

Université de Montréal

**Probabilité et temps de fixation à l'aide de processus
ancestraux**

par

Guillaume Elgbeili

Département de mathématiques et de statistique
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en Statistique

novembre 2013

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

**Probabilité et temps de fixation à l'aide de processus
ancestraux**

présenté par

Guillaume Elgbeili

a été évalué par un jury composé des personnes suivantes :

Alejandro Murua

(président-rapporteur)

Sabin Lessard

(directeur de recherche)

André Giroux

(membre du jury)

Mémoire accepté le:

26 novembre 2013

SOMMAIRE

Ce mémoire analyse l'espérance du temps de fixation conditionnellement à ce qu'elle se produise et la probabilité de fixation d'un nouvel allèle mutant dans des populations soumises à différents phénomènes biologiques en utilisant l'approche des processus ancestraux. Tout d'abord, l'article de Tajima (1990) est analysé et les différentes preuves y étant manquantes ou incomplètes sont détaillées, dans le but de se familiariser avec les calculs du temps de fixation. L'étude de cet article permet aussi de démontrer l'importance du temps de fixation sur certains phénomènes biologiques. Par la suite, l'effet de la sélection naturelle est introduit au modèle. L'article de Mano (2009) cite un résultat intéressant quant à l'espérance du temps de fixation conditionnellement à ce que celle-ci survienne qui utilise une approximation par un processus de diffusion. Une nouvelle méthode utilisant le processus ancestral est présentée afin d'arriver à une bonne approximation de ce résultat. Des simulations sont faites afin de vérifier l'exactitude de la nouvelle approche. Finalement, un modèle soumis à la conversion génique est analysé, puisque ce phénomène, en présence de biais, a un effet similaire à celui de la sélection. Nous obtenons finalement un résultat analytique pour la probabilité de fixation d'un nouveau mutant dans la population. Enfin, des simulations sont faites afin de déterminer la probabilité de fixation ainsi que le temps de fixation conditionnel lorsque les taux sont trop grands pour pouvoir les calculer analytiquement.

Mots-clé : Génétique statistique, Variabilité génétique, Inférence ancestrale, Graphe de sélection ancestral, Temps de fixation, Probabilité de fixation, Conversion génique.

SUMMARY

The expected time for fixation given its occurrence, and the probability of fixation of a new mutant allele in populations subject to various biological phenomena are analyzed using the approach of the ancestral process. First, the paper of Tajima (1990) is analyzed, and the missing or incomplete proofs are fully worked out in this Master thesis in order to familiarize ourselves with calculations of fixation times. Our study of Tajima's paper helps to show the importance of the fixation time in some biological phenomena. Thereafter, we extend the work of Tajima (1990) by introducing the effect of natural selection in the model. Using a diffusion approximation, the work of Mano (2009) provides an interesting result about the expected time of fixation given its occurrence. We derived an alternative method that uses an ancestral process that approximates well Mani's result. Simulations are made to verify the accuracy of the new approach. Finally, one model subject to gene conversion is analyzed, since this phenomenon, in the presence of bias, has a similar effect as selection. We deduce an analytical result for the probability of fixation of a new mutant in the population. Finally, simulations are made to determine the probability of fixation and the time of fixation given its occurrence when rates are too large to be calculated analytically.

Key words : Statistical genetics, Genetic variability, Ancestral inference, Ancestral selection graph, Fixation time, Probability of fixation, Genetic conversion.

TABLE DES MATIÈRES

Sommaire	v
Summary	vii
Liste des figures	xi
Liste des tableaux	xiii
Remerciements	1
Introduction	3
Chapitre 1. Analyse de l'article de Tajima 1990	5
1.1. Convolution de variables indépendantes de loi géométrique.....	9
1.2. Convolution de variables indépendantes de loi exponentielle	12
1.3. Preuves de l'article de Tajima.....	16
1.3.1. Formule de convolution.....	16
1.3.2. Explication de l'approche dans l'article.....	17
1.3.3. Polymorphisme.....	19
1.3.4. Homozygotie.....	20
1.3.5. Polymorphisme au moment de la fixation.....	21
1.3.6. Homozygotie au moment de la fixation.....	22
1.3.7. Polymorphisme au moment connu de la fixation.....	25
Chapitre 2. Introduction de l'effet de sélection et nouvelle approche ..	33
2.1. Nouvelle approche avec un événement de sélection possible.....	35
2.1.1. Espérance totale.....	42
2.1.2. Probabilité totale.....	43
2.1.3. Espérance du temps de fixation conditionnelle.....	44
2.2. Simulations.....	46

Chapitre 3. Temps de fixation avec conversion génique	49
3.1. Introduction à la conversion génique	49
3.2. Fixation d'un nouveau mutant en présence de conversion génique biaisée	50
3.2.1. Modèle du processus	50
3.2.2. Calcul de la probabilité du gamète remplaçant	52
3.2.3. Espérance conditionnelle de la fréquence des gamètes	52
3.2.4. Calcul de la probabilité de fixation	53
3.3. Simulations	59
Chapitre 4. Discussion	67
Bibliographie	69
Annexe A. Résultats complets des simulations sur les effets de la recombinaison, de la conversion et de son biais sur le temps de fixation conditionnel et la probabilité de fixation	A-i

LISTE DES FIGURES

1.1	Figure 2 de l'article de Tajima (1990) : Un exemple de relations généalogiques entre six séquences d'ADN. A_i est le i -ième plus ancien ancêtre de ces six séquences.	7
2.1	Graphe de sélection ancestral sans événement de sélection.	36
2.2	Graphe de sélection ancestral avec un événement de sélection sur une séquence mutante.	38
2.3	Graphe de sélection ancestral avec un événement de sélection sur la séquence non mutante et avec une séquence additionnelle mutante.	39
2.4	Graphe de sélection ancestral avec un événement de sélection sur la séquence non mutante, avec une séquence additionnelle non mutante et sans coalescence entre les deux.	40
2.5	Graphe de sélection ancestral avec un événement de sélection sur la séquence non mutante, avec une séquence additionnelle non mutante et avec coalescence entre les deux.	41
3.1	Représentation des états possibles pour les séquences ancêtres.	56
3.2	Nombre de fixations du mutant sur 1000000 pour $b=0$, avec c et r variables.	60
3.3	Temps de fixation conditionnel moyen du mutant sur 1000000 pour $b = 0$, avec c et r variables.	61
3.4	Nombre de fixations du mutant sur 1000000 pour $c=0,4$, avec b et r variables.	62
3.5	Nombre de fixations du mutant sur 1000000 pour $r=0,5$, avec b et c variables.	63
3.6	Temps de fixation conditionnel moyen du mutant sur 1000000 pour $r = 0,5$, avec b et c variables.	64

3.7	Nombre de fixations du mutant sur 1000000 pour $r=0$, avec b et c variables.	65
3.8	Temps de fixation conditionnel moyen du mutant sur 1000000 pour $r=0$, avec b et c variables.	66

LISTE DES TABLEAUX

2.1	Résultat des simulations sur le temps conditionnel de fixation et la probabilité de fixation en présence de sélection sur 10000000 itérations.....	47
3.1	Probabilité d'association de chaque paire de gamètes.	50
3.2	Effet de la recombinaison par paire.	51
3.3	Effet de la conversion par type de gamète.....	51
A.1	Résultats complets du nombre de fixations du mutant pour chaque valeur des taux de recombinaison, de conversion et de son biais. ...	A-ii
A.2	Résultats complets du temps de fixation conditionnel du mutant pour chaque valeur des taux de recombinaison, de conversion et de son biais.	A-iii

REMERCIEMENTS

Je tiens tout d'abord à remercier mon directeur de recherche Sabin Lessard pour son aide, son support et sa présence sans lesquels je n'aurais jamais pu mener ce mémoire à terme. Son expérience transmise par ses précieux conseils m'ont permis de traverser les impasses, mathématique, académique ou autres, auxquelles j'ai fait face au cours de ces deux importantes années. Je tiens par le fait même à remercier tous les professeurs du DMS, en particulier ceux du programme de statistique, pour m'avoir permis, par leurs connaissances et leur bon moral, à me rendre aussi loin. Finalement, j'aimerais remercier ma famille et mes amis pour leur compréhension lors des périodes plus chargées, ainsi que pour la motivation qu'ils me donnaient, et qui m'ont permis de franchir cette étape importante.

INTRODUCTION

La théorie synthétique de l'évolution, ou le néodarwinisme, est la théorie la plus largement utilisée par les scientifiques en ce qui a trait au processus d'évolution biologique. Il s'agit de l'unification de la théorie darwinienne, de la théorie de l'hérédité mendélienne ainsi que de la génétique des populations. Elle a été apportée par plusieurs chercheurs au cours des années 1930 et 1940. Cette théorie se base sur le fait que l'évolution d'une population s'effectue lorsqu'il y a une modification dans la fréquences des allèles, ou types, d'un gène. Ainsi, la population peut s'adapter à un changement environnemental puisque l'allèle apportant un trait résistant à ce changement est plus propice à être transmis dans la population par leur descendance.

Il est donc naturel de s'intéresser à la probabilité qu'un nouvel allèle mutant, apportant possiblement un trait physiologique alternatif, soit fixé dans la population, c'est-à-dire qu'il devienne le seul allèle présent. Par le fait même, l'espérance du temps pour que la fixation survienne si l'événement a lieu est un autre aspect intéressant. Plusieurs chercheurs actuels s'intéressent à cette question, et aux différents résultats qui en découlent ou qui sont influencés par ceux-ci. Par exemple, les articles de Stephan (1996) et de Piskol et Stephan (2008) démontrent et utilisent le fait que le taux de l'évolution compensatoire, c'est-à-dire le taux avec lequel la paire de nucléotides se complète après qu'une mutation frappe l'un des deux, est donné par l'inverse de l'espérance du temps de fixation d'un nouveau mutant dans la population. Il est donc évident que le temps de fixation est un élément très important à considérer. Dans ce mémoire, on travaillera plus spécialement sur l'article de Tajima (1990) qui étudie l'influence du temps de fixation sur différents phénomènes biologiques.

Cependant, plusieurs facteurs biologiques peuvent influencer le temps de fixation d'un nouvel allèle ainsi que la probabilité qu'elle survienne. L'un des plus importants en recherche est le phénomène de sélection naturelle. Il a été

étudié par plusieurs chercheurs en utilisant des méthodes telles que le processus de diffusion, par exemple dans l'article de Darden, Hudson et Kaplan (1988). Une nouvelle approche utilisant les processus ancestraux a donc été tentée dans le but d'approcher le résultat sur l'espérance du temps de fixation conditionnellement à ce qu'il survienne dans un temps fini, obtenu dans l'article de Mano (2009), et d'en trouver des résultats exacts pour l'effet du premier ordre de la sélection. Plus de détails sur cet article et sur la sélection naturelle seront donnés au début du chapitre les traitant.

Un autre facteur biologique pouvant influencer le temps et la probabilité de fixation est la conversion. Il s'agit d'un phénomène biologique, décrit plus en détail dans le chapitre qui lui est dédié, qui peut se comporter comme la sélection naturelle en présence de biais lors de la conversion. Plusieurs chercheurs au cours des dernières années se sont intéressés à ce phénomène, et ont utilisé la méthode d'approximation par un processus de diffusion afin d'étudier l'effet de la conversion, comme par exemple dans un article très récent de Mano (2013). Dans ce mémoire, une méthode utilisant les processus ancestraux est utilisée afin de déterminer la probabilité de fixation d'un nouveau mutant dans la population.

Chapitre 1

ANALYSE DE L'ARTICLE DE TAJIMA 1990

La première partie du mémoire consiste à analyser l'article de Tajima (1990) afin de se familiariser avec le calcul du temps de fixation d'un nouveau mutant dans une population. Il permet aussi de démontrer que le temps, en particulier le temps de fixation défini au premier paragraphe de la page 6, peut avoir une grande influence sur des phénomènes biologiques, tels le polymorphisme et l'homozygotie, définis à la page 7 de ce mémoire. Puisque le temps de coalescence est une somme de variables aléatoires indépendantes de loi géométrique, il faut tout d'abord calculer la convolution de telles variables ayant des taux différents. De plus, très souvent, l'approximation continue de ces variables est utilisée, soit des variables de loi exponentielle avec des taux différents. La convolution de ces variables est donc elle aussi étudiée.

Voici un résumé de l'article de Tajima (1990) spécifiant les parties qui seront détaillées dans ce chapitre du mémoire. Puisqu'on s'intéresse particulièrement à la comparaison de séquences choisies au hasard dans la population, il est important de définir le concept de séquences homologues. Deux chromosomes sont homologues s'ils ont le même contenu génétique, la même taille ainsi que la même forme. Deux séquences d'ADN sont définies comme homologues si elles se situent au même endroit, autrement dit au même locus, sur deux chromosomes homologues.

Dans l'article de Tajima (1990), on considère une population diploïde de taille N , c'est-à-dire qu'il y a $2N$ séquences d'ADN homologues dans la population, et on suppose un modèle avec une infinité de sites sans recombinaison. Les mutations, survenant à un taux ν par séquence par génération, sont considérées sélectivement neutres. On considère qu'à chaque étape, une des séquences disparaît pour être remplacée par l'une des $2N - 1$ autres séquences

homologues.

On s'intéresse au temps de fixation d'un allèle mutant dans la population, en excluant les cas où le mutant disparaît. Le temps de fixation est défini comme le temps que prend un allèle mutant à se fixer dans la population, c'est-à-dire à devenir le seul allèle présent dans toute la population. On utilise le fait que le temps que prennent $n + 1$ séquences d'ADN choisies aléatoirement dans la population pour provenir de n séquences pour la première fois a comme fonction de densité au temps t en nombre de $4N$ générations

$$f_n(t) = \lambda_n e^{\lambda_n t},$$

avec $\lambda_n = \frac{n(n+1)}{4N}$. En calculant la convolution de $f_{2N-1}(t), f_{2N-2}(t), \dots, f_1(t)$, on trouve $y(t)$, la distribution du temps en nombre de $4N$ générations que prend un mutant à se fixer dans la population, excluant les cas où le mutant disparaît, soit

$$y(t) = \sum_{i=1}^{2N-1} (2i+1)(-1)^{i+1} \prod_{j=1}^i \frac{2N-j}{2N+j} \lambda_i e^{-\lambda_i t},$$

pour tout réel $t > 0$. Cependant, la démarche pour arriver à ce résultat est manquant dans l'article de Tajima (1990). Elle sera donc donnée en détail dans ce mémoire en section 1.3.1.

Tajima (1990) présente par la suite une nouvelle approche pour arriver aux autres résultats de l'article. Considérons la population de $2N$ séquences d'ADN. Si nous prenons deux séquences au hasard parmi celles-ci, elles auront $2N - 1$ ancêtres communs possibles. Ainsi, on définit A_i comme le i -ième plus ancien ancêtre commun possible de deux séquences choisies au hasard dans la population. La figure 1.1 est tirée de la figure 2 de l'article de Tajima (1990). Elle représente la généalogie d'une petite population afin de mieux comprendre la définition de A_i . On définit aussi a_i comme la probabilité que l'ancêtre commun de deux séquences choisies au hasard soit A_i . Cette probabilité est donnée par

$$a_i = \frac{2(2N+1)}{(i+1)(i+2)(2N-1)}.$$

Cependant, la démarche pour arriver à ce résultat est partielle dans l'article de Tajima (1990). Ainsi, elle sera donnée au complet dans la section 1.3.2 de ce mémoire.

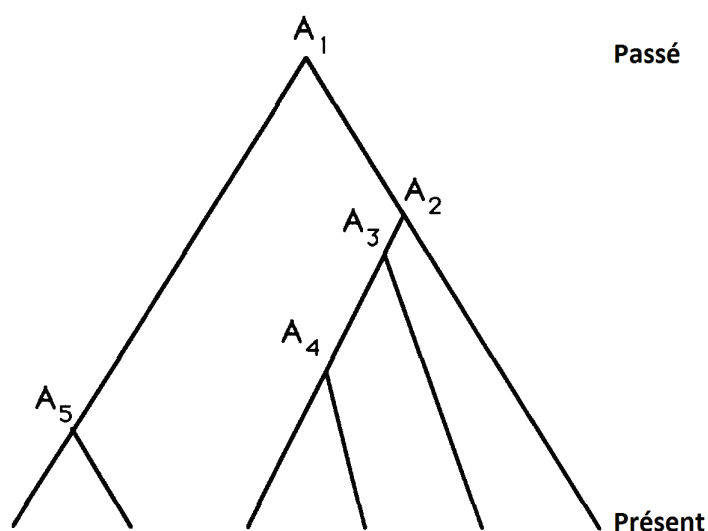


FIGURE 1.1. Figure 2 de l'article de Tajima (1990) : Un exemple de relations généalogiques entre six séquences d'ADN. A_i est le i -ième plus ancien ancêtre de ces six séquences.

Suite à cela, Tajima (1990) s'intéresse au nombre k de sites polymorphes entre deux séquences d'ADN choisies au hasard à un moment quelconque dans le temps. Un site polymorphe est une différence entre deux nucléotides situés au même endroit sur deux séquences d'ADN homologues. Watterson (1975) avait démontré que le nombre espéré de sites polymorphes était donné par

$$\mathbb{E}(k) = M,$$

avec $M = 4Nv$. Tajima (1990) arrive au même résultat avec la nouvelle approche, mais plusieurs étapes de la démarche sont manquantes. Elle sera donc donnée en détail dans ce mémoire à la section 1.3.3.

L'autre aspect important de l'article, qui est étudié par la suite, est l'homozygotie F à un moment quelconque dans le temps, soit le fait que deux séquences d'ADN homologues soient strictement identiques. Kimura et Crow (1964) ont démontré que l'homozygotie espérée, ou la probabilité que deux séquences soient homozygotes, est donnée par

$$\mathbb{E}(F) = \frac{1}{1 + M}.$$

Tajima (1990) arrive au même résultat en utilisant sa nouvelle approche, mais, encore une fois, sans démarche détaillée. La section 1.3.4 du mémoire y est donc dédiée.

Dans la suite de l'article de Tajima (1990), on considère que le mutant parvient à se fixer à un temps T inconnu. On s'intéresse d'abord au nombre espéré de sites polymorphes au moment T de la fixation, soit $\mathbb{E}(k \mid T < \infty)$, noté $\mathbb{E}(k \mid \text{fix})$ dans l'article, dans le but de le comparer avec le nombre espéré de sites polymorphes à un moment quelconque dans le temps. On peut ainsi avoir une idée de l'impact du temps de fixation sur le polymorphisme. On arrive à la formule

$$\mathbb{E}(k \mid \text{fix}) = 2M \left[\frac{\pi^2}{3} - 3 \right].$$

Ainsi, on parvient à déterminer que le nombre de sites polymorphes est environ 42% inférieur au moment inconnu de la fixation d'un nouveau mutant qu'à tout autre moment quelconque dans le temps. Cependant, encore une fois, la démarche pour y arriver est en grande partie absente dans l'article de Tajima (1990). Elle sera donc donnée en détail à la section 1.3.5.

Tajima (1990) s'intéresse par la suite à la probabilité que deux séquences choisies au hasard au moment inconnu T de la fixation soient homozygotes, soit $\mathbb{E}(F \mid \text{fix})$. On peut ainsi comparer cette valeur à la probabilité pour deux séquences choisies à un moment quelconque dans le temps, et ainsi déterminer l'impact du temps de fixation d'un nouveau mutant sur l'homozygotie. On arrive à la formule

$$\frac{1}{M} \left(1 - e^{-M+c_1M^2-c_2M^3} \right),$$

avec $c_1 = \frac{4\pi^2-39}{6}$ et $c_2 = \frac{79-8\pi^2}{3}$. On parvient donc à déterminer que la probabilité que deux séquences d'ADN choisies au hasard dans la population au moment de la fixation sera toujours plus grande qu'à un autre moment quelconque dans le temps, mais que cette différence dépend de la taille de M . La section 1.3.6 est dédiée à la démarche pour arriver à la formule proposée.

Finalement, Tajima (1990) considère que le temps de fixation T est connu. Il s'intéresse alors au nombre espéré de sites polymorphes au moment connu de la fixation, noté $\mathbb{E}(k \mid T)$. Il détermine que, même si la fixation se fait à un moment très éloigné dans le temps, le nombre espéré de sites polymorphes au moment de la fixation sera environ 33% inférieur qu'à tout autre moment dans le temps. La démarche pour y arriver est détaillée à la section 1.3.7.

1.1. CONVOLUTION DE VARIABLES INDÉPENDANTES DE LOI GÉOMÉTRIQUE

Bien que l'article de Tajima (1990) contienne seulement des variables de loi exponentielle, nous avons cru important de déterminer la formule générale de la convolution de variables indépendantes de loi géométrique, car le temps de coalescence exact est représenté par cette somme. Ainsi, soient X_1, \dots, X_n , des variables indépendantes de loi géométrique de paramètre P_i respectivement, pour $i = 1, \dots, n$. Alors, la fonction de masse de X_i est donnée par

$$f_i(x) = (1 - P_i)^{x-1} P_i,$$

pour tout entier $x \geq 1$, pour $i = 1, \dots, n$. Afin de démontrer le résultat sur la convolution de ces variables, soit que la fonction de masse de $X_1 + \dots + X_n$ est donnée par

$$f_{1,\dots,n}(k) = \sum_{i=1}^n \prod_{\substack{j=1 \\ j \neq i}}^n \frac{P_j}{P_j - P_i} P_i (1 - P_i)^{k-1} = \prod_{j=1}^n P_j \sum_{i=1}^n \frac{(1 - P_i)^{k-1}}{\prod_{\substack{j=1 \\ j \neq i}}^n (P_j - P_i)},$$

pour tout entier $k \geq n$, nous procéderons par induction.

Débutons donc par le cas où $n = 2$. Dans ce cas, on a

$$\begin{aligned} f_{12}(k) &= \sum_{l=1}^{k-1} f_1(k-l) f_2(l) \\ &= \sum_{l=1}^{k-1} (1 - P_1)^{k-l-1} P_1 (1 - P_2)^{l-1} P_2 \\ &= P_1 P_2 \sum_{l=1}^{k-1} (1 - P_1)^{k-l-1} (1 - P_2)^{l-1} \\ &= \frac{P_1 P_2}{(1 - P_1)(1 - P_2)} (1 - P_1)^k \sum_{l=1}^{k-1} \left(\frac{1 - P_2}{1 - P_1} \right)^l \\ &= \frac{P_1 P_2}{(1 - P_1)(1 - P_2)} (1 - P_1)^k \left(\frac{1 - P_2}{1 - P_1} \right) \sum_{l=0}^{k-2} \left(\frac{1 - P_2}{1 - P_1} \right)^l \\ &= \frac{P_1 P_2}{(1 - P_1)} (1 - P_1)^{k-1} \frac{1 - \left(\frac{1 - P_2}{1 - P_1} \right)^{k-1}}{1 - \left(\frac{1 - P_2}{1 - P_1} \right)} \\ &= \frac{P_1 P_2}{(1 - P_1)} (1 - P_1)^{k-1} \frac{\left(\frac{(1 - P_1)^{k-1} - (1 - P_2)^{k-1}}{(1 - P_1)^{k-1}} \right)}{\left(\frac{1 - P_1 - 1 + P_2}{1 - P_1} \right)} \end{aligned}$$

$$\begin{aligned}
&= P_1 P_2 \frac{(1 - P_1)^{k-1} - (1 - P_2)^{k-1}}{-P_1 + P_2} \\
&= \frac{P_2}{P_2 - P_1} P_1 (1 - P_1)^{k-1} + \frac{P_1}{P_1 - P_2} P_2 (1 - P_2)^{k-1}.
\end{aligned}$$

Analysons maintenant le cas où $n = 3$. On a

$$f_{123}(k) = \sum_{l=1}^{k-1} f_3(k-l) f_{12}(l).$$

Par le même raisonnement que pour le cas précédent, nous arrivons à

$$\begin{aligned}
f_{123}(k) &= \frac{P_2}{P_2 - P_1} \left(\frac{P_3}{P_3 - P_1} P_1 (1 - P_1)^{k-1} + \frac{P_1}{P_1 - P_3} P_3 (1 - P_3)^{k-1} \right) \\
&+ \frac{P_1}{P_1 - P_2} \left(\frac{P_3}{P_3 - P_2} P_2 (1 - P_2)^{k-1} + \frac{P_2}{P_2 - P_3} P_3 (1 - P_3)^{k-1} \right) \\
&= \frac{P_2 P_3}{(P_2 - P_1)(P_3 - P_1)} P_1 (1 - P_1)^{k-1} + \frac{P_1 P_3}{(P_1 - P_2)(P_3 - P_2)} P_2 (1 - P_2)^{k-1} \\
&+ \left(\frac{P_1 P_2}{(P_2 - P_1)(P_1 - P_3)} - \frac{P_1 P_2}{(P_2 - P_1)(P_2 - P_3)} \right) P_3 (1 - P_3)^{k-1} \\
&= \frac{P_2 P_3}{(P_2 - P_1)(P_3 - P_1)} P_1 (1 - P_1)^{k-1} + \frac{P_1 P_3}{(P_1 - P_2)(P_3 - P_2)} P_2 (1 - P_2)^{k-1} \\
&+ \frac{P_1 P_2}{(P_1 - P_3)(P_2 - P_3)} P_3 (1 - P_3)^{k-1} \\
&= \sum_{i=1}^3 \prod_{\substack{j=1 \\ j \neq i}}^3 \frac{P_j}{P_j - P_i} P_i (1 - P_i)^{k-1}.
\end{aligned}$$

Supposons maintenant que

$$f_{1,\dots,n-1}(k) = \sum_{i=1}^{n-1} \prod_{\substack{j=1 \\ j \neq i}}^{n-1} \frac{P_j}{P_j - P_i} P_i (1 - P_i)^{k-1} = \left(\prod_{j=1}^{n-1} P_j \right) \sum_{i=1}^{n-1} \frac{(1 - P_i)^{k-1}}{\prod_{\substack{j=1 \\ j \neq i}}^{n-1} (P_j - P_i)}.$$

On a alors

$$\begin{aligned}
f_{1,\dots,n}(k) &= (f_{1,\dots,n-1} * f_n)(k) \\
&= \left(\prod_{j=1}^n P_j \right) \sum_{i=1}^{n-1} \frac{(1 - P_n)^{k-1} - (1 - P_i)^{k-1}}{(P_i - P_n) \prod_{\substack{j=1 \\ j \neq i}}^{n-1} (P_j - P_i)} \\
&= \left(\prod_{j=1}^n P_j \right) \left(\sum_{i=1}^{n-1} \frac{(1 - P_i)^{k-1}}{\prod_{\substack{j=1 \\ j \neq i}}^n (P_j - P_i)} - \sum_{i=1}^{n-1} \frac{(1 - P_n)^{k-1}}{\prod_{\substack{j=1 \\ j \neq i}}^n (P_j - P_i)} \right).
\end{aligned}$$

Afin d'obtenir la bonne formule, il nous faut donc démontrer que

$$-\sum_{i=1}^{n-1} \frac{1}{\prod_{\substack{j=1 \\ j \neq i}}^n (P_j - P_i)} = \frac{1}{\prod_{j \neq i}^n (P_j - P_i)}.$$

En voici la preuve. On a d'abord

$$\begin{aligned} -\sum_{i=1}^{n-1} \frac{1}{\prod_{\substack{j=1 \\ j \neq i}}^n (P_j - P_i)} &= \frac{1}{\prod_{j \neq i}^n (P_j - P_i)} \\ \Leftrightarrow 0 &= \sum_{i=1}^n \frac{1}{\prod_{\substack{j=1 \\ j \neq i}}^n (P_j - P_i)} \\ \Leftrightarrow 0 &= \sum_{i=1}^n \frac{\prod_{\substack{l,m=1 \\ l \neq m \neq i}}^n (P_l - P_m)}{\prod_{\substack{l,m=1 \\ l \neq m}}^n (P_l - P_m)} \\ \Leftrightarrow 0 &= \sum_{i=1}^n \prod_{\substack{l,m=1 \\ l \neq m \neq i}}^n (P_l - P_m) \\ \Leftrightarrow 0 &= \sum_{i=1}^n \prod_{\substack{l,m=1 \\ i \neq l \neq m \neq i}}^n (P_l - P_m) \prod_{\substack{l,m=1 \\ l=i \neq m}}^n (P_l - P_m) \\ \Leftrightarrow 0 &= \pm \sum_{i=1}^n \prod_{\substack{l,m=1 \\ i \neq l > m \neq i}}^n (P_l - P_m)^2 \prod_{\substack{l,m=1 \\ i=l > m}}^n (P_l - P_m) \prod_{\substack{l,m=1 \\ i=l < m}}^n (P_l - P_m) \\ \Leftrightarrow 0 &= \pm \sum_{i=1}^n \prod_{\substack{l,m=1 \\ i \neq l > m \neq i}}^n (P_l - P_m)^2 \prod_{\substack{l,m=1 \\ i=l > m}}^n (P_l - P_m) \prod_{\substack{l,m=1 \\ l > m=j}}^n (P_l - P_m) (-1)^{n-j} \\ \Leftrightarrow 0 &= \pm \sum_{i=1}^n \prod_{\substack{l,m=1 \\ i \neq l > m \neq i}}^n (P_l - P_m) \prod_{\substack{l,m=1 \\ l > m}}^n (P_l - P_m) (-1)^{n-j} \\ \Leftrightarrow 0 &= \sum_{i=1}^n \prod_{\substack{l,m=1 \\ i \neq l > m \neq i}}^n (P_l - P_m) (-1)^j. \end{aligned}$$

Remarquons que le produit

$$\prod_{\substack{l,m=1 \\ l,m \neq i \\ l > m}}^n (P_l - P_m)$$

est le déterminant de Vandermonde suivant :

$$\begin{vmatrix} 1 & P_1 & P_1^2 & \dots & P_1^{n-2} \\ 1 & P_2 & P_2^2 & \dots & P_2^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & P_{j-1} & P_{j-1}^2 & \dots & P_{j-1}^{n-2} \\ 1 & P_{j+1} & P_{j+1}^2 & \dots & P_{j+1}^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & P_n & P_n^2 & \dots & P_n^{n-2} \end{vmatrix}.$$

On a donc que

$$\sum_{i=1}^n \prod_{\substack{l,m=1 \\ l,m \neq i \\ l>m}}^n (P_l - P_m)(-1)^j$$

est le déterminant suivant :

$$\begin{vmatrix} 1 & 1 & P_1 & P_1^2 & \dots & P_1^{n-2} \\ 1 & 1 & P_2 & P_2^2 & \dots & P_2^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & P_{j-1} & P_{j-1}^2 & \dots & P_{j-1}^{n-2} \\ 1 & 1 & P_{j+1} & P_{j+1}^2 & \dots & P_{j+1}^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & P_n & P_n^2 & \dots & P_n^{n-2} \end{vmatrix},$$

qui est égal à 0, puisqu'il contient deux colonnes identiques.

La preuve est donc complète.

1.2. CONVOLUTION DE VARIABLES INDÉPENDANTES DE LOI EXPONENTIELLE

Bien que la formule de la convolution de variables indépendantes de loi exponentielle ne soit pas présentée dans l'article de Tajima (1990), et qu'incidemment, sa preuve n'y figure pas non plus, il l'utilise manifestement dans le calcul du temps approximatif de fixation avec paramètres particuliers. Nous avons donc cru important de présenter la formule générale ainsi que sa preuve. Son utilisation sera mentionnée dans une section ultérieure de ce mémoire. Ainsi, soient X_1, \dots, X_n , des variables indépendantes de loi exponentielle de paramètre λ_i , respectivement, pour $i = 1, \dots, n$. Alors, la fonction de densité

de X_i est donnée par

$$f_i(t) = (\lambda_i)e^{-\lambda_i t},$$

pour tout réel $t > 0$, pour $i = 1, \dots, n$.

Afin de démontrer le résultat sur la convolution de ces variables, soit que la fonction de densité de $X_1 + \dots + X_n$ est donnée par

$$f_{1,\dots,n}(t) = \sum_{i=1}^n \prod_{\substack{j=1 \\ j \neq i}}^n \frac{\lambda_j}{\lambda_j - \lambda_i} \lambda_i e^{-\lambda_i t} = \prod_{j=1}^n \lambda_j \sum_{i=1}^n \frac{e^{-\lambda_i t}}{\prod_{\substack{j=1 \\ j \neq i}}^n (\lambda_j - \lambda_i)},$$

pour tout réel $t > 0$, nous procéderons par induction.

Débutons donc par le cas où $n = 2$. Dans ce cas, on a

$$\begin{aligned} f_{12}(t) &= \int_0^t f_1(t-x)f_2(x)dx \\ &= \int_0^t \lambda_1 \lambda_2 e^{-\lambda_1(t-x)-\lambda_2(x)} dx \\ &= \lambda_1 \lambda_2 e^{-\lambda_1 t} \int_0^t e^{-(\lambda_2-\lambda_1)x} dx \\ &= -\frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1} e^{-\lambda_1 t} e^{-(\lambda_2-\lambda_1)x} \Big|_{x=0}^t \\ &= -\frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1} e^{-\lambda_1 t} e^{-(\lambda_2-\lambda_1)t} + \frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1} e^{-\lambda_1 t} \\ &= \frac{\lambda_2}{\lambda_2 - \lambda_1} \lambda_1 e^{-\lambda_1 t} + \frac{\lambda_1}{\lambda_1 - \lambda_2} \lambda_2 e^{-\lambda_2 t}. \end{aligned}$$

Analysons maintenant le cas où $n = 3$. On a

$$f_{123}(x) = \int_0^t f_{12}(t-x)f_3(x)dx.$$

Par le même raisonnement que pour le cas précédent, nous arrivons à

$$\begin{aligned} f_{123}(t) &= \frac{\lambda_2}{\lambda_2 - \lambda_1} \int_0^t f_1(t-x)f_3(x)dx + \frac{\lambda_1}{\lambda_1 - \lambda_2} \int_0^t f_2(t-x)f_3(x)dx \\ &= \frac{\lambda_2}{\lambda_2 - \lambda_1} \left(\frac{\lambda_3}{\lambda_3 - \lambda_1} \lambda_1 e^{-\lambda_1 t} + \frac{\lambda_1}{\lambda_1 - \lambda_3} \lambda_3 e^{-\lambda_3 t} \right) \\ &\quad + \frac{\lambda_1}{\lambda_1 - \lambda_2} \left(\frac{\lambda_3}{\lambda_3 - \lambda_2} \lambda_2 e^{-\lambda_2 t} + \frac{\lambda_2}{\lambda_2 - \lambda_3} \lambda_3 e^{-\lambda_3 t} \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{\lambda_2 \lambda_3}{(\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1)} \lambda_1 e^{-\lambda_1 t} + \frac{\lambda_1 \lambda_3}{(\lambda_1 - \lambda_2)(\lambda_3 - \lambda_2)} \lambda_2 e^{-\lambda_2 t} \\
&+ \left(\frac{\lambda_1 \lambda_2}{(\lambda_2 - \lambda_1)(\lambda_1 - \lambda_3)} - \frac{\lambda_1 \lambda_2}{(\lambda_2 - \lambda_1)(\lambda_2 - \lambda_3)} \right) \lambda_3 e^{-\lambda_3 t} \\
&= \frac{\lambda_2 \lambda_3}{(\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1)} \lambda_1 e^{-\lambda_1 t} + \frac{\lambda_1 \lambda_3}{(\lambda_1 - \lambda_2)(\lambda_3 - \lambda_2)} \lambda_2 e^{-\lambda_2 t} \\
&+ \frac{\lambda_1 \lambda_2}{(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_3)} \lambda_3 e^{-\lambda_3 t}.
\end{aligned}$$

Dans les deux cas, nous obtenons une expression de la forme

$$\sum_{i=1}^n \prod_{\substack{j=1 \\ j \neq i}}^n \frac{\lambda_j}{\lambda_j - \lambda_i} \lambda_i e^{-\lambda_i t},$$

pour $n = 2$ et $n = 3$, respectivement. Supposons maintenant que

$$f_{1, \dots, n-1}(t) = \sum_{i=1}^{n-1} \prod_{\substack{j=1 \\ j \neq i}}^{n-1} \frac{\lambda_j}{\lambda_j - \lambda_i} \lambda_i e^{-\lambda_i t} = \left(\prod_{j=1}^{n-1} \lambda_j \right) \sum_{i=1}^{n-1} \frac{e^{-\lambda_i t}}{\prod_{\substack{j=1 \\ j \neq i}}^{n-1} (\lambda_j - \lambda_i)}.$$

On a alors

$$\begin{aligned}
f_{1, \dots, n}(t) &= (f_{1, \dots, n-1} * f_n)(t) \\
&= \left(\prod_{j=1}^n \lambda_j \right) \sum_{i=1}^{n-1} \frac{e^{-\lambda_n t} - e^{-\lambda_i t}}{(\lambda_i - \lambda_n) \prod_{\substack{j=1 \\ j \neq i}}^{n-1} (\lambda_j - \lambda_i)} \\
&= \left(\prod_{j=1}^n \lambda_j \right) \left(\sum_{i=1}^{n-1} \frac{e^{-\lambda_i t}}{\prod_{\substack{j=1 \\ j \neq i}}^n (\lambda_j - \lambda_i)} - \sum_{i=1}^{n-1} \frac{e^{-\lambda_n t}}{\prod_{\substack{j=1 \\ j \neq i}}^n (\lambda_j - \lambda_i)} \right).
\end{aligned}$$

Afin d'obtenir la bonne formule, il nous faut donc démontrer que

$$-\sum_{i=1}^{n-1} \frac{1}{\prod_{\substack{j=1 \\ j \neq i}}^n (\lambda_j - \lambda_i)} = \frac{1}{\prod_{\substack{j=1 \\ j \neq i}}^n (\lambda_j - \lambda_i)}.$$

En voici la preuve. On a d'abord

$$\begin{aligned}
-\sum_{i=1}^{n-1} \frac{1}{\prod_{\substack{j=1 \\ j \neq i}}^n (\lambda_j - \lambda_i)} &= \frac{1}{\prod_{\substack{j=1 \\ j \neq i}}^n (\lambda_j - \lambda_i)} \\
\Leftrightarrow 0 &= \sum_{i=1}^n \frac{1}{\prod_{\substack{j=1 \\ j \neq i}}^n (\lambda_j - \lambda_i)} \\
\Leftrightarrow 0 &= \sum_{i=1}^n \frac{\prod_{\substack{l,m=1 \\ l \neq m \neq i}}^n (\lambda_l - \lambda_m)}{\prod_{\substack{l,m=1 \\ l \neq m}}^n (\lambda_l - \lambda_m)}
\end{aligned}$$

$$\begin{aligned}
\Leftrightarrow 0 &= \sum_{i=1}^n \prod_{\substack{l,m=1 \\ l \neq m \neq i}}^n (\lambda_l - \lambda_m) \\
\Leftrightarrow 0 &= \sum_{i=1}^n \prod_{\substack{l,m=1 \\ i \neq l \neq m \neq i}}^n (\lambda_l - \lambda_m) \prod_{\substack{l,m=1 \\ l=i \neq m}}^n (\lambda_l - \lambda_m) \\
\Leftrightarrow 0 &= \pm \sum_{i=1}^n \prod_{\substack{l,m=1 \\ i \neq l > m \neq i}}^n (\lambda_l - \lambda_m)^2 \prod_{\substack{l,m=1 \\ i=l > m}}^n (\lambda_l - \lambda_m) \prod_{\substack{l,m=1 \\ i=l < m}}^n (\lambda_l - \lambda_m) \\
\Leftrightarrow 0 &= \pm \sum_{i=1}^n \prod_{\substack{l,m=1 \\ i \neq l > m \neq i}}^n (\lambda_l - \lambda_m)^2 \prod_{\substack{l,m=1 \\ i=l > m}}^n (\lambda_l - \lambda_m) \prod_{\substack{l,m=1 \\ l > m=j}}^n (\lambda_l - \lambda_m) (-1)^{n-i} \\
\Leftrightarrow 0 &= \pm \sum_{i=1}^n \prod_{\substack{l,m=1 \\ i \neq l > m \neq i}}^n (\lambda_l - \lambda_m) \prod_{\substack{l,m=1 \\ l > m}}^n (\lambda_l - \lambda_m) (-1)^{n-i} \\
\Leftrightarrow 0 &= \sum_{i=1}^n \prod_{\substack{l,m=1 \\ i \neq l > m \neq i}}^n (\lambda_l - \lambda_m) (-1)^i.
\end{aligned}$$

Remarquons que le produit

$$\prod_{\substack{l,m=1 \\ l,m \neq i \\ l > m}}^n (\lambda_l - \lambda_m)$$

est le déterminant de Vandermonde suivant :

$$\begin{vmatrix}
1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{n-2} \\
1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{n-2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & \lambda_{i-1} & \lambda_{i-1}^2 & \cdots & \lambda_{i-1}^{n-2} \\
1 & \lambda_{i+1} & \lambda_{i+1}^2 & \cdots & \lambda_{i+1}^{n-2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & \lambda_n & \lambda_n^2 & \cdots & \lambda_n^{n-2}
\end{vmatrix},$$

d'où que

$$\sum_{i=1}^n \prod_{\substack{l,m=1 \\ l,m \neq i \\ l > m}}^n (\lambda_l - \lambda_m) (-1)^i$$

est le déterminant

$$\begin{vmatrix} 1 & 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{n-2} \\ 1 & 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \lambda_{i-1} & \lambda_{i-1}^2 & \cdots & \lambda_{i-1}^{n-2} \\ 1 & 1 & \lambda_{i+1} & \lambda_{i+1}^2 & \cdots & \lambda_{i+1}^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \lambda_n & \lambda_n^2 & \cdots & \lambda_n^{n-2} \end{vmatrix},$$

qui est égal à 0, puisqu'il contient deux colonnes identiques.

La preuve est donc complète.

1.3. PREUVES DE L'ARTICLE DE TAJIMA

1.3.1. Formule de convolution

Nous nous intéressons ici à $y(t)$, avec t comme variable de temps continue, la distribution du nombre de générations que prend un nouveau mutant à se fixer dans la population, excluant les cas où le mutant disparaît. Ces calculs nous permettront plus tard de démontrer des résultats intéressants sur l'influence du temps de fixation T d'un nouveau mutant sur le polymorphisme. Nous considérons ici qu'à chaque changement d'état, survenant après un temps variable de loi exponentielle dont le paramètre dépend du nombre de fréquences mutantes restantes, une séquence disparaît pour être remplacée par la copie de l'une des autres séquences dans la population. Il s'agit donc de calculer la somme de variables indépendantes de loi exponentielle X_i représentant le temps pour passer de l'état comprenant $i + 1$ séquences à l'état en comprenant i . Nous avons prouvé dans la section précédente que la convolution de n variables exponentielles X_1, \dots, X_n , ayant pour paramètres $\lambda_1, \dots, \lambda_n$, respectivement, a comme fonction de densité

$$f_{1,\dots,n}(t) = \sum_{i=1}^n \left(\prod_{\substack{j=1 \\ j \neq i}}^n \frac{\lambda_j}{\lambda_j - \lambda_i} \right) \lambda_i e^{-\lambda_i t}, \quad (1.3.1)$$

pour tout réel $t > 0$. Dans l'article de Tajima (1990), le taux avec lequel on quitte l'état i a pour valeur

$$\lambda_i = \frac{(i+1)i}{4N},$$

pour $i = 1, \dots, n$. Évaluons le produit des quotients dans l'équation 1.3.1 pour ces valeurs en particulier. On a

$$\begin{aligned}
\prod_{\substack{j=1 \\ j \neq i}}^n \frac{\lambda_j}{\lambda_j - \lambda_i} &= \prod_{\substack{j=1 \\ j \neq i}}^n \frac{j(j+1)}{j(j+1) - i(i+1)} \\
&= \prod_{\substack{j=1 \\ j \neq i}}^n \frac{j(j+1)}{(j-i)(j+i+1)} \\
&= \frac{n!(n+1)!}{i(i+1)} \frac{(-1)^{i-1}}{(n-i)!(i-1)!} \frac{(2i+1)(i+1)!}{(n+i+1)!} \\
&= (2i+1)(-1)^{i+1} \frac{n!}{(n-i)!} \frac{(n+1)!}{(n+i+1)!} \\
&= (2i+1)(-1)^{i+1} \prod_{j=1}^i \frac{n+1-j}{n+1+j}.
\end{aligned}$$

Ainsi, la convolution de $2N - 1$ variables pour $i = 1, \dots, 2N - 1$ a une fonction de densité donnée par

$$y(t) = \sum_{i=1}^{2N-1} (2i+1)(-1)^{i+1} \prod_{j=1}^i \frac{2N-j}{2N+j} \lambda_i e^{-\lambda_i t},$$

pour tout réel $t > 0$.

1.3.2. Explication de l'approche dans l'article

L'article de Tajima (1990) utilise une nouvelle approche afin d'arriver aux résultats qu'il recherche. Cette section expliquera en quoi consiste cette approche. Rappelons que A_i est le i -ième plus ancien ancêtre commun possible de deux séquences choisies au hasard dans la population, tel que vu en page 6. Rappelons aussi que α_i est la probabilité que deux séquences choisies au hasard aient A_i comme ancêtre commun le plus récent. Démontrons que

$$\alpha_i = \frac{2(2N+1)}{(i+1)(i+2)(2N-1)}.$$

DÉMONSTRATION. Soit A_i , le i -ième plus vieil ancêtre d'un échantillon de n séquences et $\alpha_{i,n}$ la probabilité que A_i soit l'ancêtre commun le plus récent de deux séquences choisies au hasard parmi l'échantillon de taille n .

Il y a $\binom{n}{2}$ combinaisons de séquences possibles, mais une seule crée le plus

récent ancêtre A_{n-1} . Ainsi, on a

$$a_{n-1,n} = \frac{1}{\binom{n}{2}} = \frac{2}{n(n-1)}.$$

Supposons que la valeur de $a_{i,n-1}$ est connue. Alors, la formule de $a_{i,n}$ sachant $a_{i,n-1}$ est donnée par

$$\begin{aligned} (a_{i,n} | a_{i,n-1}) &= (1 - a_{n-1,n}) a_{i,n-1} \\ &= \frac{(n+1)(n-2)}{n(n-1)} a_{i,n-1}. \end{aligned}$$

De la même façon,

$$a_{n-2,n-1} = \frac{1}{\binom{n-1}{2}} = \frac{2}{(n-1)(n-2)},$$

et donc

$$\begin{aligned} (a_{i,n-1} | a_{i,n-2}) &= (1 - a_{n-2,n-1}) a_{i,n-2} \\ &= \frac{n(n-3)}{(n-1)(n-2)} a_{i,n-2}. \end{aligned}$$

Par récurrence, nous arrivons à

$$\begin{aligned} (a_{i,i+2} | a_{i,i+1}) &= (1 - a_{i+1,i+2}) a_{i,i+1} \\ &= \frac{(i+3)i}{(i+2)(i+1)} a_{i,i+1}, \end{aligned}$$

avec $a_{i,i+1} = \frac{2}{i(i+1)}$. Nous obtenons donc que

$$\begin{aligned} a_{i,n} &= \left(\prod_{j=i}^{n-2} \frac{(j+3)j}{(j+2)(j+1)} \right) \frac{2}{i(i+1)} \\ &= \frac{(n+1)i}{(n-1)(i+2)} \frac{2}{i(i+1)} \\ &= \frac{2(n+1)}{(i+1)(i+2)(n-1)}. \end{aligned}$$

Nous obtenons finalement le résultat suivant lorsque l'échantillon est la population complète de $2N$ séquences :

$$a_{i,2N} = \frac{2(2N+1)}{(i+1)(i+2)(2N-1)}.$$

□

1.3.3. Polymorphisme

Afin de pouvoir déterminer l'effet qu'a le temps de fixation d'un nouveau mutant sur le nombre de sites polymorphes dans la population, il est important de déterminer l'espérance de ce nombre à un moment quelconque dans le temps. Nous nous intéressons donc ici au nombre k de sites polymorphes entre deux séquences d'ADN choisies au hasard dans la population ainsi qu'à son nombre espéré $\mathbb{E}(k)$. Soit k_i le nombre de différences entre deux séquences choisies au hasard dans la population et ayant A_i comme ancêtre commun le plus récent. Rappelons que ν représente le taux de mutation sélectivement neutre par séquence par génération.

Alors, on a

$$\begin{aligned}\mathbb{E}(k_i) &= \sum_{j=i}^{2N-1} \frac{2\nu}{\lambda_j} \\ &= 2\nu \sum_{j=i}^{2N-1} \frac{4N}{j(j+1)} \\ &= 2M \left(\frac{1}{i} - \frac{1}{2N} \right) \\ &= 2M \left(\frac{1}{i} - \frac{1}{2N} \right),\end{aligned}$$

avec $M = 4N\nu$. Nous obtenons donc que

$$\begin{aligned}\mathbb{E}(k) &= \mathbb{E}(\mathbb{E}(k | A_i)) \\ &= \sum_{i=1}^{2N-1} \alpha_i \mathbb{E}(k_i) \\ &= 2M \frac{2(2N+1)}{2N-1} \left(\sum_{i=1}^{2N-1} \frac{1}{i(i+1)(i+2)} - \frac{1}{2N} \sum_{i=1}^{2N-1} \frac{1}{(i+1)(i+2)} \right) \\ &= M \frac{4(2N+1)}{2N-1} \left[\frac{1}{2} \sum_{i=1}^{2N-1} \left(\frac{1}{i} - \frac{2}{i+1} + \frac{1}{i+2} \right) - \frac{1}{2N} \sum_{i=1}^{2N-1} \left(\frac{1}{i+1} - \frac{1}{i+2} \right) \right] \\ &= M \frac{4(2N+1)}{2N-1} \left[\frac{1}{2} \left(1 - \frac{1}{2N} - \frac{1}{2} + \frac{1}{2N+1} \right) - \frac{2}{2N} \left(\frac{1}{2} - \frac{1}{2N+1} \right) \right] \\ &= M \frac{4(2N+1)}{2N-1} \left[\frac{1}{4} - \frac{1}{4N} + \frac{1}{2(2N+1)} - \frac{1}{4N} + \frac{1}{2N(2N+1)} \right]\end{aligned}$$

$$\begin{aligned}
&= M \frac{2N+1}{2N-1} \left[\frac{2N^2 + N - 2N - 1 + 2N - 2N - 1 + 2}{N(2N+1)} \right] \\
&= M.
\end{aligned}$$

Il s'agit du même résultat que celui apporté par Watterson (1975), ce qui prouve que la nouvelle approche fonctionne dans ce cas-ci.

1.3.4. Homozygotie

Comme dans la section précédente, il est important de déterminer l'équation de l'homozygotie entre deux séquences à un moment quelconque dans le temps, afin de pouvoir par la suite déterminer l'influence qu'a le temps de fixation d'un nouveau mutant sur celle-ci. Nous analysons donc ici $\mathbb{E}(F)$, l'homozygotie espérée de deux séquences choisies au hasard dans la population. Pour ce faire, nous avons besoin de la probabilité b_j qu'il n'y ait pas de mutation sur l'une des deux branches ancestrales aux deux séquences lorsqu'il y a $j+1$ branches ancestrales à la population entière, qui a pour formule

$$\begin{aligned}
b_j &= \int_0^\infty f_j(t) e^{-2vt} dt \\
&= \int_0^\infty \lambda_j e^{-\lambda_j t} e^{-2vt} dt \\
&= \int_0^\infty \lambda_j e^{-(\lambda_j + 2v)t} dt \\
&= \frac{\lambda_j}{\lambda_j + 2v}.
\end{aligned}$$

Nous avons donc que l'homozygotie espérée de deux séquences choisies au hasard dans la population qui ont A_i comme ancêtre commun le plus récent, $\mathbb{E}(F_i)$, est donnée par

$$\begin{aligned}
\mathbb{E}(F_i) &= \prod_{j=i}^{2N-1} b_j \\
&= \prod_{j=i}^{2N-1} \frac{\lambda_j}{\lambda_j + 2v} \\
&= \prod_{j=i}^{2N-1} \frac{j(j+1)}{j(j+1) + 2M} \\
&= \prod_{j=i}^{2N-1} \frac{1}{1 + \frac{2M}{j(j+1)}}
\end{aligned}$$

$$\begin{aligned}
&\approx \prod_{j=i}^{2N-1} \left(1 - \frac{2M}{j(j+1)}\right) \\
&\approx 1 - \sum_{j=i}^{2N-1} \frac{2M}{j(j+1)} \\
&= 1 - \frac{2M}{i} + \frac{2M}{2N} \\
&= 1 - \mathbb{E}(k_i),
\end{aligned}$$

lorsque M est suffisamment petit, car, tel que déjà démontré, $\mathbb{E}(k_i) = 2M \left(\frac{1}{i} - \frac{1}{2N}\right)$. Nous pouvons maintenant calculer l'homozygotie espérée approximative :

$$\begin{aligned}
\mathbb{E}(F) &= \sum_{i=1}^{2N-1} \alpha_i \mathbb{E}(F_i) \\
&\approx \sum_{i=1}^{2N-1} \alpha_i (1 - \mathbb{E}(k_i)) \\
&= \sum_{i=1}^{2N-1} \alpha_i - \sum_{i=1}^{2N-1} \alpha_i \mathbb{E}(k_i) \\
&= 1 - M \\
&\approx \frac{1}{1 + M},
\end{aligned}$$

en utilisant une approximation par série géométrique, pour un M petit. Le même résultat a été apporté par Kimura et Crow (1964), confirmant de nouveau la validité de la nouvelle approche.

1.3.5. Polymorphisme au moment de la fixation

Dans cette section, nous considérons que le mutant parvient à se fixer dans la population. Nous nous intéressons donc, de façon semblable à la section 1.3.3, au nombre de sites polymorphes entre deux séquences choisies au hasard dans la population, mais cette fois au moment inconnu T de la fixation d'un mutant, soit $\mathbb{E}(k | T < \infty)$, qui sera noté $\mathbb{E}(k | \text{fix})$ dans la suite de la section. Nous comparerons donc les résultats afin de déterminer l'influence du temps de fixation sur le phénomène. Comme précédemment, on calcule d'abord

$$\mathbb{E}(k_i | \text{fix}) = \sum_{j=i+1}^{2N-1} \frac{2\nu}{\lambda_j} = 2M \left(\frac{1}{i+1} - \frac{1}{2N} \right),$$

où k_i est le nombre de différences entre deux séquences choisies au hasard dans la population et ayant A_i comme ancêtre commun le plus récent. La somme

commence à $i + 1$ car nous suivons en plus des lignées finales la dernière lignée non mutante qui est éliminée lors de la fixation. Alors, on a

$$\begin{aligned}
\mathbb{E}(k \mid \text{fix}) &= \sum_{i=1}^{2N-2} a_i \mathbb{E}(k_i \mid \text{fix}) \\
&= \sum_{i=1}^{2N-2} \frac{2(2N+1)}{(i+1)(i+2)(2N-1)} 2M \left(\frac{1}{i+1} - \frac{1}{2N} \right) \\
&= 2M \frac{2(2N+1)}{(2N-1)} \sum_{i=1}^{2N-2} \left[\left(\frac{1}{i+1} - \frac{1}{i+2} \right) \left(\frac{1}{i+1} - \frac{1}{2N} \right) \right] \\
&= 2M \frac{2(2N+1)}{(2N-1)} \left[\sum_{i=1}^{2N-2} \frac{1}{(i+1)^2} - \frac{1}{2N} \sum_{i=1}^{2N-2} \left(\frac{1}{i+1} - \frac{1}{i+2} \right) \right. \\
&\quad \left. - \sum_{i=1}^{2N-2} \frac{1}{(i+1)(i+2)} \right] \\
&= 2M \frac{2(2N+1)}{(2N-1)} \left[\sum_{i=1}^{2N-2} \frac{1}{i^2} - 1 - \frac{2N+1}{2N} \left(\frac{1}{2} - \frac{1}{2N} \right) \right] \\
&\rightarrow 2M \left[2 \sum_{i \geq 1} \frac{1}{i^2} - 2 - 1 \right] \\
&= 2M \left[\frac{\pi^2}{3} - 3 \right] \\
&\approx 0,5797M,
\end{aligned}$$

lorsque $N \rightarrow \infty$, car $\sum_{i \geq 1} \frac{1}{i^2} = \frac{\pi^2}{6}$. Nous pouvons donc remarquer qu'en choisissant deux séquences au hasard dans la population au moment de la fixation, le nombre de sites polymorphes sera prêt de 42% inférieur à celui à un temps quelconque.

1.3.6. Homozygotie au moment de la fixation

Tout comme à la section 1.3.4, nous nous intéressons à l'homozygotie de deux séquences prises au hasard dans la population, mais cette fois-ci au moment de la fixation d'un mutant. Nous pouvons donc déterminer l'influence du temps de fixation sur ce phénomène biologique en comparant les résultats des deux sections. Pour la même raison que précédemment, nous avons

$$\mathbb{E}(F_i \mid \text{fix}) = \prod_{j=i+1}^{2N-1} b_j = \prod_{j=i+1}^{2N-1} \frac{1}{1 + \frac{2M}{j(j+1)}}.$$

Ainsi, nous obtenons que

$$\begin{aligned}\mathbb{E}(F \mid \text{fix}) &= \frac{2(2N+1)}{2N-1} \sum_{i=1}^{2N-2} \frac{1}{(i+1)(i+2)} \prod_{j=i+1}^{2N-1} \frac{1}{1 + \frac{2M}{j(j+1)}} \\ &= \sum_{i=2}^{2N-1} \frac{2(2N+1)}{2N-1} \frac{1}{i(i+1)} \prod_{j=i}^{2N-1} \frac{1}{1 + \frac{2M}{j(j+1)}}.\end{aligned}\quad (1.3.2)$$

Débutons par développer le produit au dénominateur dans l'équation 1.3.2. En multipliant les parenthèses, nous obtenons comme approximation une somme contenant tous les termes jusqu'à la troisième puissance de M , ce qui donne

$$\begin{aligned}\prod_{j=i}^{2N-1} \left(1 + \frac{2M}{j(j+1)}\right) &\approx 1 + 2M \sum_{j_1 \geq i} \frac{1}{j_1(j_1+1)} + (2M)^2 \sum_{j_1 \geq i} \frac{1}{j_1(j_1+1)} \sum_{j_2 \geq j_1+1} \frac{1}{j_2(j_2+1)} \\ &\quad + (2M)^3 \sum_{j_1 \geq i} \frac{1}{j_1(j_1+1)} \sum_{j_2 \geq j_1+1} \frac{1}{j_2(j_2+1)} \sum_{j_3 \geq j_2+1} \frac{1}{j_3(j_3+1)} \\ &= 1 + \frac{2M}{i} + (2M)^2 \sum_{j_1 \geq i} \frac{1}{j_1(j_1+1)^2} \\ &\quad + (2M)^3 \sum_{j_1 \geq i} \frac{1}{j_1(j_1+1)} \sum_{j_2 \geq j_1+1} \frac{1}{j_2(j_2+1)^2} \\ &= 1 + \frac{2M}{i} + (2M)^2 \left(\frac{1}{i} - \sum_{j_1 \geq i} \frac{1}{(j_1+1)^2} \right) \\ &\quad + (2M)^3 \sum_{j_1 \geq i} \frac{1}{j_1(j_1+1)} \left(\frac{1}{j_1+1} - \sum_{j_2 \geq 1} \frac{1}{j_2^2} + \sum_{j_2=1}^{j_1+1} \frac{1}{j_2^2} \right) \\ &= 1 + \frac{2M}{i} + (2M)^2 \left(\frac{1}{i} - \sum_{j_1 \geq 1} \frac{1}{j_1^2} + \sum_{j_1=1}^i \frac{1}{j_1^2} \right) \\ &\quad + (2M)^3 \left(\frac{1}{i} - \sum_{j_1 \geq 1} \frac{1}{j_1^2} + \sum_{j_1=1}^i \frac{1}{j_1^2} \right) \\ &\quad - \frac{(2M)^3}{i} \sum_{j_1 \geq 1} \frac{1}{j_1^2} + (2M)^3 \sum_{j_1 \geq i} \frac{1}{j_1(j_1+1)} \sum_{j_2=1}^{j_1+1} \frac{1}{(j_2)^2},\end{aligned}$$

avec

$$\begin{aligned}\sum_{j_1 \geq i} \frac{1}{j_1(j_1+1)} \sum_{j_2=1}^{j_1+1} \frac{1}{j_2^2} &= \sum_{j_1 \geq i} \left(\frac{1}{j_1} \sum_{j_2=1}^{j_1+1} \frac{1}{j_2^2} - \frac{1}{j_1+1} \sum_{j_2=1}^{j_1+1} \frac{1}{j_2^2} \right) \\ &= \frac{1}{i} \sum_{j_2=1}^{i+1} \frac{1}{j_2^2} + \sum_{j_1 \geq i+1} \frac{1}{j_1} \sum_{j_2=1}^{j_1+1} \frac{1}{j_2^2} - \sum_{j_1 \geq i+1} \frac{1}{j_1} \sum_{j_2=1}^{j_1} \frac{1}{j_2^2}\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{i} \sum_{j_2=1}^{i+1} \frac{1}{j_2^2} + \sum_{j_1 \geq i+1} \frac{1}{j_1} \frac{1}{(j+1)^2} \\
&= \frac{1}{i} \sum_{j_2=1}^{i+1} \frac{1}{j_2^2} + \frac{1}{i+1} - \sum_{j_1 \geq 1} \frac{1}{j_1^2} + \sum_{j_1=1}^{i+1} \frac{1}{j_1^2}.
\end{aligned}$$

Ainsi, l'approximation peut être réécrite comme

$$\begin{aligned}
\prod_{j=i}^{2N-1} \left(1 + \frac{2M}{j(j+1)}\right) &\approx 1 + \frac{2M}{i} + (2M)^2 \left(\frac{1}{i} - \sum_{j_1 \geq 1} \frac{1}{j_1^2} + \sum_{j_1=1}^i \frac{1}{j_1^2}\right) \\
&+ (2M)^3 \left(\frac{1}{i} - \sum_{j_1 \geq 1} \frac{1}{j_1^2} + \sum_{j_1=1}^i \frac{1}{j_1^2}\right) - \frac{(2M)^3}{i} \sum_{j_1 \geq 1} \frac{1}{j_1^2} \\
&+ (2M)^3 \left(\frac{1}{i} \sum_{j_1=1}^{i+1} \frac{1}{j_1^2} + \frac{1}{i+1} - \sum_{j_1 \geq 1} \frac{1}{j_1^2} + \sum_{j_1=1}^{i+1} \frac{1}{j_1^2}\right) \\
&= 1 + \frac{2M}{i} + 4M^2 \left(\frac{1}{i} + \sum_{j_1=1}^i \frac{1}{j_1^2}\right) - \frac{4\pi^2 M^2}{6} \\
&+ 8M^3 \left(\frac{1}{i} + \sum_{j_1=1}^i \frac{1}{j_1^2} + \frac{1}{i} \sum_{j_1=1}^{j+1} \frac{1}{j_1^2} + \frac{1}{i+1} + \sum_{j_1=1}^{i+1} \frac{1}{j_1^2}\right) \\
&- 8M^3 \frac{\pi^2}{6} \left(2 + \frac{1}{i}\right).
\end{aligned}$$

Nous pouvons donc approximer le produit dans l'équation 1.3.2 et le développer par une série géométrique, comme suit

$$\begin{aligned}
\prod_{j=i}^{2N-1} \frac{1}{1 + \frac{2M}{j(j+1)}} &= 1 / \left(1 + \frac{2M}{i} + \frac{4M^2}{i} + 4M^2 \sum_{j_1=1}^i \frac{1}{j_1^2} - \frac{4\pi^2 M^2}{6} + O(M^3)\right) \\
&= 1 - \frac{2M}{i} - \frac{4M^2}{i} - 4M^2 \sum_{j_1=1}^i \frac{1}{j_1^2} + \frac{4\pi^2 M^2}{6} + \frac{4M^2}{i^2} + O(M^3) \\
&\approx \left(1 + \frac{4\pi^2 M^2}{6}\right) - \frac{1}{i} (2M + 4M^2) + \frac{4M^2}{i^2} - 4M^2 \sum_{j_1=1}^i \frac{1}{j_1^2}.
\end{aligned}$$

Ainsi, on trouve que

$$\mathbb{E}(F \mid \text{fix}) = \sum_{i=2}^{2N-1} \frac{2(2N+1)}{2N-1} \frac{1}{i(i+1)} \prod_{j=i}^{2N-1} \frac{1}{1 + \frac{2M}{j(j+1)}}$$

$$\begin{aligned}
&\approx \sum_{i \geq 2} 2 \frac{1}{i(i+1)} \left[\left(1 + \frac{4\pi^2 M^2}{6} \right) - \frac{1}{i} (2M + 4M^2) + \frac{4M^2}{i^2} - 4M^2 \sum_{j_1=1}^i \frac{1}{j_1^2} \right] \\
&= 2 \left[\sum_{i \geq 2} \frac{1}{i(i+1)} \left(1 + \frac{4\pi^2 M^2}{6} \right) - \sum_{i \geq 2} \frac{1}{i^2(i+1)} (2M + 4M^2) \right. \\
&\quad \left. + 4M^2 \sum_{i \geq 2} \frac{1}{i^3(i+1)} - 4M^2 \sum_{i \geq 2} \frac{1}{i(i+1)} \sum_{j_1=1}^i \frac{1}{j_1^2} \right] \\
&= \left(1 + \frac{4\pi^2 M^2}{6} \right) + \left(6M + 12M^2 - \frac{2M\pi^2}{3} - \frac{4M^2\pi^2}{3} \right) \\
&\quad + \left(8M^3 \sum_{i \geq 2} \frac{1}{i^3} + 12M^2 - \frac{4M^2\pi^2}{3} \right) - \left(5M^2 + 8M^2 \sum_{i \geq 2} \frac{1}{i^3} \right) \\
&= 1 + M \left(6 - \frac{4\pi^2}{6} \right) + M^2(20 - 2\pi^2).
\end{aligned}$$

En posant $c_1 = \frac{4\pi^2 - 39}{6}$ et $c_2 = \frac{79 - 8\pi^2}{3}$, nous concluons, pour une M petit, que

$$\begin{aligned}
\mathbb{E}(F \mid \text{fix}) &= 1 - \left(c_1 + \frac{1}{2} \right) M + \left(c_1 + c_2 + \frac{1}{6} \right) M^2 \\
&= \frac{1}{M} \left[M - c_1 M^2 - \frac{M^2}{2} + \frac{c_1}{2} M^3 + \frac{c_1}{2} M^3 + c_2 M^3 + \frac{M^3}{6} \right] \\
&\approx \frac{1}{M} \left[1 - \left(1 - M + c_1 M^2 - c_2 M^3 + \frac{1}{2} (-M + c_1 M^2 - c_2 M^3)^2 \right. \right. \\
&\quad \left. \left. + \frac{1}{6} (-M + c_1 M^2 + c_2 M^3)^3 \right) \right] \\
&\approx \frac{1}{M} \left(1 - e^{-M + c_1 M^2 - c_2 M^3} \right).
\end{aligned}$$

Ce résultat est toujours plus grand que $\mathbb{E}(F) = \frac{1}{1+M}$, pour tout M , indiquant là encore que l'homozygotie est supérieure au moment de la fixation de l'allèle mutant. L'importance de la différence dépend toutefois de M .

1.3.7. Polymorphisme au moment connu de la fixation

Supposons maintenant que le moment de la fixation est connu et qu'il survient au temps T . Nous nous intéressons ici au nombre de sites polymorphes espéré entre deux séquences choisies au hasard au temps de fixation T , soit $\mathbb{E}(k \mid T)$. Sachant que la densité de probabilité du temps de fixation est donnée par la convolution de variables aléatoires indépendantes de loi exponentielle,

elle peut être exprimée sous la forme

$$y(T) = \int_{t=0}^T f_i(t)g_i(T-t)dt,$$

où $g_i(t)$ est la convolution des $f_j(t)$, pour $i = 1, \dots, l-1, l+1, \dots, 2N-1$.

Considérons tout d'abord le nombre espéré de générations entre A_{i-1} et A_i , conditionnellement à ce que la fixation survienne au temps T , soit

$$\mathbb{E}(t_i | T) = \frac{z_i(T)}{y(T)},$$

avec

$$z_i(T) = \int_{t=0}^T t f_i(t) g_i(T-t) dt.$$

La formule de la convolution pour $g_i(t)$ dans ce cas-ci est de la forme

$$g_i(t) = \sum_{\substack{j=1 \\ j \neq i}}^{2N-1} \left(\prod_{\substack{l=1 \\ l \neq j \\ l \neq i}}^{2N-1} \frac{\lambda_l}{\lambda_l - \lambda_j} \right) \lambda_j e^{-\lambda_j t}. \quad (1.3.3)$$

Débutons par déterminer la formule du produit dans l'équation 1.3.3, soit

$$\begin{aligned} \prod_{\substack{l=1 \\ l \neq j \\ l \neq i}}^{2N-1} \frac{\lambda_l}{\lambda_l - \lambda_j} &= \prod_{\substack{l=1 \\ l \neq j \\ l \neq i}}^{2N-1} \frac{l(l-1)}{(l-1)(l+j+1)} \\ &= \frac{(2N-1)!(2N)!}{j(j+1)i(i+1)} \frac{(j+1)!(2j+1)(i+j+1)}{(2N+j)!} \frac{(-1)^{j-1}(i-j)}{(2N-j-1)!(j-i)!} \\ &= \frac{(i-j)(i+j+1)}{i(i+1)} \left[(2j+1)(-1)^{j+1} \prod_{l=1}^j \left(\frac{2N-l}{2N+l} \right) \right] \\ &= \frac{\lambda_i - \lambda_j}{\lambda_i} (2j+1)(-1)^{j+1} \prod_{l=1}^j \frac{2N-l}{2N+l}. \end{aligned}$$

Ainsi, nous avons que

$$\begin{aligned} g_i(t) &= \sum_{\substack{j=1 \\ j \neq i}}^{2N-1} \prod_{\substack{l=1 \\ l \neq j \\ l \neq i}}^{2N-1} \frac{\lambda_l}{\lambda_l - \lambda_j} \lambda_j e^{-\lambda_j t} \\ &= \sum_{\substack{j=1 \\ j \neq i}}^{2N-1} \frac{\lambda_i - \lambda_j}{\lambda_i} (2j+1)(-1)^{j+1} \prod_{l=1}^j \frac{2N-l}{2N+l} \lambda_j e^{-\lambda_j t}. \end{aligned}$$

Nous pouvons alors calculer

$$\begin{aligned}
z_i(T) &= \int_0^T t f_i(t) g_i(T-t) dt \\
&= \int_0^T t \lambda_i e^{-\lambda_i t} \sum_{\substack{j=1 \\ j \neq i}}^{2N-1} \frac{\lambda_i - \lambda_j}{\lambda_i} (2j+1) (-1)^{j+1} \prod_{l=1}^j \frac{2N-l}{2N+l} \lambda_j e^{-\lambda_j (T-t)} dt \\
&= \sum_{\substack{j=1 \\ j \neq i}}^{2N-1} \lambda_j (\lambda_i - \lambda_j) (2j+1) (-1)^{j+1} \prod_{l=1}^j \frac{2N-l}{2N+l} \int_0^T t e^{-t(\lambda_i - \lambda_j)} e^{-\lambda_j T} dt.
\end{aligned}$$

En intégrant par partie avec $u = t$, $du = dt$, $dv = e^{-t(\lambda_i - \lambda_j)}$ et $v = -\frac{1}{\lambda_i - \lambda_j} e^{-t(\lambda_i - \lambda_j)}$, nous obtenons que

$$\begin{aligned}
z_i(T) &= \sum_{\substack{j=1 \\ j \neq i}}^{2N-1} e^{-\lambda_j T} \lambda_j \frac{(\lambda_i - \lambda_j)}{(\lambda_i - \lambda_j)} (2j+1) (-1)^{j+1} \prod_{l=1}^j \frac{2N-l}{2N+l} \left(-te^{-t(\lambda_i - \lambda_j)} \right) \Big|_{t=0}^T \\
&+ \sum_{\substack{j=1 \\ j \neq i}}^{2N-1} e^{-\lambda_j T} \lambda_j \frac{(\lambda_i - \lambda_j)}{(\lambda_i - \lambda_j)} (2j+1) (-1)^{j+1} \prod_{l=1}^j \frac{2N-l}{2N+l} \int_0^T e^{-t(\lambda_i - \lambda_j)} dt \\
&= - \sum_{\substack{j=1 \\ j \neq i}}^{2N-1} \lambda_j (2j+1) (-1)^{j+1} \prod_{l=1}^j \frac{2N-l}{2N+l} T e^{-T \lambda_i} \\
&+ \sum_{\substack{j=1 \\ j \neq i}}^{2N-1} (2j+1) (-1)^{j+1} \prod_{l=1}^j \frac{2N-l}{2N+l} \frac{\lambda_j}{\lambda_i - \lambda_j} (e^{-\lambda_j T} - e^{-\lambda_i T}) \\
&= - \sum_{j=1}^{2N-1} \lambda_j (2j+1) (-1)^{j+1} \prod_{l=1}^j \frac{2N-l}{2N+l} T e^{-T \lambda_i} \\
&+ \sum_{\substack{j=1 \\ j \neq i}}^{2N-1} (2j+1) (-1)^{j+1} \prod_{l=1}^j \frac{2N-l}{2N+l} \frac{\lambda_j}{\lambda_i - \lambda_j} (e^{-\lambda_j T} - e^{-\lambda_i T}) \\
&+ (2i+1) (-1)^{i+1} \prod_{l=1}^i \frac{2N-l}{2N+l} T \lambda_i e^{-T \lambda_i}.
\end{aligned}$$

Cependant, on a

$$\sum_{j=1}^{2N-1} \lambda_j (2j+1) (-1)^{j+1} \prod_{l=1}^j \frac{2N-l}{2N+l} T e^{-T \lambda_i} = 0,$$

car

$$\begin{aligned} \sum_{j=1}^{2N-1} \lambda_j (2j+1) (-1)^{j+1} \prod_{l=1}^j \frac{2N-l}{2N+l} &= \sum_{j=1}^{2N-1} \lambda_j \prod_{\substack{l=1 \\ l \neq j}}^j \frac{\lambda_l}{\lambda_l - \lambda_j} \\ &= \prod_{l=1}^{2N-1} \lambda_l \sum_{j=1}^{2N-1} \frac{1}{\prod_{\substack{l=1 \\ l \neq j}}^j (\lambda_l - \lambda_j)}, \end{aligned}$$

et nous avons démontré à la section précédente que $\sum_{j=1}^{2N-1} \frac{1}{\prod_{\substack{l=1 \\ l \neq j}}^j (\lambda_l - \lambda_j)} = 0$.

Ainsi, on trouve que

$$\begin{aligned} z_i(T) &= \sum_{\substack{j=1 \\ j \neq i}}^{2N-1} (2j+1) (-1)^{j+1} \prod_{l=1}^j \frac{2N-l}{2N+l} \frac{\lambda_j}{\lambda_i - \lambda_j} (e^{-\lambda_j T} - e^{-\lambda_i T}) \\ &\quad + (2i+1) (-1)^{i+1} \prod_{l=1}^i \frac{2N-l}{2N+l} T \lambda_i e^{-T \lambda_i}. \end{aligned}$$

Nous pouvons obtenir une approximation pour ce résultat par

$$\begin{aligned} z_i(T) &= \sum_{\substack{j=1 \\ j \neq i}}^{2N-1} (2j+1) (-1)^{j+1} \frac{\lambda_j}{\lambda_i - \lambda_j} (e^{-\lambda_j T} - e^{-\lambda_i T} e^{-2\lambda_j T}) \\ &\quad + (2i+1) (-1)^{i+1} \prod_{l=1}^i \frac{2N-l}{2N+l} T \lambda_i e^{-(T+2)\lambda_i}, \end{aligned}$$

puisque, par un développement en série de Maclaurin,

$$\begin{aligned} \log \left(\prod_{l=1}^i \frac{2N-l}{2N+l} \right) &= \sum_{l=1}^i \log \left(1 - \frac{l}{2N} \right) - \log \left(1 + \frac{l}{2N} \right) \\ &\approx \sum_{l=1}^i -\frac{l}{2N} - \frac{l^2}{2(2N)^2} - \frac{l}{2N} + \frac{l^2}{2(2N)^2} \\ &= \sum_{l=1}^i -\frac{j}{N} \\ &= -\frac{i(i+1)}{2N} \\ &= -2\lambda_i. \end{aligned}$$

Comme précédemment, nous considérons le nombre de sites polymorphes espéré entre deux séquences choisies au hasard au temps de fixation T et ayant

A_i comme ancêtre commun le plus récent, soit

$$\mathbb{E}(k_i | T) = \sum_{j=i+1}^{2N-1} 2\nu \mathbb{E}(t_j | T).$$

Nous avons alors

$$\begin{aligned} \mathbb{E}(k | T) &= \sum_{i=1}^{2N-2} \alpha_i \mathbb{E}(k_i | T) \\ &= \sum_{i=1}^{2N-2} \frac{2(2N+1)}{(i+1)(i+2)(2N-1)} \sum_{j=i+1}^{2N-1} 2\nu \mathbb{E}(t_j | T) \\ &= \frac{2(2N+1)}{(2N-1)} 2\nu \sum_{i=1}^{2N-2} \frac{1}{(i+1)(i+2)} \sum_{j=i+1}^{2N-1} \mathbb{E}(t_j | T) \\ &= \sum_{i=1}^{2N-2} \frac{i(2N+1)}{(i+2)(2N-1)} 2\nu \mathbb{E}(t_{i+1} | T) \\ &= \sum_{i=1}^{2N-2} 2\nu d_i \mathbb{E}(t_{i+1} | T), \end{aligned}$$

avec $d_i = \frac{i(2N+1)}{(i+2)(2N-1)}$. En effet, on a

$$\begin{aligned} \sum_{i=1}^{2N-2} \frac{1}{(i+1)(i+2)} \sum_{j=i+1}^{2N-1} \mathbb{E}(t_j | T) &= \sum_{i=1}^{2N-2} \frac{1}{i+1} \sum_{j=i+1}^{2N-1} \mathbb{E}(t_j | T) - \sum_{i=1}^{2N-2} \frac{1}{i+2} \sum_{j=i+1}^{2N-1} \mathbb{E}(t_j | T) \\ &= \frac{1}{2} \sum_{j=2}^{2N-1} \mathbb{E}(t_j | T) + \sum_{i=1}^{2N-3} \frac{1}{i+2} \sum_{j=i+2}^{2N-1} \mathbb{E}(t_j | T) \\ &\quad - \sum_{i=1}^{2N-2} \frac{1}{i+2} \sum_{j=i+1}^{2N-1} \mathbb{E}(t_j | T) \\ &= \frac{1}{2} \sum_{j=2}^{2N-1} \mathbb{E}(t_j | T) - \sum_{i=1}^{2N-2} \frac{1}{i+2} \mathbb{E}(t_{i+1} | T) \\ &= \sum_{i=1}^{2N-2} \left(\frac{1}{2} - \frac{1}{i+2} \right) \mathbb{E}(t_{i+1} | T) \\ &= \sum_{i=1}^{2N-2} \frac{i}{2(i+2)} \mathbb{E}(t_{i+1} | T). \end{aligned}$$

Considérons maintenant le nombre de sites polymorphes dans le cas limite où un très grand nombre de générations est nécessaire à la fixation, soit $T \rightarrow \infty$. Il

faut tout d'abord déterminer $\mathbb{E}(t_i | \infty)$. Soient

$$\alpha_j = (2j + 1)(-1)^{j+1} \prod_{l=1}^j \frac{2N - l}{2N + l} \frac{\lambda_j}{\lambda_i - \lambda_j}$$

et

$$\beta_j = (2j + 1)(-1)^{j+1} \prod_{l=1}^j \frac{2N - l}{2N + l} \lambda_j.$$

Alors, on a

$$\begin{aligned} \lim_{T \rightarrow \infty} \mathbb{E}(t_i | T) &= \lim_{T \rightarrow \infty} \frac{Z_i(T)}{y(T)} \\ &= \lim_{T \rightarrow \infty} \frac{\sum_{\substack{j=1 \\ j \neq i}}^{2N-1} \alpha_j (e^{-\lambda_j T} - e^{-\lambda_i T}) + \beta_i T e^{-\lambda_i T}}{\sum_{j=1}^{2N-1} \beta_j e^{-\lambda_j T}} \\ &= \lim_{T \rightarrow \infty} \frac{\sum_{\substack{j=1 \\ j \neq i}}^{2N-1} \alpha_j e^{-\lambda_j T} - \sum_{\substack{j=1 \\ j \neq i}}^{2N-1} \alpha_j e^{-\lambda_i T} + \beta_i T e^{-\lambda_i T}}{\sum_{j=1}^{2N-1} \beta_j e^{-\lambda_j T}} \\ &= \lim_{T \rightarrow \infty} \frac{\alpha_1 + \sum_{\substack{j=2 \\ j \neq i}}^{2N-1} \alpha_j e^{-(\lambda_j - \lambda_1)T} - \sum_{\substack{j=1 \\ j \neq i}}^{2N-1} \alpha_j e^{-(\lambda_i - \lambda_1)T} + \beta_i T e^{-T(\lambda_i - \lambda_1)}}{\beta_1 + \sum_{j=2}^{2N-1} \beta_j e^{-(\lambda_j - \lambda_1)T}} \\ &= \frac{\alpha_1}{\beta_1} \\ &= \frac{3 \frac{2N-1}{2N+1} \frac{\lambda_1}{\lambda_i - \lambda_1}}{3 \frac{2N-1}{2N+1} \lambda_1} \\ &= \frac{1}{\lambda_i - \lambda_1} \\ &= \frac{4N}{i(i+1) - 2} \\ &= \frac{4N}{(i+2)(i-1)}. \end{aligned}$$

Finalement, on obtient que

$$\begin{aligned} \mathbb{E}(k | \infty) &= \sum_{i=1}^{2N-2} 2v d_i \mathbb{E}(t_{i+1} | \infty) \\ &= \sum_{i=1}^{2N-2} 2v \frac{i(2N+1)}{(i+2)(2N-1)} \frac{4N}{(i+3)i} \\ &= 2M \frac{2N+1}{2N-1} \left(\frac{1}{3} - \frac{1}{2N+1} \right) \\ &\approx \frac{2}{3} M, \end{aligned}$$

pour N grand. Ainsi, nous remarquons que, même si le moment de la fixation est extrêmement éloigné, le nombre espéré de sites polymorphes est bien moindre au moment de la fixation d'un mutant qu'à tout autre moment.

Chapitre 2

INTRODUCTION DE L'EFFET DE SÉLECTION ET NOUVELLE APPROCHE

La sélection naturelle est un phénomène biologique permettant l'évolution et l'adaptation des espèces à un environnement au fil du temps. Il s'agit du principe selon lequel un être vivant ayant développé un trait ou une aptitude favorisant sa survie dans un certain habitat lui procure par le fait même un avantage pour procréer, principalement en raison de sa plus grande espérance de vie lui permettant d'atteindre la maturité. Son aptitude peut donc être passée à ses descendants, qui à leur tour, auront une meilleure chance d'avoir des descendants. Il existe donc certains traits génétiques, souvent acquis par mutation, qui sont plus susceptibles d'être passés à des descendants que d'autres.

Puisqu'il s'agit d'un phénomène écologique, il est très intéressant de l'analyser, entre autres de façon mathématique. C'est pourquoi certains mathématiciens ont analysé plusieurs des aspects s'y rattachant, notamment Darden et coll. (1988) ainsi que Mano (2009), pour ne nommer qu'eux. Il s'agit donc visiblement d'un sujet de recherche encore actuel.

Dans ce chapitre, deux types d'événement seront considérés mis à part la mutation initiale, soient l'événement de coalescence et l'événement de sélection. Lorsqu'un événement de coalescence survient sur un graphe de sélection ancestral et qu'il touche deux séquences identiques, l'une d'elle disparaît sur le graphe à l'état suivant. Lorsqu'un événement de sélection survient sur une séquence dans un graphe de sélection ancestral, un branchement a lieu, résultant en deux nouvelles séquences, l'une réelle et l'autre virtuelle. Nous ne savons cependant pas *a priori* laquelle est réelle. Les séquences résultantes dépendent cependant du type de séquence touchée. Si une séquence désavantagée est soumise à un événement de sélection, il y aura un branchement résultant en deux

séquences désavantagées. Cependant, si une séquence avantaée y est soumise, le branchement aura lieu, mais avec trois résultats possibles : la séquence réelle est avantaée et la séquence virtuelle ne l'est pas, la séquence virtuelle est avantaée et la séquence réelle ne l'est pas, ou bien les deux séquences sont avantaées. Dans ce mémoire, les séquences mutantes seront sélectivement désavantagées.

Dans l'article de Mano (2009), l'espérance du temps de fixation d'un nouveau gène mutant conditionnellement à l'arrivée de la fixation en un temps fini est approximée par un processus de diffusion. Pour une taille de population N très grande, dans laquelle l'allèle non mutant a un avantage sélectif $s > 0$ par rapport à l'allèle mutant, ses calculs lui donnent

$$4\pi_1 \sum_{j=0}^{\infty} \frac{H_{j+1}(4c)^j}{(j+2)!},$$

avec $H_k = 1 + \frac{1}{2} + \dots + \frac{1}{k}$, $c = Ns$ constant et

$$\pi_i = \frac{(4c)^i}{i!(e^{4c} - 1)}.$$

Ce résultat peut être approximé par $2 - \frac{4}{9}\sigma^2 + o(\sigma^2)$, où $\frac{\sigma}{2}$ est le taux de sélection par séquence par génération. Une nouvelle approche utilisant un processus ancestral a donc été avancée dans le but d'arriver aux mêmes résultats et c'est celle-ci qui sera rapportée dans cette section.

2.1. NOUVELLE APPROCHE AVEC UN ÉVÉNEMENT DE SÉLECTION POSSIBLE

La nouvelle technique consiste à calculer les espérances du temps de fixation ainsi que leur probabilité associée pour chaque scénario menant à la fixation d'un nouveau mutant lorsqu'il y a au plus un événement de sélection, afin de calculer la probabilité totale, l'espérance totale et finalement l'espérance du temps de fixation conditionnellement à la fixation du mutant. Nous considérons donc tout d'abord tous les cas possibles. Dans chacun des cas, nous considérons $n + 1$ séquences au total au départ, soit n mutantes et 1 non mutante. Le taux de sélection par séquence par génération est de $\frac{\sigma}{2}$. Nous considérons très improbable qu'une mutation touche de nouveau le même locus, et donc le taux de mutation est négligeable, sauf lors de la première mutation où il vaut μ .

Les figures pour chacun des cas qui suivent sont des graphes de sélection ancestraux, dont le fonctionnement est décrit dans l'introduction de ce chapitre. Seuls des événements de coalescence ou un événement de sélection peuvent survenir suite à la mutation. Sur les graphes, les séquences mutantes sont représentées par un point blanc et les séquences non mutantes par des points noirs. Nous considérons que chaque transition d'états est faite par un événement de coalescence touchant deux séquences mutantes, sauf lorsqu'un autre événement survient, auquel cas le nombre de séquences mutantes restantes est écrit.

Cas 1

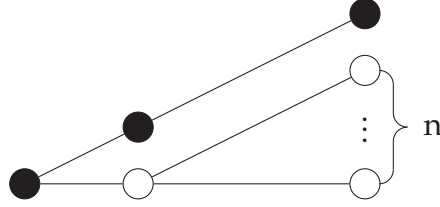


FIGURE 2.1. Graphe de sélection ancestral sans événement de sélection.

Comme les temps passés à chaque état sont des variables aléatoires de loi exponentielle indépendantes, l'espérance du temps total pour ce cas est la somme des espérances à chaque temps, soit

$$\begin{aligned}
 E &= \mathbb{E}(T) \\
 &= \sum_{i=1}^n \frac{2}{i(i+1) + (i+1)\sigma} \\
 &\approx \sum_{i=1}^n \frac{2}{i(i+1)} - \sigma \sum_{i=1}^n \frac{2}{i^2(i+1)} \\
 &= E_1 - \sigma E_2,
 \end{aligned}$$

car, en approximant par une série géométrique avec un σ petit, nous avons que

$$\sum_{i=1}^n \frac{2}{i(i+1) + (i+1)\sigma} = \sum_{i=1}^n \frac{2}{i(i+1)} \left(\frac{1}{1 + \frac{\sigma}{i}} \right) \approx \sum_{i=1}^n \frac{2}{i(i+1)} \left(1 - \frac{\sigma}{i} \right)$$

Nous avons donc que

$$E_1 = 2 \sum_{i=1}^n \left(\frac{1}{i} - \frac{1}{i+1} \right) \rightarrow 2,$$

lorsque $n \rightarrow \infty$. De même, nous avons que

$$E_2 = 2 \sum_{i=1}^n \left(-\frac{1}{i(i+1)} + \frac{1}{i^2} \right) \rightarrow \frac{\pi^2}{3} - 2,$$

lorsque $n \rightarrow \infty$, car $\sum_{i \geq 1} \frac{1}{i^2} = \frac{\pi^2}{6}$.

La probabilité totale pour ce cas est quant à elle calculée par le produit de la probabilité que chaque état transite vers le prochain état comme illustré sur la figure 2.1. Comme nous considérons que la probabilité qu'une deuxième mutation affecte le même locus est très petite, nous approximons que le taux

de mutation pour les événements subséquents est nul. La probabilité donne alors

$$\begin{aligned}
B &\approx \frac{\mu}{2 + 2\sigma + 2\mu} \prod_{i=2}^n \frac{i(i-1)}{i(i+1) + (i+1)\sigma} \\
&\approx \frac{\prod_{i=2}^n i(i-1)}{\prod_{i=2}^n (i(i+1) + (i+1)\sigma)} \times \frac{\mu}{2 + 2\sigma} \\
&= \frac{\prod_{i=1}^{n-1} i(i+1)}{\prod_{i=2}^n (i(i+1) + (i+1)\sigma)} \times \frac{\mu}{2 + 2\sigma} \\
&= \frac{\prod_{i=1}^{n-1} i(i+1)}{\prod_{i=1}^{n-1} (i(i+1) + (i+1)\sigma)} \times \frac{\mu}{n(n+1) + (n+1)\sigma} \\
&= \prod_{i=1}^{n-1} \left(\frac{1}{1 + \frac{\sigma}{i}} \right) \frac{\mu}{n(n+1) \left(1 + \frac{\sigma}{n}\right)} \\
&= \frac{\mu}{n(n+1)} \prod_{i=1}^n \frac{1}{1 + \frac{\sigma}{i}} \\
&\approx \frac{\mu}{n(n+1)} \prod_{i=1}^n \left(1 - \frac{\sigma}{i}\right) \\
&\approx \frac{\mu}{n(n+1)} \left(1 - \sigma \sum_{i=1}^n \frac{1}{i}\right) \\
&= B_1 - \sigma B_2,
\end{aligned}$$

avec $B_1 = \frac{\mu}{n(n+1)}$, $B_2 = B_1 \sum_{i=1}^n \frac{1}{i} = B_1 S_0$ et $S_0 = \sum_{i=1}^n \frac{1}{i}$.

Cas 2

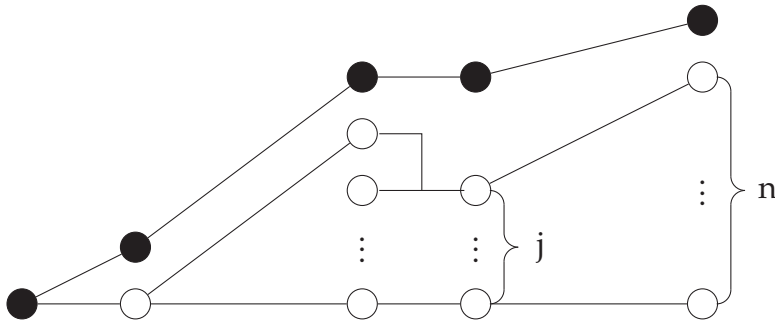


FIGURE 2.2. Graphe de sélection ancestral avec un événement de sélection sur une séquence mutante.

Les calculs dans ce cas se font de façon semblable à ceux du cas précédent. Ainsi, l'espérance du temps de fixation dans ce cas est

$$\begin{aligned}
 E_{\text{Cas2}} &= E + \frac{2}{(j+1)(j+2) + (j+2)\sigma} + \frac{2}{j(j+1) + (j+1)\sigma} \\
 &\approx E + \frac{2}{(j+1)(j+2)} \left(1 - \frac{\sigma}{j+1}\right) + \frac{2}{j(j+1)} \left(1 - \frac{\sigma}{j}\right) \\
 &\approx E + \frac{2}{(j+1)(j+2)} + \frac{2}{j(j+1)},
 \end{aligned}$$

pour $j = 1, \dots, n$. La probabilité d'un tel arbre est donnée par

$$\begin{aligned}
 B_{\text{Cas2}} &= B \times \frac{j\sigma}{j(j+1) + (j+1)\sigma} \times \frac{j(j+1)}{(j+1)(j+2) + (j+2)\sigma} \\
 &\approx \frac{Bj\sigma}{(j+1)(j+2)},
 \end{aligned}$$

pour $j = 1, \dots, n$ et avec B la probabilité totale du cas 1.

Cas 3

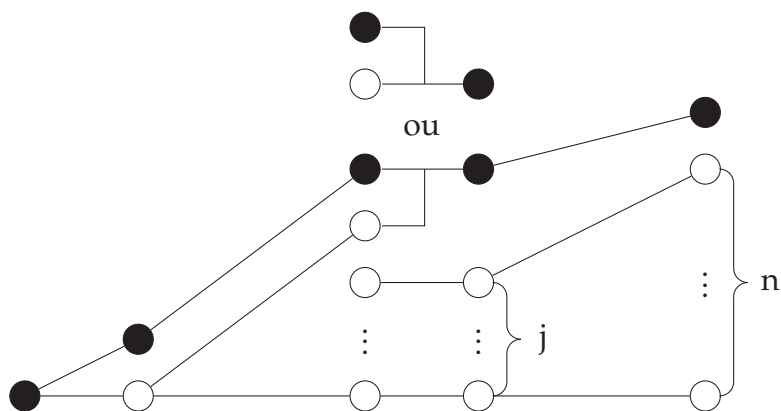


FIGURE 2.3. Graphe de sélection ancestral avec un événement de sélection sur la séquence non mutante et avec une séquence additionnelle mutante.

Les calculs dans ce cas se font de façon analogue à ceux du cas précédent. Ainsi, l'espérance du temps de fixation dans ce cas est

$$E_{\text{Cas3}} = E + \frac{2}{(j+1)(j+1) + (j+2)\sigma} + \frac{2}{j(j+1) + (j+1)\sigma}$$

$$\approx E + \frac{2}{(j+1)(j+2)} + \frac{2}{j(j+1)},$$

pour $j = 1, \dots, n$. La probabilité d'un tel arbre est donnée par

$$B_{\text{Cas3}} = B \times \frac{\sigma}{j(j+1) + (j+1)\sigma} \times \frac{j(j+1)}{(j+1)(j+2) + (j+2)\sigma} \times 2$$

$$\approx \frac{B2\sigma}{(j+1)(j+2)},$$

pour $j = 1, \dots, n$ et avec B la probabilité totale du Cas 1.

Cas 4

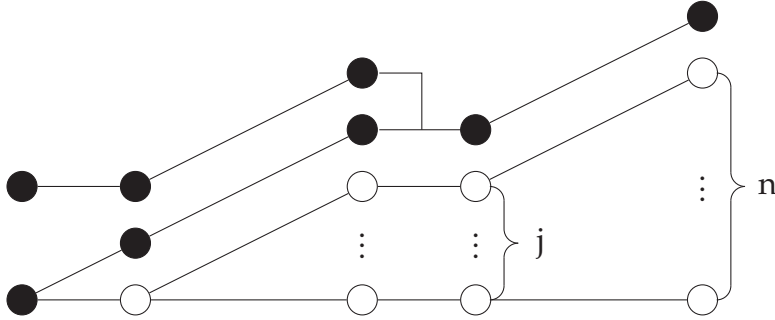


FIGURE 2.4. Graphe de sélection ancestral avec un événement de sélection sur la séquence non mutante, avec une séquence aditionnelle non mutante et sans coalescence entre les deux.

Les calculs dans ce cas se font comme ceux du cas précédent. Ainsi, l'espérance du temps de fixation dans ce cas est

$$E_{\text{Cas4}} = E + \frac{2}{(j+1)(j+1) + (j+2)\sigma} + \frac{2}{j(j+1) + (j+1)\sigma} - \frac{2}{2+2\sigma}$$

$$\approx E + \frac{2}{(j+1)(j+2)} + \frac{2}{j(j+1)} - 1,$$

pour $j = 1, \dots, n$. La probabilité d'un tel arbre est donnée par

$$B_{\text{Cas4}} = \prod_{i=j+1}^n \frac{i(i-1)}{i(i+1) + (i+1)\sigma} \times \prod_{i=2}^j \frac{i(i-1)}{(i+1)(i+2) + (i+2)\sigma}$$

$$\times \frac{\sigma}{j(j+1) + (j+1)\sigma} \times \frac{\mu}{2 \times 3 + 3\sigma}$$

$$= \frac{\prod_{i=2}^n i(i-1)}{\prod_{i=2}^n (i(i+1) + (i+1)\sigma)} \times \sigma\mu \times \frac{1}{(j+1)(j+2) + (j+2)\sigma}$$

$$\times \frac{1}{j(j+1) + (j+1)\sigma}$$

$$\approx \frac{2(1+\sigma)B}{\mu} \times \sigma\mu \times \frac{1}{(j+1)(j+2)} \times \frac{1}{j(j+1)}$$

$$\approx \frac{2\sigma B}{j(j+1)^2(j+2)},$$

pour $j = 1, \dots, n$ et avec B la probabilité totale du Cas 1.

Cas 5

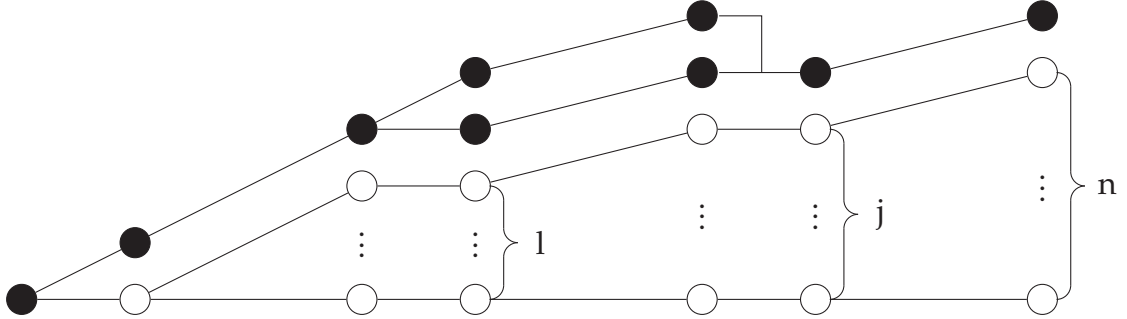


FIGURE 2.5. Graphe de sélection ancestral avec un événement de sélection sur la séquence non mutante, avec une séquence aditionnelle non mutante et avec coalescence entre les deux.

Les calculs dans ce cas se font comme ceux du cas précédent. Ainsi, l'espérance du temps de fixation est

$$E_{\text{Cas5}} = E + \frac{2}{(j+1)(j+1) + (j+2)\sigma} + \frac{2}{j(j+1) + (j+1)\sigma}$$

$$\approx E + \frac{2}{(j+1)(j+2)} + \frac{2}{j(j+1)},$$

pour $l = 1, \dots, j$ et $j = 1, \dots, n$. La probabilité d'un tel arbre est donnée par

$$B_{\text{Cas5}} = \prod_{i=j+1}^n \frac{i(i-1)}{i(i+1) + (i+1)\sigma} \times \prod_{i=l+1}^j \frac{i(i-1)}{(i+1)(i+2) + (i+2)\sigma}$$

$$\times \prod_{i=2}^l \frac{i(i-1)}{i(i+1) + (i+1)\sigma} \times \frac{\sigma}{j(j+1) + (j+1)\sigma}$$

$$\times \frac{1 \times 2}{(l+1)(l+2) + (l+2)\sigma} \times \frac{\mu}{1 \times 2 + 2\sigma}$$

$$= \frac{\prod_{i=2}^n i(i-1)}{\prod_{i=2}^n (i(i+1) + (i+1)\sigma)} \times 2\sigma\mu \times \frac{1}{(j+1)(j+2) + (j+2)\sigma}$$

$$\times \frac{1}{j(j+1) + (j+1)\sigma} \times \frac{1}{2 + 2\sigma}$$

$$\approx \frac{2(1+\sigma)B}{\mu} \times 2\sigma\mu \times \frac{1}{(j+1)(j+2)} \times \frac{1}{j(j+1)} \times \frac{1}{2}$$

$$\approx \frac{2\sigma B}{j(j+1)^2(j+2)},$$

pour $l = 1, \dots, j$ et $j = 1, \dots, n$, et avec B la probabilité totale du Cas 1.

2.1.1. Espérance totale

Nous allons calculer ici l'espérance totale du temps de fixation, c'est-à-dire la somme sur tous les cas du produit entre l'espérance du temps pour le cas et sa probabilité associée. Ainsi, nous avons

$$\begin{aligned}
E_{\text{Totale}} &\approx E \times B + \sum_{j=1}^n \left(E + \frac{2}{(j+1)(j+2)} + \frac{2}{j(j+1)} \right) \frac{Bj\sigma}{(j+1)(j+2)} \\
&+ \sum_{j=1}^n \left(E + \frac{2}{(j+1)(j+2)} + \frac{2}{j(j+1)} \right) \frac{B2\sigma}{(j+1)(j+2)} \\
&+ \sum_{j=1}^n \left(E + \frac{2}{(j+1)(j+2)} + \frac{2}{j(j+1)} - 1 \right) \frac{B2\sigma}{j(j+1)^2(j+2)} \\
&+ \sum_{j=1}^n \sum_{l=1}^j \left(E + \frac{2}{(j+1)(j+2)} + \frac{2}{j(j+1)} \right) \frac{B2\sigma}{j(j+1)^2(j+2)} \\
&\approx E_1 B_1 - \sigma E_2 B_1 - \sigma E_1 B_2 \\
&+ \sigma B_1 \sum_{j=1}^n \left(E_1 + \frac{2}{(j+1)(j+2)} + \frac{2}{j(j+1)} \right) \frac{j^2(j+1) + 2j(j+1) + 2 + 2j}{j(j+1)^2(j+2)} \\
&- \sigma B_1 \sum_{j=1}^n \frac{2}{j(j+1)^2(j+2)} \\
&= E_1 B_1 - \sigma (E_2 B_1 + E_1 B_2 - B_1 S_1 + B_1 S_2) \\
&= B_1 (E_1 - \sigma (E_2 + E_1 S_0 - S_1 + S_2)),
\end{aligned}$$

avec

$$\begin{aligned}
\frac{j^2(j+1) + 2j(j+1) + 2 + 2j}{j(j+1)^2(j+2)} &= \frac{j^2 + 2j + 2}{j(j+1)(j+2)} = \frac{1}{j} - \frac{1}{j+1} + \frac{1}{j+2}, \\
S_1 &= \sum_{j=1}^n \left(E_1 + \frac{2}{(j+1)(j+2)} + \frac{2}{j(j+1)} \right) \left(\frac{1}{j} - \frac{1}{j+1} + \frac{1}{j+2} \right),
\end{aligned}$$

et finalement

$$S_2 = \sum_{j=1}^n \frac{2}{j(j+1)^2(j+2)}.$$

Si nous détaillons S_1 davantage, nous avons

$$\begin{aligned}
S_1 &= \sum_{j=1}^n \left(E_1 + \frac{2}{(j+1)(j+2)} + \frac{2}{j(j+1)} \right) \left(\frac{1}{j} - \frac{1}{j+1} + \frac{1}{j+2} \right) \\
&= E_1 A + 2 \sum_{j=1}^n \left(\frac{j+j+2}{j(j+1)(j+2)} \right) \left(\frac{1}{j} - \frac{1}{j+1} + \frac{1}{j+2} \right) \\
&= E_1 A + \sum_{j=1}^n \left(\frac{4}{j(j+2)} \right) \left(\frac{1}{j} - \frac{1}{j+1} + \frac{1}{j+2} \right) \\
&= E_1 A + S_3,
\end{aligned}$$

$$\text{avec } A = \sum_{j=1}^n \left(\frac{1}{j} - \frac{1}{j+1} + \frac{1}{j+2} \right) \text{ et } S_3 = \sum_{j=1}^n \left(\frac{4}{j(j+2)} \right) \left(\frac{1}{j} - \frac{1}{j+1} + \frac{1}{j+2} \right).$$

2.1.2. Probabilité totale

Nous allons calculer ici la probabilité totale de fixation, soit la somme des probabilités associées à chaque cas. Nous avons donc

$$\begin{aligned}
B_{\text{Totale}} &= B + \sum_{j=1}^n \frac{Bj\sigma}{(j+1)(j+2)} + \sum_{j=1}^n \frac{B2\sigma}{(j+1)(j+2)} \\
&\quad + \sum_{j=1}^n \frac{B2\sigma}{j(j+1)^2(j+2)} + \sum_{j=1}^n \sum_{l=1}^j \frac{B2\sigma}{j(j+1)^2(j+2)} \\
&\approx B_1 - \sigma B_2 + \sigma B_1 \sum_{j=1}^n \left(\frac{j}{(j+1)(j+2)} + \frac{2}{(j+1)(j+2)} \right. \\
&\quad \left. + \frac{2}{j(j+1)^2(j+2)} + \frac{2j}{j(j+1)^2(j+2)} \right) \\
&= B_1 - \sigma B_2 + \sigma B_1 \sum_{j=1}^n \frac{j^2(j+1) + 2j(j+1) + 2 + 2j}{j(j+1)^2(j+2)} \\
&= B_1 - \sigma B_2 + \sigma B_1 \sum_{j=1}^n \left(\frac{1}{j} - \frac{1}{j+1} + \frac{1}{j+2} \right) \\
&= B_1 - \sigma(B_2 - B_1 A) \\
&= B_1(1 - \sigma(S - A)).
\end{aligned}$$

2.1.3. Espérance du temps de fixation conditionnelle

Nous allons maintenant calculer l'espérance du temps de fixation conditionnellement à ce que la fixation survienne. On a

$$\begin{aligned}
\mathbb{E}(T \mid \text{fix}) &= \frac{E_{\text{Totale}}}{B_{\text{Totale}}} \\
&\approx \frac{B_1(E_1 - \sigma(E_2 + E_1 S_0 - S_1 + S_2))}{B_1(1 - \sigma(S - A))} \\
&= \frac{E_1 - \sigma(E_2 + E_1 S_0 - S_1 + S_2)}{1 - \sigma(S - A)} \\
&\approx (E_1 - \sigma(E_2 + E_1 S_0 - S_1 + S_2))(1 + \sigma(S - A)) \\
&\approx E_1 - \sigma(E_2 + E_1 S_0 - S_1 + S_2 - E_1 S_0 + E_1 A) \\
&\approx 2 - \sigma(E_2 - S_1 + S_2 + E_1 A) \\
&= 2 - \sigma(E_2 - S_3 + S_2).
\end{aligned}$$

Afin d'être cohérent avec l'article de Mano (2009), le coefficient $(E_2 - S_3 + S_2)$ devrait être nul, ce que nous allons vérifier. Tout d'abord, exprimons S_3 sous une autre forme. On a

$$\begin{aligned}
S_3 &= \sum_{j=1}^n \left(\frac{2}{(j+1)(j+2)} + \frac{2}{j(j+1)} \right) \left(\frac{1}{j} - \frac{1}{(j+1)(j+2)} \right) \\
&= \sum_{j=1}^n \frac{2}{j(j+1)(j+2)} + \sum_{j=1}^n \frac{2}{j^2(j+1)} - \sum_{j=1}^n \frac{2}{(j+1)^2(j+2)^2} - \sum_{j=1}^n \frac{2}{j(j+1)^2(j+2)} \\
&= \sum_{j=1}^n \frac{2}{j(j+1)(j+2)} + E_2 - \sum_{j=1}^n \frac{2}{(j+1)^2(j+2)^2} - S_2.
\end{aligned}$$

Ainsi,

$$\begin{aligned}
E_2 - S_3 + S_2 &= 2S_2 + \sum_{j=1}^n \frac{2}{(j+1)^2(j+2)^2} - \sum_{j=1}^n \frac{2}{j(j+1)(j+2)} \\
&= \sum_{j=1}^n \frac{4}{j(j+1)^2(j+2)} + \sum_{j=1}^n \frac{2}{(j+1)^2(j+2)^2} - \sum_{j=1}^n \frac{2}{j(j+1)(j+2)}.
\end{aligned}$$

Développons chaque fraction en somme de fractions partielles. Pour la première fraction, nous obtenons

$$\begin{aligned}
 & \sum_{j=1}^n \frac{1}{j(j+1)^2(j+2)} \\
 &= \left(\frac{a}{j} + \frac{b}{j+1} + \frac{c}{(j+1)^2} + \frac{d}{j+2} \right) \\
 &= \frac{a(j+1)^2(j+2) + bj(j+1)(j+2) + cj(j+2) + dj(j+1)^2}{j(j+1)^2(j+2)} \\
 &= \frac{a(j^3 + 4j^2 + 5j + 2) + b(j^3 + 3j^2 + 2j) + c(j^2 + 2j) + d(j^3 + 2j^2 + j)}{j(j+1)^2(j+2)} \\
 &= \frac{j^3(a+b+d) + j^2(4a+3b+c+2d) + j(5a+2b+2c+d) + 2a}{j(j+1)^2(j+2)}.
 \end{aligned}$$

Nous obtenons donc le système d'équation

$$1a + 1b + 0c + 1d = 0;$$

$$4a + 3b + 1c + 2d = 0;$$

$$5a + 2b + 2c + 1d = 0;$$

$$2a + 0b + 0c + 0d = 1.$$

En résolvant ce système, nous obtenons que $a = \frac{1}{2}$, $b = 0$, $c = -1$ et $d = -\frac{1}{2}$, et ainsi que

$$\sum_{j=1}^n \frac{4}{j(j+1)^2(j+2)} = \sum_{j=1}^n \left(\frac{2}{j} - \frac{4}{(j+1)^2} - \frac{2}{j+2} \right).$$

Par le même procédé, nous obtenons le développement des deuxième et troisième fractions, soient

$$\sum_{j=1}^n \frac{2}{(j+1)^2(j+2)^2} = \sum_{j=1}^n \left(-\frac{4}{j+1} + \frac{2}{(j+1)^2} + \frac{4}{j+2} + \frac{2}{(j+2)^2} \right)$$

et

$$\sum_{j=1}^n \frac{2}{j(j+1)^2(j+2)} = \sum_{j=1}^n \left(\frac{1}{j} - \frac{2}{j+1} + \frac{1}{j+2} \right).$$

Finalement, nous avons que

$$\begin{aligned}
E_2 - S_3 + S_2 &= \sum_{j=1}^n \left(\frac{2}{j} - \frac{4}{(j+1)^2} - \frac{2}{j+2} - \frac{4}{j+1} + \frac{2}{(j+1)^2} \right. \\
&\quad \left. + \frac{4}{(j+2)} + \frac{2}{(j+2)^2} + \frac{1}{j} - \frac{2}{j+1} + \frac{1}{j+2} \right) \\
&= \sum_{j=1}^n \left(\frac{1}{j} - \frac{1}{j+1} \right) - \sum_{j=1}^n \left(\frac{1}{j+1} - \frac{1}{j+2} \right) \\
&\quad - 2 \sum_{j=1}^n \frac{1}{(j+1)^2} + 2 \sum_{j=1}^n \frac{1}{(j+2)^2} \\
&= 1 - \frac{1}{n+1} - \frac{1}{2} + \frac{1}{n+2} - 2 \sum_{j=1}^n \frac{1}{(j+1)^2} + 2 \sum_{j=1}^n \frac{1}{(j+1)^2} - \frac{1}{2} \\
&= 0 - \frac{1}{n+1} + \frac{1}{n+2} \\
&= -\frac{1}{(n+1)(n+2)} \\
&\rightarrow 0,
\end{aligned}$$

lorsque $n \rightarrow \infty$. La nouvelle approche est donc cohérente avec l'approche de Mano (2009).

2.2. SIMULATIONS

Plusieurs simulations ont été effectuées afin de vérifier si les résultats théoriques obtenus sont cohérents avec la réalité. Ainsi, nous avons construit un nombre N de graphes de sélection ancestraux débutant par un nombre n de séquences mutantes et 1 séquence non mutante, soumis à un taux de sélection s . Nous avons par la suite conservé seulement les graphes qui coalescaient complètement, c'est-à-dire en écartant ceux où un événement qui est impossible dans le cas de la fixation survenait. Nous nous sommes bien entendu intéressés au temps moyen de fixation, ou de coalescence complète, des graphes, mais aussi au ratio de graphes que nous avons conservé. Nous nous sommes intéressés au cas neutre, c'est-à-dire sans sélection, au cas où un seul événement de sélection peut survenir, mais aussi à celui où il y avait deux événements de sélection possibles, et ce, en présence de forte sélection. Voici les résultats des différentes simulations.

TABLEAU 2.1. Résultat des simulations sur le temps conditionnel de fixation et la probabilité de fixation en présence de sélection sur 10000000 itérations.

n	s	Nombre d'événements de sélection	Temps moyen	Ratio
5	0	0	1,667148	0,0666095
5	0,2	1	1,613100	0,0611326
5	0,2	2	1,648649	0,0596328
20	0	0	1,913564	0,0047360
20	0,2	1	1,817713	0,0043128
20	0,2	2	1.876615	0.0041751

Considérons tout d'abord le cas où n est petit, soit $n = 5$. Pour le temps de fixation dans le cas neutre, nous nous attendions à obtenir $2(1 - \frac{1}{6}) = \frac{5}{3}$. Nous considérons donc que nous sommes près du but. Dans la section 2.1.3, nous avons vu que pour un n quelconque, le coefficient devant le paramètre de sélection en puissance 1 est approximativement $-\frac{1}{(n+1)(n+2)}$. Ainsi, pour $n = 5$, en nous basant sur le résultat de Mano (2009) cité au début du chapitre, nous nous attendions à un résultat de

$$2(1 - \frac{1}{6}) - \frac{0,2}{6 * 7} - \frac{4}{9}0,2^2 = 1,644127.$$

Nous avons obtenu par simulation 1,613100 et 1,648649, ce qui est très près du résultat théorique espéré.

Considérant maintenant le cas où n est grand, soit $n = 20$, les résultats des simulations sont aussi cohérents avec la théorie. En effet, dans le cas neutre, nous obtenons 1,913564, ce qui est très près du $2(1 - \frac{1}{21}) = 1,904762$ espéré. De plus, nous nous attendions à obtenir

$$2(1 - \frac{1}{21}) - \frac{4}{9}0,2^2 = 1,886984.$$

Ce résultat est suffisamment près des résultats obtenus pour que nous en soyons satisfaits. Nous pouvons donc en conclure que les simulations sont cohérentes avec la théorie, même en présence de forte sélection.

Chapitre 3

TEMPS DE FIXATION AVEC CONVERSION GÉNIQUE

3.1. INTRODUCTION À LA CONVERSION GÉNIQUE

La conversion génique est un phénomène biologique qui, lorsqu'elle est biaisée, s'apparente fortement à la sélection naturelle. Elle survient lors de la méiose, c'est-à-dire lors de la création des cellules reproductrices chez les êtres vivants. Lors de celle-ci, il survient parfois que deux séquences d'ADN homologues se retrouvent l'une à côté de l'autre. Si l'une des séquences est endommagée, il arrive qu'elle se fasse réparer à partir de la deuxième, transformant alors la première en une copie conforme de la seconde. Il a cependant été noté que la paire de nucléotide G–C est préférée à la paire A–T lors de la réparation de l'ADN, créant ainsi un biais dans le résultat de la conversion. La conversion génique peut aussi survenir si un premier locus d'une séquence génétique se retrouve collée à un deuxième locus de la même séquence, par une sorte de boucle. C'est ce cas particulier qui sera analysé dans ce mémoire.

Plusieurs articles mathématiques ont été écrits sur ce phénomène, notamment Nagylaki et Petes (1982), Nagylaki (1984), Innan (2002), Innan (2003a), Innan (2003b) ainsi que Innan et Mano (2008) et Mano (2013). Dans plusieurs de ces articles, la conversion est neutre et survient à un taux c par séquence par génération. Dans le cas de modèles à deux loci ayant à chaque locus l'un des deux allèles A ou a , seules les séquences Aa ou aA permettent la conversion. Lorsque c'est le cas, elles deviennent aa ou AA avec la même probabilité.

Il arrive cependant qu'une conversion génique « préfère » une paire de nucléotide à une autre, comme mentionné précédemment. Dans ce cas, nous appelons ce phénomène la conversion génique biaisée. Elle est entre autres étudiée dans l'article de Mano (2013). Si la conversion a lieu sur une séquence Aa

ou aA , alors le résultat sera aa ou AA , mais avec probabilités $\frac{1+b}{2}$ et $\frac{1-b}{2}$, respectivement, où $-1 \leq b \leq 1$. C'est en raison de la préférence envers l'un des deux allèles que ce type de conversion est proche du phénomène de sélection.

C'est en raison de son actualité, démontrée par la date de parution de certains articles mentionnés, ainsi que pour sa proximité avec l'effet de sélection que la conversion génique est le sujet de cette section. Puisque les résultats sont seulement démontrés pour des taux de petite taille, des simulations ont été faites afin d'évaluer les tendances sur une plus grande échelle.

3.2. FIXATION D'UN NOUVEAU MUTANT EN PRÉSENCE DE CONVERSION GÉNIQUE BIAISÉE

3.2.1. Modèle du processus

Nous étudierons la probabilité de fixation d'un nouveau mutant en présence de conversion génique sur un modèle de Moran basé sur une population haploïde de taille $2N$ à deux loci. Il y a donc $4N$ loci. Nous considérons deux allèles, A et a . Nous considérons $0 \leq r \leq 0,5$, le taux de recombinaison par nouvelle gamète par génération, $0 < c < 1$, le taux de conversion par gamète par génération ainsi que $-1 < b < 1$, le biais de conversion lorsque le phénomène a lieu. Si b est positif, la conversion avantage le type mutant, alors que s'il est négatif, elle avantage le type non mutant. Le cas où $b = 0$ est le cas neutre. Il y a quatre types de séquence possibles : AA , Aa , aA et aa . Soient x_1 , x_2 , x_3 et x_4 les fréquences respectives de ces types dans la population actuelle. Ainsi, la première étape consiste à regrouper les gamètes par paire. Voici donc dans le tableau 3.1 les paires de gamètes ainsi que leur probabilité respective :

TABLEAU 3.1. Probabilité d'association de chaque paire de gamètes.

Paire	Probabilité
$AA - AA$	x_1^2
$AA - Aa$	$2x_1x_2$
$AA - aA$	$2x_1x_3$
$AA - aa$	$2x_1x_4$
$Aa - Aa$	x_2^2
$Aa - aA$	$2x_2x_3$
$Aa - aa$	$2x_2x_4$
$aA - aA$	x_3^2
$aA - aa$	$2x_3x_4$
$aa - aa$	x_4^2

Par la suite, les paires de gamètes sont soumises à la recombinaison dont les effets sont décrits dans le tableau 3.2.

TABLEAU 3.2. Effet de la recombinaison par paire.

Paire	gamète résultant	Probabilité
AA – AA	AA	1
AA – Aa	AA	0.5
	Aa	0.5
AA – aA	AA	0.5
	aA	0.5
AA – aa	AA	0.5(1-r)
	Aa	0.5r
	aA	0.5r
	aa	0.5(1-r)
Aa – Aa	Aa	1
Aa – aA	AA	0.5r
	Aa	0.5(1-r)
	aA	0.5(1-r)
	aa	0.5r
Aa – aa	Aa	0.5
	aa	0.5
aA – aA	aA	1
aA – aa	aA	0.5
	aa	0.5
aa – aa	aa	1

Finalement, les gamètes résultants sont soumis à la conversion biaisée, comme décrit dans le tableau 3.3.

TABLEAU 3.3. Effet de la conversion par type de gamète.

gamète mère	gamète résultant	Probabilité
Aa	AA	0.5c(1 + b)
	Aa	1 – c
	aa	0.5c(1 – b)
aA	AA	0.5c(1 + b)
	aA	1 – c
	aa	0.5c(1 – b)

3.2.2. Calcul de la probabilité du gamète remplaçant

Calculons maintenant la probabilité que le gamète choisi pour en remplacer un autre dans la prochaine génération soit de type AA. On a

$$\begin{aligned}
P_1 &= x_1^2 + 2x_1x_2 \left[\frac{1}{2} + \frac{1}{2} \left(c \frac{1+b}{2} \right) \right] + 2x_1x_3 \left[\frac{1}{2} + \frac{1}{2} \left(c \frac{1+b}{2} \right) \right] \\
&+ 2x_1x_4 \left[\frac{1-r}{2} + \frac{r}{2} \left(c \frac{1+b}{2} \right) + \frac{r}{2} \left(c \frac{1+b}{2} \right) \right] \\
&+ x_2^2 \left[c \frac{1+b}{2} \right] + 2x_2x_3 \left[\frac{r}{2} + \frac{1-r}{2} \left(c \frac{1+b}{2} \right) + \frac{1-r}{2} \left(c \frac{1+b}{2} \right) \right] \\
&+ 2x_2x_4 \left[\frac{c}{2} \frac{1+b}{2} \right] + x_3^2 \left[c \frac{1+b}{2} \right] + 2x_3x_4 \left[\frac{c}{2} \frac{1+b}{2} \right] \\
&= x_1 (x_1 + x_2 + x_3 + x_4) + c \frac{1+b}{2} [x_2 (x_1 + x_2 + x_3 + x_4) + x_3 (x_1 + x_2 + x_3 + x_4)] \\
&+ (x_2x_3 - x_1x_4) (r + -rc(1+b)).
\end{aligned}$$

Puisque $x_1 + x_2 + x_3 + x_4 = 1$, que $rc \ll r$ et que $D = x_1x_4 - x_2x_3$, nous obtenons que

$$P_1 \approx x_1 + c \frac{1+b}{2} (x_2 + x_3) - rD.$$

De façon semblable, nous obtenons les probabilités suivantes pour les types Aa, aA et aa, respectivement :

$$P_2 \approx (1-c)x_2 + rD,$$

$$P_3 \approx (1-c)x_3 + rD,$$

$$P_4 \approx x_4 + c \frac{1-b}{2} (x_2 + x_3) - rD.$$

3.2.3. Espérance conditionnelle de la fréquence des gamètes

Nous allons ici calculer la fréquence prochaine espérée des quatres types de gamètes au prochain état conditionnellement à l'état actuel. Nous avons que

$$x'_1 = \begin{cases} x_1 + \frac{1}{2N} & \text{avec probabilité } \mu_1 = (x_1 + c \frac{1+b}{2} (x_2 + x_3) - rD)(1 - x_1), \\ x_1 - \frac{1}{2N} & \text{avec probabilité } \nu_1 = (1 - (x_1 + c \frac{1+b}{2} (x_2 + x_3) - rD))x_1, \\ x_1 & \text{avec probabilité } 1 - \mu_1 - \nu_1. \end{cases}$$

Nous avons donc que

$$\begin{aligned}
\mathbb{E}(x'_1 | x_1) &= \left(x_1 + \frac{1}{2N}\right) (1 - x_1) \left(x_1 + c \frac{1+b}{2}(x_2 + x_3) - rD\right) \\
&+ \left(x_1 - \frac{1}{2N}\right) x_1 \left[1 - \left(x_1 + c \frac{1+b}{2}(x_2 + x_3) - rD\right)\right] \\
&+ x_1 \left\{1 - (1 - x_1) \left(x_1 + c \frac{1+b}{2}(x_2 + x_3) - rD\right) \right. \\
&\left. - x_1 \left[1 - \left(x_1 + c \frac{1+b}{2}(x_2 + x_3) - rD\right)\right]\right\} \\
&= x_1^2 + x_1 c \frac{1+b}{2}(x_2 + x_3) - x_1 rD - x_1^3 - x_1^2 c \frac{1+b}{2}(x_2 + x_3) + x_1^2 rD \\
&+ \frac{x_1}{2N} + \frac{c}{2N} \frac{1+b}{2}(x_2 + x_3) - \frac{rD}{2N} - \frac{x_1^2 c}{2N} \frac{1+b}{2}(x_2 + x_3) - \frac{x_1}{2N} rD \\
&+ x_1^2 - x_1^3 - x_1^2 c \frac{1+b}{2}(x_2 + x_3) + x_1^2 rD \\
&- \frac{x_1}{2N} + \frac{x_1^2}{2N} + \frac{x_1 c}{2N} \frac{1+b}{2}(x_2 + x_3) - \frac{x_1}{2N} rD \\
&+ x_1 - x_1^2 - x_1 c \frac{1+b}{2}(x_2 + x_3) + x_1 rD - x_1^2 \\
&+ 2 \left(x_1^3 + x_1^2 c \frac{1+b}{2}(x_2 + x_3) - x_1^2 rD\right) \\
&= x_1 + \frac{1}{2N} \left(c \frac{1+b}{2}(x_2 + x_3) - rD\right).
\end{aligned}$$

De façon semblable, nous obtenons que

$$\begin{aligned}
\mathbb{E}(x'_2 | x_2) &= x_2 + \frac{1}{2N}(rD - cx_2), \\
\mathbb{E}(x'_3 | x_3) &= x_3 + \frac{1}{2N}(rD - cx_3), \\
\mathbb{E}(x'_4 | x_4) &= x_4 + \frac{1}{2N} \left(c \frac{1-b}{2}(x_2 + x_3) - rD\right).
\end{aligned}$$

3.2.4. Calcul de la probabilité de fixation

Nous nous intéressons maintenant à la probabilité de fixation u d'un nouveau mutant dans la population. Considérons d'abord le vecteur des fréquences de AA , Aa , aA et aa , respectivement, à la génération t , donné par $x(t) =$

$(x_1(t), x_2(t), x_3(t), x_4(t))$. Comme déterminées à la section précédente, les fréquences espérées à la génération $t + 1$ sont alors

$$\begin{aligned}\mathbb{E}(x_1(t+1) | x_1(t)) &= x_1(t) + \frac{1}{2N} \left(c \frac{1+b}{2} (x_2(t) + x_3(t)) - rD \right), \\ \mathbb{E}(x_2(t+1) | x_2(t)) &= x_2(t) + \frac{1}{2N} (rD - cx_2(t)), \\ \mathbb{E}(x_3(t+1) | x_3(t)) &= x_3(t) + \frac{1}{2N} (rD - cx_3(t)), \\ \mathbb{E}(x_4(t+1) | x_4(t)) &= x_4(t) + \frac{1}{2N} \left(c \frac{1-b}{2} (x_2(t) + x_3(t)) - rD \right).\end{aligned}$$

Définissons $Z(t)$ comme la fréquence du mutant A au temps t , soit

$$Z(t) = x_1(t) + \frac{1}{2}x_2(t) + \frac{1}{2}x_3(t).$$

Nous avons aussi que

$$Z(T) = Z(0) + \sum_{t=1}^T \Delta Z(t),$$

avec $\Delta Z(t) = Z(t) - Z(t-1)$. Nous avons que

$$u = \mathbb{E}(Z(\infty)) = Z(0) + \sum_{t \geq 1} \mathbb{E}(\Delta Z(t)),$$

avec

$$\begin{aligned}\mathbb{E}(\Delta Z(t)) &= \mathbb{E}\left(\mathbb{E}(Z(t) - Z(t-1) | x(t-1))\right) \\ &= \mathbb{E}\left(\mathbb{E}(Z(t) | x(t-1)) - \mathbb{E}(Z(t-1) | x(t-1))\right) \\ &= \mathbb{E}\left(\mathbb{E}\left(x_1(t) + \frac{1}{2}x_2(t) + \frac{1}{2}x_3(t) | x(t-1)\right) \right. \\ &\quad \left. - \mathbb{E}\left(x_1(t-1) + \frac{1}{2}x_2(t-1) + \frac{1}{2}x_3(t-1) | x(t-1)\right)\right) \\ &= \mathbb{E}\left[x_1(t-1) + \frac{1}{2N} \left(c \frac{1+b}{2} (x_2(t-1) + x_3(t-1)) - rD \right) \right. \\ &\quad \left. + \frac{x_2(t-1)}{2} + \frac{1}{2N} \left(\frac{rD}{2} - \frac{c}{2}x_2(t-1) \right) + \frac{x_2(t-1)}{2} \right. \\ &\quad \left. + \frac{1}{2N} \left(\frac{rD}{2} - \frac{c}{2}x_2(t-1) \right) \right] \\ &\quad - x_1(t-1) - \frac{1}{2}x_2(t-1) - \frac{1}{2}x_3(t-1) \\ &= \mathbb{E}\left(\frac{cb}{4N} (x_2(t-1) + x_3(t-1)) \right).\end{aligned}$$

Nous avons donc que

$$u = Z(0) + \sum_{t \geq 0} \frac{cb}{4N} \mathbb{E}(x_2(t) + x_3(t)).$$

Supposons que nous tirons une séquence au hasard à la génération t . Alors nous pouvons soit tirer une séquence Aa , soit une séquence d'un autre type. Nous posons donc $\xi_2(t)$ comme l'indicatrice de cet événement, soit

$$\xi_2(t) = \begin{cases} 1 & \text{si on tire } Aa \text{ au temps } t, \\ 0 & \text{sinon.} \end{cases}$$

Nous avons que

$$\mathbb{E}(\xi_2(t)) = \mathbb{E}\left(\mathbb{E}(\xi_2(t) \mid x_2(t))\right) = \mathbb{E}(x_2(t)).$$

Il existe deux possibilités d'ancêtres pour les deux lignées A et a tirées à t lorsque $\xi_2(t) = 1$. La première est que les deux lignées proviennent de lignées ancêtres situées dans le même ordre que sur la séquence. Nous posons donc comme variable indicatrice

$$\eta_2(t) = \begin{cases} 1 & \text{si les deux lignées à } t \text{ ont deux lignées ancêtres à } 0 \\ & \text{dans le même ordre,} \\ 0 & \text{sinon.} \end{cases}$$

La deuxième possibilité est que les deux lignées proviennent de lignées ancêtres situées au même locus, soit au locus du nouveau mutant. Nous posons donc la variable indicatrice

$$\zeta_2(t) = \begin{cases} 1 & \text{si les deux lignées à } t \text{ ont deux lignées ancêtres à } 0 \\ & \text{situées au même locus,} \\ 0 & \text{sinon.} \end{cases}$$

Nous avons donc que

$$\sum_{t \geq 0} \mathbb{E}(\xi_2(t)) = \sum_{t \geq 0} \frac{1}{2N} \mathbb{E}(\eta_2(t)) + \sum_{t \geq 0} \frac{1}{2N} \mathbb{E}(\zeta_2(t)).$$

Si nous considérons que la séquence choisie au temps t a l'allèle représentée par un cercle au locus du haut et par une croix au locus du bas, les deux allèles pouvant être identiques, il existe huit possibilités de séquence pour ses ancêtres. La figure 3.1 présente les différents états possibles, numérotés de 1 à 8, pour les ancêtres d'une séquence choisie au temps t .

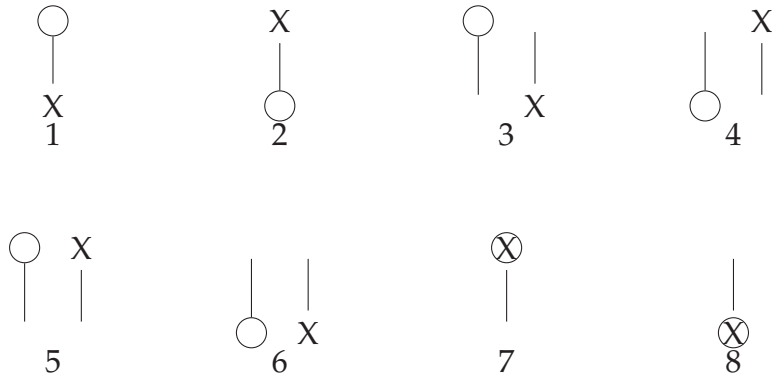


FIGURE 3.1. Représentation des états possibles pour les séquences ancêtres.

Les états 7 et 8 sont absorbants. Si nous supposons que Aa est l'état 1, nous avons que

$$\mathbb{E}(\eta_2(t)) = \Pr(1 \text{ et } 3 \text{ à } t \text{ générations antérieures})$$

et

$$\mathbb{E}(\zeta_2(t)) = \Pr(5 \text{ à } t \text{ générations antérieures}).$$

Considérons \mathbb{P} , la matrice de transition pour ces états, donnée par

$$\begin{pmatrix} 1-r-2c & 0 & r & 0 & 0 & 0 & c & c \\ 0 & 1-r-2c & 0 & r & 0 & 0 & c & c \\ \frac{1}{2N^2} & 0 & 1-\frac{1}{2N^2}-2c & 0 & c & c & 0 & 0 \\ 0 & \frac{1}{2N^2} & 0 & 1-\frac{1}{2N^2}-2c & c & c & 0 & 0 \\ 0 & 0 & c & c & 1-\frac{1}{2N^2}-2c & 0 & \frac{1}{2N^2} & 0 \\ 0 & 0 & c & c & 0 & 1-\frac{1}{2N^2}-2c & 0 & \frac{1}{2N^2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Nous avons alors que \mathbb{P}^t est la matrice de transition pour t générations. Nous

avons donc que

$$\sum_{t \geq 0} \mathbb{E}(\eta_2(t)) = \sum_{t \geq 0} [(\mathbb{P}^t)_{1,1} + (\mathbb{P}^t)_{1,3}]$$

et

$$\sum_{t \geq 0} \mathbb{E}(\zeta_2(t)) = \sum_{t \geq 0} (\mathbb{P}^t)_{1,5}.$$

Puisque les états 7 et 8 sont absorbants, nous pouvons seulement conserver la matrice \mathbb{Q} des états transitoires de \mathbb{P} donnée par

$$\begin{pmatrix} 1-r-2c & 0 & r & 0 & 0 & 0 \\ 0 & 1-r-2c & 0 & r & 0 & 0 \\ \frac{1}{2N^2} & 0 & 1-\frac{1}{2N^2}-2c & 0 & c & c \\ 0 & \frac{1}{2N^2} & 0 & 1-\frac{1}{2N^2}-2c & c & c \\ 0 & 0 & c & c & 1-\frac{1}{2N^2}-2c & 0 \\ 0 & 0 & c & c & 0 & 1-\frac{1}{2N^2}-2c \end{pmatrix}.$$

Ainsi, nous avons que

$$\begin{aligned} \sum_{t \geq 0} (\mathbb{P}^t)_{1,1} &= \left(\sum_{t \geq 0} \mathbb{P}^t \right)_{1,1} \\ &= \left(\sum_{t \geq 0} \mathbb{Q}^t \right)_{1,1} \\ &= \left((\mathbb{I} - \mathbb{Q})^{-1} \right)_{1,1}, \end{aligned}$$

puisque toutes les valeurs propres de \mathbb{Q} sont inférieures à 1 en module. De façon analogue, nous obtenons que $\sum_{t \geq 0} (\mathbb{P}^t)_{1,3} = \left((\mathbb{I} - \mathbb{Q})^{-1} \right)_{1,3}$ et $\sum_{t \geq 0} (\mathbb{P}^t)_{1,5} = \left((\mathbb{I} - \mathbb{Q})^{-1} \right)_{1,5}$. De la même façon, nous obtenons les résultats suivants pour $\mathbb{E}(x_3(t))$, soit

$$\begin{aligned} \sum_{t \geq 0} \mathbb{E}(\xi_3(t)) &= \sum_{t \geq 0} \frac{1}{2N} \mathbb{E}(\eta_3(t)) + \sum_{t \geq 0} \frac{1}{2N} \mathbb{E}(\zeta_3(t)), \\ \sum_{t \geq 0} \mathbb{E}(\eta_3(t)) &= \sum_{t \geq 0} [(\mathbb{P}^t)_{2,1} + (\mathbb{P}^t)_{2,3}] \end{aligned}$$

et

$$\sum_{t \geq 0} \mathbb{E}(\zeta_3(t)) = \sum_{t \geq 0} (\mathbb{P}^t)_{2,5}.$$

La matrice inverse étant trop complexe, seuls les éléments d'importance seront rapportés. Nous avons donc

$$\begin{aligned} \left((I - Q)^{-1} \right)_{1,1} &= \frac{1 + 32c^2n^4 + 2n^2r + 12c(n^2 + n^4r)}{2c(32c^2n^4 + (1 + 2n^2r)^2 + 12c(n^2 + 2n^4r))}, \\ \left((I - Q)^{-1} \right)_{1,3} &= \frac{n^2r(1 + 8c^2n^4 + 2n^2r + 4cn^2(2 + n^2r))}{c(32c^2n^4 + (1 + 2n^2r)^2 + 12c(n^2 + 2n^4r))}, \\ \left((I - Q)^{-1} \right)_{1,5} &= \frac{2n^4r}{1 + 8cn^2 + 2n^2r}, \\ \left((I - Q)^{-1} \right)_{2,1} &= \frac{2n^4r}{32c^2n^4 + (1 + 2n^2r)^2 + 12c(n^2 + 2n^4r)}, \\ \left((I - Q)^{-1} \right)_{2,3} &= \frac{2n^6r(2c + r)}{32c^2n^4 + (1 + 2n^2r)^2 + 12c(n^2 + 2n^4r)}, \\ \left((I - Q)^{-1} \right)_{2,5} &= \frac{2n^4r}{1 + 8cn^2 + 2n^2r}. \end{aligned}$$

Ainsi, nous avons que

$$\begin{aligned} \sum_{t \geq 0} \mathbb{E}(x_2(t)) &= \frac{1}{2N} \left(\left((I - Q)^{-1} \right)_{1,1} + \left((I - Q)^{-1} \right)_{1,3} + \left((I - Q)^{-1} \right)_{1,5} \right) \\ &= \frac{1 + 8cn^2 + 2n^2r + 8cn^4r}{4cn + 32c^2n^3 + 8cn^3r}, \end{aligned}$$

et aussi

$$\begin{aligned} \sum_{t \geq 0} \mathbb{E}(x_3(t)) &= \frac{1}{2N} \left(\left((I - Q)^{-1} \right)_{2,1} + \left((I - Q)^{-1} \right)_{2,3} + \left((I - Q)^{-1} \right)_{2,5} \right) \\ &= \frac{2n^3r}{1 + 8cn^2 + 2n^2r}. \end{aligned}$$

Nous obtenons donc finalement que la probabilité de fixation d'un nouvel allèle mutant dans la population est donnée par

$$\begin{aligned} u &= Z(0) + \sum_{t \geq 0} \frac{cb}{4N} \mathbb{E}(x_2(t) + x_3(t)) \\ &= \frac{1}{4N} + \frac{cb}{4N} \left(\frac{1 + 8cn^2 + 2n^2r + 8cn^4r}{4cn + 32c^2n^3 + 8cn^3r} + \frac{2n^3r}{1 + 8cn^2 + 2n^2r} \right) \\ &= \frac{1}{4N} + \frac{b(1 + 8cn^2)(1 + 2n^2r)}{16n^2(1 + 8cn^2 + 2n^2r)}. \end{aligned}$$

3.3. SIMULATIONS

Plusieurs simulations ont été faites afin de déterminer les tendances du temps de fixation moyen conditionnel à la fixation, ainsi que du nombre de fois sur les 1000000 itérations que l'allèle mutant s'est fixé, reflétant la probabilité de fixation. Les différents taux utilisés pour le biais b sont : -0,8 ; -0,4 ; 0 ; 0,4 et 0,8. Les taux utilisés pour le taux de conversion c sont : 0,2 ; 0,4 ; 0,6 et 0,8. Pour le taux recombinaison r , les taux utilisés sont : 0 ; 0,1 ; 0,2 ; 0,3 ; 0,4 et 0,5. Les résultats les plus intéressants sont affichés ici sous forme graphique afin de rendre plus aisée leur analyse. La totalité des résultats est donnée en annexe A.

Cas 1

Considérons tout d'abord le cas de conversion neutre, c'est-à-dire lorsque $b = 0$. Débutons par le graphique représentant le nombre d'occurrences de la fixation.

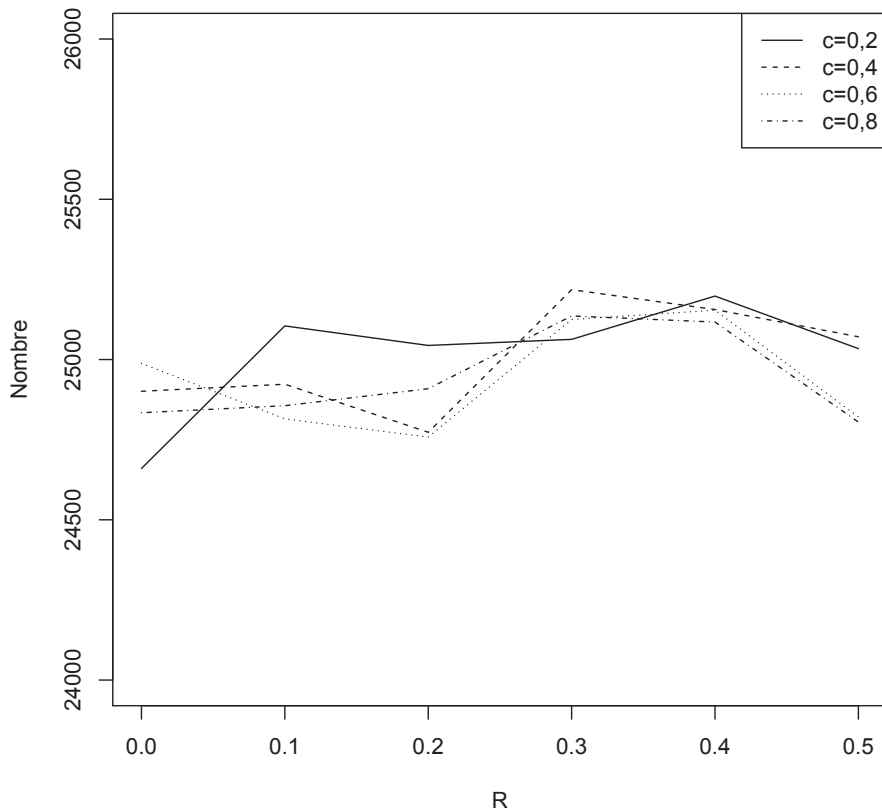


FIGURE 3.2. Nombre de fixations du mutant sur 1000000 pour $b=0$, avec c et r variables.

Nous pouvons constater à partir de la figure 3.2 que, comme nous nous y attendons, ni r ni c n'influence le nombre de fixations dans le cas neutre, qui est près de $\frac{1}{40}1000000 = 25000$.

Passons maintenant au temps de fixation conditionnel. Nous remarquons

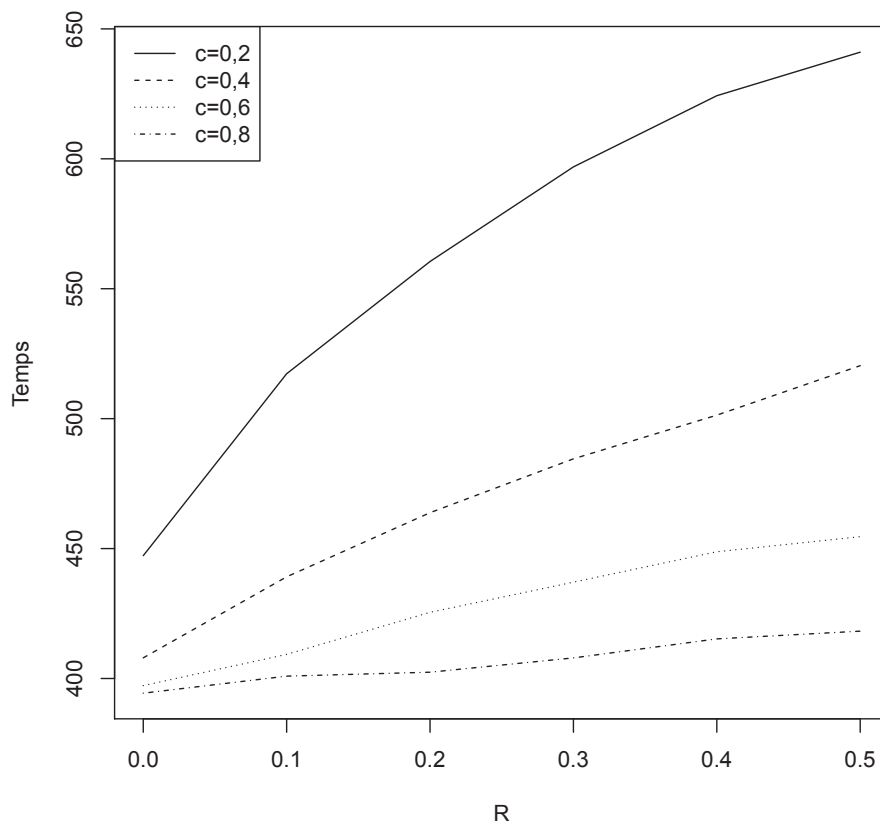


FIGURE 3.3. Temps de fixation conditionnel moyen du mutant sur 1000000 pour $b = 0$, avec c et r variables.

sur la figure 3.3 que, peu importe le c , plus r augmente, plus le temps de fixation conditionnel est long. Ceci est cohérent avec la théorie, car un plus petit taux de recombinaison préserve mieux les paires d'allèles mutants, diminuant ainsi le temps de fixation. Nous pouvons aussi constater que, peu importe le r , plus le c augmente, plus le temps de fixation diminue. En effet, si le taux de conversion augmente, alors les paires d'allèles mutants sont plus souvent créées, et ainsi la fixation survient plus rapidement, si elle a lieu.

Cas 2

Nous considérons ici le comportement de la proportion d'occurrences de la fixation pour un c fixe, par rapport à différents b et r (figure 3.4).

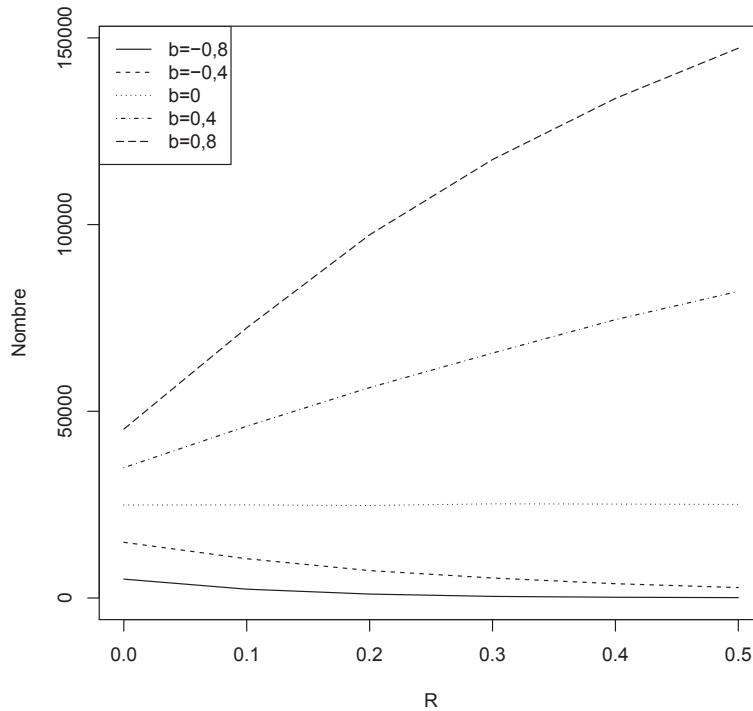


FIGURE 3.4. Nombre de fixations du mutant sur 1000000 pour $c=0,4$, avec b et r variables.

Puisqu'un b positif donne un avantage sélectif au mutant, il est normal de constater qu'une augmentation du b augmente la proportion d'occurrences de la fixation, peu importe le r . De plus, lorsque r augmente, si b est négatif, alors le nombre de fixations diminue, alors que si b est positif, celui-ci augmente. Cela s'explique par le fait que la recombinaison crée des séquences Aa , qui peuvent à leur tour subir la conversion. Ainsi, plus r augmente, plus souvent ces séquences sont créées, ce qui pousse de façon plus importante la population vers la fixation de l'allèle avantageé, soit la séquence mutante pour $b > 0$ et la non mutante pour $b < 0$. Puisque dans le cas neutre, la recombinaison n'affecte pas la probabilité de fixation, il est normal de voir un taux de changement nul par rapport à r lorsque $b = 0$.

Cas 3

Considérons maintenant le cas de la recombinaison libre, c'est-à-dire lorsque $r = 0,5$, pour b et c variable.

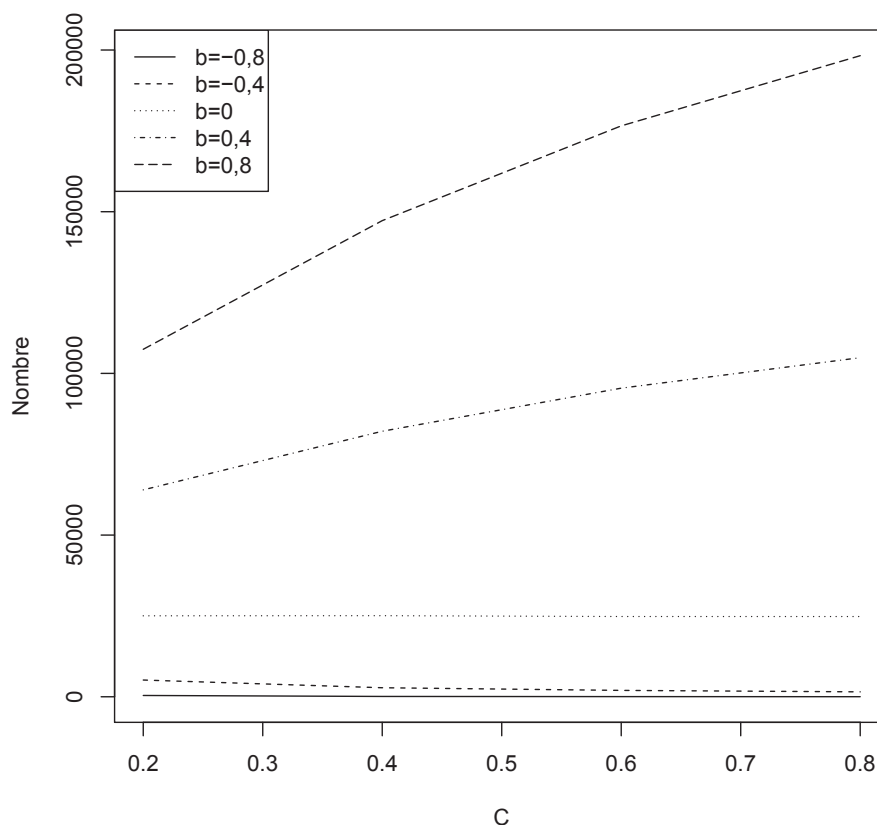


FIGURE 3.5. Nombre de fixations du mutant sur 1000000 pour $r=0,5$, avec b et c variables.

Nous remarquons dans la figure 3.5 que lorsque c augmente, le nombre de fixations du mutant augmente si $b > 0$ et diminue si $b < 0$. En effet, plus le taux de conversion est grand, plus souvent les séquences deviendront des paires d'allèles identiques, augmentant ainsi la probabilité de fixation de l'allèle favorisé. Bien entendu, nous pouvons remarquer que le taux de conversion n'affecte pas la proportion d'occurrences de la fixation dans le cas neutre. Nous pouvons aussi constater que lorsque le b augmente, alors le nombre de fixations du mutant augmente, peu importe le c .

Passons maintenant au temps de fixation conditionnel (figure 3.6).

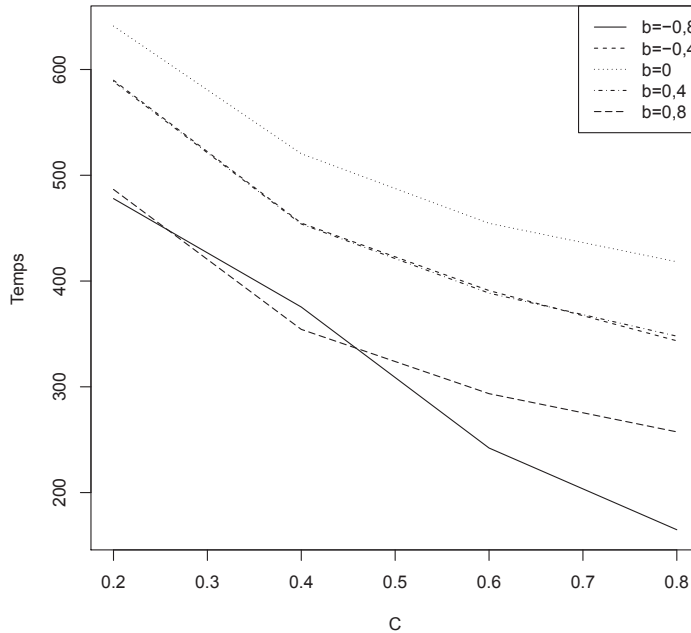


FIGURE 3.6. Temps de fixation conditionnel moyen du mutant sur 1000000 pour $r = 0,5$, avec b et c variables.

Nous remarquons tout d'abord que, peu importe le b , lorsque c augmente, le temps de fixation conditionnel diminue. On peut l'expliquer par le fait qu'un plus grand taux de conversion crée des paires d'allèles identiques plus rapidement. Ainsi, dans le cas où b est positif, l'allèle mutant est créé plus rapidement et donc le temps de fixation conditionnel diminue. Dans le cas où b est négatif, l'allèle non mutant est favorisé, et donc une augmentation du c implique que la fixation du mutant doit être plus rapide afin que celle-ci ait lieu. Le cas neutre a déjà été traité plus en détail. Ce graphique nous apprend aussi que plus b s'éloigne du 0, plus le temps de fixation conditionnel diminue, peu importe le c . En effet, en augmentant l'avantage du gène mutant lors d'un événement de conversion, les paires d'allèles mutants se créent plus souvent, diminuant le temps de fixation conditionnel. Dans le cas où on augmente l'avantage de l'allèle non mutant, la fixation doit se faire plus rapidement afin qu'il y ait fixation.

Cas 4

Considérons maintenant le cas sans recombinaison, c'est-à-dire lorsque $r = 0$, pour b et c variables (figure 3.7).

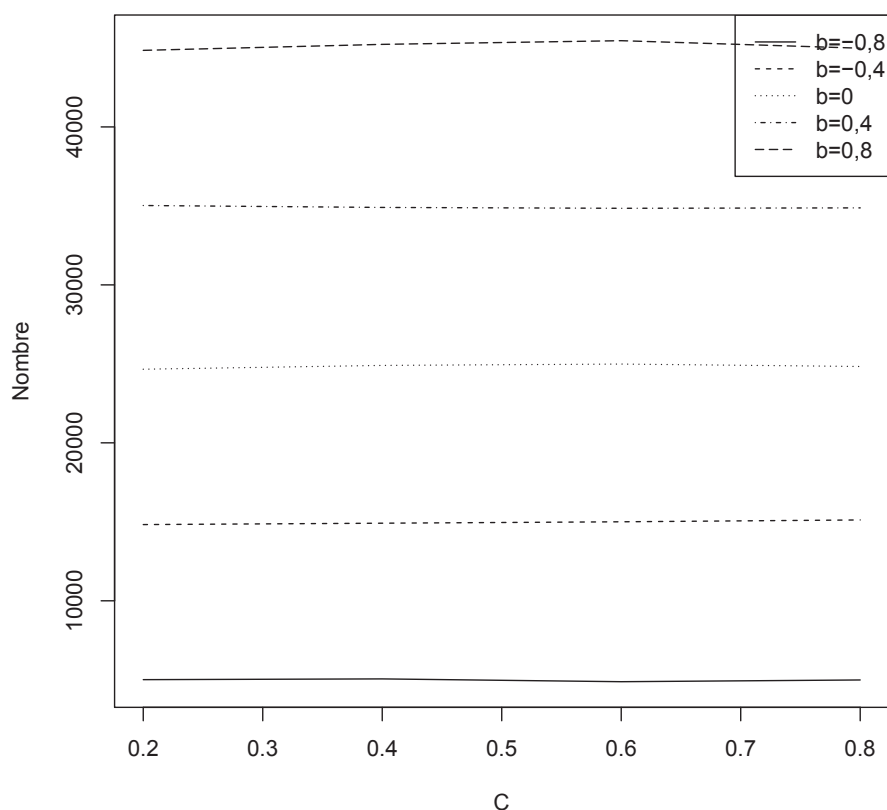


FIGURE 3.7. Nombre de fixations du mutant sur 1000000 pour $r=0$, avec b et c variables.

Puisqu'un plus grand b avantage le mutant, il n'est pas surprenant de voir que, plus b augmente, plus la proportion d'occurrences de la fixation augmente, peu importe le c . En fait, comme confirmé dans le cas précédent, peu importe le r , l'augmentation du b entraîne une augmentation du nombre de fixations. Nous remarquons cependant que le c n'a ici aucun effet sur le nombre de fixations, peu importe le b . En effet, puisqu'il n'y a aucune recombinaison possible, le nombre d'événements de conversion possibles est très limité, et ainsi l'importance de son taux apporte un effet négligeable à la probabilité de fixation.

Considérons maintenant le temps de fixation conditionnel (figure 3.8).

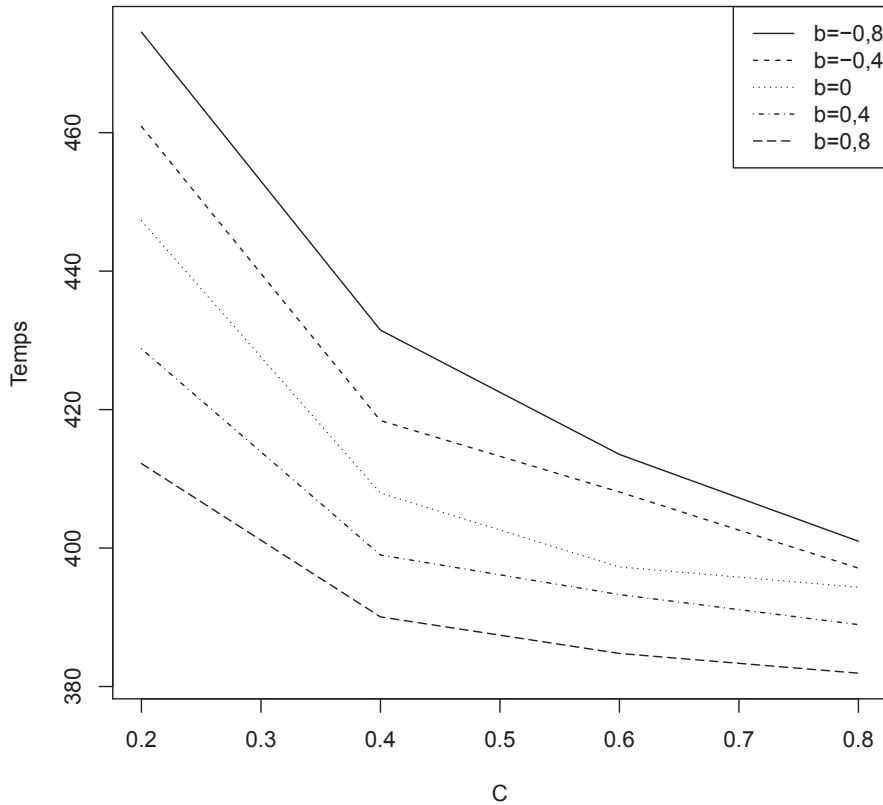


FIGURE 3.8. Temps de fixation conditionnel moyen du mutant sur 1000000 pour $r=0$, avec b et c variables.

Nous pouvons remarquer que, comme dans le cas précédent, peu importe le b , plus le c augmente, plus le temps de fixation conditionnel diminue. En fait, c'est le cas peu importe le r . Nous remarquons cependant que peu importe le c , plus b augmente, plus le temps de fixation conditionnel diminue, ce qui est différent du cas trois. En effet, puisque le nombre d'événements de conversion est restreint, la fixation n'a pas à être plus rapide pour qu'elle ait lieu. Ainsi, le temps de fixation conditionnel diminue lorsque l'avantage du gène mutant lors de la conversion augmente et augmente lorsque l'avantage du gène mutant diminue.

Chapitre 4

DISCUSSION

Tout d'abord, l'article de Tajima (1990) a été analysé, et les preuves qui y étaient partiellement ou complètement manquantes ont été faites. Cela nous a permis de constater que plusieurs phénomènes biologiques, tels que le nombre de sites polymorphes ou l'homozygotie entre deux séquences choisies au hasard dans la population, sont grandement affectées lorsqu'il y a fixation d'un nouveau mutant, et plus particulièrement lorsque le temps de cette fixation est connu. Par exemple, il y a en moyenne $\frac{1}{3}$ de sites polymorphes de moins entre deux séquences choisies au hasard au moment de la fixation connu par rapport à un moment quelconque dans le temps.

Par la suite, un effet de sélection a été ajouté au modèle. Une nouvelle méthode a été présentée afin d'approcher le résultat de Mano (2009) sur l'espérance du temps de fixation d'un nouveau mutant conditionnellement à l'arrivée de la fixation en un temps fini. Cette approche utilise le processus ancestral. Elle mène à un résultat cohérent avec celui de Mano (2009), en plus de donner un résultat non asymptotique, soit que l'effet du premier ordre de la sélection est de $-\frac{1}{(n+1)(n+2)}$, qui approche 0 de façon asymptotique. Cependant, seulement le coefficient faisant intervenir un seul événement de sélection a été trouvé théoriquement. En effet, puisque la méthode est de considérer tous les cas de graphe possibles menant à la fixation d'un nouveau mutant, le nombre de cas serait très élevé si on considérait la possibilité que deux événements de sélection ou plus surviennent, et donc les calculs pour y arriver seraient très laborieux. Des simulations ont cependant été faites afin de vérifier la théorie et de déterminer s'ils tiennent toujours en présence de forte sélection ($s = 0, 2$). Or, les résultats trouvés sont cohérents avec la théorie.

Enfin, l'effet de conversion a été considéré, puisque, lorsqu'il est biaisé, il se rapproche de l'effet de sélection. En utilisant le processus ancestral, un résultat théorique pour la probabilité de fixation d'un nouveau mutant dans la population a été déduit. Cependant, aucune formule théorique n'a été tentée pour l'espérance du temps de fixation en raison de la complexité des calculs. Des simulations ont cependant été faites afin de déterminer les tendances des variations pour l'espérance conditionnelle du temps de fixation et la probabilité de fixation par rapport aux taux de recombinaison, de conversion et au biais de conversion, lorsque ceux-ci sont de plus grande échelle, et plusieurs tendances intéressantes ont été constatées.

Bibliographie

- Darden, T., Hudson, R. et Kaplan, N. (1988). The Coalescent Process in Models with Selection. *Genetics*, **120**, 819–829.
- Innan, H. (2002). A Method for Estimating the Mutation, Gene Conversion and Recombination Parameters in Small Multigene Families. *Genetics*, **161**, 865–872.
- Innan, H. (2003a). The Coalescent and Infinite-Site Model of a Small Multigene Family. *Genetics*, **163**, 803–810.
- Innan, H. (2003b). A two-locus gene conversion model with selection and its application to the human rhce and rhd genes. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 8793–8798.
- Innan, H. et Mano, S. (2008). The Evolutionary Rate of Duplicated Genes Under Concerted Evolution. *Genetics*, **180**, 493–505.
- Kimura, M. et Crow, J. (1964). The Number of Alleles That Can Be Maintained in a Finite Population. *Genetics*, **49**, 725–738.
- Mano, S. (2009). Duality, ancestral and diffusion processes in models with selection. *Theor. Popul. Biol.*, **75**, 164–175.
- Mano, S. (2013). Ancestral graph with bias in gene conversion. *J. Appl. Probab.*, **50**, 239–255.
- Nagylaki, T. (1984). The Evolution of Multigene Families under Intrachromosomal Gene Conversion. *Genetics*, **106**, 529–548.
- Nagylaki, T. et Petes, T. (1982). Intrachromosomal Gene Conversion and the Maintenance of Sequence Homogeneity among Repeated Genes. *Genetics*, **100**, 315–337.
- Piskol, R. et Stephan, W. (2008). Analyzing the Evolution of RNA Secondary Structures in Vertebrate Introns Using Kimura's Model of Compensatory Fitness Interactions. *Mol. Biol. Evol.*, **25**, 2483–2492.
- Stephan, W. (1996). The Rate of Compensatory Evolution. *Genetics*, **144**, 419–426.

Tajima, F. (1990). Relationship Between DNA Polymorphism and Fixation Time. *Genetics*, **125**, 447–454.

Watterson, G. (1975). On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.*, **7**, 256–276.

Annexe A

RÉSULTATS COMPLETS DES SIMULATIONS SUR LES EFFETS DE LA RECOMBINAISON, DE LA CONVERSION ET DE SON BIAIS SUR LE TEMPS DE FIXATION CONDITIONNEL ET LA PROBABILITÉ DE FIXATION

A-ii

TABLEAU A.1. Résultats complets du nombre de fixations du mutant pour chaque valeur des taux de recombinaison, de conversion et de son biais.

b=-0,8

r	c			
	0,2	0,4	0,6	0,8
0	5008	5056	4878	4988
0,1	2643	2363	2304	2116
0,2	1363	1051	846	776
0,3	900	431	297	224
0,4	561	198	127	89
0,5	397	98	46	17

b=-0,4

r	c			
	0,2	0,4	0,6	0,8
0	14825	14913	15002	15118
0,1	11122	10514	10148	10137
0,2	8515	7341	7065	6658
0,3	7002	5358	4473	4192
0,4	5945	3824	3093	2543
0,5	5177	2793	1972	1522

b=0

r	c			
	0,2	0,4	0,6	0,8
0	24660	24901	24988	24834
0,1	25105	24923	24815	24856
0,2	25044	24773	24758	24909
0,3	25063	25218	25125	25136
0,4	25198	25156	25154	25117
0,5	25034	25071	24821	24805

b=0,4

r	c			
	0,2	0,4	0,6	0,8
0	35025	34903	34849	34875
0,1	43350	45993	47421	48015
0,2	50841	56352	60175	62426
0,3	56202	65592	72245	76717
0,4	60307	74544	84308	90197
0,5	63977	82113	95429	104853

b=0,8

r	c			
	0,2	0,4	0,6	0,8
0	44846	45226	45459	44984
0,1	64981	72341	76701	79588
0,2	80236	97261	107933	115315
0,3	91646	117417	134859	147570
0,4	99968	133785	157639	175777
0,5	107451	147239	176545	198249

TABLEAU A.2. Résultats complets du temps de fixation conditionnel du mutant pour chaque valeur des taux de recombinaison, de conversion et de son biais.

b=-0,8				
r	c			
	0,2	0,4	0,6	0,8
0	474,5131	431,4793	413,5236	400,9883
0,1	483,7673	430,7094	406,9507	389,7848
0,2	489,4068	398,1646	362,3733	354,0463
0,3	477,1462	368,8310	347,0281	316,5350
0,4	463,2020	362,5816	295,3448	289,9258
0,5	477,9322	375,4346	242,1333	164,8583

b=-0,4				
r	c			
	0,2	0,4	0,6	0,8
0	460,8976	418,4136	408,0973	397,0997
0,1	507,4390	438,1353	411,1510	397,6930
0,2	534,8496	446,7300	413,5696	391,4030
0,3	560,5206	450,0433	407,6257	379,5299
0,4	578,2331	452,0984	396,5928	369,7134
0,5	589,8425	454,8116	390,9908	343,4919

b=0				
r	c			
	0,2	0,4	0,6	0,8
0	447,2720	407,9447	397,2631	394,3430
0,1	517,2805	439,1284	409,2775	400,9023
0,2	560,4777	463,8091	425,4622	402,4192
0,3	596,8921	484,5358	437,0722	407,9283
0,4	624,2839	501,3431	448,7686	415,2451
0,5	641,0416	520,3843	454,5760	418,2155

b=0,4				
r	c			
	0,2	0,4	0,6	0,8
0	438,0307	401,3814	394,3701	392,5200
0,1	498,3058	428,1185	403,9654	387,1984
0,2	531,2526	444,1319	407,8016	387,6673
0,3	565,3685	453,8643	399,2247	376,2849
0,4	583,9112	457,4553	392,7110	362,8552
0,5	596,2152	443,6227	388,5410	348,5205

b=0,8				
r	c			
	0,2	0,4	0,6	0,8
0	413,8496	387,8218	384,7639	384,1294
0,1	470,7041	404,1955	390,4339	370,8780
0,2	479,0963	392,2421	365,0859	350,1947
0,3	483,8684	379,3599	339,2465	311,8316
0,4	489,3627	368,5534	313,4309	284,5711
0,5	485,4204	354,4494	292,8887	256,6802