



**HAL**  
open science

# Relation entre la complémentation et le taux de mutation des virus à ARN

Raphaël Villeneuve

► **To cite this version:**

Raphaël Villeneuve. Relation entre la complémentation et le taux de mutation des virus à ARN. Médecine vétérinaire et santé animale. 2024. dumas-04625293

**HAL Id: dumas-04625293**

**<https://dumas.ccsd.cnrs.fr/dumas-04625293>**

Submitted on 26 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# RELATION ENTRE LA COMPLEMENTATION ET LE TAUX DE MUTATION DES VIRUS A ARN

---

THESE D'EXERCICE

pour obtenir le titre de  
DOCTEUR VETERINAIRE

DIPLOME D'ETAT

*présentée et soutenue publiquement  
devant l'Université Paul-Sabatier de Toulouse*

*par*

**VILLENEUVE Raphaël, Achille,  
François**

**Directeur de thèse : M. Romain VOLMER**

---

## JURY

PRESIDENT :

**M. Stéphane BERTAGNOLI**

Professeur à l'Ecole Nationale Vétérinaire de TOULOUSE

ASSESEURS :

**M. Romain VOLMER**

Professeur à l'Ecole Nationale Vétérinaire de TOULOUSE

**Mme Christelle CAMUS**

Professeure à l'Ecole Nationale Vétérinaire de TOULOUSE



**Ministère de l'Agriculture et de l'Alimentation  
ÉCOLE NATIONALE VÉTÉRINAIRE DE TOULOUSE**

**Liste des directeurs/assesseurs de thèse de doctorat vétérinaire**

**Directeur : Professeur Pierre SANS**

**PROFESSEURS CLASSE EXCEPTIONNELLE**

- M. **BAILLY Jean-Denis**, *Hygiène et industrie des aliments*
- M. **BERTAGNOLI Stéphane**, *Pathologie infectieuse*
- M. **BOUSQUET-MELOU Alain**, *Pharmacologie, thérapeutique*
- M. **BRUGERE Hubert**, *Hygiène et industrie des aliments d'origine animale*
- M. **CONCORDET Didier**, *Mathématiques, statistiques, modélisation*
- M. **DELVERDIER Maxence**, *Anatomie pathologique*
- M. **FOUCRAS Gilles**, *Pathologie des ruminants*
- Mme **GAYRARD-TROY Véronique**, *Physiologie de la reproduction, endocrinologie*
- M. **GUERIN Jean-Luc**, *Aviculture et pathologie aviaire*
- Mme **HAGEN-PICARD Nicole**, *Pathologie de la reproduction*
- M. **MEYER Gilles**, *Pathologie des ruminants*
- M. **SCHELCHER François**, *Pathologie médicale du bétail et des animaux de basse-cour*
- Mme **TRUMEL Catherine**, *Biologie médicale animale et comparée*

**PROFESSEURS 1<sup>ère</sup> CLASSE**

- Mme **BOULLIER Séverine**, *Immunologie générale et médicale*
- Mme **BOURGES-ABELLA Nathalie**, *Histologie, anatomie pathologique*
- Mme **CADIERGUES Marie-Christine**, *Dermatologie vétérinaire*
- M. **DUCOS Alain**, *Zootecnie*
- M. **GUERRE Philippe**, *Pharmacie et toxicologie*
- M. **JACQUIET Philippe**, *Parasitologie et maladies parasitaires*
- Mme **LACROUX Caroline**, *Anatomie pathologique, animaux d'élevage*
- Mme **LETRON-RAYMOND Isabelle**, *Anatomie pathologique*
- M. **LEFEBVRE Hervé**, *Physiologie et thérapeutique*
- M. **MAILLARD Renaud**, *Pathologie des ruminants*
- Mme **MEYNADIER Annabelle**, *Alimentation animale*

**PROFESSEURS 2<sup>ème</sup> CLASSE**

- Mme **CAMUS Christelle**, *Biologie cellulaire et moléculaire*
- M. **CORBIERE Fabien**, *Pathologie des ruminants*
- Mme **DIQUELOU Armelle**, *Pathologie médicale des équidés et des carnivores*
- M. **MATHON Didier**, *Pathologie chirurgicale*
- M. **MOGICATO Giovanni**, *Anatomie, imagerie médicale*
- M. **NOUVEL Laurent**, *Pathologie de la reproduction*
- Mme **PAUL Mathilde**, *Epidémiologie, gestion de la santé des élevages avicoles*
- M. **VOLMER Romain**, *Microbiologie et infectiologie*

## MAITRES DE CONFERENCES HORS CLASSE

- M. **BERGONIER Dominique**, *Pathologie de la reproduction*  
Mme **BIBBAL Delphine**, *Hygiène et industrie des denrées alimentaires d'origine animale*  
M. **JAEG Jean-Philippe**, *Pharmacie et toxicologie*  
M. **LYAZRHI Faouzi**, *Statistiques biologiques et mathématiques*  
Mme **PALIERNE Sophie**, *Chirurgie des animaux de compagnie*  
Mme **PRIYMENKO Nathalie**, *Alimentation*

## MAITRES DE CONFERENCES CLASSE NORMALE

- M. **ASIMUS Erik**, *Pathologie chirurgicale*  
Mme **BRET Lydie**, *Physique et chimie biologiques et médicales*  
Mme **BOUHSIRA Emilie**, *Parasitologie, maladies parasitaires*  
M. **CARTIAUX Benjamin**, *Anatomie, imagerie médicale*  
M. **COMBARROS Daniel**, *Dermatologie vétérinaire*  
M. **CONCHOU Fabrice**, *Imagerie médicale*  
Mme **DANIELS Hélène**, *Immunologie, bactériologie, pathologie infectieuse*  
Mme **DAVID Laure**, *Hygiène et industrie des aliments*  
M. **DIDIMO IMAZAKI Pedro**, *Hygiène et industrie des aliments*  
M. **DOUET Jean-Yves**, *Ophthalmologie vétérinaire et comparée*  
M. **FERCHIOU Ahmed**, *Economie de santé des bovins*  
Mme **FERRAN Aude**, *Physiologie*  
M. **FUSADE-BOYER Maxime**, *Microbiologie et infectiologie*  
M. **GAIDE Nicolas**, *Anatomie pathologique*  
Mme **GRANAT Fanny**, *Biologie médicale animale*  
Mme **JOURDAN Géraldine**, *Anesthésie, analgésie*  
M. **JOUSSERAND Nicolas**, *Médecine interne des animaux de compagnie*  
Mme **LALLEMAND Elodie**, *Chirurgie des équidés*  
Mme **LAVOUE Rachel**, *Médecine Interne*  
M. **LE LOC'H Guillaume**, *Médecine zoologique et santé de la faune sauvage*  
M. **LIENARD Emmanuel**, *Parasitologie et maladies parasitaires*  
Mme **MEYNAUD-COLLARD Patricia**, *Pathologie chirurgicale*  
Mme **MILA Hanna**, *Elevage des carnivores domestiques*  
M. **VERGNE Timothée**, *Santé publique vétérinaire, maladies animales réglementées*  
Mme **WASET-SZKUTA Agnès**, *Production et pathologie porcine*

## INGENIEURS DE RECHERCHE

- M. **AUMANN Marcel**, *Urgences, soins intensifs*  
M. **AUVRAY Frédéric**, *Santé digestive, pathogénie et commensalisme des entérobactéries*  
M. **CASSARD Hervé**, *Pathologie des ruminants*  
M. **CROVILLE Guillaume**, *Virologie et génomique cliniques*  
Mme **DIDIER Caroline**, *Anesthésie, analgésie*  
M. **DELPONT Mattias**, *Clinique Aviaire*  
Mme **DUPOUY GUIRAUTE Véronique**, *Innovations thérapeutiques et résistances*  
Mme **GAILLARD Elodie**, *Urgences, soins intensifs*  
Mme **GEFFRE Anne**, *Biologie médicale animale et comparée*  
Mme **GRISEZ Christelle**, *Parasitologie et maladies parasitaires*  
Mme **JEUNESSE Elisabeth**, *Bonnes pratiques de laboratoire*  
Mme **LAYSSOL-LAMOUR Catherine**, *Imagerie Médicale*  
Mme **POUJADE Agnès**, *Anatomie pathologique Vétérinaire*  
Mme **PRESSANTI Charline**, *Dermatologie vétérinaire*  
M. **RAMON PORTUGAL Felipe**, *Innovations thérapeutiques et résistances*  
M. **REYNOLDS Brice**, *Médecine interne des animaux de compagnie*  
Mme **ROUCH BUCK Pétra**, *Médecine préventive*  
Mme **SAADA Chloé**, *Gestion intégrée de la santé des ruminants*

## **Remerciements**

Que soit ici exprimée toute ma reconnaissance à

**Monsieur le Professeur Stéphane BERTAGNOLI,**

Professeur classe exceptionnelle de l'École Nationale Vétérinaire de Toulouse

*Pathologie infectieuse*

Qui m'a fait l'honneur d'accepter la présidence du jury de cette thèse

Hommages respectueux.

**Monsieur le Professeur Romain VOLMER,**

Professeur 2<sup>ème</sup> classe de l'École Nationale Vétérinaire de Toulouse

*Microbiologie - Infectiologie*

Qui m'a proposé le sujet de cette thèse, qui m'a guidé tout au long de ce travail avec bienveillance et m'a transmis un grand intérêt pour sa discipline.

Sincères remerciements.

Enfin, j'adresse mes remerciements et toute ma sympathie à

**Madame la Professeure Christelle CAMUS,**

Professeure 2<sup>ème</sup> classe de l'École Nationale Vétérinaire de Toulouse

*Biologie cellulaire et moléculaire*

Qui a aimablement accepté de faire partie du jury de cette thèse en tant qu'assesseur

Remerciements chaleureux.



## Table des matières

Table des matières .....	1
Table des figures .....	4
Table des tableaux .....	6
Introduction .....	9
1) Détermination du taux de mutation : n'évaluer que les mutations survenues au cours d'un unique événement de copie.....	10
2) Détermination du taux de mutation : prendre en compte le biais de la sélection naturelle .....	12
Objectif du travail de thèse : .....	15
Partie I : Les méthodes de détermination du taux de mutation .....	16
1) Les trois méthodes disponibles pour la détermination expérimentale d'un taux de mutation viral, avec focus sur les techniques de séquençage .....	16
2) Première méthode de détermination du taux de mutation : le séquençage génomique, avec correction des erreurs artéfactuelles par rétro-transcription circulaire.....	17
3) Deuxième méthode de détermination du taux de mutation : la méthode d'accumulation des mutations .....	19
4) Troisième méthode de détermination du taux de mutation : le test de fluctuation de Luria-Delbrück .....	20
a) Principe du test .....	20
b) Limites du test .....	22
5) Récapitulatifs des différentes méthodes de détermination du taux de mutation .....	23
Partie II : Étude d'une possible influence de la segmentation génomique sur la tolérance à des taux de mutation élevés .....	24
1) De grands écarts de taux de mutation entre les différentes espèces de virus à ARN.....	24
2) Analyse d'un premier cas de divergence du taux de mutation $\mu$ : le virus de la gravure du tabac et le virus de la mosaïque de la luzerne.....	26
3) Analyse d'un deuxième cas de divergence du taux de mutation $\mu$ : le virus de la rougeole et le virus influenza A.....	27
a) Présentation du cas.....	27
b) Analyse d'un premier biais de comparaison potentiel : la taille du génome .....	28
c) Analyse d'un deuxième biais de comparaison potentiel : la région du génome.....	29
Partie III : Bénéfices possibles de la complémentation.....	30
1) Définition .....	30
2) État des lieux de l'effet d'une mutation sur la fitness chez les virus à ARN.....	32
3) Bénéfices et inconvénients de la complémentation pour une population virale .....	34



a) Réflexions théoriques sur les bénéfices et inconvénients de la complémentation .....	34
b) Interaction entre la complémentation et la sélection naturelle : une fitness populationnelle diminuée à pression de sélection faible, mais augmentée à pression de sélection forte .....	35
c) Une explication possible à l'augmentation de la robustesse vis-à-vis de la sélection par la complémentation : favoriser une grande diversité génétique dans la population .....	37
4) Conditionnalité des bénéfices/inconvénients de la complémentation à un paramètre : le taux de mutation ( $\mu$ ) .....	39
5) Évaluation d'un biais dans l'analyse des effets de la complémentation .....	42
6) Comment font les virus à ARN qui ne peuvent pas bénéficier de la complémentation génomique ?.....	44
Conclusion.....	46
Bibliographie :.....	49



## Table des figures

<b>Figure 1</b>	Taux de mutation de certains organismes en fonction de la taille de leur génome (Elena, Sanjuán 2005)	9
<b>Figure 2</b>	Méthode de comparaison de séquences pour remonter aux mutations vraies (Yasukawa et al. 2017)	11
<b>Figure 3</b>	Principe de l'expérience de Tromas et Elena de 2010 visant à imposer la neutralité sélective	13
<b>Figure 4</b>	Protocole pour s'affranchir de la sélection naturelle dans la détermination d'un taux de mutation viral à l'aide d'une séquence exogène (Pauly, Procario, Lauring 2017)	14
<b>Figure 5</b>	Utilisation du barcoding et du séquençage circulaire pour discriminer mutations vraies et dues au séquençage/PCR/rétro-transcription (Lou et al. 2013)	18
<b>Figure 6</b>	Principe de la complémentation virale	31
<b>Figure 7</b>	Interaction entre la complémentation et la sélection, avec $\mu = 0,01$ , $\varepsilon = 0,0$ , $\eta = 0,0$ et S variant comme indiqué sur la légende (selon les données de (Gao, Feldman 2009))	36
<b>Figure 8</b>	Interaction entre la complémentation et le taux de mutation, avec $S = 0,2$ , $\varepsilon = 0,0$ , $\eta = 0,0$ et $\mu$ variant comme indiqué sur la légende (selon les données de (Segredo-Otero, Sanjuán 2019))	40
<b>Figure 9</b>	Relation entre la prévalence des signaux d'empaquetage (Ps) et le taux de mutation ( $\mu$ ) chez les virus à ARN (selon les données de Hutchinson et al., 2010)	43



## Table des tableaux

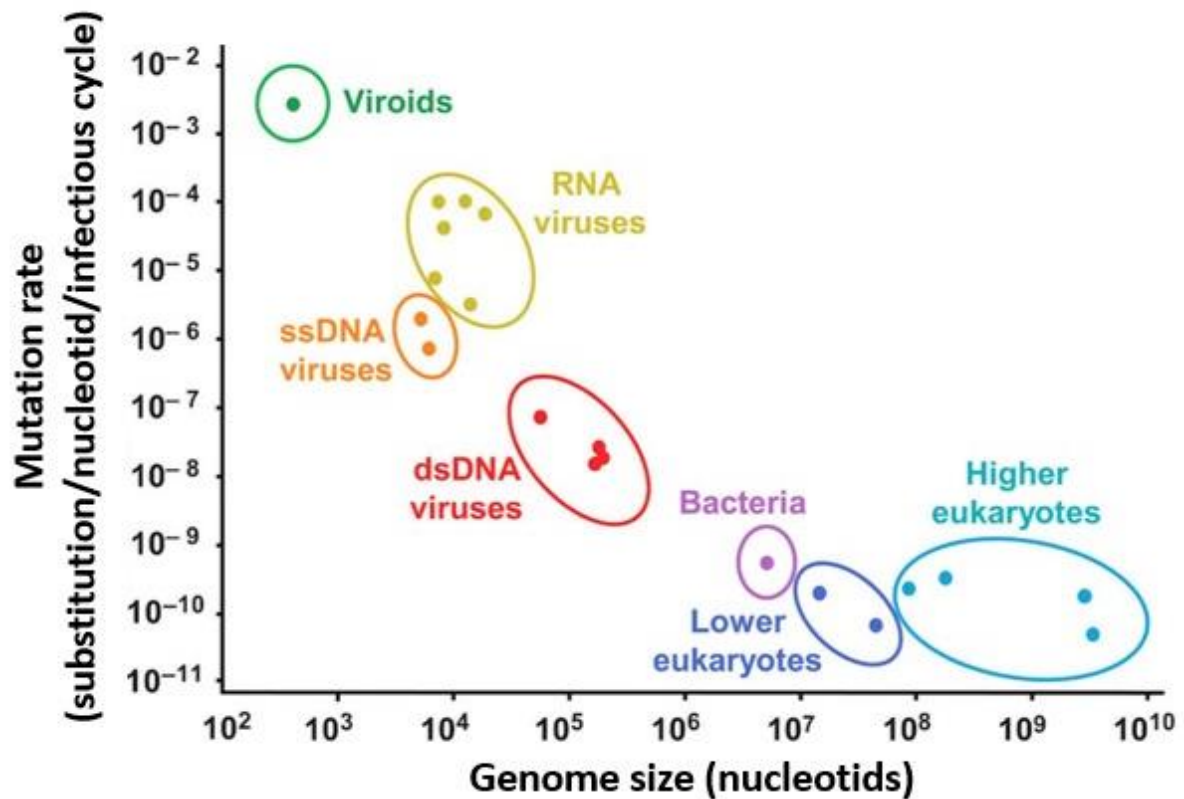
<b>Tableau 1</b>	Avantages/inconvénients des différentes méthodes de détermination du taux de mutation	23
<b>Tableau 2</b>	Quelques exemples de taux de mutation de différents virus à ARN	24
<b>Tableau 3</b>	Résultats sur la distribution de la fitness conféré par les mutations chez 3 virus à ARN (d'après Sanjuán et al. 2004a, Carrasco et al. 2007 et Deyde et al. 2006)	33
<b>Tableau 4</b>	Parties des équations permettant de prendre en compte l'effet de l'épistase et des mutations (Gao et Feldman, 2009)	35





## Introduction

Les virus à ARN sont caractérisés par un taux de mutation très important comparé aux virus à ADN, et *a fortiori* aux autres organismes vivants. Cela implique une grande variabilité génétique, qui facilite notamment l'échappement à la réponse immunitaire et les rend redoutables d'un point de vue épidémiologique. En conséquence, l'étude des dynamiques mutationnelles de ces virus, particulièrement la mise en évidence des facteurs favorisant les mutations, apparaît indispensable pour lutter efficacement contre les maladies dont ils peuvent être responsables.



*Figure 1 : Taux de mutation de certains organismes en fonction de la taille de leur génome (Elena, Sanjuán 2005)*

Sept groupes d'organismes vivants sont représentés sur ce graphique : les taux de mutation représentés sur ce graphique sont parmi les plus divers trouvés dans la littérature scientifique pour un groupe donné. En ordre de grandeur, le taux de mutation d'un organisme a tendance à diminuer lorsque la taille et la complexité de l'organisme augmente, et/ou que l'organisme utilise l'ADN par rapport à l'ARN comme support génétique. Également, le groupe des virus à ARN (notre groupe d'intérêt) présente un taux de mutation compris entre  $10^{-6}$  et  $10^{-4}$  s/n/cycle infectieux (parmi les plus élevés de tous les organismes vivants ici représentés).

Chez les virus à ARN, le taux de substitution par nucléotides par cycle infectieux est globalement compris entre  $10^{-6}$  et  $10^{-4}$  s/n/cycle infectieux (Figure 1). Or les mutations par substitution sont de loin les plus fréquentes chez les virus à ARN, elles sont quatre fois plus fréquentes que les mutations indel notamment (définies par l'insertion ou la délétion d'un ou de plusieurs nucléotides dans une séquence d'ARN par rapport à une séquence de référence,



celle du type sauvage en général) (Sanjuán et al. 2010). On peut donc approximer avec une bonne précision le taux de mutation chez ces virus par leur taux de substitution.

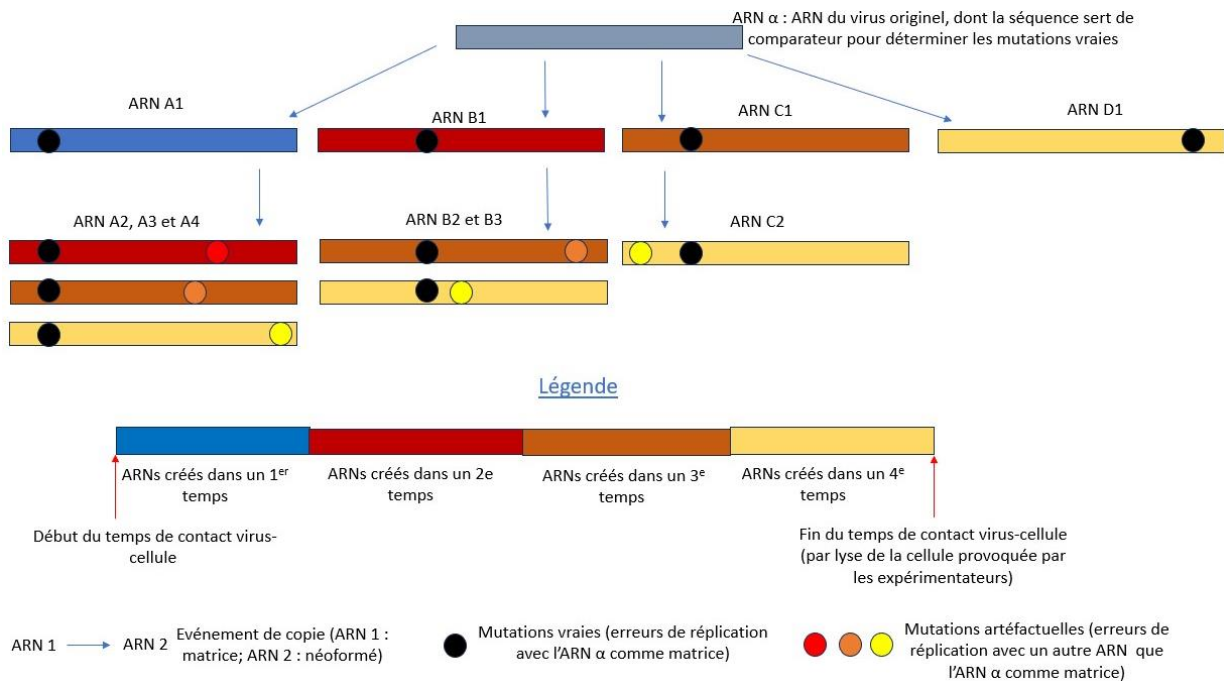
Une fois ceci fait, deux problèmes majeurs persistent dans l'évaluation actuelle des taux de mutation : décompter uniquement les mutations survenues au cours d'un seul événement de copie (1)) et prendre en compte le biais de la sélection naturelle (2)).

### 1) Détermination du taux de mutation : n'évaluer que les mutations survenues au cours d'un unique événement de copie

Les virus peuvent se répliquer suivant deux modes : mode linéaire (« stamping-machine mode ») ou binaire (Sanjuán et al. 2010). Suivant le mode linéaire, un seul brin matrice (issu du virion qui a infecté la cellule) est utilisé pour créer de multiples brins d'ARN, et les brins d'ARN néoformés ne deviennent des matrices que lorsque les virions issus de cet événement de copie (=réplication) infectent une autre cellule. *A contrario*, suivant le mode binaire, les brins néoformés deviennent immédiatement des brins matrices pour l'ARN-polymérase virale. Dans la réalité, les virus ont un mode de répllication intermédiaire entre ces deux modèles théoriques, plus ou moins proches du mode linéaire (bactériophages  $\phi 6$  et  $\phi X174$ ) ou binaire (poliovirus-1) suivant leur espèce (Sanjuán et al. 2010).

Or, pour calculer un taux de mutation qui permette véritablement une comparaison entre virus, on doit s'exprimer en nombre de mutations par unique événement de copie. Mais en comparant directement le génome du virus originel qui a infecté la cellule et ceux des particules virales qui sortent de cette dernière après sa lyse ou qui s'accumulent dans la cellule, on observe (sauf si le mode de répllication du virus est strictement linéaire) des mutations accumulées au cours de plusieurs événements de copie, ce qui ne permet pas de calculer strictement un taux de mutation. Le taux issu de cette comparaison directe est mesuré en mutations/cycle infectieux, et il peut être de une fois (mode de répllication très proche du linéaire : bactériophage  $\phi 6$ , selon Sanjuán et al. 2010) à dix fois (mode de répllication très proche du binaire : poliovirus-1, selon Sanjuán et al. 2010) plus élevé que le taux en nombre de mutations par événement de copie.

- Pour remonter aux taux en mutations/ événement de copie à partir des séquences d'ARN obtenues à l'issue de l'étape de multiplication *in vivo*, on peut utiliser une méthode de comparaison de séquences (Yasukawa et al. 2017).



*Figure 2 : Méthode de comparaison de séquences pour remonter aux mutations vraies (Yasukawa et al. 2017)*

L'ARN  $\alpha$  est copié dans un premier temps pour former l'ARN A1 (en bleu), puis l'ARN  $\alpha$  et l'ARN A1 sont copiés pour former l'ARN B1 et A2 respectivement (en rouge), et ainsi de suite avec les événements de copie générant les ARN indiqués en orange puis en jaune dans la légende de la figure. Or les mutations pertinentes pour calculer le taux de mutation de ce virus (à partir de la séquence de l'ARN  $\alpha$ , dont on dispose), sont uniquement celles apparues pour les événements de copie ayant utilisé l'ARN  $\alpha$  comme matrice. Pour exclure les autres mutations (mutations artéfactuelles), on utilise une étape de traitement informatique qui vise à reconstruire les « familles » (A, B, C et D) d'ARN à partir des proportions d'ARN présentant telle ou telle mutation (ce sont ces proportions qui détermineront si une mutation est considérée comme vraie ou pas)

La méthode Yasukawa (représentée ici pour un virus au mode de réplication parfaitement binaire) se base sur la proportion d'ARN avec telle ou telle mutation pour déterminer s'il s'agit ou non d'une mutation obtenue à partir de l'ARN du virus ayant infecté la cellule (A1, B1, C1, D1 sur le schéma), ou d'un autre ARN séparé de l'ARN originel par plus d'un événement de copie (tous les autres ARN du schéma). Selon l'étude de Sanjuán et al. de 2010, les virus à ARN simple brin ont un mode de réplication très proche du binaire dans la majorité des cas, on peut donc considérer les proportions de mutations vraies et de mutations artéfactuelles, présentées dans un cas simple sur ce schéma, comme valide. En utilisant un traitement informatique et en garantissant expérimentalement un nombre global d'évènements de copie faible (via une période de temps brève entre début de mise en contact de l'échantillon viral originel et lyse de la cellule par les expérimentateurs), on peut réussir à établir les liens de « filiation » des différents ARN recueillis, en se basant sur ces proportions. On remonte donc au nombre de mutations vraies générées au cours du nombre d'évènements de copie de l'ARN  $\alpha$  ayant eu lieu, et donc au taux de mutation du virus à ARN utilisée.

- On peut également calculer un taux en nombre de mutation par événement de copie en utilisant un test de fluctuation de Luria-Delbrück. Il s'agit de la méthode dont nous discuterons le plus dans la suite de cette thèse (cf notamment partie I, 4)). Cette méthode, employée par Sanjuán et al. dans leur étude de 2010, ne comporte pas d'étape de séquençage. Elle permet donc de s'affranchir des erreurs qu'impliquent celui-ci (détails en partie I, 2)), et ne nécessite pas d'étape de traitement informatique complexe comme celle de l'étude de 2017 de Yasukawa et al.

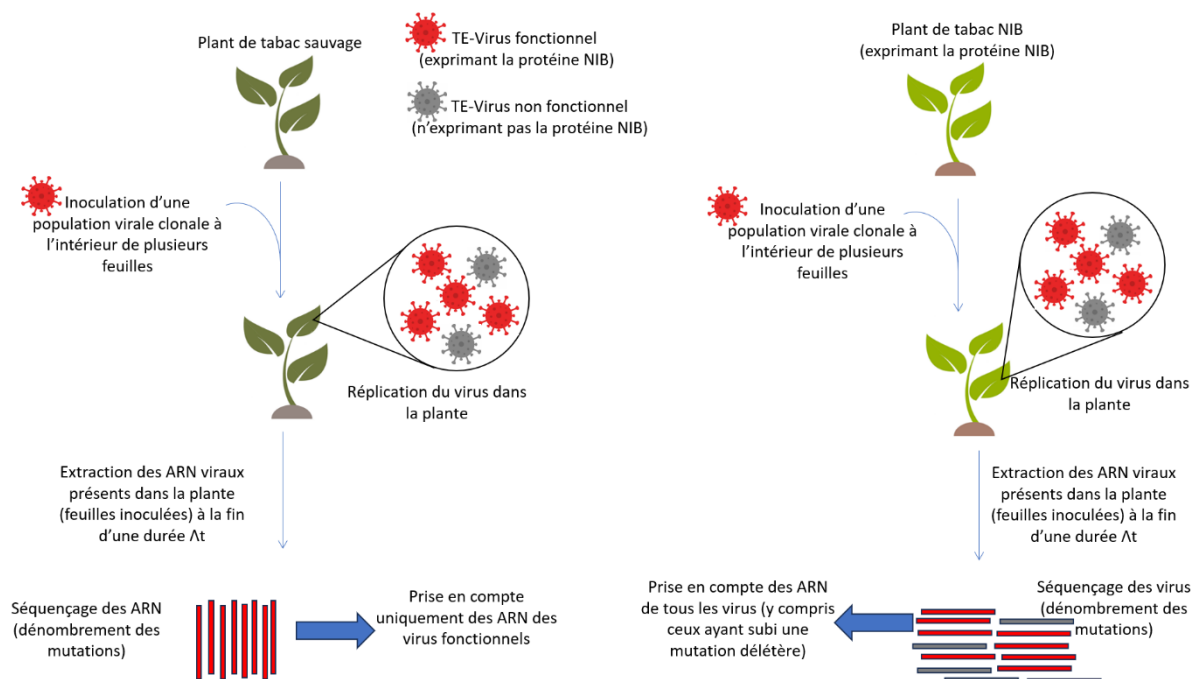
Pour récapituler sur ces moyens d'évaluer seulement les mutations survenues au cours d'un unique événement de copie, la méthode Yasukawa présente l'avantage de ne pas imposer de protocole expérimental très précis, elle permet de s'adapter aux besoins des expérimentateurs en utilisant une étape de traitement informatique *a posteriori*. Mais comme dans la population virale initiale, il est quasiment impossible d'obtenir 2 virions avec des génomes identiques (c'est-à-dire pour le traduire sur la figure 2, avec des molécules d'ARN  $\alpha$  identiques), il existe une imprécision inhérente à la méthode. *A contrario*, le test de fluctuation de Luria-Delbrück impose de partir d'une population virale clonale avec un phénotype précis, elle constitue un protocole expérimental à elle toute seule. Cependant elle ne nécessite pas de connaître la séquence exacte de chacun des virions de la population initiale, seules les séquences des *loci* codant pour le phénotype d'intérêt ont besoin d'être connues avec précision, ce qui présente un grand avantage vu qu'il n'existe quasiment pas 2 génomes viraux identiques dans une population virale, aussi clonale qu'on essaye de la générer. C'est la principale raison pour laquelle il s'agit de la méthode de détermination du taux de mutation la plus utilisée dans la littérature scientifique en général (Sanjuán et al. 2010).

## 2) Détermination du taux de mutation : prendre en compte le biais de la sélection naturelle

Comme on désire évaluer le taux de mutation du virus à un moment très ponctuel (un seul événement de copie), en excluant donc son taux d'évolution sur une période temporelle plus importante (qui s'exprimerait par exemple en substitution/nucléotide/an), un deuxième problème se pose : s'affranchir du biais de la sélection naturelle, qui aboutit à l'élimination de certains mutants.

En effet, toutes les expériences visant à mesurer des taux de mutation viraux *in vivo* passent par une étape de multiplication : une petite population virale clonale est inoculée à un organisme vivant dans lequel ces virions se multiplient et acquièrent progressivement des mutations par rapport à la séquence génomique de la population initiale. Comme les virus se répliquent très vite, il est souvent difficile de n'avoir que les virions issus d'une seule génération virale après cette étape de multiplication, il y a souvent deux ou trois (voire plus) générations. Dans des conditions expérimentales normales, les virus qui présentent des mutations délétères à court terme se répliquent donc moins que les autres (sélection naturelle), il y aura donc moins de virions à l'issue de la multiplication *in vivo*, et le taux de mutation calculé se retrouvera donc sous-estimé. D'où l'importance de mettre en place des conditions expérimentales pour lesquelles les virus ayant subi des mutations délétères peuvent se répliquer au même titre que ceux ayant conservé la séquence sauvage et ceux ayant acquis une mutation avantageuse : neutralité sélective. Deux exemples de moyens d'y parvenir sont présentés ci-dessous.

- Un exemple de telles conditions est détaillé dans un article de Tromas et Elena paru en 2010.



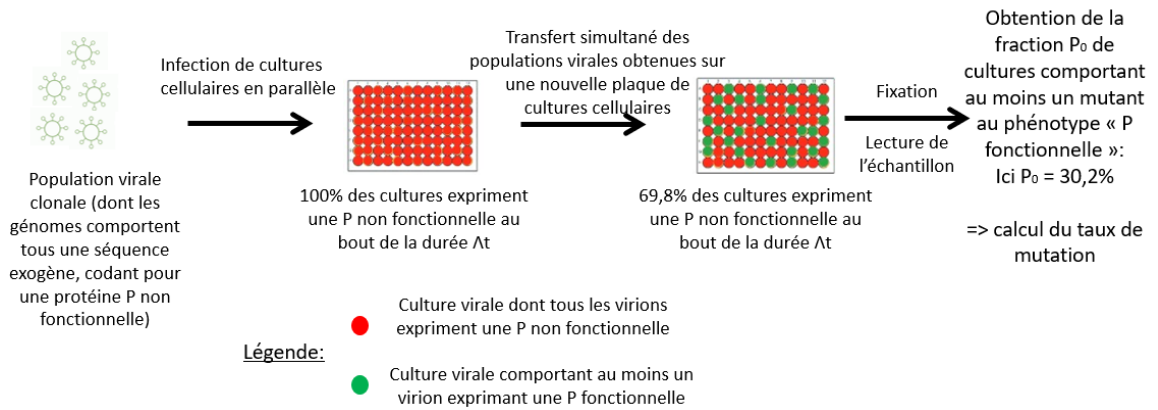
*Figure 3 : Principe de l'expérience de Tromas et Elena de 2010 visant à imposer la neutralité sélective*

A gauche : Le virus est inoculé dans un plant de tabac sauvage, en conséquence les virus n'exprimant pas la protéine Nuclear Inclusion B (NIB) ne se trouvent pas dans des conditions favorables pour se répliquer => ils ne serviront pas de matrice pour former de nouveaux ARN et seront dégradés => Seuls les ARN ne présentant pas de mutations délétères sur le gène NIB seront comptabilisés dans le calcul du taux de mutation du virus => Imprécision

A droite : Le virus est inoculé dans un plant de tabac exprimant la protéine NIB, en conséquence les virus n'exprimant pas celle-ci trouveront quand même des conditions favorables à leur répllication => ils pourront servir de matrice pour former de nouveaux ARN, et ne seront pas dégradés => Tous les ARN, y compris ceux présentant des mutations délétères sur le gène NIB seront comptabilisés dans le calcul du taux de mutation du virus => Elimination de la précédente source d'imprécision, neutralité sélective

Dans cet article, les manipulateurs ont créé une souche de tabac transgénique, dans laquelle le gène codant pour la protéine NIB (essentielle à la répllication du virus de la gravure du tabac) avait été inséré dans le génome végétal. Ainsi les virus ayant subi une mutation délétère concernant la protéine NIB pouvaient quand même survivre et être pris en compte dans les calculs par la suite. La neutralité sélective a bien été mise en place ici, en se focalisant sur la protéine la plus essentielle dans le cycle de vie de ce virus.

- On peut aussi imposer la neutralité sélective en introduisant, dans le génome du virus cette fois, une séquence exogène neutre sélectivement parlant pour le virus (comme le gène codant pour la Green Fluorescent Protein par exemple). Celle-ci sera copiée au fur et à mesure des répllications virales, et les expérimentateurs pourront ensuite mesurer les mutations survenues dans cette séquence en particulier.



*Figure 4 : Protocole pour s'affranchir de la sélection naturelle dans la détermination d'un taux de mutation viral à l'aide d'une séquence exogène (Pauly, Procaro, Lauring 2017)*

La population virale de départ (la plus clonale possible) présente dans son génome un transgène codant pour une protéine P (donnant lieu à un phénotype facilement évaluable et neutre pour le virus) non fonctionnelle, à une mutation près. On met ensuite en culture des aliquots de cette population virale dans des puits contenant des cellules réceptives et sensibles (64 ici), et on effectue plusieurs arrêts de réplication et transfert de plaque en plaque séquentiels à des intervalles de temps réguliers et brefs  $\Delta t$  (pour avoir le moins d'événements de réplication possibles entre chaque transfert= limitation de la sélection naturelle). Au bout d'un certain nombre (pré-défini) de transferts, un certain pourcentage de cultures virales aura subi au moins une fois la mutation aboutissant à un retour de P à un phénotype fonctionnel (30,2% ici), qui permettra d'obtenir, à partir du nombre de transferts et de la quantité estimée de virions dans chaque culture cellulaire, le taux de mutation du dit-virus affranchi de la sélection naturelle.

(Figure 4) Pour ce faire, il faut prendre de petits échantillons viraux pour les transférer de plaque en plaque au bout d'une durée  $\Delta t$  (qu'on choisit identique et brève), et s'intéresser aux cultures exprimant une protéine P fonctionnelle (qui sont les cultures dans lesquelles une mutation de la séquence exogène introduite a eu lieu), codée par la séquence exogène introduite à une mutation près, à la fin des étapes de transfert. La sélection naturelle est ici prise en compte, car la séquence exogène est neutre sélectivement parlant pour le virus, et les transferts multiples de plaque en plaque de petits échantillons viraux permettent de diminuer fortement la compétition entre les différents variants viraux présents au sein de chaque population.

## Objectif du travail de thèse :

Pour permettre une comparaison, les taux de mutation dans la suite de cette thèse seront exprimés en mutation/événement de copie. (Sanjuán et al. 2010) Les substitutions constituant la très grande majorité des mutations chez les virus à ARN, on effectuera l'approximation suivante pour des raisons de commodité : taux de mutation  $\sim$  taux de substitution. Enfin, les virus à ARN ayant des tailles de génome différentes, il faut bien entendu ramener ce taux de mutation global au taux de mutation par nucléotide.

On exprimera par la suite les taux de mutation des différents virus considérés en nombre de substitutions/nucléotide/événement de copie. Une fois cette base posée, des questions surviennent naturellement : comment les chercheurs déterminent-ils les taux de mutation des virus à ARN et, au-delà, quels sont les facteurs favorisant un taux de mutation plus ou moins important au sein d'une espèce virale ou d'une famille virale ? Parmi ces facteurs potentiels, nous allons nous intéresser tout particulièrement à la segmentation du génome.

En effet, certains virus à ARN présentent un génome segmenté. Un premier avantage de la segmentation est qu'un génome fragmenté en plusieurs morceaux occupe un volume moins important qu'un génome de même taille mais codé en une seule molécule d'ARN (Ojosnegros et al. 2011). Cela permet de diminuer les contraintes exercées de l'intérieur sur la capsid du virus, ce qui est favorable d'un point de vue énergétique.

Le deuxième avantage, sur lequel on se focalisera par la suite, est que les brins d'ARN issus de la réplication du génome des virions peuvent se mélanger et être échangés au sein d'une même cellule. Ainsi, lors de la co-infection d'une même cellule, un virion présentant une mutation bénéfique sur l'un des segments constitutifs de son génome, et une mutation délétère sur un autre segment, peut ainsi échanger ce dernier avec un segment indemne en provenance d'un autre virion : son infectivité s'en trouvera rétablie, c'est le phénomène de **complémentation virale** (cf partie III, 1)).

Cependant, la segmentation présente aussi des désavantages pour un virus. Le plus évident est la difficulté à emballer au moins un exemplaire de chaque fragment dans les particules virales formées ; or un génome non complet se traduira le plus souvent par un virion non viable (Nakatsu et al. 2016).

L'objectif de ce travail de thèse est donc finalement d'étudier la question suivante : est-ce que le caractère segmenté d'un virus à ARN permettrait de tolérer un taux de mutation plus élevé, grâce à la possible complémentation des segments génomiques ?

## Partie I : Les méthodes de détermination du taux de mutation

### 1) Les trois méthodes disponibles pour la détermination expérimentale d'un taux de mutation viral, avec focus sur les techniques de séquençage

Avant toute chose, il convient d'introduire une distinction importante, pour bien définir le taux de mutation qui sera notre objet d'étude. Le taux de mutation (« mutation rate ») est la probabilité à laquelle les mutations se produisent dans un génome viral en particulier ; alors que la fréquence des mutations (« mutation frequency ») est la proportion de mutants dans une population virale (Peck, Lauring 2018).

On exprimera par la suite le taux de mutation en substitution/nucléotide/copie (cf introduction). De plus, le taux de mutation étant variable selon les régions du génome (Visher et al. 2016), on indiquera dans toute la suite le chiffre correspondant à la moyenne des taux calculés sur l'ensemble du génome d'une espèce virale.

Ce cadre défini, étudions plus avant les trois principales méthodes utilisées par les chercheurs actuellement pour la détermination d'un taux de mutation chez les virus à ARN : le séquençage génomique, la méthode d'accumulation des mutations et le test de fluctuation de Luria-Delbrück (Peck, Lauring 2018).

Concernant la première méthode, on peut distinguer deux principales méthodes de séquençage génomique utilisées par les chercheurs, avec leurs avantages et inconvénients respectifs :

- le séquençage Sanger : il s'agit de la méthode la plus précise, avec 99,9% de fiabilité (Sevilla, de la Torre 2006). Elle est surtout adaptée pour séquencer de petites molécules d'ARN (aux alentours de  $10^3$  nucléotides) mais comme les protocoles de séquençage impliquent souvent de fragmenter le génome viral (d'une taille bien inférieure à un génome eucaryote), cela ne constitue pas un inconvénient de taille. Par contre, comme elle ne permet le séquençage que d'un seul génome viral à la fois, il est donc obligatoire de reproduire de nombreuses étapes afin de séquencer un nombre suffisant de génomes viraux obtenus à la suite de l'étape expérimentale de multiplication *in vivo*, ce qui est très chronophage.

- Séquençage haut débit : cela recouvre plusieurs méthodes un peu moins précises de séquençage (99,5% de fiabilité pour la majorité d'entre elles), mais qui sont beaucoup plus rapides, et permettent l'analyse simultanée de nombreuses molécules (et donc de séquencer des génomes beaucoup plus longs). Pour cette raison, ce sont les plus utilisées actuellement, mais comme moins précises que la méthode Sanger, elles nécessitent obligatoirement une étape de correction des erreurs de séquençage (Glenn 2011).

Il convient donc d'évaluer quelles sont les sources d'erreurs possibles lorsqu'on utilise le séquençage pour quantifier le taux de mutation d'un virus, et comment les prendre en compte.

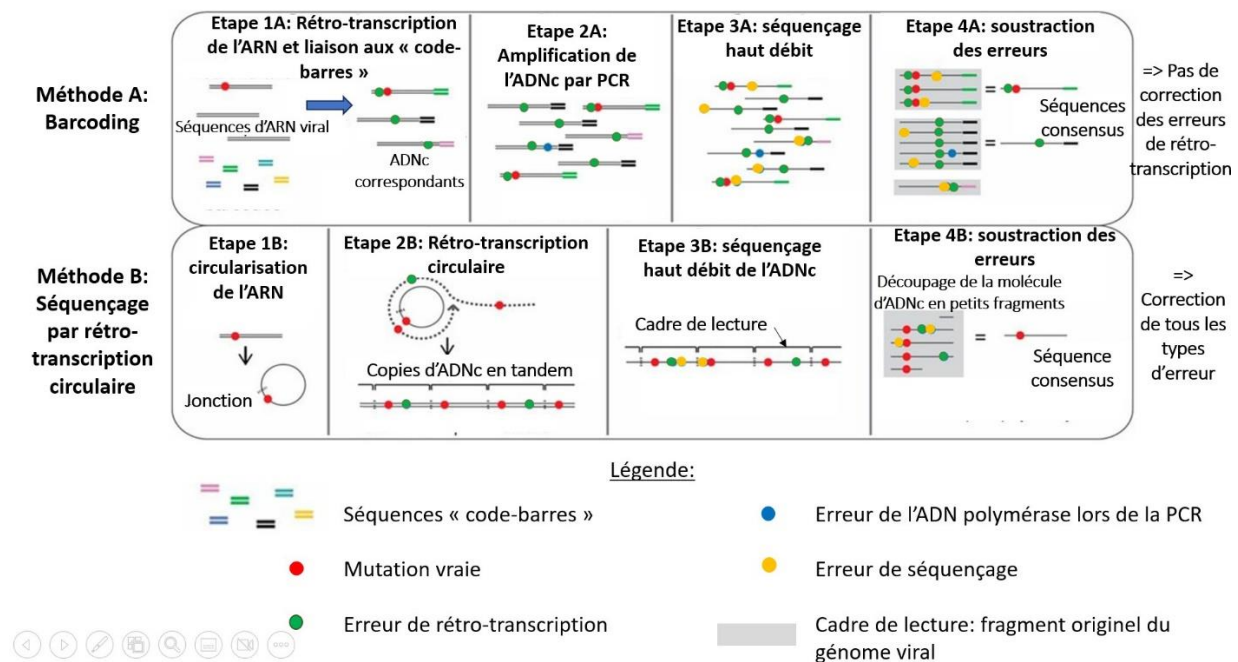
## 2) Première méthode de détermination du taux de mutation : le séquençage génomique, avec correction des erreurs artéfactuelles par rétro-transcription circulaire

Le séquençage est la méthode la plus directe, au cours de laquelle l'ARN des virus issu du nombre le plus faible possible d'événements de copie (au moyen d'un temps de contact très faible entre virus et cellules) est recueilli, puis subi une étape de RT-PCR (Rétro-Transcription, puis amplification par Réaction de Polymérisation en Chaîne) et de séquençage proprement dit (Smith et Inglis, 1987). Cependant, il présente plusieurs sources d'imprécision de mesure, parmi lesquelles :

- ❖ séquençage uniquement des génomes avec une fitness importante (au moins équivalente à celle du type sauvage), peu de séquençage des virions avec des mutations délétères, pas de séquençage des virions avec des mutations létales.
  
- ❖ mesure du taux de mutation en substitution/nucléotide/cycle infectieux, ce qui implique d'utiliser une méthode de comparaison de séquence sur les ARN viraux obtenus expérimentalement pour le convertir en substitution/nucléotide/événement de copie. Parmi ces méthodes de comparaison, nous avons déjà détaillé la méthode Yasukawa en introduction, figure 2. Celle-ci permet bien, en comparant directement le génome du virus originel qui a infecté la cellule et ceux des virions qui sortent de cette dernière ou après sa lyse (plusieurs événements de copie subis ; obtention du taux en substitution/nucléotide/cycle infectieux), de remonter au taux en substitution/nucléotide/événement de copie. Mais comme elle se base sur une population virale initiale qui, bien que clonale, présente des différences de génome entre les virions, il existe forcément une imprécision lorsque l'on calcule un taux de mutation en comparant la séquence initiale (en réalité plusieurs séquences non identiques entre les virions initiaux) et les séquences génomiques des virions obtenus en fin d'expérience.
  
- ❖ nécessité d'une étape de transcription inverse (étant donné que nous étudions des virus à ARN) puis d'amplification par PCR et de séquençage, qui sont toutes sources d'erreurs supplémentaires (interférant donc avec l'identification des mutations vraies). Selon l'étude de Peck et Lauring de 2018, il y a autant de mutations introduites par ces trois étapes que de mutations vraies produites par les polymérases virales (observation réalisée grâce à l'utilisation de rétrotranscription circulaire, cf ci-dessous).

Pour pallier cela, il existe deux solutions expérimentales, le barcoding et le séquençage circulaire. La figure 5 détaille le principe de ces deux techniques, appliqués à un virus à ARN simple brin :





*Figure 5 : Utilisation du barcoding et du séquençage circulaire pour discriminer mutations vraies et dues au séquençage/PCR/rétro-transcription (Lou et al. 2013)*

(A) Barcoding: On identifie les segments d'ARN viraux via une étape de rétro-transcription et de liaison à des séquences code-barre. Puis ces segments sont amplifiés et séquencés. L'étape finale de traitement informatique permet d'avoir accès directement aux séquences des segments présents dans le virus, mais ne permet pas de corriger les erreurs de rétro-transcription, qui sont générées avant l'étape d'amplification et donc indétectables.

(B) Séquençage par rétro-transcription circulaire: Les segments d'ARN viraux sont circularisés individuellement, puis servent de matrice à une rétro-transcriptase spéciale qui génère une longue molécule d'ADNc avec des copies de l'ADNc correspondant au segment viral côte-à-côte. Puis cette longue molécule sera segmentée aléatoirement et séquencée. L'étape de traitement informatique devra donc reconstruire le segment viral originel à partir de la fragmentation aléatoire de l'étape 3B, mais permettra d'isoler les mutations vraies de toutes les sources d'erreur artéfactuelles (l'étape de rétro-transcription et d'amplification ayant lieu en même temps dans cette méthode)

La méthode du barcoding, comme on le voit sur ce schéma, présente l'inconvénient de ne pas corriger les erreurs de la transcriptase inverse (erreurs introduites dès l'étape 1A). Ces erreurs seront par la suite impossibles à distinguer des mutations vraies (les erreurs de rétro-transcription persistent dans la séquence consensus de l'étape 4A), ce qui est problématique pour évaluer de façon fiable un taux de mutation.

C'est pourquoi nous allons nous intéresser à la méthode la plus efficace pour détecter les erreurs des trois sources sus-citées : le séquençage par rétrotranscription circulaire. Selon celui-ci, l'ARN génomique viral obtenu est fragmenté afin d'obtenir des fragments de taille compatible avec les techniques de séquençage utilisées (fragments originels). Les fragments d'ARN sont circularisés (1B), puis servent de matrice à une transcriptase inverse avec activité de déplacement de brin, ce qui permet la synthèse d'une longue molécule d'ADNc composée de nombreuses répétitions côte-à-côte de la séquence correspondant au fragment génomique viral (2B). Il faut considérer qu'à l'étape 2B du schéma, il y a déjà des erreurs par rapport à la séquence génomique du virus récolté, mais seules les erreurs qui sont présentes sur l'ensemble des segments d'ADNc ainsi formés correspondront aux véritables mutations : c'est la clé de voûte de toute la méthode. Cette longue molécule d'ADNc sera ensuite séquencée (3B), ce qui

nécessite de la fragmenter aléatoirement. Enfin, l'étape de traitement informatique (4B) permettra, à partir de ces fragments aléatoires, d'identifier les fragments originels du génome viral (« read » sur le schéma), de reconstruire la séquence virale récoltée (« consensus sequence » sur le schéma) et de déterminer ce qui correspond aux mutations véritables présentes à l'origine sur l'ARN viral : ce sont les erreurs (par rapport au génome du virus de type sauvage utilisé pour infecter la cellule) présentes sur absolument tous les fragments originels. Cette technique permet de réduire considérablement le nombre d'erreurs artéfactuelles (Lou et al. 2013), mais elle est très lourde à mettre en place.

### 3) Deuxième méthode de détermination du taux de mutation : la méthode d'accumulation des mutations

En raison de tous les inconvénients qu'implique le séquençage, et de la lourdeur de protocole expérimental nécessaire pour les compenser, il peut être judicieux d'utiliser une méthode de détermination indirecte.

La méthode d'accumulation des mutations, elle, implique de soumettre des populations virales à des « goulets d'étranglement » de manière séquentielle (en faisant des transferts de « clones viraux » d'une plaque de lyse donnée à une nouvelle population de cellules par ex), pour éviter la sélection vis-à-vis des mutations délétères (Peck, Lauring 2018). Un transfert de virus récoltés d'une plaque de lyse implique de sélectionner une seule plaque de manière aléatoire pour inoculer une très faible quantité de clones viraux au lot de cellules suivant.

On définit une unité virale formant colonie (UvFC) comme le nombre de plaques de lyse apparaissant sur une gélose présentant une concentration cellulaire donnée. Il s'agit d'une unité de mesure de la quantité de particules virales infectieuses présente dans un échantillon. En sélectionnant continuellement une seule UvFC à chaque cycle de passage, on peut fixer les mutations délétères dans la population virale, et ainsi les détecter par séquençage. Cependant, bien que cette méthode permette de détecter les mutations délétères, ce n'est toujours pas le cas pour les mutations létales.

En connaissant le nombre de cycle infectieux nécessaires à la formation d'une plaque de lyse (et le mode de réplication du virus, souvent proche du binaire), cette méthode permet d'obtenir le taux de mutation en substitution/nucléotide/cycle infectieux. Cette méthode est cependant peu fiable pour des comparaisons entre virus, comme vu en introduction.

#### 4) Troisième méthode de détermination du taux de mutation : le test de fluctuation de Luria-Delbrück

##### a) Principe du test

Le test de fluctuation de Luria-Delbrück, autre méthode indirecte, mesure directement le taux de mutation en substitution/nucléotide/événement de copie, quel que soit le mode de répllication (Luria, Delbrück 1943). Il consiste à ensemencer à partir de la même source un grand nombre de cultures parallèles à l'aide d'un petit inoculum et à analyser le retour de la population des différentes cultures à un phénotype spécifique. L'inoculum doit contenir le nombre le plus faible possible de virus, et leur génome doit être connu autant que possible (même si en pratique, c'est avant tout la séquence des *loci* aboutissant au phénotype d'intérêt qu'il faut absolument connaître).

La version la plus utilisée de cette expérience consiste à partir d'une population virale au phénotype altéré. Elle repose le plus souvent sur l'utilisation d'un gène rapporteur (gène codant pour des protéines liées à un phénotype mesurable : résistance à un médicament, à un anticorps monoclonal, fluorescence...). Ce gène rapporteur doit idéalement être neutre sélectivement parlant, mais le test procure une bonne estimation du taux de mutation même si le gène ne l'est pas tout à fait (Sanjuán et al. 2010). Prenons l'exemple d'un test pour lequel le gène rapporteur est celui codant pour la Green Fluorescent Protein (GFP) : dans ce cas, dans la population virale initiale (la plus clonale possible dans l'idéal), tous les virus devront avoir intégré en leur génome un gène de GFP comportant une (ou plusieurs) substitution précise par rapport au gène GFP sauvage, qui rend la GFP produite non fluorescente (phénotype altéré). La population initiale est donc non fluorescente, et les substitutions sont considérées comme représentatives des mutations dans leur ensemble (Sanjuán et al. 2010). Le nombre de sites de substitutions qui sépare le gène introduit du gène GFP fonctionnel sera noté  $T_s$ . Il faut également connaître le nombre de substitutions sur chacun de ces sites, parmi les trois possibles, qui aboutissent à un retour du gène GFP à une séquence fonctionnelle (noté  $N_s$ , idéalement identique entre tous les sites). Une fois cette population virale initiale générée, on inocule comme décrit ci-dessus des échantillons de cette population de tailles identiques et connus à un grand nombre de cultures cellulaires en parallèle (étape de multiplication *in vivo*). Puis on attend qu'il y ait au moins une culture comportant au moins un virion au phénotype non altéré, donc au moins un virion fluorescent dans notre exemple. On stoppe alors la répllication virale artificiellement dans toutes les cultures simultanément, puis on dénombre alors le nombre de virions final présents dans les cultures (par PCR quantitative par exemple).

Ensuite, pour calculer le taux de mutation du virus utilisé, Luria et Delbrück ont utilisé la propriété suivante : bien que la distribution du nombre de mutants par culture dépende du mode de répllication et de la sélection naturelle, la fraction de cultures comportant au moins un mutant au phénotype non altéré  $P_0$  (c'est-à-dire au moins un mutant fluorescent dans notre exemple) n'en dépend pas. Étant donné que les mutations sont des événements rares et aléatoires, leur nombre par culture doit suivre une distribution de Poisson. Elles se produisent donc avec une probabilité  $P_0 = e^{-m(N_1-N_0)}$ , où  $m$  est le taux de mutation amenant à la réversion du phénotype altéré par copie de brin et  $N_1-N_0$  est la croissance de la population à cette génération. En obtenant  $P_0$  et  $N_1-N_0$  à l'issue de l'expérience de fluctuation, on peut calculer  $m$  (en

substitution/nucléotide/événement de copie). Enfin, en prenant en compte que seules deux substitutions, parmi les trois possibles, aboutissent à une réversion de la non-fluorescence ici ( $N_s=2$ ), on peut calculer  $\mu_{s/n/copy} = 2m/T_s = -2 \times \ln(P_0) / ((N_1-N_0) \times T_s)$

Exemple théorique : Si on considère que dans le gène de la GFP introduit dans une certaine expérience, 14 substitutions séparent la séquence de ce dit-gène de celle d'un gène codant pour une GFP fonctionnelle, qu'on a  $N_s=3$  sur chacune de ces 14 substitutions, qu'on réalise 100 cultures de virus, que notre nombre initial de virus introduit est de 100 particules virales, qu'on a stoppé la réplication virale dans les cultures dès l'apparition de la première culture fluorescente ( $P_0= 1/100=0,01$ ), et que le nombre moyen final de virions obtenu dans les cultures est de  $1 \times 10^6$  particules virales, on obtient alors un taux de mutations pour ce virus de :  $\mu = -3 \times \ln(0,01) / (14 \times (10^6-100)) = 4,29 \times 10^{-7}$  substitutions/nucléotide/copie.

Une autre version du test, celle utilisé dans l'expérience originale de 1943 (Luria, Delbrück 1943), consiste à prendre des virus sans phénotypes particuliers au départ, et à suivre ensuite la fraction de cultures ne comportant que des mutants non viables (c'est-à-dire ayant subi une mutation létale).

### b) Limites du test

Ce test s'affranchit de tous les inconvénients des précédentes méthodes, ce qui permet de mesurer précisément le véritable taux de mutation. Cette méthode présente le désavantage de n'évaluer les mutations qu'en un *locus* précis du génome du virus (celui du gène rapporteur), alors qu'il est avéré que les taux de mutation sont différents selon les régions du génome étudiées (Visher et al. 2016). Mais cela peut être pallié en introduisant plusieurs gènes rapporteurs en différents *loci* du génome (par exemple plusieurs gènes altérés codant pour une protéine fluorescente), et à mesurer ensuite le degré d'expression de la protéine d'intérêt (le degré de fluorescence dans l'exemple précédent). Parmi les autres conditions d'utilisation de cette méthode, il faut connaître précisément tous les *loci* qui peuvent aboutir au résultat phénotypique recherché (directement ou indirectement), et connaître leur état chez les virus générés avant l'étape de multiplication *in vivo* (évident dans le cas de la fluorescence). Également, il faut s'assurer que les gènes codant pour l'ARN polymérase des virions considérés ne présentent aucune mutation par rapport au type sauvage, car cela influera inévitablement sur le taux de mutation de ces virions (Smith et Inglis, 1987). Par exemple, au cours de la réplication de l'ARN qui s'effectue de l'extrémité 5' vers l'extrémité 3' (pour le brin néoformé), certaines ARN-polymérase ARN-dépendantes de virus possèdent une activité de correction (dite exonucléase 3'-5'), qui influe bien entendu sur le taux de mutation de ce virus (te Velthuis et al. 2010). Une ARN-polymérase dont la sous-unité à activité exonucléasique présente une mutation peut voir son fonctionnement altéré, et donc le taux de mutation calculé sur ce virus expérimental ne sera pas représentatif du virus de type sauvage. Enfin, il faut bien sûr veiller, lors de l'étape de multiplication *in vivo*, à mettre en contact initialement les cellules avec un petit échantillon de clones viraux, dont le génome doit être le plus proche possible les uns des autres, et surtout comporter une séquence du gène rapporteur parfaitement identique. Cette population virale initiale peut être obtenue par génétique inverse, en transfectant une cellule avec des plasmides de séquence connue et identique, qui correspond à l'intégralité de la séquence génomique virale plus à la séquence du gène rapporteur modifiée comme il se doit.

Si on se place toujours dans l'approximation définie en introduction (à savoir que le taux de mutation des virus à ARN est équivalent au taux de substitutions par nucléotide), une autre limite est à prendre en compte : chaque substitution (A->U, C->G, C->A ...) se produit avec une probabilité propre.

Dans l'article de Pauly, Procario et Lauring de 2017, c'est pour cette raison que les manipulateurs ont effectué un test de fluctuation avec douze sous-groupes de virus Influenza A, aux génomes viraux construits de telle manière qu'à chaque fois c'était une substitution différente qui amenait à la réversion de la non-fluorescence. Le gène rapporteur utilisé codait pour la Green Fluorescent Protein (GFP), ce qui constitue un bon choix car le caractère « fluorescent/non fluorescent » est neutre sélectivement pour un virus (cf deuxième paragraphe). Les auteurs ont également pris soin de placer les gènes rapporteurs dans des zones représentatives (au regard du taux de mutation) de la majorité du génome de l'Influenza (selon une analyse *in silico* de la structure secondaire de son ARN). Par cette méthode, ils ont déterminé un taux de mutation pour le virus influenza A de  $\mu = 2 \times 10^{-5}$  s/n/copie environ.

## 5) Récapitulatifs des différentes méthodes de détermination du taux de mutation

Pour résumer, voici l'ensemble des différents avantages et inconvénients de ces méthodes de détermination (Tableau 1).

Méthode	Avantages	Inconvénients
Séquençage haut débit	<ul style="list-style-type: none"> <li>- Permet de distinguer les taux de mutation des différentes régions du génome</li> <li>- Permet de connaître précisément la séquence génomique du virus, et donc d'identifier les mutations autres que les substitutions comme telles</li> </ul>	<ul style="list-style-type: none"> <li>- Nécessite une étape de comparaison de séquences pour remonter au taux de mutation en s/n/copie</li> <li>- Nécessite une étape permettant d'identifier les mutations artéfactuelles (barcoding/séquençage circulaire)</li> <li>- Biaisé vis-à-vis des mutations délétères et létales (nécessite une adaptation de protocole pour prendre en compte l'effet de la sélection naturelle)</li> </ul>
Méthode d'accumulation des mutations	<ul style="list-style-type: none"> <li>- Non biaisé vis-à-vis des mutations délétères</li> <li>- Permet de connaître précisément la séquence génomique du virus, et donc d'identifier les mutations autres que les substitutions comme telles</li> </ul>	<ul style="list-style-type: none"> <li>- Nécessite une étape de comparaison de séquences pour remonter au taux de mutation en s/n/copie</li> <li>- Nécessite une étape permettant d'identifier les mutations artéfactuelles (barcoding/séquençage circulaire)</li> <li>- Biaisée vis-à-vis des mutations létales</li> </ul>
Test de fluctuation de Luria-Delbrück	<ul style="list-style-type: none"> <li>- Pas d'erreurs de RT-PCR, ni de séquençage</li> <li>- Indépendant de l'effet de la sélection naturelle</li> <li>- Obtention directe du taux de mutation en s/n/copie (permet de comparer avec d'autres espèces virales)</li> </ul>	<ul style="list-style-type: none"> <li>- Requier un gène rapporteur séquencé, donnant lieu à un phénotype mesurable</li> <li>- N'évalue le taux de mutation que dans une seule région du génome</li> <li>- Ne permet d'identifier que les substitutions (qui sont cela dit les plus fréquentes des mutations)</li> </ul>

Tableau 1 : Avantages/inconvénients des différentes méthodes de détermination du taux de mutation

Au bilan, ces méthodes présentent toutes des avantages et des inconvénients, mais la plus fiable, celle qui fait office de référence pour le calcul des taux de mutation, reste le test de fluctuation de Luria-Delbrück.

Conclusion partie 1 : Ainsi, ils existent différentes méthodes de détermination, présentant chacune plus ou moins de limites à prendre en compte pour calculer un taux de mutation fiable. Grâce à ces outils, il est possible d'étudier quel est l'impact d'une augmentation ou d'une diminution du taux de mutation sur la « fitness » (valeur évolutive) d'un virus, et ainsi de savoir quels sont les facteurs expliquant que certains virus soient plus tolérants que d'autres à une augmentation de leur taux de mutation.

## Partie II : Étude d'une possible influence de la segmentation génomique sur la tolérance à des taux de mutation élevés

### 1) De grands écarts de taux de mutation entre les différentes espèces de virus à ARN

Intéressons-nous maintenant aux facteurs qui expliquent la tolérance aux mutations des virus à ARN, autrement dit les facteurs expliquant les différences de taux de mutations observées entre certaines espèces virales. Comme vu dans la partie précédente, il existe trois grandes méthodes de détermination du taux de mutation, avec de nombreuses déclinaisons possibles au sein de chaque méthode. Voici les taux de mutation  $\mu$  déterminés sur plusieurs espèces virales au moyen d'une de ces méthodes:

Virus considéré	Génome segmenté	Taux de mutation $\mu$ estimé (en s/n/copie)	Méthode de détermination du taux
Virus de la chorioméningite lymphocytaire (ARN +)	oui	$2,6 \times 10^{-6}$ (Combe, Sanjuán 2014)	Test Luria (résistance à un anticorps monoclonal dirigé contre la glycoprotéine d'enveloppe G) *
Virus de la mosaïque de la luzerne (ARN +)	oui	$6,7 \times 10^{-5}$ (van Vloten-Doting, Bol, Cornelissen 1985)	Séquençage direct (méthode Sanger)
Virus de la gravure du tabac (ARN +)	non	$4,75(\pm 1,7) \times 10^{-6}$ (Tromas, Elena 2010; Sevilla, de la Torre 2006)	Séquençage direct (haut débit)
Virus de la stomatite vésiculeuse (ARN -)	non	$6 \times 10^{-6}$ (Sanjuán et al. 2010)	Méthode d'accumulation des mutations
Virus influenza A (ARN -)	oui	$2,1(\pm 0,5) \times 10^{-5}$ (Pauly, Procaro, Lauring 2017)	Test Luria (réversion de la fluorescence)
Virus de la rougeole (ARN -)	non	$2(\pm 0,2) \times 10^{-6}$ (Zhang et al. 2013)	Test Luria (réversion de la fluorescence)

Le phénotype d'intérêt utilisée dans le calcul de  $\mu$  est mentionné directement dans le tableau.

Tableau 2 : Quelques exemples de taux de mutation de différents virus à ARN

\* Pour faire un test de fluctuation de Luria-Delbrück en utilisant la résistance à un anticorps comme phénotype d'intérêt, il faut faire se répliquer un virus de type sauvage dans de nombreux puits de culture (où se trouvent des cellules) en parallèle, pendant environ deux jours, puis récolter les virus. Ensuite, on expose ces virus à une solution de l'anticorps utilisé, puis on filtre cette solution avec une matrice constituée de billes recouvertes d'anticorps complémentaires des anticorps monoclonaux (qui permet de recueillir uniquement les virus dont le gène codant pour la protéine complémentaire de l'anticorps n'a pas muté). Enfin, on dénombre les virus retenus par le filtre et ceux qui ont réussi à passer, et on obtient alors la fraction de virus ayant subi une mutation conférant la résistance à l'anticorps dans chaque culture, et donc la proportion  $P_0$  de cultures dont 100% des virus sont résistants à l'anticorps. A l'aide des formules détaillées en partie I, 4), a), on peut finalement calculer le taux de mutation  $\mu$ .

On constate des différences de taux de mutation importantes, jusqu'à un ordre de grandeur. Mais ces résultats peuvent parfois s'expliquer par des différences de méthode utilisées.

Considérons par exemple, dans l'article de Sevilla et de la Torre de 2006, deux taux de mutation viraux en particulier : celui du virus de la chorioméningite lymphocytaire ( $5.5 \times 10^{-6}$  s/n/copie, selon cet article), et celui du virus de la stomatite vésiculeuse ( $7,3 \times 10^{-6}$  s/n/copie, selon ce même article). Les taux sus-cités ont été déterminé par méthode directe, c'est-à-dire une étape de RT-PCR (transcription inverse, puis amplification par réaction de polymérisation en chaîne), puis une étape de séquençage haut débit.

Ces deux mêmes virus ont également vu leur taux de mutation déterminé dans l'article de Combe et Sanjuán de 2014. Et dans ce dernier, il est mentionné un taux de mutation de  $2.6 \times 10^{-6}$  s/n/copie pour le virus de la chorioméningite lymphocytaire, et un taux de  $6,15 \times 10^{-6}$  s/n/copie pour le virus de la stomatite vésiculeuse. Ces taux ont été mesurés avec un test de Luria-Delbrück, avec comme phénotype évalué la résistance à un anticorps monoclonal dirigé contre la glycoprotéine d'enveloppe du virus étudié.

On voit bien que sur une même espèce virale, un simple changement de méthode suffit à modifier le taux de mutation calculé : le  $\mu$  calculé via un séquençage est supérieur au  $\mu$  issu d'un test de fluctuation de Luria-Delbrück (à cause des sources d'erreurs non prises en compte par le séquençage, cf partie I, 2)). Pour identifier les facteurs intrinsèques aux virus influant sur la tolérance aux mutations, il convient donc de comparer des études utilisant des méthodes de détermination les plus proches possibles, et bien entendu des virus assez proches phylogénétiquement parlant (ne pas comparer un virus à ARN de polarité différente (+/-) par exemple).



## 2) Analyse d'un premier cas de divergence du taux de mutation $\mu$ : le virus de la gravure du tabac et le virus de la mosaïque de la luzerne

A partir du tableau 2, intéressons-nous au cas de deux virus à ARN positif, dont le taux de mutation a été déterminé par séquençage pour tous deux. Dans l'article de Tromas et Elena publié en 2010, les auteurs ont déterminé le taux de mutation du virus de la gravure du tabac (*Tobacco etch potyvirus* : TEV), un virus à ARN(+) non-segmenté. Ils ont utilisé une méthode de transcription inverse suivi d'un séquençage haut débit avec barcoding et traitement informatique. Cela permet d'exclure facilement les mutations artéfactuelles dues à l'amplification et au séquençage proprement dit, mais les erreurs de transcription inverse demeurent non prises en compte (cf partie 1, 2)). D'autre part, ils ont inoculé diverses souches de virus de la gravure du tabac à des plants de tabac transgéniques : ceux-ci expriment la protéine Nib (Nuclear inclusion protein b), nécessaire à la réplication du TEV. De cette manière, la protéine Nib est présente en toutes circonstances dans le cytoplasme des cellules végétales, ce qui permet aux virions de TEV présentant une mutation délétère sur le gène codant pour la protéine Nib de pouvoir survivre malgré tout, et d'être donc séquencés ensuite. C'est une manière de s'affranchir du biais de la sélection naturelle (cf première partie). A l'issue de leurs expériences, le taux de mutation obtenu était de  $\mu_{TEV} = 4,75(\pm 1,7) \times 10^{-6}$  s/n/copie.

L'article de van Vloten-Doting, Bol et Cornelissen de 1985, lui, étudie le virus de la mosaïque de la luzerne (Alfalfa Mosaic Virus : AMV), qui est un virus à ARN (+) segmenté en trois parties. Les auteurs ont également utilisé une méthode de transcription inverse suivi d'un séquençage direct. Ils ont grâce à cela déterminé un taux de mutation de  $\mu_{AMV} = 6,7 \times 10^{-5}$  s/n/copie, donc plus élevé d'au moins un ordre de grandeur que son parent le TEV. On aurait donc tendance à conclure que c'est un facteur intrinsèque à ces virus : par exemple, le fait que l'AMV soit un virus au génome segmenté alors que ce n'est pas le cas du TMV. Mais cette comparaison est biaisée, notamment par le fait que le séquençage utilisé par van Vloten-Doting, Bol et Cornelissen suivait la méthode Sanger (une des rares techniques disponibles à l'époque de l'étude), *a contrario* de Tromas et Elena qui ont utilisé un séquençage haut débit. De plus, le protocole de van Vloten-Doting, Bol et Cornelissen ne comprenait aucun des mécanismes de correction des erreurs artéfactuelles (transcription inverse, PCR, séquençage) mis en place par Tromas et Elena (qui ne corrigent pas non plus les erreurs de transcription inverse), ni de moyen de prendre en compte ou de s'affranchir de l'effet de la sélection naturelle. Or, nous avons vu en première partie que tous ces aspects constituent une grande source de biais expérimentaux, ce qui empêche d'identifier de façon valide un facteur intrinsèque à ces virus expliquant leur grande différence de taux de mutation.

### 3) Analyse d'un deuxième cas de divergence du taux de mutation $\mu$ : le virus de la rougeole et le virus influenza A

#### a) Présentation du cas

Comparons maintenant deux autres espèces virales aux taux de mutation mentionnés dans le tableau 2 : le virus de la rougeole (Measles Virus : MSV) et le virus influenza A (InfAV). Il s'agit de deux virus à ARN (-), ce facteur intrinsèque n'entrera donc pas en jeu par la suite.

Dans les articles de (Pauly, Procario, Lauring 2017) et (Zhang et al. 2013), les auteurs ont introduit des gènes rapporteurs codant pour la GFP au moyen de plasmides dans des virus Influenza A et de la rougeole respectivement. Ils ont ensuite effectué un test de fluctuation de Luria-Delbrück en étudiant le retour au phénotype fluorescent (caractère « fluorescent/non fluorescent » neutre sélectivement) comme moyen pour déterminer le taux de mutation. Il s'agit de la famille de méthodes la plus fiable comme vu dans la précédente partie, et toutes les conditions de validité de ce test ont été respectées dans ces deux études (nombre de substitutions donnant lieu à la réversion de la fluorescence connue, etc...). Les auteurs ont ainsi déterminé un taux de mutation de respectivement  $\mu_{\text{InfAV}} = 2,1 (\pm 0,5) \times 10^{-5}$  s/n/copie pour le virus influenza A et  $\mu_{\text{MSV}} = 2 (\pm 0,2) \times 10^{-6}$  s/n/copie pour le virus de la rougeole.

L'utilisation par ces auteurs d'une méthode strictement identique pour déterminer le taux de mutation de ces deux virus dans ces études permet de se pencher plus avant sur les autres facteurs (ou autres potentiels biais) donnant lieu à cette différence de  $\mu$  conséquente (1 ordre de grandeur environ).

b) Analyse d'un premier biais de comparaison potentiel : la taille du génome

Un facteur intrinsèque aux virus reconnu comme influant sur leur taux de mutation est la taille du génome. On constate globalement une légère corrélation négative entre taille du génome et taux de mutation chez les virus à ARN (Sanjuán et al. 2010) (Peck, Lauring 2018) .

Prenons l'exemple de la famille des Nidovirales (coronavirus, torovirus, ronivirus), dont les membres présentent les tailles de génome les plus élevées de tous les virus à ARN (de 20 à 40 kb environ). Il s'agit d'une des rares familles de virus à ARN à posséder une ARN-polymérase ARN dépendante (ApAd) présentant une activité correctrice (sous-unité NSP12 (Non-Structural Protein) pour les coronavirus, parmi les 16 sous-unités que compte leur ApAd), et donc des taux de mutations bien plus faibles (te Velthuis et al. 2010). Eckerle et al., dans leur étude de 2010, évalue notamment le taux de mutation d'une souche de SARS-CoV (Severe Acute Respiratory Syndrom–Coronavirus) au moyen d'un séquençage circulaire haut débit, à  $\mu \sim 5 \times 10^{-7}$  s/n/copie. Ils ont préalablement déterminé, au moyen d'essais de culture et d'établissement d'une courbe étalon, le nombre d'événement de copie ayant lieu chez la souche de SARS-CoV utilisée en fonction du temps ; sachant cela, ils ont mis en place un nombre d'événements de copie très faible pour leur expérience d'évaluation du taux de mutation (par un temps de contact bref) entre virus et cellules, ce qui limite grandement l'influence de la sélection naturelle, et permet de se dispenser de l'utilisation d'une méthode de comparaison de séquence. Les inconvénients du séquençage pour déterminer le taux de mutation décrits en première partie étaient ainsi compensés.

Concernant le facteur taille du génome, le virus de la rougeole a un génome de 15 894 nucléotides de long (Zhang et al. 2013), celui du virus Influenza A fait 13 588 nucléotides de long (Pauly, Procario, Lauring 2017) et celui du SARS-CoV utilisé fait 29 903 nucléotides de long (Eckerle et al. 2010). On constate qu'on observe quasiment le même écart de taux de mutation (facteur 10) entre l'InfAV et le MsV qu'entre le MsV et le SARS-CoV, alors que les tailles de l'InfAV et du MsV sont quasi-similaires, contrairement aux tailles du MsV et du SARS-CoV entre lesquelles il y a un facteur 2. Le facteur taille ne suffit donc pas à expliquer la grande différence de taux de mutation entre MsV et InfAV mise en évidence plus haut.

c) Analyse d'un deuxième biais de comparaison potentiel : la région du génome

Un autre facteur expérimental qui pourrait expliquer la différence de taux obtenue entre le MsV et l'InfAV est la région du génome dans laquelle a été insérée le transgène de la GFP. En effet, peut-être une différence de taux de mutation est-elle présente au sein même du génome de ces deux virus suivant les régions de leur génome considérées, ce qui peut constituer un biais.

Concernant l'expérience de Zhang et al. de 2013, ce biais a été pris en compte par les expérimentateurs. Ils ont construit expérimentalement plusieurs populations clonales de départ de virus de la rougeole, pour lesquels le transgène de la GFP a été inséré à cinq endroits (à chaque fois différents) au sein du cadre de lecture de la molécule d'ARN du MsV. Il s'agit d'une des sources d'incertitude expliquant le taux de  $\mu_{\text{MsV}} = 2(\pm 0,2) \times 10^{-6}$  s/n/copie déterminé dans leur expérience. Globalement, le taux de mutation déterminé reste toujours compris entre  $1,8 \times 10^{-6}$  et  $2,2 \times 10^{-6}$  s/n/copie, il apparaît donc qu'il n'y a pas spécialement de « régionalisation » du taux de mutation au sein du cadre de lecture du génome du virus de la rougeole.

Pour l'expérience de Pauly, Procaro et Lauring de 2017, les expérimentateurs ont également pris en compte ce biais d'une manière similaire. Ils ont pris soin d'insérer le transgène dans deux régions différentes au sein de chacun des huit segments d'ARN du virus Influenza A. Cela explique l'incertitude plus élevée que dans l'expérience de Zhang et al., mais le  $\mu_{\text{InfAV}}$  reste cantonné entre  $1,6 \times 10^{-5}$  et  $2,6 \times 10^{-6}$  s/n/copie, ce qui est toujours significativement différent du  $\mu_{\text{MsV}}$ . Ainsi donc, l'effet de « régionalisation » ne suffit pas non plus à expliquer la différence de taux de mutation observée entre ces deux virus.

Au final, au vu de la très grande similitude entre les protocoles de détermination du taux de mutation du MsV et de l'InfAV détaillés plus haut, et de l'absence de biais évident contribuant à cet écart conséquent de taux de mutation, un facteur intrinsèque de différence entre ces deux virus est ici le plus probable. Parmi ces facteurs intrinsèques, celui qui semble le plus à même d'influer sur le taux de mutation de ces deux virus est la segmentation du génome. En effet, le virus Influenza A présente un génome segmenté en huit fragments, tandis que le génome du virus de la rougeole est d'un seul tenant. Les virus à ARN (-) segmentés sembleraient donc capables sélectivement parlant de supporter un taux de mutation plus élevé que les virus à ARN (-) non-segmentés.

Conclusion partie 2 : Il existe donc bien des virus à ARN plus robustes que d'autres aux variations de leur taux de mutation (et notamment à son augmentation). Un facteur explicatif en particulier semble se dégager, qu'il convient d'investiguer plus en profondeur : la segmentation ou non du génome du virus.

## Partie III : Bénéfices possibles de la complémentation

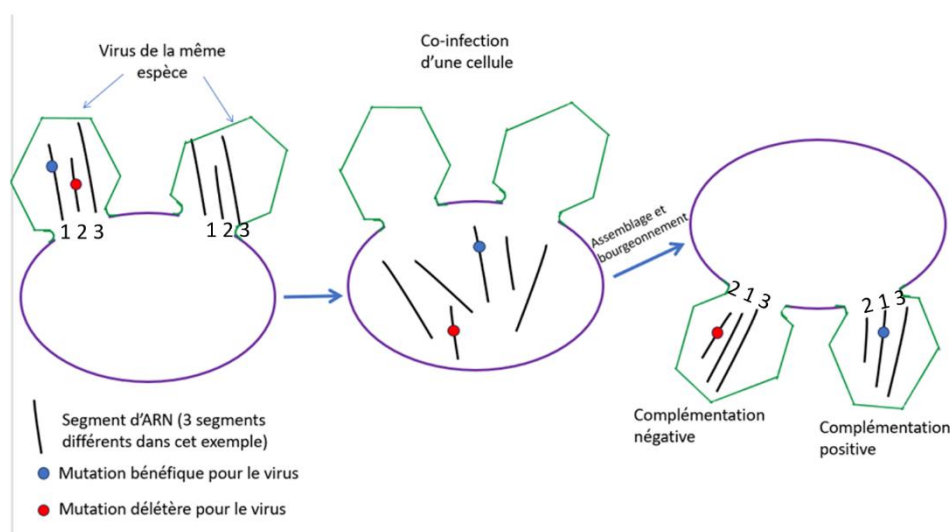
### 1) Définition

La segmentation est un caractère particulièrement important dans l'étude des génomes viraux. En effet, en découpant son génome en plusieurs fragments, le virus recrée en quelque sorte les chromosomes multiples des eucaryotes. Cela permet l'existence chez les virus de phénomènes génétiques décrits à l'origine dans ce taxon, dont l'un nous intéresse tout particulièrement : la complémentation.

La complémentation, telle que définie chez les eucaryotes (Fincham, 1966), fait référence au processus génétique selon lequel deux variants d'un même organisme avec différentes mutations homozygotes récessives qui donnent lieu au même phénotype mutant (par rapport au type sauvage), lorsqu'accouplés, produisent une progéniture avec le phénotype sauvage.

Adaptée aux virus, on pourrait décrire la **complémentation** comme suit : processus génétique par lequel deux virus d'une même espèce porteurs de certaines mutations délétères sur des segments génomiques différents, lorsqu'infectant une même cellule, parviennent à recréer un virus fonctionnel.

Pour bien comprendre l'ambivalence de la complémentation, voici une situation concrète :



*Figure 6 : Principe de la complémentation virale*

Deux virions de la même espèce infectent une même cellule (ici à 3 segments). L'un de ces virions présente dans son génome un segment avec une mutation bénéfique et un autre segment avec une mutation délétère (par rapport au type sauvage), l'autre virion ne présente aucune mutation délétère par rapport au type sauvage sur aucun de ses segments. Lors de la co-infection, les segments d'ARN de ces 2 virions (bien plus nombreux dans la réalité) vont se trouver mélangés dans la cellule. A partir de ce pool commun de segments d'ARN, des virions néo-formés qui présentent une mutation bénéfique (ici sur le segment 1) et aucune mutation délétère (ici sur le segment 2) pourront émerger (=virions avantagés sélectivement) ; à l'inverse, des virions néo-formés qui présentent une mutation délétère (ici sur le segment 2) et aucune mutation bénéfique particulière sur un autre segment émergeront également (désavantagés sélectivement). Ces 2 aspects font partie d'un même phénomène génétique : la complémentation virale.

Ainsi, la complémentation peut permettre de créer des virus avec un génome ne présentant plus aucune mutation délétère, uniquement des mutations bénéfiques, ce qui constitue un indéniable avantage. Mais elle génère aussi des virions ne présentant que des mutations délétères, qui seront donc totalement inutiles et même néfastes lors d'une co-infection future. Il convient donc d'analyser quel est l'impact global de la complémentation sur une population de virus à ARN, à la lumière de l'effet des mutations chez ces virus et du taux auxquels elles se produisent notamment.

## 2) État des lieux de l'effet d'une mutation sur la fitness chez les virus à ARN

En préambule, il convient de comprendre qu'analyser la relation entre la complémentation et le taux de mutation d'un virus à ARN, c'est analyser l'effet de cette interaction sur la fitness. Il faut donc séparer tout d'abord l'effet du taux de mutation sur la fitness (2)), puis l'effet de la complémentation sur la fitness (3)) pour étudier enfin l'effet de l'interaction mutation/complémentation sur une population de virus à ARN (4)).

Dans un premier temps, étudions donc l'effet d'une mutation aléatoire (dont la fréquence augmente lorsque le taux de mutation augmente par définition) sur la fitness d'une population de virus à ARN.

Pour ce faire, nous allons nous pencher sur l'étude de Sanjuán de 2010. Il s'agit d'une analyse de travaux expérimentaux sur le Virus de la Stomatite Vésiculaire (VSV ; Sanjuán et al. 2004a), sur le virus de la mosaïque du tabac (TEV ; Carrasco et al. 2007) et sur le virus de la fièvre hémorragique de Crimée-Congo (CCHFV ; Deyde et al. 2006). Parmi ces virus, seul le CCHFV présente un génome segmenté (en trois fragments). Ces auteurs ont cherché à calculer  $E(s)$ , l'effet moyen sur la fitness d'une unique substitution nucléotidique aléatoire chez un virus à ARN, sur leurs virus respectifs.

- Dans ce but, ils ont généré aléatoirement (*loci* calculés par un algorithme type random) des mutations réparties en de multiples endroits (plus de cinquante *loci* dans ces trois études) du génome du virus d'intérêt (pour le CCHFV, dont le génome est segmenté, des mutations ont été générées sur tous les segments). Ils ont ensuite mis en compétition une population de virus mutés (mutant 1, mutant 2, mutant 3, ... mutant k) avec une population de virus de type sauvage. La taille des populations était identique au départ (c'est-à-dire  $\text{Fréquence}_{t_0}(\text{mutant } k) / \text{Fréquence}_{t_0}(\text{Type Sauvage}) = 1$ ), et elles ont été mise en contact toutes deux avec une population de cellules au sein du même milieu de culture. Les expérimentateurs stoppent ensuite la répllication virale à l'issue d'une durée  $t_f$ , qu'ils choisissent brève de sorte que l'effet d'apparition de nouvelles mutations au cours des cycles de répllication viraux, inévitable, n'influe que peu sur les proportions finales des deux variants viraux. En dénombrant les quantités des deux variants dans le milieu à la fin de  $t_f$ , ils obtiennent finalement la fitness du mutant k considéré :  $s_k = \text{Fréquence}_{t_f}(\text{mutant } k) / \text{Fréquence}_{t_f}(\text{Type Sauvage})$ . Ils ont ensuite regroupé les résultats obtenus pour tous les différents mutants d'une espèce virale et ont obtenu l'effet moyen d'une mutation sur la fitness chez cette espèce  $E(s)$ . Il apparaît que cet effet moyen est globalement bien conservé chez les trois virus étudiés ici, avec  $E(s_{\text{VSV}}) = 0,1$ ,  $E(s_{\text{TEV}}) = 0,13$  et  $E(s_{\text{CCHFV}}) = 0,2$ .
- De plus, ces trois auteurs ont également classé les k mutations étudiées en fonction de la variation de fitness qu'elles occasionnaient. Ils ont ainsi obtenu un certain nombre de mutations de chaque catégorie pour le total des cinquante ou plus mutations étudiées, et en divisant par le nombre total de mutations étudiées, la probabilité p qu'une mutation aléatoire tombe dans telle ou telle catégorie. Leurs résultats sont résumés dans le tableau 3 :

Catégorie de mutation	Intervalles de fitness correspondants décidés par les auteurs	Probabilité pour une mutation aléatoire d'appartenir à la catégorie		
Létale	$s < 0,01$	$p_L \text{ CCHFV} = 0,2$	$p_L \text{ VSV} = 0,18$	$p_L \text{ TEV} = 0,19$
Fortement délétère	$0,01 \leq s \leq 0,3$	$p_D \text{ CCHFV} = 0,07$	$p_D \text{ VSV} = 0,09$	$p_D \text{ TEV} = 0,11$
Moyennement délétère	$0,3 < s \leq 0,7$	$p_{md} \text{ CCHFV} = 0,27$	$p_{md} \text{ VSV} = 0,25$	$p_{md} \text{ TEV} = 0,28$
Faiblement délétère	$0,7 < s \leq 0,95$	$p_{fd} \text{ CCHFV} = 0,26$	$p_{fd} \text{ VSV} = 0,29$	$p_{fd} \text{ TEV} = 0,23$
Neutre	$0,95 < s \leq 1,05$	$p_n \text{ CCHFV} = 0,15$	$p_n \text{ VSV} = 0,12$	$p_n \text{ TEV} = 0,14$
Bénéfique	$s > 1,05$	$p_b \text{ CCHFV} = 0,05$	$p_b \text{ VSV} = 0,07$	$p_b \text{ TEV} = 0,04$

*Tableau 3 : Résultats sur la distribution de la fitness conféré par les mutations chez 3 virus à ARN (d'après Sanjuán et al. 2004a, Carrasco et al. 2007 et Deyde et al. 2006)*

On observe, en additionnant les probabilités des catégories autres que neutre et bénéfique, que la probabilité qu'une mutation aléatoire soit délétère chez un virus à ARN est comprise entre 78% et 80% ; à titre de comparaison, chez les eucaryotes, on estime qu'il y a 70% de probabilité qu'une mutation soit délétère (Eyre-Walker et al. 2002). Ainsi, les mutations sont beaucoup plus souvent délétères que bénéfiques chez les virus à ARN.

- Toujours selon le tableau 3, les mutations à effet modéré sur la fitness (que ce soit en bien ou en mal) sont plus fréquentes que celles avec un effet important : on peut voir que  $p_{fd} + p_n > p_L + p_D + p_b$  pour les trois espèces étudiées. Une hypothèse explicative de Sanjuán de 2010 est que si un effet fort des mutations sur la valeur évolutive du virus (Mutational Fitness Effect) présente l'avantage de favoriser une plus grande variabilité phénotypique, et une élimination plus rapide des mutations délétères, cela empêche l'accumulation de mutants neutres potentiellement bénéfiques par la suite.

Au bilan, la donnée la plus importante à retenir est que les mutations ont un effet moyen sur la fitness majoritairement délétère chez les virus à ARN ( $E(s)$  compris entre 0,1 et 0,2). Parmi les rares moyens via lesquels les virus peuvent se débarrasser de ces fréquentes mutations délétères, la complémentarité entre particules virales apparaît comme le plus évident. Il convient donc d'analyser l'effet de la complémentarité sur la fitness d'une population de virus à ARN.



### 3) Bénéfices et inconvénients de la complémentation pour une population virale

#### a) Réflexions théoriques sur les bénéfices et inconvénients de la complémentation

Dans un deuxième temps, nous allons nous pencher sur l'effet de la complémentation sur la fitness d'une population de virus à ARN.

Dans la revue de Hutchinson et al. de 2010, les auteurs ont fait la synthèse de nombreux articles sur les virus influenza A, visant au départ à étudier l'empaquetage des segments génomiques dans les nouvelles particules virales. Deux avantages à créer un virion à partir d'un pool de segments d'ARN sont mis en exergue.

D'une part, la capacité à reconstituer un génome sans erreur à partir d'une sélection de segments mal copiés et correctement copiés, grâce à des « checkpoints » lors de l'assemblage, et à un mécanisme de sexualité (=échange de matière génétique) au sens large du terme. Cela permet d'échapper au "Muller's ratchet" (l'accumulation irréversible de mutations délétères dans un génome issu d'une reproduction asexuée). Comme le disent bien les auteurs, cet aspect est particulièrement important pour les virus à ARN, qui ont un taux de mutation très important, et une forte probabilité qu'une mutation aléatoire soit délétère (cf (Sanjuán 2010), partie III, 2)). Et la reconstruction d'un génome entier fonctionnel, après une mutation délétère acquise de quelque manière que ce soit (altération du génome par un agent externe (UV, ...), erreur de réplication, ...), est permise par un phénomène majoritairement : la complémentation (Sugiura, Tobita, Kilbourne 1972).

D'autre part, cette complémentation permet au virus d'acquérir des allèles bénéfiques selon les auteurs. Dans le cas du virus Influenza A, des assemblages génomiques de virus humains et animaux ont été décrits, ce qui génère de nouvelles souches génétiques au potentiel pandémique très important. Et innover rapidement est crucial dans la course de vitesse qu'est l'évolution ; comme le disait métaphoriquement le scientifique et auteur Lewis Carroll : « Here [confronté à l'évolution], you see, it takes all the running you can do to stay in the same place. » (Lewis Carroll, « Alice Through the Looking-glass », 1871).

Cependant, la segmentation génomique nécessite la transmission de chacun des segments au nouveau virion pour qu'il soit infectieux. Cela ne pose pas de problèmes lors d'infections à haute multiplicité (transmission par des vecteurs biologiques (arthropodes...) ou contamination verticale) : les auteurs en veulent pour preuve l'existence de virus avec une stratégie de génomes multipartites (particulièrement pour les virus de plantes), qui ne peuvent exister qu'en cas d'échange génétique. Mais en cas d'infection à basse multiplicité (transmission par aérosol, contact de l'eau), toutes les particules virales n'incorporent pas forcément tous les segments génomiques fonctionnels, et seront donc non infectieuses.

Ainsi, le phénomène de complémentation souffre d'une première limite : il ne peut se mettre en place que lors d'une infection à haute multiplicité. Pour donner un ordre de grandeur, selon l'article de von Magnus de 1954, cette multiplicité d'infection doit être d'au moins  $10^2$  particules virales dans la population initiale. Cette étude a été réalisée en utilisant des virus influenza A comme modèle, mis au contact d'une colonie de cellules embryonnaires de souris : le virus

avait donc directement accès à une population de cellules sensibles (ce qui n'est pas toujours le cas, par exemple lors de franchissement de la barrière d'espèce).

Dans la suite de cette thèse, la complémentation sera régulièrement assimilée à la multiplicité d'infection, car elle est directement proportionnelle à celle-ci (Marshall et al. 2013).

*b) Interaction entre la complémentation et la sélection naturelle : une fitness populationnelle diminuée à pression de sélection faible, mais augmentée à pression de sélection forte*

Évaluons maintenant les effets potentiellement bénéfiques ou délétères de la complémentation via l'expérimentation. Dans l'étude de Gao et Feldman de 2009, les chercheurs ont construit un modèle mathématique, suivi d'une simulation informatique, pour explorer les mécanismes reliant la complémentation et d'autres facteurs évolutifs : mutation, recombinaison et épistase. L'épistase est définie comme l'interaction entre les mutations qui se produisent.

Dans leur modèle mathématique, ces chercheurs sont partis d'une population clonale de virus à ARN avec trois segments. Ils ont mis en équation le nombre de virus issus de chaque génération avec zéro, un, deux ou trois segments mutés en fonction des quatre forces évolutives mentionnées ci-dessus (mutation, recombinaison, épistase et complémentation). Donc du point de vue de leur modèle, quatre segments mutés (impossible, vu que le virus théorique qu'ils considèrent n'a que trois segments) équivalent à une perte de fonctionnalité totale pour le virus dont c'est le génome.

Pour prendre en compte l'effet de l'épistase sur la fitness, ils ont utilisé des coefficients d'épistase  $\epsilon$  (passage d'un à deux segments mutés) et  $\eta$  (passage de deux à trois segments mutés) dans leurs formules. Pour prendre en compte celui de la mutation, ils ont utilisé une valeur moyenne de l'effet de la mutation sur la fitness d'un virus à ARN, notée  $S$ , variant entre  $S=0,1$  et  $S=0,2$  suivant leurs simulations (chiffres tirés de l'étude de Sanjuán de 2010, cf partie III, 2), avec  $S$  équivalent à  $E(s)$ .

Nombre de mutations du virus par rapport au type sauvage	0	1	2	3	4 ou +
Fitness	1	$e^{-S}$	$e^{-2S(1+\epsilon)}$	$e^{-3S(1+\epsilon+\eta)}$	0

*Tableau 4 : Parties des équations permettant de prendre en compte l'effet de l'épistase et des mutations (Gao et Feldman, 2009)*

Les équations permettant de prendre en compte l'effet de la complémentation et de la recombinaison ne seront pas détaillées ici, en raison de leur complexité. On peut juste en retenir que la complémentation est traduite par la multiplicité d'infection (cf partie III, 3), a)).

A l'issue de manipulations, ils obtiennent une formule calculant la fréquence de virus avec zéro, un, deux ou trois segments d'ARN mutés après chaque événement de copie en fonction des

fréquences du cycle précédent et des différentes forces évolutives susmentionnées. Nous allons nous intéresser plus particulièrement à leur comparaison de populations virales au sein desquelles la complémentation et la valeur  $S$  varient, résumée sur la figure 7. Le coefficient  $S$  correspond à l'effet moyen de la mutation sur la fitness d'un virus, autrement dit à l'effet de la sélection naturelle sur la population virale. Plus  $S$  est faible, plus une mutation aléatoire s'avèrera délétère, plus l'effet de la sélection naturelle est fort sur la population virale.



*Figure 7 : Interaction entre la complémentation et la sélection, avec  $\mu = 0,01$ ,  $\varepsilon = 0,0$ ,  $\eta = 0,0$  et  $S$  variant comme indiqué sur la légende (selon les données de (Gao, Feldman 2009))*

La fitness moyenne de populations virales théoriques en fonction de la fraction de co-infection virale (proportionnelle à la complémentation comme expliqué plus haut) a été évaluée, pour une sélection naturelle faible ( $S=0,1$ ), modérée ( $S=0,15$ ) et forte ( $S=0,2$ ). L'interaction entre les mutations est considérée nulle ( $\varepsilon = 0$ ,  $\eta = 0$  dans le modèle mathématique).

Sur le graphique 7, on observe que plus la sélection naturelle a tendance à éliminer d'individus viraux de la population (c'est-à-dire plus  $S$  est faible), plus ce sont les populations virales où la complémentation est prévalente (c'est-à-dire où les infections multiples sont fréquentes) qui ont tendance à avoir la fitness la plus élevée.

On dénote également qu'à court terme la complémentation a un effet légèrement négatif sur la fitness globale de la population : légère décroissance de la fitness populationnelle sur la figure 7 pour  $S=0,2$ . Par contre, elle a un effet bénéfique sur le long terme, où les conditions environnementales seront sans doute amenées à changer, c'est-à-dire pour le traduire sur la figure 7 que le coefficient  $S$  aura tendance à diminuer : légère croissance de la fitness populationnelle sur la figure 7 pour  $S=0,15$ , et *a fortiori* pour  $S=0,1$ .

La conclusion des auteurs est qu'une forte complémentation diminue légèrement la fitness d'une population virale, mais accroît significativement sa robustesse vis-à-vis de la sélection naturelle. En effet, les différences de fitness observées sur la figure 7 peuvent paraître minimes (fitness toujours comprise entre 0,9 et 1 dans n'importe lequel des cas de figure étudiés), mais sur le long terme, ces différences peuvent s'accumuler et produire un effet biologique important. Parmi les mécanismes qui peuvent expliquer cet effet sur une période de temps longue de la complémentation, il est possible qu'elle entretienne un haut niveau de diversité génétique dans la population virale.

c) Une explication possible à l'augmentation de la robustesse vis-à-vis de la sélection par la complémentation : favoriser une grande diversité génétique dans la population

Pour vérifier si la complémentation est effectivement favorable à une grande diversité génotypique d'une population virale, nous allons nous pencher sur l'étude de (Frank, 2001). Celle-ci s'intéresse à la relation entre divergence génétique et divers facteurs, qui peuvent être reliés à la complémentation.

Les chercheurs prennent comme modèle un virus à deux segments, A et B, et sur chaque segment est considéré un gène à deux allèles, 1 et 2 (le gène situé sur le segment B et présentant l'allèle 1 sera donc noté B<sub>1</sub>). Avec ce modèle, ils dégagent une formule quantifiant la pénalité causée par les hybridations génétiques, qui dépend des corrélations respectives entre la fréquence de l'allèle A<sub>x</sub> et celle de l'allèle B<sub>x</sub> et de la taille effective de la population, N<sub>e</sub>.

Les corrélations entre fréquence d'allèles sont quantifiées ici par quatre coefficients, compris entre -1 et 1 : Corr(F<sub>A1</sub>,F<sub>B1</sub>), Corr(F<sub>A1</sub>,F<sub>B2</sub>), Corr(F<sub>A2</sub>,F<sub>B2</sub>), et Corr(F<sub>A2</sub>,F<sub>B1</sub>). A titre d'exemple, si on considère que les gènes A<sub>1</sub> et B<sub>1</sub> ont un effet fortement délétère sur la fitness du virus lorsque présents ensemble dans un même génome, on aura Corr(F<sub>A1</sub>,F<sub>B1</sub>) proche de -1.

N<sub>e</sub>, elle, est reliée à la multiplicité d'infection par la formule  $N_e = C \times m_h$  (C=nombre de cellules hôtes ; m<sub>h</sub>= moyenne de la multiplicité d'infection sur la population). Or la multiplicité d'infection (nombre de virus infectant la même cellule en moyenne) traduit directement la complémentation. La complémentation est donc proportionnelle à la taille effective de la population virale.

A partir de ce modèle, les auteurs ont ensuite utilisé une simulation informatique pour quantifier la facilité avec laquelle la population diverge d'une situation initiale en considérant les deux génotypes  $A_1B_1$  et  $A_2B_2$  comme favorables (et les génotypes  $A_1B_2$  et  $A_2B_1$  comme défavorables) et en partant d'une population composée à 70% de  $A_1B_1$  et 30% de  $A_2B_2$ . Ils ont calculé le temps que mettaient la population à inverser ses proportions, c'est-à-dire à passer à 70% de  $A_2B_2$  et 30% de  $A_1B_1$ . Cela requérait de traverser des « vallées » où la fitness est mauvaise ( $\text{Corr}(F_{A_1}, F_{B_1})$  et  $\text{Corr}(F_{A_2}, F_{B_2}) > 0$  ;  $\text{Corr}(F_{A_1}, F_{B_2})$  et  $\text{Corr}(F_{A_2}, F_{B_1}) < 0$ ). Ils ont finalement constaté que, pour  $m_h \geq 2$  (c'est-à-dire à partir du moment où au moins deux virus infectent la même cellule), le temps d'attente entre l'inversion de fréquence du génotype  $A_1B_1$  au génotype  $A_2B_2$  diminue beaucoup lorsque la multiplicité d'infection augmente.

Le modèle suggère donc qu'une multiplicité d'infection accrue, c'est-à-dire une complémentarité accrue augmente la facilité d'une population à diverger de fréquences génotypiques initiales. Ainsi, la diversification génétique de la population virale s'en trouve bien favorisée. Les auteurs suggèrent que ce serait en augmentant les perturbations stochastiques (synonymes de la dérive génétique) dans la population que la complémentarité favoriserait le changement des fréquences génotypiques initiales au cours des générations.

Mais il ne s'agit que d'une conjecture. Maintenant que nous avons établi les effets de la complémentarité sur la fitness de la population virale (notamment l'effet bénéfique à long terme : favorise la robustesse vis-à-vis de la sélection naturelle en entretenant une haute diversité génétique), nous allons nous pencher sur les facteurs qui modulent ces effets de la complémentarité, en particulier le taux de mutation  $\mu$ .

#### 4) Conditionnalité des bénéfices/inconvénients de la complémentation à un paramètre : le taux de mutation ( $\mu$ )

Dans un troisième temps, nous allons enfin étudier l'effet de l'interaction entre taux de mutation et complémentation sur la fitness d'une population de virus à ARN.

L'étude de Segredo-Otero et Sanjuán de 2019 considère un virus théorique à deux segments d'ARN. A l'aide d'un modèle mathématique, ils analysent deux facteurs en relation avec la complémentation : l'appartenance à une unité d'infection collective, et le taux de mutation  $\mu$ . D'une part, les auteurs montrent que la complémentation au sein d'une unité d'infection collective (un ensemble de virus associés capable d'infecter une même cellule, comme par exemple des virus adsorbés à la surface d'une bactérie) a un effet délétère sur la fitness globale de la population sur le court terme, à sélection naturelle constante : c'est-à-dire avec un effet des mutations sur la fitness constant (S ou E(s) pour reprendre les notations précédentes).

En effet, les multiplicités d'infection élevées favorisent l'évolution des particules interférentes défectueuses (DIP) et d'autres types de virus défectueux qui se répliquent aux dépens des virus "auxiliaires" entièrement fonctionnels ((Marriott, Dimmock 2010), (Rezelj, Levi, Vignuzzi 2018)). Étant donné que les DIP ont des séquences plus courtes et bénéficient de virus auxiliaires sans réciproque, un phénomène d'erreur-catastrophe se produit : il y a invasion de la population virale par des virus incompetents qui sont dépendants des autres (« tricheurs sociaux »), et cela réduit considérablement l'aptitude (fitness) de la population virale (Turner et Chao, 1999) (Díaz-Munoz et al., 2017). Par conséquent, les avantages de la complémentation, s'ils existent, pourraient être compensés par les coûts issus de l'émergence des DIP (qu'elle promeut).

Les auteurs nuancent néanmoins leurs propos en admettant que les effets délétères de la complémentation sont peu marqués lorsque les virus membres de la même unité d'infection sont issus d'une même cellule.

D'autre part, ils mettent en lumière deux aspects très importants, sur la relation entre la complémentation et le taux de mutation chez un virus à ARN. Ceux-ci sont visibles sur la figure 8 (Segredo-Otero, Sanjuán 2019).



*Figure 8 : Interaction entre la complémentation et le taux de mutation, avec  $S = 0,2$ ,  $\varepsilon = 0,0$ ,  $\eta = 0,0$  et  $\mu$  variant comme indiqué sur la légende (selon les données de (Segredo-Otero, San-juán 2019))*

La fitness moyenne de populations virales théoriques en fonction de la fraction de co-infection virale (proportionnelle à la complémentation comme expliqué plus haut) a été évaluée, pour un taux de mutation par événement de copie nul ( $\mu=0$ ), élevé ( $\mu=0,005$ ) et très élevé ( $\mu=0,01$ ). L'interaction entre les mutations est considérée nulle ( $\varepsilon = 0$ ,  $\eta = 0$  dans le modèle mathématique).

Premièrement, lorsque le taux de mutation  $\mu$  de ce virus au génome bisegmenté est très faible, la complémentation n'a pas d'effet sur la fitness de la population : visible sur la figure 8 pour  $\mu=0$ . C'est logique, car si les virus ont tous un génome quasiment identique, une redistribution des segments d'ARN entre virus n'aura que peu d'influence sur le génotype des futures particules virales.

Deuxièmement, pour des taux de mutations élevés, plus il y a de complémentation au sein de la population virale, plus la fitness globale de la population augmente.

Au bilan, selon ce modèle mathématique, plus le taux de mutation du virus est élevé, plus le phénomène de complémentation s'avère bénéfique pour lui. On peut néanmoins nuancer en remarquant que les taux de mutation de 0,005 et 0,01 s/n/c présentés sur ce graphique sont beaucoup plus élevés que les taux de mutation *in vivo* des virus à ARN (cf partie II, 1) pour exemples).

Il convient de valider ce résultat théorique par une expérience sur une population de virus à ARN présentant un  $\mu$  élevé. Parmi eux, le virus Influenza A est un modèle de choix, avec son taux de  $\mu_{\text{InfAV}} = 2,1(\pm 0,5) \times 10^{-5}$  s/n/copie (selon Pauly, Procario, Lauring 2017, cf partie 2).

L'étude de Brooke et al. de 2013 s'intéresse à l'éventuel intérêt évolutif de la complémentation sur cette espèce-modèle. Comme vu auparavant (partie III, 2), dernier paragraphe), la segmentation du génome des virus influenza A permet des réassortiments génétiques rapides, mais complique également le fait de pouvoir assembler un génome viral en entier. Les auteurs ont recherché l'expression de quatre des douze protéines du virus (HA, NA, NP et NS1) à l'issue d'infections de cellules MDCK (cellules de reins de chien) à différentes MOI. Résultat : seulement une ou deux protéines exprimées sur les trois essentielles au virus (HA, NA et NP) pour une MOI de 4, contre 100% d'expression des trois protéines pour une MOI de  $5 \times 10^2$ .

Ces protéines virales étaient recherchées via une méthode d'immunofluorescence, avec des anticorps monoclonaux associés à des fluorochromes. Leurs expériences montrent de façon plus large que plus la MOI augmente, plus la proportion de virions qui expriment toutes leurs protéines augmente aussi (preuve du bénéfice de la complémentation), quel que soit le type de cellule utilisé (ils ont également expérimenté avec des œufs, des cellules de souris et des cellules de cochon d'Inde) et la souche d'influenza A utilisée. Ainsi, très peu de virions sont infectieux à faible MOI, c'est-à-dire sans complémentation.

Ces chercheurs concluent à l'issue de cette étude que les virus influenza A ne sont en fait qu'un essaim de virions semi-infectieux pour lesquels la complémentation est non seulement bénéfique, mais indispensable à leur survie dans la nature. Ainsi, évaluer l'infectivité d'un VIA à une faible MOI revient à sous-estimer grandement le potentiel infectieux de la population virale dans son ensemble.



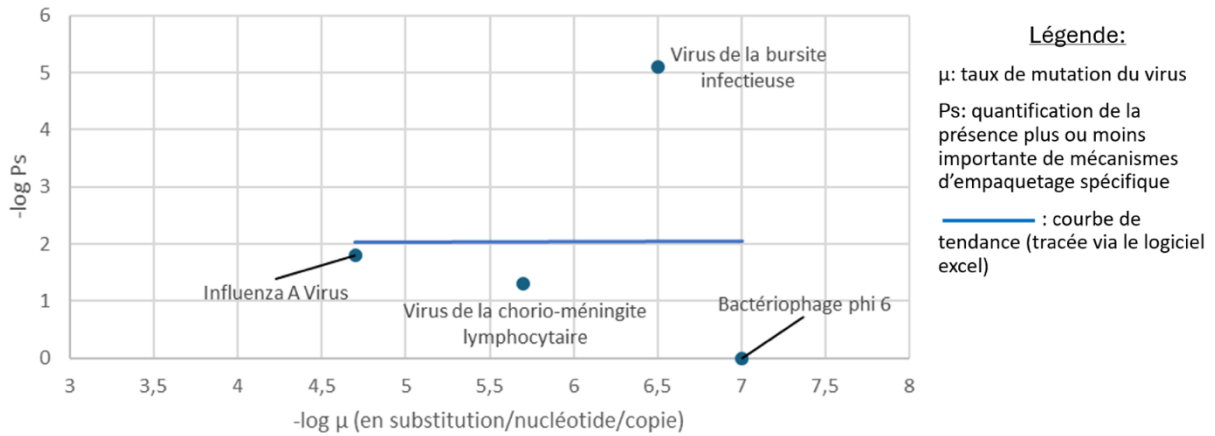
## 5) Évaluation d'un biais dans l'analyse des effets de la complémentation

Il convient cependant de vérifier qu'il n'existe pas d'autres facteurs que la complémentation qui sont proportionnels à la multiplicité d'infection (auquel cas ce ne serait pas forcément celle-ci qui serait responsable, à  $\mu$  élevé, de l'augmentation de la fitness de la population virale). Segredo-Otero et Sanjuán, dans tous les articles qu'ils ont rassemblés pour leur synthèse de 2019, n'en identifie qu'un seul qui remplirait cette condition : le Ps, décrit par Hutchinson et al. dans leur article de 2010.

Cette valeur Ps (« packaging-signals ») est en fait une quantification de la présence plus ou moins importante de mécanismes d'empaquetage spécifique : plus Ps est proche de 1, plus cela signifie qu'il y a des mécanismes permettant d'empaqueter à chaque fois un exemplaire de chaque segment du génome de ce virus (elle est comprise entre 0 et 1). A titre d'exemple, les auteurs ont calculé une valeur Ps de 1 pour le bactériophage phi 6, de  $10^{-5}$  pour le Gumboro Virus, de  $1,6 \times 10^{-2}$  pour le virus influenza A et de  $5 \times 10^{-2}$  pour le virus de la chorioméningite lymphocytaire. Le génome des quatre virus évoqués est séparé en respectivement 6, 2, 8 et 2 segments d'ARN.

Cela signifie que le GumbV (virus de la bursite infectieuse) est plus proche du modèle aléatoire, où un mécanisme existe pour distinguer les segments d'ARN du génome viral des autres segments d'ARN présents dans la cellule avant de les incorporer au virion, mais c'est tout. Dans ce cas, plus le virus empaquète de segments, plus la probabilité qu'il y ait au moins un exemplaire de chacun des deux segments du génome viral est grande. Le bactériophage phi 6, lui, est très proche du modèle spécifique : des mécanismes d'assemblage spécifiques existent et garantissent qu'un unique exemplaire de chacun des segments est empaqueté avant le bourgeonnement du virus.

Pour savoir si c'est bien la présence plus ou moins importante de signaux d'empaquetage spécifiques qui est responsable des effets attribués précédemment à la complémentation (c'est-à-dire plus  $\mu$  est élevé, plus ce facteur présente un effet bénéfique important sur la fitness de la population virale), les auteurs ont mis en relation les taux de mutation  $\mu$  des quatre virus étudiés (extraits de Sevilla, de la Torre 2006 ; Pauly, Procario, Lauring 2017 ; Combe, Sanjuán 2014) avec leur valeur Ps déterminée. Ces taux ne sont pas toujours déterminés suivant la même méthode, mais tous les biais inhérents à ces différentes méthodes ont été pris en compte dans ces études, et les écarts conséquents d'ordre de grandeur entre les valeurs (comme on peut le voir sur la figure 9 ci-dessous) permettent quand même de tirer des conclusions.



**Figure 9 : Relation entre la prévalence des signaux d’empaquetage ( $P_s$ ) et le taux de mutation ( $\mu$ ) chez les virus à ARN (selon les données de Hutchinson et al., 2010)**

La prévalence des signaux d’empaquetage a été calculée par les auteurs en analysant plusieurs articles quantifiant la proportion du nombre total de segments d’ARN du virus considéré qui étaient présents par particule virale au bout d’une certaine durée de culture. Sur ce graphique,  $-\log(P_s)$  est inversement proportionnel à cette prévalence de signaux d’empaquetage. Les valeurs de taux de mutation ( $\mu$ , en s/n/c) retenues par les auteurs correspondent à des valeurs issues de plusieurs articles, et obtenues via des tests de fluctuation de Luria-Delbrück. Sur ce graphique,  $-\log(\mu)$  est inversement proportionnel au taux de mutation de ces virus. Aucune corrélation significative entre ces deux paramètres n’est mise en évidence par ce graphique (la courbe de régression tracée informatiquement est plate).

En observant la figure 9, on se rend compte que la  $P_s$  est indépendante du taux de mutation chez les virus à ARN considérés (courbe de tendance presque plate). Donc son effet ne varie pas suivant les variations du taux de mutation, contrairement à celui de la complémentarité.

Ainsi donc, le  $P_s$ , bien que proportionnel à la multiplicité d’infection, n’explique pas les effets susmentionnés sur la fitness. Ceux-ci pourraient donc être attribuables à la complémentarité.

## 6) Comment font les virus à ARN qui ne peuvent pas bénéficier de la complémentation génomique ?

Pour finir, afin d'être sûr de bien attribuer les effets bénéfiques de long terme décrits plus haut à la complémentation chez les virus à ARN et uniquement à ce trait, nous allons nous pencher sur un modèle de virus à ARN non segmenté (donc chez lequel la complémentation n'est pas possible) : le VSV (Vesicular Stomatitis Virus).

Dans l'étude de Wilke, Reissig, et Novella de 2004, les auteurs ont étudié une population de VSV présentant deux génotypes : une proportion  $P_{1i}$  de la population est de type sauvage (WT), et une proportion  $P_{2i}$  est de type MARM (virus résistants à un anticorps monoclonal). Ils ont ensuite exposé des inocula de cette population à une population de cellules BHK-21, dans un milieu de culture tout à fait classique (ne contenant pas d'anticorps monoclonaux auxquels les virus MARM sont résistants). Ils ont fait varier les proportions P1 et P2 initiales entre leurs différentes séries d'expériences (avec toujours  $P_{2i} > P_{1i}$ ), ainsi que la quantité initiale de virus dans l'inoculum (c'est-à-dire la MOI).

Ils ont constaté que les proportions finales  $P_{1f}$  et  $P_{2f}$ , à l'issue de cinq cycles infectieux, étaient globalement toujours les mêmes quelles que soient les proportions initiales, et très en faveur du virus WT ( $P_{1f} \sim 80\%$  pour une MOI de 10 particules,  $P_{1f} \sim 70\%$  pour une MOI de 1000 particules). Ils ont également calculé la fitness relative du variant MARM par rapport au WT : cette fitness était globalement toujours inférieure à celle du type sauvage, mais la différence était moindre au fur et à mesure que la MOI augmentait (un mécanisme explicatif sera évoqué plus bas).

Ainsi, l'absence de segmentation, et donc de complémentation, chez un virus à ARN permet d'atteindre rapidement un optimum vis-à-vis de la sélection naturelle. Cependant, il est très compliqué de changer ensuite les fréquences génotypiques de cette population : sa rapidité d'évolution s'en trouve donc considérablement affaiblie. Cela peut s'avérer très préjudiciable, surtout dans un cas comme celui-ci, où un allèle qui se trouve être faiblement délétère dans la situation actuelle s'avère tout à fait bénéfique lorsque l'environnement change (apparition d'une réponse immunitaire dans ce cas précis).

Cependant, on constate quand même une moins grande différence de proportions finales entre les deux variants WT et MARM du VSV à haute MOI qu'à basse MOI. Pour l'expliquer, nous allons nous pencher sur une autre étude. Celle réalisée en 2004 par Novella, Ball et Wertz exploite une particularité génétique du VSV : l'expression de ses gènes dépend de la position du gène considéré sur l'unique chromosome viral, en l'occurrence plus le gène est loin de l'extrémité 3' du génome, moins il est transcrit. Dans le virus sauvage, c'est donc les protéines nécessaires en plus grande quantité dont les gènes correspondants sont les plus proches de l'extrémité 3'. L'ordre de disposition des virus sauvages (WT) est 3'-NPMGL-5' (avec N la protéine de capsid, P la phosphoprotéine, M la protéine de matrice, G la petite sous-unité de l'ARN polymérase ARN-dépendante du VSV et L la grande sous-unité de l'ARN polymérase).

Les auteurs ont mis en compétition des variants viraux avec les positions des gènes N et L conservées par rapport au WT, mais avec d'autres dispositions des trois autres gènes au milieu : GMP, GPM, MPG

## Conclusion

Pour conclure, les taux de mutation particulièrement élevés des différents virus à ARN peuvent être mesurés de plusieurs manières différentes, qui ont chacune leurs avantages et leurs limites. Parmi elles, le test de fluctuation de Luria-Delbrück, lorsqu'on respecte ses conditions de validité, s'impose comme la méthode la plus fiable, permettant de prendre en compte au mieux les biais de la sélection naturelle, et d'éviter la lourdeur protocolaire nécessaire pour corriger les erreurs inhérentes au séquençage.

Grâce à cette méthode, les scientifiques ont été en mesure de mettre à jour d'importantes différences de taux de mutation entre les différentes espèces de virus à ARN : ces valeurs s'étendent de  $10^{-5}$  à  $10^{-7}$  substitution/nucléotide/événement de copie. Et le facteur le plus probable pour expliquer ces importantes différences de taux de mutation est la segmentation ou non du génome de l'espèce virale considérée.

Le caractère segmenté a tendance à favoriser un taux de mutation plus important chez les virus à ARN. Ce taux de mutation plus élevé pourrait être lié au phénomène de complémentation génétique. Celle-ci présente *a priori* de nombreux inconvénients chez ces virus : diminution de la fitness globale de la population, invasion de celle-ci par des virus « tricheurs », ... Cependant, ces effets négatifs perdent beaucoup de leur ampleur plus le taux de mutation du virus considéré est important. De nombreux effets bénéfiques sont alors visibles : favorisation de la divergence génotypique à partir d'une population initiale (augmentation de sa rapidité d'évolution en quelque sorte), accroissement important de sa diversité et de sa robustesse vis-à-vis de la sélection naturelle, purge plus rapide des mutations délétères sur le moyen terme.

Au final, pour des virus à ARN à taux de mutation élevé, comme le virus influenza, la complémentation s'avère même indispensable, avec une infectivité dérisoire d'une particule virale considérée individuellement devant celle de la population infectante dans son ensemble.

Cependant, un autre phénomène que la complémentation pourrait expliquer le fait qu'une population de virus à ARN segmentés présentant des mutations délétères sur certains de leurs segments génomiques demeure apte sélectivement parlant : certains de ces virus sont capables de polypléidie. Cela a été mis en évidence notamment sur le virus de la rougeole (Rager et al. 2002) : ces virus présentent donc en leur sein plusieurs copies de génome complète, et seraient donc capables de produire en toutes circonstances des lots de protéines différents suivant les copies de génome intégrées, ce qui leur conférerait une grande capacité d'adaptation à leur milieu sur le long terme.

Ainsi, pour analyser plus en détail les effets de la complémentation chez les virus à ARN, il pourrait être intéressant de segmenter artificiellement le génome chez une espèce virale dans laquelle il n'est pas segmenté, et de pouvoir ainsi individualiser le plus fidèlement possible les effets de ce phénomène génétique, par comparaison avec une population virale sauvage.





## Bibliographie :

BROOKE, Christopher B., INCE, William L., WRAMMERT, Jens, AHMED, Rafi, WILSON, Patrick C., BENNINK, Jack R. et YEWDELL, Jonathan W., 2013. Most Influenza A Virions Fail To Express at Least One Essential Viral Protein. *Journal of Virology*. 15 mars 2013. Vol. 87, n° 6, pp. 3155-3162. DOI 10.1128/JVI.02284-12.

COMBE, Marine et SANJUÁN, Rafael, 2014. Variation in RNA Virus Mutation Rates across Host Cells. LAURING, Adam (éd.), *PLoS Pathogens*. 23 janvier 2014. Vol. 10, n° 1, pp. e1003855. DOI 10.1371/journal.ppat.1003855.

DEYDE, Varough M., KHRISTOVA, Marina L., ROLLIN, Pierre E., KSIAZEK, Thomas G. et NICHOL, Stuart T., 2006. Crimean-Congo hemorrhagic fever virus genomics and global diversity. *Journal of Virology*. septembre 2006. Vol. 80, n° 17, pp. 8834-8842. DOI 10.1128/JVI.00752-06.

ECKERLE, Lance D., BECKER, Michelle M., HALPIN, Rebecca A., LI, Kelvin, VENTER, Eli, LU, Xiaotao, SCHERBAKOVA, Sana, GRAHAM, Rachel L., BARIC, Ralph S., STOCKWELL, Timothy B., SPIRO, David J. et DENISON, Mark R., 2010. Infidelity of SARS-CoV Nsp14-Exonuclease Mutant Virus Replication Is Revealed by Complete Genome Sequencing. *PLOS Pathogens*. 6 mai 2010. Vol. 6, n° 5, pp. e1000896. DOI 10.1371/journal.ppat.1000896.

ELENA, Santiago F. et SANJUÁN, Rafael, 2005. Adaptive Value of High Mutation Rates of RNA Viruses: Separating Causes from Consequences. *Journal of Virology*. 15 septembre 2005. Vol. 79, n° 18, pp. 11555-11558. DOI 10.1128/jvi.79.18.11555-11558.2005.

EYRE-WALKER, Adam, KEIGHTLEY, Peter D., SMITH, Nick G. C. et GAFFNEY, Daniel, 2002. Quantifying the Slightly Deleterious Mutation Model of Molecular Evolution. *Molecular Biology and Evolution*. 1 décembre 2002. Vol. 19, n° 12, pp. 2142-2149. DOI 10.1093/oxfordjournals.molbev.a004039.

FINCHAM, J. R. S., 1966. *Genetic Complementation*. W.A. Benjamin.

FRANK, Steven A, 2001. Multiplicity of infection and the evolution of hybrid incompatibility in segmented viruses. *Heredity*. novembre 2001. Vol. 87, n° 5, pp. 522-529. DOI 10.1046/j.1365-2540.2001.00911.x.

GAO, Hong et FELDMAN, Marcus W, 2009. Complementation and Epistasis in Viral Coinfection Dynamics. *Genetics*. 1 mai 2009. Vol. 182, n° 1, pp. 251-263. DOI 10.1534/genetics.108.099796.

GLENN, Travis C., 2011. Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*. 2011. Vol. 11, n° 5, pp. 759-769. DOI 10.1111/j.1755-0998.2011.03024.x.

HUTCHINSON, E. C., VON KIRCHBACH, J. C., GOG, J. R. et DIGARD, P., 2010. Genome packaging in influenza A virus. *Journal of General Virology*. 1 février 2010. Vol. 91, n° 2, pp. 313-328. DOI 10.1099/vir.0.017608-0.

LOU, Dianne I., HUSSMANN, Jeffrey A., MCBEE, Ross M., ACEVEDO, Ashley, ANDINO, Raul, PRESS, William H. et SAWYER, Sara L., 2013. High-throughput DNA



- sequencing errors are reduced by orders of magnitude using circle sequencing. *Proceedings of the National Academy of Sciences*. 3 décembre 2013. Vol. 110, n° 49, pp. 19872-19877. DOI 10.1073/pnas.1319590110.
- LURIA, S. E. et DELBRÜCK, M., 1943. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics*. novembre 1943. Vol. 28, n° 6, pp. 491-511. DOI 10.1093/genetics/28.6.491.
- MARRIOTT, A. C. et DIMMOCK, N. J., 2010. Defective interfering viruses and their potential as antiviral agents. *Reviews in Medical Virology*. 2010. Vol. 20, n° 1, pp. 51-62. DOI 10.1002/rmv.641.
- MARSHALL, Nicolle, PRIYAMVADA, Lalita, ENDE, Zachary, STEEL, John et LOWEN, Anice C., 2013. Influenza Virus Reassortment Occurs with High Frequency in the Absence of Segment Mismatch. *PLOS Pathogens*. 13 juin 2013. Vol. 9, n° 6, pp. e1003421. DOI 10.1371/journal.ppat.1003421.
- NAKATSU, Sumiho, SAGARA, Hiroshi, SAKAI-TAGAWA, Yuko, SUGAYA, Norio, NODA, Takeshi et KAWAOKA, Yoshihiro, 2016. Complete and Incomplete Genome Packaging of Influenza A and B Viruses. *mBio*. 6 septembre 2016. Vol. 7, n° 5, pp. 10.1128/mbio.01248-16. DOI 10.1128/mbio.01248-16.
- NOVELLA, Isabel S., BALL, L. Andrew et WERTZ, Gail W., 2004. Fitness Analyses of Vesicular Stomatitis Strains with Rearranged Genomes Reveal Replicative Disadvantages. *Journal of Virology*. 15 septembre 2004. Vol. 78, n° 18, pp. 9837-9841. DOI 10.1128/JVI.78.18.9837-9841.2004.
- OJOSNEGROS, Samuel, GARCÍA-ARRIAZA, Juan, ESCARMÍS, Cristina, MANRUBIA, Susanna C., PERALES, Celia, ARIAS, Armando, MATEU, Mauricio García et DOMINGO, Esteban, 2011. Viral Genome Segmentation Can Result from a Trade-Off between Genetic Content and Particle Stability. *PLoS Genetics*. 17 mars 2011. Vol. 7, n° 3, pp. e1001344. DOI 10.1371/journal.pgen.1001344.
- PAULY, Matthew D, PROCARIO, Megan C et LAURING, Adam S, 2017. A novel twelve class fluctuation test reveals higher than expected mutation rates for influenza A viruses. *eLife*. 9 juin 2017. Vol. 6, pp. e26437. DOI 10.7554/eLife.26437.
- PECK, Kayla M. et LAURING, Adam S., 2018. Complexities of Viral Mutation Rates. SULLIVAN, Christopher S. (éd.), *Journal of Virology*. 15 juillet 2018. Vol. 92, n° 14, pp. e01031-17. DOI 10.1128/JVI.01031-17.
- RAGER, Monika, VONGPUNSAWAD, Sompong, DUPREX, William Paul et CATTANEO, Roberto, 2002. Polyploid measles virus with hexameric genome length. *The EMBO Journal*. 15 mai 2002. Vol. 21, n° 10, pp. 2364-2372. DOI 10.1093/emboj/21.10.2364.
- REZELJ, Veronica V, LEVI, Laura I et VIGNUZZI, Marco, 2018. The defective component of viral populations. *Current Opinion in Virology*. 1 décembre 2018. Vol. 33, pp. 74-80. DOI 10.1016/j.coviro.2018.07.014.
- SANJUÁN, Rafael, 2010. Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*. 27 juin 2010. Vol. 365, n° 1548, pp. 1975-1982. DOI 10.1098/rstb.2010.0063.

- SANJUÁN, Rafael, NEBOT, Miguel R., CHIRICO, Nicola, MANSKY, Louis M. et BELSHAW, Robert, 2010. Viral Mutation Rates. *Journal of Virology*. octobre 2010. Vol. 84, n° 19, pp. 9733-9748. DOI 10.1128/JVI.00694-10.
- SEGREDO-OTERO, Ernesto et SANJUÁN, Rafael, 2019. The effect of genetic complementation on the fitness and diversity of viruses spreading as collective infectious units. *Virus Research*. juillet 2019. Vol. 267, pp. 41-48. DOI 10.1016/j.virusres.2019.05.005.
- SEVILLA, N. et DE LA TORRE, J. C., 2006. Arenavirus Diversity and Evolution: Quasispecies In Vivo. In : DOMINGO, Esteban (éd.), *Quasispecies: Concept and Implications for Virology* [en ligne]. Berlin, Heidelberg : Springer Berlin Heidelberg. pp. 315-335. Current Topics in Microbiology and Immunology. [Consulté le 15 mars 2023]. ISBN 978-3-540-26395-1. Disponible à l'adresse : [http://link.springer.com/10.1007/3-540-26397-7\\_11](http://link.springer.com/10.1007/3-540-26397-7_11)
- SMITH, Donald B. et INGLIS, Stephen C. YR 1987, [sans date]. The Mutation Rate and Variability of Eukaryotic Viruses: An Analytical Review. *Journal of General Virology*. Vol. 68, n° 11, pp. 2729-2740. DOI 10.1099/0022-1317-68-11-2729.
- SUGIURA, A., TOBITA, K. et KILBOURNE, Edwin D., 1972. Isolation and Preliminary Characterization of Temperature-Sensitive Mutants of Influenza Virus. *Journal of Virology*. octobre 1972. Vol. 10, n° 4, pp. 639-647. DOI 10.1128/jvi.10.4.639-647.1972.
- TE VELTHUIS, Aartjan J. W., ARNOLD, Jamie J., CAMERON, Craig E., VAN DEN WORM, Sjoerd H. E. et SNIJDER, Eric J., 2010. The RNA polymerase activity of SARS-coronavirus nsp12 is primer dependent. *Nucleic Acids Research*. 1 janvier 2010. Vol. 38, n° 1, pp. 203-214. DOI 10.1093/nar/gkp904.
- TROMAS, Nicolas et ELENA, Santiago F, 2010. The Rate and Spectrum of Spontaneous Mutations in a Plant RNA Virus. *Genetics*. 1 juillet 2010. Vol. 185, n° 3, pp. 983-989. DOI 10.1534/genetics.110.115915.
- VAN VLOTEN-DOTING, L., BOL, J. F. et CORNELISSEN, B., 1985. Plant-virus-based vectors for gene transfer will be of limited use because of the high error frequency during viral RNA synthesis. *Plant Molecular Biology*. septembre 1985. Vol. 4, n° 5, pp. 323-326. DOI 10.1007/BF02418253.
- VON MAGNUS, Preben, 1954. Incomplete Forms of Influenza Virus. In : SMITH, Kenneth M. et LAUFFER, Max A. (éd.), *Advances in Virus Research* [en ligne]. Academic Press. pp. 59-79. [Consulté le 11 septembre 2023].
- WILKE, Claus O, REISSIG, Daniel D et NOVELLA, Isabel S, 2004. Density-Dependent Selection in Vesicular Stomatitis Virus. *J. VIROL*. 2004. Vol. 78.
- YASUKAWA, Kiyoshi, IIDA, Kei, OKANO, Hiroyuki, HIDESE, Ryota, BABA, Misato, YANAGIHARA, Itaru, KOJIMA, Kenji, TAKITA, Teisuke et FUJIWARA, Shinsuke, 2017. Next-generation sequencing-based analysis of reverse transcriptase fidelity. *Biochemical and Biophysical Research Communications*. 14 octobre 2017. Vol. 492, n° 2, pp. 147-153. DOI 10.1016/j.bbrc.2017.07.169.
- ZHANG, Xiaomeng, RENNICK, Linda J., DUPREX, W. Paul et RIMA, Bert K., 2013. Determination of Spontaneous Mutation Frequencies in Measles Virus under Nonselective Conditions. *Journal of Virology*. mars 2013. Vol. 87, n° 5, pp. 2686-2692. DOI 10.1128/JVI.02146-12.







---

## Relation entre la complémentation et le taux de mutation des virus à ARN

---

**Auteur:** Raphaël Villeneuve

### Résumé

Les virus à ARN présentent un des taux de mutation les plus élevés de tous les organismes vivants ; or certaines mutations peuvent engendrer une perte de fonctionnalité. Ce travail s'axe sur un mécanisme propre aux virus à génome segmenté pouvant compenser les mutations délétères : la complémentation.

Cette thèse bibliographique établit l'unité substitution/nucléotide/événement de copie comme celle qui permet une comparaison fiable entre les taux de mutation des différentes espèces virales. Plusieurs méthodes de mesure du taux de mutation sont ensuite détaillées, la façon dont elles prennent en compte le biais de la sélection naturelle et les erreurs artéfactuelles notamment. Si elles ont toutes leurs avantages et leurs inconvénients respectifs, la plus fiable reste le test de fluctuation de Luria-Delbrück.

Fort de ce constat, deux comparaisons, entre à chaque fois deux virus à ARN dont l'un présente un génome segmenté et l'autre non, ont été réalisées. Le but était d'isoler la corrélation entre segmentation/non-segmentation et taux de mutation. Ces comparaisons permettent de mettre en évidence un taux de mutation des virus à ARN segmenté plus élevé d'un ordre de grandeur que celui des virus à ARN non-segmentés. La comparaison entre une estimation expérimentale du taux de mutation du virus de la rougeole et une de celle du virus influenza A en particulier, utilisant toutes deux un test de fluctuation et ne présentant aucun biais à l'analyse, incite à analyser cette constatation plus en profondeur.

Enfin, plusieurs pistes de réflexion sur les mécanismes permettant d'expliquer cette plus grande tolérance aux mutations des virus à ARN segmentés par rapport aux espèces non segmentés sont analysés. A l'issue de l'examen de travaux expérimentaux théoriques, la complémentation apparaît comme un processus tout à fait pertinent pour expliquer cet écart de taux de mutation entre ces deux catégories de virus à ARN.

**Mots-clés :** ARN, Complémentation, Mutation, Virus

---

## Relationship between complementation and mutation rate in RNA viruses

---

**Author:** Raphaël Villeneuve

### Abstract

RNA viruses have one of the highest mutation rates of all living organisms; however, these frequent mutations can cause a loss of functionality. This bibliographic thesis focuses on a mechanism specific to viruses with segmented genomes that can compensate for deleterious mutations: complementation.

This work first establishes the substitution/nucleotide/copy event unit as the one that allows a reliable comparison between the mutation rates of different viral species. Several methods for measuring the mutation rate are then detailed, the way in which they take into account the bias of natural selection and artifactual errors in particular. While they all have their respective advantages and disadvantages, the most reliable remains the Luria-Delbrück fluctuation test.

Based on this observation, two comparisons, each time between two RNA viruses, one of which has a segmented genome and the other not, were carried out. The goal was to isolate the correlation between segmentation/non-segmentation and mutation rate. These comparisons highlight a mutation rate of segmented RNA viruses that is an order of magnitude higher than the one of non-segmented RNA viruses. The comparison between an experimental estimate of the mutation rate of the measles virus and the one of the influenza A virus in particular, both using a fluctuation test and presenting no bias in the analysis, encourages us to analyze this observation more thoroughly. Finally, several avenues for reflection on the mechanisms explaining this greater tolerance to mutations in segmented RNA viruses compared to non-segmented species are analyzed. After examining theoretical experimental work, complementation appears to be a completely relevant process to explain this difference in mutation rates between these two categories of RNA viruses.

**Keywords:** RNA, Complementation, Mutation, Virus