



## Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>  
Eprints ID : 15544

**To cite this version :**

Cadiou, Yannis. *La théorie des quasi-espèces : concepts, application à la dynamique des populations de virus à ARN, implications biologiques et limites*. Thèse d'exercice, Médecine vétérinaire, Ecole Nationale Vétérinaire de Toulouse - ENVT, 2016, 218 p.

Any correspondance concerning this service should be sent to the repository administrator: [staff-oatao@inp-toulouse.fr](mailto:staff-oatao@inp-toulouse.fr).

# LA THEORIE DES QUASI-ESPECES : CONCEPTS, APPLICATION A LA DYNAMIQUE DES POPULATIONS DE VIRUS A ARN, IMPLICATIONS BIOLOGIQUES ET LIMITES

---

THESE  
pour obtenir le grade de  
DOCTEUR VETERINAIRE

DIPLOME D'ETAT

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

*par*

**CADIEU Yannis**

Né, le 13 juillet 1989 à Blois (41)

---

**Directeur de thèse : M. Stéphane BERTAGNOLI**

---

## JURY

PRESIDENT :

**M. Christophe PASQUIER**

Professeur à l'Université Paul-Sabatier de TOULOUSE

ASSESEURS :

**M. Stéphane BERTAGNOLI**

**M. Gilles MEYER**

Professeur à l'Ecole Nationale Vétérinaire de TOULOUSE

Professeur à l'Ecole Nationale Vétérinaire de TOULOUSE



**Ministère de l'Agriculture de l'Agroalimentaire et de la Forêt**  
**ECOLE NATIONALE VETERINAIRE DE TOULOUSE**

**Directrice** : **Madame Isabelle CHMITELIN**

**PROFESSEURS CLASSE EXCEPTIONNELLE**

- M. **AUTEFAGE André**, *Pathologie chirurgicale*
- Mme **CLAUW Martine**, *Pharmacie-Toxicologie*
- M. **CONCORDET Didier**, *Mathématiques, Statistiques, Modélisation*
- M. **DELVERDIER Maxence**, *Anatomie Pathologique*
- M. **ENJALBERT Francis**, *Alimentation*
- M. **FRANC Michel**, *Parasitologie et Maladies parasitaires*
- M. **MARTINEAU Guy**, *Pathologie médicale du Bétail et des Animaux de Basse-cour*
- M. **PETIT Claude**, *Pharmacie et Toxicologie*
- M. **REGNIER Alain**, *Physiopathologie oculaire*
- M. **SHELCHER François**, *Pathologie médicale du Bétail et des Animaux de Basse-cour*

**PROFESSEURS 1° CLASSE**

- M. **BERTAGNOLI Stéphane**, *Pathologie infectieuse*
- M. **BERTHELOT Xavier**, *Pathologie de la Reproduction*
- M. **BOUSQUET-MELOU Alain**, *Physiologie et Thérapeutique*
- M. **DUCOS Alain**, *Zootecnie*
- M. **FOUCRAS Gilles**, *Pathologie des ruminants*
- Mme **GAYRARD-TROY Véronique**, *Physiologie de la Reproduction, Endocrinologie*
- Mme **HAGEN-PICARD, Nicole**, *Pathologie de la reproduction*
- M. **LEFEBVRE Hervé**, *Physiologie et Thérapeutique*
- M. **MEYER Gilles**, *Pathologie des ruminants*
- M. **SANS Pierre**, *Productions animales*
- Mme **TRUMEL Catherine**, *Biologie Médicale Animale et Comparée*

**PROFESSEURS 2° CLASSE**

- M. **BAILLY Jean-Denis**, *Hygiène et Industrie des aliments*
- Mme **BENARD Geneviève**, *Hygiène et Industrie des Denrées alimentaires d'Origine animale*
- Mme **BOURGES-ABELLA Nathalie**, *Histologie, Anatomie pathologique*
- M. **BRUGERE Hubert**, *Hygiène et Industrie des aliments d'Origine animale*
- Mme **CHASTANT-MAILLARD Sylvie**, *Pathologie de la Reproduction*
- M. **GUERRE Philippe**, *Pharmacie et Toxicologie*
- M. **GUERIN Jean-Luc**, *Aviculture et pathologie aviaire*
- M. **JACQUIET Philippe**, *Parasitologie et Maladies Parasitaires*
- M. **LIGNEREUX Yves**, *Anatomie*
- M. **PICAVET Dominique**, *Pathologie infectieuse*

**PROFESSEURS CERTIFIES DE L'ENSEIGNEMENT AGRICOLE**

- Mme **MICHAUD Françoise**, *Professeur d'Anglais*
- M. **SEVERAC Benoît**, *Professeur d'Anglais*

#### **MAITRES DE CONFERENCES HORS CLASSE**

- M. **BERGONIER Dominique**, *Pathologie de la Reproduction*
- Mme **BOUCLAINVILLE-CAMUS Christelle**, *Biologie cellulaire et moléculaire*
- Mlle **BOULLIER Séverine**, *Immunologie générale et médicale*
- Mlle **DIQUELOU Armelle**, *Pathologie médicale des Equidés et des Carnivores*
- M. **DOSSIN Olivier**, *Pathologie médicale des Equidés et des Carnivores*
- M. **JOUGLAR Jean-Yves**, *Pathologie médicale du Bétail et des Animaux de Basse-cour*
- Mme **LETRON-RAYMOND Isabelle**, *Anatomie pathologique*
- M. **LYAZRHI Faouzi**, *Statistiques biologiques et Mathématiques*
- M. **MATHON Didier**, *Pathologie chirurgicale*
- Mme **PRIYMENKO Nathalie**, *Alimentation*

#### **MAITRES DE CONFERENCES (classe normale)**

- M. **ASIMUS Erik**, *Pathologie chirurgicale*
- Mme **BENNIS-BRET Lydie**, *Physique et Chimie biologiques et médicales*
- Mlle **BIBBAL Delphine**, *Hygiène et Industrie des Denrées alimentaires d'Origine animale*
- Mme **BOUHSIRA Emilie**, *Parasitologie, maladies parasitaires*
- Mlle **CADIERGUES Marie-Christine**, *Dermatologie*
- M. **CONCHOU Fabrice**, *Imagerie médicale*
- M. **CORBIERE Fabien**, *Pathologie des ruminants*
- M. **CUEVAS RAMOS Gabriel**, *Chirurgie Equine*
- Mme **DANIELS Hélène**, *Microbiologie-Pathologie infectieuse*
- Mlle **DEVIERS Alexandra**, *Anatomie-Imagerie*
- M. **DOUET Jean-Yves**, *Ophthalmologie vétérinaire et comparée*
- Mlle **FERRAN Aude**, *Physiologie*
- M. **JAEG Jean-Philippe**, *Pharmacie et Toxicologie*
- Mlle **LACROUX Caroline**, *Anatomie Pathologique des animaux de rente*
- Mlle **LAVOUE Rachel**, *Médecine Interne*
- M. **LIENARD Emmanuel**, *Parasitologie et maladies parasitaires*
- M. **MAILLARD Renaud**, *Pathologie des Ruminants*
- Mme **MEYNADIER Annabelle**, *Alimentation*
- Mme **MEYNAUD-COLLARD Patricia**, *Pathologie Chirurgicale*
- M. **MOGICATO Giovanni**, *Anatomie, Imagerie médicale*
- M. **NOUVEL Laurent**, *Pathologie de la reproduction (en disponibilité)*
- Mlle **PALIERNE Sophie**, *Chirurgie des animaux de compagnie*
- Mlle **PAUL Mathilde**, *Epidémiologie, gestion de la santé des élevages avicoles et porcins*
- Mme **PRADIER Sophie**, *Médecine interne des équidés*
- M. **RABOISSON Didier**, *Productions animales (ruminants)*
- M. **VOLMER Romain**, *Microbiologie et Infectiologie*
- M. **VERWAERDE Patrick**, *Anesthésie, Réanimation*
- Mme **WARET-SZKUTA Agnès**, *Production et pathologie porcine*

#### **MAITRES DE CONFERENCES et AGENTS CONTRACTUELS**

- M. **DAHAN Julien**, *Médecine Interne*

#### **ASSISTANTS D'ENSEIGNEMENT ET DE RECHERCHE CONTRACTUELS**

- Mme **COSTES Laura**, *Hygiène et industrie des aliments*
- Mme **LALLEMAND Elodie**, *Chirurgie des Equidés*
- M. **TANIS Jean-Benoît**, *Anatomie – Imagerie Médicale*

# REMERCIEMENTS

## **A Monsieur le Professeur Christophe Pasquier**

Professeur des Universités – Praticien hospitalier

Doyen de la Faculté des Sciences Pharmaceutiques

Microbiologie – Virologie

Pour m'avoir fait l'honneur d'accepter la présidence de ce jury de thèse,

Qu'il trouve ici sincères remerciements et hommages respectueux.

## **A Monsieur le Professeur Stéphane Bertagnoli**

Professeur de l'Ecole Nationale Vétérinaire de Toulouse

Virologie – Infectiologie

Pour m'avoir fait l'honneur d'être mon directeur de thèse. Pour votre disponibilité, votre gentillesse, votre patience et votre bonne humeur permanente. Pour toutes ces discussions posées et réfléchies sur ce travail mais pas que ! Pour votre aide malgré votre emploi du temps plus que surchargé. Sincères remerciements.

## **A Monsieur le Professeur Gilles Meyer**

Professeur de l'Ecole Nationale Vétérinaire de Toulouse

Pathologie des ruminants

Pour m'avoir fait l'honneur d'accepter d'être l'assesseur de cette thèse,

Qu'il trouve ici sincères remerciements et hommages respectueux.



# TABLE DES MATIERES

TABLE DES ANNEXES .....	16
TABLE DES ILLUSTRATIONS.....	17
LISTE DES ABREVIATIONS .....	19
INTRODUCTION.....	21
Partie 1    Rappels structuraux, cycles et variabilité génétique chez les virus à ARN.....	23
1.1.    Structure et constituants viraux .....	23
1.1.1.    L'acide nucléique viral .....	23
1.1.2.    Les protéines virales .....	24
1.1.3.    Les lipides viraux .....	24
1.2.    L'architecture virale .....	24
1.2.1.    Les capsides ou nucléocapsides.....	24
1.2.2.    Les enveloppes virales.....	25
1.3.    Le cycle viral .....	25
1.3.1.    Phase initiale.....	25
1.3.2.    Phase de multiplication.....	26
1.3.3.    Phase de libération.....	26
1.3.4.    Diversité du cycle viral et de l'expression génétique .....	27
1.4.    Dogme central de la biologie moléculaire chez les virus .....	28
1.4.1.    Première ébauche du dogme central.....	28
1.4.2.    Extension du dogme central aux virus.....	28
1.4.3.    Le code génétique.....	29
1.5.    Familles de virus faisant l'objet de l'étude.....	29
1.6.    Variabilité et instabilité génétique chez les virus à ARN .....	30
1.6.1.    Mise en évidence précoce de la variabilité et de l'instabilité virale .....	30
1.6.2.    Enzymes de réplication et génération des mutants .....	31
1.6.3.    Les mutations, mécanisme de variabilité génétique .....	31
1.6.3.1.    Les mutations ponctuelles .....	32
1.6.3.2.    Hypomutations et hypermutations.....	33
1.6.3.3.    Les expansions répétées .....	34
1.6.3.4.    Les duplications.....	34
1.6.3.5.    Conséquences des microlésions.....	34
1.6.3.5.1.    Conséquences d'une substitution en séquence codante.....	35



1.6.3.5.2.	Conséquences des décalages du cadre de lecture en séquence codante .....	35
1.6.3.5.3.	Conséquences des microlésions en région non-codante.....	36
1.6.3.5.4.	Conséquences sur l'aspect fonctionnel.....	36
1.6.3.5.5.	Conséquences en fonction du contexte.....	36
1.6.4.	Les recombinaisons .....	36
1.6.5.	Les réarrangements/réassortiments.....	38
1.6.6.	La segmentation du génome.....	39
1.6.7.	Conclusion sur la variabilité.....	40
1.7.	Conclusion.....	40
Partie 2	Principes darwiniens chez des populations virales à ARN.....	41
2.1.	Théorie de l'évolution, un premier pas vers l'évolution des populations.....	41
2.2.	La théorie synthétique de l'évolution .....	42
2.2.1.	Loi de Hardy-Weinberg.....	42
2.2.2.	Caractéristiques des facteurs .....	43
2.2.3.	Variation génétique .....	43
2.2.3.1.	Taux de mutation.....	43
2.2.3.2.	Fréquence de mutation .....	44
2.2.3.3.	Taux d'évolution .....	45
2.2.4.	La compétition.....	45
2.2.5.	La sélection.....	46
2.2.5.1.	Extension au milieu extracellulaire .....	46
2.2.5.2.	Limite floue entre sélections positive et négative.....	46
2.2.5.3.	Contraintes sélectives .....	47
2.2.6.	Effets fondateurs et dérive génétique .....	47
2.2.6.1.	Effets fondateurs.....	47
2.2.6.2.	Dérive génétique.....	48
2.2.7.	Les migrations .....	48
2.2.8.	Bilan des forces évolutives appliquées aux populations virales.....	49
2.3.	La valeur sélective virale.....	50
2.3.1.	La valeur sélective en génétique des populations.....	50
2.3.2.	La valeur sélective appliquée aux virus.....	50
2.3.3.	Mesure de la valeur sélective chez les virus.....	51
2.3.4.	Intérêts des mesures de la valeur sélective .....	52
2.3.5.	La valeur sélective moyenne .....	53
2.4.	Le paysage adaptatif .....	53

2.4.1.	Topographie des paysages.....	54
2.4.2.	Déplacements au sein des paysages et shifting balance theory .....	56
2.4.2.1.	Selon la séquence nucléotidique.....	56
2.4.2.2.	Selon l'environnement.....	56
2.4.2.3.	Shifting balance theory.....	56
2.4.3.	Un paysage mouvant et complexe.....	57
2.5.	Conclusion.....	58
Partie 3	Les quasi-espèces virales : expériences historiques, modèles théoriques et expérimentaux	59
3.1.	Le modèle mathématique d'Eigen.....	59
3.1.1.	Les expériences de Spiegelman et l'origine des passages en série .....	59
3.1.2.	La théorie mathématique d'Eigen et Schuster.....	60
3.1.2.1.	Approche intuitive .....	60
3.1.2.2.	Equations fondamentales de la théorie mathématique.....	61
3.1.3.	Simplification du modèle de la quasi-espèce .....	62
3.1.3.1.	La balance sélection-mutation chez un modèle répliatif présentant des erreurs..	62
3.1.3.2.	Le seuil d'erreur et l'erreur catastrophique .....	64
3.1.3.3.	Le seuil d'extinction .....	68
3.1.3.4.	Ajout d'autres génotypes au modèle .....	69
3.1.3.5.	Introduction des mutations retours .....	70
3.1.3.6.	Les réseaux neutres.....	71
3.1.4.	Le modèle d'Eigen .....	73
3.2.	Applications du modèle à des virus à ARN.....	73
3.2.1.	Mise en évidence d'un réservoir de mutants .....	74
3.2.2.	Une distribution chez le bactériophage Q $\beta$ évoquant une quasi-espèce.....	74
3.2.3.	Un modèle applicable mais limité .....	75
3.3.	Les quasi-espèces chez les virus à ARN .....	75
3.3.1.	Organisation du spectre de mutants.....	77
3.3.1.1.	Définition du spectre de mutants .....	77
3.3.1.2.	La séquence maîtresse .....	77
3.3.1.3.	La séquence consensus .....	77
3.3.1.4.	Un spectre organisé de mutants.....	77
3.3.2.	Formation du spectre de mutants et exploration de l'espace séquence .....	78
3.3.2.1.	L'espace séquence .....	79
3.3.2.1.1.	Représentation en deux dimensions .....	79

3.3.2.1.2.	Représentation en trois dimensions .....	80
3.3.2.2.	Exploration de l'espace séquence.....	81
3.3.3.	Seuil d'erreur et l'erreur catastrophique chez les virus à ARN .....	82
3.3.3.1.	Seuil d'erreur et erreur catastrophique chez les virus.....	82
3.3.3.2.	Application à l'exploration de l'espace séquence .....	84
3.4.	Mesure de la complexité du spectre .....	84
3.4.1.	Paramètres utilisés pour estimer la complexité du spectre .....	84
3.4.2.	Méthodes de séquençage classique.....	85
3.4.2.1.	Clonages biologique et moléculaire .....	85
3.4.2.2.	Le séquençage de Sanger.....	86
3.4.3.	Séquençage haut débit appliqué aux études des quasi-espèces virales.....	88
3.4.3.1.	Déroulement général du séquençage haut-débit .....	88
3.4.3.2.	Technologies utilisant le séquençage haut-débit .....	90
3.4.3.3.	L'analyse des données.....	90
3.4.3.4.	Apports du séquençage haut débit dans l'étude des quasi-espèces .....	91
3.4.4.	Méthodes de classification .....	94
3.4.4.1.	Partition Analysis of Quasispecies .....	94
3.4.4.2.	Une grande variété de méthodes d'analyses et d'échantillonnages.....	94
3.5.	Conclusion.....	95
Partie 4	Paramètres et dynamique des quasi-espèces virales.....	97
4.1.	Hétérogénéité dans les populations virales à ARN .....	97
4.1.1.	Une fidélité de réplication modulable .....	97
4.1.2.	Le fort taux de mutation, source d'hétérogénéité de la population .....	98
4.1.3.	Remplissage du spectre de mutants : première approche de la dynamique des quasi-espèces	99
4.1.3.1.	Remise en question de l'intérêt de l'étude de la séquence consensus .....	99
4.1.3.2.	Une dynamique basée sur la composition du spectre de mutants.....	100
4.1.3.3.	Evolution du spectre de mutant sans changement de la séquence consensus.....	101
4.1.3.4.	Evolution du spectre de mutant avec changement de la séquence consensus .....	102
4.1.4.	Conclusion.....	103
4.2.	Taille du génome viral.....	103
4.2.1.	Une meilleure exploration de l'espace séquence.....	103
4.2.2.	Relation du seuil d'erreur et taille du génome.....	104
4.2.3.	Un génome soumis à des contraintes fonctionnelles .....	104
4.2.4.	Segmentation du génome .....	105

4.3.	Impact de la taille de la population.....	106
4.3.1.	Approche théorique à l'aide de l'espace séquence .....	106
4.3.2.	Taille et hétérogénéité de la population virale.....	106
4.3.3.	Etudes expérimentales des valeurs sélectives.....	107
4.3.3.1.	Perte de valeur sélective au sein d'une population virale.....	107
4.3.3.1.1.	Simulations expérimentales des goulots d'étranglement .....	107
4.3.3.1.2.	Applications expérimentales du cliquet de Müller .....	108
4.3.3.1.3.	Maintien de valeur adaptative et échappement à l'extinction .....	109
4.3.3.1.4.	Une conséquence positive de la sélection négative .....	109
4.3.3.2.	Augmentation de la valeur sélective au sein d'une population virale et atteinte de l'équilibre	109
4.3.3.2.1.	Définition de l'équilibre en matière de valeur sélective.....	109
4.3.3.2.2.	Mise en évidence expérimentale et interprétation .....	110
4.3.3.2.3.	Avantage sélectif de la taille et de l'hétérogénéité.....	110
4.3.3.3.	Limites de la taille du spectre.....	111
4.3.4.	Le taux d'évolution, autre approche de l'importance de la taille et de l'hétérogénéité de la population .....	111
4.4.	Apport des paysages adaptatifs : paysages accidentés et survie du plus plat .....	111
4.4.1.	Barrières génétiques et phénotypiques .....	111
4.4.2.	Un paysage adaptatif accidenté chez les virus.....	112
4.4.3.	La survie du plus plat sur des paysages adaptatifs .....	113
4.4.4.	Observations expérimentales.....	116
4.4.5.	Difficultés d'interprétation .....	116
4.4.6.	Conclusion.....	116
4.5.	Les facteurs liés à l'hôte .....	117
4.6.	Conclusions .....	118
4.6.1.	Conclusion sur la dynamique des quasi-espèces .....	118
4.6.2.	Paramètres viraux et impact sur l'adaptabilité du virus.....	121
Partie 5	Implications biologiques de la dynamique des quasi-espèces.....	123
5.1.	Résistance et mutants d'échappement .....	123
5.1.1.	Les contraintes sélectives appliquées lors de traitements.....	123
5.1.2.	Définition des mutants d'échappement .....	124
5.1.3.	Résistance aux inhibiteurs antiviraux .....	124
5.1.3.1.	Les inhibiteurs antiviraux .....	124
5.1.3.2.	Mécanismes et caractéristiques généraux des résistances .....	124

5.1.3.3.	Sélection de mutants résistants .....	125
5.1.3.3.1.	Barrière génétique .....	125
5.1.3.3.2.	Barrière phénotypique .....	125
5.1.3.3.3.	Concentrations et caractéristiques de l'inhibiteur .....	126
5.1.3.3.4.	Apparition de résistance sans exposition aux inhibiteurs.....	126
5.1.3.4.	Dynamique d'apparition des résistances .....	127
5.1.3.5.	Un nouveau mécanisme de résistance .....	128
5.1.4.	Résistance à la réponse immunitaire et à la vaccination.....	129
5.1.4.1.	Mutants d'échappement face à la réponse immunitaire .....	129
5.1.4.1.1.	Immunité innée.....	129
5.1.4.1.2.	Immunité adaptative .....	129
5.1.4.1.3.	Multiple exemples de mutants d'échappement.....	130
5.1.4.2.	Mutants d'échappement dans le cadre de la vaccination.....	130
5.1.4.2.1.	Principes de la vaccination.....	130
5.1.4.2.2.	Echappement à la vaccination .....	131
5.1.4.3.	Orientation de l'évolution virale par la vaccination .....	132
5.2.	Modification du tropisme cellulaire .....	133
5.2.1.	Modulation par les protéines structurales.....	133
5.2.1.1.	Les récepteurs cellulaires .....	133
5.2.1.2.	Interactions cellules-virus via les récepteurs .....	133
5.2.1.3.	Conséquences des substitutions d'acides aminés présents dans les sites de reconnaissance.....	134
5.2.1.3.1.	Changement de tropisme cellulaire .....	134
5.2.1.3.2.	Expansion du tropisme cellulaire .....	135
5.2.1.3.3.	Modification de la pathogénèse.....	135
5.2.2.	Modulation par les protéines non structurales.....	136
5.2.3.	Dynamique de la modulation du tropisme.....	136
5.2.4.	La coévolution du tropisme et de l'antigénicité .....	137
5.2.4.1.	Variation antigénique et modification du tropisme <i>in vitro</i> .....	138
5.2.4.2.	Variation antigénique et modification du tropisme <i>in vivo</i> .....	139
5.2.4.3.	Dynamique de la coévolution .....	139
5.2.5.	Conclusion.....	140
5.3.	La mémoire moléculaire.....	140
5.3.1.	L'hypothèse des génomes mémoires.....	141
5.3.2.	Mise en évidence d'une mémoire moléculaire chez les virus à ARN .....	141
5.3.3.	Caractéristiques des génomes mémoires .....	142

5.3.4.	Proposition schématique de la dynamique de la mémoire moléculaire.....	142
5.3.5.	Proposition de mécanisme moléculaire .....	143
5.3.6.	Implications de la mémoire moléculaire .....	144
5.4.	Les quasi-espèces, unité de sélection soumise à des interactions internes .....	145
5.4.1.	Complémentarité .....	145
5.4.1.1.	Approches expérimentales.....	145
5.4.1.2.	Caractéristiques de la complémentarité.....	146
5.4.2.	Interférence.....	147
5.4.2.1.	Approches théoriques et expérimentales .....	147
5.4.2.2.	Caractéristiques des interférences .....	147
5.4.2.3.	Sélection densité-dépendante et auto-organisation.....	148
5.4.3.	Mécanismes moléculaires de la complémentarité et de l'interférence .....	149
5.4.3.1.	Deux mécanismes liés .....	149
5.4.3.2.	Différences avec les mécanismes inhibiteurs généraux.....	150
5.4.3.3.	Mécanisme de complémentarité et d'interférence.....	150
5.4.3.4.	Difficultés pour établir les mécanismes d'interactions.....	152
5.4.3.5.	Interactions des quasi-espèces en tant que groupe .....	152
5.5.	La mutagénèse létale .....	154
5.5.1.	Bases de la stratégie antivirale .....	154
5.5.2.	Principes de la mutagénèse létale .....	155
5.5.3.	Les analogues nucléotidiques et nucléosidiques .....	155
5.5.4.	Mécanismes de la mutagénèse létale.....	156
5.5.4.1.	Caractéristiques de la mutagénèse létale .....	156
5.5.4.2.	Etapas de la mutagénèse létale .....	157
5.5.5.	Limites de la mutagénèse létale.....	159
5.5.5.1.	Termes utilisés.....	159
5.5.5.2.	Effets bénéfiques de la mutagénèse létale pour le virus .....	159
5.5.5.3.	Résistance à la mutagénèse létale.....	160
5.5.5.3.1.	Plusieurs mécanismes de résistance .....	160
5.5.5.3.2.	Pertinence de l'apparition des résistances .....	162
5.5.5.4.	Effets secondaires des agents mutagènes .....	163
5.5.6.	Conclusion.....	163
5.6.	Mise en place de protocoles combinant mutagénèse létale et d'autres stratégies antivirales	164
5.6.1.	Autres stratégies de lutte antivirale .....	164
5.6.1.1.	La thérapie combinée .....	164

5.6.1.2.	Un traitement en deux étapes.....	165
5.6.1.3.	L'utilisation de médicaments dirigés contre les fonctions cellulaires .....	165
5.6.1.4.	Stimulation de l'immunité innée .....	165
5.6.1.5.	Combinaison d'immunothérapie et de chimiothérapie.....	166
5.6.2.	Utilisation de mutagènes pour une thérapie combinée .....	166
5.6.3.	Utilisation de mutagènes de manière successive ou simultanée.....	167
5.6.4.	Limites d'application.....	167
5.6.5.	Conclusion.....	168
5.7.	Conclusion des implications biologiques .....	168
5.7.1.	Application de la dynamique des quasi-espèces virales <i>in vivo</i> .....	168
5.7.2.	Application de la dynamique des quasi-espèces sur l'évolution sur le long terme .....	169
Partie 6	Limites et extensions de la théorie des quasi-espèces .....	171
6.1.	Une théorie déterministe.....	171
6.1.1.	L'intervention d'éléments stochastiques .....	171
6.1.2.	Des exemples isolés de comportements déterministes au sein des populations virales 172	
6.1.3.	Nuances du comportement déterministe.....	172
6.1.4.	Une démarche dite classique .....	173
6.1.5.	Conclusion.....	174
6.2.	Limite des méthodes de séquençage.....	174
6.2.1.	Limites d'échantillonnage des clonages biologique et moléculaire .....	174
6.2.1.1.	Echantillons parfois peu représentatifs et erreurs de copies .....	174
6.2.1.2.	Echantillonnage faible .....	175
6.2.2.	Limites de l'amplification et du séquençage .....	175
6.2.3.	Corrections des limites .....	176
6.2.3.1.	Corrections des erreurs dues à l'amplification .....	176
6.2.3.2.	Circular Sequencing .....	177
6.2.3.3.	Méthodes de correction .....	178
6.2.4.	Conclusion.....	178
6.3.	Limites des mesures de la valeur sélective .....	179
6.3.1.	Des limites dues aux interactions d'une multitude de facteurs.....	179
6.3.2.	Limites retrouvées pour les protocoles expérimentaux .....	179
6.3.3.	Difficultés de l'application de la valeur sélective <i>in vivo</i> .....	180
6.3.4.	Conclusion.....	181
6.4.	Remise en cause de la théorie.....	181

6.4.1.	Mauvaise utilisation du terme .....	181
6.4.2.	Remise en cause des caractéristiques du comportement des quasi-espèces .....	181
6.4.3.	Un modèle fait d'exemples.....	182
6.4.4.	Plusieurs modèles complémentaires.....	183
6.5.	Extension de la théorie à d'autres organismes.....	183
6.5.1.	Extension aux virus à ADN.....	183
6.5.2.	Extension aux cellules : comportement collectif chez des bactéries .....	184
6.5.3.	Extension aux cellules cancéreuses : hétérogénéité et comportement en groupe.....	185
6.5.4.	Extension aux hypothèses des origines de la vie.....	186
6.5.5.	Extension à des systèmes de prions.....	186
6.6.	Conclusion.....	187
CONCLUSION .....		189
BIBLIOGRAPHIE .....		190
WEBOGRAPHIE.....		207
ANNEXES .....		208



## TABLE DES ANNEXES

ANNEXE 1 : FREQUENCES DE MUTATIONS CHEZ DES VIRUS A ARN.....	208
ANNEXE 2 : IMPACT DE MUTATIONS RETOURS .....	209
ANNEXE 2 (suite) : IMPACT DE MUTATIONS RETOURS .....	210
ANNEXE 3 : EXEMPLES D'APPLICATION DES NGS EN VIROLOGIE.....	211
ANNEXE 4 : CARACTERISTIQUES DES PLATE-FORMES NGS .....	212
ANNEXE 5 : TAUX D'EVOLUTION POUR LES VIRUS A ARN .....	213
ANNEXE 6 : FREQUENCE DE RESISTANCE AUX ANTICORPS.....	215
ANNEXE 7 : MODIFICATION DU TROPISME CELLULAIRE .....	216
ANNEXE 8 : EXEMPLES DE CHEVAUCHEMENT DE SITES ANTIGENIQUES ET DE SITE DE RECONNAISSANCE DES RECEPTEURS CHEZ LES VIRUS A ARN.....	217
ANNEXE 9 : LOGICIELS DE RECONSTRUCTION D'UNE QUASI-ESPECE VIRALE .....	218

## TABLE DES ILLUSTRATIONS

Figure 1: Résumé du cycle viral productif .....	27
Figure 2 : Code génétique .....	29
Figure 3 : Résumé des différents types de mutations ponctuelles .....	33
Figure 4 : Représentation schématique simplifiée des phénomènes de recombinaison .....	37
Figure 5 : Représentation schématique simplifiée du mécanisme de réassortiment .....	39
Figure 6 : Représentation schématique simplifiée de la segmentation du génome .....	40
Figure 7 : Représentation schématique de la différence entre fréquence et taux de mutation .....	45
Figure 8 : Schéma bilan simplifié des forces évolutives .....	49
Figure 9 : Représentation schématique de vecteurs de valeur sélective, sur un exemple de 4 transferts .....	52
Figure 10 : Exemple d'un paysage adaptatif accidenté en trois dimensions de Kaufmann et Levin ...	54
Figure 11 : Paysage adaptatif en 2 dimensions appliqué aux populations virales en fonction de la position dans l'espace séquence .....	55
Figure 12 : Paysage adaptatif en 2 dimensions appliqué aux populations virales en fonction du changement d'environnement .....	55
Figure 13 : Description schématique du passage en série .....	60
Figure 14 : Modèle d'évolution d'une quasi-espèce avec deux génotypes .....	63
Figure 15 : Approche dynamique de l'équilibre des quasi-espèces .....	64
Figure 16 : Proportion de A1 en fonction du taux de mutation et erreur catastrophique .....	65
Figure 17 : Le seuil d'erreur .....	66
Figure 18 : Différences entre seuil d'erreur et seuil d'extinction .....	68
Figure 19 : Modèle avec plusieurs génotypes .....	69
Figure 20 : Modèle de deux génotypes avec des mutations-retours .....	70
Figure 21 : Valeurs absolues en fonction du taux de mutation .....	71
Figure 22 : Modèle avec 3 génotypes et un réseau neutre simple de deux génotypes .....	72
Figure 23 : Deux représentations des quasi-espèces virales .....	76
Figure 24 : Distribution des quasi-espèces .....	78
Figure 25 : Représentation schématique des connections entre des génomes hypothétiques .....	79
Figure 26 : Représentation schématique de la formation d'une quasi-espèce virale à partir d'un génome matrice sur un espace séquence en deux dimensions .....	80
Figure 27 : Représentation théorique de l'occupation de l'espace séquence en trois dimensions .....	81
Figure 28 : Distribution de la population au sein d'une quasi-espèce de PV et seuil d'erreur .....	83
Figure 29 : Représentation schématique des étapes de clonage moléculaire et biologique utilisées pour l'analyse des quasi-espèces virales .....	86
Figure 30 : Représentation des deux méthodes d'électrophorèse .....	87
Figure 31 : Etapes d'un séquençage haut-débit .....	89
Figure 32 : Les 3 approches de l'identification des variants viraux et de leurs fréquences .....	91
Figure 33 : Mise en évidence d'événements de sélection transitoire grâce aux NGS .....	93
Figure 34 : Représentation schématique de l'évolution d'une quasi-espèce virale, sans modification de la séquence consensus .....	101
Figure 35 : Représentation schématique de l'évolution d'une quasi-espèce virale, avec possible modification de la séquence consensus .....	102
Figure 36 : La survie du plus plat du point de vue d'un paysage adaptatif .....	114

Figure 37 : La survie du plus plat en fonction du taux de mutation .....	115
Figure 38 : Représentation schématique de l'évolution des populations de virus à ARN .....	119
Figure 39 : représentation schématique de l'effet de la taille de la population sur l'évolution.....	120
Figure 40 : Cas de coévolution chez le FMDV .....	139
Figure 41 : Représentation schématique de la mémoire des quasi-espèces et de l'effet des goulots d'étranglement sur la mémoire.....	143
Figure 42 : Modèle de complémentarité et d'interférence médiée par des interactions au sein d'une protéine polymérique .....	151
Figure 43 : Représentation schématique des interactions de complémentarité et de défection au sein d'un spectre de mutants. ....	153
Figure 44 : Effets de l'augmentation de la mutagenèse dans une population de virus.....	157
Figure 45 : Evaluation schématique actuelle des événements impliqués dans l'extinction virale .....	158
Figure 46 : Distribution d'une population au sein d'une quasi-espèce de PV-G64S et seuil d'erreur	161
Figure 47 : Représentation de l'évolution temporelle des quasi-espèces virales .....	173
Figure 48 : Représentation schématique de la procédure CirSeq .....	177

## **LISTE DES ABREVIATIONS**

5-AZC = 5-Azacytidine  
5-FU = 5-Fluorouracil  
5-OH-dC = 5-hydroxydéoxycytidine  
A = Adénine  
ADN = Acide désoxyribonucléique  
ADNc = ADN complémentaire  
ARN = Acide Ribonucléique  
ARNm = Acide ribonucléique messenger  
BHK-21 = Baby hamster kidney cell  
C = Cytosine  
CBV3 = Cocksackievirus B3  
CCSS = Classical Molecular Cloning and Sanger sequencing  
CHO = Chinese Hamster Ovary  
CMV = Cytomegalovirus  
DI = Defective Interfering  
EMCV = EncephaloMyoCarditis virus  
FMDV = Foot-and-Mouth Disease Virus  
G = Guanine  
GU = Guanidine Hydrochloride  
HAV = Hepatitis A Virus  
HBV = Hepatitis B Virus  
HCV = Hepatitis C Virus  
HDV = Hepatitis Delta Virus  
HIV-1 = Human Immunodeficiency Virus 1  
HRV = Human Rhinovirus  
IFN = Interféron  
IV = Influenza virus  
Kb = kilobase = 1000 bases nucléotidiques  
LCMV = Lymphocytic ChorioMeningitis Virus  
MAR = Monoclonal Antibody Resistant  
MeV = Measles Virus  
MOI = Multiplicity Of Infection  
NGS = Next-Generation Sequencing  
PAQ = Partition Analysis of Quasispecies  
PCR = Polymeras Chain Reaction  
PNS = Protéines Non Structurales  
PRRSV = Porcine Reproductive and Respiratory Syndrome virus  
PS = Protéines Structurales  
PV = Poliovirus  
R = Ribavirine  
RT-PCR = Reverse transcription Polymeras Chain Reaction  
RSV = Respiratory Syncytial Virus  
U = Uracile  
UDS = Ultra-Deep Sequencing  
VSV = Vesicular Stomatitis Virus



## INTRODUCTION

Les notions d'une nature en pleine évolution ont toujours intrigué la réflexion humaine, que ce soit pour savoir qu'elles étaient les origines de l'homme ou pour l'appliquer en agriculture ou en élevage via la sélection artificielle. Bien que tardive, la théorie de l'évolution a ouvert la marche à de nombreux autres modèles afin d'appréhender les forces remaniant constamment les espèces qui nous entourent.

L'apprentissage des forces évolutives, afin d'appréhender au mieux l'évolution des espèces, est passé par de nombreuses observations paléontologiques, mais aussi par de nombreux modèles utilisant des microorganismes. Les virus en sont un exemple et sont des piliers de la recherche évolutive. Nombreuses sont les expériences qui tentent de percer le fonctionnement de ces populations si particulières. Les événements majeurs intervenant dans la biologie des virus sont en cours d'investigation, tels que leur capacité à changer leur tropisme cellulaire, à changer d'hôte, à surpasser les contraintes sélectives, à échapper à des traitements antiviraux et à provoquer des récurrences chez certains patients. Même l'émergence de nouvelles souches virales pathogènes est due à ce comportement propre aux populations virales. Cela explique les difficultés majeures rencontrées dans la prévention des maladies et la mise en place des traitements. Cette dynamique est encore de nos jours difficilement compréhensible, et donc prédictible.

Un mathématicien, Eigen, a émis un modèle mathématique pour tenter de prédire le comportement d'entités auto-répliquatives avec un génome de petite taille et présentant de nombreuses erreurs lors de leur réplication, ces entités pouvant correspondre à des virus à acide ribonucléique (ARN). De nombreux résultats expérimentaux et théoriques ont trouvé en ce modèle une tentative d'explication, ce qui a tôt fait de convaincre une grande partie du domaine de la recherche en virologie. Pour qualifier ce comportement biologique des virus à ARN, Eigen a posé les bases de la théorie des quasi-espèces. Cette dernière a pour but de mettre en avant des caractéristiques du comportement d'une population virale dans son environnement, et n'a eu de cesse d'évoluer en parallèle de la virologie.

Il est proposé ici de visiter les bases théoriques et expérimentales qui composent cette théorie, les paramètres de ces quasi-espèces ainsi que les implications que la théorie a sur la compréhension du comportement des virus à ARN. Cela passera par une synthèse bibliographique des différents éléments qui s'articulent autour de la théorie des quasi-espèces, afin de démontrer ou non la concordance de tous ces éléments avec les caractéristiques des virus à ARN. Les résultats théoriques et pratiques qui semblent s'intéresser à cette théorie chez certains virus sont nombreux, sans doute trop pour être tous répertoriés : le but n'est donc pas d'être exhaustif, mais d'apporter suffisamment de matériel à la bonne compréhension des propriétés supposées de ces populations.

Un rappel de la structure et des cycles viraux à ARN va permettre de développer les caractéristiques de l'objet d'étude. Etant donné que les quasi-espèces sont des populations virales, il semble nécessaire d'aborder les tenants et aboutissants de la génétique des populations, dont la théorie des quasi-espèces se veut être une extension. Cela permettra d'exposer le modèle de la quasi-espèce, les protocoles d'études à disposition, sa dynamique et sa complexité. Tous ces paramètres ont des implications biologiques qu'il convient d'étudier, telles que les interactions au sein d'une population, l'apparition de résistance, la mémoire moléculaire ou la mise en place de nouveaux traitements. Enfin, les limites et possibles extensions à d'autres modèles biologiques seront passées en revue.



## **Partie 1    Rappels structuraux, cycles et variabilité génétique chez les virus à ARN**

Pendant longtemps, on a considéré que la transformation opérée chez les espèces suite au processus de l'évolution n'était pas observable à notre échelle temporelle. L'étude des virus, en particulier des virus à ARN, a complètement réformé cette vision. L'émergence de nouvelles espèces virales en un laps de temps très court, avec des conséquences dramatiques qui sont dues à leurs propriétés, en sont la preuve. C'est d'ailleurs ce qui rend aujourd'hui la lutte contre les nouvelles maladies virales extrêmement difficile : des traitements, qui offraient au premier abord une bonne protection, peuvent provoquer l'apparition de virus résistants de manière très rapide.

Ainsi, si les principes de l'évolution et les mécanismes moléculaires de variations sont les mêmes pour toutes les formes de vie, c'est la période nécessaire à des événements évolutifs majeurs qui diffère énormément d'un organisme à un autre. Le temps de génération des virus étant beaucoup plus court et la fréquence des phénomènes de variation étant beaucoup plus important, leur évolution est un processus perceptible à notre échelle de temps. De plus, leur génome de taille réduite, leur population très élevée et leurs cascades d'étapes chimiques relativement plus simples à investiguer en font un modèle d'étude très intéressant. L'étude des virus permet alors de comprendre des mécanismes que l'on ne pourrait même pas appréhender sur d'autres êtres vivants.

Il apparaît comme essentiel de rappeler les caractéristiques structurelles et fonctionnelles des virus à ARN, qui sont les acteurs de la théorie des quasi-espèces. Nous pourrions ensuite nous intéresser à la variabilité qui touche ces familles, avant d'étudier les forces évolutives qui interviennent lors de l'évolution.

### **1.1.            Structure et constituants viraux**

#### **1.1.1.            L'acide nucléique viral**

Le génome viral est constitué d'acide ribonucléique (ARN) chez les virus à ARN (175). Il s'agit de molécules d'acides nucléiques, caractérisées par une structure et une séquence nucléotidique propres. Ce génome présente énormément de caractéristiques différentes : il est en règle générale monocaténaire (à simple brin) mais peut être bicaténaire (à double brin, chez les *Reoviridae* par exemple), linéaire ou circulaire. Il peut s'agir d'une molécule simple ou segmentée, c'est-à-dire formée de plusieurs fragments uniques, comme c'est le cas pour les virus Influenza (Influenza Virus, IV).

Sa taille est très réduite comparée à celle des génomes cellulaires et varie de 1,7 kilobases (kb) (pour le virus de l'hépatite D (HDV)) à 32 kb (pour le coronavirus) pour les ARN simple brin et de 7 à 27 kb pour les ARN double brin. Ils sont définis par leur polarité :

- Brin de polarité positive qui sera directement traduit une fois dans la cellule.
- Brin de polarité négative qui sera transcrit en brin positif avant d'être traduit dans la cellule.
- Brin ambisens, associant des gènes encodés positivement ou négativement.

Il contient souvent des gènes chevauchants qui assurent un recouvrement partiel des cadres de lectures afin d'augmenter leur capacité de codage. Ils sont aussi caractérisés par une variabilité génétique élevée (voir partie 1.6.).



### **1.1.2. Les protéines virales**

Les protéines structurales (S) sont des protéines incorporées dans la particule virale, ayant différentes fonctions (175) :

- Protection de l'acide nucléique sous forme d'une capsidie,
- Reconnaissance d'un récepteur à la surface cellulaire,
- Agglutination autour de la capsidie pour former des cores ou nucléoïdes,
- Protéines de la matrice ou du tégument permettant de faire le lien entre la capsidie et l'enveloppe
- Activité enzymatique (par exemple l'ARN polymérase ARN-dépendante)

Les protéines non structurales (NS) ne sont en général pas incorporées dans la particule virale et sont produites pendant le cycle afin de pouvoir être utilisées au cours de ce dernier, avec plusieurs fonctions (175) :

- Activité enzymatique
- Régulation du cycle viral
- Facteur de virulence. Ces protéines ont la capacité de réguler les défenses immunitaires de l'hôte.

### **1.1.3. Les lipides viraux**

Les virus enveloppés sont enfermés dans une trame lipidique d'origine cellulaire, formant une membrane plasmique dans laquelle sont incorporées des protéines virales de surface (spicules, peplomères).

## **1.2. L'architecture virale**

### **1.2.1. Les capsides ou nucléocapsides**

Elles sont composées de protéines structurales afin de protéger l'acide nucléique. Le capsomère est l'unité morphologique de la capsidie et est composé d'un nombre de protéines structurales limité. Elles peuvent être nue ou posséder une enveloppe, ce qui crée des caractères physico-chimiques différents entre ces deux catégories (175).

Elles peuvent être de symétrie hélicoïdale, icosaédrique (dit cubique) ou complexe :

- Icosaédrique : c'est une structure polyédrique composée de 12 sommets, 20 faces triangulaires et 30 arêtes. Le nombre total de capsomères est variable d'un virus à l'autre, mais reste constant pour une même famille de virus, avec un nombre d'arrangement des capsomères unique et indépendant de la présence d'un acide nucléique qui se forme par auto-assemblage.
- Hélicoïdale : le capsomère est formé d'une seule protéine, permettant une structuration sous forme d'hélice. L'auto-assemblage des capsomères nécessite la présence d'un ARN obligatoirement et d'une enveloppe.
- Symétrie complexe : quand la symétrie n'est pas conforme aux deux premières catégories. Dans le cas du poxvirus par exemple, il y a 2 doubles feuilletts lipidiques, avec entre eux un maillage protéique.

## **1.2.2. Les enveloppes virales**

Elles varient beaucoup d'un virus à l'autre. Elles sont issues du bourgeonnement de membranes cellulaires plasmiques ou de membranes internes (appareil de Golgi, du réticulum endoplasmique ou de l'enveloppe nucléaire). Sur celles-ci viennent s'enchaîner des protéines virales et parfois des protéines cellulaires. Une matrice protéique est souvent associée à cette enveloppe (si elle est située en dessous, elle a alors un rôle de stabilisation).

Elles ont un rôle important dans le caractère infectieux du virus : il peut y avoir perte de ce caractère s'il y a retrait de cette enveloppe et ainsi provoquer une fragilité du virus. La sensibilité de l'enveloppe conditionne aussi le traitement.

## **1.3. Le cycle viral**

Les virus sont des entités vraiment particulières : en plus de la taille de leur génome, qui est plus petite que la plupart des organismes cellulaires primitifs, leur organisation génétique est unique. Ce sont des parasites intracellulaires qui se multiplient au sein de la cellule (175).

### **1.3.1. Phase initiale**

Sous la forme particulière décrite précédemment, le virus ne peut pas se répliquer, il ne peut que se fixer à un récepteur disposé sur une cellule hôte potentielle, enclenchant ainsi l'invasion du virus dans cette cellule. Le virus reconnaît le récepteur cellulaire grâce à des domaines protéiques spécifiques. Cela constitue le déterminant important du tropisme viral et donc la spécificité de l'hôte (voir partie 5.2.1.). Beaucoup de virus peuvent utiliser de nombreux récepteurs et co-récepteurs différents dont l'efficacité varie selon le virus, la cellule, les conditions de l'environnement. De plus, la présence à la surface de la cellule n'est pas le seul déterminant et peut nécessiter des molécules additionnelles, car les récepteurs fonctionnels pour un virus sont exprimés dans beaucoup plus de types de tissus et de cellules que ceux qui sont en réalité infectés par un virus. D'autres mécanismes peuvent intervenir, comme un bloc intracellulaire qui peut en effet empêcher la progression de l'infection suite à l'entrée du virus dans la cellule.

Des substitutions d'acides aminés au niveau des sites de reconnaissance des récepteurs des virus peuvent modifier l'usage du récepteur. Ces changements sont connus pour contribuer à la modification de la spécificité d'hôte des virus et ainsi à l'émergence de nouveaux pathogènes viraux. L'importance de ces changements sera discutée plus loin (voir partie 5.2.1).

Suite à cette reconnaissance du récepteur, le virus entre dans la cellule par endocytose et/ou fusion des membranes virus-hôte (voir figure 1). Ce mécanisme est un phénomène actif qui demande de l'énergie et dépend des conditions environnementales (pH, température). Il y a alors relargage du génome viral nu ou de la nucléocapside dans le cytoplasme cellulaire. Dans ce dernier cas, la capsidite doit être capable de se désassembler une fois entrée dans la cellule afin de libérer le génome.

### **1.3.2. Phase de multiplication**

Une fois à l'intérieur, le virus utilise les métabolites de l'hôte pour décoder son génome viral et l'amplifier, lui et ses produits d'expression : c'est une phase d'amplification. Ces produits sont présents sous forme de protéines structurales, en quantité stœchiométriques, et non structurales en quantité plus réduites. Ils sont nécessaires dès le début de la phase de multiplication.

L'ARN viral participe à trois étapes fondamentales (voir partie 1.4. et figure 1) : la synthèse protéique (via la transcription et la traduction), la réplication, et la formation de nouveaux virions. Il possède différents loci responsables de la régulation des étapes de multiplication, des sites de fixations ribosomiaux ou des sites de fixation pour la formation de la capsid virale.

Cela va permettre la formation de nouveaux virions par assemblage de tous ces éléments structuraux, puis l'encapsidation de la « descendance » du génome viral. Certaines interactions du virus avec la cellule peuvent induire une infection non productive, pendant laquelle le génome viral peut être perdu, ou intégré dans l'acide désoxyribonucléique (ADN) cellulaire comme pour les rétrovirus. Dans ces cas-là, les cellules peuvent mourir ou survivre, mais avec une altération des propriétés de développement, conduisant parfois à une transformation (processus oncogénique).

### **1.3.3. Phase de libération**

Les nouveaux virions produits seront ensuite libérés dans le milieu extracellulaire afin d'infecter une nouvelle cellule. Cette sortie se fait après lyse cellulaire pour les virus non enveloppés, et par bourgeonnement à partir des membranes plasmiques ou intracellulaires pour les virus enveloppés (figure 1). Cette libération est manifestée par des effets cytopathiques sur les cellules, des formations de syncytia ou de corps d'inclusion (agrégats de macromolécules à l'intérieur de la cellule). Ces effets cytopathiques pourraient être dus aux effets secondaires de la réplication virale, et non pas seulement aux possibles effets toxiques de certains produits de l'expression virale. Les mécanismes classiques de la mort cellulaire, l'apoptose et la nécrose cellulaire, peuvent tous les deux la provoquer dans cette situation.

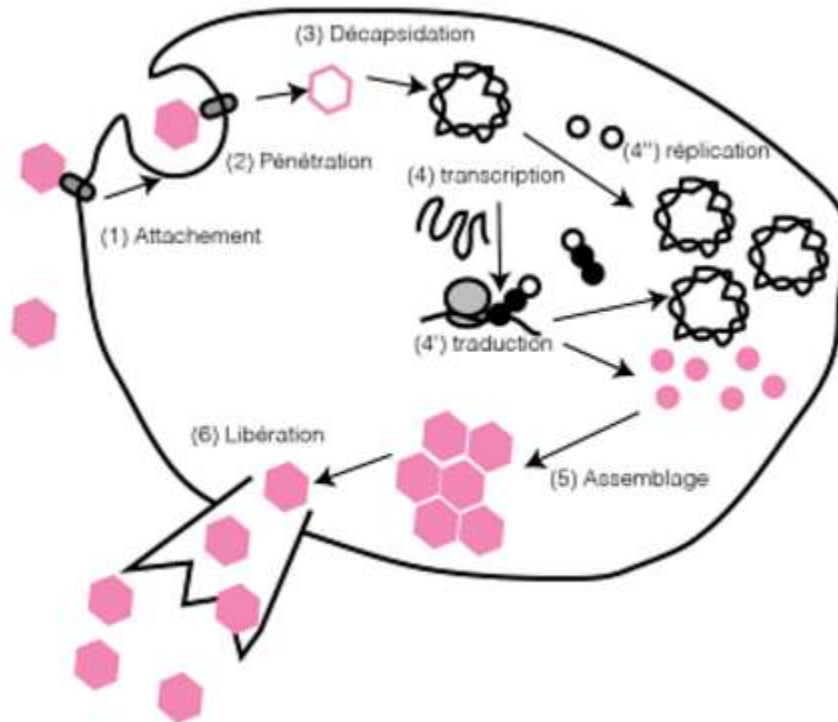
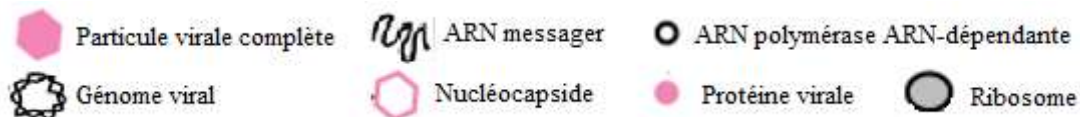


Figure 1: Résumé du cycle viral productif (175)



### 1.3.4. Diversité du cycle viral et de l'expression génétique

Cependant l'organisation du cycle viral reste très flexible. Certains virus associent leur génome au génome de la cellule hôte et entre en dormance comme pour les herpèsvirus. D'autres créent, à partir de leur ARN, un ADN complet en utilisant la transcriptase inverse contenu dans le virion, comme pour les rétrovirus qui s'intègrent ensuite dans le génome cellulaire sous forme de provirus.

Le phénomène d'amplification nécessite aussi une lecture de l'ARN viral dans le sens positif ou négatif, dépendant de la polarité du brin d'ARN. Cette polarité participe aussi à la diversité du cycle (voir partie 1.5.).

La variété de ces stratégies de multiplication virale contraste cependant avec des caractéristiques communes, comme une réplication rapide qui a pour conséquence une production d'un grand nombre d'individus, afin de maintenir le plus grand nombre d'entre eux. L'abondance, la variabilité, mais aussi la stabilité, semblent être des caractéristiques qui ont été fortement sélectionnées. Sans faire de finalisme, il s'agit dans tous les cas d'assurer la transmission du génome à la génération suivante. La maladie provoquée par l'infection virale n'est qu'une conséquence. Sur le court terme, le virus doit être capable de répondre à des contraintes particulières imposées par un individu en évoluant, donc en présentant une certaine variabilité. Cependant, sur le long terme, ce virus doit être suffisamment stable et doit aussi garder la capacité de pouvoir se transmettre de manière efficace à de nouveaux hôtes, sans quoi on se dirigerait vers un cul de sac évolutif.

## **1.4. Dogme central de la biologie moléculaire chez les virus**

### **1.4.1. Première ébauche du dogme central**

Le dogme central de la biologie moléculaire, établi en 1958, par Crick, décrit l'expression de l'information génétique depuis son support jusqu'aux protéines (55). Il énonce que l'information est stockée dans un génome sous forme d'une séquence de nucléotides qui va être décodée. Le premier dogme central décrivait ainsi les différentes étapes de l'expression de l'information génétique :

- La réplication est un mécanisme permettant de copier l'ADN, grâce à l'ADN polymérase.
- L'information nécessaire à un instant  $t$ , afin de remplir une certaine tâche, est copiée suite à un processus appelé la transcription, donnant ainsi des copies de l'information sous forme d'ARN messenger (ARNm).
- La séquence copiée est ensuite traduite en une séquence d'acides aminés grâce aux ribosomes, pendant le phénomène de traduction selon les règles du code génétique.
- Ces acides aminés s'assemblent ensuite sous forme de protéines afin de servir de composants structuraux ou métaboliques.

Ce dogme fonctionne comme un programme informatique : afin qu'il soit exécuté, il faut le logiciel approprié et un système opérateur qui décodent et interprètent les informations et fournissent les commandes et opérations de base. Ce système opérateur complet ainsi que le logiciel peuvent être inclus au sein de la plupart des organismes vivants pour la bonne exécution du programme (64).

### **1.4.2. Extension du dogme central aux virus**

Cependant, si l'on considère cette caractéristique comme nécessaire à la vie, certains organismes ne rentrent pas dans ce modèle : ce sont les virus. Ceux-ci sont dépendants du logiciel et du système opérateur de l'hôte et doivent utiliser les deux composants de celui-ci. Cela en fait des organismes fortement hôte-spécifique et hôte-dépendant. Différents mécanismes permettent cependant l'utilisation du système opérateur et du logiciel et permettent d'intégrer les virus au dogme central, qui a plusieurs fois été affiné (64).

Selon la polarité du brin d'ARN, différentes stratégies sont appliquées au dogme :

- Chez les virus à brin positif, le génome viral est utilisé directement en tant qu'ARNm, et l'ARN viral est répliqué par une ARN-polymérase ARN dépendante.
- Chez des virus à brin négatif, ils sont transcrits par une ARN-polymérase ARN dépendante en différentes molécules d'ARNm et en un brin positif, qui est ensuite utilisé comme modèle pour produire des brins négatifs qui vont être empaquetés dans le virion.
- Chez les rétrovirus, l'ARN viral est transcrit grâce à une enzyme, la transcriptase inverse, en un ADN double brin, qui peut être intégré dans le génome de l'hôte. A partir de cet ADN, différents ARNm et des brins positifs peuvent être générés.

### 1.4.3. Le code génétique

Une étape clé dans l'expression de l'information génétique est cette traduction d'une séquence nucléotidique en une séquence d'acides aminés, sous forme protéique. La traduction fonctionne sur une base de triplets de nucléotides appelés codons ; 3 nucléotides codent donc pour l'un des 20 acides aminés. Le code génétique correspond donc à cette association codon-acide aminé (figure 2).

Ce code a trois propriétés :

- Il s'applique à tous les êtres vivants, à quelques exceptions près. Il est dit universel.
- A un codon est associé un seul et unique acide aminé : il est dit non ambigu.
- A un acide aminé est associé plusieurs codons : il est dit dégénéré. En effet, il existe  $4^3$  codons possibles pour 20 acides aminés connus.

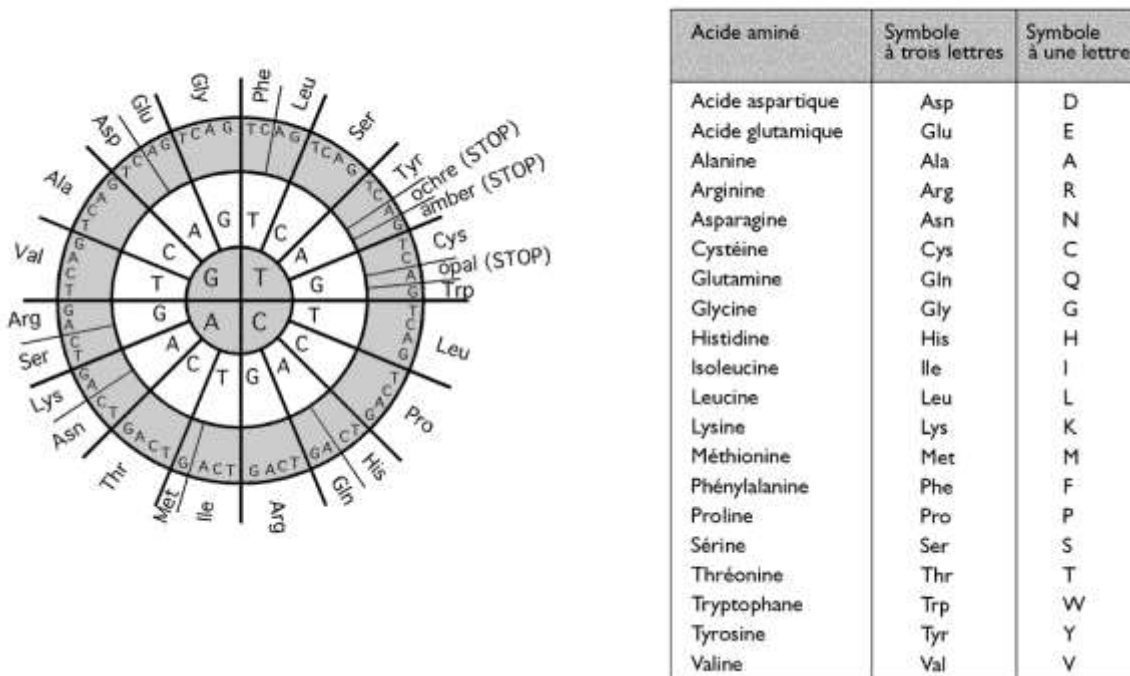


Figure 2 : Code génétique (2')

### 1.5. Familles de virus faisant l'objet de l'étude

Plusieurs familles sont retrouvées dans la plupart des cas d'émergences et de réémergences de virus pathogènes à travers le monde (63). Celles-ci sont favorisées par la plasticité génétique qui caractérise les virus à ARN et par les altérations qui ont lieu dans l'environnement. Elles forment ainsi des familles idéales d'étude de la variabilité génétique virale et de l'évolution chez les populations de virus. Ce sont donc chez ces virus que la théorie des quasi-espèces a été étudiée :

- La famille des *Picornaviridae*, tels que le virus de la fièvre aphteuse (Foot-and-Mouth Disease Virus, FMDV) et le poliovirus (PV) principalement, mais aussi le virus de l'hépatite A (HAV) ou le virus de l'encéphalomyocardite (EncephaloMyoCarditis virus, EMCV).

- La famille des *Retroviridae*, dont le virus de l'immunodéficience humaine de type 1 (Human Immunodeficiency Virus 1, HIV-1).
- La famille des *Arenaviridae* avec le virus de la chorioméningite lymphocytaire (Lymphocytic ChorioMeningitis Virus, LCMV).
- La famille des *Rhabdoviridae* avec le virus de la stomatite vésiculeuse (Vesicular Stomatitis Virus, VSV).
- La famille des *Leviviridae* avec le bactériophage Q-β, l'un des premiers à avoir été étudié.
- D'autres familles telles que les *Hepadnaviridae* (avec le virus de l'hépatite B (HBV)), les *Flaviviridae* (le virus de l'hépatite C (HCV)), les *Paramyxoviridae* (le virus respiratoire syncytial (RSV) et le virus de la rougeole (Measles Virus, MeV)) et les *Orthomyxoviridae* (IV).

## 1.6. Variabilité et instabilité génétique chez les virus à ARN

Le taux de mutation qui apparaît en moyenne lors de la réplication virale chez les virus à ARN est de  $10^{-4}$  substitutions par nucléotide copié (63). Les ARN viraux possédant  $10^4$  nucléotides en moyenne, la descendance obtenue dans une cellule infectée contiendra donc en moyenne 1 mutation par ARN. Par conséquent, la plupart des génomes de la descendance différera de ses modèles parentaux (63, 77).

Nous allons maintenant aborder cette variabilité et exposer ses mécanismes. Cette étape est importante car la variabilité des virus à ARN est à la base de leur dynamique. L'investigation dont elle fait preuve est un pas vers la compréhension des mécanismes qui permettent à ces virus d'échapper à la réponse immunitaire et de résister aux traitements antiviraux mis en place. Cette variabilité va enfin servir de base à la théorie des quasi-espèces. Les virus utilisent tous les mécanismes de variation génétique connus qui agissent sur le génome cellulaire : mutation, recombinaison homologue et non homologue, réassortiment de segments de génomes, duplication de génomes (63).

### 1.6.1. **Mise en évidence précoce de la variabilité et de l'instabilité virale**

La mise en évidence de l'instabilité génétique des virus à ARN a été précoce, car elle a eu lieu bien avant l'apparition des techniques de séquençage moléculaire, chez plusieurs virus : de hautes fréquences de mutants apparus spontanément ou de révertants du type sauvage ont été mises en évidence par Granoff pour le virus de la maladie de Newcastle, par Fields et Joklik avec les réovirus, par Pelhalm avec le virus de la mosaïque de la tomate. L'une des plus importantes reste l'expérience de Valentine chez les bactériophages Qβ (239). On observait l'apparition de révertants, variant viraux ayant subi des réversions et donc un retour vers le type sauvage initial. Des variants mutants étaient aussi obtenus sur des préparations cellulaires avec des virus du type sauvage, alors qu'aucun processus de mutagenèse n'avait été mis en place.

L'importance de ces observations apparaît vraiment lorsque des techniques de séquençage ont pu être appliquées aux bactériophages à ARN. Après que les séquences du premier bactériophage à ARN soient obtenues, Weissmann a rapporté que « un stock de phages apparemment homogènes au niveau phénotypique pourrait contenir de multiples variations à différents sites de l'ARN » (251, 252). A l'époque, le concept de variation génétique est bien établie, mais les notions de mutations, de compétition et de sélection en tant que source de cette variation n'étaient pas encore bien assimilées et n'apparaîtront que progressivement.

### 1.6.2. Enzymes de réplication et génération des mutants

La réplication de l'ARN est une étape de copie d'un brin d'ARN matrice en un brin complémentaire, qui va lui-même servir d'intermédiaire de réplication pour la synthèse de génomes à ARN à intégrer dans de nouvelles particules virales. C'est une étape fondamentale pour le maintien de l'information génétique et pour la propagation de copies portant cette information de manière plus ou moins similaire.

Les acteurs de cette étape servent à la reconnaissance du brin matrice et à l'élongation du brin complémentaire. Il s'agit des ARN polymérases ARN-dépendante chez les virus à ARN, aussi appelée réplisase, et des transcriptases inverses rétrovirales, d'ailleurs particulièrement étudiées lors des dernières décennies (52). Ce sont des enzymes constituées de protéines codées uniquement par le génome viral, comme la transcriptase inverse du HIV-1 (52) ou par un mélange de ces dernières et de protéines codées par l'hôte, comme la réplisase Q $\beta$  (une sous-unité codée par le virus et de trois autres sous-unités codées par l'hôte). Cette dernière a beaucoup servi à caractériser des systèmes de réplication *in vitro* et à comprendre leurs mécanismes grâce aux travaux de Spiegelman puis de Weissman (251, 252).

La comparaison des structures de ces enzymes en 3 dimensions (celles de la transcriptase inverse, de l'ADN polymérase ADN dépendante des *E.coli* et de la réplisase du PV par exemple) a mis en évidence des caractéristiques communes, comme par exemple une forme conservée avec des domaines de catalyse identiques et une forte spécificité pour l'ARN viral (224). Mais cela a aussi démontré l'absence d'activité exonucléase 3'-5' chez la transcriptase inverse et la réplisase à ARN. Elle permet une relecture de la séquence répliquée pour détecter des nucléotides mal insérés, ainsi que d'autres mécanismes de réparation post-répliatif. Ces activités sont présentes chez les polymérases à ADN virale et cellulaire (224). Cette absence a pour conséquence une réplication à la précision limitée chez les virus à ARN et contribue à un fort taux d'erreur. Ce taux caractérise la « fidélité » des enzymes de réplication : si le taux d'erreur est élevé, on parle de fidélité faible de la polymérase, et inversement. Nous allons voir que cette fidélité peut être modulée lors de mutations qui touchent ces enzymes, même en l'absence d'activité 3'-5' exonucléase (voir partie 4.1.1.).

Il s'agit de la base moléculaire de ce qui est appelé la « error-prone replication » dans la littérature, une réplication sujette à l'erreur (71, 224). Elle explique en majeure partie le fort taux de mutation chez les virus à ARN, car ces enzymes sont en première ligne, en reconnaissant et copiant les acides nucléiques (voir partie 2.2.3.).

### 1.6.3. Les mutations, mécanisme de variabilité génétique

Les premières observations de mutations ont été rapportées par De Vries en 1901, au niveau phénotypique, chez l'onagre de Lamarck. De nouvelles formes de ces espèces étaient viables et pouvaient transmettre leur nouveau caractère à leur descendance (3'). Mais à l'époque, on considérait qu'il s'agissait d'un phénomène discontinu, épisodique comparable à un effet saltatoire. L'évolution était alors considérée comme le résultat d'apparitions de variations discontinues, les mutations, et de leur transmission chez des individus d'une population de génération en génération.

La génétique moléculaire actuelle nous a permis de mettre en évidence que c'était inexact : le support de l'information génétique est constamment soumis à des agressions pouvant induire l'apparition de mutations. Ces agressions peuvent être exogènes (radiations...), endogènes (radicaux libres), ou être dues à des erreurs de réplication ou de recombinaisons accidentelles. La plupart du temps, ces agressions sont corrigées par un mécanisme de réparation, mais le taux de correction n'est pas parfait. Les mutations



proviennent donc de cet échappement à la réparation. Ce sont des changements au sein de la séquence nucléotidique codée par l'ARN ou l'ADN, qui vont affecter la séquence d'un gène et parfois altérer sa fonction, mais de manière non systématique (121).

Elles sont localisées à deux échelles : l'une à l'échelle du chromosome, appelée macrolésion et ne s'appliquant donc qu'aux organismes pourvus de cet élément, l'autre à l'échelle du gène, appelée microlésion. Elles sont aussi appelées mutations ponctuelles. Seule cette dernière nous intéresse dans le cadre de l'étude de la variation génétique chez les virus. Les mécanismes de ces mutations à l'échelle moléculaire ne seront pas abordés.

#### 1.6.3.1. Les mutations ponctuelles

Il faut distinguer plusieurs types de mutations (figure 3) :

- Les substitutions consistent en un remplacement d'un nucléotide par un autre et représentent environ 70% des mutations observées (121), ce qui en fait le plus fréquent des remaniements du génome. Elles sont dues à des erreurs de réplication ayant échappé au système de réparation, à des erreurs du dit système ou à un agent exogène ou endogène. On distingue les transitions et les transversions.
  - Transition : remplacement d'une purine par une autre (adénine (A) en guanine (G) et inversement) ou d'une pyrimidine par une autre (cytosine (C) en thymine ou en uracile (U) et inversement).
  - Transversion : remplacement d'une purine par une pyrimidine et inversement.
- Les insertions de nucléotides sont le gain d'un ou plusieurs nucléotides par rapport à la séquence initial. Cet ajout a lieu lors du phénomène de réplication, notamment au niveau de certaines séquences répétées. Lorsqu'elles se produisent dans une portion codante de la séquence nucléotidique, elles décalent le cadre de lecture et conduisent à une protéine anormale.
- Les délétions de nucléotides sont la perte d'un ou plusieurs nucléotides, et ont les mêmes effets que les insertions. Une section d'ARN est ici perdue ou détruite (93). C'est ce mécanisme qui intervient dans la formation de particules « defective interfering » (DI) défectueuses, qui ont la particularité d'interférer avec la réplication des particules virales homologues non mutées (148).

Ces deux derniers événements sont aussi appelés des « indels » en génétique et en bio-informatique.

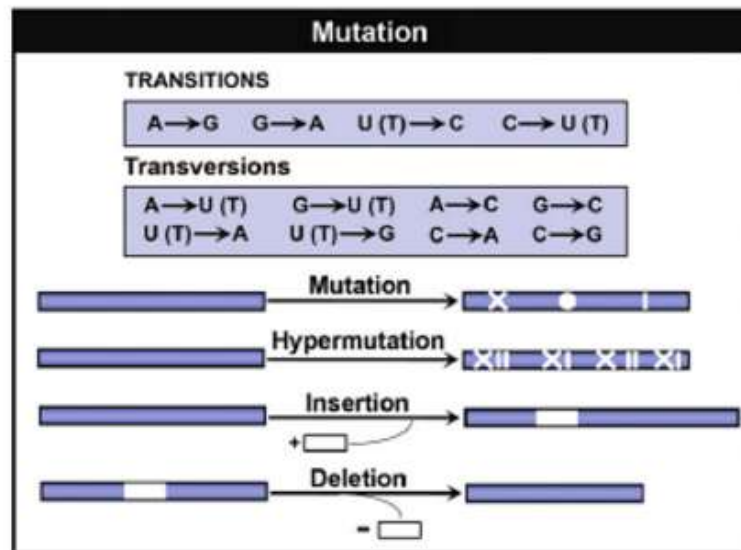


Figure 3 : Résumé des différents types de mutations ponctuelles (72)

- Séquence nucléotidique
- Mutations
- Portion de génome à insérer ou retirer

Ces phénomènes peuvent avoir lieu avec un petit nombre de nucléotides mais peuvent aller jusqu'à plusieurs centaines. Dans ce deuxième cas, il fait suite à une réparation incomplète des lésions subies par le génome ou à des anomalies de recombinaisons ou de réplication. Ces mutations peuvent affecter une portion plus ou moins grande d'ARN et, en fonction de leur localisation dans le génome, peuvent avoir ou non des effets phénotypiques.

### 1.6.3.2. Hypomutations et hypermutations

Le taux de mutation se situe entre  $10^{-3}$  et  $10^{-5}$  substitutions par nucléotide copié lors de la réplication (63, 64). Il se peut cependant que certains sites du génome soient soumis à un taux plus élevé ou plus bas, on parle alors respectivement d'hyper- et d'hypomutation.

Des cas de génomes à ARN hypermutés se retrouvent chez les ARN (DI) des VSV (voir partie 5.4.3.1.) (164), chez des variants du MeV persistant dans le système nerveux central de l'homme, chez le virus parainfluenza humain ou encore des rétrovirus, et chez certains variants échappant aux anticorps neutralisants du RSV (64). Il se peut que ce phénomène puisse contribuer à la persistance du virus sur le long terme, en interrompant la formation de la particule et en minimisant ainsi son exposition à une réponse immunitaire.

Chez le MeV, on suppose aussi que ces hypermutations sont apparues suite au relâchement des contraintes fonctionnelles sur des régions du génome qui servent à la maturation et au caractère infectieux mais pas à la réplication (64). Elles peuvent aussi être dues à l'activité de désaminases cellulaires telles que « l'Apolipoprotein B mRNA and the Editing Complex » (APOBEC) ou à celle de « l'Adenosine Deaminase Acting on double-stranded RNA » (ADAR) (72). Ce sont deux fonctions régulatrices et correctrices cellulaires. Elles participent ainsi à la réponse immunitaire innée de la cellule en provoquant ces hypermutagénèses (226).

Ces mutations sont des outils intéressants, une fois induites par mutagenèse, pour étudier la fonction nucléotidique et protéique.

#### 1.6.3.3. Les expansions répétées

Les motifs nucléotidiques sont des courtes séquences d'ARN qui sont répétées un certain nombre de fois. Par exemple, un motif trinuécléotidique est composé de séquences de 3 bases, et un motif tétranuécléotide est composé de séquences de 4 bases, toutes répétées  $n$  fois.

Les mutations instables, aussi appelées expansions répétées, interviennent au niveau de ces régions du génome (71). En effet, ces motifs ont tendance à voir leur nombre modifié au cours de la réplication. Le plus fréquemment, il s'agira d'une expansion, et très rarement d'une contraction. Le nombre de répétition d'un motif de base sera souvent variable mais toujours en dessous d'un seuil. Tant que le nombre de répétition de ce motif reste inférieur à ce seuil, la transmission à la génération suivante sera toujours stable. Mais si le nombre de répétition est au deçà, deux cas de figure se posent :

- Si le nombre de répétition est modérément supérieur, on parle de « prémutation », normalement sans effet pathogène.
- Si le nombre de répétition est très supérieur (avec une valeur limite fixée), on parle de mutation complète et un effet pathogène apparaît pour le génome.

#### 1.6.3.4. Les duplications

Une partie du génome est anormalement copiée une ou plusieurs fois : la protéine résultante est altérée dans sa fonction (71). Ces événements sont cependant supposés rares chez ces derniers, car ils augmentent la taille du génome, facilitant le franchissement du seuil d'erreur (voir partie 3.3.3. et 112).

#### 1.6.3.5. Conséquences des microlésions

Une séquence nucléotidique située sur un ARN va déterminer l'enchaînement d'acides aminés composant une protéine selon les lois du code génétique (voir partie 1.4.3.). Le premier effet des microlésions va être de modifier cette séquence (121). Ainsi, selon son type, les lésions vont avoir des conséquences différentes. Mais le processus de synthèse protéique et de réplication ne consiste pas qu'en une lecture linéaire de la séquence nucléotidique modèle. Il fait intervenir des mécanismes de régulation de l'expression génétique, l'organisation dynamique de loci, la transcription, la traduction, la maturation protéique, toutes contrôlées par des séquences régulatrices en dehors ou dans la séquence codante. Toute atteinte de l'une de ces étapes peut résulter en effet délétère.

Il peut y avoir perte de fonction (121) : il peut alors s'agir d'une diminution ou de l'abolition de protéine sous sa forme active, produite par la séquence touchée. On peut parler d'effet délétère sur le plan quantitatif (quantité de protéines produites) ou qualitatif (fonctionnalité de la protéine, produite mais non fonctionnelle ou encore produite mais instable et dégradée). Mais il peut aussi s'agir d'un effet délétère au niveau de l'ARNm, avec un effet sur la régulation de la transcription, la réplication, la maturation ou même la stabilité de la séquence nucléotidique, aboutissant généralement à sa destruction.

Il peut y avoir gain de fonction (121) : il peut s'agir de l'acquisition d'une nouvelle fonction, qui sera délétère ou bénéfique pour l'individu. L'effet délétère peut se situer au niveau de la protéine produite par la séquence mutée, ou au niveau de l'ARNm.

#### *1.6.3.5.1. Conséquences d'une substitution en séquence codante*

Une substitution peut changer un codon en un autre, qui code pour un autre acide aminé et provoque un léger changement dans la protéine produite. On appelle cela une mutation de « faux-sens » ou non-synonyme (121). En fonction de la localisation, il peut y avoir perte ou gain de fonction : altération des sites d'interaction protéique, modification du repliement protéique ou de la stabilité protéique, etc. Ces mutations semblent plutôt avoir un impact sur la fonctionnalité protéique.

Une substitution peut changer le codon en un codon qui code pour le même acide aminé et ne provoque pas de changement chez la protéine. Elles sont appelées des mutations silencieuses, synonymes ou isosémantiques. Cela est dû au fait que le code génétique est dégénéré. Ces mutations participent au polymorphisme génétique. S'il n'y a pas modification directe de la séquence nucléotidique, il y a cependant des effets délétères possibles si la séquence est impliquée dans des mécanismes de régulation de l'expression génétique : elles ont ainsi un impact sur le fonctionnement de l'ARN lui-même. Ce point distingue bien les génomes à ARN des génomes cellulaires. Dans certains cas, ces mutations silencieuses peuvent même avoir plus de conséquence sur la structure de l'ARN que les mutations faux-sens (62).

Enfin, elle peut changer le codon en un codon stop et provoque l'arrêt de la traduction, et donc crée une protéine incomplète. Cela peut avoir des sérieuses conséquences sur la protéine qui ne pourra certainement pas fonctionner. Elles sont appelées des mutations de non-sens.

#### *1.6.3.5.2. Conséquences des décalages du cadre de lecture en séquence codante*

Tous ces mécanismes d'insertion et de délétion se basent sur le « frameshift », aussi appelé déphasage. Etant donné que l'ARN est divisé en codons de 3 bases, les insertions et les délétions altèrent le gène de telle manière que le message n'est plus correctement analysé. Ces changements sont appelés des déplacements du cadre de lecture.

Si ces mécanismes insèrent ou retirent un multiple de trois nucléotides, il n'y a pas à proprement parler décalage du cadre de lecture, mais gain ou perte d'acides aminés au sein de la chaîne d'acides aminés de la protéine. L'impact sur la fonction protéique est alors variable en fonction de la localisation.

S'il y a insertion ou délétion d'un non-multiple de trois nucléotides, il y a le plus souvent décalage du cadre de lecture avec apparition prématurée d'un codon stop. Les conséquences sont donc les mêmes que celles des mutations de non-sens.

#### 1.6.3.5.3. *Conséquences des microlésions en région non-codante.*

Les mutations possibles en régions non codantes sont les mêmes que celles en régions codantes. Elles sont essentielles de par leurs séquences régulatrices de l'expression des gènes. Leurs mutations peuvent donc avoir des effets délétères sur la régulation de la transcription, la maturation, la stabilité de l'ARN, en ayant des conséquences directement sur l'ARN ou sur les protéines traduites (conséquences quantitatives ou qualitatives). Cela peut en effet avoir des conséquences sur les structures secondaires, tertiaires et quaternaires de ces molécules.

#### 1.6.3.5.4. *Conséquences sur l'aspect fonctionnel*

Il existe des mutations neutres qui n'ont aucun effet phénotypique sur l'organisme (121). Il ne faut pas confondre le caractère d'une mutation, c'est-à-dire son caractère silencieux, faux-sens et non-sens, et ses conséquences, c'est-à-dire le fait que la mutation soit neutre, létale ou bénéfique. Une substitution silencieuse n'est par exemple pas synonyme de mutation neutre. Cette dernière est en fait un phénomène rare chez les virus (voir partie 4.4.1.).

Les mutations non neutres ont un large panel d'effets sur la valeur adaptative des virus : cela peut aller d'une complète tolérance à la létalité.

La plupart des protéines sont multifonctionnelles. Une mutation non synonyme peut ainsi affecter une ou plusieurs fonctions mais pas certaines : tout dépend de la zone touchée. Les conséquences d'une mutation vont aussi dépendre de sa localisation, car on sait que le génome à ARN présente des gènes chevauchants. Cela peut avoir un impact sur la coévolution de certaines fonctions (voir partie 5.2.4.)

#### 1.6.3.5.5. *Conséquences en fonction du contexte*

Pour compliquer encore l'affaire, il se peut qu'une mutation synonyme soit par exemple neutre ou quasi-neutre dans un environnement, mais puisse contribuer à la sélection dans un autre car les effets phénotypique sur l'ARN ne sont pas les mêmes.

Il faut aussi tenir compte des autres mutations sur le même génome, mais aussi des autres génomes dans l'environnement. Ce dernier point sera plus amplement expliqué lorsque le spectre de mutants sera abordé.

### **1.6.4. Les recombinaisons**

Un ARN recombiné est un génome qui a été généré à partir d'un ou plusieurs segments de génomes appartenant à des génomes parentaux différents (153) (figure 4). La recombinaison peut aussi impliquer différents sites au sein d'un même génome, permettant des insertions et délétions. C'est ce mécanisme qui intervient dans la formation des génomes DI (164). Ces événements sont décrits comme menant à de nouvelles combinaisons de matériel génétique indépendamment des mutations décrites plus haut.

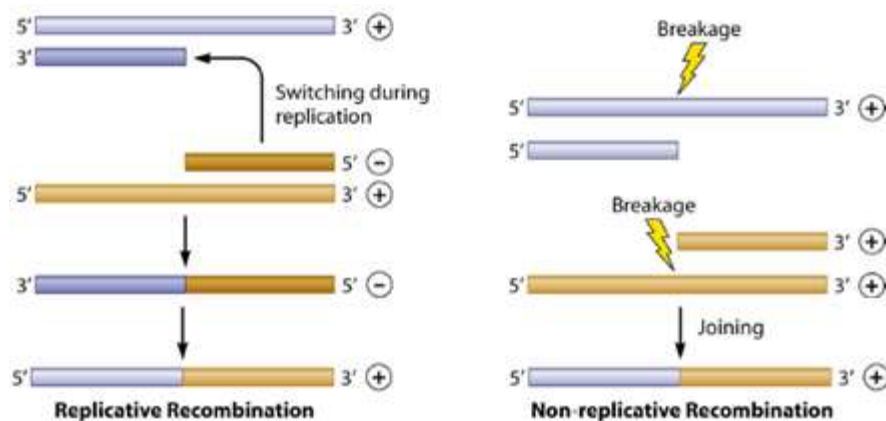
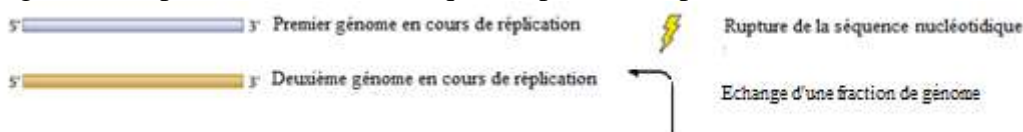


Figure 4 : Représentation schématique simplifiée des phénomènes de recombinaison (71)



Elles peuvent être obtenues par deux mécanismes. La recombinaison des virus à ARN est plus fréquemment le résultat du « copy-choice » ou « template switching model » : c'est la recombinaison répllicative durant la synthèse d'ARN ou d'ADNc (153). Il y a échange d'une fraction de génome entre le brin positif et le brin négatif durant la répllication, au niveau d'un site de la séquence appelé « crossover point » (3, 44, 153). Cet échange est notamment dû à la processivité qui caractérise les enzymes de la répllication chez les virus à ARN : elles réalisent une copie continue de la même molécule qui sert de matrice. La recombinaison est aussi possible hors répllication lorsqu'il y a cassure et réunion des molécules d'ARN et est alors dite non répllicative (153). Il est enfin possible que cette cassure du brin d'ARN ait lieu lors de la répllication, dans ces cas-là on rejoint le modèle du « template switching » (153).

Des recombinaisons homologues, c'est-à-dire d'ARN viraux apparentés, se produisent à des fréquences variables et parfois hautes chez beaucoup de virus à ARN à polarité positive, que ce soit *in vivo* ou sur culture cellulaire (72). Il y a par exemple apparition de recombinants de PV inter-typiques (entre des virus de différents sérotypes) et intra-typiques (entre des virus d'un même sérotype) suite à la répllication de souches vaccinales atténuées (145). La recombinaison est aussi assez active chez les rétrovirus et semble par exemple avoir participé à l'apparition d'un sérotype du HIV-1 (215, 235). Les recombinaisons sont en revanche peu fréquentes chez les virus à ARN négatif non segmenté, mais semblent intervenir dans la formation de DI, comme chez le VSV (153). Cette différence pourrait être due aux propriétés de la répllicase différentes entre un ARN à brin positif et négatif.

Il est aussi possible que cette recombinaison ne soit pas homologue et ait eu lieu entre deux génomes de différentes familles. Cela expliquerait les similarités structurales retrouvées entre des individus des familles des *Flaviviridae* et des *Picornaviridae* lors de l'étude du rôle de la recombinaison dans la spéciation des virus (3, 44). D'autres recombinaisons ont été observées en milieu extracellulaire chez le bactériophage Q $\beta$  (44). Enfin, cela a été décrit entre des génomes de virus à ARN et des ARNs cellulaires (acides ribonucléiques ribosomiaux, de transfert et messagers), formant ainsi un ARN chimérique (64). Il se peut que l'ARN chimérique formé soit défectueux ou viable, comme lors de la formation de particules DI se révélant contenir une séquence presque identique à celle d'un ARNt cellulaire (148).

- Ce mécanisme joue donc un rôle majeur dans l'évolution virale (56, 152, 153) :
- L'avantage sélectif fourni par ces procédés de recombinaison consiste donc en un réarrangement de fragments d'ARN qui leur permet de générer de nouvelles molécules fonctionnelles à partir de fragments préexistants.
    - Il permet, lors de la recombinaison amenant au type sauvage appelée la réversion, de rétablir des fonctions de gènes défectueux. Cela en fait une force conservatrice qui élimine l'accumulation de mutations délétères.
    - Il permet la création de nouveaux gènes, ce qui modifie la virulence. On a alors affaire à l'apparition d'événements évolutifs saltatoires chez des génomes ARN, compatibles avec l'émergence de nouveaux virus pathogènes, comme on le suspecte pour l'apparition du virus de l'encéphalite équine de l'ouest (67).
    - Il est cependant possible que beaucoup d'événements de recombinaison produisent une descendance non viable, notamment à cause de l'épistasie. Ce phénomène se réfère à l'effet combiné des différentes mutations au sein d'un même génome. L'épistasie est positive dans le cas où les mutations se renforcent en regard de la fonction virale, et négative si elles interfèrent entre elles. Ainsi, si la recombinaison associe des mutations qui créent une épistasie négative, la descendance ne sera pas viable et une sélection négative aura lieu.
  - Il permet le transfert de « modules » (ensemble de gènes impliqués dans une fonction, comme la réplication, le mouvement etc.) entre génomes éloignés, de manière plus rapide que lors d'une évolution dite graduelle, par mutation.

#### **1.6.5. Les réarrangements/réassortiments**

Un ARN réassorti est un génome où un ou plusieurs segments de génomes au sein d'un génome multipartite ont été remplacés par le segment correspondant d'un virus différent (figure 5). Dans ce processus, il n'y a pas d'attachement covalent de parties de molécules d'ARN distinguables (64, 71). Ce mécanisme peut faire penser à la reproduction sexuée et aux réassortiments chromosomiques.

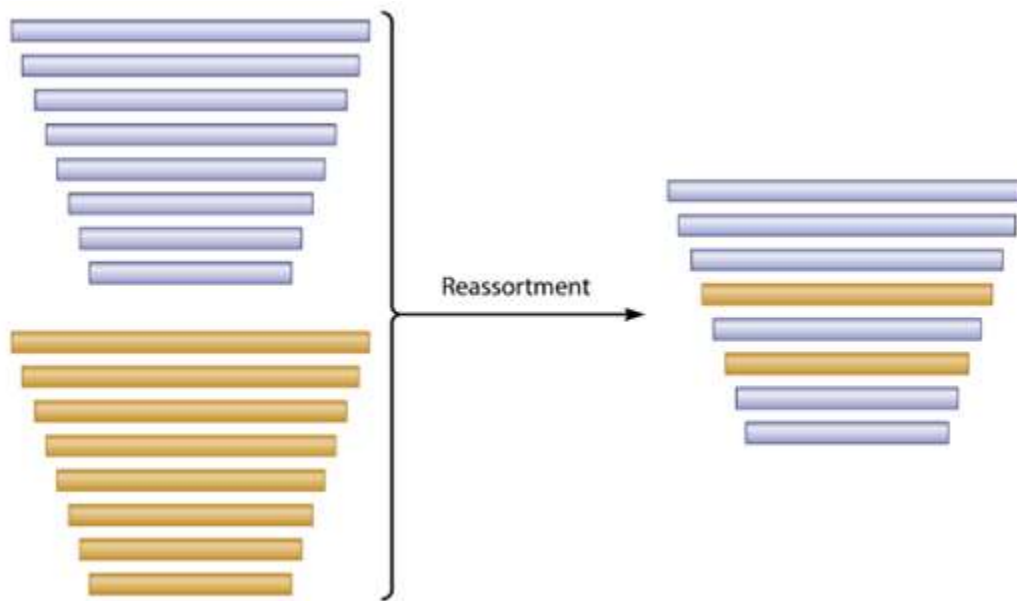
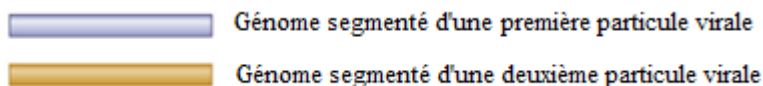


Figure 5 : Représentation schématique simplifiée du mécanisme de réassortiment (71)



Les réarrangements de segments de génome apparaissent chez les virus avec des génomes segmentés dont certains composants peuvent faire preuve d'une autonomie suffisante pour être fonctionnels s'il y a présence de composants viraux voisins (112). Ces mécanismes ont des conséquences épidémiologiques importantes : ils peuvent produire de nouveaux phénotypes et sont par exemple supposés responsables de pandémies nouvelles de grippe (18, 64, 71, 75). Même pour les virus à ARN à génome monopartite, les analyses de séquences tendent à montrer que certains d'entre eux se seraient développés à partir d'autres virus par échange et réarrangement de séquences génomiques (71).

Pour que la recombinaison et le réassortiment se produisent, il faut que les génomes parentaux en coinfection se répliquent dans la même cellule. Cependant certains virus peuvent engendrer un phénomène d'exclusion de super-infection : quand une cellule est engagée dans un processus de réplication pour un virus, elle ne peut pas commencer la réplication pour un autre. Cela limite ainsi les possibilités de recombinaison et de réassortiment qui nécessitent une coinfection synchronisée et simultanée des deux génomes parentaux. La situation se complique encore plus si les segments du génome pour un possible réassortiment sont encapsidés dans différentes particules virales.

### **1.6.6. La segmentation du génome**

Une haute multiplicité de passages infectieux, par exemple chez le FMDV, résulte en une segmentation de génomes (93, 168). Son apparition et les hypothèses concernant ses implications et ses mécanismes seront détaillés (voir partie 4.2.4.). Bien qu'expérimenté sur un seul virus, cette segmentation pourrait être mise en parallèle avec les conséquences de la théorie des quasi-espèces sur l'évolution virale, notamment vis-à-vis de la taille du génome, comme dans le cas de la duplication (figure 6).



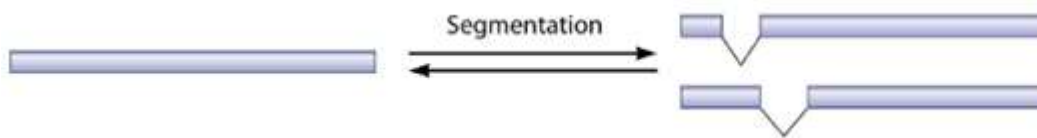
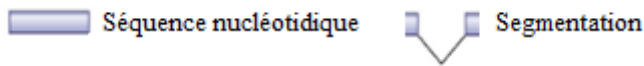


Figure 6 : Représentation schématique simplifiée de la segmentation du génome (71)



### 1.6.7. Conclusion sur la variabilité

Les mutations qui apparaissent durant la réplication semblent être le processus principal qui détermine le répertoire de variants que l'on peut identifier dans les populations virales. Mais il ne faut pas négliger les autres mécanismes présentés ici car ils participent à l'adaptabilité et la variabilité au sein de ces populations. Les mutations sont cependant un prérequis indispensable à la recombinaison et au réassortiment, ce qui les place à juste titre à la tête du raisonnement de la théorie des quasi-espèces.

L'une des questions que soulèvent ces mécanismes est de savoir si les mutations, recombinaisons et réassortiments apparaissent par hasard ou préférentiellement sur certains sites génétiques, ou s'ils sont soumis à un certain type de régulation. Il semblerait que, dans certains cas, comme chez le FMDV *in vitro* et *in vivo*, les mutations se concentrent à des positions particulières sur la séquence génomique, appelés hot-spots (80). Pour la recombinaison, elle a l'air de se concentrer sur les sites « crossover » (153). Enfin, pour le réassortiment, la situation se complique : il s'agirait de déterminer quelle combinaison de segments pourrait être la plus adaptée. Elle est difficile à appréhender, car rien qu'avec 2 particules virales possédant chacune  $n$  segments de génomes, on a  $2^n$  combinaisons de segments possibles. Cependant, s'il semble y avoir des sites préférentiels d'apparition, l'existence de telle ou telle mutation va finalement être conditionnée par une autre force, la sélection.

## 1.7. Conclusion

La principale source de ce fort taux de mutation est donc la faible fidélité des enzymes répliquatives. Mais d'autres événements épisodiques tels que la recombinaison ou le réassortiment peuvent en être la cause. Cependant les séquences retrouvées lors d'échantillonnages au sein de populations ne présentent pas une accumulation aveugle de mutation. La variabilité génétique est l'une des forces évolutives qui va guider la formation de population adaptée à l'environnement, mais beaucoup d'autres interviennent dans le processus de l'évolution de populations virales.

## Partie 2 Principes darwiniens chez des populations virales à ARN

La descendance virale va en moyenne différer d'un nucléotide au sein de sa séquence par cycle de réplication (63). Selon les virus, le rendement moyen des particules virales par cellule infectée est très variable. Mais il est possible de retrouver un grand nombre de particules, allant de  $10^{11}$  à  $10^{12}$  unités infectieuses chez un hôte infecté (63, 71). La population virale obtenue après réplication devrait donc présenter un nombre incroyable de mutations comparé à la séquence nucléotidique d'origine. Cependant, les échantillons de séquences d'ARN présentent un nombre modeste de mutations. Pourquoi ? Qu'est-ce qui prévient l'accumulation de mutations ? Comment expliquer cette stabilité chez les virus alors qu'ils possèdent un taux de mutation si élevé pour certains ? Qu'est-ce qui permet le maintien de l'identité du virus lors de l'infection, des transmissions, etc. ?

Afin de comprendre la dynamique qui s'exerce au sein d'une population virale, il est nécessaire de définir les différents termes et les différentes lois qui s'appliquent dans une étude évolutive. Ces notions évolutives et génétiques sont la base de la génétique des populations et de la théorie synthétique de l'évolution. L'approche intuitive de cette théorie, sans aborder les simulations mathématiques qui y sont rattachées, permettra une première compréhension du comportement évolutif d'une population virale.

### 2.1. Théorie de l'évolution, un premier pas vers l'évolution des populations

Il a fallu attendre le XIX<sup>ème</sup> siècle pour que la notion d'une nature changeante, en constante évolution, soit acceptée et commence à être mise en évidence. La transformation des espèces vivantes se manifeste par des changements de leurs caractères physiques et génétiques au cours des générations. Quels mécanismes interviennent dans ce processus ? Comment apparaissent les nouveaux caractères au sein d'une population ? Pourquoi certains sont-ils sélectionnés par rapport à d'autres ? Et comment fonctionne cette sélection ?

Le concept de l'origine des espèces et de la théorie de l'évolution par la sélection naturelle, formulé pour la première fois par Charles Darwin en 1859, a créé un réel bouleversement pour les biologistes (64). Cette théorie, bien qu'incomplète, a été très vite acceptée par le milieu scientifique, en se basant sur cinq points essentiels :

- Sans limitation, une population s'accroît de façon exponentielle.
- Années après années, les populations demeurent relativement stables, car des facteurs limitants empêchent l'accroissement exponentiel.
- Il existe des variations entre les individus : à chaque génération, les descendants diffèrent de leurs parents et diffèrent entre eux.
- Si une différence s'avère avantageuse face à l'environnement de l'organisme, les individus qui la possèdent ont plus de chances de survivre que ceux de la population qui ne la possèdent pas. Ils ont donc plus de chances de se reproduire et de transmettre à leur tour leurs caractères. C'est le concept de sélection naturelle et de la survie du plus apte.
- Le caractère avantageux a donc tendance à se répandre jusqu'à ce que ceux qui le possèdent surpassent et remplacent ceux qui ne l'ont pas.

La théorie de l'évolution a ainsi apporté un mécanisme essentiel à la compréhension de l'évolution : la sélection naturelle. Cependant de nombreux points restaient à explorer à l'époque. Darwin ne savait pas d'où provenaient les nouveaux caractères qui étaient

sélectionnés. Il introduisit alors le phénomène de mutation. Celui-ci était très peu compris à l'époque car on ne savait rien des lois de la génétique, de son support et de son transfert à la descendance.

Ce n'est qu'à la suite d'expériences et d'observations initiées à la fin du XIX<sup>ème</sup> siècle et se prolongeant durant tout le XX<sup>ème</sup> siècle que ces éléments seront apportés. Une miriade de chercheurs va participer à la découverte de la génétique que l'on connaît aujourd'hui : lois de Mendel, de l'hérédité mendélienne, la théorie chromosomique de l'hérédité, mais aussi la découverte de l'ADN comme support de l'information génétique, puis de l'ARN, et celle de l'impact des mutations sur la variabilité génétique, et enfin l'apport plus récent de la biologie moléculaire. Il en découlera l'émission du code génétique, du dogme central moléculaire, les notions d'homozygotie, hétérozygotie, dominance, récessivité, de phénotype, de génotype et de bien d'autres notions. On peut d'ailleurs décrire brièvement ces deux derniers termes car ils seront utilisés très largement par la suite :

- Le phénotype : caractères observables chez un individu, et qui permettaient d'aborder l'évolution à ses balbutiements. Ils interviennent au niveau clinique, cellulaire, moléculaire.
- Le génotype : il s'agit de l'ensemble des gènes qu'un individu possède en un ou plusieurs loci qu'il convient de définir. Il est caractérisé par une séquence précise de nucléotides.

## **2.2. La théorie synthétique de l'évolution**

Dans la nature, les populations sont avant tout composées d'organismes génétiquement différents. La génétique des populations a pour objectif d'évaluer l'importance, le maintien ou l'évolution de la diversité au sein de cette population, en étudiant la fréquence des gènes et des génotypes retrouvés chez les populations d'individus et les facteurs, appelés forces évolutives, susceptibles de modifier ces fréquences au cours de générations suivantes. Historiquement il s'agit d'une synthèse entre les lois de la génétique et les applications de la théorie de l'évolution et de la sélection naturelle, appliquée à des populations d'individus. Avec l'apport supplémentaire de la biologie moléculaire, ces études ont pu déboucher sur la théorie synthétique de l'évolution, dite Néodarwinisme.

Elle permet d'appréhender les différents paramètres qui vont intervenir dans l'évolution génétique au sein d'une population virale. Pour cela, il est nécessaire de présenter les différentes forces évolutives applicables dans le domaine de la virologie évolutive, au fur et à mesure que les virus se répliquent chez l'hôte.

### **2.2.1. Loi de Hardy-Weinberg**

Pour supporter la génétique des populations, une loi est proposée en 1908, celle de Hardy-Weinberg : « Dans une population de dimension infinie, où les unions se font au hasard (on parle alors de panmixie), où il n'existe ni migration, ni sélection contre un phénotype particulier, et où le taux de mutations est constant, les proportions des différents génotypes restent constantes d'une génération à l'autre» (213).

Cependant, le nombre et l'importance des conditions posent cette loi comme un modèle théorique non applicable, ce qui n'empêche pas les études pratiques d'être en accord avec elle. Certaines conditions sont en fait applicables, mais sur une courte échelle de temps.

Elle a tout de même le mérite de mettre en évidence les facteurs qui influencent les fréquences génétiques, qui serviront de forces évolutives : la mutation et son taux, ou plus largement la variation génétique, la taille de la population, la sélection, la dérive génétique et

les migrations. Ces paramètres, liés aux conditions que l'on retrouve dans le modèle théorique, peuvent modifier la composition génétique d'une population. Il faudra cependant considérer qu'il en existe d'autres telles que la recombinaison, les effets fondateurs, ou encore la compétition.

A l'échelle de l'évolution, ce modèle n'est donc pas conforme à la réalité : il faut considérer que la mutation, la sélection, la dérive génétique, les migrations, voire les écarts à la panmixie, combinent leurs effets. Cette combinaison a pour originalité le fait de combiner des facteurs déterministes et stochastiques.

### **2.2.2. Caractéristiques des facteurs**

Un facteur est dit déterministe s'il y a possibilité de définir l'état final de la composition génétique et la vitesse exacte du processus. Cela caractérise la mutation, la sélection et la migration (213). Cependant il faut noter que l'apparition des mutations a lieu de manière aléatoire, même si leurs effets sont bien déterministes.

Un facteur est stochastique quand il y a possibilité de définir l'état final de la composition génétique sous forme d'un ensemble d'états, affectés d'une probabilité (213). Un seul état seulement sera réalisé, sans tenir compte de l'importance de la probabilité qui lui est liée. Cela caractérise la dérive génétique et l'effet fondateur.

### **2.2.3. Variation génétique**

Cette force évolutive intervient en premier, lors de l'infection ou de la transfection du matériel génétique viral dans la cellule-hôte. Les différents types de mécanismes ont déjà été abordés dans la partie 1.6. Ces variations au sein du génome sont l'un des éléments importants de la diversité génétique interindividuelle. Les variations dites « non pathogènes », dont la fréquence est élevée (>1% de la population) participent au polymorphisme génétique et sont elles-mêmes de différents types. L'étude de la variation a mis en évidence plusieurs paramètres en génétique des populations, notamment en ce qui concerne les mutations. On s'attardera sur les paramètres de ces dernières car ils ont été bien caractérisés.

#### **2.2.3.1. Taux de mutation**

Le taux de mutation est le taux d'apparition d'une mutation durant la réplication. Elle quantifie le nombre d'incorporations erronées par site nucléotidique copié et par cycle répliatif, et est exprimée en substitution par nucléotide copié (64). Ce phénomène a lieu indépendamment de la concurrence qu'il peut ensuite y avoir entre les variants au sein d'une population. Il ne prend donc pas en compte le phénomène de sélection, ce qui le distingue de la fréquence de mutation.

Les taux de mutations sont déterminés grâce aux techniques de génétique inverse mise en place par Weissman. Ils sont très variables, de l'ordre de  $10^{-3}$  pour les plus forts, à  $10^{-10}$ , pour les plus faibles. Ces erreurs d'incorporations sont dues aux enzymes intervenant dans la réplication. Pour les ADN polymérase ADN-dépendantes, on l'estime entre  $10^{-10}$  et  $10^{-7}$ , car celles-ci possèdent une fonction de contrôle de lecture et de réparation des erreurs de réplication. Pour les ARN polymérase ARN-dépendantes et les rétrotranscriptases, il est estimé entre  $10^{-3}$  et  $10^{-5}$  (donc  $10^{-4}$  en moyenne), car celles-ci n'ont pas ces fonctions (63, 64). Ainsi, pour chaque copie d'un génome de taille 10 kb, on sera susceptible d'y retrouver une erreur.

Il faut cependant rester prudent avec ces valeurs de taux de mutation, car elles ne peuvent être considérées comme absolues. Même pour un échantillon donné, une position spécifique, une enzyme de copiage définie, un taux de mauvaise incorporation nucléotidique va varier avec la concentration en ion ou la concentration relative des quatre substrats nucléotidiques. Ces dernières vont elles-mêmes varier en fonction du type cellulaire, et leur état métabolique. Même avec des conditions environnementales identiques, on aura une incertitude quant à l'apparition de mauvaise incorporation. Malgré l'indétermination concernant le taux de mutation précis de chaque nucléotide et la nature stochastique de l'apparition des mutations, la génération continue de mutants viraux dans la culture cellulaire et dans des organismes infectés a été amplement documenté et est vu comme une stratégie adaptative clé chez les virus à ARN.

### 2.2.3.2. Fréquence de mutation

Il s'agit de la proportion de mutation sur les séquences d'individus dans la population. Elle est exprimée en substitution par nucléotide copié (64). Elle intègre tous les aspects biochimiques de la relation entre le virus et son hôte, à un instant donné et dans un environnement donné, ainsi que le taux de mutation. C'est une proportion de la population qui sera influencée par la capacité répliquative compétitive des génomes mutants comparé aux génomes parentaux. Elle fait donc intervenir la sélection, contrairement au taux de mutation. Ces deux termes ne doivent pas être confondus, même si la distinction est parfois difficile à faire dans la littérature. En effet, un mutant peut être généré à un haut taux mais sa fréquence pourrait devenir très basse si la mutation diminue la capacité répliquative relative de la molécule qui la porte.

Selon l'exemple apporté par la figure 7, un nucléotide A sur un brin matrice est répliqué, et devrait normalement donner U. Cependant une mauvaise incorporation de nucléotide a lieu et on obtient un C ou un A, une fois sur 10000 lorsque la copie est réalisée. On a ici un taux de mutation de  $10^{-4}$  substitutions par site nucléotidique copié et par cycle répliquatif. Le brin complémentaire obtenu contenant C est bien plus efficace que le brin contenant A en regard de la fixation d'un facteur protéique P requis pour d'autres répliquations. On obtient ainsi une plus haute fréquence de produits obtenus avec G qu'avec U. On obtient une plus haute fréquence de la mutation de A en G.

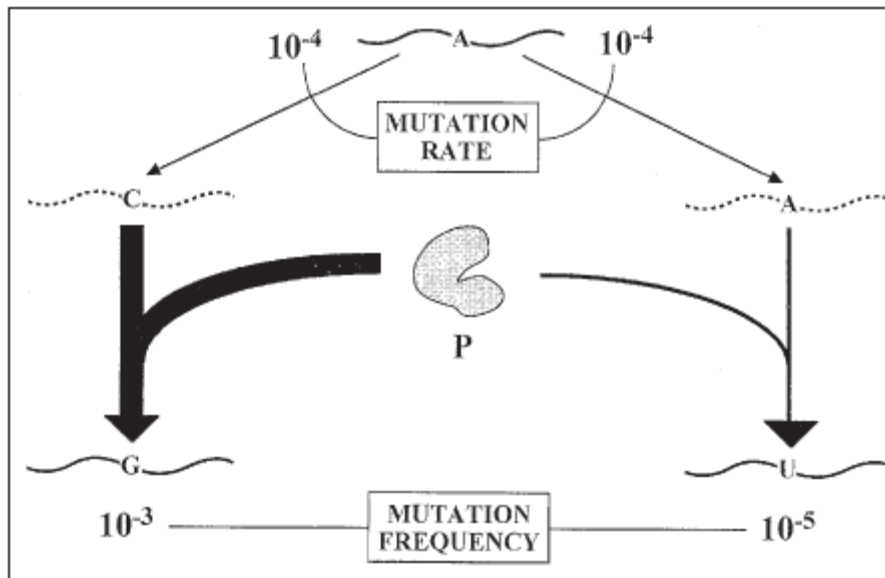
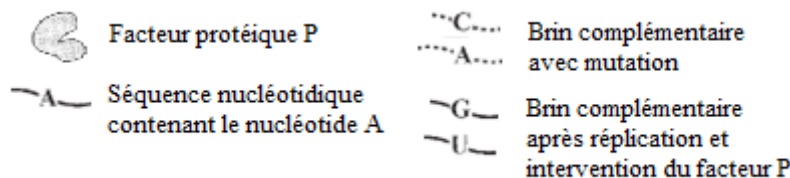


Figure 7 : Représentation schématique de la différence entre fréquence et taux de mutation (64)



Le terme de fréquence est parfois utilisé pour faire référence à la fréquence d'un type spécifique de mutant (résistant à un anticorps, un médicament etc.) dans une population de génomes viraux. La fréquence de mutation d'un virus à ARN est évaluée comme  $10^5$  à  $10^6$  fois plus importante que chez une cellule eucaryote qui l'héberge (64) (Annexe 1).

Un parallèle intéressant peut être fait dans le cadre de la recombinaison. Toute recombinaison ne donne en effet pas forcément un génome viable. Lors de l'étude de possible recombinaison, il faut donc faire la différence entre la recombinaison qui peut avoir lieu, comme pour le taux de mutation, et les recombinaisons que l'on observe, après sélection, comme lorsqu'on étudie la fréquence de mutation.

### 2.2.3.3. Taux d'évolution

On peut ainsi faire intervenir le taux d'évolution : le nombre de mutation qui s'accumulent dans un génome viral en fonction du temps. La différence clé par rapport au taux de mutation et à la fréquence de mutation est que l'on inclut le facteur temps. Il est souvent exprimé en substitution par site et par an. Cela peut se référer à la répllication intra-hôte ou inter-hôte des populations de virus (63) (voir partie 4.3.4.).

### 2.2.4. **La compétition**

Il s'agit de la seconde phase, intervenant lors de l'infection, après la pénétration du matériel génétique viral dans la cellule et après répllication. Elle peut cependant intervenir lors de coinfection sur une même cellule. Il s'agit d'un processus continu dû aux ressources

limitées dans le milieu : le nombre de cellules à disposition lors de l'infection d'un organisme, les structures intracellulaires, les compartiments cellulaires chez des organismes différenciés, les métabolites et autres molécules (72, 213).

Cette force va réaliser une première sélection au sein de la population. Cette sélection est un peu particulière car elle a lieu suite à des contraintes sélectives imposées par la population elle-même, par ces composants. De plus, il y aurait des interactions internes, de coopération ou d'interférence, au sein des composants du spectre de mutants, la plupart du temps à travers leurs produits d'expression. Ces interactions vont être décrites plus en détail dans la partie 5.4.

### **2.2.5. La sélection**

Cela correspond à la survie et la reproduction différentielle des organismes, en fonction de leur capacité à s'adapter aux variations du milieu ou à en tirer avantage (213). Alors que la variation génétique affecte les génotypes, la sélection affecte les phénotypes, produits de l'expression des génomes. La sélection est un facteur déterministe et a lieu lors de l'intervention de contraintes, qui seront dites sélectives, au sein d'une population. Celles-ci peuvent être, dans le cadre de populations virales, une réponse immunitaire, un changement d'hôte, un traitement antiviral, mais aussi l'arrivée d'une autre population virale par exemple. La « survie du plus apte » sélectionne les plus adaptés à ces contraintes, qui pourront laisser un plus grand nombre de descendants. Cela induit donc une modification de la diversité génétique, pour tendre vers une augmentation de la fréquence d'individus les mieux adaptés, une diminution de la fréquence de ceux qui le sont moins, et une transformation de l'espèce.

#### **2.2.5.1. Extension au milieu extracellulaire**

La lignée virale peut envahir les cellules voisines au sein d'une culture *in vitro*, d'un tissu ou d'un organe *in vivo*. Il y a donc non seulement un premier niveau de sélection et de compétition au sein de la cellule, mais un deuxième niveau extracellulaire.

#### **2.2.5.2. Limite floue entre sélections positive et négative**

La sélection positive est un processus par lequel un génotype ou un ensemble de génotypes devient dominant dans une population en évolution, car les traits phénotypiques exprimés par l'individu (ou l'ensemble d'individus) ont apporté un avantage sélectif (71). Une distribution de mutants, dont les composants partagent un trait de sélection, devient alors dominante. Cela provoque un enrichissement d'individus, qui affichent des valeurs d'adaptabilité relativement hautes.

La sélection négative est un processus par lequel un génotype ou un ensemble de génotype est éliminé ou réduit, dans une population en évolution, suite à l'évaluation négative des traits phénotypiques exprimé par l'individu ou l'ensemble d'individus (71). La sélection négative n'est pas nécessaire pour éliminer des sous-ensembles de génomes, mais va avoir pour conséquence leur maintien à basse fréquence. Ces génomes à basse fréquence peuvent être détectables, cela dépendra des procédures d'analyses utilisées (voir partie 3.4.3.).

Il faut noter que la distinction entre la sélection positive et négative peut devenir floue car, dans un processus de compétition, la sélection négative d'un sous-ensemble de génomes pourrait avoir le même résultat que la sélection positive du sous-ensemble de génomes restant et vice-versa.

### 2.2.5.3. Contraintes sélectives

Les virus sont constamment soumis à des contraintes structurelles, fonctionnelles et environnementales. Les deux premières ont même lieu lorsque la réplication se réalise dans un environnement relativement constant comme une culture cellulaire (64).

- Les contraintes inhérentes à la nature du virus et aux conditions nécessaires à la réplication.
- Les contraintes intrinsèques à la cellule-hôte : la membrane cellulaire et les récepteurs reconnus par les particules virales, la composition ionique, en autres ARN, en nucléotides, les facteurs de l'hôte utilisés pendant la réplication virale, les mécanismes d'exclusion de super-infection.
- Les contraintes de l'hôte en tant qu'organisme, c'est-à-dire en tant qu'ensemble complexe de différents tissus et organes composés d'une multitude de cellules. Cela fait intervenir la réponse immunitaire, mais aussi les perturbations métaboliques ou une hétérogénéité phénotypique propre à chaque cellule.
- Les contraintes sélectives qui sont exercées sur l'hôte : la vaccination, l'administration d'antiviraux, les traitements immunosuppresseurs...

L'ensemble de ces contraintes sont autant de mécanismes qui empêchent une infection massive de l'organisme par les virus mais permettent ainsi la survie de l'hôte et du virus sur le long terme.

### 2.2.6. **Effets fondateurs et dérive génétique**

#### 2.2.6.1. Effets fondateurs

Si, suite à un événement aléatoire, une sous-partie de la population est sélectionnée et isolée de la population initiale de taille beaucoup plus grande, elle ne sera pas forcément le reflet de la composition génétique d'origine. Il y aura une variation d'échantillonnage affectant la composition génétique d'une population : c'est un effet fondateur (213).

Il a lieu lors de goulots d'étranglement démographiques, aussi appelé « bottleneck events » dans la littérature. Ces événements ont lieu lors de changements de contraintes dans le milieu, comme la baisse brutale des ressources nutritives du milieu, le passage vers un nouvel hôte, l'augmentation de la réponse immunitaire etc. Ils entraînent une réduction du nombre de variants génétiques composant la population suite à la disparition d'un grand nombre de ces derniers. Cela a pour conséquence une dérive génétique au sein de la population. Quelques générations de dérive génétique peuvent avoir lieu avant de revenir à une taille suffisamment grande pour retrouver une composition stable.

Ces goulots d'étranglement vont nous permettre d'étudier la dynamique et la complexité des populations virales. Expérimentalement, ils peuvent être reproduits grâce à des repiquages en série, en transférant d'un milieu à un autre une population de taille très faible (voir partie 3.1.1.). Ces expériences vont nous permettre de mettre en évidence les différents paramètres pertinents quant à la théorie des quasi-espèces (voir partie 4.3.3.1.).



### 2.2.6.2. Dérive génétique

Elle concerne l'évolution de la fréquence d'un allèle d'un gène considéré, au sein d'une population, mais causée par des phénomènes stochastiques et imprédictibles. La loi de Hardy-Weinberg prédit qu'un équilibre des fréquences alléliques et génotypiques est atteint dans le cas d'une population de taille importante, infinie. Certains allèles ont alors moins de risque de disparaître et vont tendre vers un point de fixation, c'est-à-dire qu'ils vont atteindre une fréquence stable (213). Mais quand l'effectif de la population est limité suite à un effet fondateur, la part des événements stochastiques va être très importante. Il s'instaure alors un processus de dérive génétique. Plus la taille de la population sera réduite, plus cette dérive sera rapide. Elle est caractérisée par :

- Une fluctuation aléatoire des fréquences alléliques d'une génération à une autre. Celle-ci va provoquer la fixation d'un allèle au détriment des autres, ce qui va entraîner une baisse de la diversité génétique sur le long terme.
- Une augmentation récurrente de la consanguinité, malgré la panmixie, en raison de la restriction de l'effectif, ce qui va provoquer aussi une baisse de la diversité génétique. Cet effet n'est pas à prendre en compte chez des modèles d'entités auto-répliquatives.

La dérive génétique aboutit donc à une réduction du polymorphisme génétique des populations, par la perte, pour certains gènes, de tous les allèles sauf un, celui qui est fixé. On peut même imaginer qu'avec le temps, on obtiendra une souche pure par dérive. Mais si on se place sur une espèce divisée en plusieurs populations indépendantes, chacune d'entre elles va dériver de manière indépendante et aléatoire. Le polymorphisme de l'espèce ne sera pas altéré mais seulement réparti : la diversité inter-populationnelle aura augmenté au détriment de la diversité intra-populationnelle. De plus, il faut aussi considérer le fait que l'on s'intéresse, dans la plupart des études, aux allèles d'un gène et non pas à la séquence nucléotidique complète du génome. L'augmentation ou la réduction du polymorphisme génétique concerne donc une portion bien particulière et définie du génome et non pas le génome dans son intégralité.

### 2.2.7. **Les migrations**

Les migrations permettent le mélange de plusieurs populations présentant des différences génétiques plus ou moins importantes. De ce fait, les populations diffèrent moins entre elles après les migrations qu'avant (213).

Si ces migrations sont unidirectionnelles, la composition génétique de la population recevant les migrants se verra transformée et deviendra identique à celle de la population d'origine de ces derniers. A l'échelle de l'histoire génétique de l'humanité, les migrations ont pu jouer un rôle considérable par rapport aux mutations. Cependant, en biologie moléculaire et en virologie, la vitesse de ce processus est généralement négligée dans les expériences. C'est donc la valeur du taux de mutation qui agirait sur l'équilibre et la composition génétique d'une population.

Lors de migrations réciproques, on peut considérer que celles-ci se compensent. Un possible équilibre génétique dépendrait donc exclusivement du taux de mutation, qui est un paramètre faible. L'équilibre est donc atteint à une vitesse très lente, et ne peut jamais être véritablement atteint en pratique car de nouveaux allèles apparaissent et vont faire tendre vers un nouvel équilibre.

Dans le contexte des populations virales, ce paramètre est donc peu abordé. Il introduit cependant les notions de compétitions et d'interactions entre les individus de différentes populations.

### 2.2.8. Bilan des forces évolutives appliquées aux populations virales

On peut considérer qu'une population virale produit, à chaque cycle de réplication, un certain nombre de mutants comparé à la séquence « parentale ». Ces mutants sont soumis à un processus de compétition pour la réplication et de sélection selon leur valeur sélective, ce qui évite une accumulation aveugle de mutations sur un génome nouvellement répliqué. Leur trajectoire évolutive est donc soumise à l'effet conjugué des facteurs déterministes (mutations, sélection) et stochastiques (dérive, effets fondateurs) (figure 8). Ces derniers interviennent sur cette trajectoire si leur effet est plus puissant que celui des facteurs déterministes, c'est-à-dire si la population est considérée comme « petite » suite à un effet fondateur.

Le variant initial (figure 8A) subissant une variation génétique suite à la réplication. Il va donner divers mutants. Ceux-ci vont ensuite subir une compétition et une sélection, ce qui va faire tendre la population vers une distribution de variants les plus adaptés.

Si un effet fondateur (figure 8B) a lieu après la variation (que ce soit avant ou après la compétition), il y a dérive génétique et fixation du variant isolé par l'effet fondateur, qui n'est pas le plus adapté.

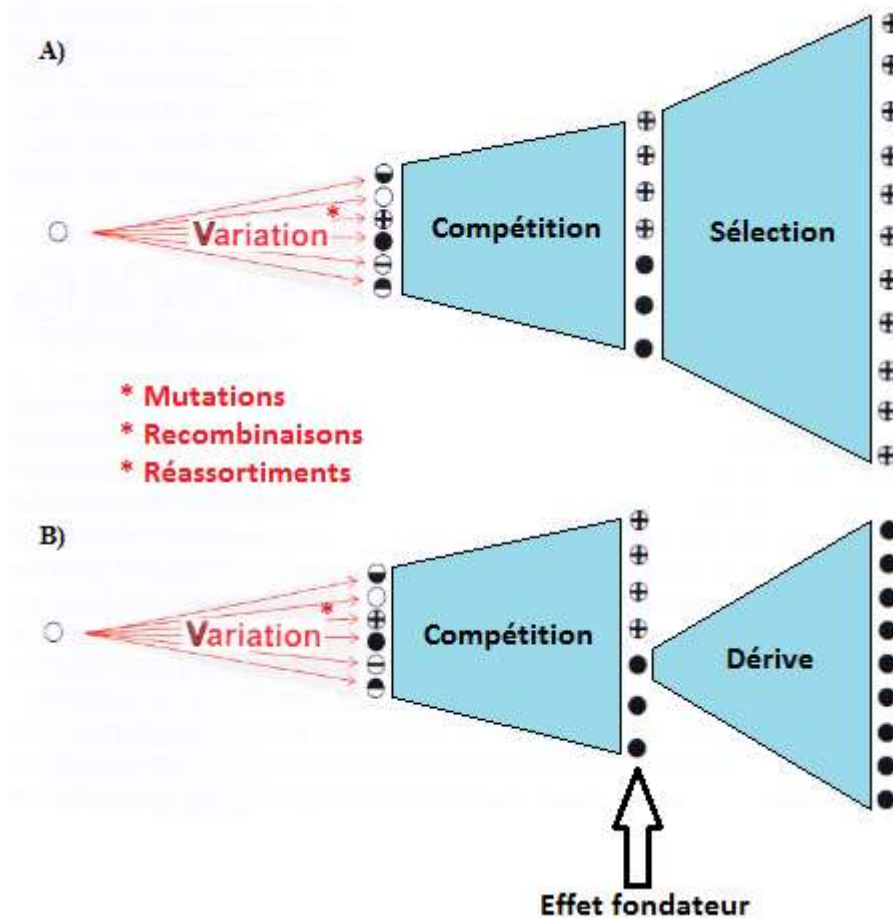


Figure 8 : Schéma bilan simplifié des forces évolutives (d'après 72)

- Variant initial
- ⊕●⊖ Variant mutant

Ce schéma est une vision simplifiée des événements sélectifs pour distinguer les cas où les événements déterministes ou stochastiques agissent de manière prépondérante. Une vision plus réaliste présente beaucoup plus de nuances et surtout de nombreux cycles impliquant ces forces évolutives.

Ce phénomène est plus facilement visualisable en culture cellulaire *in vitro*, car les virus les plus rapides auront fini leur réplication, vont quitter la cellule et en infecter d'autres, que ce soit dans la même ou sur une autre boîte. Ces virus sont sélectionnés pour de multiples raisons : parce qu'ils rentrent rapidement dans la cellule, se décapsident rapidement, traduisent leurs protéines de manière plus efficace, se répliquent plus vite, ou plus généralement se comportent plus efficacement à une étape limitante lors du cycle viral. A tout cela, il faut ajouter une part de facteurs stochastiques, qui interviennent à toutes les étapes, même lors de l'apparition de mutations. Leurs effets seront plus ou moins importants selon la taille de la population virale. On peut prédire que les mutants qui ne sont pas sélectionnés seront éliminés ou maintenus à une fréquence tellement basse qu'ils n'ont pas été détectables pendant longtemps avec les techniques de séquençage classique.

Toutes les espèces voient leur composition génétique soumise à l'effet conjugué de ces facteurs, jusqu'à atteindre un équilibre. L'étude de la valeur sélective et du paysage adaptatif d'une population permet d'apporter une vision synthétique du comportement évolutif et de la composition génétique d'une population soumis aux facteurs déterministes et stochastiques.

## **2.3. La valeur sélective virale**

### **2.3.1. La valeur sélective en génétique des populations**

Le concept de valeur sélective est importé des premières bases de la théorie de l'évolution et intimement connecté à la variation génétique et à des événements évolutifs (213). Il s'agit de la capacité de contribution relative d'individus composant une population à la prochaine génération. La composition génétique d'une population évolue selon les relations d'ordre entre ces valeurs sélectives des génotypes considérés. On retrouve plusieurs termes pour s'y référer : la « fitness », la capacité adaptative, ou encore la valeur adaptative. En clair, si une population a une valeur sélective élevée, elle aura tendance à être sélectionnée. Il s'agit donc d'un paramètre en lien étroit avec les événements évolutifs qui ont lieu au sein d'une population lors de la sélection.

### **2.3.2. La valeur sélective appliquée aux virus**

Dans la littérature, on retrouve le terme de « viral fitness ». Appliquée aux virus, elle indique la capacité de réplication relative des membres d'une population, comprenant même les interactions au sein de la population (63). Elle quantifie donc la capacité de la population virale à produire une descendance à caractère infectieux, quelle que soit sa complexité, dans un environnement donné (192, 250) : on parle de valeur sélective répliquative. On suppose que cette valeur sélective mesurée sera maintenue dans les générations suivantes, à condition que l'environnement reste invariant. Cependant la composition de la population est en continuel changement, à cause des apparitions au hasard des mutations et des incertitudes inhérentes au concept d'évolution même si l'environnement reste constant.

Ce sont les variations de ce paramètre qui vont nous donner un aperçu des modifications de la capacité répliquative au sein d'une population (64). La valeur sélective peut être déterminée pour un virus entier, un génome viral ou des régions de régulation insérée sur un clone infectieux. Etant donné que l'étude de la population portera plutôt sur la comparaison de ces valeurs pour chaque sous-population, on parlera de valeur sélective relative. La quantification par « Reverse transcription Polymerase Chain Reaction » en temps réel (RT-qPCR) permet cependant de déterminer des valeurs sélectives absolues (124, 250).

### **2.3.3. Mesure de la valeur sélective chez les virus**

Plusieurs méthodes sont utilisées pour mesurer la valeur sélective. Il faut pouvoir estimer la charge virale. La méthode utilisée traditionnellement est la quantification des variants viraux formés via les « plaque forming units » (PFU). Les techniques de « Polymerase Chain Reaction » (PCR) quantitative, « enzyme-linked immunosorbent assay » (ELISA), l'étiquetage de sondes fluorescentes ou les nouvelles techniques de séquençage (Next-Generation Sequencing, NGS) estiment aussi la charge virale, avec une grande capacité de traitement et une grande sensibilité (250). La valeur sélective peut ensuite être exprimée en déterminant la pente d'un vecteur exprimant sa valeur.

Pour ce faire, on détermine cette valeur sélective virale relative grâce à des expériences visant à mettre en compétition les populations étudiées. On réalise une co-infection au sein d'un hôte (culture cellulaire, tissu animal ou végétal) avec un mélange contenant des proportions connues de deux compétiteurs : un variant viral à tester, et un variant servant de référence (en règle général, le type sauvage). Pour une plus grande fiabilité, on peut tester parallèlement différentes proportions initiales des deux compétiteurs. Etant donné le caractère stochastique de l'apparition de mutation, il faut aussi procéder à un triplicat de chaque expérience, en parallèle et limiter le nombre de transferts à 4, étant donné que les virus peuvent varier au fur et à mesure des essais, sous un panel bien contrôlé de conditions expérimentales (250).

Après cette première infection, on utilise la descendance produite pour infecter de nouveaux hôtes dans les mêmes conditions. On détermine ensuite la proportion des deux virus relativement à celle retrouvée dans le mélange utilisé initialement. Cela donne un rapport, que l'on passe en fonction logarithmique et que l'on exprime en fonction du nombre de transferts. L'antilogarithme de la pente vecteur trouvée permet ainsi une quantification de la valeur sélective relative entre le virus de référence et le virus testé (73, 110, 192, 216).

Lors de l'étude, les deux compétiteurs peuvent être différenciés par des marqueurs phénotypiques ou génotypiques :

- Les marqueurs phénotypiques sont tous les traits mesurables qui diffèrent entre deux virus : résistance à des anticorps monoclonaux ou à des inhibiteurs, morphologie ou taille des plages de lyses...
- Les marqueurs génotypiques consistent à faire la différence entre le génome de référence et le virus testé grâce à leurs séquences nucléotidiques, ou plus généralement une portion, celle qui concerne le caractère étudié. Elles sont généralement étudiées grâce au séquençage de Sanger ou aux NGS, voire grâce à des amorces spécifiques utilisées pour discriminer les deux génomes en passant par la RT-PCR (19, 209, 221).

On obtient ainsi un graphique du type de la figure 9 :

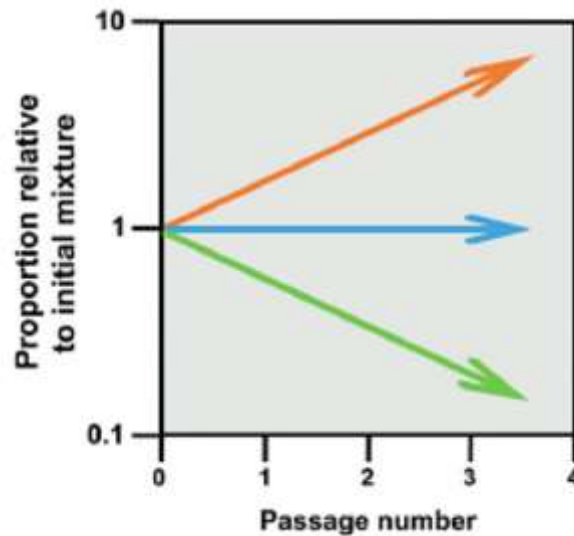


Figure 9 : Représentation schématique de vecteurs de valeur sélective, sur un exemple de 4 transferts (72)

- En abscisse, on retrouve le nombre de passage en série.
- En ordonnée, on retrouve les proportions relatives des différents variants en fonction du mélange initial.

- Variant présentant un avantage sélectif par rapport au variant de référence
- Variant d'égale valeur sélective par rapport au variant de référence
- Variant de valeur sélective inférieure relativement à celle du variant de référence

La nature des environnements utilisés pour réaliser la réplication virale (travaux *in vitro*, *in vivo*, ou *ex vivo*) va avoir un impact qu'il ne faut pas négliger sur la valeur sélective (250). L'utilisation et le développement d'un modèle expérimental précis dépend largement de la nature et de la maîtrise des virus étudiés.

Il faut bien entendu définir précisément les conditions d'expérimentation et les reproduire le plus exactement possible afin d'obtenir des mesures précises et comparatives : la « Multiplicity Of Infection » (MOI), nombre de cellules, nombre et durée des passages, densité cellulaire, temps entre chaque collection des virus, etc. (voir partie 3.2.3. et partie 5.4.2.3.). Si on réalise les expériences *in vivo*, il faut utiliser des organismes déjà bien caractérisés. Toutes ces procédures devraient minimiser des facteurs de confusion qui résultent des changements génétiques et environnementaux.

#### 2.3.4. Intérêts des mesures de la valeur sélective

Il s'agit d'un outil important en génétique virale car elle apporte une représentation de la capacité globale du virus à se répliquer. La capacité répliquative d'un virus détermine la charge virale présente chez des individus infectés, cette dernière exerçant une influence majeure dans la progression de l'infection et sur les manifestations cliniques de la maladie. La valeur sélective est influencée par chaque étape du cycle viral, elle peut donc servir à tester les différences d'efficacité d'étapes spécifiques dans le cycle viral : le mécanisme d'entrée du virus dans la cellule par la comparaison des cinétiques d'internalisation de deux virus mutants, la synthèse de protéines virales, le relargage de virus à partir des cellules, etc. Mais le seul fait de déterminer cette valeur est suffisant pour la compréhension générale du

comportement des virus. L'un des objectifs en biologie moléculaire est de déterminer quelle étape du cycle viral est responsable de l'augmentation de la valeur sélective chez un virus par rapport à un autre.

Les valeurs trouvées sont environnement-dépendantes (63, 124, 172, 192) et il n'y a pas de valeur universelle pour un variant ou une population. La valeur sélective *in vivo* est la résultante de multiples influences dues à un environnement hétérogène fourni par des organismes différenciés. *In vivo*, elle est donc aussi utile pour comparer le potentiel répliatif et pathogène d'un virus.

Les limites liées à ce paramètre, comme tous les autres étudiés pour les quasi-espèces virales, ont un impact non négligeable dans l'interprétation des résultats théoriques et expérimentaux (voir partie 6.3.).

### **2.3.5. La valeur sélective moyenne**

La plupart des études de la valeur sélective répliatif mesure la valeur moyenne de multiples particules virales de populations sur culture cellulaire, *in vitro*, et sur des tissus *in vivo* (250). Des analyses de cette valeur à l'échelle d'une seule particule virale sont cependant possibles, en calculant le nombre de descendants produits par celle-ci lors de l'infection d'une cellule sensible (250).

L'unité de la sélection naturelle est une question toujours remise en cause en biologie moléculaire. Il se peut que la valeur sélective du groupe soit plus haute que la moyenne arithmétique de la valeur de capacité sélective de chaque composant au sein d'une population, on parle dans ces cas-là de sélection d'un groupe en tant que tout. La population formerait ainsi une structure coopérative qui évolue comme une seule unité (125).

## **2.4. Le paysage adaptatif**

Afin de mieux appréhender les possibilités multiples de trajectoires évolutives, notamment lorsque des causes déterministes et aléatoires agissent simultanément, une vision graphique a été développée. C'est un paysage adaptatif (63, 213). Il est utilisé pour représenter les transitions de populations virales d'une forte valeur sélective vers une valeur basse.

Cette description a été apportée pour la première fois par Wright en 1931 (119). Il s'agit d'un paysage semblable à des cartes IGN, parfois imprimées en 2 ou 3 dimensions. Dans sa formulation initiale, la valeur adaptative de la population est représentée en ordonnée, en fonction du gène étudié et ses séquences possibles en abscisse. Cela revient à assigner une valeur adaptative à chaque séquence nucléotidique. Cela donne un paysage avec une topographie spécifique, et le variant est représenté par un point à la surface du paysage ou la population virale est représentée par un nuage de points. Selon les modifications de la séquence du gène étudié, le variant ou la population va parcourir la surface en relief du paysage.

Il s'agit de la version la plus simplifiée car seule une portion du génome, donc d'une séquence nucléotidique, est ainsi représentée. Il est possible de former un paysage adaptatif en fonction du génome entier d'un virus et de l'ensemble des séquences nucléotidiques possibles, l'espace séquence (voir figure 10 et partie 3.3.2.1.). Ces systèmes sont compliqués à analyser mais permettent de mieux visualiser les événements évolutifs chez les populations virales.

Mais certaines représentations expriment en abscisse les conditions de l'environnement dans lequel évolue une population pour une séquence précise (72). Ces

paysages peuvent donc être interprétés de deux façons : selon le déplacement dans l'espace séquence (figure 11) ou selon la modification de l'environnement (figure 12). Ces deux représentations permettent d'appréhender différents aspects des trajectoires évolutives.

#### 2.4.1. Topographie des paysages

Ce paysage est composé de ce que l'on pourrait comparer à des pics, vallées, plaines en fonction d'événements évolutifs. Les « pics adaptatifs » sont des zones de forte valeur sélective, sur lesquelles les populations seront ainsi qualifiées de très adaptées, comme juchées sur un « optimum évolutif » grâce à leur séquence nucléotidique. Mais la moindre mutation modifie la séquence nucléotidique et peut provoquer la « chute » de ces populations vers des « flancs de pic » et des « collines », voire des « vallées adaptatives » de valeur sélective plus modérée (213). Cette topographie et ces termes permettent de mieux appréhender les forces évolutives et apportent une vision imagée de l'évolution d'une population.

La topographie du paysage dépend de l'élément étudié : s'il s'agit de l'espace séquence d'un génome ou des variations de l'environnement, elle aura tendance à être différente (figures 10, 11 et 12).

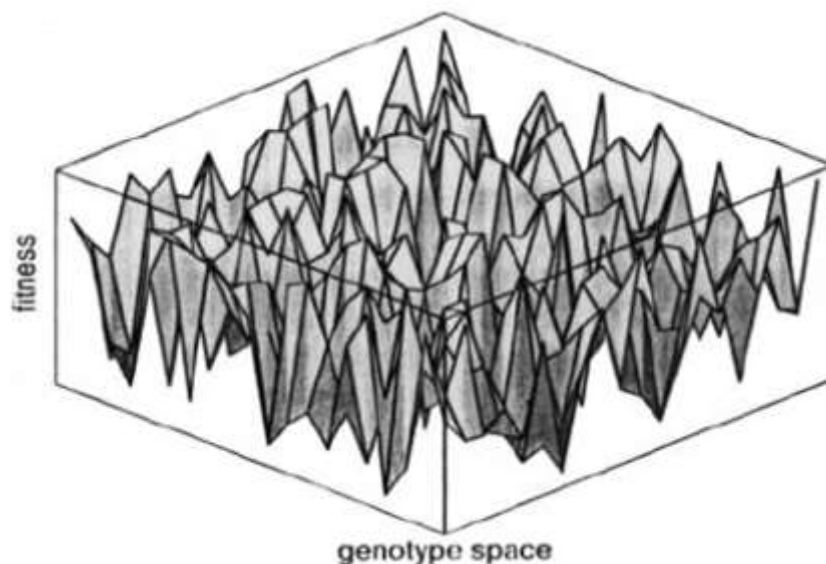


Figure 10 : Exemple d'un paysage adaptatif accidenté en trois dimensions de Kaufmann et Levin (119)

- En abscisse, on retrouve l'espace séquence du génotype (voir partie 3.3.2.1.), c'est-à-dire l'ensemble des séquences nucléotidiques possibles pour le génotype étudié.
- En ordonnée, on retrouve la valeur sélective associée à chacune des séquences du génotype.

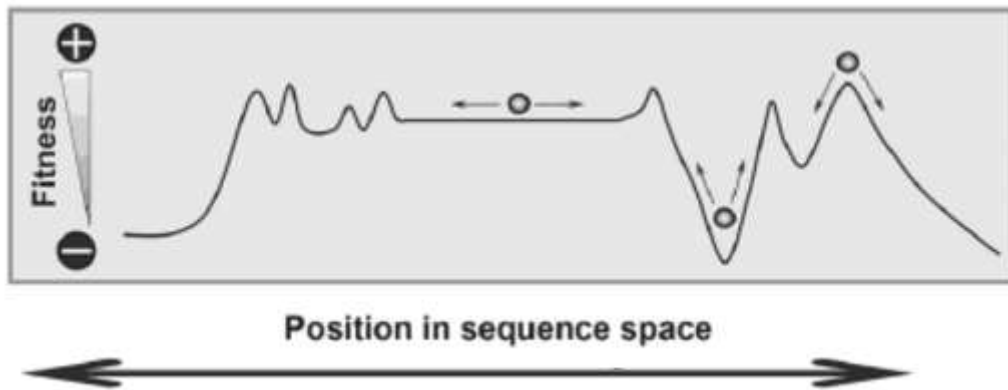


Figure 11 : Paysage adaptatif en 2 dimensions appliqué aux populations virales en fonction de la position dans l'espace séquence (72)

← ● → Variant viral et ses possibles variations ("déplacement" dans le paysage)

- En abscisse, on retrouve l'espace séquence.
- En ordonnée, on retrouve la valeur sélective en fonction de la position dans l'espace séquence

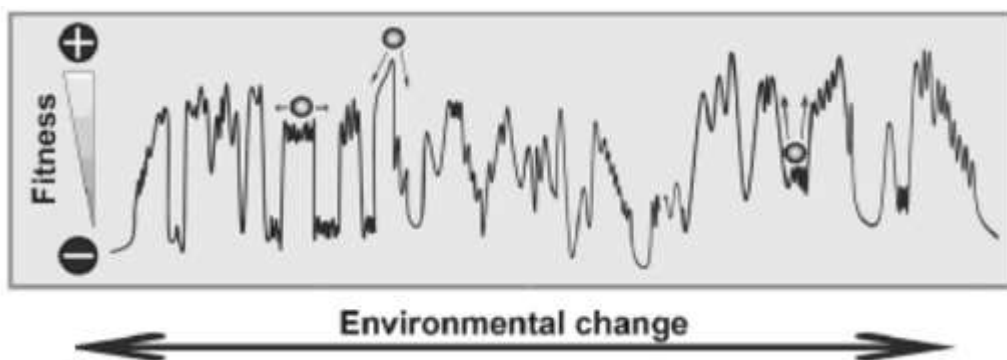


Figure 12 : Paysage adaptatif en 2 dimensions appliqué aux populations virales en fonction du changement d'environnement (72)

← ● → Variant viral et ses possibles variations ("déplacement" dans le paysage)

- On retrouve en abscisse le changement d'environnement.
- On retrouve en ordonnée la valeur sélective en fonction du changement d'environnement.

On considère ici que la valeur adaptative dépend de la position dans l'espace séquence de la population virale (dont la modification dépend de modifications génétiques) (figures 10 et 11), ou dépend d'un changement dans l'environnement si l'on considère une position fixe dans l'espace séquence (figure 12). En fonction de l'endroit où est situé le virus, on peut remarquer que ces variations n'auront pas les mêmes impacts quant à la valeur adaptative.



## 2.4.2. Déplacements au sein des paysages et shifting balance theory

### 2.4.2.1. Selon la séquence nucléotidique

La trajectoire du point au sein du paysage représente le changement de valeur adaptative d'un variant en fonction de la séquence nucléotidique de ce dernier. Elle dépend des facteurs déterministes, la sélection et la mutation, chez les populations de grande taille. L'apparition d'une mutation (une seule substitution par exemple) va modifier la séquence et donc déplacer ce virus sur le paysage. Il est possible que la mutation dirige le virus, s'il est situé sur un pic abrupt, vers une vallée adaptative : la mutation a alors ce qu'on appelle un coût de valeur adaptative (64). Les conséquences sur la valeur sélective, et donc la sélection qui va ensuite avoir lieu, dépendent donc de l'aspect du paysage et de la position de nuage de points correspondant à la population : si celle-ci est située sur un pic abrupt, les conséquences ne seront pas les mêmes que si elle est située sur une colline plus plate (voir partie 4.4.3.).

Mais si la population est de taille réduite, les effets des facteurs stochastiques peuvent alors se rajouter (voir partie 2.2.8.). Sous les effets combinés de la sélection, des mutations et de la dérive génétique, on observera une trajectoire erratique au sein du paysage. Le point peut traverser une vallée entre deux pics adaptatifs et passer d'un flanc de pic à un autre, empêchant parfois la progression vers un pic. Si les effets stochastiques redeviennent minimales, la trajectoire évolutive de la population va la conduire vers un nouveau sommet.

### 2.4.2.2. Selon l'environnement.

En supposant une séquence nucléotidique stable (donc une population à la variabilité limitée), le déplacement au sein du paysage adaptatif peut être abordé comme une modification de l'environnement. Un exemple a été apporté chez le FMDV, qui démontre une grande capacité adaptative chez les cellules « Baby hamster kidney cell (BHK-21) », mais une basse capacité chez les cellules « Chinese Hamster Ovary (CHO) ». Le déplacement dans le paysage se réfère donc à une même séquence transportée vers un autre type cellulaire (des BHK-21 aux CHO) : il s'agit ici d'un changement de cellules (172). La population, qui était sur un pic adaptatif chez les BHK-21, se retrouve dans une vallée adaptative chez les CHO.

Mais ce déplacement peut aussi être vu comme un changement d'environnement intracellulaire : si, par exemple, un génome capable de se répliquer dans une cellule est soumis à un traitement antiviral avec un inhibiteur viral répliatif, sa capacité d'adaptation va fortement diminuer et être transportée d'un pic adaptatif vers une vallée adaptative. A l'inverse, s'il y a acquisition d'une mutation de résistance aux inhibiteurs, la valeur adaptative en présence des inhibiteurs augmentera et le virus va passer d'une vallée à un pic adaptatif (51, 192). Le changement intracellulaire de l'environnement est ici vu comme la présence ou l'absence de médicaments.

### 2.4.2.3. Shifting balance theory

La logique voudrait que les forces sélectives poussent les populations vers un optimum adaptatif, représenté par ces pics de forte valeur sélective, correspondant ainsi à une séquence qui sera la plus adaptée pour un environnement constant. Il s'agira du pic sur le flanc duquel le point ou le nuage de point est situé à un instant  $t$ , tant que les effets stochastiques sont négligeables. Mais que se passe-t-il quand une population se retrouve coincée sur un pic, qui

est la zone la plus haute de la région, mais pas le plus haut pic possible ? Pour répondre à cette question, Wright a développé la « shifting balance theory » (119).

Cette théorie s'expose comme un avantage sélectif, qui permettrait aux populations « bloquées » sur des pics optimums sélectifs de maintenir ses gains sélectifs, c'est-à-dire de se maintenir sur ces pics, tout en étant encore capable d'explorer l'espace alentour comportant une valeur sélective plus faible. Cela se ferait de manière aléatoire, pour atteindre d'autres pics de valeur sélective potentiellement plus élevés. Wright émet l'hypothèse que ces deux buts, que l'on peut penser au préalable opposés, étaient réalisables grâce à une subdivision de la population. Selon lui, beaucoup d'espèces étaient subdivisées en populations plus petites, des sous-populations, qui n'échangeaient que peu de migrants entre eux, mais n'étaient pas complètement isolées. C'est ce genre de raisonnement précoce qui a amené à émettre des hypothèses comme celle de la quasi-espèce (voir partie 3.)

Cette théorie a cependant été énormément décriée, spécialement par Fischer qui a émis l'hypothèse que les facteurs environnementaux provoquaient un changement des paysages trop rapidement pour que les populations puissent se déplacer entre les pics, en tout cas en les étudiant à notre échelle de temps (119).

### **2.4.3. Un paysage mouvant et complexe.**

La séquence nucléotidique et les conditions de l'environnement sont donc deux facteurs qui vont conditionner la valeur adaptative d'un virus, et qui pourraient en fait agir conjointement et simultanément, avec d'autres paramètres. Des facteurs liés à l'individu étudié vont entrer en ligne de mire, comme l'épistasie, ou encore l'apparition stochastique des recombinaisons. Beaucoup de facteurs extérieurs à l'individu sont aussi à prendre en compte : celui de l'hôte comme on l'a déjà vu dans le cas des virus, de l'environnement (pH, ressources, compétition), etc. Ces paramètres évoluent indépendamment ou de manière liée. Dans tous les cas, cela va tendre à compliquer l'analyse et l'utilisation du paysage adaptatif (212). Cette complication peut s'exprimer sous la forme d'un facteur de mouvance, dépendant du temps et modifiant le relief du paysage.

En effet, il faut considérer l'instant auquel est étudiée la population. Une population qui était sur le flanc d'un pic ou à son sommet au temps  $t$ , peut se retrouver au fond d'une vallée ou sur un nouveau pic dans un nouveau paysage adaptatif au temps  $t + \Delta t$ .

Si cet intervalle de temps  $\Delta t$  est faible, la mouvance est imperceptible et le paysage semble présenter un état de stabilité et on se retrouve dans la situation précédemment décrite. Une telle vision du paysage adaptatif rejoint ainsi la vision d'une évolution continue darwinienne et réaliste, cependant remaniée avec l'introduction des facteurs stochastiques et de leur rôle dans la trajectoire évolutive d'une population.

Si cet intervalle de temps  $\Delta t$  est grand (mais ne tend pas vers l'infini), on ne peut plus ignorer la mouvance du paysage. Cette variation de relief dépend alors d'un nombre si important de facteurs, dont certains sont stochastiques, qu'elle est largement imprévisible. Dans ces cas-là, la trajectoire évolutive, déjà soumise à un ensemble complexe de facteurs, se retrouve soumise à une mouvance difficile à appréhender. Il ne faut toutefois pas en conclure que l'évolution est le fruit du hasard. Ce dernier joue un rôle, mais ce sont les effets déterministes qui, par leur nombre et leur diversité et leurs combinaisons, génèrent le hasard sur la longue durée des temps géologiques.

## 2.5. Conclusion

Les mutations vont ainsi introduire des variants, adaptés ou non, dans le milieu à chaque nouvelle génération. La sélection naturelle va augmenter la fréquence des variants au détriment de ceux qui le sont moins. A terme, cela donne naissance à une distribution de variants en équilibre, selon une balance mutation-sélection.

L'impact d'un taux de mutation basal ou de recombinaison varie d'une population virale à une autre, et d'une espèce à ARN à une autre. Des effets non expliqués par les forces évolutives seules sont mis en évidence chez les populations virales. Des résistances aux traitements antiviraux et à la réponse immunitaire apparaissent. Des réémergences de maladies virales sont observées. Des fréquences de sous-populations présentant une faible valeur sélective sont maintenues à des valeurs inattendues. Ces mêmes génomes semblent même interférer dans le fonctionnement d'autres génomes. De nouvelles méthodes de séquençage démontrent une grande variabilité des séquences dans une même population, et les nouvelles méthodes permettent même de mettre en évidence des génomes à des fréquences très basses alors qu'on les croyait éliminées.

Cela indique que des caractéristiques spécifiques à ces virus et à leur hôte pourraient intervenir : le taux de réplication du génome, la taille de la population du virus, la nature des cellules cibles potentielles dans l'organisme de l'hôte, la migration du virus d'un site de l'organisme à un autre, la réponse immunitaire et inflammatoire, et beaucoup d'autres influences. Afin de comprendre les interactions de ces paramètres potentiels, une théorie a été formulée autour d'entités auto-répliquatives pour tenter de construire une dynamique logique qui expliquerait tous ces événements. Il s'agit de la théorie des quasi-espèces virales.

## **Partie 3 Les quasi-espèces virales : expériences historiques, modèles théoriques et expérimentaux**

Manfred Eigen, encouragé par Francis Crick, s'est intéressé en 1971 à un traitement quantitatif de l'évolution de macromolécules biologiques. L'étude initiale avait pour but d'expliquer les événements évolutifs qui pouvaient être impliqués dans les origines de la vie sur Terre. Il s'agit d'une description quantitative et un modèle mathématique essayant d'appréhender l'évolution de réplicons primitifs (76). Pour cela, il a décrit un système caractérisé par un phénomène de réplication à basse fidélité, générant régulièrement des erreurs dans ses copies à partir d'une molécule servant de matrice (76).

### **3.1. Le modèle mathématique d'Eigen**

Il s'est inspiré des premières expériences sur la réplication de l'ARN *in vitro* du bactériophage Q $\beta$  par Spiegelman (72).

#### **3.1.1. Les expériences de Spiegelman et l'origine des passages en série**

Ces expériences reposent sur des "serial passages" (repiquages en série, ou passages en série). Le principe était basique à l'époque, mais a ouvert la voie à de nombreuses expériences permettant l'étude des organismes à ARN *in vitro*, en culture cellulaire ou sur animaux vivants, utilisées aujourd'hui (figure 13).

Le tube 1 (figure 13A) contient le mélange réactionnel complet, avec l'ARN Q $\beta$  à amplifier, la réplicase à ARN, les facteurs de l'hôte, des nucléosides triphosphates et une solution tampon avec du Mg<sup>2+</sup>. Après un temps donné, une aliquote du tube 1 est transféré dans le tube 2 qui contient le mélange réactionnel complet, sans l'ARN du phage. Cela permet d'éviter l'épuisement des ressources du milieu. Le processus est répété le nombre de fois voulu ( $n$  fois), et permet une amplification importante de l'ARN présent dans le tube, de telle manière que les molécules servant de matrice ne soient plus présentes à la fin du passage en série.

Le passage en série de virus se réalise cette fois-ci sur une monocouche de cellules ou sur des cellules en suspension (figure 13B). Une fois que l'infection initiale est complétée sur la plaque 1, le virus est titré et la quantité désirée est utilisée pour infecter les cellules de la plaque 2 (que ce soit une grande quantité de virus, ou un seul comme nous le verrons dans la partie 4.3.).

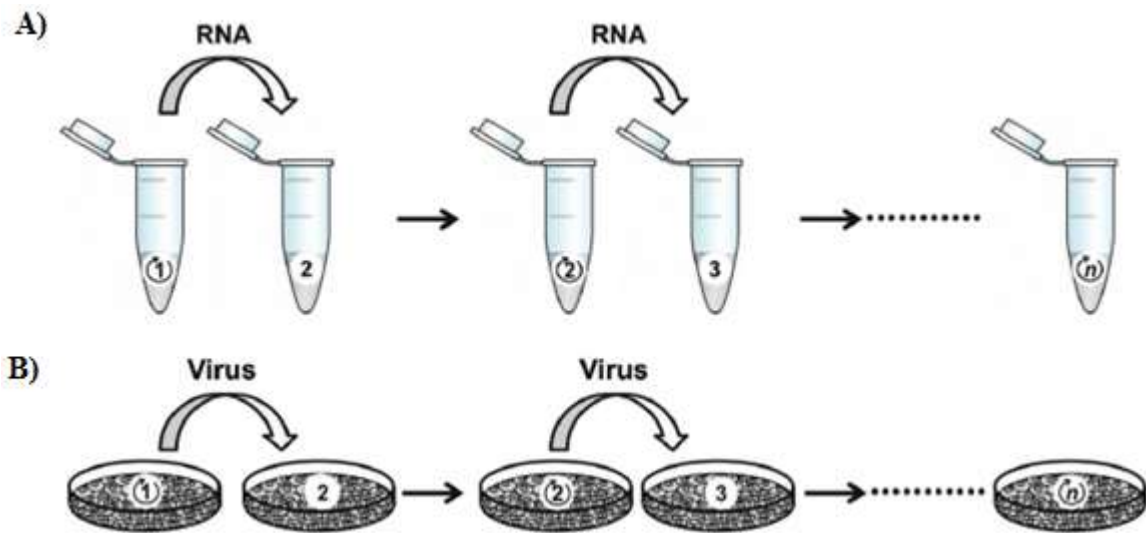




Figure 13 : Description schématique du passage en série (72)

 Passage numéro  $n$  et multiplication virale

 Transfert de l'aliquot ou du virus

Les expériences de Spiegelman n'impliquaient pas de cellules vivantes mais l'amplification de l'ARN des phages Q $\beta$  dans des tubes. Le bactériophage Q $\beta$  est un virus à ARN infectant *E.coli*, qui est capable de se répliquer continuellement en brin négatif et positif d'ARN dans les tubes réunissant les éléments nécessaires à la réplication.

Les résultats de Spiegelman ont permis de mettre en évidence la production, la compétition et la sélection de mutants au sein d'une population de virus à ARN. Ce sont tout simplement des principes darwiniens appliqués à une population virale qui ont ici été mis en évidence, et c'est ce qui a intéressé Eigen.

### 3.1.2. La théorie mathématique d'Eigen et Schuster

#### 3.1.2.1. Approche intuitive

L'intérêt de cette théorie est qu'elle introduit un regard nouveau sur l'impact du taux de mutation et sur la dynamique de l'évolution, comme une extension de la balance mutation-sélection décrite par les principes darwiniens. Les populations de réplicons sont composées d'un ensemble de génotypes présentant cette balance entre mutation et sélection (143). Cet ensemble, si l'on suit les principes darwiniens, devrait être dominé essentiellement par des individus ayant une séquence nucléotidique maîtresse, la plus adaptée, ce que l'on appelle le type sauvage. Selon Eigen, le type sauvage ne serait pas forcément le plus commun. Les populations tendraient ainsi vers un assemblage en équilibre de génotypes mutants et d'un type sauvage plutôt rare (32).

Ces géotypes mutants s'expliqueraient par les caractéristiques du phénomène de réplication. La copie maîtresse d'une molécule auto-répliquative produit des copies présentant des erreurs, versions mutantes d'elle-même, avec une certaine probabilité. La production de ces copies est dépendante d'un facteur de qualité qui décompose le copiage en deux fractions :

- Une fraction du processus de copiage qui tend à donner une copie exacte.
- Une fraction du processus de copiage qui tend à donner une copie présentant des erreurs.

Cela donne un ensemble de variants qui ressemble à la copie maîtresse, mais ne lui est pas identique. Eigen utilise alors le terme de « queue de comète » pour se référer à ces erreurs de copie, qui correspond au terme "spectre de mutants" utilisé aujourd'hui, proposé plus tard (77).

Dans ce contexte, la population ne va plus consister en un ensemble de séquences identiques se répliquant plus ou moins vite mais en une distribution de variants plus ou moins proches génétiquement organisés autour d'une séquence maîtresse. Cette distribution se comporte « quasiment » comme le type sauvage de l'espèce étudiée, c'est une « quasi-espèce ». Ce terme quasi-espèce a aussi été proposé plus tard pour décrire ces entités primitives auto-répliquatives et auto-organisées (77).

### 3.1.2.2. Equations fondamentales de la théorie mathématique

Des années de travaux et d'approfondissement, apportés notamment par Biebricher, Schuster et d'autres, ont fini par apporter deux équations fondamentales de la théorie des quasi-espèces : celle qui décrit la dynamique de la production des copies avec erreurs, et de la relation du seuil d'erreur (71).

La dynamique de la production de copies avec erreurs décrit la génération des mutants au sein de la population, en exprimant la concentration de mutants  $i$  en fonction du temps :

$$\frac{dx_i}{dt} = (A_i Q_i - D_i)x_i + \sum_{k=1, k \neq i}^n W_{ik} x_k - \Phi_i$$

- $x_i(t)$  la concentration de mutants  $i$  en fonction du temps,
- $x_k(t)$  la concentration de mutants  $k$  en fonction du temps,
- $A_i$ , le paramètre pour la réplication de  $i$
- $D_i$ , le paramètre pour la dégradation de  $i$
- $W_i$  le paramètre pour la synthèse susceptible de présenter des erreurs (production de  $i$  en  $k$ )
- *Le facteur  $Q_i$*  la fraction de réplication correcte, donc produisant bien  $i$  lors de la copie de l'échantillon  $i$ ,
- $\Phi_i$  la fonction qui décrit le flux de molécules comme une conséquence de l'implantation du système de réplication-mutation dans un environnement.

L'une de ces conséquences est qu'une simple augmentation du taux d'erreurs réalisées pendant la production de copies peut provoquer un changement fondamental dans la composition des géotypes de la population. Cette composition n'est alors plus à l'équilibre et va devoir évoluer pour y retourner. L'évolution est alors vue comme le déséquilibre de cette composition.

Cela établit aussi qu'au-delà d'un certain taux d'erreurs, un changement bien particulier de la composition de la population intervient. Ce taux est appelé le seuil d'erreur.

Le changement de composition est, quant à lui, appelé erreur catastrophique, et correspond à une perte d'information, une perte des génotypes favorisés, via des mutations délétères fréquentes (32).

La relation du seuil d'erreur est présentée sous cette forme (71) :

$$v < v_{\max} = \frac{\ln \sigma_0}{1-q} = \frac{\ln \sigma_0}{p} \quad \text{and} \quad \bar{p} < \bar{p}_{\max} = \frac{\ln \sigma_0}{v}$$

- $v_{\max}$  la complexité génétique maximale qui peut être maintenue durant la réplication,
- $\sigma_0$  la sélectivité ou la supériorité de la séquence maîtresse par rapport aux autres séquences du spectre de mutants,
- $p$  est la fidélité de copie moyenne du système de réplication
- $1-q$  est le taux d'erreur moyen par site et par réplication

Cette dernière établit une condition nécessaire pour assurer la conservation stable de l'information génétique. Pour une quantité d'information génétique non redondante transmise par un système répliatif, il y a un taux maximal d'erreur qui sera compatible avec le maintien de cette information génétique. L'organisme considéré ne peut pas muter au-delà d'un certain seuil d'erreur, sans quoi ces mutations ne seront plus compatibles avec la survie d'un variant considéré. Ce seuil est différent du seuil d'extinction, qui est une perte complète de la population suite à l'apparition de mutations létales.

Les quasi-espèces telles qu'elles sont décrites par les équations d'Eigen sont des distributions de taille infinie, en équilibre, de génomes liés entre eux, en constante production, dominés par une « séquence maîtresse » et suivant un modèle déterministe (77). La distribution est caractérisée par la fréquence des différents variants. Cette fréquence ne dépend pas seulement de la capacité répliatif du variant considéré mais aussi de la probabilité de produire ce variant à partir d'autres variants (ayant eux-mêmes des fréquences propres). Si une mutation apparaît et offre un caractère avantageux, ses porteurs vont se développer et inévitablement surpasser la distribution précédente. Ces nouveaux éléments font que l'on considère que c'est la quasi-espèce en elle-même qui sert de cible au processus de mutation-sélection.

Parmi tous les modèles présents dans la littérature, l'exemple de Bull (32) se base sur une approche avec seulement quelques génotypes, afin de comprendre les concepts de quasi-espèces et d'erreur catastrophique. Le modèle est extrêmement simplifié mais explique concrètement les concepts apportés par la théorie des quasi-espèces.

### 3.1.3. Simplification du modèle de la quasi-espèce

#### 3.1.3.1. La balance sélection-mutation chez un modèle répliatif présentant des erreurs

Un modèle simple peut être émis, avec deux génotypes,  $A_1$  et  $A_2$ , pour un même génome. Nous allons étudier les conséquences des mutations à un loci particulier de la séquence génomique, dont l'allèle ne peut prendre que deux états : un état muté, donnant ainsi le génotype  $A_2$ , ou le type sauvage, donnant ainsi  $A_1$  (figure 14).

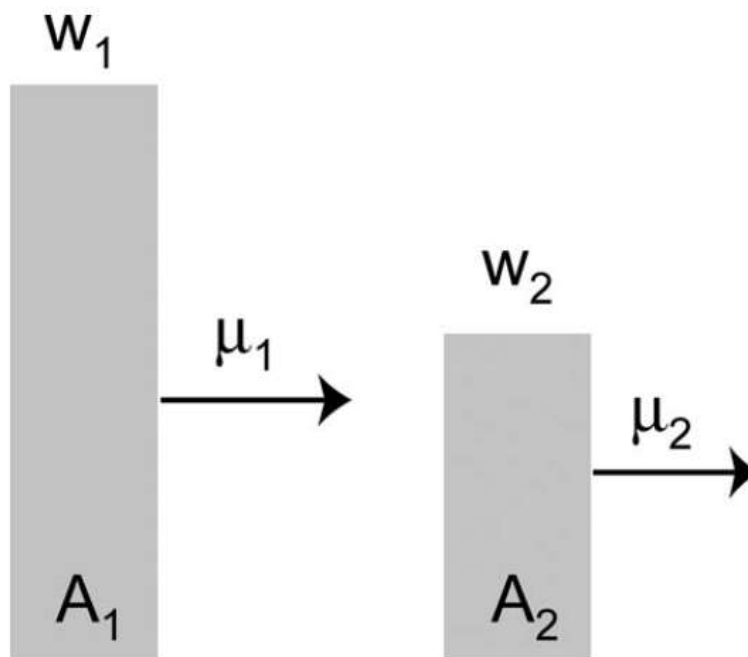


Figure 14 : Modèle d'évolution d'une quasi-espèce avec deux génotypes (32)

- Soient  $w_1$  la valeur sélective du génotype  $A_1$  et  $w_2$  la valeur sélective du génotype  $A_2$ .
- $w_1 > w_2$ . Dans un environnement donné et fixé,  $A_1$  est donc décrit comme plus adapté.
- $\mu_1$  est le reflet du taux de mutation chez  $A_1$ . La descendance de  $A_1$  sera composée, pour une fraction  $\mu_1$ , du génotype  $A_2$ , suite aux mutations à partir de  $A_1$ . Elle sera aussi composée pour une fraction  $1 - \mu_1$  de génotypes  $A_1$ , ce qui correspond à la fidélité de copiage que l'on retrouve lors de la réplication des virus à ARN.
- De même,  $\mu_2$  correspond au taux de mutation qui touche  $A_2$ . Le génotype  $A_2$  reproduit son génotype dans sa descendance avec une fraction de  $1 - \mu_2$ . On obtient une fraction  $\mu_2$  de mutants avec des mutations létales, donc provoquant la disparition des variants.
- On ignore ici le principe de « back mutation » ou mutation retour, c'est-à-dire l'apparition possible de mutation chez les descendants de  $A_2$  qui pourraient donner le génotype  $A_1$ . Cela permet de simplifier le modèle.
- Pour les figures suivantes (15, 16, 17, 18, 19, 20, 21, 22), ces paramètres sont conservés. Si l'un d'eux venait à changer, cela sera expliqué dans la légende.

$A_1$  correspond donc au type sauvage et  $A_2$  au nuage de mutants qui persiste avec le type sauvage, les deux formant ainsi une quasi-espèce, où la balance mutation-sélection est atteinte.  $A_1$  produit constamment des mutants  $A_2$  et la sélection naturelle « purge » constamment  $A_2$  en faveur de  $A_1$  (figure 14). Certains auteurs parlent cependant uniquement de quasi-espèce dans le cas d'une balance où le type sauvage serait rare. Cela viendrait du fait que pour de forts taux de mutations, la distribution des quasi-espèces en tant que tout et non en tant qu'individus est la cible de la sélection.

Le point sur lequel il faut insister est que la quasi-espèce vise l'atteinte de l'équilibre, c'est-à-dire une distribution finale de génotypes dans la population jusqu'à un état stable. Il importe peu que cet équilibre soit atteint ou non (ce sera rarement le cas), car c'est la dynamique des événements qui est intéressante. La mutation et la sélection naturelle dirigent la population vers une distribution en équilibre, quelle que soit la distribution initiale des génotypes. Sur la figure 15, deux populations de compositions complètement différentes au départ vont pouvoir éventuellement atteindre le même équilibre. Cela peut se faire sous réserve d'avoir les mêmes conditions environnementales, les mêmes taux de mutations et en



gardant les mêmes valeurs sélectives pour les deux génotypes, donc sans qu'aucun élément stochastique n'intervienne.

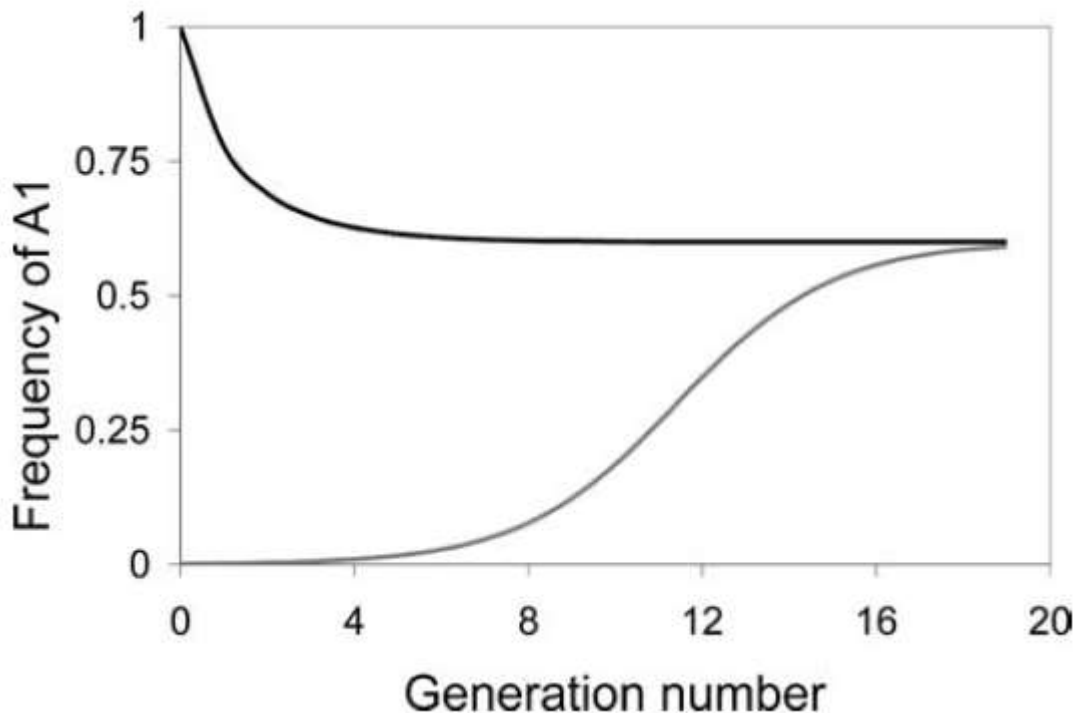


Figure 15 : Approche dynamique de l'équilibre des quasi-espèces (32)

- Exemple avec  $\mu_1 = 0.7$ ,  $\mu_2 = 0.6$ ,  $w_1 = 1.5$ ,  $w_2 = 1.0$ .
- L'axe des abscisses correspond au nombre de générations.
- L'axe des ordonnées correspond à la fréquence du génotype  $A_1$ .
- La partie complètement horizontale des deux courbes représente la possible balance mutation-sélection, qui sera atteinte seulement après un nombre de générations tendant vers l'infini (ce qui est simplifié sur ce schéma).
- Les deux courbes représentent deux points de départ différents : le plus haut avec 100% de génotypes  $A_1$ , le plus bas avec 100% de génotype  $A_2$

Inversement, si le taux de mutation ou la valeur adaptative changent, la distribution d'équilibre va ainsi changer. Les courbes vont converger vers une autre valeur de fréquence du génotype  $A_1$  (en ordonnée sur la figure). Si on diminue par exemple la valeur adaptative  $w_1$  ou la fidélité  $1-\mu_1$  de  $A_1$  par rapport à celles de  $A_2$ , la distribution d'équilibre va se déplacer pour atteindre des valeurs de fréquence plus basses de  $A_1$ .

### 3.1.3.2. Le seuil d'erreur et l'erreur catastrophique

La distribution d'équilibre va parfois être déplacée jusqu'à atteindre une fréquence de 0 pour  $A_1$  (figure 16).  $A_1$  étant le type sauvage, ce point 0 va mener à l'erreur catastrophique, un changement particulier de l'équilibre. Le seuil d'erreur et l'erreur catastrophique sont donc des propriétés découlant de l'équilibre mutation-sélection.

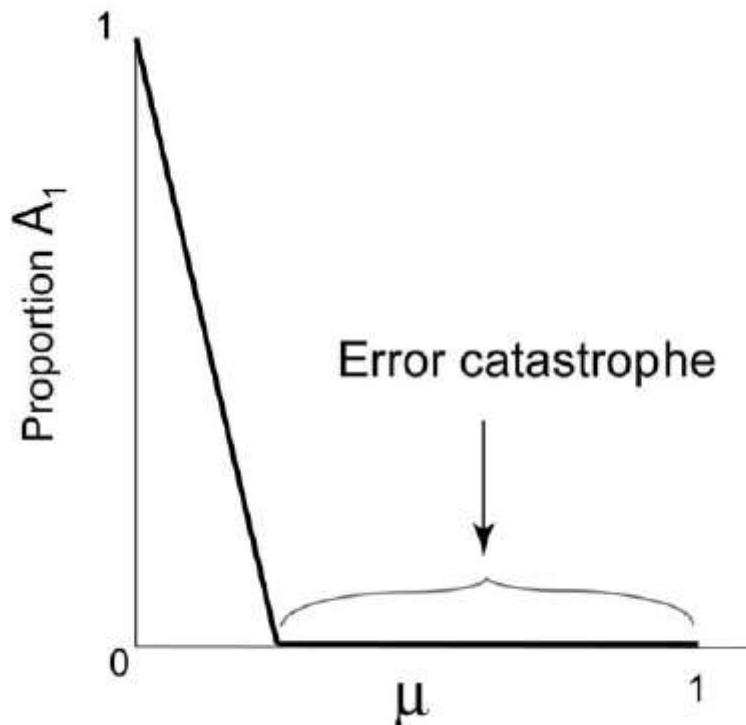


Figure 16 : Proportion de  $A_1$  en fonction du taux de mutation et erreur catastrophique (32)

- L'axe des abscisses décrit le taux de mutation du génotype  $A_1$ .
- L'axe des ordonnées décrit la proportion de génotypes  $A_1$  dans la population.

La fréquence de  $A_1$  diminue lorsque le taux de mutation augmente, jusqu'au seuil d'erreur ( $\mu$  pour lequel la proportion de  $A_1$  équivaut à 0). A un fort taux de mutation dépassant le seuil, seul  $A_2$  est présent, on est dans l'erreur catastrophique pour  $A_1$ . La représentation de la diminution de fréquence est ici linéaire mais peut être concave, convexe, en fonction des valeurs des paramètres.

Pour étudier la relation du seuil d'erreur, il faut considérer les paramètres suivants :

- Le produit  $w_1 \cdot (1 - \mu_1)$  est le nombre de descendants  $A_1$  d'un parent  $A_1$ , que l'on appellera le taux de remplacement du génotype  $A_1$ . Pour plus de simplicité, ce nombre est considéré comme une quantité absolue qui ne change pas avec la densité de la population. Par conséquent, la population peut s'étendre pour un temps infini, dans un espace infini.
- Le produit  $w_1 \cdot \mu_1$  correspond au nombre de descendants  $A_2$  d'un parent  $A_1$ , au taux d'apparition du génotype  $A_2$  à partir de  $A_1$ .
- Le produit  $w_2 \cdot (1 - \mu_2)$  correspond au taux de remplacement du génotype  $A_2$ , selon le même principe que pour  $A_1$ .
- Le produit  $w_2 \cdot \mu_2$  correspond au taux d'apparition d'un autre génotype mutant, avec des mutations létales, à partir du génotype  $A_2$ .

Dans notre exemple,  $w_1 > w_2$  et sont deux valeurs constantes. Si  $\mu_1$  est très petit, de telle sorte que  $1 - \mu_1$  tend vers 1, cela équivaut à décrire l'inégalité :

$$w_1 \cdot (1 - \mu_1) > w_2 \cdot (1 - \mu_2)$$

Si cette inégalité persiste, les deux génotypes resteront en équilibre :  $A_1$  car il a un taux de remplacement plus élevé que  $A_2$ , et  $A_2$  car il apparaît suite aux mutations à partir de  $A_1$ . Mais si le taux de mutation  $\mu_1$  augmente, au point que la fidélité ( $1 - \mu_1$ ) diminue par rapport à

celle de  $A_2$ , alors  $w_1*(1-\mu_1)$  va diminuer. Un point où les deux taux de remplacement sont identiques va finir par apparaître : c'est le seuil d'erreur.

$$w_1*(1-\mu_1) = w_2*(1-\mu_2)$$

Si  $\mu_1$  continue d'augmenter, l'inégalité du taux de remplacement va s'inverser : c'est l'erreur catastrophique.

$$w_1*(1-\mu_1) < w_2*(1-\mu_2)$$

$A_1$  n'est plus maintenu et  $A_2$  a donc un plus haut taux de remplacement que  $A_1$ , ce qui crée un déséquilibre.  $A_2$  est remplacé à partir de génotypes  $A_2$  lors de la réplication, mais  $A_1$  est absent de tous les équilibres atteints car il ne peut être recréé à partir de mutations apparaissant chez  $A_2$ . En exprimant la fréquence de  $A_1$  en fonction du taux de mutation  $\mu_1$ , on voit apparaître clairement le seuil d'erreur et l'erreur catastrophique pour le génotype  $A_1$ .

En exprimant le taux de remplacement de  $A_1$  et de  $A_2$  en fonction du taux de mutation  $\mu_1$  et  $\mu_2$ , on obtient la figure 17 qui permet de visualiser facilement le seuil d'erreur dans le contexte des deux génotypes.

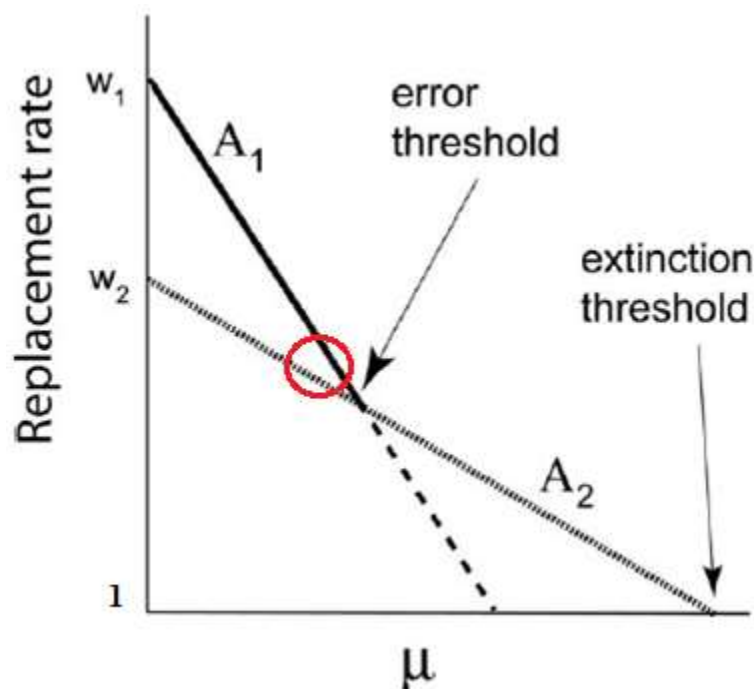


Figure 17 : Le seuil d'erreur (32)

- - - Absence de  $A_1$  au-delà du seuil d'erreur (error threshold)
- Taux de remplacement de  $A_1$  en fonction du taux de mutation
- ..... Taux de remplacement de  $A_2$  en fonction du taux de mutation
- Population décrivant une quasi-espèce

- Le taux de mutation est exprimé en abscisse. Afin que les taux de mutation de chaque génotype puissent être représentés sur le même schéma, on a émis  $\mu_2=\mu$  et  $\mu_1=k\mu$ , avec  $k>1$ .
- Les taux de remplacement de  $A_1$  et  $A_2$  sont exprimés en ordonnée en fonction du taux de mutation  $\mu$ . Si  $w_2*(1-\mu_2)$  passe sous la valeur 1, on atteint le seuil d'extinction (extinction threshold).

Puisque  $w_1 > w_2$ , si  $\mu$  est petit, le taux de remplacement de  $A_1$  est plus grand que celui de  $A_2$ . Cependant cette tendance s'inverse si  $\mu$  augmente, ce qui explique la pente négative. Plus il augmente, plus les taux de remplacements se rapprochent. On dit que le taux de mutation se rapproche du seuil d'erreur. Une fois dépassé, on a affaire à l'erreur catastrophique :  $A_1$  est absent. Malgré la terminologie utilisée, la diminution de la fréquence d'équilibre de  $A_1$  est progressive et non brutale : l'erreur catastrophe n'en représente que le point culminant. Une analogie intéressante est faite avec l'érosion que subissent les montagnes : elles atteignent l'état de plaine graduellement, jusqu'à ne laisser que cette plaine, sans trace de montagne. Juste avant le seuil d'erreur, on peut considérer que la population est dominée par un nuage de mutants  $A_2$  avec une infime fraction de type sauvage  $A_1$ . Cela remplit les critères standards d'une quasi-espèce pour certains auteurs. Cela correspond au cercle rouge de la légende.

Au-delà du seuil, l'effet délétère des mutations est retardé : le génotype  $A_1$ , sensible aux mutations, a été « remplacé » par un génotype  $A_2$  plus robuste aux mutations. L'erreur catastrophique réduit la vitesse à laquelle le taux de remplacement diminue lorsque le taux de mutation augmente. Cela est représenté par la pente moins raide de la fonction du taux de remplacement de  $A_2$ . Si le taux de remplacement diminue encore à cause de l'augmentation du taux de mutation, la population dominée par  $A_2$  court vers l'extinction : elle atteint le seuil d'extinction.

L'erreur catastrophique est donc un changement particulier dans la distribution d'équilibre des génotypes, qui apparaît lorsque le taux de mutation est élevé. Si ce taux est augmenté et passe au dessus de ce seuil, il y aura alors une perte dynamique de  $A_1$  lorsque la population s'approchera d'un nouvel équilibre. Ces propriétés ont plusieurs conséquences :

- Un seuil d'erreur existe parce que des mutations délétères ont un impact plus fort sur certains génotypes que d'autres : dans ce modèle, le seuil est dépassé seulement si  $A_1$  subit une perte mutationnelle potentiellement plus forte que  $A_2$ . En terme mathématique, cela est représenté par le fait que la pente négative de  $A_1$  est plus forte que celle de  $A_2$ . C'est une condition nécessaire, sans quoi ces fonctions ne se croiseront jamais, mais pas suffisante car il peut très bien ne pas exister, en tout cas en terme mathématique (voir le troisième point).
- L'erreur catastrophique n'est pas synonyme d'extinction de la population, seulement de la disparition d'un génotype. Le taux de remplacement d'un génotype autonome en surpasse un autre et par conséquent le remplace dans l'équilibre. Si le taux de mutation est trop grand, la population ne peut donc maintenir un génotype qui a la plus haute valeur sélective,  $A_1$ . Il y aura donc perte du génotype avec la plus forte valeur sélective en faveur de génotypes avec une plus faible valeur sélective mais une plus forte robustesse aux mutations.
- L'extinction a lieu quand le génotype avec la meilleure valeur sélective dans la population ne peut pas se reproduire afin de maintenir une taille de population minimale, donc si  $w_i \cdot (1 - \mu_i) < 1$ , pour tous les génotypes (figure 17).

Ce modèle est le plus simple qui puisse être décrit et implique déjà beaucoup de propriétés différentes. Il n'y a que deux génotypes considérés pour un seul allèle avec des valeurs sélectives stables, avec un seul seuil d'erreur, sans seuil d'extinction précédant un seuil d'erreur, sans mutations retours, une taille de population infinie, un temps infini de réalisation de l'équilibre, sans recombinaison. Nous allons aborder quelques uns de ces points pour montrer qu'à chaque paramètre rajouté, le modèle se complique très fortement.

### 3.1.3.3. Le seuil d'extinction

La différence entre le seuil d'erreur et le seuil d'extinction est présentée sur la figure 18. Pour une population avec un seul génotype, toutes les mutations sont létales, ainsi aucune erreur catastrophique n'est possible car aucun autre génotype n'existe. Mais l'extinction est possible (figure 18 B).

Pour une population avec deux génotypes, l'erreur catastrophique est possible.  $A_2$  ne peut disparaître que suite à une extinction car il n'existe pas d'autres génotypes, il n'y a donc pas de seuil d'erreur au-delà de  $A_2$ . Le génotype avec la plus haute valeur sélective (ici,  $A_1$ ) peut disparaître suite à une erreur catastrophique, car la descendance de  $A_1$  peut « survivre en tant que  $A_2$  », ou suite à une extinction de la population. Deux cas de figure sont possibles (figure 18 A et 18 B) :

On peut voir le seuil d'erreur atteint à un taux de mutation plus bas que le seuil d'extinction (figure 18 A).  $A_1$  sera remplacé par  $A_2$  lors d'une erreur catastrophique avant que le taux de remplacement de chaque génotype ne tombe en dessous de 1.

Le seuil d'erreur est supérieur au seuil d'extinction et est atteint après (figure 18 B). La population totale  $A_1 + A_2$  sera éteinte avant même qu'une véritable erreur catastrophique ne puisse apparaître.

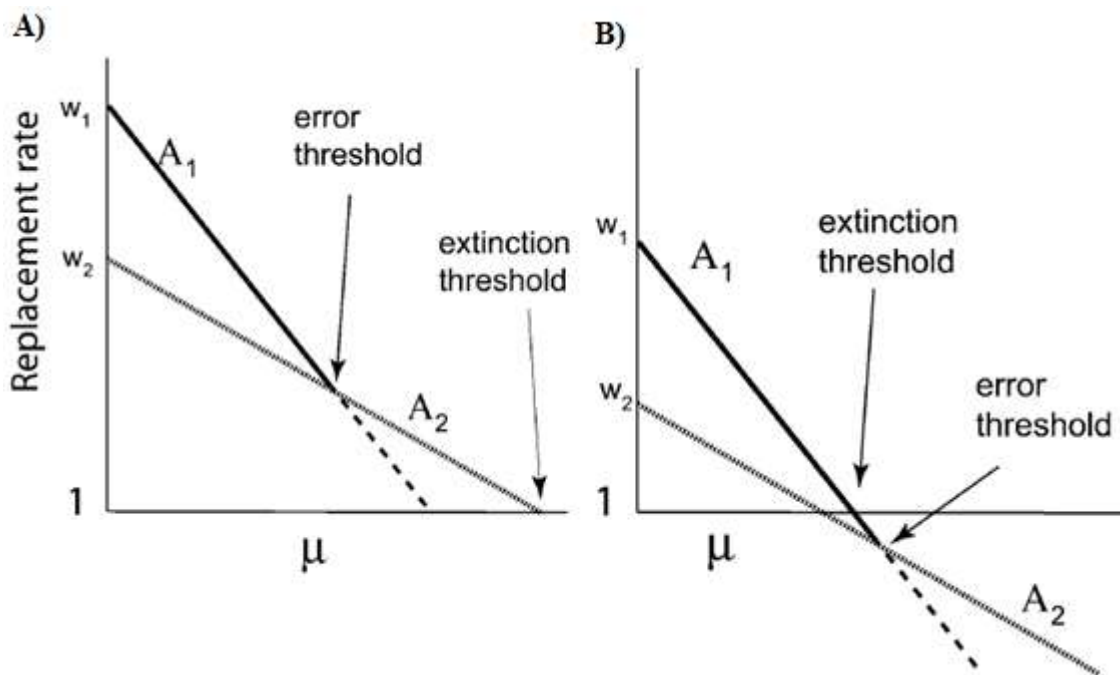


Figure 18 : Différences entre seuil d'erreur et seuil d'extinction (32)

- - - Absence de  $A_1$  au-delà du seuil d'erreur (error threshold)

— Taux de remplacement de  $A_1$  en fonction du taux de mutation

..... Taux de remplacement de  $A_2$  en fonction du taux de mutation

- Le taux de mutation est exprimé en abscisse. Afin que les taux de mutation de chaque génotype puissent être représentés sur le même schéma, on a émis  $\mu_2 = \mu$  et  $\mu_1 = k\mu$ , avec  $k > 1$ .
- Les taux de remplacement de  $A_1$  et  $A_2$  sont exprimés en ordonnée en fonction du taux de mutation  $\mu$ . Si  $w_2 \cdot (1 - \mu_2)$  passe sous la valeur 1, on atteint le seuil d'extinction (extinction threshold).

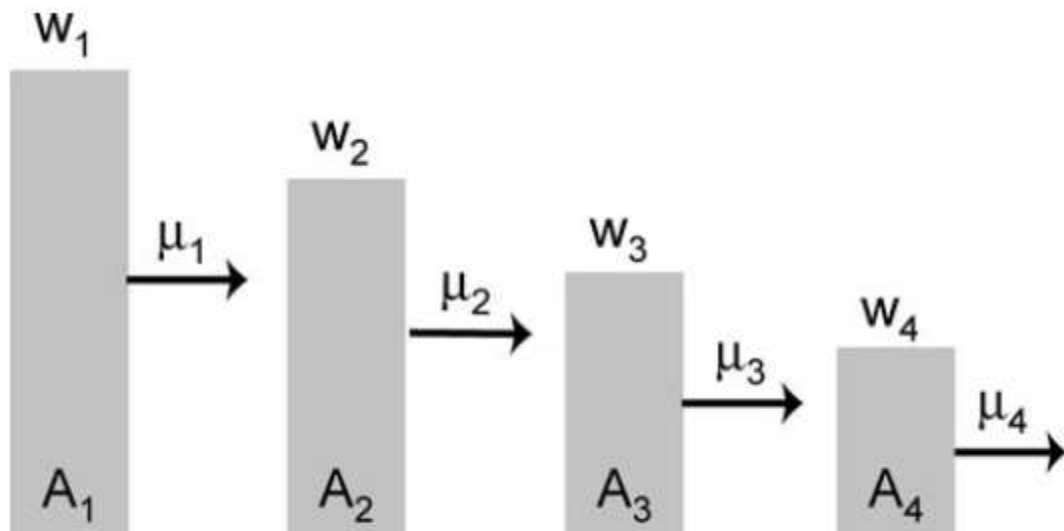
Le seuil d'erreur dépend donc des valeurs de  $\mu_1$ ,  $\mu_2$ ,  $w_1$  et  $w_2$ . Le seuil d'extinction dépend quant à lui du taux de remplacement absolu du génotype. L'interaction entre le seuil d'erreur et le seuil d'extinction complique ainsi le modèle.

### 3.1.3.4. Ajout d'autres génotypes au modèle

En intégrant plusieurs autres génotypes à ce modèle (nommés  $A_i$ ), avec chacun leur valeur adaptative  $w_i$  et leur taux de mutation  $\mu_i$ , on peut supposer que les seuils d'erreur peuvent apparaître par pallier, de manière successive (figure 19). Plus le taux de mutation augmente, plus la population a tendance à passer d'un génotype  $A_i$  au suivant, et ainsi de suite.

Il est aussi possible que l'erreur catastrophique procède à un saut d'un génotype  $A_i$ . Par exemple, si un taux de mutation pour le seuil d'erreur entre  $A_1$  et  $A_2$ , excède le taux de mutation pour le seuil d'erreur entre  $A_2$  et  $A_3$ , alors une transition aura lieu entre  $A_1$  et  $A_3$ . L'étape intermédiaire dans laquelle  $A_1$  disparaît en faveur de  $A_2$  va sauter.

L'interaction entre les erreurs catastrophiques pour plusieurs génotypes, rajoutée aux erreurs d'extinction, peut de même compliquer de manière intéressante le modèle.



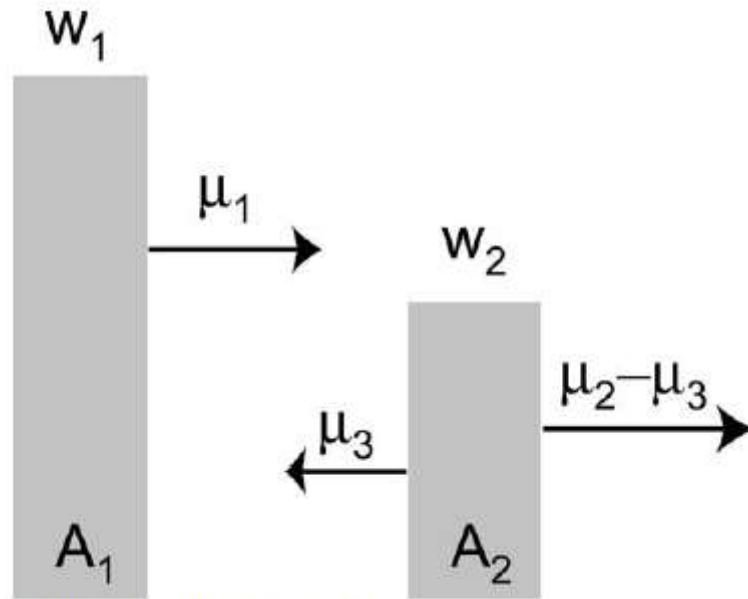
DOI: 10.1371/journal.pcbi.0010061.g005

Figure 19 : Modèle avec plusieurs génotypes (32)

- La légende est la même que pour la figure 14.
- $w_1 > w_2 > w_3 > w_4$  respectivement pour les génotypes  $A_1, A_2, A_3, A_4$ .
- $\mu_1, \mu_2, \mu_3$  et  $\mu_4$  sont respectivement les taux de mutation pour  $A_1, A_2, A_3$  et  $A_4$ .

### 3.1.3.5. Introduction des mutations retours

Des mutations retours impliquent que le génotype  $A_1$  peut réapparaître à partir du génotype  $A_2$ , avec un troisième taux de mutation  $\mu_3$  (figure 20).

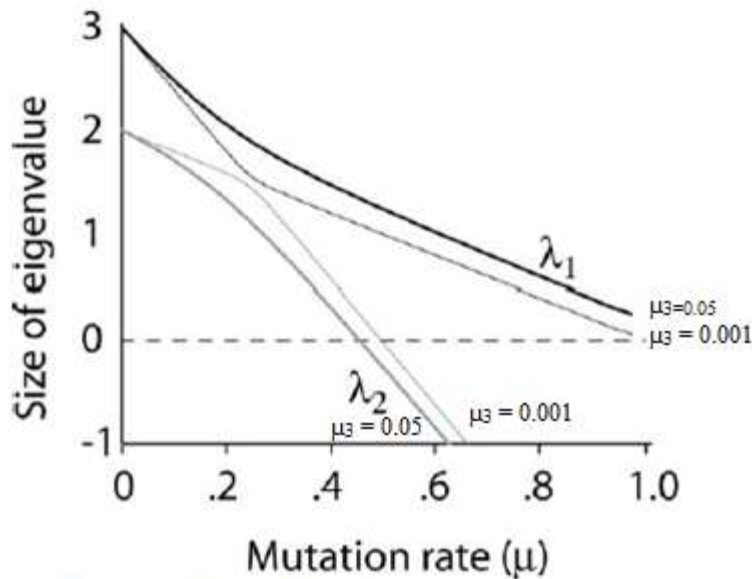


DOI: 10.1371/journal.pcbi.0010061.g006

Figure 20 : Modèle de deux génotypes avec des mutations-retours (32)

- La légende est la même que pour la figure 14.
- On obtient une fraction  $\mu_3$  de génotype
- Le taux de mutation total de  $A_2$  est toujours de  $\mu_2$ , de telle manière que  $\mu_3 < \mu_2$ . La descendance de  $A_2$  sera composée d'une fraction  $\mu_2 - \mu_3$  de mutants avec des mutations létales, pour une fraction  $\mu_3$  du génotype  $A_1$  et pour une fraction  $1 - \mu_2$  du génotype  $A_2$ .

Cette insertion a des conséquences sur l'apparition du seuil d'erreur. Le but n'est pas ici de rentrer dans les détails du calcul qui pourront être vus plus longuement dans l'annexe 2 et impliquent l'utilisation de systèmes matriciels. Dans ce système, ce sont les valeurs absolues  $\lambda_1$  et  $\lambda_2$  qui représentent les taux de remplacements de  $A_1$  et  $A_2$  et intègrent le taux de mutation  $\mu_3$ . L'insertion de ces mutations va modifier la fonction de ces taux de remplacements en fonction du taux de mutation  $\mu$  (figure 21).



DOI: 10.1371/journal.pcbi.0010061.g007

Figure 21 : Valeurs absolues en fonction du taux de mutation (32)

$\mu_3=0.05$  —  $\lambda_1$   
                   —  $\lambda_2$   
 $\mu_3=0.001$  —  $\lambda_1$   
                   —  $\lambda_2$

- Le taux de mutation est exprimé en abscisse. Afin que les taux de mutation de chaque génotype puissent être représentés sur le même schéma, on a émis  $\mu_2=\mu$  et  $\mu_1=k\mu$ , avec  $k>1$ .
- Les taux de remplacement de  $A_1$  et  $A_2$  sont exprimés en ordonnée en fonction du taux de mutation  $\mu$ .

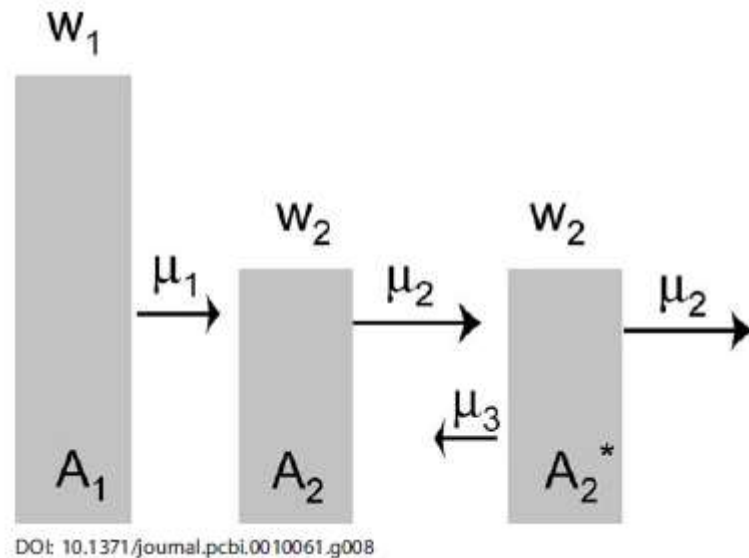
Comme on peut le voir, si  $\mu_3$  tend vers 0, les deux valeurs tendent à se rapprocher et à ressembler à la figure 20, la discontinuité mathématique ne sera alors pas véritablement visible. Le système étudié se rapprochera de manière criante du modèle décrit auparavant. Mais ce n'est pas le cas si un grand nombre de ces mutations a lieu. Un taux  $\mu_3$ , même proche de 0, a pour conséquence qu'aucun seuil d'erreur stricte n'est atteint (figure 21). Cela vaut bien entendu pour des valeurs  $w_1$ ,  $w_2$ ,  $\mu_1$  et  $\mu_2$  fixes. Le modèle s'en retrouve encore modifié et compliqué en fonction d'une autre valeur.

### 3.1.3.6. Les réseaux neutres

Un même phénotype peut être produit par des génomes multiples et liés par des mutations. On parle de mutations neutres dans ces cas-là, qui ne vont pas avoir d'impact sur le phénotype mais bien sur la séquence nucléotidique. On parle de réduction du taux de mutation phénotypique, mais sans altération du taux de mutation génétique correspondant. L'ensemble de ces génotypes qui peuvent caractériser le même phénotype sont appelés un réseau neutre.



Nous pouvons donc encore compliquer le modèle : on suppose que cette fois-ci deux génotypes  $A_2$  et  $A_2^*$  donnent le même phénotype, avec deux taux de mutations identiques  $\mu_2$  et deux valeurs adaptatives  $w_2$  identiques (figure 22). Cela va permettre d'étudier l'impact de la neutralité sur le modèle en exposant un réseau neutre simple.



DOI: 10.1371/journal.pcbi.0010061.g008

Figure 22 : Modèle avec 3 génotypes et un réseau neutre simple de deux génotypes (32)

- La légende est la même que pour la figure 14.
- Les génotypes  $A_2$  et  $A_2^*$  forment un réseau neutre, avec une même valeur sélective  $w_2$  et un même taux de mutation.

Si le  $\mu_3$  est nul, le seuil d'erreur n'est pas affecté par le réseau neutre et on revient à un modèle avec un seul génotype  $A_2$  : on retrouve exactement les mêmes propriétés que dans la figure 17, donc le seuil d'erreur entre  $A_1$  et  $A_2$  est le même.

Si le  $\mu_3$  augmente, donc si des mutations retours ont lieu, le second génotype  $A_2^*$  va donner une descendance  $A_2$ . Les deux génotypes étant connectés, le réseau de  $A_2$  devient ainsi plus compétitif que le  $A_1$ . Le taux de remplacement global de  $A_2$  et  $A_2^*$  va surpasser celui de  $A_1$ , les distributions d'équilibre sont modifiées. La valeur du seuil d'erreur s'en trouve donc modifiée. Plus  $\mu_3$  augmente, plus cet effet augmente et plus la disparition de  $A_2$  diminue.

On peut encore étendre ce modèle pour comprendre l'impact de la taille du réseau neutre sur le seuil d'erreur, en ajoutant d'autres génotypes  $A_2$  dans le réseau neutre, chacun avec les mêmes taux de mutation  $\mu_2$  et taux de mutation retour  $\mu_3$  (32). Plus le nombre de génotypes donnant un même phénotype augmente, plus le taux de remplacement du réseau de  $A_2$  augmente, déplaçant ainsi le seuil d'erreur en faveur du  $A_2$  et le protégeant contre l'extinction.

En termes de paysage adaptatif, on parle d'élargissement ou d'extension du réseau neutre, qu'il ait lieu en incluant d'autres génotypes produisant le même phénotype ou suite à une diminution des mutations au sein de ce réseau. Plus il sera étendu, plus on parlera de « robustesse » vis-à-vis des mutations (79). Une population dite robuste présentera un ensemble de phénotypes constants malgré les mutations apparaissant sur la séquence nucléotidique des variants qui la composent. Cette robustesse a ainsi plusieurs conséquences (210):

- Une sensibilité réduite face aux mutations délétères, donc un phénotype constant.
- Une augmentation de l'hétérogénéité génétique.

- Une dispersion moins importante de la distribution individuelle de la valeur sélective. Cela a pour conséquence une valeur moyenne plus proche de la somme de toutes celles des individus.

Le seuil d'erreur au-dessus duquel un type sauvage sera perdu dépendra de la valeur sélective de ce type, relativement à la valeur sélective et l'étendue du réseau.

Ce modèle rajoute un point de vue intéressant quant à l'impact des mutations et l'ajout de plusieurs génotypes, mais complique encore l'interprétation. Cependant ce modèle se base sur des mutations neutres, qui semblent être des phénomènes assez rare, comme on le voit en partie 1.4.3.5.4. et 6.1.4.

### **3.1.4. Le modèle d'Eigen**

Bull étend encore ce modèle afin d'appréhender l'étendue de sa complexité, avec un génome de longueur 40, donc à 40 loci, sans tenir compte des mutations retours (32). On obtient potentiellement 40 génotypes différents. Le système s'en trouve énormément compliqué : cela fait intervenir plusieurs réseaux neutres, de différentes valeurs sélectives et avec un plus grand nombre de génotype donnant le même phénotype, avec une plus ou moins grande robustesse, plusieurs seuils d'erreurs et plusieurs états d'erreurs catastrophiques en augmentant le taux de mutation. On peut facilement imaginer qu'une simple approche graphique, comme celle qui a été réalisée jusqu'ici, est très difficilement utilisable et interprétable.

Les résultats et l'interprétation se compliquent particulièrement dès que l'on rajoute un élément. Il faut donc imaginer ce que cela représente dans un modèle qui décrit le processus de mutation et de sélection appliqué à des populations de taille infinie, de molécules de longueur  $N$  auto-répliquatives qui servent de support à l'information génétique, pouvant donner une multitude de génotypes (voir partie 3.1.2.2.). Ce processus a lieu selon un modèle binaire où les bases qui composent les molécules sont soit des purines soit des pyrimidines, et subissent une mutation à un taux  $\mu$  à chaque locus.

Il décrit alors une transition de phase (78), basées sur deux phases distinctes. Il y a d'abord apparition d'une phase auto-organisée décrivant un équilibre au sein d'une collection de mutants non identiques mais génétiquement liés, les quasi-espèces. Une autre phase de déséquilibre apparaît lorsque le taux de mutation touchant les molécules lors de la réplication dépasse un taux critique, le seuil d'erreur. Cette théorie du seuil d'erreur décrit ainsi une randomisation de la composition des quasi-espèces, vue comme l'exploration de l'espace séquence. La transition de phase dépend, comme on a pu le voir, de la valeur sélective qui est décrite comme une fonction quadratique linéaire.

## **3.2. Applications du modèle à des virus à ARN**

Parallèlement à l'établissement de ce modèle, l'étude des populations virales à ARN apporte des résultats que la génétique des populations seule peine à expliquer. A l'époque, la théorie des quasi-espèces va assez vite attirer l'attention des virologistes, car les conditions et les paramètres d'étude de cette théorie peuvent se rapprocher de ce que l'on retrouve chez les virus à ARN.

### 3.2.1. Mise en évidence d'un réservoir de mutants

Durant la seconde moitié du XX<sup>ème</sup> siècle, les virus à ARN, comme le bactériophage Q-β, s'avèrent être un système particulièrement intéressant pour étudier la réplication du génome à ARN, la régulation de l'expression génétique et la formation des particules virales à ARN. En 1969, Valentine et ses collègues ont rapporté l'apparition d'un échantillon de bactériophage Q-β contenant 8% de mutants sensibles à la chaleur. Ainsi, 3 hypothèses ont été formulées (239) :

- Certains des traits phénotypiques des virus peuvent varier au sein d'une population,
- Des erreurs de réplication remplissent constamment un réservoir de mutants,
- Ces erreurs correspondent à environ une base mal lue sur  $3 \cdot 10^4$  lors de la réplication.

Comme on l'a vu en ce qui concernait la variabilité génétique des populations à ARN, l'équipe de Weissman et Billeter a mis en évidence deux séquences différentes sur un échantillon de bactériophages Q-β apparemment homogènes au niveau phénotypique. Cet échantillon pouvait donc contenir des variants multiples de ce virus présentant pourtant le même phénotype, d'où l'existence d'un probable réservoir de mutants (251, 252). Le parallèle avec ce que décrit la théorie mathématique commence à se dessiner.

### 3.2.2. Une distribution chez le bactériophage Qβ évoquant une quasi-espèce

Des clones de variants mutés de phages à ARN se montraient capables de se diversifier en de multiples autres mutants lors de série de passage sur *E.coli* (61), démontrant ainsi une hétérogénéité des séquences des virus au sein de la population. Cette diversité était forte mais stable, organisée autour d'une séquence consensus. La conclusion de l'époque fut que « la population de bactériophage Q-β était en équilibre dynamique et présentait des génomes à haut taux de réplication et étant soumis à une forte sélection. Le génome a alors commencé à être abordé non pas comme une structure et une séquence unique, mais comme une moyenne d'un grand nombre de séquences différentes » (61).

Des expériences *in vitro* réalisées par Briebricher se sont intéressées à la réplication d'ARN dérivés de l'ARN du bactériophage Q-β de type sauvage. Elles ont permis d'apporter des supports expérimentaux supplémentaires pour établir quantitativement le processus de mutation, de compétition et de sélection (22, 24). Elles ont ainsi décrit que la distribution obtenue à l'équilibre sélection-mutation évoquait celle d'une quasi-espèce, où les mutants étaient présents selon leur valeur sélective et le taux de mutation.

Un nombre incalculable d'études avec d'autres virus à ARN, à ADN, ou même des réplicons subviraux ont rapporté des résultats qui pouvaient évoquer une distribution de quasi-espèces. Malgré les caractères plutôt abstraits de ce modèle, les virologistes ont pu y voir un parallèle à faire avec les virus dont le génome est composé d'une molécule d'ARN. Le modèle soutient bien que le taux important d'erreurs lors de la réplication des virus à ARN est l'un des mécanismes à l'origine de hauts niveaux de variation de séquences observés chez beaucoup de virus à ARN. Cette variation semble encourager l'évolution des populations virales, l'apparition de résistance et d'échappement à la réponse immunitaire, grâce à ce répertoire de variants qui diffèrent du type sauvage.

### 3.2.3. Un modèle applicable mais limité

Cependant ce modèle a plusieurs limites. Il faut en effet considérer l'ensemble des loci au sein d'un génome à ARN, pour lesquels il existe plusieurs génotypes possibles. Il ne faut pas limiter un locus à deux états (type sauvage et muté), même si cela est réalisé pour l'étude d'une mutation en particulier. Il faut aussi considérer les séquences étudiées comme des polynucléotides dont les différents niveaux de structures peuvent jouer un rôle (228). A cela s'ajoutent les mutations retours (32, 174), les possibles effets de la recombinaison (32, 152), les effets stochastiques dus à des populations finies (174), la relation entre la valeur sélective et les mutations, qui n'évoluent pas forcément comme une fonction quadratique.

De plus, le modèle suppose que chaque cellule est infectée par au plus une particule virale. Ce facteur peut être mesuré grâce à la MOI. Il correspond au nombre moyen de particules virales qui infectent une cellule. Cela peut s'exprimer comme le ratio du nombre de virus par rapport au nombre de cellules cibles présentes dans un espace défini. A forte MOI, plusieurs virions peuvent co-infecter la cellule, faisant ainsi intervenir recombinaison et réassortiment. Il faut donc considérer aussi la densité de particules virales au sein d'une même cellule (257).

Au fur et à mesure des avancées de la recherche, de nombreux modèles *in silico* ont tenté d'incorporer ces éléments pour se rapprocher de ce qui se retrouve dans la réalité pour les virus à ARN (32, 152, 174, 228, 257). De nombreuses banques de données, l'utilisation d'outils informatiques et des modèles mathématiques ont grandement aidé à l'application de la théorie en virologie. L'une des implications de la théorie des quasi-espèces, le seuil d'erreur, a notamment fait l'objet d'études *in silico*, de modèles théoriques pour comprendre son application aux virus (23, 155, 160, 165, 230). Il en ressort que ce seuil aurait une implication médicale très importante : l'extinction des populations virales en augmentant le taux de mutation afin d'atteindre l'erreur catastrophique et l'extinction. C'est la mutagenèse létale (voir partie 5.5.).

L'application du modèle ne correspond ainsi pas de manière parfaite aux virus à ARN, et beaucoup de points ont encore besoin d'être revisités. Mais il implique certaines caractéristiques qui pourraient expliquer ce que l'on retrouve dans certains résultats expérimentaux et *in vivo* : le spectre de mutants, le seuil d'erreur, les événements évolutifs appliqués à un ensemble, une grande variabilité qui accorde un potentiel d'adaptation aux virus (143).

### 3.3. Les quasi-espèces chez les virus à ARN

Le terme de quasi-espèces est aujourd'hui utilisé dans plusieurs domaines :

- En physique, les quasi-espèces sont vues comme un nuage dans un espace séquence.
- En chimie, elles sont une distribution de séquences de nucléotides ayant un lien de parenté mais non identiques.
- En biologie, elles sont la cible de la sélection, sans modifier le concept des espèces.
- Pour les épidémiologistes, ce sont des colonies de mutants viraux auxquels nous sommes confrontés lors des thérapies antivirales.

Les virologistes utilisent le terme de quasi-espèces virales pour parler de « distribution de génomes non-identiques mais liés, générés par un processus continu de variation génétique, de sélection et de compétition » (71, 179), présent chez un individu ou dans un environnement donné à l'instant t. Chez les virus à ARN hautement sujets à des mutations,

cette distribution est composée d'un spectre de mutants, dominé par une ou plusieurs séquences maîtresses et des séquences consensus, dont la prépondérance peut être transitoire ou étendue, en fonction, en partie, de la stabilité environnementale.

On peut voir la quasi-espèce de deux façons : selon le trait phénotypique exprimé par les différents mutants de la distribution (figure 23 A) ou selon la séquence des génomes de la distribution (figure 23 B).

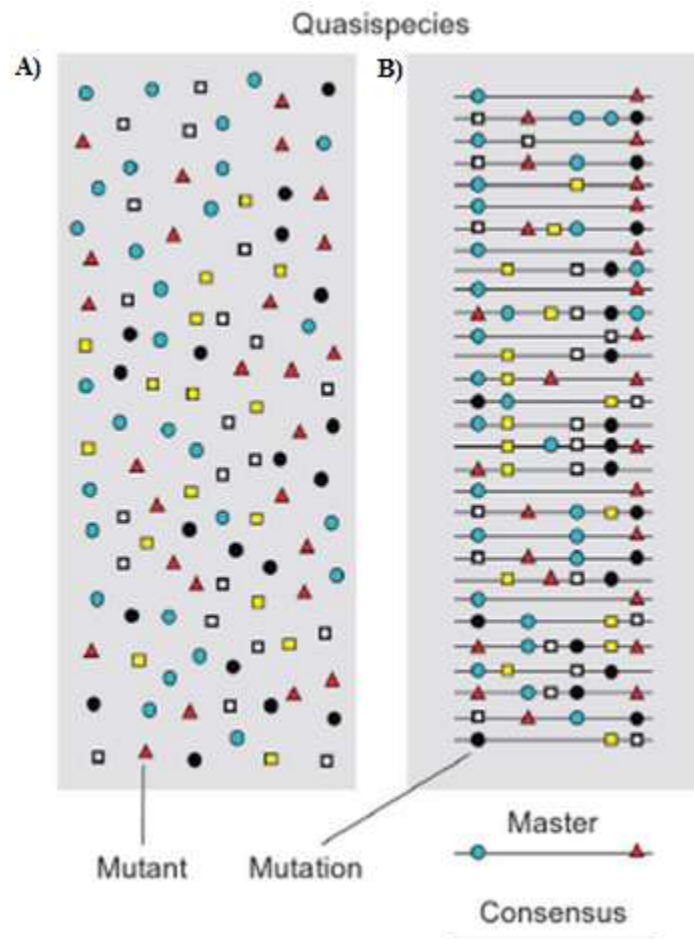


Figure 23 : Deux représentations des quasi-espèces virales (72)

— Séquence nucléotidique  
 ■ ● ▲ □ Trait phénotypique différent ou mutation dans la séquence nucléotidique

A) Représentation selon les traits phénotypiques. Les différents symboles représentent différents mutants dans la population exprimant un trait phénotypique quelconque. Les symboles identiques expriment un même phénotype mais diffèrent par leur séquence nucléotidique.

B) Représentation selon les séquences nucléotidiques

Des termes alternatifs aux quasi-espèces, tels que variation intra-hôte, diversité intra-hôte, mixe de mutants, dégénérescence des nucléotides, hyperploïdies, hétéroespèces ou hétéropopulations ont été proposés (71). Toutes les terminologies sont acceptables, du moment que cela transmet la véritable nature des populations virales. Mais ils échouent pour faire reconnaître la théorie des quasi-espèces en tant qu'origine conceptuelle pour le spectre de mutants.

### **3.3.1. Organisation du spectre de mutants**

#### **3.3.1.1. Définition du spectre de mutants**

C'est ce spectre de mutants, constituant la population de virus à ARN, qui est la cible d'événements évolutifs, et non les génomes avec une même séquence nucléotidique. Il comprend un ou plusieurs variants majeurs quantitativement prédominants, et des variants mineurs, tous issus de la réplication et des erreurs lors de celle-ci. Ces variants viraux doivent présenter un certain degré de similarité de la séquence nucléotidique. Cette similarité est par exemple estimée à au moins 90% chez le HCV (271). Leur présence dans la population assure à tout moment un potentiel d'expression de nouveaux et différents génotypes et phénotypes, qui peuvent devenir dominants ou s'éteindre.

#### **3.3.1.2. La séquence maîtresse**

Il s'agit de la séquence nucléotidique qui présente les mutations les plus fréquentes au sein de la distribution (71). Cela est dû à sa valeur adaptative, plus grande que celles des autres génomes (figure 23). Elle peut cependant être en minorité, comparé à d'autres variants ayant une fréquence plus petite. Dans beaucoup d'études, il peut s'agir du type sauvage.

#### **3.3.1.3. La séquence consensus**

Il s'agit de la séquence nucléotidique qui possède, à chaque position de nucléotides ou d'acides aminés (si l'on considère la protéine associée), le résidu le plus abondamment représenté à la position correspondante, dans un ensemble de séquences alignées (figure 23). Les séquences maîtresse et consensus peuvent être identiques, mais ce n'est pas toujours le cas (figure 23). Elle ne représente aucune vraie séquence, il s'agit d'une moyenne (71).

A n'importe quel moment durant la réplication d'une quasi-espèce virale, celle-ci est constituée d'un spectre de mutants, duquel dérive une séquence consensus. L'utilisation du terme consensus n'a pas ici de caractère universel : des populations indépendantes, ou même successives, d'un même virus, peuvent être caractérisées par des séquences consensus différentes. Au contraire, une même séquence consensus peut caractériser différents spectres de mutants affichant des comportements différents, ou même un spectre mutant avec des complexités différentes. Il ne suffit donc pas à caractériser un spectre de mutants.

#### **3.3.1.4. Un spectre organisé de mutants**

En s'intéressant à l'ensemble de la distribution de mutants, on pourrait s'attendre à une population désorganisée et complexe (figure 24). Mais par groupement des analyses faites par séquençage haut-débit et par « Partition Analysis of Quasispecies » (PAQ), il a été possible de regrouper des génomes proches génétiquement en « sous-population » et de mettre en évidence une distribution de quasi-espèces, notamment chez le HIV-1 et le virus de l'anémie infectieuse équine (figure 24) (13, 14) (voir partie 3.4.4.1.). Cela a notamment montré une coexistence de sous-populations au sein d'une quasi-espèce avec la persistance de variants mineurs avec une faible valeur sélective et de variants majeures. Cela peut être aussi utile pour mettre en évidence une extension ou au contraire une diminution de la quasi-espèce, suite à des traitements utilisant des mutagènes par exemple.

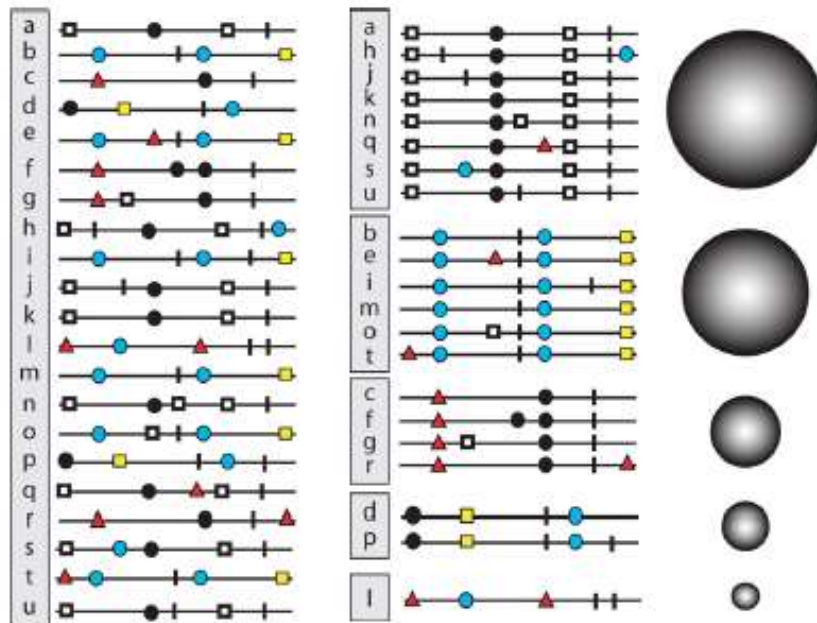
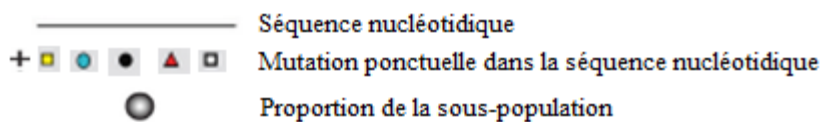


Figure 24 : Distribution des quasi-espèces (71)



Ainsi, une distribution apparemment complexe peut accueillir des groupes distincts. Dans cet exemple, on décrit les classes de mutants avec des sphères de tailles proportionnelles au nombre de génome dans chaque classe, à un instant donné.

### 3.3.2. Formation du spectre de mutants et exploration de l'espace séquence

En considérant le fort taux de mutation que l'on retrouve chez les virus à ARN, il semble peu probable de produire une descendance virale à ARN identique à son modèle (63, 77). Et étant donné la taille très importante des populations virales et la taille moyenne du génome viral à ARN, on estime qu'il est probable qu'une mutation ponctuelle voire deux mutations soient générées à chaque cycle de réplication (243).

Ainsi, une matrice virale va générer de nombreux variants si les conditions du milieu sont favorables à sa réplication. Cela va former une population, organisée autour d'une séquence consensus et d'une séquence maîtresse. Lors des répliquations successives, chaque variant pourra à son tour être répliqué, amenant une distribution en constante évolution de variants ayant des séquences de plus en plus éloignées de la séquence d'origine (125).

De nombreux modèles graphiques ont été apportés afin de mieux appréhender cette dynamique de manière schématique. Ce sont autant d'outils qu'il est intéressant d'utiliser lors de l'interprétation de résultats expérimentaux et lors de la mise en place de modèles théoriques.

### 3.3.2.1. L'espace séquence

Afin de mieux comprendre cette distribution de mutants et sa dynamique, le spectre peut être visualisé comme un espace séquence, c'est-à-dire une représentation théorique de tous les variants possibles de la séquence considérée au sein d'une population (71, 76). Cette utilisation permet ici de mieux appréhender l'ensemble de la population virale et sa réaction face à des événements évolutifs (sélection, mutation, compétition...).

Il peut se référer aux séquences de nucléotides ou aux acides aminés. Il est extrêmement large, dépendant du nombre d'unités utilisées pour construire une macromolécule correspondant à la séquence étudiée. Par exemple, l'espace séquence d'un génome viral de 10 000 nucléotides sera de  $4^{10000}$ , un nombre défiant l'imagination. C'est pour cela que l'on ne représente généralement cet espace qu'en 2 (figure 25 et 26) ou 3 dimensions (figure 27).

#### 3.3.2.1.1. Représentation en deux dimensions

La figure 25 permet d'illustrer la structure d'espaces séquences simples, en mettant en évidence la connectivité entre les points de l'espace et l'augmentation de la complexité avec la longueur du génome. L'espace séquence pour un véritable virus est impossible à décrire avec les termes utilisés dans ces trois diagrammes.

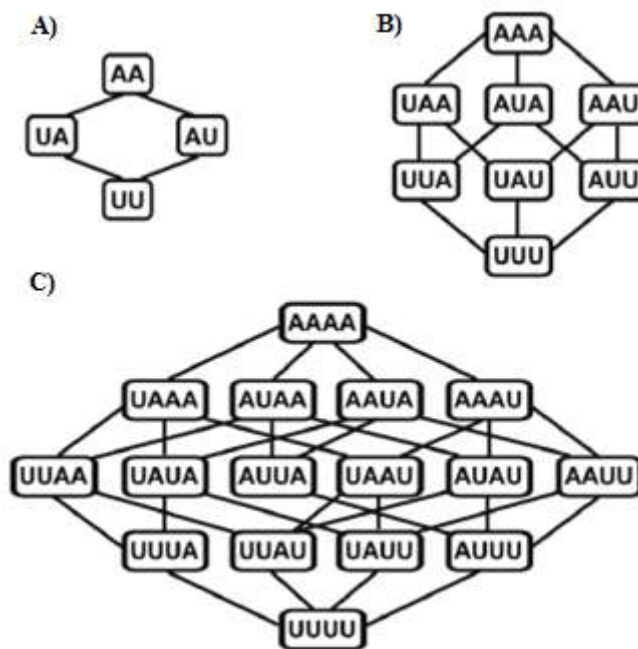


Figure 25 : Représentation schématique des connexions entre des génomes hypothétiques (72)

- Liaison entre deux variants différents d'un nucléotide
- AA** Séquences nucléotidiques composées de deux nucléotides (A, U) de longueur 2.
- AAA** Séquences nucléotidiques composées de deux nucléotides (A, U) de longueur 3
- AAAA** Séquences nucléotidiques composées de deux nucléotides (A, U) de longueur 4

- A) Espace séquence pour des variants de longueur 2 avec deux nucléotides possibles (A, U).
- B) Espace séquence pour des variants de longueur 3 avec deux nucléotides possibles (A, U).
- C) Espace séquence pour des variants de longueur 4 avec deux nucléotides possibles (A, U).



Le variant initial, du fait des erreurs commises lors des cycles de réplication successifs, est à l'origine d'une quasi-espèce virale complexe après plusieurs cycles réplicatifs (figure 26).

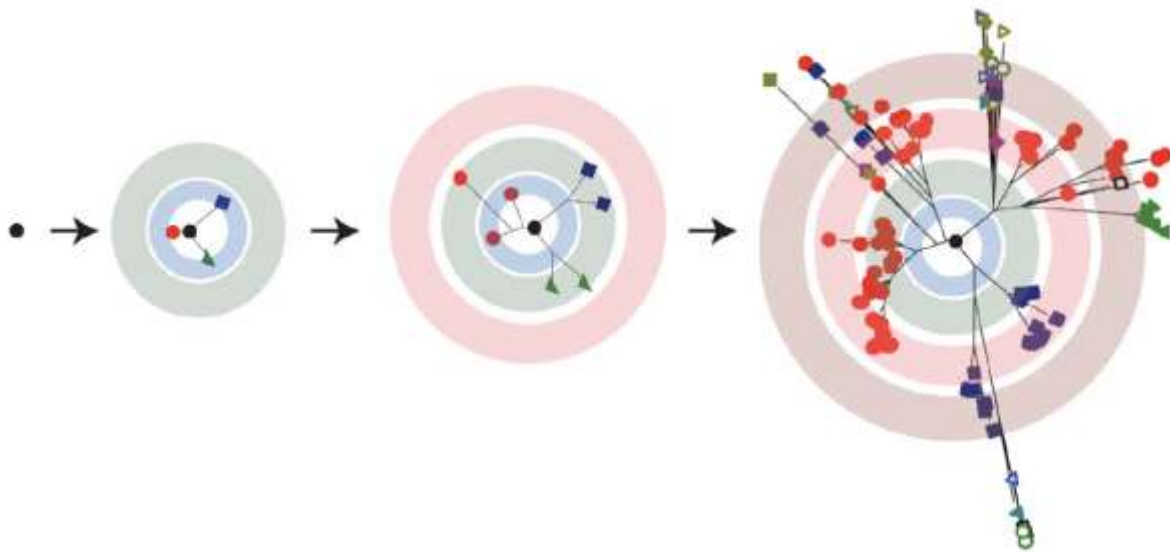
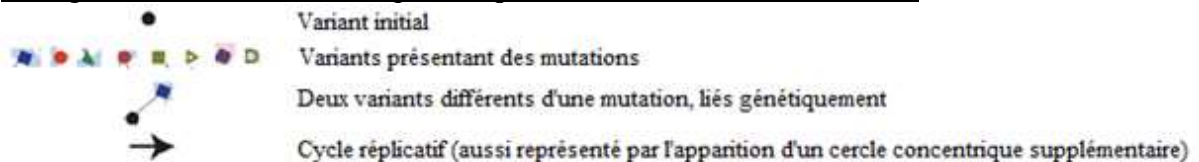


Figure 26 : Représentation schématique de la formation d'une quasi-espèce virale à partir d'un génome matrice sur un espace séquence en deux dimensions (125)



### 3.3.2.1.2. Représentation en trois dimensions

La représentation d'un espace séquence viable (c'est-à-dire capable de se répliquer) d'un virus avec 4 nucléotides possibles et d'une longueur entre 1700 et 32000 bases pour le génome à ARN, sera en forme de sphère, de taille proportionnelle à l'occupation de l'espace séquence. Cette occupation dépend de la taille de la population. On peut cependant remarquer que l'occupation est faible comparée à l'espace séquence théorique, c'est-à-dire tous les génomes possibles, viables et non viables (figure 27). Cette représentation est l'une des plus simples, mais il se peut que l'espace séquence occupé ne soit pas complètement sphérique, voire asymétrique (72).

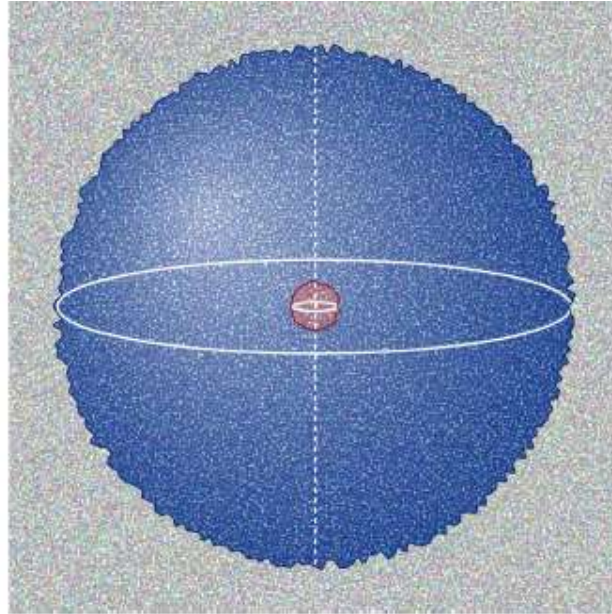





Figure 27 : Représentation théorique de l'occupation de l'espace séquence en trois dimensions (71)

-  Espace séquence théorique
-  Espace séquence occupé par une population de grande taille
-  Espace séquence occupé par une population de petite taille

### 3.3.2.2. Exploration de l'espace séquence

En étudiant ces deux représentations possibles, on peut donc affirmer que la population virale, plus exactement son spectre de mutants, correspond à un nuage de points dans une région de l'espace séquence. Chaque point correspond à une séquence nucléotidique précise de cette population, et deux points séparés par une mutation sont liés par une ligne. Cela met en évidence l'une des caractéristiques principales de cet espace qui correspond bien à la définition d'une quasi-espèce : sa grande connectivité (71). Cela implique bien entendu que deux génomes doivent être séparés par un nombre de mutations qui ne doit jamais excéder la longueur totale du génome.

Plusieurs événements vont permettre de « guider » ce nuage de points à travers l'espace séquence et modifier sa « forme ». Ils ont déjà été vus et peuvent être appliqués à un espace séquence (voir partie 2.) :

- Des mutations (figure 26) : des erreurs sont commises lors des cycles de réplication, le variant répliqué engendre donc d'autres séquences mutées. Cette mutagenèse crée une expansion, au moins transitoire, dans l'espace séquence. Cela rejoint l'idée du spectre de mutants représenté dans l'espace séquence, et permet d'atteindre des points voisins de l'espace sans gradient de valeur sélective. Il arrive parfois qu'une perte irréversible de l'information génétique survienne, quand de nombreux points non fonctionnels sont atteints.

- Une pression sélective : un changement d'environnement va modifier la position de ce nuage de points dans l'espace (par sélection positive et négative), ou un goulot d'étranglement ou un effet fondateur vont le réorganiser en en reformant un nouveau, suite au phénomène de mutation, selon leur capacité d'adaptation (77). Cela se retrouve par exemple lors de l'apparition de mutations résistantes aux médicaments dans les populations virales se répliquant chez des patients infectés. Ces mutations sont souvent sélectionnées en présence du principe actif, qui va jouer ce rôle d'agent sélectif qui conduit au gradient de valeur sélective. Il faut cependant noter que ces mutations qui confèrent cette résistance peuvent être trouvées dans le spectre mutant de virus dont les populations n'ont jamais été exposées au médicament (voir partie 5.1.3.3.4.).

Ce sont les différences de valeurs adaptatives qui vont guider les mouvements des nuages de mutants à travers l'espace séquence et permettent d'atteindre des points distants, avec des conséquences parfois importantes pour l'adaptation du virus. Une fois que la population virale a atteint un point distant, le nuage produit par des pressions de mutation et de sélection va se retrouver dans un nouveau domaine d'exploration, et réitérer le processus. Cela permet au virus d'augmenter sa valeur sélective dans chaque environnement donné.

Si certains points atteints sont incompatibles avec le caractère infectieux du génome, ils peuvent cependant maintenir une capacité répliquative et peuvent guider l'ensemble du génome vers des régions de l'espace qui sont compatibles avec leur survie et leur réplification. Un exemple frappant se trouve dans des « résurrections » virales parfois observées. Des virus qui ont presque perdu leur caractère infectieux à cause de l'effet d'un mutagène peuvent tout de même regagner leur valeur adaptative quand le mutagène est retiré : c'est l'effet Lazarus, la mémoire moléculaire. Si des patients sont soignés d'une infection, parfois un virus peut réapparaître des jours ou des mois plus tard (voir partie 5.3.).

### **3.3.3. Seuil d'erreur et l'erreur catastrophique chez les virus à ARN**

#### **3.3.3.1. Seuil d'erreur et erreur catastrophique chez les virus**

Les effets d'agents mutagènes ont depuis longtemps été démontrés (109) : ils augmentent la fréquence de mutations chez les virus à ARN en regard de certains sites de la séquence nucléotidique. Cette augmentation peut provoquer l'accumulation de mutations létales qui vont diminuer la charge virale (voir partie 5.5.). Nombres d'agents mutagènes comme le 5-Fluorouracil (5-FU), le 5-Azacytidine (5-AZC), ou la Ribavirine (R) sont des analogues nucléotidiques et nucléosidiques utilisés dans de nombreuses expériences.

Des études menées sur le VSV ont montré que le 5-FU et le 5-AZC réduisaient généralement la charge virale et l'adaptabilité du virus (127). Chez le PV et le VSV, l'augmentation de la fréquence de mutation observée au niveau de certains sites de la séquence est très faible mais a les mêmes effets (109, 243). Les virologistes ont très vite fait le rapprochement entre cette augmentation du taux de mutation (qui a un impact sur la fréquence de mutation) et le franchissement d'un seuil d'erreur qui est décrit dans le modèle. Ce franchissement serait ici provoqué par cette mutagénèse chimique.

L'équation du seuil d'erreur décrite par le modèle théorique des quasi-espèces et appliquée aux virus montrent que la stabilité de l'information génétique durant la réplification dépend de deux paramètres (71) (voir partie 3.1.2.2.) :

- Elle dépend de la quantité d'information génétique. Il existe alors une longueur de séquence génomique maximale pour une précision constante de réplication (voir partie 4.2.).
- Elle dépend également du taux d'erreurs réalisées lors de la réplication. Il existe un taux de mutation maximal  $\mu_{\max}$  pour une longueur constante de séquence, en fonction de la fréquence de mutation des espèces concernées.

Ainsi, en augmentant très peu cette fréquence de mutation chez le PV et le VSV (de l'ordre de 2 à 3 fois seulement) (109, 127), on semble dépasser ce seuil d'erreur et induire une erreur catastrophique. On peut donc supposer que le taux d'erreur lors de la réplication chez les virus à ARN est proche de ce seuil (figure 28). Cela apporte mêmes des preuves que des virus de plus basse valeur sélective se répliquent plus prêt du seuil d'erreur, car la supériorité de la séquence maîtresse est encore diminuée au sein du spectre de mutants (voir figure 17).

Dans la distribution d'une population normale (normal population), certains virus, comportant un nombre trop grand de mutations, sont considérés comme étant au niveau ou au-delà du seuil d'erreur, donc entrés dans la phase d'erreur catastrophique. Les génomes ne sont alors plus viables.

Si l'on augmente le taux de mutation, donc la fréquence de mutations retrouvées dans la population (mutagenized population), la distribution entière se retrouve déplacée au-delà du seuil d'erreur, dans la phase d'erreur catastrophique. La majorité de la population virale du type sauvage va disparaître, laissant la place à une autre distribution. Cela n'est pas synonyme d'extinction de la totalité de la population (voir partie 3.1.3.2. et partie 5.5.).

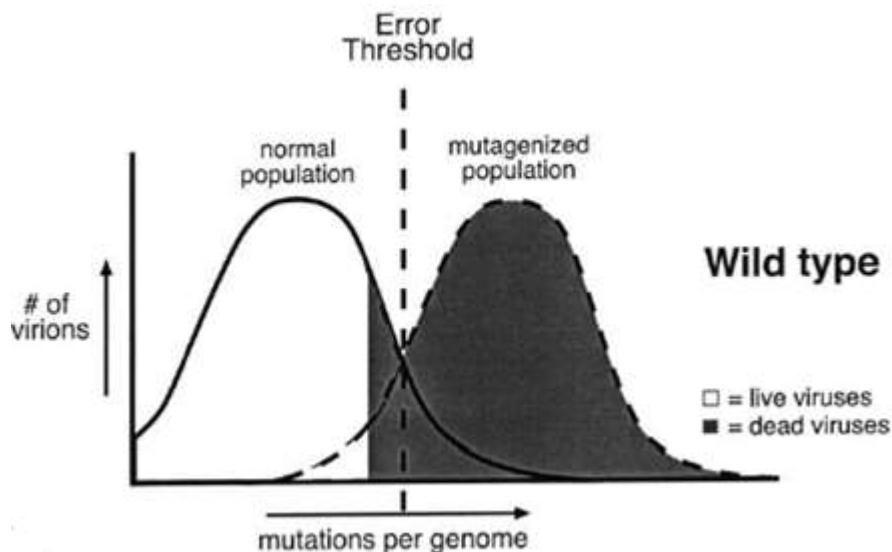


Figure 28 : Distribution de la population au sein d'une quasi-espèce de PV et seuil d'erreur (243)

- L'axe des abscisses correspond au nombre de mutation par génome.
- L'axe des ordonnées correspond à la distribution des virions.
- ! Seuil d'erreur
- = Virus viables
- = Virus non viables

En excédant ces valeurs limites, on arrive à une rupture de l'héritage porté par la séquence. La manipulation du taux de mutation des virus à ARN pourrait donc être une méthode intéressante et efficace pour combattre les infections virales (voir partie 5.5.).

Grâce à cette description et l'équation décrite par la théorie des quasi-espèces, on peut déduire que la valeur du seuil d'erreur dépend de la précision de la réplication, de la valeur adaptative de la séquence maîtresse comparée à celle des copies présentant des erreurs, et de la taille du génome.

### 3.3.3.2. Application à l'exploration de l'espace séquence

Pour aborder cette notion de manière plus imagée, on peut mettre en parallèle cette équation et la représentation d'une population dans l'espace séquence (voir figure 27). On remarque que celle-ci en occupe une portion mineure pour des raisons biologiquement évidentes. Des régions du génome doivent être conservées pour rester fonctionnelles : il doit coder pour des signaux de régulation, présente des contraintes structurelles importantes (de nombreuses structures secondaires et tertiaires dans l'ARN ont des rôles fonctionnels), et les cadres de lecture ouverte doivent coder les protéines fonctionnelles, qui interagissent avec d'autres virus ou des protéines de l'hôte. Tous ces paramètres font qu'il y a énormément de contraintes et de limites à ce qui est réalisable biologiquement (71).

Les séquences génomiques virales ne peuvent donc pas muter de telle sorte qu'elles n'aient plus rien à voir avec la séquence originale (on parle alors d'expansion dans l'espace séquence) et qu'elles perdent leur information biologique. La sélection négative élimine continuellement les génomes non adaptés. Le spectre de mutants représenterait donc une source d'adaptabilité pour le virus. Il s'agirait d'un répertoire dynamique, en constante évolution, de variants génotypiques et phénotypiques viraux, mais ayant des limites.

## 3.4. Mesure de la complexité du spectre

L'application du modèle des quasi-espèces chez les virus a pour conséquence la mise en évidence du caractère complexe de ces populations. La notion de complexité fait ici référence à l'hétérogénéité de la population sous forme d'une multitude de variants possédant des variations dans leur séquence nucléotidique. En effet, la distribution des différents variants que l'on retrouve est largement imprévisible et va dépendre de multiples variables (voir partie 4. et 5.). Il existe un grand nombre de variables, de paramètres et de méthodes qui sont disponibles et peuvent permettre d'appréhender cette complexité. Cependant aucune d'entre elles n'en capture tous les aspects. Une analyse de ces méthodes va nous permettre de comprendre leurs implications quant à la mesure de la complexité du spectre de mutants.

### 3.4.1. Paramètres utilisés pour estimer la complexité du spectre

- La fréquence de mutation minimale : le nombre de différentes mutations observées dans un ensemble de séquences, divisé par le nombre total de nucléotides séquencés (exprimé en substitution/nucléotides)
- La fréquence maximale de mutation : nombre total de mutations (répétées ou non) dans un ensemble de séquences, divisé par le nombre total de nucléotides séquencés (exprimé en substitution/nucléotides).
- Quand la fréquence de mutation se réfère à une mutation spécifique, elle est appelée fréquence mutante.
- L'entropie de Shannon : la proportion de séquences différentes dans un ensemble de séquence. La valeur 0 signifie que toutes les séquences comparées sont identiques ; la valeur 1 signifie que chaque séquence comparée est différente des autres.

- La distance de Hamming ou distance génétique moyenne : le nombre moyen de mutation qui distingue deux séquences dans l'ensemble étudié.
- Paramètres obtenus par NGS : nombre de différentes mutations, nombres de sites polymorphiques, nombre d'haplotypes.

Ces indices de la complexité et leurs propriétés statistiques (104), évalués soit par les méthodes de clonage moléculaire et séquençage de Sanger (Classical Molecular Cloning and Sanger sequencing, CCSS) ou NGS, peuvent cependant présenter des biais, qui ont été étudiés via des échantillonnages *in silico* (104). Gregori propose ainsi des méthodes de traitements de données afin de minimiser au mieux les biais de ces indices, en fonction de l'utilisation des NGS ou des CCSS.

### 3.4.2. Méthodes de séquençage classique

#### 3.4.2.1. Clonages biologique et moléculaire

La première étape pour évaluer la composition de la distribution est de réaliser un clonage moléculaire ou biologique. On obtient ainsi des ARN ou ADN issus de clones biologiques (donc la descendance de génomes viables) ou des ADN ou ADNc issus de clones moléculaires (donc la descendance de génomes copiés et amplifiés *in vitro*). Ces deux méthodes d'obtention de tels clones vont ensuite permettre de caractériser le spectre de mutants afin de comprendre la pathogénèse et l'évolution virale (figure 29).

Le clonage biologique (9) (figure 29 A). Une population virale estensemencée sur une monocouche cellulaire afin d'obtenir des virus issus d'une seule plage de lyse (des clones dits biologiques). Les virus se répliquent d'abord au sein d'une cellule, puis se propagent aux cellules voisines jusqu'à former une plage visible. Ils se propagent jusqu'à l'épuisement des ressources, l'inactivation du virus ou l'interruption de la formation des plages. L'ADN ou l'ARN est extrait, et la séquence virale amplifiée par PCR ou RT-PCR, respectivement. L'ADN amplifié est lié à un vecteur et le produit obtenu est utilisé pour transformer *E.coli*. L'ADN viral, inséré dans les colonies bactériennes, est encore amplifié et séquencé via la méthode de Sanger, afin de fournir un échantillon des séquences génomiques présentes dans les populations virales de départ. Le résultat obtenu est une séquence consensus ou moyenne de la population échantillonnée, qui ne peut être affectée par les possibles erreurs d'incorporation dues à la basse fidélité de copie des enzymes utilisées pour l'amplification *in vitro* (sauf si le nombre de molécules servant de matrice est très petit, de l'ordre de quelques molécules) (72). Cette méthode exclut cependant les génomes de particules virales non infectieuses.

Le clonage moléculaire (9) (figure 29 B). L'ADN ou l'ARN est extrait d'un échantillon biologique et la séquence virale est amplifiée par PCR ou RT-PCR, respectivement. Les produits obtenus sont liés à un vecteur ADN et la procédure poursuivie comme pour le clonage biologique. Cette méthode permet de caractériser le spectre de mutants qui ne se développent pas ou peu sur culture cellulaire. Elle permet d'analyser des génomes infectieux mais aussi non infectieux voire défectueux, mais la séquence obtenue est soumise à des erreurs des enzymes utilisées lors de l'amplification *in vitro*. Il faut aussi s'assurer que les matrices d'ADN ou d'ARN soient en quantité suffisante pour que le répertoire de séquences obtenues après séquençage soit à peu près équivalent à celui présent dans la population que l'on étudie, sans quoi le résultat ne sera pas représentatif.

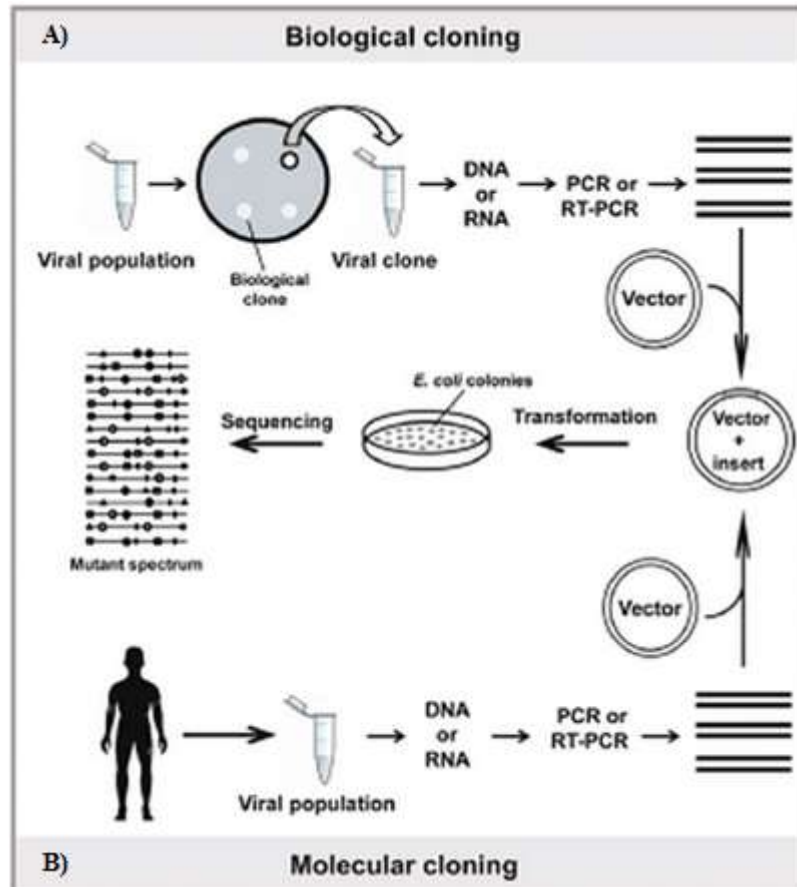
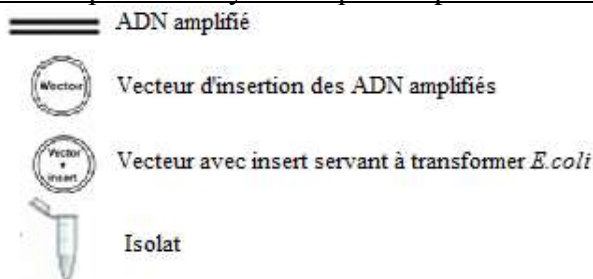


Figure 29 : Représentation schématique des étapes de clonage moléculaire et biologique utilisées pour l'analyse des quasi-espèces virales (72)



Les analyses de clones biologiques et moléculaires ont permis, chez le FMDV, des approches équivalentes pour fournir une représentation d'une quasi-espèce en regard de certains types de mutations (ponctuelles, insertions et délétions) (9). Elles ont aussi mis en évidence une possible sous-structuration de la distribution des mutations, avec notamment l'existence de génomes mémoires (voir partie 5.3.) voire de génomes qui sont, par la suite, amenés à devenir dominants (9).

### 3.4.2.2. Le séquençage de Sanger

Les deux méthodes de clonage incluent une étape de séquençage de type Sanger. Le séquençage de l'ADN a commencé dans les années 70 avec l'avènement de deux méthodes, le séquençage de Maxam-Gilbert et celui de Sanger. Ce dernier est devenu la méthode de choix



et aujourd'hui utilisé en routine dans la plupart des laboratoires. Elle est le plus souvent utilisée pour séquencer une région bien précise d'un génome d'intérêt (figure 30).

Le séquençage de Sanger est une méthode de synthèse enzymatique afin de déterminer l'ordre des quatre nucléotides sur un brin d'ADN (209). Elle part du principe qu'en copiant l'ADN et en observant quels nucléotides sont ajoutés, on peut trouver sa séquence. On place dans un même milieu un brin d'ADN à séquencer, une séquence d'amorce constituée d'un petit oligonucléotide complémentaire à une partie du fragment d'ADN à séquencer ; un mélange des quatre désoxyribonucléotides (dATP, dCTP, dGTP, dTTP) ; une faible concentration de quatre didésoxynucléotides (ddATP, ddCTP, ddGTP ou ddTTP) empêchant la poursuite de l'élongation une fois incorporé dans le brin synthétisé ; et l'ADN-polymérase dépourvue d'activité exonucléase dont le rôle est d'incorporer les nouvelles bases de nucléotides lors de la réplication.

Quatre réactions différentes sont effectuées, comprenant chacune un nucléotide modifié qui constituera le dernier nucléotide intégré à la chaîne de l'ADN. On obtient ainsi un mélange de fragments d'ADN de tailles croissantes, se terminant à toutes les positions dans la séquence. L'ensemble des brins d'ADN produits sont ensuite soumis à une électrophorèse sur gel qui séparera ces derniers, à l'aide d'un courant électrique, selon leur poids moléculaire, et donc leur taille. Ils sont ensuite lus à l'aide de rayon X ou lumière ultraviolette. Chaque colonne ayant un poids moléculaire différent et finissant par un nucléotide différent à l'extrémité, on peut ainsi lire la séquence nucléotidique suite à un traitement bio-informatique qui reconstruit le génome entier.

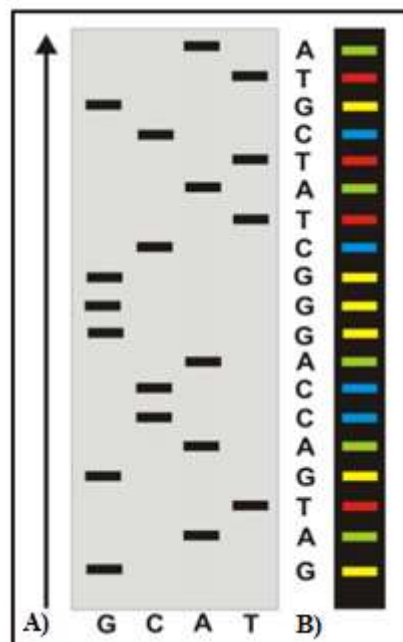


Figure 30 : Représentation des deux méthodes d'électrophorèse (2')

- Brin d'ADN soumis à l'électrophorèse
- ■ Brins d'ADN soumis à l'électrophorèse à l'aide de fluorophores
- A) Electrophorèse en gel standard
- B) Electrophorèse à l'aide de fluorophores



L'utilisation de fluorophores, des composés chimiques dégageant une lumière colorée, associé à chaque didésoxynucléotide, permet le séquençage en une seule réaction avec une seule colonne de gel (221). L'automatisation de cette méthode a permis le séquençage d'un grand nombre de brins d'ADN d'organismes variés, allant du premier génome bactérien  $\Phi$ X174 au génome humain.

Le marquage fluorescent, associé à l'automatisation et l'électrophorèse capillaire, ont ouvert la voie vers le séquençage haut débit. Le seul frein est la nécessité de cloner l'ADN à séquencer, étape qui ne peut être totalement automatisée. De nouvelles pistes ont été explorées, notamment augmenter le débit de lecture de séquences, ce qui a multiplié les possibilités en techniques de génétique moléculaire.

Cette méthode est idéale pour détecter un nombre faible de substitutions au sein de gènes connus, mais elle ne peut être utilisée pour séquencer de grandes portions de génomes de végétaux ou d'animaux ou une grande quantité de génomes. Cependant les populations comme le HIV, le HBV et le HCV sont composées d'un grand nombre de séquences différentes, même au sein d'un même patient. Les CCSS permettent seulement d'identifier la séquence consensus d'une quasi-espèce à un instant donné. Mais celle-ci présente souvent des variants possédant des mutations à de très faibles fréquences, qui ne seront donc pas détectées. Les produits de la PCR peuvent être clonés et séquencés pour identifier des sous-populations virales dites « intermédiaires », mais cette approche est fastidieuse et chronophage.

### **3.4.3. Séquençage haut débit appliqué aux études des quasi-espèces virales**

Les techniques de séquençage de nouvelle génération, aussi appelé NGS, sont globalement découpées en plusieurs étapes communes. Le séquençage haut-débit (Ultra-Deep Sequencing, UDS) est basée sur les technologies utilisant les NGS et palie aux limites du CCSS.

L'analyse des quasi-espèces virales s'effectue par une technique dite de reséquençage, utilisée lorsque la séquence du génome de référence est déjà connue. Le séquençage à haut débit va alors être utilisé afin de connaître les variations génomiques d'un échantillon donné par rapport à la séquence de référence (19, 20). Il peut cependant être utilisé pour détecter des virus inconnus *de novo* (19).

#### **3.4.3.1. Déroulement général du séquençage haut-débit**

Le séquençage haut-débit se déroule selon trois étapes résumées sur la figure 31 :

- La première étape consiste en l'obtention d'une banque d'ADN simple brin associé à des ligands afin de pouvoir l'amplifier lors de la seconde étape. Après extraction, l'ADN à analyser est fragmenté aléatoirement et fixé à l'aide de ligands à une surface solide ou un support de billes.
- La seconde étape est l'amplification clonale de l'ADN, réalisée par PCR en émulsion ou par PCR avec formation de ponts sur une surface. Ces deux techniques permettent la production d'une grande quantité d'amplicons (ADN amplifié) qui vont pouvoir subir un séquençage. Avant cette amplification, une rétro-transcription est réalisée chez les virus à ARN afin d'obtenir un ADN complémentaire (ADNc).

- Le séquençage peut être réalisé par polymérisation, synthèse ou ligation selon les technologies à haut débit utilisées (19, 20). Il va être combiné avec la détection de pyrophosphate, de protons, qui sont libérés lors de l'ajout d'une base lors du séquençage, vont faire diminuer le pH, et ainsi provoquer une impulsion électrique qui sera lue ; ou de signaux fluorescents lors d'utilisation de fluorophores pour certaines technologies. La lecture va se faire suite à l'émission de ces signaux luminescents, électriques ou fluorescents, grâce à différents algorithmes.

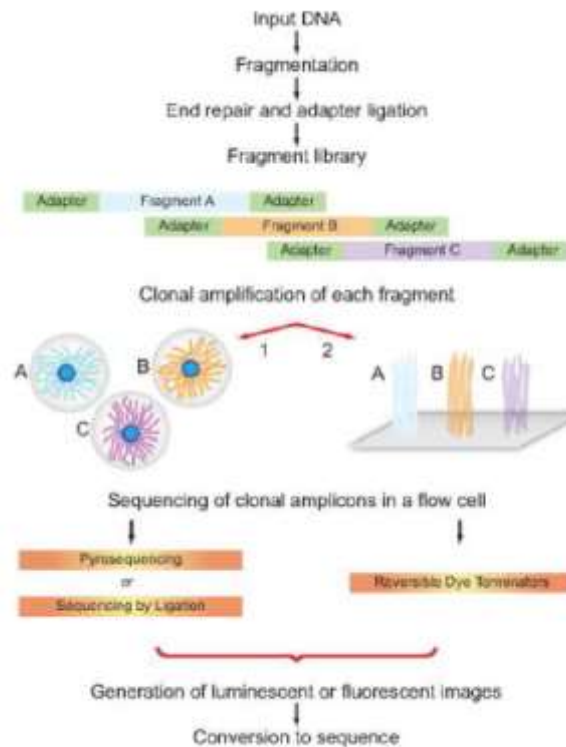


Figure 31 : Etapes d'un séquençage haut-débit (247)

- Fragment A
- Fragment B
- Fragment C
- Ligands
- Emulsion
- Formation de ponts avec une surface

### 3.4.3.2. Technologies utilisant le séquençage haut-débit

Trois technologies à haut-débit sont apparues de façon quasi-simultanée (Annexes 3 et 4) : le pyroséquençage, le séquençage avec des terminateurs réversibles et le séquençage par ligation. Elles utilisent les trois étapes fondamentales, mais diffèrent par la façon dont ces étapes se déroulent (19, 20, 247). Si la méthode de Sanger nécessite d'abord de générer des brins d'ADN de différentes longueurs pour être lus, ces méthodes reconstituent directement les brins d'ADN en déterminant quels éléments sont intégrés (115).

### 3.4.3.3. L'analyse des données

Une fois le séquençage réalisé, il faut analyser les données afin de reconstituer la population des quasi-espèces virales. Les séquences générées par reséquençage vont être comparées au génome de référence grâce à l'alignement de séquences biologiques par des cartes de séquences courtes ou des algorithmes d'alignement.

Il s'agit d'une technique bio-informatique par laquelle on identifie des séries de caractères similaires, apparaissant dans le même ordre sur les deux séquences comparées. Cela permet de conclure quant aux modifications qui se sont produites d'une séquence à l'autre et d'en déduire si deux ou plusieurs séquences sont potentiellement homologues, et donc partagent un ancêtre commun.

Il faut enfin identifier les variants viraux et leur fréquence au sein d'une population. Trois approches sont alors possibles (20) (figure 32):

- L'analyse repose sur la comparaison d'un seul nucléotide (SNV) : elle porte sur les SNP (Single Nucleotide Polymorphism) de chaque séquence, qui constitue la forme la plus abondante de variation génétique, avec variation d'une seule base (pour l'ARN) ou paire de base (pour l'ADN).
- L'analyse repose sur une approche locale qui définit une fenêtre d'étude, c'est-à-dire un nombre donné de nucléotides et de séquences analysées. La comparaison des séquences va permettre de les regrouper selon leur degré de similarité.
- L'analyse repose sur une approche globale qui intéresse l'ensemble des séquences. On aligne les séquences en fonction des chevauchements et les haplotypes vont correspondre à l'ensemble des chemins qu'il est possible de parcourir sur le graphe.

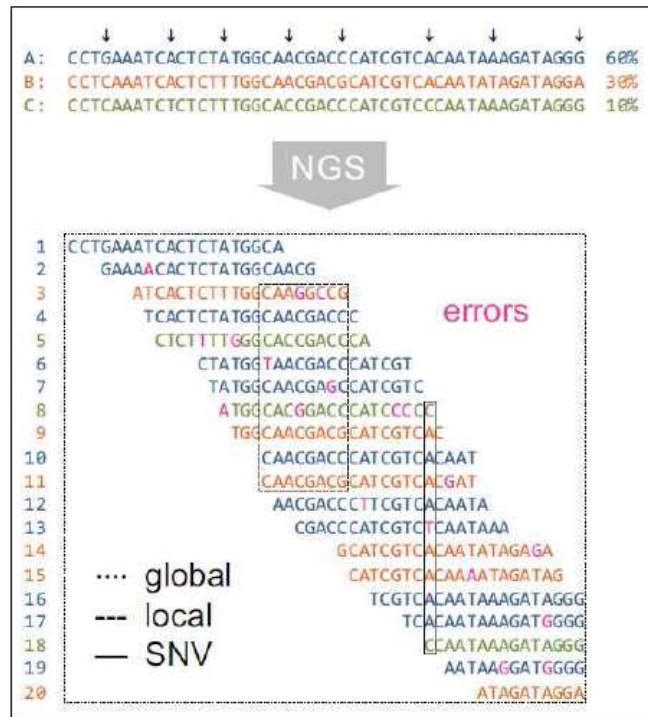


Figure 32 : Les 3 approches de l'identification des variants viraux et de leurs fréquences (20)

- A) A : CCTGAA... Première séquence nucléotidique, retrouvée à une fréquence de 60%
- B : CCTCAA... Deuxième séquence nucléotidique, retrouvée à 30%
- C : CCTCAA... Troisième séquence nucléotidique, retrouvée à 10%
- ↓ Appartenance d'une mutation dans la séquence B et/ou C

- B) Numérotation de 1 à 20 : séquences fragmentées alignées afin de réaliser les trois approches
- A Erreurs détectées grâce à l'alignement des séquences
- ... Approche globale
- Approche locale
- Approche par rapport à un seul nucléotide (SNV)

#### 3.4.3.4. Apports du séquençage haut débit dans l'étude des quasi-espèces

Les méthodes de clonage et de séquençage de Sanger sont limitées, notamment par le nombre de séquences obtenues, souvent moins d'une centaine par échantillon. L'avènement des NGS représente un important progrès dans tous les domaines de la génétique. Ils permettent de pallier ces limites par l'automatisation des étapes, l'intégration de plusieurs systèmes (comme la PCR ou le pyroséquençage), la parallélisation des réactions de séquençage, la miniaturisation et la grande quantité de données générées en une seule réaction (jusqu'à plusieurs millions d'échantillons de petites tailles en une seule fois). Ils permettent aussi une réduction des coûts et du temps de main d'œuvre (19).


Ils fournissent au minimum 10000 à 100000 séquences à partir d'un même échantillon, qu'il soit d'origine biologique (tissu, sécrétion, excrétion, tumeur, animaux entiers, plantes...) ou environnemental (eau, sol...) (19, 20). Des études ont comparé les méthodes classiques de séquençage et de clonage avec les NGS pour caractériser des populations virales (Influenza A, HIV-1, FMDV, HCV) (104, 123, 241, 263, 270). Elles ont permis de confirmer et de renforcer ce concept d'hétérogénéité génétique généralisée aux populations virales et le comportement caractéristique des quasi-espèces virales (voir partie



4.), notamment chez des pathogènes très souvent étudiés, comme le HBV, le HCV, Influenza A, le HIV-1, mais aussi d'autres comme Ebola.

Ils permettent des attributions fiables des mutations présentes à des fréquences de l'ordre de 0.01-1%, en fonction des protocoles de copiage et d'amplification de génome utilisés, ainsi que des opérations de correction et de contrôle réalisées (2, 43, 104, 186, 241, 270). Leur utilisation conjointe avec les séquençages plus classiques pourront permettre d'affiner les valeurs de fréquence et de paramètres mesurés. L'appréhension de la composition du spectre n'en devient alors que plus précise, permettant une véritable dissection à un niveau encore jamais atteint et mettant à jour une multitude de variants minoritaires de basse fréquence qui pourront devenir dominants à des étapes plus tardives de l'évolution (voir partie 5.3.). Cela permet d'aborder les génomes à une autre échelle : ceux autrefois classés dans la catégorie « avec une mutation létale », car de valeur adaptative trop basse, peuvent en fait démontrer une capacité de réplication très faible mais existante.

Les NGS sont largement utilisés pour étudier la diversité virale, principalement dans le contexte de la résistance de certains virus aux traitements médicaux, comme chez le HIV, le HBV et le HCV (20). Ils apportent un regard intéressant sur la cinétique de dominance de sous-populations de mutants échappant aux traitements antiviraux. Ils permettent de quantifier les mutations de résistance aux traitements du virus chez des patients traités et non traités, voire chez des patients avant traitement (223). Ils permettent de détecter des infections mixtes avec différents génotypes du même virus, qui seraient passées inaperçues avec les analyses standards de clones (137, 191). Enfin, ils ont révélé de nombreuses voies de sélection auparavant non remarquées lors d'utilisation des autres modèles de séquençage, comme par exemple les voies d'échappement aux anticorps durant des épisodes de sélection immunitaire (voir partie 5.1.)

Certaines voies de sélection sont seulement transitoires, d'autres deviennent ensuite dominantes (72) (figure 33).

Les génomes avec une mutation  sont directement sélectionnés, car celle-ci confère un phénotype adéquat pour l'environnement existant (figure 33 A).

Les génomes avec une mutation  sont sélectionnés de manière transitoire, jusqu'à être surpassés par les génomes avec une mutation  (figure 33 B). L'utilisation des NGS permet d'appréhender cette mutation qui pourtant ne modifie pas la séquence consensus.

L'utilisation de ces nouvelles méthodes permet donc d'appréhender des événements de sélection transitoires pour une mutation qui, par exemple, fournit un avantage sélectif modeste et pourra être surpassé par une mutation fournissant un avantage sélectif fort.

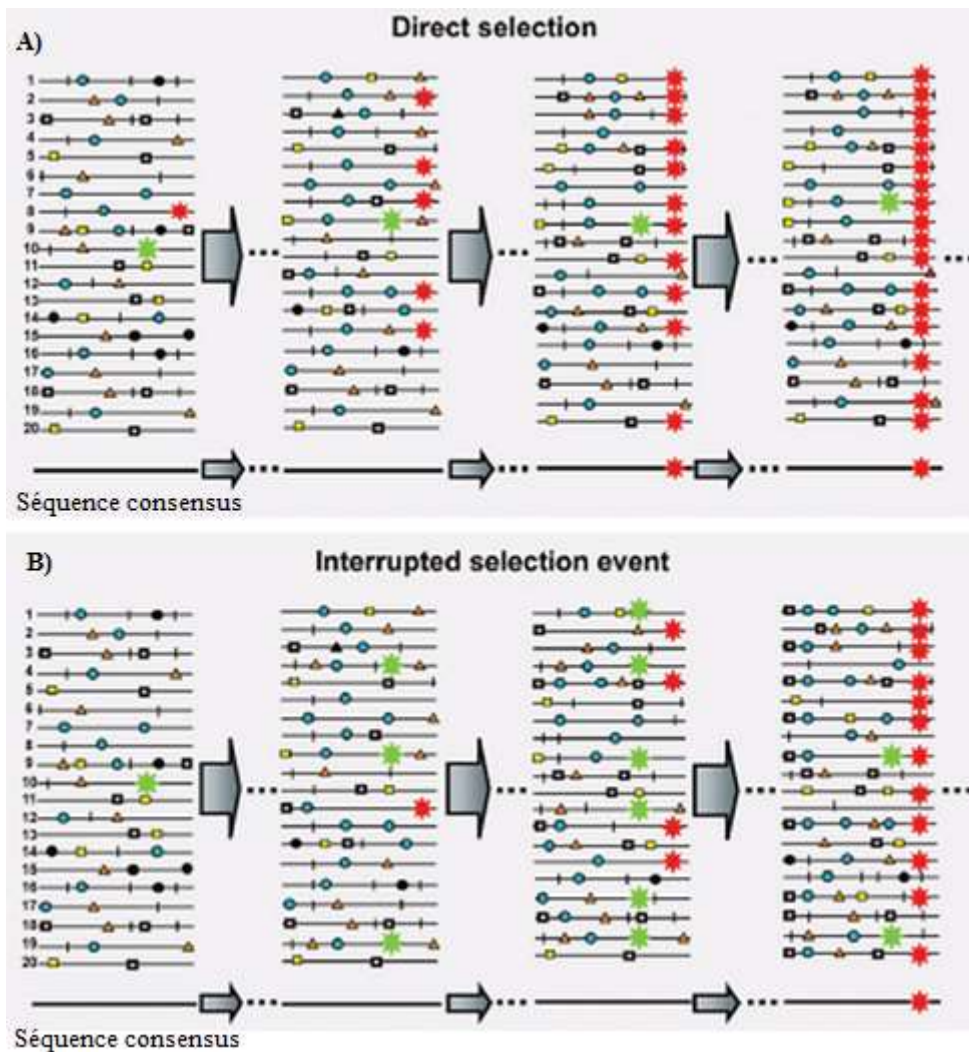
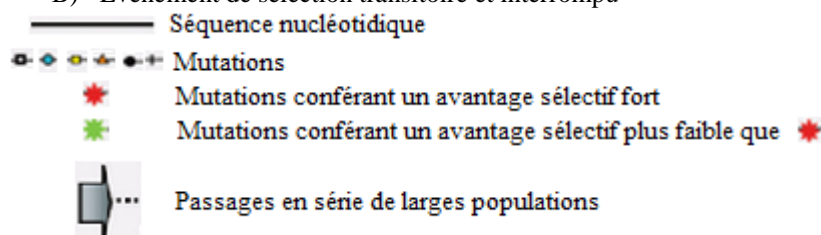


Figure 33 : Mise en évidence d'événements de sélection transitoire grâce aux NGS (72)

A) Événement de sélection direct

B) Événement de sélection transitoire et interrompu



La « pénétration » sans précédent de la composition génétique du spectre de mutants aura des implications cliniques importantes pour la gestion des patients. Le but est à terme de pouvoir adapter les traitements à chaque patient, répondant à leur spécificité génétique et d'appréhender de nouvelles stratégies thérapeutiques et antivirales, en évitant par exemple les résistances croisées ou la sélection rapide de mutants résistants.

Mais ces nouvelles techniques n'ont pas encore pu mettre en évidence de nouveaux concepts. Elles ont de plus des limites dues à une haute fréquence d'erreurs d'origine techniques (voir partie 6.2.). Il s'agit donc d'être très prudent lors de l'interprétation de ces résultats et de mettre en place des systèmes correctifs toujours plus performants pour améliorer ces processus.

### **3.4.4. Méthodes de classification**

#### **3.4.4.1. Partition Analysis of Quasispecies**

L'analyse de la composition complexe et dynamique des populations virales demande aussi de nouvelles approches de classification. Une méthode appelée PAQ a été développée avant l'utilisation en routine de l'UDS. Les PAQ regroupent ces séquences virales sous forme de clades (regroupements des parents et des descendants supposés), issus d'un seul isolat chez un hôte, d'isolats successifs chez un hôte, ou d'hôtes différents. Le programme utilise la distance de Hamming pour mesurer la distance génétique entre les différentes séquences et une méthode de regroupement non-hiérarchique. Elle se base sur le principe que les séquences séparées par les plus courtes distances génétiques possibles sont les plus similaires et devraient donc être groupées (54). Elles permettent ainsi d'organiser le spectre de mutants à un instant donné (figure 24). En rajoutant des codes couleurs, on peut décrire les différents clades en fonction du temps (54). Le PAQ a par exemple été appliqué pour l'analyse de sous-populations du virus de l'anémie infectieuse équine et du HIV-1 (13, 14), des séquences de régulation du virus de la diarrhée virale bovine (120), des isolats viraux de l'hépatite A (54) ou pour qualifier l'évolution du FMDV soumis à une mutagénèse augmentée (166).

#### **3.4.4.2. Une grande variété de méthodes d'analyses et d'échantillonnages**

D'autres algorithmes et méthodes informatiques sont constamment développés afin d'organiser et interpréter un nombre toujours plus croissant de variants minoritaires découverts chez les quasi-espèces virales (voir partie 6.2.). Ils ont été développés pour comparer les lectures, issues de séquençages massifs, avec une séquence référence (19, 75, 142, 186, 237). Ils permettraient de découvrir quelle information est biologiquement et médicalement pertinente et vaut le coup d'être incluse dans les banques de données. Le domaine de la recherche bio-informatique, notamment en ce qui concerne les virus à ARN, commence à attirer l'attention des chercheurs mais reste largement peu exploitée en regard des intérêts médicaux qu'il pourrait apporter (142).

Des nombreuses autres procédures d'échantillonnage rapide comparent de très grands nombres d'échantillons via des procédés automatisés, mais manquent de précision si elles ne sont pas couplées avec des méthodes de séquençage. Les micro-puces à ADN sont notamment devenues des technologies prometteuses pour le diagnostic virologique et pour quantifier la complexité des spectres de mutants (125).

### **3.5. Conclusion**

L'application de la théorie des quasi-espèces à des populations virales à ARN semble imparfaite mais implique certaines notions qui peuvent servir à comprendre la dynamique et la complexité de ces populations. Le spectre de mutants fournit un grand nombre de variants ayant le potentiel d'exprimer différents phénotypes et de devenir dominants si les conditions de l'environnement changent. Cela semble cohérent avec la forte adaptabilité que l'on retrouve chez ces virus. Le seuil d'erreur conditionne la taille du génome et la complexité de l'information génétique qu'il porte, ce qui pourrait expliquer la faible taille des génomes à ARN. L'existence de réseaux neutres prédit aussi la survie du plus plat et l'importance d'un processus évolutif abordant une population comme un ensemble (143). Il convient de définir les paramètres viraux qui peuvent se recouper avec les notions qu'apporte le modèle des quasi-espèces.





## **Partie 4 Paramètres et dynamique des quasi-espèces virales**

La mise en évidence de l'organisation du spectre de mutants d'une quasi-espèce nous permet maintenant d'en étudier la dynamique. Celle-ci se reflète dans la variation de fréquence de ses composants en réponse à des contraintes sélectives. Dans une population, cette variation est mesurée en comparant la composition de la séquence consensus avec celle des autres composants du spectre de mutants. Cependant nous allons voir que ce paramètre est limité dans son interprétation (voir partie 4.1.3.1.).

Ces variations de fréquences nous donnent une idée de la capacité de la population à s'adapter à un environnement, car si une sous-population est fréquente, elle est à priori adaptée, même si la situation s'avère en fait plus complexe (71). Cette capacité d'adaptation se rapporte à la valeur adaptative relative décrite plus haut. Son étude permet donc d'établir la dynamique au sein d'une population virale : la diminution ou augmentation de certaines sous-populations en fonction de leur capacité à s'adapter à un milieu donné. A l'aide de différents outils, nous allons voir comment l'application de la théorie des quasi-espèces peut être cohérente et pertinente vis-à-vis des paramètres viraux et de la dynamique des populations virales.

### **4.1. Hétérogénéité dans les populations virales à ARN**

Comme nous l'avons vu dans les sources de variabilité ainsi que dans le fonctionnement de la réplication, un important paramètre chez les virus à ARN servant de pilier à la théorie des quasi-espèces est le taux d'erreur lié à la machinerie répliquative virale. Plusieurs expériences emblématiques ont permis de comprendre l'impact de la modification de ce fort taux d'erreur sur la dynamique et la complexité.

#### **4.1.1. Une fidélité de réplication modulable**

Un même mutant 3D-G64S chez le PV a été obtenu indépendamment dans deux laboratoires différents et code une substitution d'un acide aminé dans l'ARN-polymérase ARN dépendante du virus (245). Cette substitution provoque une diminution du taux d'erreur lors du processus de réplication : le brin copié est plus fidèle au brin matrice. Ce taux étant diminué, le taux d'incorporation de la ribavirine triphosphate (RTP) dans le brin copié d'ARN lors de la réplication est ainsi diminué : on parle de résistance du mutant G64S à la R (40, 184, 243, 245).

Les conséquences de la modification de l'action de la polymérase sont donc une diminution du taux de mutation et de la fréquence de mutation, ainsi qu'une résistance à l'agent antiviral R. Mais cette résistance a un prix : ce virus devient ainsi incapable de causer des troubles neuropathiques dans le cerveau des souris car il n'arrive pas à atteindre cette zone lorsqu'il est injecté dans un membre. Il est moins pathogène que son homologue de type sauvage, qui peut diffuser efficacement jusqu'au cerveau. Il reste cependant capable de s'y répliquer comme le type sauvage (184, 243). Sur culture cellulaire, cette croissance est cependant légèrement moins efficace (184).

D'autres études ont été réalisées sur les transcriptases inverses chez le HIV-1 et l'effet de substitutions d'acides aminés au sein de ces enzymes, comme celle de la mutation M184V (249, 254). Elle a pour conséquences l'acquisition d'une résistance au traitement avec le 2',3'-dideoxy-3'-thiacytidine (3TC), une augmentation de la fidélité de ces enzymes lors de la réplication, mais ainsi une diminution de la capacité adaptative du virus HIV-1. Cette

substitution semble en effet provoquer une diminution de la processivité ou encore une difficulté à initier la réplication virale.

Les coronavirus sont les virus à ARN avec le plus grand génome jamais découvert, entre 28000 et 32000 bases. Aucune fonction de relecture ou de correction n'a été identifiée chez leurs polymérases ou chez d'autres virus à ARN avec un génome plus court, bien que certains mécanismes indirects de réparation de l'ARN aient été décrits. Un site exonucléase 3'-5' fonctionnel ExoN, situé sur une protéine non structurale NSP 14, a pu être mis en évidence (60, 74, 146). Des mutants pour cette protéine ont montré soit un défaut de synthèse d'ARN (146), soit une augmentation importante du taux d'erreurs commises lors de la réplication (60, 74, 75), tout en restant viables. Le fait que cette protéine soit mutée semble donc avoir un impact sur la fonction correctrice : elle ne fonctionne plus bien, le nombre d'erreurs lors de la réplication augmente et le spectre de mutants devient ainsi plus complexe. Ce trait semble reproductible de génération en génération. Bien que la longueur de ces génomes en fasse un exemple un peu particulier, on retrouve ici les mêmes traits que décrits auparavant.

Plusieurs points sont cependant à revoir : il faut se demander si les mutants présentent la même adaptabilité *in vivo*, et si l'impact de leur mutation sur l'adaptabilité est bien dû à la mutation de l'activité exonucléase et non à celles d'autres rôles de la protéine, car celles-ci sont le plus souvent multifonctionnelles.

#### **4.1.2. Le fort taux de mutation, source d'hétérogénéité de la population**

Toutes ces études, parmi beaucoup d'autres, ont permis de révéler plusieurs implications quant à la dynamique des quasi-espèces :

- Un taux d'erreur élevé réalisé lors de la réplication à cause des enzymes joue un rôle fondamental dans le grand nombre de mutations retrouvées dans une population : il joue sur la complexité, l'hétérogénéité du spectre de mutants.
- Une hétérogénéité limitée semble présenter un désavantage sélectif pour les virus.
- La taille d'un spectre de mutant peut être maintenue de manière stable après plusieurs repiquages cellulaires.

Si la fidélité de réplication augmente, le taux d'erreur devient ainsi plus faible car la polymérase a tendance à incorporer moins de mauvais nucléotide lors de la polymérisation d'un nouveau brin d'ARN (27). Le taux de mutation à l'origine d'une distribution large de variants est donc limité. Dans le cas du PV, la capacité de la population à générer une quasi-espèce est limitée et la souris chez laquelle il est inoculé joue le rôle d'un environnement complexe et changeant lors de la propagation virale. Le spectre de mutants deviendrait ainsi plus restreint, moins complexe, ce qui aurait pour conséquence une restriction du tropisme cellulaire du virus. La croissance et la propagation du virus s'en trouvent diminuées : la virulence, le taux d'évolution mais aussi la vitesse de réplication du mutant 3D-G64S sont diminués chez la souris (184, 243). De tels variants sont donc supposés être plus sensibles à des pressions sélectives apportées notamment lors de traitements antiviraux. Chez le HIV, sous ces conditions, le spectre de potentielles mutations compensatrices est restreint. C'est un autre exemple de l'intérêt biologique de spectre de variants multiples et variés. C'est l'effet inverse pour les coronavirus, qui voient la complexité de leur population augmenter.

Une distribution de variants présentant un large panel de mutations peut donc permettre à la population une croissance et une adaptation plus rapide pour répondre aux

changements de l'environnement. La quasi-espèce et son spectre de mutants sont dits larges et complexes, ce qui peut expliquer l'hétérogénéité ainsi retrouvée chez les virus à ARN. Si la distribution de variants est limitée, on parle alors de spectre de mutants restreint. Enfin, la quasi-espèce ne doit pas entrer dans l'erreur catastrophique, au risque de perdre l'information génétique contenue par la séquence de ces variants. Ces points sont cohérents avec la théorie des quasi-espèces.

Ces expériences ont aussi permis de mettre en évidence des « fidelity mutants » (27, 40, 60, 245). Ces mutants présentent des taux d'erreurs commises lors de la réplication modifiés (augmentés ou diminués). Ils sont des objets d'étude idéaux pour établir des connexions entre la structure des polymérase et leurs mécanismes d'action, pour comprendre comment est modulée la fidélité (donc le rôle de certains acides aminés et de leur localisation spatiale) et pour appréhender l'influence des erreurs lors de la réplication sur la capacité adaptative des populations virales. Il faudrait cependant fournir une approche plus quantitative de leur implication, notamment *in vivo*.

Ainsi les ARN polymérase ARN-dépendantes montrent de nombreux mécanismes pour moduler leur réplication. Les substitutions d'acides aminés semblent modifier les activités de relecture et de réparation ainsi que la capacité intrinsèque de la polymérase à distinguer les nucléotides. Le fait que ce nombre d'erreurs commises lors de la réplication soit modulable suggère que de forts taux de mutation ont été sélectionnés positivement durant l'évolution des virus à ARN. La capacité d'adaptation des virus semble très grande grâce à ces nombreuses erreurs. Il s'agit bien d'un point clé de la diversité, de la virulence et de l'adaptation des virus à ARN, mais ce point est aussi soumis à des contraintes. On estime une « fenêtre » de fidélité de réplication allant de deux à quatre fois la fidélité (60, 74), sans quoi on perd l'information génétique supportée par le génome à ARN. Il s'agit encore d'un point en accord avec la théorie des quasi-espèces.

#### **4.1.3. Remplissage du spectre de mutants : première approche de la dynamique des quasi-espèces**

A l'importance du grand nombre de mutations dues aux erreurs commises lors de la réplication peut s'ajouter un fort taux de réplication chez les virus à ARN. On parle d'un turn-over rapide, avec, par exemple, une production de particules virales pouvant aller jusqu'à  $10^{11}$  par jour pour le HIV-1 et un temps de demi-vie des virions très court (51). Ces deux paramètres conjugués peuvent permettre le remplissage rapide du spectre de mutants, le tout articulé autour de la séquence consensus. Comment peut-on alors qualifier ce remplissage ? Quelles conséquences va-t-il avoir ? La séquence consensus est-elle suffisante pour caractériser la population ?

##### **4.1.3.1. Remise en question de l'intérêt de l'étude de la séquence consensus**

En génétique des populations, les variations ou l'équilibre sont en général évalués grâce aux séquences nucléotidiques consensus de l'organisme étudié. En virologie, elles correspondent à celles qui ont été le plus souvent déterminées et vont servir à l'identification du virus, à l'établissement des relations phylogénétiques, à l'interprétation des processus de sélection et de dérive aléatoire. Cela suppose que l'on n'ait pas besoin d'information sur le

spectre de mutants sous-jacents de la population, et qu'un possible équilibre soit atteint si la séquence consensus ne change pas.

Cependant, pour l'interprétation de l'évolution virale, la pathogénèse et la relation phénotype-génotype, il est essentiel de considérer que la séquence consensus est une moyenne de toutes les différentes séquences, et qu'un génome identique à la séquence consensus n'existe pas dans la population. Elle est souvent confondue avec la séquence maîtresse, celle du composant majeur de la population, alors que ça n'est pas forcément le cas. De plus, les génomes les plus abondants à un instant  $t$  ne sont pas nécessairement les ancêtres des génomes qui domineront la population à un instant  $t+\Delta t$  (voir partie 2.4.3.). Ces ancêtres pourraient très bien dériver d'une minorité cachée à une phase précédente de l'évolution d'une même lignée (31) (voir partie 5.3.).

Des conditions proches de l'équilibre ont été recréées expérimentalement avec un repiquage de larges portions d'une population de virus à ARN. On a ainsi remarqué dans un grand nombre d'expérience qu'une séquence consensus pouvait rester invariable malgré un spectre de mutant complexe et en constant changement (22, 24, 26, 61). Cela n'apparaît que lorsqu'on analyse les composants de la population, chose plus aisément réalisable avec l'application des NGS.

De même, l'augmentation de la mutagenèse via le 5-FU chez le FMDV et le LCMV a été utilisée pour étudier la mutagenèse létale et les interactions d'interférence au sein d'une population virale (voir partie 5.1.2.). Dans chaque cas, la séquence consensus n'a pas varié, mais le spectre de mutants était altéré (96, 97, 102). Ceci s'observe aussi dans le cas d'un état de pré-extinction lors de la mutagenèse létale (voir partie 5.5.4.).

L'étude de la séquence consensus à elle seule ne suffit alors plus (112). Si elle peut être utilisée, notamment lors de grandes phases évolutives, il faut passer par l'étude du spectre dynamique de mutants, qui semble affecter la biologie virale.

#### 4.1.3.2. Une dynamique basée sur la composition du spectre de mutants

La situation est compliquée par la cinétique d'évolution virale. Le fort taux de mutation permet en effet une évolution virale rapide, mais ce n'est pas une obligation. Un même virus peut alterner une évolution rapide et une stase évolutive, le choix d'un état ou d'un autre dépendant surtout de l'environnement biologique dans lequel il évolue, et depuis combien de temps. Pour donner un exemple plus concret, ce genre d'équilibre évolutif est fréquemment observé pour le virus influenza (71). Chez leurs hôtes aviaires naturels, ils ne causent pas de maladie. Mais lors du passage chez des hôtes aviaires inhabituels et des mammifères, les virus semblent montrer une évolution rapide lorsqu'ils se répliquent et peuvent provoquer l'apparition de symptômes, sans avoir forcément un changement de la séquence consensus.

Quand des événements stochastiques ou une sélection positive ont lieu, des composants du spectre de mutants augmentent en fréquence et remplacent la précédente distribution, provoquant parfois un changement de la séquence consensus selon l'ampleur et la nature de la modification. Ces deux éléments, qui perturbent la composition du spectre, peuvent avoir lieu sur le court ou le long terme, en fonction de la tendance de l'ensemble à retourner dans un état proche de l'équilibre. Cette tendance dépend de la valeur adaptative relative des génomes parentaux et de leur descendance. Si celle-ci est grande, il y a de grandes chances qu'un état proche de l'équilibre soit retrouvé.

Une complication supplémentaire se retrouve dans la confusion inappropriée entre taux de mutation, fréquence de mutation, et taux d'évolution, souvent réalisée au cours des études sur les quasi-espèces (71). Par exemple, le taux de mutation continue d'être utilisé de manière incorrecte, aussi bien en étude de l'évolution qu'en virologie évolutive : il contribue à la croyance, erronée, qu'une séquence consensus invariable est due à une absence de mutation. Il faudrait donc aborder l'évolution comme un déséquilibre du spectre de mutants, et non comme un changement de séquence consensus.

Avec tous ces éléments, on peut établir deux modèles de dynamique des quasi-espèces, l'une avec le remplissage du spectre de mutants sans changement de la séquence consensus (figure 34), et l'autre avec changement de cette même séquence (figure 35).

#### 4.1.3.3. Evolution du spectre de mutant sans changement de la séquence consensus

On peut se baser sur l'exemple de la figure 34 : le génome 2 de la distribution 1 génère les génomes 2 et 3 de la distribution 2, et fait de même à la génération suivante (distribution 3) jusqu'à ce qu'un excès de mutations ou l'apparition d'une mutation létale (sur le génome 2 dans la 3<sup>ème</sup> distribution) entrave sa répllication. Les génomes incorporent ainsi des mutations durant la répllication. On observe ici une évolution constante du spectre de mutant avec néanmoins la conservation de la même séquence consensus.

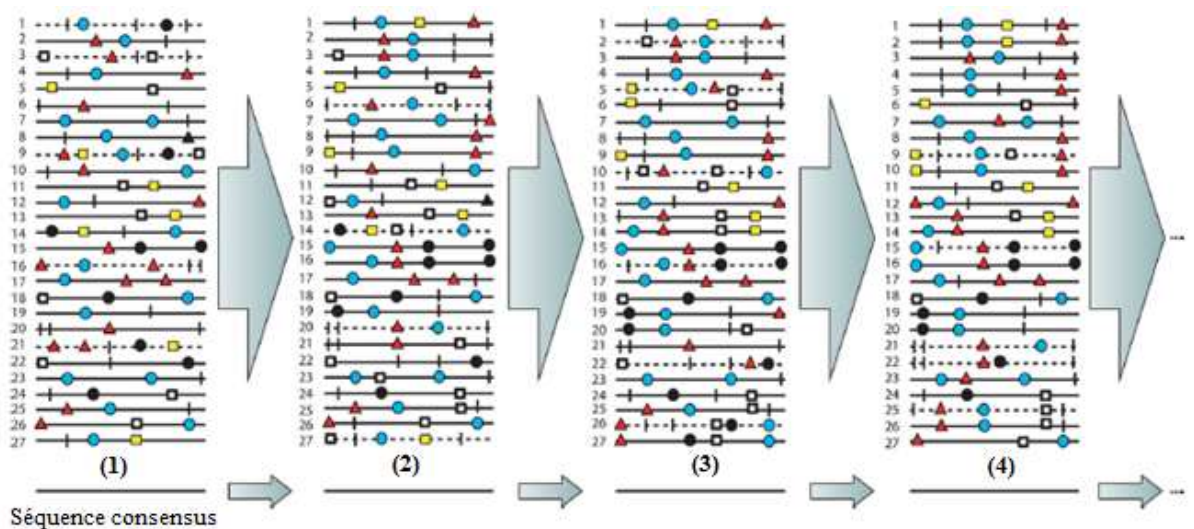
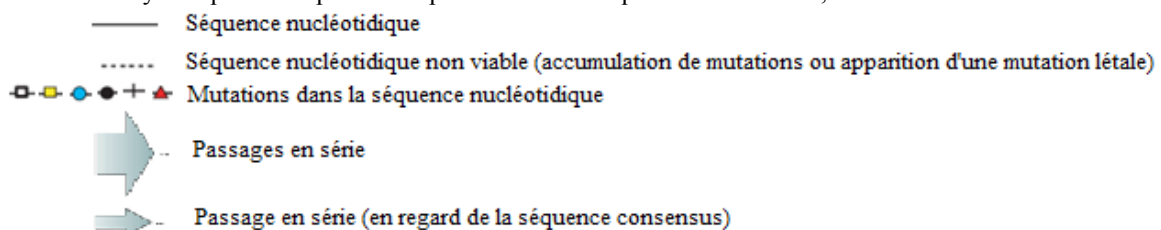


Figure 34 : Représentation schématique de l'évolution d'une quasi-espèce virale, sans modification de la séquence consensus (71)

- La population est représentée ici par 27 variants (numérotés de 1 à 27), subissant des passages en série. L'analyse à quatre temps donnés permet d'obtenir quatre distributions, numérotées de 1 à 4.



Si l'on applique cela à une population virale entière, le nombre de génomes différents dans une même unité de réplication chez une cellule infectée peut être beaucoup plus grand. Cela implique l'existence d'un spectre de mutants hautement dynamique et indéterminé, mais qui peut conserver la même séquence consensus.

#### 4.1.3.4. Evolution du spectre de mutant avec changement de la séquence consensus

La séquence consensus peut changer suite à un événement aléatoire ou à une modification de l'équilibre au sein du spectre de mutants (Figure 35)

L'apparition de la mutation \* confère un fort avantage sélectif (figure 35 A). Cela a pour résultat la dominance de la mutation après un certain nombre de tour de réplication et la modification de la séquence consensus à la distribution 3.

La même mutation au génome 8 confère un avantage sélectif modeste (figure 35 B). Malgré le fait que sa fréquence augmente dans la population, il ne devient pas dominant à la 3<sup>ème</sup> distribution et la séquence consensus reste inchangée. On retrouve le cas étudié dans la figure 33B.

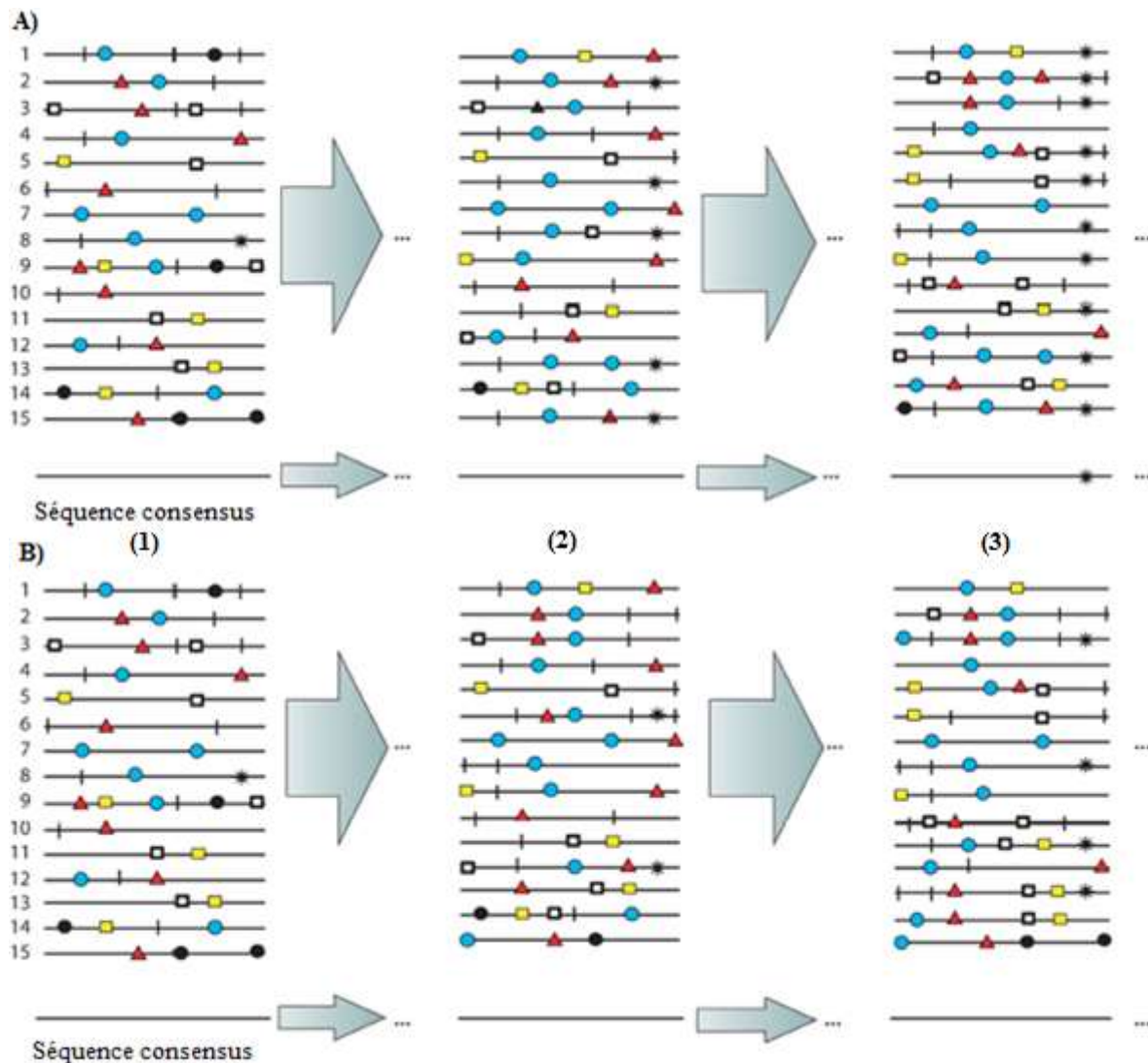
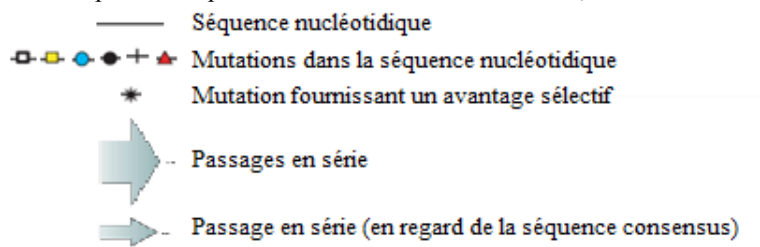


Figure 35 : Représentation schématique de l'évolution d'une quasi-espèce virale, avec possible modification de la séquence consensus (71)

- La population virale est représentée par 15 variants selon le modèle de la figure précédente. L'analyse à trois temps donnés permet d'obtenir trois distributions, numérotées de 1 à 3.



Cela illustre le fait que l'augmentation de la valeur sélective accordée par une mutation sur un génome ne va pas forcément se conclure en une modification de la séquence consensus. De plus, des événements de sélection peuvent passer inaperçus si le spectre de mutants n'est pas analysé, et s'il ne l'est pas au bon moment. L'analyse de la population pourra alors être faussée. Il est donc parfois très difficile d'analyser correctement tous ces événements et leurs implications sur la population virale.

#### 4.1.4. Conclusion

Ce remplissage constant du spectre de mutants est basé sur un fort taux de réplication et d'erreurs lors de celle-ci, et est à la base de la dynamique des quasi-espèces. Cela semble pertinent vis-à-vis du modèle des quasi-espèces, et favoriserait un spectre de mutants complexe, donc une hétérogénéité qui serait plus à même de fournir un avantage sélectif à la population virale. Il faut garder en tête que l'étude de cette hétérogénéité restera toujours limitée si le temps entre l'infection initiale et l'analyse n'est pas connue. Le facteur temps joue ici un rôle qu'il convient de maîtriser dans les protocoles expérimentaux.

Cependant il semble que d'autres paramètres touchant à la population virale et à son environnement interviennent et modulent ce spectre, et vont avoir un effet sur sa capacité à s'adapter.

## 4.2. Taille du génome viral

La taille d'un génome viral à ARN est comprise en 1,7 et 32 kb. C'est l'une des caractéristiques principales de ces espèces. Ce génome semble condenser une information génétique de manière importante. La théorie des quasi-espèces apporte un point de vue intéressant sur la compréhension et l'intérêt de cette taille restreinte.

### 4.2.1. Une meilleure exploration de l'espace séquence

La diversité au sein du génome est une caractéristique générale de toutes les entités répliquatives, que ce soit des virus ou des bactéries. On est en droit de se demander pourquoi l'accent est alors souvent mis sur l'hétérogénéité et l'adaptabilité génétique des virus dans la plupart des études et des modèles théoriques, alors qu'il ne semble pas y avoir de différence avec les autres organismes.

Si l'on prend l'exemple d'un génome viral de 10000 nucléotides de long, il présente une possibilité maximale, donc théorique, de  $3 \cdot 10^4$  mutants, même si beaucoup ne survivront pas ou ne contribueront pas directement à une nouvelle génération car ayant une valeur adaptative trop faible. Ce nombre est bien en dessous de la taille de nombreuses populations virales *in vivo*. Pour un génome de mammifère, le nombre total de mutants possibles est



d'environ  $10^{10}$ , une valeur bien supérieure à la taille des populations estimée chez tous les mammifères (71).

Cela signifierait donc que la capacité à explorer l'espace séquence est bien meilleure chez les virus (71). Un génome de mammifère présente de nombreux nucléotides qui sont destinés à rester invariants quelles que soient les conditions de l'environnement. Il pourra explorer d'autres solutions d'adaptation via la recombinaison, mais pas les mutations. Contrairement à eux, les virus ont le potentiel d'explorer les effets du changement de chaque nucléotide à chaque position. Cela permet une exploitation efficace et rapide d'opportunités phénotypiques qui se présentent lors de cette exploration, en étant à la limite du seuil d'erreur sans cependant le dépasser (77, 110).

#### **4.2.2. Relation du seuil d'erreur et taille du génome**

La relation du seuil d'erreur a plusieurs conséquences :

- La taille du génome est inversement proportionnelle aux taux de mutation qui vont avoir lieu sur cette séquence (111). Cela est cohérent avec la faible taille et le fort taux de mutation lors de la réplication.
- Il y a une limite haute à la taille du génome (76). Cela s'expliquerait par le fait qu'il est impossible de répliquer de grandes portions d'ARN, sans avoir une trop grande accumulation de mutations délétères et perte de l'information génétique. Cela se confirme en pratique, car la taille des génomes à ARN est d'un ordre de grandeur de  $10^4$  bases. On peut supposer que cette taille de génome est mise en place par le seuil d'erreur. L'une des exceptions est le génome du coronavirus, qui est d'environ 32 kb, et qui pourrait sans doute évoluer moins vite que les autres virus à ARN (60, 74, 75, 146).
- Tout événement qui peut augmenter la taille du génome est très limité chez les virus à ARN (111) : la duplication est un phénomène rare et il y a peu de capture de parties du génome de l'hôte par les virus à ARN.

Toutes les fonctions virales se retrouvent ainsi codées par des portions de génomes confinées. Celles-ci ont ainsi des fonctions diverses et multiples, et l'impact des mutations va être d'autant plus important.

#### **4.2.3. Un génome soumis à des contraintes fonctionnelles**

Ce taux de mutation très haut accorde une capacité de variation génétique idéale pour s'adapter à un changement d'environnement, mais paradoxalement cela limite son adaptabilité en limitant la taille de son génome. Cette condensation de l'information génétique rajoute des contraintes :

- Des cadres de lecture se chevauchent souvent chez les virus à ARN, afin de maximiser les informations génétiques disponibles. Des mutations dans ces zones vont être soumises à des compromis complexes, qui affectent les taux de mutations et les relations phénotypiques (111). Cela se retrouve par les effets de mutations qui affectent à la fois le tropisme cellulaire et l'antigénicité (voir partie 5.2.).
- Certaines positions dans la séquence nucléotidique semblent très fortement conservées et soumises à des contraintes. Par exemple, une substitution à un point de la séquence peut accorder un échappement aux lymphocytes T cytotoxiques, mais n'apparaît pas avant plusieurs années d'infection dans le cas d'infection chronique (111). Les traits

phénotypiques majeurs, comme par exemple les mécanismes de transmissions, ne changent que très rarement entre individus d'une même famille. L'évolution limitée de ces traits, en tout cas sur le court terme, prouve qu'ils sont soumis à d'importants compromis.

- Cela soumet le génome viral à un effet appelé le « no free lunch » (72), qui explique qu'il peut être difficile voire impossible pour les virus d'acquérir tous les caractères nécessaires pour maximiser l'impact de la maladie sur les populations humaines. Ainsi, l'amélioration de la virulence aura pour conséquence une diminution de la transmissibilité (111). L'apparition d'une résistance accordée par des mutations peut empêcher le virus d'être résistant à d'autres voies de traitements, ce qui peut expliquer sans doute l'intérêt d'un traitement combinant plusieurs médicaments. Plus le nombre de type cellulaire qu'un virus peut atteindre est grand, plus il sera soumis à des compromis évolutifs.
- Ces contraintes structurales et génétiques semblent cependant être assez variables. On pourrait en effet s'attendre à ce qu'elles empêchent le franchissement de barrières inter-espèces, et pourtant il s'agit d'un caractère retrouvé chez certains virus à ARN, dont les IV (71, 112) et les arbovirus (111). Il a lieu malgré les effets délétères que peut avoir un changement de site de reconnaissance d'un récepteur cellulaire spécifique d'une espèce.

La marge de manœuvre des virus est ainsi limitée : cela empêche tous les virus de s'adapter à toutes les conditions environnementales. Même si l'exploration d'opportunités phénotypiques est favorisée grâce à cette taille réduite qui permet un fort taux de mutation, il faut que cette exploration soit compatible avec la viabilité. Cela rejoint l'exploration de l'espace séquence et ses limites (voir partie 3.3.3.2.).

#### **4.2.4. Segmentation du génome**

En soumettant le FMDV C-S8c1 à une très grande série de transfert de milieu en milieu à forte MOI, on observe l'apparition des délétions internes au niveau du génome viral, jusqu'à obtenir un phénomène de segmentation (93). Ces génomes étaient tout de même infectieux, capables de tuer les cellules par complémentarité (voir partie 5.4.) et semblent avoir surpassé les virus possédant une taille de génome standard. Le génome de taille standard n'était en effet plus détectable (donc de fréquence  $10^{-4}$  fois plus faible que celle des génomes avec délétion). Suite à un repiquage à faible MOI, une réversion du génome de taille standard a eu lieu, suite à une recombinaison entre des génomes avec délétions. D'autres exemples de délétions sont décrits dans la littérature, chez le virus de l'encéphalite équine du Venezuela, chez le virus Coxsackie B3 (CBV3), ou certaines souches du virus « Porcine Reproductive and Respiratory Syndrome » (71).

Ces études ont permis de montrer que des conditions de réplication qui favorisent l'augmentation de la valeur adaptative (ici, le repiquage avec de larges populations) peuvent donner naissance à des modifications génétiques énormes, même si l'environnement physique et biologique est constant. Dans notre cas, il s'agit d'une segmentation du génome, qui aurait pu être favorisée au cours de l'évolution car elle permet de contrer les effets des mutations délétères et d'apporter un avantage sélectif. Deux mécanismes sont proposés pour expliquer son apparition :

- Une sélection de génomes plus courts est favorisée car leur réplication est réalisée dans un laps de temps plus court que le génome de longueur standard. Les délétions seraient apparues au hasard, auraient commencé à se retrouver chez plusieurs

descendants, et lorsqu'une combinaison de délétions internes a permis une complémentarité entre génomes segmentés, ceux-ci ont commencé à surpasser les formes standards. L'une des conditions est que cette combinaison puisse mener à une recombinaison. Cette hypothèse est soutenue par le fait que l'on a retrouvé de nombreux génomes avec des délétions différentes à basse fréquence (93).

- La stabilité de la capsidie serait augmentée chez les génomes segmentés et leur conférerait ainsi un avantage sélectif sur le génome de taille standard. En effet, un génome segmenté, donc plus court, diminuerait possiblement les contraintes d'emballage de la capsidie qui se retrouvent pour le génome standard (168).

Le modèle évolutif observé est particulièrement intéressant : il inclut ici des phases de gradualisme, par accumulation de mutations ponctuelles, et des phases de sauts évolutifs, aussi appelées phénomène de saltation, suite à cette segmentation du génome. Celle-ci permet de contourner les effets négatifs de l'augmentation du contenu génétique due à l'augmentation de la longueur du génome.

### **4.3. Impact de la taille de la population**

Les virus à ARN ont tendance à avoir des populations de forte taille, certains pouvant aller jusqu'à  $10^{12}$  particules virales au sein d'un organisme infecté (71). Même lors des études en laboratoire, le nombre de particules infectieuses excède généralement  $10^3$  en culture cellulaire (72). Il est donc normal de considérer que ce paramètre puisse influencer sur la capacité d'adaptation d'une quasi-espèce virale et sur sa dynamique d'adaptation. Qui dit grande taille de population dit grand nombre de variants, et étant donné la grande variabilité qui touche les particules virales, l'étude de la taille ne peut qu'être liée à celle de l'hétérogénéité, donc la complexité.

#### **4.3.1. Approche théorique à l'aide de l'espace séquence**

En étudiant la représentation en 3 dimensions de l'espace séquence (figure 27), on remarque que cette minuscule occupation de l'espace par les populations virales peut permettre une extension aux points voisins lors du processus d'adaptation. La sphère bleue a un rayon dix fois plus important que la sphère rouge, ce qui signifie que les populations diffèrent dans leur occupation de l'espace. Les virus dans la plus grande sphère possèdent un nombre de contacts potentiels directs plus grand avec l'espace potentiel voisin, que les virus de la petite sphère (72). Ces nouvelles positions voisines peuvent être atteintes par un nombre limité d'étapes de mutations. Malgré cette version très simplifiée, on peut donc émettre l'hypothèse qu'une sphère plus grande, donc une population plus grande, a un nombre bien plus important de possibilités de « progresser » dans l'espace séquence, c'est-à-dire d'évoluer.

#### **4.3.2. Taille et hétérogénéité de la population virale**

Un exemple simple permet d'aborder l'importance de la taille (72). Il faut considérer la fréquence d'une mutation permettant d'acquérir un caractère de résistance aux anticorps, de  $10^{-4}$ , pour laquelle on néglige le coût en termes de valeur sélective. Pour un génome d'une taille de  $10^4$  nucléotides, il faudra une population d'au moins  $10^4$  particules virales afin d'observer l'apparition de cette résistance. Même si trois populations de taille différente présentent cette même fréquence, seules celles respectant cette condition pourront voir

apparaître cette résistance. La composition d'un spectre de mutants dépend donc de la taille de la population.

Domingo et Perales mettent en relation la complexité et la taille de la population sur un autre exemple (71). En augmentant le taux de mutation, une population de  $2 \cdot 10^{11}$  particules virales inclut  $3.3 \cdot 10^9$  génomes avec une mutation et  $2.6 \cdot 10^3$  génomes avec 20 mutations. Une population de  $1.10^8$  particules inclut  $3.3 \cdot 10^6$  mutants simples et seulement 26 génomes avec 20 mutations.

Une grande taille et un haut taux de mutation favoriseraient ainsi l'hétérogénéité au sein d'une population virale. Cependant ces applications théoriques ne sont pas suffisantes pour pouvoir conclure. L'étude de l'impact des goulots d'étranglement, effets fondateurs qui réduisent considérablement la taille de la population, va permettre de déterminer avec plus de précision la dynamique d'évolution et les paramètres qui la caractérisent.

### **4.3.3. Etudes expérimentales des valeurs sélectives**

Le repiquage d'un virus sur culture cellulaire constitue un exemple d'événement de sélection : il va provoquer la croissance d'une sous-population, et donc la modification de la composition génétique de la population virale. Le virus doit « s'adapter » à une culture cellulaire, comme il le ferait lors de l'infection d'une nouvelle cellule *in vivo*. Comme l'a dit Wain-Hobson, « cultiver, c'est déranger » (71). En observant les variations de valeur sélective, il est possible d'appréhender la dynamique et les paramètres de la quasi-espèce.

#### **4.3.3.1. Perte de valeur sélective au sein d'une population virale**

Les premiers événements à considérer et décrits en génétique des populations sont les goulots d'étranglements. Une modélisation de ces derniers est possible grâce à des repiquages en série d'une seule particule virale *in vitro*, et va permettre d'appréhender les différents critères impliqués dans la dynamique des quasi-espèces grâce à la mesure de la valeur sélective.

##### **4.3.3.1.1. *Simulations expérimentales des goulots d'étranglement***

La première diminution de la valeur sélective d'une population a été rapportée sur le bactériophage  $\phi 6$  (41). A l'époque, les mécanismes sous-jacents n'avaient pas été mis en évidence. Les chercheurs se sont rendu compte, au fur et à mesure des avancées scientifiques, que le repiquage en série avait pour conséquence une accumulation de mutations dans le génome viral, avec une modification de la séquence consensus. Elle s'est retrouvée chez beaucoup de virus à ARN dans les mêmes expériences de repiquage en série (48, 73, 83, 157, 158, 268). Cependant les effets biochimiques de cette accumulation et de la raison de la diminution de la valeur sélective ont été très rarement identifiés.

Parmi ces expériences, certaines chez le FMDV et le VSV ont tenté de comprendre les mécanismes moléculaires impliqués dans cette perte de valeur sélective (83, 158). Des centaines de repiquages chez le FMDV ont permis l'apparition de nombreuses mutations inhabituelles, encore jamais observées *in vivo* ni lorsque les populations repiquées étaient plus larges (83). Une telle accumulation de mutations a été décrite pour la première fois par Yin et son équipe, en travaillant sur le bactériophage T7 et en réalisant des repiquages en série sur monocouches cellulaires (265). Chez le FMDV, cela a même provoqué l'apparition de formes

non cytopathiques, affichant une réduction de son caractère infectieux (82, 83). Ces formes étaient cependant capables de se répliquer, d'infecter des cellules BHK-21 sans phase préalable de cytopathogénie et, malgré une valeur adaptative basse, de ne pas subir d'extinction. L'accumulation de mutations, suite à ces repiquages en série, est associée avec une diminution de la valeur sélective en deux phases : une diminution puis un état stationnaire, avec des grandes fluctuations autour d'une moyenne constante de valeur sélective (126).

#### 4.3.3.1.2. *Applications expérimentales du cliquet de Müller*

Toutes ces expériences ont permis de mettre en parallèle deux paramètres : le repiquage en série d'un nombre très petit de génomes provoquerait une accumulation de mutations dans les populations repiquées, et ainsi une diminution de la valeur sélective au sein de la population.

Le repiquage consiste en un prélèvement d'une très petite portion de la population, parfois peu représentative de la population. A chaque repiquage, seule une petite portion sélectionnée est replacée dans un nouveau milieu : la taille et l'hétérogénéité de la nouvelle population sont diminuées, et même la séquence consensus s'en retrouve modifiée. Sous ces conditions, les chances que la séquence la plus mutée soit sélectionnée sont notablement augmentées. Celle-ci présente souvent les mutations qui ont été sélectionnées lors du transfert vers le nouveau milieu. Mais surtout, d'autres mutations vont apparaître sur le génome des membres nouvellement répliqués à cause des erreurs de la réplication virale, d'où l'accumulation de mutations.

Cette situation imite des événements que l'on peut retrouver dans le cycle viral, comme dans le cas d'un transfert à une nouvelle cellule, un nouvel hôte, un épisode immunitaire. Il faut noter que les repiquages sont un cas extrême de goulot d'étranglement, puisqu'une seule particule virale initie la réplication dans un nouveau milieu. Comme on peut s'y attendre selon les prédictions théoriques, dérivées de modèles où les mutations compensatrices ne sont pas prises en compte, on peut supposer qu'une répétition d'événements d'étranglement puisse conduire à l'extinction de la population (32).

Cette évolution de la population est une démonstration expérimentale du cliquet de Müller chez les virus à ARN (41, 48, 82, 83, 157, 158, 268). Müller a proposé que les organismes asexués, avec une population de petite taille et un haut taux de mutation, tendent à incorporer des mutations, qui vont forcément se retrouver chez les descendants suite à leur mode de réplication (150). Ainsi, si une erreur de copie cause une mutation désavantageuse et apparaît dans cette petite population peu variée, elle sera transmise à toute la descendance et avec elle d'autres qui peuvent apparaître aux générations suivantes. La qualité génétique générale de la lignée aura alors tendance à se détériorer avec le temps. Ce processus est comparable à un cliquet dont les dents n'autorisent le mouvement que dans un seul sens. Ces effets ont aussi été étudiés et décrits chez des virus à ADN, des protozoaires et des bactéries voire même des organismes pluricellulaires.

Ces expériences sont des démonstrations fortes de l'influence de la taille et de l'hétérogénéité de la population, même s'il faut rester prudent sur leur généralisation car seules quelques familles de virus sont ici utilisées. De même, elles apportent des exemples intéressants de l'impact des mutations et de leurs caractères neutre, létal ou bénéfique.

#### 4.3.3.1.3. *Maintien de valeur adaptative et échappement à l'extinction*

L'extinction que l'on aurait pu s'attendre à trouver n'a pas été observée (41, 48, 73, 82, 83, 157, 158), exceptée chez le HIV-1 (268). Cette exception peut être liée au taux de réplication, qui est probablement plus bas que chez les autres virus, et à la compétition et la sélection au sein des plages de HIV-1 qui sont plus limitées que dans le cas du FMDV.

Cette évolution de valeur adaptative en deux phases, avec une fluctuation autour d'une valeur basse, démontre ainsi une résistance à l'extinction. Cette résistance chez les virus à ARN peut notamment être due à des mutations compensatrices qui apparaissent lors du processus d'accumulation de mutations (81, 126), voire lors de recombinaisons, dans un environnement toujours changeant. La population emploie alors des mutations pour atteindre un état le plus stable possible par rapport à ces changements.

Il pourrait être intéressant de voir si cette tolérance à l'extinction est un fait général ou une particularité de ces virus. Il faut continuer à soumettre des systèmes viraux à un nombre encore plus grand de repiquage et avec des populations toujours plus différentes.

#### 4.3.3.1.4. *Une conséquence positive de la sélection négative*

Le goulot d'étranglement va procéder à une sélection négative dans notre cas. Un contexte particulier peut cependant favoriser la survie d'une sous-population minoritaire, et la sélection négative peut alors avoir une conséquence positive.

Des populations de FMDV soumises à des repiquages en série ont montré posséder une fréquence de mutation de  $1.5 \cdot 10^{-2}$ , ce qui est plus grand que la fréquence trouvée dans le cas d'une même population soumis à une mutagenèse létale (de  $4 \cdot 10^{-4}$  à  $3 \cdot 10^{-3}$ ). Ces dernières avaient d'ailleurs subi une extinction (82, 110 (chez le VSV), 172). En regard des fréquences de mutation, on pourrait s'attendre à une extinction lorsque l'on fait subir à la population ce repiquage en série. Mais durant ces expériences, on « sauve » une faible portion de la population parmi une majorité qui a perdu ses capacités répliquatives. Un goulot d'étranglement, perçu au premier abord comme un événement qui restreint l'hétérogénéité et la taille d'une population, permettrait le sauvetage de celles-ci lors de contraintes particulières.

#### 4.3.3.2. Augmentation de la valeur sélective au sein d'une population virale et atteinte de l'équilibre

La dynamique au sein des quasi-espèces implique aussi que la valeur adaptative au sein de la population puisse augmenter : la population atteint un état proche de l'équilibre dans un environnement donné, car il est adapté à ces conditions.

##### 4.3.3.2.1. *Définition de l'équilibre en matière de valeur sélective*

Une population est considérée comme adaptée dans un environnement donné quand les processus de mutation, de compétition et de sélection ont agi pendant un laps de temps suffisamment long afin qu'un équilibre, proche de l'optimum, soit atteint sans qu'un événement ne vienne le perturber (qui cependant n'est pas foncièrement atteignable en réalité) (71). Cela devrait se traduire par une augmentation de la valeur adaptative au sein de la population. Cependant, les mutations sont inhérentes à la réplication, et, dans une population

bien adaptée, beaucoup d'entre elles vont avoir un effet délétère en matière de valeur adaptative, donc devrait avoir tendance à la diminuer. Le processus d'accumulation de mutations est particulièrement important quand le taux de mutation est élevé, donc chez les virus à ARN. Ce n'est que s'il est compensé par la sélection des mutations apportant un avantage sélectif, ou les moins délétères, que l'on peut parler de quasi-équilibre.

#### 4.3.3.2.2. *Mise en évidence expérimentale et interprétation*

Des expériences similaires aux repiquages en série ont été réalisés, mais avec des transferts larges de population (48, 132, 156). Il en résulte une augmentation de la valeur adaptative, qui peut être vu comme un état d'équilibre entre mutations et sélection à son optimum pour un environnement donné.

Ce transfert de larges populations doit être vu comme un changement d'environnement dont les contraintes ne sont pas trop sélectives. L'augmentation de valeur adaptative peut être expliquée par la taille relativement grande ( $10^9$  à  $10^{12}$  particules) des populations virales retrouvée chez un hôte infecté. Plus la population est grande, plus il y a de chances qu'une grande partie de cette population survive lors d'un changement d'environnement, et plus elle sera hétérogène et complexe. Cela est mimé par les transferts réalisés dans ces expériences.

Cette taille est intimement liée à son hétérogénéité et à la part de la population véritablement « fonctionnelle ». Mais il est difficile de déterminer la proportion de la population effective, c'est-à-dire capable de réplication. Une particule pourrait ne pas pouvoir initier une réplication ou une infection dans un environnement donné, mais pouvoir le faire si elle est exposée à un environnement différent. Cela est le cas lors d'un transfert large de population. La valeur adaptative est donc environnement-dépendante (63, 124, 172, 192).

#### 4.3.3.2.3. *Avantage sélectif de la taille et de l'hétérogénéité*

Ces expériences mettent en évidence deux paramètres qui interviennent dans la dynamique des populations virales : la taille et l'hétérogénéité des populations virales ; et un nouveau facteur, celui de l'hôte. Ces paramètres sont aussi concordants avec la théorie des quasi-espèces et ce qu'elle implique.

Les goulots d'étranglement sont des événements fréquents lors du cycle viral. Les populations qui présentent un grand panel de mutants, capables de s'adapter rapidement suite à un tel événement, vont très certainement surpasser celles qui ne peuvent se propager qu'à un sous-ensemble de cellule. Une population avec un spectre complexe (large et hétérogène) peut présenter un avantage sélectif sur un spectre plus restreint car il contient un plus grand nombre de sous-ensemble de variants bien spécifiques avec un grand potentiel d'adaptation, que l'on a moins de chance de retrouver si le spectre est plus petit (125, 245). De plus, les interactions de complémentarité sont augmentées au sein de son répertoire de variants (244, voir partie 5.4.1.).

#### 4.3.3.3. Limites de la taille du spectre

La plupart des études s'articulent en faveur d'un virus à ARN avec un spectre de mutants de grande taille. Mais un spectre restreint peut présenter des avantages : il peut permettre d'exclure des interactions d'interférence mais aussi d'enrichir la part d'un sous-ensemble spécifique, qui serait noyé dans un spectre trop grand (voir partie 5.4.2.).

#### 4.3.4. **Le taux d'évolution, autre approche de l'importance de la taille et de l'hétérogénéité de la population**

L'hétérogénéité semble apporter un avantage sélectif cohérent avec la dynamique des quasi-espèces et la nature des virus. Si le répertoire de variants est grand, la sélection d'un sous-ensemble de génomes peut alors être très rapide : les taux d'évolution de la séquence consensus peuvent atteindre, *in vivo*, jusqu'à  $10^{-2}$  substitutions par site et par an, soit  $10^6$  à  $10^7$  fois plus important que les taux d'évolution moyens retrouvés lors d'évolution sur le long terme chez les cellules (64) (Annexe 5). Si ce spectre de mutants est, de plus, complexe, il présente un double potentiel :

- Une atteinte rapide de l'équilibre, avec un taux d'évolution autour de  $10^{-4}$  substitutions par site et par an, voire moins. On parle alors de stase évolutive. Mais cela n'implique pas qu'il y ait un taux de mutation plus bas que celui observé lors de la réplication habituelle. Dans certains cas comme celui du HDV, cela peut correspondre à la phase chronique de l'infection (128).
- Une évolution rapide de la séquence consensus, avec un taux d'évolution de  $10^{-2}$  à  $10^{-1}$  substitutions par site et par an. Tout comme pour une stase, cette évolution rapide n'implique pas un taux de mutation plus élevé que la normale. Cela peut correspondre à une phase aiguë et à des changements dans la présentation clinique, comme chez le HDV (128).

#### 4.4. **Apport des paysages adaptatifs : paysages accidentés et survie du plus plat**

L'apport des études de la valeur adaptative nous amène à considérer les paysages adaptatifs. La théorie des quasi-espèces accorde une certaine importance à des réseaux neutres, représentant des situations dans l'espace séquence ou dans des variations de l'environnement qui ne provoquent pas de modification de la valeur sélective. La théorie intègre ainsi une nouvelle notion, la survie du plus plat, en référence à la survie du plus apte selon les principes darwiniens.

##### 4.4.1. **Barrières génétiques et phénotypiques**

On utilise le terme de barrière pour décrire les limites que le virus doit surpasser pour atteindre un phénotype donné (64). Ces deux notions seront reprises pour l'apparition des résistances virales mais sont essentielles pour comprendre le paysage adaptatif et son relief (voir partie 5.1.3.3.).



Une barrière génétique se réfère au nombre de mutations qui seront nécessaires pour que le virus s'adapte, donc pour l'apparition d'un phénotype donné (72).

- Si deux ou plusieurs mutations peuvent conférer le même phénotype, la probabilité d'apparition de celle-ci va correspondre à la somme des probabilités d'apparition de chaque mutation. En effet, si l'une de ces mutations apparaît, elle confère le phénotype de résistance, les autres mutations ne sont donc pas nécessaires. La barrière génétique est, dans ce cas, faible.
- Si deux ou plusieurs mutations doivent avoir lieu en même temps pour permettre l'apparition du phénotype, sa probabilité d'apparition est égale au produit des probabilités d'apparition de chaque mutation. Dans ce cas-là intervient aussi l'épistasie positive et négative. La barrière génétique augmente encore.
- Pour le dernier cas, on se place dans une situation où le caractère conféré par les mutations est une résistance. Si un génotype confère une résistance à plusieurs médicaments antiviraux, sans résistance croisée, la probabilité de son apparition correspond au produit des probabilités d'apparition de chaque mutation conférant une résistance pour un inhibiteur spécifique. La barrière génétique à franchir est ici assez haute.

On parle donc de barrière génétique basse si, par exemple, une simple mutation de transition permet l'apparition d'un phénotype. Si elle requiert deux mutations simultanées, comme une transition et une transversion, ce qui est moins probable, la barrière génétique sera dite plus haute. Les réseaux neutres correspondent à des séquences nucléotidiques où la barrière génétique est haute : une modification de la séquence n'entraînera pas l'apparition d'un nouveau phénotype.

Une barrière phénotypique se réfère aux conséquences qu'aura une substitution d'acide aminé sur la capacité adaptative du virus (64). Cela rejoint les notions de caractères des mutations : si elles sont neutres, létales ou bénéfiques en regard de certains critères du virus. Une forte barrière phénotypique signifie que les conséquences de la mutation seront importantes pour le virus, que la valeur sélective va diminuer et que le virus aura plus de difficultés à franchir cette barrière pour survivre. On dit alors que la mutation implique un coût de valeur sélective. Elle aura un effet sur l'ARN, la protéine virale exprimée, ou les deux.

#### **4.4.2. Un paysage adaptatif accidenté chez les virus**

Il apparaît que le paysage utilisé dans la figure 12 est le plus à même de décrire l'évolution des virus (64). On aurait donc affaire à un paysage très accidenté, fait de pics abrupts et de vallées assez étroites. Plusieurs arguments sont en faveur de cette hypothèse :

- Les génomes capables d'interférences qui sont responsables de la perte du caractère infectieux d'un virus peuvent cesser d'être interférent juste à cause d'une mutation (176) (voir partie 5.4.2.). Du point de vue du paysage, cela signifie qu'une seule substitution peut faire basculer un virus d'un pic adaptatif vers une vallée profonde. La barrière génétique est dite faible et le coût de valeur sélective est fort.
- Les mutations sont, de manière générale, rarement neutres chez les virus (71, 79). La barrière phénotypique est considérée comme forte, notamment parce qu'une grande quantité d'information génétique est concentrée sur une petite portion de génome. Cela aura pour conséquence une modification de la valeur adaptative du mutant présentant ce changement, qui apparaîtra de manière fréquente. Il peut en résulter des valeurs adaptatives très variables d'un point à un autre du paysage, d'où son aspect accidenté.

- Les variants d'un même virus mais isolés à partir d'organes différents ont généralement des capacités adaptatives différentes (voir partie 6.3.3.). On peut donc supposer que cette grande diversité de valeurs sélectives soit représentée sur le paysage.
- La valeur adaptative est environnement-dépendante (63, 124, 172, 192). Le changement d'environnement rapide, auquel participe la composition elle-même de la population virale, rend ce paysage transitoire (voir partie 2.4.).

Les paysages accidentés sont ceux qui représentent le mieux la réalité de l'évolution virale (72). Sur la figure 11, on remarque un plateau au sein du paysage : il n'y a pas ou peu de variation de valeur sélective malgré le mouvement dans l'espace séquence ou la modification de l'environnement. La barrière génétique est dite plus grande et les populations qui peuplent ce plateau sont dites robustes (79, voir partie 3.1.3.6.). .

#### **4.4.3. La survie du plus plat sur des paysages adaptatifs**

Ces zones de robustesses (les réseaux neutres) sont décrites par la théorie des quasi-espèces (voir partie 3.1.3.6.). Elle prédit que des organismes présentant cette robustesse ont un avantage sélectif, mais son application aux populations virales semblent être difficile étant donnée la barrière génétique faible qui les caractérise. Le principe de Darwin de la survie du plus apte suggère en effet que les virus se répliquant le plus vite vont surpasser les autres. Mais en tenant compte des erreurs commises pendant la réplication, il se peut que ce ne soit pas toujours le cas.

Dans les paysages adaptatifs appliqués aux espèces virales, on retrouve un paysage accidenté, qui alterne des pics adaptatifs très hauts mais étroits, des vallées adaptatives qui font la jonction entre ces derniers et des pics plus plats, sous forme de collines, donc moins hautes mais plus larges (125). La plupart des modèles *in silico* étudiant les quasi-espèces ne réussissent pour l'instant qu'à modéliser un paysage simple, ne contenant qu'un pic adaptatif ou une colline (figure 36).

La population 1 a une haute valeur adaptative, mais est piégée dans un espace séquence car chaque mutation va conduire à une perte très importante de valeur sélective. Cela rejoint le fait que les mutations, même pour une simple substitution, ne sont pas neutres, et que le paysage adaptatif peut être très accidenté dans le cas de populations virales.

Pour la population 2, une mutation va entraîner une plus faible perte de valeur sélective. Les populations les plus robustes sont idéalement situées afin de pouvoir se déplacer dans l'espace séquence et ainsi accéder à d'autres pics des environs, via un ensemble de mutations voisines.

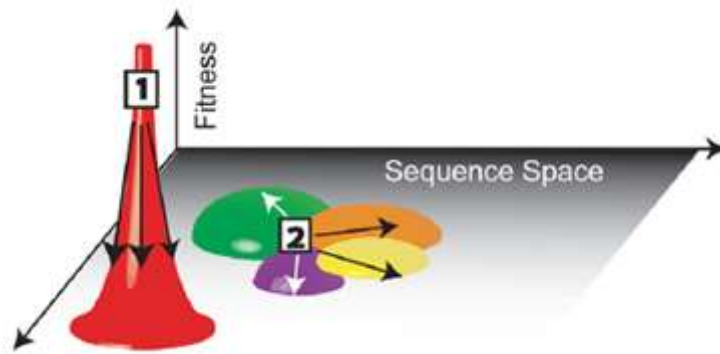


Figure 36 : La survie du plus plat du point de vue d'un paysage adaptatif (125)

- Pic adaptatif sur lequel se trouve la population 1, de valeur sélective forte
- ■ ■ Colline adaptative sur laquelle se trouve la population 2, de valeur sélective plus basse
- ↔ Mutation chez un variant de la population = déplacement dans l'espace séquence

- Sur le plan horizontal se trouve l'espace séquence d'une population virale.
- L'axe des ordonnées correspond à la valeur sélective exprimée pour chaque génotype retrouvé dans l'espace séquence.

Considérons ici deux populations, l'une se répliquant très vite (population 1) et l'autre plus lentement (population 2), étudiées par Lauring (125, Figure 37).

Avec un taux de mutation bas, la population 1, qui se réplique vite, va surpasser les autres car sa descendance sera génétiquement proche et générée plus vite (figures 37 A et C). Ses variants vont être génétiquement stables. En termes de paysage adaptatif, cette population va occuper un pic adaptatif haut et étroit, où l'on retrouve peu de diversité génétique mais une très grande valeur adaptative, avec un variant qui surpasse les autres, la séquence maîtresse, dite aussi type sauvage. La valeur sélective moyenne de cette population surpasse celle d'une population 2, qui se réplique plus lentement. En termes de paysage, cette dernière est située sur une colline plus plate et moins haute.

Chez les virus à ARN, le taux de mutation est élevé : les variants de la population 1 vont se déployer très rapidement autour du pic adaptatif et vont perdre leurs avantages sélectifs (figures 37 B et D). Si une population se réplique vite, sa descendance va présenter une grande diversité, mais de nombreux descendants vont montrer une capacité d'adaptation moins importante que leurs parents. La population 2, située sur une colline plus plate, possède des variants qui vont rester plus près de leur optimum d'adaptation. Si le taux de mutation augmente fortement, les variants de la population 2 vont présenter une capacité d'adaptation moyenne plus haute que la population 1.

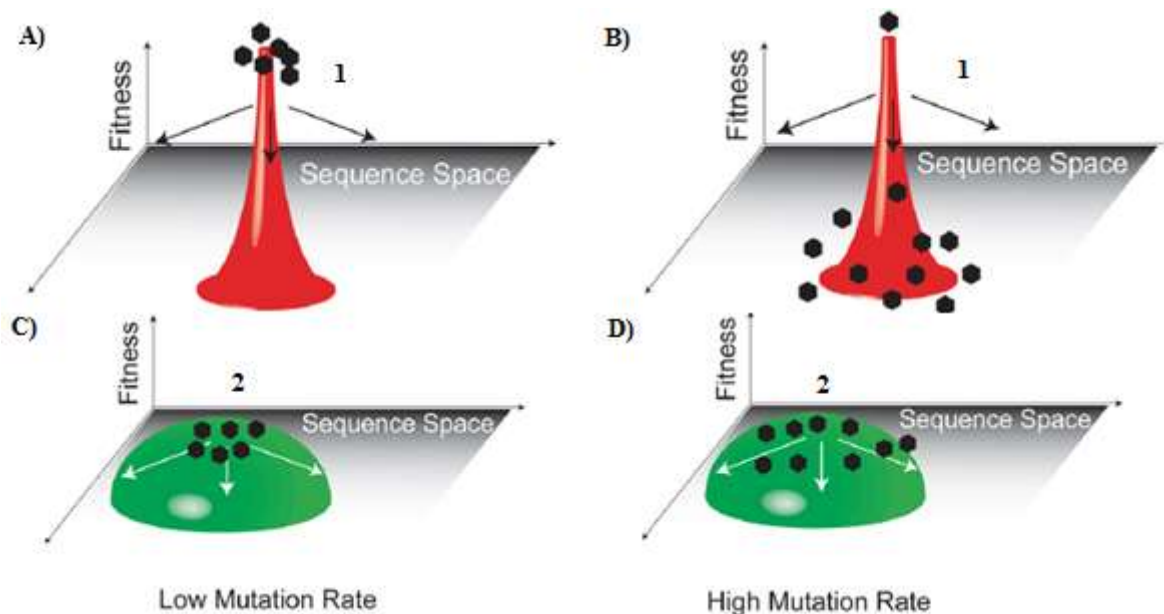


Figure 37 : La survie du plus plat en fonction du taux de mutation (125)

- Pic adaptatif sur lequel se trouve la population 1, de valeur sélective forte
- Colline adaptative sur laquelle se trouve la population 2, de valeur sélective plus basse
- ✂ Mutation chez un variant de la population = déplacement dans l'espace séquence
- Variant des populations 1 et 2

- Sur le plan horizontal se trouve l'espace séquence d'une population virale.
- L'axe des ordonnées correspond à la valeur sélective exprimée pour chaque génotype retrouvé dans l'espace séquence.
- Les figures 37A et 37C décrivent l'évolution des populations 1 et 2 selon un faible taux de mutation.
- Les figures 37B et 37D décrivent l'évolution des populations 1 et 2 selon un fort taux de mutation.

La théorie des quasi-espèces prédit que la population se répliquant moins vite sera favorisée, si elle donne une descendance avec une capacité adaptative moyenne assez haute et si elle est plus robuste vis-à-vis des mutations délétères (50). Ces populations occupent des régions plates sur les paysages adaptatifs : c'est la survie du plus plat.

Les populations évoluant sur des portions plates d'un paysage adaptatif sont plus robustes vis-à-vis des mutations, qui ont un coût moins important de valeur sélective. Ces quasi-espèces forment ainsi une population avec un réseau large et sélectivement neutre de variants dans l'espace séquence. Ce répertoire peut explorer de vastes régions de l'espace séquence sans conséquence et peut ainsi s'adapter à un rapide changement de l'environnement, comme l'a décrit la « shifting balance theory » (79, 119). Cette exploration peut mener à l'occupation de ces vastes régions, donc amener à former ce réseau neutre de variants, mais aussi à occuper des pics adaptatifs voisins, selon l'environnement. La population agit ainsi comme un ensemble de variants agissant en collectivité (143).

#### 4.4.4. Observations expérimentales

Cet effet a été observé chez des organismes digitaux auto-réplicatifs étudiés *in silico* (256), des viroïdes infectant des plantes et soumis à l'effet de rayonnements ultraviolets (50) et chez le VSV soumis à des mutagènes (79, 210). Des études chez le HIV-1 et IV de type A laissent supposer que ceux-ci pourraient se mouvoir vers une région plus plate du paysage pour explorer l'espace séquence (79, 192). Cela pourrait expliquer les alternances de périodes de stases évolutives et de changements phénotypiques soudains.

Ces observations expérimentales vont dans le sens d'une robustesse, qui pourrait notamment être responsable d'une diminution de l'efficacité du traitement utilisant le principe de la mutagénèse létale (50, 79, 100).

#### 4.4.5. Difficultés d'interprétation

Les mécanismes de la robustesse restent en général très mal compris et on manque cruellement de données expérimentales (79). On peut supposer que des mécanismes extrinsèques interviennent : les virus peuvent par exemple recruter des chaperons cellulaires afin de tamponner les effets des mutations. A cela peuvent s'ajouter des mécanismes intrinsèques aux populations (79) :

- La sélection positive élimine les variants avec des mutations délétères si la population est assez grande.
- La polyplôidie retrouvée chez les virus à haute MOI peut augmenter la diversité de la population et assurer une complémentarité (voir partie 5.4.3.).
- L'épistasie négative entre les différentes mutations délétères permettrait de les éliminer.
- Le mode spécifique de réplication des virus pourrait aussi intervenir.

On sait juste que les « fidelity mutants » n'en sont pas une preuve car ils agissent au niveau de l'action de la polymérase. Ils réduisent ainsi le taux de mutation, mais il ne s'agit pas d'un trait phénotypique non altéré par l'acquisition de nouvelles mutations. L'une des explications est que, naturellement, un virus aurait une robustesse plus grande qu'un autre, même entre virus d'un même genre comme le CBV3 et le PV (100). Une autre serait que le virus, lors de sa transition vers l'extinction, atteint des portions plus plates du paysage adaptatif, et est ainsi rendu plus robuste aux mutations, offrant un mécanisme de résistance à la mutagénèse létale (32, 79, 210). Cette hypothèse a été réfutée par une expérience chez le LCMV (139) et un modèle théorique (163).

Une autre difficulté se retrouve dans le fait que la plupart des études théoriques utilisant les paysages adaptatifs en décrivent des très et trop simples. On considère généralement un paysage constitué d'un simple pic adaptatif, dans un environnement constant. Cela ne tient pas compte du caractère mouvant des paysages (voir partie 2.4.).

#### 4.4.6. Conclusion

Malgré le fait que la version accidenté des paysages adaptatifs soit plus réaliste pour les virus, des aires plates de l'espace séquence peuvent être visualisées et sont concordantes avec ce modèle. Ces aires modélisent ainsi des virus qui sont plus résistants aux effets de la mutation. Il se peut donc simplement que la dynamique des quasi-espèces englobe ces deux modèles, avec des populations qui suivent la survie du plus apte, tandis que dans certaines

conditions environnementales et dans le cadre de l'évolution sur le long terme, la survie du plus plat prévaut.

Ainsi, quand le taux de mutation augmente, deux cas sont possibles. Au-delà d'un certain seuil de taux de mutation, des populations présentant un taux de réplication plus bas auraient un avantage sélectif, car la réplication présentera moins d'erreur si elle se réalise moins vite. Et des populations les plus robustes présenteraient une valeur sélective moyenne plus grande (50). Le modèle étudié et les preuves expérimentales apportées prouvent qu'il existe un compromis entre ces deux caractéristiques pour certains modèles subviraux et viraux (125). Si le taux de mutation est trop haut, les populations se répliquant le plus vite sont surpassées par les populations se répliquant moins vite mais étant plus robustes vis-à-vis des mutations. Encore une fois, le spectre de mutants et sa composition sont d'une importance centrale. Cela est bien décrit par le modèle des quasi-espèces.

Ce modèle apporte aussi une observation intéressante : des populations avec une distribution qui ne diffèrent pas grandement quant à leur valeur sélective peuvent parfois surpasser des populations, avec des valeurs sélectives bien plus hautes. Cela pourrait être une conséquence de l'action de la population à l'échelle de l'ensemble plutôt que de l'individu, ce qui est soutenu par la théorie des quasi-espèces.

La survie du plus plat pourrait cependant être considérée comme une exception dans le cadre de l'application de la théorie aux virus. Comme cela est expliqué sur la figure 37, tout dépend du taux de réplication et de mutation. Il se pourrait donc que la survie du plus plat ne se retrouve que dans des conditions particulières, comme par exemple lors de la mutagenèse induite par traitement antiviral ou par l'hôte (210).

#### **4.5. Les facteurs liés à l'hôte**

Comme nous l'avons vu lors de l'augmentation de la capacité d'adaptation, la valeur adaptative est environnement-dépendante (63, 124, 172, 192). A l'échelle d'un organisme, c'est encore plus complexe : il s'agit d'une mosaïque d'environnements (cellules, tissus, organes et conditions physiologiques) dans laquelle la valeur adaptative peut varier constamment. On peut ainsi avoir des virus spécialisés dans un tissu spécifique, dans lequel la valeur adaptative augmentera très vite, mais diminuera s'il ne s'agit pas du tissu adéquat. On peut aussi cependant trouver des virus qui ciblent plusieurs tissus, pour lesquels on peut alors s'attendre à retrouver une compartimentation des valeurs adaptatives (72).

L'environnement va intervenir dans toutes les étapes du cycle viral, et complique bien souvent l'interprétation de l'adaptabilité de la population. Cela s'observe notamment lorsqu'on s'intéresse à l'occupation de l'espace séquence par cette population et son extension lors de la réplication (voir partie 3.3.2.2.), elles-mêmes dépendent des paramètres du virus et de l'environnement de réplication. Les paramètres de l'environnement n'ont pas encore été tous identifiés mais reposent sur, d'une part, les contraintes de l'hôte, d'autre part, la participation de l'hôte (71).

- Des contraintes plus importantes chez un hôte peuvent éliminer une large proportion de mutants chez le virus, alors que cela ne sera pas le cas chez un autre hôte dont les contraintes sont moindres.
- La « participation » de l'hôte correspond à la participation des protéines de l'hôte (dans les complexes avec la polymérase par exemple) et peut ainsi modifier la fidélité de la matrice copiée.

La complexité du spectre de mutants reste ainsi largement imprédictible, étant donnée la multitude de facteurs génétiques et biochimiques impliqués dans la fidélité de la

réplication, et les événements sélectifs. Les interactions complexes et intrinsèques de tous ces facteurs impliquent ainsi forcément une part de stochasticité qu'il ne faut pas non plus exclure.

Des échantillons de génomes, les calculs de distances génétiques, de fréquences de mutations et des entropies de Shannon, des séquençages haut-débit plus poussés permettent de quantifier cette complexité au fur et à mesure, mais il s'agit là d'un travail de longue haleine.

## 4.6. Conclusions

### 4.6.1. Conclusion sur la dynamique des quasi-espèces

Les expériences de repiquage et les conséquences de ces événements d'étranglement fournissent une preuve supplémentaire quant à l'intérêt de l'hétérogénéité et de la grande taille des quasi-espèces virales. Il est remarquable de voir à quel point ces événements ont un impact sur la capacité d'adaptation et les altérations biologiques des virus. Cependant, le transfert réalisé ici est un cas extrême de goulots d'étranglement, car on ne prélève qu'un seul génome. *In vivo*, les virus sont sujets à ces événements d'intensité différente, qui dépendent du nombre de génomes de la population et des conditions de l'environnement.

Ce spectre est un véritable réseau complexe de variants interconnectés, qui présentent des mutations neutres, de manière assez rare, permettant l'exploration de l'espace séquence lorsqu'ils sont situés dans une portion plate et large de l'espace dans le cas de la survie du plus plat ; ou des mutations où une seule d'entre elle va provoquer une baisse brutale de la capacité d'adaptation, car la population se trouve sur un pic adaptatif dans le cas de la survie du plus apte.

L'existence d'un vaste spectre de mutants dans une quasi-espèce virale permet au virus de s'adapter rapidement à des pressions sélectives ou de changer de tropisme tissulaire. De plus, les agents infectieux secondaires peuvent induire une réponse de l'hôte, causant ainsi de nouvelles conditions environnementales et donc une sélection de certains variants de cette quasi-espèce. Ces pressions de sélection peuvent éliminer ou restreindre la réplication de la plupart des quasi-espèces virales au sein d'une population, laissant ainsi une fraction mineure de variants adaptés capables de se reproduire. Elles vont aussi altérer le précédent ratio de variants, permettant ainsi l'expression d'un phénotype nouveau et différent de la quasi-espèce originelle, à cause de la nouvelle distribution. L'émergence de variants viraux potentiellement dangereux peut être tenue en échec par cette dynamique complexe des quasi-espèces.

L'étude de l'augmentation et de la diminution de la valeur adaptative des quasi-espèces permet de proposer un modèle évolutif, en tenant compte des différents paramètres vus précédemment :

Un événement sélectionnant une large partie de la population va permettre, de manière très probable, une augmentation de sa valeur adaptative (figure 38 A). Une grande taille et une grande hétérogénéité se retrouvent dans une nouvelle population, qui possède une nouvelle composition mais avec conservation de la séquence consensus. Ces deux paramètres fournissent une plus grande capacité d'adaptation.

Au contraire, un événement sélectionnant une très faible portion de la population (ici, un seul virion) va diminuer l'hétérogénéité et la taille de la population (figure 38 B). Lors des réplifications suivantes, la séquence consensus est très largement modifiée et la capacité d'adaptation de la population s'en retrouve très fortement diminuée, car elle ne présente pas d'hétérogénéité.



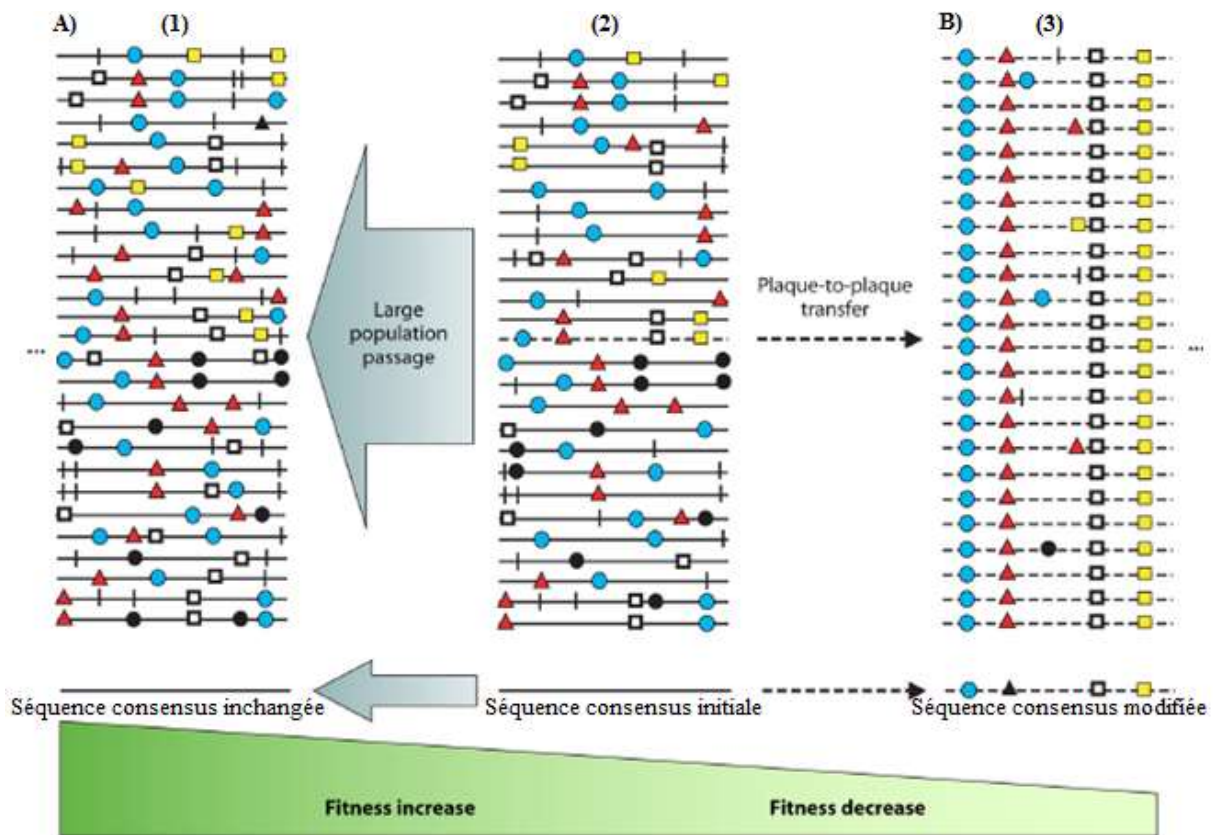


Figure 38 : Représentation schématique de l'évolution des populations de virus à ARN (71)

- Séquence nucléotidique
- - - - - Génome viral sélectionné lors d'un repiquage en série (goulot d'étranglement)
- - ● - + - ▲ Mutations dans la séquence nucléotidique
- \* Mutation fournissant un avantage sélectif
- ➡ Passages en série (transfert d'une large population)
- ➡ Passage en série (en regard de la séquence consensus)
- > Repiquage d'un seul virion
- Valeur sélective évoluant en fonction de la distribution de la quasi-espèce



La fraction sélectionnée est d'une très grande importance pour la suite de l'évolution de la population, comme l'illustre la représentation schématique de la figure 39.

Lorsqu'on prélève une petite portion de la population, on ne trouvera pas de variants résistants (figure 39 A). De plus amples réplifications virales du spectre mutant seront nécessaires pour générer des mutants résistants.

Si l'on prélève une population de taille intermédiaire, deux des variants résistants seront trouvés lors de l'analyse (figure 39 B).

Si l'on prélève une population de grande taille, tous les variants sont représentés si on l'analyse (figure 39 C).

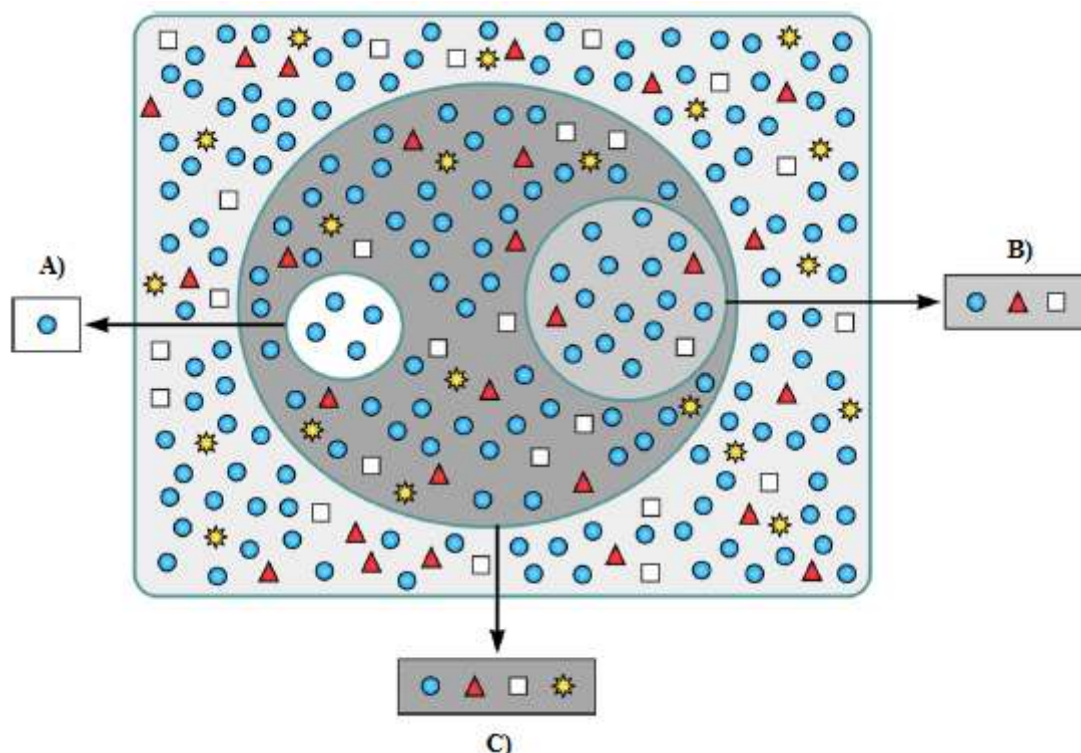


Figure 39 : représentation schématique de l'effet de la taille de la population sur l'évolution (71)

- Le rectangle bleu représente une quasi-espèce virale.
- Chaque symbole (même ceux qui ont la même couleur et la même forme) décrit un variant présentant une séquence nucléotidique légèrement différente.
- Variant viral de la population exprimant un même phénotype
- ▲ ★ □ Variant viral de la population comportant des mutations conférant une résistance
- Prélèvement d'une part faible de la population
- Prélèvement d'une part intermédiaire de la population
- Prélèvement d'une part importante de la population

Cette représentation schématique apporte deux points importants :

- La taille et l'hétérogénéité de la population virale peuvent conditionner la transmission virale intra-hôte ou inter-hôte ou l'évolution expérimentale. Une population plus grande et plus hétérogène présentera une plus grande variété de mutants, donc une plus grande probabilité de pouvoir s'adapter à des changements de conditions environnementales (71).
- Chez de telles populations, il faut prélever un plus grand nombre de clones pour maintenir la valeur adaptative au sein de la nouvelle population (22). Cela peut être expliqué par la probabilité de prendre, lors du transfert, des variants avec une valeur adaptative plus petite, égale ou supérieure à la fitness de la population initiale. Plus la valeur sélective de la population initiale est grande, plus le nombre de variants subissant un goulot d'étranglement pour maintenir une valeur sélective égale à celle de la population de départ.

#### **4.6.2. Paramètres viraux et impact sur l'adaptabilité du virus**

Nous pouvons en conclure que l'adaptabilité du virus va dépendre des paramètres de la population virale, de l'environnement (c'est-à-dire de l'hôte), desquels on ne peut exclure un élément fort de l'évolution virale sur le court terme (intra-hôte) ou au long terme (inter-hôte) : la stochasticité.

Les paramètres viraux qui ressortent des observations sont (71) :

- La taille de la population : elle va déterminer le nombre et les types de variants retrouvés dans la population, ainsi que son caractère infectieux.
- L'hétérogénéité (le nombre moyen de mutations par génome calculé pour les composants d'un spectre viral). Il est d'une fourchette de 1 à 100 mutations par génome, qui dépend du taux d'erreur de réplication et de la distance par rapport à la séquence d'origine dans l'espace séquence. Il est important de retenir que cette hétérogénéité est une conséquence des événements de mutations et de sélection, et n'équivaut pas au taux de mutation en lui-même.
- La fécondité, dit rendement moyen des particules virales par cellule infectée. Une grande fourchette de valeur ( $<1$  à  $10^6$  particules produites par cellule) a été rapportées dans le cas d'infections dans la nature. Une haute fécondité favorise l'exploration de l'espace séquence. C'est un paramètre qui est intimement lié aux deux premiers.
- La taille du génome : un génome de petite taille permet une exploration et une occupation plus efficace de l'espace séquence, sans subir le franchissement du seuil d'erreur.
- Le nombre de mutations nécessaire à un changement phénotypique. Il est important de retenir qu'une ou même très peu de mutations peuvent être suffisantes pour modifier les traits phénotypiques d'un virus. Si ce nombre était élevé (de l'ordre d'une centaine), ces paramètres sur lesquels reposent le principe des quasi-espèces ne seraient pas pertinents. Dans certaines conditions de réplication, la population peut faire preuve d'une certaine robustesse et ces mutations n'ont pas l'impact phénotypique escompté : c'est la survie du plus plat.

Les paramètres de l'hôte vont exercer une influence sur les paramètres viraux, via :

- Les contraintes de l'hôte.
- La participation de l'hôte.

Tous ces paramètres sont des facteurs qui sont cohérents avec la théorie des quasi-espèces. Le fait qu'ils soient interconnectés, pour certains facilement mesurables, d'autres plus difficilement, compliquent encore l'interprétation de la dynamique d'évolution des populations virales. Tous déterminent la complexité du spectre de mutants.

## **Partie 5 Implications biologiques de la dynamique des quasi-espèces**

Les paramètres et la dynamique des quasi-espèces décrits jusqu'ici apportent des éléments importants : un panel important de variants, la notion de seuil d'erreur qui limite la taille du génome et est étroitement lié aux taux de mutation lors de la réplication, et des mutations qui ont des conséquences différentes selon le contexte et l'environnement. Mais leurs implications biologiques vont bien au-delà d'une population constituée d'un simple ensemble de variants soumis à des mutations avec un génome de petite taille.

Les points qui manquaient à la génétique des populations pour expliquer et comprendre le comportement d'une population virale pourraient bien être apportés par la théorie des quasi-espèces. La composition du spectre de mutants, notamment les sous-populations retrouvées à basses fréquences, pourrait jouer un rôle dans l'apparition de résistances, dans le changement de tropisme cellulaire ou encore dans les épisodes d'émergences et de réémergences. De possibles interactions d'interférence et de complémentarité pourraient expliquer le maintien de certains variants qui semblent pourtant moins adaptés, voire moduler la virulence virale. Cette dynamique suppose même l'existence d'une mémoire au niveau moléculaire et une sélection densité-dépendante. Enfin, le seuil d'erreur pourrait bien avoir des implications biologiques pour la mise en place d'une nouvelle stratégie thérapeutique.

### **5.1. Résistance et mutants d'échappement**

Beaucoup de virus à ARN dont les rétrovirus n'ont besoin que de quelques semaines pour échapper à la réponse immunitaire ou présenter des variants qui peuvent résister aux molécules utilisées lors des traitements antiviraux (86). L'importance du spectre de mutants décrit par la théorie des quasi-espèces appliquée aux virus à ARN, et des variants exprimés en minorité, prend ici tout son sens.

#### **5.1.1. Les contraintes sélectives appliquées lors de traitements**

La dynamique d'une quasi-espèce décrite jusqu'ici présentait la réponse d'une population virale à des contraintes sélectives. Dans le cadre de l'infection d'un nouvel hôte, le virus doit aussi subir une série de pressions sélectives. Ces contraintes sélectives peuvent être apportées par des traitements thérapeutiques, tels que des médicaments antiviraux ou la vaccination, ou par des composants de la réponse immunitaire innée ou acquise. Ces contraintes vont ainsi avoir affaire à un spectre de mutants dynamique et complexe.

La vaccination permet de prévenir une atteinte virale, tandis que les traitements de chimiothérapie (les antiviraux) ou d'immunothérapie permettent de traiter l'infection virale déjà présente. Il existe globalement deux stratégies afin de résister à ces contraintes. Le pathogène peut les moduler, c'est-à-dire altérer les composants de cette contrainte via ses produits d'expression, ce qui permet de faciliter sa persistance et sa transmission. Mais il peut aussi « fuir », échapper aux contraintes en résistant à l'inactivation qu'il subit, en se modifiant. Etant donné la forte variabilité et la compaction de l'information génétique, c'est principalement cette deuxième stratégie que l'on retrouve chez les virus à ARN.

La littérature rapporte un nombre incroyable de sélection de mutants résistants à tous ces procédés sensés lutter contre l'atteinte virale, si ce n'est pour chaque traitement mis en place (72). Ce sont des mutants d'échappement. Ces résistances ont pour conséquence un échec du traitement sur le court terme, mais aussi une sélection à l'échelle des populations touchées sur le long terme. Cela a donc des conséquences épidémiologiques non négligeables.

### 5.1.2. Définition des mutants d'échappement

Les mutants d'échappement sont des sous-ensembles de variants présents dans le spectre de mutants, qui peuvent se répliquer malgré l'action d'une contrainte sélective qui entrave la réplication des génomes qui représentent la majorité de la population (71). Ces mutants présentent des substitutions d'acides aminés, dues à des mutations dans la séquence génomique, qui vont leur permettre d'échapper à la contrainte imposée. Les apparitions de tels mutants semblent être nombreuses *in vitro* et *in vivo* et sont une manifestation pertinente de la présence et de l'apparition de variants capables d'échapper à des contraintes sélectives particulières, dans le spectre de mutants chez des populations virales. On retrouve ces mutants d'échappement aussi bien pour les inhibiteurs antiviraux que pour la réponse immunitaire, et donc aussi lors de la vaccination.

### 5.1.3. Résistance aux inhibiteurs antiviraux

#### 5.1.3.1. Les inhibiteurs antiviraux

Un nombre toujours plus important d'agents antiviraux sont développés en se basant sur la structure tridimensionnelle des protéines et la formation de complexes avec des ligands aussi bien naturels que synthétiques (72). Ils peuvent intervenir à tous les niveaux du cycle réplcatif viral afin de diminuer la charge virale : la fixation à la membrane cellule, la pénétration dans la cellule, le démantèlement de la capsid, ou le virion lui-même, empêchant sa réplication, transcription ou traduction. Un seul inhibiteur cible cependant une étape.

Leur diversité structurale est assez conséquente. Cela peut aller de simples analogues nucléotidiques à des composés hétérocycliques complexes. Cependant l'utilisation de ces inhibiteurs mènera souvent à une sélection de mutants multi-résistants si l'exposition aux médicaments est prolongée, et ainsi à l'échec du traitement (98, 198, 211).

#### 5.1.3.2. Mécanismes et caractéristiques généraux des résistances

Les substitutions d'acides aminés peuvent rendre l'inhibiteur inefficace selon plusieurs mécanismes (72) :

- Elles peuvent diminuer l'affinité de la protéine pour la molécule inhibitrice. Pour les enzymes polymérase, cela équivaut à la diminution de la sélectivité des nucléotides.
- Si la cible de l'inhibition agit elle-même une autre structure protéique, virale ou cellulaire (agissant ainsi en *trans*), une modification de cette dernière peut aussi provoquer une résistance (225, 238).
- Chez les polymérase virales, des mutations peuvent provoquer l'excision d'un nucléotide terminateur, à l'extrémité 3' du brin synthétisé.

### 5.1.3.3. Sélection de mutants résistants

Leur sélection dépend de plusieurs facteurs : des barrières génotypique et phénotypique à surpasser pour obtenir une mutation conférant une résistance, des interactions qui peuvent avoir lieu au sein du spectre de mutants (voir partie 5.4.), de la taille de la population et sa charge répliquative, et de la nature et la concentration des inhibiteurs utilisés et surtout de la population étudiée et des précédentes contraintes qu'elle a pu subir (voir partie 5.3.). Il est très difficile de comprendre comment ces différents facteurs permettent d'obtenir le niveau de résistance que l'on peut observer pour une particule virale donnée, dans une cellule donnée située dans un organe ou un tissu donné d'un individu bien particulier, à un instant donné.

La résistance est un phénomène général, mais prouve la spécificité des traitements antiviraux, car elle passe par une sélection spécifique de variants résistants à un type de traitement particulier (196).

#### 5.1.3.3.1. *Barrière génétique*

La notion de barrière a déjà été abordée dans la partie 4.4.1. L'apparition du phénotype auquel on se raccorde ici est la résistance. En fonction de cette barrière génétique, la probabilité d'apparition de la mutation conférant une résistance va être différente. Ce concept, qui intervient déjà dans une sélection de mutations conférant une résistance difficile à appréhender, est lui-même assez complexe. Il est difficile à quantifier, car beaucoup de facteurs s'ajoutent à son établissement et n'a rien d'universel. La barrière génétique peut en effet être différente même en considérant un même médicament. Elle dépend aussi de la présence des autres traitements mis en place, de la concentration des nucléotides dans le milieu cellulaire ou de l'abondance des ARN de transfert.

Un paramètre déjà traité revêt ici toute son importance (voir parties 4.2. et 4.6.) : la taille de la population semble aussi jouer un rôle et va permettre d'augmenter les chances de surpasser la barrière génétique. Pour reprendre l'approche intuitive présentée précédemment, on peut considérer une fréquence de mutation conférant une résistance de  $10^{-4}$ . Si la population, pour une raison ou pour une autre, est incapable d'atteindre une taille de  $10^4$  particules, il est peu probable que cette résistance apparaisse (180).

#### 5.1.3.3.2. *Barrière phénotypique*

La barrière phénotypique correspond au coût de la mutation en termes de valeur sélective. Si le coût pour acquérir une résistance est fort, il y aura une probable réversion vers le type sauvage quand les variants se répliqueront en l'absence de la contrainte sélective, ou apparition de mutations compensatrices afin que la valeur sélective des variants résistants augmente (106, 154). Ces deux voies ne s'excluent d'ailleurs pas et participent à l'établissement de multiples voies d'évolution différentes.

Cette barrière est souvent établie de manière empirique grâce à la fréquence d'apparition des mutations lors de cultures cellulaires ou chez des patients traités avec l'inhibiteur. Mais pour le moment, seules les mesures de valeur sélective lors d'expériences de compétition entre le type sauvage et le virus permettent de l'appréhender de manière correcte (72). Elles vont donner un aperçu de la force sélective d'une mutation : si cette force est grande, cela signifie que la mutation accorde un grand avantage sélectif, mais qu'elle a un coût de valeur sélective très fort. Le principe est le même que présenté dans la partie 2.3.3.,

avec le virus de type sauvage et le virus résistant à l'inhibiteur, en présence et en l'absence de l'inhibiteur, ce qui donne des vecteurs de valeur sélective (voir figure 9). Une approche intuitive passe par l'observation des pentes de ces deux vecteurs : s'ils s'éloignent de manière très forte, on parlera de grande force sélective. Si leur pente est moins prononcée et qu'ils s'éloignent moins vite, elle sera faible. Pour obtenir une valeur chiffrée, on peut faire le rapport des mesures de valeur sélective du type sauvage et du type muté au même passage. Encore une fois, si la valeur de ce rapport est grande, la force sélective est dite grande. Si la mutation n'a pas de coût de valeur sélective, la mesure de la valeur adaptative correspond à la force sélective.

#### 5.1.3.3.3. *Concentrations et caractéristiques de l'inhibiteur*

Lors de l'utilisation d'un inhibiteur antiviral, il faut considérer que l'organisme est compartimenté. Il se peut donc que les concentrations en inhibiteur ne soit pas les mêmes d'un compartiment cellulaire à un autre. La fréquence dépend de cette concentration : plus cette dernière augmente, plus la fréquence de la mutation diminue (194). Cela met aussi en avant les sites servant de refuge (« sanctuary sites » dans la littérature (194)), où la capacité répliquative du virus est plus grande, même lors de traitement. Leur implication n'est cependant pas tout à fait comprise quant à l'émergence de résistance dans la population. A l'échelle de l'organisme entier, ce qui prend en compte tous les compartiments, on obtient donc une fréquence moyenne. L'analyse de la fréquence des mutations des particules circulantes peut donner une bonne approche de cette fréquence moyenne.

Les caractéristiques des inhibiteurs sont testées sur culture cellulaire (72) :

- On teste la toxicité de la molécule sur les cellules hôtes infectées. On mesure pour cela le pourcentage de cellules survivantes en fonction d'une concentration d'inhibiteurs. On cherche ainsi à déterminer la concentration pour laquelle 50% des cellules survivent : la concentration cytotoxique ( $CC_{50}$ ).
- On teste la capacité d'inhibition de la production de particules virales. On mesure ainsi le pourcentage de particules inhibées en fonction d'une concentration d'inhibiteurs. On peut déterminer la concentration inhibitrice pour laquelle 50% des particules sont inhibées : la concentration inhibitrice ( $IC_{50}$ ).
- Le rapport  $CC_{50}/IC_{50}$  donne l'indice thérapeutique du médicament.
  - Si  $CC_{50}/IC_{50} > 100$ , on dit qu'il est excellent.
  - Si  $10 < CC_{50}/IC_{50} < 100$ , on dit qu'il est acceptable.
  - Si  $CC_{50}/IC_{50} < 10$ , on dit qu'il est d'efficacité limitée.

Cette approche est très simplifiée : les caractéristiques dépendent encore une fois des populations virales sur lesquels les inhibiteurs sont utilisés, de l'environnement etc. Mais elles permettent tout de même d'aider à déterminer les concentrations d'utilisation.

#### 5.1.3.3.4. *Apparition de résistance sans exposition aux inhibiteurs*

Des cas d'apparitions de résistance sans exposition aux agents concernés ont été relevés lors de certaines expériences, notamment pour le HIV-1 (118) ou le HCV (198, 216), parmi beaucoup d'autres, grâce aux NGS. Un modèle mathématique basé sur la dynamique d'évolution du HIV-1 démontre que l'échec presque universel de tous les traitements pourrait être dû à l'existence de sous-populations de variants résistants, à basse fréquence, dans le spectre de mutants, avant même la mise en place des traitements (194). L'observation récente de la cinétique d'émergence d'une résistance à un inhibiteur chez le HCV à l'aide de lignées

cellulaires et de marquages fluorescents oriente aussi vers l'existence de variants préexistants dans la population virale (198).

Il s'agit d'une conséquence directe de la dynamique des quasi-espèces et du fort taux de mutation et de variabilité que l'on retrouve au sein d'une population virale. Un spectre de mutants large permet encore une fois l'exploration de l'espace séquence et l'acquisition de variants permettant de se confronter à des contraintes sélectives. Il s'agit d'un véritable réservoir phénotypique qui permet d'anticiper ces contraintes. L'un des objectifs est cependant de déterminer, à l'aide des NGS, les fréquences basales de ces mutations accordant une résistance, afin d'indiquer quand il faut éviter l'utilisation d'un inhibiteur.

#### 5.1.3.4. Dynamique d'apparition des résistances

Tous ces paramètres peuvent laisser penser que la barrière pour atteindre une résistance est assez haute. Cependant cela ne semble pas être un phénomène rare dans la dynamique d'évolution des espèces virales, qu'elles soient à ARN ou ADN (86, 196). Dans la littérature étudiée, on en retrouve des cas *in silico*, *in vitro* et *in vivo*, chez le HIV-1 (118, 194, 196, 235), le HCV (187, 198, 211), le FMDV (172, 173). De même, des résistances ont été décrites pour chacune des molécules qui ont été utilisées pour traiter les infections humaines à influenza (71, 196). Plusieurs références rapportent une fréquence de mutants résistants aux inhibiteurs de l'ordre de  $10^{-3}$  à  $10^{-6}$  (70, 196), même si nous avons vu que ce paramètre était limité.

Il est possible que ces résistances puissent être dues à un couplage entre des faibles barrières génétique et phénotypique, et au fait que les voies évolutives menant à un même phénotype sont multiples. Des mutations différentes ont par exemple permis d'acquérir une résistance à l'interféron  $\alpha$  (IFN- $\alpha$ ) chez le HCV (182), ou une résistance du rhinovirus humain (Human Rhinovirus, HRV) à l'inhibiteur WIN 52084 (107). Le cas de la résistance à la ribavirine a aussi été étudié (voir partie 5.5.5.3.). En substance, bien que s'appliquant aux agents mutagènes, cela reflète parfaitement le fait qu'une cascade de mutations puisse induire des substitutions d'acides aminés différentes, mais mener à un même phénotype, la résistance à la ribavirine. Cela peut s'expliquer par le fait qu'il existe des interactions de courte et de longue portée au sein d'une même molécule protéique. De multiples combinaisons de substitutions d'acides aminés chez cette dernière sont donc possibles pour diminuer son affinité pour le ligand, c'est-à-dire la molécule antivirale (225, 238). Cependant, si les voies évolutives sont multiples, il est possible que l'efficacité de résistance offerte soit nivelée, comme cela a été vu chez le HRV (107). Enfin, quand plusieurs substitutions sont impliquées dans l'apparition d'une résistance, certaines mutations n'interviendraient pas directement dans la perte d'affinité pour l'antiviral mais permettraient d'augmenter la valeur sélective des variants, comme une mutation compensatrice.

Ces mutations ont de grandes chances d'être présentes dans le spectre de mutants d'une quasi-espèce (118, 194, 196, 198), et même si ce n'est pas le cas, l'acquisition de cette résistance peut avoir lieu lors de répliquations suivantes *de novo*, chez les variants n'ayant pas été éliminés par les inhibiteurs car leur efficacité n'est pas parfaite ou qu'il y a une mauvaise observance (194). L'apparition *de novo* serait cependant moins probable (194, 198). Cette acquisition semble être facilitée par le fait que plusieurs mutations puissent permettre d'arriver au même résultat. En présence de l'inhibiteur, les variants présentant ce phénotype de résistance vont être sélectionnés et augmenter de fréquence, voir acquérir des mutations supplémentaires permettant d'augmenter le caractère de résistance (154). Les mécanismes qui accordent la résistance augmentent considérablement la valeur sélective en présence des antiviraux, mais celle-ci diminue lorsqu'ils disparaissent du milieu, c'est-à-dire à l'arrêt du



traitement. On parle de pharmacodépendance (70, 154). Leur fréquence va diminuer de nouveau si la barrière phénotypique est trop grande, ou rester constante si le coût de valeur sélective est faible ou si des mutations compensatrices apparaissent dans les phases de réplication suivantes (106, 154, 196). Ce n'est donc pas parce que des mutations apparaissent qu'elles vont ensuite contribuer à la diversité antigénique de la population virale.

Les implications de cette dynamique au sein du spectre se retrouvent dans les domaines de santé publique : de nombreux inhibiteurs, nouvellement confectionnés et rapportés comme très efficaces, offrent généralement de grandes promesses au moment de leur arrivée sur le marché. Mais avec le recul de leur utilisation, on peut le plus souvent observer l'apparition de résistance. La prise en compte de l'apparition de ces résistances est maintenant essentielle dans le développement des inhibiteurs et dans la mise en place d'un protocole thérapeutique et sa gestion clinique (196). Ces implications poussent la recherche antivirale vers l'utilisation des thérapies combinées et de nouvelles stratégies utilisant d'autres propriétés des quasi-espèces, afin d'augmenter les barrières phénotypiques et génétiques à surpasser pour obtenir une résistance (voir parties 5.5. et 5.6.).

#### 5.1.3.5. Un nouveau mécanisme de résistance

Un nouveau trait accordant un mécanisme de résistance à de nombreux inhibiteurs a été mis en évidence chez le HCV et semble en lien direct avec la valeur sélective (216). La résistance du HCV à l'IFN- $\alpha$  a été étudiée en soumettant une population à 100 repiquages en série en présence ou en l'absence d'IFN- $\alpha$ . Dans tous les cas, les populations ont démontré une résistance (partielle mais significative en l'absence d'IFN- $\alpha$ ) (182). Ces résultats ont poussé à essayer de comprendre si cette résistance était due à la présence de composants minoritaires qui auraient été sélectionnés par la dynamique des quasi-espèces (216). La même expérience, mais sans IFN- $\alpha$ , a été mise en place, et les populations à 0 repiquage, 45 repiquages et 100 repiquages ont été analysées. Pour les deux dernières, une résistance à quatre autres inhibiteurs de la réplication chez le HCV a été enregistrée, ainsi qu'une augmentation de la valeur sélective, ce qui pourrait effectivement être due à la présence de variants résistants exprimés à basse fréquence. Mais une analyse de la population par séquençage n'a pas mis en évidence de mutants résistants spécifiques, même à très basse fréquence. D'autres expériences, notamment le clonage biologique issu des populations à 0 et 100 repiquages, ont assuré l'élimination des variants minoritaires en mimant un goulot d'étranglement. Pourtant, les clones démontraient exactement la même valeur sélective et une résistance correspondant à celle de la population parentale.

Il existe donc un élément en lien avec la valeur sélective qui permet d'accorder une résistance à plusieurs molécules inhibitrices chez le HCV. Il est encore trop tôt pour le généraliser à d'autres virus, mais il est possible que l'augmentation de la valeur sélective, qui est en lien direct avec la capacité répliquative, permette l'augmentation des capacités des unités répliquatives, donc leur concentration. Cette augmentation permettrait de lutter contre les inhibiteurs lors d'infections chroniques chez le HCV (216).

## 5.1.4. Résistance à la réponse immunitaire et à la vaccination

Des mutants d'échappement sont décrits dans le cadre de la réponse immunitaire, mais aussi lors de la mise en place de protocoles de vaccination, qui stimule cette immunité. Encore une fois, la dynamique des quasi-espèces virales et l'importance des variants au sein du spectre de mutants vont jouer un rôle dans l'évolution à l'échelle de la population virale.

### 5.1.4.1. Mutants d'échappement face à la réponse immunitaire

L'immunité consiste en un ensemble de mécanismes de défense d'un organisme vivant face à des agents étrangers qui menacent son bon fonctionnement. Le système immunitaire est constitué d'un ensemble de cellules, tissus et molécules qui servent à la communication et la coordination de ces derniers. Lors de la reconnaissance d'antigènes reconnus comme étrangers à l'organisme, les antigènes du non-soi, il va déclencher une réaction coordonnée de ses composantes : c'est la réponse immunitaire. Lors d'une infection virale, elle cible donc plusieurs étapes du cycle, contrairement à un traitement antiviral qui va cibler seulement une étape bien spécifique. Cela se traduit par de multiples contraintes imposées au virus. Elles sont de deux origines : l'immunité innée, ou naturelle, et l'immunité adaptative ou acquise (71).

#### 5.1.4.1.1. *Immunité innée*

Elle est non spécifique de l'agent pathogène mais est immédiate. Elle repose sur des mécanismes moléculaires et cellulaires (129).

- On retrouve le système du complément, les cytokines (IFN- $\gamma$ , TNF- $\alpha$ ), les chémokines, les protéines de l'inflammation dans les mécanismes moléculaires.
- La composante cellulaire consiste en l'utilisation de cellules à fonction phagocytaires : polynucléaires neutrophiles, les monocytes, les macrophages, ou encore les cellules Natural Killer.

#### 5.1.4.1.2. *Immunité adaptative*

Cette réponse se développe lors de l'exposition à un agent pathogène infectieux et va permettre une protection de l'individu si une deuxième rencontre avec le même agent a lieu. Elle est mise en place plus tardivement, mais est spécifique du pathogène. Elle possède une composante cellulaire et humorale.

- L'immunité humorale est dirigée contre des antigènes de surface, à l'aide d'anticorps neutralisant sécrétés par les plasmocytes, différenciation finale des cellules de la lignée B. Ces anticorps facilitent l'activité phagocytaire de macrophages, de neutrophiles ou de mastocytes recrutés sur place.
- La réponse à médiation cellulaire fait intervenir les lymphocytes B, T et les cellules présentatrices d'antigènes.

Les effecteurs tardifs participent aussi au recrutement de composants de la réponse immunitaire innée, contribuant au nettoyage de l'infection virale.

#### 5.1.4.1.3. *Multiples exemples de mutants d'échappement*

Bien que le mécanisme d'action des molécules antivirales soit différent de la réponse immunitaire, la dynamique des quasi-espèces a aussi un impact sur la survie du virus et sur la réplication lors de la réponse immunitaire de l'hôte. De nombreux variants mutants présentant une résistance ont été isolés et sont utilisés afin de comprendre certains paramètres de la dynamique d'une population virale. Parmi eux, on retrouve des mutants résistants aux anticorps monoclonaux, variants appelés « Monoclonal Antibody resistant », dits « Mab-resistants » (MAR ou MARL selon les sources bibliographiques) chez le VSV (110) et chez le FMDV (16, 28, 140, 167, 201, 204). Ils sont étudiés pour leurs conséquences sur la modification du tropisme cellulaire (voir partie 5.2.4.). Lors d'infections chroniques retrouvées chez le LCMV (47, 129) ou chez le HCV (182, 187), des variants, résistants aux lymphocytes T-cytotoxiques et à l'IFN- $\alpha$  respectivement, sont décrits. Ils vont tous jouer un rôle important dans l'établissement ou dans le maintien de la persistance virale *in vivo* et *in vitro*.

#### 5.1.4.2. Mutants d'échappement dans le cadre de la vaccination

La vaccination est un moyen de contrôler certaines maladies souvent très graves et de prévenir l'apparition de leurs symptômes. Cette méthode a prouvé son efficacité, par l'éradication de certaines maladies comme la variole. Cependant certaines maladies, comme le syndrome d'immunodéficience acquise (235), sont réfractaires à ce traitement. D'autres, comme l'influenza ou la fièvre aphteuse, nécessitent des mises à jour constantes de leurs vaccins.

##### 5.1.4.2.1. *Principes de la vaccination*

La vaccination vise à stimuler activement l'immunité d'un individu de façon à le protéger contre un agent pathogène ou une maladie. Le principe est d'essayer de mimer la présence d'un agent pathogène chez l'hôte afin de mettre en place une réponse immunitaire sans déclencher l'apparition des signes cliniques associés (145).

Pour être efficace, la réponse immunitaire qui va être induite par le vaccin doit posséder plusieurs caractéristiques. Elle doit être spécifique, c'est-à-dire qu'elle ne reconnaît spécifiquement que certains antigènes du pathogène et la réponse va être adaptée à ces sites antigéniques en particulier. Ces derniers sont dits immunogènes. Enfin, elle doit mettre en place une mémoire immunitaire suite à la primo-infection qui a été mimée par le vaccin. Cette mémoire se traduit par l'apparition d'effecteurs qui persistent plus ou moins longtemps et qui vont induire une réponse plus rapide et plus efficace lors d'une nouvelle rencontre avec l'agent pathogène présentant le site antigénique concerné. Elle peut alors soit éliminer le virus complètement, soit limiter l'apparition de signes cliniques mais en laissant le virus se répliquer.

Il peut y avoir circulation d'un ou plusieurs types antigéniques pour un même virus. Il possède en effet de nombreux sites antigéniques, chacun d'entre eux composés d'épitopes, chevauchants ou non, au sein de protéines S et NS. Si l'on veut une protection optimale, un vaccin doit correspondre à la composition antigénique virale. Ainsi, si le virus possède un seul sérotype comme le HAV (7), un vaccin monovalent peut suffire. Mais si plusieurs sérotypes sont présents, comme pour le FMDV, les vaccins vont être multivalents. La fabrication du vaccin passe donc par une analyse de la composition antigénique des virus circulants. Mais

elle passe aussi par une connaissance structurale des virus et de ses propriétés, de sa stabilité ou de son instabilité antigénique, de l'efficacité et des risques des types de vaccins que l'on peut utiliser (vaccin vivant atténué, inactivé, peptidique...) (145).

#### 5.1.4.2.2. *Echappement à la vaccination*

La dynamique générale est la même que pour l'apparition d'une résistance vis-à-vis d'un inhibiteur. Au sein d'une quasi-espèce virale, un spectre de mutants complexe va contenir de multiples variants, dont certains vont présenter un site antigénique modifié qui est censé être la cible de la réponse immunitaire. Si ce n'est pas le cas, la forte variabilité intra-quasi-espèce pourra permettre l'apparition de ces modifications. Suite à une pression sélective, par exemple des anticorps monoclonaux précédemment stimulés par un vaccin dirigé contre la population virale, ces variants vont être sélectionnés et augmenter de fréquence. Une fois la contrainte sélective passée, c'est-à-dire le retrait des anticorps, la variation de fréquence va être la même que celle décrite lors de la résistance aux inhibiteurs. L'échappement décrit va ainsi avoir plusieurs conséquences : il va bien entendu réduire l'efficacité du vaccin, mais aussi augmenter la persistance virale, et surtout orienter l'évolution virale, indirectement grâce au vaccin.

L'efficacité des vaccins utilisés va donc dépendre des mécanismes d'apparition de mutations au site antigénique visé et conférant une résistance à la réponse immunitaire induite par le protocole vaccinal. Cela dépend entièrement des barrières phénotypique et génétique d'une mutation, qui confère par exemple un échappement à la reconnaissance des lymphocytes T-cytotoxiques ou à la neutralisation par les anticorps. En fonction des barrières génétique et phénotypique d'une mutation, on devrait s'attendre à des variations de fréquences des mutations de résistance selon les virus. Par exemple, on peut s'attendre à une fréquence très élevée chez des virus présentant de nombreux sérotypes. L'annexe 6, qui référence différentes fréquences de mutants d'échappement, montre qu'elles sont en fait toutes situées entre  $10^{-3}$  et  $10^{-6}$ . Il se peut donc que ces fréquences sous-estiment le taux d'apparition réel de ces substitutions d'acides aminés. Elles ne permettent pas d'appréhender le caractère évolutif des mutations, mais l'étude de la valeur sélective le peut (voir partie 5.1.3.).

De nombreuses approches thérapeutiques sont basées sur l'induction de réponses immunitaires visant des épitopes hautement conservés de protéines virales. Un exemple bien décrit a été rapporté chez le HAV et son unique sérotype (7, 208). On pense qu'il y a une faible variabilité au niveau de sa capsidie car il y aurait maintien des codons dans l'ORF (Open reading Frame) chez le virus, et des contraintes structurales protéiques agissant au niveau protéique. Or les séquences correspondant aux épitopes des sites antigéniques du virus se trouvent à proximité de ces régions conservées du génome. Cela devrait normalement prévenir l'apparition de nouveaux variants antigéniques suffisamment éloignés génétiquement pour que l'on parle de nouveau sérotype. Malgré cela, des mutants d'échappement à une stimulation immunitaire du HAV ont été décrits, présentant une faible valeur sélective et des substitutions d'acides aminés près des sites antigéniques (7, 183), et rappelant une dynamique de mutation et de sélection que l'on retrouve chez les quasi-espèces (208). Les vaccins peptidiques utilisés chez le FMDV sont un autre exemple de répertoire épitopique très restreint, et un échappement à la réponse immunitaire a aussi été observé (229, 231). Cela montre qu'il est finalement peu probable que la population virale concernée soit éliminée sous de telles pressions immunitaires. En effet, il est possible que le spectre de mutants présente déjà des variants plus tolérants à la réponse immunitaire (235). Et si ce n'est pas le cas, la complexité du spectre de mutants permet de découvrir de nombreuses autres routes

d'échappement via l'exploration de l'espace séquence, par l'apparition de substitutions sur des sites viraux jusqu'alors invariables et conservés.

Dans l'établissement d'un vaccin, il faut ainsi chercher à créer des barrières génétiques et phénotypiques hautes. Cela peut être atteint en ciblant de multiples épitopes et de nombreux sites antigéniques. Ainsi, le nombre de mutations nécessaires pour empêcher tous les sites antigéniques visés d'être atteints, ainsi que leur coût en termes de valeur sélective, seront trop importants. C'est encore une fois l'utilisation des NGS qui permettrait de mettre en évidence les sites antigéniques à viser, qui sont exprimées par le spectre de mutants. Il pourrait être intéressant d'inclure des vaccins visant ces sites antigéniques avec les traitements antiviraux classiques, en prenant bien entendu en compte la dynamique des quasi-espèces.

#### 5.1.4.3. Orientation de l'évolution virale par la vaccination

Sur le long terme et à l'échelle d'une population, la vaccination peut très bien ne pas être satisfaisante. En effet, si le virus n'est pas éliminé, celui-ci peut être transmis d'un hôte vacciné à un hôte non vacciné. Il se peut qu'il en résulte une sélection de variants présentant un antigène altéré sur lequel la réponse immunitaire induite par le vaccin ne fonctionne plus. Plus le temps passe, et plus cela augmente les risques d'apparition de mutations compensatrices.

Des vaccins peptidiques confèrent généralement une immunité partielle car ils présentent un répertoire antigénique limité au système immunitaire de l'hôte. Chez le FMDV, cela se présente sous la forme de peptides de synthèse correspondant à la séquence VP1 (229, 231). L'efficacité de tels vaccins contre le FMDV a été testée, en mettant des individus vaccinés en contact avec le virus. Certains animaux vaccinés ont développé des lésions vésiculaires et une virémie. Ainsi, l'immunisation partielle semble permettre une réplication virale et une charge virale suffisante pour promouvoir l'émergence de variants antigéniques et un changement de tropisme cellulaire (229, 231).

De plus, les vaccins sont inégaux dans leur efficacité. Selon la localisation géographique, les virus peuvent présenter des différences antigéniques avec la souche vaccinale utilisée. Le temps peut être insuffisant entre la mise en place de la vaccination et l'exposition au virus. Un mauvais protocole vaccinal peut avoir été mis en place, ce que l'on retrouve parfois en médecine vétérinaire (72). L'état immunitaire de l'individu joue aussi un rôle important : des variants antigéniques du HAV isolés, présentés plus haut, provenaient d'individus infectés par le HIV-1. L'immunodépression provoquée par l'infection a pu permettre au HAV d'échapper à la réponse immunitaire et peut aussi être corrélée à des vaccinations incomplètes (183). Ces variants ont ensuite pu se propager à des individus correctement vaccinés.

Cette possible orientation de l'évolution virale est un domaine à documenter. On ne sait par exemple pas si la dérive génétique qui a lieu lors de goulots d'étranglement (qui peuvent avoir lieu lors d'une réponse vaccinale) peut jouer un rôle. On ne sait pas non plus avec précision quel est le rôle joué par la réversion des vaccins vivants atténués ou la recombinaison. Il est encore difficile d'évaluer les conséquences de la vaccination étant donné qu'il s'agit d'un phénomène récent.

## 5.2. Modification du tropisme cellulaire

La sélection de composants minoritaires présents dans le spectre de mutants peut jouer un rôle dans les interactions entre le virus et sa cellule-hôte. Le tropisme cellulaire et la spécificité de l'hôte dépendent de facteurs extracellulaires, comme la reconnaissance d'un récepteur à la surface cellulaire par le virus, mais aussi de facteurs intracellulaires, qui vont favoriser la multiplication virale dans la cellule (17). Ces caractéristiques vont donc être modulées lors de substitutions d'acides aminés au niveau des protéines S et NS.

### 5.2.1. Modulation par les protéines structurales

Ce sont les protéines qui interviennent dans la structure virale (voir partie 1.1.2.). Bien qu'elles aient différentes fonctions, celles intervenant dans la reconnaissance des récepteurs cellulaires jouent un rôle très important dans les interactions entre l'hôte et le virus. Ce sont les mieux décrites dans la littérature, bien qu'il ne faille pas négliger les autres fonctions d'autres protéines structurales.

#### 5.2.1.1. Les récepteurs cellulaires

Les récepteurs reconnus par les virus appartiennent à différentes familles moléculaires : protéines, lipoprotéines, glycoprotéines, glycosaminoglycane, etc. Ils sont souvent organisés sous forme de structures complexes et d'assemblage de ces différentes familles (17). Ils sont en très grand nombre à la surface cellulaire, remplissent une grande variété de fonctions (adhésion, signalisation, reconnaissance ou même inflammation) et sont nécessaires à l'entrée du virus dans la cellule mais parfois pas suffisant (17). Certains servent à faciliter l'approche de la particule virale, d'autres servent de corécepteurs.

#### 5.2.1.2. Interactions cellules-virus via les récepteurs

L'interaction entre un virus et la cellule-cible se fait grâce au site de reconnaissance d'un récepteur cellulaire présent sur la particule virale. Elle va permettre l'internalisation du virus dans la cellule. Cependant ces interactions présentent des subtilités (17) :

- Un même virus peut utiliser des récepteurs et corécepteurs différents, comme par exemple pour le HCV, qui peut en utiliser au moins deux (17), le FMDV au moins trois (16, 21) ou le HRV plusieurs types (246).
- Des virus d'une même famille, ou associés à des manifestations cliniques similaires, peuvent utiliser des récepteurs différents. Les virus de l'hépatite ont bien tous le même organe cible, pourtant le HAV, le HBV et le HCV utilisent tous des récepteurs différents (17).
- Des virus complètement différents peuvent partager le même récepteur : le « coxsackievirus adenovirus receptor » (CAR) est utilisé par les adénovirus 2 et 5, ainsi que par les coxsackievirus (allant de B1 à B6) (17, 21).
- Il y a une compartimentation cellulaire chez les organismes différenciés. En effet, les différents tissus et organes n'expriment pas les mêmes macromolécules de surface. Cette compartimentation revient à appliquer *in vivo* la dynamique des quasi-espèces virales, en sélectionnant les sous-

populations capables de se fixer aux récepteurs spécifiques de tels ou tels tissus (21).

- Un même tissu peut être ciblé par des virus différents utilisant un récepteur à chaque fois différent. Cela confirme l'existence probable de récepteurs alternatifs présents à la surface cellulaire (17).

Toutes ces caractéristiques font que les interactions entre cellules et virus restent encore très difficiles à interpréter. La dynamique et les mécanismes moléculaires de reconnaissance des récepteurs cellulaires et de leur modulation sont des domaines qui nécessitent encore beaucoup d'investigations. Suite à l'émergence ou la découverte d'un nouveau virus, il est très difficile de prédire le type et le nombre de récepteurs cellulaires qu'il peut utiliser. L'utilisation d'un récepteur chez un virus ne semble en effet avoir aucune corrélation évidente avec la position phylogénétique d'un virus.

### 5.2.1.3. Conséquences des substitutions d'acides aminés présents dans les sites de reconnaissance

Lors de la dynamique d'évolution des espèces virales, l'apparition de mutations va avoir pour conséquences une modification des séquences d'acides aminés. Les mutations, rapportées comme provoquant ces substitutions d'acides aminés, sont principalement des transitions et des transversions. Ce sont les événements les plus fréquents, mais des insertions et des délétions pourraient avoir le même effet. Une ou quelques substitutions au sein d'une capsid virale ou d'une protéine de surface, vont modifier la capacité de reconnaissance cellulaire (17). Elles vont donc avoir plusieurs conséquences.

#### 5.2.1.3.1. *Changement de tropisme cellulaire*

Plusieurs études chez le HIV-1 ont montré que des mutations ponctuelles spécifiques au sein du génome ont eu des conséquences sur des régions protéiques virales, par exemple sur les régions V3, V1 de la glycoprotéine de surface gp120. Elles ont modifié à chaque fois le tropisme cellulaire du virus (30). Il s'agissait même parfois d'une seule mutation ponctuelle. Dans d'autres cas comme, celui du virus de la bursite infectieuse, c'est la combinaison de deux mutations apparues simultanément qui déterminent la virulence et le tropisme (188).

Les récepteurs et corécepteurs utilisés chez le HIV-1 changent au cours de l'infection : les isolats issus d'un début d'infection chez l'homme sont généralement spécifiques du corécepteur CCR5. Lors de la progression de l'infection, des variants utilisant le corécepteur CXCR4 deviennent peu à peu dominants et permettent l'infection de lymphocytes naïfs. Une transition de R5 à X4 se retrouve donc dans les phases plus tardives de l'infection, est associée à un déclin accéléré des cellules T-CD4<sup>+</sup> et démontre, ici aussi, un changement de tropisme cellulaire (17, 193). De même, le FMDV devient capable d'infecter les cellules dendritiques présentatrices d'antigènes lorsqu'il forme un complexe immun en présence d'anticorps, mais il perd alors sa capacité à cibler les cellules qu'il infecte normalement (197).

Une seule substitution d'acide aminé au niveau d'une glycoprotéine chez le LCMV altère quant à elle son affinité pour le récepteur  $\alpha$ -dystroglycane. Si l'affinité est haute, il y a immunosuppression et persistance virale chez la souris. Si l'affinité est basse, il y a diminution de l'infection (17).

Chez l'IV, la nature de la liaison entre l'acide sialique et le galactose ( $\alpha$ 2,3 et  $\alpha$ 2,6) va déterminer la spécificité d'hôte (respectivement les espèces aviaires et l'homme). Cette spécificité est conférée par l'hémagglutinine du virus (17). Deux acides aminés situés sur la protéine S chez le SARS coronavirus (Syndrome Respiratoire Aigu Sévère) modulent la fixation à l'enzyme de conversion de l'angiotensine 2. En fonction de leur nature, cette fixation sera effective chez l'homme ou chez la civette palmiste (171).

Une seule substitution d'acides aminés, voire quelques-unes, suffisent ainsi à changer le tropisme cellulaire d'un virus ou le moduler. Les exemples présentés ne sont qu'une fraction du nombre rapporté de modification du tropisme (annexe 7). Ces substitutions peuvent même expliquer la spécificité d'hôte, et un simple changement peut faire changer le virus d'hôte (17).

#### 5.2.1.3.2. *Expansion du tropisme cellulaire*

Un exemple d'expansion de tropisme cellulaire est bien décrit chez le FMDV *in vitro*. Le site de reconnaissance d'un récepteur à intégrine est constitué d'un triplet d'acides aminés RGD, situé sur une boucle protéique VP1 de la capsid virale, pour infecter les cellules, par exemple les BHK-21 (16, 20, 141). Le fait qu'on l'ait retrouvé chez de nombreux isolats a laissé supposer qu'il était fortement conservé (141). Des passages sur culture cellulaire ont cependant sélectionné des variants mutants avec des substitutions de quelques acides aminés de la surface de la capsid virale, autour ou sur le motif RGD. Cela a rendu la nécessité de la reconnaissance de l'intégrine superflue (141). La particule virale a alors pu provoquer une infection par une voie alternative, en utilisant les héparanes sulfates situés à la surface cellulaire (16, 17). Le FMDV s'est donc montré capable d'utiliser des récepteurs alternatifs. Cependant il a aussi acquis la capacité d'infecter les cellules humaines K-562 ou encore les cellules CHO, qui n'expriment ni l'intégrine ni les héparane sulfates à leur surface cellulaire. Ainsi un troisième récepteur, inconnu, est utilisé (16, 17, 21). De plus, les deux autres récepteurs étaient encore utilisables par le virus (17).

Les récepteurs à intégrine semblent être utilisés *in vivo* chez les bovins. Les variants utilisant les autres récepteurs sont cependant présents en tant que composants minoritaires au sein du spectre de la quasi-espèce (17, 263), et ces substitutions d'acides aminés au niveau du site RGD ou aux alentours ont été isolées chez des bovins protégés partiellement par un vaccin peptidique (229). On a ici affaire à une véritable expansion du tropisme cellulaire et de la gamme d'hôte pour le FMDV.

#### 5.2.1.3.3. *Modification de la pathogénèse*

Un variant LCMV, qui diffère du type sauvage au niveau de deux acides aminés, affiche un tropisme cellulaire altéré : il présente un plus haut niveau d'infection pour les cellules présentatrices d'antigènes de la rate, notamment les cellules dendritiques. Celles-ci sont alors détruites précocement par les lymphocytes T cytotoxiques CD8<sup>+</sup>. Cela se traduit par une immunosuppression généralisée chez la souris (29).

De même, une substitution d'acide aminé au niveau de la capsid chez l'EMCV affecterait le tropisme cellulaire du virus, provoquant la destruction des cellules pancréatiques. Alors qu'il est normalement apathogène chez les rongeurs, il peut donc être diabétogène chez des souris à cause de ce changement de tropisme (15).



### 5.2.2. Modulation par les protéines non structurales

Bien que n'intervenant pas directement dans la formation de la particule virale, ces protéines font partie intégrante du cycle viral, notamment via leurs fonctions régulatrices. Elles peuvent donc moduler le tropisme cellulaire et la spécificité d'hôte indirectement. Ceci a été bien étudié pour l'IV et le FMDV.

Deux substitutions d'acides aminés sur la sous-unité PB2 de la polymérase du l'IV-A ont été décrites comme permettant l'adaptation du virus des espèces aviaires aux mammifères. Mais ces substitutions ne sont pas retrouvées chez le H1N1. Ce dernier présente cependant une substitution d'acide aminé qui compense cette absence, permettant ainsi l'infection chez les mammifères. Cette même substitution est d'ailleurs retrouvée chez le H5N1 et augmente la létalité chez la souris. Ici aussi, une seule substitution d'acide aminé a permis d'acquérir un tropisme cellulaire par une autre voie évolutive (264).

Pour une réplication efficace de l'IV-A dans la cellule-hôte, la sous-unité PB2 ainsi que la nucléoprotéine NP requièrent une protéine cellulaire, l'importine- $\alpha$ , qui permet le transport du génome viral dans le noyau, où ont lieu la réplication et la transcription virales. Cependant cette protéine est spécifique pour chaque espèce : l'importine- $\alpha 3$  chez les espèces aviaires, l'importine- $\alpha 7$  chez les mammifères. Pour une transmission inter-espèces, les deux sont nécessaires, comme chez le virus H1N1. La spécificité de ces protéines joue un rôle essentiel dans la détermination de la spécificité d'hôte, et l'IV est capable de passer d'une forme protéique à une autre pour surpasser la barrière inter-espèce (91).

L'adaptation du FMDV chez le cochon d'inde à partir du porc a été provoquée il y a plusieurs dizaines d'années, afin d'établir un système convenable d'étude de la réponse immunitaire et tester l'efficacité des vaccins. Des études expérimentales ont tenté d'investiguer les raisons de cette adaptation. L'analyse des virus découverts dans les lésions secondaires qui avaient été engendrés par l'adaptation ont pu mettre en évidence une substitution d'acide aminé sur la protéine NS 3C, une autre sur la protéine 3A et une dernière sur le site antigénique de la protéine VP1 de la capsid. Seule la deuxième était réellement essentielle. Cela a démontré que quelques substitutions sur des protéines non structurales permettaient l'adaptation d'un virus à un nouvel hôte (161). La poursuite de ces études a apporté des résultats intéressants : un variant mutant présentant une quatrième substitution d'acide aminé bien spécifique n'était pas détectable dans la population virale d'origine chez le porc, mais est apparu lors de son adaptation à un hôte inhabituel, le cochon d'inde. Il s'agissait de variants minoritaires, devenus dominants lors du transfert de retour à l'hôte d'origine, le porc. Une minorité de variants, émergeant chez un hôte inhabituel, pourrait ainsi devenir dominant suite à des réinfections de l'espèce hôte d'origine (162).

### 5.2.3. Dynamique de la modulation du tropisme

L'apparition d'un nombre minimal de ces substitutions d'acides aminés peut donc modifier le tropisme cellulaire via la capacité de reconnaissance cellulaire, en rendant parfois superflus certains récepteurs, via les protéines non structurales. Cela peut même modifier la pathogénèse du virus et favoriser le passage d'une espèce à une autre en modifiant la spécificité virale.

Les forces sélectives qui enclenchent ce changement de tropisme sont assez difficiles à identifier, cependant on peut suspecter que c'est la disponibilité des récepteurs, dans le cas des protéines structurales, qui sélectionnent une sous-population de mutants au sein de la distribution (17). Il ne faut cependant pas négliger la dynamique qui touche les cellules elles-mêmes. En effet, certaines d'entre elles sont plus permissives que d'autres à différentes

formes de virus, et des effets viro-induits peuvent faire affecter certaines cellules ou promouvoir la survie d'autres (72).

Le seul fait d'avoir à disposition un récepteur cellulaire ou des protéines intracellulaires, différents de ceux utilisés par la majorité des composants du spectre de mutants, semble favoriser la sélection de composants du spectre qui utilisent ce récepteur ou ces protéines non structurales pour conduire à une infection lors de contraintes sélectives particulières (72). De nouveau, on se retrouve dans une dynamique de sélection décrite chez les quasi-espèces virales, associées cette fois-ci à une modification du tropisme et de la spécificité de l'hôte. On retrouve aussi une implication biologique à la présence de sous-populations minoritaires, qui peuvent être sélectionnées à condition qu'une molécule réceptrice correspondante soit présente. De manière générale, ce changement de tropisme est en lui-même une conséquence pertinente du spectre de mutants et de la dynamique des quasi-espèces.

Le concept de barrières génétiques et phénotypiques intervient ici aussi lors de ces modifications de la reconnaissance cellulaire. Connaissant le fort taux de mutations chez les virus à ARN, une simple substitution d'acide aminé semble être un événement fréquent. Mais cet événement a de forte chance d'avoir des effets négatifs sur le site de reconnaissance des récepteurs à cause d'une barrière phénotypique forte, donc de subir une sélection négative. De plus, les substitutions requises peuvent avoir des effets négatifs sur d'autres éléments, comme la capside, la stabilité des protéines S, etc. Cela peut même varier selon l'ordre d'apparition des substitutions. Mais un nouveau site de reconnaissance d'un récepteur devrait avoir un effet positif sur la valeur sélective : il permet au virus de pénétrer dans un nouveau type cellulaire ou de surpasser une contrainte. Tous ces éléments vont avoir un impact sur la sélection (71).

Comme dans le cas des résistances, un spectre de mutants large et complexe permet de contribuer à surmonter ces barrières durant la sélection virale, et va intervenir dans la probabilité d'apparition de nouveaux sites de reconnaissance de récepteurs cellulaires. Encore une fois, la taille de la population est un paramètre d'une importance capitale. Cela va permettre d'étendre la possibilité de modification du tropisme cellulaire ou de spécificité d'hôte et diminuer le concept de barrières inter-espèces. Sur le plan épidémiologique, cela va résulter en l'émergence de pathogènes dont le tropisme a été modifié, ciblant de nouveaux tissus ou de nouvelles espèces.

#### **5.2.4. La coévolution du tropisme et de l'antigénicité**

La coévolution est une force évolutive qui intervient aussi dans la modification du tropisme cellulaire. Il s'agit de l'adaptation génétique et réciproque d'un agent pathogène et de l'organisme dans lequel il se développe, due à leurs interactions sur une longue période de temps (231). On peut retrouver cela à l'échelle du virus même, où il peut y avoir des influences réciproques entre différents sites. Les domaines subissant une coévolution peuvent se retrouver sur l'acide nucléique ou au sein des protéines codées par le virus, comme entre un site de reconnaissance d'un récepteur cellulaire et un site antigénique. Le chevauchement de ces domaines favorise encore plus cette force évolutive et a été largement rapporté dans divers expériences (72). Il semble bien qu'il s'agisse d'une règle plutôt que d'une exception. De nombreux exemples sont présentés dans l'annexe 8.

#### 5.2.4.1. Variation antigénique et modification du tropisme *in vitro*

L'hypothèse dite du canyon suppose que certains sites de reconnaissance des récepteurs sont protégés des variations au niveau des sites antigéniques, si ceux-ci sont liés (144). Chez le FMDV, le site de reconnaissance du récepteur à intégrine RGD est situé sur une boucle VP1, qui sert aussi de déterminant antigénique majeur (200). Le triplet RGD est en effet localisé au niveau de l'épitope SD6, qui est impliqué dans la fixation aux anticorps neutralisants (242). Il y a dans ce cas-ci chevauchement des deux sites (dans ce cas, le site de reconnaissance est même englobé dans le site antigénique) (figure 40A).

Des virus mutants C-S8c1, résistants aux anticorps (les MAR), ont été isolés et ont montré des substitutions d'acides aminés sur l'épitope SD6. Le site de reconnaissance vis-à-vis d'anticorps a été donc modifié. Mais si cette modification vient à toucher le site de reconnaissance des récepteurs à intégrine, cela risque d'abolir cette reconnaissance. Si des variants présentent de telles modifications (comme par exemple le triplet RGG (200)), ils seront soumis à une sélection négative à cause des contraintes liées au site de reconnaissance du récepteur, à moins qu'ils ne soient sauvés par complémentarité (via des capsides mosaïques, voir partie 5.4.1.) ou réversion (200). L'avantage sélectif espéré, accordé par l'échappement aux anticorps, ne sera donc pas possible. Le site de reconnaissance participe à l'établissement de la barrière phénotypique concernant l'épitope SD6, et on ne semble retrouver ici que les MAR ne présentant pas de modification au niveau du triplet RGD et ses alentours (figure 40B).

Après 100 repiquages en série des MAR, des mutants CS8c1p100 issus de la population nouvellement obtenue présentaient des substitutions supplémentaires, cette fois-ci dans le RGD (figure 40C). L'une des hypothèses expliquant ces substitutions est que la contrainte imposée par le RGD a disparu. Cela n'est possible que si ce site est rendu superflu et qu'un autre peut être utilisé, ce qui a été décrit (16, 17). Ainsi, le répertoire des MAR s'est étendu par rapport au répertoire obtenu avec le virus du type sauvage, car des substitutions ont permis de rendre un autre site de reconnaissance de récepteur effectif.

Une altération de la spécificité des récepteurs va avoir un impact sur la liaison aux anticorps et vice-versa. On a ici un exemple parfait d'événement coévolutif.

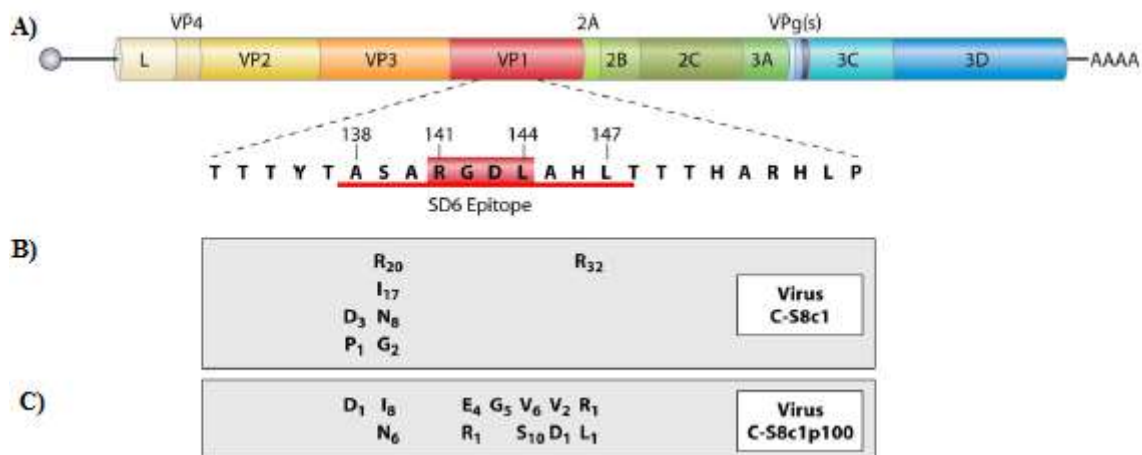
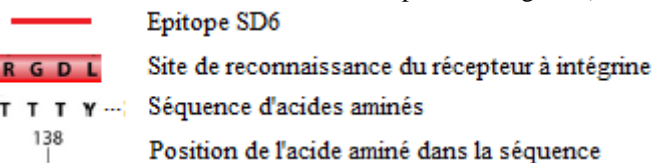


Figure 40 : Cas de coévolution chez le FMDV (71)

A) Génome du FMDV et sa séquence d'acides aminés, incluant le site antigénique majeur (épitope SD6) et le site de reconnaissance du récepteur à intégrine (RGDL).



B) Substitutions d'acides aminés obtenues suite aux mutations pour les MAR C-S8c1, se localisant sur l'épitope mais pas sur le site RGDL.

C) Substitutions d'acides aminés obtenues suite aux mutations pour les MAR C-S8c100, se localisant sur l'épitope et sur le site RGDL.

**X<sub>n</sub>** Substitution de l'acide aminé situé sur le figure 40A par l'acide aminé X, retrouvé n fois lors d'événements sélectifs indépendants

#### 5.2.4.2. Variation antigénique et modification du tropisme *in vivo*

Chez les animaux vaccinés à l'aide de vaccin peptidique et mis en contact avec le FMDV, des lésions ont tout de même été observées (voir partie 5.1.4.3.). Les virus isolés sur ces lésions possédaient des substitutions d'acides aminés dans la région du RGD ou autour (229, 231). L'analyse de ces variants FMDV se répliquant *in vivo* chez des vaches pourtant immunisées montre un cas remarquable de coévolution. Et il est donc possible de retrouver des cas de mutants d'échappement à la réponse immunitaire *in vivo* ainsi que les mutations qui ne semblaient décrites qu'*in vitro*.

#### 5.2.4.3. Dynamique de la coévolution

Deux caractéristiques semblent favoriser la coévolution :

- Des modifications des sites antigéniques ou de reconnaissance cellulaire semblent nécessiter très peu de mutations, donc un faible nombre de substitutions d'acides aminés. On parle ici encore de barrière génétique faible. Cela rend la nature des quasi-espèces virales cohérentes biologiquement.
- Les acides aminés composant la surface des sites semblent moins soumis à des contraintes que les composants internes des capsides ou des enveloppes virales. Il s'agit d'une barrière phénotypique faible. Cependant la tolérance des sites antigéniques aux mutations semble moins grande. La barrière phénotypique peut dans certains cas être grande.

On observe une fois de plus une dynamique des quasi-espèces virales. Celles-ci présentent un répertoire de mutants ou de génomes recombinés qui ont le potentiel d'utiliser des récepteurs alternatifs. Ce répertoire de mutants d'échappement présents dans le spectre mutant peut cependant varier en fonction de certaines caractéristiques biologiques liées à d'autres sites, certains chevauchant même le site d'intérêt. Ici encore, c'est un spectre de mutants composé de nombreux variants qui va permettre de surpasser ces barrières phénotypiques. Il y a une probabilité réelle que cette classe de mutants soit présente dans la quasi-espèce virale, soit parce que les mutations n'ont pas un coût significatif de valeur sélective, soit parce que les mutants peuvent être maintenus par complémentarité.

### **5.2.5. Conclusion**

La plasticité phénotypique que l'on retrouve chez les virus, mais aussi cette coévolution de l'antigénicité et du tropisme ont pour conséquences l'augmentation de l'éventail des variants viraux, et permettent de modifier tropisme cellulaire et caractère antigénique. Une population virale peut alors sélectionner des mutants surpassant les contraintes de l'hôte, permettant ainsi des changements de cibles (tissus, organes), voire de franchir des barrières inter-espèces. Ce franchissement des barrières d'espèces a pour conséquence que des variants minoritaires d'un spectre de mutants chez un hôte ont le potentiel de devenir dominant chez un autre hôte, donc sous d'autres contraintes de l'environnement. Cela contribue fortement à l'émergence de nouvelles maladies virales, avec des impacts forts sur la santé publique. De façon concomitante, les influences socio-culturelles et environnementales ont aussi un rôle dans la transmission et l'émergence de nombreuses maladies virales zoonotiques. Par exemple, les virus qui se répliquent chez les rongeurs vont générer des spectres de mutants avec des variants minoritaires capables de devenir dominants chez d'autres hôtes, suite aux déplacements des animaux avec lesquels ils rentrent en contact et suite à l'augmentation des populations de rongeurs dues aux pratiques de l'agriculture.

Une analyse de la composition du spectre de mutants grâce aux nouvelles techniques de séquençage, combinée à des études structurales, pourrait permettre de prédire, dans une certaine mesure, les conséquences évolutives liées aux modifications des sites de reconnaissance des récepteurs. Ces analyses pourraient dévoiler la présence de multiples types de virus dans un certain nombre d'habitats écologiques. Les analyses structurales tridimensionnelles des complexes protéiques d'attachement aux récepteurs et des récepteurs cellulaires pourraient quant à elles permettre de mieux comprendre les mécanismes d'adaptation de ces protéines.

### **5.3. La mémoire moléculaire**

L'étude des génomes présents à une basse fréquence au sein du spectre de mutants a longtemps été ignorée, en partie car les techniques de séquençage ne permettaient pas de les aborder (voir partie 3.4.3.4.). Il semble maintenant reconnu que ces sous-populations puissent jouer un rôle biologique important, avec des conséquences cliniques non négligeables.

### 5.3.1. L'hypothèse des génomes mémoires

La théorie des quasi-espèces étant un système d'adaptation complexe, l'hypothèse a été émise que le spectre de mutants des quasi-espèces pouvait inclure une sous-population spécifique des génomes qui représentent un enregistrement du passage de génomes dominants dans les phases précédentes d'une même lignée (201). Ils sont nommés les génomes mémoires.

Cette hypothèse a été inspirée du système immunitaire, qui est aussi un système d'adaptation complexe à des contraintes du milieu. En effet, en réponse à un antigène, des génomes vont subir recombinaison et hypermutation, et les lignées cellulaires présentant la plus grande affinité pour les antigènes sont produites en grande quantité afin de les éliminer. Suite à ces événements, des cellules mémoires sont générées et confèrent ainsi une immunité sur le long terme contre un antigène spécifique, permettant une réponse rapide et efficace lors d'un prochain contact avec celui-ci. Cela fait penser à la sélection d'une sous-population virale (201).

### 5.3.2. Mise en évidence d'une mémoire moléculaire chez les virus à ARN

Un exemple d'une telle mémoire se retrouve chez le FMDV (10, 201, 204). Des sous-ensembles de génomes, appelés RED, sont devenus dominants en présence d'anticorps monoclonaux, suite à la sélection de mutants présentant une résistance pour ces derniers (les variants MAR). Lorsque les anticorps ont été retirés du milieu de culture, les variants ont ensuite été surpassés par d'autres génomes révertants, les RGD, de la même population. Cependant, ils présentent des fréquences plus grandes (de  $1.10^{-2}$  à  $5.10^{-1}$ ) que des fréquences basales (de  $4.10^{-3}$  à  $5.10^{-4}$ ), nommées le niveau de mémoire (31, 201). Ces mutants présentent des caractéristiques de génome mémoire en évitant l'extinction et en restant dans la population pendant une longue période de temps. Ils caractérisent l'histoire évolutive de la population et reflètent ainsi la précédente dominance de génomes dans une population virale (31).

Les génomes mémoires ont aussi pu être mis en évidence chez le HIV-1 *in vivo* où l'on semble pouvoir distinguer deux types de mémoires, même si la limite entre les deux est parfois floue (31) :

- Une mémoire répliquative similaire à celle décrite chez le FMDV.
- Une mémoire cellulaire, dite réservoir, due à l'intégration de l'ADN viral dans le génome ADN cellulaire lors des cycles des rétrovirus (31). Ces ADN se retrouvent donc dans les chromosomes cellulaires, fournissant ainsi un réservoir de séquences.

En parallèle d'analyses phylogénétiques, un modèle mathématique *in silico*, basé sur la théorie des quasi-espèces, a été développé pour tester si la mémoire moléculaire pouvait se réaliser et servir de stratégie évolutive pour l'émergence de virus comme le HIV-1 (31). Il a non seulement été résolu mais a rapporté des données en accord avec les résultats expérimentaux. Cependant ce modèle a prédit qualitativement cette évolution, mais pas quantitativement : il peut y avoir des différences de valeur de fréquence chez ces génomes avec celles observées lors des études expérimentales (31, 257).

### 5.3.3. Caractéristiques des génomes mémoires

Plusieurs observations expérimentales ont permis de mettre en évidence certaines caractéristiques de cette mémoire moléculaire.

- Les goulots d'étranglement éliminent cette mémoire (201), ce qui est concordant avec l'évolution des quasi-espèces en tant qu'ensemble et non pas en tant qu'individus indépendants au sein de ce dernier.
- La fréquence qui sera atteinte par un génome mémoire dépend de la capacité adaptative relative du virus de ce génome (202) et du taux de mutation. En comparant les niveaux de mémoire de mutants FMDV, l'un d'eux a montré une valeur adaptative 7,6 fois plus élevée que les autres, donc une fréquence 30 à 100 fois plus élevée quand il est devenu un génome mémoire (31). Cela en fait une propriété collective chez une quasi-espèce entière (201), et une conséquence de la dynamique des quasi-espèces.
- Différents niveaux de mémoire peuvent être atteints. Certains ont été relevés à des fréquences entre 0.1% et 20% du nombre de total de génomes dans la population (31).
- Les génomes mémoires évoluent en parallèle des génomes les plus fréquents au sein d'une même quasi-espèce (10) : si ces derniers gagnent en valeur adaptative, ils le feront aussi. Ces génomes fonctionnent bien en accord avec l'hypothèse de la Reine Rouge. Cette dernière stipule que, dans des populations en compétition, il y a une sélection continue des génomes les plus adaptés.
- Le niveau de mémoire peut être durable mais se perd graduellement dans des populations en réplication. Et cette perte a suivi son cours de manière strictement identique chez des lignées parallèles de taille finie et hors équilibre (204). Cette caractéristique suggère un comportement déterministe (voir partie 6.1.2.).
- La mémoire réservoir retrouvée chez le HIV-1 peut fournir des génomes qui évoluent plus lentement que les génomes circulants, étant donné qu'ils se répliquent en tant que « gènes cellulaires », avec une fidélité de copie plus importante (31). Ces provirus intégrés peuvent être maintenus pendant une période plus longue : de l'ordre du mois ou de l'année. Ce réservoir peut donc agir comme une sorte d'archive, à tel point que les variants qu'il code peuvent différer fortement des variants circulants à un instant donné.

### 5.3.4. Proposition schématique de la dynamique de la mémoire moléculaire

La figure 41 présente une dynamique d'évolution d'une quasi-espèce faisant intervenir la mémoire moléculaire :

La population est constituée de multiples variants, formant plusieurs sous-populations et exprimés à diverses fréquences (figure 41 A).

Des variants, exprimés à basse fréquence, sont sélectionnés suite à une contrainte sélective (figure 41 B). Ils vont dominer la population avec une fréquence importante. Cette phase correspond par exemple à la sélection d'une résistance lors de l'arrivée d'anticorps monoclonaux.

Si la réplication devient extensive, c'est-à-dire que la contrainte sélective disparaît, les révertants, qui étaient dominants avant l'intervention de la contrainte sélective, redeviennent dominants (figure 41 C). Si, après le processus d'amplification sélective, les variants dominants figure 41B ont atteint un niveau de fréquence supérieure aux variants parentaux (de la figure 41A), ils vont atteindre un niveau de fréquence supérieur à leur niveau basal, appelé le niveau de mémoire. Ils verront ainsi leur fréquence augmenter dans la nouvelle distribution.

Si un goulot d'étranglement a lieu après la réplication, les révertants redeviennent dominants et les variants sélectionnés retombent à basse fréquence (figure 41 D). Il y a perte de mémoire moléculaire.

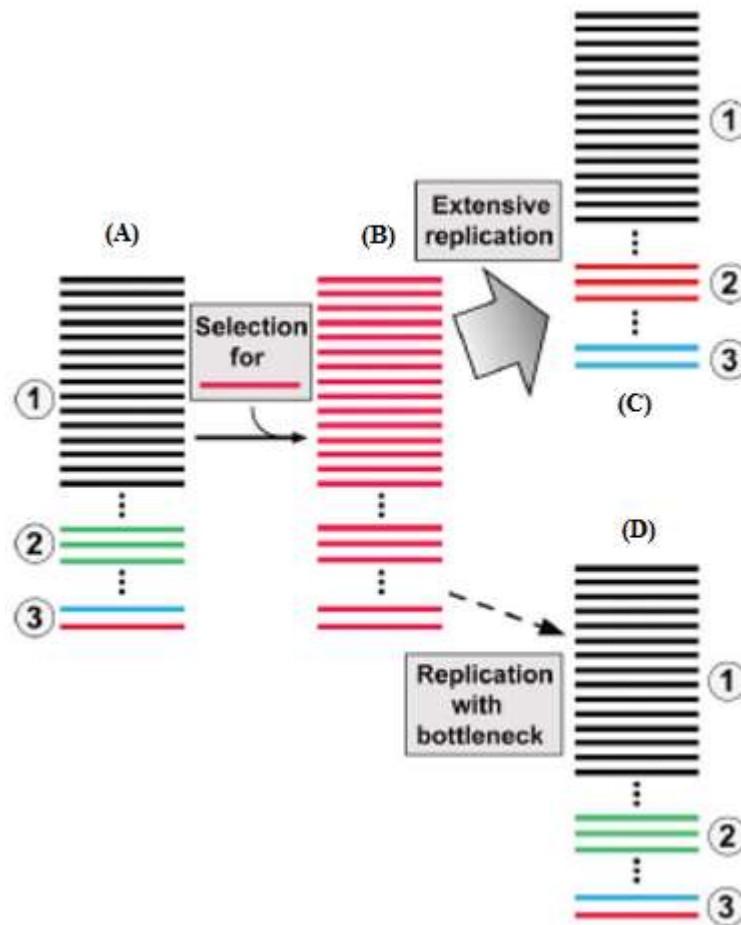


Figure 41 : Représentation schématique de la mémoire des quasi-espèces et de l'effet des goulots d'étranglement sur la mémoire (72)

- Ⓝ Niveau de fréquence des différentes sous-populations : le numéro 1 représente la sous-population de génomes la plus fréquente le numéro 2 la sous-population classée 2<sup>ème</sup> en terme de fréquence, etc.
- Variants de la population, retrouvés à différentes fréquences, formant une sous-population
- Variant sélectionné lors de la mise en place d'une contrainte sélective
- Intervention d'une contrainte sélective
- Réplication avec disparition de la contrainte sélective (extensive replication) = passage en série avec une large population
- Réplication avec un goulot d'étranglement

### 5.3.5. Proposition de mécanisme moléculaire

Un mécanisme supposé de cette mémoire correspondrait à une forme de complémentarité bien spécifique appelée le « mixing and hiding » (159, 257). Un modèle *in silico* a supposé que la différence de valeur sélective entre le type sauvage et les variants RED du FMDV est due à l'incapacité de ces derniers à pénétrer dans la cellule-hôte. Ainsi, si la dynamique qui touche le spectre de mutants du virus de type sauvage ne permet pas de recréer les variants RED suite à des mutations et ainsi maintenir sa fréquence au sein de la population, le modèle prédit qu'il est possible qu'une fraction de génomes mutants soit



encapsidée avec les protéines du type sauvage, leur permettant d'être aussi infectieux que ce dernier. Le génome mutant et les protéines du type sauvage sont « mélangés » et permettent au génome de se « cacher ». Cela permettrait de retarder l'extinction et d'augmenter la fréquence des variants mutants. Cet effet serait plus prononcé à forte MOI, bien qu'existant à faible MOI, et expliquerait la fréquence à un niveau mémoire, plus élevé que ce que prédit la théorie des quasi-espèces (257). Cependant aucune preuve de ce mécanisme n'a été apportée chez le système FMDV RED-RGD, le mieux compris et le mieux maîtrisé (204). Cela n'est pas suffisant non plus pour l'écarter (257).

### 5.3.6. Implications de la mémoire moléculaire

La mémoire moléculaire facilite ainsi la réponse d'une collectivité de variants à une contrainte sélective. Si cette contrainte implique la sélection d'une sous-population de génome, une fois que la force de sélection diminue, le stockage de la sous-population peut avoir lieu à un niveau mémoire, afin de préparer l'ensemble de la population à une confrontation à une prochaine contrainte identique ou ressemblante. Cette préparation est cependant à tempérer étant donné que les goulots d'étranglement semblent diminuer la mémoire. Ces goulots étant des événements reconnus comme abondants dans les cycles viraux, la mémoire peut ne pas avoir une influence aussi forte qu'on ne l'aurait cru sur le comportement de la population. Mais cela trouverait son application dans le cas d'infections virales chroniques.

*In vivo*, on peut supposer que lorsqu'un agent antiviral est utilisé pour traiter une infection, un mutant résistant à cet antiviral peut être sélectionné. Il peut ensuite être surpassé par d'autres composants du spectre, suite à l'interruption d'un traitement ou à la mise en œuvre d'un traitement alternatif. Cela peut avoir pour conséquence une conversion en génome mémoire. C'est ce que l'on suspecte dans le cas du HIV-1 chez des patients infectés, favorisant ainsi la réémergence de génomes ancestraux (31), et l'on peut supposer que cela s'applique à d'autres virus (HBV et HCV). On se retrouve là face à un dilemme médical : lorsque le traitement est arrêté car des résistances aux inhibiteurs antiviraux apparaissent et deviennent dominantes, la sous-population qui était sensible aux inhibiteurs remplit le spectre de mutants (si la résistance impliquait un coût en termes de valeur sélective) et redevient dominante. Les mutants résistants occupent ainsi un niveau de mémoire plutôt que leur niveau basal, et si aucun traitement alternatif n'est utilisé chez les patients pendant ce laps de temps, et qu'on l'administre de nouveau, alors la résistance aura lieu. L'arrêt du traitement semblait donc indiqué, mais aura tout de même des conséquences sur la suite.

La situation se complique encore avec l'existence de cette mémoire réservoir chez le HIV-1 (31) : les traitements antiviraux, visant la transcriptase inverse ou les effecteurs de pré-intégration, n'affectent pas les provirus intégrés au génome cellulaire. Les virus du réservoir peuvent ainsi ré-émerger et perturber ou remplacer la quasi-espèce circulante quand celle-ci est soumise à des contraintes sélectives, bien après l'intégration du provirus. Il s'agit d'une caractéristique dont il faut tenir compte lors de la mise en place d'un traitement.

Certaines mutations maintenues à un niveau mémoire peuvent servir pour prédire une dominance précédente dans une lignée évolutive : ce sont les « harbinger mutations », observées par exemple chez le FMDV (10, 71). La mise en évidence de ces mutations pourrait avoir lieu grâce à l'utilisation croissante des NGS.

Des sous-populations, parfois à des fréquences très basses, ont donc le potentiel d'influer sur l'évolution intra-hôte de virus pouvant présenter cette caractéristique (31, 140). Il est donc légitime de penser qu'il serait intéressant de soumettre la population virale à une analyse par séquençage haut-débit ou par micro-puces afin d'identifier des marqueurs de

mémoire et d'adapter et de préparer au mieux les traitements pour chaque patient (204). On pourrait alors éviter d'administrer des médicaments pour lesquels des mutations sont apparues au sein de la population virale, ralentir l'apparition de résistance et augmenter les bénéfices du traitement.

#### **5.4. Les quasi-espèces, unité de sélection soumis à des interactions internes**

Les génomes viraux ne sont pas confinés à différents compartiments pour chaque étape de la réplication. Si l'infection est bien avancée, de nombreuses copies sont présentes en même temps, avec leurs produits d'expression. Cela renforce l'idée que la pression de sélection agit sur l'ensemble de la distribution du génome, étant donné que cette population sera toujours présente chez l'hôte.

Le modèle du spectre de mutants en constante réplication laisse supposer qu'il existe des interactions entre les génomes composant ce spectre, le plus probablement par leurs produits d'expression. Mais en quoi consisteraient ces interactions ? Quelles influences ont-elles sur la population et son évolution ?

##### **5.4.1. Complémentarité**

###### **5.4.1.1. Approches expérimentales**

Des observations expérimentales ont rapportées, de manière assez précoce, que des clones biologiques issus d'une population, non clonée, présentaient une valeur sélective moyenne plus basse que celle de cette population dont ils étaient issus (61). L'hypothèse formulée était que cette différence pouvait provenir d'une interaction entre les composants de la population, impossible lorsqu'ils étaient isolés lors du clonage biologique. Puis d'autres expériences ont conforté cette hypothèse.

Le mutant PV G64S étudié dans le cadre des « fidelity mutants » n'arrive à se répliquer dans le système nerveux central de souris que s'il est inoculé avec le PV neuropathique de type sauvage (176, 184, 244, 245) (voir partie 4.1.1.). Ce dernier possède un nombre de variants plus grand, un spectre de mutants plus étendu, et permet à une population plus restreinte de devenir infectieuse. Le PV semble donc fonctionner comme un ensemble de composants interactifs au sein d'une quasi-espèce, bien que le mécanisme soit mal connu.

Un variant mutant de MeV a la caractéristique de posséder deux types de génomes au sein d'une même particule virale : l'une d'eux code pour une protéine de fusion n'ayant subi aucune substitution, tandis que l'autre code pour la protéine de fusion avec une substitution d'un acide aminé. Les deux protéines ensemble présentent une activité de fusion améliorée, suite à la formation d'un hétéro-oligomère (217).

Chez la souris, la coexistence de variants HBV (SH et SH-DPS) semble améliorer la réplication virale et provoquer des réponses plus fortes de lymphocytes T cytotoxiques intrahépatiques et d'anticorps spécifiques des protéines de surface (35). Ainsi, ces interactions intra-quasi-espèce ne seraient pas seulement un mécanisme pour améliorer la réplication virale, mais pourraient aussi augmenter les dommages causés au foie par la réponse immunitaire, introduisant un nouveau mécanisme d'immunopathogénèse induite par le virus.

Enfin, certains modèles théoriques comme celui décrit par Arbiza tentent d'intégrer la théorie des niches écologiques afin de mieux caractériser les interactions entre les variants

viraux, ce dont ne tient pas compte la théorie des quasi-espèces (8). L'utilisation des ressources devrait en effet être un déterminant primaire de la ségrégation écologique entre autant d'individus dans un même organisme, si on considère ce dernier comme un même milieu écologique. Il démontre ainsi qu'une coopération, en plus de la compétition, est nécessaire entre les variants d'une quasi-espèce pour comprendre sa dynamique et l'appréhender en tant qu'entité.

Des analyses de populations virales par NGS rapportent de hautes fréquences de génomes ayant visiblement perdu leur caractère infectieux ou étant défectueux, c'est-à-dire produisant des particules ou des protéines défectueuses ou fonctionnant de manière non optimale (71, 72, 164, 167, 261). L'une des hypothèses de leur maintien est l'action en *trans* de protéines non défectueuses.

La littérature regorge d'autres exemples qui prônent l'existence d'une interaction entre les différents composants du spectre de mutants d'une quasi-espèce. Les quelques exemples présentés, les plus étudiés et les mieux connus, démontrent les formes diverses que peut prendre cette interaction positive entre composants, la complémentarité.

#### 5.4.1.2. Caractéristiques de la complémentarité

La complémentarité permettrait le maintien de composants dominants ou présentant des mutations délétères dans un spectre de mutants, au sein d'une population virale. Tandis qu'elle augmenterait la valeur adaptative moyenne de la population en promouvant une coopération entre variants (61), d'autres mutants, incapables d'établir une infection à eux seuls, peuvent survivre grâce à la complémentarité. L'intervention de virus infectieux dit « helper » semble nécessaire à la survie des variants mutants sur plusieurs cycles cellulaires si les cellules sont infectées avec un grand nombre de particules (42, 109). Ces interactions auraient lieu via le produit d'expression des composants dominants. Le génome du virus « helper », qui exprime une protéine fonctionnelle donnée, peut permettre la réplication d'un autre génome génétiquement proche, dont la protéine correspondante est défectueuse ou fonctionne de manière non-optimale. Cela rappelle le modèle proposé pour la mémoire moléculaire : le phénotype mixing and hiding (159, 257).

A une forte MOI, ce modèle laissait supposer que deux variants différents, ou plus, échangent librement leur séquences génétiques ou leurs protéines de surface à l'intérieur de la cellule avant d'être libérés ; à un faible MOI, la descendance mutante est encapsidée dans les protéines parentales non mutées (257). L'effet de ce phénotype est donc augmenté avec la MOI. De même, la valeur sélective de variants mutants chez le VSV semble augmenter par rapport à celle du type sauvage, avec l'augmentation de la MOI (159). Un modèle théorique suppose que deux virus co-infectant une même cellule se partagent leurs produits d'expression, et semble valider ces données expérimentales (159).

D'un point de vue évolutif, l'importance de la MOI a été rapportée pour la première fois sur des sous-populations au sein d'une quasi-espèce de FMDV (214). Ces expériences ont démontré que des sous-populations, S ou L, peuvent être sélectionnées préférentiellement en fonction de la MOI. A faible MOI, la sous-population S est favorisée car semble présenter une plus forte affinité pour le récepteur héparane sulfate présent sur les cellules à infecter, les BHK-21. Mais à forte MOI, un équilibre entre les deux sous-populations est observé car la sélection favorise aussi les variants présentant des moyens de fixation alternatifs aux cellules, les récepteurs étant surpassés par le grand nombre de particules virales. La sous-population L peut dans ces cas-là se développer (214). Certaines sous-populations démontrent donc un avantage sélectif MOI-dépendant par rapport à une autre sous-population proche génétiquement, et ici aussi le rôle d'une complémentarité ou d'interférence est supposé.

Une interprétation possible est que les chances de complémentarité augmentent avec le nombre de particules virales infectant la cellule sensible. La complémentarité est donc directement dépendante de la densité de particules virales (159). De plus, la complémentarité peut permettre le maintien de certains variants mutants : c'est ce que l'on retrouve dans le cas de la mémoire moléculaire. Ces deux phénomènes peuvent donc être liés.

Dans certains cas, cette complémentarité pourrait aussi agir pour augmenter la pathogénicité, la charge virale et la sévérité d'une maladie des virus à ARN, comme c'est le cas chez les cytomégalovirus (CMV) qui est un virus à ADN (46). Une co-infection de plusieurs souches assombrit en effet grandement le pronostic chez les patients infectés. Enfin, il est aussi possible que la complémentarité stimule la réponse immunitaire chez l'hôte (35).

Certaines études font le parallèle avec les mutations récessives retrouvées chez des organismes diploïdes, qui évitent la sélection négative quand elles sont portées chez un hétérozygote avec un allèle dominant normal (71).

## **5.4.2. Interférence**

### **5.4.2.1. Approches théoriques et expérimentales**

Une étude théorique basée sur le modèle des quasi-espèces, modélisant des réplicons à courte chaîne nucléotidique soumis à des erreurs de réplifications, a démontré qu'il était possible qu'un trait dominant exprimé par un génome soit éliminé dans un spectre de mutants, même s'il est plus adapté à son environnement (228).

Une première expérience sur le VSV a obtenu des résultats allant dans ce sens (58). Des variants Clone 1 et Clone 4 ont été isolés à partir d'une population et ont présenté une valeur sélective relativement forte comparée à celle de la population dont ils étaient issus. Cependant au sein de celle-ci, ces variants se retrouvaient seulement à basse fréquence, voire étaient supprimés. On retrouve une sous-population à forte valeur sélective supprimée par une quasi-espèce complexe avec une valeur moyenne plus basse, ce qui n'est pas sans rappeler la survie du plus plat. De même, une sous-population FMDV de variants résistants aux anticorps a supprimé une sous-population de variants présentant une forte valeur adaptative (28). A l'inverse, chez le PV, une proposition de nouveau traitement antiviral se base sur le fait que ce sont des mutants résistants aux inhibiteurs qui sont la cible d'interférence de la part du spectre de mutants du PV de type sauvage (232).

Ici encore, la littérature regorge d'exemples similaires, comme chez le LCMV (103, 234), le FMDV (93, 96, 176, 214), le bactériophage Q $\beta$  (61), le VSV (164) ou même chez des vaccins vivants contre le PV (96). Il semblerait donc qu'une interaction d'interférence empêche la répllication et la génération de certains variants, à cause d'autres variants du spectre de mutants.

### **5.4.2.2. Caractéristiques des interférences**

Lors d'expériences où l'on soumettait une population à une mutagenèse induite (FMDV et LCMV) (respectivement 96, 103), ces effets interférents semblent accentués, mais ceux-ci semblent aussi avoir lieu en l'absence d'un traitement mutagène (176). Il semble donc que l'un des mécanismes de cette interférence soit lié au haut taux de mutation, qui peut être induit ou non. Chez le FMDV, l'augmentation de ce taux a permis la formation d'un spectre de mutants hautement complexe, qui interfère avec la répllication des ARN viables (96, 176). Les ARN interférents sont dits de pré-extinction, en référence à la phase induite lors de

l'augmentation de la mutagenèse (voir partie 5.5.). Ils sont présents en excès et suppriment ainsi de manière synergique le caractère infectieux de la population (176). Chez le LCMV, ces génomes étaient défectueux, et il est possible que leur nombre ait été sous-estimé (103).

Cette capacité d'interférence ne serait pas liée, en tout cas chez le FMDV, au caractère infectieux ou non de la particule virale produite par les mutants interférents (176). L'importance des génomes défectueux est donc à prendre en compte lors des études des populations virales : ils peuvent servir de modulateur de la réplication virale. Ce rôle potentiel revêt d'autant plus d'importance que ces génomes semblent montrer des degrés différents de défektivité, permettant une régulation assez fine au sein d'une population. Ils sont de plus retrouvés en grande quantité *in vivo*, par exemple chez le HIV-1 (207). Enfin, la famille de gènes APOBEC-3G stimule la libération de particules non infectieuses du rétrovirus Friend, stimulant ainsi une réponse immunitaire d'anticorps neutralisants antiviraux (220). Il faudrait savoir si les mécanismes impliquent réellement des phénomènes d'interférence ou s'il s'agit d'une réponse immunitaire plus générale.

L'interférence chez le FMDV a été observée lorsque les ARN de pré-extinction étaient de taille complète, donc capable d'exprimer des protéines (96). Une diminution du taux de réplication du virus FMDV du type sauvage a aussi été observée lors d'un mélange avec des protéines de capsides et des polymérases issues de variants mutants (176). Cette interférence pourrait donc se produire grâce à des interactions inter-protéiques et au sein des complexes formées lors de la réplication.

L'interférence, aussi appelée défection, est l'opposé de la complémentarité. Il semblerait donc que ce mécanisme soit exercé par des génomes viables, se répliquant de manière non optimale ou codant pour une protéine fonctionnelle ou défectueuse, suite à l'acquisition de mutations (suite à l'action d'agents mutagènes ou suite à l'exploration de l'espace séquence). Ces mutations ont des conséquences sur les protéines produites, qui fonctionnent moins bien dans les conditions normales de développement du virus, voire deviennent défectueuses. Celles-ci, via des interactions au sein d'un complexe protéique ou du complexe enzymatique de réplication, vont alors affecter le fonctionnement des autres protéines non mutées (176). Grâce à ces mécanismes, un génome avec une valeur adaptative plus basse peut acquérir un avantage sélectif par rapport à un génome mieux adapté, selon la présence d'un spectre mutant favorable. Mais il semble pouvoir aller plus loin, allant jusqu'à pouvoir auto-organiser une population pour limiter sa virulence.

#### 5.4.2.3. Sélection densité-dépendante et auto-organisation

Chez le FMDV, des variants résistants aux anticorps, nommés MARLS, ont été isolés (167). Ces derniers affichaient une grande valeur adaptative et surtout une forte virulence, car ils provoquaient la destruction des cellules utilisées pour la culture (BHK-21). Cependant ces variants ne sont jamais devenus dominants au sein de la population. Le clonage biologique d'isolats issus de celle-ci a mis en évidence deux sous-populations, les clones p200, similaires aux populations de départ non clonées, et les clones MARLS-like, hautement virulents et avec une forte capacité adaptative. Ces clones ont été respectivement qualifiés de compétiteurs et de colonisateurs. L'une des hypothèses est que la population a pu se diversifier et permettre l'apparition de sous-populations génétiquement proches mais distinctes : la sous-population MARLS-like aurait été supprimée par la sous-population p200, laissant supposer une dynamique de compétition et de colonisation similaire à celle retrouvée dans des systèmes écologiques classiques (236). Une dynamique de cette co-infection a été proposée comme étant densité-dépendante (MOI-dépendante) pour les deux sous-populations, ce qui a été

confirmé expérimentalement (167). Les colonisateurs, très adaptés et hautement virulents, permettent une atteinte cellulaire rapide et efficace à faible MOI, dans la phase précoce de l'infection où la densité cellulaire est basse. Lorsque celle-ci augmente, leur réplication est limitée par les compétiteurs qui modulent ainsi la virulence en cas de besoin à forte MOI. Un équilibre entre ces deux sous-populations peut à terme être atteint.

Cette étude a permis l'établissement des simulations théoriques qui démontrent une modulation de la virulence grâce aux interactions de sous-populations virales lors de co-infections cellulaires (59).

Chez le LCMV, il existe des variants qui induisent le Growth Hormone Deficiency Syndrome (GHDS-positive WE C2.5) chez des souriceaux nouveau-nés, mais certains isolats, en minorité dans la population, n'ont pas cette capacité à causer la maladie (GHDS-negative WE c54) (234). Un large excès de ces variants GHDS-negative WE c54 peut restreindre la réplication des GHDS-positive WE C2.5 dans les cellules productrices des hormones de croissance. Cela prévient ainsi le développement du syndrome à partir d'un certain seuil (234).

Le FMDV et le LCMV apportent donc des exemples d'évolution des quasi-espèces vers des sous-ensembles qui affichent des nouveaux traits phénotypiques, afin de s'auto-organiser pour moduler la destruction cellulaire. Les composants d'un spectre de mutant agissent ici en tant qu'ensemble pour commander le comportement général de la population, dans une dynamique de compétition-colonisation retrouvée dans la théorie des niches écologiques. L'émergence de nouveaux virus potentiellement pathogènes pourrait bien être prévenue par une telle dynamique d'interférence au sein de la population (167, 234). Seuls des goulots d'étranglement vont ainsi provoquer la sélection de ces variants hautement pathogènes à la valeur sélective plus grande (58, 103, 234).

### **5.4.3. Mécanismes moléculaires de la complémentarité et de l'interférence**

#### **5.4.3.1. Deux mécanismes liés**

Ces deux mécanismes sont intimement liés. En effet, la complémentarité exprime le caractère d'une protéine produite par un sous-ensemble de génome ayant la capacité de « sauver » une version de cette protéine exprimée par un autre sous-ensemble, mais défectueuse ou dont la fonction n'est pas optimale. Mais si l'expression de ces protéines défectueuses dépasse un certain niveau, elles vont empêcher les protéines non défectueuses de fonctionner normalement : il s'agira alors d'un phénomène d'interférence. L'étude la plus intéressante sur le sujet concerne les particules DI, rendues non infectieuses par des délétions et des mutations ponctuelles chez le VSV, bien qu'il s'agisse d'un exemple un peu particulier d'interférence (164). De telles particules DI ont été trouvées chez le FMDV (42).

Ces particules sont incapables de coder certains facteurs de réplication ou codent des facteurs défectueux, ce qui explique qu'ils soient éliminés dans la plupart des cas. Cependant les VSV DI peuvent faire concurrence aux virus standards quant aux complexes de réplication qui sont disponibles dans le milieu. Ils peuvent alors se répliquer et même devenir dominants sous un MOI élevé car leur réplication aura lieu au dépend de celle des VSV standards. Mais la dominance des VSV DI est inévitablement limitée, étant donné que la concentration de facteurs nécessaires à leur réplication sera décroissante si les DI deviennent dominants. Avant que la décroissance des VSV standards ne devienne trop importante, des mutations chez ces

derniers peuvent provoquer l'apparition d'un phénotype accordant une certaine résistance aux particules DI, nommé Sdi<sup>-</sup> (164). L'une des principales hypothèses est que l'acquisition de ces mutations chez le variant Sdi<sup>-</sup> agirait au niveau du site d'initiation de la réplication et permettrait aux facteurs de réplication de différencier les VSV DI des VSV standards, avec une sensibilité réduite pour les premiers. Les standards retrouveront alors leur dominance, jusqu'à ce que d'autres DI soient produits, plus adaptés à eux, etc. Cela établit une alternance cyclique de VSV et de leur DI correspondant qui illustre le changement continu possible en évoluant rapidement (108) et cela rejoint l'hypothèse de la Reine rouge.

C'est un exemple d'alternance d'interférence et de complémentarité, de coévolution à l'échelle des génomes viraux, mais aussi de rétrocontrôle, car la dominance de la forme défectueuse ou de la forme standard dépend de la génération régulière de mutations dans l'ARN du virus, ces formes agissant l'une sur l'autre. L'observation de ce VSV standard et de ses DI engagés dans une dynamique continue de mutation et de compétition a conduit les chercheurs à reconnaître la valeur de la dynamique des quasi-espèces pour comprendre le comportement des virus, qui s'applique très bien dans ce cas précis (108).

#### 5.4.3.2. Différences avec les mécanismes inhibiteurs généraux

Plusieurs mécanismes biologiques expliquent des activités inhibitrices que l'on retrouve sur les virus (267) : l'induction d'interféron, la réponse immunitaire, la production de polypeptides défectueux qui entrent dans un complexe multimérique, une fonction défectueuse qui peut séquestrer un facteur nécessaire pour la réplication de deux virus, la compétition pour les facteurs de l'hôte. Ce sont des mécanismes d'inhibition généraux.

Pour établir les véritables mécanismes d'interférence, il faut vérifier la spécificité de ces interactions d'interférences. Par exemple, lors d'expériences avec le FMDV, les mutants avec des capsides altérées et des polymérases altérées inhibaient la réplication du FMDV mais pas des autres picornavirus comme l'EMCV (176). Une co-transfection d'une même quantité d'ARN différents (HIV-1 ou même ARN d'*E.coli*) avec des ARN de pré-extinction, qui démontraient pourtant une action d'interférence sur le FMDV, n'a apporté aucun signe d'interférence entre ceux-ci (96). Via d'autres expériences, l'action d'ARN de petite taille, l'introduction d'inhibiteurs intracellulaires (96) ou la compétition directe pour les facteurs de l'hôte (234) n'ont pas été retenus comme mécanismes possibles. De même, aucune activité des interférons, ainsi que toute activité antivirale générale associée à une réponse immunitaire innée, n'ont été observées chez le FMDV de la part des ARN de pré-extinction (96). Une réponse immunitaire a bien été observée dans le cas du Friend virus (220) ou chez le HBV (35), mais il n'est pas dit qu'elle soit liée directement à l'action d'interférence. On peut ainsi exclure qu'il s'agisse de mécanismes inhibiteurs généraux évoqués plus haut.

#### 5.4.3.3. Mécanisme de complémentarité et d'interférence

Ainsi, les mécanismes d'interaction semblent se situer à un niveau plus intime, entre protéines produites par les variants au centre de ce phénomène, lors du cycle viral (176, 217, 232). Une manière d'aborder la complémentarité et l'interférence est de s'attarder sur l'étude de protéines agissant en tant qu'homo- ou hétéropolymères (217), par exemple des capsides virales dont la stabilité dépend d'une bonne interaction entre ses sous-unités, ou des complexes de réplication incluant des protéines virales et de l'hôte. Ainsi, une sous-unité exprimée par un génome défectueux sera défectueuse et va compromettre la fonction de

l'ensemble des protéines. Cela va mener à une capsidie instable ou un complexe de réplication non optimal (176, 232) (figure 42).

Un même gène, codant pour une sous-unité protéique participant à un hexamère, peut subir des mutations (figure 42 A). Cela peut rendre le génome défectueux et coder pour une protéine défectueuse qui va compromettre l'ensemble protéique.

Plus l'hexamère est composé de sous-unités défectueuses, moins il est fonctionnel, donc plus on a affaire à une interférence (figure 42 B). On peut ainsi remarquer des degrés différents d'interférence en fonction de la sous-unité intégrée dans l'hexamère.

Il faut noter que l'on peut considérer ce mécanisme de deux façons : la sous-unité bleue diminue l'activité de l'hexamère fonctionnel jaune, ou sauve partiellement l'activité de la protéine hexamérique non fonctionnelle rouge. Ce sont deux mécanismes liés. La diminution progressive de l'activité de la protéine se rapporte aux différents degrés de défectuosité retrouvés *in vivo* (207).

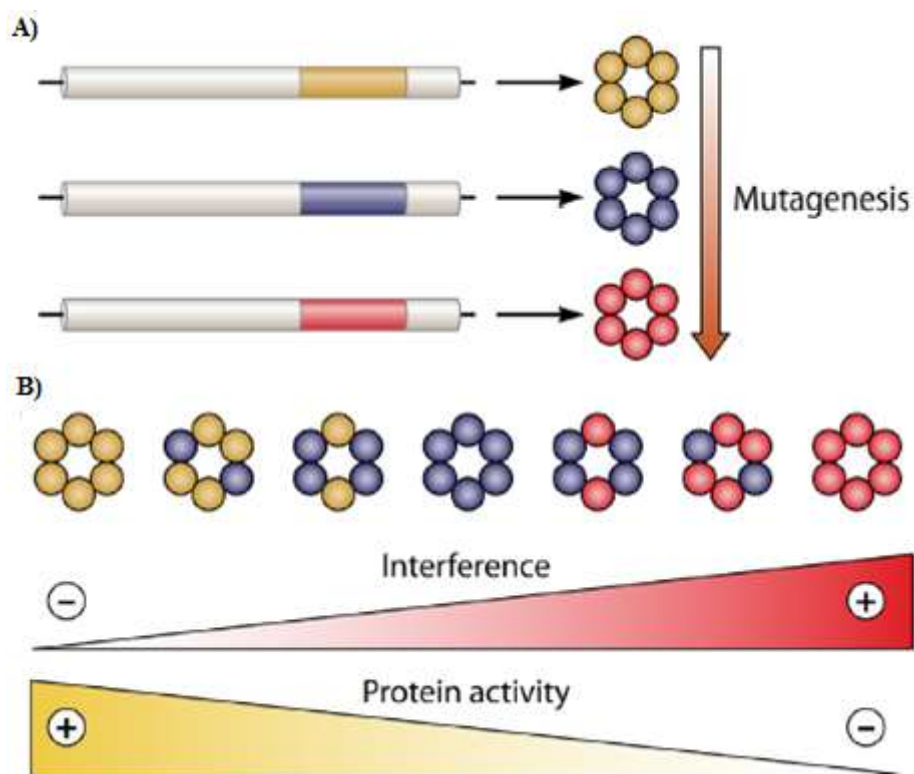









Figure 42 : Modèle de complémentarité et d'interférence médiée par des interactions au sein d'une protéine polymérique (71)

-  Gène standard et sous-unité protéique appartenant à un hexamère, exprimée par ce gène.
-  Gène défectueux apparu suite à l'augmentation de la mutagenèse, et sous-unité protéique appartenant à un hexamère, exprimée par ce gène.
-  Gène défectueux apparu suite à l'augmentation de la mutagenèse, et sous-unité protéique appartenant à un hexamère, exprimée par ce gène.
-  Augmentation de la mutagenèse.
-  Hexamères d'activité protéique décroissante (de gauche à droite).
-  Interférence exercée par les sous-unités protéiques au sein de l'hexamère.
-  Estimation de l'activité protéique des hexamères formés à partir des différentes sous-unités protéiques.



Une nouvelle approche antivirale est basée sur ce mécanisme, à l'aide d'un inhibiteur V-073 chez le PV, *in vitro* sur culture cellulaire et *in vivo* chez des souris. Ce dernier cible les constituants viraux d'une capsid virale, dont l'assemblage est comparable à celui d'un oligomère chimérique. Cet assemblage se fait à l'aide de protéines codées par des génomes résistants et sensibles à l'inhibiteur, présents tous les deux dans la cellule suite à la dynamique de sélection qui a lieu au sein d'une quasi-espèce virale. Ce sont cependant les génomes sensibles à l'inhibiteur qui dominent et sont la cible préférentielle. La croissance virale a ainsi été inhibée et l'apparition de variants résistants a été évitée, car ils sont constamment éliminés à cause de ces capsides chimériques (232).

#### 5.4.3.4. Difficultés pour établir les mécanismes d'interactions

Des difficultés d'interprétation se retrouvent à deux niveaux : à celui de la protéine, si l'on suppose que c'est bien à ce niveau que les interactions ont lieu ; et à celui de la cellule.

On peut supposer que ces interférences sont renforcées par la nature multifonctionnelle des protéines virales (72). Ainsi, une protéine défectueuse va compromettre chaque étape du cycle viral où elle intervient, sauf si des domaines bien distincts de cette protéine sont impliqués dans des étapes bien précises du cycle et que la mutation affecte seulement l'un de ces sites. De plus, des protéines peuvent présenter de nombreuses isoformes à un instant t, ce qui complique encore les possibilités d'interaction. De nouveaux outils devront être développés afin de clarifier ces mécanismes d'interaction intra-quasi-espèces.

Pour qu'un spectre de mutants interfère avec une classe spécifique de variants, il doit être généré dans la même cellule que cette classe ou infecter la cellule dans laquelle la classe se réplique. Cette infection peut être réalisée simultanément, ou de manière successive, s'il n'y a pas de mécanisme d'exclusion de super-infection de la part de la cellule. Bien que des infections doubles ou multiples semblent improbables, il se peut qu'elles soient plus fréquentes que ce que l'on peut s'attendre à trouver, car on peut la retrouver chez des virus à ADN comme les CMV (46). De plus, l'altération d'une classe de variants par une autre est possible via l'expression altérée de facteurs de régulation diffusibles (46). Etant donné que ces co-infections ne semblent pas si rares, cela complique l'interprétation des interactions, car celles-ci peuvent avoir lieu en même temps, directement, indirectement, voire engager la population dans une dynamique continue de mutation, sélection et compétition comme dans le cas du VSV (108, 164).

#### 5.4.3.5. Interactions des quasi-espèces en tant que groupe

Les interactions mutuelles au sein de ce spectre et entre les produits de leur génome définissent les quasi-espèces comme une unité de sélection. Une modulation des traits phénotypiques exprimés est réalisée par une population entière, et ces derniers ne peuvent pas être attribués à des génomes spécifiques de cette population. Les deux mécanismes d'interaction peuvent être vus comme une propriété collective de la population virale. Cela implique que le comportement d'un ensemble n'est pas forcément prédictible grâce au comportement des individus.

Ces interactions intra-quasi-espèces peuvent être exercées au niveau intracellulaire par les variants déjà présents, ou par des mutants d'origine exogène qui pénètrent dans la cellule, introduits par co-infection, réinfection ou même polypléidie chez certains virus. Dans tous les cas, le spectre de mutants de la quasi-espèce peut moduler l'expression de ses composants. Ce

spectre joue encore un rôle central dans le comportement répliatif de la population, ce qui est bien prédit par la théorie des quasi-espèces (69).

Le taux de mutation semble un des points clés du mécanisme des interactions. En augmentant ce taux artificiellement (96, 103), des variants défectueux apparaissent de plus en plus et augmentent les possibilités d'interférences au sein de la population. Celle-ci subit une transition vers l'extinction (voir partie 5.5.). Ce modèle a ouvert la voie à de nouvelles possibilités thérapeutiques sans doute prometteuses : la mutagénèse létale.

Un modèle a proposé d'intégrer les deux formes d'interaction, en fonction du taux pour comprendre cette transition, sur la figure 43 :

Les interactions de complémentarité dominent quand l'ARN viral se réplique avec un taux de mutation standard (figure 43A). Sous de telles conditions, la valeur adaptative de l'ensemble de la population est plus grande que celle des individus qui composent la population. Cela permet de maintenir des génomes défectueux.

Quand le taux de mutation augmente, les interactions d'interférence dominent car le nombre de variants avec un génome défectueux va devenir trop grand et avoir un impact sur ceux au génome standard (figure 43 B). Quand l'interférence semble prévaloir, la valeur adaptative des individus viables sauvés tend à être plus grande que la valeur adaptative de la population parentale (96).

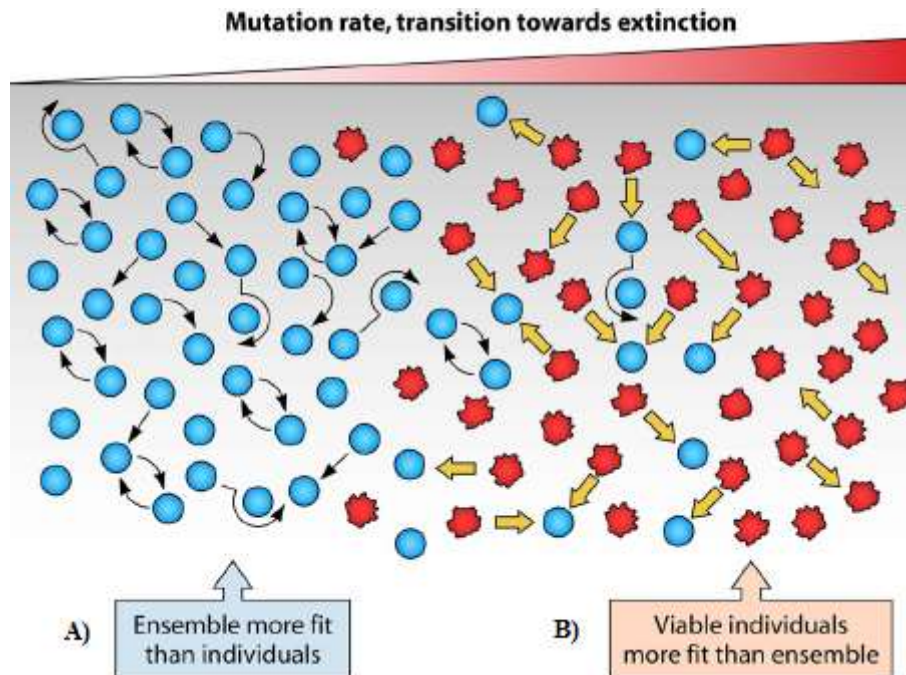


Figure 43 : Représentation schématique des interactions de complémentarité et de défection au sein d'un spectre de mutants (71 (adapté de (70))).

- Chaque symbole (même de la même forme) dépeint un variant possédant un génome avec une séquence génomique légèrement différente.
- ▲ Augmentation du taux de mutation
- Variant possédant un génome standard
- Variant possédant un génome défectueux, réalisant donc une interférence.
- ↔ Interaction de complémentarité
- ➡ Interaction d'interférence

Ce niveau d'interaction a été moins exploré, chez les virus, que les interactions au sein d'une même molécule d'ARN viral, l'épistasie. Certains chercheurs parlent d'épistasie positive quand la valeur adaptative moyenne d'un ensemble de virus de la population est plus grande que celle de ses individus pris à part, et d'épistasie négative dans le cas contraire (71).

Ce modèle démontre une certaine dualité dans la population virale, reposant sur le taux de mutation, mais aussi sur l'intensité et la spécificité des contraintes sélectives. Une fois un certain seuil dépassé, une sélection à l'échelle de l'individu peut avoir lieu car ceux-ci sont plus viables que l'ensemble. Une telle dualité pourrait fournir un avantage sélectif aux virus, en leur fournissant une certaine flexibilité pour répondre aux contraintes sélectives. Il s'agit cependant pour l'instant de notions théoriques et très floues qui méritent la mise en place de nombreux modèles théoriques et des confirmations expérimentales.

## **5.5. La mutagénèse létale**

L'un des apports majeurs de la théorie des quasi-espèces appliquée aux virus à ARN est la relation du seuil d'erreur, une autre conséquence du fort taux d'erreurs réalisées lors de la réplication. C'est donc naturellement que les chercheurs se sont interrogés sur son application pratique et les conséquences de l'augmentation du taux de mutation sur une population virale, à des fins thérapeutiques.

### **5.5.1. Bases de la stratégie antivirale**

L'objectif d'une stratégie antivirale est d'atteindre une charge virale basse, diminuant ainsi souvent la pathogénicité et apportant au système immunitaire une opportunité de détruire le virus, voire de prévenir les signes cliniques de la maladie. Les nombreuses études sur l'efficacité de ces stratégies portent sur la valeur adaptative des virus sur lesquels sont testés les traitements, car une faible charge virale est liée à une faible capacité adaptative (49).

Depuis le siècle dernier, elles n'ont cessé de se développer : immunothérapie, chimiothérapie, vaccination, utilisées séparément ou conjointement. Cependant, pour chacune d'entre elles, il est possible de décrire des limites claires, qui sont liées au potentiel adaptatif des virus. Celui-ci a été largement ignoré lors de la mise en place de protocoles thérapeutiques. Ce n'est que récemment qu'il a été considéré comme nécessaire d'incorporer des principes évolutifs et écologiques pour comprendre le développement des maladies infectieuses, qu'elles soient virales ou non. La dynamique des quasi-espèces représente l'un des aspects des problèmes liés à la prévention et au contrôle des maladies (68, 70, 86).

Ils se définissent notamment par la sélection de variants résistants aux traitements au sein du spectre de mutants. Si l'on considère la figure 28, celle-ci ne représente que 30 variants et présente déjà une certaine complexité. On peut facilement imaginer ce que cela représente à l'échelle d'une population virale *in vivo* : pour décrire une quasi-espèce entière, il faudrait reprendre cette figure des millions de fois. En ajoutant le fait que cela va changer d'un patient à l'autre, et d'un instant  $t$  à un instant  $t+\Delta t$  chez un même patient, la population virale est une véritable cible mouvante et l'impact de cette dynamique est presque imprédictible. Il est quasiment certain qu'au sein de ce spectre va se retrouver au moins un variant qui peut échapper à la stratégie qui pourra être mise en place. Plusieurs types de stratégies ont été développés ou proposés pour essayer de minimiser ou ralentir leur sélection. L'une d'entre elle est la mutagénèse létale (49, 90).

### 5.5.2. Principes de la mutagénèse létale

De nombreux modèles théoriques ont tenté d'expliquer la mutagénèse létale : ils sont divers, parfois peu intuitifs, et peuvent parfois se reposer non pas sur la théorie des quasi-espèces ni sur la notion de seuil d'erreur, mais plutôt sur un seuil de mutation dit critique. Perales et Domingo ont tenté de résumer les données apportées par les résultats expérimentaux et de les confronter aux modèles les plus réalistes qui sont rapportés dans la littérature, afin d'apporter une proposition de mécanisme se rapprochant le plus possible de ce qui peut se trouver dans la réalité (71).

Bien que remis en question, dans le contexte de la théorie des quasi-espèces, la stabilité de l'information génétique dépend de la longueur de la séquence nucléotidique, du taux d'erreurs réalisées pendant la réplication, et de la supériorité relative de la séquence maîtresse au sein de son spectre de mutants. Il peut donc être intéressant de viser l'une de ces variables (90). La mutagénèse létale exploite le concept de violation du seuil d'erreur et des interactions d'interférence entre les composants d'un spectre de mutants dans une population virale, décrits par la théorie des quasi-espèces (5, 66, 99). Cette stratégie vise ici à provoquer un excès de mutation dans les génomes viraux, amenant la population à franchir un seuil critique de mutation. Via l'accumulation de mutations qui sont souvent délétères et l'apparition de génomes défectueux interférents, les séquences obtenues seront tellement éloignées de la séquence initiale que l'on obtient des génomes non fonctionnels. L'objectif final est d'atteindre l'extinction virale, via la mutagénèse ou en diminuant suffisamment la charge virale pour que l'hôte puisse prendre le relais dans l'élimination du virus.

### 5.5.3. Les analogues nucléotidiques et nucléosidiques

L'augmentation du taux de mutation peut être accomplie grâce à des agents mutagènes sous forme d'analogues nucléotidiques et nucléosidiques. L'action de la mutagénèse a été démontrée pour la première fois lors d'expériences sur le VSV et le PV (109, 127). Lors de la découverte de la perte du potentiel répliatif du HIV-1 causée par un analogue, le 5-hydroxydéoxycytidine (5-OH-dC), il a alors été proposé que des analogues mutagènes soient utilisés en tant qu'agents anti-HIV-1 (131). Depuis, de nombreuses études ont rapporté l'extinction de virus à ARN *in vitro*, qui emploient pourtant des stratégies différentes de réplication (FMDV, LCMV, HCV, HIV-1, Hantaan virus, GB virus B, West Nile virus, IV), en présence de ces analogues, notamment la R, le 5-AZC et le 5-FU (71, 203).

Ces agents sont convertis sous la forme d'un nucléoside triphosphate une fois dans la cellule. Ils sont ensuite incorporés à l'ARN lors de la réplication, ce qui va provoquer des mutations ponctuelles lors de la réplication suivante, et parfois des interruptions de l'ARN synthétisé. Cette incorporation va être due à un couplage anomal entre la base de l'agent mutagène et celle des nucléotides standards, spécifique de chaque analogue nucléotidique (72). Par exemple, le 5-FU a tendance à se coupler avec les bases A et G, alors que la R a tendance à se coupler avec les bases U et le C. Ce couplage va provoquer une cascade de mauvaises incorporations de nucléotides lors des répliations suivantes, dans la séquence synthétisée de polarité positive ou négative. L'étude de ces cascades permet d'évaluer les types de mutations que l'on peut espérer trouver dans le spectre de mutants lors de l'utilisation des analogues nucléotidiques. Ainsi, l'incorporation de 5-FU semble provoquer un excès de transitions  $A \rightarrow G$  et  $U \rightarrow C$  (101, 203), et la R provoque un excès de transitions  $G \rightarrow A$  et  $C \rightarrow U$  (4, 151, 180, 266). Cependant de telles transitions ont été retrouvées chez des populations de FMDV, qu'elles aient été mises en présence de R ou non (181). Cela démontre que l'effet majeur des analogues nucléotidique (tout du moins celui de R) serait d'accélérer la

formation de génomes hypermutés, donc d'augmenter la fréquence de ces transitions (180, 181). Ce phénomène serait délétère chez le FMDV par exemple, et pourrait provoquer son extinction.

En dehors de modèles théoriques ou expérimentaux *in vitro*, la mutagenèse létale semble être active et efficace *in vivo* face au LCMV chez des souris (203). Un premier essai clinique a été réalisé en 2011 dans le cadre de la « viral decay acceleration » ou mutagenèse létale chez le HIV-1, stratégie thérapeutique utilisant un analogue mutagène, le KP1461. Il a montré une augmentation de la fréquence de mutation du HIV-1 chez des patients atteints comme le feraient les enzymes de la famille APOBEC (151). Même si aucune diminution de la charge virale n'a encore été enregistrée, l'analogue s'est montré efficace chez les patients traités. Ces deux exemples montrent des aspects pratiques et surtout prometteur de l'utilisation de cette méthode.

La recherche d'agents mutagènes ou non mutagènes sous forme d'analogues nucléotidiques est devenue un domaine très actif de la recherche. De nouveaux analogues nucléotidiques sont constamment recherchés et surtout importés du domaine de la thérapie anticancéreuse et antibactérienne.

## 5.5.4. Mécanismes de la mutagenèse létale

### 5.5.4.1. Caractéristiques de la mutagenèse létale

Plusieurs caractéristiques sont mises en évidence dans les modèles théoriques et résultats expérimentaux :

- Une mutagenèse est bien entendu nécessaire. Elle augmente de manière considérable la fréquence de mutation, et diminue la proportion d'ARN infectieux. Cette augmentation semble être centrale et est soutenue par le fait que l'extinction n'a pu être atteinte via l'utilisation d'inhibiteurs non mutagènes seuls chez le FMDV (172, 173, 180). Elle est provoquée par les agents mutagènes et donne des génomes hypermutés.
- La mutagenèse est corrélée à la diminution du caractère infectieux de la population (131). Cette perte serait due à des interactions d'interférence de la part de virus possédant un génome non infectieux mais capable de se répliquer (102, 103). L'augmentation de la fréquence de mutation diminue donc la proportion d'ARN viral infectieux (figure 44). La perte du caractère infectieux précède la perte de la capacité de répllication de l'ARN (103).
- La séquence consensus est invariable durant la transition vers l'extinction (97, 102).
- On considère que la complexité du spectre de mutants augmente (101, 180). Cela revient à l'exploration de régions de l'espace séquence qui ne sont normalement pas explorées à cause des contraintes intrinsèques au virus (181) (voir partie 3.3.2.2. et figure 27). Dans certains cas, ce mouvement dans l'espace séquence peut avoir lieu avec une apparition précoce et transitoire de génomes hypermutés (102, 166, 180, 181). Il y a alors expansion puis compression du spectre de mutants suite à une sélection négative (166).
- L'efficacité de la mutagenèse létale peut être favorisée par la nature multifonctionnelle des protéines et des domaines au sein de celles-ci. Une seule substitution peut avoir de multiples effets à différentes étapes du cycle viral (69, voir aussi partie 5.4.3.4.).
- La diminution de la charge virale n'est pas le mécanisme de la mutagenèse, elle en est une conséquence (71)

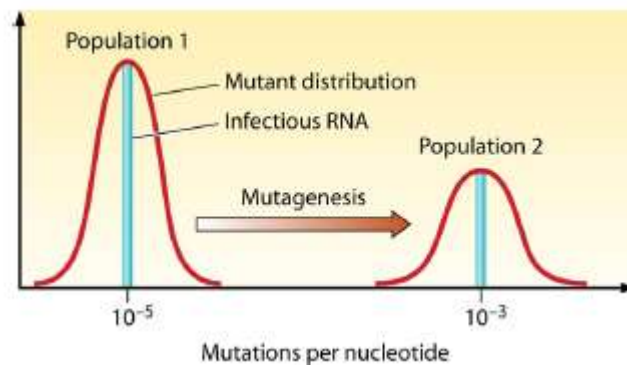
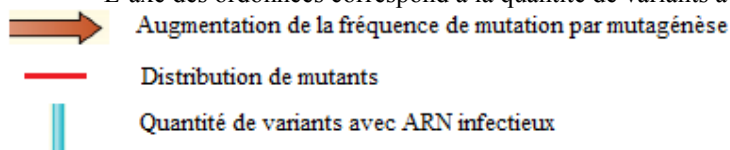


Figure 44 : Effets de l'augmentation de la mutagenèse dans une population de virus (71)

- L'axe des abscisses correspond à la fréquence de mutation, que l'on augmente par un facteur 100.
- L'axe des ordonnées correspond à la quantité de variants avec ARN infectieux.



Sur la figure 44, la mutagenèse augmente 100 fois la fréquence moyenne de mutation dans le spectre de mutant. Cela provoque une diminution du nombre d'ARN infectieux et une expansion du spectre de mutants.

#### 5.5.4.2. Etapes de la mutagenèse létale

Domingo et Perales proposent un mécanisme de la mutagenèse létale en prenant en compte les caractéristiques mises en avant ci-dessus (71). Elle consiste principalement en trois étapes et sont résumées sur la figure 45A, tandis que la figure 45B met en avant les différents éléments pouvant faire pencher la balance vers une conservation ou une perte de l'information génétique.

Une augmentation du taux de mutation chez un virus à ARN va, d'une part, favoriser l'accumulation de mutations qui sont pour la plupart délétères, et, d'autre part, augmenter la fraction de génomes défectueux (figure 45 A1). Mais certains sont encore capables de se répliquer en utilisant les facteurs de variants standards. Ils sont appelés des « defectors » (ou transfuges) (102, 103) ou ARN de pré-extinction (176). Ceux-ci interfèrent avec la réplication du variant standard du virus au sein de la population (voir parties 5.4.2.2 et 5.4.3.2.). Ce modèle est à la fois supporté par des résultats expérimentaux (70, 96, 103, 135, 176, 180) et des modèles théoriques (33). Il est appelé le modèle de la défection létale et constitue la première phase supposée de transition vers l'extinction virale par la mutagenèse létale.

Si le taux de mutation continue d'être aussi élevé ou augmente encore, les mutations létales deviennent de plus en plus importantes et participent directement à l'extinction du virus en augmentant le nombre de variants défectueux, qui vont être amenés à disparaître (figure 45 A2). Il s'agit ici de la deuxième phase supposée, dite de létalité ouverte (71).

Cela contribue à terme à la suppression de l'ensemble des virus de la population en dépassant le seuil d'erreur (figures 45 A3 et A4). Cette dernière phase peut être vue comme une transition d'une étape où il y a encore une information génétique cohérente vers une autre où toutes les séquences génomiques virales dégénèrent en séquences aléatoires. On se retrouve dans une phase de « non-information » (71). Il y a perte de l'information génétique.

Plusieurs événements peuvent faire pencher la balance vers une fidélité 1 ou 0 (figure 45 B):

- Les mutations létales et les interférences conduisent les populations virales vers une perte de l'information.
- La sélection de mutations conférant une résistance à un agent mutagène préserve l'identité des quasi-espèces.

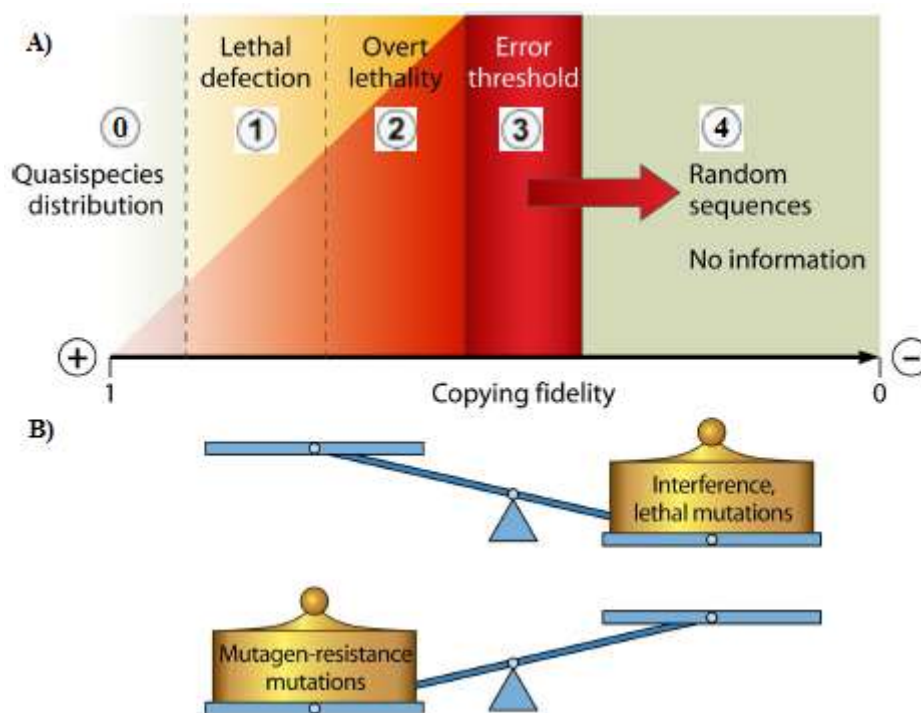



Figure 45 : Evaluation schématique actuelle des événements impliqués dans l'extinction virale (71)

A) Les différentes étapes de la mutagénèse en fonction du taux de mutation (ou de la fidélité de copie, allant de 1 pour une copie exacte, à 0 pour une copie complètement différente).

 Augmentation de la proportion de mutations létales et de "defectors"

① Distribution d'une quasi-espèce

② Défection létale

③ Létalité ouverte

④ Seuil d'erreur

⑤ Ensemble de séquences aléatoires : perte de l'information génétique

B) Événements faisant pencher la balance

Cependant la cinétique d'accumulation des mutations d'interférence ou de mutations létales n'est pas encore connue, et il est probable que le modèle doive s'étayer avec la participation d'autres virus ou même de l'hôte. De même, on ne peut encore définir de seuil de mutation universel qui signera l'extinction chez toutes les populations virales à ARN, ou pour le même virus sous des conditions différentes. La définition de ce seuil est d'autant plus complexe que la transition entre défection létale et létalité ouverte est floue.

On rencontre encore une fois la dynamique que la théorie des quasi-espèces a apportée à l'étude de l'évolution virale à ARN : le remplacement d'une sous-population de génome par une autre. Ici, une sous-population de génomes standards est peu à peu remplacée par une



population de génomes défectueux et interférents. En terme d'espace séquence, la mutagenèse létale provoque une relocalisation de manière incontrôlée, sans suivre un quelconque gradient de valeur sélective, d'une population vers des zones normalement non compatibles avec le maintien de l'information génétique.

## 5.5.5. Limites de la mutagenèse létale

### 5.5.5.1. Termes utilisés

Comme exposé dans le modèle simplifié de Bull (voir partie 3.1.3.), il y a une distinction à faire entre erreur catastrophique et extinction, ainsi que seuil d'extinction et seuil d'erreur. Dans les études qui visent à comprendre l'implication de ces notions quant aux mécanismes de la mutagenèse létale, la distinction n'est généralement pas faite, ou l'utilisation du terme peut être mauvaise (33). Beaucoup d'auteurs assimilent la mutagenèse létale avec le franchissement du seuil d'erreur et l'entrée dans une phase d'erreur catastrophique, ce qui n'est pas exactement la même chose. Cette phase correspond en effet à un remaniement de la distribution dans la population virale, avec perte de supériorité du type sauvage, mais pas à l'extinction de la population. L'erreur catastrophique pourrait même retarder l'extinction grâce à la dominance de génotypes plus robustes (32, 33).

Cependant la mutagenèse létale peut être vue comme plusieurs phases d'erreur catastrophique successives, qui mèneront en effet à l'extinction de la population. Le franchissement du seuil d'extinction n'est pas une condition nécessaire de la mutagenèse : toute diminution de la charge virale et de sa valeur sélective peut permettre au système immunitaire d'agir et d'arriver à atteindre cette extinction (33). Finalement, seul le principe de franchissement d'un seuil critique de mutation est important ici, et les nombreuses expériences rapportent bien une efficacité de la mutagenèse pour atteindre l'extinction. La théorie des quasi-espèces a servi de concept de base à cette stratégie antivirale mais ne lui correspond pas avec exactitude, et aucun modèle théorique sous-jacent précis n'existe pour le qualifier entièrement, d'où la proposition de Domingo et Perales (71).

### 5.5.5.2. Effets bénéfiques de la mutagenèse létale pour le virus

L'augmentation du taux de mutation via l'utilisation d'agents mutagènes a des conséquences négatives sur les virus à ARN, car il y a accumulation de mutations délétères et génération de mutants défectueux. Mais il peut y avoir certaines circonstances où la capacité répliquative des virus peut être augmentée, grâce à l'apparition de mutations permettant de contrer les effets des mutagènes, lorsque le génotype possède une faible valeur sélective (135). Le 5-AZC a amélioré la capacité répliquative chez le bactériophage Q $\beta$  s'il était soumis à des goulots d'étranglement, mais a conduit ce même virus à l'extinction dans d'autres conditions de développement (12, 38). L'intervention d'un événement de goulot d'étranglement permet d'éliminer la plupart des génomes « defectors », limitant ainsi les interactions d'interférence et prolongeant la réplication au sein de la population soumise à la mutagenèse (38, 135). Cela peut avoir un impact sur l'utilisation de la mutagenèse *in vivo* étant donné que ces événements sont considérés comme fréquents, comme par exemple dans le cas d'une nouvelle infection vers une nouvelle cellule.

Théoriquement, il est donc très difficile, voire impossible, de prédire l'orientation vers l'extinction ou la survie de la population lors de la mutagenèse (180). Il semblerait que cela dépende de l'état de la population (135). A l'équilibre, elle va augmenter le nombre de



mutations délétères et diminuer la valeur sélective moyenne de la population. Mais dans une population en constante évolution, un taux de mutation élevé va augmenter la variabilité au sein de la population, permettant une adaptation plus rapide (38) voire favoriser l'apparition de formes résistantes (177).

A cela s'ajoutent la charge virale et l'historique évolutif de la population, qui vont avoir un impact sur la possible apparition de mutations conférant des résistances aux processus utilisés (110, 180). On suppose que l'élément déterminant quant au comportement de la population virale se trouve dans la proximité de la population vis-à-vis de son seuil d'erreur.

#### 5.5.5.3. Résistance à la mutagenèse létale

Comme tous les procédés thérapeutiques antiviraux, on peut supposer que des mutations conférant une résistance peuvent apparaître au sein du spectre de mutants. Les mutants résistants à la R ont par exemple été décrits pour plusieurs virus, comme le PV, le HCV, le FMDV ou encore le bactériophage Q $\beta$  (12, 84, 85, 177, 185, 243, 266). Toutes ces mutations semblent être en lien avec le mécanisme d'action supposé de la R et vont limiter l'augmentation de la fréquence de mutation au sein de la population virale (90).

##### 5.5.5.3.1. *Plusieurs mécanismes de résistance*

Les mêmes variants G64S chez le PV ont été obtenus dans deux laboratoires différents, dont la polymérase démontrait une fidélité augmentée et une résistance à R, mais aussi une complexité du spectre de mutants diminuée avec les conséquences que l'on connaît (184, 243). Cela peut laisser imaginer qu'il existe peu de mécanismes différents pour qu'un variant résistant aux agents mutagènes soit sélectionné au sein d'un spectre de mutants de picornavirus.

Vignuzzi propose ainsi une interprétation graphique des conséquences de cette forte fidélité de réplication sur la résistance à la mutagenèse chez le PV, sur la figure 46. Pour bien la comprendre, il faut reprendre la figure 28 (243).

Si l'on augmente le taux de mutation (figure 46 A), donc la fréquence de mutation, via un traitement mutagène agissant sur une population de type sauvage (mutagenized population), la distribution entière se retrouve déplacée au-delà du seuil d'erreur, dans la phase d'erreur catastrophique. La majorité de la population virale du type sauvage va disparaître, laissant la place à une autre distribution (voir figure 28).

Une forte fidélité de réplication a pour conséquence un spectre de mutants plus étroit, mais qui se réplique plus loin du seuil d'erreur (figure 46 B, normal population). Un traitement mutagène va résulter en une moindre augmentation de la fréquence de mutation chez la population G64S. Il en résulte une expansion puis une compression de la complexité du spectre, sans dépasser le seuil d'erreur et atteindre la phase d'erreur catastrophique (mutagenized population).

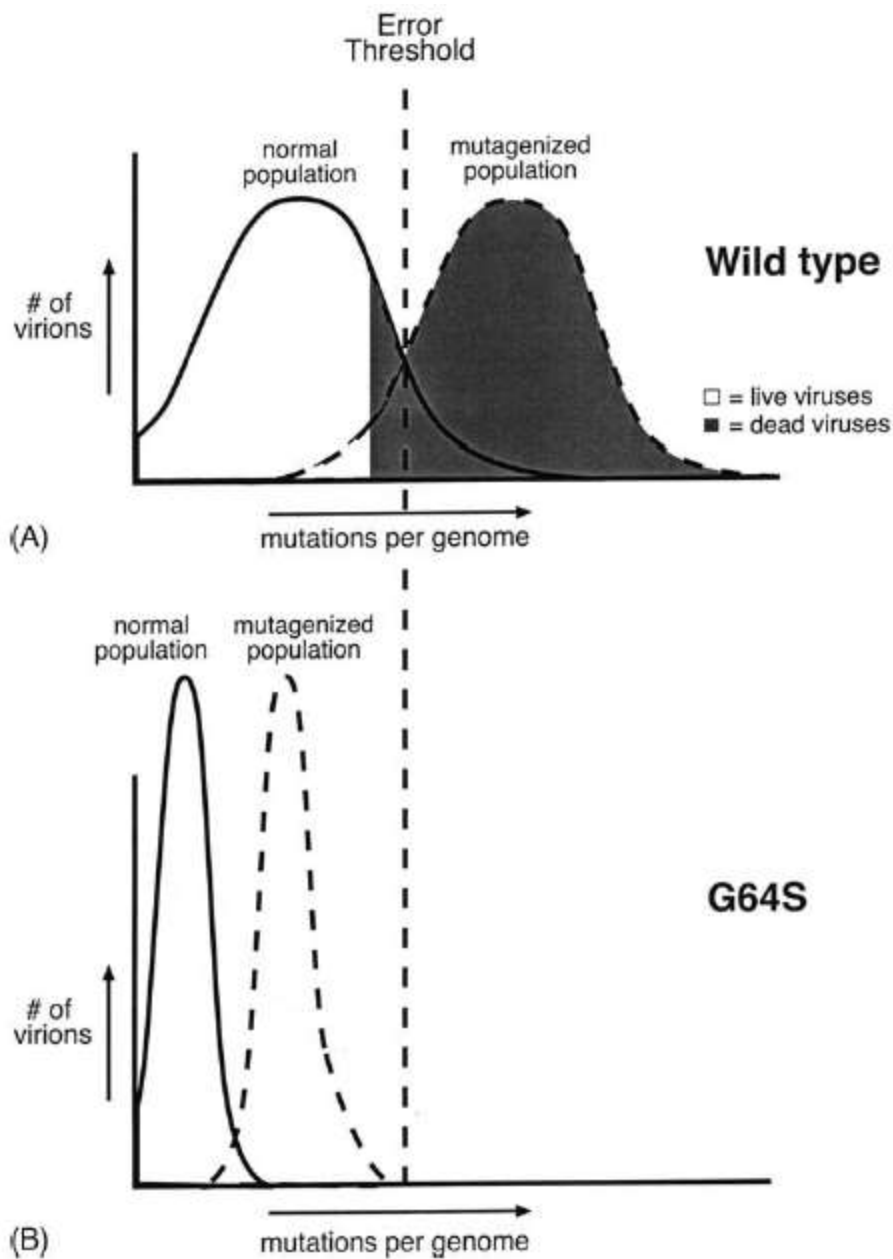


Figure 46 : Distribution d'une population au sein d'une quasi-espèce de PV-G64S et seuil d'erreur (243)

- L'axe des abscisses correspond au nombre de mutation par génome.
- L'axe des ordonnées correspond à la distribution des virions.

⋮ Seuil d'erreur

□ = Virus viables

■ = Virus non viables

Mais la situation semble plus complexe. Le FMDV, appartenant aussi à cette famille, possède des variants présentant une substitution équivalente à G64S, appelée G62S, qui apporte aussi une résistance vis-à-vis de la R. Cet équivalent n'a pas du tout le même impact sur l'ARN-polymérase (85, 219) :

- La reconstitution par mutagénèse dirigée a montré que G62S présente un important désavantage sélectif par rapport au FMDV de type sauvage, qui est partiellement compensé si la substitution d'acide aminé M296I est présente chez l'ARN-polymérase

du FMDV. Lors de passages sur culture cellulaire, on a même pu observer des phénomènes de réversions de la polymérase vers le type sauvage. G62S semble en effet nuire à la fixation et la polymérisation de l'ARN, même si elle limite l'incorporation de la forme triphosphate de l'analogue nucléotidique (85).

- En présence de R, une sélection de variants dans une population de FMDV a eu lieu *in vitro*, présentant la substitution M296I plutôt que celle avec G62S (11, 219). Le mutant M296I présente une forte valeur sélective en présence de la R et a diminué l'incorporation d'analogue dans l'ARN viral, mais sans altération significative de la fidélité de l'enzyme polymérase, et sans autre conséquence sur la complexité du spectre de mutants (11, 85, 177, 184, 243). En l'absence de R, mais aussi en présence de 5-FU ou de 5-AZC, cette sélection n'a pas eu lieu (177).

La sélection de ces substitutions M296I et G62S est donc spécifiquement liée à la présence de R (177). Elles sont toutes les deux localisées sur la polymérase mais de manière assez éloignée. Malgré cela, elles sont connectées par un réseau d'interactions internes pouvant atteindre le site actif de la polymérase. La substitution G62S restreint la flexibilité de la boucle où se situe M296I et ainsi la capacité de l'enzyme à se lier au brin matrice d'ARN. La substitution G62S restreint ainsi l'incorporation de R, mais elle crée un tel désavantage qu'elle ne peut être sélectionnée (85). M296I confère un avantage sélectif en présence de R mais pas en augmentant la fidélité générale de l'ARN-polymérase (177, 219).

On peut retrouver des mutants de FMDV avec un plus haut niveau de résistance à la R. Deux mutations supplémentaires sont apparues successivement chez le FMDV avec M296I, donnant les substitutions d'acides aminés P44S et P169S. La polymérase possédant ce triplet de substitutions est appelée le SSI (4). Son apparition a pour conséquence une baisse modeste de la valeur sélective si la R est absente et une très forte augmentation en sa présence. Tandis que P44S, conjointement à M296I, atténue les conséquences de l'activité de la R en maintenant une balance au sein des quatre types de transitions [ratio de  $(G \rightarrow A) + (C \rightarrow U)$  et de  $(A \rightarrow G) + (U \rightarrow G)$ ], P169S permet l'augmentation de la valeur sélective en présence de R. Cette balance semble prévenir les mutations délétères de la R, ce qui explique l'échappement à l'extinction de ces variants SSI (4). La fidélité de la polymérase ne semble pas affectée, et la complexité du spectre de mutants ne semble pas diminuée.

Enfin, l'accumulation de mutations et le changement de lignées cellulaires contenant les réplicons du HCV *in vitro* permettraient l'apparition de variants résistants à la R (185). Cette résistance serait due à l'import limité de R ou l'export très important dans des cellules comportant des composants viraux résistants à la R. Cela suppose un dérèglement des récepteurs permettant l'incorporation de l'analogue.

Tous ces exemples suggèrent de nombreuses voies pour atteindre une résistance, allant jusqu'à deux substitutions complètement distinctes chez deux virus génétiquement proches (85, 219). Mais seulement quelques-unes sont compatibles avec le maintien d'une valeur adaptative suffisante pour le virus. Ces résultats soulignent d'ailleurs l'intérêt d'utiliser la valeur sélective comme approche générale lors de l'étude des traitements antiviraux.

#### 5.5.5.3.2. Pertinence de l'apparition des résistances

Certaines expériences, comme celles avec le PV, mettent en situation des augmentations graduelles de R dans le milieu (243). Cela pourrait avoir favorisé l'apparition de mutants G64S résistants. Lors de la mise en place de traitements mutagènes à haute concentration sans phase d'augmentation graduelle, il se peut donc que l'échappement viral,

par l'apparition de mutations conférant une résistance, n'ait pas le temps de se mettre en place. L'interprétation des données est ainsi compliquée par l'obtention de ces mutants à différents passages selon les études.

Cependant, ces expériences sont réalisées à partir de population du type sauvage ayant déjà été utilisées dans des expériences de repiquage (243). Connaissant la richesse et la complexité du spectre de mutants qui joue un rôle dans l'apparition de résistances, il est donc normal de penser que la quasi-espèce PV possède dans ce spectre des variants mutants résistants, à basse fréquence. Pour le FMDV, SSI se retrouve bien dans une population n'ayant pas subi de passage avec la R (181). C'est ce qui se rapprocherait le plus du cas d'une infection *in vivo*, et cela suggère que l'apparition de résistance est probable, même au sein d'une population n'ayant jamais été exposée. C'est ce qui est rapporté dans le cadre d'une monothérapie à la R chez des patients infectés par le HCV (266). Cela fait le parallèle avec la sélection de variants mutants résistants aux inhibiteurs chez des patients n'ayant jamais été exposés à ces derniers (118, 198, 211). La R imposerait donc des contraintes sélectives qui sélectionnent les variants résistants, présents dans le spectre de mutants, et augmenterait leur fréquence (181).

Dans la même idée, on peut supposer qu'un historique de mise en contact avec un agent mutagène va favoriser la conservation de génomes résistants à un niveau mémoire (voir partie 5.3.).

#### 5.5.5.4. Effets secondaires des agents mutagènes

Les effets du 5-FU et de la R sur la cellule-hôte et l'organisme doivent être pris en compte lors de leur utilisation (177). Tous les deux ont en effet de nombreuses propriétés. Le 5-FU, par exemple, inhibe les enzymes impliquées dans le métabolisme nucléotidique, peut être incorporé dans l'ADN et l'ARN cellulaire, affectant l'expression de la cellule-hôte (177). Le R a une action inhibitrice vis-à-vis de l'inosine monophosphate déshydrogénase (IMPDH), qui convertit l'inosine monophosphate (IMP) en guanosine triphosphate (GTP), ce qui provoque la diminution de GTP intracellulaire. Cela réduit ainsi la transcription et la réplication virale (243). Elle agit aussi directement sur la réplication virale en inhibant l'activité de l'ARN-polymérase et potentialiserait même la réponse antivirale stimulée par l'IFN (177, 243). Ces multiples propriétés accentuent encore l'activité antivirale de ces analogues et ne s'excluent pas mutuellement.

Ces différents mécanismes peuvent avoir de multiples influences sur l'hôte, ce qui requiert une utilisation et un dosage prudent, qu'ils soient utilisés seuls ou conjointement, et encore plus lors de leur possible utilisation avec des inhibiteurs antiviraux.

#### 5.5.6. **Conclusion**

La résistance aux inhibiteurs antiviraux est un phénomène qui a été décrit de nombreuses fois. Tous les éléments décrits précédemment indiquent que la sélection de virus résistants à des agents mutagènes n'est pas non plus un événement rare. La dynamique des quasi-espèces et la notion de seuil d'erreur jouent ici un rôle dans la création d'un protocole antiviral mais aussi dans l'apparition de résistance à ce dernier. Il reste encore à déterminer à quel point ces résistances peuvent compromettre les traitements antiviraux basés sur ce processus. Encore une fois, les NGS pourront être des outils intéressants afin de planifier des protocoles antiviraux à base d'agents mutagènes.

Il existe cependant d'autres stratégies utilisées comme traitements antiviraux. Leurs apports pourraient permettre de pallier ou retarder l'apparition de résistances lors de l'utilisation d'agents mutagènes.

## **5.6. Mise en place de protocoles combinant mutagénèse létale et d'autres stratégies antivirales**

L'extinction par la mutagénèse létale est favorisée par une faible valeur sélective et une faible charge virale (173). L'intérêt de combiner plusieurs stratégies antivirales serait donc de mettre en place un protocole qui permet cette faible valeur et cette charge au moment de la mise en place du traitement de la mutagénèse létale, via une combinaison de traitements inhibiteur ou mutagène. Il faut aussi que ces protocoles limitent la sélection de mutations permettant l'acquisition de résistances (173).

### **5.6.1. Autres stratégies de lutte antivirale**

#### **5.6.1.1. La thérapie combinée**

Le principe de cette stratégie repose sur l'utilisation de plusieurs médicaments en même temps pour atteindre une charge virale basse. De plus en plus de protocoles thérapeutiques font intervenir une combinaison de différents traitements, notamment pour les inhibiteurs antiviraux.

La thérapie combinée reste souvent utilisée pour la prise en charge des maladies virales dues à des virus connus pour leur grande variabilité, pour les patients atteints par le HIV-1 et HBV par exemple (154, 187). Le HAART (highly active antiretrovirale therapy), exemple bien connu pour le traitement du HIV-1, se base sur une tri-thérapie, mais se heurte toujours à la réserve mémoire du provirus intégré dans l'ADN cellulaire (240).

Cette combinaison diminue ainsi la probabilité de générer un variant viral susceptible d'être résistant ou de sélectionner des variants résistants déjà présents à basse fréquence dans l'organisme. Il faudrait en effet, un nombre de mutations plus important pour que tous les variants puissent échapper à plusieurs antiviraux utilisés en même temps, augmentant ainsi la barrière génétique de résistance (68, 194, 198). Ce raisonnement a été fondé en se basant sur la dynamique des quasi-espèces et confère un avantage certain sur la monothérapie. Cela rejoint l'idée de Ho qui stipulait pour le HIV-1 « Time to hit HIV-1, early and hard » (72). Sans faire de finalisme, la population virale va en effet se développer suite à n'importe quelle opportunité qui lui permettrait de le faire : interruption d'un traitement chronique suite aux effets secondaires, monothérapie, augmentation par pallier de la dose d'antiviraux utilisés. Tous ces éléments vont favoriser l'apparition potentielle de variants résistants. D'un point de vue purement évolutif, il semble donc nécessaire de frapper le plus vite possible, avant que la complexité de la population virale ne soit trop grande et que des variants résistants ne puissent se propager ; et de frapper le plus fort possible. Cela vaudrait pour tous les patients, étant donné les conséquences que l'on connaît.

Cependant cela est tempéré par d'autres points dont il faut absolument tenir compte, comme les effets secondaires sur les patients ou l'existence de certaines phases asymptomatiques chez des personnes infectées à vie.

#### 5.6.1.2. Un traitement en deux étapes

Il s'agit d'une proposition thérapeutique, développée sur la base de considérations théoriques *in silico* (248), qui n'a pas encore été testée expérimentalement. Elle a été élaborée pour répondre au fait que lors de l'échec d'un traitement, l'antiviral contré par les résistances doit être retiré, ce qui permet aux virus résistants de se développer et même de mettre en place une mémoire moléculaire. Cela exclut l'utilisation du composé utilisé pour de futurs traitements.

Elle divise le traitement en deux étapes : une phase d'induction visant à réduire la charge virale et la variété de mutants générés, suivie par un régime de maintien, pour maintenir une charge virale basse et limiter l'apparition de résistances aux antiviraux utilisés lors de la première phase (248). Si le principe reste intéressant, il faut encore déterminer à quel moment et à quelle charge virale critique le régime de maintien doit être mis en place, car il doit être mis en place avant le développement de résistance.

#### 5.6.1.3. L'utilisation de médicaments dirigés contre les fonctions cellulaires

Cette stratégie repose sur le fait que de nombreuses fonctions cellulaires sont essentielles à la réalisation du cycle viral et donc à la multiplication des virus. Viser ainsi les protéines de l'hôte permettrait l'inhibition du cycle. Cela permettrait aussi de limiter les possibilités d'apparition de résistance, car les protéines visées sont cellulaires et présentent donc une variabilité moindre comparée à celle des virus. Ainsi, divers éléments cellulaires peuvent être ciblés, par exemple les complexes enzymatiques cellulaires comme l'hélicase à ARN DDX3 (92) ou la cyclophiline A et son activité peptidyl-propyl isomérase (PPIase) qui jouent un rôle dans la réplication du HCV et sont ciblées par la cyclosporine A (114). Les protéines chaperonnes qui servent à la maturation des protéines de la capsid virale peuvent aussi être inhibées (95). Enfin des récepteurs cellulaires peuvent être la cible d'inhibiteurs, comme les récepteurs tyrosine kinase, qui semblent intervenir dans le cycle répliatif de certains virus (HIV-1, IV, ou le genre *Arenavirus*) (122). Deux problèmes peuvent potentiellement apparaître :

- Une toxicité pour les cellules de l'hôte, car la fonction cellulaire est inhibée. Cela pourrait se manifester par des effets secondaires non négligeables pour le patient.
- La sélection de variants viraux pouvant tout de même utiliser le complexe formé entre le médicament et la protéine, pouvant utiliser un récepteur alternatif à celui qui est inhibé, ou pouvant rendre inefficace indirectement l'inhibition cellulaire, comme cela est décrit par exemple chez le HCV et l'analogue de la cyclosporine SCY-635 (114).

#### 5.6.1.4. Stimulation de l'immunité innée

L'intérêt de stimuler l'immunité innée est que plusieurs branches de la réponse immunitaire sont ainsi mises en jeu. La population virale doit donc faire face à une réponse antivirale multifactorielle qui augmente la barrière génétique et phénotypique de résistance. Cela restreint la sélection de mutants. C'est ce qui est recherché lors de protocoles combinés où intervient l'IFN- $\alpha$  (182, 187).

#### 5.6.1.5. Combinaison d'immunothérapie et de chimiothérapie

La thérapie combinée peut aussi s'appliquer à deux domaines thérapeutiques différents. La diminution de la charge virale pourrait être atteinte suite à l'utilisation conjointe d'anticorps neutralisants ou de la vaccination, qui stimulent la réponse immunitaire acquise, avec des agents antiviraux. C'est aussi un domaine qui nécessite des essais cliniques, comme l'étude utilisant la R couplée à la réponse avec des anticorps neutralisants visant le LCMV, qui a été réalisée chez des souris transgéniques H25 et a apporté des résultats encourageants (212).

#### 5.6.2. **Utilisation de mutagènes pour une thérapie combinée**

Les différentes stratégies virales montrent qu'il y a un avantage certain à utiliser une thérapie combinée, mais on ne sait pas si cela peut être applicable si un agent mutagène rentre en jeu. Des modèles théoriques existent et proposent plusieurs mécanismes possibles pour la mutagenèse létale, mais peu d'études expérimentales et théoriques abordent la thérapie combinée comportant des agents mutagènes ou un inhibiteur antiviral. Elles se font de plus chez peu de virus (HIV-1, FMDV) et avec un nombre limité d'inhibiteurs et de mutagènes (57, 116, 172, 173, 177, 178, 180, 203, 233).

Ainsi, chez le FMDV, malgré le fait qu'il ait permis l'extinction d'une population présentant une valeur sélective basse à modérée, un traitement à base d'inhibiteurs Guanidine Hydrochloride (GU) et d'héparine a provoqué l'apparition de résistances au GU. L'ajout de 5-FU au traitement a permis de l'éviter et d'éteindre des populations de plus grandes valeurs adaptatives (173). De même, une combinaison de 5-FU et de GU a été efficace contre des populations de valeurs sélectives moyennes, mais l'ajout d'un autre inhibiteur, l'héparine, a été nécessaire pour permettre l'extinction lorsque la valeur sélective était augmentée (172). Une combinaison d'inhibiteurs et de mutagènes semblent ainsi présenter des résultats intéressants. L'apparition de résistance à des agents mutagènes (R) a aussi été décrite chez le FMDV. Mais à haute concentration de R, suivie par une combinaison de 5-FU et de GU, l'extinction de la population a été possible (177). Ici, c'est la combinaison de plusieurs mutagènes qui est avancée.

Chez le HIV-1, la combinaison d'une multitude d'inhibiteurs utilisés dans le cadre du « potent AntiRetroviral Therapy » (ART), combinés avec l'hydroxyurée, permet une forte augmentation de la fréquence de mutation, évoquant aussi fortement une combinaison d'antiviraux avec un mutagène (136). De même, une élimination systématique a été atteinte grâce à une combinaison d'azidothymidine (AZT) et de 5-OH-dC, alors que l'utilisation du mutagène seul ne la permettait qu'épisodiquement (233).

Cela suggère plusieurs possibilités:

- Une combinaison d'agents mutagènes et d'inhibiteurs antiviraux permettrait l'extinction rapide d'une population virale, là où l'utilisation d'inhibiteurs seuls ou combinés à d'autres inhibiteurs échouent à provoquer l'extinction ou à empêcher l'apparition de variants présentant des résistances (172, 173).
- Deux agents mutagènes ou plus, proposant différents mécanismes d'action, pourraient agir conjointement contre des virus (177). Le concept serait donc le même que celui utilisé lors de thérapies basées sur l'utilisation de plusieurs agents inhibiteurs.
- Il serait possible d'utiliser des combinaisons d'agents soit de manière simultanée, soit successivement (177).

### 5.6.3. Utilisation de mutagènes de manière successive ou simultanée

La question est de savoir ce qui, de leur utilisation combinée successive ou simultanée, est le plus efficace. L'étude des différentes combinaisons possibles s'est faite pour le FMDV grâce au GU et à la R (178). Il se trouve que c'est l'utilisation successive de GU puis de R qui semble avoir été plus efficace que leur utilisation combinée simultanée. Un modèle théorique dans cette même étude prédit aussi cet avantage de la thérapie successive. Un autre modèle *in silico* décrit la réponse virale face à des doses différentes de R et de GU, afin de déterminer quelle combinaison et quelles doses de traitement sont les plus efficaces et comment éviter au mieux l'apparition de mutations conférant une résistance. Il montre essentiellement que l'avantage d'une thérapie successive sur une thérapie combinée simultanée se manifeste pour une grande concentration d'inhibiteur, et que l'utilisation de l'inhibiteur puis du mutagène est préférable (116).

Deux facteurs peuvent expliquer l'avantage de cette combinaison successive. L'administration simultanée d'un inhibiteur et d'un mutagène peut potentiellement provoquer des comportements non voulus, étant donné que deux mécanismes opposés agissent en même temps (135) :

- Les agents mutagènes vont générer une grande quantité de génomes défectueux, les « defectors ». Leur action d'interférence peut être potentiellement inhibée par la présence d'un inhibiteur qui va limiter la réplication virale et ainsi réduire de manière significative la production de descendants viraux et la diversité au sein de la population (178).
- En plus de la génération de génomes défectueux, le mutagène augmente la diversité au sein de la population et la fréquence de mutation, favorisant ainsi l'apparition de formes résistantes aux inhibiteurs s'ils sont utilisés conjointement (178). L'étude sur le FMDV de Perales (180) vis-à-vis de l'interaction entre R et GU a suggéré qu'une charge virale suffisamment basse pourrait entraver la sélection des mutants viraux résistants aux inhibiteurs. Il faut cependant garder en tête que certains facteurs contribuent à augmenter la probabilité de sélection de mutants résistants, comme une faible barrière génétique et phénotypique, un fort taux de mutation, et la possibilité d'apparition de mutations compensatrices (106).

### 5.6.4. Limites d'application

Cependant certains éléments sont à prendre en compte :

- Il se peut qu'un agent mutagène affiche des activités mutagène et inhibitrice, liées ou non. Par exemple, la R était considérée comme un inhibiteur de la réplication des *Arenavirus*, jusqu'aux expériences chez le LCMV qui ont mis en évidence une activité mutagène uniquement à des concentrations où la R n'exerçait pas d'action inhibitrice (149, 203). On ne sait cependant pas si les activités inhibitrices et mutagènes sont liées pour cette molécule. Il faut donc rester prudent quant à l'interprétation des résultats obtenus avec ces molécules.
- Il se peut aussi qu'un agent inhibiteur ait une activité mutagène, ce qui semble le cas pour l'hydroxyurée, dont l'une des propriétés serait d'augmenter la sensibilité du support de l'information génétique à la mutagenèse (136). Les conséquences sont les mêmes que pour le cas précédent.
- Une multitude de facteurs environnementaux (la charge virale, des goulots d'étranglement, l'hétérogénéité des environnements, la compartimentation de



l'infection, le mode de répllication virale etc.) peuvent affecter les interactions inhibiteur-mutagène (116, 180).

L'action combinée de deux médicaments, qu'ils soient mutagènes ou non, ne peut ainsi être réduite à la somme de leurs effets pris indépendamment. Une multitude de facteurs doit être prise en compte dans leur utilisation, et c'est pour cela que de plus amples essais cliniques, modèles théoriques et expérimentaux doivent être formulés, visant notamment à s'assurer de l'innocuité des molécules utilisées. Ces résultats ont permis d'ouvrir de nouvelles perspectives pour de nouveaux protocoles antiviraux impliquant un agent mutagène.

### **5.6.5. Conclusion**

La prochaine étape est maintenant de confirmer l'application de ces résultats à d'autres modèles viraux, sur culture cellulaire et *in vivo*. L'objectif est, à terme, de trouver de véritables applications médicales, en passant notamment par la mise en place de thérapie combinés. On peut supposer que c'est grâce à la compréhension de la dynamique d'une quasi-espèce (ici, la notion de seuil d'erreur et la dynamique d'apparition de résistances) qu'il sera possible d'appréhender et de construire de nouveaux traitements antiviraux.

## **5.7. Conclusion des implications biologiques**

Tous ces éléments sont autant de caractéristiques apportées par le spectre de mutants d'une quasi-espèce et qui semblent influencer le comportement viral de manière importante, faisant d'elle une population diversifiée et auto-organisée.

### **5.7.1. Application de la dynamique des quasi-espèces virales *in vivo***

Tous les concepts relatifs à la dynamique des quasi-espèces ont été élaborés à l'aide de protocoles utilisant des cultures cellulaires, l'organisme naturel d'un hôte ou des modèles *in silico*. D'autres critères sont nés d'observations de la dynamique des populations virales sur le terrain. Jusqu'à quel point les conclusions de ces études sur l'évolution expérimentale reflètent-elles la dynamique des populations pour des virus dans leur environnement naturel ? Les principaux éléments que nous avons étudiés jusqu'ici vont permettre de faire le bilan sur les caractéristiques de cette dynamique :

- La base de la dynamique des quasi-espèces repose sur une répllication sujette à réaliser des erreurs lors de son déroulement. Ce paramètre est reconnu comme une caractéristique partagée par de nombreux virus à ARN, sur culture cellulaire et *in vivo*, grâce à l'étude des « fidelity mutants » par exemple (27, 184, 245).
- La taille de la population, de celle du génome, le taux de mutation, la fécondité et la nature non neutre des mutations chez un virus à ARN sont des faits reconnus chez la plupart des virus à ARN et notamment ceux qui ont aidé à l'établissement des expériences qui ont été présentées.
- Le spectre de mutant augmente le potentiel adaptatif des virus parce que des variants phénotypiques sont présents dans la population ou peuvent être rapidement générés. La complexité de ce spectre, via son hétérogénéité et sa dynamique, peut influencer l'émergence d'une infection virale et l'apparition de résistance aux traitements. Cette capacité d'adaptation conférée par le spectre de mutant a été observée *in vivo* et validé par de nombreuses observations expérimentales pour le FMDV, le VSV et le PV par exemple (37, 65, 67, 70, 94,

162, 184). Toutes ces caractéristiques sont cohérentes avec le comportement des virus étudiés et expliquent notamment les autres points.

- La mémoire moléculaire a été décrite chez le FMDV (201) et le HIV-1 *in vivo* (31). Elle pourrait expliquer de nombreuses réémergences de séquences dites ancestrales.
- Les génomes qui composent le spectre de mutants peuvent se comporter en fonction d'autres génomes. Que ce soit les interactions positives ou négatives, elles peuvent moduler le comportement de l'ensemble. Des grandes différences de valeur adaptative peuvent être la conséquence d'altérations dans la composition du spectre sans modification de la séquence consensus, incluant des événements drastiques comme la transition vers l'extinction par mutagenèse létale (96, 97, 103).
- Des modèles tels que celui de la mutagenèse létale, basés sur la théorie des quasi-espèces et la notion de seuil d'erreur, semblent être applicable en matière de traitement antiviral, avec un avantage, démontré pour le FMDV, des thérapies combinées successives. Des modèles analogues ainsi que des paramètres adaptés à une autre stratégie de réplication seront nécessaires pour étendre ces hypothèses au virus à ADN.

En regard de tous les résultats expérimentaux obtenus et des modèles théoriques construits, on peut avancer qu'il n'y pas de contradictions fondamentales entre le comportement des virus tel qu'il est décrit dans les expériences et celui retrouvé dans les infections naturelles. De plus, la théorie des quasi-espèces et son application aux populations virales n'entrent pas en contradiction avec une dynamique des populations dite classique (259). Elle semble même expliquer des comportements viraux que la théorie synthétique de l'évolution seule ne permet pas de comprendre.

Il faut cependant garder en tête que les modèles exprimés sont toujours en cours d'évaluation, et que les observations expérimentales sont restreintes à un ou quelques virus. Il serait bien imprudent de tirer une généralisation des observations retrouvées chez une seule population. La recherche est en constante évolution afin de savoir si telle donnée s'applique bien à telle population, et le travail est loin d'être terminé.

### **5.7.2. Application de la dynamique des quasi-espèces sur l'évolution sur le long terme**

Ces éléments rendent le comportement des populations virales encore plus complexe : il n'existe pas de relations simples et prédictibles entre les différentes propriétés qui ont été ici mises en évidence. De plus, ces propriétés prouvent bien que le comportement d'une population ne peut être prédit en étudiant chaque virion qui la compose (ce qui est de toute façon très difficile).

L'évolution des virus sur le long terme et *in vivo*, étudiée en épidémiologie, tient compte de variables écologiques, sociales et environnementales qui interviennent toutes pour expliquer l'infection d'un nouvel hôte, une émergence ou une réémergence virale. Ces variables sont fortement complexes. La nature même des quasi-espèces virales, si elle est bien cohérente, introduit donc une incertitude supplémentaire quand on se penche sur l'évolution d'une population virale au niveau épidémiologique. Par exemple, il est difficile de déterminer quelle sous-population de la quasi-espèce va pouvoir être transmise à de nouveaux hôtes. Cela rend l'émergence d'une maladie virale imprévisible, malgré une surveillance épidémiologique

accrue. Etudier l'incorporation de ces éléments dans la dynamique des quasi-espèces virales pourrait faire l'objet d'un autre travail.

Enfin, l'évolution des virus à ARN sur le terrain était auparavant vue comme une horloge moléculaire, c'est-à-dire une théorie neutre de l'évolution (72). Cela implique que les populations évoluent selon une suite d'accumulation linéaire de mutations, proportionnellement au temps géologique. Ces mutations sont reflétées par le taux d'évolution. Cependant l'étude de la dynamique des quasi-espèces et ses applications *in vivo* montrent que ces populations virales évoluent suite à un déséquilibre, ce qui a pour conséquence un renouvellement de la distribution des variants. De plus, il a été vu que le taux d'évolution était très variable, avec des phases de stases évolutives alternées avec des phases d'évolution rapide. L'évolution chez les virus devrait donc plutôt être vue comme un remplacement continu d'une sous-population par une autre.

## Partie 6 Limites et extensions de la théorie des quasi-espèces

La théorie est loin de faire l'unanimité. Nous allons mettre en évidence certaines limites dues à ses paramètres et ses caractéristiques. Il semble en effet important de nuancer certains propos tenus jusque-là concernant le comportement déterministe, les différents protocoles méthodologiques, et le paramètre central qu'est la valeur sélective. Si certains chercheurs réfutent même complètement l'application du modèle mathématique aux espèces virales à ARN, d'autres voient en lui une possible extension à d'autres entités biologiques.

### 6.1. Une théorie déterministe

La théorie des quasi-espèces, comme étant formulée initialement, est une théorie déterministe. Des valeurs numériques sont ainsi assignées à un certain nombre de paramètres et le système décrit par l'équation qui relate ces différents paramètres est prédictible. Les modèles théoriques déterministes sont élaborées habituellement en premier car ils peuvent être résolus mathématiquement.

#### 6.1.1. L'intervention d'éléments stochastiques

En considérant son application à la dynamique des populations virales, certaines incohérences peuvent être mises en évidence. En effet, il est évident que les virus ainsi que les réplicons primitifs, que la théorie des quasi-espèces veut représenter, ne constituent pas des populations de taille infinie, en équilibre constant quand ils se répliquent chez leurs hôtes. Un modèle déterministe ne correspondrait donc pas à un système réaliste pour ces entités, comme pour toute unité répliquative.

De même, la mutagénèse résulte de fluctuations au niveau atomique : par exemple, ce sont des fluctuations stochastiques qui vont déterminer la présence d'une forme tautomérique (des couples d'isomères de constitution inter-convertibles) ou d'une autre pour un nucléotide, à un instant  $t$  lorsque ce dernier atteint le site catalytique du complexe d'élongation de la polymérase. De même, la conformation précise d'une protéine au niveau de son site catalytique et des résidus l'entourant dépendent aussi de ces fluctuations (63). Elles interviennent aussi dans le processus de compétition : celui-ci est soumis à des fluctuations statistiques, en fonction de la position spatio-temporelle des individus en compétition, ce qui peut altérer les résultats prévus lors d'une théorie déterministe.

Lors de la sélection, un mutant nouvellement apparu et démontrant un avantage sélectif ne va pas nécessairement surpasser les autres dans une population virale, car il dépend de la composition du spectre de mutants. Cela rend la prédiction de l'évolution d'une sous-population encore plus difficile, même si elle semble présenter un avantage sélectif sur le papier.

Il semble difficile de se reposer sur un modèle déterministe excluant des éléments stochastiques. Certains exemples *in vitro* permettent malgré tout de nuancer l'inadaptation d'un modèle déterministe.

### **6.1.2. Des exemples isolés de comportements déterministes au sein des populations virales**

Certaines expériences rapportent des comportements évolutifs qui orientent fortement vers le déterminisme.

L'apparition du déclin de la mémoire moléculaire semble décrire un comportement déterministe étant donné la cinétique identique que l'on retrouve sur différentes lignées, même si ce phénomène se retrouve chez des populations finies hors équilibre (31, 204). De tels cas se retrouvent aussi dans des études concernant d'autres caractéristiques, comme le déclin de la valeur sélective ou dans la sensibilité à la température (83).

Lors d'une étude compétitive entre le VSV de type sauvage et un variant présentant une mutation neutre (189), il a été observé qu'à partir d'un mélange initial de ces deux variants, il a été possible de reproduire une même dynamique de compétition entre eux, pendant une certaine période de temps, presque constante, jusqu'à l'atteinte d'un « point critique ». Au-delà de ce point, les trajectoires évolutives finissaient par diverger. Ainsi, malgré le caractère stochastique des apparitions des mutations, le comportement était hautement reproductible. Des perturbations environnementales introduites pendant ces compétitions ont de plus favorisé la dominance du type sauvage (190). Une hypothèse pour expliquer ce comportement est que l'apparition de ces mutations neutres (qui ont d'ailleurs modifié la séquence consensus) a déplacé la population virale dans une autre portion de l'espace séquence, où elle présente une valeur adaptative identique au type sauvage dans des conditions de croissance normales. Mais en cas d'événements sélectifs, la population s'adapterait moins rapidement que dans le cas du type sauvage. C'est donc cette neutralité relative qui expliquerait le comportement déterministe.

### **6.1.3. Nuances du comportement déterministe**

Une population virale peut avoir un comportement proche du déterministe si elle se tend vers une grande taille et un quasi-équilibre :

- Nous avons vu qu'il pouvait y avoir alternance entre des périodes de réplication de larges populations et des épisodes de diminution drastique de celles-ci, notamment lors de goulots d'étranglement. Lors de périodes de larges populations, on peut retrouver un comportement déterministe, mais c'est cette alternance entre larges et faibles populations qui accorde aux virus leur caractère imprédictible, donc qui échappe au déterminisme.
- Une autre question est de savoir si la formation des mutants et leur orientation vers un état d'équilibre, même fugace, ont lieu plus rapidement que les changements de l'environnement auquel la distribution de mutant est confrontée (72). Il est probable de retrouver ces conditions pour des réplicons à ARN avec un fort taux de mutation. La modification du spectre de mutants peut alors être approchée comme une succession de très courtes étapes en équilibre, chacune d'entre elles suivant les conditions d'une théorie déterministe. Une représentation graphique présentée sur la figure 47 permet de mieux comprendre cette approche.

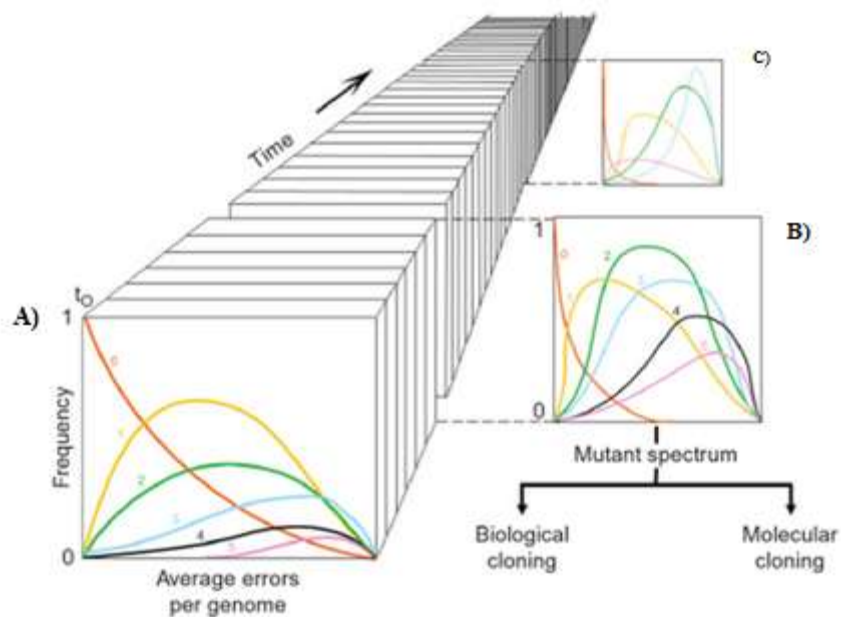


Figure 47 : Représentation de l'évolution temporelle des quasi-espèces virales (72)

- L'échelle de temps est représentée par un ensemble de « tranches » de temps, dans laquelle on retrouve une distribution de mutants à chaque fois différente.
- La distribution est exprimée par la fréquence de génomes avec un nombre défini de mutations en ordonnée, en fonction du nombre d'erreur moyen par génome en abscisse.

→ Echelle de temps

- Distribution de génomes avec 0 mutation
- Distribution de génomes avec 1 mutation
- Distribution de génomes avec 2 mutations
- Distribution de génomes avec 3 mutations
- Distribution de génomes avec 4 mutations
- Distribution de génomes avec 5 mutations

- A) Distribution du spectre de mutants à l'instant  $t_0$
- B) Distribution du spectre de mutants à l'instant  $t$
- C) Distribution du spectre de mutants à l'instant  $t+\Delta t$

Comme le temps s'écoule, les différentes catégories de génomes varient en fréquence, et le processus de changement peut être considéré comme une succession d'étapes en équilibre. Les clonages biologique et moléculaire fournissent un cliché à un instant  $t$  de la distribution de mutants, qui est en réalité en constant changement. Cela démontre aussi les limites des techniques utilisées pour caractériser ces populations.

#### 6.1.4. Une démarche dite classique

Il est décrit qu'il s'agit d'une procédure normale de partir d'un modèle déterministe, plus facile à démontrer, pour ensuite y incorporer des événements stochastiques, ce qui permet d'appréhender vraiment l'évolution darwinienne appliquée à des populations virales à ARN. La même question s'était posée lors de l'intégration de la Loi de Hardy Weinberg dans la génétique des populations : elle est établie avec des conditions extrêmement larges et décriées, non applicables si l'on veut que ces modèles reflètent les conditions réelles pour les populations diploïdes (260). Mais on peut malheureusement dire la même chose de tous les

modèles théoriques de la sélection, de théories physiques, chimiques ou biologiques (76). Chaque modèle impose en effet des conditions spécifiques pour établir et comprendre l'objet d'étude, et celles-ci sont nécessaires pour apporter des données interprétables.

Comme cela a été fait pour la génétique des populations, Eigen et de nombreux autres auteurs ont donc étendu la théorie, et, dans notre contexte, à des populations finies de réplicons hors équilibre (78, 165, 174, 206, 230, 255). Ces modèles, s'ils permettent en effet d'étendre la théorie des quasi-espèces, sont cependant très simplifiés et ne correspondent pas à ce que l'on s'attend à retrouver *in vivo* selon certains auteurs (113).

### **6.1.5. Conclusion**

Les mécanismes moléculaires de ce comportement déterministe pour des populations finies de virus à ARN restent très difficiles à décrire. Il faudrait déterminer quels paramètres, intervenant sur la complexité et la dynamique des populations virales (taille de la population, hétérogénéité...), vont favoriser des caractères déterministes ou stochastiques (126, 199). On sait déjà que la taille de la population étudiée joue un rôle prépondérant, mais on a vu que sa relation avec les autres paramètres est complexe. Sur ce facteur, aucune publication scientifique ne semble apporter de données chiffrées d'une taille critique, qui définirait la transition d'un comportement déterministe à un comportement stochastique (donc un passage de la sélection à la dérive génétique).

Il s'agit donc d'une procédure standard d'établir un modèle déterministe. Cependant ce n'est pas parce que ces méthodes sont utilisées dans le monde de la recherche qu'il ne faut pas les remettre en question. C'est la constante recherche de la validité d'une hypothèse qui permet à la recherche d'avancer et d'élaborer des modèles, comme ceux des quasi-espèces, disponibles aujourd'hui.

## **6.2. Limite des méthodes de séquençage**

Les nombreuses étapes nécessaires à la réalisation de l'analyse des séquences au sein des populations apportent autant de risque de fausser les résultats et leurs interprétations. Les différentes techniques présentent ainsi des erreurs, des biais d'utilisation pour certains paramètres (voir partie 3.4.1.), ayant un impact sur l'analyse des populations. Même le choix des protocoles d'extraction des génomes viraux et d'élimination des possibles ARN et ADN contaminants est important (20). Sachant cela, des modèles correctifs ont été progressivement élaborés.

### **6.2.1. Limites d'échantillonnage des clonages biologique et moléculaire**

#### **6.2.1.1. Echantillons parfois peu représentatifs et erreurs de copies**

Le clonage biologique oriente l'information en faveur des clones infectieux dans les cellules, car on prélève les plages de lyse visibles qui sont produites uniquement par ces clones. Les génomes issus des particules non infectieuses sont donc exclus de l'analyse, sauf si la complémentarité intervient au sein de la plage en croissance, ce que l'on ne peut pas exclure. Avec l'avancée des NGS, il a été démontré que ces génomes non infectieux

participaient pleinement à la complexité et à la diversité des populations virales, il faut donc tenir compte de ce biais lors de l'utilisation de clones biologiques (72).

Le clonage moléculaire peut quant à lui examiner la composition des préparations virales indépendamment de leur caractère infectieux dans un organisme infecté ou en culture cellulaire. Un biais est cependant introduit à cause des erreurs introduites lors de la copie de la matrice par les enzymes utilisées lors de l'amplification *in vitro* (72). Ce problème peut être résolu en utilisant des polymérases avec un taux d'erreur moins important, par des conditions d'amplification correctes, et par des expériences de contrôle pour déterminer le taux d'erreur basal inhérent au protocole d'amplification *in vitro*.

#### 6.2.1.2. Echantillonnage faible

Une limite importante et commune aux deux méthodes couplées à un séquençage de Sanger touche l'échantillonnage : celui-ci peut fournir environ 100 particules virales, ce qui représente une infime fraction du nombre total de particules infectieuses (de  $10^{10}$  à  $10^{12}$ ) ou infectieuses et non infectieuses (de  $10^{11}$  à  $10^{16}$ ) chez un hôte infecté ou en culture cellulaire. Cela représenterait 0.000001% à 0.0000000001% de la population. Le séquençage haut débit permettrait d'augmenter cette proportion entre 0.0001% à 0.000000001%, une fraction encore minuscule (72).

Cependant, l'utilisation du clonage moléculaire et biologique et du séquençage de Sanger pour analyser la grande majorité des spectres mutants semble avoir apporté des résultats convaincants, si l'on considère que valider des concepts sur la complexité des populations virales est suffisant (72). De plus, les résultats obtenus par ces méthodes semblent avoir été confirmés par les NGS. Il semblerait donc que l'échantillonnage ne soit pas une limite, si l'on formule l'hypothèse que 100 génomes échantillonnés ne sont pas une représentation biaisée de la hiérarchisation au sein de la population virale, dans le sens où aucun génome minoritaire n'est surreprésenté, et aucun génome dominant n'est sous-représenté.

#### 6.2.2. **Limites de l'amplification et du séquençage**

Avant l'amplification des séquences à analyser, une rétro-transcription est réalisée chez les virus à ARN afin d'obtenir un ADN complémentaire (ADNc), seul support possible de l'amplification. Cette étape constitue donc un désavantage pour le séquençage pour plusieurs raisons (20) :

- Les enzymes utilisées (les transcriptases inverses) sont sujettes à des erreurs, ne possédant pas d'activité de relecture. Ces erreurs sont difficilement discernables des véritables mutations des variants viraux.
- Des ADNc de courte taille et incomplets peuvent se dissocier de la matrice d'ARN et agir en tant qu'amorces lors de la PCR, provoquant ainsi une recombinaison *in vitro*. Cela peut se produire lors de RT-PCR visant à amplifier des fragments de plusieurs kb. Cependant l'utilisation d'une enzyme appelée « RNase H-negative RT » pourrait permettre de minimiser ce phénomène.

Il est décrit que les procédures d'amplification par PCR biaiserait la représentativité des séquences obtenues (20). La PCR peut introduire des mutations ponctuelles, des indels, à cause des erreurs de réplication dues au manque de précision des ADN-polymérases. Un mauvais appariement des amorces peut aussi avoir lieu. Une recombinaison *in vitro*, comme



décrite précédemment, peut aussi provoquer l'apparition de séquences recombinantes ou chimériques. Un phénomène de « template switching » peut aussi produire ces ADNc chimériques artificiels (133). Le gros problème de ces erreurs est que leur mécanisme d'apparition *in vitro* est le même que celui touche les quasi-espèces *in vivo*. Il est donc très difficile de séparer les vraies mutations de celles induites par la PCR.

Selon les technologies de séquençage utilisées, on peut aussi retrouver des erreurs de séquences (115). Celles majoritairement détectées sont des insertions et des délétions dues aux régions homopolymères (des répétitions identiques de la même base) lors du pyroséquençage. Pour le séquençage à l'aide de terminateurs réversibles, il s'agit plutôt d'erreurs de séquençage aux extrémités des séquences et des substitutions d'une base par une autre lors de la lecture. D'autres erreurs sont provoquées par un phénomène, le « Carry Forward/Incomplete Extension » (138). Il est dû soit à la persistance de nucléotides dans les puits des plaques utilisées pour le pyroséquençage, malgré leur rinçage, soit à un défaut d'activité de la polymérase, qui réalise alors un séquençage incomplet. Certaines technologies, comme le séquençage par ligation, incluent cependant un système de correction des erreurs d'incorporation, ce qui les rend particulièrement fiables.

Les plateformes de séquençages sont elles-mêmes une source intrinsèque de variabilité, avec un impact important sur la reproductibilité, la précision, le taux d'erreur, la spécificité et la sensibilité. Ces nouvelles technologies, malgré l'engouement qu'elles suscitent, doivent encore faire face à beaucoup de limitations dans leur utilisation des études des quasi-espèces : problèmes techniques, données en quantité importante et parfois difficilement traitables, recours à la bio-informatique et donc intervention d'un intermédiaire supplémentaire pour l'analyse et les gestions des données, et mise en évidence délicate des erreurs potentielles (115). Ces limites orientent de plus en plus vers un système de « cloud computing », qui correspond à un réseau informatique dématérialisé permettant d'avoir accès aux données à distance et de les traiter.

### **6.2.3. Corrections des limites**

#### **6.2.3.1. Corrections des erreurs dues à l'amplification**

Une étude a démontré que le séquençage direct de l'ARN sans passage par un ADNc était possible via l'utilisation de polyoligoribonucléotides fixant l'ARN à séquencer (170). Cette approche permettrait de s'affranchir des erreurs engendrées par la synthèse d'ADNc, et de séquencer tous types d'ARN, quelle que soit leur taille.

Le mauvais appariement des amorces peut être évité en utilisant des amorces dégénérées ou en utilisant des sites de liaisons de l'amorce correspondant à des régions conservées du génome viral. Les recombinaisons et la formation de chimères peuvent être diminuées en augmentant le temps d'élongation, diminuant le nombre de cycles et en enlevant l'étape d'extension finale. Cela revient à affiner et améliorer l'étape de PCR par définition. L'utilisation d'ADN polymérases avec une forte fidélité de réplication peut permettre de diminuer le taux d'erreurs (20).

Même en appliquant toutes ces précautions, il est impossible d'éliminer toutes les erreurs dues à la PCR et de les différencier des erreurs retrouvées réellement chez les variants. En partant du principe que l'apparition des erreurs d'amplification est aléatoire, il est possible de réaliser plusieurs PCR en parallèle pour les comparer. L'utilisation d'identificateurs d'amorces (« primer identifiants » (IDs)) permet d'étiqueter chaque molécule d'ADNc mais possède aussi des limites (117).

L'utilisation d'ORP (overlapping reading pairs) réduirait significativement le taux d'erreur de séquençage mais permettrait aussi d'augmenter la précision de la détection de variants rares dans une population (des mutations avec une fréquence inférieure à 0.05%) (43).

### 6.2.3.2. Circular Sequencing

Une autre solution a été proposée pour différencier les variations génétiques véritables des erreurs de séquençage : diminuer très fortement les erreurs dues au séquençage (1). Cette méthode appelée Circular Sequencing (ou CirSeq) repose sur le fait de créer des ADNc répétées et en tandem, en utilisant des fragments d'ARN circularisés en continu comme matrice (figure 48). De courts fragments de génomes d'environ 90b sont circularisés pour servir de matrice à une synthèse d'ADNc par cercle roulant. Les répétitions en tandem obtenues sont ensuite séquencées en une seule lecture. Les répétitions peuvent ainsi être identifiées et alignées. La véritable mutation apparaît sur chaque répétition parce qu'elle est présente sur la matrice. Les erreurs de séquençage ont très peu de chance d'apparaître sur 3 répétitions à la fois, et peuvent donc être identifiées et exclues de la séquence consensus.

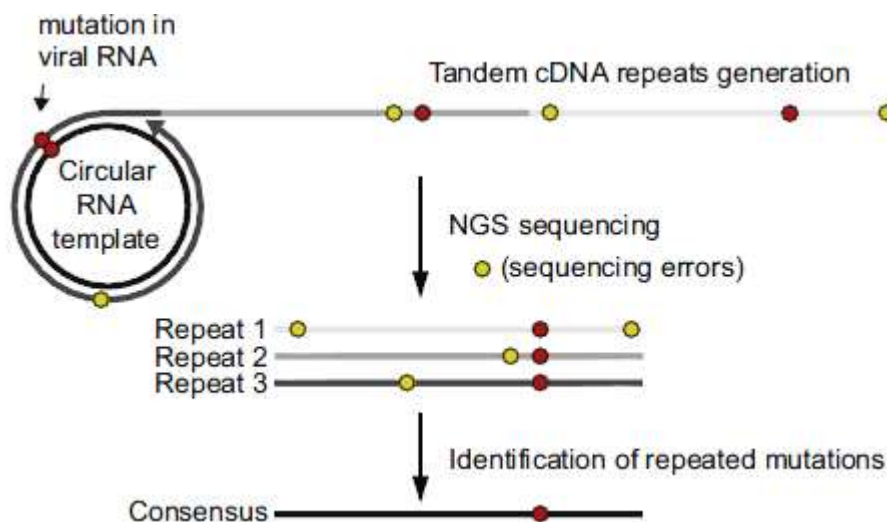
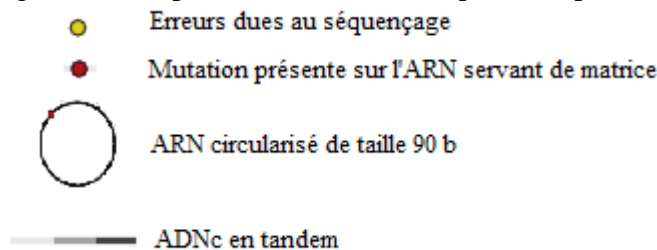


Figure 48 : Représentation schématique de la procédure CirSeq (6)



Cette méthode permet une détection rapide et précise des erreurs d'amplification et de séquençage, par exemple une diminution de la probabilité théorique d'erreurs réalisées avec une plateforme de séquençage Illumina d'un ordre de grandeur de 8 (de  $10^{-3}$  à  $10^{-11}$  par base). Les implications de telles méthodes sont importantes : elles permettraient d'estimer des changements de fréquence de mutation à une échelle très petite, voire de définir des paysages adaptatifs à la résolution d'un simple nucléotide.

### 6.2.3.3. Méthodes de correction

Plusieurs autres méthodes sont disponibles afin de corriger les données apportées par l'UDS (186). Cela implique plusieurs étapes : le filtrage, l'alignement, la correction des erreurs de lecture, l'inférence des haplotypes et l'estimation des fréquences. Ce ne sont pas des étapes indépendantes et certaines méthodes utilisées englobent une ou plusieurs de celles-ci (19, 20).

Le filtrage et l'alignement consistent à retirer les lectures de basse qualité de la base de données. Ces étapes peuvent être réalisées grâce des algorithmes ou des analyses cartographiques et informatiques, en visant notamment certaines régions des séquences particulièrement touchées par les méthodes utilisées (19). L'alignement permet notamment d'identifier les indels qui peuvent provoquer un décalage dans le cadre de lecture (20).

La correction des erreurs techniques de séquençage, notamment dues à l'utilisation de la PCR, se base sur le fait que leurs apparitions sont aléatoires et rares (comme c'est le cas pour la méthode CirSeq). Ce n'est pas le cas des véritables mutations présentes dans la population, normalement échantillonnées proportionnellement à leurs fréquences dans la population. Différents algorithmes de regroupement permettent de réduire le taux de ces erreurs (19, 20).

L'inférence locale d'haplotype permet aussi de corriger certaines erreurs. La reconstruction de type globale des haplotypes, qui est aussi utilisée, est cependant plus compliquée. Des graphiques de lecture combinés à des algorithmes de combinaisons et à des méthodes d'assemblage *de novo* peuvent être utilisés pour proposer un sous-ensemble, le plus simple possible, d'haplotypes qui peut expliquer toutes les lectures qui sont observées (19, 20, 269).

Certaines de ces méthodes ont été implantées dans des ensembles de logiciels (Annexe 9) et sont testées sur des données de contrôle expérimentales et de simulation (19). De nouvelles méthodes informatiques (241), spécifiques à chaque plate-forme de NGS, sont continuellement apportées afin d'augmenter la précision du séquençage et de diminuer les erreurs inhérentes aux protocoles utilisées (2, 241, 269, 270).

Les scientifiques ont de plus en plus de recul sur cette technique de séquençage. Plusieurs éléments sont de mieux en mieux maîtrisés, comme la correction des erreurs de séquençage, et les différents modèles statistiques à appliquer. Cependant il faut rester prudent quant à l'interprétation des résultats et toujours reconsidérer ces différentes limites.

### 6.2.4. **Conclusion**

Il faut toujours avoir en tête que la représentativité de l'échantillonnage est inévitablement conditionnée par les filtres méthodologiques. Quelle que soit la méthode de séquençage utilisée, elle aura des limites et des biais qu'il convient d'analyser, de compenser ou de corriger. L'important est de les avoir déterminées avec précision et de les garder lors de l'interprétation des résultats. Ces problèmes s'appliquent à tous les domaines de la recherche, pas seulement lorsqu'il s'agit de l'établissement de la théorie des quasi-espèces.

### **6.3. Limites des mesures de la valeur sélective**

#### **6.3.1. Des limites dues aux interactions d'une multitude de facteurs**

Il existe plusieurs limites à l'utilisation de la valeur sélective : elle dépend de l'environnement dans laquelle elle est mesurée, et un hôte complet est l'équivalent d'un environnement très complexe, hétérogène et compartimenté. Ainsi, lors du processus d'infection, le virus est soumis inévitablement à tout un éventail de différentes conditions. Un virus peut ainsi avoir une grande valeur adaptative dans un type de cellule et une très faible dans un autre type. Ce concept de valeur sélective cellule-dépendante a été utilisé pour démontrer que la perte de valeur sélective est un facteur décisif de l'extinction virale lors de sa transition au-delà du seuil d'erreur, par exemple chez le FMDV (172). La hiérarchisation des valeurs sélectives que l'on peut ainsi faire en culture cellulaire ou *ex vivo* peut donc ne pas du tout refléter les différences de valeurs que l'on retrouve chez l'hôte naturel du virus, comme cela a été démontré pour le HIV-1 *in vivo* (192).

Dans le cas des quasi-espèces virales, le terme d'environnement et son influence doivent être pris dans son sens le plus large, avec de nombreux facteurs participatifs. Les paramètres physiques et les paramètres biologiques des cellules font partie de cet environnement. Si deux virus en compétition montrent des stabilités thermiques différentes, parce qu'ils diffèrent de plusieurs acides aminés sur la capsidie ou sur les protéines de surface, cela signifie que la température affecte la valeur sélective. L'effet va être dépendant du temps passé dans le milieu extracellulaire pour le virion, par rapport au temps passé dans l'environnement intracellulaire, avant la sortie de la cellule. Le temps dans l'environnement extracellulaire doit donc être inclus comme une étape additionnelle du cycle de réplication qui peut affecter la valeur adaptative.

Les influences agissant sur la valeur adaptative s'étendent jusqu'aux caractéristiques du virus lui-même, étant donné que le spectre de mutant d'un virus doit être considéré comme une partie de l'environnement ; sa composition (soumise aux indéterminations dérivées de l'apparition au hasard des mutations et des recombinaisons) peut moduler le comportement de la quasi-espèce entière. Une population virale peut modifier sa valeur adaptative sans altération de la séquence consensus.

#### **6.3.2. Limites retrouvées pour les protocoles expérimentaux**

Ces limites s'appliquent même aux expériences réalisées en culture cellulaire, car certains paramètres clés peuvent modifier les valeurs adaptatives. Si des cellules sont infectées à faible MOI, les deux virus utilisés pour mesurer la valeur sélective peuvent se répliquer dans des cellules différentes. Toutefois, si le « rendement » des deux virus est suffisamment haut et que l'on n'interrompt pas l'infection, un autre cycle de réplication peut avoir lieu avant que l'on transfère les deux compétiteurs dans un nouveau milieu, ce qui va fausser le résultat de la valeur sélective. On se retrouve ainsi avec deux vagues de réplication : l'une sans compétition intracellulaire et l'autre avec compétition pour les ressources de la cellule. La situation se complique encore plus si le rendement d'un virus est plus haut que l'autre lors de la première vague, et si l'importance de la compétition intracellulaire de la deuxième vague affecte la capacité de réplication relative des deux virus. De plus, la compétition intracellulaire peut faire intervenir l'interférence ou la complémentarité entre les nuages de variants produits par les deux virus (179, 258).

Des tests alternatifs peuvent être cependant mis en place pour minimiser les phases répliquatives :

- On peut utiliser des vecteurs viraux et des réplicons sous-génomiques qui limitent la réplication à un seul cycle, permettant d'éviter les secondes vagues de réplication.
- On peut aussi mener des expériences d'infections en parallèle, chaque virus pris indépendamment, sans réaliser de compétition intracellulaire entre eux. Le titrage viral à des instants  $t$  après le début de l'infection remplace ainsi la valeur sélective. L'équivalence de ces deux procédures a été démontrée chez le FMDV, avec un résultat similaire obtenu à chaque fois (81).
- Enfin, on peut aussi limiter la durée d'infection afin d'empêcher la phase de forte multiplication.

Dans tous les cas, toutes ces alternatives posent de nombreuses difficultés techniques qui alourdissent le travail. Comme on l'a vu précédemment, il faut donc définir précisément les conditions d'expérimentation. Mais ce point s'applique en fait à tout protocole expérimental *in vitro* et *in vivo*. La rigueur d'analyse et de mise en place d'un protocole est à la base de tout raisonnement scientifique.

### **6.3.3. Difficultés de l'application de la valeur sélective *in vivo***

Les limites de détermination de la valeur sélective sont accentuées dans les essais réalisés *in vivo* à cause des multiples étapes et changements d'environnement lors de la progression de l'infection virale dans un organisme vivant (animal ou végétal). La dérive génétique joue en effet un rôle important en modifiant la composition de la population virale par rapport à la composition qu'elle aurait si la sélection était la seule force évolutive.

Des expériences sur le FMDV ont permis une mesure expérimentale de valeur sélective *in vivo* et ont démontré qu'il existait des variations de ces valeurs entre les différents composants du spectre de mutants d'une population se répliquant chez un porc (36). Ces variations différaient même entre les lésions aphteuses retrouvées chez un même individu. Cela démontrait que chaque lésion était caractérisée par un spectre de mutants de composition différente, insistant ainsi sur l'impact du hasard et de la dérive génétique dans ces études.

Des résultats similaires ont été trouvés chez le HCV (182, 216) (voir partie 5.1.3.5). Il a en effet été possible d'obtenir des populations de HCV de la même lignée mais avec des valeurs sélectives différentes. La plus haute valeur correspondait à des variants HCV avec une résistance significative à de nombreux inhibiteurs anti-HCV. Le fait que la valeur sélective puisse être liée à la réponse à un traitement *in vivo*, rend pertinent le fait de chercher des moyens de mesurer ou d'estimer la valeur sélective du HCV chez des patients infectés, car cela pourra apporter des informations pour les traitements à mettre en place. La découverte de possibles paramètres mesurables *in vivo*, qui permettraient d'évaluer la valeur sélective, seraient plus que bienvenue, à moins qu'une méthode efficace pour cultiver des isolats de HCV *ex vivo* ne soit disponible d'ici-là.

#### **6.3.4. Conclusion**

Malgré les nombreuses limites et difficultés liées à ce paramètre, la valeur adaptative est un pilier de la recherche pour déterminer la dynamique des populations virales en culture cellulaire, *in vivo* et *ex vivo*. Ces essais quantitatifs permettent maintenant d'étudier la connexion entre cette valeur et la charge virale, le coût de valeur sélective de mutations accordant des résistances aux médicaments, l'identification de mutations compensatrices afin de gagner en capacité adaptative.

#### **6.4. Remise en cause de la théorie**

Tout au long de l'exposition des différentes propriétés de la théorie des quasi-espèces et de son application aux virus à ARN, nos propos ont tenté, de manière générale, d'être nuancés et objectifs. Nous abordons ici des éléments de discussion plus généraux.

Certains auteurs réfutent la tendance à ériger la théorie des quasi-espèces en une règle universelle, généralisant ainsi le modèle à toutes les espèces à ARN, voire à l'étendre à d'autres systèmes biologiques. Si la plupart acceptent le modèle mathématique, leur principal argument est le manque de données expérimentales qui peuvent appuyer les modèles théoriques articulés autour de cette théorie (113).

##### **6.4.1. Mauvaise utilisation du terme**

Ceux qui remettent en question cette théorie pointent du doigt le fait que le terme de « quasi-espèce » est utilisé à outrance, ce qui est en effet démontré par l'afflux de publications se référant à ce terme sans vraiment en considérer la dynamique et les conséquences. Le terme semble en effet être utilisé à n'importe quel type de variation intra-hôte (113). L'observation de ces variations est certes une condition nécessaire mais pas suffisante pour démontrer que la population virale adopte un comportement de quasi-espèce. Il faut en effet tenir compte de toutes les caractéristiques étudiées jusqu'à présent et ne pas résumer une quasi-espèce à des variations de séquence au sein d'une population.

Les notions d'interactions, notamment de complémentarité et d'interférences, sont aussi remises en question étant donné le manque de preuves à l'appui du mécanisme impliqué. Là aussi, on peut considérer que le terme de coopération est abusif et qu'il s'agit en fait d'une forme de balance entre les mutations et les sélections au sein d'une population.

##### **6.4.2. Remise en cause des caractéristiques du comportement des quasi-espèces**

L'une des autres caractéristiques de ce comportement est l'action de la population en tant qu'ensemble et non pas en tant qu'individus. Il existerait en fait un biais dû au très haut taux de mutations au sein des espèces virales à ARN : ce dernier est tellement haut qu'il ne pourrait résulter qu'en un « couplage des mutations » entre les variants, expliquant ainsi qu'ils évoluent comme un groupe (113). De plus, la plupart des études se concentrent sur la génération de variants mutants présentant un avantage sélectif (une résistance, un échappement...): on part d'une sélection qui certes s'applique au groupe, mais pour finalement privilégier les individus plus aptes. De ce point de vue, on est bien dans le cadre de la survie du plus apte.

La pertinence du modèle des quasi-espèces dans la nature et celle des conditions expérimentales, qui seraient source de biais, sont aussi remises en cause. Outre les limites des conditions expérimentales décrites plus haut, il est plus exactement décrit que les taux de mutations sont augmentés de manière artificielle, et que ce comportement est mis en évidence uniquement de manière transitoire dans la plupart des cas. Un exemple se retrouve chez des viroïdes infectant des plants de chrysanthèmes qui semblaient présenter un comportement allant dans le sens de la survie du plus plat seulement quand le taux de mutations était artificiellement augmenté, comme cela a pu aussi être fait avec le VSV (50, 210). Même si ce comportement reste donc plausible, il se peut très bien que les taux de mutations induits soient si élevés qu'ils ne soient pas applicables dans la nature. Il se pourrait alors très bien que les populations à ARN fassent preuve d'un fort taux de mutation comparé aux autres entités biologiques, mais pas assez haut pour décrire le comportement des quasi-espèces. Cela s'applique notamment au virus considéré comme le pionnier des quasi-espèces, le HIV-1.

Si l'existence du seuil d'erreur n'est pas remise en cause et est compatible avec la théorie des quasi-espèces, elle peut aussi être expliquée par la génétique des populations dite classique (111). Holmes soutient que toute forme de vie peut être soumise à une « crise », où beaucoup de mutations interviennent, comme si elles dépassaient un seuil d'erreur. Cela reste cependant difficile à déterminer pour les formes de vie sexuées.

Un débat touche aussi l'origine de la mutagenèse létale, bien que son existence et son application ne soient pas remise en cause (113). Il est proposé que cette dernière soit due au franchissement d'un seuil d'extinction, qui ne nécessite pas l'intervention des modèles des quasi-espèces, et est un parallèle au seuil épidémiologique d'extinction d'une maladie. Ce mécanisme est basé sur le taux de mutation mais aussi des composantes écologiques (32). La confusion entre seuil d'erreur et seuil d'extinction est souvent faite et est discutée Partie 5.5.5.1.

### **6.4.3. Un modèle fait d'exemples**

Toutes les données rapportées dans la littérature scientifique se réfèrent généralement à un ou deux virus étudié *in vitro* ou *in vivo* afin de décrire une dynamique évolutive. On pourrait grossièrement dire que c'est un ensemble de cas observés chez différents virus qui ont permis de comprendre la théorie des quasi-espèces appliquée aux virus à ARN, et d'établir des modèles évolutifs.

La littérature scientifique échoue à dresser un index universel et compréhensible de la complexité des quasi-espèces, une loi générale pouvant s'appliquer à tous les virus à ARN sans exception. Cela peut être expliqué par les nombreuses variables qui interviennent dans la dynamique des quasi-espèces, et aucune d'elles ne suffisent, indépendamment, à capturer tous les aspects de cette complexité. Ces multiples variables interagissent fortement et sont parfois incroyablement complexes. On en revient à être finalement très descriptif et très sommaire pour comprendre la dynamique d'évolution des populations virales.

Cela met en évidence un parallèle intéressant entre théorie et vérification expérimentale. Des protocoles expérimentaux tentent de vérifier des modèles théoriques comme celui des quasi-espèces, mais certains résultats inattendus apportent de nouvelles idées pour d'autres modèles théoriques, qui doivent être à leur tour testés. C'est ce que la théorie des quasi-espèces a continuellement apporté à la recherche sur l'évolution virale.

#### **6.4.4. Plusieurs modèles complémentaires**

Ce « chiisme » entre ceux qui ont accepté les quasi-espèces et d'autres qui jugent cette théorie inutile dure depuis des dizaines d'années et semble encore continuer de nos jours. Si certains chercheurs ne sont pas pour cette qualification de quasi-espèce, ils ne peuvent nier l'impact de ces théories et modèles expérimentaux, qui ont apporté des idées innovantes et importantes du point de vue de l'évolution des populations virales. Ils ont permis de mettre en place de nouveaux protocoles expérimentaux, de nouveaux modèles *in silico* pour comprendre les mécanismes évolutifs, et pas uniquement chez les virus.

D'après Wilke, il n'y a tout simplement pas de conflit entre les quasi-espèces et les modèles classiques de génétique des populations et de balance de sélection-mutation (259). Ce modèle aurait été créé pour répondre aux limites de la génétique des populations. Les virologistes ont en effet longtemps considéré que la génétique des populations ne suffisait pas à expliquer les observations faites lors l'étude de l'évolution virale. Les micro-organismes étudiés n'étaient pas considérés comme des modèles satisfaisants pour une approche correcte de cette évolution. Les outils apportés n'étaient tout simplement pas suffisants pour comprendre les observations faites chez les populations virales, comme la présence du spectre mutant ou le potentiel évolutif rapide des virus à ARN. La théorie des quasi-espèces est apparue assez tôt et présente donc un modèle alléchant, car elle était la première à s'intéresser à une hétérogénéité génétique et phénotypique forte, due à une réplication présentant beaucoup d'erreur, caractère qui se retrouvait pour les travaux chez les virus à ARN. La théorie semble apporter des éléments de réponse, d'autant plus qu'elle a évolué en parallèle de la recherche en virologie. Elle s'inspire elle-même de la génétique des populations, et afin d'être plus complète, elle peut aussi viser l'utilisation de la théorie des niches écologiques. Cette dernière prend en compte la position spatiale occupée par les populations virales dans l'écosystème qu'est l'organisme. Ce n'est pas une unique théorie qui peut permettre d'appréhender l'évolution chez les virus, mais plusieurs, parfois complémentaires, car elles vont aborder avec plus de précision les différents caractères de cette évolution.

#### **6.5. Extension de la théorie à d'autres organismes**

Les nombreuses limites proposées, bien que la plupart puissent être résolues, doivent nous faire relativiser l'utilisation du terme des quasi-espèces et les applications qui en découlent. Cependant, de nombreuses recherches ont tenté de savoir si cette théorie pouvait s'appliquer à d'autres systèmes biologiques. Si cela est le cas, ces extensions seraient probablement dues à deux caractéristiques que l'on retrouve chez les quasi-espèces : le taux de mutation élevé et son impact, ainsi que le comportement collectif de l'ensemble des quasi-espèces.

##### **6.5.1. Extension aux virus à ADN**

Il est naturel de penser qu'un comportement similaire à une quasi-espèce peut se retrouver chez les virus à ADN possédant une polymérase avec une faible fidélité. Les conséquences pour leur dynamique, suite à de multiples erreurs lors de la réplication, seront les mêmes que pour des virus à ARN (169). Certains virus à ADN, comme les CMV, semblent en effet présenter des caractéristiques retrouvées chez les virus à ARN, comme la complémentarité. Plusieurs souches du CMV peuvent se retrouver chez des patients (46). Ces souches, prises individuellement, sont pathogènes, mais une co-infection semble améliorer leur dissémination et leur réplication. Si, en premier lieu, l'hypothèse était que le mécanisme



consistait en des recombinaisons ou toute autre action synergique des différents variants, celle retenue est que les virus en co-infection agissent via des interactions positives. Ici, la complémentarité pourrait agir pour augmenter la pathogénicité, la charge virale et la sévérité d'une maladie. Une co-infection de plusieurs souches assombrit en effet grandement le pronostic chez les patients infectés. Des exemples de mécanismes d'interférence ont aussi été décrits chez les virus à ADN (169).

Le HBV est un cas particulier de virus à ADN, car son acide nucléique se réplique par l'intermédiaire d'un ARN à l'aide d'une transcriptase inverse. C'est pourquoi on retrouve une diversité et une dynamique similaires à celles des virus à ARN (137, 143). Certains virus à ADN simple brin, comme le parvovirus canin, possèdent un taux d'évolution similaire à celui des virus à ARN, et des mécanismes de réplication dont les fonctions de relecture et de correction sont moins efficaces (143). Des thérapies combinées commencent à être évaluées pour traiter de tels virus à ADN comme c'est le cas pour leurs homologues à ARN.

Cependant la question se pose pour le cas des virus à ADN dont la réplication est réalisée par des polymérases dont la fidélité est forte. Comme on l'a vu précédemment, plus un génome est complexe, plus la capacité d'exploration de l'espace séquence est limitée, ce qui signifierait que la capacité d'adaptation des virus à ADN est moindre comparée à celle des virus à ARN. Or cela ne semble pas le cas en pratique, même si elle ne l'est pas autant que pour une quasi-espèce à ARN. Ils sont en effet capables de s'adapter à différents environnements, mais forment-ils des nuages de mutants, comme les virus à ARN ? L'étude de cette capacité d'adaptation des virus à ADN en comparaison de celle des virus à ARN pourrait faire l'objet d'un autre travail.

### **6.5.2. Extension aux cellules : comportement collectif chez des bactéries**

Les mêmes arguments s'appliquent aux cellules, qui ont des génomes d'un ordre de grandeur bien supérieur à ceux des virus. Une tendance des individus à agir de manière collective a ainsi été décrite chez les bactéries (88).

Ce comportement serait médié par des signaux intercellulaires fonctionnant de manière densité-dépendante, appelé la détection du quorum (quorum sensing QS) (205). Cette dernière est basée sur un «two-components phosphorelay signal transduction» ou «two-component regulatory system» (TCS), qui permet aux microorganismes de ressentir la densité cellulaire et de produire une réponse coordonnée en synchronisant l'expression de certains gènes, agissant ainsi comme un ensemble. Des études de 2009 ont montré que ces interactions de coopération peuvent moduler le niveau de virulence. Des infections mixtes chez des souris, composées de bactéries sensibles au QS et de mutants n'y répondant pas, ont montré une virulence diminuée par rapport à une infection ne contenant pas de mutants (205). Ces mutants agissent en fait comme des fraudeurs, comparables à des «defectors» chez les quasi-espèces, présentent une virulence plus faible et une plus grande sensibilité à la réponse immunitaire (205), et exploitent ces signaux de coopération pour augmenter de ce fait leur fréquence. Ils provoquent ainsi la rupture de la coopération, ce qui rappelle fortement la mutagenèse létale chez les virus, et permettrait d'ailleurs une meilleure réponse au traitement. De plus, on peut supposer qu'une diversité élevée au sein de la population, avec la sélection de sous-populations bactériennes, aurait pour conséquence des plus grands niveaux de coopération et de modulation de la virulence, rappelant fortement ce qui se retrouve chez les quasi-espèces virales (244, 245).

Ainsi, les bactéries se comporteraient comme des collectivités d'individus, utilisant des moyens de communications intercellulaires pour faciliter leur adaptation à des

changements environnementaux et moduler leur virulence, leur motilité, la résistance à des antibiotiques ou encore la formation de biofilms (205). Cela s'accorde souvent avec une pathogénèse augmentée (169).

La formation de biofilms chez ces bactéries est une forme un peu particulière de comportement collectif. Il s'agit de communautés de microorganismes qui s'agrègent sur une surface, inerte ou biologique, au sein d'une matrice adhésive, protectrice et extracellulaire d'exopolymères. Les cellules qui le composent et s'y développent ont une composition physiologiquement différente et sont fortement hétérogènes comparé à celle des particules en suspension qui évoluent de manière isolée dans un milieu liquide. Ces biofilms ont d'importantes implications cliniques pour les agents infectieux, en leur conférant notamment une résistance aux antibiotiques. Les mécanismes expliquant cette dernière ne sont pas encore bien connus, mais on suppose qu'il y a une diffusion limitée du principe actif, que ce microenvironnement altère le fonctionnement des antibiotiques ou qu'il y a présence de sous-populations microbiennes au sein de ce biofilm qui forment un état phénotypique hautement protégé (227). Il n'est pas exclu que ce soit une interaction de tous ces éléments qui explique cette résistance. Cependant la dernière hypothèse rappelle encore une fois l'hétérogénéité décrite pour les populations de virus à ARN.

Une stratégie proposée pour combattre ces biofilms serait l'utilisation d'infection par des bactériophages. Cependant, une étude a démontré qu'un phage Pf4 est essentiel pour de nombreuses étapes de formation du biofilm et contribue de manière significative à la virulence de *P. aeruginosa in vivo*, démontrant ainsi que ces communautés sont fortement complexes (195). Un modèle a, de plus, proposé que les biofilms soient considérés comme la cible des événements sélectifs, tout comme les quasi-espèces (34).

### **6.5.3. Extension aux cellules cancéreuses : hétérogénéité et comportement en groupe**

Le déterminant important de la grande diversité et variabilité virale à ARN est son fort taux de mutation. Mais les cellules eucaryotes peuvent atteindre ce même niveau d'hétérogénéité grâce à d'autres mécanismes qui leur sont propres, comme dans le cas des maladies cancéreuses. Une très forte hétérogénéité sera alors à la base des importants désordres cliniques retrouvés lors de ces maladies (105). Les groupes de cellules touchées ont alors une croissance incontrôlée, envahissent les tissus adjacents voire les tissus éloignés, situation où l'on parle de métastases.

Des changements génétiques sont à l'origine de l'apparition de ces propriétés : des mutations ponctuelles plus nombreuses sur l'ADN cellulaire, affectant notamment les protéines intervenant dans la liaison à l'ADN, la régulation de la transcription ou les signaux intracellulaires ; des réarrangements chromosomiques, et des modifications épigénétiques (105). Ces éléments apportent aux cellules un avantage quant à leur croissance et leur multiplication.

Le séquençage haut-débit permet l'identification des mutations impliquées dans le développement de cellules cancéreuses, de mettre en évidence une grande diversité de cancers et l'implication d'un plus grand répertoire de génomes que ce que l'on pouvait imaginer (45, 115). Les populations de cellules cancéreuses semblent ainsi présenter une hétérogénéité génétique et une complexité rappelant le modèle des quasi-espèces. Ces critères semblent être d'importants déterminants de leur adaptabilité, incluant la capacité à envahir des organes distants dans le cas des métastases, à échapper à la réponse immunitaire et à devenir résistant aux médicaments anti-cancéreux (105). Des études théoriques traitant de l'instabilité génétique des cellules cancéreuses extrapolent le concept d'erreur catastrophique à ces dernières (89) et soutiennent aussi l'importance de ces critères.

Un modèle mathématique pousse le parallèle encore plus loin, en incluant le concept de cellules souches cancéreuses, une sorte d'analogue de la séquence maîtresse chez les quasi-espèces (222). Ainsi, la tumeur aurait deux composantes : l'une qui permettrait l'exploration de phénotypes permettant la croissance et la persistance de la tumeur, et l'autre, plus petite mais plus robuste, composée de cellules souches, qui joueraient le rôle de réservoir stable. Cette stratégie accorderait une « assurance-vie » à la tumeur. Un phénotype de mutation chez les cellules cancéreuses pourrait ainsi agir comme un moteur évolutif afin d'augmenter la probabilité de trouver le phénotype le plus avantageux pour la croissance de la tumeur et sa dispersion, permettant un parallèle intéressant avec la dynamique des quasi-espèces et l'exploration de l'espace séquence (25). Cela rappelle aussi la notion de pan-génome retrouvée chez les populations de bactéries (147) : dans une espèce bactérienne donnée, toutes ses composantes partagent un composant central, et des sous-représentations de cette même espèce diffèrent dans une partie de leur contenu génétique. Cela accorde une certaine flexibilité afin de s'adapter à des changements d'environnement.

Cette variabilité et adaptabilité laisse prévoir que les médicaments anti-cancéreux vont échouer s'ils sont administrés seuls, comme dans le cas d'inhibiteurs antiviraux. La mise en place d'un équivalent des thérapies antivirales telles que le HAART chez le HIV-1 se retrouve donc dans les thérapies anticancéreuses, comme la mise en place du « highly active anti-cancer treatments (HAACT) (134). Le but de ce traitement doit être d'empêcher la sélection de cellules cancéreuses résistantes. Ils sont qualifiés d'orthogonaux et agissent de manière synergique, avec deux voies d'abord anticancéreuse, afin que l'apparition d'une mutation qui limite l'action de l'une des voies n'affectent pas l'autre. Ces traitements successifs, comme chez le HIV-1, pourraient cependant conduire à des échecs thérapeutiques (134). La plupart des protocoles visent en effet l'ADN, ce qui augmente le taux d'émergence de mutants résistants. Pour tenter de contrer cela, la mutagénèse létale, adaptée aux cellules cancéreuses, a été proposée comme nouvelle approche thérapeutique vis-à-vis des tumeurs solides (89).

#### **6.5.4. Extension aux hypothèses des origines de la vie**

La théorie des quasi-espèces telle que formulée par Eigen en 1971 est aussi un modèle d'auto-organisation de formes de vie précoce sur Terre. Elle décrit l'évolution d'entités auto-répliquatives présentant des erreurs nombreuses lors de la réplication et que l'on suspecte comme étant à l'origine de la vie. Cette théorie a été proposée comme un mode d'évolution précellulaire et rappelle les interactions que l'on retrouve dans un spectre de mutants, mais dans un monde cellulaire à ARN supposé (87, 262). Ce mode d'évolution implique notamment que des échanges fréquents de matériel génétique aient lieu, à une étape précédant l'établissement des barrières que nous connaissons aujourd'hui, et définissent les lignées génétiques qui auraient évolué jusqu'à donner celles qu'on connaît aujourd'hui. On aurait donc affaire à une évolution rapide selon les principes de Darwin, où des lignées hautement variables disparaissent, apparaissent et interagissent comme c'est le cas dans un répertoire de mutants avec un grand spectre, variable.

#### **6.5.5. Extension à des systèmes de prions**

La théorie des quasi-espèces se cantonne à des entités construites à base d'acides nucléiques. Une extension intéressante et intrigante a été faite quant à l'hétérogénéité des conformations de prions, comme un déterminant de la spécificité d'hôte (130).

Les prions sont des agents infectieux composés uniquement de protéines et sont responsables de maladies neurodégénératives comme l'encéphalopathie spongiforme bovine.

Il s'agit de formes anormales, notées PrP<sup>SC</sup>, de la protéine PrP<sup>C</sup>, impliquée dans le fonctionnement normal de la cellule. Ils se propagent en transmettant un état protéique anormal chez une protéine normale, codée par des cellules (18, 39). Les souches de prions sont définies par leurs différences de conformation de la PrP<sup>C</sup>.

Des prions clonés sur culture cellulaire sont progressivement devenus hétérogènes, en accumulant des « mutants » protéiques avec des conformations différentes, et ont été soumis à des forces sélectives, aboutissant parfois à la sélection de variants présentant des résistances à des médicaments (130). Cela démontre non seulement que des systèmes non basés sur des acides nucléiques peuvent former des populations de grande taille et présenter de nombreuses conformations différentes, donc une forte hétérogénéité, mais aussi qu'ils peuvent subir des événements sélectifs. Ces caractéristiques rappellent un comportement évolutif darwinien.

Cela suggère donc que ces prions consistent en une quasi-espèce de conformations et que ces conformations se répliquent (c'est-à-dire se convertissent en une forme agrégée qui est pathologique) et sont sélectionnées plus efficacement. Dans cette extension des quasi-espèces, la « mutation » n'est pas un changement héritable dans un acide nucléique mais un changement de conformation de la protéine du prion. Selon ce modèle, un prion consiste en une conformation majeure et un spectre de variants de conformations (53, 130, 253). La sélection de mutants résistants aux inhibiteurs et l'existence préalable de ces derniers dans la population de prions qui n'avait jamais été exposée à un inhibiteur ont été observés (218, 253) et forment un parallèle surprenant avec le comportement des quasi-espèces virales. L'adaptation peut ici être vue comme un mouvement dans l'espace de conformation, au lieu d'un espace séquence.

Enfin, il peut être surprenant de faire le parallèle entre un système génétique, et une conformation protéique. Mais en considérant les bases moléculaire des erreurs de répllication, on remarque que les mutations sont les conséquences de fluctuations moléculaires soumises à une incertitude, tout comme la conformation d'une protéine est due aux interactions entre les acides aminés, elles-mêmes soumises à des fluctuations moléculaires, généralement dépendantes de l'environnement.

Comme chez les quasi-espèces virales, la compréhension du comportement des prions et de leurs dynamiques permettraient d'ouvrir la voie à de nouvelles approches thérapeutiques, bien qu'aucune thérapie efficace n'existe pour le moment.

## **6.6. Conclusion**

Malgré les limites inhérentes aux protocoles expérimentaux, aux simulations *in silico*, aux paramètres et aux méthodes d'analyses utilisés pour décrire la dynamique des quasi-espèces, le modèle des quasi-espèces semble s'appliquer aux virus à ARN et ne semble pas être en désaccord avec d'autres approches visant à décrire l'évolution des populations. De nombreux points restent cependant à revoir, car la littérature scientifique est pour l'instant dans l'incapacité d'apporter un modèle général et bien défini s'appliquant à toutes les populations virales à ARN. Le fait que l'on retrouve des similitudes entre ces populations et d'autres organismes biologiques nous pousse à croire que la théorie des quasi-espèces est encore vouée à évoluer.

Peut-être que la réponse la plus objective se trouve dans cette citation : « Qu'elle soit énoncée comme quasi-espèce ou non, nous avons affaire à une structure de population unique, qui défie les traitements antiviraux malgré les énormes investissements économiques mis en jeu, et ce terme, présentement, nous aide à interpréter le comportement des virus et à mettre en place de nouveaux moyens de contrôler les maladies » (169,179).



## CONCLUSION

Les populations virales présentent bien une évolution selon des critères darwiniens : la variabilité retrouvée *in vivo* peut être expliquée par les variations génétiques touchant la population et subissant une sélection, le tout étant modulé par des événements stochastiques.

Cette étude bibliographique permet de faire le point sur les paramètres qui semblent caractériser une quasi-espèce virale à ARN et apparaissent comme pertinents en regard de l'hétérogénéité phénotypique et génétique observée chez des virus. Ces multiples paramètres sont interconnectés et la compréhension de leurs interactions est complexe.

Le principal argument retrouvé passe par un spectre de mutants, dont l'existence a été démontrée malgré les erreurs dues aux amplifications et aux séquençages. Ce spectre est caractérisé par une complexité, une hétérogénéité et une composition qui détermine le comportement, la dynamique des populations virales, c'est-à-dire sa capacité d'adaptation et sa pathogénicité. Ces derniers sont définis par un taux de mutation élevé et une taille du génome évoluant à la limite du seuil d'erreur, un taux de réplication important, une forte taille de la population, mais aussi de mutations ponctuelles peu nombreuses conjuguée à une faible barrière phénotypique. Ils déterminent l'exploration de l'espace séquence, vue comme une capacité à trouver la sous-population avec une séquence génomique lui fournissant les outils pour se confronter aux contraintes sélectives de l'environnement. C'est une importante force d'adaptabilité, qui résulte souvent en une charge virale forte et une sévérité importante de la maladie. Tous ces paramètres sont autant de critères dont la modification va moduler l'adaptation d'une population virale : les goulots d'étranglement vont diminuer la taille et la complexité, des éléments mutagènes vont augmenter ou diminuer le taux de mutation, les facteurs de l'hôte vont éliminer une partie de la population. Mais cette corrélation entre toutes les caractéristiques n'est pas absolue car les facteurs de l'hôte et les interactions intra-populations de complémentarité et d'interaction ont aussi leur rôle à jouer.

Si la théorie des quasi-espèces semble concorder avec le comportement des populations virales, elle remet en cause le principe darwinien avec la survie du plus plat. Elle explique aussi l'échec des monothérapies et de certains vaccins dans la lutte antivirale, tout en apportant des moyens théoriques de contrer la capacité d'adaptation virale en utilisant les paramètres viraux, par exemple le principe du seuil d'erreur lors de la mutagénèse létale. Les futurs protocoles de traitement prendraient donc en compte, dans l'idéal, le profil immunitaire du patient, la pharmacocinétique des molécules antivirales, les compatibilités des médicaments, le génotype viral, la composition du spectre de mutants.

Il faut cependant rester prudent quant à l'interprétation de ces paramètres et des propriétés apportées par la théorie. Nombre des implications biologiques comme les mécanismes d'interactions, la mémoire moléculaire ou l'apparition de mutants ont été retrouvés chez certains virus mais pas tous. De plus, les limites inhérentes aux protocoles expérimentaux, aux outils utilisés et à l'interprétation des résultats sont des éléments à toujours garder en considération. Enfin, le fait que la théorie ait évolué conjointement à la virologie dans son ensemble rend encore plus délicate l'étude de sa pertinence.

**AGREMENT SCIENTIFIQUE**  
**En vue de l'obtention du permis d'imprimer de la thèse de doctorat vétérinaire**

Je soussigné, **Stéphane BERTAGNOLI**, Enseignant-chercheur, de l'Ecole Nationale Vétérinaire de Toulouse, directeur de thèse, certifie avoir examiné la thèse de **CADIEU Yannis** intitulée « *La théorie des quasi-espèces : Concepts, applications à la dynamique des populations de virus à ARN, implications biologiques et limites* » et que cette dernière peut être imprimée en vue de sa soutenance.

Fait à Toulouse, le 17 novembre 2015  
Professeur Stéphane BERTAGNOLI  
Enseignant chercheur  
de l'Ecole Nationale Vétérinaire de Toulouse



Vu :  
La Directrice de l'Ecole Nationale  
Vétérinaire de Toulouse  
Isabelle CHMIFELIN



Vu :  
Le Président du jury :  
Professeur Christophe PASQUIER



Vu et autorisation de l'impression :  
Administrateur Provisoire de  
l'Université  
Paul Sabatier  
Monsieur Amal SAYAH



L'Administrateur Provisoire



Amal SAYAH

**M. CADIEU Yannis**  
a été admis(e) sur concours en : 2010  
a obtenu son diplôme d'études fondamentales vétérinaires le : 26/06/2014  
a validé son année d'approfondissement le : 24/09/2015  
n'a plus aucun stage, ni enseignement optionnel à valider.

## BIBLIOGRAPHIE

- (1) ACEVEDO A, ANDINO R (2014). Library preparation for highly accurate population sequencing of RNA viruses. *Nat. Protoc.* 9(7), 1760–1769.
- (2) ACEVEDO A, BRODSKY L, ANDINO R (2014). Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* 505, 686-690.
- (3) AGOL VI (2010). Picornaviruses as a model for studying the nature of RNA recombination, p 239–252. In Ehrenfeld E, Domingo E, Roos RP (ed), *The Picornaviruses*. ASM Press, Washington, DC.
- (4) AGUDO R, FERRER-ORTA C, ARIAS A, DE LA HIGUERA I, PERALES C, PEREZ-LUQUE R, ET AL. (2010). A multi-step process of viral adaptation to a mutagenic nucleoside analogue by modulation of transition types leads to extinction-escape. *PLoS Pathog* 6(8): e1001072.
- (5) ANDERSON JP, DAIFUKU R, LOEB LA (2004). Viral error catastrophe by mutagenic nucleosides. *Annu. Rev. Microbiol.* 58:183–205.
- (6) ANDINO R, DOMINGO E (2015). Viral quasispecies. *Virology* 479-480, 46-51.
- (7) ARAGONES L, BOSCH A, PINTO RM (2008). Hepatitis A virus mutant spectra under the selective pressure of monoclonal antibodies: codon usage constraints limit capsid variability. *J. Virol.* 82:1688 –1700.
- (8) ARBIZA J, MIRAZO S, FORT H (2010). Viral quasispecies profiles as the result of the interplay of competition and cooperation. *BMCEvol. Biol.* 10:137.
- (9) ARIAS A, LÁZARO E, ESCARMIS C, DOMINGO E (2001). Moleculaire intermediates of fitness gain of an RNA virus: characterization of a mutant spectrum by biological and molecular cloning. *J.Gen. Virol.* 82, 1049-1060.
- (10) ARIAS A, RUIZ-JARABO CM, ESCARMIS C, DOMINGO E (2004). Fitness increase of memory genomes in a viral quasispecies. *J. Mol. Biol.* 339: 405–412.
- (11) ARIAS A, ARNOLD JJ, SIERRA M, SMIDANSKY ED, DOMINGO E, CAMERON CE (2008). Determinants of RNA-dependent RNA polymerase (in)fidelity revealed by kinetic analysis of the polymerase encoded by a foot-and-mouth disease virus mutant with reduced sensitivity to ribavirin. *J. Virol.* 82:12346 – 12355.
- (12) ARRIBAS M, CABANILLAS L, LAZARO E (2011). Identification of mutations conferring 5 azacytidine resistance in bacteriophage Qbeta. *Virology* 417:343–352.
- (13) BACCAM P, THOMPSON RJ, FEDRIGO O, CARPENTER S, CORNETTE JL (2001). PAQ: partition analysis of quasispecies. *Bioinformatics* 17:16 –22.
- (14) BACCAM P, ET AL. (2003). Subpopulations of equine infectious anemia virus Rev coexist in vivo and differ in phenotype. *J. Virol.* 77:12122–12131.
- (15) BAE YS, YOON JW (1993). Determination of diabetogenicity attributable to a single amino acid, Ala776, on the polyprotein of encephalomyocarditis virus. *Diabetes* 42:435– 443.
- (16) BARANOWSKI E, RUIZ-JARABO CM, SEVILLA N, ANDREU D, BECK E, DOMINGO E (2000). Cell recognition by foot-and-mouth disease virus that lacks the RGD integrin-binding motif: flexibility in aphthovirus receptor usage. *J. Virol.* 74:1641–1647.
- (17) BARANOWSKI E, RUIZ-JARABO CM, DOMINGO E (2001). Evolution of cell recognition by viruses. *Science* 292:1102–1105.



- (18) BARRIA MA, MULHERJEE A, GONZALEZ-ROMERO D, MORALES R, SOTO C, (2009). De novo generation of infectious prions *in vitro* produces a new disease phenotype. *PLoS Pathog.* 5 (5), e1000421.
- (19) BEERENWINKEL N, ZAGORDI O. (2011). Ultra-deep sequencing for the analysis of viral populations. *Curr Opin Virol.* 1(5), 413-8
- (20) BEERENWINKEL N, GUNTARD HF, ROTH V, METZNER KJ. (2012). Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. *Front Microbiol.* 3, 1-16
- (21) BERGELSON JM (2010). Receptors, p 73–86. In Ehrenfeld E, Domingo E, Roos, RP (ed), *The picornaviruses*. ASM Press, Washington, DC.
- (22) BIEBRICHER CK, EIGEN M, GARDINER WC (1991). Quantitative analysis of selection and mutation in self-replicating RNA, p 317–337. In Peliti L (ed), *Biologically inspired physics*. Plenum Press, New York, NY.
- (23) BIEBRICHER CK, EIGEN M (2005). The error threshold. *Virus Res.* 107:117– 127.
- (24) BIEBRICHER CK (2008). Mutation, competition, and selection as measured with small RNA molecules, p 65–86. In Domingo E, Parrish CR, Holland JJ (ed), *Origin and evolution of viruses*, 2nd ed. Elsevier, Oxford, United Kingdom
- (25) BIELAS JH, LOEB KR, RUBIN BP, TRUE LD, LOEB L.A, (2006). Human cancers express a mutator phenotype *Proc. Natl. Acad. Sci. U. S. A.*, 103 (48) (2006), pp. 18238–18242
- (26) BORDERIA AV, LORENZO-REDONDO R, PERNAS M, CASADO C, ALVARO T, DOMINGO E, LOPEZ-GALINDEZ C (2010). Initial fitness recovery of HIV-1 is associated with quasispecies heterogeneity and can occur without modifications in the consensus sequence. *PLoS One* 5:e10319.
- (27) BORDERIA AV, ROZEN-GAGNON K, VINUZZI M, (2015). Fidelity variants and RNA quasispecies. *Curr. Top. Microbiol. Immunol.*
- (28) BORREGO B, NOVELLA IS, GIRALT E, ANDREU D, DOMINGO E (1993). Distinct repertoire of antigenic variants of foot-and-mouth disease virus in the presence or absence of immune selection. *J. Virol.* 67:6071– 6079.
- (29) BORROW P, EVANS CF, OLDSTONE MB (1995). Virus-induced immunosuppression: immune system-mediated destruction of virus-infected dendritic cells results in generalized immune suppression. *J. Virol.* 69:1059– 1070.
- (30) BOYD MT, SIMPSON GR, CANN AJ, JOHNSON MA, WEISS RA (1993). A single amino acid substitution in the V1 loop of human immunodeficiency virus type 1 gp120 alters cellular tropism. *J. Virol.* 67:3649 –3652.
- (31) BRIONES C, DOMINGO E (2008). Minority report: hidden memory genomes in HIV-1 quasispecies and possible clinical implications. *AIDS Rev.* 10:93–109.
- (32) BULL JJ, MEYERS LA, LACHMANN M (2005). Quasispecies made simple. *PLoS Comput. Biol.* 1:e61.
- (33) BULL JJ, SANJUAN R, WILKE CO (2007). Theory of lethal mutagenesis for viruses. *J. Virol.* 81:2930 –2939
- (34) CALDWELL DE, COSTERTON JW (1996). Are bacterial biofilms constrained to Darwin's concept of evolution through natural selection? *Microbiologia*, 12 (3) (1996), pp. 347–358
- (35) CAO L, WU C, SHI H, GONG Z, ZHANG E, ET AL (2014). Coexistence of hepatitis B virus quasispecies enhances viral replication and the ability to induce host antibody and cellular immune response. *J. Virol.* 88, 8656-8666.

- (36) CARILLO C, BORCA M, MOORE D, MORGAN DO, SOBRINO F, (1998). *In vivo* analysis of the stability and fitness of variants recovered from foot-and-mouth disease virus quasispecies. *J. Gen. Virol.* 79, 1699-1706.
- (37) CARRILLO C, ET AL. (2007). Genetic and phenotypic variation of foot-and-mouth disease virus during serial passages in a natural host. *J. Virol.* 81:11341–11351.
- (38) CASES-GONZALEZ C, ARRIBAS M, DOMINGO E, LAZARO E (2008). Beneficial effects of population bottlenecks in an RNA virus evolving at increased error rate. *J. Mol. Biol.* 384:1120 –1129.
- (39) CASTILLA, J, MORALES R, SAA P, BARRIA M, GAMBETTI P, SOTO C (2008). Cell-free propagation of prion strains. *EMBO J.* 27 (19), 2557-2566.
- (40) CASTRO C, ARNOLD JJ, CAMERON CE (2005). Incorporation fidelity of the viral RNA-dependent RNA polymerase: a kinetic, thermodynamic and structural perspective. *Virus Res.* 107:141–149.
- (41) CHAO L (1990). Fitness of RNA virus decreased by Muller's ratchet. *Nature* 348:454–455.
- (42) CHARPENTIER N, DAVILA M, DOMINGO E, ESCARMIS C (1996). Long-term, large-population passage of aphthovirus can generate and amplify defective noninterfering particles deleted in the leader protease gene. *Virology* 223:10 –18.
- (43) CHEN-HARRIS H, BORUCKI MK, TORRES C, SLEZAK TR, ALLEN JE (2013). Ultra-deep mutant spectrum profiling: improving sequencing accuracy using overlapping read pairs. *BMC Genomics* 14, 96.
- (44) CHETVERIN AB, CHETVERINA HV, DEMINDENKO AA, ET AL (1997). Nonhomologous RNA recombination in a cell-free system: Evidence for a transesterification mechanism guided by secondary structure. *Cell* 1997; 88:503-513.
- (45) CHIN L, HAHN WC, GETZ G, MEYERSON M (2011). Making sense of cancer genomic data. *Genes Dev.* 25:534 –555.
- (46) CICIN-SAIN L, PODLECH J, MESSERLE M, REDDEHASE MJ, KOSZINOWSKI UH (2005). Frequent coinfection of cells explains functional *in vivo* complementation between cytomegalovirus variants in the multiply infected host. *J. Virol.* 79, 9492-9502.
- (47) CIUREA A, KLENERMAN P, HUNZIKER L, HORVATH E, SENN BM, ET AL. (2000). Viral persistence *in vivo* through selection of neutralizing antibody-escape variants. *Proc. Natl. Acad. Sci. U. S. A.* 97: 2749–2754.
- (48) CLARKE DK, DUARTE EA, MOYA A, ELENA SF, DOMINGO E, HOLLAND J (1993). Genetic bottlenecks and population passages cause profound fitness differences in RNA viruses. *J. Virol.* 67:222–228.
- (49) CLEMENTI M (2008). Perspectives and opportunities for novel antiviral treatments targeting virus fitness. *Clin. Microbiol. Infect.* 14:629–631.
- (50) CODONER FM, DARO S J-A, SOLE RV, ELENA SF (2006) The fittest versus the flattest: experimental confirmation of the quasispecies effect with subviral pathogens. *PLoS Pathog* 2: e136.
- (51) COFFIN JM (1995). HIV population dynamics *in vivo*: implications for genetic variation, pathogenesis and therapy. *Science* 267, 483-489.
- (52) COFFIN J, HUGHES SH, VARMUS HE, EDS (1997). Reverse Transcriptase and the Generation of Retroviral DNA. In *Retroviruses*. New York: Cold Spring Harbor Laboratory Press 1997.
- (53) COLLINGE J, CLARKE AR (2007). A general model of prion strains and their pathogenicity. *Science* 318:930 –936.

- (54) COSTA-MATTIOLI M, DOMINGO E, CRISTINA J (2006). Analysis of sequential Hepatitis A virus strains reveals coexistence of distinct viral subpopulations. *J. Gen. Virology* 87, 115-118.
- (55) CRICK, F (1970). Central dogma of molecular biology. *Nature* 227(5258): 561-3
- (56) CUERVO NS, ET AL. (2001). Genomic features of intertypic recombinant sabin poliovirus strains excreted by primary vaccinees. *J. Virol.* 75:5740- 5751.
- (57) DAPP MJ, CLOUSER CL, PATTERSON S, MANSKY LM (2009). 5-Azacytidine can induce lethal mutagenesis in human immunodeficiency virus type 1. *J. Virol.* 83:11950 -11958.
- (58) DE LA TORRE J.C., HOLLAND J.J (1990). RNA virus quasispecies populations can suppress vastly superior mutant progeny. *J. Virol.* 64, 6278-6281. 320.
- (59) DELGADO-ECKERT E, OJOSNEGROS S, BEERENWINKEL N (2011). The evolution of virulence in RNA viruses under a competition-colonization trade-off. *Bull. Math. Biol.* 73:1881-1908.
- (60) DENISON MR, GRAHAM RL, DONALDSON EF, ECKERLE LD, BARIC RS (2011). Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. *RNA Biol.* 8:270 -279.
- (61) DOMINGO E, SABO D, TANIGUCHI T, WEISSMANN C (1978). Nucleotide sequence heterogeneity of an RNA phage population. *Cell* 13:735-744.
- (62) DOMINGO E, HOLLAND J.J (1994). Mutation rates and rapid evolution of RNA viruses. In: Lirse SS, (ed.) *Evolutionary Biology of viruses*. New York: Raven Press 1994 :161-184.
- (63) DOMINGO E, HOLLAND J.J (1997) RNA virus mutations and fitness for survival. *Annu. Rev. Microbiol.* 51:151-178
- (64) DOMINGO E, BIEBRICHER C, EIGEN M, HOLLAND J.J (2001). *Quasispecies and RNA virus evolution. Principles and consequences*. 173p. Landes Bioscience, Austin TX, Unit 14.
- (65) DOMINGO E, RUIZ-JARABO CM, ARIAS A, GARCIA-ARRIAZA JF, ESCARMÍS C. (2004). Quasispecies dynamics and evolution of foot-and-mouth disease virus, p 261-304. In Sobrino F, Domingo E (ed), *Foot-and-mouth disease*. Horizon Bioscience, Wymondham, United Kingdom.
- (66) DOMINGO E (ED). (2005). Virus entry into error catastrophe as a new antiviral strategy. *Virus Res.* 107:115-228.
- (67) DOMINGO E, ET AL. (2008). Viral quasispecies: dynamics, interactions and pathogenesis, p 87-118. In Domingo E, Parrish C, Holland J.J (ed), *Origin and evolution of viruses*, 2nd ed. Elsevier, Oxford, United Kingdom.
- (68) DOMINGO E, GRANDE-PEREZ A, MARTIN V (2008). Future prospects for the treatment of rapidly evolving viral pathogens: insights from evolutionary biology. *Expert Opin. Biol. Ther.* 8:1455-1460.
- (69) DOMINGO E, HORZINEK MC (2008). Animal virology: a showcase of evolution, p 523-531. In Mettenleiter TC, Sobrino F (ed), *Animal viruses. Molecular biology*. Caister Academic Press, Norfolk, United Kingdom.
- (70) DOMINGO E, ET AL. (2010). Mutation, quasispecies and lethal mutagenesis, p 197-211. In Ehrenfeld E, Domingo E, Roos RP (ed), *The picornaviruses*. ASM Press, Washington, DC.
- (71) DOMINGO E, SHELDON J, CELIA PERALES C, (2012). Viral quasispecies evolution. *Microbiology and Molecular Biology Reviews.* 76(2):159-216

- (72) DOMINGO E (2015). *Virus as Populations: Composition, Complexity, Dynamics, and Biological Implications*. U.S.A: Elsevier. 399p. ISBN 978-0-12-800837-9.
- (73) DUARTE E, CLARKE D, MOYA A, DOMINGO E, HOLLAND J (1992). Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. *Proc. Natl. Acad. Sci. U. S. A.* 89:6015–6019.
- (74) ECKERLE LD, LU X, SPERRY SM, CHOI L, DENISON MR (2007). High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. *J. Virol.* 81:12135–12144.
- (75) ECKERLE LD, BECKER MM, HALPIN RA, LI K, VENTER E, LU X, SCHERBAKOVA S, GRAHAM RL, BARIC RS, STOCKWELL TB, SPIRO DJ, DENISON MR (2010). Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Pathog.* 6:e1000896.
- (76) EIGEN M. (1971). Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58:465–523.
- (77) EIGEN M, BIEBRICHER CK. (1988). Sequence space and quasispecies distribution, p 211–245. In Domingo E, Ahlquist P, Holland JJ (ed), *RNA genetics*, vol 3. CRC Press, Boca Raton, FL.
- (78) EIGEN M (2000). Natural selection: a phase transition? *Biophys. Chem.* 85:101–123.
- (79) ELENA SF (2012). RNA virus genetic robustness: possible causes and some consequences. *Curr. Opin. Virol.* 2, 525-530.
- (80) ESCARMÍS C, DÁVILA M, CHARPENTIER N, BRACHO A, MOYA A, DOMINGO E (1996). Genetic lesions associated with Muller's ratchet in an RNA virus. *J. Mol. Biol.* 264:255–267
- (81) ESCARMÍS C, GÓMEZ-MARIANO G, DÁVILA M, LÁZARO E, DOMINGO E (2002). Resistance to extinction of low fitness virus subjected to plaque-toplaque transfers: diversification by mutation clustering. *J. Mol. Biol.* 315: 647–661.
- (82) ESCARMÍS C, LAZARO E, ARIAS A, DOMINGO E (2008). Repeated bottleneck transfers can lead to non-cytocidal forms of a cytopathic virus: implications for viral extinction. *J. Mol. Biol.* 376:367–379.
- (83) ESCARMIS C, PERALES C, DOMINGO E (2009). Biological effect of Muller's ratchet: distant capsid site can affect picornavirus protein processing. *J. Virol.* 83:6748–6756.
- (84) FEIGELSTOCK DA, MIHALIK KB, FEINSTONE SM (2011). Selection of hepatitis C virus resistant to ribavirin. *J. Virol.* . 8:402.
- (85) FERRER-ORTA C, SIERRA M, AGUDO R, DE LA HIGUERA I, ARIAS A, PEREZ-LUQUE R, ESCARMIS C, DOMINGO E, VERDAGUER N (2010). Structure of foot-and-mouth disease virus mutant polymerases with reduced sensitivity to ribavirin. *J. Virol.* 84: 6188–6199.
- (86) FIGLEROWICZ M, ALEJSKA M, KURZYNSKA-KOKORNIAK A, FIGLEROWICZ M (2003). Genetic variability: the key problem in the prevention and therapy of RNA-based virus infections. *Med. Res. Rev.* 23:488–518.
- (87) FORTERRE P (2010). The universal tree of life and the last universal cellular ancestor: revolution and counterrevolutions, p 43–62. In Caetano-Anollés G (ed), *Evolutionary genomics and systems biology*. Wiley- Blackwell, Hoboken, NJ.
- (88) FOSTER KR (2011). The sociobiology of molecular systems. *Nat. Rev. Genet.* 12 (3), 193-203.

- (89) FOX EJ, LOEB L (2010). Lethal mutagenesis: targeting the mutator phenotype in cancer. *Semin. Cancer Biol.*, 20 (5) (2010), pp. 353–359
- (90) FREISTADT MS, MEADES GD, CAMERON CE (2004). Lethal mutagens: broad spectrum antivirals with limited potential for development of resistance? *Drug Resist. Updat.* 7:19–24.
- (91) GABRIEL G, KLINGEL K, OTTE A, THIELE S, HUDJETZ B, ARMAN-KALCEK G, SAUTER M ET AL. (2011). Differential use of importin-alpha isoforms governs cell tropism and host adaptation of influenza virus. *Nat. Commun.* 2:156.
- (92) GARBELLI A, RADI M, FALCHI F, BEERMANN S, ZANOLI S, MANETTI F, DIETRICH U, BOTTA M, MAGA G (2011). Targeting the human DEAD-box polypeptide 3 (DDX3) RNA helicase as a novel strategy to inhibit viral replication. *Curr. Med. Chem.* 18:3015–3027.
- (93) GARCÍA-ARRIAZA J, OJOSNEGROS S, DÁVILA M, DOMINGO E, ESCARMIS C (2006). Dynamics of mutation and recombination in a replicating population of complementing, defective viral genomes. *J. Mol. Biol.* 360:558–572.
- (94) GEBAUER F, DE LA TORRE JC, GOMES I, MATEU MG, BARAHONA H, TIRABOSCHI B, BERGMANN I, DE MELLO PA, DOMINGO E (1988). Rapid selection of genetic and antigenic variants of foot-and-mouth disease virus during persistence in cattle. *J. Virol.* 62:2041–2049.
- (95) GELLER R, VIGNUZZI M, ANDINO R, FRYDMAN J. (2007). Evolutionary constraints on chaperone-mediated folding provide an antiviral approach refractory to development of drug resistance. *Genes Dev.* 21:195–205
- (96) GONZALEZ-LOPEZ C, ARIAS A, PARIENTE N, GOMEZ-MARIANO G, DOMINGO E (2004). Preextinction viral RNA can interfere with infectivity. *J. Virol.* 78:3319–3324.
- (97) GONZÁLEZ-LÓPEZ C, GÓMEZ-MARIANO G, ESCARMÍS C, DOMINGO E. (2005). Invariant aphthovirus consensus nucleotide sequence in the transition to error catastrophe. *Infect. Genet. Evol.* 5:366–374.
- (98) GORMAN OT, BEAN WJ, WEBSTER RG. (1992). Evolutionary processes in influenza viruses: divergence, rapid evolution, and stasis. *Curr. Top. Microbiol. Immunol.* 176:75–97.
- (99) GRACI JD, CAMERON CE. (2008). Therapeutically targeting RNA viruses via lethal mutagenesis. *Future Virol.* 3:553–566.
- (100) GRACI J.D., GNADIG N.F., GALARRAGA J.E., CASTRO C., VIGNUZZI M. ET AL. (2012). Mutational robustness of an RNA virus influences sensitivity to lethal mutagenesis. *J. Virol.* 86, 2869-2873.
- (101) GRANDE-PEREZ A, SIERRA S, CASTRO MG, DOMINGO E, LOWENSTEIN PR (2002). Molecular indetermination in the transition to error catastrophe: systematic elimination of lymphocytic choriomeningitis virus through mutagenesis does not correlate linearly with large increases in mutant spectrum complexity. *Proc. Natl. Acad. Sci. U. S. A.* 99:12938–12943.
- (102) GRANDE-PEREZ A, GOMEZ-MARIANO G, LOWENSTEIN PR, DOMINGO E. (2005). Mutagenesis-induced, large fitness variations with an invariant arenavirus consensus genomic nucleotide sequence. *J. Virol.* 79:10451–10459.
- (103) GRANDE-PÉREZ A, LAZARO E, LOWENSTEIN P, DOMINGO E, MANRUBIA SC. (2005). Suppression of viral infectivity through lethal defection. *Proc. Natl. Acad. Sci. U. S. A.* 102:4448–4452.
- (104) GREGORI J., SALIERU M., DOMINGO E., SANCHEZ A., ESTEBAN J.I. ET AL (2014). Inference with quasispecies diversity indices: clonal and NGS approaches. *Bioinformatics* 30, 1104-1111.

- (105) HANAHAN D, WEINBERG RA (2011). Hallmarks of cancer: the next generation. *Cell* 144:646–674.
- (106) HANDEL A, REGOES RR, ANTIA R (2006). The role of compensatory mutations in the emergence of drug resistance. *PLoS Comput. Biol.* 2:e137.
- (107) HEINZ BA, RUECKERT RR, SHEPARD DA, DUTKO FJ, MCKINLAY MA, FANCHER M ET AL.(1989). Genetic and molecular analyses of spontaneous mutants of human rhinovirus 14 that are resistant to an antiviral compound. *J. Virol.* 63:2476–2485.
- (108) HOLLAND JJ (1984). Continuum of change in RNA virus genomes, p 137–143. In Notkins AL, Oldstone MBA (ed), *Concepts in viral pathogenesis*. Springer-Verlag, New York, NY.
- (109) HOLLAND JJ, DOMINGO E, DE LA TORRE JC, STEINHAUER DA (1990). Mutation frequencies at defined single codon sites in vesicular stomatitis virus and poliovirus can be increased only slightly by chemical mutagenesis. *J. Virol.* 64:3960–3962.
- (110) HOLLAND JJ, DE LA TORRE JC, CLARKE DK, DUARTE E (1991). Quantification of relative fitness and great adaptability of clonal populations of RNA viruses, *J. Virol.* 65, 2960-2967.
- (111) HOLMES EC (2003) Error thresholds and the constraints to RNA virus evolution. *Trends Microbiol* 11: 543–546.
- (112) HOLMES EC (2009). RNA virus genomics: a world of possibilities. *J. Clin. Invest.* 119:2488–2495.
- (113) HOLMES EC (2010). RNA virus quasispecies: fact or fiction? *J. Mol. Biol.* 400, 271–273
- (114) HOPKINS S, SCORNEAUX B, HUANG Z, MURRAY MG, WRING S, SMITLEY C, HARRIS R, ERDMANN F, FISCHER G, RIBEILL Y (2010). SCY-635, a novel nonimmunosuppressive analog of cyclosporine that exhibits potent inhibition of hepatitis C virus RNA replication in vitro. *Antimicrob. Agents Chemother.* 54:660–672.
- (115) ILIE M, LONG E, HOFMAN V, LESPINETV, BORDONE O, WASHETINE K, GAVRIC-TANGA V, HOFMAN P (2014). Les méthodes de séquençage de « nouvelle génération » (NGS) et le cancer broncho-pulmonaire: principales technologies, applications et limites actuelles en pathologie. *Revue Francophone des Laboratoires* 2014, 51-58.
- (116) IRANZO J, PERALES C, DOMINGO E, MANRUBIA SC (2011). Tempo and mode of inhibitor-mutagen antiviral therapies: a multidisciplinary approach. *Proc. Natl. Acad. Sci. U. S. A.* 108:16008–16013.
- (117) JABARA CB, JONES CD, ROACH J, ANDERSON JA, SWANSTROM R (2011). Accurate sampling and deep-sequencing of the HIV-1 protease gene using a Primer ID. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20166-20171.
- (118) JOHNSON JA, LI JF, WEI X, LIPSCOMB J, IRLBECK D, CRAIG C, SMITH A, BENNETT DE, MONSOUR M, SANDSTROM P, LANIER ER, HENEINE W. (2008). Minority HIV-1 drug resistance mutations are present in antiretroviral treatment-naïve populations and associate with reduced treatment efficacy. *PLoS Med.* 5:e158.
- (119) JOHNSON N., Ph.D. (*University of Massachusetts, Amherst, MA*) (2008). Nature Education Citation: Johnson, N. (2008) Sewall Wright and the development of shifting balance theory. *Nature Education* 1(1):52.

- (120) JONES LR, ZANDOMENI R, WEBER EL (2002). Quasispecies in the 5' untranslated genomic region of bovine viral diarrhoea virus from a single individual. *J. Gen. Virol.* 83, 2161-2168.
- (121) KRAHN M. (2010-2011). *Bases moléculaires des mutations et Bases moléculaires du mode de transmission des maladies génétiques*. En ligne sur le site du Collège National des Enseignants et Praticiens de Génétique Médicale, Département de Génétique Médicale - Hôpital Timone Enfants, AP-HM, Et INSERM UMR910 - Faculté de Médecine, Université de la Méditerranée Marseille. <http://campus.cerimes.fr/genetique-medicale/enseignement/genetique26/site/html/cours.pdf>
- (122) KUMAR N, LIANG Y, PARSLOW TG, LIANG Y (2011). Receptor tyrosine kinase inhibitors block multiple steps of influenza A virus replication. *J. Virol.* 85:2818–2827.
- (123) KURODA MN, KATANO H, NAKAJIMA N, TOBIUME M, AINAI A, ET AL. (2010). Characterization of quasispecies of pandemic 2009 influenza A virus (A/H1N1/2009) by de novo sequencing using a next-generation DNA sequencer. *PLoS One* 5, e10256.
- (124) LALIC J, CUEVAS JM ELENA SF (2011). Effect of host species on the distribution of mutational fitness effects for an RNA virus. *PLoS Genet.* 7:e1002378.
- (125) LAURING AS ANDINO R (2010). Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog.* 6. E1001005.
- (126) LÁZARO E, ESCARMIS C, PEREZ-MERCADER J, MANRUBIA SC, DOMINGO E (2003). Resistance of virus to extinction on bottleneck passages: study of a decaying and fluctuating pattern of fitness loss. *Proc. Natl. Acad. Sci. U. S. A.* 100:10830–10835.
- (127) LEE CH, GILBERTSON DL, NOVELLA IS, HUERTA R, DOMINGO E, HOLLAND JJ (1997). Negative effects of chemical mutagenesis on the adaptive behavior of vesicular stomatitis virus. *J. Virol.* 71:3636–3640.
- (128) LEE CM, BIH FY, CHAO YC, GOVINDARAJAN S, LAI MM (1992). Evolution of hepatitis delta virus RNA during chronic infection. *Virology* 188:265–273.
- (129) LEWICKI H, TISHON A, BORROW P, EVANS CF, GAIRIN JE, HAHN KM, JEWELL DA, WILSON IA, OLDSTONE MB (1995). CTL escape viral variants. I. Generation and molecular characterization. *Virology* 210:29–40.
- (130) LI J, BROWNING S, MAHAL SP, OELSCHLEGEL AM, WEISSMANN C (2010). Darwinian evolution of prions in cell culture. *Science* 327:869–872.
- (131) LOEB LA, MULLINS JI (2000). Lethal mutagenesis of HIV by mutagenic ribonucleoside analogs. *AIDS Res. Hum. Retroviruses* 16:1–3
- (132) LORENZO-REDONDO R, BORDERIA AV, LOPEZ-GALINDEZ C (2011). Dynamics of *in vitro* fitness recovery of HIV-1. *J. Virol.* 85:1861–1870.
- (133) LUO G, TAYLOR J (1990). Template Switching by Reverse Transcriptase during DNA Synthesis. *J Virol.* 64(9), 4321-4328
- (134) LUO J, SOLIMINI NL, ELLEDGE S.J (2009). Principles of cancer therapy: oncogene and non-oncogene addiction. *Cell*, 136 (5) (2009), pp. 823–837
- (135) MANRUBIA SC, DOMINGO E, LAZARO E (2010). Pathways to extinction: beyond the error threshold. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365:1943–1952.
- (136) MANSKY LM (2003). Mutagenic outcome of combined antiviral drug treatment during human immunodeficiency virus type 1 replication. *Virology* 307:116–121.

- (137) MARGERIDON-THERMET S, SHULMAN NS, AHMED A, SHAHRIAR R, LIU T, ET AL (2009). Ultra-deep pyrosequencing of hepatitis B virus quasispecies from nucleoside and nucleotide reverse-transcriptase inhibitor (NRTI)-treated patients and NRTI-naïve patients. *J. Infect. Dis.* 199, 1275-1285.
- (138) MARGULIES M, EGHOLM M, ALTMAN WE, ATTIYA S, BADER JS, BEMBEN LA ET AL. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437(7057), 376-80.
- (139) MARTIN V, GRANDE-PEREZ A, DOMINGO E (2008). No evidence of selection for mutagenesis robustness during lethal mutagenesis of lymphocytic choriomeningitis virus. *Virology* 378, 185-192.
- (140) MARTINEZ MA, CARRILLO C, GONZALEZ-CANDELAS F, MOYA A, DOMINGO E, SOBRINO F (1991). Fitness alteration of foot-and-mouth disease virus mutants: measurement of adaptability of viral quasispecies. *J. Virol.* 65:3954 – 3957.
- (141) MARTINEZ MA, VERDAGUER N, MATEU MG, DOMINGO E (1997). Evolution subverting essentiality: dispensability of the cell attachment Arg-Gly-Asp motif in multiply passaged foot-and-mouth disease virus. *Proc. Natl. Acad. Sci. U. S. A.* 94:6798–6802.
- (142) MARZ M, BEERENWINKEL N, DROSTEN C, FRICKE M, FRISHMAN D, ET AL. (2014). Challenges in RNA virus bioinformatics. *Bioinformatics* 30, 1793-1799.
- (143) MAS A, LOPEZ-GALINDEZ C, CACHO I, GOMEZ J, MARTINEZ MA. (2010). Unfinished stories on viral quasispecies and Darwinian views of evolution. *J. Mol. Biol.* 397:865– 877.
- (144) MATEU MG, VALERO ML, ANDREU D, DOMINGO E. (1996). Systematic replacement of amino acid residues within an Arg-Gly-Asp-containing loop of foot-and-mouth disease virus and effect on cell recognition. *J. Biol. Chem.* 271:12814 – 12819.
- (145) MINOR P.D (1992). The molecular biology of poliovaccines. *J. Gen. Virol.* 1992 ; 73 :3065-77
- (146) MINSKAIA E, HERTZIG T, GORBALENYA AE, CAMPANACCI V, CABBILLAU C, CANARD B, ZIEBUHR J (2006). Discovery of an RNA virus 3'-5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc. Natl. Acad. Sci. U. S. A.* 103:5108 –5113.
- (147) MIRA A, MARTIN-CUADRADO AB, D'AURIA G, RODRIGUEZ-VALERA F (2010). The bacterial pan-genome: a new paradigm in microbiology. *Int. Microbiol.*, 13 (2) (2010), pp. 45–57
- (148) MONROE SS, SCHLESINGER S (1983). RNAs from two independently isolated defective interfering particles of Sindbis virus contain cellular tRNA sequences at their 5' ends. *Proc. Natl. Acad. Sci. USA* 1983 ; 80 :3279-83.
- (149) MORENO H, GALLEGO I, SEVILLA N, DE LA TORRE JC, DOMINGO E, MARTIN V (2011). Ribavirin can be mutagenic for arenaviruses. *J. Virol.* 85:7246 – 7255.
- (150) MULLER HJ (1964). The relation of recombination to mutational advance. *Mut. Res.* 1:2–9.
- (151) MULLINS JI, HEATH L, HUGHES JP, KICHA J, STYRCHAK S, WONG KG, RAO U, HANSEN A, ET AL. (2011). Mutation of HIV-1 genomes in a clinical population treated with the mutagenic nucleoside KP1461. *PLoS One* 6:e15135.



- (152) MUÑOZ E, PARK JM, DEEM MW (2008). Quasispecies theory for horizontal gene transfer and recombination. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 78:061921
- (153) NAGY PD, SIMON AE (1997). New insights into the mechanisms of RNA recombination. *Virology* 1997; 235:1-9.
- (154) NIJHUIS M, VAN MAARSEVEEN NM, BOUCHER CA (2009). Antiviral resistance and impact on viral replication capacity: evolution of viruses under antiviral pressure occurs in three phases. *Handb. Exp. Pharmacol.* 2009: 299–320.
- (155) NILSSON M, SNOAD N (2000). Error threshold for quasispecies in dynamics fitness landscapes. *Phys. Rev. Lett.* 84:191–194.
- (156) NOVELLA IS, DUARTE EA, ELENA SF, MOYA A, DOMINGO E, AND HOLLAND J J (1995). Exponential increases of RNA virus fitness during large population transmissions. *Proc. Natl. Acad. Sci. U. S. A.* 92:5841–5844.
- (157) NOVELLA IS (2004). Negative effect of genetic bottlenecks on the adaptability of vesicular stomatitis virus. *J. Mol. Biol.* 336:61– 67
- (158) NOVELLA IS, EBENDICK-CORPUS BE (2004). Molecular basis of fitness loss and fitness recovery in vesicular stomatitis virus. *J. Mol. Biol.* 342:1423–1430.
- (159) NOVELLA IS, REISSIG DD, WILKE CO (2004). Density-dependent selection in vesicular stomatitis virus. *J. Virol.* 78:5799 –5804.
- (160) NOWAK MA, SCHUSTER P (1989). Error thresholds of replication in finite populations mutation frequencies and the onset of Muller’s ratchet. *J. Theor. Biol.* 137:375–395.
- (161) NÚÑEZ JI, BARANOWSKI E, MOLINA N, RUIZ-JARABO CM, SÁNCHEZ C, DOMINGO E, SOBRINO F (2001). A single amino acid substitution in nonstructural protein 3A can mediate adaptation of foot-and-mouth disease virus to the guinea pig. *J. Virol.* 75:3977–3983.
- (162) NÚÑEZ JI, ET AL. (2007). Guinea pig-adapted foot-and-mouth disease virus with altered receptor recognition can productively infect a natural host. *J. Virol.* 81:8497– 8506
- (163) O’DEA EB, KELLER TE, WILKE CO (2009). Does mutational robustness inhibit extinction by lethal mutagenesis in viral populations? *PLoS Comput. Biol.* 6, e1000811
- (164) O’HARA PJ, HORODYSKI FM, NICHOL ST ET AL. (1984). Vesicular stomatitis virus mutants resistant to defective-interfering particules accumulate stable 5’-terminal and fewer 3’-terminal mutations in a stepwise manner. *J. Virol.* 1984 ;49 :793-798
- (165) OCHOA G. (2006). Error thresholds in genetic algorithms. *Evol. Comput.* 14:157–182.
- (166) OJOSNEGROS S., AGUDO R., SIERRA M., BRIONES C., SIERRA S., ET AL. (2008). Topology of evolving, mutagenized viral populations: quasispecies expansion, compression and operation of negative selection. *BMC Evol. Biol.* 8, 207.
- (167) OJOSNEGROS S, BEERENWINKEL N, ANTAL T, MARTIN A. NOWAK MA, ESCARMÍSC, DOMINGO E (2010). Competition-colonization dynamics in an RNA virus. *Proc. Natl. Acad. Sci. U. S. A.* 107:2108 –2112.
- (168) OJOSNEGROS S, GARCÍA-ARRIAZA J, ESCARMÍS E, MANRUBIA SC, PERALES C, ARIAS A, MATEU MG, DOMINGO E (2011). Viral genome segmentation can result from a trade-off between genetic content and particle stability. *PLoS Genet.* 7:e1001344.
- (169) OJOSNEGROS S, PERALES C, MAS A, DOMINGO E (2011). Quasispecies as a matter of fact: viruses and beyond. *Virus Res.* 162:203–215.

- (170) OZSOLAK F, PLATT AR, JONES DR, REIFENBERGER JG, SASS LE, MC INERNEY P ET AL. (2009). Direct RNA sequencing. *Nature*. 461, 814-8
- (171) PACCIARINI F, ET AL. (2008). Persistent replication of severe acute respiratory syndrome coronavirus in human tubular kidney cells selects for adaptive mutations in the membrane protein. *J. Virol.* 82:5137–5144.
- (172) PARIENTE N, SIERRA S, LOWENSTEIN PR, DOMINGO E (2001). Efficient virus extinction by combinations of a mutagen and antiviral inhibitors. *J. Virol.* 75:9723–9730
- (173) PARIENTE N, AIRAKSINEN A, DOMINGO E (2003). Mutagenesis versus inhibition in the efficiency of extinction of foot-and-mouth disease virus. *J. Virol.* 77:7131–7138
- (174) PARK JM MUNOZ E, DEEM MW (2010). Quasispecies theory for finite populations. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 81:011902.
- (175) PASQUIER C, BERTAGNOLI S, DUNIA D, IZOPE J (2013). *Virologie humaine et zoonoses: Cours et fiches de synthèse*. Paris: Dunod. 270 p. ISBN 978-2-10-059110-7.
- (176) PERALES C, MATEO R, MATEU MG, DOMINGO E (2007). Insights into RNA virus mutant spectrum and lethal mutagenesis events: replicative interference and complementation by multiple point mutants. *J. Mol. Biol.* 369:985–1000
- (177) PERALES C, AGUDO R, DOMINGO E (2009). Counteracting quasispecies adaptability: extinction of a ribavirin-resistant virus mutant by an alternative mutagenic treatment. *PLoS One* 4:e5554.
- (178) PERALES C, AGUDO R, TEJERO H, MANRUBIA SC, DOMINGO E (2009). Potential benefits of sequential inhibitor-mutagen treatments of RNA virus infections. *PLoS Pathog.* 5:e1000658.
- (179) PERALES C, LORENZO-REDONDO R, LOPEZ-GALINDEZ C, MARTINEZ MA, DOMINGO E (2010). Mutant spectra in virus behavior. *Future Virol.* 5:679–698.
- (180) PERALES C, AGUDO R, MANRUBIA SC, DOMINGO E (2011). Influence of mutagenesis and viral load on the sustained low-level replication of an RNA virus. *J. Mol. Biol.* 407:60–78.
- (181) PERALES C, HENRY M, DOMINGO E, WAIN-HOBSON S, VARTANIAN JP (2011). Lethal mutagenesis of foot-and-mouth disease virus involves shifts in sequence space. *J. Virol.* 85:12227–12240.
- (182) PERALES C., BEACH N.M., GALLEGO I., SORIA M.E., QUER J., ET AL. (2013). Response of hepatitis C virus to long-term passage in the presence of alpha interferon: multiple mutations and a common phenotype. *J. Virol.* 87, 7593-7607.
- (183) PEREZ-SAUTU U, COSTAFREDA MI, CAYLÀ J, TORTAJADA C, LITE J, BOSCH A, PINTÓ RM (2011). Hepatitis a virus vaccine escape variants and potential new serotype emergence. *Emerg. Infect. Dis.* 17:734–737.
- (184) PFEIFFER JK., KIRKEGAARD K (2005a). Increased fidelity reduces poliovirus fitness under selective pressure in mice. *PLoS pathog.* 1, 102-110.
- (185) PFEIFFER JK, KIRKEGAARD K (2005b). Ribavirin resistance in hepatitis C virus replicon containing cell lines conferred by changes in the cell line or mutations in the replicon RNA. *J. Virol.* 79:2346–2355.
- (186) POH WT, XIA E, CHIN-INMANU KN, WONG LP, CHENG AY, ET AL.. (2013). Viral quasispecies interference from 454 pyrosequencing. *BMC Bioinform* 14, 355.
- (187) POL S, ET AL. (1999). A randomized trial of ribavirin and interferon-alpha vs. interferon-alpha alone in patients with chronic hepatitis C who were non-responders to

- a previous treatment. *Multicenter Study Group under the Coordination of the Necker Hospital, Paris, France. J. Hepatol.* 31: 1–7.
- (188) QI X, GAO H, GAO Y, QIN L, WANG Y, GAO L, WANG X (2009). Naturally occurring mutations at residues 253 and 284 in VP2 contribute to the cell tropism and virulence of very virulent infectious bursal disease virus. *Antiviral Res.* 84:225–233.
- (189) QUER J, HUERTA R, NOVELLA IS, TSIMRING L, DOMINGO E, HOLLAND JJ (1996). Reproducible nonlinear population dynamics and critical points during replicative competitions of RNA virus quasispecies. *J. Mol. Biol.* 264:465–471.
- (190) QUER J, HERSHEY CL, DOMINGO E, HOLLAND JJ, NOVELLA IS (2001) Contingent neutrality in competing viral populations. *J. Virol.* 75: 7315–7320.
- (191) QUER J, GREGORI J, RODRIGUEZ-FRIAS F, BUTI M, MADEJON A, ET AL. (2014). High resolution hepatitis C virus subtyping using NS5B deep sequencing and phylogeny, an alternative to current methods. *J. Clin. Microbiol.* 53, 219–226.
- (192) QUIÑONES-MATEU ME, ARTS E. (2006). Virus fitness: concept, quantification, and application to HIV population dynamics. *Curr. Top. Microbiol. Immunol.* 299:83–140.
- (193) REGOES RR, BONHOEFFER S. (2005). The HIV coreceptor switch: a population dynamical perspective. *Trends Microbiol.* 13:269–277.
- (194) RIBEIRO RM, BONHOEFFER S (2000). Production of resistant HIV mutants during antiretroviral therapy. *Proc. Natl. Acad. Sci. U. S. A.* 97:7681–7686.
- (195) RICE SA, TAN CH, MIKKELSEN PJ, KUNG V, WOO J, TAY M, HAUSER A, MCDUGALD V, J.S. WEBB JS, KJELLEBERG S, (2009). The biofilm life cycle and virulence of *Pseudomonas aeruginosa* are dependent on a filamentous prophage. *ISME J.*, 3 (3), pp. 271–282
- (196) RICHMAN DD (2006). Antiviral drug resistance. *Antiviral Res* 71: 117–121.
- (197) ROBINSON L, WINDSOR M, MCLAUGHLIN K, HOPE J, JACKSON T, CHARLESTON B (2011). Foot-and-mouth disease virus exhibits an altered tropism in the presence of specific immunoglobulins, enabling productive infection and killing of dendritic cells. *J. Virol.* 85:2212–2223.
- (198) ROBINSON M, TIAN Y, DELANEY WE, GREENSTEIN AE. (2011). Preexisting drug-resistance mutations reveal unique barriers to resistance for distinct antivirals. *Proc. Natl. Acad. Sci. U. S. A.* 108:10290–10295.
- (199) ROUZINE IM, RODRIGO A, COFFIN JM. (2001). Transition between stochastic evolution and deterministic evolution in the presence of selection: general theory and application to virology. *Microbiol. Mol. Biol. Rev.* 65: 151–185.
- (200) RUIZ-JARABO CM, ET AL. (1999). Antigenic properties and population stability of a foot-and-mouth disease virus with an altered Arg-Gly-Asp receptor-recognition motif. *J. Gen. Virol.* 80:1899–1909.
- (201) RUIZ-JARABO CM, ARIAS A, BARANOWSKI E, ESCARMÍS C, DOMINGO E. (2000). Memory in viral quasispecies. *J. Virol.* 74:3543–3547.
- (202) RUIZ-JARABO CM, SEVILLA N, DÁVILA M, GÓMEZ-MARIANO G, BARANOWSKI E, DOMINGO E (2002). Duration and fitness dependence of quasispecies memory. *J. Mol. Biol.* 315:285–296
- (203) RUIZ-JARABO CM, LY C, DOMINGO E, DE LA TORRE JC. (2003). Lethal mutagenesis of the prototypic arenavirus lymphocytic choriomeningitis virus (LCMV). *Virology* 308:37–47.

- (204) RUIZ-JARABO CM, MILLER E, GÓMEZ-MARIANO G, DOMINGO E. (2003). Synchronous loss of quasispecies memory in parallel viral lineages: a deterministic feature of viral quasispecies. *J. Mol. Biol.* 333:553–563.
- (205) RUMBAUGH KP, DIGGLE SP, WATTERS CM, ROSS-GILLESPIE A, GRIFFIN AS, WEST SA, (2009). Quorum sensing and the social evolution of bacterial virulence. *Curr. Biol.* 19 (4), 341-345.
- (206) SAAKIAN DB, BIEBRICHER CK, HU CK (2009). Phase diagram for the Eigen quasispecies theory with a truncated fitness landscape. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 79:041905.
- (207) SALAZAR-GONZALEZ JF, SALAZAR MG, KEELE BF, LEARN GH, GIORGI EE, LI H, ET AL. (2009). Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *J. Exp. Med.* 206:1273–1289.
- (208) SANCHEZ G, BOSCH A, GOMEZ-MARIANO G, DOMINGO E, PINTO RM (2003). Evidence for quasispecies distributions in the human hepatitis A virus genome. *Virology* 315:34–42.
- (209) SANGER F., NICKLENS S., COULSON AR (1977). DNA sequencing with chainterminating inhibitors. *Proc Natl Acad Sci U S A.* 74(12), 5463-7
- (210) SANJUAN R., CUEVAS J.M., FURIO V., HOLMES E.C., MOYA A (2007). Selection for robustness in mutagenized RNA viruses. *PLoS Genet.* 3, e93.
- (211) SARRAZIN C, ZEUZEM S (2010). Resistance to direct antiviral agents in patients with hepatitis C virus infection. *Gastroenterology* 138:447– 462.
- (212) SEILER P, ET AL. (2000). Additive effect of neutralizing antibody and antiviral drug treatment in preventing virus escape and persistence. *J. Virol.* 74: 5896–5901.
- (213) SERRE JL (2006). *Génétique des populations*, Paris, Dunod. ISBN 2-10-049620 4
- (214) SEVILLA N, RUIZ-JARABO CM, GÓMEZ-MARIANO G, BARANOWSKI E, DOMINGO E (1998). An RNA virus can adapt to the multiplicity of infection. *J. Gen. Virol.* 79:2971–2980.
- (215) SHARP PM, ROBERTSON DI, GAO F, HAHN BH (1994). Origins and diversity of human immunodeficiency viruses. *AIDS* 1994; 8:S27-S42.
- (216) SHELDON J, BEACH N, MORENO E, GALLEGO I, PIÑEIRO D, ET AL. (2014). Increased replicative fitness can lead to decreased drug sensitivity of hepatitis C virus. *J. Virol.* 88, 12098-12111.
- (217) SHIROGANE Y, WATANABE S, YANAGI Y (2012). Cooperation between different RNA virus genomes produces new phenotype. *Nat. Commu.* 3, 1235.
- (218) SHORTER J (2010). Emergence and natural selection of drug-resistant prions. *Mol. Biosyst.* 6:1115–1130.
- (219) SIERRA M, AIRAKSINEN A, GONZÁLEZ-LÓPEZ C, AGUDO R, ARIAS A, DOMINGO E. (2007). Foot-and-mouth disease virus mutant with decreased sensitivity to ribavirin: implications for error catastrophe. *J. Virol.* 81:2012–2024
- (220) SMITH DS, GUO K, BARRETT BS, HEILMAN KJ, EVANS LH, HASENKRUG KJ, GREENE WC, SANTIAGO ML (2011). Noninfectious retrovirus particles drive the apobec3/ rfv3 dependent neutralizing antibody response. *PLoS Pathog.* 7:e1002284.
- (221) SMITH L, SANDERS JZ, KAISER RJ, HUGHES P, DODD C, CONNELL CR. ET AL. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature.* 321, 674-9

- (222) SOLÉ RV, RODRIGUEZ-CASO C, DEISBOECK TS, SALDANA J (2008). Cancer stem cells as the engine of unstable tumor progression. *J. Theor. Biol.*, 253 (4) (2008), pp. 629–637
- (223) SOLMONE M, VINCENTI D, PROSPERI MC, BRUSELLES A, IPPOLITO G ET AL., (2009). Use of massively parallel ultradeep pyrosequencing to characterize the genetic diversity of hepatitis B virus in drug-resistant and drug-naïve patients and to detect minor variants in reverse transcriptase and hepatitis B S antigen. *J. Virol.* 83, 1718-1726.
- (224) SOUSA R (1996). Structural and mechanistic relationships between nucleic acid polymerases. *Trends Biochem Sci* 1996; 21:186-90
- (225) SPYRAKIS F, BIDONCHANAL A, BARRIL X, LUQUE FJ (2011). Protein flexibility and ligand recognition: challenges for molecular modeling. *Curr. Top. Med. Chem.* 11:192–210.
- (226) STAVROU S, CRAWFORD D, BLOUCH K, BROWNE EP, KOHLI RM, ROSS SR (2014) Different modes of retrovirus restriction by human apobec3a and apobec3g in vivo. *PLoS Pathog* 10(5): e1004145.
- (227) STEWART PS, FRANKLIN MJ (2008). Physiological heterogeneity in biofilms. *Nat. Rev. Microbiol.*, 6 (3) (2008), pp. 199–210
- (228) SWETINA J, SCHUSTER P (1982). Self-replication with errors. A model for polynucleotide replication. *Biophys. Chem.* 16:329–345.
- (229) TABOGA O, TAMI C, CARRILLO E, NÚÑEZ JI, RODRÍGUEZ A, SAÍZ JC, BLANCO E, ET AL. (1997). A large-scale evaluation of peptide vaccines against foot-and-mouth disease: lack of solid protection in cattle and isolation of escape mutants. *J. Virol.* 71:2606–2614.
- (230) TAKEUCHI N, HOGEWEG P (2007). Error-threshold exists in fitness landscapes with lethal mutants. *BMC Evol. Biol.* 7:15. (Author reply, 7:15.)
- (231) TAMI C, TABOGA O, BERINSTEIN A, NÚÑEZ JI, PALMA EL, DOMINGO E, SOBRINO F, CARRILLO E (2003). Evidence of the coevolution of antigenicity and host cell tropism of foot-and-mouth disease virus in vivo. *J. Virol.* 77:1219–1226.
- (232) TANNER EJ, LIU H-M, OBERSTE MS, PALLANSCH M, COLLET MS, KIRGEGAARD K, (2014). Dominant drug targets suppress the emergence of antiviral resistance. *eLife* 3 :e03830
- (233) TAPIA N, FERNÁNDEZ G, PARERA M, GÓMEZ-MARIANO G, CLOTET B, QUIÑONES-MATEU M, DOMINGO E, MARTÍNEZ MA (2005). Combination of a mutagenic agent with a reverse transcriptase inhibitor results in systematic inhibition of HIV-1 infection. *Virology* 338:1–8.
- (234) TENG MN, OLDSTONE MB, DE LA TORRE JC. (1996). Suppression of lymphocytic choriomeningitis virus-induced growth hormone deficiency syndrome by disease-negative virus variants. *Virology* 223:113–119.
- (235) THOMSON MM, PEREZ-ALVAREZ L, NAJER AR (2002). Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. *Lancet Infect. Dis.* 2:461–471
- (236) TILMAN D (1994). Competition and biodiversity in spatially structured habitats. *Ecology* 75:2–16.
- (237) TOPFER A, MARSCHALL T, BULL RA, LUCIANI F, SCHONHUTH A, ET AL. (2014). Viral quasispecies assembly via maximal clique enumeration. *PLoS Comput. Biol.* 10, e1003515.
- (238) TZENG SR, KALODIMOS CG. (2011). Protein dynamics and allostery: an NMR view. *Curr. Opin. Struct. Biol.* 21:62–67.

- (239) VALENTINE RC, WARD R, STRAND M. 1969. The replication cycle of RNA bacteriophages. *Adv. Virus Res.* 15:1–59.
- (240) VAN VAERENBERGH K, HARRER T, SCHMIT JC, CARBONEZ A, FONTAINE E, KUROWSKI M, GRÜNKE M ET AL. (2002). Initiation of HAART in drug-naive HIV type 1 patients prevents viral breakthrough for a median period of 35.5 months in 60% of the patients. *AIDS Res. Hum. Retroviruses* 18:419–426.
- (241) VANDERBROUCKE I, VAN MARCK H, VERSHASSELT P, THYS K, MOSTMANS W, ET AL. (2011). Minor variant detection in amplicons using 454 massive parallel pyrosequencing: experiences and considerations for successful applications. *BioTechniques* 51, 167-177.
- (242) VERDAGUER N, SEVILLA N, VALERO ML, STUART D, BROCCHE, ANDREU D ET AL.(1998). A similar pattern of interaction for different antibodies with a major antigenic site of foot-and-mouth disease virus: implications for intratypic antigenic variation. *J. Virol.* 72:739 –748.
- (243) VIGNUZZI M, STONE JK, ANDINO R (2005). Ribavirin and lethal mutagenesis of poliovirus: molecular mechanisms, resistance and biological implications. *VirusRes* 107: 173–181.
- (244) VIGNUZZI M, STONE JK, ARNOLD JJ, CAMERON CE, ANDINO R (2006). Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 439:344 –348.
- (245) VIGNUZZI M, ANDINO R (2010). Biological implications of picornavirus fidelity mutants, p 213–228. In Ehrenfeld E, Domingo E, Roos RF (ed), *The picornaviruses*. ASM Press, Washington, DC.
- (246) VLASAK M, GOESLER I, BLAAS D (2005). Human rhinovirus type 89 variants use heparan sulfate proteoglycan for cell attachment. *J. Virol.* 79:5963–5970.
- (247) VOELKERDING KV, DAMES S, DURTSCHI JD (2010). Next Generation Sequencing for clinical diagnostics-principles and application to targeted resequencing for hypertrophic cardiomyopathy: a paper from the 2009 william beaumont hospital symposium on molecular pathology. *J Mol Diagn.* 12(5), 539-51
- (248) VON KLEIST M, ET AL. (2011). HIV quasispecies dynamics during pro-active treatment switching: impact on multi-drug resistance and resistance archiving in latent reservoirs. *PLoS One* 6:e18204.
- (249) WAINBERG MA, ET AL. 1996. Enhanced fidelity of 3TC-selected mutant HIV-1 reverse transcriptase. *Science* 271:1282–1285.
- (250) WARGO AR, KURATH G (2012). Viral fitness: definition, measurement and current insights. *Curr. Opin. Virol.* 2, 538-545.
- (251) WEISSMANN C, BILLETTER MA, GOODMAN HM, HINDLEY J, WEBER H (1973). Structure and function of phage RNA. *Annu. Rev. Biochem.* 2:303–328.
- (252) WEISSMANN C (1974). The making of a phage. *FEBS Lett.* 40(Suppl.):S10–S18.
- (253) WEISSMANN C, LI J, MAHAL SP, BROWNING S (2011). Prions on the move. *EMBO Rep.* 12:1109 –1117.
- (254) WHITNEY JB, OLIVEIRA M, DETORIO M, GUAN Y, WAINBERG MA (2002). The M184V mutation in reverse transcriptase can delay reversion of attenuated variants of simian immunodeficiency virus. *J. Virol.* 76: 8958 – 8962.
- (255) WILKE CO, RONNEWINKEL C, MARTINETZ T (2001). Dynamic fitness landscapes in molecular evolution. *Phys. Rep.* 349:395– 446.

- (256) WILKE CO, WANG JL, OFRIA C, LENSKI RE, ADAMI C (2001) Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* 412: 331–333.
- (257) WILKE CO, NOVELLA IS (2003). Phenotypic mixing and hiding may contribute to memory in viral quasispecies. *BMC Microbiol.* 3:11.
- (258) WILKE CO REISSIG DD, NOVELLA IS (2004). Replication at periodically changing multiplicity of infection promotes stable coexistence of competing viral populations. *Evolution* 58, 900-905
- (259) WILKE CO (2005). Quasispecies theory in the context of population genetics. *BMC Evol. Biol.* 5, 44.
- (260) WILSON ED (1998). Consilience. Abacus, Little, Brown and Co. U.K.
- (261) WIMMER E, PAUL AV (2010). The making of a picornavirus genome, p 33–58. In Ehrenfeld E, Domingo E, Roos RP (ed), *The picornaviruses*. ASM Press, Washington, DC.
- (262) WOESE CR (2002). On the evolution of cells. *Proc. Natl. Acad. Sci. U. S. A.* 99:8742– 8747
- (263) WRIGHT C.F., MORELLI M.J., THEBAUD G., KNOWLES N.J., HERZYK P., ET AL. 2011. Beyond the consensus : dissecting within-host viral population diversity of foot-and-mouth disease virus by using next-generation genome sequencing. *J. Virol.* 85, 2266-2275.
- (264) YAMADA S, HATTA M, STAKER BL, WATANABE S, IMAI M, SHINYA K, SAKAI-TAGAWA Y, ET AL. (2010). Biological and structural characterization of a host-adapting amino acid in influenza virus. *PLoS Pathog.* 6:e1001034.
- (265) YIN J. (1993). Evolution of bacteriophage T7 in a growing plaque. *J. Bacteriol.* 175:1272–1277. 885.
- (266) YOUNG KC, LINDSAY KL, LEE KJ, LIU WC, HE JW, MILSTEIN SL, LAI MM (2003). Identification of a ribavirin-resistant NS5B mutation of hepatitis C virus during ribavirin monotherapy. *Hepatology* 38:869–878
- (267) YOUNGNER JS, WHITAKER-DOWLING P (1999). Interference, p 850–854. In Granoff A, Webster RG (ed), *Encyclopedia of virology*, vol 2. Academic Press, San Diego, CA.
- (268) YUSTE E, SÁNCHEZ-PALOMINO S, CASADO C, DOMINGO E, LÓPEZ-GALÍNDEZ C (1999). Drastic fitness loss in human immunodeficiency virus type 1 upon serial bottleneck events. *J. Virol.* 73:2745–2751.
- (269) ZAGORDI O, GEYRHOFFER L, ROTH V, BEERENWINKEL N (2010). Deep sequencing of a genetically heterogeneous sample: local haplotype reconstruction and read error correction. *J Comput Biol.* 17(3), 417-28
- (270) ZAGORDI O, KLEIN R DAUMER M, BEERENWINKEL N (2010). Error connection of next-generation sequencing data and reliable estimation of HIV quasispecies. *Nucleic Acids RES.* 38, 7400-7409
- (271) ZEIN NN (2000). Clinical significance of hepatitis C virus genotypes. *Clin Microbiol Rev* 13(2),223-35.

## WEBOGRAPHIE

(1') <http://acces.ens-lyon.fr/evolution/logiciels/anagene/documentation/5%20-%20Themes%20Premiere%20-%20Genotype,%20phenotype,%20environnement.pdf>

(2') [http://www.google.fr/imgres?imgurl=http%3A%2F%2Ffr.cdn.v5.futura-sciences.com%2Fbuilds%2Fimages%2Frte%2FRTEmagicC\\_7289\\_codegenetique\\_dr\\_txdam19416\\_9dd4e4.gif&imgrefurl=http%3A%2F%2Fwww.futura-sciences.com%2Fmagazines%2Fsante%2Finfos%2Fdossiers%2Fd%2Fgenetique-gene-adn-proteines-1130%2Fpage%2F5%2F&h=377&w=400&tbnid=M9QIIfwKF-XRGM%3A&docid=iO8na4G\\_RuFp2M&ei=8Iw6VsfDN8n8adnZuIgJ&tbm=isch&iact=rc&uact=3&dur=707&page=1&start=0&ndsp=18&ved=0CEUQrQMwBWoVChMIx\\_vbyO73yAIVSX4aCh3ZLA6R](http://www.google.fr/imgres?imgurl=http%3A%2F%2Ffr.cdn.v5.futura-sciences.com%2Fbuilds%2Fimages%2Frte%2FRTEmagicC_7289_codegenetique_dr_txdam19416_9dd4e4.gif&imgrefurl=http%3A%2F%2Fwww.futura-sciences.com%2Fmagazines%2Fsante%2Finfos%2Fdossiers%2Fd%2Fgenetique-gene-adn-proteines-1130%2Fpage%2F5%2F&h=377&w=400&tbnid=M9QIIfwKF-XRGM%3A&docid=iO8na4G_RuFp2M&ei=8Iw6VsfDN8n8adnZuIgJ&tbm=isch&iact=rc&uact=3&dur=707&page=1&start=0&ndsp=18&ved=0CEUQrQMwBWoVChMIx_vbyO73yAIVSX4aCh3ZLA6R)

(3') [http://www.pasteur.fr/infosci/biblio/expos/evolution/textes\\_evolution/V5Le%20mutatinnisme.pdf](http://www.pasteur.fr/infosci/biblio/expos/evolution/textes_evolution/V5Le%20mutatinnisme.pdf)



## ANNEXES

### ANNEXE 1 : FREQUENCES DE MUTATIONS CHEZ DES VIRUS A ARN (40)

Table 1  
Mutation frequencies<sup>a</sup> of selected RNA viruses

Virus	Mutation frequency	Assay	References
PV-1	$5.4 \times 10^{-3}$ – $7 \times 10^{-4}$	Biochemical assay <sup>b</sup>	Ward et al. (1988)
PV-1	$2.1 \times 10^{-4}$	GuaD–GuaR <sup>c</sup>	De la Torre et al. (1990)
PV-1	$7.8 \times 10^{-5}$ – $4 \times 10^{-8}$	GuaD–GuaR	Pincus et al. (1986)
Coxsackie A9	$1 \times 10^{-4}$	2-HBB resistance <sup>d</sup>	Eggers and Tamm (1965)
Influenza A	$1.5 \times 10^{-5}$	Plaque assay <sup>e</sup>	Parvin et al. (1986)
VSV	$1 \times 10^{-3}$ – $1 \times 10^{-4}$	Plaque assay	Steinhauer et al. (1989)
PV-1	$2.1 \times 10^{-4}$	Sequencing <sup>f</sup>	Crotty et al. (2001)
HAV	$1 \times 10^{-3}$ – $1 \times 10^{-4}$	Sequencing	Sanchez et al. (2003)
TMV	$3 \times 10^{-4}$	Sequencing	Kearney et al. (1999)
HCV	$\sim 1 \times 10^{-3}$	Sequencing	Young et al. (2003)
FMDV	$1.4 \times 10^{-4}$	Sequencing	Airaksinen et al. (2003)
HTNV	$1.1 \times 10^{-3}$	Sequencing	Severson et al. (2003)
LCMV	$3.6 \times 10^{-4}$	Sequencing	Ruiz-Jarabo et al. (2003)

<sup>a</sup> Mutation frequency is defined as the number of mutations per replication event.

<sup>b</sup> Amount of a noncomplementary nucleotide incorporated divided by the total amount of complementary and noncomplementary nucleotide incorporated using homopolymeric RNA templates.

<sup>c</sup> Reversion of guanidine dependant to guanidine resistant.

<sup>d</sup> Development of resistance to 2-( $\alpha$ -hydroxybenzyl)-benzimidazole (2-HBB).

<sup>e</sup> Mutations were determined by sequencing different plaques developed after infection with a single clone.

<sup>f</sup> Viral population samples from the host organism or cell culture were isolated, amplified and sequenced.

## ANNEXE 2 : IMPACT DE MUTATIONS RETOURS (32)

### Box 1. Back Mutation

Our model of two genotypes ignored back mutation, hence  $A_1$  could not be regenerated once lost from the population (see Figure 1). The absence of back mutations is often assumed in models of population genetics (e.g., the “infinite alleles” and “infinite sites” models of population genetics [42,43]). This assumption offers analytical convenience and is perhaps an approximate interpretation of asexual genomes, in which an “allele” can be considered the entire suite of mutations in a genome; a new mutation, which can occur anywhere in the genome, is unlikely to recreate a former allelic state. Nonetheless, since back mutations can occur in many systems, it is useful to study their effect on the error threshold. In particular, they present a special challenge to the concept of an error threshold: when back mutations occur, the wild-type genome is never strictly lost in a deterministic sense. What, then, is the counterpart to a strict error threshold? To approach this question, we return to the simple model in Figure 6.

At this point we use simple matrix algebra to explore the problem. Assuming discrete time steps, let  $n_i$  be the number of individuals with genotype  $A_i$ ; mutation precedes reproduction, so the fitness  $w_i$  accrues to an individual after its genotype experiences any mutation (the model is just as easily constructed with mutation occurring after reproduction). Two equations describe the system:

$$n'_1 = n_1(w_1[1 - \mu_1]) + n_2(w_1\mu_3) \quad (1)$$

$$n'_2 = n_1(w_2\mu_1) + n_2(w_2[1 - \mu_2]) \quad (2)$$

The system can then be described with the aid of the transition matrix:

$$M = \begin{pmatrix} w_1(1 - \mu_1) & w_1\mu_3 \\ w_2\mu_1 & w_2(1 - \mu_2) \end{pmatrix} \quad (3)$$

so that  $(n'_1, n'_2) = (n_1, n_2)M$ . The only new parameter with respect to our earlier model is  $\mu_3$ , which allows for back mutation and cannot exceed  $\mu_2$ . If we set  $\mu_3 = 0$ , then these equations describe our earlier model exactly (in Figure 1). To evaluate the long term behavior of the system, we consider  $n_1(t)$  and  $n_2(t)$  as  $t$  becomes arbitrarily large; that is, we calculate  $(n_1, n_2) M^t$  for large  $t$ . The standard method for evaluating long-term behavior in such a linear system is to consider the eigenvalues of the system, found as the values of  $\lambda$  satisfying

## ANNEXE 2 (suite) : IMPACT DE MUTATIONS RETOURS (32)

$$\lambda^2 - \lambda(w_1[1 - \mu_1] + w_2[1 - \mu_2]) + (w_1w_2[1 - \mu_1][1 - \mu_2] - w_1w_2\mu_1\mu_2) = 0 \quad (4)$$

To establish a baseline, we first assume no back mutation ( $\mu_3 = 0$ ). Equation 4 then yields the solutions

$$\lambda_1 = w_1(1 - \mu_1) \quad (5)$$

$$\lambda_2 = w_2(1 - \mu_2) \quad (6)$$

These two eigenvalues are the respective replacement rates of  $A_1$  and  $A_2$  described in Figure 3. Importantly,  $\lambda_1$  is larger for some mutation rates, but  $\lambda_2$  is larger for others, and it is the intersection of the eigenvalues that gives rise to the error threshold. This simple picture appears to change fundamentally with even a small level of back mutation. When  $\mu_3 > 0$ , the two eigenvalues become

$$\lambda_1 = \frac{x + y + \sqrt{(x - y)^2 + 4w_1w_2\mu_1\mu_3}}{2} \quad (7)$$

$$\lambda_2 = \frac{x + y - \sqrt{(x - y)^2 + 4w_1w_2\mu_1\mu_3}}{2} \quad (8)$$

where  $x = w_1(1 - \mu_1)$  and  $y = w_2(1 - \mu_2)$ .

Whereas a transition from  $\lambda_1$  to  $\lambda_2$  as the dominant eigenvalue occurred at  $x = y$  in the absence of back mutation, now  $\lambda_1$  remains superior at all values of  $x$  and  $y$ . As  $\mu_3$  approaches arbitrarily close to zero, the eigenvalues approach each other but do not cross. Thus, there is no strict error threshold unless  $\mu_3$  equals zero exactly. Yet as  $\mu_3$  approaches zero, the system appears more and more like that with no back mutation (Figure 7). So the mathematical discontinuity between systems with and without back mutation does not obviously translate into a meaningful biological distinction, but large amounts of back mutation do change the picture.



## ANNEXE 3 : EXEMPLES D'APPLICATION DES NGS EN VIROLOGIE (20)

Table 2 | Applications of 454/Roche pyrosequencing and Illumina NGS technologies in clinical virology.

Virus	Study	NGS platform	NGS approach	Basis of analysis	References
CMV	Epidemiology	454/Roche	Amplicon-based	Reads	Gorzer et al., 2010
CMV	Epidemiology	454/Roche	Shotgun	Consensus sequence	Jung et al., 2011
EBV	Epidemiology	Illumina	Shotgun	SNV consensus sequence	Liu et al., 2011
EBV	Epidemiology	Illumina	Shotgun (amplicons)	SNV	Kwok et al., 2012
HBV	Drug resistance	454/Roche	Amplicon-based	Reads, SNV	Solmoné et al., 2009; Homs et al., 2011; Rodríguez-Frías et al., 2012
HBV	Drug resistance	454/Roche	Amplicon-based	SNV	Margadon-Thermet et al., 2009; Yu et al., 2012; Sede et al., 2012
HBV	Drug resistance	Illumina	Shotgun	SNV	Nishijima et al., 2012
HCV	Drug resistance	454/Roche	Amplicon-based	Reads	Bolico et al., 2012; Fonseca-Coronado et al., 2012
HCV	Drug resistance	Illumina	Shotgun (cDNA)	SNV	Hiraga et al., 2011
HCV	Drug resistance	454/Roche	Shotgun (amplicons)	SNV consensus sequences	Lauck et al., 2012
HCV	Drug resistance	Illumina	Paired-end (amplicons)	SNV	Nsai et al., 2011
HCV	Drug resistance	454/Roche	Amplicon-based	SNV	Powdrill et al., 2011
HCV	Epidemiology	454/Roche	Amplicon-based	Reads	Escobar-Gutiérrez et al., 2012; Forbi et al., 2012
HCV	Epidemiology	Illumina	Shotgun (cDNA)	SNV consensus sequences	Ninomiya et al., 2012
HIV	Drug resistance	454/Roche	Amplicon-based	SNV	Hoffmann et al., 2007; Wang et al., 2007; Mitsuya et al., 2008; Le et al., 2009; Simen et al., 2009; Virghese et al., 2009; Latalade et al., 2010, 2012; Alteri et al., 2011; D'Aquila et al., 2011; Debbel et al., 2011; Ganala et al., 2011; Ji et al., 2011; Kozal et al., 2011; Moorthy et al., 2011; Stebl et al., 2011; Fisher et al., 2012; Messiaen et al., 2012
HIV	Drug resistance	454/Roche	Amplicon-based	Reads, SNV	Hedskog et al., 2010; Ji et al., 2010; Mild et al., 2011; Mukherjee et al., 2011; Armenia et al., 2012
HIV	Epidemiology	454/Roche	Shotgun (amplicons)	Consensus sequence	Bruselles et al., 2009
HIV	Epidemiology	454/Roche	Amplicon-based	Consensus sequence	Eshleman et al., 2011
HIV	Epidemiology	454/Roche	Amplicon-based	Reads	Redd et al., 2012
HIV	Tropism	454/Roche	Amplicon-based	Reads	Archer et al., 2009; Rozen et al., 2009; Abbate et al., 2011; Swenson et al., 2010; Vandembroucke et al., 2010; Baatz et al., 2011; Bunnik et al., 2011; Raymond et al., 2011; Safioui et al., 2011; Svidner et al., 2011; Swenson et al., 2011a, b; Vandekerckhove et al., 2011
Influenza A virus	Epidemiology	Illumina	Shotgun (amplicons)	SNV	Kuroda et al., 2010; Kampmann et al., 2011
Influenza A virus	Epidemiology	454/Roche	Shotgun (amplicons)	SNV	Bartolini et al., 2011
Influenza A virus	Epidemiology	454/Roche	Shotgun	Reads	Lorusso et al., 2011
norovirus	Epidemiology	454/Roche	Shotgun (amplicons)	SNV haplotype reconstruction	Bulle et al., 2012
rhinovirus	Epidemiology	Illumina	Shotgun (amplicons)	SNV consensus sequences	Tapparel et al., 2011
rotavirus	Epidemiology	454/Roche	Shotgun (cDNA)	Consensus sequences	Jere et al., 2011
VZV	Epidemiology	454/Roche	Shotgun (amplicons)	Consensus sequences	Zell et al., 2012

BAL, bronchoalveolar lavage; CMV, cytomegalovirus; EBV, Epstein Barr virus; HBV, hepatitis B virus; HCV, hepatitis C virus; HIV, human immunodeficiency virus; SNV, single nucleotide variant; VZV, varicella zoster virus

## ANNEXE 4 : CARACTERISTIQUES DES PLATE-FORMES NGS (247)

**Table 1.** Characteristics of Next Generation Sequencing Platforms

Platform	Template preparation	Chemistry	Read length (bases)	Run time (days)*	Gb per run <sup>†</sup>
Roche 454					
GS FLX Titanium	Clonal-ePCR	Pyrosequencing	400 <sup>‡</sup>	0.42	0.40–0.60
GS Junior	Clonal-ePCR	Pyrosequencing	400 <sup>‡</sup>	0.42	0.035
Illumina					
HiSeq 2000	Clonal Bridge Amplification	Reversible Dye Terminators	35–100	2–4	30–100
Genome Analyzer IIX	Clonal Bridge Amplification	Reversible Dye Terminators	35–100	2–4	9–25
Genome Analyzer IIE	Clonal Bridge Amplification	Reversible Dye Terminators	35–100	2–5	3.5–10
iScanSQ	Clonal Bridge Amplification	Reversible Dye Terminators	35–75	2.5–5	4–10
Life Technologies					
SOLiD 4	Clonal-ePCR	Oligonucleotide Probe Ligation	35–50	4–7	35–50
Helicos Biosciences					
HeliScope	Single Molecule	Reversible Dye Terminators	35 <sup>‡</sup>	8	25
Pacific Biosciences					
SMRT	Single Molecule	Phospholinked Fluorescent Nucleotides	800–1000	0.02	Pending

Run times\* and gigabase (Gb) output<sup>†</sup> per run for single-end sequencing are denoted by a double asterisk and a single dagger, respectively. Run times and outputs approximately double when performing pair-end sequencing. <sup>‡</sup>Average read lengths for the Roche 454 and Helicos Biosciences platforms.

## ANNEXE 5 : TAUX D'EVOLUTION POUR LES VIRUS A ARN (71)

Table 6.3. Rates of evolution for RNA viruses

VIRUS	GENOMIC REGION	SUBSTITUTIONS PER NUCLEOTIDE PER YEAR	REFERENCE
Foot-and-mouth disease virus	Average over genome	$<4 \times 10^{-4}$ - $4 \times 10^{-2}$	Sobrino et al 1986; Domingo et al 1990
	VP1 <sup>a</sup>	$9.0 \times 10^{-4}$ - $7.4 \times 10^{-2}$	Gebauer et al 1988
	VP1	$6.5 \times 10^{-3}$	Villaverde et al 1991
Poliovirus	VP1	$1.4 \times 10^{-4}$	Martinez et al 1991
	Average over genome	$6.9 \times 10^{-4}$ - $1.4 \times 10^{-2}$	Kew et al 1990
Enterovirus 70	VP1 <sup>b</sup>	$3 \times 10^{-2}$	Kew et al 1998
	VP1	$5 \times 10^{-3}$	Takeda et al 1994
Eastern equine encephalomyelitis virus	26S RNA	$1.4 \times 10^{-2}$	Weaver et al 1991
Hepatitis C virus	Average over genome <sup>c</sup>	$1.4 \times 10^{-3}$	Okamoto et al 1992
	Structural proteins <sup>b</sup>	$2 \times 10^{-3}$	Smith et al 1997
	5'NC, C, E, NS1, 2, 3, 5 <sup>d</sup>	$9 \times 10^{-4}$	Abe et al 1992
Hepatitis G virus	5'NC <sup>d,e</sup>	$2 \times 10^{-3}$	Ogata et al 1991
	NS3 <sup>d,e</sup>	$4 \times 10^{-4}$	Gimenez-Barcons et al, 1998
Hepatitis delta virus	Average over genome	$3.9 \times 10^{-4}$	Nakao et al 1997
	Delta antigen <sup>d</sup>	$6 \times 10^{-4}$	Imazequi et al 1990
Hemorrhagic septicemia virus (fish rhabdovirus)	Average over genome	$3.0 \times 10^{-2}$ - $3.0 \times 10^{-2}$	Lee et al 1992
	G	$1.6 \times 10^{-4}$ (slow phase) $4.5 \times 10^{-3}$ (fast phase)	Benmansour et al 1997
Influenza virus, Type A	NS	$1.9 \times 10^{-3}$	Smith & Palese 1988
	HA(H3)	$6.7 \times 10^{-3}$	Smith & Palese 1988
	HA(H3)	$5.7 \times 10^{-3}$	Fitch et al, 1997
	HA(H1) <sup>f</sup>	$5.8 \times 10^{-3}$ - $1.7 \times 10^{-2}$	Rocha et al 1991
	NA(N2)	$3.2 \times 10^{-3}$	Smith & Palese 1988
	M1	$8 \times 10^{-4}$	Ito et al 91
	M2	$1.4 \times 10^{-3}$	Ito et al 91
	NP	$1.6 \times 10^{-3}$ - $2.2 \times 10^{-3}$	Altmuller et al 1989
			Gorman et al 1990
Influenza virus, Type B	NP <sup>g</sup>	$3.5 \times 10^{-3}$ - $2.4 \times 10^{-2}$	Rocha et al 1991
	HA	$3.4 \times 10^{-3}$ - $4.7 \times 10^{-3}$	Rota et al, 1990
Visna virus	HA, NS	$1.1 \times 10^{-3}$	Yamashita et al 1988
Visna-related	Average over genome	$1.7 \times 10^{-3}$	Braun et al 1987
	Average over genome (nonsynonymous subst.)	$8 \times 10^{-4}$	Querat et al 1990

**Table 6.3. continued**

VIRUS	GENOMIC REGION	SUBSTITUTIONS PER NUCLEOTIDE PER YEAR	REFERENCE
Human immuno-deficiency virus	env	$3.2 \times 10^{-4}$ - $1.6 \times 10^{-2}$	Hahn et al 1986
Simian immuno-deficiency virus	gag	$3.7 \times 10^{-4}$ - $1.8 \times 10^{-1}$	Burns and Desrosiers 1991
	gp120 <sup>b</sup>	$8.5 \times 10^{-3}$	
Equine infectious anemia virus	Env <sup>h</sup>	$10^{-1}$ - $10^{-2}$	Clements et al 1988
HTLV-II	LTR <sup>i</sup>	$1 \times 10^{-4}$ - $3 \times 10^{-5}$	Salemi et al 1998
Canine parvovirus	VP1-VP2	$1.7 \times 10^{-4}$	Parrish et al 1991
Polyomavirus JC	Intergenic region	$1 \times 10^{-7}$ - $3 \times 10^{-7}$	Sugimoto et al 1997
Papillomavirus	E6	$3 \times 10^{-8}$	Van Ranst et al 1995
Cellular genes of hosts		$10^{-8}$ - $10^{-9}$	Britten, 1986
			Weissmann and Weber, 1986

<sup>a</sup> Persistent infection established experimentally *in vivo*  
<sup>b</sup> Synonymous mutations; immunodeficient patient  
<sup>c</sup> Human isolate administered to chimpanzee  
<sup>d</sup> Inpatient evolution  
<sup>e</sup> Average of many patients coinfecting with HIV-1  
<sup>f</sup> Immunodeficiency child persistently infected with IV  
<sup>g</sup> Infections with a molecular clone of SIV  
<sup>h</sup> Chronically infected ponies  
<sup>i</sup> Drug users

## ANNEXE 6 : FREQUENCES DE RESISTANCE AUX ANTICORPS (72)

Virus	Monoclonal Antibody-resistant Mutant (MARM) Frequencies	References
Poliovirus	$10^{-4}$ to $10^{-5}$	Emini et al. (1982) and Minor et al. (1983, 1986)
Mengovirus	$3 \times 10^{-3}$ to $5 \times 10^{-3}$	Boege et al. (1991)
Foot-and-mouth disease virus	$10^{-4}$ to $10^{-5}$ (continuous epitopes) $10^{-4}$ to $10^{-7}$ (discontinuous epitopes)	Martínez et al. (1991a) Lea et al. (1994)
Rhinovirus	$10^{-4}$ to $10^{-5}$	Sherry et al. (1986)
Hepatitis A virus	$3 \times 10^{-3}$	Stapleton and Lemon (1987)
Vesicular stomatitis virus	$0.5 \times 10^{-4}$ to $1 \times 10^{-4}$	Holland et al. (1990)
Rabies virus	$10^{-4}$	Wiktor and Koprowski (1980)
Measles virus	$9 \times 10^{-5}$	Schrag et al. (1999)
Sindbis virus	$10^{-3}$ to $10^{-5}$	Stec et al. (1986)
Canine parvovirus	$10^{-34}$ to $10^{-54}$	Smith and Inglis (1987)
Herpes simplex virus	$1 \times 10^{-5}$	Smith and Inglis (1987)



## ANNEXE 7 : MODIFICATIONS DU TROPISME CELLULAIRE (72)

<b>Observation</b>	<b>References</b>
Host restriction of avian polyomavirus budgerigar fledgling disease virus	Stoll et al. (1994)
Colonic tropism and persistence of murine norovirus	Nice et al. (2013)
N versus B tropism of murine leukemia virus	Jung and Kozak (2000)
Tissue tropism of adeno-associated viruses	Wu et al. (2006)
Cellular tropism of feline immunodeficiency virus	Verschöor et al. (1995), Vahlenkamp et al. (1997), Lerner and Elder (2000)
Loss of enteric tropism of transmissible gastroenteritis coronavirus by two amino acid substitutions in the spike protein	Ballesteros et al. (1997)
HIV-1 tropism by a single amino acid substitution in gp120	Takeuchi et al. (1991), Boyd et al. (1993)
SARS coronavirus recognition of ACE2 receptor (see also text)	Li et al. (2005)
Receptor preferences of herpes simplex virus	Spear et al. (2000)
Hemagglutinin residues in influenza virus tropism (see also text)	Rogers et al. (1983), Connor et al. (1994)
Substitutions in poliovirus capsid expand receptor recognition	Colston and Racaniello (1995)
Conversion of encephalomyocarditis D into a diabetogenic variant through altered cell tropism	Bae and Yoon (1993)
High-affinity binding of measles virus to CD46	Hsu et al. (1998)
Decreased neurovirulence of Sindbis virus through impaired receptor recognition in neural cells	Tucker and Griffin (1991), Lee et al. (2002)
Parvovirus host range (see also text and Figure 4.5)	Hueffer et al. (2003)
Change in receptor recognition by foot-and-mouth disease virus <i>in vivo</i> and in cell culture (see also text and Figure 4.6)	Baranowski et al. (2000, 2001), Ruiz-Jarabo et al. (2004)

## ANNEXE 8 : EXEMPLES DE CHEVAUCHEMENT DE SITES ANTIGENIQUES ET DE SITE DE RECONNAISSANCE DES RECEPTEURS CHEZ LES VIRUS A ARN (72)

**Table 4.3 Examples of Overlap Between Antigenic Sites and Receptor Recognition Sites in RNA Viruses**

Virus	Observation	References
Poliovirus	Receptor recognition influenced by residues of antigenic sites Critical role of VP1 BC loop in receptor interaction	Murray et al. (1988), Harber et al. (1995) Yeates et al. (1991)
Human rhinovirus	Neutralizing antibody to HRV14 penetrates the receptor-binding canyon Exposed VP1 BC- and HI-loops covered by footprint of very low-density lipoprotein receptor	Smith et al. (1996) Hewat et al. (2000)
Theiler's encephalomyelitis virus	Neutralizing antibodies map close to putative receptor-binding site Substitutions of adaptation to some cells map in antigenic sites	Sato et al. (1996) Jnaoui and Michiels (1998)
Foot-and-mouth disease virus	Overlap of integrin- and antibody-binding sites (Additional studies presented in the text)	Verdaguer et al. (1995)
Human influenza virus	Amino acid residues of the sialic acid-binding pocket are accessible to neutralizing antibodies (Additional studies presented in the text)	Stewart and Nemerow (1997)
Newcastle disease virus	Monoclonal antibodies to HN glycoprotein prevent virus attachment	Iorio et al. (1989)
Rabies virus	Anti-idiotypic antibodies bind to cellular receptors Residues critical for neurotropism are involved in antibody binding	Hanham et al. (1993) Coulon et al. (1998)
Bovine viral diarrhea virus	Anti-idiotypic antibodies bind to cellular receptors	Xue and Minocha (1993), Minocha et al. (1997)
Dengue virus	Residues critical for mouse neurovirulence are involved in antibody binding	Hiramatsu et al. (1996)
Yellow fever virus	Residues critical for neurotropism are involved in antibody binding	Jennings et al. (1994)
Murine coronavirus	Overlap between epitopes and receptor-binding sites	Kubo et al. (1993, 1994)
Middle East respiratory syndrome (MERS) coronavirus	Antibodies bind to receptor recognition site	Ying et al. (2014)
Sindbis virus	Anti-idiotypic antibodies bind to cellular receptors	Ubol and Griffin (1991), Wang et al. (1991), Strauss et al. (1994)
Ross River virus	Binding of antibodies to cell-receptor recognition regions	Smith et al. (1995)
Reovirus	Anti-idiotypic antibodies bind to cellular receptors	Co et al. (1985), Gaulton et al. (1985), Williams et al. (1988, 1989, 1991)
Bluetongue virus	Anti-idiotypic antibodies bind to cellular receptors	Xu et al. (1997)

## ANNEXE 9 : LOGICIELS DE RECONSTRUCTION D'UNE QUASI-ESPECE VIRALE (19)

**Table 1**

**Computational methods for viral quasispecies reconstruction.**

Method	Error correction	Global assembly	Confidence values	Software available	Applications	References
ShoRAH	Yes	Yes	Yes	Yes <sup>a</sup>	HIV, HCV	[37**,38,39*]
ViSpA	Yes	Yes	No	Yes <sup>b</sup>	HIV, HCV	[43*,44]
Jojic <i>et al.</i>	Yes	Yes	No	No	HIV	[40]
AmpliconNoise	Yes	No	No	Yes <sup>c</sup>	16S rRNA, HCV	[31,33*]
PredictHaplo	Yes	Yes	Yes	Yes <sup>d</sup>	HIV	[41]
Prosperi <i>et al.</i>	No	Yes	No	No	HBV	[42]

Listed are all published methods to date that have been applied to at least one real virus population. Each method is characterized by its ability to correct read errors, to assemble global haplotypes, and to provide confidence values with its predictions. Software is available for four of the methods. Most applications concern HIV, HBV, and HCV.

<sup>a</sup> <http://www.cbg.ethz.ch/software/shorah>.

<sup>b</sup> <http://alla.cs.gsu.edu/~software/VISPA/vispa.html>.

<sup>c</sup> <http://code.google.com/p/ampliconnoise/>.

<sup>d</sup> [http://www.cs.unibas.ch/personen/roth\\_volker/HivHaploTyper](http://www.cs.unibas.ch/personen/roth_volker/HivHaploTyper).

**NOM : CADIEU**

**PRENOM : YANNIS**

**TITRE : LA THEORIE DES QUASI-ESPECES : CONCEPTS, APPLICATION A LA DYNAMIQUE DES POPULATIONS DE VIRUS A ARN, IMPLICATIONS BIOLOGIQUES ET LIMITES**

**RESUME**

La théorie des quasi-espèces a été développée à partir d'un modèle mathématique pour caractériser le comportement des populations virales à ARN. Cette synthèse bibliographique vise dans un premier temps à rappeler la structure et les cycles viraux, ainsi qu'à définir les différents critères de la génétique des populations qui pourraient être utilisés dans l'étude des populations virales.

Dans un second temps, les bases du modèle sont définies et les méthodes d'analyses de séquences au sein des populations sont présentées. Ces concepts permettent d'aborder la complexité et la dynamique qui caractérisent les quasi-espèces. Ils sont recensés dans les résultats expérimentaux et théoriques de la littérature. Ils ont de possibles implications quant au comportement biologique et évolutif des populations virales, et apportent un nouveau point de vue sur la thérapeutique antivirale. Enfin, la pertinence de ces propriétés est nuancée en exposant les limites mais aussi les extensions possibles à d'autres systèmes biologiques.

**MOTS CLES**

THEORIE DES QUASI-ESPECES – VIRUS A ARN – SPECTRE DE MUTANTS – MUTATION – SEQUENCE CONSENSUS – VALEUR SELECTIVE – PAYSAGE ADAPTATIF – ESPACE SEQUENCE – MUTAGENESE LETALE – RESISTANCE

**TITLE : QUASISPECIES THEORY: CONCEPTS, APPLICATION TO POPULATION DYNAMICS FOR RNA VIRUSES, BIOLOGICAL IMPLICATIONS AND LIMITS**

**ABSTRACT**

The quasispecies theory was elaborated from a mathematical model in order to characterize RNA viral populations' behavior. In a first part, this bibliographic review aims to recall what viral structure and life cycles are, and to define the different factors in population genetics that are likely to be used in the study of viral population.

In a second part, the bases of the theoretical model are defined, and analytical and sequencing methods are described. These concepts allow to approach the complexity and the dynamics that characterize viral quasispecies. They are described in experimental and theoretical results hereby presented. They have possible implications on the biological and evolutionary behavior of viral populations and offer a new approach to antiviral treatments. Finally, the relevance of those properties is discussed by showing their limits but also their possible extensions to other biological systems.

**KEY WORDS**

QUASISPECIES THEORY – RNA VIRUS – MUTANT SPECTRUM – MUTATION – CONSENSUS SEQUENCE – FITNESS – FITNESS LANDSCAPE – SEQUENCE SPACE – LETAL MUTAGENESIS – RESISTANCE