y en a qu'un, puis ici, ils sont deux, ils sont 3, passent à 4, on suit la fréquence du nombre de ces individus au cours du temps. Effectivement ils se fixent et se remplacent les uns les autres.

[Temps = 28 minutes et 7 secondes]

3.6 Le processus de coalescence - The coalescent process

[Diapo 19]

Depuis les années 80, on a regardé exactement le même modèle, mais à l'envers, [en échantillonnant] tous les individus, à un moment donné, et en regardant leur ancestralité. C'est le même modèle, mais au lieu de le regarder vers l'avant, on le regarde vers l'arrière.

La littérature de modélisation autour de ce modèle est vraiment extrêmement généreuse et foisonnante. On a calculé beaucoup de choses [à ce sujet] qu'on maîtrise très bien.

[Temps = 28 minutes et 42 secondes]

3.7 Le paradigme actuel - The current paradigm

[Diapo 20]

Et le paradigme dans lequel on travaille aujourd'hui est de dire : « Ce modèle-là représente une certaine forme de réalité, où la nature n'échantillonne pas mais le nombre d'individus est fini, chaque personne, par hasard, donne des descendants ou n'en donne pas et tout cela fait varier les fréquences des allèles au cours du temps ». C'est le cadre théorique dans lequel on travaille. Il n'y a aucune place pour la sélection naturelle ici, puisqu'on a retiré toute idée que les personnes avaient un nombre de descendants différent en fonction de leur couleur de cheveux par exemple.

[Temps = 29 minutes et 44 secondes]

3.8 Diversité génétique - Genome-wide = expected diversity

[Diapo 24]

Sous ce modèle, on peut calculer une sorte de parenté entre les individus, une parenté stochastique. On échantillonne des individus à un temps donné et on peut remonter leur généalogie dans le temps et essayer de voir ce qu'elle est. C'est un processus aléatoire. On peut caractériser la distribution de cet arbre, c'est-à-dire, en moyenne, combien de temps il faut remonter dans le passé pour trouver un ancêtre commun entre deux individus. C'est une question vraiment très typique. De nombreux calculs ont été faits et on peut montrer qu'entre deux chromosomes, pris au hasard dans une population, sous ce modèle, une certaine diversité est attendue et vaut deux fois la taille de la population fois le taux de mutation ($2 \times$ taille de la population \times taux de mutation).

Après, on regarde la diversité qu'il y a le long des génomes. [Diapo 25] On imagine que c'est un chromosome et la diversité qu'on observe le long du chromosome [est représentée sur cette diapositive]. On observe typiquement ce genre de choses : un peu plus de diversité à cet endroit-là, un peu moins à cet endroit-là [donc] une certaine variation, mais à certains endroits, il n'y en a pas.

Puis la diversité revient.

Comment lit-on ce graphique ? Les gens adorent faire ce type d'expériences : on prend [un grand nombre] d'individus, on séquence, on regarde la diversité entre [eux] le long du chromosome et on va chercher ces endroits-là (locus). Pourquoi va-t-on les chercher justement ? Parce que ces endroits sont différents de la ligne de base de diversité qu'il y a le long du génome. Comment lit-on cela ? On imagine que cet endroit n'est justement pas neutre mais a été sélectionné. C'est à dire qu'à un moment donné, -je vais revenir au premier schéma [Diapo 9]- cet endroit a subi un événement sélectif qui a nettoyé toute la diversité. Un variant avantageux est apparu et a retiré toute la diversité.

[Temps = 32 minutes et 00 seconde]

[Il s'agit] de l'histoire normale qu'on [retrouve] partout dans le génome ; ce processus stochastique qui est très lent mais qui a beaucoup de variances qui [font] que parfois on se ressemble plus ou moins. Puis, à certains endroits du génome, un variant bénéfique est apparu et a envahi toute la population. Donc, pour cet endroit précis du génome, il n'y a plus de diversité et c'est ce qui [provoque] la réduction [sur ce graphique].

Pour les endroits intermédiaires entre la ligne de base et l'endroit où il y a eu le balayage sélectif, ils ont été affectés modérément par cet événement de sélection. Modérément pourquoi ? Parce que cet endroit [précis] est à côté d'un endroit qui a été sélectionné et, comme vous le savez, les chromosomes, tous les morceaux, sont physiquement liés, donc il y a une influence. S'il se passe quelque chose à un endroit, cela va influencer tout ce qui est génétiquement lié à cet endroit. [Ce sont] des situations intermédiaires.

Si vous regardez la diversité le long de génomes de n'importe quelle espèce, typiquement, on trouve [ce genre] de choses et les gens [ont eu comme] jeu favori d'aller chasser ces endroits pour dire : "Ce *locus* a été soumis à la sélection". On a placé la sélection naturelle comme une exception à la règle. La règle générale est qu'il n'y a pas de sélection naturelle, mais certains endroits dans le génome sont soumis à sélection naturelle et c'est ceux-là qu'on va aller regarder.

[Temps = 33 minutes et 46 secondes]

Par rapport aux années 20-30, c'est complètement différent ; on a un cadre non sélectif dans lequel, bien sûr, la sélection existe mais joue un rôle mineur dans l'explication de la diversité. Chez les humains les endroits [locus à diversité réduite, et donc une sélection très forte, sont par exemple], les fameux exemples de la lactase qui est une enzyme qui dégrade le lait. Elle est présente chez les nourrissons et jeunes enfants [et leur] permet de digérer le lait maternel.

En général, les adultes chez la plupart des espèces, ne boivent pas de lait. Chez les humains, c'est [également le cas]. Cependant, on a remarqué dans certaines populations dont les populations européennes qu'il [y aurait] eu une mutation avantageuse [permettant] de digérer le lait à l'état adulte. [Ce variant avantageux] a été mis en relation avec la pasteurisation, l'apparition de l'agriculture. Chez les humains, le fait de pouvoir boire le lait des vaches permettait de vivre mieux, même en cas de disette.

Cette histoire est très jolie, mais il y a évidemment des problèmes, notamment lorsqu'on a commencé à séquencer les Gaulois et qu'on a vu qu'ils ne pouvaient pas digérer le lait.

Une autre histoire qu'on aime bien : les systèmes immunitaires.

Typiquement, les systèmes immunitaires sont des endroits particulièrement exposés à réagir aux pathogènes et qui se trouvent être des endroits où il y a une trace de sélection très forte aussi. On peut faire une liste mais [elle] n'est pas très longue. Le

nombre de régions où il y a une dépression de diversité est assez faible dans le génome humain.

[Temps = 35 minutes et 48 secondes]

The Ne(des)illusion

[Diapo 26]

Je voudrais aborder ce concept tout à fait intéressant qu'utilisent les généticiens des populations, dont moi-même, et qu'on appelle « la taille efficace ». Je vais passer un peu de temps à expliquer ce que c'est. Je pense que c'est un exemple tout à fait intéressant de l'absurdité dans laquelle on peut se projeter. On a un cadre théorique de travail qu'on ne peut plus lâcher puisque on a [basé] tout notre travail dessus alors, même s'il est faux, on va essayer de se raccrocher à quelque chose et finalement, on va lui donner du sens. Mais en fait, je pense que ça n'a pas de sens.

[Temps = 36 minutes et 26 secondes]

Let's Pause and Ponder

[Diapo 27]

[Beaucoup] de gens se sont posé cette question : « Peut-on évaluer la pertinence du modèle neutre ? ». Au moment où le modèle neutre a été proposé, il y a eu beaucoup de débats et de gens qui étaient complètement opposés à l'idée que ce modèle neutre allait expliquer la diversité.

[Temps = 36 minutes et 51 secondes]

4 Est-ce que ce modèle neutre rend compte de la diversité entre les individus ?

Puisqu'on a dit que, s'il y avait de la sélection naturelle, il n'y avait pas assez de diversité, il faudrait mettre des taux de mutation trop grands, ce qui n'a pas de sens. On a donc été obligé d'invoquer la dérive génétique. Mais je propose l'exercice inverse. Admettons que la dérive génétique est vraie, quelle est la diversité qu'on devrait obtenir chez les espèces ? Est-elle compatible avec ce qu'on observe ? Pendant longtemps, on a regardé qu'un petit morceau dans le génome : le gène X qu'on observe chez plein d'individus. Le problème est que, sous ce modèle, il y a une très grande variance du processus. Ce qui fait qu'on ne peut pas rejeter ce modèle neutre sur la base d'un seul endroit dans le génome. On n'y arrive pas. Mais depuis quelques années, on séquence des génomes complets, donc on n'a pas qu'un seul endroit dans le génome, on a tout le génome. On a séquencé plein d'humains, plein de bovins, etc., on séquence tout et plus on attend, plus les secondes passent, [de plus en] plus de génomes sont déposés dans les bases. On peut commencer à réinterroger vraiment cette question avec beaucoup plus de données.

[Temps = 38 minutes et 5 secondes]

4.1 Du modèle à la population réelle

[Diapo 28]

Qu'est-ce que cette « taille efficace » ?

[Voici] un modèle qui est ma façon de voir le monde. La population est de taille constante, il n'y a pas de sélection et à une « panmixie » stricte. Si je prends deux chromosomes, au hasard, dans la population, l'échelle de temps pour laquelle je vais trouver un ancêtre commun est à peu près la taille de la population. Vous avez 100000 individus dans la population, il faudrait que je remonte en arrière d'environ 100000 générations pour trouver un ancêtre commun entre deux bouts de chromosomes homologues dans la population.

C'est le monde des modèles et on peut [y] calculer plein de choses. La diversité moyenne entre des individus, c'est deux fois la taille de la population fois le taux de mutation (diversité moyenne = 2 x taille de la population x taux de mutation).

Puis, il y a le monde réel, qui est dehors, celui où on ne sait pas comment ça se passe, mais on aimerait bien savoir. C'est un monde où il y a évidemment de la sélection, de la taille variable, de la structuration, de la démographie, des choses compliquées (vous pouvez parier que ça y est aussi). Toutes ces choses qui s'[y] passent sont complètement différentes du modèle. Ce qu'on va appeler la taille efficace (en tout cas, ce que je vais appeler la taille efficace), c'est le temps qu'il faut pour trouver un ancêtre commun entre deux bouts de chromosomes pris dans une population réelle. Il ne s'agit plus d'un modèle [mais] de la réalité. Bien sûr, je séquence deux bouts de chromosome parmi les humains. Je n'ai pas de photos de ce qui s'est passé il y a quatre cent mille ans, donc je ne sais pas combien de temps il faut remonter, mais par contre, ce que je sais, c'est combien il y a de différences entre ces deux bouts de chromosome. Et comme je sais combien de mutations se produisent à chaque génération, je peux régler mon petit problème ; je n'ai plus qu'une équation à une inconnue, je tourne le problème, je connais la diversité, le taux de mutation, alors je peux estimer le temps qu'il faut pour retrouver un ancêtre commun entre deux individus. Je vais appeler cela la taille efficace. Si ce modèle avait un sens parfait, ces 2 nombres devraient être complètement égaux.

4.2 Evaluation de la diversité des espèces - Assessing species diversity

[Diapo 29]

C'est une question qui se pose depuis longtemps, depuis le [moment] où l'on a commencé à générer des données à l'échelle populationnelle dont les pionniers sont Lewontin et Hubby, en 1966. On a mesuré la diversité dans les populations.

[Temps = 40 minutes et 35 secondes]

Ça continue pour des raisons diverses et variées. En 2003, Lynch and Connery ont estimé la taille efficace fois le taux de mutation.

[Dans cet article,] il y a des bactéries, des eucaryotes vertébrés, des plantes. En 2012 encore un [article de] revue : on estime la diversité chez tout un tas d'espèces. Très vite, il a été noté que la diversité n'augmente pas linéairement avec le nombre d'individus qui existent dans la population. L'une des personnes qui n'était pas complètement emballée par la théorie neutraliste, Lewontin, a tout de suite mis sur la table qu'il y avait peut-être un souci. « Pourquoi la diversité n'augmente-t-elle pas avec le nombre d'individus dans une espèce alors qu'elle [le] devrait si le modèle était vrai ? ».

Voici quelques chiffres [Diapo 30]

Je mets l'homme en premier parce qu'il n'y a pas besoin d'expliquer ce que c'est comme espèce. On est entre 7 et 8 milliards, et si on regarde la taille efficace, le nombre de générations qu'il faut remonter en arrière n'est pas tout à fait de l'ordre du milliard, mais de 10 000.

Si je regarde les gorilles : aujourd'hui ils sont 10^5 mais la taille efficace est [de] 1000. Donc j'ai encore un facteur 100. J'ai des exemples intéressants :

* la drosophile -je ne parle que d'espèces échantillonnées dans la nature-. Il faudrait remonter un million de générations en arrière. Mais combien y a-t-il de drosophiles ? Probablement plus d'un million dans un seul verger en plein été, donc je ne sais pas combien il y en a dans le monde entier, mais vraiment beaucoup.

* Pour une bête qui vit dans l'antarctique, de la famille des étoiles de mer, j'ai calculé la taille efficace [et elle est de] 10. Dix individus dans l'antarctique, ce n'est pas possible.

* De plus en plus intéressant, les bactéries coli [dont] la taille efficace dans la nature est 10^8 . Mais ce qui est intéressant, c'est que dans un millilitre de culture, il y en a déjà dix fois plus. Combien y en a-t-il sur la planète ? Je ne sais pas, mais certainement sans aucune mesure avec ces 10^8 .

* Le HIV [sur lequel] j'ai pas mal travaillé. A l'intérieur d'un patient, on regarde la taille efficace de cette « bête » et on trouve 1000, mais en fait il y a 10^{10} virus. Là [même chose], il y a un facteur de 10 millions à expliquer qu'il manque. Je veux bien qu'on discute d'un facteur de 100, de 10 et aller jusqu'à 1000, mais quand on arrive à 10 millions, c'est un peu difficile à passer sous le tapis, en disant que ça marche, que je n'ai qu'à mettre 1000 au lieu de 10^{10} [pour que] mon modèle marche très bien. Je ne suis pas complètement convaincu.

Alors pourquoi cette taille efficace n'a aucun sens ? Elle n'a pas de relation avec le nombre d'individus. Les neutralistes purs et durs vont dire que c'est [dû] à la démographie, [parce] qu'avant les humains étaient moins nombreux, donc c'est ce

qu'on voit aujourd'hui [dans la « taille efficace » avec un décalage temporel]. Les humains étaient moins nombreux. Mais toutes les autres bêtes du monde étaient-elles également moins nombreuses avant ? Est-ce que toutes les bêtes de toutes les espèces sont toutes en croissance partout ? Je ne le sais pas.

[Temps = 44 minutes et 32 secondes]

4.3 The genetic draft

[Diapo 42]

Une explication possible : [il s'agit] d'un modèle intéressant [qui] remet la sélection naturelle un peu plus au-devant de la scène aujourd'hui. Est-ce que la plupart des variants sont neutres ? C'est à peu près [d'accord]. Les variants qui apparaissent dans les enzymes, les protéines, ont probablement un effet. Tous les variants qui apparaissent dans des séquences non codantes, non fonctionnelles, sont neutres. Je n'ai aucun problème avec cela.

Le modèle marche [de cette façon] : on imagine qu'on ait un bout de génomes neutre qui ne code à rien, qui n'a aucune fonction, mais qui est toujours lié physiquement -le génome est toujours lié linéairement, physiquement, les molécules sont liées à des locus qui sont sous sélection-. Une fois qu'on [en] a pris conscience, on peut se dire que peut-être les événements de sélection qui se passent dans ces gènes, même s'ils sont dispersés un peu partout dans le génome, pourraient affecter la diversité à ces locus neutres par le mécanisme [suivant] : « Je suis uniquement le locus neutre puisque les locus sélectionnés sont transitoirement polymorphes [puis] fixés, [donc] je les perds et ne les vois jamais ». Que va-t-il se passer ? Je suis ce locus neutre et pour tous ceux qui sont très proches du locus sélectionné, imaginons que c'est le dernier variant avantageux qui est apparu à ce moment-là et qui a envahi la population. Si tout le monde a hérité de ce variant avantageux, cela veut dire que tout le monde a pour ancêtre commun ce variant avantageux. Donc tous ceux-là sont directement hérités de ce variant avantageux ; ils ont été liés génétiquement avec lui. Mais comme la liaison génétique n'est pas parfaite, cela vient de ce qu'on appelle la recombinaison, c'est à dire qu'à chaque génération, il y a un petit mélange entre les deux génomes des parents-, ce qui fait qu'il y a un certain lien génétique dans la transmission mais il y a une petite mosaïque qui découple l'histoire de ces deux locus. Donc que se passe-t-il ?

Dernier événement de sélection : tous les locus neutres qui, à cette époque, n'étaient pas encore liés avec celui-là repassent par le variant bénéfique, mais les autres n'en sont pas affectés. Par contre, encore avant, ils étaient potentiellement liés à ce locus sélectionné. Ce qui ferait cette parenté entre les individus n'a rien à voir avec de la dérive génétique, mais pourrait [être dû] à des événements de sélection naturelle, avec un petit peu de recombinaison entre

les deux. Donc les variants sont bien neutres mais ce qui crée la parenté entre les variants serait peut-être la sélection naturelle.

Ce type de modèle s'applique extrêmement bien à tout ce qui est micro-organisme typiquement. Pourquoi y a-t-il une si pauvre diversité dans le *e. coli* alors qu'il y a un si grand nombre d'individus ? Ce sont probablement des explications qui tiennent.

[Temps = 47 minutes et 47 secondes]

[Actuellement,] on assiste à une sorte d'explosion d'applications de ce modèle ou variations sur ce modèle qui est en train de remettre finalement la sélection naturelle au-devant de la scène, même pour expliquer la parenté entre les individus.

Merci beaucoup.

Transcription réalisée par Véronique Luec et Julien Gargani.

CENTRE D'ALEMBERT

Centre Interdisciplinaire d'Étude de l'Évolution des Idées, des Sciences et des Techniques

Bâtiment 407 - 91405 ORSAY Cedex

Tél. : 01.69.15.61.90

Courriel : centre.dalembert@universite-paris-saclay.fr

Web : <http://www.centre-dalembert.universite-paris-saclay.fr>

