THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (Toulouse INP)

Discipline ou spécialité :

Infectiologie, Physipoathologie, Toxicologie, Génétique et Nutrition

---

Présentée et soutenue par :

M. GABRIEL DUPRÉ

le vendredi 23 juin 2023

Titre :

Virus influenza aviaires : analyse du risque d'évolution vers les formes hautement pathogènes

---

Ecole doctorale :

Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB)

Unité de recherche :

Interactions Hôtes - Agents Pathogènes (IHAP)

Directeurs de Thèse :

M. ROMAIN VOLMER

M. STEPHANE BERTAGNOLI

Rapporteurs :

MME CATHERINE ISEL-GRIFFITHS, INSTITUT PASTEUR

MME SOPHIE LE PODER, ENV MAISON ALFORT

Membres du jury :

M. CHRISTOPHE PASQUIER, CHU TOULOUSE, Président

M. ROMAIN VOLMER, ECOLE NATIONALE VETERINAIRE DE TOULOUSE, Membre

M. SEBASTIAN LEQUIME, UNIVERSITY OF GRONINGEN, Membre

M. STEPHANE BERTAGNOLI, ECOLE NATIONALE VETERINAIRE DE TOULOUSE, Membre

Remerciements

Tout d'abord, je remercie mon directeur de thèse, le Docteur Romain Volmer, pour m'avoir confié ce projet si passionnant, pour sa confiance, sa gentillesse et sa disponibilité.

Je remercie également mon co-directeur de thèse, le Docteur Stéphane Bertagnoli, pour les discussions que nous avons eues sur certains points du projet.

Je remercie les Docteurs Catherine Isel-Griffiths et Sophie Le Poder, qui ont accepté d'être les rapporteures de ce manuscrit.

Je remercie le Docteur Sebastian Lequime, qui a accepté d'être l'examinateur de mon jury de thèse.

Je remercie le Professeur Christophe Pasquier, qui a accepté de présider le jury.

Enfin, je remercie tous les membres des deux équipes virologie de l'unité IHAP, mes collègues ou plutôt amis du fameux « Poulailler », du nouveau « 10.2 », mes proches, ma famille, pour tous les bons et moins bons moments passés ensemble, ainsi que leur soutien inestimable.

Merci.

Liste des abréviations

ADNc	**ADN complémentaire**
ARNc	ARN complémentaire
ARNm	ARN messager
ARNv	ARN viral
CMV	*Cucumber mozaic virus* ou virus de la mosaïque du concombre
EBOV	*Ebolavirus* ou virus Ebola
HA	Hémagglutinine
HCV	*Hepatitis C virus* ou virus de l'hépatite C
HIV-1	*Human immunodeficiency virus type I* ou virus de l'immunodéficience humaine de type I
HPAIV	*Highly Pathogenic Avian Influenza Virus* ou virus influenza aviaire hautement pathogène
IFN	Interféron
indels	Insertions et délétions
IPVI	Indice de pathogénicité par voie intraveineuse
ISG	*Interferon Stimulating Genes* ou gènes activés par l'interféron
LPAIV	*Low Pathogenic Avian Influenza Virus* ou virus influenza aviaire hautement pathogène
MBCS	*Multibasic cleavage site motif* ou site de clivage polybasique
NA	Neuraminidase
NCR	*Non-coding region* ou région non codante
NP	Nucléoprotéine
Nt	Nucléotide
ORF	*Open Reading Frame* ou cadre de lecture
PA	Protéine acide (sous-unité du complexe polymérase viral)
PB1	Protéine basique 1 (sous-unité du complexe polymérase viral)
PB2	Protéine basique 2 (sous-unité du complexe polymérase viral)
PolyA	Polyadénylée
RIG-I	*Retinoic acid-inducible gene I* ou gène I inductible par l'acide rétinoïque
RNPv	Ribonucléoparticule virale
RNPc	Ribonucléoprotéine virale contenant la forme complémentaire du génome virale
TuMV	*Turnip mosaic virus* ou virus de la mosaïque du navet

Table des matières

Remerciements 2
Liste des abréviations 3
Table des matières 4
Tables des illustrations – Figures 6
Table des illustrations – Tableaux 8
Introduction bibliographique 9
I. Classification et morphologie des virus influenza A 10
II. Cycle cellulaire 13
1. L'entrée 13
2. Import nucléaire des RNPv et métabolisme de l'ARN viral 14
3. Export nucléaire et synthèse des nouveaux virions 20
III. Les structures secondaires d'ARN présentes sur le génome des virus influenza A 22
IV. Ecologie virale des virus influenza A 26
V. Les différents phénotypes d'influenza aviaires 30
VI. Plasticité du génome des virus influenza A 33
1. Concepts généraux sur l'évolution des virus influenza A 33
2. Effets des mutations et leurs interactions 34
3. L'évolution des virus influenza chez les oiseaux et l'Homme 36
VII. Evolution des virus influenza aviaires vers les formes hautement pathogènes 36
1. L'apparition du site de clivage polybasique 36
2. Les autres mutations localisées à distance de la région codant le site de clivage HA et associées au phénotype hautement pathogène 38
3. Pathogénicité, acquisition du MBCS et autres sous-types HA 39
4. Epidémiologie 41
5. Phénomène d'émergence reproduit en laboratoire 44
VIII. Problématique et objectifs du projet de thèse 47
Résultats 49
I. Influence de l'environnement génomique dans l'évolution des virus 50
II. Analyse phylodynamique de la tige-boucle d'ARN conservée du segment HA et son implication dans l'émergence HPAIV 59
III. Génération de virus génétiquement modifiés pour l'étude de l'acquisition du MBCS et le phénomène d'émergence HPAIV 77
Discussion et perspectives 119

I. Rôle de la structure secondaire d'ARN conservée au niveau de HA1/HA2 dans l'émergence de HPAIV : 120

II. Le phénomène de recombinaison chez les virus influenza aviaires 123

III. La formation du motif homopolymérique de six adénines, nécessaire à « l'accélération » de l'évolution de la région HA1/HA2 125

Conclusion 127

Annexes 129

I. Données supplémentaires, article : "Phylodynamic study of the conserved RNA structure encompassing the hemagglutinin cleavage site encoding region of H5 and H7 low pathogenic avian influenza virus" , Virus Evolution, 2021, 7(2), 1−16 130

II. Données supplémentaires, article (manuscrit) : « Design of engineered viruses to study the influence of genomic properties on the furin-cleavage-site motif acquisition in avian influenza viruses » (manuscrit) 145

III. Revue « La lutte contre l'émergence des virus influenza aviaires hautement pathogènes : un des défis du XXIe siècle ?», Virologie 2022, 26 (5) : 343-354 155

Références bibliographiques 168

Tables des illustrations – Figures

Figure 1 : Organisation interne du virion chez virus influenza A (à gauche); images de microscopie électronique montrant des virions influenza A sphériques (A), filamenteux (B) et pléiomorphes (C) (à droite), (d'après Krammer et al., 2018 et Noda, 2012). 11
Figure 2 : Organisation du génome des virus influenza A et les stratégies d'expression des protéines virales (ViralZone). 11
Figure 3 : Schéma de l'entrée des virus influenza A au sein de sa cellule cible (d'après Dou et al. 2018) 13
Figure 4 : Schéma simplifié des différentes étapes du cycle cellulaire de AIV (d'après Nuñez et Ross, 2019)..... 15
Figure 5 : Métabolisme de l'ARN du virus Influenza. (A) Image de microscopie électronique en coloration négative de RNPv purifiées depuis des particules virales. (B) Schéma de la transcription et la réplication du génome viral des virus influenza A. (C) Implication des étapes du cycle correspondant à la transcription / réplication du génome viral des virus influenza A (d'après Pflug et al, 2017)..... 17
Figure 6 : Représentation de l'architecture de la polymérase virale des virus influenza A (PDB : 4WSB). Représentation en trois dimensions et dans plusieurs orientations des domaines de la sous-unité PB1 (apparentée au modèle de la main droite) : motifs A-F, la boucle d'amorçage (a), les doigts (b), la paume, le pouce et les doigts (b). Les structures des domaines composant la sous-unités PA (c) et PB2 (d) ont été coloriées. Les extrémités 5' et 3' de l'ARNv sont représentés en noir et en ocre respectivement sur les différentes figures précédentes (a-d). Modèles du complexe polymérase viral à l'état « inactif » (e) et de « pré-initiation de la transcription » (f), montrant le réarrangement conformationnel du domaine de fixation flexible de l'extrémités capée des ARN cellulaire, PB2-627 et le domaine endonucléase de la sous-unité PA (Endo). Les canaux d'entrée/sortie de la matrice, d'entrée des NTP, de sortie de l'ARN néosynthétisé, la boucle d'amorçage près du site actif, les terminaisons 5′ et 3′ de l'ARNv, la séquence polyU en 5' et l'emplacement des nucléoprotéines (NP) sont illustrées (d'après te Velthuis et Fodor 2016)..... 18
Figure 7 : **Modèles de transcription (synthèse d'ARNm) et de réplication** (synthèse d'ARNc et d'ARNv) **virale chez les virus influenza A**. **Un modèle de la transcription** (a). La polymérase est dans l'état de « pré-initiation de la transcription ». L'ARN cellulaire coiffé de l'hôte est lié par le domaine de fixation de la coiffe de PB2 et clivé par le domaine de l'endonucléase de PA. Le domaine de fixation de la coiffe change de conformation pour permettre l'insertion de l'extrémité 3′ de l'ARN cellulaire jouant le rôle au sein du site actif de PB1 via le canal de sortie des produits de synthèse. L'extrémité 3′ de la matrice (ARNv) est également insérée au niveau du site actif et les NTP entrent par le canal d'entrée des NTP. La transcription est initiée par l'ajout d'un GTP à l'extrémité 3′ de l'ARN cellulaire capé. Durant l'élongation, l'ARNm néosynthétisé et la matrice sont séparés et sortent par leurs canaux de sortie respectifs. Lorsque la matrice est entrainée dans le site actif de PB1, les NP se détachent de l'ARNv, sont transloquées sur la surface de la polymérase et se refixent sur l'ARNv lorsqu'il sort par le canal de sortie. La polyadénylation de l'ARNm est réalisée par bégaiement de la polymérase sur une séquence polyU au niveau de l'extrémité 5' de la matrice (ARNv). **Modèle de synthèse de l'ARNc** (b). La polymérase est présentée dans la même conformation que pour la synthèse de l'ARNm. L'extrémité 3′ de l'ARNv s'insère dans le site actif et les NTP entrent par le canal d'entrée des NTP. L'initiation de novo se produit au premier résidu de l'ARNv et le nucléotide initiateur est stabilisé par la boucle d'amorçage. Pendant l'élongation, l'ARNc et la matrice sont séparés et sortent par leurs canaux de sortie respectifs. Les NP transloquent au niveau de la surface de la polymérase pour rejoindre la matrice (ARNv), à mesure qu'elle sort du canal de sortie. L'extrémité 5′ du produit ARNc est liée par une deuxième polymérase lorsqu'elle sort du canal de sortie et le reste de l'ARN s'associe avec des NP pour former le RNPv. L'extrémité 5′ de l'ARNv est libérée de son site de liaison et est entrainée dans le site actif de la polymérase tandis que l'extrémité 3′ se retient à la surface de la polymérase. **Modèle de la synthèse de l'ARNv** (c). La polymérase est représentée dans

l'état « inactif », ce qui correspond à la conformation de la polymérase liée à l'extrémité 5′ de l'ARNc. Pour que l'initiation de la réplication ait lieu, la polymérase doit interagir avec une seconde polymérase transactivatrice. L'extrémité 3′ de l'ARNc s'insère au niveau du site actif et les NTP entrent par leur canal d'entré. L'initiation de novo se produit aux 4ième et 5ième résidus de la matrice (ARNc) sans la participation de la boucle d'amorçage. Le dinucléotide pppApG qui en résulte est déplacé vers l'extrémité 3′ de l'ARNc et l'élongation commence. La terminaison se déroulent comme décrit pour la synthèse de l'ARNc. L'extrémité 5′ du produit d'ARNv est liée par la polymérase transactivante et le reste de l'ARNv s'associe avec des NP pour former le RNPv (d'après te Velthuis et Fodor 2016). 19
Figure 8 : Schéma représentant les différentes étapes composant l'assemblage et la libération des virions chez les virus influenza A : le recrutement des segments et l'empaquetage du génome viral (A), le bourgeonnement de la surface de la membrane plasmique au niveau de radeaux lipidiques (B, C) et la libération des virions néosynthétisés (C) (d'après Rossman et Lamb, 2011). 21
Figure 12 : Les régions promotrices des ARN viraux des virus influenza A, d'après (Ferhadian et al. 2018)..... 23
Figure 13 : Cartes du réseau d'intéractions ARN-ARN entre les segments de la souche PR8 (H1N1) (a), Udorn (H3N2) (b) et le réassortant PR8::Udorn (PB1+NA) (c), d'après (Dadonaite et al. 2019) . 25
Figure 9 : Diversité des échanges virologiques au sein des oiseaux et des espèces de mammifères susceptibles et sensibles à l'infection par les virus influenza A. Les différents sous-types concernés par chaque espèce sont indiqués (modifiée, d'après J. S. Long et al., 2019). 28
Figure 10 : Différences protéiques au niveau du site de clivage de la HA entre les virus influenza aviaires faiblement (LPAIV, à gauche) et hautement pathogènes (HPAIV, à droite), puis les conséquences cliniques chez le poulet (d'après Horimoto et al., 2005). E, T, K et R représentent des acides aminés dont les deux derniers sont basiques : respectivement les résidus lysine (K) et arginine (R)..... 32
Figure 11 : Mécanismes d'évolution des les virus influenza A. (A) Accumulation progressives de mutations au niveau du génome des virus influenza A, conduisant à la génération de nouveaux variants : dérive antigénique, responsable des épidémies samsonières chez l'Homme (« antigenic drift »). (B) Echange ou réassortiment des segments entre deux souches virus influenza A dans le cadre d'une coïnfection d'une cellule peut entrainer l'émergence d'une nouvelle souche (nouveau sous-type) : cassure antigénique (« antigenic shift »). Ce dernier mécanisme est observé dans le cas de pandémies (d'après Mostafa et al, 2018). 34
Figure 15 : Arbres phylogénétiques (horloge moléculaire) des séquences HA5 (A.) et H7 (B.) représentant les principaux évènements d'émergences HPAIV (en rouge), d'après (D.-H. Lee, Criado, and Swayne 2020) 44
Figure 16 : Schéma conceptuel sur les facteurs génomiques influençant l'émergence des HPAIV. Certaines propriétés du génome au niveau de la région codant le site de clivage HA comme des structures ARN (A.) et/ou motifs nucléotidiques (B.) peuvent être favorable à l'apparition de mutation et à l'émergence. 48
Figure 14 : Schéma conceptuel de l'évolution des génomes viraux. Les mutations de novo apparaissent sur le génome viral via plusieurs mécanismes incluant les erreurs de la polymérase virale, les mécanismes de réparation, et l'édition médiée par les enzymes cellulaires ou virales, ainsi que les dommages physiques directs aux acides nucléiques (espèces réactives de l'oxygène). Des caractéristiques intrinsèques du génome viral (structures, motifs nucléotidiques, marques épigénétiques) peuvent influencer l'édition médiée par l'hôte et la fidélité de la polymérase virale. In fine, certaines mutations vont être sélectionnées par leur conséquence sur le fitness viral. Les évènements de recombinaison et les réassortiments (dans le cas des virus diploïdes, segmentés ou multipartites) peuvent aussi influencer la diversité génomique via des échanges plus larges de matériel génétique. 50

Table des illustrations – Tableaux

Tableau 1 : Listes des protéines virales codées par le génome des virus influenza A et leur(s) fonction(s) associées..... 12
Tableau 2 : Liste des évènements d'émergence HPAIV naturels décrits. 43
Tableau 3 : Evènements d'émergence naturels par recombinaison 124

Introduction bibliographique

I. Classification et morphologie des virus influenza A

Les virus influenza A appartiennent à la famille des *Orthomyxoviridae,* composée par quatre genres *Influenza* : A, B, C, D, les *Thogotovirus*, les *Isavirus* et *Quaranjavirus* (d'après *International Commitee on Taxonomy of Viruses* ou ICTV).

Une nomenclature précise est utilisée afin d'identifier chaque souche virale. Un isolat est caractérisé par le genre de la souche (A, B, C ou D pour les genres *Influenzavirus*), l'espèce chez laquelle la souche a été isolée pour la première fois (non précisée s'il s'agit d'un virus infectant l'homme), la localisation géographique de l'échantillon, la référence et l'année de l'isolement. Exclusivement pour les virus influenza A ces derniers sont classés en différents sous-types en fonction de la combinaison des deux glycoprotéines majoritaires de surface du virus : l'hémagglutinine (HA) et la neuraminidase (NA). Les sous-types de HA et de NA apparaissent entre parenthèse sous la forme de HxNy. Au sein des virus influenza A, 18 HA et 11 NA différentes ont été décrites (Krammer et al. 2018). Les glycoprotéines HA et NA permettent respectivement la reconnaissance des récepteurs de type acide sialique situés sur la cellule cible et l'entrée du virus, puis la libération des virions nouvellement produits lors de la réplication virale. Par ailleurs, HA et NA sont les protéines virales les plus antigéniquement variables et ciblées par la réponse immunitaire de type humorale (Kosik and Yewdell 2019).

Les virus influenza A sont des virus enveloppés, dont la membrane lipidique provient de la cellule hôte depuis laquelle ils ont bourgeonné. La particule virale est pléomorphe : de forme globalement sphérique (**Figure 1**), ou filamenteuses dans certains cas, avec un diamètre variant entre 80 et 120 nm (Krammer et al. 2018; Noda 2011).

Le génome, situé au cœur du virion, est constitué de 8 segments d'ARN simple brin de polarité négative (sb -, ARNv) dont la longueur totale est de 13,5 kB (**Figure 2**). Douze à quatorze protéines virales sont codées et produites en fonction des souches. Chaque segment d'ARN est recouvert par des nucléoprotéines (NP). Le complexe de réplication viral constitué par les protéines PA (polymérase acide), PB1 (polymérase basique 1) et PB2 (polymérase basique 2) nécessaires à la transcription et réplication du génome viral est lié au niveau des deux extrémités 5' et 3' de chaque segment, qui sont hybridés partiellement (Te Velthuis and Fodor 2016). Ensemble, l'ARNv recouvert de NP et le complexe PA-PB1-PB2 constituent la ribonucléoparticule virale (RNPv), l'unité de la transcription et la réplication virale. Plusieurs protéines sont retrouvées enchâssées au niveau de l'enveloppe du virus, à savoir les glycoprotéines HA et NA constituant les principaux déterminants antigéniques (avec un ratio

nombre de protéines HA/NA ~ 6,5) puis le canal ionique M2. La protéine M1 quant à elle est retrouvée sur la face interne de la membrane lipidique et constitue la matrice (Krammer et al. 2018; Harris et al. 2006).

D'autres protéines telles que PB2-S1, PB1-F2, PA-X et NS sont codées sur des cadres de lectures alternatifs (**Figure 2** et **Tableau 1**).

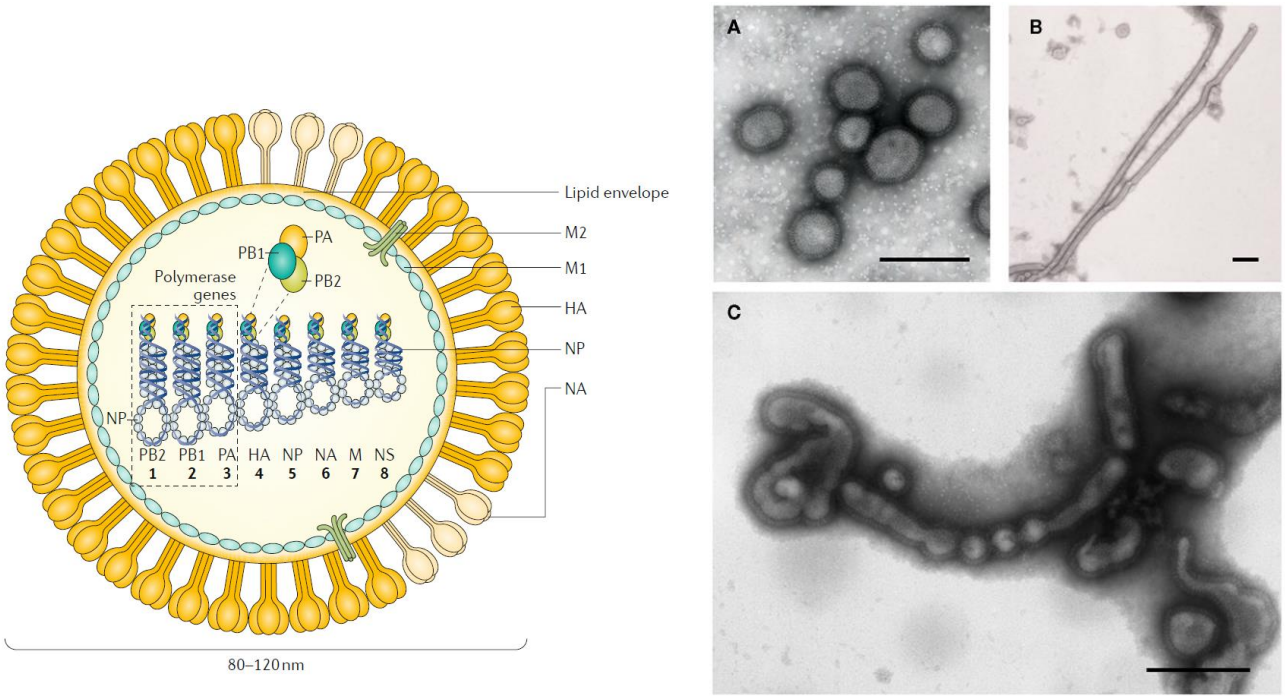


Figure 1 *: Organisation interne du virion chez virus influenza A (à gauche); images de microscopie électronique montrant des virions influenza A sphériques (A), filamenteux (B) et pléiomorphes (C) (à droite), (d'après Krammer et al., 2018 et Noda, 2012).*

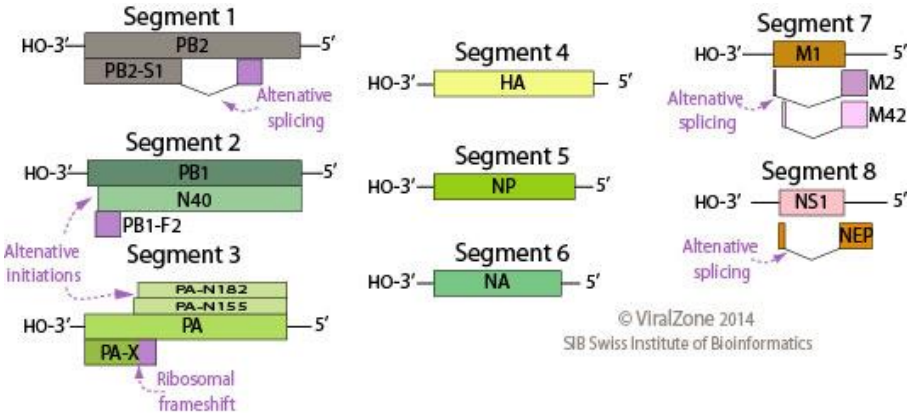


Figure 2 *: Organisation du génome des virus influenza A et les stratégies d'expression des protéines virales (ViralZone).*

***Tableau 1** : Listes des protéines virales codées par le génome des virus influenza A et leur(s) fonction(s) associées*

Segment d'ARNv	Protéine(s)	Fonction(s) de la ou des protéines
1	1. PB2 2. PB2-S1	1. Sous-unité du complexe polymérase, reconnaissance de la coiffe 5' des ARNm de l'hôte dans le cadre de la transcription virale (vol de coiffe) (Fodor 2013; Pflug et al. 2017; Ruigrok et al. 2010; Te Velthuis and Fodor 2016) 2. Forme tronquée de PB2 issue d'un épissage alternatif, inhibiteur de RIG-I (Yamayoshi et al. 2015)
2	1. PB1 2. PB1-F2 3. PB1-N40	1. Sous-unité du complexe polymérase en charge de la synthèse d'ARN (Fodor 2013; Pflug et al. 2017; Ruigrok et al. 2010; Te Velthuis and Fodor 2016) 2. Facteur de virulence, induction de la réponse inflammatoire - recrutement des neutrophiles, activation de l'apoptose (Kamal, Alymova, and York 2017) 3. Rôle dans la réplication de certaines souches (Q. Wang et al. 2019)
3	1. PA 2. PA-X 3. PA-N155 4. PA-N182	1. Sous-unité du complexe polymérase virale, permet le clivage de l'ARNm lors de la transcription (Fodor 2013; Pflug et al. 2017; Ruigrok et al. 2010; Te Velthuis and Fodor 2016) 2. Facteur de virulence, régule la réponse immunitaire de l'hôte – répression de gènes cellulaires de l'hôte (via son activité endonucléase) (Hu, Ma, and Liu 2018; Nogales et al. 2018) 3. Rôles dans la réplication (Muramoto et al. 2013) 4. Rôles dans la réplication (Muramoto et al. 2013)
4	HA	Glycoprotéine de surface, rôle dans l'attachement des virions aux récepteurs cellulaires, la fusion, facteur de virulence principale des souches influenza aviaires hautement pathogènes, cible principale des anticorps neutralisant de l'hôte (Abdelwhab, Veits, and Mettenleiter 2013; Gamblin and Skehel 2010; Kosik and Yewdell 2019; Nao et al. 2017; Steinhauer 1999)
5	NP	Liaison aux ARNv et ARNc, import nucléaire des RNPv (signaux NLS), régulation de la transcription/réplication des ARNv (Neumann, Castrucci, and Kawaoka 1997; Pflug et al. 2017; Te Velthuis and Fodor 2016)
6	NA	Glycoprotéine de surface, rôle dans la libération des virions, cible majoritaire des anticorps neutralisants (Kosik and Yewdell 2019; Jeremy S. Rossman and Lamb 2011)
7	1. M1 2. M2 3. M42	1. Protéine de matrice, rôle dans l'export des RNPv (Brunotte et al. 2014; Cros and Palese 2003) 2. Canal ionique, rôle dans l'étape d'entrée du virus (fusion, acidification des endosomes) et de la maturation au niveau du trans-Golgi (Pielak and Chou 2011) 3. Forme alternative et fonctionnelle de M2 (épissage alternatif) (Wise et al. 2012)
8	1. NS1 2. NEP (NS2)	1. Facteur de virulence, régulation de l'activation du système IFN (Ayllon and García-Sastre 2015; Nogales et al. 2018) 2. Rôle dans l'export des RNPv (Brunotte et al. 2014; Neumann, Castrucci, and Kawaoka 1997)

II. Cycle cellulaire

1. L'entrée

Le contact du virion avec la cellule cible s'effectue par la mise en jeu de récepteurs spécifiques de type acide sialique avec la HA virale trimérique (Dou et al. 2018; Gamblin and Skehel 2010). Il existe une spécificité d'hôte selon le type de liaison glycosidique entre l'acide sialique terminal et le galactose adjacent qui détermine le spectre d'hôte des virus influenza A. Ainsi les acides sialiques formant une liaison de type alpha 2-6 glycosidique avec le galactose sont ciblés par des virus virus influenza A « humains », tandis que le récepteur cellulaire décrivant une liaison alpha 2-3 est requis pour l'infection avec des souches d'origine aviaire. Il a été cependant démontré la présence de ces récepteurs aviaires au niveau de certains tissus restreints au niveau de l'arbre respiratoire profonds (bronchioles, alvéoles pulmonaires) chez l'Homme, expliquant une possible susceptibilité aux souches virus influenza aviaires après une exposition prolongée (Dou et al. 2018; Van Riel et al. 2006). Les porcs expriment au niveau de leurs tissus les deux types de récepteurs, ce qui les rend susceptibles tant aux virus d'origine aviaire qu'à des virus humains.

Le virus entre dans la cellule par différentes voies, majoritairement par endocytose clathrine-dépendante (Dou et al. 2018; Edinger, Pohl, and Stertz 2014) (**Figure 3** et **Figure 4**). La macropinocytose a aussi été décrite comme voie d'entrée alterne des virus de forme filamenteuse (J. S. Rossman, Leser, and Lamb 2012). Ensuite, le virion est adressé vers le

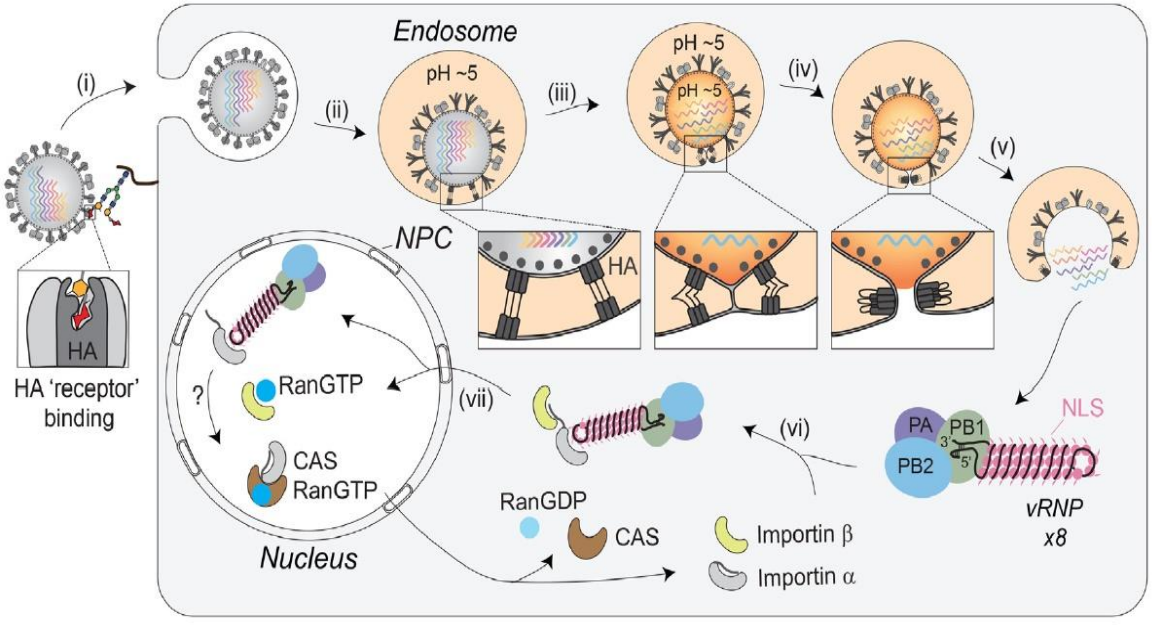


***Figure 3** : Schéma de l'entrée des virus influenza A au sein de sa cellule cible (d'après Dou et al. 2018)*

compartiment endosomal tardif. L'abaissement progressif du pH va provoquer une succession de changements conformationnels de la HA, ayant pour résultat l'exposition du peptide de fusion situé au niveau de l'extrémité N terminale (N-ter) de HA2 (obtenue après maturation protéolytique de HA0 par des protéases cellulaires). Ce dernier va être projeté vers la membrane cible endosomale. Les membranes lipidiques virales et cellulaires vont progressivement se rapprocher puis fusionner, faisant apparaître le diaphragme d'hémi fusion qui va ensuite s'étirer, laissant finalement communiquer les compartiments intra virionique et cytoplasmique (mécanisme de fusion de type I) (Dou et al. 2018). Lors de l'acidification du compartiment endosomal, les canaux ioniques M2 vont permettre l'influx d'ions H+ au sein de la particule virale, ce qui va déstabiliser les liaisons préexistantes entre les RNPv et la matrice M1. Les RNPv libres sont ensuite transloquées vers le compartiment cytoplasmique.

2. Import nucléaire des RNPv et métabolisme de l'ARN viral

La totalité du métabolisme de l'ARN viral dépend des fonctions nucléaires. Les RNPv sont importées dans le noyau grâce à des signaux de localisation intranucléaires NLS (*nuclear localization signal*) présents sur les protéines NP, permettant leur interaction avec les kariophérines α et β cellulaires (Dou et al. 2018; Weber et al. 1998; J. Li et al. 2015).

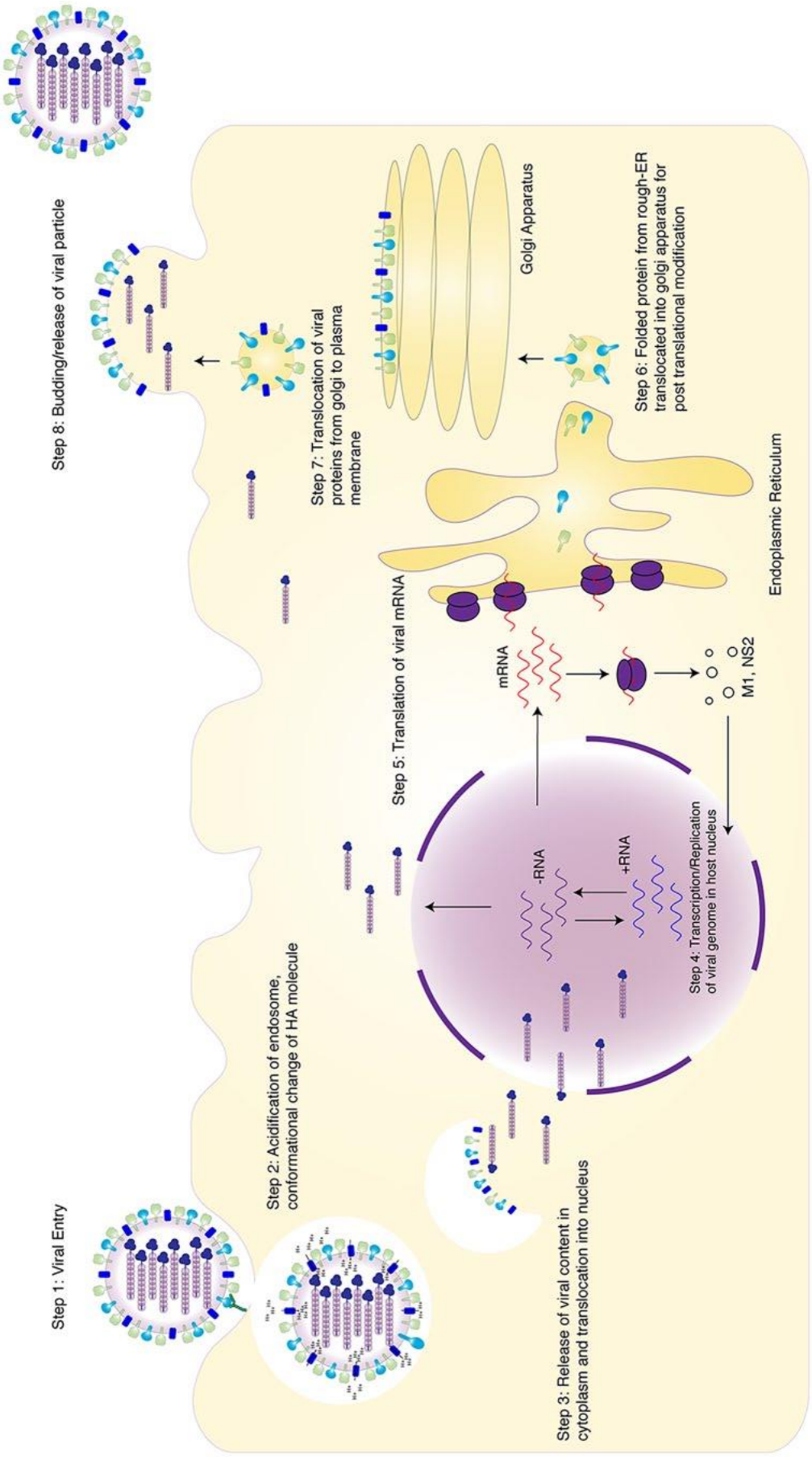


Figure 4 *: Schéma simplifié des différentes étapes du cycle cellulaire de AIV (d'après Nuñez et Ross, 2019).*

Les RNPv vont être ensuite utilisées à la fois pour la transcription et la réplication du génome viral (Te Velthuis and Fodor 2016; Pflug et al. 2017) (**Figure 5**, **Figure 6** et **Figure 7**).
Le complexe de réplication en cis de chaque segment ou ARNv, constitué par les sous-unités PB1, PB2 et PA permet la formation des ARN messagers (ARNm) viraux. Un mécanisme particulier de « vol de coiffe » (*cap snatching*), permet leur production. PB2 lie l'extrémité capée des ARN messagers cellulaires. En effet, il a été montré que PA interagit avec l'ARN polymérase II cellulaire via son domaine C terminal (CTD), permettant au complexe polymérase de se localiser dans une zone riche en ARN messagers cellulaires (Martínez-Alonso, Hengrung, and Fodor 2016). La matrice virale (sous la forme de RNPv) est fixée à la sous-unité PB1 au niveau de ces extrémités ARN 3' et 5' terminales. La protéine PA possède une activité endonucléasique et clive l'ARN cellulaire 12 à 15 nucléotides en aval de la coiffe. Le fragment 5' coiffé cellulaire créé est ensuite transféré sur l'ARNv, jouant le rôle d'amorce pour la transcription (Pflug et al. 2017). Pour finir, une queue polyA est ajoutée au niveau de l'extrémité 3' de l'ARN+ viral coiffé par bégaiement de l'ARN polymérase virale (encore appelé : *polymerase stuttering*) au niveau d'une séquence répétée en uraciles. Une structure de type épingle à cheveux située à 7 nucléotides en amont de la séquence polyU serait à l'origine de mécanisme de bégaiement polymérase virale, induisant une pause de la progression de celle-ci sur le brin matrice, un retour en arrière et la copie réitérative de la base uracile (Pflug et al. 2017; Luo et al. 1991).

Après leur traduction, les protéines NP, les sous-unités PA, PB1 et PB2 sont importées au sein du noyau grâce à la présence de signaux NLS afin de pouvoir réaliser la réplication du génome viral (J. Li et al. 2015). Cette étape correspond à la phase d'amplification du matériel génétique du virus.
Les RNPv composées du segment d'ARNv(-) sont utilisés comme matrices, permettant ainsi la synthèse des intermédiaires génomiques : ARN complémentaires de polarité positive (ARNc (+)) (Pflug et al. 2017). Ces derniers correspondent à la copie exacte de chaque ARNv (hormis les erreurs de la polymérase virale). L'initiation de la polymérisation du brin complémentaire est réalisée via la polymérase locale sans amorce (*de novo)*, par l'ajout d'un dinucléotide adénosine triphosphate-guanosine monophosphate (pppApG). Ce dernier va s'hybrider avec les deux premiers nucléotides en 3' terminale de l'ARNv, permettant la synthèse de l'ARNc lors de l'élongation. Le brin formé est ensuite associé avec un second complexe polymérase et les NP pour former les ribonucléoprotéines contenant l'ARNc viral (RNPc) (Pflug et al. 2017).
La synthèse de l'ARNv présente quelques différences. Un second complexe polymérase « transactivateur » doit interagir avec la polymérase dite « résidente » (liée aux extrémités 3' et

5’ de l’ARNc) pour démarrer la synthèse du brin ARNv. Le dinucléotide initiateur pppApG est ajouté au niveau des positions 4 et 5, puis réaligné dans un second temps aux deux premiers nucléotides sur l’ARNc. Ensuite le brin ARNv est allongé et encapsidé pour reformer le RNPv (Pflug et al. 2017). Les étapes de la transcription et de la réplication virales sont détaillées et schématisées plus bas (**Figure 6** et **Figure 7**).

Les facteurs régulant le déclenchement soient de la transcription ou de la réplication du génome virale restent encore mal connus. Les quantités stœchiométriques des différents acteurs moléculaires nécessaires à la transcription / réplication et la localisation du complexe réplicatif complet dans le noyau pourraient expliquer le choix de tel chemin métabolique. Par ailleurs, malgré leur très forte similarité, la structure des régions promotrices de l’ARNc et ARNv (extrémités 3’ et 5’) semblent pouvoir être discriminée par la polymérase virale, permettant une initiation et synthèse différente de l’ARN complémentaire correspondant.

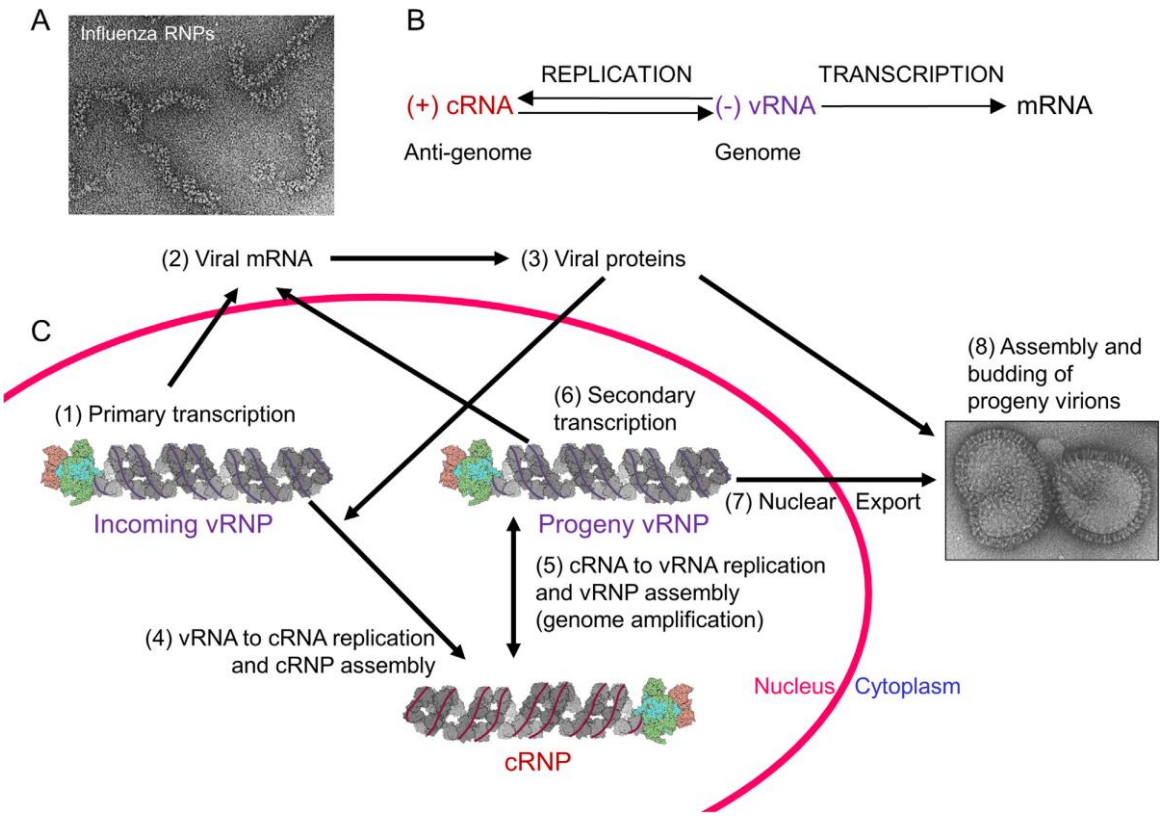


Figure 5 *: Métabolisme de l'ARN du virus Influenza. (A) Image de microscopie électronique en coloration négative de RNPv purifiées depuis des particules virales. (B) Schéma de la transcription et la réplication du génome viral des virus influenza A. (C) Implication des étapes du cycle correspondant à la transcription / réplication du génome viral des virus influenza A (d’après Pflug et al, 2017).*

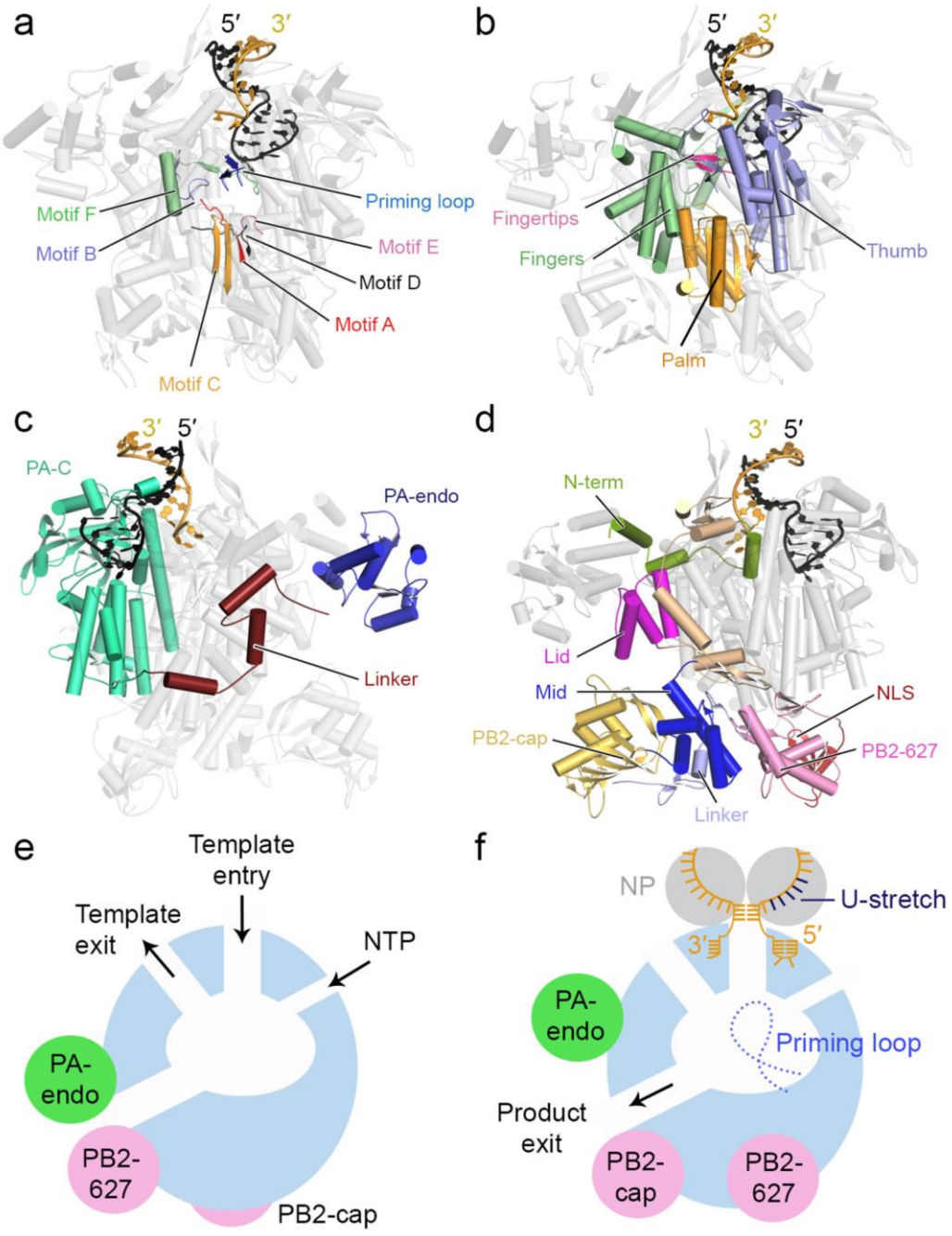


Figure 6 *: Représentation de l'architecture de la polymérase virale des virus influenza A (PDB : 4WSB). Représentation en trois dimensions et dans plusieurs orientations des domaines de la sous-unité PB1 (apparentée au modèle de la main droite) : motifs A-F, la boucle d'amorçage (a), les doigts (b), la paume, le pouce et les doigts (b). Les structures des domaines composant la sous-unités PA (c) et PB2 (d) ont été coloriées. Les extrémités 5' et 3' de l'ARNv sont représentés en noir et en ocre respectivement sur les différentes figures précédentes (a-d). Modèles du complexe polymérase viral à l'état « inactif » (e) et de « pré-initiation de la transcription » (f), montrant le réarrangement conformationnel du domaine de fixation flexible de l'extrémités capée des ARN cellulaire, PB2-627 et le domaine endonucléase de la sous-unité PA (Endo). Les canaux d'entrée/sortie de la matrice, d'entrée des NTP, de sortie de l'ARN néosynthétisé, la boucle d'amorçage près du site actif, les terminaisons 5′ et 3′ de l'ARNv, la séquence polyU en 5' et l'emplacement des nucléoprotéines (NP) sont illustrées (d'après te Velthuis et Fodor 2016).*

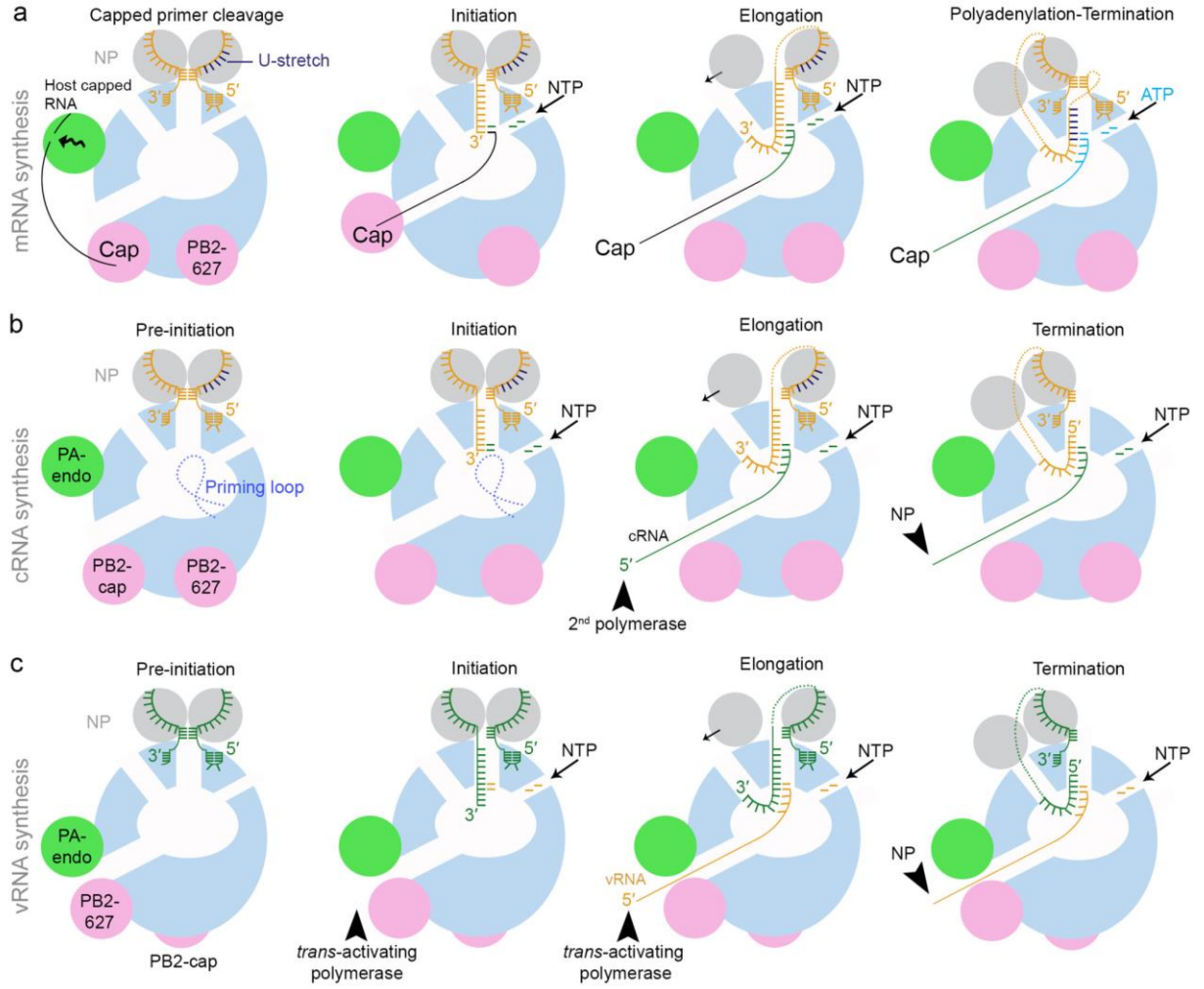


Figure 7 *: **Modèles de transcription (synthèse d'ARNm) et de réplication** (synthèse d'ARNc et d'ARNv) **virale chez les virus influenza A**. **Un modèle de la transcription** (a). La polymérase est dans l'état de « pré-initiation de la transcription ». L'ARN cellulaire coiffé de l'hôte est lié par le domaine de fixation de la coiffe de PB2 et clivé par le domaine de l'endonucléase de PA. Le domaine de fixation de la coiffe change de conformation pour permettre l'insertion de l'extrémité 3' de l'ARN cellulaire jouant le rôle au sein du site actif de PB1 via le canal de sortie des produits de synthèse. L'extrémité 3' de la matrice (ARNv) est également insérée au niveau du site actif et les NTP entrent par le canal d'entrée des NTP. La transcription est initiée par l'ajout d'un GTP à l'extrémité 3' de l'ARN cellulaire capé. Durant l'élongation, l'ARNm néosynthétisé et la matrice sont séparés et sortent par leurs canaux de sortie respectifs. Lorsque la matrice est entrainée dans le site actif de PB1, les NP se détachent de l'ARNv, sont transloquées sur la surface de la polymérase et se refixent sur l'ARNv lorsqu'il sort par le canal de sortie. La polyadénylation de l'ARNm est réalisée par bégaiement de la polymérase sur une séquence polyU au niveau de l'extrémité 5' de la matrice (ARNv). **Modèle de synthèse de l'ARNc** (b). La polymérase est présentée dans la même conformation que pour la synthèse de l'ARNm. L'extrémité 3' de l'ARNv s'insère dans le site actif et les NTP entrent par le canal d'entrée des NTP. L'initiation de novo se produit au premier résidu de l'ARNv et le nucléotide initiateur est stabilisé par la boucle d'amorçage. Pendant l'élongation, l'ARNc et la matrice sont séparés et sortent par leurs canaux de sortie respectifs. Les NP transloquent au niveau de la surface de la polymérase pour rejoindre la matrice (ARNv), à mesure qu'elle sort du canal de sortie. L'extrémité 5' du produit ARNc est liée par une deuxième polymérase lorsqu'elle sort du canal de sortie et le reste de l'ARN s'associe avec des NP pour former le RNPv. L'extrémité 5' de l'ARNv est libérée de son site de liaison et est entrainée dans le site actif de la polymérase tandis que l'extrémité 3' se retient à la surface de la polymérase. **Modèle de la synthèse de l'ARNv** (c). La polymérase est représentée dans l'état « inactif », ce qui correspond à la conformation de la polymérase liée à l'extrémité 5' de l'ARNc. Pour que l'initiation de la réplication ait lieu, la polymérase doit interagir avec une seconde polymérase transactivatrice. L'extrémité 3' de l'ARNc s'insère au niveau du site actif et les NTP entrent par leur canal d'entré. L'initiation de novo se produit aux 4ième et 5ième résidus de la matrice (ARNc) sans la participation de la boucle d'amorçage. Le dinucléotide pppApG qui en résulte est déplacé vers l'extrémité 3' de l'ARNc et l'élongation commence. La terminaison se déroulent comme décrit pour la synthèse de l'ARNc. L'extrémité 5' du produit d'ARNv est liée par la polymérase transactivante et le reste de l'ARNv s'associe avec des NP pour former le RNPv (d'après te Velthuis et Fodor 2016).*

3. Export nucléaire et synthèse des nouveaux virions

Les ARNm et RNPv vont être ensuite exportés hors du noyau. L'action conjointe de la protéine NEP et de la protéine de matrice M1 permettrait l'export des RNPv néosynthétisées. M1 recouvrant les signaux NLS des NP (Dou et al. 2018), NEP interagit avec la machinerie d'export de la cellule et provoque alors la sortie des RNPv hors du noyau. Au niveau cytoplasmique, les ARNm viraux sont pris en charge par les ribosomes cellulaires qui vont ensuite produire les protéines virales. Les protéines NP, PB1, PB2, PA et NS1 sont traduites. Elles ont des actions spécifiques au niveau nucléaire comme la régulation de la transcription, de la réplication du génome viral et de l'inhibition de l'export des ARN cellulaires (Lo, Tang, and Shaw 2018; Dou et al. 2018; Neumann, Castrucci, and Kawaoka 1997). La protéine NS1 a de multiples intéractants cellulaires. Elle est considérée comme un facteur de virulence majeur (Ayllon and García-Sastre 2015).

Les ARNm viraux codant pour HA, NA et le canal ionique M2 sont adressés vers le réticulum endoplasmique (RE) et traduits. Les protéines synthétisées sont alors enchâssées au niveau de la membrane lipidique du RE. Ces dernières sont ensuite adressées vers la membrane plasmique de la cellule en utilisant le système de sécrétion de la cellule et vont s'accumuler au niveau de radeaux lipidiques (Dou et al. 2018; Nayak, Hui, and Barman 2004). Il a été montré que les RNPv interagissent avec la GTPase Rab11, ce qui pourrait expliquer le transport de ces derniers vers la membrane cytoplasmique et des premières interactions inter segments, en préparation de l'empaquetage du génome viral complet (De Castro Martin et al. 2017). Les RNPv exportées hors du noyau sont transportées vers les radeaux lipidiques via le cytosquelette d'actine et le réseau vésiculaires issue du réticulum endoplasmique. Les huit RNPv constituant le génome complet du virus vont être adressées au niveau de ces sites de formation de nouveaux virions. Des interactions fines au niveau de zones moins denses en NP entre les segments d'ARN vont permettre l'empaquetage correct du génome viral complet. Ce processus est critique pour la formation de virions infectieux et semble suivre un mécanisme dirigé ou sélectif. Certaines structures secondaires d'ARN au niveau des extrémités 5' et 3' des différents segment ainsi que des résidus spécifiques de la protéine NP seraient impliqués (Dadonaite et al. 2019; Dawson, Lazniewski, and Plewczynski 2017; Williams et al. 2018; Ferhadian et al. 2018). Après la déformation de la membrane plasmique induite par la composition biochimique des radeaux lipidiques et la présence locale des protéines virales (HA, NA, M1, M2 et NP), les nouveaux virions se forment par bourgeonnement. Ces derniers restent fixés à la surface de la cellule du fait des interactions entre les acides sialiques et les protéines HA trimériques. Ce n'est qu'après

le clivage de ces liaisons, médié par la protéine NA, que les virus seront libérés hors de la cellule infectée (**Figure 8**) (Dou et al. 2018; Jeremy S. Rossman and Lamb 2011).

Les virions néosynthétisés deviennent infectieux après l'action de protéases cellulaires qui vont cliver ou maturer la protéine HA0, entrainant la formation du dimère HA1 / HA2 fonctionnel (Gamblin and Skehel 2010). Deux groupes de protéases sont responsables de la maturation de HA0. Le premier groupe constitué des tryptases cellulaires et de la plasmine reconnait une arginine et clive la protéine après ce résidu. Le second groupe de protéases est composé de protéases intracellulaires de type furine et PC6. Ces dernières reconnaissent un motif minimal dibasique : Q-R/K-X-R/K-R (Horimoto et al. 1994). Un lien direct entre la capacité de HA0 à pouvoir être maturée et la virulence a été démontré (Horimoto and Kawaoka 1994; Garten and Klenk 1999). Ce point sera explicité dans un paragraphe suivant, page 30.

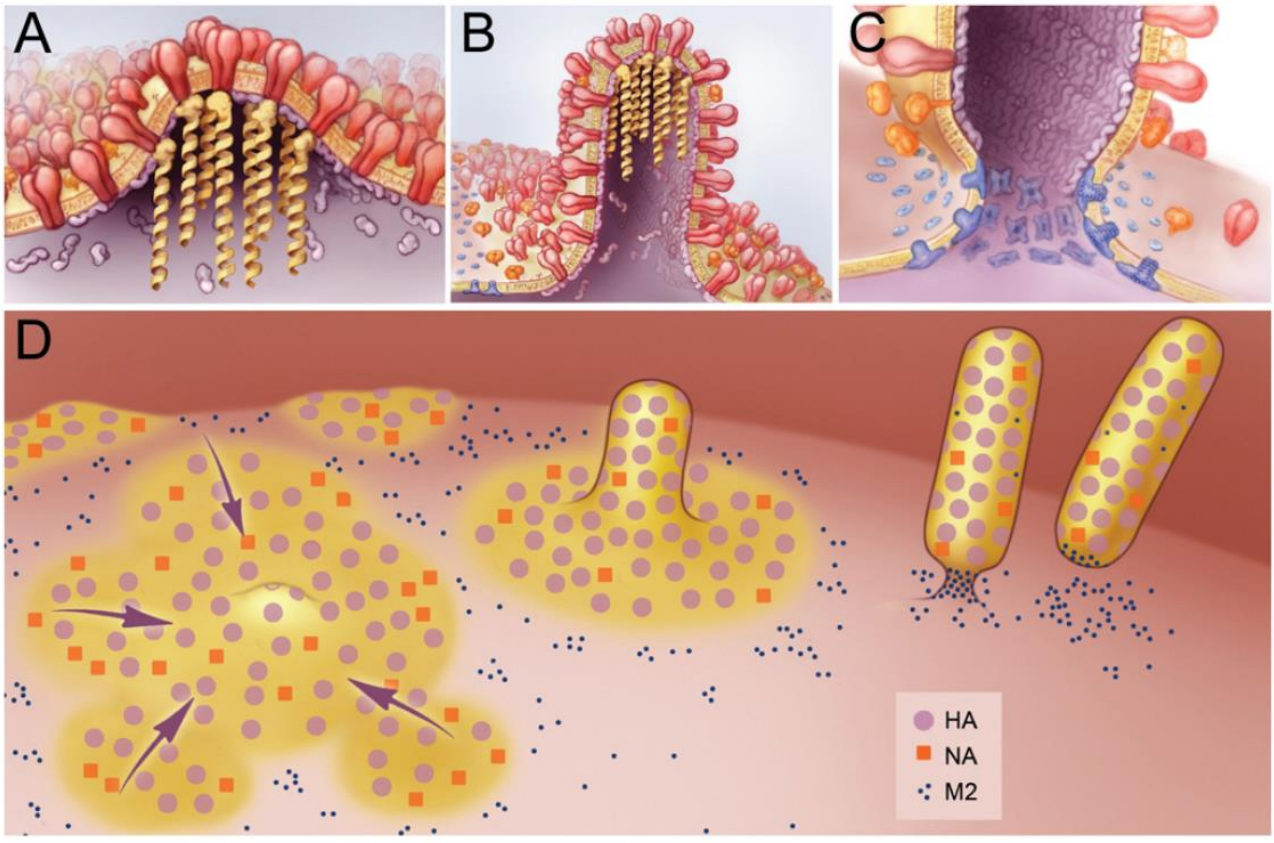


Figure 8 *: Schéma représentant les différentes étapes composant l'assemblage et la libération des virions chez les virus influenza A : le recrutement des segments et l'empaquetage du génome viral (A), le bourgeonnement de la surface de la membrane plasmique au niveau de radeaux lipidiques (B, C) et la libération des virions néosynthétisés (C) (d'après Rossman et Lamb, 2011).*

III. Les structures secondaires d'ARN présentes sur le génome des virus influenza A

Depuis plusieurs décennies, la structure de l'ARN est progressivement décrite comme jouant un rôle majeur dans la régulation de plusieurs voies métaboliques chez beaucoup d'organismes dont les virus. La réplication virale, l'empaquetage du génome, l'édition de l'ARN et son épissage sont finement régulés par ces structures (Mehedi et al. 2013; Kolakofsky 2016; Lim and Brown 2018; Jinwei Zhang and Landick 2016).

Ces découvertes ont été rendues possibles via le développement de plusieurs méthodes : des modèles de prédictions de structures d'ARN ainsi que des techniques de détection directe de ces structures. Deux grands types de méthodes peuvent être utilisées pour la prédiction de structure secondaires d'ARN. La première est fondée sur des concepts thermodynamiques comme la minimisation de l'énergie libre. Plusieurs structures sont calculées et celles étant les plus stables sont retenues comme les plus probables d'exister dans la nature (SantaLucia 1998). L'autre méthode de prédiction est basée sur l'alignement multiple de séquences relativement homologues et l'analyse de covariation (Parsch, Braverman, and Stephan 2000). Une substitution primaire suivie d'une mutation compensatrice au niveau de bases non adjacentes correspond potentiellement à deux positions dîtes « covariantes » ou évoluant conjointement, et donc potentiellement impliquées dans un appariement (dans la tige d'une structure de type épingle à cheveux ou tige-boucle). Ce procédé donne des informations sur des zones de contraintes au niveau d'une séquence nucléotidique donnée.

La cartographie chimique de l'ARN a permis d'identifier directement la présence de motifs d'ARN chez plusieurs virus dans des conditions *in vitro* et cellulaires (Williams et al. 2018; Lin and Miles 2019; Flynn et al. 2016; Merino et al. 2005). Le principe de ces méthodes de cartographie repose sur l'utilisation d'agents chimiques qui vont interagir avec certaines bases de l'ARN. Après l'utilisation de ces agents et des traitements spécifiques supplémentaires, la base modifiée induira une mutation ponctuelle lors de la synthèse d'ADN complémentaire (ADNc) ou la coupure du brin d'ARN. Seules les bases accessibles ou non incluses dans des appariements vont pouvoir interagir avec ces agents, et ainsi, une carte du repliement de l'ARN étudié va pouvoir être générée. La revue suivante décrit en détail les différentes méthodes chimiques disponibles ainsi que leur développement (Gilmer et al. 2021).

Chez les virus influenza aviaires, de multiples structures au niveau de leur génome, tant sur l'ARNv, l'ARNc que sur l'ARNm ont été observées et jouent des rôles variés au niveau du cycle viral. Les principaux motifs sont décrits dans cette revue en détails (Ferhadian et al. 2018). Au niveau des huit segments composant le génome viral des virus influenza A, les séquences terminales 5' et 3' (de 13 et 12 nucléotides) sont très conservées chez toutes les souches du même genre. En effet, ces dernières vont remplir le rôle de promoteurs et réguler la réplication du génome viral ainsi que la transcription (te Velthuis and Fodor 2016). C'est à ce niveau précis où le complexe polymérase viral va se fixer et former les RNPv, composés aussi de la molécule d'ARN viral imparfaitement recouverte de NP. Au cours de ces trente dernières années, plusieurs modèles de topologie du promoteur viral ont été suggérés (**Figure 9**). L'ARNv nu décrit au niveau de ses promoteurs forme une queue de poêle (modèle *panhandle*) dans lequel les bases sont majoritairement appariées jusqu'au position 15 et 14 (depuis les extrémités 5' et 3' respectivement). En revanche la fixation du complexe polymérase au niveau du promoteur entraine un changement de topologie. Une tige-boucle apparait en 5' terminal entre les nucléotides 1-10 appelée « crochet » (modèle *hook*). L'extrémité 3' terminale libre simple brin va pouvoir entrer au sein de la sous-unité PB1 et arriver au niveau du site catalytique. Plusieurs études ont montré que la région duplex du promoteur est primordiale, de telle façon que des mutations induites à ce niveau par mutagenèse dirigée perturbent l'empaquetage correct des segments et/ou la réplication (Ferhadian et al. 2018; Pflug et al. 2014; J. Wang et al. 2017).

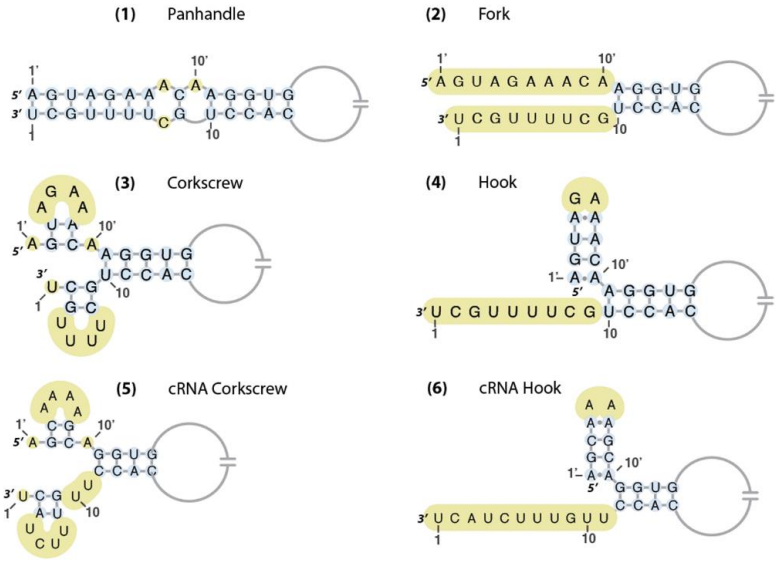


Figure 9 *: Les régions promotrices des ARN viraux des virus influenza A, d'après* (Ferhadian et al. 2018)

Les structures secondaires régulent aussi la transcription virale chez les virus influenza A. Il a été montré que la structure en épingle à cheveu au niveau 5' du promoteur est essentielle à

l'activité endonucléase de PA lors du mécanisme de vol de coiffe pour la formation de l'ARNm (Leahy et al. 2001). De plus, une autre structure de type tige-boucle secondairement induite au niveau 3' terminale de chaque segment participerait aussi à cette activité endonucléasique (Leahy, Dobbyn, and Brownlee 2001). La formation de la queue polyA est aussi sous la dépendance de structures secondaires d'ARN au niveau du promoteur via un mécanisme de bégaiement de la polymérase, entrainant la copie réitérative d'uraciles présents entre les positions 15 et 17 depuis l'extrémités 3' terminal (De Vlugt, Sikora, and Pelchat 2018; Pflug et al. 2017; 2014; Luo et al. 1991).

D'autre part, le promoteur est impliqué aussi dans la reconnaissance des virus influenza A par le système immunitaire innée. La protéine RIG-I est majoritairement activée en présence de ARNv, induisant la production d'IFN. La structure en manche (*panhandle*) du promoteur et l'extrémité 5'triphosphate sont tous deux des signaux agonistes nécessaires pour l'activation de RIG-I (M.-K. Lee et al. 2016; Ferhadian et al. 2018).

Outre la région duplex proche du promoteur en 5' et 3' terminales, d'autres zones au niveau interne des gènes viraux sont impliquées dans des interactions entre les segments, permettant un empaquetage viral efficace. Une étude réalisée sur le segment NP a montré la présence de structure de type pseudo-nœud (Gultyaev et al. 2014). D'autres structures de type épingle à cheveux ont été prédites sur le segment M, dont la perturbation par mutagenèse dirigée entraine une diminution de la capacité réplicative chez la souche A/Puerto Rico/8/34 (PR8, H1N1), suggérant un rôle dans l'empaquetage du génome (Kobayashi et al. 2016). Ces interactions inter-segments semblent être sous-type-spécifiques et pourraient en partie expliquer les mécanismes de réassortiment préférentiels entre plusieurs souches. De telles interactions ont été décrites entre deux virus de sous-type H3N2 et H5N2 (Gavazzi et al. 2013). De plus, ces interactions semblent se réaliser au niveau de régions particulières. Chez H5N2, les segments NS et PB1 interagissent au niveau de deux structures tiges boucles, à l'instar du mécanisme de *kissing-loops* observé chez les rétrovirus. Malgré la spécificité de ces mécanismes (Gavazzi et al. 2013), plusieurs autres interactions impliquant des structures d'ARN différentes peuvent avoir lieu (**Figure 10**), témoignant une capacité d'adaptation et une certaine flexibilité des virus influenza A pour l'empaquetage efficace de leur génome ainsi que les réassortiments lors de coïnfection (Dadonaite et al. 2019).

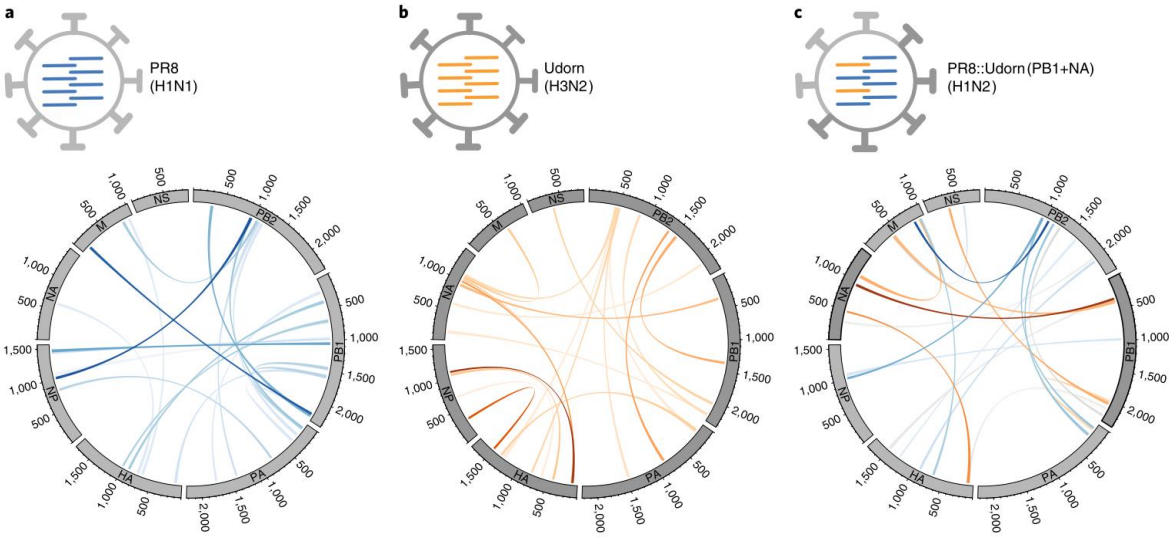


Figure 10 *: Cartes du réseau d'intéractions ARN-ARN entre les segments de la souche PR8 (H1N1) (a), Udorn (H3N2) (b) et le réassortant PR8::Udorn (PB1+NA) (c), d'après* (Dadonaite et al. 2019)

La régulation du métabolisme de l'ARNm notamment l'épissage est aussi soumis à la présence de structures de secondaires. De tels mécanismes régulant la production d'ARNm alternatifs codant pour NS1 - NS2 (ou NEP) et M1 – M2 ont été décrits sur les segments NS et M, respectivement (Ferhadian et al. 2018; Jiang et al. 2016).

D'autre part, certaines régions structurées au niveau du segment HA semble être soumises à une pression de sélection positive et joueraient potentiellement un rôle dans l'évolution. Des études bio-informatiques combinant des prédictions de structures secondaires d'ARN et des analyses de covariations indiquent qu'au niveau de la région codant le site de clivage de HA une structure de types épingle à cheveux serait présente et conservée chez les tous les sous-type Influenza A. Des analyses de cartographie de l'ARN montrent que ces zones sont plutôt faiblement recouvertes de NP, corroborant ainsi avec les données de prédictions de structures. Cependant, des études supplémentaires mettant directement en évidence la présence de ces structures prédites sont nécessaires afin de déterminer leur impact réel dans l'évolution des virus influenza aviaires (Williams et al. 2018; Gultyaev et al. 2016; Dawson, Lazniewski, and Plewczynski 2017; Dadonaite et al. 2019; Ferhadian et al. 2018).

IV. Ecologie virale des virus influenza A

Les virus influenza A infectent une très grande diversité d'espèces animales. Ils ont été détectés chez plus d'une centaine d'espèces d'oiseaux répartis dans 13 ordres différents (Munster and Fouchier 2009; Vandegrift et al. 2010; Ohuchi et al. 1989a; Webster et al. 1992). Les oiseaux sauvages sont infectés par la majorité des virus influenza A (**Figure 11**). Une multitude de virus issus de la combinaison entre les sous-types HA1-16 et NA 1-9 ont été isolés chez ces hôtes et sont dénommés virus influenza aviaires (Mostafa et al. 2018b).

Les oiseaux aquatiques de l'ordre des Ansériformes (incluant les canards, les oies et les cygnes) suivis par les Charadriiformes (les mouettes, les goélands, les sternes, les bécasseaux, les pluviers) constituent les réservoirs naturels des virus influenza aviaires. D'autres oiseaux aquatiques comme les espèces appartenant aux ordres de Ciconiiformes, Gaviiformes, Gruiformes, Pelecaniformes, Podicipediformes et Procellariiformes sont moins fréquemment infectés. D'autres oiseaux terrestres sont occasionnellement infectés par des virus influenza A, tels que les espèces des ordres des Columbiformes, Falconiformes, Passeriformes, Piciformes et Piscittaciformes. Cela suggère que ces derniers jouent un rôle de « *spill-over* » de l'épizootie (impliqués dans la propagation secondaire du virus chez d'autres espèces : effet de « débordement ») après un contact avec des volailles ou des oiseaux aquatiques infectés (Olsen et al. 2006; Wahlgren 2011; Vandegrift et al. 2010; França, Stallknecht, and Howerth 2013).

Chez les Ansériformes, les canards barboteurs (genre *Anas* classés au sein de la famille des *Anatidae*) sont les plus fréquemment infectés par les virus influenza aviaires, en particulier l'espèce *Anas platyrhynchos* (canard col-vert). La quasi-totalité des souches issues des combinaisons entre HA 1-16 et NA 1-9 ont été retrouvées chez cet hôte. Une circulation plus intense de ces virus (prévalence entre 10 et 60%) est observée avant et pendant l'automne en Eurasie et en Amérique du Nord chez *Anas platyrhynchos*. Une plus faible prévalence d'individus infectés est observée durant le printemps et l'été (moins de 10%)(Vandegrift et al. 2010; Spackman 2009).

Chez les oiseaux classés au sein des Charadriiformes, composés par les oiseaux de rivage, les goélands et les sternes, le schéma épidémiologique des virus influenza aviaires diffère par rapport aux Ansériformes. Deux sous-types de HA : HA13 et H16, ont été isolés presque exclusivement chez les Charadriiformes. De plus, les oiseaux de rivage révèlent un pic de prévalence d'infection durant le printemps et l'été. Ce décalage temporel en termes de prévalence d'infection par rapport aux Ansériformes suggère que les Charadriiformes sont

d'importantes espèces dans le maintien des virus influenza aviaires en dehors de la période hivernale et leur persistance chez les oiseaux sauvages au niveau des zones de reproduction (Vandegrift et al. 2010; Spackman 2009; Olsen et al. 2006).

Certaines caractéristiques liées aux niches écologiques des oiseaux aquatiques sauvages et aux virus influenza aviaires permettent d'expliquer une circulation importante et efficace de ces virus au sein de ces populations d'animaux. Une très forte densité et variété d'espèces sont retrouvés dans les lieux de rassemblements d'un grand nombre d'espèces d'oiseaux sauvages (Olsen et al. 2006). De plus, le virus est excrété en de très grandes quantités via l'appareil respiratoire et surtout au niveau des selles. Ces facteurs favorisent les infections interindividuelles et la maintenance de ces virus au sein des populations d'oiseaux. Un brassage génétique important des virus influenza aviaires est alors observé (Mostafa et al. 2018b; Olsen et al. 2006; Wahlgren 2011). D'autres part, des caractéristiques virologiques et environnementales favorisent la propagation et la persistance de ces virus dans le biotope de ses hôtes naturels. En effet, les virus influenza A peuvent persister plusieurs semaines dans un environnement humide et froid (Brown et al. 2009; Stallknecht et al. 1990).
Compte tenu du spectre d'hôte naturel des virus influenza A, la distribution et la transmission de ces virus est dépendant des mouvements de populations d'espèces d'oiseaux, qui, pour une large partie d'entre-elles sont migratrices. Différents couloirs majeurs de migrations ont pu être décrits (Olsen et al. 2006).

.

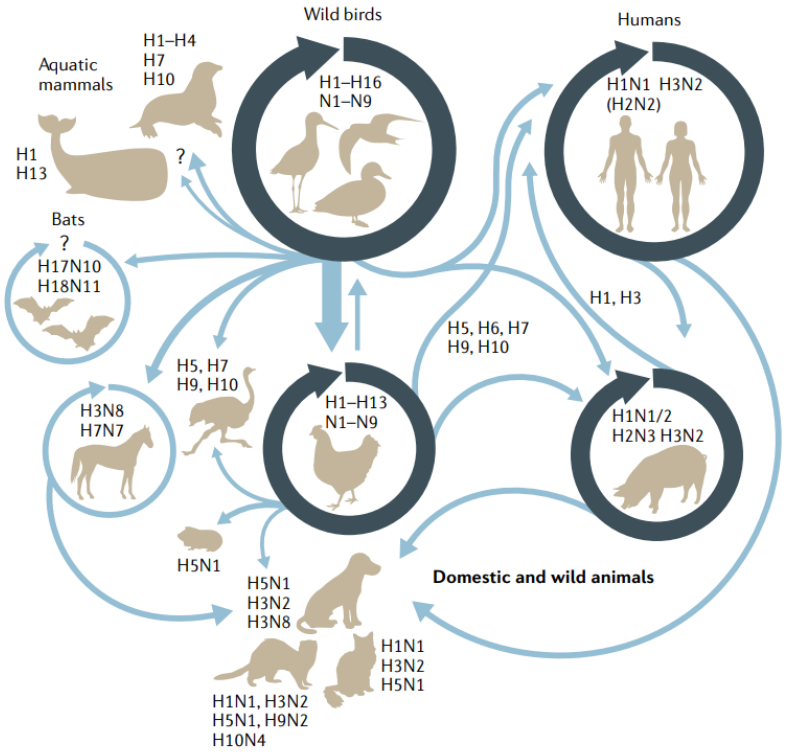


Figure 11 *: Diversité des échanges virologiques au sein des oiseaux et des espèces de mammifères susceptibles et sensibles à l'infection par les virus influenza A. Les différents sous-types concernés par chaque espèce sont indiqués (modifiée, d'après J. S. Long et al., 2019).*

Les volailles (oiseaux domestiqués, utilisés pour la production de viande ou d'œufs destinés à la consommation humaine, pour la reproduction de ce type d'oiseaux, ou pour la production de gibier de repeuplement ou d'autres produits commerciaux, constituées par des Gallinacées (dindes, poulets, cailles, poules, pintades, faisans) et des palmipèdes (canards domestiques, oies)) sont infectés par contacts directs ou indirects (transmission via l'environnement : sol / eau contaminé ou fomites) avec des oiseaux sauvages contaminés (Wahlgren 2011; Vandegrift et al. 2010; Webster et al. 1992). L'infection démarre comme chez ces derniers au niveau de l'appareil respiratoire supérieur et se disperse au niveau du tube digestif (**Figure 12**). En comparaison aux Ansériformes et Charadriiformes, les dindes, poulets et cailles sont généralement plus sensibles à l'infection aux virus influenza aviaires . Certains virus influenza aviaires sont à l'origine d'épidémies dramatiques au sein des volailles terrestres. Les virus H5N1 et H7N9 hautement pathogènes apparus lors des années 1996 et 2013 respectivement en Chine en font partie (Guan and Smith 2013; Liu et al. 2018; D. H. Lee et al. 2017; Qi et al. 2017) . Ces derniers ont conduit à un taux de mortalité très élevé chez le poulet, qui sera explicité plus en détails dans un paragraphe suivant **(**Les différents phénotypes d'influenza aviaires**).**

D’autres espèces animales autres que les oiseaux sont aussi sujets à l’infection aux virus influenza A, notamment les mammifères marins, les porcs, les furets, les humains, les chevaux, les chats et les chiens. De plus, les chauves-souris constituent un autre réservoir de virus influenza aviaires à prendre en compte (Wahlgren 2011; Webster et al. 1992). Une épidémie d’influenza aviaire H10N7 a frappé la population de phoques au nord de L’Europe, accompagnée d’une augmentation significative de la mortalité. Le virus responsable a la capacité d’infecter des furets par aérosols et son hémagglutinine lie préférentiellement les acides sialiques α2,6 (Herfst et al. 2020).

Par ailleurs, un rapport récent a décrit une épidémie d’influenza aviaire hautement pathogène H5N1 dans une exploitation de vison au nord-ouest de l’Espagne, accompagnée d’une augmentation de la mortalité (Agüero et al. 2023). L’analyse moléculaire a montré que la souche responsable appartient au clade 2.3.4.4b, ce qui correspond à l’épizootie actuellement active en Europe. De plus, une mutation peu commune (T271A) a été identifiée sur le segment PB2 des virus isolés sur les visons, dont les implications au niveau de son adaptation chez l’homme doivent être surveillées. Cette mutation particulière n’est pas présente dans la majorité des souches H5N1 circulantes en Europe, suggérant une apparition spontanée chez le vison.

En outre, lors d’épidémies importantes d’influenza aviaire chez la volaille, des transmissions directes de virus aviaires vers l’Homme ont été décrits, notamment lors de contacts rapprochés. Les infections ont été accompagnés par des signes cliniques neurologiques et/ou pulmonaires sérieuses, avec des issues parfois fatales (F. Zhang et al. 2017; Subbarao et al. 1998; Claas et al. 1998). Bien que la transmission interhumaine de ces virus semble pour l’instant être limitée, ces observations ainsi que les adaptations chez des plusieurs hôtes mammifères montrent bien le risque zoonotique élevé que constituent les virus influenza aviaires. Dans ce sens, une surveillance rapprochée de ces agents infectieux est primordiale. La plupart des virus influenza A circulants chez l’Homme proviennent provient de virus ayant circulés dans un premier temps chez le canard de surface (ou barboteurs) et ensuite chez le porc (Wahlgren 2011; Webster et al. 1992; Mostafa et al. 2018a). Ces derniers sont devenus endémiques au sein de la population humaine et constituent les grippes saisonnières.

V. Les différents phénotypes d'influenza aviaires

Les virus influenza aviaires peuvent être classés en deux catégories en fonction de leur pathogénicité chez les volailles : les virus influenza aviaires faiblement pathogènes (LPAIV : *low pathogenic avian influenza virus*) et hautement pathogènes (HPAIV : *highly pathogenic avian influenza virus*).

Généralement, les LPAIV causent peu ou pas de symptômes, que ce soit chez les oiseaux sauvages aquatiques (Ansériformes et Charadriiformes), captifs ou d'élevage. Les LPAIV se répliquent principalement au niveau de l'appareil respiratoire supérieur et des cellules épithéliales du tube digestif des oiseaux. De très grandes quantités de virus sont excrétés via les fèces, expliquant une contamination importante de l'environnement et une transmission principalement oro-fécale chez ces hôtes (Olsen et al. 2006; Wahlgren 2011). Malgré une clinique majoritairement inapparente, certaines souches LPAIV peuvent perturber le comportement migratoire chez les canards (Vandegrift et al. 2010).

Dans le cadre d'une infection avec une souche LPAIV, l'infection sera restreinte au niveau de l'appareil respiratoire et du tube digestif. Le tableau clinique d'une infection est peu marqué. Classiquement, 5% en moyenne de mortalité est observée chez ces individus et des signes de morbidité est identifié dans 50% des cas. Les symptômes les plus retrouvés sont des signes respiratoires comme des éternuements, la toux, des râles, des secrétions oculaires, ainsi que de la diarrhée (Steinhauer 1999).

Lors d'une infection avec une souche HPAIV chez un Galliforme, un tableau clinique plus sévère est observé. Des hémorragies, des œdèmes, des nécroses et le dysfonctionnement de la majorité des organes sont observés (Steinhauer 1999). Des symptômes respiratoires et/ou neurologiques majeurs caractérisent le tableau clinique, dont l'issue est souvent fatale. L'infection est associée à une sécrétion importante de cytokines pro-inflammatoires responsable en grande partie des signes cliniques (Kuribayashi et al. 2013). Chez le poulet et la dinde, les HPAIV causent une infection systémique létale pouvant provoquer 100 % de mortalité au cours des 48 heures suivant l'infection (David E. Swayne 2007; D E Swayne and Pantin-Jackwood 2006). L'Organisation Mondiale de la Santé Animale (OMSA) prévoit une définition pour la classification des souches H5 et H7 influenza aviaire suivante :

« Les virus responsables de l'influenza aviaire à déclaration obligatoire peuvent être classés en deux catégories : virus hautement pathogènes et virus faiblement pathogènes :

Les virus responsables de l'influenza aviaire à déclaration obligatoire hautement pathogènes ont un indice de pathogénicité par voie intraveineuse supérieur à 1,2 chez les poulets âgés de six semaines, ou entraînent une mortalité d'au moins 75 % chez les poulets âgés de quatre à huit semaines infectés par voie intraveineuse. Les virus de sous-types H5 et H7 dont l'indice de pathogénicité par voie intraveineuse n'est pas supérieur à 1,2 ou qui entraînent une mortalité inférieure à 75 % lors d'un test de létalité par voie intraveineuse, doivent être séquencés pour déterminer si de multiples acides aminés basiques sont présents sur le site de clivage de la molécule d'hémagglutinine (HA0) ; si la séquence d'acides aminés est similaire à celle observée chez d'autres virus responsables de l'influenza aviaire à déclaration obligatoire hautement pathogènes isolés précédemment, il doit être considéré qu'il s'agit du virus responsable de la forme hautement pathogène de la maladie. Les virus responsables de l'influenza aviaire à déclaration obligatoire faiblement pathogènes comprennent tous les virus influenza de type A appartenant aux sous-types H5 et H7, qui ne sont pas des virus responsables de la forme hautement pathogène de la maladie. »
L'indice de pathogénicité représente le score clinique moyen quotidien pour chaque animal observé à l'issue une période de 10 jours après inoculation intraveineuse chez 10 poulets âgés de quatre à huit semaines. Un score de zéro est attribué à un animal non malade, un pour un animal malade, deux pour un animal très malade ou paralysé et trois pour un animal mort.

L'isolement d'une souche HPAIV au sein des Charadriiformes reste rare. Dans la majorité des cas, une infection avec une souche HPAIV induit peu de mortalité et de morbidité chez les hôtes appartenant aux ordres des Ansériformes et des Charadriiformes (Vandegrift et al. 2010).

On note très fréquemment une différence de sensibilité à l'infection avec une souche HPAIV entre l'hôte canard et poulet. Celle-ci semble être multifactorielle, impliquant notamment la différence de tropisme des virus et la réponse immunitaire de l'hôte (Steinhauer 1999; Evseev and Magor 2019). La réponse immunitaire innée caractérisée par la sécrétion d'inférons béta (IFN-α et β) et la synthèse des ISGs (*Interferon Stimulating Genes*) est beaucoup plus précoce, efficace et régulée chez le canard. L'absence de la protéine RIG-I et une réponse inflammatoire massive chez le poulet expliqueraient en partie sa vulnérabilité accrue face à l'infection (Evseev and Magor 2019).

Au niveau moléculaire, la HA d'un LPAIV va être clivée ou maturée par des protéases extracellulaires de type trypsine reconnaissant un motif de clivage monobasique. Ces enzymes sont exclusivement exprimées au niveau de l'appareil respiratoire et du tube digestif de l'hôte.

Les souches LPAIV ont un potentiel d'évolution vers des formes HPAIV. Cette évolution a été observée naturellement chez les sous-types H5 et H7. Elle est caractérisée par l'acquisition d'acides aminées basiques au niveau du site de clivage de HA (Abdelwhab, Veits, and Mettenleiter 2013). L'obtention d'un motif polybasique (MBCS) va considérablement changer le tropisme du virus (**Figure 12**). La maturation protéolytique de HA va pouvoir être effectuée par d'autres protéases golgiennes de type furine dont l'expression est ubiquitaire. Le virus va alors pouvoir infecter d'autres types cellulaires, tels que les cellules endothéliales et provoquer une extension de l'infection au niveau de la circulation sanguine et lymphatique, expliquant le tableau clinique sévère avec une infection généralisée dans le cas d'un HPAIV (Steinhauer 1999).

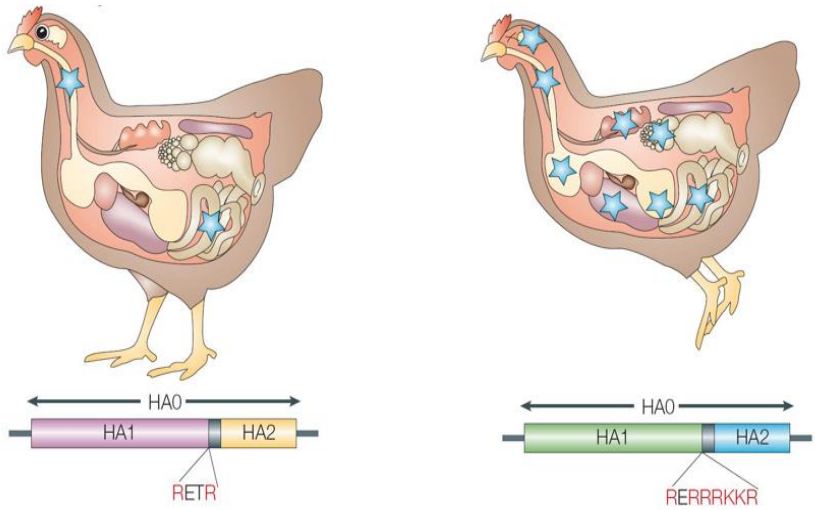


***Figure 12** : Différences protéiques au niveau du site de clivage de la HA entre les virus influenza aviaires faiblement (LPAIV, à gauche) et hautement pathogènes (HPAIV, à droite), puis les conséquences cliniques chez le poulet (d'après Horimoto et al., 2005). E, T, K et R représentent des acides aminés dont les deux derniers sont basiques : respectivement les résidus lysine (K) et arginine (R).*

VI. Plasticité du génome des virus influenza A

1. Concepts généraux sur l'évolution des virus influenza A

Les virus influenza évoluent selon deux types de mécanisme, que sont le réassortiment génétique et l'accumulation progressive de mutations.

Lors de la coinfection d'une même cellule, les virus du même genre peuvent échanger leurs segments de leur génome respectif et ainsi former de nouveaux virus via l'association de ces gènes par réassortiment (**Figure 13**). Ce phénomène peut être responsable de l'apparition de nouveaux virus au sein d'une population immunologiquement naïve par changement radicale des glycoprotéines virales, et pourra être à l'origine des grandes pandémies de grippe aviaire. Cette « cassure antigénique » a eu lieu à plusieurs reprises dans l'histoire : la grippe espagnole (1918-1920, causée par une souche H1N1), asiatique (1957-1958, H2N2), de Hong Kong (1968-1970, H3N2) et porcine (2009-2011, H1N1), présentant ce pathogène comme un problème majeur en santé humaine et animale (Bjørnstad and Viboud 2016; Monne et al. 2014; Nelson and Holmes 2007).

L'ARN polymérase virale ARN dépendante des virus influenza A est peu fidèle. Un taux de mutations pouvant atteindre 10^{-4} substitutions/nucléotide/matrice répliquée en moyenne a été calculé chez certaines souches (Pauly, Procario, and Lauring 2017). Cette forte variabilité entraine la formation de populations de génomes différents, appelées quasi-espèces (Bragstad, Nielsen, and Fomsgaard 2008; Barbezange et al. 2018). Les épidémies saisonnières touchant la population humaine sont principalement à l'origine d'acquisitions de mutations (substitutions) au niveau des glycoprotéines virales permettant d'échapper à la pression du système immunitaire et responsables d'un « glissement antigénique » progressif (Nelson and Holmes 2007) (**Figure 13**). Les virus influenza A humains circulants comme H3N2 acquièrent en moyenne 3 à 4 changements d'acides aminés par an au niveau de la protéine HA, témoignant de l'évolution rapide de ces virus et de la nécessité de réviser continuellement la stratégie de vaccination antigrippale (Fitch et al. 1997; Bhatt, Holmes, and Pybus 2011).

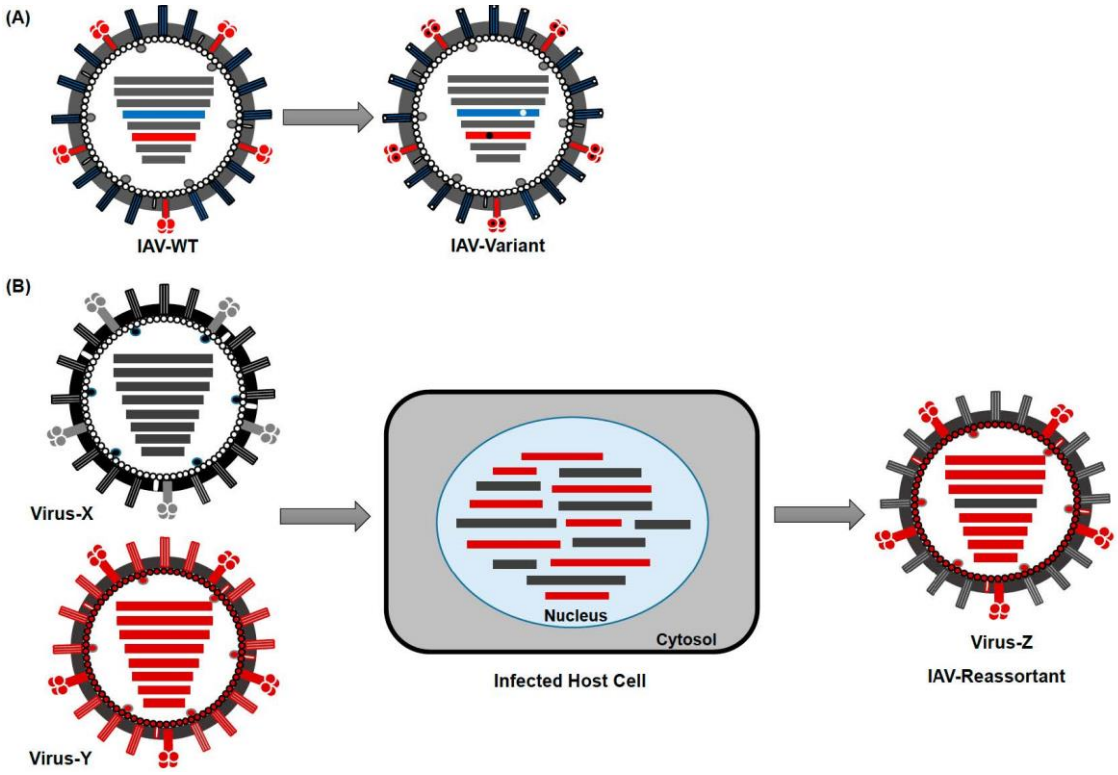


Figure 13 *: Mécanismes d'évolution des les virus influenza A. (A) Accumulation progressives de mutations au niveau du génome des virus influenza A, conduisant à la génération de nouveaux variants : dérive antigénique, responsable des épidémies samsonières chez l'Homme (« antigenic drift »). (B) Echange ou réassortiment des segments entre deux souches virus influenza A dans le cadre d'une coïnfection d'une cellule peut entrainer l'émergence d'une nouvelle souche (nouveau sous-type) : cassure antigénique (« antigenic shift »). Ce dernier mécanisme est observé dans le cas de pandémies (d'après Mostafa et al, 2018).*

2. Effets des mutations et leurs interactions

Au vu de la très grande plasticité du génome des virus influenza A, il semble que ces derniers aient des capacités d'adaptation, d'échappement immunitaire, et plus largement d'évolution, illimitées. Cependant, il est important de noter que la majorité des mutations observées ont un impact négatif sur la capacité réplicative (ou *fitness)* de ces virus.

A l'échelle de mutations uniques, il a été estimé qu'environ 40 % d'entre-elles sont délétères et 50 % ont un effet neutre. Moins de 10 % seulement des mutations uniques sont bénéfiques (Lyons and Lauring 2018). En accord avec ce qui est largement admis, les mutations non-synonymes sont plus délétères. Par ailleurs, les parties non-codantes des segments des virus influenza A sont beaucoup moins tolérantes aux mutations. Deux tiers de celles-ci sont létales, ce qui reflète les fonctions critiques que détiennent ces séquences promotrices. Effectivement,

elles constituent les zones de fixation du complexe polymérase virale, permettent l'enclenchement de la réplication/transcription et contrôle également l'empaquetage du génome (Dawson, Lazniewski, and Plewczynski 2017; Visher et al. 2016). Cette vulnérabilité génomique est présente aussi chez d'autres virus à ARN et ADN, ce qui renvoie à des contraintes génétiques communes au sein de ces organismes pathogènes (Elena, Solé, and Sardanyés 2010; Sanjuán and Elena 2006).

Des différences au niveau de la distribution des mutations ont été remarquées au niveau des gènes viraux. Les segments HA, NA et NS (séquence codante de NS1) semblent avoir une plus grande tolérance aux mutations. Plus précisément, on distingue des zones très polymorphiques au niveau de la zone codant la tête globulaire de HA, correspondant aux domaines antigéniques majoritaires (J. M. Lee et al. 2018; Heaton et al. 2013; Plotkin and Dushoff 2003). Ces différences importantes de variabilité de séquences selon les régions reflètent une intensité de pression de sélection différentielle, principalement médiée par la réponse immunitaire et des rôles fonctionnels des protéines codées par ces séquences.

Les mutations peuvent apparaître sur des contextes génomiques très différents. De plus, à cause du fort taux d'erreur de l'ARN polymérase virale, plusieurs mutations peuvent apparaître chez un même virus, voir sur un même segment. Des interactions négatives ou positives complexes entre ces mutations se forment, et ont des conséquences notables sur le *fitness* du virus. Ce phénomène est référé sous le nom d'épistasie (Elena, Solé, and Sardanyés 2010), et conditionne fortement l'évolution des virus grippaux (Lyons and Lauring 2018). Théoriquement, l'épistasie est définie comme la différence observée entre i. le *fitness* d'un génome x avec la somme des mutation considérés, ii. et le *fitness* « prédit » du génome x avec le même jeu de mutation, dont leur effet est pris en compte individuellement (Mani et al. 2008). Ainsi, plusieurs interactions ont pu être décodées chez les virus influenza A, principalement au niveau de leur impact sur l'échappement au niveau de la réponse immunitaire et de l'émergence de résistances aux antiviraux. La revue écrite par Daniel M. Lyons et Adam S. Lauring offre une vue détaillée du mécanisme d'épistasie chez les virus influenza A (Lyons and Lauring 2018). De façon étonnante, il s'avère qu'une grande partie des mutations, même neutres vis-à-vis du *fitness* viral, interagissent et conditionnent fortement l'adaptation et l'évolution du génome viral (Koelle et al. 2006; Tria, Pompei, and Loreto 2013; Nshogozabahizi, Dench, and Aris-Brosou 2017).

3. L'évolution des virus influenza chez les oiseaux et l'Homme

Plusieurs études phylogénétiques suggèrent que les virus influenza A circulants chez la plupart des mammifères dérivent d'un virus ancestral d'origine aviaire (Bean et al. 1992; Owen T Gorman et al. 1991; 1990; Ito et al. 1991; O T Gorman et al. 1990). L'évolution des virus influenza A au sein des hôtes aviaires et humains se distingue sur plusieurs points (Webster et al. 1992) .

Tout d'abord, les virus influenza aviaires ont conjointement évolué avec leurs hôtes naturels, rendant ces virus très adaptés aux oiseaux aquatiques sauvages (Webster et al. 1992). Ces virus évoluent plus lentement chez les oiseaux en comparaison des souches humaines, du fait d'une politique de vaccination des oiseaux beaucoup moins importante, entrainant une pression de sélection immunitaire moins intense.

Chez l'Homme, la vaccination contre les virus influenza (A et B) est plus importante. Par conséquent, une plus forte pression de sélection induite par la réponse immunitaire est exercée sur ces virus (J. M. Lee et al. 2018; Xue et al. 2018), ce qui intensifie l'évolution de ces virus chez l'Homme. Par ailleurs, il est bon de rappeler que nous sommes en contact avec ces virus que relativement récemment d'un point de vue de l'évolution virale (Webster et al. 1992).

VII. Evolution des virus influenza aviaires vers les formes hautement pathogènes

1. L'apparition du site de clivage polybasique

Comme énoncée précédemment (page 30), nous pouvons distinguer deux groupes de virus influenza chez les oiseaux : les virus influenza aviaire faiblement (LPAIV) et hautement pathogènes (HPAIV). Les HPAIV sont naturellement issus de LPAIV qui ont évolué, aboutissant à l'apparition d'un motif protéique polybasique au niveau du site de clivage de HA (Abdelwhab, Veits, and Mettenleiter 2013). L'apparition de ce motif au niveau de HA est considérée comme le marqueur de pathogénicité principal.

Différents mécanismes génétiques ont été proposés comme étant responsables de l'apparition du motif polybasique au niveau du site de clivage de HA : la recombinaison génétique avec des gènes cellulaires ou viraux, des erreurs de la polymérase virale (substitutions) et des insertions

de nucléotides (Abdelwhab, Veits, and Mettenleiter 2013; Abolnik 2017; Lowen 2017; Nao N, Yamagishi J et al. 2017; Pasick et al. 2005).

Le phénomène de glissement ou de bégaiement de la polymérase virale (*polymerase slippage* ou *stuttering*) participerait à l'évolution des génomes viraux en provoquant l'insertion de nucléotides par duplication de codon (García et al. 1996; M L Perdue et al. 1997; Kolakofsky 2016; Nao N, Yamagishi J et al. 2017). Certaines régions du génome riches en base en répétée seraient potentiellement favorables à ce phénomène. Dans le cas des LPAIV de sous-type H5 et H7, la région codant le site de clivage est caractérisé par une présence importante en A (ou en U sur l'ARNv) et correspond à l'emplacement d'une structure secondaire d'ARN conservée de type tige-boucle (cf page 22), dont le lien entre l'ampleur de la boucle apicale prédite et les évènements d'insertions a été suggéré, chez les virus influenza de sous-type H5 et H7 (Abolnik 2017; Nao N, Yamagishi J et al. 2017). Cependant, cette étude n'a pas pris en compte l'impact que peut avoir le complexe polymérase au cours de la réplication sur l'intégrité de cette structure (elle-même prédite), ce qui limite l'interprétation du lien entre la structure réelle présente au niveau de la région codant le site de clivage de HA, restant encore inconnue à ce jour, et les événements d'insertions. Pour certaines émergences, il a aussi été d'observé des acquisitions de site de clivage HA polybasique combinant plusieurs mécanismes, dans lesquels certaines substitutions ont précédés des évènements d'insertions de nucléotides, suggérant un modèle d'évolution « progressif », concernant la transition LPAIV-HPAIV (García et al. 1996; Abolnik 2017; Bashiruddin, Gould, and Westbury 1992; Michael L Perdue et al. 1997; Abbas et al. 2010). Une étude récente a mise en évidence la présence importante de bases de type purine au niveau de la région codant le site de clivage HA chez les séquences HA de sous-type H5 et H7 par rapport à des séquences HA d'autres sous-types (Funk et al. 2022). De plus, seulement quelques évènements de substitution sont nécessaires pour la création d'un site de clivage polybasique chez les séquences HA de sous-type H5 et H7 (Funk et al. 2022) en comparaison aux autres séquences, ce qui peut expliquer la prédisposition naturelle des LPAIV-H5 et H7 de pouvoir évoluer vers un phénotype hautement pathogène.

Les virus influenza aviaires sont capables d'évoluer par recombinaison, malgré le fait que ce dernier mécanisme soit rare chez l'ensemble des virus influenza A (Boni et al. 2008; Han and Worobey 2011). De façon surprenante, seuls des événements de recombinaisons non-homologues ont été décrits chez des virus influenza aviaires de sous-type H7 (Sun, Belser, and Tumpey 2016; Pasick et al. 2005; Abdelwhab, Veits, and Mettenleiter 2013). Les raisons qui expliqueraient les restrictions qu'invoquent ces types de mutations sont encore inconnues. Des

recombinaisons entre des gènes cellulaires (ARN ribosomique 28S) ou d'autres segments viraux (NP et M) et le segment HA ont été décrites chez un nombre conséquent de virus influenza aviaires de sous-type H7 (Pasick et al. 2005; Abdelwhab, Veits, and Mettenleiter 2013). Du fait de la présence de structure de type tige boucle conservée au niveau du site de clivage de HA, un mécanisme de saut de polymérase et changement de matrice au niveau de cette structure a été proposé comme favorisant la recombinaison, à l'image de ce qui peut être observé au sein des rétrovirus (Moumen et al. 2003; Simon-Loriere et al. 2010). Dans une étude récente, il a été montré que chez la majorité des segments HA dérivant de virus HPAIV-H7 ayant évoluées par recombinaison, les séquences insérées exogènes en amont de la région codant pour le site de clivage constituent une cible de petits ARN snoRNA, impliqués dans les régulations épigénétiques des gènes de l'hôte (Gultyaev et al. 2021). La piste d'un lien potentiel entre la liaison spécifique de certains snoRNA et la présence de la tige boucle d'ARN conservée au niveau de la région codant le site de clivage de HA serait intéressante à investiguer.

2. Les autres mutations localisées à distance de la région codant le site de clivage HA et associées au phénotype hautement pathogène

Bien que l'acquisition du motif polybasique de clivage sur HA constitue le facteur de pathogénicité principal, la transition phénotypique LPAIV-HPAIV n'est pas strictement expliquée par l'acquisition d'un tel motif sur HA. D'autres mutations ont été décrites comme contribuant à l'augmentation de la pathogénicité chez les hôtes aviaires, suggérant une signature polygénique possible de l'évolution LPAIV-HPAIV.

Les facteurs modifiant la conformation de la protéine HA influencent sa stabilité, sa fonction importante au niveau de l'entrée, du tropisme cellulaire/tissulaire, et par extension, la pathogénicité du virus. D'un point de vue structural, le site de clivage de HA constitue une boucle accessible aux protéases cellulaires. Au niveau du trimère HA0 immature replié, cette boucle est positionnée à proximité de l'Asn11 où un site potentiel de glycosylation est présent (Steinhauer 1999).

Les sites de glycosylation jouent un rôle important dans la régulation de l'accès des protéases environnantes au niveau du motif de clivage de HA. La mutation T13K responsable d'une perte d'une N-glycosylation a été observé chez un virus H5N2 isolé en Pennsylvanie en 1983 qui a été responsable d'une épidémie d'influenza aviaire accompagnée d'une très grande mortalité (Kawaoka and Webster 1985; Ohuchi et al. 1989b; Abdelwhab, Veits, and Mettenleiter 2013).

D'autres mutations ayant lieu au niveau des positions 10-12 ont été décrites comme pouvant augmenter la pathogénicité. D'autre part, certains points de glycosylation influencent la réplication virale et la transmission, participant *in fine* à la virulence. L'histidine en position 184 régule la dépendance au pH, et par conséquent, impacte le changement conformationnel de HA. Des mutations au niveau de ce résidu ou son entourage participent à l'adaptation de certaines souches vers d'autres hôtes. De même, la mutation E216R a aussi été observé chez certains virus H5N1 hautement pathogène (Mair et al. 2014).
De plus, La présence d'une glycosylation sur le résidu Asn188 de la souche A/chicken/Victoria/76 (H7N7) augmente sa pathogénicité observée chez le poulet en comparaison avec un variant non glycosylé (Michael L. Perdue, Latimer, and Crawford 1995).

Par ailleurs, d'autres mutations sont apparues à distance du site de clivage et sur d'autres gènes de certaines souches HPAIV. Ces dernières sont retrouvées principalement sur le segment HA et ceux codant les sous-unités PB1 et PB2 du complexe polymérase. Une étude utilisant une approche phylogénétique et statistique a permis de mettre en évidence certaines mutations significativement associées au phénotype hautement pathogène (Escalera-Zamudio et al. 2020). Chez les virus de sous-type H7, les mutations A143T, M/L274I, V438I, D384N ayant apparues sur le segment HA ont été significativement associés au phénotype hautement pathogène. Les mutations R355K, V480I sur PB2 ainsi que S152L sur PB1 ont aussi été relevées. Concernant les virus de sous-type H5, les mutations K379R, M242I, S145L (ou délétion), N154I/L/Q, F127L, I167T, T204I sur HA, A674T et E627K sur PB2, ainsi que V113I sur PB1 ont été identifiées comme significativement associées au phénotype hautement pathogène. Certaines de ces mutations ont été validées via une approche structurale, par inférence de l'effet de ces mutations sur la stabilité de la protéine virale impliquée et sa fonction. Cependant, il est encore difficile de déterminer si ces mutations sont motrices de l'évolution LPAIV-HPAIV, ou ne constituent que le résultat d'adaptation ou de compensations de l'évolution vers un HPAIV. Des études expérimentales ayant pour but de vérifier l'impact de ces mutations sur l'évolution LPAIV-HPAIV devraient être envisagées.

3. Pathogénicité, acquisition du MBCS et autres sous-types HA

Comme énoncé précédemment, la pathogénicité des virus HPAIV n'est pas exclusivement apportée par l'acquisition du MBCS. Un exemple bien documenté est celui de l'épidémie ayant

eu lieu en Pennsylvanie, aux Etats-Unis en 1983. L'analyse des séquences des premières souches isolées au début de l'épidémie a montré la présence du MBCS suivant : KKKR|GLF, malgré un indice de pathogénicité par voie intraveineuse (IPVI) nul. Les analyses effectuées sur d'autres virus isolés plus tard dans l'épidémie ont mis évidence une augmentation significative de la pathogénicité (IPVI > 2), mais le site de clivage HA restait inchangé. Ce changement de pathogénicité a été attribué à une substitution en position 22 sur la protéine HA, entrainant une perte d'un site de N-glycosylation, rendant potentiellement accessible le motif de clivage HA aux protéases cellulaires (Kawaoka and Webster 1985; Ohuchi et al. 1989b; Abdelwhab, Veits, and Mettenleiter 2013).

D'autre part, certains virus de sous-type non H5/7 ont aussi naturellement acquis un MBCS. Cependant, l'obtention d'un tel motif n'est pas forcément accompagnée d'un changement au niveau de la pathogénicité. Un exemple de virus de sous-type H4N2 a été caractérisé avec un site de clivage HA suivant : PEKRRTR, bien qu'il soit faiblement pathogène (IPVI = 0) (Wong et al. 2014). L'extension du motif de clivage polybasique n'a montré aucun effet sur sa pathogénicité. En revanche, l'ajout de gènes internes d'un virus HPAIV-H5 a permis de rendre ce virus modestement indépendant à la trypsine (Gischke et al. 2020). Par ailleurs, plusieurs virus H9N2 issus de la sous-lignée Y280 et G1 présentent un motif de clivage polybasique, malgré un phénotype restant faiblement pathogène (Jiahao Zhang et al. 2021).

Afin de mieux comprendre le lien entre acquisition de MBCS, sa compatibilité éventuelle avec d'autre segment HA et la pathogénicité, des études de génétique inverse ont été réalisées. Plusieurs virus non H5/H7 ont pu être produits avec l'ajout artificiel d'un MBCS au sein d'un segment HA de sous –type H1, H2, H3, H4, H6, H8, H11, H14 et H15 dans un contexte avec les sept autres gènes dérivant d'un virus LPAIV H9N2 ou HPAIV H5N1. Alors que tous les réassortants ont présenté une indépendance à la trypsine *in vitro*, seulement certains d'entre-eux possédant un contexte génétique H5N1 HPAIV ont décrit une pathogénicité élevé *in vivo* (Veits et al. 2012).

L'acquisition du MBCS semble être le marqueur principal de pathogénicité chez les virus influenza aviaires. Néanmoins, d'autre facteurs génétiques encore inconnus semblent modifier cette règle, rendant la classification de ces virus et le mécanisme d'émergence plus complexe.

4. Epidémiologie

L'évènement d'émergence d'un HPAIV est un phénomène relativement rare. Quinze et vingt évènements d'émergences indépendants HPAIV-H5 et HPAIV-H7, respectivement, ont été recensés à ce jour (Richard et al. 2017; D.-H. Lee, Criado, and Swayne 2020; Gultyaev et al. 2021) (**Tableau 2** et **Figure 14**), pour un nombre total de 1143 séquences ADNc HA-H5 et 1942 séquences ADNc HA-H7 (données relevées le 7 février 2023 sur NIAID *Influenza Research Database* (IRD)).

Des analyses phylogénétiques de plusieurs génomes H5/H7-HPAIV ont montré que les évènements d'émergences ne sont pas regroupés dans un clade commun. Au contraire, les transitions LPAIV-HPAIV décrivent des linéages différents prenant origine au niveau d'ancêtres distincts (**Figure 14**), convergeant vers un phénotype identique « hautement pathogène » et partageant des mutations communes sur leur génome : l'acquisition d'un site de clivage HA polybasique. Cette répétition d'évènements suggère une évolution parallèle, classiquement retrouvée chez des organismes à taux de réplication élevé, mutant rapidement et possédant un génome généralement contraint spatialement (Escalera-Zamudio et al. 2020; D.-H. Lee, Criado, and Swayne 2020; Gutierrez, Escalera-Zamudio, and Pybus 2019).

La description de plusieurs épidémies HPAIV a permis de relever certaines caractéristiques importantes concernant le contexte d'émergence : type d'infrastructures touchée, hôtes, densité aviaire, ainsi que leur diffusion et durée. L'hôte joue très probablement un rôle majeur dans l'émergence LPAIV-HPAIV. Le poulet apporte un environnement favorable à la sélection de souches HPAIV, corrélant avec une réponse immunitaire innée faible par rapport à ce qui est observé chez le canard (Bessière et al. 2022). La majorité des foyers HPAIV apparait chez des oiseaux domestiques, principalement le poulet et la dinde. Cependant, il est important de noter que d'autres espèces sont aussi susceptibles d'être infectées par ces virus, comme le canard, l'oie, la pintade, la perdrix, l'émeu, le pigeon, le faisan et la caille (Richard et al. 2017). Bien que l'introduction de HPAIV par la faune sauvage constitue un risque possible important, son implication pour certains événements de nouveau HPAIV (au sens génétique) reste difficile à statuer, souvent expliqué par un manque de données ou l'absence de séquence HA avec un motif polybasique provenant de ces hôtes, sur le terrain. Avant la transition vers une forme HPAIV au sein d'une population d'oiseau, la circulation du virus ancestral LPAIV peut varier de manière considérable : de 10 jours à 24 mois. D'autre part, aucune corrélation n'a été mise en évidence entre la taille de l'exploitation ou le nombre d'individu la constituant (densité aviaire) et l'émergence HPAIV. Le tableau ci-dessous recense toutes les émergences HPAIV

décrites jusqu'à ce jour, en précisant le mécanisme génétique responsable de l'acquisition du site de clivage HA polybasique.

Tableau 2 *: Liste des évènements d'émergence HPAIV naturels décrits.*

Localisation	Année de l'évènement d'émergence	Sous-type HA	Numéro d'accession	Nom de la souche	Mécanisme génétique responsable de l'apparition du MBCS
Italie	1997	H5	EF597263	A/chicken/Italy/312/1997/H5N2	insertions
Mexique	1994	H5	AB558473	A/chicken/Puebla/8623/607/1994/H5N2	insertions
Mexique	1995	H5	AB558474	A/chicken/Queretaro/14588/19/1995/H5N2	insertions
Irlande	1983	H5	GU052853	A/duck/Ireland/113/1983/H5N8	insertions
USA	1993	H5	AY444750	A/emu/Texas/39442/1993/H5	insertions
Chine	1996	H5	AF144305	A/goose/Guangdong/1/1996/H5N1	insertions
Afrique du Sud	2011	H5	JX069081	A/ostrich/South/Africa/AI2114/2011/H5N2	insertions
Afrique du Sud	2004	H5	FJ519983	A/ostrich/South/Africa/N227/2004/H5N2	insertions
UK	1991	H5	GU052510	A/turkey/England/50/92/1991/H5N1	insertions
France	2016	H5	KX014902	A/chicken/France/160013g/2016/H5N2	substitutions
Mexique	2005	H5	KM368285	A/chicken/Hidalgo/7637/05/2005/H5N2	substitutions
USA	1983	H5	CY015073	A/chicken/Pennsylvania/1/1983/H5N2	substitutions
Taiwan	2008	H5	AB507264	A/chicken/Taiwan/A703/1/2008/H5N2	substitutions
USA	2004	H5	GU052644	A/chicken/Texas/298313/2/2004/H5N2	substitutions
Canada	1966	H5	M30122	A/turkey/Ontario/7732/1966/H5N9	substitutions
Australie	1976	H7	CY024786	A/chicken/Victoria/1976/H7N7	insertions
Allemagne	1934	H7	CY077420	A/chicken/Rostock/45/1934	insertions
USA	1994	H7	GU052922	A/Pekin/robin/California/30412/1994/H7N1	insertions
USA	2016	H7	KU558906	A/turkey/Indiana/16/001403/1/2016/H7N8	insertions
China	2016	H7	MF280190	A/chicken/Guangdong/Q1/2016/H7N9	insertions
Allemagne	1979	H7	U20459	A/chicken/Leipzig/79/H7N7	insertions
UK	1963	H7	U20462	A/turkey/England/1963/H7N3	insertions
Australie	2007	H7	CY061610	A/duck/Victoria/512/2007/H7N6	substitutions
USA	2002	H7	AY240908	A/GuineaFowl/MA/148081/11/02/H7N2	substitutions
Pays-Bas	2005	H7	CY077008	A/mallard/Netherlands/9/2005/H7N7	substitutions
Japon	2009	H7	AB538457	A/quail/Aichi/2/2009/H7N6	substitutions
Corée	2009	H7	KC609770	A/wild/duck/Korea/CSM27/12/2009/H7N6	substitutions
Chili	2002	H7	AY303631	A/chicken/Chile/4322/02/H7N3	recombinaison

Canada	2004	H7	AY611524	A/chicken/British/Columbia/04/H7N3	recombinaison
Canada	2007	H7	EU500860	A/chicken/SK/HR/00011/2007/H7N3	recombinaison
Mexique	2012	H7	JX908509	A/chicken/Jalisco/12283/2012/H7N3	recombinaison
USA	2017	H7	KY818811	A/chicken/Tennessee/17/007147/2/2017/H7N9	recombinaison
Italie	1999	H7	KF493066	A/chicken/Italy/4845/1999/H7N1	insertions (origine inconnue)
Pays-Bas	2003	H7	EPI_ISL_391798	A/chicken/Netherlands/03008927/2003	insertions (origine inconnue)
Espagne	2009	H7	GU121458	A/chicken/Spain/6279/2/2009/H7N7	recombinaison

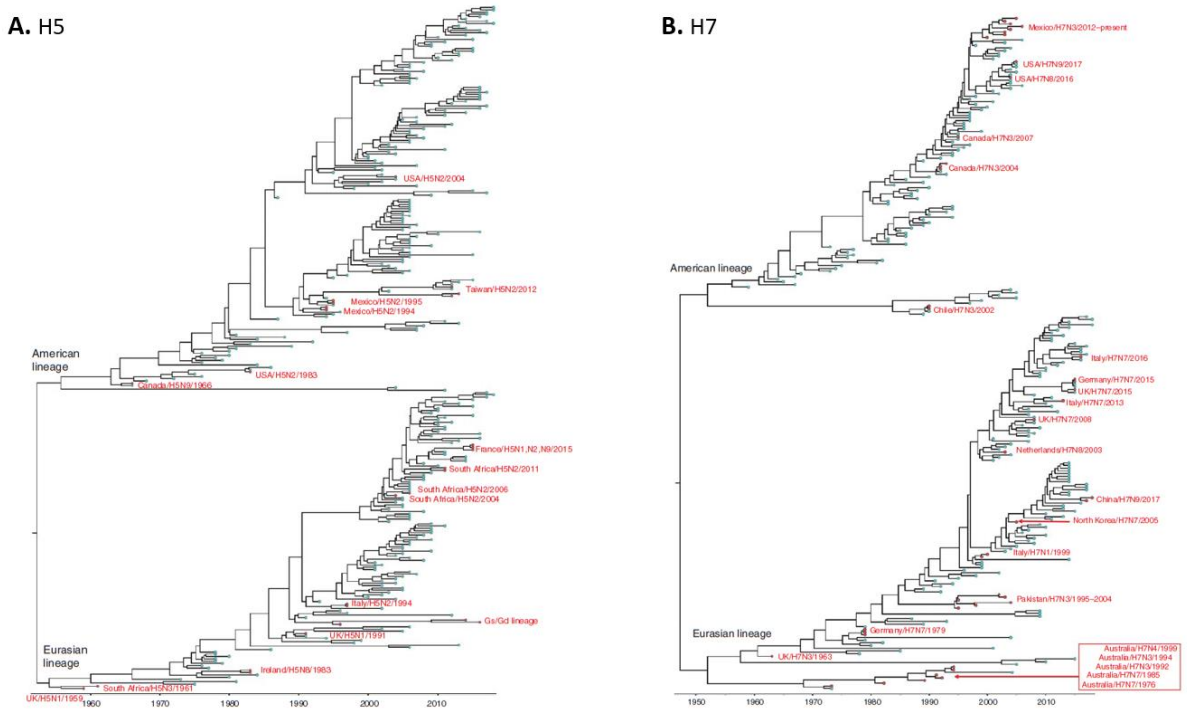


Figure 14 *: Arbres phylogénétiques (horloge moléculaire) des séquences HA5 (A.) et H7 (B.) représentant les principaux évènements d'émergences HPAIV (en rouge), d'après* (D.-H. Lee, Criado, and Swayne 2020)

5. Phénomène d'émergence reproduit en laboratoire

Le phénomène d'émergence d'HPAIV a été étudié et reproduit à plusieurs occasions de manière contrôlée, en laboratoire, à partir d'une souche LPAIV. Ces évolutions expérimentales ont été obtenues par passages successifs sur des lignées cellulaires, sur œufs ou encore *in vivo*.

Une souche H7N7 provenant d'un phoque (A/seal/Mass/1/80) a été adaptée par passages sériés en absence de trypsine sur des cellules immortalisées rénales de chiens (MDCK) et sur des cellules embryonnaires de poulet (CEC). Après plusieurs passages, certains variants isolés ont acquis la capacité de croitre indépendamment à la trypsine. Chez un variant en particulier, issus de CEC infectées, 9 nucléotides supplémentaires ont été insérés au niveau du motif de clivage HA, formant ainsi 4 acides aminés basiques consécutifs (S. Q. Li, Orlich, and Rott 1990).

Une autre étude a pu mettre en évidence la possibilité de créer un virus HPAIV à partir d'un virus sous-type H5 faiblement pathogène. La souche A/swan/Shimane/499/83 (ShimH5) isolée à partir d'un oiseau sauvage a été amplifiée sur cellules et a subi des passages sériés par voie respiratoire (24 passages réalisés via l'utilisation de sacs aériens) et des infections intracérébrales (2 passages) sur des poussins. A l'issu de ces passages, le virus est devenu hautement pathogène (variant ShimH5-24a2b), responsable d'une augmentation considérable de la mortalité (100%) chez des poulets infectés, et possède un tropisme viral étendu vers d'autres organes (cerveau, trachée, poumon, reins, pancréas). Sa protéine HA est maturée en absence de trypsine. La région codant le motif de clivage de HA du variant ShimH5-24a2b a subi des substitutions successives vers des adénines au cours des passages réalisés, amenant à l'insertion d'adénines supplémentaires et la formation d'un site de clivage devenu polybasique RKKR|GLF (Nao N, Yamagishi J et al. 2017). De manière intéressante, le réassortant issu de génétique inverse, composé du segment HA-ShimH5-24a2b et de tous les autres gènes de la souche H1N1 A/Puerto Rico/8/1934 (PR8) décrit une fréquence d'insertions de nucléotides en amont de la séquence codant le site de clivage de HA 1000x supérieure en comparaison avec le réassortant composé du segment HA sauvage (ShimH5) : 0,003% et 3% pour ShimH5 et ShimH5-24a2b respectivement (Nao N, Yamagishi J et al. 2017). Cette différence notable en terme de fréquence d'insertions pourrait être expliquée par l'influence de certains motifs présent sur l'ARN viral (environnement en bases, structures secondaires d'ARN) et/ou des substitutions préalables.

Dans une autre étude, des œufs embryonnés de poulets ont été infectés par un virus H5N2 et H7N1 isolés à partir d'autruches et plusieurs passages ont été réalisés. La virulence de ces derniers a augmenté progressivement au cours des passages successifs. Plusieurs mutations localisées sur d'autres gène, hormis le segment HA, ont pu être détectées par séquençage haut débit. Au bout du 11ième et 13ième passage pour la souche H7N1 et H5N2 respectivement, un motif de clivage HA polybasique est apparu. De plus, une insertion de 42 nucléotides correspondant à un fragment du segment NP a été détectée au 17ième passage sur œuf du virus H5N2, constituant la premier évènement de recombinaison non-homologue obtenu expérimentalement chez un virus de sous-type H5 (Laleye and Abolnik 2020).

De manière intéressante, des expériences de génétique inverse suivies de passages *in ovo* sur des embryons de poulets avec un virus LPAIV monobasique H7N7 (correspondant à l'ancêtre de l'émergence HPAIV anglaise de 2008) et son variant rare dibasique (également isolé durant la même épizootie) ont montré une évolution spontanée du variant dibasique vers la forme hautement pathogène (Seekings et al. 2020). Ces résultats suggèrent un continuum et un lien direct entre le virus LPAIV ancestral, l'apparition d'un variant intermédiaire et l'émergence de virus hautement pathogène.

VIII. Problématique et objectifs du projet de thèse

Les HPAIV constituent une menace majeure pour l'industrie avicole, ainsi qu'un danger potentiel pour la santé humaine, comme le montrent les nombreux cas d'infections fatales humaines, suite à une infection avec des virus aviaires HPAIV de sous-type H5 et H7. Chez l'Homme, la létalité a été estimée à 60 % pour les infections par le virus HPAIV H5N1 (Abdel-Ghafar et al. 2008) et à 40 % pour le virus HPAIV H7N9 (Yang et al. 2019). Ces HPAIV ont le potentiel d'évoluer vers des souches adaptées à l'homme, ce qui pourrait conduire à des pandémies exceptionnellement meurtrières. Mieux comprendre les facteurs qui influencent l'émergence des HPAIV est donc d'une importance capitale pour la santé animale et humaine, en accord total avec le concept "One Health".

Le mécanisme d'évolution des virus influenza aviaires vers les formes hautement pathogènes reste encore mal compris, et pose des réflexions fondamentales sur l'évolution des génomes viraux.

Malgré l'intensité de la recherche mise en œuvre dans ce domaine, des questions subsistent :

i. Est-ce que ce mécanisme d'émergence de HPAIV est strictement dû au hasard ?
ii. Si cette évolution n'est pas due au hasard, est-ce que des éléments présents sur le génome des LPAIV permettent d'expliquer l'évolution de certaines souches vers les formes hautement pathogènes ?

Connaître les facteurs pouvant influencer l'évolution des LPAIV permettra d'améliorer la gestion du risque d'émergence des HPAIV. S'il est possible de déterminer des facteurs génétiques (motifs, structures secondaires) favorables au phénomène d'émergence LPAIV-HPAIV, l'élaboration d'un système de surveillance active basée sur le séquençage d'échantillon de terrain et la classification de virus en fonction de leur risque évolutif pourrait être envisagée.

Comme discuté plus haut (page 36), plusieurs travaux suggèrent le rôle possible de certaines propriétés du génome viral dans l'apparition de mutation et le phénomène d'émergence. Dans le cadre de cette thèse, nous avons testé l'hypothèse de la présence d'éventuels motifs (environnements nucléotidiques, structures secondaires d'ARN) présents sur le segment génomique HA des virus influenza aviaires de sous-type H5 et H7 qui pourraient être favorables à l'augmentation du taux d'erreur de la polymérase virale et donc, à l'évolution vers les formes hautement pathogènes (**Figure 15**).

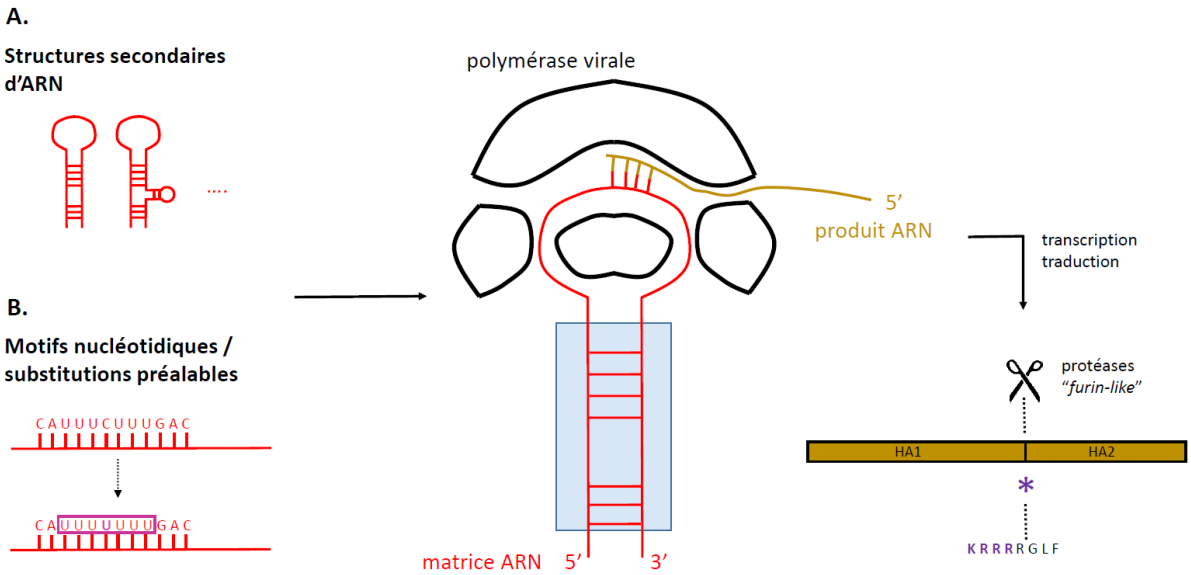


Figure 15 *: Schéma conceptuel sur les facteurs génomiques influençant l'émergence des HPAIV. Certaines propriétés du génome au niveau de la région codant le site de clivage HA comme des structures ARN (A.) et/ou motifs nucléotidiques (B.) peuvent être favorable à l'apparition de mutation et à l'émergence.*

Résultats

I. Influence de l'environnement génomique dans l'évolution des virus

L'apparition de mutation *de novo* constitue la première étape de l'évolution des organismes, dont les virus. Au cours de la réplication, le génome va accumuler des mutations dont l'origine peut être d'ordre physique (comme les dommages directs aux acides nucléiques), cellulaire ou viral. Le taux d'erreur élevé des polymérases virales, généralement observé chez les virus à ARN (Sanjuan et al. 2010; Sanjuán and Domingo-Calap 2016), permet la production d'un nuage de nouveaux virus ou variants, chacun d'entre eux différent d'un point de vue génétique. En fonction de l'impact des mutations sur le *fitness* viral, certains variants seront sélectionnés. En apparente opposition avec la théorie classique du Darwinisme, dans lequel les mutations apparaissent de manière aléatoire sur le génome avant le phénomène de sélection, plusieurs observations sur des virus de familles différentes montrent que certaines caractéristiques de la matrice génomique peuvent influencer le taux d'erreur des polymérases virales, et de ce fait, peuvent orienter en partie l'apparition des mutations plus spécifiquement sur certains loci (**Figure 16**).

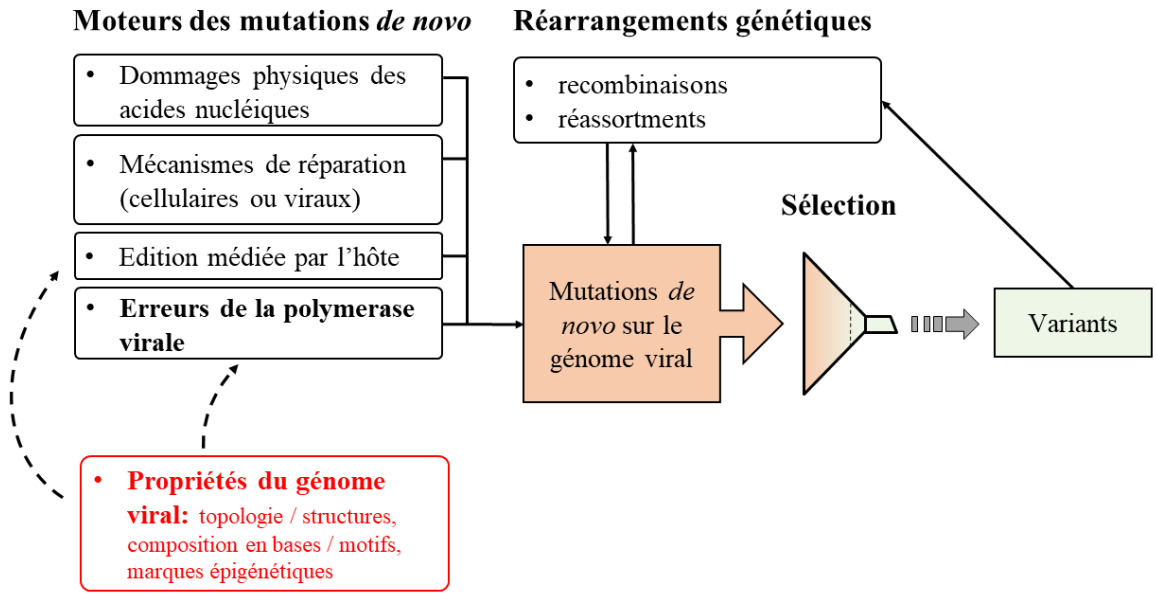


Figure 16 *: Schéma conceptuel de l'évolution des génomes viraux. Les mutations de novo apparaissent sur le génome viral via plusieurs mécanismes incluant les erreurs de la polymérase virale, les mécanismes de réparation, et l'édition médiée par les enzymes cellulaires ou virales, ainsi que les dommages physiques directs aux acides nucléiques (espèces réactives de l'oxygène). Des caractéristiques intrinsèques du génome viral (structures, motifs nucléotidiques, marques épigénétiques) peuvent influencer l'édition médiée par l'hôte et la fidélité de la polymérase virale. In fine, certaines mutations vont être sélectionnées par leur conséquence sur le fitness viral. Les évènements de recombinaison et les réassortiments (dans le cas des virus diploïdes, segmentés ou multipartites) peuvent aussi influencer la diversité génomique via des échanges plus larges de matériel génétique.*

En effet, les événements de recombinaisons, ou plus précisément, de saut de polymérase et changement de matrice, semblent être sous la dépendance de caractéristiques structurelles du génome. Il a été montré que les événements d'insertions et de délétions (indels) observés sur le virus SARS-CoV-2 ont lieu précisément sur des régions du génome possédant des structures secondaires d'ARN importantes (Chrisman et al. 2021). Une observation similaire a été faite chez certains virus de plante (Figlerowicz 2000).
D'autre part, des événements d'insertions par bégaiement ou glissement de la polymérase lors de réplication semblent aussi être dépendants de certains motifs présents sur le génome viral. Un motif correspondant à la répétition de plusieurs adénines consécutives chez le virus de la mosaique du navet (*Turnip mosaic virus,* TuMV, membre de la famille des *Potyviridae*) est responsable d'insertions multiples lors de la transcription (Olspert et al. 2015). Chez le virus Ebola (EBOV, membre de la famille des *Filoviridae*), un motif correspondant à une répétition de sept ou huit uridines sur le génome viral ainsi que les séquences avoisinantes est responsable de l'insertion de 1 ou 2 nucléotides au cours de la transcription (ou *RNA editing*), permettant la production de nouveau ORF et la production de protéines agissant comme leurres contre la réponse immunitaire. Les insertions observées ont aussi lieu au cours de la réplication du génome viral, impactant 10% des nouveaux génomes synthétisés (Shabman et al. 2014). Dans le cas du virus de la grippe, des variants possédant 2 à 13 nucléotides supplémentaires au niveau de la séquence polyU à l'extrémités 5'NCR ont été identifiés (Zheng et al. 1999), suggérant un mécanisme d'insertion similaire à ce qui a été observé chez TuMV et EBOV.

Des études utilisant des modèles permettant d'éviter le biais de mutations dû à la sélection (au niveau protéique) ont confirmé ces précédentes observations. Grâce à l'utilisation d'un système minigénome ou miniréplicon, la région C2 et *Rev responsive element* (RRE) du génome du virus de l'immunodéficience humaine acquise de type 1 (HIV-1) sont plus sujets aux substitutions et aux recombinaisons (Geller et al. 2015), et correspondent à des zones très repliées de l'ARN viral (Simon-Loriere et al. 2010; 2009; Tomezsko et al. 2020). Par le biais d'un système similaire, une autre étude a montré que le taux de substitution varie très fortement le long du génome de l'hépatite C (HCV) jusqu'à l'échelle de bases adjacentes et dépend de l'environnement nucléotidique (Geller et al. 2016).

Les caractéristiques du génome n'influencent pas seulement le taux d'erreur des polymérases virales, mais aussi celui d'autres réplicases d'origine eucaryote. Dans une étude récente, il a été montré que l'ARN polymérase ADN dépendante II (Pol II) humaine accumule des insertions au niveau de certains motifs nucléotidiques (Chung et al. 2023).

Plusieurs études montrent que certaines caractéristiques intrinsèques du génome viral, telles que certaines structures, la composition en base ou des motifs nucléotidiques peuvent directement influencer la fidélité des polymérases virales, et donc partiellement diriger le chemin évolutif entreprit par les virus. Une revue de la littérature ainsi qu'une discussion plus détaillée sur ce sujet ont constitué la rédaction et la publication du papier suivant (Dupré and Volmer 2022) afin de placer en relation ce travail de thèse au sien de l'état de l'art actuel. Le concept suivant dans lequel des éléments génétiques seraient favorables aux événements de mutation responsable à l'apparition du MBCS chez les virus influenza aviaires et donc à l'émergence de HPAIV constitue une hypothèse intéressante sur l'évolution de ces virus.

Trends in Genetics

CellPress OPEN ACCESS

Opinion

Influence of viral genome properties on polymerase fidelity

Gabriel Dupré [1,*] and Romain Volmer [1,*]

The first step of viral evolution takes place during genome replication via the error-prone viral polymerase. Among the mutants that arise through this process, only a few well-adapted variants will be selected by natural selection, renewing the viral genome population. Viral polymerase-mediated errors are thought to occur stochastically. However, accumulating evidence suggests that viral polymerase-mediated mutations are heterogeneously distributed throughout the viral genome. Here, we review work that supports this concept and provides mechanistic insights into how specific features of the viral genome could modulate viral polymerase-mediated errors. A predisposition to accumulate viral polymerase-mediated errors at specific loci in the viral genome may guide evolution to specific pathways, thus opening new directions of research to better understand viral evolutionary dynamics.

Viral polymerase errors during genome copy and the filter of selection

It is straightforward to define an evolving virus by its capacity to create numerous genome variants in its progeny. The cloud of viral genome descendants, initially described in the case of RNA viruses because of the high error rate of viral RNA polymerases, is defined as viral quasispecies [1,2]. This property has profound consequences on viral evolutionary dynamics, regulating viral fitness, immune escape, drug resistance, virulence, host adaptation, and 'social' interactions between variants, detailed in the following reviews [3–5]. Specific errors or mutations are fixed by natural selection and then passed down, after the action of multiple filters or bottlenecks that reduce the size of the mutant spectrum population [6]. Parts of the genome encoding for more plastic protein domains are more tolerant to mutations, favouring mutation accumulation in these loci. Moreover, parts of the viral particle are subjected to strong immune response-mediated bottlenecks, which tend to select immune-escaping viruses harbouring new antigenic sites. Consequently, these antigenic sites are associated with high nucleotide variability. Viral glycoproteins (GPs), which constitute major targets of the immune response, evolve more rapidly than other viral proteins [7–10]. On the contrary, some internal proteins are less tolerant to mutations and are more evolutionary stable, probably because they are hidden from neutralising antibodies and are composed of domains that are structurally and/or functionally more constrained [8,9,11–13].

However, accumulating evidence also suggest that the viral genome template itself can modulate mutation events. Host enzymes have been shown to modify genomes of several RNA and DNA viruses [14–23]. These host editing enzymes target specific genome locations enriched in particular base composition or RNA/DNA structures, revealing the influence of the viral genome properties on host-dependent editing [24–26]. Host editing enzymes will not be further described here. By contrast, this review will focus on recent findings indicating that the viral RNA genome template properties also modulate viral polymerase fidelity, causing nonrandom mutation events, which ultimately could influence viral evolution.

Highlights

Evolution can be defined as a two-step process. The first step is the accumulation of mutations in the viral genome thanks to error-prone viral polymerases. The second step is the selection of descendants that harbour the genetic pool conferring the highest fitness. Thus, specific parts of the viral genome are under intense selection pressure and often appear as the most plastic loci because they accumulate more mutations than other parts of the genome.

However, various studies indicate that the base composition and the structure of the genome sequence can themselves modulate viral polymerase fidelity, resulting in the non-homogenous accumulation of mutation events along the viral genome.

A better understanding of the influence of the template genome properties on the viral polymerase fidelity may improve our understanding of the evolutionary trajectories taken by viruses.

[1]Ecole nationale vétérinaire de Toulouse, ENVT, INRAE, IHAP, UMR 1225, Université de Toulouse, 23 Chemin des Capelles, Toulouse 31076, France

*Correspondence:
gabriel.dupre@envt.fr (G. Dupré) and
romain.volmer@envt.fr (R. Volmer).

© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Impact of the genomic structure and sequence on viral polymerase fidelity

Beta-coronaviruses evolve through different genetic mechanisms, including recombination events or, more precisely, polymerase jumps followed by template switching. Several **recombination breakpoints** (see Glossary) and crossing overs have been described on both structural and non-structural genes, mostly on the spike encoding gene [27,28]. Recent chemical mappings revealed the presence of extensive RNA structures in the genome of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that colocalise with hotspots of template switching, characterised by insertion and deletion (indel) scars in the viral genome, suggesting that the RNA structures present in the viral genome could promote polymerase template switching [11]. In line with this observation, the brome mosaic virus, which belongs to the *Bromoviridae* family, undergoes non-homologous recombination events that seem to be dependent on both the primary sequence of the viral genome and its folding [29]. These findings support early reports suggesting that RNA structures present on influenza A viral ribonucleoproteins could trigger polymerase-jumps and promote the production of defective genome segments [30].

In addition, viral polymerase-mediated nucleotide mutations, consisting of small indels, have also been shown to be influenced by the viral genome properties. The cucumber mosaic virus of the *Bromoviridae* family is known to harbour parasitic RNA called satellite RNA (satRNA), which are small well-structured elements with specific functions depending on nucleic acid folding. Their replication and persistence are strictly dependent on the helper virus [31]. Indel events have been detected predominantly in very structured regions on the satRNA of cucumber mosaic virus [32]. Moreover, viral polymerase-mediated RNA editing, leading to the generation of additional mRNA and new open reading frames (ORFs) by the insertion of one or several nontemplate nucleotides, has been shown to depend on the RNA template properties. In the case of the Turnip mosaic virus, belonging to the *Potyviridae* family, a conserved sequence consisting of a stretch of six adenosines preceded by one guanosine induces insertion events via polymerase stuttering [33]. Similarly, Ebola virus (EBOV) of the *Filoviridae* family produces its GP after the insertion of additional adenosine at a precise location during transcription of the GP-ORF. When this genetic mechanism is absent, only truncated forms of the GP are produced (secreted and small secreted GPs: sGP and ssGP), which are involved in neutralising antibodies. Cis-acting elements corresponding to a stretch of seven/eight consecutive uridines and flanking sequences present on the genome control RNA editing [34–36]. Similar findings were observed for viruses of the *Paramyxoviridae* family involving different editing sequences [34,37,38]. Importantly, RNA editing also occurs during viral genome replication, modifying around 10% of the Zaire EBOV newly synthesised genomes [39] and so could potentially influence viral evolution. In the case of influenza A virus, variants harbouring two to 13 supplementary U/C insertions at the poly-U stretch in the 5′ noncoding region were identified, suggesting a stuttering mechanism comparable with the RNA editing in filoviruses and paramyxoviruses [40].

All the aforementioned studies point to an influence of the viral RNA genome structure or base composition on the viral polymerase fidelity. However, it is warranted to note that these observations were made in the context of selective pressure on the functions of proteins encoded by the viral genome. As a consequence, mutation events associated with a deleterious effect on protein function are counter selected. To strictly assess the effect of genomic properties on mutation events, grasp their diversity, and progress in the understanding of the mechanisms, the use of experimental systems allowing selective pressure to be bypassed is recommended.

Driving a wedge between genome properties and protein function to grasp the full diversity of mutation events

Using a shuttle vector experimental system, freeing the sequence of interest from selection pressure, the most external part of the HIV viral GP, corresponding to the V1–V5 loops, was found to be less

Glossary

Nucleotide insertions: nucleoside insertions or deletions are genetic mutations where one or several nucleotides/nucleosides are inserted/deleted from a viral DNA or an RNA genome.

Nucleotide substitution: genetic mutation where a single nucleotide or nucleoside base is changed from a viral DNA or an RNA genome, respectively. This modification can have no effect (synonymous mutation) or modify the protein sequence (non-synonymous mutation).

Recombination breakpoints: location in the viral genome where the genetic material has been exchanged from the parental sequence to another sequence ('donor') during genome replication.

subject to spontaneous mutations than another part of the genome, suggesting the existence of hypo-mutative sequences or genetic elements reducing mutation rate [41]. Conversely, this same study found that the C2 region and the Rev responsive element (RRE) of the HIV-1 *env* gene were significantly more subjected to substitutions and recombination or template-switching events than other parts of the genome [41]. Interestingly, extended and complex RNA secondary structures have been described in the C2 region and RRE, suggesting that the topology of the viral genome could impact viral polymerase fidelity [42–44].

In the case of the human hepatitis C virus (HCV), deep sequencing of HCV genome fragments obtained via a replicon system, freed from selection pressure on protein function, indicates that the HCV genome exhibits non-homogenous mutation rates along its genome. Substitution biases were found to be dependent on the genome base composition and can largely differ between two adjacent genomic sites [45]. However, whether the RNA structures may play a role in HCV polymerase-mediated mutation events remains unclear [45].

Avian influenza viruses of the H5 and H7 subtypes, initially low pathogenic avian influenza viruses (LPAIV), can become highly pathogenic (HPAIV) through the acquisition of a multi-basic cleavage site (MBCS) motif on the hemagglutinin (HA), ultimately responsible for increased pathogenicity in poultry (Figure 1) [46–48]. **Nucleotide substitutions**, insertions (codon duplication), and/or polymerase jump-template switching have been shown to be responsible for MBCS acquisition [48–53]. Work using a minigenome system, again freeing the sequence of interest from selection pressure, showed that the HA nucleotide sequence had a profound effect on nucleotide insertion rates, a mechanism possibly leading to the acquisition of a MBCS [52]. To our knowledge, this study represents the first demonstration of a possible link between nucleotide composition and/or RNA topology of the viral genome and a probability to acquire a phenotypic change, here acquisition of a HA MBCS, thus linking genome properties to an evolutionary pathway [52]. Interestingly, several studies highlighted the presence of a conserved predicted RNA stem-loop encompassing the HA cleavage site coding sequence, that may play a role in LPAIV-to-HPAIV transition through polymerase slippage/stuttering and polymerase jump-template switching [48,50–57], even though the detailed mechanisms involved in these modifications remain to be identified (Figure 1).

Concluding remarks

Here, we compiled evidence that specific features of the viral genome, including its folding and the presence of specific nucleotide motifs, directly influence the rate and type of viral RNA-dependent RNA polymerase mutations. Interestingly, a recent unreviewed report also showed that the propensity of eukaryotic DNA-dependent RNA polymerase II to cause **nucleotide insertions** depended on the DNA template nucleotide composition [58], thus extending the concept from viruses to eukaryotes. To the best of our knowledge, there is currently no strong evidence that the nucleotide template properties modulate the fidelity of DNA-dependent DNA polymerases. The only studies found to refer to possible mutation hotspots in DNA viruses at precise genomic motifs were obtained under strong selective pressures [59,60]. A straightforward and probable explanation is the proofreading activity of DNA-dependent DNA polymerases that will considerably reduce/correct mutation load [1]. Interestingly, however, a recent study performed on *Arabidopsis thaliana* demonstrated that epigenetic modifications and physical genomic features are linked, resulting in mutation rate biases, which impact genome evolution [61]. Whether this phenomenon can also be translatable to viral genomes is unknown and constitutes an exciting route for future research.

Strong evidence suggests that RNA structures in viral genomes modulate viral polymerase-induced mutation events [41,52,57]. These observations provide an impetus to further investigate

Outstanding questions

What are the genetic elements that can influence mutation rates?

What are the mechanisms causing increased viral polymerase-mediated mutation events?

What is the extent of this evolutionary force in both RNA and DNA viruses?

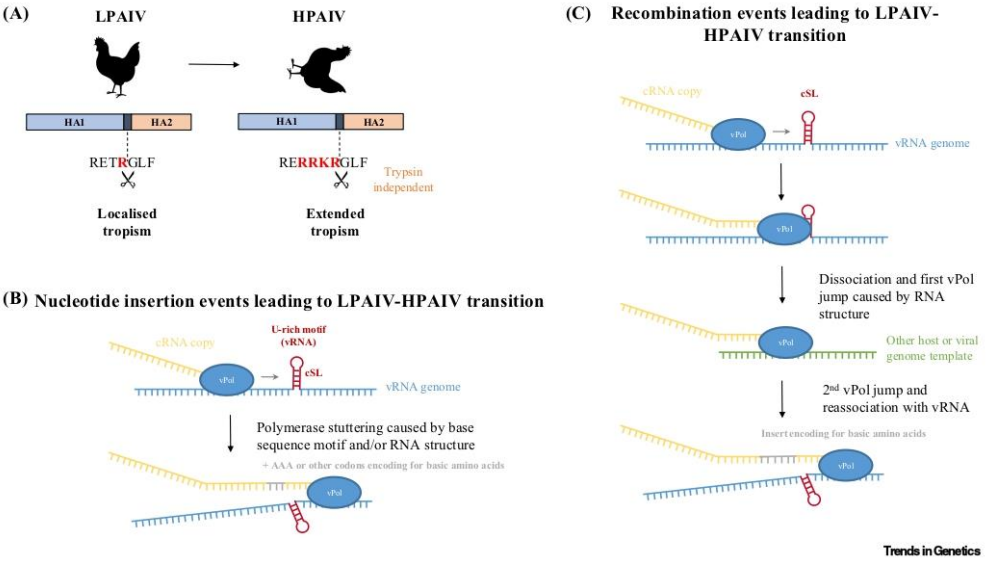


Figure 1. Possible contribution of the hemagglutinin (HA) genome properties to avian influenza A virus evolution from low pathogenic avian influenza viruses (LPAIV) to highly pathogenic avian influenza viruses (HPAIV). (A) Schematic representation of the influenza A HA protein of LPAIV and HPAIV. The HA of LPAIV is proteolytically matured by trypsin-like proteases located in the respiratory and digestive tract of birds, which determine its tropism for these tissues. HPAIV originate from LPAIV that evolve through the acquisition of a multibasic HA cleavage site motif via nucleotide substitutions, insertions, and/or non-homologous recombination. This enables the HA protein to be matured by ubiquitous intracellular furin-like proteases, leading to an extended viral tropism and high pathogenicity in poultry. (B) Model of the nucleotide insertion events leading to LPAIV–HPAIV transition. In the HA1/HA2 encoding region, a specific U-rich stretch and/or the conserved stem-loop structure (cSL) could promote polymerase stuttering during genome replication, leading to the insertions of one or multiple bases in the neo-synthetised strand. (C) Model of recombination or polymerase jump/template switching leading to LPAIV-to-HPAIV transition. In this model, the viral polymerase copies the viral template until it reaches a cSL at the HA1/HA2 boundary region. The cSL, with the action of possible cofactors, induces the dissociation of the polymerase from the template. The polymerase then lands on another template, which can be a host RNA or another viral RNA, and copies a portion of this new template. Eventually, the polymerase goes back to the initial viral template and elongates the remaining part of the viral genome template.

this association. However, it is warranted to note that despite the power of predicting algorithms available [62], *in vitro*, *in cellulo*, and *in virion* chemical structural probing is required to verify the authenticity of predicted RNA structures [44,62–68]. In addition, because there is currently no method capable of assessing RNA structures in the context of the genome replication process, and more precisely during template copy in the catalytic site of viral polymerases, the interpretation of a suggested link between RNA structures and mutability has to remain prudent.

Studies using replicon-based systems, which free the genome from selection pressure on the protein functions, provided evidence in favour of a possible link between mutation events and RNA structure constraints and/or nucleotide motifs [41,45,52]. Replicon-based experiments are powerful in multiple ways, including their relative simplicity of setup, their flexibility to assess various combinations of factors (different viral polymerases, nucleotide sequences), and the absence of infectious-related risks. However, replicon systems do not allow accurate recapitulation of the exponential production of new genome variants by multiple infection cycles, due to their 'non-infectious' nature. Thus, the design of viable infectious viruses, in which part of their genome can be freed from selective pressures, could help to more accurately investigate how the genetic environment could impact mutation rates. A better understanding of the mechanisms

modulating viral polymerase-mediated mutations and their potential consequences on viral evolution represents an interesting challenge in viral evolutionary biology (see Outstanding questions).

Declaration of interests

The authors declare no competing interests.

References

1. Sanjuán, R. and Domingo-Calap, P. (2016) Mechanisms of viral mutation. *Cell. Mol. Life Sci.* 73, 4433–4448
2. Sanjuán, R. *et al.* (2010) Viral mutation rates. *J. Virol.* 84, 9733–9748
3. Andino, R. and Domingo, E. (2015) Viral quasispecies. *Virology* 479–480, 46–51
4. Domingo, E. *et al.* (2021) Historical perspective on the discovery of the quasispecies concept. *Annu. Rev. Virol.* 8, 51–72
5. Sanjuán, R. (2021) The social life of viruses. *Annu. Rev. Virol.* 8, 183–199
6. McCrone, J.T. and Lauring, A.S. (2018) Genetic bottlenecks in intraspecies virus transmission. *Curr. Opin. Virol.* 28, 20
7. Lee, J.M. *et al.* (2018) Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. *Proc. Natl. Acad. Sci. U. S. A.* 115, E8276–E8285
8. Thyagarajan, B. and Bloom, J.D. (2014) The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *eLife* 2014, e03300
9. Lyons, D.M. and Lauring, A.S. (2018) Mutation and epistasis in influenza virus evolution. *Viruses* 10, 407
10. Heaton, N.S. *et al.* (2013) Genome-wide mutagenesis of influenza virus reveals unique plasticity of the hemagglutinin and NS1 proteins. *Proc. Natl. Acad. Sci. U. S. A.* 110, 20248–20253
11. Chrisman, B.S. *et al.* (2021) Indels in SARS-CoV-2 occur at template-switching hotspots. *BioData Min.* 14, 1–16
12. Munis, A.M. *et al.* (2021) Genomic diversity of SARS-CoV-2 in Oxford during United Kingdom's first national lockdown. *Sci. Rep.* 11, 21484
13. Harvey, W.T. *et al.* (2021) SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* 19, 409–424
14. Harris, R. and Dudley, J. (2015) APOBECs and virus restriction. *Virology* 479–480, 131–145
15. Hirose, Y. *et al.* (2018) Within-host variations of human papillomavirus reveal APOBEC signature mutagenesis in the viral genome. *J. Virol.* 92, e00017-18
16. Suspène, R. *et al.* (2005) Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo. *Proc. Natl. Acad. Sci. U. S. A.* 102, 8321–8326
17. Vartanian, J. *et al.* (2008) Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. *Science* 320, 230–233
18. Sadler, H.A. *et al.* (2010) APOBEC3G contributes to HIV-1 variation through sublethal mutagenesis. *J. Virol.* 84, 7396–7404
19. Suspène, R. *et al.* (2011) Double-stranded RNA adenosine deaminase ADAR-1-induced hypermutated genomes among inactivated seasonal influenza and live attenuated measles virus vaccines. *J. Virol.* 85, 2458
20. Martínez, I. and Melero, J. (2002) A model for the generation of multiple A to G transitions in the human respiratory syncytial virus genome: predicted RNA secondary structures as substrates for adenosine deaminases that act on RNA. *J. Gen. Virol.* 83, 1445–1455
21. Cuevas, J. *et al.* (2016) Human norovirus hyper-mutation revealed by ultra-deep sequencing. *Infect. Genet. Evol.* 41, 233–239
22. Atkinson, N.J. *et al.* (2014) The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. *Nucleic Acids Res.* 42, 4527–4545
23. Cattaneo, R. *et al.* (1988) Biased hypermutation and other genetic changes in defective measles viruses in human brain infections. *Cell* 55, 255–265
24. Holtz, C. *et al.* (2013) APOBEC3G cytosine deamination hotspots are defined by both sequence context and single-stranded DNA secondary structure. *Nucleic Acids Res.* 41, 6139–6148
25. Sharma, S. and Baysal, B.E. (2017) Stem-loop structure preference for site-specific RNA editing by APOBEC3A and APOBEC3G. *PeerJ* 5, e4136
26. Kuttan, A. and Bass, B.L. (2012) Mechanistic insights into editing-site specificity of ADARs. *Proc. Natl. Acad. Sci. U. S. A.* 109, E3295–E3304
27. Rowe, C. *et al.* (1997) Generation of coronavirus spike deletion variants by high-frequency recombination at regions of predicted RNA secondary structure. *J. Virol.* 71, 6183–6190
28. Forni, D. *et al.* (2020) Recombination and positive selection differentially shaped the diversity of betacoronavirus subgenera. *Viruses* 12, 1313
29. Figlerowicz, M. (2000) Role of RNA structure in non-homologous recombination between genomic molecules of brome mosaic virus. *Nucleic Acids Res.* 28, 1714
30. Jennings, P.A. *et al.* (1983) Does the higher order structure of the influenza virus ribonucleoprotein guide sequence rearrangements in influenza viral RNA? *Cell* 34, 619–627
31. Palukaitis, P. (2016) Satellite RNAs and satellite viruses. *Mol. Plant-Microbe Interact.* 29, 181–186
32. Pita, J.S. *et al.* (2007) Environment determines fidelity for an RNA virus replicase. *J. Virol.* 81, 9072–9077
33. Olspert, A. *et al.* (2015) Transcriptional slippage in the positive-sense RNA virus family *Potyviridae*. *EMBO Rep.* 16, 995–1004
34. Le Mercier, P. and Kolakofsky, D. (2019) Bipartite promoters and RNA editing of paramyxoviruses and filoviruses. *RNA* 25, 279–285
35. Volchkova, V.A. *et al.* (2011) Genomic RNA editing and its impact on Ebola virus adaptation during serial passages in cell culture and infection of guinea pigs. *J. Infect. Dis.* 204, S941–S946
36. Mehedi, M. *et al.* (2013) Ebola virus RNA editing depends on the primary editing site sequence and an upstream secondary structure. *PLoS Pathog.* 9, e1003677
37. Welch, B.D. *et al.* (2012) Structure of the cleavage-activated prefusion form of the parainfluenza virus 5 fusion protein. *Proc. Natl. Acad. Sci. U. S. A.* 109, 16672–16677
38. Kolakofsky, D. (2016) Paramyxovirus RNA synthesis, mRNA editing, and genome hexamer phase: a review. *Virology* 498, 94–98
39. Shabman, R.S. *et al.* (2014) Deep sequencing identifies noncanonical editing of Ebola and Marburg virus RNAs in infected cells. *MBio* 5, e02011
40. Zheng, H. *et al.* (1999) Influenza A virus RNA polymerase has the ability to stutter at the polyadenylation site of a viral RNA template during RNA replication. *J. Virol.* 73, 5240
41. Geller, R. *et al.* (2015) The external domains of the HIV-1 envelope are a mutational cold spot. *Nat. Commun.* 6, 1–9
42. Simon-Loriere, E. *et al.* (2009) Molecular mechanisms of recombination restriction in the envelope gene of the human immunodeficiency virus. *PLoS Pathog.* 5, e1000418
43. Simon-Loriere, E. *et al.* (2010) RNA structures facilitate recombination-mediated gene swapping in HIV-1. *J. Virol.* 84, 12675–12682
44. Tomezsko, P.J. *et al.* (2020) Determination of RNA structural diversity and its role in HIV-1 RNA splicing. *Nature* 582, 438–442
45. Geller, R. *et al.* (2016) Highly heterogeneous mutation rates in the hepatitis C virus genome. *Nat. Microbiol.* 1, 1–7
46. Swayne, D.E. and Pantin-Jackwood, M. (2006) Pathogenicity of avian influenza viruses in poultry. *Dev. Biol. (Basel)* 124, 61–67
47. Lee, D.-H. *et al.* (2020) Pathobiological origins and evolutionary history of highly pathogenic avian influenza viruses. *Cold Spring Harb. Perspect. Med.* 11, a038679
48. Abdelwhab, E.-S.M. *et al.* (2013) Genetic changes that accompanied shifts of low pathogenic avian influenza viruses toward higher pathogenicity in poultry. *Virulence* 4, 441–452

49. Perdue, M.L. and Suarez, D.L. (2000) Structural features of the avian influenza virus hemagglutinin that influence virulence. *Vet. Microbiol.* 74, 77–86
50. Perdue, M.L. *et al.* (1997) Virulence-associated sequence duplication at the hemagglutinin cleavage site of avian influenza viruses. *Virus Res.* 49, 173–186
51. García, M. *et al.* (1996) Heterogeneity in the haemagglutinin gene and emergence of the highly pathogenic phenotype among recent H5N2 avian influenza viruses from Mexico. *J. Gen. Virol.* 77, 1493–1504
52. Nao, N. *et al.* (2017) Genetic predisposition to acquire a polybasic cleavage site for highly pathogenic avian influenza virus hemagglutinin. *MBio* 8, 1–15
53. Gulyaev, A.P. *et al.* (2021) Insertions of codons encoding basic amino acids in H7 hemagglutinins of influenza A viruses occur by recombination with RNA at hotspots near snoRNA binding sites. *RNA* 27, 123–132
54. Dietze, K. *et al.* (2018) From low to high pathogenicity—characterization of H7N7 avian influenza viruses in two epidemiologically linked outbreaks. *Transbound. Emerg. Dis.* 65, 1576–1587
55. Beerens, N. *et al.* (2020) Emergence and selection of a highly pathogenic avian influenza H7N3 virus. *J. Virol.* 94, e01818-19
56. Gulyaev, A.P. *et al.* (2019) Conserved structural RNA domains in regions coding for cleavage site motifs in hemagglutinin genes of influenza viruses. *Virus Evol.* 5, 1–11
57. Dupré, G. *et al.* (2021) Phylodynamic study of the conserved RNA structure encompassing the hemagglutinin cleavage site encoding region of H5 and H7 low pathogenic avian influenza viruses. *Virus Evol.* 7, 1–16
58. Chung, C. *et al.* (2022) The fidelity of transcription in human cells. *bioRxiv* Published online May 10, 2022. https://doi.org/10.1101/2022.05.10.491385
59. Wilson, J.B. *et al.* (1986) A frameshift at a mutational hotspot in the polyoma virus early region generates two new proteins that define T-antigen functional domains. *Cell* 44, 477–487
60. Sasadeusz, J.J. *et al.* (1997) Homopolymer mutational hot spots mediate herpes simplex virus resistance to acyclovir. *J. Virol.* 71, 3872–3878
61. Monroe, J.G. *et al.* (2022) Mutation bias reflects natural selection in *Arabidopsis thaliana*. *Nature* 602, 101–105
62. Zhao, Q. *et al.* (2021) Review of machine learning methods for RNA secondary structure prediction. *PLoS Comput. Biol.* 17, e1009291
63. Smyth, R.P. *et al.* (2018) RNA structure-a neglected puppet master for the evolution of virus and host immunity. *Front. Immunol.* 9, 2097
64. Gilmer, O. *et al.* (2021) Chemical and enzymatic probing of viral RNAs: from infancy to maturity and beyond. *Viruses* 13, 1894
65. Lim, C.S. and Brown, C.M. (2018) Know your enemy: successful bioinformatic approaches to predict functional RNA structures in viral RNAs. *Front. Microbiol.* 8, 2582
66. Merino, E.J. *et al.* (2005) RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* 127, 4223–4231
67. Rausch, J.W. *et al.* (2018) Probing the structures of viral RNA regulatory elements with SHAPE and related methodologies. *Front. Microbiol.* 8, 2634
68. Dadonaite, B. *et al.* (2019) The structure of the influenza A virus genome. *Nat. Microbiol.* 4, 1781–1789

II. Analyse phylodynamique de la tige-boucle d'ARN conservée du segment HA et son implication dans l'émergence HPAIV

Plusieurs études ont émis l'hypothèse d'un lien entre la présence d'une structure secondaire d'ARN conservée de type tige-boucle, présente sur le segment HA au niveau de la région codant le site de clivage et le phénomène d'émergence (cf p 36). Cependant, ces travaux ont soit : i. utilisés un nombre relativement réduit de virus/séquences génétiques associées à des prédictions de structures, ou, ii. réalisés leur analyse sans prendre en considération la diversité génétique des séquences HA-LPAIV, iii. ni le lien phylogénétique entre les virus émergents (HPAIV) et leur ancêtre LPAIV. Nous avons émis l'hypothèse que certaines caractéristiques structurelles de la tige-boucle conservée : sa stabilité thermodynamique ou des topologies particulières pouvaient être identifiées chez les séquences HA (de sous-type H5 et H7) des pro géniteurs HPAIV connus, et pourraient être utilisées comme marqueurs de risque de l'évolution LPAIV-HPAIV.

Les banques de données publiques (NIAID Influenza Research Database (IRD) et GISAID) permettent le stockage et le partage des séquences génétiques virales issues d'échantillons cliniques et environnementaux, associées à des métadonnées (données d'échantillonnage, pathogénicité, localisation géographique, date de la collection, hôte d'origine), formant ainsi une source d'information précieuse pour l'étude de la diversité génétique virale. Afin d'éviter les biais potentiels ou un manque de sensibilité dans l'analyse (énoncés en amont), nous avons réalisé une étude phylodynamique sur l'ensemble des séquences HA disponibles dans les banques de données, dérivant de LPAIV et des HPAIV connus. Après avoir construit les arbres phylogénétiques des séquences HA (de sous-type H5 et H7), les séquences HA ancestrales ont été estimées et nous avons comparé l'ensemble des séquences virales ancestrales ayant évolué vers l'acquisition d'un MBCS par rapport à celles n'ayant pas évolué vers un phénomène d'émergence. Ces séries de séquences HA ancestrales dirigées vers un évènement phylogénétique final connu (séquence HA-LPAIV ou HPAIV connue, feuille de l'arbre) constituent les différents chemins évolutifs. Nous avons réalisé des prédictions de structure secondaire d'ARN sur l'ensemble des séquences HA au niveau de la région codant le site de clivage afin d'identifier une éventuelle corrélation entre des caractéristiques de la tige-boucle d'ARN conservée au niveau du segment HA (stabilité thermodynamique ou la topologie) et le

phénomène d'émergence. Ce travail constitue une première étude visant à étudier le lien entre la structure d'ARN prédite et le phénomène d'émergence HPAIV, en prenant en compte l'histoire évolutive des séquences virales.

VIRUS EVOLUTION

Virus Evolution, 2021, **7(2)**, 1–16
DOI: https://doi.org/10.1093/ve/veab093
Advance access publication date: 1 November 2021
Research Article

Phylodynamic study of the conserved RNA structure encompassing the hemagglutinin cleavage site encoding region of H5 and H7 low pathogenic avian influenza viruses

Gabriel Dupré,[1,*,†] Claire Hoede,[2,‡] Thomas Figueroa,[1,§] Pierre Bessière,[1] Stéphane Bertagnoli,[1] Mariette Ducatez,[1] Christine Gaspin,[2] and Romain Volmer[1,*]

[1] Ecole nationale vétérinaire de Toulouse, ENVT, INRAE, IHAP, UMR 1225, Université de Toulouse, 23 chemin des Capelles, Toulouse 31076, France and [2] INRAE, UR875 Mathématiques et Informatique Appliquées Toulouse, Plateforme GenoToul BioInfo, Chemin de Borde-Rouge BP 52627, Castanet-Tolosan 31326, France
[†] https://orcid.org/0000-0003-2862-8863
[‡] https://orcid.org/0000-0001-5054-7731
[§] https://orcid.org/0000-0002-4141-271X
*Corresponding authors: E-mail: gabriel.dupre@envt.fr; romain.volmer@envt.fr

Abstract

Highly pathogenic avian influenza viruses (HPAIVs) evolve from low pathogenic avian influenza viruses (LPAIVs) of the H5 and H7 subtypes. This evolution is characterized by the acquisition of a multi-basic cleavage site (MBCS) motif in the hemagglutinin (HA) that leads to an extended viral tropism and severe disease in poultry. One key unanswered question is whether the risk of transition to HPAIVs is similar for all LPAIVs H5 or H7 strains, or whether specific determinants in the HA sequence of some H5 or H7 LPAIV strains correlate with a higher risk of transition to HPAIVs. Here, we determined if specific features of the conserved RNA stem-loop located at the HA cleavage site-encoding region could be detected along the LPAIV to HPAIV evolutionary pathway. Analysis of the thermodynamic stability of the predicted RNA structures showed no specific patterns common to HA sequences leading to HPAIVs and distinct from those remaining LPAIVs. However, RNA structure clustering analysis revealed that most of the American lineage ancestors leading to H7 emergences via recombination shared the same viral RNA (vRNA) structure topology at the HA1/HA2 boundary region. Our study thus identified predicted secondary RNA structures present in the HA of H7 viruses, which could promote genetic recombination and acquisition of a multibasic cleavage site motif (MBCS).

Key words: highly pathogenic avian influenza virus; evolution; RNA structure; influenza

1. Introduction

Influenza viruses are widely distributed among animals, including humans. The aquatic birds, predominately of the orders *Anseriformes* and *Charadriiformes*, constitute their major natural reservoirs (Webster et al. 1992). A wide variety of hemagglutinin (HA)/neuraminidase (NA) combinations of HA1-16 and NA1-9 avian influenza viruses (AIVs) are circulating among aquatic birds, especially ducks (Webster et al. 1992; Vandegrift et al. 2010; Wahlgren 2011). One major determinant of viral pathogenicity in poultry is the cleavability of HA, which splits AIVs into two main pathotypes: low and highly pathogenic avian influenza viruses (LPAIVs and HPAIVs, respectively) (Horimoto et al. 1994, 1995; Garten and Klenk 1999; Steinhauer 1999). Most of the currently circulating AIVs among birds are LPAIVs and are distributed worldwide (Vandegrift et al. 2010; Wahlgren 2011). The HA cleavage site (HA-CS) motif of LPAIVs has non-consecutive dibasic amino acids and is cleaved by trypsin-like proteases located in the lung and intestinal tract of birds. This restricted HA maturation is responsible for mild or quasi-unapparent clinical signs depending on the viral strain and the host species. Some LPAIVs can acquire a multi-basic cleavage site (MBCS) motif that can be processed by ubiquitous furin-like proteases present in all cells of the organism. As a consequence, the virus can infect a broader range of tissues and cause a systemic infection. This shift in viral tropism is responsible for severe disease associated with HPAIV infection in poultry (Swayne and Pantin-Jackwood 2006; Abdelwhab, Veits, and Mettenleiter 2013; Lee, Criado, and Swayne 2020). HPAIVs do not only represent a major threat to the poultry industry worldwide: many human cases of H5 and H7 HPAIV infections have been described and were unfortunately fatal (Wan et al. 2011; Bui et al. 2016; Simon et al. 2016; Wu et al. 2017; Long et al. 2019). HPAIVs evolve from an LPAIV progenitor virus through genetic mutations that trigger the MBCS appearance. Nucleotide substitutions, insertions, and non-homologous recombination between the host (ribosomal RNA 28S) and viral genes (matrix and nucleoprotein segments) have been observed to be responsible for MBCS

© The Author(s) 2021. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/ve/article/7/2/veab093/6415083 by Institut National Polytechnique Toulouse user on 06 February 2023

acquisition (García et al. 1996; Perdue et al. 1997; Perdue and Suarez 2000; Abdelwhab, Veits, and Mettenleiter 2013; Nao et al. 2017; Gultyaev et al. 2021). In nature, LPAIV-to-HPAIV evolution has only been described for viruses belonging to the H5 and H7 subtype (with the exception of one H4 subtype virus that acquired a MBCS but did not exhibit a HPAIV phenotype) (Wong et al. 2014). This subtype restriction remains a mystery, as genetic manipulations of the HA-CS region using reverse genetics conferred higher pathogenicity in other subtypes (H2, H4, H6, H8, H9, and H14), demonstrating that these HA subtypes do not harbour structural incompatibilities with artificial MBCS insertion (Munster et al. 2010; Soda et al. 2011; Veits et al. 2012). Moreover, no other nucleotide or protein motifs elsewhere on the viral genome have yet been associated with the pioneer step of HPAIV emergence (Escalera-Zamudio et al. 2020). A major, yet unachieved, goal is to determine if genetic factors might be present on the HA segment from H5/H7 AIV and could explain their potency to acquire MBCS and finally become more pathogenic. Additionally, it is still undetermined if all the H5/H7 LPAIVs have the same probability to evolve towards HPAIVs or whether specific genetic determinants in the HA sequence correlate with a higher risk of HPAIV transition for some H5/H7 LPAIVs.

Elements on the viral genome of several viruses have been suggested to be responsible for further genomic edition during replication. Viruses of the Paramyxoviridae and Filoviridae families are negative-stranded RNA viruses and share many similarities with AIV in terms of their ribonucleoprotein and viral polymerase structures. It has been shown that specific sequences named as editing sites control mRNA edition at the local level (Le Mercier and Kolakofsky 2019). Upstream secondary structures have been suggested to influence nucleotide insertion events (Mehedi et al. 2013). Interestingly, more than 10 per cent of the Ebola virus antigenome is also edited during viral replication (Shabman et al. 2014; Le Mercier and Kolakofsky 2019). Similar findings were also observed in positive-stranded RNA viruses. Secondary structures present on the viral genome are known to influence human immunodeficiency viruses-1 (HIV-1) and hepatitis C virus (HCV) mutation rates in a site-specific manner (Pathak and Temin 1992; Pita et al. 2007; Simon-Loriere et al. 2009, 2010; Geller et al. 2015, 2016; Krammer et al. 2018). Overall, these observations suggest that RNA structures and specific nucleotide motifs may impact the genetic evolution of a broad spectrum of RNA viruses.

Focusing on LPAIV to HPAIV evolution, the propensity of H5 or H7 LPAIV to evolve into an HPAIV has been linked to the presence of a predicted secondary RNA stem-loop that encompasses the HA-CS coding sequence (Garcia et al. 1996; Perdue et al. 1997; Gultyaev et al. 2016; Nao et al. 2017; Dietze et al. 2018; Beerens et al. 2020). The apical loop of this secondary RNA structure composed of six to twelve unpaired bases has been proposed to promote nucleotide insertions through polymerase slippage (Nao et al. 2017). However, a large-scale bioinformatics analysis has revealed that this predicted RNA stem-loop is widely conserved among AIVs, even in the HA of influenza virus subtypes that have never been shown to evolve to HPAIVs (Gultyaev et al. 2019). This study analysed representative sequences from different HA subtypes without considering that within H5 and H7 subtypes only a few HA sequences have evolved to HPAIVs. As a consequence, it may have failed to identify specific features of this conserved RNA stem-loop (cSL) of HA sequences that only appear at specific stages along the LPAIV to HPAIV evolutionary pathway.

To identify whether the cSL structure of HPAIV ancestors followed a specific evolutionary pathway, we combined predicted RNA structure analyses with phylogenetic analyses on all available HA5/HA7 sequences (see the summarized analysis pipeline in Fig. 1). We chose an analytic window of eighty nucleotides encompassing the sequence encoding HA cleavage to be consistent with possible intra-segment interactions between the strands incoming and exiting the polymerase complex (Chang et al. 2015; Gultyaev et al. 2019; Wandzik et al. 2020). HA-tree inferences, molecular clock, and ancestral sequence reconstructions were performed to investigate the evolution of RNA structures as a function of the genetic distance from every HPAIV emergence event. A comparative analysis was performed with selected LPAIVs unrelated to any known HPAIV emergence event. We also performed RNA structure clustering analysis to identify potential RNA structure topology homologies along the LPAIV-to-HPAIV evolutionary pathway.

2. Methods

2.1 Phylogeny of HA segment and constitution of sequence groups

2.1.1 *Selection of HA sequences and multiple alignment*

All available HA coding sequences (complementary DNA (cDNA)) have been uploaded from the NIAID Influenza Research Database (IRD) through http://www.fludb.org on 15 April 2020. Duplicates and laboratory-originating sequences were removed from the data set. In the case of H1–H16 subtypes, only sequences from avian hosts were selected. Moreover, HPAIV sequences descending from the unique H5 emergence in Guangdong in 1996 were also removed from the analysis to keep only independent H5 HPAIV emergence events. Only the HPAIV sequences with the accession numbers AF144305 and AF148678 were retained to represent this emergence event.

Multiple alignments of HA cDNA sequences per subtype were performed by using MUSCLE program (3.8.31 version). More precisely, HA nucleotide sequences were downloaded, translated, and then subjected to alignments based on the amino acid sequence. Finally, codons of each sequence from the original nucleotide sequences were retrieved and replaced the corresponding amino-acid residues of the protein alignment. Referring to H5 and H7 HPAIV emergence mechanisms, specific HA sequences underwent nucleotide insertions and recombination with other genes, making the HA1/HA2 region alignment difficult. The consideration of indels remains challenging in phylogeny and is absent in commonly used DNA models, which potentially constitute a source of bias for tree inference. Thus, cDNA alignments of HA sequences were manually edited to remove nucleotide insertions corresponding to codon duplication and recombination (H7 HPAIV) events at the HA-CS encoding region. The minimal consensus motifs of H5 and H7 LPAIV-HA-CS motif, 'QRETR|GLF' and 'XPKGR|GLF', respectively, have been preserved during the alignment editing. HA sequences were annotated as HPAIV or LPAIV-HA depending on the HA-CS motif on non-edited alignment. HPAIV/LPAIV-HA assignment was verified from the literature (Abdelwhab, Veits, and Mettenleiter 2013; Richard et al. 2017; Lee, Criado, and Swayne 2020).

2.1.2 *Model selection, HA5/HA7 DNA tree inferences*

The phylogenetic analysis was conducted with the IQ-Tree-2.0.6 software using a maximum-likelihood (ML) approach (Nguyen et al. 2015; Minh et al. 2020).

The best DNA models were estimated and selected via Akaike Information Criterion (Kalyaanamoorthy et al. 2017). HA5 and HA7 phylogenetic trees were constructed using a general time-reversible nucleotide substitution model with a relaxed-Gamma

Downloaded from https://academic.oup.com/ve/article/7/2/veab093/6415083 by Institut National Polytechnique Toulouse user on 06 February 2023

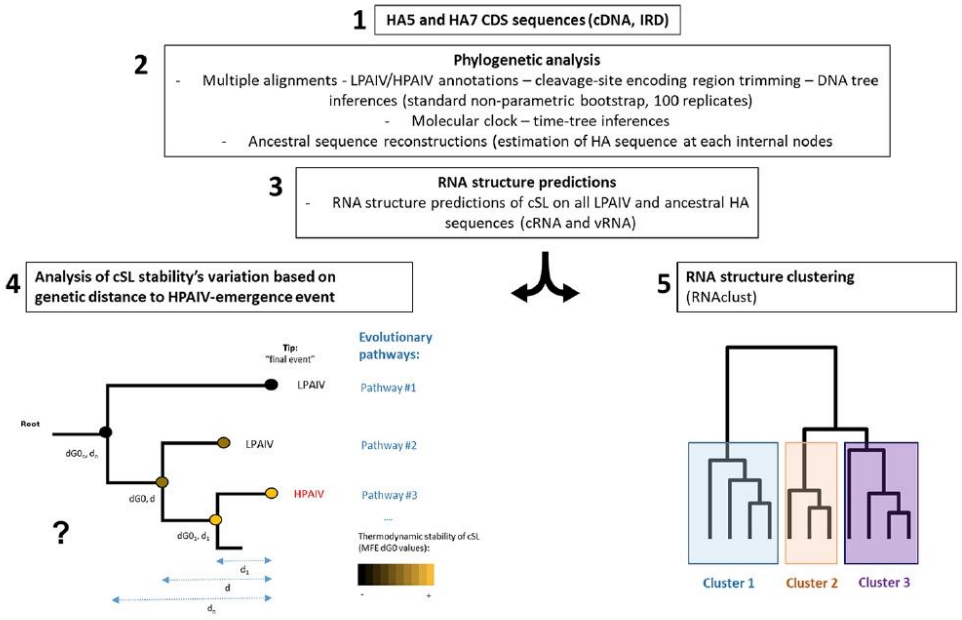


Figure 1. Analysis pipeline to the study of the possible relation between cSL stability evolution and the HPAIV emergence event. The analytic procedure is constituted of five steps. First, HA5 and HA7 encoding cDNA sequences were downloaded from the NIAID IRD (Step 1). Sequences were subjected to multiple alignments, classified as LPAIV or HPAIV-HA sequence, thanks to data from literature/cleavage site motif. DNA trees were inferred by using the best evolutionary model. Generated alignments and trees were subsequently subjected to molecular clock analysis to create final DNA time-trees. Afterward, HA sequences at internal nodes were estimated (Step 2). Finally, multiple secondary RNA structure predictions were performed at the encoding region of the cSL on every H5/H7 LPAIVs and estimated ancestral sequences (Step 3). Two complementary approaches were performed to study the RNA structure of the HA1/HA2 encoding region. The evolution of the cSL structure thermodynamic stability based on the genetic distance (d) to any evolutionary pathways (ending to final LPAIV or HPAIV event) was analysed (Step 4). A schematic phylogenetic tree is shown, representing our hypothesis about a progressive cSL stabilization preceding the HPAIV emergence event. Finally, clustering of different RNA structures families was performed on every H5/H7 LPAIV-HA sequences and their estimated ancestral sequences by using RNAclust software (Step 5). dGO: free energy (kcal/mol); MFE: minimal free energy structure; D: genetic distance; cSL: previously described conserved stem-loop at the HA1/HA2 boundary region.

distribution: GTR + F + R6 and GTR + F + R5 for HA5 and HA7 trees, respectively (GTR—Free Rate Model). Topological support was estimated by the standard non-parametric bootstrap with 100 replicates.

2.1.3 *Molecular clock*

Molecular clock analysis using the least square dating (LSD2) method was performed (To et al. 2016). The collection date of every HA5 and HA7 DNA sequence was retrieved from the IRD website (uploaded on 15 April 2020). A home-designed Python script edited the date format to fulfil the IQ-Tree-command requirements. The best-used DNA model and previous ML-DNA trees were used to speed up calculation and fix the new time-tree global topology. The HA5 and HA7 time-trees with estimated date of every internal node (complemented with the confidence interval at 0.95) were generated and compared to previous trees and data from literature. The pairwise genetic distances from generated time-trees were computed and used as data for the analysis of RNA structure stability evolution based on the genetic distance to HPAIV emergence events.

2.1.4 *Ancestral sequence reconstruction of internal nodes of HA5/HA7 trees*

Every internal node of the HA5 and HA7 time-scaled trees was estimated by the empirical Bayesian method. Only HA-estimated sequences of internal nodes that meet the following criteria were retained for RNA structure predictions: (1) every nucleotide position around the HA-CS encoding region of the estimated HA sequence had to have a posterior probability superior to 0.80, and (2) the estimated HA sequence did not harbour a MBCS motif in its amino-acid sequence.

2.2 Analysis of the cSL thermodynamic stability and clustering of RNA structures based on phylogenetic information

2.2.1 *RNA structure prediction*

The aim of the study was to analyse the predicted RNA structures encompassing the conserved HA-CS encoding region of all H5/H7 LPAIV and HA sequences from other subtypes. During genome replication, the polymerase may be locked between the pairing of entering and exiting RNA strands that constitute the cSL, which can partially re-hybridize (Gultyaev et al. 2019). Moreover, a recent analysis has shown that during the complementary RNA synthesis, twenty-five to twenty-seven nucleotides of the template RNA are single-stranded in the polymerase elongation complex (Wandzik et al. 2020). In agreement with these observations, eighty-nucleotide and 100-nucleotide windows centred at the HA-CS were chosen for our RNA structure predictions. The RNAfold program from the ViennaRNA-2.4.3 package was used to assess the predicted RNA structure of each HA sequence (Lorenz

Downloaded from https://academic.oup.com/ve/article/7/2/veab093/6415083 by Institut National Polytechnique Toulouse user on 06 February 2023

et al. 2011). RNA structure predictions were performed on HA sequences from non-H5/H7 subtypes, H5/H7 LPAIVs, and the estimated internal nodes of H5 and H7 AIV from the HA phylogenetic trees. After the predictions were made, dG0 values (kcal/mol) of the minimal free energy (MFE) structures were extracted.

2.2.2 *Analysis of the link between the RNA structure stability and the genetic distance to emergence event*

We used an analytic pipeline to address the question of how structural features of the RNA secondary stem-loop evolve along the evolutionary pathway of HA sequences. Most of the codes were realized with the R-software 4.0.2 (R Core Team 2020) and Python 3.7 (Von Rossum and Drake 2009). First, phylogenetic trees of HA5 and HA7 sequences were analysed with the Ape 5.4 package (Paradis and Schliep 2019). A matrix containing the pairwise genetic distances between all terminal and estimated internal nodes was computed. By using a slightly modified version of the Ape function 'seq_root2tip', all lists of tree-nodes from the root to every tip were calculated, referred as 'evolutionary pathways'. Data from other variables were also extracted: the list of every node from the tree-root to a specified terminal node (HPAIV or LPAIV-HA sequence or taxon), the HA subtype (H5 or H7), the evolutionary group membership (mechanism of genetic evolution involved for the evolutionary pathways leading to HPAIV emergence event or LPAIV group when the last sequence from the list of internal nodes was an LPAIV), the MFE-dG0 values (kcal/mol) from RNA structure predictions, and finally the genetic distance between every node to the specified terminal node (HPAIV or LPAIV-HA sequence). Line plots of RNA structure stabilities based on the genetic distance to tree-tip and regressions using the 'lowess' method were generated with ggplot2 3.3.2 and tidyverse 1.3.0 packages (Wickham 2009).

2.2.3 *Clustering analysis of predicted RNA structures from ancestors*

A complementary approach was performed to evaluate structural differences or similarities in the predicted RNA structures encompassing the conserved HA-CS encoding region. We performed a clustering of RNA structures from all LPAIV-HA sequences uploaded from the NIAID IRD and from the predicted ancestors (both on the vRNA and cRNA) by RNAclust perl-script (Will et al. 2007). In brief, groups of sequences that share a common secondary RNA structure motif were identified. The base pair probability matrix of each sequence–secondary structure distribution was computed (by using RNAfold from Vienna RNA package). A sequence–structure alignment was subsequently calculated using LocARNA. Eventually, a hierarchical clustertree derived by WPGMA clustering of the pairwise alignment distances was generated, as previously described (Will et al. 2007). The significance level used to determine the optimal number of clusters was set to $k = 1$.

3. Results

3.1 Stability of predicted RNA structure encompassing the HA-CS

To test if the ability of H5 and H7 AIV to acquire MBCS correlated with specific differences in the cSL structure, we performed an analysis of the thermodynamic stability of the predicted cSL structure focusing on a region of eighty nucleotides. The thermodynamic stability of the predicted MFE structures was calculated and expressed in dG0 value (in kcal/mol): higher dG0 values indicated lower thermodynamic stabilities. As vRNA and cRNA can exhibit different RNA structures, we performed the analysis on both strands. All available HA-encoding sequences from avian origin were downloaded from NIAID IRD. Groups were designed in terms of the HA subtype. H17 and H18 HA CDS sequences (from bat origin) were also included in the analysis. On the eighty-nucleotide window, the thermodynamic stability of the predicted MFE structure was variable within and between HA subtypes (Fig. 2 and Table 1). The predicted MFE values were globally lower in the positive strand (cRNA) than in the negative strand (vRNA), with specificities amongst HA subtypes. Mean dG0 of −16.66 and −11.60 kcal/mol were observed for all HA sequences on the cRNA and vRNA, respectively. Lower means of dG0 were observed for sequences belonging to H4, H17, H8, H11, and H13 subtypes, ranging from −20.61 to −18.96 kcal/mol on cRNA and from −20.35 to −15.09 on vRNA. Neither HA5 nor HA7 sequences showed the lowest dG0 compared to all HA subtypes: −11.40 and −12.30 kcal/mol for H5 and H7 subtype, respectively. It is important to note that only one H4 AIV had been observed to naturally acquire an MBCS on its HA segment (Wong et al. 2014) and that the H4 group shows very low dG0. Altogether this analysis that took into account all the LPAIV's diversity confirmed that it is not possible to differentiate HA subtypes for their ability to experience an MBCS acquisition on their HA segment solely via the cSL predicted thermodynamic stability in the eighty-nucleotide analytic window, in agreement with previous data (Gultyaev et al. 2019). Similar results were found in the 100-nucleotide analytic window (data not shown).

3.2 Phylogenetic reconstruction of H5 and H7 HA segment and analysis of cSL stability evolution based on the genetic distance to every emergence event

Previous work (Gultyaev et al. 2019) and the results presented above analysed all the viral strains within the H5 and H7 subtypes as a homogenous group, with no distinction made between the lineages that gave rise to HPAIV emergences and the lineages for which no HPAIV emergence event has been described yet. We raised the hypothesis that taking into account the evolutionary trajectory could reveal a specific cSL structure pattern for LPAIVs that gave rise to HPAIVs.

As a first step towards this objective, we generated phylogenetic trees with 1,116 and 1,928 HA sequences for the H5 and H7 subtypes, respectively, corresponding to all HA sequences from H5/H7 LPAIV and known HPAIV emergences (detailed in Section 2). We performed phylogenetic tree reconstructions combined with molecular-clock analysis (Supplementary Figs S1 and S2). The major Eurasian and American lineages were both retrieved in our two inferred phylogenetic trees, as observed in a previous study (Lee, Criado, and Swayne 2020). Thirty-three documented HPAIV emergence events were found from our data sets, listed in Table 2. Fifteen of them were of H5 subtype and eighteen were of H7 subtype (Lee, Criado, and Swayne 2020). These HPAIV emergence events correspond to independent events. They individually appeared in specific clades, at different time and location, supporting a parallel-evolutionary LPAIV–HPAIV transition process (Escalera-Zamudio et al. 2020). As expected, seventeen HPAIV emergence events by insertion mechanism were retrieved in the phylogenetic trees (nine H5 and eight H7 HPAIVs). In the American lineages, only three H5 and two H7 insertion-driven HPAIV events were identified. For H5 AIVs, four over six HPAIVs that had arisen from substitutions appeared in the American lineage. This was the case for one over five event for H7 AIVs. Well-described emergence events were also retrieved, clustering with their most probable

Downloaded from https://academic.oup.com/ve/article/7/2/veab093/6415083 by Institut National Polytechnique Toulouse user on 06 February 2023

Table 1. Summary of descriptive statistics of the cSL structure stability among HA subtype on the eighty-nucleotide (A) and 100-nucleotide (B) windows.

HA subtype	*N*	cRNA							vRNA						
		Min	Max	Median	Mean	SD	SE	CI	Min	Max	Median	Mean	SD	SE	CI
A															
H1	806	−26.1	−12.6	−17.1	−17.57	2.07	0.07	0.14	−24.3	−8.7	−18.3	−17.57	2.61	0.09	0.18
H2	50	−22.8	−11.5	−15.3	−16.16	2.89	0.41	0.82	−17.1	−7.7	−10.1	−11.46	2.99	0.42	0.85
H3	1964	−24.7	−3.5	−15.1	−16.29	3.93	0.09	0.17	−20.8	−3.7	−13.8	−14.02	2.77	0.06	0.12
H4	50	−28.5	−15.9	−18.1	−20.61	4.09	0.58	1.16	−22.5	−12.2	−15.6	−16.41	2.72	0.38	0.77
H5	999 (over 1116 total HA sequences)	−23.8	−2.6	−9.4	−11.03	5.12	0.16	0.32	−21.1	−6.1	−11.4	−11.71	2.24	0.07	0.14
H6	1797	−24.5	−8.4	−13.6	−13.78	2.23	0.05	0.1	−23.3	−4.1	−9.6	−10.05	2.52	0.06	0.12
H7	1691 (over 1928 total HA sequences)	−25.9	−9.8	−17	−17.35	2.31	0.06	0.11	−22	−4.6	−12.3	−12.1	4	0.1	0.19
H8	162	−25.6	−16.4	−19.4	−20.3	2.16	0.17	0.34	−21.8	−12.4	−16.3	−15.9	1.86	0.15	0.29
H9	6178	−28.6	−8.2	−17.6	−17.79	2.55	0.03	0.06	−19.3	−0.4	−8.9	−9.16	1.87	0.02	0.05
H10	932	−23.1	−1.7	−17.3	−14.78	4.96	0.16	0.32	−17.5	−2.2	−11.5	−9.53	3.69	0.12	0.24
H11	749	−31.7	−13.2	−20.1	−20.02	1.95	0.07	0.14	−22.1	−6.77	−15.4	−15.36	2.27	0.08	0.16
H12	330	−27.5	−12.9	−19.1	−18.96	2.11	0.12	0.23	−23	−7.5	−16.3	−15.82	2.93	0.16	0.32
H13	384	−27.4	−11.3	−20.3	−19.9	3.91	0.2	0.39	−23.9	−8.9	−17.3	−15.93	3.14	0.16	0.32
H14	34	−23	−10.5	−12.9	−13.79	2.48	0.43	0.87	−21	−13.9	−15.4	−15.94	1.62	0.28	0.56
H15	16	−10.6	−8	−10.6	−9.63	1.17	0.29	0.62	−7.5	−6.2	−7.25	−6.94	0.61	0.15	0.32
H16	207	−23.6	−10	−13.1	−13.46	1.82	0.13	0.25	−16.1	−6.4	−10.3	−10.33	1.59	0.11	0.22
H17	2	−22.7	−18.4	−20.55	−20.55	3.04	2.15	27.32	−20.4	−20.3	−20.35	−20.35	0.07	0.05	0.64
H18	2	−13.3	−11.9	−12.6	−12.6	0.99	0.7	8.89	−11	−11	−11	−11	0	0	0
B															
H1	806	−31.6	−15.8	−26.5	−25.94	2.05	0.07	0.14	−26.3	−9.5	−21.7	−20.5	2.56	0.09	0.18
H2	50	−25.3	−18.8	−22.9	−22.86	1.41	0.2	0.4	−18.3	−8.7	−12.25	−12.89	2.99	0.42	0.85
H3	1964	−30.9	−6.5	−19.1	−19.99	3.72	0.08	0.17	−26.2	−5.5	−16	−16.78	3.28	0.07	0.15
H4	50	−35.4	−21.8	−25	−27.74	4.79	0.68	1.36	−25.9	−16.2	−19.5	−20.62	2.6	0.37	0.74
H5	999	−34.4	−10.6	−19.9	−20.71	4.99	0.16	0.31	−33.5	−11.5	−19.3	−19.89	2.73	0.09	0.17
H6	1797	−33.5	−12.9	−20.3	−20.22	3.08	0.07	0.14	−26.4	−5	−11.6	−12.1	2.91	0.07	0.14
H7	1691	−32.6	−15.6	−24.7	−24.2	2.61	0.06	0.12	−25.3	−9	−16.2	−16.58	3.26	0.08	0.16
H8	162	−35.5	−22.3	−27.8	−28.15	2.72	0.21	0.42	−25.9	−14.7	−18.1	−18.22	2.25	0.18	0.35
H9	6178	−39.1	−12.7	−23.9	−24.56	3.04	0.04	0.08	−25.4	−3.4	−14	−14.86	2.68	0.03	0.07
H10	932	−31.4	−1.7	−23.6	−20.68	6.22	0.2	0.4	−17.7	−2.2	−11.8	−11.51	2.22	0.07	0.14
H11	749	−36.8	−18	−26.7	−26.77	2.29	0.08	0.16	−26.6	−8.37	−20	−20.02	2.47	0.09	0.18
H12	330	−35.2	−18.8	−26.1	−26.2	2.52	0.14	0.27	−28.3	−13.2	−21.2	−20.53	3.07	0.17	0.33
H13	384	−35.6	−17.9	−26.2	−26.06	3.01	0.15	0.3	−33.6	−14.6	−21	−21.83	3.58	0.18	0.36
H14	34	−35.4	−20.2	−26.5	−25.99	2.9	0.5	1.01	−28	−18.6	−19.9	−21.09	2.29	0.39	0.8
H15	16	−18.1	−14.6	−17.1	−16.76	1.17	0.29	0.62	−18.6	−11	−12.5	−14.4	2.98	0.75	1.59
H16	207	−37.5	−16	−19.7	−20.53	4.1	0.29	0.56	−23.9	−7	−14.6	−14.13	2.9	0.2	0.4
H17	2	−30.4	−27.8	−29.1	−29.1	1.84	1.3	16.52	−27.6	−26.8	−27.2	−27.2	0.57	0.4	5.08
H18	2	−20.5	−20.3	−20.4	−20.4	0.14	0.1	1.27	−18.8	−16.9	−17.85	−17.85	1.34	0.95	12.07

'*N*': number of analysed sequences, 'min': minimum, 'max', maximum, 'SD': standard deviation of the mean, 'SE': standard error of the mean, 'CI': 95 per cent confidence interval of the mean.

Downloaded from https://academic.oup.com/ve/article/7/2/veab093/6415083 by Institut National Polytechnique Toulouse user on 06 February 2023

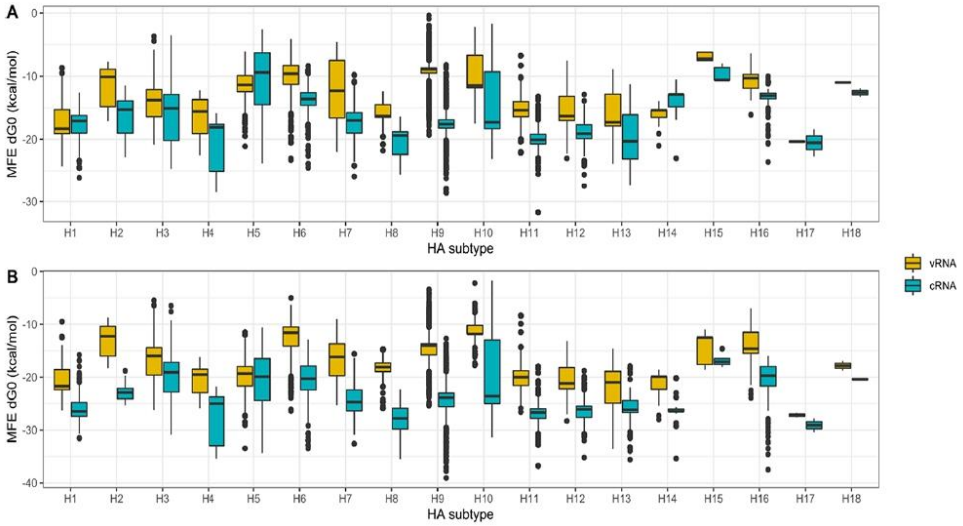


Figure 2. Box-plots of cSL structure stability (MFE dG0, kcal/mol) based on every HA subtype. RNA secondary structure predictions were performed and the cSL structure stability (MFE dG0, kcal/mol) calculated on two different analytic windows (80 and 100 nt encompassing the HA1/HA2 boundary region, panels A and B respectively), on both vRNA and cRNA.

co-circulating-related progenitor (Table 2; Supplementary Figs S1 and S2). HPAIV emergence events were not equally distributed among lineages: six over nine H5 HPAIV events by insertion were clustered into the Eurasian lineage, and the others were positioned into one clade of the American lineage. Four over six HPAIV emergence events via substitutions were within the American lineage. All the HA sequences related to recombination events clustered into the American lineage.

The time-trees were used to analyse the cSL structure pattern along the evolutionary pathways leading to final phylogenetic events (Supplementary Figs S1 and S2). Our objective was to determine if a specific cSL structure pattern was found along evolutionary pathways that gave rise to HPAIV emergences ('LPAIV-to-HPAIV' group) and whether this cSL structure pattern was absent from evolutionary pathways leading to LPAIVs as the final phylogenetic event ('LPAIV-only' control groups). As this work aimed to investigate the possible influence of cSL on the emergence, we decided to only select sequences that display non-consecutive dibasic codons in their cleavage site encoding regions, to remove potential compensatory and adaptive mutation events that appeared after the evolutionary LPAIV–HPAIV transition. Thus, the 'LPAIV-to-HPAIV' group only contained HA sequences from predicted ancestors and was terminated by the HPAIV precursor that harboured a monobasic HA-CS as the final phylogenetic event. In the same way, the 'LPAIV-only' control group contained HA sequences from predicted ancestors of LPAIVs located in different clades (in both Eurasian and American lineages). Admittedly, the control group may contain a fraction of LPAIVs en route towards a future HPAIV emergence. The 'LPAIV-only' control group was designed by selecting terminal LPAIV nodes in order to cover LPAIV nucleotide sequence diversity (list of HA-LPAIV sequence id in Supplementary Table S1). In this analysis, we distinguished 'LPAIV-only' group in two subgroups, based on the major lineage membership, to form the Eurasian and American 'LPAIV-only' groups. Indeed, notable differences may be present between HA sequences from different major lineages. HA nucleotide sequence of every internal tree-node from the tips to the root was estimated via ancestral sequence reconstruction. The MFE-dG0 were extracted from RNA structure prediction of each ancestor to analyse cSL stability along the individual evolutionary pathways as a function of the genetic distance to every final phylogenetic event, being either the HPAIV precursors that harbour a monobasic HA-CS or the members of the 'LPAIV-only' control groups: the x-axis displays the genetic distance from tree-tip ($x = 0$) towards the root of the phylogenetic tree (represented by the highest value on the x-axis) (Figs 3–6). As the phylogenetic HA5 and HA7 trees were non-ultrametric, lineages had variable total lengths of branches from the tree-root, which can be visualized as various line lengths.

Individual evolutionary pathways in the H5 'LPAIV-only' group showed significant variability in their cSL structure stability (Fig. 3). No apparent distinct pattern could be observed in the 'LPAIV-to-HPAIV' group (irrespective of the genetic mechanism involved in the MBCS acquisition). Nevertheless, specific HA vRNA sequences from American lineage ancestors that led to HPAIV event by insertions appeared to display increasing cSL stability values on the vRNA when getting closer to HP emergence, compared to the other estimated sequences (Fig. 3B). The genetically closest HA sequences leading to insertions had dG0 of around −15 kcal/mol, while HA sequences evolving to known LPAIVs had a mean-dG0 of −11.13 kcal/mol (fraction of the data corresponding to the 112 last sequences en route to the selected American-LP events: HA sequences from internal nodes located between 0 and 0.01 in terms of genetic distance to the tree-tips). When data from all evolutionary pathways were compiled and fitted using the loess regression method, we observed that the pattern of cSL structure stability evolved differently depending on the sequence sense (vRNA or cRNA) and as a function of the major lineage

Downloaded from https://academic.oup.com/ve/article/7/2/veab093/6415083 by Institut National Polytechnique Toulouse user on 06 February 2023

Table 2. HA sequences from known HPAIV emergence events retrieved from our phylogenetic analysis. For each HPAIV emergence, we supplied the country and the date (year) of isolation, the HA subtype, the HA-HPAIV sequence accession number, the sequence id, the phylogenetic-tree lineage membership, the closest LP HA sequence identified, the estimated LPAIV-HA sequence ancestor, and the genetic mechanism involved into the LPAIV-to-HPAIV transition.

Location (country)	Year of HP emergence	HA sub-type	HP-HA sequence accession number	HP sequence id	Genetic mechanism for LPAIV-HPAIV transition	Closest LP HA sequence accession number	Estimated ancestor (internal node id)	Major lineage membership
Italy	1997	H5	EF597263	A_chicken_Italy_312_1997_H5N2	Insertions	EF597262	Node786	Eurasian
Mexico	1994	H5	AB558473	A_chicken_Puebla_8623_607_1994_H5N2	Insertions	GU186573	Node150	American
Mexico	1995	H5	AB558474	A_chicken_Queretaro_14588_19_1995_H5N2	Insertions	GU186573	Node42	American
Ireland	1983	H5	GU052853	A_duck_Ireland_113_1983_H5N8	Insertions	CY021381	Node768	Eurasian
USA	1993	H5	AY444750	A_emu_Texas_39442_1993_H5	Insertions	U67783	Node154	American
China	1996	H5	AF144305	A_goose_Guangdong_1_1996_H5N1	Insertions	EU564114	Node1091	Eurasian
South Africa	2011	H5	JX069081	A_ostrich_South_Africa_AI2114_2011_H5N2	Insertions	DQ387854	Node947	Eurasian
South Africa	2004	H5	FJ519983	A_ostrich_South_Africa_N227_2004_H5N2	Insertions	EF591757	Node945	Eurasian
UK	1991	H5	GU052510	A_turkey_England_50_92_1991_H5N1	Insertions	KF435066	Node1101	Eurasian
France	2016	H5	KX014902	A_chicken_France_160013g_2016_H5N2	Substitutions	KF462362	Node956	Eurasian
Mexico	2005	H5	KM368285	A_chicken_Hidalgo_7637_05_2005_H5N2	Substitutions	KM368300	Node50	American
USA	1983	H5	CY015073	A_chicken_Pennsylvania_1_1983_H5N2	Substitutions	CY179779	Node721	American
Tawain	2008	H5	AB507264	A_chicken_Taiwan_A703_1_2008_H5N2	Substitutions	CY006040	Node121	Eurasian
USA	2004	H5	GU052644	A_chicken_Texas_298313_2_2004_H5N2	Substitutions	EF607872	Node160	American
Canada	1966	H5	M30122	A_turkey_Ontario_7732_1966_H5N9	Substitutions	CY087808	Node26	American
Australia	1976	H7	CY024786	A_chicken_Victoria_1976_H7N7	Insertions	Unknown[a]	Unknown[a]	Eurasian
Germany	1934	H7	CY077420	A_chicken_Rostock_45_1934	Insertions	Unknown[a]	Unknown[a]	Eurasian
USA	1994	H7	GU052922	A_Pekin_robin_California_30412_1994_H7N1	Insertions	MH574749	Node1032	American
USA	2016	H7	KU558906	A_turkey_Indiana_16_001403_1_2016_H7N8	Insertions	KY550896 or KY684303	Node1297	American
China	2016	H7	MF280190	A_chicken_Guangdong_Q1_2016_H7N9	Insertions	MF630285	Node219	Eurasian
Germany	1979	H7	U20459	A_chicken_Leipzig_79_H7N7	Insertions	KJ889439	Node1076	Eurasian
UK	1963	H7	U20462	A_turkey_England_1963_H7N3	Insertions	HM346485	Node40	Eurasian
Australia	2007	H7	CY061610	A_duck_Victoria_512_2007_H7N6	Substitutions	Unknown[a]	Unknown[a]	Eurasian
USA	2002	H7	AY240908	A_GuineaFowl_MA_148081_11_02_H7N2	Substitutions	EU742912	Node1830	American
Netherlands	2005	H7	CY077008	A_mallard_Netherlands_9_2005_H7N7	Substitutions	EF467826	Node931	Eurasian
Japan	2009	H7	AB538457	A_quail_Aichi_2_2009_H7N6	Substitutions	AF202235	Node1064	Eurasian
Korea	2009	H7	KC609770	A_wild_duck_Korea_CSM27_12_2009_H7N6	Substitutions	KC609767	Node598	Eurasian
Chile	2002	H7	AY303631	A_chicken_Chile_4322_02_H7N3	Recombination	AY303630	Node29	American
Canada	2004	H7	AY611524	A_chicken_British_Columbia_04_H7N3	Recombination	AY650270	Node1687	American
Canada	2007	H7	EU500860	A_chicken_SK_HR_00011_2007_H7N3	Recombination	KF573723	Node1557	American
Mexico	2012	H7	JX908509	A_chicken_Jalisco_12283_2012_H7N3	Recombination	KR077940	Node1175	American
USA	2017	H7	KY818811	A_chicken_Tennessee_17_007147_2_2017_H7N9	Recombination	MG266064	Node1271	American
Italy	1999	H7	KF493066	A_chicken_Italy_4845_1999_H7N1	Insertions of unknown origin	KF492994	Node712	Eurasian
Netherlands	2003	H7	EPI_ISL_391798	A_chicken_Netherlands_03008927_2003	Insertions of unknown origin	GU053030	Undetermined[b]	Eurasian
Spain	2009	H7	GU121458	A_chicken_Spain_6279_2_2009_H7N7	Recombination	MF575189	Undetermined[b]	Eurasian

[a]For some emergence events, the closest LPAI strain and/or the estimated ancestor (internal node harboring monobasic HA-CS motif) were impossible to be identified. The phylogenetic clade that includes the corresponding HPAI strain was only constituted of HPAI viruses.
[b]At the time of the data collection, the HPAI sequence was not present in the used public database for this work. The corresponding HPAI sequence for this emergence was added to our initial data set, and only the closest LPAI strain was identified for the RNA structure clustering analysis.

Downloaded from https://academic.oup.com/ve/article/7/2/veab093/6415083 by Institut National Polytechnique Toulouse user on 06 February 2023

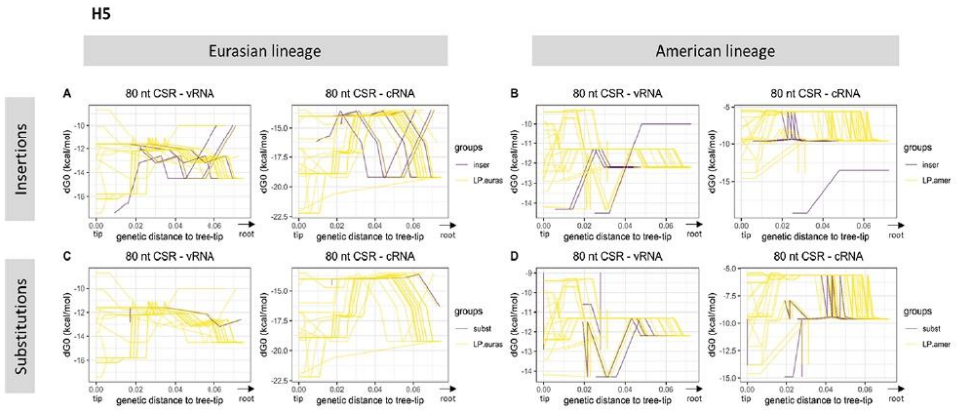


Figure 3. Variation of H5 cSL structure stability along the evolutionary pathways. Each related ancestral H5 HA sequences from the tree-root to a specific final phylogenetic event (HP emergence or not), also named as an evolutionary pathway, is represented. Variation of cSL structure stability (MFE dG0, kcal/mol) based on the genetic distance to the tree-tip (or the final phylogenetic event) is shown. Evolutionary pathways leading to an HP event or not are drawn in purple and yellow lines, respectively. The analysis has been split in terms of genetic mechanism involved in the HPAIV–LPAIV evolution (insertion, substitution) and major lineage membership (panels A–D). RNA structure predictions were performed on eighty nucleotides around the cleavage site encoding region ('80 CSR').

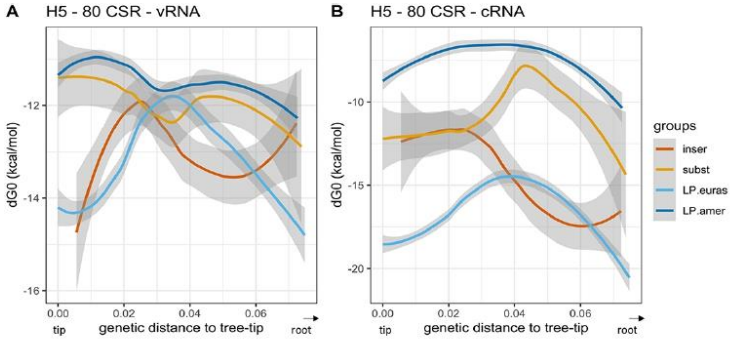


Figure 4. Local polynomial regression fitting of H5 cSL structure stability along the evolutionary pathways. The analysis was performed on both vRNA and cRNA (panels A and B, respectively). Pooled data from the different studied H5 evolutionary pathways are represented. Variation of cSL structure stability (MFE dG0, kcal/mol) based on the genetic distance to the tree-tip (or the final phylogenetic event) is shown. Local polynomial regression fitting (loess) smoothing method was used (confidence interval of 0.95 is shown in grey areas). Groups were designed in terms of the type of evolutionary pathway membership and the genetic mechanism involved in the LPAIV–HPAIV transition ('inser' and 'subst' groups in the legend box stand for evolutionary routes leading to HP event via nucleotide insertions or substitutions, respectively). Data from ancestral sequences that are not involved in any HP emergence event were divided in terms of the major lineage membership (LP.euras and LP.amer for evolutionary pathways leading to LP event from Eurasian and American lineage, respectively). RNA structure predictions were performed on eighty nucleotides around the cleavage site encoding region ('80 CSR').

membership (Fig. 4). The pattern of vRNA cSL stability was similar between the Eurasian 'LPAIV-only' group and 'LPAIV-to-HPAIV' group by insertion. Similarly, the pattern of vRNA cSL stability was also comparable between the American 'LPAIV-only' group and 'LPAIV-to-HPAIV' group by substitutions. A slight cSL stabilization could be noticed in ancestors that led to American HPAIV events via insertions, but the difference of cSL stability evolution in the 'LPAIV-only' American lineage appeared to be negligible (Fig. 4A). On the cRNA (Fig. 4B), stability variations were different between the 'LPAIV-only' groups belonging to different lineages. We also noted that stability pattern 'LPAIV-to-HPAIV' groups (by insertion and substitutions) reached a common stability value when getting closer to the HPAIV event. Similar results were found in the 100-nucleotide analytic window (Supplementary Fig. S3).

In the case of H7 sequences, evolutionary pathways that did not lead to any known emergence event also showed high variability in terms of MFE stability values (Fig. 5). We did not detect any pattern of cSL stability evolution specific to the 'LPAIV-to-HPAIV' group or the 'LPAIV-only' group, nor in the American or in the Eurasian lineages. Compiled and fitted data on H7 sequences revealed a considerable variation in cSL stability on vRNA (Fig. 6A). On the vRNA, ancestors in the American lineage that led to HP event by recombination evolved in the same way as sequences from the 'LPAIV-only' group and showed the highest cSL stability

Downloaded from https://academic.oup.com/ve/article/7/2/veab093/6415083 by Institut National Polytechnique Toulouse user on 06 February 2023

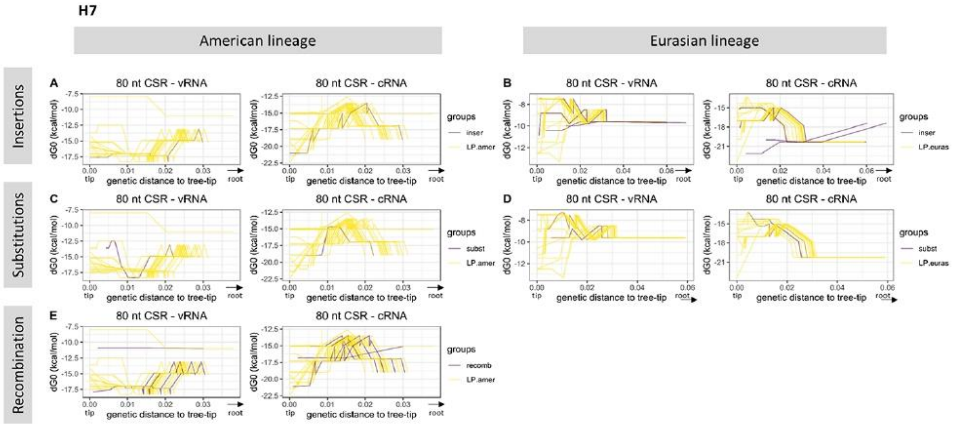


Figure 5. Variation of H7 cSL structure stability along the evolutionary pathways. Each related ancestral H7 HA sequences from the tree-root to a specific final phylogenetic event (HP emergence or not), also named as an evolutionary pathway, is represented. Variation of cSL structure stability (MFE dG0, kcal/mol) based on the genetic distance to the tree-tip (or the final phylogenetic event) is shown. Evolutionary pathways leading to an HP event or not are drawn in purple and yellow lines, respectively. The analysis has been split in terms of genetic mechanism involved in the HPAIV–LPAIV evolution (insertions, substitutions, and recombination) and major lineage membership (panels A–E). RNA structure predictions were performed on eighty nucleotides around the cleavage site encoding region ('80 CSR').

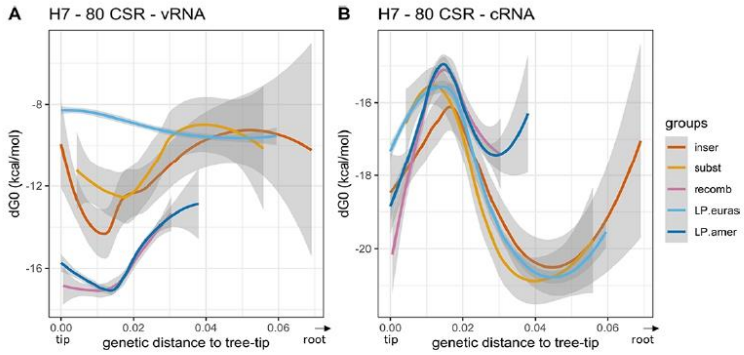


Figure 6. Local polynomial regression fitting of H7 cSL structure stability along the evolutionary pathways. The analysis was performed on both vRNA and cRNA (panels A and B, respectively). Pooled data from the different studied H7 evolutionary pathways are represented. Variation of cSL structure stability (MFE dG0, kcal/mol) based on the genetic distance to the tree-tip (or the final phylogenetic event) is shown. Local polynomial regression fitting (loess) smoothing method was used (confidence interval of 0.95 is shown in grey areas). Groups were designed in terms of the type of evolutionary pathway membership and the genetic mechanism involved in the LPAIV–HPAIV transition ('inser', 'subst', and 'recomb' groups in the legend box stand for evolutionary routes leading to HP event via nucleotide insertions, substitutions, or recombination, respectively). Data from ancestral sequences that are not involved in any HP emergence event were divided in terms of the major lineage membership (LP.euras and LP.amer for evolutionary pathways leading to LP event from Eurasian and American lineage, respectively). RNA structure predictions were performed on 80 nucleotides around the cleavage site encoding region ('80 CSR').

values (around −17 to −16 kcal/mol). Similarly, sequences leading to HP events by insertions or substitutions displayed the same evolutionary pattern as sequences from the 'LPAIV-only' group, with an intermediate cSL stability score (−12 to −10 kcal/mol). Interestingly, cSL stability evolution from cRNA predictions was almost identical for all groups (Fig. 6B). Altogether, this analysis revealed that no apparent differences could be observed in the evolution of cSL structure thermodynamic stability between evolutionary pathways en route to HPAIV compared to evolutionary pathways en route to LPAIV events. Similar results were found in the 100-nucleotide analytic window (Supplementary Fig. S4).

3.3 Clustering of predicted RNA structures from LPAIV-HA sequences and predicted ancestors

RNA molecules with similar thermodynamic stabilities can have different secondary conformations. Thus, to determine if specific RNA structure topologies could be identified among HA sequences evolving towards HPAIV, we analysed RNA structures using the RNA structure clustering software RNAclust. The analysis was performed on all LPAIV-HA sequences and the HA sequences estimated at each internal node of the time-trees, including the HPAIV progenitors. RNA structure clusters that contain ancestors specifically leading towards HPAIV

Downloaded from https://academic.oup.com/ve/article/7/2/veab093/6415083 by Institut National Polytechnique Toulouse user on 06 February 2023

emergence events were highlighted by a blue rectangle on the computed cluster trees (Figs 7, 9 and 10). Two HPAIV-HA sequences that were not present in the NIAID IRD at the time of initial data retrieval were subsequently downloaded from the IRD or GISAID databases and included in the RNA clustering analysis. For A/chicken/Spain/6279-2/2009 (H7N7) (Spain 2009) and A/chicken/Netherlands/03008927/2003 (H7N3) (NL 2003), the phylogenetically and genetically closest LPAIV-HA nucleotide sequences were considered as their direct ancestors (Table 2).

In the case of HA5 sequences, we identified eight different clusters containing HPAIV ancestors on the vRNA and eight clusters containing HPAIV ancestors on the cRNA (Table 3). All the ancestors directed towards HPAIV events via insertions were grouped in three and two different clusters on the vRNA

Table 3. Additional information on the computed groups from H5 sequence clustering.

Evolutionary direction of the HP-ancestor(s) contained in the cluster	Cluster id	SCI	Consensus MFE (dG0, kcal/mol)	Number of leaves
A				
To emergence events via insertions	63	0.9411	−16.81	7
	57	0.4062	−6.48	35
To emergence events via substitutions	615	0.7796	−8.07	45
	256	0.8737	−10.93	180
	990	0.9647	−12.84	53
	822	0.9612	−9.9	208
	372	0.8722	−13.4	19
To different emergence events via insertions or substitutions	931	0.9719	−14.14	49
B				
To emergence events via insertions	66	0.4918	−9.47	44
To emergence events via substitutions	643	0.835	−12.4	2
	658	0.8839	−14.36	15
	350	0.8136	−13.01	283
	619	0.9452	−13.8	33
	1010	0.6408	−5.45	349
	640	0.9227	−13.29	12
To different emergence events via insertions or substitutions	1316	0.6489	−6.61	55

The computed groups from H5 sequence clustering on vRNA (A) and cRNA (B) with additional information are shown:
- The unique evolutionary pathway that the HP-ancestor present in the cluster is following.
- Cluster id (which also corresponds to the multiple alignment file id in the RNAClust outputs, with significance level for clustering $k = 1$).
- Minimal free energy of the consensus secondary structure of the grouped sequences.
- Number of leaves in the cluster.

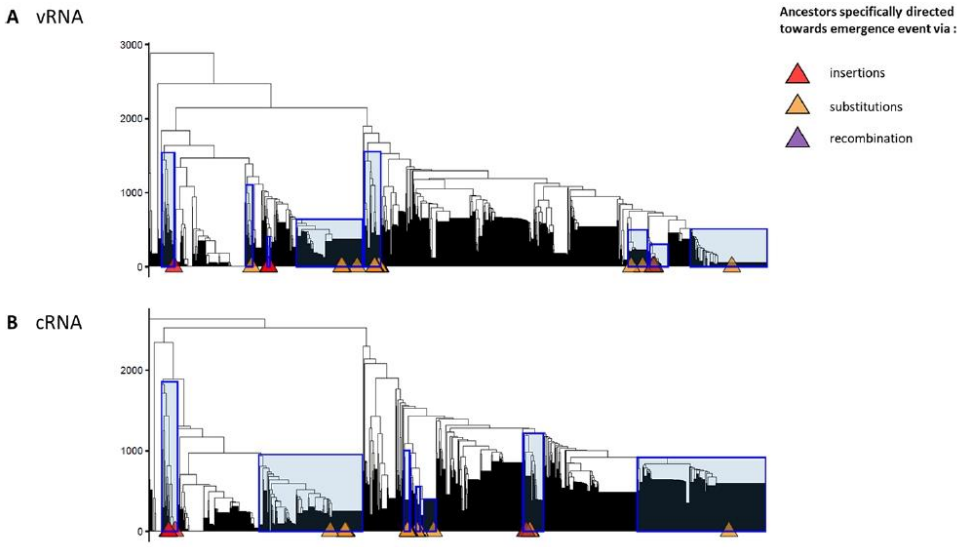


Figure 7. Cluster trees of H5-RNA secondary structures. Cluster trees of H5-RNA secondary structures on vRNA (A) and cRNA (B) have been computed via RNAclust with the RNAsoup option (with the significance level $k = 1$). Determined clusters containing ancestor(s) that is (are) specifically leading towards one type of emergence or any HP event are highlighted on the tree with blue rectangles. The tips annotated with coloured triangles indicate the HPAIV-directed ancestors (see legend). The y-axis displays the distance between RNA structures.

Downloaded from https://academic.oup.com/ve/article/7/2/veab093/6415083 by Institut National Polytechnique Toulouse user on 06 February 2023

and cRNA, respectively. The ancestors directed towards HPAIV events via substitutions were grouped in six and seven distinct clusters on the vRNA and cRNA (Fig. 7). Moreover, the clusters of RNA structures containing HPAIV ancestors were also composed of ancestors leading towards LPAIVs as final phylogenetic events. HPAIV ancestor clusters had different consensus MFE values (Table 3) and various consensus MFE structures (Fig. 8 and Supplementary Fig. S5). On the vRNA, we observed a simple stem-loop without multi-branching junctions with the presence of varying numbers and size of internal bulges in the stem (cluster # 63, 372, 822, 931, and 990). More complex stem-loop structures were also noticed like two-way (cluster # 615) three-way (cluster # 57), and four-way junctions (cluster # 256), with branching secondary stem-loops. Similar structures were observed among clusters based on cRNA (Supplementary Fig. S5). As mentioned in previous studies (Gultyaev et al. 2016, 2019; Nao et al. 2017), the two triplets encoding the cleavage-site motif R-G were positioned at the predicted cSL's apical loop (e.g. consensus of vRNA MFE structure of cluster # 822, 990, and 931). However, the R-G encoding codons could be positioned outside the apical loop in a bulging region or in some base pairings (e.g. cluster # 63 and 372).

In the case of HA7 sequences, we identified ten different clusters containing HPAIV ancestors on the vRNA and twelve clusters containing HPAIV ancestors on the cRNA (Table 4). Ancestors that were specifically leading towards HPAIV events via insertions were grouped in six different clusters on the vRNA and cRNA (Fig. 9). The ancestors directed towards H7 HPAIV events via substitutions were grouped in three clusters on the vRNA and cRNA. As observed with HA5 sequences, consensus MFE RNA structures of clusters that contained H7 HPAIV ancestors via substitutions or insertions exhibited various topologies (Fig. 11 and Supplementary Fig. S6). On the vRNA, we observed a simple stem-loop without multi-branching junctions (cluster # 2451, 2373, and 1831) and a more complex stem-loop with three-way junctions (cluster # 1313), with consensus MFE ranging from 7.40 to 10 kcal/mol. Similar structures were observed among clusters based on cRNA (Supplementary Fig. S6). The ancestors directed towards H7 HPAIV events via recombination were grouped in three clusters on the vRNA and five clusters on the cRNA (Figs 9 and 10). For the Eurasian H7 lineage HP emergence events, the ancestor of the H7N7-HP emergence in Spain (2009) did not cluster with the other ancestors of recombination events and was located in cluster # 1608. The vRNA structure of the ancestors of H7N1-Italian (1999) and H7N3-Dutch (2003) HP emergence events for which the genetic mechanism for LPAIV-HPAIV transition is currently debated (insertion or recombination) was located in distinct groups that did not contain ancestors of recombination events (Gultyaev et al. 2021). Cluster # 802 contained vRNA structures from four over five HPAIV progenitors via recombination belonging to the American H7 lineage, suggesting a possible evolutionary convergence for this specific genetic mechanism of HPAIV emergence. To verify that the vRNA structure clustering was specific to H7 HPAIV precursors via recombination, we checked the position of RNA structure from LPAIVs of the American lineage 'LPAIV only' group on the HA7-vRNA cluster tree. The American lineage 'LPAIV-only' sequences were evenly distributed among the different clusters (Fig. 10). The ancestor of the H7N3-HP emergence in Chile (accession number AY303631) that did not cluster with the other ancestors of recombination events was located in cluster # 1181. We noted that the MFE consensus structure of this cluster appeared similar in shape to cluster #802 (Fig. 11).

This analysis thus revealed that most of the American H7 ancestors (four over five events of recombination belonging to the American lineage) leading towards HPAIV emergence events via recombination shared similar cSL RNA structure topologies.

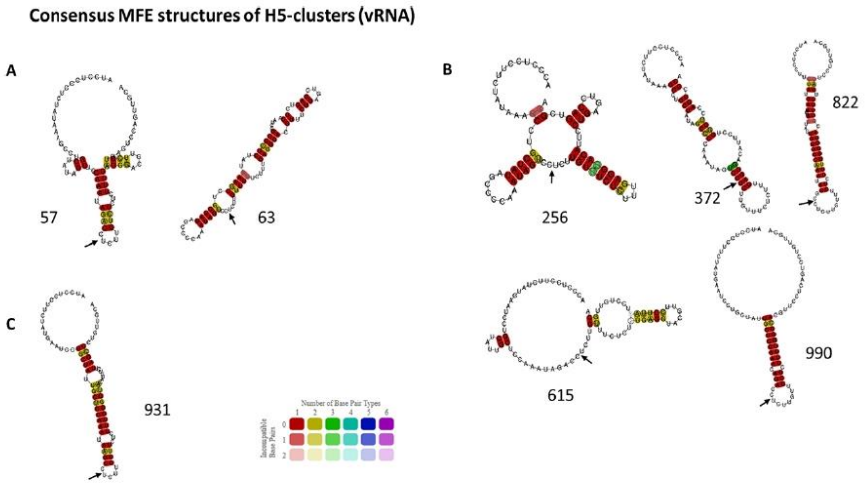


Figure 8. Consensus MFE structures of H5 clusters. Consensus MFE structures of H5 clusters from vRNA sequences that contain ancestors directed towards HP event via insertions (A), via substitutions (B), or both types of emergence (C) were generated by LocARNA. Cluster id is indicated next to the corresponding RNA structure. The HA-CS is indicated by a black arrow placed between the R/G encoding codons.

Downloaded from https://academic.oup.com/ve/article/7/2/veab093/6415083 by Institut National Polytechnique Toulouse user on 06 February 2023

Table 4. Additional information on the computed groups from H7 sequence clustering.

Evolutionary direction of the HP ancestor(s) contained in the cluster	Cluster id	SCI	Consensus MFE (dG0, kcal/mol)	Number of leaves
A				
To emergence events via insertions	2451	0.7872	−7.4	19
	1831	0.7458	−8.86	126
	2373	0.797	−7.49	500
	1313	0.8993	−10	7
To emergence events via substitutions	1524	0.6843	−8.73	162
	197	0.8716	−12.94	170
To emergence events via recombination	1181	0.9816	−11.19	7
	1608	0.9613	−12.7	83
To different emergence events via recombination or insertions	802	0.8847	−16.48	466
To different emergence events via insertions or substitutions	1307	0.7928	−8.99	88
B				
To emergence events via insertions	754	0.8358	−19.25	94
	2201	0.875	−18.9	23
	2076	0.8173	−14.87	36
	1816	0.9049	−17.06	500
	2173	0.9279	−16.04	97
To emergence events via substitutions	1281	0.9073	−19.81	62
	630	1.0059	−19.01	191
To emergence events via recombination	1028	0.9287	−17.13	275
	199	0.9827	−15.3	150
	355	1.0213	−21.28	19
	1218	1.0368	−16.77	4
To different emergence events via insertions, substitutions, or recombination	2522	0.8667	−14.95	304

The computed groups from H7 sequences clustering on vRNA (A) and cRNA (B) with additional information are shown:
- The unique evolutionary pathway that the HP ancestor present in the cluster is following.
- Cluster id (which also corresponds to the multiple alignment file id in the RNAClust outputs, with significance level for clustering $k = 1$).
- Minimal free-energy of the consensus secondary structure of the grouped sequences.
- Number of leaves in the cluster.

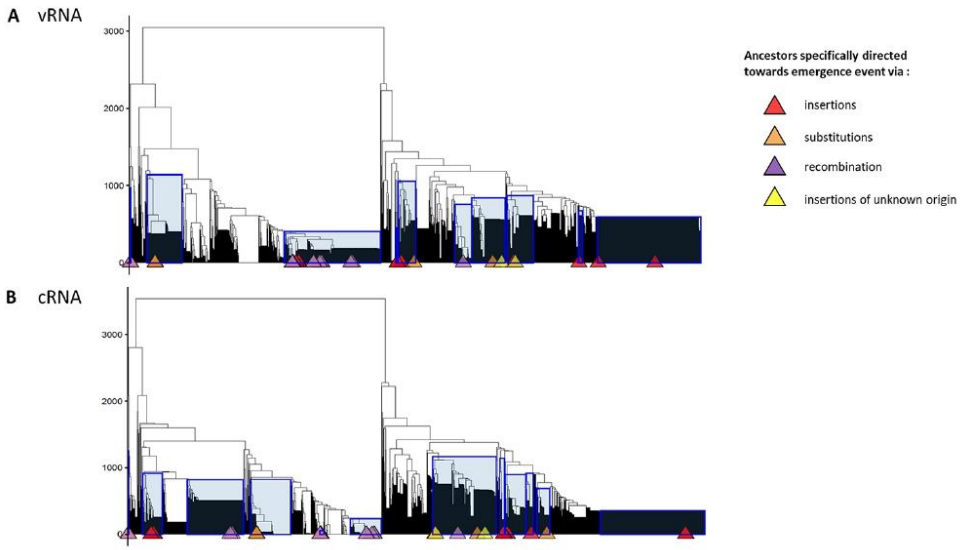


Figure 9. Cluster trees of H7-RNA secondary structures. Cluster trees of H7-RNA secondary structures on vRNA (A) and cRNA (B) have been computed via RNAclust with the RNAsoup option (with the significance level $k = 1$). Determined clusters containing ancestor(s) that is (are) specifically leading towards one type of emergence or any HP event are highlighted on the tree with blue rectangles. The tips annotated with coloured triangles indicate the HPAIV-directed ancestors (see legend). The y-axis displays the distance between RNA structures.

Downloaded from https://academic.oup.com/ve/article/7/2/veab093/6415083 by Institut National Polytechnique Toulouse user on 06 February 2023

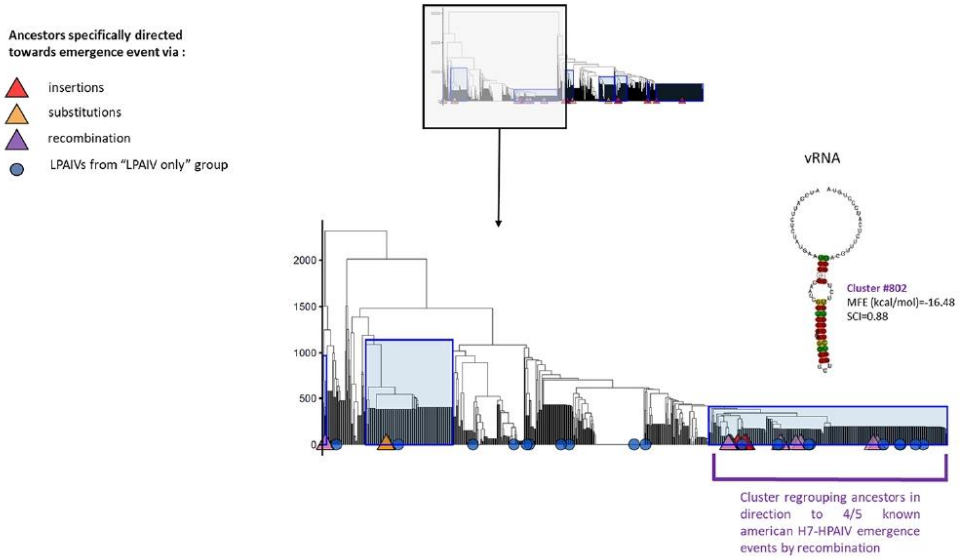


Figure 10. Zoom on the vRNA H7-RNA secondary structures clusters containing American lineage H7 ancestors of HPAIV via recombination. Zooming on the left part of Fig. 9A to display a more detailed view on the H7-RNA secondary structures clusters containing American lineage H7 ancestors of HPAIV via recombination. Determined clusters containing ancestor(s) that is (are) specifically leading towards one type of emergence or any HP event are highlighted on the tree in blue rectangles. The tips annotated with coloured triangles indicate the HPAIV-directed ancestors. The consensus MFE structure generated by LocARNA corresponding to Cluster # 802 is shown above this cluster.

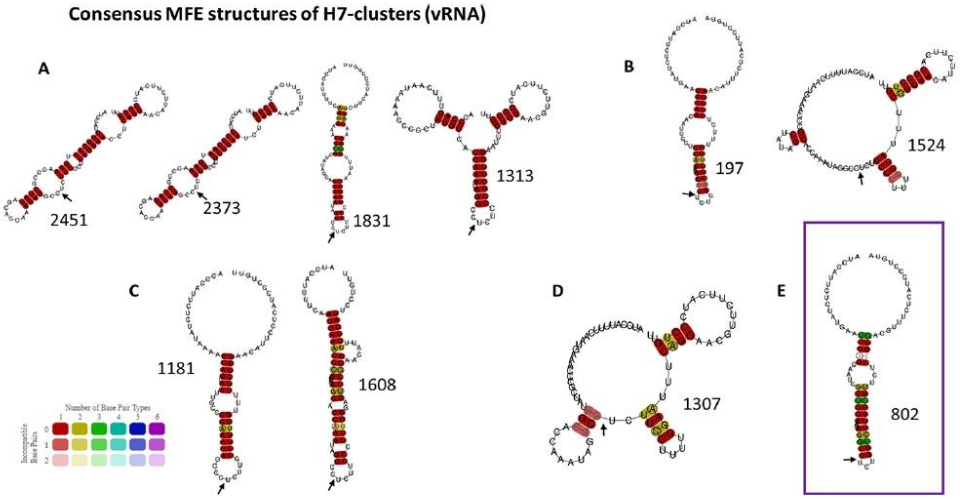


Figure 11. Consensus MFE structures of H7 clusters. Consensus MFE structures of H7 clusters from vRNA sequences that contain ancestors directed only towards HP event via insertions (A), via substitutions (B), via recombination (C), via insertions or substitutions (D), and via recombination or insertions (E) were generated by LocARNA. Cluster id is indicated next to the corresponding RNA structure. A violet rectangle frames the cluster's consensus RNA structure regrouping the HPAIV ancestors of most of the H7 recombination events. The HA-CS is indicated by a black arrow placed between the R/G encoding codons.

Downloaded from https://academic.oup.com/ve/article/7/2/veab093/6415083 by Institut National Polytechnique Toulouse user on 06 February 2023

4. Discussion

Previous studies suggested a role of the cSL structure in acquiring a MBCS and thus in the evolution towards HPAIVs in the case of H5/H7 AIV (Garcia et al. 1996; Perdue et al. 1997; Gultyaev et al. 2016, 2019; Nao et al. 2017; Dietze et al. 2018; Beerens et al. 2020). However, these studies suffered from various limitations. First, most of them only described the RNA topology of the HA1/HA2 boundary region of one HPAIV or one LPAIV-derived HA sequence (Garcia et al. 1996; Perdue et al. 1997; Dietze et al. 2018; Beerens et al. 2020). Secondly, RNA structure predictions were mostly performed on a short sequence window and focused on the cSL's apical loop (Garcia et al. 1996; Perdue et al. 1997; Nao et al. 2017; Dietze et al. 2018; Beerens et al. 2020), making any conclusion dubious because of possible physical incompatibilities of such RNA refolding inside the polymerase catalytic site (Wandzik et al. 2020). Finally, no study took into account the natural history of the HA sequence and its genetic distance from the known emergence event. Previous results (Escalera-Zamudio et al. 2020) and the present phylodynamic analysis showed that H5/H7 LPAIV–HPAIV emergence was a parallel evolutionary process, as several emergences appeared in different clades. Therefore, taking into account the phylogenetic relationships to explore the cSL evolution and its influence in LPAIV-HPAIV transition appeared necessary. By taking into account the evolutionary history of the HA sequences, the present work addressed the following question: is there any feature in the cSL that would explain why some H5/H7 LPAIVs have the ability to acquire a MBCS?

Analysis of the cSL thermodynamic stability was insufficient to explain why only specific H5 and H7 LPAIVs could acquire the MBCS pathogenic motif. No specific feature in the cSL was detected for the HPAIV emergence events via nucleotide insertions or substitutions. Interestingly, we observed a specific clustering of vRNA structures in HA1/HA2 encoding region from most of the American H7 HPAIV ancestors via recombination events. We therefore hypothesize that a putative evolutionary convergence mechanism involving the HA1/HA2 vRNA topology contributes to H7 HPAIV emergence via recombination (Figs 10 and 11). Comparable findings were observed in the HIV genome, where a connexion could be made between specific RNA structured encoding regions (C2 hairpin and RRE element) and hotspots for recombination events (Simon-Loriere et al. 2009, 2010). Moreover, another study showed that recombination events in H7 AIVs are more likely to occur at specific locations in ribosomal, transfer, and viral RNAs, corresponding to small nucleolar RNA (snoRNA)-binding regions (Gultyaev et al. 2021). A possible link between snoRNA binding, recombination by template switching, and the involvement of such RNA structure must be more fully explored.

Influenza virus evolution is driven by several viral and environmental factors, making these biological processes complex. Therefore, LPAIV-to-HPAIV evolution might result from multiple interacting and cooperating factors, involving other known pathogenicity factors (Ayllon and García-Sastre 2015; Wang et al. 2008; Nogales et al. 2018). Moreover, trans-activating factors have also been described as a necessary cofactor for RNA editing, as it was observed in the case of Ebola virus (Mehedi et al. 2013). In this study, the viral protein VP30 was shown to interact with a specific predicted-RNA structure located upstream of the nucleotide-insertion site, which eventually controls RNA editing. Trans-acting proteins might influence in some ways, directly or indirectly, the evolution of the HA encoding segment. Several mutations have been described in the HA segment (out of the cleavage site encoding region) and in other genes (PB1, NS1, NP, M1), which were suggested to be associated with LPAIV-HPAIV evolution (Horimoto and Kawaoka 1995; Dundon et al. 2006; Monne et al. 2014; Richard et al. 2017; Gutierrez, Escalera-Zamudio, and Pybus 2019; Suttie et al. 2019; Laleye and Abolnik 2020). However, it is still difficult to assess whether these changes are necessary to the HPAIV emergence or whether they derived from subsequent adaptations. To precisely limit this bias, we performed our phylogenetically informed analyses with HA sequences that derived from LPAIVs and predicted ancestors, which all harbour monobasic HA-CS motif.

The biological relevance of the shared RNA structure identified amongst H7 HPAIV ancestors via recombination must be further explored to establish a functional link between RNA structure and evolution via recombination. At least, these predicted structures constitute a new marker of HA sequences that evolved towards HPAIV via recombination, which may be used to generate a new classification of viral sequences. Moreover, predicted RNA structures in the HA1/HA2 boundary region must be compared with RNA structures determined chemically in the context of the NP and the polymerase complex (Ferhadian et al. 2018). Other structured regions of the vRNA have been shown to be detectable by such method (Kobayashi et al. 2016; Ferhadian et al. 2018; Dadonaite et al. 2019), indicating that the predicted RNA structures could exist and may have a role in influenza virus evolution.

As the cSL's biological role is still under study, it may have other roles beyond evolution. Other RNA structures located at the internal region of other vRNA (M and NS segments) have been linked to genome packaging (Gavazzi et al. 2013). Moreover, a recent study has shown that different AIVs describe extensive intra- and inter-segment interactions that could control genome-packaging and influence reassortments (Dadonaite et al. 2019). An outstanding question is whether the HA cSL structure could be under positive selection pressure during the LP-to-HP transition to maintain intra- or inter-segment interactions that may be important for genome packaging. Nevertheless, synonymous mutations affecting the predicted cSL structure on the HA segment did not result in a detectable reduction in viral replication, thus questioning the role of cSL structure in genome packaging (Gultyaev et al. 2016). As a consequence, further investigations have to be set forward to investigate the HA cSL's biological role.

4.1 Major findings

- Based on the analysis of the cSL across HA subtypes and along the distinct evolutionary pathways, we conclude that the thermodynamic stability of the cSL cannot be used to explain why specific H5/H7 HA sequences can acquire the MBCS pathogenic motif.
- Using RNA clustering analyses, we observed that most of the American H7 emergence events via recombination (four over five) shared the same cSL structure, suggesting a possible role of RNA topology in polymerase jump and template switching.

Data availability

Data and codes will be made available upon request.

Supplementary data

Supplementary data is available at *Virus Evolution* online.

Downloaded from https://academic.oup.com/ve/article/7/2/veab093/6415083 by Institut National Polytechnique Toulouse user on 06 February 2023

Acknowledgement

The authors thank Roland Marquet (Institut de Biologie Moléculaire et Cellulaire, Strasbourg, France) for insightful discussions.

Funding

This work was funded by a grant from the Agence Nationale de la Recherche (ANR-16-CE35-0005-01) to Romain Volmer. Pierre Bessière was supported by a Ph.D. scholarship funded by the Region Occitanie (France) and by the Chaire de Biosécurité at the École Nationale Vétérinaire de Toulouse (French Ministry of Agriculture). Gabriel Dupré is supported by Ph.D. scholarship funded by the French Ministry of Research and Education.

Conflict of interest: None declared.

References

Abdelwhab, E.-S. M., Veits, J., and Mettenleiter, T. C. (2013) 'Genetic Changes that Accompanied Shifts of Low Pathogenic Avian Influenza Viruses toward Higher Pathogenicity in Poultry', *Virulence*, 4: 441.

Ayllon, J., and Adolfo, G.-S. (2015) 'The NS1 Protein: A Multitasking Virulence Factor'. In: Oldstone M. B. A., and Compans R. W. (eds) *Influenza Pathogenesis and Control - Volume II*, pp. 73–107. Cham: Springer International Publishing.

Beerens, N. et al. (2020) 'Emergence and Selection of a Highly Pathogenic Avian Influenza H7N3 Virus', *Journal of Virology*, 94: e01818–19.

Bui, C. et al. (2016) 'A Systematic Review of the Comparative Epidemiology of Avian and Human Influenza A H5N1 and H7N9 - Lessons and Unanswered Questions', *Transboundary and Emerging Diseases*, 63: 602–20.

Chang, S. et al. (2015) 'Cryo-EM Structure of Influenza Virus RNA Polymerase Complex at 4.3 Å Resolution', *Molecular Cell*, 57: 925–35.

Dadonaite, B. et al. (2019) 'The Structure of the Influenza A Virus Genome', *Nature Microbiology*, 4: 1781–9.

Dietze, K. et al. (2018) 'From Low to High Pathogenicity—Characterization of H7N7 Avian Influenza Viruses in Two Epidemiologically Linked Outbreaks', *Transboundary and Emerging Diseases*, 65: 1576–87.

Dundon, W. G. et al. (2006) 'Progressive Truncation of the Non-Structural 1 Gene of H7N1 Avian Influenza Viruses following Extensive Circulation in Poultry', *Virus Research*, 119: 171–6.

Escalera-Zamudio, M. et al. (2020) 'Parallel Evolution in the Emergence of Highly Pathogenic Avian Influenza A Viruses', *Nature Communications*, 11: 5511.

Ferhadian, D. et al. (2018) 'Structural and Functional Motifs in Influenza Virus RNAs', *Frontiers in Microbiology*, 9: 559.

García, M. et al. (1996) 'Heterogeneity in the Haemagglutinin Gene and Emergence of the Highly Pathogenic Phenotype among Recent H5N2 Avian Influenza Viruses from Mexico', *Journal of General Virology*, 77: 1493–504.

Garten, W., and Klenk, H. D. (1999) 'Understanding Influenza Virus Pathogenicity', *Trends in Microbiology*, 7: 99–100.

Gavazzi, C. et al. (2013) 'An in Vitro Network of Intermolecular Interactions between Viral RNA Segments of an Avian H5N2 Influenza A Virus: Comparison with A Human H3N2 Virus', *Nucleic Acids Research*, 41: 1241–54.

Geller, R. et al. (2015) 'The External Domains of the HIV-1 Envelope are a Mutational Cold Spot', *Nature Communications*, 6: 1–9.

——et al. (2016) 'Highly Heterogeneous Mutation Rates in the Hepatitis C Virus Genome', *Nature Microbiology*, 1: 1–7.

Gultyaev, A. P. et al. (2016) 'Subtype-Specific Structural Constraints in the Evolution of Influenza A Virus Hemagglutinin Genes', *Scientific Reports*, 6: 38892.

——et al. (2019) 'Conserved Structural RNA Domains in Regions Coding for Cleavage Site Motifs in Hemagglutinin Genes of Influenza Viruses', *Virus Evolution*, 5: vez034.

——et al. (2021) 'Insertions of Codons Encoding Basic Amino Acids in H7 Hemagglutinins of Influenza A Viruses Occur by Recombination with RNA at Hotspots near SnoRNA Binding Sites', *RNA*, 27: 123–32.

Gutierrez, B., Escalera-Zamudio, M., and Pybus, O. G. (2019) 'Parallel Molecular Evolution and Adaptation in Viruses', *Current Opinion in Virology*, 34: 90–6.

Horimoto, T. et al. (1994) 'Proprotein-Processing Endoproteases PC6 and Furin Both Activate Hemagglutinin of Virulent Avian Influenza Viruses', *Journal of Virology*, 68: 6074–8.

——et al. (1995) 'Origin and Molecular Changes Associated with Emergence of a Highly Pathogenic H5N2 Influenza Virus in Mexico', *Virology*, 213: 223–30.

Horimoto, T., and Kawaoka, Y. (1995) 'Molecular Changes in Virulent Mutants Arising from Avirulent Avian Influenza Viruses during Replication in 14-Day-Old Embryonated Eggs', *Virology*, 206: 755–9.

Kalyaanamoorthy, S. et al. (2017) 'ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates', *Nature Methods*, 14: 587–9.

Kobayashi, Y. et al. (2016) 'Computational and Molecular Analysis of Conserved Influenza A Virus RNA Secondary Structures Involved in Infectious Virion Production', *RNA Biology*, 13: 883–94.

Krammer, F. et al. (2018) 'Influenza', *Nature Reviews Disease Primers 2018 4:1*, 4: 1–21.

Laleye, A. T., and Abolnik, C. (2020) 'Emergence of Highly Pathogenic H5N2 and H7N1 Influenza A Viruses from Low Pathogenic Precursors by Serial Passage in Ovo. Edited by Camille Lebarbenchon', *PLoS One*, 15: e0240290.

Le Mercier, P., and Kolakofsky, D. (2019) 'Bipartite Promoters and RNA Editing of Paramyxoviruses and Filoviruses', *RNA*, 25: 279–85.

Lee, D.-H., Criado, M. F., and Swayne, D. E. (2020) 'Pathobiological Origins and Evolutionary History of Highly Pathogenic Avian Influenza Viruses', *Cold Spring Harbor Perspectives in Medicine*, 11: a038679.

Long, J. S. et al. (2019) 'Host and Viral Determinants of Influenza A Virus Species Specificity', *Nature Reviews Microbiology*, 17: 67–81.

Lorenz, R. et al. (2011) 'ViennaRNA Package 2.0', *Algorithms for Molecular Biology*, 6: 26.

Mehedi, M. et al. (2013) 'Ebola Virus RNA Editing Depends on the Primary Editing Site Sequence and an Upstream Secondary Structure', *PLoS Pathogens*, 9: e1003677.

Minh, B. Q. et al. (2020) 'IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era', *Molecular Biology and Evolution*, 37: 1530–4.

Monne, I. et al. (2014) 'Emergence of a Highly Pathogenic Avian Influenza Virus from a Low-Pathogenic Progenitor', *Journal of Virology*, 88: 4375–88.

Munster, V. J. et al. (2010) 'Insertion of a Multibasic Cleavage Motif into the Hemagglutinin of a Low-Pathogenic Avian Influenza H6N1 Virus Induces a Highly Pathogenic Phenotype', *Journal of Virology*, 84: 7953–60.

Nao, N. et al. (2017) 'Genetic Predisposition to Acquire a Polybasic Cleavage Site for Highly Pathogenic Avian Influenza Virus Hemagglutinin', *MBio*, 8: 1.

Downloaded from https://academic.oup.com/ve/article/7/2/veab093/6415083 by Institut National Polytechnique Toulouse user on 06 February 2023

Nguyen, L. T. et al. (2015) 'IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies', *Molecular Biology and Evolution*, 32: 268–74.

Nogales, A. et al. (2018) 'Modulation of Innate Immune Responses by the Influenza A NS1 and PA-X Proteins', *Viruses*, 10: 12.

Paradis, E., and Schliep, K. (2019) 'Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R', *Bioinformatics*, 35: 526–8.

Pathak, V. K., and Temin, H. M. (1992) '5-Azacytidine and RNA Secondary Structure Increase the Retrovirus Mutation Rate', *Journal of Virology*, 66: 3093–100.

Perdue, M. L. et al. (1997) 'Virulence-Associated Sequence Duplication at the Hemagglutinin Cleavage Site of Avian Influenza Viruses', *Virus Research*, 49: 173–86.

—— and Suarez, D. L. (2000) 'Structural Features of the Avian Influenza Virus Hemagglutinin that Influence Virulence', *Veterinary Microbiology*, 74: 77–86.

Pita, J. S. et al. (2007) 'Environment Determines Fidelity for an RNA Virus Replicase', *Journal of Virology*, 81: 9072–7.

Richard, M. et al. (2017) 'Mechanisms and Risk Factors for Mutation from Low to Highly Pathogenic Avian Influenza Virus.', *EFSA Supporting Publications*, 14: 10.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>.

Shabman, R. S. et al. (2014) 'Deep Sequencing Identifies Noncanonical Editing of Ebola and Marburg Virus RNAs in Infected Cells', *MBio*, 5: 6.

Simon, P. F. et al. (2016) 'Avian Influenza Viruses that Cause Highly Virulent Infections in Humans Exhibit Distinct Replicative Properties in Contrast to Human H1N1 Viruses', *Scientific Reports*, 6: 24154.

Simon-Loriere, E. et al. (2009) 'Molecular Mechanisms of Recombination Restriction in the Envelope Gene of the Human Immunodeficiency Virus', *PLoS Pathogens*, 5: e1000418.

—— et al. (2010) 'RNA Structures Facilitate Recombination-Mediated Gene Swapping in HIV-1', *Journal of Virology*, 84: 12675–82.

Soda, K. et al. (2011) 'H9N2 Influenza Virus Acquires Intravenous Pathogenicity on the Introduction of a Pair of Di-Basic Amino Acid Residues at the Cleavage Site of the Hemagglutinin and Consecutive Passages in Chickens', *Virology Journal*, 8: 64.

Steinhauer, D. A. (1999) 'Role of Hemagglutinin Cleavage for the Pathogenicity of Influenza Virus', *Virology*, 258: 1–20.

Suttie, A. et al. (2019) 'Inventory of Molecular Markers Affecting Biological Characteristics of Avian Influenza A Viruses', *Virus Genes*, 55: 739–68.

Swayne, D. E., and Pantin-Jackwood, M. (2006) 'Pathogenicity of Avian Influenza Viruses in Poultry', *Developments in Biologicals*, 124: 61–7.

To, T. H. et al. (2016) 'Fast Dating Using Least-Squares Criteria and Algorithms', *Systematic Biology*, 65: 82–97.

Vandegrift, K. J. et al. (2010) 'Ecology of Avian Influenza Viruses in a Changing World', *Annals of the New York Academy of Sciences*, 1195: 113–28.

Veits, J. et al. (2012) 'Avian Influenza Virus Hemagglutinins H2, H4, H8, and H14 Support a Highly Pathogenic Phenotype', *Proceedings of the National Academy of Sciences of the United States of America*, 109: 2579–84.

Von Rossum, G., and Drake, F. L. (2009) *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Wahlgren, J. (2011) 'Influenza A Viruses: An Ecology Review', *Infection Ecology & Epidemiology*, 1: 6004.

Wan, X.-F. et al. (2011) 'Indications that Live Poultry Markets are a Major Source of Human H5N1 Influenza Virus Infection in China', *Journal of Virology*, 85: 13432–8.

Wandzik, J. M. et al. (2020) 'A Structure-Based Model for the Complete Transcription Cycle of Influenza Polymerase', *Cell*, 181: 877–93.e21.

Wang, L. et al. (2008) 'Characterization of Influenza Virus Variants with Different Sizes of the Non-Structural (NS) Genes and Their Potential as a Live Influenza Vaccine in Poultry', *Vaccine*, 26: 3580–6.

Webster, R. G. et al. (1992) 'Evolution and Ecology of Influenza A Viruses', *Microbiological Reviews*, 56: 152–79.

Wickham, H. (2009) *Ggplot2*. New York: Springer.

Will, S. et al. (2007) 'Inferring Noncoding RNA Families and Classes by Means of Genome-Scale Structure-Based Clustering', *PLoS Computational Biology*, 3: 680–91.

Wong, S. S. et al. (2014) 'Characterization of an H4N2 Influenza Virus from Quails with a Multibasic Motif in the Hemagglutinin Cleavage Site', *Virology*, 468–470: 72–80.

Wu, Z. Q. et al. (2017) 'Comparative Epidemiology of Human Fatal Infections with Novel, High (H5N6 and H5N1) and Low (H7N9 and H9N2) Pathogenicity Avian Influenza a Viruses', *International Journal of Environmental Research and Public Health*, 14: 263.

Downloaded from https://academic.oup.com/ve/article/7/2/veab093/6415083 by Institut National Polytechnique Toulouse user on 06 February 2023

III. Génération de virus génétiquement modifiés pour l'étude de l'acquisition du MBCS et le phénomène d'émergence HPAIV

Dans une deuxième étude, nous avons développé un modèle expérimental pour étudier l'effet de l'environnement nucléotidique sur les évènements de mutation pouvant être responsable de l'acquisition d'un MBCS, et donc l'émergence. Comme discuté précédemment, l'émergence HPAIV apparaît comme un phénomène progressif, nécessitant des étapes successives avec l'apparition de formes intermédiaires pouvant évoluer rapidement vers l'acquisition d'un MBCS, et donc l'émergence (Seekings et al. 2020; García et al. 1996; Abolnik 2017). Ces formes intermédiaires correspondent aux génomes contenant des mutations transitoires, ayant un impact potentiellement négatif sur le fitness général du virus, mais pouvant servir de matrices pour l'acquisition de mutations supplémentaires, primordiales à l'émergence HPAIV. Les études cherchant à reproduire le phénomène d'émergence HPAIV en laboratoire souffrent d'un problème de sensibilité et/ou de biais de sélection (au niveau protéique) limitant l'étude de facteurs génétiques potentiellement favorables à l'émergence. Nous avons proposé l'élaboration d'un système de génétique inverse permettant la génération de virus génétiquement modifiés pour lesquels une séquence génomique HA est affranchie de la pression de sélection au niveau protéique. A partir de la souche A/turkey/Italy/977/1999 (H7N1), nous avons généré des virus génétiquement modifiés, possédant un transgène qui a été inséré à la fin de la partie codante du segment PA. Ce transgène correspond au segment HA d'un virus H5N3 (A/whistling swan/Shimane/499/83, nommé Shim) et d'autre virus de sous-type H5 décrits comme pro géniteurs d'évènements d'émergence connus. Nous avons fait varier la séquence nucléotidique environnant la région codant le site de clivage du transgène HA, à la fois en terme du motif nucléotidique et de la structure secondaire d'ARN (prédite). Plusieurs virus avec un transgène unique ont ainsi été générés et caractérisés. Des infections expérimentales, suivies du séquençage à haut débit des virions produits ont permis d'analyser le spectre de mutation au niveau de la région codant le site de clivage HA en fonction de caractéristiques inhérentes de la séquence génétique : motifs nucléotidiques et structures secondaires d'ARN.

Title:

Design of engineered viruses to study the influence of genomic properties on the furin-cleavage-site motif acquisition in avian influenza viruses

Gabriel Dupré[1]*, Bertille Pouget[1], Aldair Martín Martínez Pineda[1], Charlotte Foret-Lucas[1], Thomas Figueroa[1], Claire Hoede[2], Christine Gaspin[2], Roland Marquet[3] and Romain Volmer[1]*

[1] Université de Toulouse, ENVT, INRAE, UMR 1225, Toulouse, France.

[2] Plateforme GenoToul BioInfo UR875 MIAT, Castanet-Tolosan, France.

[3] Unistra, IBMC, CNRS, UPR 9002 ARN, Strasbourg, France.

* Corresponding authors:

Dr Romain Volmer, Gabriel Dupré

romain.volmer@envt.fr, gabriel.dupre@envt.fr

Keywords: Evolution, avian influenza virus, hemagglutinin, cleavage site motif, highly pathogenic avian influenza virus, emergence, genomic properties, base-sequence motifs, RNA structure, high-throughput-sequencing

Abstract:

Certain low pathogenic influenza viruses (LPAIV) of the H5 and H7 subtypes can evolve into highly pathogenic avian influenza viruses (HPAIV). This evolution is characterised by the acquisition of viral mutations that will give rise to a multi-basic cleavage site motif in the hemagglutinin (HA), leading to an extended viral tropism and severe disease in poultry. The genetic mechanism underlying this evolutionary process is poorly characterised, and some questions remain unanswered. Within the H5 and H7 subtypes, have all the LPAIV the same probability of evolving towards a highly pathogenic phenotype? If the probability of HPAIV emergence among LPAIV is significantly different, are there specific features on the HA genomic segment that can influence the evolution to HPAIV? In this work, we investigated the possible presence of viral genomic elements on the HA segment that can affect the viral mutation rates, and potentially HPAIV emergence. We developed a reverse-genetic system that generates influenza viruses harbouring the complete HA sequence as a transgene at the end of the PA encoding segment. In this model, the HA transgene cannot be translated into a protein and is released from selection pressure on protein function. We successfully rescued and characterised infectious viruses containing transgene with various HA cleavage site encoding motifs. We tested how different nucleotide environments could influence viral polymerase errors using this system and deep sequencing. We identified a minimal nucleotide motif consisting of a six consecutive adenine stretch (viral cRNA orientation) that could be present in transitory variants and that favours insertion events, a critical step in MBCS acquisition and HPAIV emergence. Moreover, recombination events with both ribosomal RNA and viral segments have been observed in our system.

Introduction:

Influenza viruses infect a vast number of different animal species, including humans. The aquatic birds, predominately of the orders *Anseriformes* and *Charadriiformes*, constitute their major natural reservoirs (Webster, Bean, Gorman, Chambers, & Kawaoka, 1992). These viruses are classified into subtypes, depending on the two major viral glycoproteins: hemagglutinin (HA) and neuraminidase (NA), they harbour at their surface. Several combinations of sixteen and nine different HA and NA proteins, have been described among avian influenza viruses, with a standing-out diversity observed in ducks (Vandegrift, Sokolow, Daszak, & Kilpatrick, 2010; Wahlgren, 2011; Webster et al., 1992). Avian influenza viruses can also be classified in two groups, depending of their pathogenicity in poultry: Low and Highly Pathogenic Avian Influenza Viruses (LPAIV and HPAIV, respectively) (Garten & Klenk, 1999; Horimoto, Nakayama, Smeekens, & Kawaoka, 1994; Horimoto et al., 1995; Steinhauer, 1999). The major pathogenicity determinant is due to HA cleavability. The HA cleavage site motif of LPAIV has non-consecutive dibasic amino acids and can be cleaved by trypsin-like proteases located in birds' lung and intestinal tract. At the opposite, HPAIV harbour a Multi-Basic Cleavage Site (MBCS) motif that can be processed by ubiquitous furin-like proteases present in all cells of the organism, responsible of systemic infection, multiple organ failure and finally the host death (Abdelwhab, Veits, & Mettenleiter, 2013; Lee, Criado, & Swayne, 2020; Swayne & Pantin-Jackwood, 2006). HPAIV do not only represent a major threat to the poultry industry worldwide: many human cases of H5 and H7 HPAIV infections have been described and were unfortunately fatal (Bui et al., 2016; Long, Mistry, Haslam, & Barclay, 2019; Simon et al., 2016; Wan et al., 2011; Wu et al., 2017).

HPAIV evolves from an LPAIV progenitor virus through genetic mutations that trigger the MBCS appearance. Nucleotide substitutions, insertions, and non-homologous recombination between the host (ribosomal RNA 28S) or viral genes (matrix and nucleoprotein segments) have been observed in nature to be responsible for MBCS acquisition (Abdelwhab et al., 2013; Gultyaev, Spronken, Funk, Fouchier, & Richard, 2021; Lee et al., 2020; Richard et al., 2017). LPAIV-to-HPAIV evolution has only been described for viruses belonging to the H5 and H7 subtypes (except for a H4-subtype virus that acquired a MBCS but did not exhibit a HPAIV phenotype) (Wong et al., 2014). This subtype restriction remains a mystery, as genetic manipulations of the HA cleavage site region using reverse genetics conferred higher pathogenicity in other subtypes (H2, H4, H6, H8, H9, and H14), demonstrating that these HA

subtypes do not harbour structural incompatibilities with artificial MBCS insertion (Munster et al., 2010; Soda, Asakura, Okamatsu, Sakoda, & Kida, 2011; Stech et al., 2009; Veits et al., 2012). One recent study showed that the region encoding for the HA cleavage site of H5 and H7 avian influenza viruses contains a high number of adenine and/or purines, which could explain their natural propensity to acquire MBCS by additional mutations (Funk, de Bruin, Spronken, Gultyaev, & Richard, 2022). Moreover, the ability of H5 or H7 LPAIV to evolve into an HPAIV has been linked to the presence of a predicted secondary RNA stem loop that encompasses the HA cleavage site coding sequence (Beerens et al., 2020; Dietze et al., 2018; Garcia, Crawford, Latimer, Rivera-Cruz, & Perdue, 1996; Gultyaev et al., 2016; Nao et al., 2017; Perdue, García, Senne, & Fraire, 1997). The apical loop of this secondary RNA structure composed of 6 to 12 unpaired bases has been proposed to promote nucleotide insertions through polymerase slippage (Nao et al., 2017). However, a large scale bioinformatics analysis has revealed that this predicted RNA stem loop is widely conserved among avian influenza viruses, even in the HA of influenza virus subtypes that have never been shown to evolve to HPAIV (Gultyaev, Richard, Spronken, Olsthoorn, & Fouchier, 2019). No other nucleotide or protein motifs elsewhere on the viral genome have yet been associated with the pioneering step of HPAIV emergence (Escalera-Zamudio et al., 2020). Thus, it is still unknown if genetic signatures present on the HA segment can favour HPAIV emergence. Additionally, it is still undetermined if all the H5-H7 LPAIV have the same probability of evolving towards HPAIV or whether specific genetic determinants in the HA sequence correlate with a higher risk of HPAIV transition for some H5-H7 LPAIV.

To better understand the evolution of avian influenza viruses to highly pathogenic forms and investigate the possible presence of possible favouring genome properties on the HA segment, we generated H7N1-engineered viruses harbouring one supplementary HA sequence on their PA segment (HA transgene) that can freely evolve and accumulate any kind of mutations during replication cycle. Several H7N1-engineered viruses were produced and characterised, harbouring various HA sequences in terms of base-sequence and predicted RNA structure. Experimental infection on different cell-lines combined with deep-sequencing analysis permitted to identify specific genomic elements that favour mutation events, potentially responsible of MBCS acquisition, and so, emergence of HPAIV.

Materials and methods:

I. Cells

MDCK, HEK-293T and DF-1 (ATCC® CRL1590) cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% calf serum, 100 U/ml penicillin, and 0.1 mg/ml streptomycin. All cells were incubated at 37°C in a 5% CO2 incubator.

II. Virus productions and characterizations

1. Plasmids

Each cDNA of the height viral genes from A/turkey/Italy/977/1999 (H7N1 subtype, "977"), low pathogenic version of A/mallard duck/France/171201g/2017 (H5N8 LP) (Bessière et al., 2022), the HA encoding segment of A/whistling swan/Shimane/499/1983 (H5N3 subtype, "Shim"), HA segment of A/whistling swan/Shimane/499/1983 mutant 24a2b ("Shim24a2b") were cloned in pHW2000 vectors. RNAfold program from the ViennaRNA-2.4.3 package was used to assess the predicted RNA structure of each HA sequence (Lorenz et al., 2011).

2. Cloning and design of 977.PA-HA viruses

To produce 977PA-HA derived mutants: 977 virus with its fused PA segment-HA sequence, different HA sequences (full segment) were cloned right after the PA stop codon in a pHW2000 vector. More precisely, a DNA construction containing the last 60 nucleotides of PA cDNA followed by a linker cloning site (LCS) harbouring two BsmbI restriction sites in a pUC57 plasmid was synthetized and cloned in a pUC57 plasmid (synthetized and clone by Genscript®, pUC57-PA-LCS). Reverse-genetic pHW2000-PA(977) and pUC57-PA-LCS plasmids were double-digested by PvuII. The linearized plasmid (digested pHW2000-PA(977)) and the produced insert (PA-LCS fragment) were agarose gel-purified, and ligated by a T4 DNA ligase. Competent cells (Stellar™ Competent Cells from In-Fusion® HD Cloning kit from Takara®) were transformed. After selecting the correct clone with the desired genetic construction, the corresponding bacteria were amplified into selective LB-agar petri-dishes containing 50 µg/mL

135 Ampicillin and subjected to plasmid-purification. It is important to note that the 60-nucleotides-
136 end of PA-ORF was duplicated and flanks the LCS region. This specific sequence repetition
137 was realized to conserved the contiguous packaging signals located both at the CDS and NCR
138 at the PA segment 5' end (Tran, Moser, Poole, & Mehle, 2013). HA inserts from different
139 strains were cloned in pHW2000 (Shim, Shim24a2b, ShimA1-to-A4, HA sequences from
140 A/turkey/Ontario/6213/1966, A/ostrich/South Africa/AI1160/2006, A/chicken/Mexico/31381-
141 7/1994 and their relative "6-A" mutants). Insert from each HA sequence consisting of the full
142 HA segment were created by PCR amplification using primers that bind to the NCRs
143 extremities (full-length fragments) with a high-fidelity DNA polymerase (Phusion High-
144 Fidelity PCR Kit from Thermo Scientific). Primers carry on their 5' ends BsmbI restriction sites
145 which enable us to clone produced HA inserts into the phW2000-PA-LCS vector after BsmbI
146 digestion of both amplicons and vector. Digested and gel-purified HA-inserts and phW2000-
147 PA-LCS were ligated with T4 DNA ligase, and competent cells were transformed by the
148 ligation product. Selected clone with the correct construction without mutations was checked
149 by sequencing, amplified et subjected to plasmid purification.

150 As it mentioned in a previous work, repeats of the PA end may induce genetic instability via
151 homologous recombination or possible double crossing-over in the plasmid and at the
152 eukaryotic level. To get rid of this possible source of genetic instability, 18 silent mutations
153 were introduced at the PA-ORF end (PA-SWAP) via small insert exchange using the In-
154 Fusion® HD Cloning kit from Takara® kit. Complementary oligonucleotides corresponding to
155 the PA-SWAP were annealed and used as template for PCR amplification with primer
156 harbouring 5'extremities homologous to the pHW2000-PA_HA plasmids. After linearization
157 of the each pHW2000-PA_HA via reverse-PCR, the receiving vector and amplified insert were
158 subjected to ligation. Finally, Stellar® competent cells were transformed by the ligation
159 products and clones were amplified and sanger sequenced to check the presence of the
160 following construction at the end of the PA-ORF: "PAswap-HA segment".

161 Various HA-ShimH5 mutants (ShimH5mutA1-to-A4) were design in order to harbour different
162 predicted RNA structure and base-sequence environment at the HA1/HA2 boundary region.
163 RNAfold program from the ViennaRNA-2.4.3 package was used to assess the predicted RNA
164 structure of each HA sequence (Lorenz et al., 2011).

165

166 3. Virus production

Wild-type A/turkey/Italy/977/1999 (H7N1, 977), H5N8-reassortant with the HA segment from A/whistling swan/Shimane/499/83 (Shim) or HA-Shim24a2b (Shim24a2b), DB H5N8 viruses and mutant 977.PA-HA viruses were produced by reverse genetics. For this purpose, viral segments of each virus are cloned into a phW2000 vector with bidirectional promoters for the production of both mRNA and vRNA (Erich Hoffmann, Neumann, Kawaoka, Hobom, & Webster, 2000). Co-cultures of MDCK and HEK-293T or DF-1 and HEK-293T cells were transfected by 500 ng of each of the eight-phW2000 plasmids and maintained in Viral Growth Medium (VGM): Opti-MEM I Reduced Serum Medium (Opti-MEM) supplemented with 100 U/ml penicillin. L-1-tosylamido-2-phenylethyl chloromethyl ketone (TPCK)-treated trypsin was added 20 hours post-transfection to 1 µg/mL final concentration. Seventy-two-hours post-transfection, supernatants containing rescued virions were collected and clarified to remove cell debris. Virus stocks were created by amplifying viruses obtained post co-transfection on MDCK (stocks obtained from MDCK/293T cocultures) or on DF-1 cells (stocks obtained from cocultures of DF-1/HEK-293T cells), maintained in VGM for three days. The fifty-percent tissue culture infective dose ($TCID_{50}$) of virus stocks were determined on the same cell type using the Spearman-Karber calculation-method.

4. PA.977-HA virus characterization

Viral genome extraction and sequencing

First stock of each virus was subjected to vRNA extraction using the NucleoSpin® RNA Virus from Macherey-Nagel®. cDNA synthesis of each viral segment was realized using the Universal primer Uni12 (E. Hoffmann, Stech, Guan, Webster, & Perez, 2001) and the RevertAid reverse transcriptase kit (Thermo Scientist ®). The PA (CDS) – HA transgene segment of each PA-HA virus was PCR-amplified using PA segment specific primers Uni12 and Uni13, extracted on 1% agarose-gel and Sanger-sequenced by Eurofins® (see list of primers S#).

Growth kinetics

To compare viral growth kinetics between the wild-type and the engineered-977PA-HA viruses, MDCK cells were sub cultivated in 6-wells plates and infected with wild-type 977 (977), 977PA-HAShim or 977PA-HAShim24a2b viruses, at moi of 0.0001 $TCID_{50}$. Infections

were carried out to 72 hours. Part of the infection supernatant was collected at different time points and subjected to viral titrations ($TCID_{50}$).

HA5 protein expression of the H7N1(977)-HA viruses:

HA5 subtype protein expression was assessed by western-blot. MDCK were sub-cultivated in 6-well-plates and infected at a 0.1 MOI for 24 hours by Shim, Shim24a2b, or 977PA-HA viruses from passage #1. Cells were collected and lysed in RIPA buffer (50 mM Tris-HCl, pH 8.0, with 150 mM sodium chloride, 1.0% Igepal CA-630 (NP-40), 0.5% sodium deoxycholate, and 0.1% sodium dodecyl sulfate) with antiproteases-cocktail from (cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail from Roche). Dosage of total proteins in cell lysates was realized by Bradford assays (sigma, B6916). Nuclear and cytoplasmic proteins were analysed in 12% SDS-PAGE gels (Mini-PROTEAN® TGX Stain-Free™ from Biorad). After transfer, nitrocellulose membranes were blocked in 5% milk TBS-T buffer for 1h30 and blotted with primary antibodies directed to HA5 (BEI NR-163, 1:1000 in 3% milk-TBS-T), NP (a kind gift of Dr Gary R Whittaker, Cornell University, Ithaca, USA) and NS1 of 977 (kindly provided by Daniel Marc, INRAE Tours France) proteins. Secondary antibodies (anti-goat IgG for HA membranes (sigma, A-5420), sigma anti-rabbit IgG for NS1 and NP membranes (sigma, 0545)) coupled with Horseradish Peroxidase (HRP) were used with the compatible first antibody (1:4000 in 5% milk TBS-T). Biorad® Chemidoc was used to read the membranes.

HA-transgene stability:

A fraction of collected virus stock at each cell-passage were analysed for the HA-transgene stability (located at the PA-segment 5'end) by quantitative RT-PCR (qRT-PCR). Viral RNA extraction was performed by using the NucleoSpin® RNA Virus (Macherey-Nagel®) from 150 µL of supernatant. Specific cDNA amplification form PA vRNA were performed using primer targeting the 3'NCR of the PA segment: Bm-PA-1 (E. Hoffmann et al., 2001). Reverse-transcription reactions were performed using RevertAid reverse transcriptase kit (Thermo Scientist ®). qPCR reactions were performed by specifically amplifying specific regions on the PA CDS sequence and the HA transgene using the QuantiFast SYBR® Green PCR Kit. Serial dilutions of the phW2000-PA.977-HA constructions were used as standards for absolute quantification of vRNA copies. To analyse HA-transgene stability, the ratio HA-transgene vRNA copies/PA vRNA copies were calculated.

III. HA segment evolution

1. Infections with Shim recombinant viruses and PA-HA viruses

Cells were sub-cultivated into 25 cm^2 flasks and infected with Shim, Shim24a2b and PA977-HA mutant viruses at a MOI of 0.001 $TCID_{50}$, using the following viral growth media (VGM): Opti-MEM I Reduced Serum Medium (Opti-MEM) supplemented with 100 U/ml penicillin, and 0.1 mg/ml streptomycin, in presence of 1 µg/mL TPCK-treated trypsin. Infections were carried out between 48 to 72 hours, depending on the percentage of remaining cell-monolayer. Until only 30% of adherent cells remain, supernatant was collected and clarified. Virus stocks were tittered on MDCK by $TCID_{50}$ before each passage.

2. Illumina sequencing of the HA cleavage site encoding region and data processing

Mutability of HA cleavage site encoding region between viruses were analysed by high throughput sequencing using similar methods detailed in this previous work (Doud & Bloom, 2016) with few modifications.

Library preparation:

vRNA extractions were performed using the NucleoSpin® RNA Virus from Macherey-Nagel®. For each sample; $5x10^6$ molecules of HA or PA-HA vRNA were subjected to cDNA synthesis using the AccuScript High Fidelity Reverse Transcriptase (Agilent®) and primer targeting the PA segment harbouring the HA transgene (977PA-HA viruses) or the HA segment (wild-type viruses) (see list of primers#). Experiments were performed in triplicates and are independent from the viral rescue.

Then, a three-step amplification were realized to produced our viral DNA libraries, named as "PCR0", "PCR1" and "PCR2". First, viral cDNA of each sample was pre-amplified by PCR ("PCR0") using the High-Fidelity DNA polymerase form Master Mix pour PCR Platinum™ SuperFi II (ThermoFisher®) (see list of primers S#). For the PCR reaction, 5 µL of cDNA sample (corresponding to a theoretical number of $1.25x10^6$ of ssDNA molecule if the reverse-transcription were fully efficient) was added to a solution containing 2.5 µL of each primer (10

μM, see list of primer S#), 25 μL of PCR master mix and 15 μL of nuclease-free water. The following cycling conditions was used: 1. 98 °C for 30 s; 2. 98 °C for 0:10 s; 3. 56 °C for 00:10 s; 4. 72 °C for 0:10 s; 5. Go to step 2. 21 times; 6. 72 °C for 5 min; and 4 °C hold. A fraction of PCR product were analysed on a 1.5% agarose gel to check the amplicons size and purity. The remaining of PCR products were bead-purified using the AMPure XP beads (Beckman Coulter®) at a ratio of 1:1 in volume of sample:beads. PCR0 amplicons were quantified with Quant-iT™ PicoGreen™ dsDNA Assay Kit (ThermoFisher®). Then, PCR0 amplicons were used as template for PCR1 that will bring Unique Molecular Identifiers (UMIs) or barcodes and partial Illumina adaptors. 8 μL of PCR0 amplicons at 0.5 ng/μL were PCR amplified with 1 μL of 10 μM forward and reverse primers that contain the complementary sequence to the template, height random nucleotides (constituting the random barecode), and the partial Illumina adapter's sequence in their 5'tails (see list of primer S#). Different cycling conditions were used: 1. 98 °C for 30 s; 2. 98 °C for 0:10 s; 3. 56 °C for 00:10 s; 4. 72 °C for 0:10 s; 5. Go to step 2. 8 times; 6. 72 °C for 5 min; 7. 98 °C for 1 min and 4 °C hold. PCR1 products were also bead-purified (see above), quantified with Quant-iT™ PicoGreen™ and diluted to 70 pg/μL. For the PCR2 that aimed to add the remaining part of the Illumina's adaptors and indexes for sample multiplexing, template DNA (PCR1 product) was further diluted to reach precisely $7.2x10^5$ ssDNA molecules/μL. In doing so, the number of template's molecules used was less than the anticipated sequencing depth to ensure the creation of barcode-families or multiple reads per unique barcode. The PCR mix contained 1 μL of PCR1 template (at $7x10^5$ ng/μL, corresponding to $7.2x10^5$ ssDNA molecules), 2 μL of Universal Illumina forward primer and reverse primer with 6-base-long index (see list of primers S#). The following cycling was used: 1. 98 °C for 30 s; 2. 98 °C for 0:10 s; 3. 56 °C for 00:10 s; 4. 72 °C for 0:10 s; 5. Go to step 2. 30 times; 6. 72 °C for 5 min; and 4 °C hold. Final PCR products were bead-purified, analysed on a 1% agarose and quantified (see above). DNA libraires of 242 and 244 bp (depending of the sample, see S#) were obtained. Each sample were mixed together in stoichiometric manner and sent to Genewiz® for Illumina sequencing in a HiSeq 2500 using 2×250 pb configuration.

Bioinformatics and mutability analysis:

Raw fastq reads were de-multiplexed and controlled for their quality using the Fastqc software. After preparing for each DNA libraries the corresponding reference genome with BWA and Samtools sotwares, we used part of the code from the "Duplex-Sequencing" pipeline (Kennedy et al., 2014) (github link: https://github.com/Kennedy-Lab-UW/Duplex-Sequencing) that have

been slightly modified for our datasets (see the code in S#). We grouped all reads that contain the same UMI (barcode-family) into one "tag family" and calculated the consensus sequence. Then adapter-trimming of consensus read1 and read2 sequences were performed independently and alignments of each trimmed consensus reads to the reference genome were realized. Mpileup files were generated, and mutation counting and localisation were finally realized. In the case of insertion events, the insert sequence and its frequency were also analysed. Mutation spectra of each virus was analysed. Position were significantly higher mutation rates than the mutation average (fixed at 10^{-4} (Pauly, Procario, & Lauring, 2017)), were identified via a binomial test with Bonferroni correction. Mutation frequencies per nucleotide and mean mutation frequencies at the 1015-1026 nucleotide-window of the HA sequence (cRNA numbering, starting at the first base of the start codon) were also determined.

3. Minigenome system and luciferase assay

Base-sequence characteristics (nucleotide motifs, repetition and/or RNA structure) can possibly influence error rates from viral polymerases, cellular polymerases and reverse transcriptase during Illumina library preparation (Chung et al., 2023; Geller et al., 2015; Gout et al., 2017). To evaluate the proportion of insertion events from reverse transcriptase errors, we used a reporter assay consisting of a minireplicon system inserted in pGL4.10 expression vector. The minigenome construction is composed of HA cleavage site sequence from Shim or Shim24a2b (see Figure **S2A**) that begins with a start codon. This inserted sequence or linker varies in length: 28, 29 or 30 nucleotide-long. The coding sequence of the Nanoluciferase (Nluc) is placed directly after the linker, without its start codon. The following construction: [atg – linker (HA Shim or Shim24a2b)$_{28,\ 29\ \text{or}\ 30\text{nt}}$ – Nluc] is flanked by NCR from the HA segment of A/chicken/Germany-MV/R10048/2016 (H5N8). The whole minireplicon has been placed between the pol I terminator (ribozyme motif from Hepatitis delta virus (HDV)) and the chicken polymerase I promoter to produce the pseudo-viral RNA of [NCR – linker – Nluc – NCR] that will be both replicated and transcribed by the viral polymerase, added *in trans* in DF-1 chicken cells via pcDNA3.1 expression vectors encoding the polymerase subunits of A/WSN/1933 (WSN). The Firefly luciferase placed after the mouse PGK promoter also present in pGL4.10 was used as an internal control to normalise the Nluc signal.

DF1 cells grown in 24-well plates ($1x10^6$ cells per well) were co-transfected with the minigenome system (100 ng) and plasmids encoding the subunits of the viral polymerase

complex (500 ng each), according to the protocol of the Lipofectamine® LTX & PLUS™ Reagent kit (InvitrogenTM). Twenty-four hours post-transfection, the cells were lysed using Passive Lysis Buffer (Promega®). After 15 min of lysis, 40 µL of each sample was plated onto a 96-well plate (Greiner bio-one). Readings of Nluc and Fluc luciferase enzymatic activities were performed respectively with the Nano-Glo and One-Glo kits from Promega® on a Clariostar instrument (BMG Labtech; optical parameters: 470-480 nm filter for Nluc readings, no filter for Fluc readings, acquisition time: 10 s). The Nluc signal was normalised by the Fluc signal and the Nluc/Fluc ratios were calculated to give the absolute normalised signal values for each well.

Results:

Engineered PA-HA viruses to bypass selection and to study the base-sequence influence on viral mutation events (characterisation).

To strictly analyse the influence of the genome properties on viral mutation rates in the context of infection, we designed a new system consisting of engineered H7N1 viruses with modified PA segment. We placed different HA segments from various H5-subtype avian influenza A viruses, at the end of the PA-CDS sequence (**Figure 1**). This supplementary HA sequence referred as the HA transgene, was placed out of the PA open reading frame (ORF) after supplementary stop codons to avoid any possible HA transgene translation. To preserve correct packaging of this modified PA segment in the new particles, the last sixty nucleotides of the PA CDS sequence were duplicated and placed directly after the HA transgene, and before the PA 5'NCR sequence. Theoretically, the HA transgene could accumulate any mutations and freely evolve during the replication cycles, and possibly accumulate mutations, as it is not involved in any protein production nor function.

A previous study showed that two PR8-reassortants harbouring the HA segment of Shim (H5N3 virus) and its mutant version Shim24a2b exhibit very different insertion frequencies at the HA1/HA2 boundary region. PR8-reassortant with HA of Shim24a2b accumulates one thousand times more insertion events that its HA-wild-type (Shim) version (Nao et al., 2017). To validate the capacity of our 977PA-HA system to detect such mutation events and different mutation spectra between viruses, we generated 977PA-HA viruses with HA transgenes derived from Shim and Shim24a2b, performed series of basic characterisation and compare the mutation

spectra with H5N8 (LP) HA-reassortants harbouring the same HA segment of Shim or Shim24a2b (Shim, Shim24a2b).

After generating 977PA-HAShim and 977PA-HAShim24a2b viruses, we first checked that genomic engineering did not enable HA transgene translation. We infected MDCK cells with the produced viruses at high multiplicity of infection (MOI, 0.1 TCID50) during 24 hours, collected infected cells, and checked the Shim-HA protein expression by western-blot (**Figure 2A**). The obtained blots showed clear expression of HA5 protein in the positive controls: H5N8 (LP) –reassortant virus harbouring the Shim-HA segment and H5N8 (LP). However, HA5 protein was not detected in experimental conditions with 977PA-HA viruses, confirming the absence of HA transgene protein expression in the 977PA-HA system.

After one passage on MDCK or DF-1 cells (after reverse genetics), we checked that the HA transgene of 977PA-HAShim and 977PA-HAShim24a2b was genetically stable in the new progenies. We designed a specific RT-qPCR targeting the PA segment. A couple of primers that bind inside the PA CDS sequence and at the end of the HA transgene was used to amplify the end of the modified PA segment. After migration of PA-HA amplicons on agarose gels, we were able to detect a major band at around 2 kb (**Figure 2B**), which corresponds to the full PA-HA genetic construction (2097 bp-long amplicon expected). Other amplicons of smaller sizes were also detected with a lowest signal on gel, showing minor genetic instability and the presence of PA segment with possible deletion on the HA transgene. One amplicon with a lower size on gel was detected for 977, which correspond of the amplification of the end of the PA segment that did not contain any transgene. We quantified PA segment harbouring the HA transgene with intact HA1/HA2 boundary region by qPCR and primer flanking this region (**Figure 2C**). Between 30 to 55% of the new progenies possessed the HA transgene with intact HA1/HA2 boundary region. 977PA-HAShim virus seemed to exhibit better HA transgene stability, with a mean HA transgene / PA segment vRNA copies at 1.2. No significant difference was observed between viruses from different infected cell lines.

We also investigated the replication of 977PA-HA viruses in cell culture. We infected MDCK cells at low MOI (0.0001 TCID50) with 977PA-HAShim, 977PA-HAShim24a2b, 977, and performed viral growth kinetics (**Figure 2D-E**). In comparison to the original H7N1 (977), 977PA-HA viruses replicated and produced similar amount of infectious particles until reaching a plateau at 48 hours post infection (**Figure 2D**). One log difference was observed at 72 hours post infection in comparison to 977. The PA vRNA production was also monitored. Different rates of genome replication were observed between 977PA-HA viruses and the original 977

388 virus. The 977 parent virus replicated more efficiently than 977PA-HA viruses, with ten times
389 more vRNA copies at 24h post infection. The difference in vRNA copies increased at 48 h (2-
390 log difference) and remained stable until 72 h post infection (**Figure 2D**). No difference in
391 terms of infectious particles production and PA segment replication was observed between
392 977PA-HAShim and 977PA-HAShim24a2b. The difference observed in terms of infectious
393 particles production and PA segment replication could be due to the HA transgene, which can
394 slightly affect replication efficiency and/or other step of the viral cycle. Yet, 977PA-HA viruses
395 efficiently replicated in cell culture, with the production of sufficient genetic material (approx.
396 $1x10^6$ copies of PA vRNA / µL) for mutations detection.

397

398 Comparison of mutation events using the PA-HA system or wild-type viruses.

399 First, we analysed the mutation spectra of the HA1/HA2 boundary region of the HA segment
400 of Shim, Shim4a2b, and the HA transgene from 977PA-HAShim or HAShim24a2b originated
401 from MDCK infected cells.

402 Shim and Shim24a2b exhibited very different insertion frequencies (**Figure 3A**). In the
403 sequence window between bases 1015 and 1026 (on the cRNA HA ORF numbering),
404 Shim24a2b exhibit mean insertion frequency of $4.92x10^{-3}$, approximatively three thousand
405 times higher than the mean insertion frequency observed with Shim ($2.08x10^{-6}$) (**Figure 3E**).
406 This first observations confirmed previous results obtained by Nao et al. with PR8 – HAShim
407 reassortants' viruses (Nao et al., 2017). We observed similar insertions frequencies with
408 977PA-HA viruses (**Figure 3A**). 977PA-HAShim24a2b exhibited mean insertion frequency of
409 $1.80x10^{-2}$, which represents almost ten thousand times more events observed in comparison
410 with 977PA-HAShim ($2.25x10^{-6}$) (**Figure 3E**). To distinguish viral mutations events from
411 possible errors that could possibly occur during HTS library preparation, mutation spectra from
412 pseudo-viral RNA of HA-Shim and HA-Shim24a2b produced by the cellular polymerase I (pol
413 I) in transfected 293T cells were analysed. Position #1016 was found to significantly exhibit
414 higher insertion, deletion and substitution frequencies (**S1A-B**). However, the observed
415 frequencies dependent to pol I and the reverse –transcriptase (used for HTS library preparation)
416 remained 100x lower than what were observed with Shim, Shim24a2b, 977PA-HAShim and
417 977PA-HAShim24a2b (**Figure 3**), which indicates that most of the observed mutation came
418 from the viral polymerase activity. Moreover, to better characterise the origin of the observed
419 insertions, we used a reporter assay based on the expression of nanoluciferase (Nluc) thanks to

insertion of one or two nucleotides in a minireplicon system (**S2A**) We observed a significant increase in Nluc activity with the HA sequence from Shim24a2b (**S2B**) and no augmentation of Nluc activity with the HA sequence from Shim in comparison to plasmid control with no HA sequence. This results show that most of the observed insertions are not due to reverse-transcriptase errors. Deletions events were observed at the same nucleotide position with similar frequencies to previously observed insertions (**Figure 3B**). In terms of substitution events, Shim24a2b tended to experience slight more mutations (5x times more, $7.25x10^{-4}$) than Shim (**Figure 3C, E**). Similar observation was found for 977PA-HA viruses. 977PA-HAShim24a2b exhibited more substitutions on its transgene than 977PA-HAShim. Interestingly, most of the previously observed substitutions in Shim24a2b and 977PA-HAShim24a2b were substitutions giving rise to adenines, with observed mean frequencies of $4.34x10^{-4}$ and $1.40x10^{-3}$ respectively (**Figure 3D-E**).

Certain viral polymerase complexes could be possibly more prone to errors and influence mutation frequencies. To test this hypothesis, we compared the mutation spectra from viral polymerase complex of A/WSN/33 (WSN) and 977 strains in a minigenome system (**S3**). Similar mutation spectra were observed in HA-Shim using the viral polymerase complex from 977 (minig977-HAShim) or WSN (minigWSN-HAShim). Same observations were made with HA-Shim24a2b (minig977-HAShim an minigWSN-HAShim). However, for the same HA sequence, around 10x higher frequencies were observed with the used of WSN polymerase, in comparison to 977 polymerase. This difference could possibly be explained by the higher processivity of WSN polymerase.

Moreover, we noticed that more nucleotide positions on the HA1/HA2 boundary region stood out with higher *indels* and substitution frequencies in 977PA-HAShim and 977PA-HAShim24a2b in comparison to Shim and Shim24a2b, respectively (**Figures 3A-D, 3F**). Combining the two experimental replicates, a total of 36 nucleotide positions exhibited significantly higher substitution frequencies in the 977PA-HAShim. However, Shim only exhibited 20 nucleotide positions with significantly higher substitutions frequencies. One nucleotide position significantly exhibited more insertion events in the case of 977PA-HAShim. At the opposite, no nucleotide position in HA sequence of Shim was involved in insertion events (**Figure 3F**). Similar results were observed between Shim24a2b and 977PA-HAShim24a2b. More nucleotide positions were involved in insertions and deletions events in the case of 977PA-HAShim24a2b (17 and 13 respectively for 977PA-HAShim24a2b, in comparison to 13 and 10 for Shim24a2b) (**Figure 3F**). Moreover, 977PA-HAShim24a2b exhibited a higher

453 frameshifting-insertion frequency than Shim24a2b (with means at 92% and 82% of observed
454 insertions in 977PA-HAShim24a2b and Shim24a2b, respectively) (**Figure 3G**). These data
455 demonstrate that 977PA-HA system can accumulate more mutations than H5N8 - HAShim
456 reassortants' viruses, thus, permitting to observe greater events' diversity. More detailed
457 information about the inserted sequences per virus and replicate are provided in the
458 supplementary data section (**S4-5**).

459 977PA-HA viruses can be used to efficiently detect viral mutation events (insertions, deletions,
460 substitutions) influenced by any kind of base-sequence environments, in the context of infection
461 and independently of the protein selection pressure. The data presented above demonstrate that
462 only intrinsic characteristics of the HA-Shim24a2b sequence are responsible for the observed
463 insertions and substitutions responsible to adenine enrichment in the HA1/HA2 boundary
464 region. As the 977-PA-HA system permit to infer conclusions between the sole base-sequence
465 environment and mutation event, we further used 977-PA-HA viruses for this study.

466

467 The host environment does not impact mutation frequencies.

468 The host-cell environment can vary between species and possibly affect viral mutation rates.
469 To take into account any possible influence of specific host-cell factors on mutation events we
470 conducted our experiments on both MDCK and DF-1 cells with viruses that were produced
471 from MDCK/HEK-293T and DF-1/HEK-293T reverse genetics, respectively.

472 Mutation spectra between 977PA-HAShim and 977PA-HAShim24a2b were virtually similar
473 and no significant difference were observed between the tested cell-lines (**Figure 4**). As it was
474 observed previously, similar *indels* and substitutions frequencies were observed in the 1015-
475 1026 nucleotide-window (cRNA, HA-ORF).

476 The host-cell environment does not have influence on the observed mutation events, at least
477 between the two tested cell-lines.

478

479 The base-sequence environment at the HA1/HA2 boundary region influence viral mutation
480 events.

481 To better understand how genome properties can influence viral mutation rates, we generated
482 four supplementary 977PA-HA viruses with Shim-derived HA transgene containing various
483 substitutions at the HA1/HA2 region: 977PA-HAShimA1-to-A4 (**Figure 5**), which constituted

all the possible combinations of adenine-(A) content at the HA1/HA2 transgene region (in the 1015-1026 nucleotide-window, HA-ORF, cRNA). By analysing these variants, potential nucleotide motifs or preliminary substitutions might be identified as potential milestones for further mutation events and MBCS acquisition. Moreover, all the six 977PA-HAShim viruses could be classified in two groups depending on the HA1/HA2 RNA structure prediction (performed on the viral cRNA). 977PA-HAShim24a2b, ShimA1 and ShimA3 shared the same predicted RNA structure. Similarly, 977PA-HAShim, ShimA2 and ShimA4 exhibited a common, but different RNA structure. Thus, using 977PA-HAShim viruses could permit to identify which RNA structure and/or the base-sequence (primary sequence) could influence mutation events. MDCK cells and DF-1 cells were infected at a MOI of 0.001 TCID50 with the generated 977PA-HAShim viruses obtained from transfected MDCK/293T co-cultures or DF-1/293T co-cultures respectively. Infections were stopped 72h after virus inoculation, and supernatants from infected culture were collected for viral RNA extractions. HA transgene stability of the supplementary 977PA-HAShim viruses was checked before designing the Illumina libraries (**S6A-B**).

Very different insertion spectra were found between viruses. In the 1015-1026 nucleotide-window, 977PA-HAShim24a2b, 977PA-HAShimA3 and 977PA-HAShimA4 significantly exhibited higher insertion frequencies, with means of $1.21x10^{-2}$, $9.27x10^{-4}$ and $1.08x10^{-3}$ respectively, than 977PA-HAShim, 977PA-HAShimA1 and 977PA-HAShimA2 which exhibited low insertion frequencies, ranging from $1.19x10^{-5}$ to $2.31x10^{-6}$ (**Figure 6A**). Interestingly, similar pattern was observed in terms of deletion spectra (**Figure 6B**). High deletion frequencies were observed with 977PA-HAShim24a2b, 977PA-HAShimA3 and 977PA-HAShimA4 (mean frequency of $3.67x10^{-3}$, $3.15x10^{-4}$ and $4.64x10^{-5}$ respectively), which showed a relative high genetic plasticity conferred by the Shim24a2b, HAShimA3 and HAShimA4 sequences. Moreover, high substitution frequency, that mostly gave rise to new adenines (**Figure 6C-D**) was observed with 977PA-HAShim24a2b ($1.28x10^{-3}$) and 977PA-HAShimA3 (at position #1009 specifically). It is important to note that viruses sharing the same mutation spectra, as 977PA-HAShim24a2b, 977PA-HAShimA3 and 977PA-HAShimA4, did not have similar predicted RNA structure at the HA1/HA2 boundary region. Similar observation was made with 977PA-HAShim, 977PA-HAShimA1 and 977PA-HAShimA2.

Thus, we can conclude that the base-sequence environment (primary sequence) does have a significant influence on mutation events. No predicted-RNA structure has been found to be associated to mutation events. Interestingly, 977PA-HAShim24a2b, 977PA-HAShimA3 and

977PA-HAShimA4 that significantly experience more insertion events than the other viruses possess at least 6-consecutives adenines stretch at the observed insertion hotspots. These 6-long adenines-stretch (or longer) is absent in the HA sequence of the other viruses. We thus inferred that 6-consecutive adenine-stretch (6A) could constitute the minimal motif required for insertion accumulation at the HA1/HA2 boundary region.

A 6A motif and its role in step-wise evolution to MBCS acquisition.

As mentioned before, LPAIV-HPAIV evolution appears to be a step-wise process, involving the rise of transitory variants. The 6A motif is underrepresented in the current sequence database, suggesting its transient presence. To test the hypothesis that the apparition of a 6A motif is the major event required in to trigger insertion events and the potential apparition of a MBCS, supplementary 977PA-HA viruses with different HA sequences were then produced. The HA sequences of three documented H5-HPAIV progenitors: A/turkey/Ontario/6213/1966, A/ostrich/South Africa/AI1160/2006, A/chicken/Mexico/31381-7/1994 (de Bruin et al., 2022) were used as HA transgenes. For each transgene, a mutant version containing the 6A motif at the HA1/HA2 boundary region was generated through directed mutagenesis (**Figure 7A**). Every 977PA-HA virus containing the HA sequence harbouring the 6A motif exhibited high insertion frequency (**Figure 7B**), as it was previously observed for 977PA-HAShim24a2b, 977PA-HAShimA3 and 977PA-HAShimA4. 977PA-HA viruses harbouring the wild-type HA sequence as transgene did not exhibit significantly higher insertion frequency to the insertion background. This results show that the 6A motif is important to trigger insertion accumulation and this effect is independent to the HA sequence context outside the HA1/HA2 boundary region.

Detection of recombination events.

Using our experimental model, we were able to detect various insertions types with the engineered 977PA-HA and wild-type viruses. Interestingly, large insertions between 23 to 43 nucleotides long were also identified in some 977PA-HA viruses, at the HA1/HA2 boundary regions (**Figure 8**). Six independent large insertions were detected, corresponding to recombination events with other viral segments and ribosomal RNA (rRNA) after BLAST-nucleotide analysis on the observed inserts. 977PA-HAShim24a2b and 977PA-HAShimA3 recombined with different part of the HA(7) segment. 977PA-HAShim recombined with part of the PB2 segment, 5S and 28S rRNA. Moreover, 977PA-HAShim4 also recombined with 28S

rRNA. These events were rare as they exhibit low frequency, between 2-4x10^{-6}. It is important to note that recombination events were only detected in 977PA-HA viruses, indicating that this system has the capacity to detect all mutation types, even for very rare events, in comparison to other experimental models. The observed insertions occur at any location in the HA transgene, around the HA1/HA2 boundary region. As the rRNA is single-stranded, we were able to identify the directionality of insertion of rRNA in the HA transgene. Thus, rRNA insertions occurred during the viral cRNA synthesis. Moreover, all the inserts (viral or cellular origins) were integrated in antisense in the PA segment.

Discussion:

Different studies showed that some elements present in the viral genome can influence viral mutation rates. Viruses of the *Paramyxoviridae* and *Filoviridae* families are negative-stranded RNA viruses and share many similarities with avian influenza A virus in terms of their ribonucleoprotein and viral polymerase structures. It has been shown that specific sequences named as editing sites control mRNA-edition at the local level (Le Mercier & Kolakofsky, 2019). Upstream secondary structures have been suggested to influence nucleotide insertion events (Mehedi et al., 2013). Interestingly, more than ten percent of the Ebola virus antigenome is also edited during viral replication (Le Mercier & Kolakofsky, 2019; Shabman et al., 2014). Similar findings were also observed in positive-stranded RNA viruses. Secondary structures and base sequence motifs present on the viral genome are known to influence HIV-1 and HCV mutation rates in a site-specific manner (Geller et al., 2015, 2016; Krammer et al., 2018; Pathak & Temin, 1992; Pita, de Miranda, Schneider, & Roossinck, 2007; Simon-Loriere et al., 2009; Simon-Loriere, Martin, Weeks, & Negroni, 2010). Overall, these observations suggest that properties on the viral genome can directly impact viral mutation rates and potentially bend virus evolution to specific trajectories. Similar genome elements could be present on the HA segment and might potentially explain the emergence of HPAIV in nature. In the case of avian influenza virus evolution to highly pathogenic forms, it has been suggested that a conserved RNA stem-loop at the HA1/HA2 boundary region could influence the acquisition of MBCS (Beerens et al., 2020; Dietze et al., 2018; Garcia et al., 1996; Gultyaev et al., 2019, 2016; Nao et al., 2017; Perdue et al., 1997). However, these previous studies suffered from various limitations. Despite the great flexibility and convenient usage of minireplicon system (Nao et al., 2017) to test other factors, as different viral polymerase complexes, its use does not allow

581 to detect specific mutation events, like recombination events with other viral segments, which
582 are important in LPAIV-HPAIV transition in nature. Moreover, it is difficult to transpose viral
583 genome replication from minireplicon model to natural infection. Indeed, due to its non-
584 infectious nature, minireplicon systems might not be able to reproduce the exponential rounds
585 of replication cycles in the context of an infection, and might have limited vRNA production.
586 Moreover, it is important to note that cell transfection with minigenome plasmid can result in
587 plasmid contamination, even after DNAse treatments, which can potentially interfere with
588 minority variant analysis. Another limitation detected in some of these studies is the use of
589 experimental models that are under selection at the protein level (full wild-type viruses or
590 reassortants), which could potentially hide interesting transitory genome variants, possibly
591 involved in the LPAIV-HPAIV transition (Beerens et al., 2020; Dietze et al., 2018; Garcia et
592 al., 1996; Gultyaev et al., 2016; Nao et al., 2017; Perdue et al., 1997). Experimental work on
593 genomic factors influencing viral mutation events in influenza viruses is very limited. The
594 present work aimed to identify possible genomic elements present in the HA1/HA2 boundary
595 region that can influence viral mutation events that could eventually lead to the MBCS
596 acquisition, taking care of the limitations discussed above. Thus, we developed a new
597 experimental system consisting of engineered H7N1 viruses harbouring different transgenes
598 placed at the end of the PA segment (**Figure 1**). These transgenes consisted of a full HA
599 segment with various sequences at the HA1/HA2 region on the nucleotide motif and/or
600 predicted RNA structure. This present work constitutes at the time of writing the manuscript
601 the first study providing data on the strict influence of the viral genome environment on any
602 kind of mutation events, in the context of infection.
603 These modified viruses (977PA-HA viruses) could efficiently replicate in cell culture, allowing
604 the HA transgene to freely evolve during multiple replication cycles (**Figure 2D-E**). The HA
605 transgene remained stable during infection (**Figure 2B-C, S6**) and was not translated into any
606 protein (**Figure 2A**), dismissing artificial protein selection pressure. Similar insertion
607 frequencies were observed HA sequences from Shim and Shim24a2b in our 977PA-HA system
608 (**Figure 3**), as it was already described previously with reassortants' viruses (Nao et al., 2017).
609 However, 977PA-HA viruses were able to catch more mutation events than the reassortant's
610 viruses, suggesting that the 977PA-HA system was more sensible for detecting mutations, most
611 likely due to the fact that the HA transgene was not subjected to the protein selection pressure
612 (**Figure 3F-G**). Different mutation spectra were identified and found to be dependent on the
613 base-sequence environment. More precisely, viruses harbouring at least a 6-A-stretch on the
614 cRNA (HA sequence from Shim24a2b, ShimA3 and ShimA4) experienced significantly more

615 insertions than the other viruses which did not exhibit such nucleotide motif (**Figure 6A**). The
616 influence of the 6-A stretch in inducing insertion events was further confirmed with three other
617 HA sequence contexts that correspond to HA sequence deriving from documented HPAIV
618 progenitors (**Figure 7**). Similar findings were observed with TuMV (from *Potyviridae* family),
619 where the GAAAAAA homopolymeric sequence induces transcriptional frameshifting,
620 resulting in around 2% of insertions (Olspert, Chung, Atkins, Carr, & Firth, 2015). Similar
621 observations were also made with EBOV, in which a 7/8-U stretch triggers RNA editing during
622 both transcription and genome replication (Shabman et al., 2014). The previous studies and the
623 present data better specify the effect of such nucleotide motif on genome of viruses from
624 different families. A 6-A motif at the HA1/HA2 boundary region is exclusively found in dibasic
625 viruses, which have been characterised with low pathogenic-like and dependence to trypsin,
626 preceding the HPAIV emergence with MBCS motif (Ito et al., 2001; Seekings et al., 2020).
627 Combining these observations and results from the present study, we suggest that LPAIV-
628 HPAIV transition is a step-wise process involving the rise of intermediary mutations or specific
629 motifs, that can subsequently favour the apparition of other mutations. The 6-A nucleotide motif
630 triggers insertion events that can potentially enrich the HA1/HA2 boundary region with
631 additional basic amino acids. Indeed, virtually all observed inserts corresponded to additional
632 A-rich sequences and duplication of arginine codons. No specific sequence property was
633 identified to favour substitutions events to A, apart the high frequency observed with the HA-
634 Shim24a2b sequence, which is already A-rich (**Figure 5, 6**). We thus hypothesised the
635 following model of evolution for MBCS acquisition. First, prior substitutions can randomly
636 occur at the HA1/HA2 boundary region, and, possibly enrich the viral genome in A-content.
637 The 6-A motif is underrepresented in the sequence databases, suggesting its transitory nature.
638 The formation of a 6-A motif constitutes the principal milestone for evolution acceleration, and
639 could rapidly trigger polymerase slippage or stuttering in a directed manner, resulting in
640 nucleotide insertions and MBCS acquisition.
641
642 According to the present data from the different sequences derived from HA-Shim (HA-Shim,
643 Shim24a2b, ShimA1-A4), no relation was found between the predicted RNA structure on the
644 HA1/HA2 boundary region and mutation events. It is important to highlight that it is very
645 challenging to assess viral genome topology in the context of genome replication. The only
646 available methods to study RNA structures are indirect, and based on assessing the
647 thermodynamic stability of predicted RNA structures and base covariation using *in silico*
648 approaches (Parsch, Braverman, & Stephan, 2000; SantaLucia, 1998; Zhao et al., 2021), or

identifying accessible sequence positions to chemical reagents (Mailler, Paillart, Marquet, Smyth, & Vivet-Boudou, 2019). These methods do not take into the account the influence of a full active viral polymerase complex and its impact on the viral template topology. Moreover, it is possible that cellular factors can also influence viral genome topology. The lack of technology permitting to take into account possible *trans*-acting factors makes the interpretation of the possible influence of RNA structure on mutation events difficult.

In this work, we identified large insertion events that were observed in the HA1/HA2 region of the HA transgene of the 977PA-HAShim viruses (**Figure 8**). These insertions correspond to recombination with both viral segments and rRNA. The analysis showed that recombination events can occur and any location at the HA1/HA2 boundary region and always involved frameshifting insertions. Using the 977PA-HA system enabled us to detect several independent recombination events, even with genetically close base-sequence environments (HAShim, HAShim24a2b, HAShimA3 and HAShimA4 sequences, **Figure 8**). At the moment, no clear genomic factors (base-sequence motif, RNA structure) were identified as influencing recombination events in influenza viruses, probably due to a lack of data and the low frequency of such genetic events. Previous work suggested the presence of palindromic sequences and sequence complementary at recombination breakpoints, in both influenza virus and coronavirus (Herman et al., 2022; Khatchikian, Orlich, & Rott, 1989; Maurer-Stroh, Lee, Gunalan, & Eisenhaber, 2013). In our study, we did not detect such complementary and/or palindromic sequences. Additionally, a link between recombination events and snoRNA has also been proposed. In a previous work, we identified a common-shared predicted RNA structure between most of the progenitors of emergence events via recombination, suggesting a possible participation of such RNA structure in LPAIV-HPAIV transition (Dupré et al., 2021). These hypotheses are interesting and should be more explored.

Data availability:

Data and codes about Illumina sequencing libraries and analysis will be made available via a gitlab repository.

Supplementary data:

Supplementary data are provided: file.

Acknowledgement:

Funding:

This work was funded by a grant from the Agence Nationale de la Recherche (ANR-16-CE35-0005-01) to Romain Volmer. Gabriel Dupré is supported by PhD scholarship funded by the French Ministry of Research and Education.

Conflict of interest:

None declared.

References:

Abdelwhab, E.-S. M., Veits, J., & Mettenleiter, T. C. (2013). Genetic changes that accompanied shifts of low pathogenic avian influenza viruses toward higher pathogenicity in poultry. *Virulence*, *4*(6), 441. https://doi.org/10.4161/VIRU.25710

Beerens, N., Heutink, R., Harders, F., Bossers, A., Koch, G., & Peeters, B. (2020). Emergence and selection of a highly pathogenic avian influenza H7N3 virus. *Journal of Virology*. https://doi.org/10.1128/JVI.01818-19

Bessière, P., Figueroa, T., Coggon, A., Foret-Lucas, C., Houffschmitt, A., Fusade-Boyer, M., … Volmer, R. (2022). Opposite Outcomes of the Within-Host Competition between High- and Low-Pathogenic H5N8 Avian Influenza Viruses in Chickens Compared to Ducks. *Journal of Virology*, *96*(1). https://doi.org/10.1128/JVI.01366-21

Bui, C., Bethmont, A., Chughtai, A. A., Gardner, L., Sarkar, S., Hassan, S., … MacIntyre, C. R. (2016). A Systematic Review of the Comparative Epidemiology of Avian and Human Influenza A H5N1 and H7N9 - Lessons and Unanswered Questions. *Transboundary and Emerging Diseases*, *63*(6), 602–620. https://doi.org/10.1111/tbed.12327

Chung, C., Verheijen, B. M., Zhang, X., Huang, B., Coakley, A., McGann, E., … Vermulst, M. (2023). The fidelity of transcription in human cells. *Proceedings of the National Academy of Sciences*, *120*(5), e2210038120. https://doi.org/10.1073/PNAS.2210038120

de Bruin, A. C. M., Funk, M., Spronken, M. I., Gultyaev, A. P., Fouchier, R. A. M., & Richard, M. (2022). Hemagglutinin Subtype Specificity and Mechanisms of Highly Pathogenic Avian Influenza Virus Genesis. *Viruses*, *14*(7). https://doi.org/10.3390/V14071566

Dietze, K., Graaf, A., Homeier-Bachmann, T., Grund, C., Forth, L., Pohlmann, A., … Harder, T. (2018). From low to high pathogenicity—Characterization of H7N7 avian influenza viruses in two epidemiologically linked outbreaks. *Transboundary and Emerging Diseases*, *65*(6), 1576–1587. https://doi.org/10.1111/tbed.12906

Doud, M., & Bloom, J. (2016). Accurate Measurement of the Effects of All Amino-Acid Mutations on Influenza Hemagglutinin. *Viruses*, *8*(6), 155. https://doi.org/10.3390/v8060155

720 Dupré, G., Hoede, C., Figueroa, T., Bessière, P., Bertagnoli, S., Ducatez, M., … Volmer, R.
721 (2021). Phylodynamic study of the conserved RNA structure encompassing the
722 hemagglutinin cleavage site encoding region of H5 and H7 low pathogenic avian
723 influenza viruses. *Virus Evolution*, *7*(2), 1–16. https://doi.org/10.1093/VE/VEAB093
724 Escalera-Zamudio, M., Golden, M., Gutiérrez, B., Thézé, J., Keown, J. R., Carrique, L., …
725 Pybus, O. G. (2020). Parallel evolution in the emergence of highly pathogenic avian
726 influenza A viruses. *Nature Communications*, *11*(1), 5511.
727 https://doi.org/10.1038/s41467-020-19364-x
728 Funk, M., de Bruin, A. C. M., Spronken, M. I., Gultyaev, A. P., & Richard, M. (2022). In
729 Silico Analyses of the Role of Codon Usage at the Hemagglutinin Cleavage Site in
730 Highly Pathogenic Avian Influenza Genesis. *Viruses*, *14*(7).
731 https://doi.org/10.3390/V14071352
732 Garcia, M., Crawford, J. M., Latimer, J. W., Rivera-Cruz, E., & Perdue, M. L. (1996).
733 Heterogeneity in the haemagglutinin gene and emergence of the highly pathogenic
734 phenotype among recent H5N2 avian influenza viruses from Mexico. *Journal of General*
735 *Virology*, *77*(7), 1493–1504. https://doi.org/10.1099/0022-1317-77-7-1493
736 Garten, W., & Klenk, H. D. (1999, March 1). Understanding influenza virus pathogenicity.
737 *Trends in Microbiology*, Vol. 7, pp. 99–100. https://doi.org/10.1016/S0966-
738 842X(99)01460-2
739 Geller, R., Domingo-Calap, P., Cuevas, J. M., Rossolillo, P., Negroni, M., & Sanjuán, R.
740 (2015). The external domains of the HIV-1 envelope are a mutational cold spot. *Nature*
741 *Communications*, *6*(1), 1–9. https://doi.org/10.1038/ncomms9571
742 Geller, R., Estada, Ú., Peris, J. B., Andreu, I., Bou, J. V., Garijo, R., … Sanjuán, R. (2016).
743 Highly heterogeneous mutation rates in the hepatitis C virus genome. *Nature*
744 *Microbiology*, *1*(7), 1–7. https://doi.org/10.1038/nmicrobiol.2016.45
745 Gout, J.-F., Li, W., Fritsch, C., Li, A., Haroon, S., Singh, L., … Vermulst, M. (2017). The
746 landscape of transcription errors in eukaryotic cells. *Science Advances*, *3*(10).
747 https://doi.org/10.1126/sciadv.1701484
748 Gultyaev, A. P., Richard, M., Spronken, M. I., Olsthoorn, R. C. L., & Fouchier, R. A. M.
749 (2019). Conserved structural RNA domains in regions coding for cleavage site motifs in
750 hemagglutinin genes of influenza viruses. *Virus Evolution*, *5*(2), vez034.
751 https://doi.org/10.1093/ve/vez034
752 Gultyaev, A. P., Spronken, M. I., Funk, M., Fouchier, R. A. M., & Richard, M. (2021).
753 Insertions of codons encoding basic amino acids in H7 hemagglutinins of influenza A

viruses occur by recombination with RNA at hotspots near snoRNA binding sites. *RNA*, *27*(2), 123–132. https://doi.org/10.1261/rna.077495.120

Gultyaev, A. P., Spronken, M. I., Richard, M., Schrauwen, E. J. A., Olsthoorn, R. C. L., & Fouchier, R. A. M. (2016). Subtype-specific structural constraints in the evolution of influenza A virus hemagglutinin genes. *Scientific Reports*, *6*(1), 38892. https://doi.org/10.1038/srep38892

Herman, C., Baylor, C. B., Baylor, A. G., Wang Baylor, C., Cooke Baylor, M., Kohrn, B., … Lichtarge Baylor, O. (2022). *RNA polymerase inaccuracy underlies SARS-CoV-2 variants and vaccine heterogeneity*. https://doi.org/10.21203/RS.3.RS-1690086/V1

Hoffmann, E., Stech, J., Guan, Y., Webster, R. G., & Perez, D. R. (2001). Universal primer set for the full-length amplification of all influenza A viruses. *Archives of Virology*, *146*(12), 2275–2289. https://doi.org/10.1007/s007050170002

Hoffmann, Erich, Neumann, G., Kawaoka, Y., Hobom, G., & Webster, R. G. (2000). A DNA transfection system for generation of influenza A virus from eight plasmids. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(11), 6108–6113. https://doi.org/10.1073/pnas.100133697

Horimoto, T., Nakayama, K., Smeekens, S. P., & Kawaoka, Y. (1994). Proprotein-processing endoproteases PC6 and furin both activate hemagglutinin of virulent avian influenza viruses. *Journal of Virology*, *68*(9), 6074–6078. https://doi.org/10.1128/jvi.68.9.6074-6078.1994

Horimoto, T., Rivera, E., Pearson, J., Senne, D., Krauss, S., Kawaoka, Y., & Webster, R. G. (1995). Origin and Molecular Changes Associated with Emergence of a Highly Pathogenic H5N2 Influenza Virus in Mexico. *Virology*, *213*(1), 223–230. https://doi.org/10.1006/viro.1995.1562

Ito, T., Goto, H., Yamamoto, E., Tanaka, H., Takeuchi, M., Kuwayama, M., … Otsuki, K. (2001). Generation of a Highly Pathogenic Avian Influenza A Virus from an Avirulent Field Isolate by Passaging in Chickens. *Journal of Virology*, *75*(9), 4439–4443. https://doi.org/10.1128/JVI.75.9.4439-4443.2001

Kennedy, S. R., Schmitt, M. W., Fox, E. J., Kohrn, B. F., Salk, J. J., Ahn, E. H., … Loeb, L. A. (2014). Detecting ultralow-frequency mutations by Duplex Sequencing. *Nature Protocols*, *9*(11), 2586–2606. https://doi.org/10.1038/nprot.2014.170

Khatchikian, D., Orlich, M., & Rott, R. (1989). Increased viral pathogenicity after insertion of a 28S ribosomal RNA sequence into the haemagglutinin gene of an influenza virus. *Nature*, *340*(6229), 156. https://doi.org/10.1038/340156a0

788 Krammer, F., Smith, G. J. D., Fouchier, R. A. M., Peiris, M., Kedzierska, K., Doherty, P. C.,
789 … García-Sastre, A. (2018). Influenza. *Nature Reviews Disease Primers 2018 4:1*, *4*(1),
790 1–21. https://doi.org/10.1038/s41572-018-0002-y
791 Le Mercier, P., & Kolakofsky, D. (2019). Bipartite promoters and RNA editing of
792 paramyxoviruses and filoviruses. *Rna*, *25*(3), 279–285.
793 https://doi.org/10.1261/rna.068825.118
794 Lee, D.-H., Criado, M. F., & Swayne, D. E. (2020). Pathobiological Origins and Evolutionary
795 History of Highly Pathogenic Avian Influenza Viruses. *Cold Spring Harbor Perspectives*
796 *in Medicine*, a038679. https://doi.org/10.1101/cshperspect.a038679
797 Long, J. S., Mistry, B., Haslam, S. M., & Barclay, W. S. (2019, January 1). Host and viral
798 determinants of influenza A virus species specificity. *Nature Reviews Microbiology*, Vol.
799 17, pp. 67–81. https://doi.org/10.1038/s41579-018-0115-z
800 Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., &
801 Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology*,
802 *6*(1), 26. https://doi.org/10.1186/1748-7188-6-26
803 Mailler, E., Paillart, J.-C., Marquet, R., Smyth, R. P., & Vivet-Boudou, V. (2019). The
804 evolution of RNA structural probing methods: From gels to next-generation sequencing.
805 *Wiley Interdisciplinary Reviews: RNA*, *10*(2), e1518. https://doi.org/10.1002/wrna.1518
806 Maurer-Stroh, S., Lee, R. T. C., Gunalan, V., & Eisenhaber, F. (2013). The highly pathogenic
807 H7N3 avian influenza strain from July 2012 in Mexico acquired an extended cleavage
808 site through recombination with host 28S rRNA. *Virology Journal*, *10*(1), 139.
809 https://doi.org/10.1186/1743-422X-10-139
810 Mehedi, M., Hoenen, T., Robertson, S., Ricklefs, S., Dolan, M. A., Taylor, T., … Feldmann,
811 H. (2013). Ebola Virus RNA Editing Depends on the Primary Editing Site Sequence and
812 an Upstream Secondary Structure. *PLOS Pathogens*, *9*(10), e1003677.
813 https://doi.org/10.1371/journal.ppat.1003677
814 Munster, V. J., Schrauwen, E. J. A., de Wit, E., van den Brand, J. M. A., Bestebroer, T. M.,
815 Herfst, S., … Fouchier, R. A. M. (2010). Insertion of a Multibasic Cleavage Motif into
816 the Hemagglutinin of a Low-Pathogenic Avian Influenza H6N1 Virus Induces a Highly
817 Pathogenic Phenotype. *Journal of Virology*, *84*(16), 7953–7960.
818 https://doi.org/10.1128/JVI.00449-10
819 Nao, N., Yamagishi, J., Miyamoto, H., Igarashi, M., Manzoor, R., Ohnuma, A., … Takada, A.
820 (2017). Genetic Predisposition To Acquire a Polybasic Cleavage Site for Highly
821 Pathogenic Avian Influenza Virus Hemagglutinin. *MBio*, *8*(1).

822 https://doi.org/10.1128/mBio.02298-16
823 Olspert, A., Chung, B. Y., Atkins, J. F., Carr, J. P., & Firth, A. E. (2015). Transcriptional
824 slippage in the positive-sense RNA virus family *Potyviridae*. *EMBO Reports*, *16*(8),
825 995–1004. https://doi.org/10.15252/embr.201540509
826 Parsch, J., Braverman, J. M., & Stephan, W. (2000). Comparative sequence analysis and
827 patterns of covariation in {RNA} secondary structures. *Genetics*, *154*(2), 909–921.
828 Pathak, V. K., & Temin, H. M. (1992). 5-Azacytidine and RNA secondary structure increase
829 the retrovirus mutation rate. *Journal of Virology*, *66*(5), 3093–3100.
830 https://doi.org/10.1128/jvi.66.5.3093-3100.1992
831 Pauly, M. D., Procario, M. C., & Lauring, A. S. (2017). A novel twelve class fluctuation test
832 reveals higher than expected mutation rates for influenza A viruses. *ELife*, *6*, e26437.
833 https://doi.org/10.7554/eLife.26437
834 Perdue, M. L., García, M., Senne, D., & Fraire, M. (1997). Virulence-associated sequence
835 duplication at the hemagglutinin cleavage site of avian influenza viruses. *Virus Research*,
836 *49*(2), 173–186. https://doi.org/10.1016/S0168-1702(97)01468-8
837 Pita, J. S., de Miranda, J. R., Schneider, W. L., & Roossinck, M. J. (2007). Environment
838 Determines Fidelity for an RNA Virus Replicase. *Journal of Virology*, *81*(17), 9072–
839 9077. https://doi.org/10.1128/jvi.00587-07
840 Richard, M., Fouchier, R., Monne, I., Kuiken, T., Monne, I., & Kuiken, T. (2017).
841 Mechanisms and risk factors for mutation from low to highly pathogenic avian influenza
842 virus. *EFSA Supporting Publications*, *14*(10). https://doi.org/10.2903/sp.efsa.2017.EN-
843 1287
844 SantaLucia, J. (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA
845 nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences of*
846 *the United States of America*, *95*(4), 1460–1465. https://doi.org/10.1073/pnas.95.4.1460
847 Seekings, A. H., Howard, W. A., Nuñéz, A., Slomka, M. J., Banyard, A. C., Hicks, D., …
848 Brown, I. H. (2020). The emergence of H7N7 highly pathogenic avian influenza virus
849 from low pathogenicity avian influenza virus using an in ovo embryo culture model.
850 *Viruses*, *12*(9). https://doi.org/10.3390/v12090920
851 Shabman, R. S., Jabado, O. J., Mire, C. E., Stockwell, T. B., Edwards, M., Mahajan, M., …
852 Basler, C. F. (2014). Deep Sequencing Identifies Noncanonical Editing of Ebola and
853 Marburg Virus RNAs in Infected Cells. *MBio*, *5*(6). https://doi.org/10.1128/mbio.02011-
854 14
855 Simon-Loriere, E., Galetto, R., Hamoudi, M., Archer, J., Lefeuvre, P., Martin, D. P., …

Negroni, M. (2009). Molecular Mechanisms of Recombination Restriction in the Envelope Gene of the Human Immunodeficiency Virus. *PLOS Pathogens*, *5*(5), e1000418. https://doi.org/10.1371/journal.ppat.1000418

Simon-Loriere, E., Martin, D. P., Weeks, K. M., & Negroni, M. (2010). RNA Structures Facilitate Recombination-Mediated Gene Swapping in HIV-1. *Journal of Virology*, *84*(24), 12675–12682. https://doi.org/10.1128/JVI.01302-10

Simon, P. F., De La Vega, M. A., Paradis, É., Mendoza, E., Coombs, K. M., Kobasa, D., & Beauchemin, C. A. A. (2016). Avian influenza viruses that cause highly virulent infections in humans exhibit distinct replicative properties in contrast to human H1N1 viruses. *Scientific Reports*, *6*. https://doi.org/10.1038/srep24154

Soda, K., Asakura, S., Okamatsu, M., Sakoda, Y., & Kida, H. (2011). H9N2 influenza virus acquires intravenous pathogenicity on the introduction of a pair of di-basic amino acid residues at the cleavage site of the hemagglutinin and consecutive passages in chickens. *Virology Journal*, *8*, 64. https://doi.org/10.1186/1743-422X-8-64

Stech, O., Veits, J., Weber, S., Deckers, D., Schröer, D., Vahlenkamp, T. W., … Stech, J. (2009). Acquisition of a Polybasic Hemagglutinin Cleavage Site by a Low-Pathogenic Avian Influenza Virus Is Not Sufficient for Immediate Transformation into a Highly Pathogenic Strain. *Journal of Virology*, *83*(11), 5864. https://doi.org/10.1128/JVI.02649-08

Steinhauer, D. A. (1999). Role of Hemagglutinin Cleavage for the Pathogenicity of Influenza Virus. *Virology*, *258*(1), 1–20. https://doi.org/10.1006/viro.1999.9716

Swayne, D. E., & Pantin-Jackwood, M. (2006). Pathogenicity of avian influenza viruses in poultry. *Developments in Biologicals*, *124*, 61–67. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16447495

Tran, V., Moser, L. A., Poole, D. S., & Mehle, A. (2013). Highly Sensitive Real-Time In Vivo Imaging of an Influenza Reporter Virus Reveals Dynamics of Replication and Spread. *Journal of Virology*, *87*(24), 13321–13329. https://doi.org/10.1128/jvi.02381-13

Vandegrift, K. J., Sokolow, S. H., Daszak, P., & Kilpatrick, A. M. (2010). Ecology of avian influenza viruses in a changing world. *Annals of the New York Academy of Sciences*, Vol. 1195, pp. 113–128. https://doi.org/10.1111/j.1749-6632.2010.05451.x

Veits, J., Weber, S., Stech, O., Breithaupt, A., Gräber, M., Gohrbandt, S., … Stech, J. (2012). Avian influenza virus hemagglutinins H2, H4, H8, and H14 support a highly pathogenic phenotype. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(7), 2579–2584. https://doi.org/10.1073/pnas.1109397109

Wahlgren, J. (2011). Influenza A viruses: an ecology review. *Infection Ecology & Epidemiology*, *1*(1), 6004. https://doi.org/10.3402/iee.v1i0.6004

Wan, X.-F., Dong, L., Lan, Y., Long, L.-P., Xu, C., Zou, S., … Shu, Y. (2011). Indications that Live Poultry Markets Are a Major Source of Human H5N1 Influenza Virus Infection in China. *Journal of Virology*, *85*(24), 13432–13438. https://doi.org/10.1128/jvi.05266-11

Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M., & Kawaoka, Y. (1992). Evolution and ecology of influenza A viruses. *Microbiological Reviews*, Vol. 56, pp. 152–179. https://doi.org/10.1111/j.1751-0813.1992.tb09940.x

Wong, S. S., Yoon, S. W., Zanin, M., Song, M. S., Oshansky, C., Zaraket, H., … Crossley, B. (2014). Characterization of an H4N2 influenza virus from Quails with a multibasic motif in the hemagglutinin cleavage site. *Virology*, *468–470*, 72–80. https://doi.org/10.1016/j.virol.2014.07.048

Wu, Z. Q., Zhang, Y., Zhao, N., Yu, Z., Pan, H., Chan, T. C., … Liu, S. L. (2017). Comparative epidemiology of human fatal infections with novel, high (H5N6 and H5N1) and low (H7N9 and H9N2) pathogenicity avian influenza a viruses. *International Journal of Environmental Research and Public Health*, *14*(3), 263. https://doi.org/10.3390/ijerph14030263

Zhao, Q., Zhao, Z., Fan, X., Yuan, Z., Mao, Q., & Yao, Y. (2021). Review of machine learning methods for RNA secondary structure prediction. *PLOS Computational Biology*, *17*(8), e1009291. https://doi.org/10.1371/JOURNAL.PCBI.1009291

Figures:

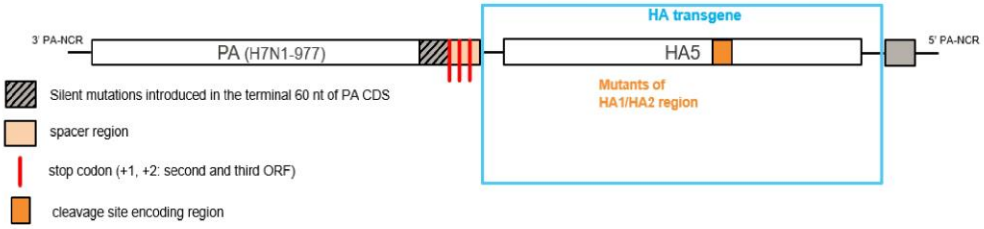


Figure 1: Schematic of the 977PA-HA construct. Different HA5 sequences (entire genomic segment, HA transgene) were inserted at the end of the PA ORF of 977. Additional modifications were added to the construction to ensure genetic stability and proper packaging of the modified PA segment. First, the last sixty nucleotides of the PA CDS sequence were duplicated and positioned after the HA transgene (gray box). The PA 5'NCR was placed right after the duplicated sequence. Secondly, the first copy of the sixty PA CDS end was modified, and silent mutations were introduced (hashed gray box). The HA transgene was placed out of frame compared to the PA ORF. To better ensure the absence of the HA transgene translation, additional stop codons were added in a spacer region between the PA CDS sequence and the HA transgene (red vertical lines).

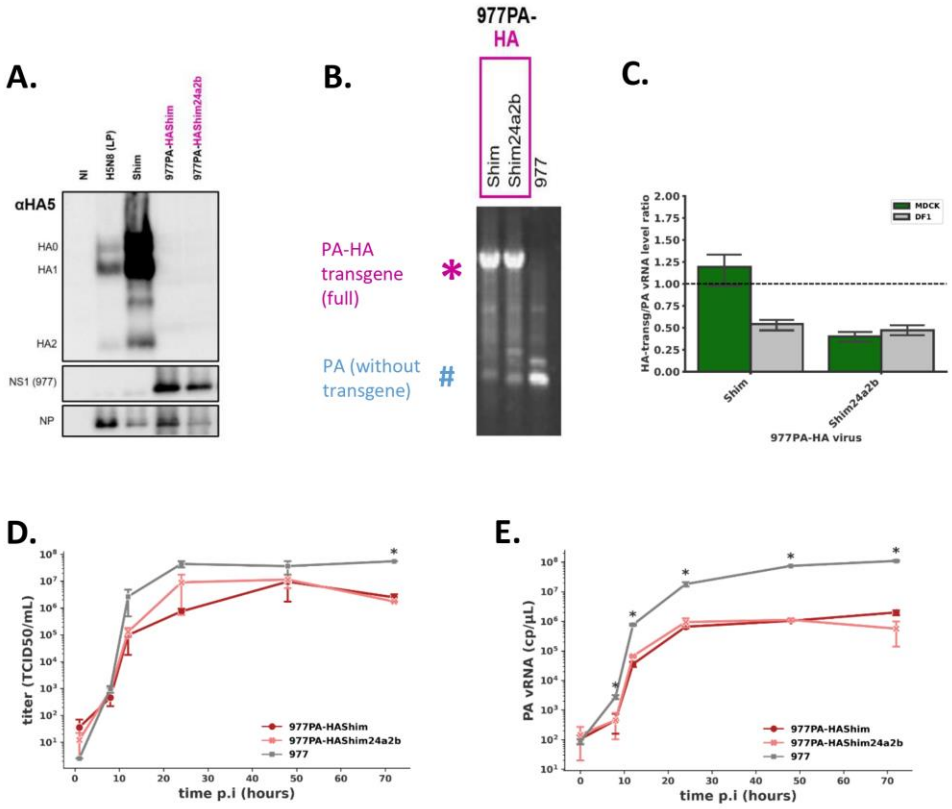
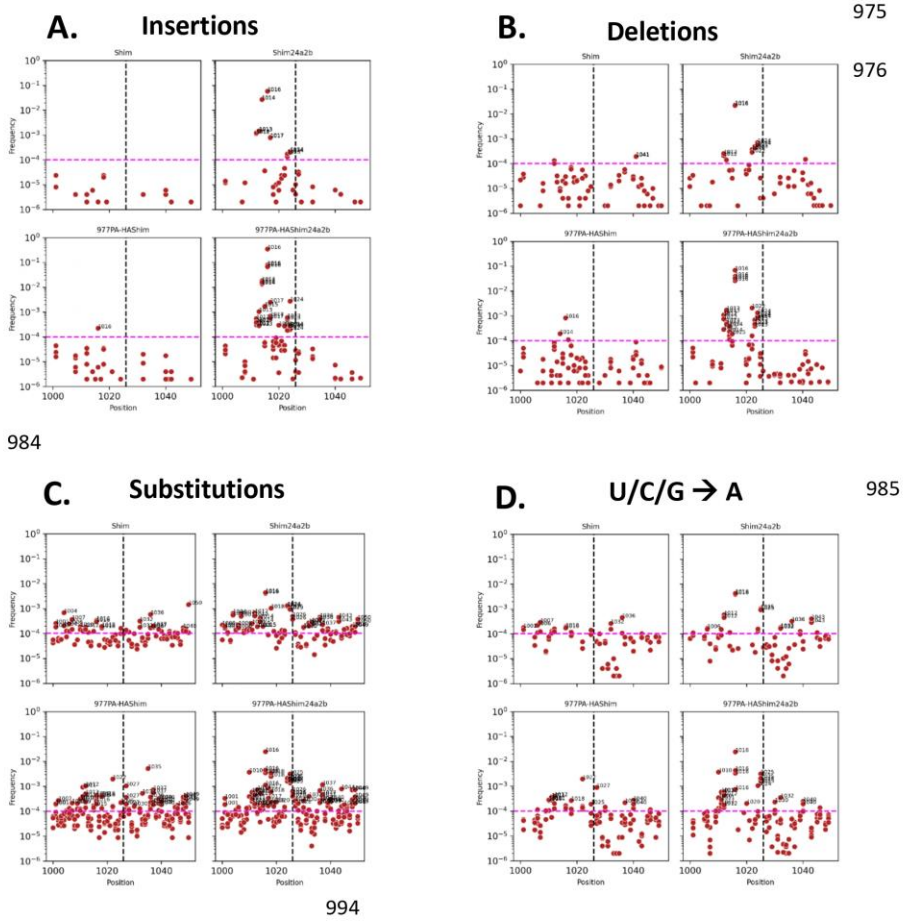


Figure 2: Characterization of the 977PA-HA system. A) Analysis of HA5 expression in the produced 977PA-HA viruses. Western blots were realized on MDCK-infected cells with 977PA-HAShim, 977PA-HAShim24a2b, H5N8-LP and Shim at an MOI of 0.1 for 24 hours. Cells were collected, lysed, and cytoplasmic contents were purified. HA5, NS1 (977), and NP protein expression were assessed using primary antibodies. The secondary antibodies coupled with Horseradish Peroxidase (HRP) were used. B) Specific viral PA-segment RT-PCR was performed on RNA recovered from viral stocks of passage 1 (used for deep-sequencing analysis) on both MDCK and DF-1 cells. A set of primers that bind at the end pf the PA CDS sequence and the end of HA transgene were used for the PCR. The PCR amplicons were analyzed by agarose gel electrophoresis (data from infected MDCK cells are shown). Expected product sizes for native PA (977) and the 977PA-HA (full-length HA transgene) are indicated. C) Quantification of viral RNA from the unmodified part of the PA segment and the HA

966 transgene was realized by qPCR. Ratios of viral RNA from the HA transgene and the PA
967 sequence were calculated. (D-E) Viral growth kinetics were realized on infected-MDCK cells
968 (MOI of 0.0001) with 977PA-HAShim, 977PA-HAShim24a2b and 977. The production of
969 infectious particles (D) and viral PA segment replication (E) were assessed by TCID50 titration
970 and specific viral PA-segment RT-qPCR from collected supernatants at different time points.
971 Two independent experiments were carried-out, starting at the production of virus from the
972 reverse genetics step. Standard deviation are shown. Comparison of means were done using
973 Student test (*: p-value < 0.05).

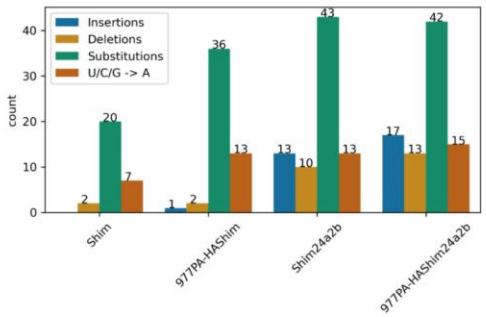
974



E.

Virus	Replicate	Insertions	Deletions	Substitutions	U/C/G → A
Shim	1	2.33E-06	2.25E-05	1.02E-04	3.27E-05
	2	1.83E-06	1.50E-05	9.99E-05	3.98E-05
Shim24a2b	1	4.85E-03	2.11E-03	7.22E-04	4.23E-04
	2	5.00E-03	1.98E-03	7.28E-04	4.44E-04
977PA-HAShim	1	1.17E-06	1.18E-05	1.15E-04	4.62E-05
	2	3.33E-06	1.05E-05	1.40E-04	4.87E-05
977PA-HAShim24a2b	1	6.69E-03	2.35E-03	7.53E-04	4.81E-04
	2	2.93E-02	2.87E-03	2.84E-03	2.32E-03

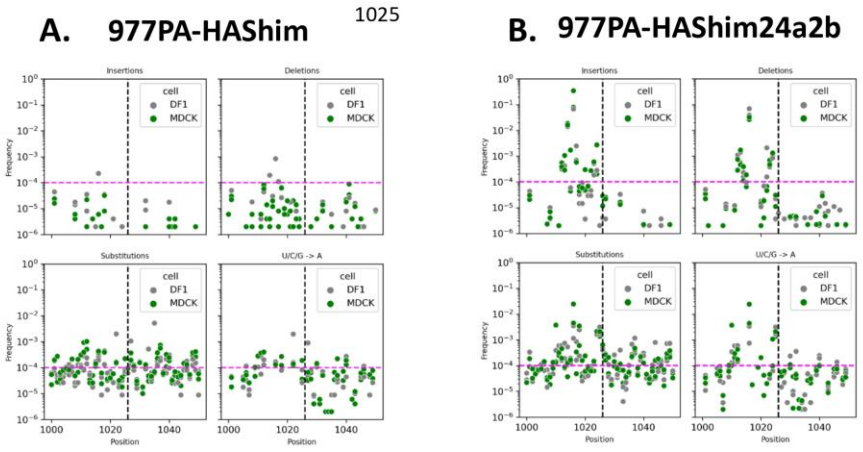
F.



G.

Virus	Replicate	frameshifting	non-frameshifting
977PA-HAShim24a2b	1	0.9209	0.0791
	2	0.9244	0.0756
Shim24a2b	1	0.8282	0.1718
	2	0.8181	0.1819

Figure 3: Evaluation of obtained mutations' information with 977PA-HA system. Mutation spectra of the HA1/HA2 boundary region from the HA segment of Shim, Shim24a2b, or from the HA transgene of 977PA-HAShim and 977PA-HAShim24a2b, were analysed. Results from infected MDCK cells at an MOI of 0.001 TCID50 are shown. (A-D) All insertions, deletions, substitutions and transitions/transversions to adenine events around the HA1/HA2 boundary region are represented in scatter plots. Frequencies of each event are shown. Nucleotide positions with significantly higher mutation frequency to 10^(-4) are annotated on the plots. The vertical-hashed line represents position 1026, corresponding to the R codon's last nucleotide in the RGLF protein motif at the HA1/HA2 region. (E) Mean mutation frequencies were calculated at the 1015-1026 nucleotide window (on the cRNA HA-ORF). (F) Counts of observed mutation events per virus were shown. Two independent experiments were realised. (G) Frequencies of frameshifting (non multiple of 3) or non-frameshifting insertion events in 977PA-HAShim24a2b and Shim24a2b.



C.

Virus	Replicate	Cell	Insertions	Deletions	Substitutions	U/C/G → A
977PA-HAShim	1	DF1	2.13E-05	8.95E-05	2.52E-04	2.00E-04
		MDCK	1.17E-06	1.18E-05	1.15E-04	4.62E-05
	2	DF1	2.94E-06	3.67E-06	7.78E-05	3.60E-05
		MDCK	3.33E-06	1.05E-05	1.40E-04	4.87E-05
977PA-HAShim24a2b	1	DF1	6.47E-03	5.91E-03	1.06E-03	5.46E-04
		MDCK	6.69E-03	2.35E-03	7.53E-04	4.88E-04
	2	DF1	5.87E-03	3.55E-03	4.73E-04	2.08E-04
		MDCK	2.93E-02	2.87E-03	2.84E-03	2.32E-03

Figure 4: Analysis of the host-cell environment impact on mutation events. Mutation spectra of the HA1/HA2 boundary region from the HA transgene of 977PA-HAShim and 977PA-HAShim24a2b were analysed. Results from two independent experiments with both infected MDCK or DF-1 cells at an MOI of 0.001 TCID50 are shown. All insertions, deletions, substitutions and transitions/transversions to adenine events around the HA1/HA2 boundary region of 977PA-HAShim (A) and 977PA-HAShim24a2b (B) are represented in scatter plots. Frequencies of each event are shown. Nucleotide positions with significantly higher mutation frequency to 10^(-4) are annotated on the plots. The vertical-hashed line represents position 1026, corresponding to the R codon's last nucleotide in the RGLF protein motif at the HA1/HA2 region. (C) Mean mutation frequencies were calculated at the 1015-1026 nucleotide window (on the cRNA HA-ORF).

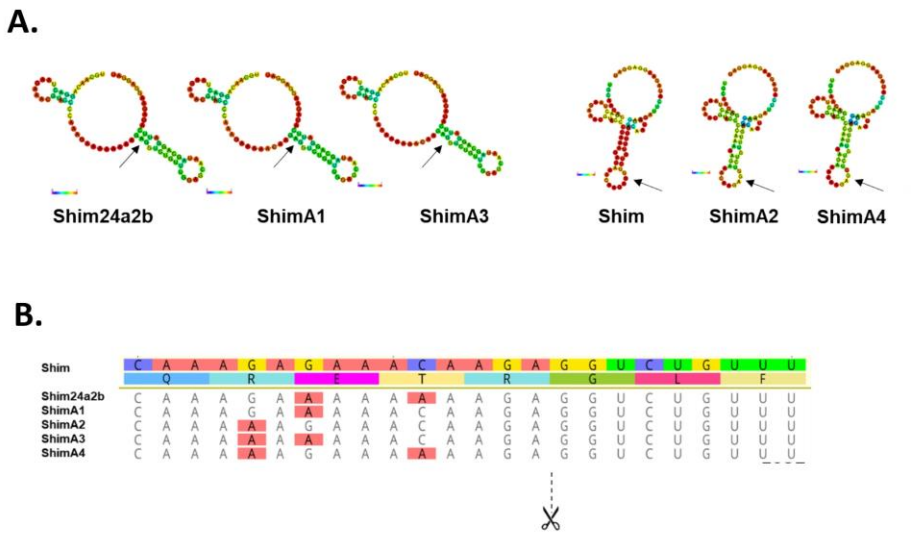
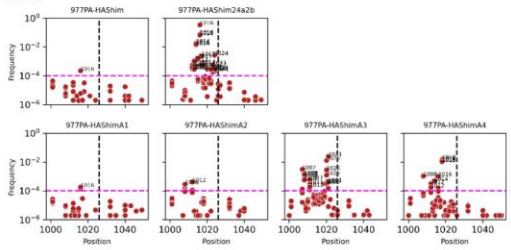
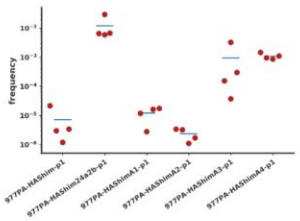


Figure 5: 977PA-HA viruses with HAShim-derived sequences. Different 977PA-HA viruses with different HA transgene derivating from the HA segment of Shim and Shim24a2b were generated via reverse genetics. The supplementary HAShim transgenes (ShimA1-A4) were generated by site-directed mutagenesis. Substitutions to A were realised to create all the HA1/HA2 possible variants in the 1015-1026 nucleotide window (cRNA, HA ORF) numbering. A) Predicted-RNA structures of the HAShim cRNA sequences on the 80 nucleotide window were realised, clustering the HAShim sequences in two groups. B) Alignment of the HAShim cRNA sequences with highlighted substitutions is depicted.

1064

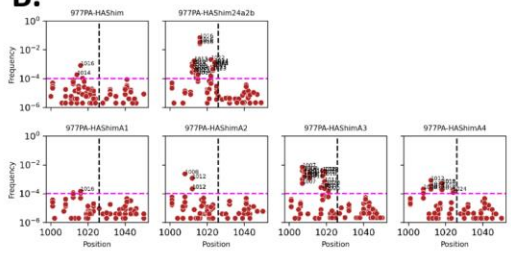
A. insertions

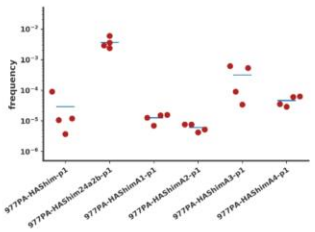




1070

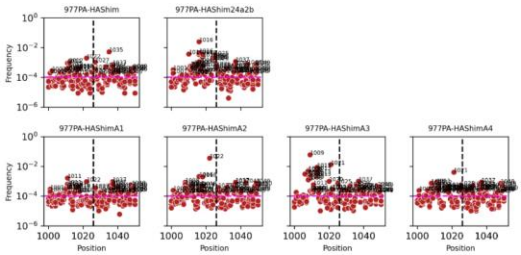
B. deletions

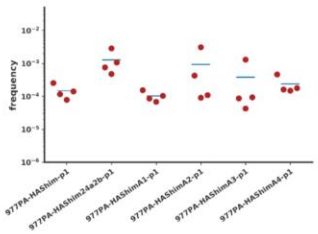




C. substitutions

1077

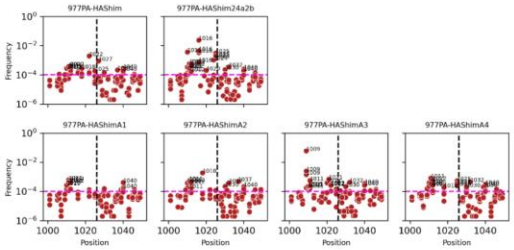


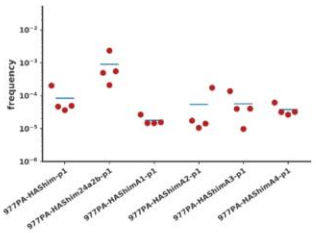


1083

D. U/C/G → A

1084





1086

1087

1088

Figure 6: Analysis of the influence of base-sequence properties on mutation events. Mutation spectra of the HA1/HA2 boundary region from the HA transgene of 977PA-HAShim, 977PA-HAShim24a2b, 977PA-HAShimA1-A4 were analysed. Results from two independent experiments with both infected MDCK or DF-1 cells at an MOI of 0.001 TCID50 are shown. All insertions (A), deletions (B), substitutions (C) and transitions/transversions to adenine (D) events around the HA1/HA2 boundary region of the specific HA transgene are represented in scatter plots. Frequencies of each event are shown. Nucleotide positions with significantly higher mutation frequency to 10^(-4) are annotated on the plots. The vertical-hashed line represents position 1026, corresponding to the R codon's last nucleotide in the RGLF protein motif at the HA1/HA2 region. Mean mutation frequencies were also calculated at the 1015-1026 nucleotide window (on the cRNA HA-ORF).

A. 977PA-HA viruses with HA sequences from documented HPAIV-progenitors

Emergence	HA sequence of HPAIV progenitor	977PA-HA virus	HA transgene version	HA1/HA2 boundary region
1966 Canada H5N9	CY087808	977PA-HACY	wt form	CAAAGAGAAA**C**AAGAGGTCTGTTT
		977PA-HACY6A	6A mutant	CAAAGAG**AAAAAA**GAGGTCTGTTT
2006 South Africa H5N2	EF591757	977PA-HAEF	wt form	CAAAGAGAAACAAGAGGATTATTT
		977PA-HAEF6A	6A mutant	CAAAGAG**AAAAAA**GAGGATTATTT
1994–1995 Mexico H5N2	GU186573	977PA-HAGU	wt form	CAAAGAGAAA**C**AAGAGGCCTATTT
		977PA-HAGU6A	6A mutant	CAAAGAG**AAAAAA**GAGGCCTATTT

B. Insertion frequency

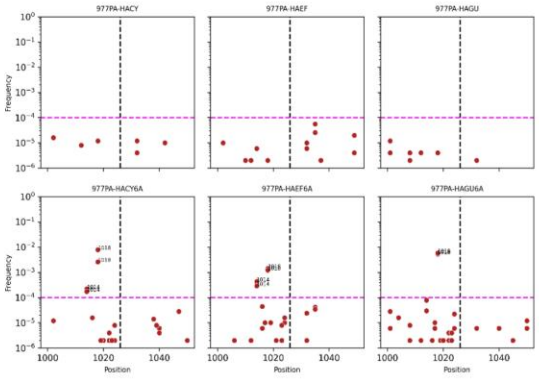


Figure 7: Analysis of the influence of the 6A stretch in insertion events. (A.) 977PA-HA viruses with HA transgenes deriving from the HA sequence of three known H5-HPAIV progenitors: A/turkey/Ontario/6213/1966, A/ostrich/South Africa/AI1160/2006, A/chicken/Mexico/31381-7/1994 were generated. A "6A mutant" version containing the 6A motif at the HA1/HA2 boundary region of the HA transgene was generated through directed mutagenesis for each virus. (B.) Insertion frequency per nucleotide position of the HA1/HA2 boundary region from the HA transgene of the newly produced 977PA-HA viruses were analysed. Results from two independent experiments with infected MDCK cells at an MOI of 0.001 TCID50 are shown. Nucleotide positions with significantly higher mutation frequency to 10^(-4) are annotated on the plots. The vertical-hashed line represents position 1026, corresponding to the R codon's last nucleotide in the RGLF protein motif at the HA1/HA2 region.

Virus	Position of the insertion in HA transgene (cRNA)	Insert	Insert length	Match	Positions of the insert in match	Score	Expect	Identity	Gaps	Frequency
977PA-HA Shim24a2b	1021	TATTTTGCATT GCCTCTTCCCT GTACTTGCTGT GATC	37	segment 4 HA7	1526-1490	69.4 bits(37)	1.00E-11	37/37 (100%)	0	2.31E-06
977PA-HAShim	1041	ATACTCTGTTC CGGTGGGGCT GCTGC	26	segment 1 PB2	544-519	52.0 bits(26)	2.00E-06	26/26 (100%)	0	2.00E-06
	1018	AGAGATCAGA CGAGATCGGG CGCGTTCAGG GTGGTATGGC CGT	43	5S rNA	46-5	78.7 bits(42)	3.00E-13	43/43 (100%)	0	2.00E-06
	1042	TAACTTTTCTG ACACCTCCTG CT	23	28S rRNA	4089-4067	46.1 bits(23)	5.00E-04	23/23 (100%)	0	4.00E-06
977PA-HAShimA3	1013	CATCTTTGTAG CCGCTGCTTA GTTTGACTGG GT	33	segment 4 HA7	1587-1555	62.1 bits(33)	2.00E-09	33/33 (100%)	0	2.00E-06
977PA-HAShimA4	1040	AAAACTTTTCT GACACCTCCT GCTT	25	28S rRNA	4088-4066	46.1 bits(23)	7.00E-04	23/25 (92%)	0	2.00E-06

Figure 8: Recombination events detected in the 977PA-HA system. Analysis of large insertion observed in the deep-sequencing libraries. BLAST-nucleotide analysis confirmed the exogenous origin (viral or cellular) of these insertions. The sequence of the insert integrated into the viral genome (cRNA), its length, frequency, and origin are provided.

Discussion et perspectives

I. Rôle de la structure secondaire d'ARN conservée au niveau de HA1/HA2 dans l'émergence de HPAIV :

Plusieurs études ont montré la présence d'une structure secondaire d'ARN prédite, de type tige-boucle, au niveau de HA1/HA2 (M L Perdue et al. 1996; Michael L. Perdue et al. 1997; Nao et al. 2017; Dietze et al. 2018; Beerens et al. 2020; Gultyaev et al. 2016; 2019). Cette structure serait présente sur la RNPv, comme le suggère cette étude qui montre que la zone codant le site de clivage est faiblement recouverte en NP (Williams et al. 2018). Cependant, cette structure prédite est conservée au sein de tous sous-type influenza aviaires et chez les autres genres Influenza, remettant en cause son implication dans l'évolution LPAIV-HPAIV (Gultyaev et al. 2019). Néanmoins, il est possible que des différences plus subtiles puissent être identifiées dans ces structures et expliquer pourquoi le phénomène d'évolution vers l'acquisition du MBCS et quasi-exclusivement observé chez les séquences HA de sous-type H5/7, et également déterminer si seulement certaines séquences HA-5/7 sont plus particulièrement propices à l'évolution vers les HPAIV. Nous avons testé cette hypothèse dans une première étude (cf page 59, (Dupré et al. 2021)), en nous intéressant à la stabilité thermodynamique de la structure prédite, son évolution en fonction du lien phylogénétique de la séquence HA à l'événement d'émergence : séquence évoluant ou non vers l'acquisition d'un MBCS. De plus, nous avons également cherché à identifier au sein de la diversité des séquences HA-5/7 des structures pouvant être partagées parmi les progéniteurs des émergences HPAIV (déjà caractérisés et/ou inférés grâce à l'analyse phylogénétique). L'analyse de la stabilité thermodynamique de la structure conservée n'a pas montré de différence significative entre les types de chemins évolutifs. Cependant, l'analyse par *clustering* des structures d'ARN prédites a permis d'identifier une structure conservée chez la majorité des progéniteurs connus d'émergence par évènement de recombinaison. Des observations similaires ont été observées sur le génome du HIV, sur lequel des zones très structurées correspondent à des point chauds de recombinaison (Simon-Loriere et al. 2009; 2010; Tomezsko et al. 2020). Au vu des résultats de cette étude, nous émettons l'hypothèse que cette structure partagée serait favorable au mécanisme de recombinaison chez les virus de sous-type H7. Compte tenu du faible nombre de cas d'évènement d'émergence, il est difficile d'étudier ce phénomène expérimentalement. Toutefois, des études supplémentaires doivent être envisagées afin de confirmer le rôle biologique d'un tel motif de type structure d'ARN sur l'évènement de recombinaison et l'émergence HPAIV.

Il est important de noter que la majorité des études menées à ce jour sur le sujet des structures secondaires d'ARN chez les virus influenza reposent sur des méthodes de prédictions et/ou la détection de points de covariation. Ces travaux ne prennent pas en considération l'influence de facteurs extérieurs à la molécule d'ARN, comme la fixation de la NP, du complexe polymérase, de facteurs cellulaires, rendant l'interprétation du rôle éventuelle de ces structures prédites difficile. Un moyen de pouvoir mieux appréhender l'intégrité de ces structures est l'utilisation de méthodes de cartographie chimique du génome viral (Mailler et al. 2019). Coupler les données d'accessibilité des bases à certains réactifs chimiques avec les algorithmes de prédiction de repliement des séquences permet de renforcer ces analyses, et doit être encouragé. Il n'existe que très peu d'information sur l'accessibilité des bases. En effet, les analyses de cartographie chimique des structures du génome influenza ont été réalisées jusqu'à présent seulement sur des génomes de virus humains de sous-type H1N1 (A/Puerto Rico/8/34) et H3N2 (A/Udorn/72) (Dadonaite et al. 2019). A quel point la topologie du génome de virus influenza reste conservée entre les différents sous-types et les origines humaine ou aviaire reste encore indéterminée. Une étude similaire réalisée sur des représentants de virus influenza aviaires et des progéniteurs connus d'émergence HPAIV devrait être envisagée.

Outre la prise en compte nécessaire de facteurs extérieurs à la séquence elle-même pouvant aussi influencer le repliement du génome viral, tels que le complexe polymérase complet (l'environnement RNPv) et certaines protéines cellulaires, les analyses de prédiction de structures d'ARN sont dépendantes du choix de la fenêtre d'analyse. En effet, des différences notables sur les structures d'ARN (topologies différentes, valeurs de stabilité thermodynamique de la structure la plus probable) ont pu être observées, suivant le choix de la longueur de la fenêtre de prédiction (M L Perdue et al. 1996; Nao et al. 2017; Gultyaev et al. 2019; Dupré et al. 2021).

De plus, aucune méthode ne permet encore à ce jour d'évaluer l'intégrité de la structure secondaire conservée prédite au cours de la réplication du génome viral. Avec les données de structure du complexe polymérase viral actuelles, il est possible d'envisager des prédictions de structure d'ARN sur les parties de la matrice considérées comme « libres » et à distance de la région enclavée dans la sous-unité PB1 (French et al. 2022). Des systèmes permettant d'appréhender l'activité spatio-temporelle des polymérases à de très hautes résolutions ont été développés, utilisant les technologies de transfert d'énergie par résonance de fluorescence, de pinces optiques ou magnétiques (Dulin et al. 2015; 2017). Ces méthodes sont intéressantes pour identifier les signatures cinétiques des polymérases virales, les possibles mécanismes d'erreurs

et pourraient être combinées aux analyses de structures afin d'identifier des corrélats entre des motifs de la séquence virale et l'activité de la polymérase virale.

Il est possible que la structure secondaire d'ARN prédite au niveau de HA1/HA2 soit impliquée dans un tout autre rôle biologique, au-delà l'acquisition du MBCS et l'évolution vers les formes hautement pathogènes. En effet, d'autres structures localisées au niveau interne d'autres segments génomiques ont été identifiées et décrites comme impliquées dans l'empaquetage du génome viral (Gavazzi et al. 2013; Dadonaite et al. 2019). Ces structures semblent aussi être plastiques et impliquées dans de multiples interactions inter et intra segments (Dadonaite et al. 2019). Une étude précédente ayant réalisée des mutations synonymes au niveau de la région codant HA1/HA2 des virus A/PR/8/34 (H1N1), A/Bilthoven/16190/68 (H3N2) et A/Anhui/1/2005 (H5N1) n'a pas mis en évidence une quelconque réduction de la réplication virale, par lecture du diamètre de plage de lyse en culture cellulaire (Gultyaev et al. 2016). Ces résultats suggèrent une faible contrainte de cette structure conservée sur la réplication virale. Cependant, il aurait été intéressant d'évaluer ces mutants en cinétique de réplication, afin de vérifier si des différences de réplication virale peuvent être observées à des temps plus précoces, et ainsi potentiellement identifier certains mutants ayant un avantage réplicatif. L'hypothèse que la structure secondaire d'ARN conservée puisse participer à l'empaquetage du génome viral et être sélectionnée par l'évolution pour son effet positif sur la réplication virale reste à être testée. En revanche, il est encore difficile d'attribuer un rôle clair de cette structure dans l'émergence LPAIV-HPAIV. Des études supplémentaires sur ce sujet restent nécessaires.

II. Le phénomène de recombinaison chez les virus influenza aviaires

Des évènements de recombinaisons avec des gènes cellulaires (ARNr 28S) ou segments viraux (M, NP) ont été observés chez les virus influenza aviaires de sous-type H7, et ont permis l'émergence naturelle de six à huit événements HPAIV indépendants (**Tableau 3**).

Dans notre précédente étude (Dupré et al. 2021), nous avons pu identifier une structure partagée chez la majorité des progéniteurs connus des émergences H7-HPAIV par recombinaison, suggérant une influence probable de la topologie de l'ARN dans le mécanisme de recombinaison. Les observations réalisées sur le génome du HIV-1 corroborent avec cette hypothèse (Simon-Loriere et al. 2009; 2010). D'autre part, une étude récente a mis en évidence le lien potentiel entre certains petits ARN nucléolaires ou snoRNA (*small nucleolar RNA*) et l'évènement d'émergence par recombinaison (Gultyaev et al. 2021). Il a été montré sur cinq évènements de recombinaison avec des gènes cellulaires ou viraux que la séquence insérée correspond à un *hotspot* de fixation pour certains snoRNA. En particulier, deux snoRNA ont été identifiés chez la majorité des émergences décrites : SNORD22 and SNORD49. Dans notre deuxième étude (cf manuscrit p 77), nous avons pu détecter six évènements de recombinaison avec des ARN ribosomiques ou viraux au niveau de la séquence codant le site de clivage HA, produits expérimentalement. Des séquences insérées d'origine virale (segments HA(7) et PB2) ou cellulaire (ARNr 28s et l' ARNr 5S) ont pu être observées au sein de certains virus 977PA-HA. Il a été possible de montrer la directionnalité de l'insertion des ARNr dans les deux analyses grâce à la nature simple-brin de ces ARN (Gultyaev et al. 2021 et notre étude : manuscrit p 77). Les séquences ARN ribosomiques sont strictement insérées lors de la synthèse de l'ARNc viral. Contrairement à ce qui est observé dans la nature, les évènements de recombinaisons identifiés chez les virus 977PA-HA de notre étude ont eu lieu au sein de transgène, correspondant à des séquences HA de sous-type H5 (dérivées de la souche Shim). Cela suggère donc que l'évènement de recombinaison n'est pas contraint par des caractéristiques de séquence génomique spécifique au sous-type. Dans notre étude (manuscrit p 77), les évènements de recombinaison n'ont été observés que sur des virus 977PA-HA produits sur MDCK. La raison pour laquelle aucune recombinaison n'a été observé sur des virus produit en cellules DF-1 reste inconnue. Ceci pourrait être expliqué par le fait que la quantité de particules infectieuses produites en DF-1 était moins importante en comparaison à des infections sur MDCK, du fait d'une sensibilité accrue des DF-1 à la trypsine ajoutée dans le

milieu de d'infection. Le mécanisme moléculaire et les raisons pour lesquelles les recombinaisons sont observées uniquement chez les virus de sous-types H7 dans la nature sont encore inconnus. La piste d'un lien potentiel entre les trois acteurs suivants : certains snoRNA, la séquence « donneuse » ou potentiel insert (ARNr ou segment viral), ainsi que le segment HA (séquence « acceptatrice ») serait intéressante à investiguer. Nous pouvons émettre l'hypothèse d'un modèle d'interaction suivant, dans lequel le snoRNA jouerait le rôle de « pont » moléculaire entre la séquence donneuse et acceptatrice via une complémentarité de séquence snoRNA/insert potentiel et des interactions de type « *kissing loops* » entre la structure secondaire du snoRNA et la tige boucle conservée présente au niveau de HA1/HA2 sur la séquence acceptatrice.

Tableau 3 *: Evènements d'émergence naturels par recombinaison*

Localisation	Année	Numéro d'accession	Nom de la souche	Origine de l'insertion	Références
Chili	2002	AY303631	A/chicken/Chile/4322/02/H7N3	Segment NP	(Suarez et al. 2004)
Canada	2004	AY611524	A/chicken/British/Columbia/04/H7N3	Segment M	(Hirst et al. 2004)
Canada	2007	EU500860	A/chicken/SK/HR/00011/2007/H7N3	ARNr 28S	(Berhane et al. 2009)
Mexique	2012	JX908509	A/chicken/Jalisco/12283/2012/H7N3	ARNr 28S	(Maurer-Stroh et al. 2013)
USA	2017	KY818811	A/chicken/Tennessee/17/007147/2/2017/H7N9	ARNr 28S	(D.-H. Lee et al. 2017)
Italie	1999	KF493066	A/chicken/Italy/4845/1999/H7N1	origine inconnue, ARNt ?	(Gultyaev et al. 2021)
Pays-Bas	2003	EPI_ISL_391798	A/chicken/Netherlands/03008927/2003	origine inconnue, ARNt ?	(Gultyaev et al. 2021)
Espagne	2009	GU121458	A/chicken/Spain/6279/2/2009/H7N7	ARNr 28S	(Iglesias et al. 2010; Gultyaev et al. 2021)

III. La formation du motif homopolymérique de six adénines, nécessaire à « l'accélération » de l'évolution de la région HA1/HA2

La transition LPAIV-HPAIV apparaît comme un phénomène nécessitant certaines étapes ou l'apparition de variants intermédiaires, avant d'évoluer vers l'acquisition d'un MBCS, et donc l'émergence (Garcia et al. 1996; Abolnik 2017; Seekings et al. 2020). Ces variants intermédiaires incluent également les génomes contenant un cortège de mutations intermédiaires, dont l'impact potentiellement négatif sur le fitness général du virus, les rendent difficilement observables à cause du filtre de la sélection. Ces génomes peuvent être rapprochés au groupe des génomes interférents par leur caractère défavorable au fitness viral. Cependant, leur rôle dans l'évolution pourrait être primordial. En effet, ils peuvent servir de matrices transitoires, continuer à évoluer au cours des cycles de réplication du génome, et enfin permettre d'obtenir la constellation « optimale » de mutations pour être sélectionnés, et permettre l'émergence d'un HPAIV. C'est précisément dans l'objectif de mieux capturer et étudier les évènements de mutations précédant l'apparition du MBCS que le système de virus génétiquement modifiés 977PA-HA a été créé (cf manuscrit p 77). Ce model expérimental permet d'analyser les évènements de mutation apparaissant sur un transgène, ajouté dans le génome d'un virus influenza aviaire, non soumis à la pression de sélection protéique, et pouvant donc évoluer librement au cours des cycles réplicatifs.

En utilisant ce système d'évolution libre (virus PA-HA), nous avons pu mettre en évidence que la quasi-totalité des insertions observées au niveau de la région codant HA1/HA2 sont dépendantes d'un motif homopolymérique précis. Les insertions observées enrichissent considérablement la teneur en A au niveau de la région codant HA1/HA2 et favorisent l'apparition d'acide-aminé basique supplémentaires, et donc, à l'apparition potentielle d'un MBCS. Des observations comparables ont été réalisées précédemment chez d'autres virus. Un site de *RNA editing* a été bien caractérisé sur le génome de EBOV, et consiste en un homopolymère de 7 ou 8 uraciles consécutives. Il a été montré qu'au cours d'une infection (en culture cellulaire), plus de 10% de l'antigénome synthétisé est sujet à des insertions d'un nucléotide, au niveau du site homopolymérique (Shabman et al. 2014; Le Mercier and Kolakofsky 2019). TuMV, un virus membre de la famille des *Potyviridae*, possède aussi dans son génome un motif homopolymérique favorable au phénomène d'insertion. La séquence

conservée « GAAAAAA » située dans la région P3 du génome de TuMV provoque l'insertion d'adénines supplémentaires à une fréquence observée de 2% au niveau de ce motif, permettant ainsi l'apparition de nouveaux transcrits et l'expression de la protéine de mouvement P3N-PIPO, essentielle pour le cycle viral de ce virus (Olspert et al. 2015). Il semblerait que ce phénomène soit aussi conservé chez l'ensemble des ARN polymérases. En effet, l'ARN polymérase II de la levure et humaine commettent des insertions à la rencontre de séquence homopolymériques (Gout et al. 2017; Chung et al. 2023). Grâce à notre seconde étude, nous avons pu montrer que le mécanisme de bégaiement ou glissement de la polymérase virale impliqué dans l'évolution des virus influenza aviaires survient précisément après la formation d'un homopolymère 6-A (ou 6-U sur l'ARN génomique viral). Cette séquence « seuil » déclenche l'accumulation d'adénines supplémentaires au niveau de la région codant le site de clivage (ARNc viral) et la duplication de codon basique. Par contre, quand et sur quel type de matrice (au cours de la synthèse de l'ARNc viral ou génomique) ont lieu ces insertions restent encore des questions en suspens, méritant des travaux supplémentaires pour y répondre. Dans le cadre du mécanisme génétique d'acquisition du MBCS par glissement ou bégaiement de la polymérase virale, nous proposons le modèle d'évolution suivant : la séquence d'un virus LPAIV peut acquérir des substitutions aléatoires amenant à la formation d'un motif homopolymérique, constitué de 6A consécutifs (ou 6U sur l'ARNv génomique) au niveau de la région codant le site de clivage HA. A partir de l'obtention d'un tel motif, l'évolution de la séquence HA n'est plus aléatoire et l'homopolymère force la polymérase virale à commettre des erreurs, notamment des insertions (A supplémentaire, duplication de codon basique), amenant à la formation d'un MBCS, et donc l'émergence d'HPAIV.

Conclusion

Ce travail de thèse a été organisé sur deux axes complémentaires. Le premier concernait l'étude de la structure d'ARN prédite conservée au niveau de la région codant le site de clivage de HA par le biais de méthodes bio-informatiques. Le deuxième axe concernait l'élaboration d'un système expérimental afin de tester différentes propriétés du génome viral quant à leur influence sur les évènements d'erreur de la polymérase virale et l'acquisition du MBCS.

Plusieurs éléments montrent que des propriétés du génome viral, telles que certaines structures secondaires et motifs nucléotidiques peuvent influencer les évènements de mutation. De ce fait, caractériser ces facteurs génomiques et mieux comprendre comment ces derniers peuvent guider l'évolution viral vers certaines trajectoires est primordial, notamment pour l'amélioration de notre préparation au risque d'émergence et épidémique.

L'évolution des virus influenza aviaires vers les formes hautement pathogènes constitue un sujet très intéressant, dans lequel l'apparition d'un motif MBCS conditionne de façon majeure la pathogénicité de ces virus. Bien que plusieurs évènements génétiques (insertions/duplication de codon, substitutions et recombinaisons) à l'origine de l'apparition d'un motif de pathogénicité ont été observés, les facteurs influençant leur apparition restaient encore mal connus.

Il semble que l'étude de la structure secondaire conservée au niveau de la région codant le site de clivage HA ne permet pas de distinguer les séquences HA des LPAIV en cours d'évolution vers l'acquisition d'un MBCS de celles qui restent « stables » ou monobasique. Néanmoins, des études complémentaires doivent être poursuivies, notamment au sujet des évènements de recombinaison, aussi impliqués dans le phénomène d'émergence, pour lequel le repliement de l'ARN viral pourrait jouer un rôle plus important. Nous avons pu identifier qu'un motif minimal de six A consécutifs (sur l'ARNc viral) permet d'orienter l'évolution de la séquence codant la région HA1/HA2 vers l'obtention d'un MBCS par insertion. Ce dernier motif est retrouvé chez certains virus de sous-type H5 et H7 dans la nature, dont le site de clivage HA est dibasique. De plus, ces virus sont caractérisés par une faible pathogénicité. Une attention toute particulière pourrait être mise en place pour ces virus possédant ce motif homopolymérique 6A en amont de la région codant le site de clivage HA, qui peuvent être considérés comme à potentiel évolutif élevé vers des formes hautement pathogènes. L'ensemble de ces données ont permis de mieux appréhender le rôle de certaines propriétés du génome viral dans l'émergence des virus influenza aviaires vers les formes hautement pathogènes, et peuvent constituer une base pour des travaux futurs sur l'évolution génétique des virus influenza, et d'autres virus.

Annexes

I. Données supplémentaires, article : “Phylodynamic study of the conserved RNA structure encompassing the hemagglutinin cleavage site encoding region of H5 and H7 low pathogenic avian influenza virus” , Virus Evolution, 2021, 7(2), 1–16

Supplementary data

HA subtype	Major lineage membership	accession number	sequence id
H5	eurasian	GU361206	A_aquatic_bird_Korea_w164_2007_H5N2
H5	eurasian	MK358001	A_duck_Moscow_4182_C_2017_H5N3
H5	eurasian	KX978398	A_mallard_duck_Netherlands_20_2009_H5N3
H5	eurasian	AY995803	A_Mallard_Sweden_49_02_H5N9
H5	eurasian	AB530992	A_mallard_Hokkaido_24_2009_H5N1
H5	eurasian	AB516418	A_northern_pintail_Aomori_385_2008_H5N3
H5	eurasian	CY022629	A_turkey_Italy_1325_2005_H5N2_1-1695
H5	eurasian	MN171442	A_duck_Jiangsu_51665_2015_H5N3
H5	eurasian	KJ162764	A_duck_Taiwan_4374_2013_H5N2_1-1689
H5	eurasian	EF597250	A_duck_Hong_Kong_698_1979_H5N3_1-1695
H5	eurasian	KC785049	A_duck_Taiwan_DV1237_2009_H5N2
H5	eurasian	JX566053	A_mallard_Sweden_101224_2009_H5N9
H5	eurasian	KR265552	A_duck_Japan_11OG1038_2011_H5N2
H5	eurasian	KR265544	A_duck_Japan_11OG1032_2011_H5N2
H5	eurasian	HM849028	A_mallard_PT_29497_10_2006_H5N2
H5	eurasian	GU361205	A_aquatic_bird_Korea_w163_2007_H5N2
H5	eurasian	GU052802	A_duck_Singapore_F119_3_1997_H5N3
H5	eurasian	MK592573	A_duck_Mongolia_194_2011_H5N3
H5	eurasian	GU361198	A_aquatic_bird_Korea_w113_2006_H5N2
H5	eurasian	GU361210	A_aquatic_bird_Korea_w200_2007_H5N2
H5	american	MF046544	A_Canada_goose_Delaware_Bay_601_2016_H5N1
H5	american	CY178623	A_mallard_Wisconsin_487_1983_H5N2
H5	american	AB275404	A_chicken_Ibaraki_6_2005_H5N2
H5	american	GQ923461	A_tundra_swan_Alaska_450959_2006_H5N2
H5	american	EF607865	A_mallard_Maryland_789_2002_H5N2
H5	american	KP674447	A_chicken_Pennsylvania_7659_1985_H5N2
H5	american	CY213584	A_sandpiper_Southcentral_Alaska_16MB01259_2016_H5N2
H5	american	KY130574	A_green_winged_teal_Alaska_143_2011_mixed
H5	american	CY176973	A_mallard_California_1418_2013_H5N6
H5	american	MH546443	A_emperor_goose_Alaska_16_041335_20_2016_H5N2
H5	american	KR271015	A_thick_billed_murre_Greenland_9045_2K_2014_H5N1
H5	american	GQ923485	A_mallard_Washington_456277_2_2006_H5N2
H5	american	GQ923565	A_mallard_Montana_458329_2_2006_H3N3
H5	american	MH546731	A_mallard_Pennsylvania_16_035727_1_2016_H5N6
H5	american	KM368298	A_chicken_Mexico_876_06_2006_H5N2
H5	american	CY166752	A_gadwall_Mississippi_11OS5840_2011_H5N3
H5	american	KM368275	A_chicken_Coahuila_2675_05_2005_H5N2
H5	american	MH546307	A_Muscovy_duck_New_York_09_005059_002_2009_H5N2
H5	american	CY166298	A_mallard_Illinois_11OS5576_2011_H5N3
H5	american	KX351551	A_American_wigeon_California_COL041_2014_H5N2
H7	eurasian	KP416312	A_chicken_Dongguan_1314_2014_H7N9
H7	eurasian	KF258977	A_duck_Jiangxi_25134_2009_H7N6
H7	eurasian	KP417025	A_duck_Jiangxi_5465_2014_H7N3
H7	eurasian	MF630333	A_chicken_Shanghai_S4100_2015_H7N9
H7	eurasian	CY095592	A_shoveler_Italy_2698_3_2006_H7N7
H7	eurasian	MF630277	A_chicken_Jiangsu_S0012_2015_H7N9
H7	eurasian	MF630397	A_chicken_Zhejiang_S4071_2013_H7N9
H7	eurasian	KP415618	A_chicken_Dongguan_1051_2014_H7N9
H7	eurasian	MF147461	A_mallard_duck_Georgia_3_2012_H7N3
H7	eurasian	KP416484	A_silkie_chicken_Dongguan_1002_2014_mixed
H7	eurasian	KU646960	A_Tadorna_tadorna_Belgium_3441_P3_2009_H7N1
H7	eurasian	KP413738	A_chicken_Huzhou_3916_2013_H7N9
H7	eurasian	KP417829	A_chicken_Jiangxi_13530_2014_H7N9
H7	eurasian	KP185949	A_chicken_Henan_F248_2013_H7N9
H7	eurasian	GU052999	A_turkey_Italy_977_1999_H7N1
H7	eurasian	MF630229	A_chicken_Hunan_S0019_2014_H7N9
H7	eurasian	KC609783	A_wild_duck_Korea_CSM42_34_2011_H7N9
H7	eurasian	KP413198	A_chicken_Shenzhen_898_2013_H7N9
H7	eurasian	AY999999	A_Mallard_Sweden_105_02_H7N7
H7	eurasian	KP415720	A_chicken_Dongguan_1424_2014_mixed
H7	american	MK827750	A_Blue_winged_teal_Guatemala_CIP049_H121_09_2014_H7N7
H7	american	MH579249	A_sanderling_Delaware_Bay_294_2006_H7N3
H7	american	CY133649	A_northern_shoveler_Mississippi_11OS145_2011_H7N9
H7	american	CY076261	A_northern_shoveler_Washington_44249_664_2006_H7N3
H7	american	KF445404	A_mallard_MN_AI07_4315_2007_H7N3
H7	american	CY039580	A_northern_shoveler_California_HKWF1026_2007_H7N3
H7	american	CY035922	A_Guinea_fowl_New_York_88291_9_2005_H7N2
H7	american	KU740204	A_pheasant_New_Jersey_26996_2_2014_H7N3
H7	american	KU290204	A_mute_swan_Rhode_Island_A00325125_2008_H7N3
H7	american	MN179383	A_chicken_California_B1801923_2018_H7N8
H7	american	MH134779	A_yellow_billed_pintail_Chile_C14719_2016_H7N3
H7	american	KU290239	A_northern_shoveler_Mississippi_A00602284_2009_Mixed
H7	american	GU051717	A_ruddy_turnstone_Delaware_887_2006_H7N3
H7	american	KY550802	A_cinnamon_teal_California_AH0079011_2016_H7N9
H7	american	CY139937	A_blue_winged_teal_Texas_578588_2002_H7N1
H7	american	KU290156	A_mallard_Oklahoma_A00449455_2009_H7N3
H7	american	EF576989	A_duck_AB_AFLBs8734c16_2002_H7
H7	american	KU289991	A_domestic_duck_West_Virginia_A00140915_2008_H7N3
H7	american	AY240909	A_GuineaFowl_NJ_13246_9_98_H7N2
H7	american	CY141702	A_mallard_Interior_Alaska_10BM08874R0_2010_H7N3

Supplementary Table S1: LPAIV sequence ids of LP.euras and LP.amer evolutionary groups

List of randomly selected LPAIV-taxa that have been used to designed the evolutionary routes leading to non-HPAIV-emergence event. LPAIV sequences are grouped in terms of their major lineage membership (Eurasian or American) and HA-subtype (H5 and H7). Both sequence ids (used in the multiple alignment and phylogenetic trees) and accession numbers are provided. These sequences were checked to be located into distinct clades within each major-lineage in order to better represent the LPAIVs variability.

Supplementary figures legends

Supplementary Figure S1

A) DNA phylogenetic tree of H5-HA nucleotide sequences. The tree-image was generated and mapped by iTOL (Ivica Letunic, Peer Bork, Interactive Tree Of Life (iTOL) v4: recent updates and new developments, *Nucleic Acids Research*, Volume 47, Issue W1, 02 July 2019, Pages W256–W259). Coloured ranges indicate the HA sequence location. Known-HPAIV emergence events are shown by the purple symbols.

B) DNA phylogenetic tree of H5-HA nucleotide sequences (with bootstraps). The tree-image was generated and mapped by iTOL (Ivica Letunic, Peer Bork, Interactive Tree Of Life (iTOL) v4: recent updates and new developments, *Nucleic Acids Research*, Volume 47, Issue W1, 02 July 2019, Pages W256–W259). Coloured ranges indicate the HA sequence location. Bootstraps values at internal nodes are represented by the colour graduation between red (weak-bootstrap value) and green (high bootstrap value). Known-HPAIV emergence events are shown by the purple symbols.

C) DNA phylogenetic time-tree of H5-HA nucleotide sequences. The tree-image was generated and mapped by iTOL (Ivica Letunic, Peer Bork, Interactive Tree Of Life (iTOL) v4: recent updates and new developments, *Nucleic Acids Research*, Volume 47, Issue W1, 02 July 2019, Pages W256–W259). Coloured ranges indicate the HA sequence location. Known-HPAIV emergence events are shown by the purple symbols. Estimated-mean date of every internal node of the tree is shown at the middle-position of the corresponding branch.

Supplementary Figure S2

A) DNA phylogenetic tree of H7-HA nucleotide sequences. The tree-image was generated and mapped by iTOL (Ivica Letunic, Peer Bork, Interactive Tree Of Life (iTOL) v4: recent updates and new developments, *Nucleic Acids Research*, Volume 47, Issue W1, 02 July 2019, Pages W256–W259). Coloured ranges indicate the HA sequence location. Known-HPAIV emergence events are shown by the purple symbols.

B) DNA phylogenetic tree of H7-HA nucleotide sequences (with bootstraps). The tree-image was generated and mapped by iTOL (Ivica Letunic, Peer Bork, Interactive Tree Of Life (iTOL) v4: recent updates and new developments, *Nucleic Acids Research*, Volume 47, Issue W1, 02 July 2019, Pages W256–W259). Coloured ranges indicate the HA sequence location. Bootstraps values at internal nodes are represented by the colour graduation between red (weak-bootstrap value) and green (high bootstrap value). Known-HPAIV emergence events are shown by the purple symbols.

C) DNA phylogenetic time-tree of H7-HA nucleotide sequences. The tree-image was generated and mapped by iTOL (Ivica Letunic, Peer Bork, Interactive Tree Of Life (iTOL) v4: recent updates and new developments, *Nucleic Acids Research*, Volume 47, Issue W1, 02 July 2019, Pages W256–W259). Coloured ranges indicate the HA sequence location. Known-HPAIV emergence events are shown by the purple symbols. Estimated-mean date of every internal node of the tree is shown at the middle-position of the corresponding branch.

Supplementary Figure S3. Evolution of sCL-structure stability based on evolutionary pathways for H5-HA sequences (100 nt-analysis-window centered at the HA cleavage site encoding region)

Pooled-data from the different studied H5-evolutionary pathways is represented. Variation of cSL-stability structure (MFE dG0, kcal/mol, vRNA) based on the genetic distance to the tree-tip (or the final phylogenetic event) is shown. Local Polynomial Regression Fitting (loess) smoothing method was used (confidence interval of 0.95 is shown in gray areas). Groups were designed in terms of the type of evolutionary pathway membership and the genetic mechanism involved in the LPAIV-HPAIV transition. Concerning data from ancestral sequences that are not involved in any HP emergence event, the evolutionary group was divided in terms of the major lineage membership (LP.euras and LP.amer for evolutionary pathways leading to LP event from Eurasian and American lineage, respectively). RNA structure predictions were performed on 100 nucleotides around the cleavage site encoding region ("100 CSR"), on both vRNA (A.) and cRNA (B.).

Supplementary Figure S4. Evolution of sCL-structure stability based on evolutionary pathways for H7-HA sequences (100 nt-analysis-window centered at the HA cleavage site encoding region)

Pooled-data from the different studied H7-evolutionary pathways is represented. Variation of cSL-stability structure (MFE dG0, kcal/mol, vRNA) based on the genetic distance to the tree-tip (or the final phylogenetic event) is shown. Local Polynomial Regression Fitting (loess) smoothing method was used (confidence interval of 0.95 is shown in gray areas). Groups were designed in terms of the type of evolutionary pathway membership and the genetic mechanism involved in the LPAIV-HPAIV transition. Concerning data from ancestral sequences that are not involved in any HP emergence event, the evolutionary group was divided in terms of the major lineage membership (LP.euras and LP.amer for evolutionary pathways leading to LP event from Eurasian and American lineage, respectively). RNA structure predictions were performed on 100 nucleotides around the cleavage site encoding region ("100 CSR"), on both vRNA (A.) and cRNA (B.).

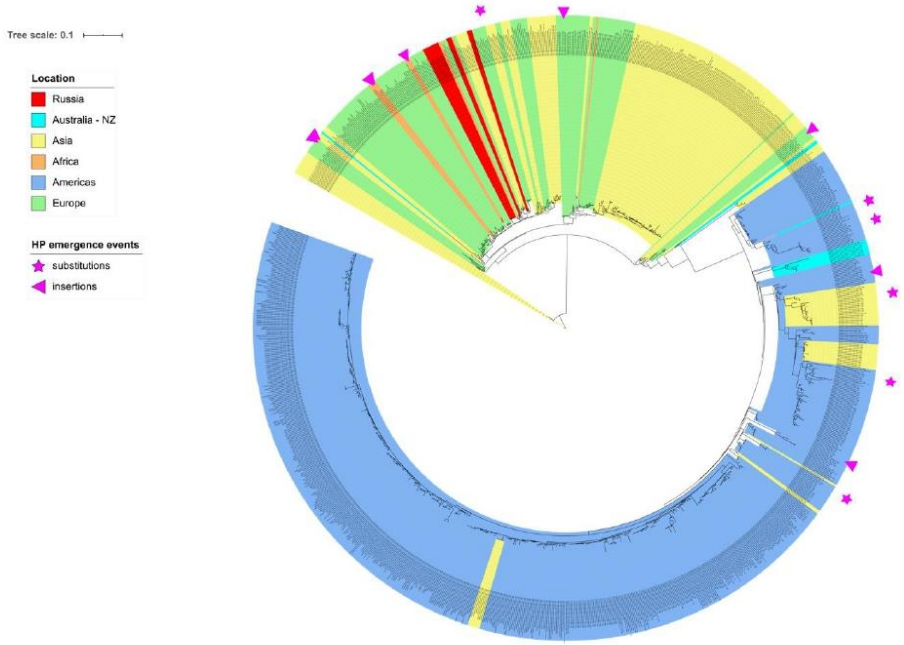
Supplementary Figure S5

Consensus MFE structures of H5-clusters from cRNA sequences that contain ancestors directed only towards HP event via insertions (A.), via substitutions (B.) or both type of emergence (C.) (generated by locarna). Cluster id is indicated next to the corresponding RNA structure.

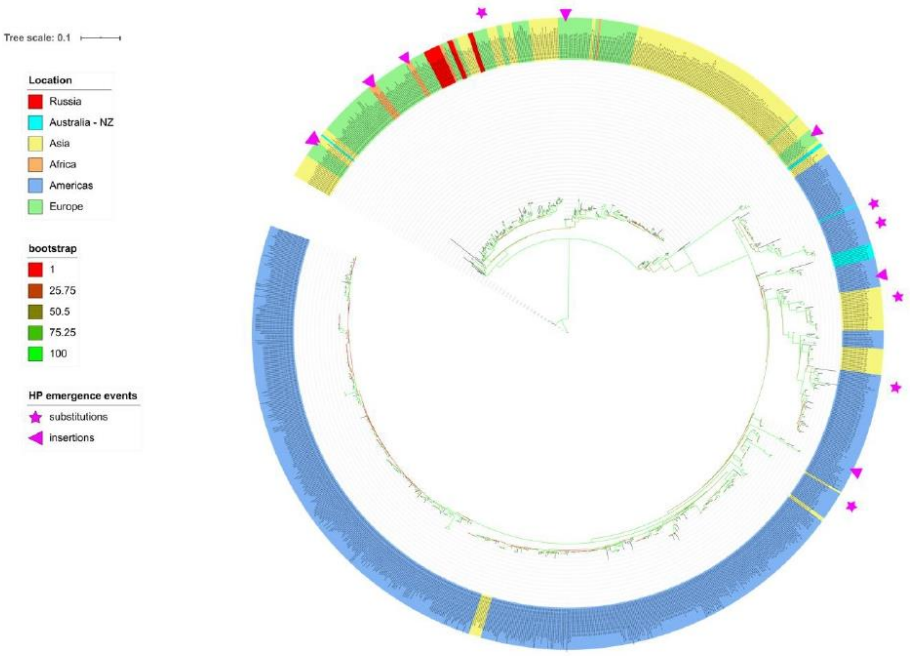
Supplementary Figure S6

Consensus MFE structures of H7-clusters from cRNA sequences that contain ancestors directed only towards HP event via insertions (A.), via substitutions (B.), via recombination (C.), and via insertions or substitutions (D.) (generated by locarna). Cluster id is indicated next to the corresponding RNA structure.

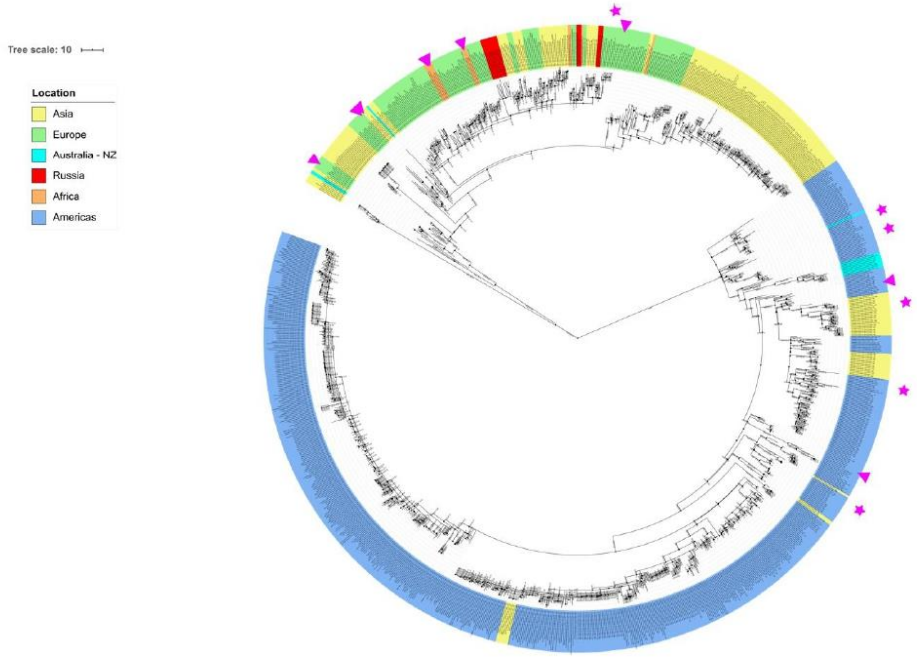
Supplementary figures



Supplementary Figure S1A



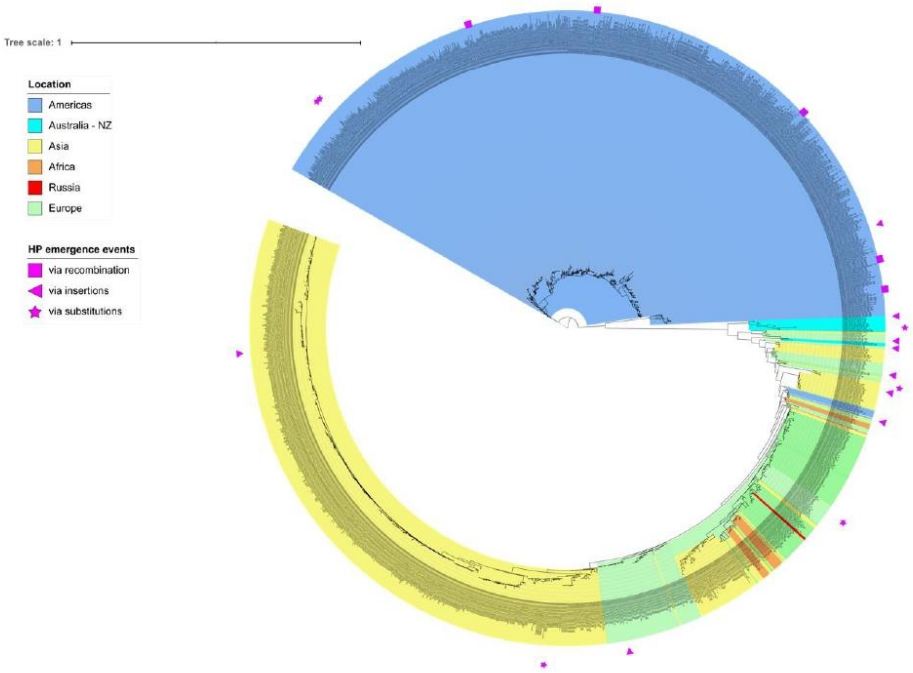
Supplementary Figure S1B



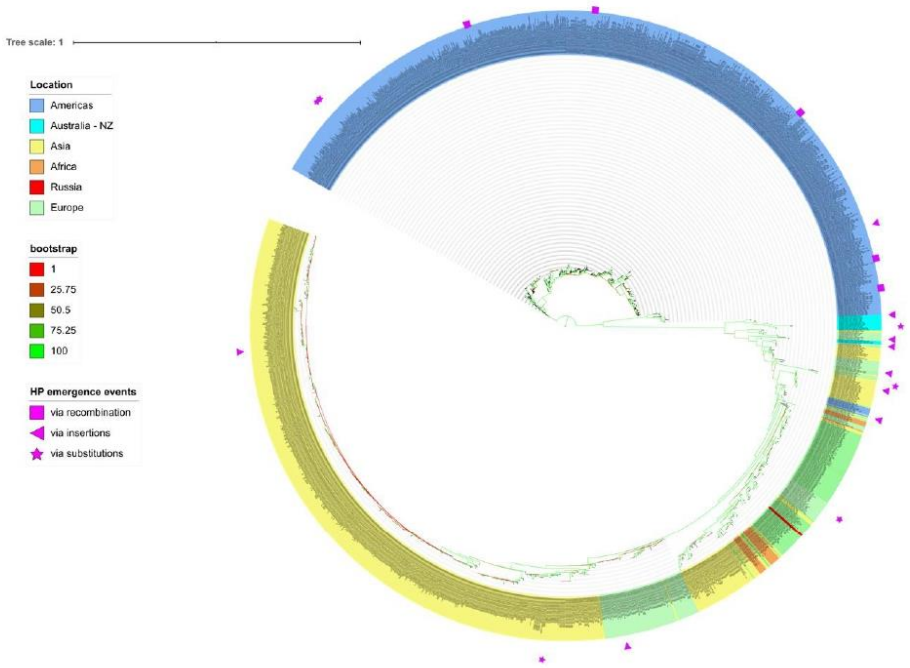
105

106 **Supplementary Figure S1C**

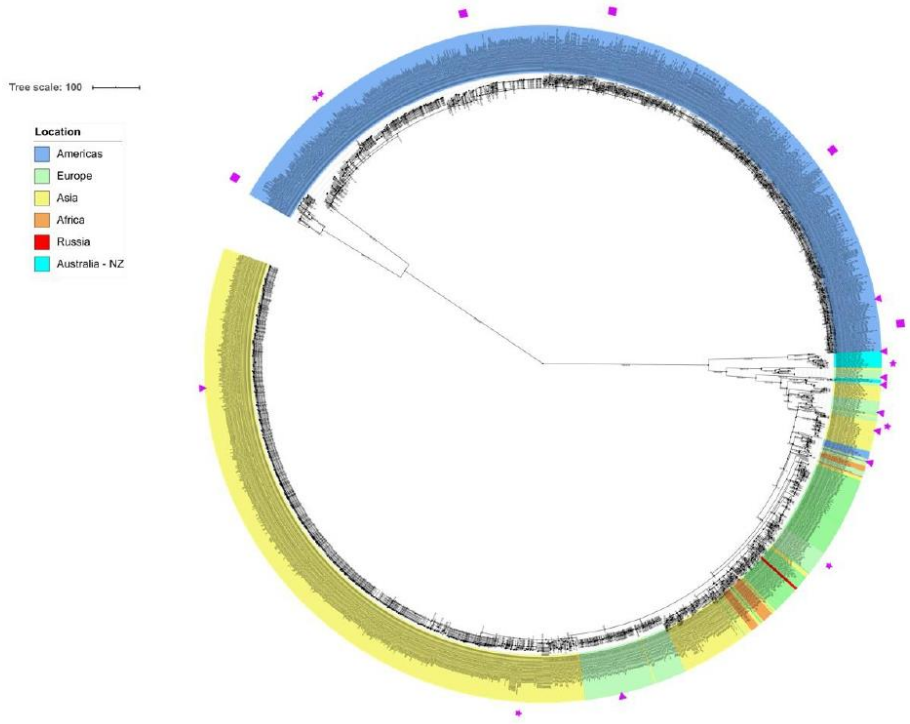
107



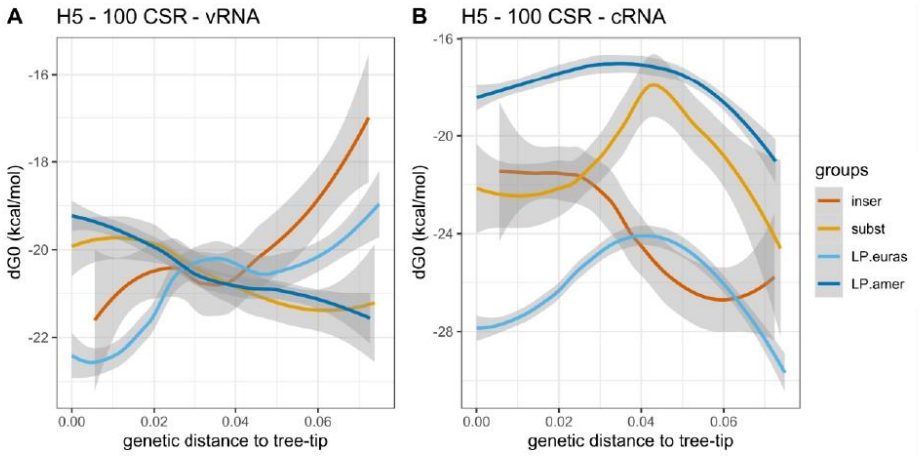
Supplementary Figure S2A



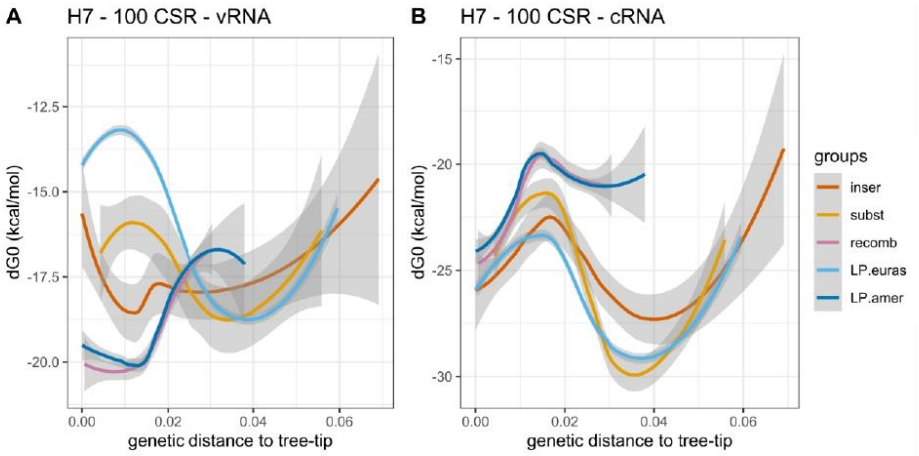
Supplementary Figure S2B



Supplementary Figure S2C



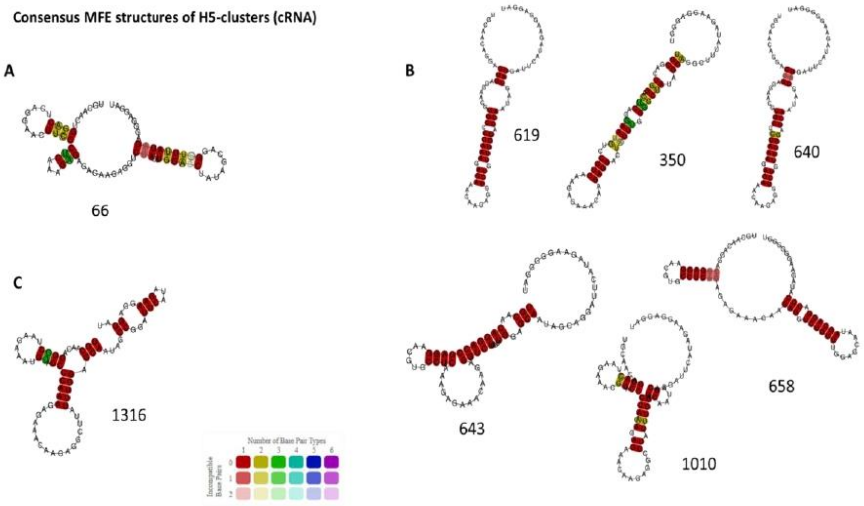
Supplementary Figure S3



120

121 **Supplementary Figure S4**

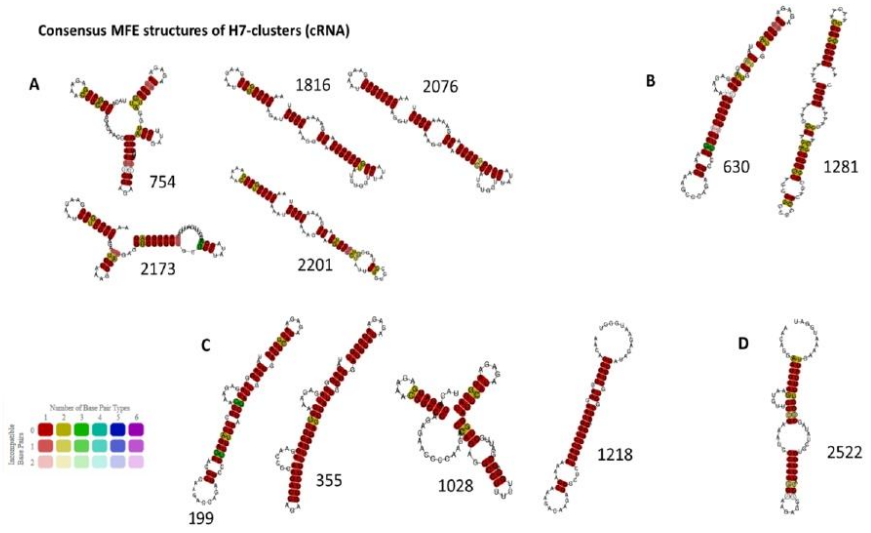
122



123

124 **Supplementary Figure S5**

125



Supplementary Figure S6

II. Données supplémentaires, article (manuscrit) : « Design of engineered viruses to study the influence of genomic properties on the furin-cleavage-site motif acquisition in avian influenza viruses » (manuscrit)

Supplementary data:

Design of engineered viruses to study the influence of genomic properties on the furin-cleavage-site motif acquisition in avian influenza viruses

Gabriel Dupré[1]*, Bertille Pouget[1], Aldair Martín Martínez Pineda[1], Charlotte Foret-Lucas[1], Thomas Figueroa[1], Claire Hoede[2], Christine Gaspin[2], Roland Marquet[3] and Romain Volmer[1]*

[1] Université de Toulouse, ENVT, INRAE, UMR 1225, Toulouse, France.

[2] Plateforme GenoToul BioInfo UR875 MIAT, Castanet-Tolosan, France.

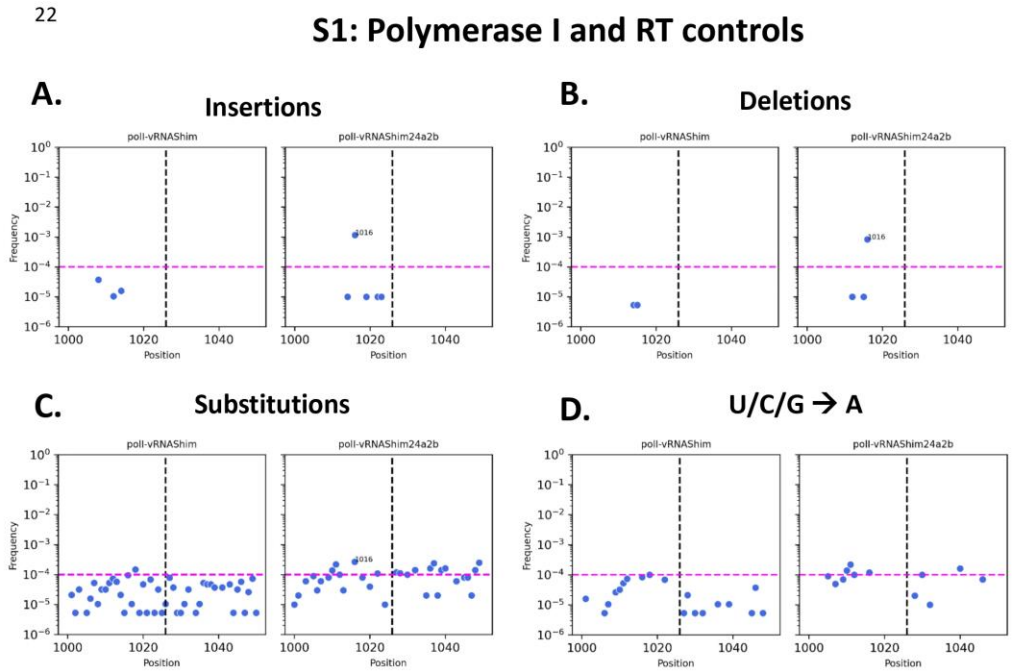
[3] Unistra, IBMC, CNRS, UPR 9002 ARN, Strasbourg, France.

* Corresponding authors:

Dr Romain Volmer, Gabriel Dupré

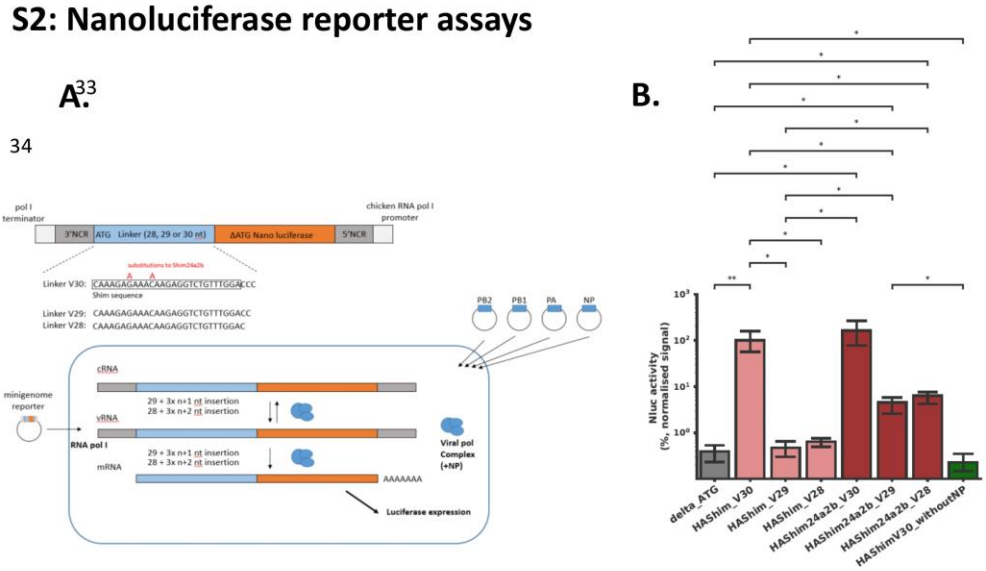
romain.volmer@envt.fr, gabriel.dupre@envt.fr

Keywords: Evolution, avian influenza virus, hemagglutinin, cleavage site motif, highly pathogenic avian influenza virus, emergence, genomic properties, base-sequence motifs, RNA structure, high-throughput-sequencing



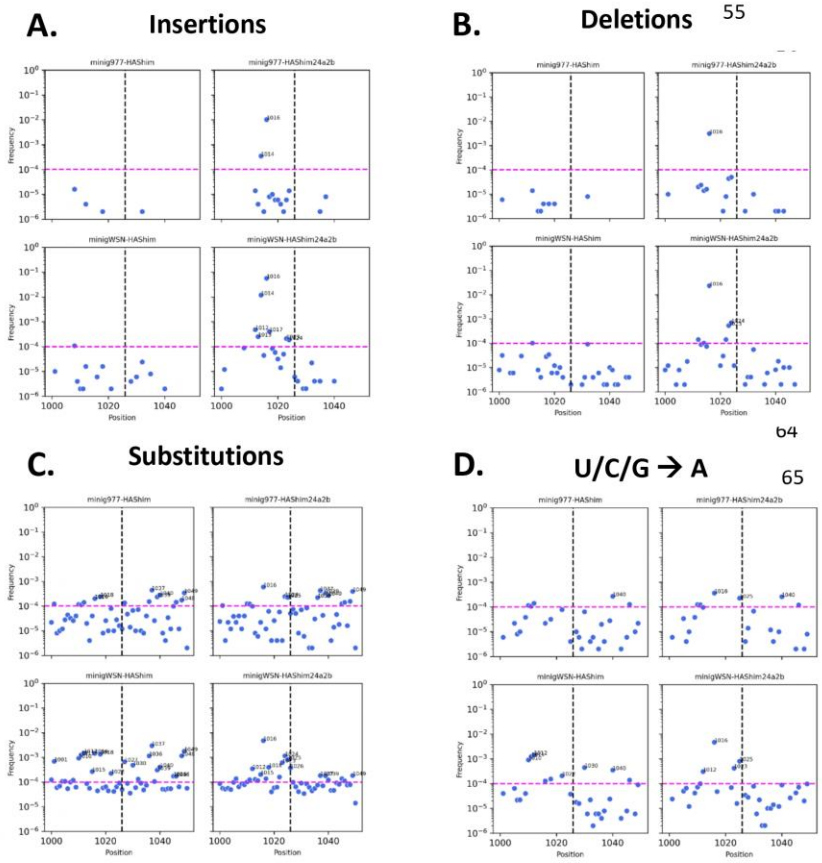
Supplementary Figure 1: Evaluation of mutation errors from cellular polI and reverse transcriptase. Mutation spectra of the HA1/HA2 boundary region from the pseudo viral RNA of HA-Shim or Shim24a2b produced by cellular pol I and reverse-transcriptase. (A-D) All insertions, deletions, substitutions and transitions/transversions to adenine events around the HA1/HA2 boundary region are represented in scatter plots. Frequencies of each event are shown. Nucleotide positions with significantly higher mutation frequency to 10^(-4) are annotated on the plots. The vertical-hashed line represents position 1026, corresponding to the R codon's last nucleotide in the RGLF protein motif at the HA1/HA2 region.

S2: Nanoluciferase reporter assays



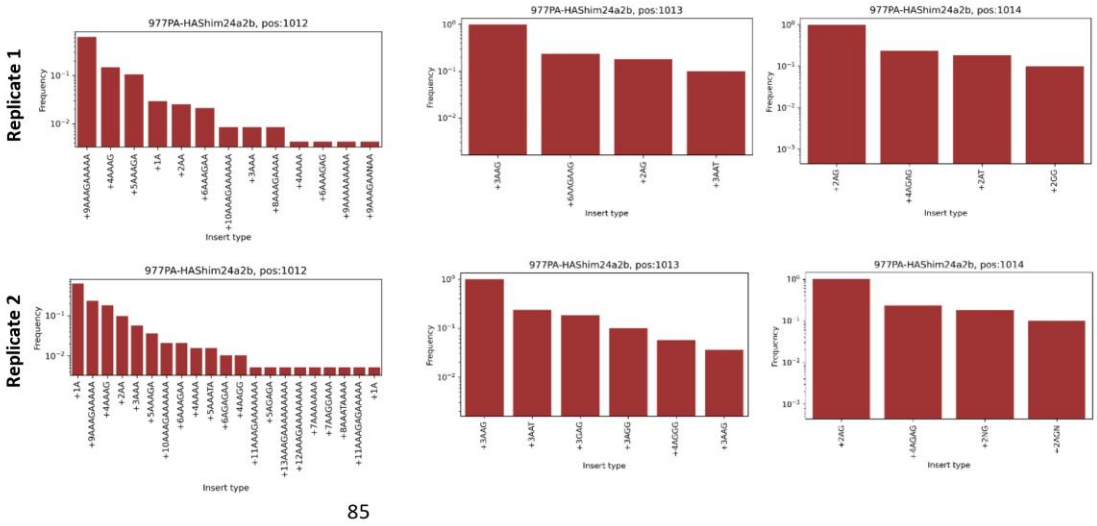
Supplementary Figure 2: Schematic diagram of the Nluc reporter assays. (A) The reporter plasmid contained the chicken RNA polymerase I promoter, mouse RNA polymerase I terminator, H5N8 HA segment-derived NCR, and firefly luciferase gene. Linkers of 28, 29 or 30 nt in length (V28, V29 and V30) that originated from the RNA sequence encoding the HA1/HA2 region of Shim or Shim24a2b were cloned between a start codon and the Nluc gene that lacks its start codon. The point substitutions characterizing the Shim24a2b HA sequence are indicated in red. DF-1 cells were transfected with the reporter plasmid, and a mixture of plasmids expressing the full WSN polymerase complex. Pseudo-vRNA templates are produced by cellular Pol I, and then replicated/transcribed by the WSN polymerase. The transcripts containing V28 or V29 linkers produce mRNAs that are not able to permit Nluc protein expression. When one or two nucleotides are inserted into the V28 and V29 linker region of mRNA, cRNA, and/or vRNA, the Nluc gene can be expressed into a protein. (B) Nluc activities of the different reporters with the different linkers were measured and standardized with the Firefly luciferase internal control. Nluc activity was expressed relative to the reporter plasmid harboring the V30 linker. Representative data from four independent experiments are shown. Standard deviations and statistical significance (using a Mann-Whitney-Wilcoxon test,* and **: p-value <= 0.05 and 0.01 respectively) are shown.

S3: Minigenome assays

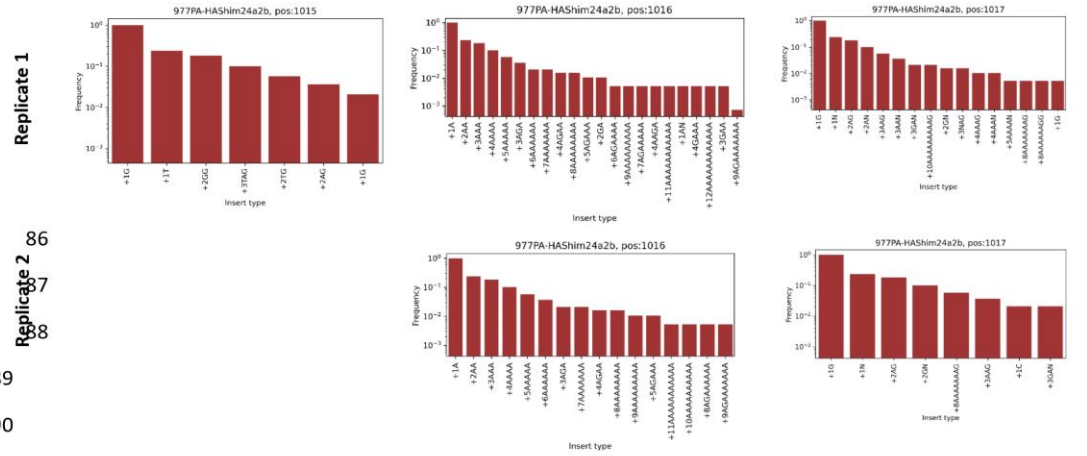


Supplementary Figure 3: Evaluation of mutation spectra in minigenome system. Mutation spectra of the HA1/HA2 boundary region from the HA segment of Shim or Shim24a2b in minigenome systems were analysed. Two different minigenome system were used: the viral polymerase complexe from A/WSN/33 (minigWSN) and 977 (minig977) strains were compared. (A-D) All insertions, deletions, substitutions and transitions/transversions to adenine events around the HA1/HA2 boundary region are represented in scatter plots. Frequencies of each event are shown. Nucleotide positions with significantly higher mutation frequency to 10^(-4) are annotated on the plots. The vertical-hashed line represents position 1026, corresponding to the R codon's last nucleotide in the RGLF protein motif at the HA1/HA2 region.

S4: Insertion diversity in 977PA-HAShim24a2b



85



86

87

88

89

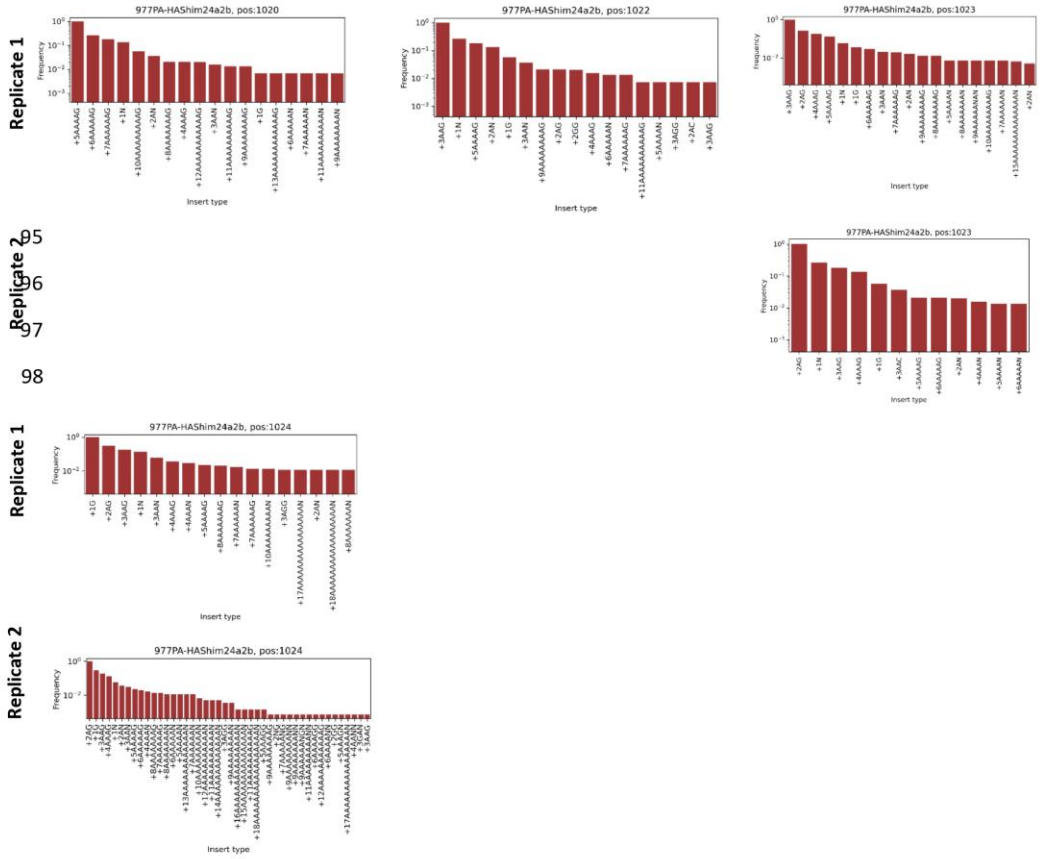
90

91

92

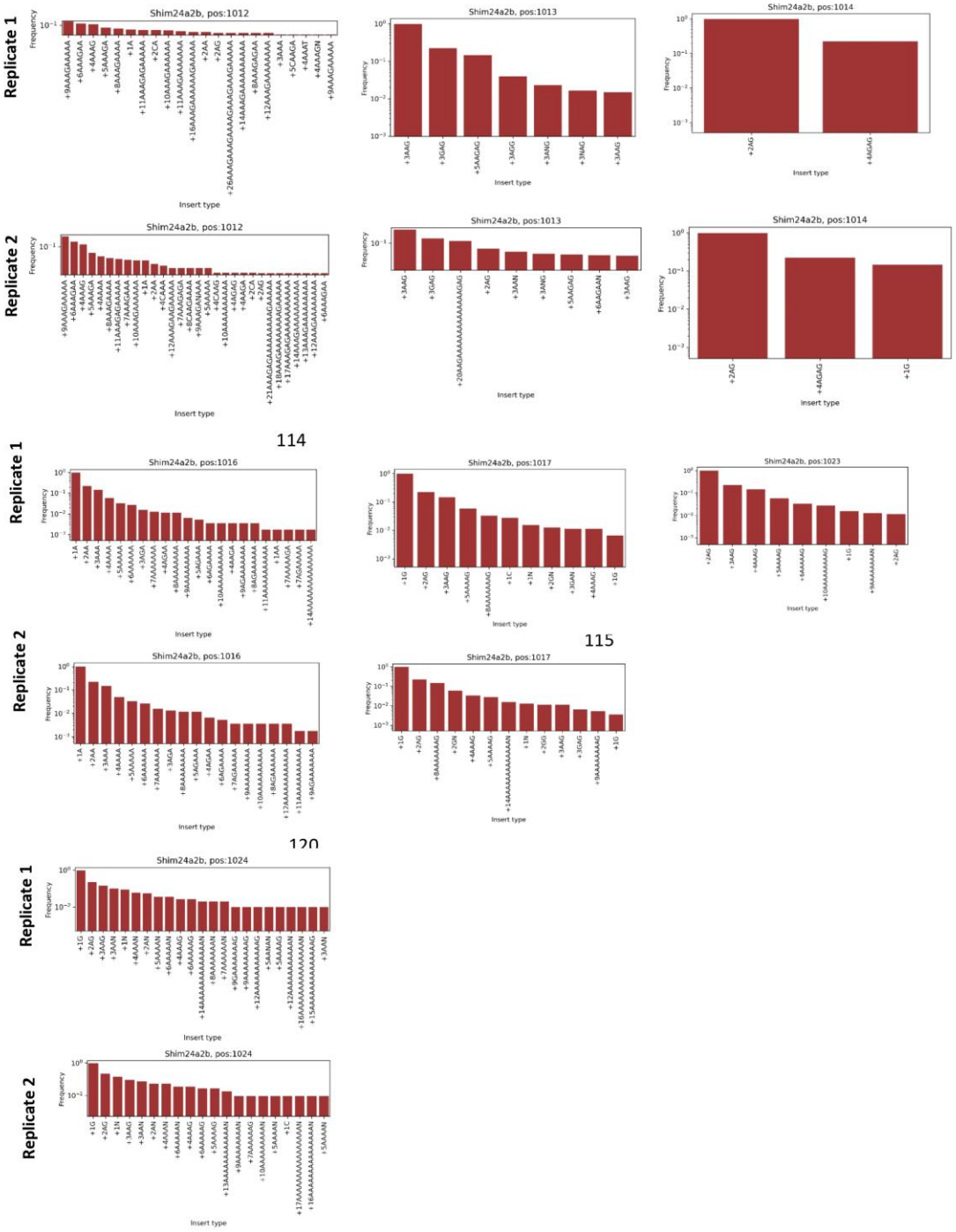
93

94



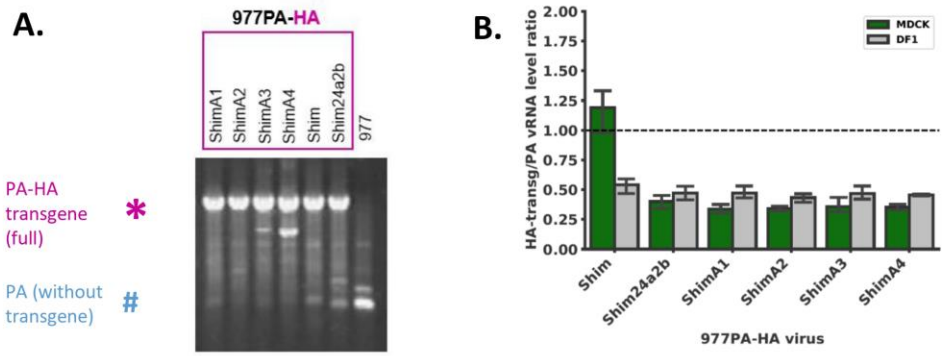
Supplementary Figure 4: Analysis of inserted sequences in 977PA-HAShim24a2b. Observed inserts at significant insertion hotspots of 977PA-HAShim24a2b were analyzed. Each unique insertion type with its apparition frequency is shown. Only events that were found to be significantly more frequent than 10^(-4) according to the binomial test are represented.

S5: Insertion diversity in Shim24a2b



Supplementary Figure 5: Analysis of inserted sequences in Shim24a2b. Observed inserts at significant insertion hotspots of Shim24a2b were analyzed. Each unique insertion type with its apparition frequency is shown. Only events that were found to be significantly more frequent than 10^(-4) according to the binomial test are represented.

S6: HA transgene stability of the 977PA-HAShim viruses



Supplementary Figure 6: HA transgene stability of the 977PA-HAShim viruses. A) Specific viral PA-segment RT-PCR was performed on RNA recovered from viral stocks of passage 1 (used for deep-sequencing analysis) on both MDCK and DF-1 cells. A set of primers that bind at the end pf the PA CDS sequence and the end of HA transgene were used for the PCR. The PCR amplicons were analyzed by agarose gel electrophoresis (data from infected MDCK cells are shown). Expected product sizes for native PA (977) and the 977PA-HA (full-length HA transgene) are indicated. B) Quantification of viral RNA from the unmodified part of the PA segment and the HA transgene was realized by qPCR. Ratios of viral RNA from the HA transgene and the PA sequence were calculated.

III. Revue « La lutte contre l'émergence des virus influenza aviaires hautement pathogènes : un des défis du XXIe siècle ?», Virologie 2022, 26 (5) : 343-354

Virologie 2022, 26 (5) : 343-354

La lutte contre l'émergence des virus influenza aviaires hautement pathogènes : un des défis du XXIe siècle ?

Controlling the emergence of highly pathogenic avian influenza viruses: one of the challenges of the 21st century?

Pierre Bessière
Gabriel Dupré
Romain Volmer

École nationale vétérinaire de Toulouse, Université de Toulouse, ENVT, INRAE, IHAP, UMR 1225, Toulouse, France

Article accepté le 29 septembre 2022

Copyright © 2022 John Libbey Eurotext. Téléchargé par M DAMIEN VITOUR le 28/11/2022.

Résumé. Des millions d'oiseaux d'élevage abattus, des centaines de fous de Bassan morts sur le littoral, des mammifères marins agonisants atteints de troubles neurologiques : ces événements font régulièrement la une de l'actualité. Leur point commun ? Les virus influenza aviaires hautement pathogènes (VIAHP). Les VIAHP sont des virus capables de se répliquer de manière systémique, engendrant des infections asymptomatiques comme une mortalité foudroyante, en fonction de la sensibilité des espèces hôtes. Connus depuis maintenant plusieurs décennies, ces virus ont vu leur circulation particulièrement augmenter ces dernières années et ont été responsables d'épizooties massives sur plusieurs continents. Aux effets dévastateurs qu'ils peuvent causer dans les filières avicoles et dans la faune sauvage, les VIAHP sont également capables de franchir la barrière d'espèce. Mieux les connaître et mieux maîtriser leur diffusion revêt donc plusieurs objectifs : protéger la santé publique, garantir la sécurité alimentaire, préserver la biodiversité ou encore l'économie des filières avicoles. Cet article de synthèse fait un état de l'art des connaissances actuelles sur les VIAHP : de leur épidémiologie aux mécanismes d'émergence en passant par les mesures de lutte.

Mots-clés : influenza, émergence, épizootie, aviaire

Abstract. Millions of farmed birds culled, hundreds of gannets dead on the coast, dying marine mammals suffering from neurological disorders: these events regularly make the headlines. What do they have in common? Highly pathogenic avian influenza viruses (HPAIV). HPAIVs are viruses capable of replicating systemically, causing both asymptomatic infections and devastating mortality, depending on the susceptibility of the host species. Known for several decades now, these viruses have seen their circulation particularly increased in recent years, and have been responsible for massive epizootics on several continents. In addition to the devastating effects they can cause in poultry and wildlife, HPAIVs are also capable of crossing the species barrier. Improving knowledge about these viruses and better control of their spread therefore has several objectives: to protect public health, to guarantee food safety, to preserve biodiversity and the economy of the poultry industry. This article reviews the current state of knowledge on HPAIVs: from their epidemiology to the mechanisms of emergence and control measures.

Key words: influenza, emergence, epizootics, avian

Introduction

Les virus influenza A (VIA) sont des virus enveloppés à ARN segmenté de polarité négative (*figure 1*). En remontant aux origines de tous VIA qui circulent à ce jour, on trouve des ancêtres aviaires. Oiseaux et VIA entretiennent des relations étroites, avec une diversité génétique virale considérable chez les espèces aviaires, notamment chez les oiseaux aquatiques sauvages, qui sont considérés comme le réservoir des VIA. Sur les 18 types d'hémagglutinines (HA) et les 11 types de neuraminidases (NA) recensés,

Correspondance : P. Bessière
<pierre.bessiere@envt.fr>

doi:10.1684/vir.2022.0971

Copyright © 2022 John Libbey Eurotext. Téléchargé par M DAMIEN VITOUR le 28/11/2022.

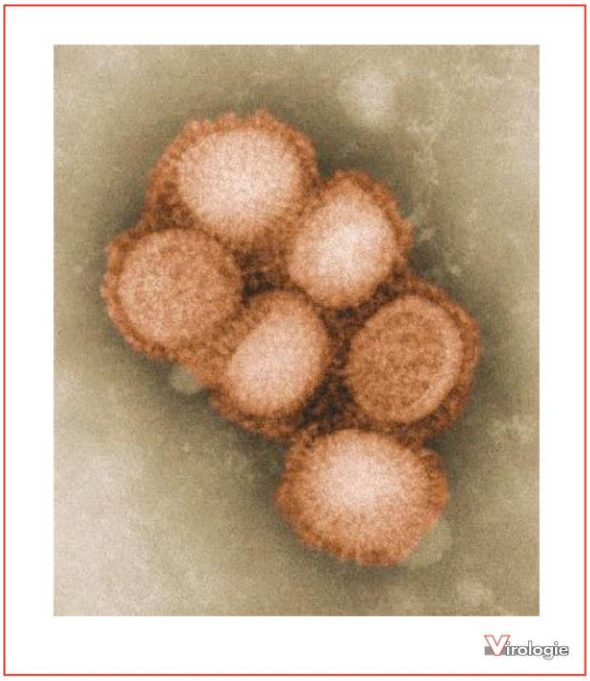


Figure 1. Virus influenza de sous-type H1N1 observés en microscopie électronique. Les protéines de surfaces (hémagglutinine et neuraminidase) sont visibles, enchâssées dans la membrane virale.
C. S. Goldsmith and A. Balish. Photo libre de droits, https://phil.cdc.gov/details.aspx?pid=11214.

respectivement 16 et 9 ont été détectés chez des oiseaux sauvages [1]. Du fait de leur potentiel pandémique, ces virus représentent une menace pour l'homme, avec le risque d'introduction de gènes viraux d'origine aviaire chez des virus humains (comme ce fut par exemple le cas en 1968, lors de la grippe de Hong-Kong et l'émergence d'un virus H3N2 pandémique) ou encore celui de l'acquisition de mutations adaptatives rendant possibles une transmission interhumaine d'un virus aviaire (comme ce fut très probablement le cas en 1918 avec la grippe espagnole) [2]. Au-delà de ces considérations de santé publique, les VIA mettent en péril sécurité alimentaire et bien-être animal. Alors que la viande de poulet est la première source de protéines animales pour l'alimentation humaine, les Galliformes sont particulièrement sensibles à l'infection par les VIA aviaires. Ces derniers sont classés en deux phénotypes : les virus influenza aviaires faiblement pathogènes (VIAFP) et les virus influenza aviaires hautement pathogènes (VIAHP). Si l'infection d'un Galliforme par un VIAFP est souvent asymptomatique, celle par un VIAHP peut entraîner une mortalité foudroyante [3]. Au cours de cet article, seront développées les caractéristiques des VIAFP et des VIAHP, en s'intéressant notamment à l'émergence de ces derniers à partir de précurseurs VIAFP.

Épidémiologie des VIAHP

Au cours de ces dernières années, la circulation des VIAHP de sous-types H5 a connu une augmentation considérable, notamment dans la faune sauvage [4]. Si la première épizootie confirmée à VIAHP H5Ny remonte à 1959, ce n'est qu'en 1996 qu'une des toutes premières souches d'un tel virus à avoir été décrite a été isolée [5]. Elle a été identifiée en Chine et porte le nom A/Goose/Guangdong/1/1996 (H5N1). Depuis, les virus H5Nx de cette lignée n'ont cessé d'évoluer et ont été classés en clades et sous-clades, sur la base d'analyses phylogénétiques et de comparaisons de

Copyright © 2022 John Libbey Eurotext. Téléchargé par M DAMIEN VITOUR le 28/11/2022.

la séquence nucléotidique de HA. Par exemple, dans cette classification, dès lors que la proportion de nucléotides différents excède 1,5 % en comparaison à un clade connu, un nouveau clade (ou sous-clade) peut être créé. On obtient ainsi un code numérique décrivant le chemin évolutif suivi par les virus. Ainsi, le clade 2.1.2 appartient au clade 2.1, lui-même appartenant au clade 2. Si de nombreux clades ont alors été créés, depuis 2010, les virus H5Ny de clade 2.3.4, et particulièrement le sous-type H5N8, sont progressivement devenus prédominants à l'échelle de la planète, en étant transportés sur de longues distances par les oiseaux migrateurs. Des épizooties causées par des virus H5N8 de clade 2.3.4 ont été rapportées par au moins 46 pays, dont la France, durant les hivers 2016-2017 et 2020-2021 [6-8]. Des infections humaines par des virus H5Ny ont déjà eu lieu, avec des virus appartenant à divers clades, mais particulièrement pour le clade 2.3.4, avec l'implication des sous-types H5N1 et H5N6 [9, 10]. La toute première infection humaine par un virus H5N8 a récemment été décrite, suite à une épizootie à VIAHP H5N8 de clade 2.3.4.4b dans un élevage de volaille en Russie [11]. Ainsi, les programmes de surveillance virologique revêtent un intérêt majeur, à la fois pour la santé animale et la santé humaine.

Compte tenu de l'intensification de leur circulation et de l'impact économique absolument considérable que ces virus ont causé en France et en Europe – rappelons qu'entre décembre 2021 et mai 2022, plus de 16 millions d'animaux ont été abattus en France seulement – la possibilité d'utiliser des vaccins, jusque-là jamais envisagée, a été mise sur la table [12].

Classification VIAFP/VIAHP

La classification VIAFP/VIAHP repose sur deux critères : l'index de pathogénicité intraveineuse (IPIV) et la séquence nucléotidique du site de clivage de HA.

Si, suivant l'inoculation intraveineuse de virus (préalablement amplifié sur œuf) chez des poulets âgés de quatre à huit semaines, la mortalité sur dix jours excède les 75 %, le virus sera classifié comme VIAHP. Il en sera de même si l'inoculation intraveineuse chez dix poulets âgés de quatre à huit semaines donne un index de pathogénicité supérieur à 1,2. Cet index est calculé de la façon suivante : un score est attribué à chaque animal à chacun des dix jours post-infection, avec une valeur de 0 si l'animal est cliniquement sain, une valeur de 1 s'il est malade (ne présente qu'un seul signe clinique), une valeur de 2 s'il est sévèrement malade (en présente plusieurs), et une valeur de 3 s'il est mort. Le calcul du score moyen par animal et par jour permet de déterminer l'index [13]. Indépendamment de la pathogénicité, tout virus de sous-type H5 ou H7 ayant un site de clivage dont la séquence est déjà connue pour exister chez des VIAHP, sera lui aussi traité comme un VIAHP [13]. Cette définition se restreint à ces deux sous-types, car, à ce jour, tous les VIAHP à avoir été identifiés étaient soit des H5, soit des H7.

Au cours du cycle viral, l'hémagglutinine sous sa forme immature HA0 est clivée par des protéases en deux sous-unités, HA1 et HA2 : alors seulement le virus devient infectieux. Les VIAFP possèdent, dans la quasi-totalité des cas, un site de clivage monobasique : la plupart du temps une unique arginine, parfois une lysine [14] (*figure 2*). Un nombre limité de protéases *trypsin-like*, sécrétées quasi exclusivement par les cellules des épithéliums digestif et respiratoire, peut reconnaître et cliver ce site. Des protéases bactériennes pourraient également intervenir [15]. La réplication virale est donc limitée aux appareils digestifs et respiratoires, ces protéases n'étant pas exprimées par tous les types cellulaires [16] (*figure 3*). Chez le canard, la multiplication virale peut être particulièrement intense dans l'intestin, avec de fortes charges virales excrétées dans les fèces [17]. Les signes cliniques causés par l'infection varient selon la souche virale, l'espèce hôte, l'âge, la présence d'autres microorganismes et les paramètres environnementaux [3, 18]. La symptomatologie est généralement modérée et peut-être inexistante dans certains cas, notamment chez les canards et les autres espèces réservoirs [19]. Si un VIAFP doit être cultivé *in vitro*, la présence de trypsine dans le milieu de culture est nécessaire à l'obtention de cycles de réplication multiples et de particules infectieuses [20]. Des variations de tropisme sont néanmoins possibles : sans qu'il y ait de réplication systémique, certains VIAFP peuvent se répliquer en dehors des appareils respiratoire et digestif, par exemple dans les reins [21] ou dans l'oviducte [22] des poulets.

Le site de clivage des VIAHP est, lui, polybasique, avec plusieurs arginines ou lysines sur la séquence protéique (*figure 2*). Ce site de clivage polybasique est reconnu par des protéases comme les furines ou la proprotéine convertase 6 (PC6), qui sont ubiquitaires et permettent ainsi au virus de se répliquer non plus uniquement dans les appareils respiratoire et digestif, mais de manière systémique [23, 24] (*figure 3*). Ce clivage est d'autant plus efficace qu'il a lieu de manière intracellulaire [25], pendant l'étape d'assemblage du cycle viral [26]. Ainsi, les particules virales sont déjà infectieuses lorsqu'elles bourgeonnent. Les signes cliniques varient ici aussi selon l'espèce, l'âge, la souche virale et les paramètres environnementaux. On peut aller d'une mortalité brutale sans signes annonciateurs, pouvant atteindre les 100 % en 24 h chez les espèces les plus sensibles, à une symptomatologie plus modérée, faisant intervenir plusieurs présentations cliniques. Ces dernières peuvent inclure des signes cliniques respiratoires

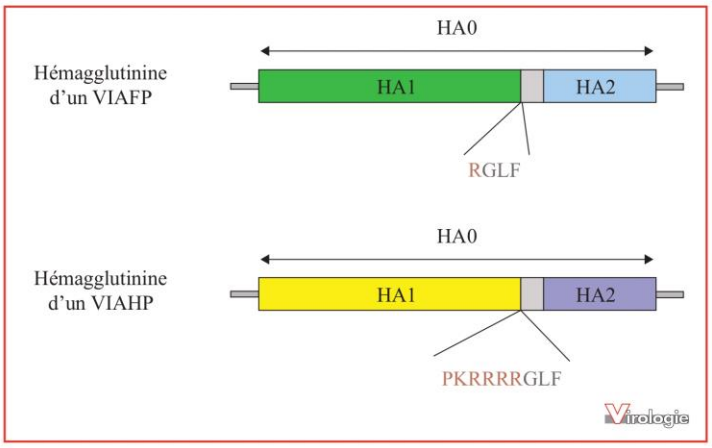


Figure 2. Hémagglutinines des virus influenza aviaires faiblement et hautement pathogènes. Le clivage protéolytique de la molécule précurseur HA0 en deux sous-unités HA1 et HA2 se fait au niveau d'un site de clivage qui est monobasique pour les VIAFP et polybasique pour les VIAHP.

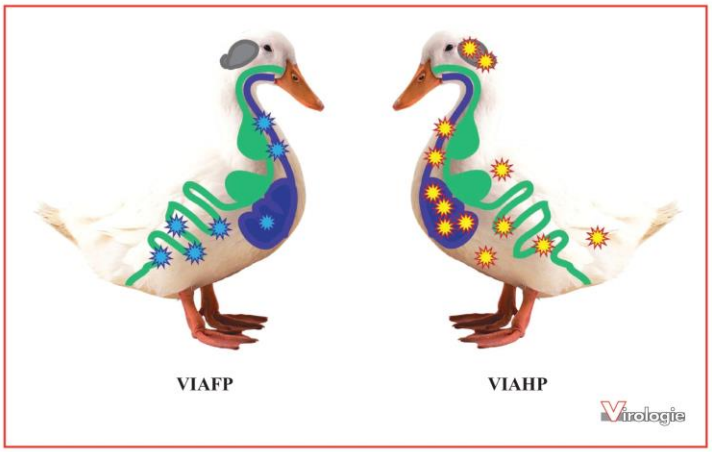


Figure 3. Réplication des virus influenza aviaires faiblement et hautement pathogènes chez les oiseaux. Les étoiles indiquent les sites de réplication virale. La réplication des VIAFP est généralement limitée aux appareils respiratoires et digestifs, tandis que les VIAHP peuvent se répliquer de manière systémique.

(écoulements nasaux, toux, dyspnée), neurologiques (torticolis, ataxie), digestifs (diarrhée, anorexie) ou plus généraux (abattement, diminution des vocalises). Aucun d'entre eux n'est pathognomonique. Une morbidité élevée s'accompagne généralement d'une forte mortalité [13]. Les canards peuvent, selon la souche virale, ne présenter aucun signe clinique, comme présenter une mortalité élevée. Prenons l'exemple de deux virus parmi ceux responsables des épizooties ayant frappé la France au cours des hivers 2015-2016 et 2016-2017, de sous-types respectivement H5N9 et H5N8 : l'infection de canards Pékin par le premier passe cliniquement inaperçue [27], tandis que celle par le deuxième cause d'importants troubles neurologiques, associés à une mortalité forte [28, 29].

Copyright © 2022 John Libbey Eurotext. Téléchargé par M DAMIEN VITOUR le 28/11/2022.

Copyright © 2022 John Libbey Eurotext. Téléchargé par M DAMIEN VITOUR le 28/11/2022.

Cette classification a toutefois ses limites, car certains virus atypiques existent et en voici quelques exemples. Deux virus appartenant au sous-type H10, H10N4 et H10N5, ont été isolés et remplissaient le critère concernant la pathogénicité, avec un IPIV supérieur à 1,2, malgré un site de clivage monobasique. Toutefois, inoculés par voie intranasale, ils n'entraînaient pas de mortalité et ont été classifiés comme VIAFP [30]. De même, un H4N2 isolé chez des cailles possédait un site de clivage polybasique, mais un IPIV nul, le classant comme VIAFP [31]. L'existence de sous-types H5 avec un site de clivage polybasique, mais un IPIV faible, a également été rapportée [32]. Comme dit précédemment, les phénomènes de recombinaison au niveau de la séquence du site de clivage peuvent affecter la pathogénicité du virus : l'insertion de courts fragments à cet endroit peut résulter en l'acquisition d'un site de clivage polybasique, avec une séquence atypique, comme cela a été le cas au Chili en 2002 [33] et au Canada en 2004 [34]. Enfin, une récente étude a montré que certains virus HxN6 avaient, tout en gardant un site de clivage monobasique et un IPIV bas, la capacité de se répliquer de manière systémique *in vivo* et en l'absence de trypsine *in vitro*, grâce à la neuraminidase N6 et un site de clivage atypique, reconnu par la thrombine [35].

Obtention d'un site de clivage polybasique

Les VIAHP ne forment pas de groupes phylogénétiques distincts, mais ont systématiquement pour ancêtres communs des VIAFP [36-38]. L'étude épidémiologique des épizooties à VIAHP documentées au cours de ces dernières décennies semble indiquer que les VIAHP émergent suite à la réplication et à la transmission d'un VIAFP au sein des élevages de Galliformes [39]. En 1994, au Mexique, un VIAFP de sous type H5N2 causant des symptômes respiratoires modérés a été isolé et des analyses sérologiques ont suggéré que le virus circulait au moins depuis l'année passée. Dans certains élevages, une augmentation sévère de la virulence du virus a été observée et les prélèvements réalisés ont révélé la présence d'un VIAHP H5N2. Les analyses phylogénétiques basées sur la séquence de HA ont montré que ce VIAHP avait émergé à partir du VIAFP au cours de deux événements distincts, à Puebla en 1993 et à Queretaro en 1994, et ont également indiqué que le VIAFP avait probablement été introduit dans le pays par les oiseaux migrateurs [40]. Ainsi, l'hypothèse qu'un VIAHP pouvait émerger suite à l'introduction d'un VIAFP au sein des galliformes était prouvée. D'autres scénarios similaires ont suivi : en octobre 2003, au Pakistan, un VIAFP H7N3 a été isolé chez des poules pondeuses et en novembre de la même année, un VIAHP H7N3 a émergé et s'est répandu en quelques semaines à travers le pays, jusqu'au contrôle de l'épizootie en juin 2004. Les analyses génétiques ont montré que le VIAHP avait émergé chez le poulet suite à l'insertion de plusieurs acides aminés basiques dans le site de clivage de HA du VIAFP [41]. On peut également citer les émergences des VIAHP H7N1 en Italie en 1999 [37, 42] et H7N3 au Canada en 2004 [34], pour lesquelles le VIAFP précurseur était connu. Néanmoins, la circulation d'un VIAFP dans les élevages de galliformes préalablement à l'émergence d'un VIAHP est un paramètre inconnu dans la plupart des épizooties [43]. Une raison pouvant expliquer cela est que la circulation des VIAFP a fréquemment lieu à bas bruit, ne causant que peu à pas de signes cliniques, les virus passant ainsi souvent inaperçus. Dans certains cas, l'implication des oiseaux sauvages dans l'introduction d'un VIAFP chez les Galliformes a été démontrée, comme en 1979, au Royaume-Uni, où des oiseaux sauvages ont été contraints de venir s'alimenter à proximité d'élevages de dindes du fait d'un hiver particulièrement rigoureux [44]. Dans la majorité des cas cependant, le rôle joué par la faune sauvage reste inconnu [43].

Plusieurs mécanismes génétiques sont responsables de l'apparition du motif polybasique au niveau du site de clivage de HA (*figure 4*). Des insertions de nucléotides, des substitutions et des événements de recombinaison génétique avec des gènes cellulaires et viraux ont été observés [34, 39, 45-47]. Comme chez d'autres virus, des phénomènes de glissement ou de bégaiement des polymérases virales (*polymerase slippage* ou *stuttering*) participeraient à l'évolution du génome des VIAFP en provoquant l'insertion de nucléotides au niveau de la région HA1/HA2. Des duplications de codons basiques ont été également observées [45, 47, 48]. Les matrices à ARN composées de régions riches en répétitions d'adénosines (A) ou uridines (U) sont favorables à ces phénomènes, et décrivent également des structures secondaires d'ARN de type épingle à cheveux ou tige-boucle, pour lesquelles a été proposée l'induction d'un ralentissement de la progression de la polymérase virale lors de la rencontre de ces structures. Une tige-boucle d'ARN très conservée a été identifiée au niveau de la région HA1/HA2, et pour laquelle un rôle dans l'acquisition d'un site de clivage polybasique a été attribué [47, 49]. Ce mécanisme impliquant l'environnement nucléotidique de la séquence codant le site de clivage de HA chez les AIV pourrait expliquer une grande partie des émergences VIAHP connues jusqu'à ce jour, comme la souche HPAIV H5N2 qui est apparue au Mexique en 1995 [50, 51]. Pour certaines émergences, des acquisitions de site de clivage HA polybasique combinant plusieurs mécanismes ont été observées, avec des substitutions ayant précédé des événements d'insertions de nucléotides, suggérant un modèle d'évolution « progressif » possible [45, 50-53]. Par ailleurs,

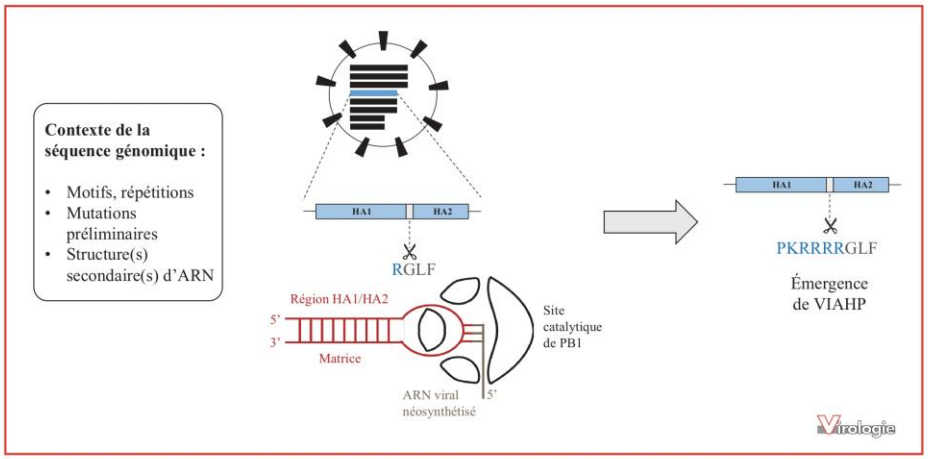


Figure 4. Prédisposition génétique à l'émergence de VIAHP. Schéma conceptuel représentant les différentes hypothèses concernant les facteurs responsables de l'acquisition d'un site de clivage HA polybasique. Des éléments génétiques locaux présents sur le segment HA ou agissant à distance (localisés sur un autre segment non HA) tels que la composition en base, des motifs ou certaines structures secondaires d'ARN pourraient influencer les erreurs de la polymérase virale, potentiellement génératrices de l'acquisition d'un site de clivage HA polybasique.

des recombinaisons entre des gènes cellulaires (ARN ribosomique 28S) ou d'autres segments viraux (NP et M) et le segment HA ont été décrites chez un nombre conséquent de virus de sous-type H7 [34, 39, 54]. Du fait de la présence de structure de type tige-boucle conservée au niveau du site de clivage de HA, un mécanisme de saut de polymérase et changement de matrice au niveau de cette structure a été proposé, à l'image de ce qui peut être observé chez certains rétrovirus [55, 56].

Malgré d'importants efforts de recherche depuis le début des années 90, il est encore aujourd'hui difficile de mettre en lumière un ou des déterminants génétiques précis à l'émergence vers les VIAHP. Une équipe japonaise a proposé que la taille de la boucle flexible de la structure tige-boucle présente au niveau de la région HA1/HA2 serait déterminante dans les événements d'insertions [47]. Néanmoins, ce modèle implique la reformation d'une boucle d'ARN d'une longueur de 6 à 8 nucléotides de long au sein du site catalytique du complexe polymérase (sous-unité PB1), ce qui est pour l'instant peu compatible avec les données structurales actuelles du complexe de réplication. En effet, la matrice correspondant à l'ARN viral semblerait être sous forme simple d'une longueur de 25-27 bases au sein du complexe polymérase [57]. Dans un autre modèle, des interactions partielles entre les brins entrant et sortant de la matrice virale au cours de la réplication du génome viral ont été suggérées, impliquant également la structure secondaire conservée [49, 58]. Aucun lien n'a été clairement identifié entre certaines caractéristiques de la structure tige-boucle et les événements d'insertions ou de substitutions au niveau de HA1/HA2. Cependant, une structure tige boucle particulière a été identifiée chez la majorité des ancêtres des émergences américaines par recombinaison, mais son importance biologique reste encore à être prouvée expérimentalement [58]. Par ailleurs, il a été montré que chez la majorité des segments HA dérivant de VIAHP ayant évolué par recombinaison, les séquences insérées exogènes en amont de la région codant le site de clivage constituent la cible de petits ARN nucléolaires : snoRNA, impliqués dans les régulations épigénétiques des gènes de l'hôte [54]. Le mécanisme exact et les facteurs expliquant pourquoi des recombinaisons sont observées uniquement chez les virus de sous-types H7 restent encore énigmatiques. La piste d'un lien potentiel entre la liaison spécifique de certains snoRNA et la présence de la tige boucle d'ARN conservée au niveau de la région codant le site de clivage de HA serait intéressante à investiguer.

Hormis le rôle que peut avoir l'environnement génomique correspondant à la région HA1/HA2, il est fortement probable que l'émergence vers les formes hautement pathogènes puisse impliquer d'autres mutations à distance de la zone codant le site de clivage, et également touchant les autres segments. Une analyse de toutes les séquences provenant des huit segments de virus de sous-types H5 et H7 disponibles sur NCBI a montré que l'apparition d'un site de clivage polybasique était significativement corrélée

Copyright © 2022 John Libbey Eurotext. Téléchargé par M DAMIEN VITOUR le 28/11/2022.

à l'apparition d'autres mutations, présentes sur le segment HA, mais également sur les segments codant PB1 et PB2. Si cette analyse permettait d'établir si ces mutations étaient apparues avant ou après l'émergence du VIAHP, certaines d'entre elles auraient pu être répertoriées comme favorables à l'évolution vers le phénotype hautement pathogène [59].

Quasi-espèces virales

Du fait du fort taux d'erreur de la polymérase virale et des phénomènes de réassortiment, les virus influenza sont d'une très grande variabilité. Pour un sous-type donné, comme H5N8, il existe un très large panel de souches différentes. Une souche donnée n'existe pas en tant qu'entité individuelle, mais sous la forme d'un nuage de variants, intimement proches sur le plan génétique, avec de légères variations : on parle de quasi-espèce virale [60]. En infectant un tapis cellulaire ou un animal avec un seul variant, au fur et à mesure des cycles de réplication, des mutations vont inexorablement apparaître et la population virale qui se forme perd progressivement son homogénéité génomique. Ce concept est particulièrement connu pour le VIH [61], mais a été rapporté pour de nombreux autres virus à ARN, comme le virus de l'hépatite C [62]. Il confère au virus des avantages pour l'adaptation à l'hôte, la virulence, la transmission ou l'échappement à la réponse immunitaire [63].

Lorsque, par le biais des mécanismes décrits précédemment, un VIAFP acquiert un site de clivage polybasique, le VIAHP nouvellement formé ne représente vraisemblablement qu'une fraction minime de la quasi-espèce. Pour que la modification génétique constitue un succès évolutif, le VIAHP nouvellement formé doit surpasser le VIAFP pour devenir un variant prédominant et ainsi être transmis à d'autres individus. En dépit de l'avantage sélectif majeur que possède le VIAHP, s'il ne parvient pas à cela, il ne peut y avoir émergence : la sélection positive des variants minoritaires de phénotype HP à partir du nuage de la quasi-espèce de variants FP est également un des mécanismes d'émergence. L'émergence réussie d'un VIAHP passerait donc par deux étapes successives : l'acquisition d'un site de clivage polybasique, puis la capacité à devenir prédominant au sein de la quasi-espèce virale (*figure 5*). En réalisant des infections simultanées avec un VIAHP et un précurseur VIAFP à des ratios différents chez le poulet et chez le canard, des travaux récents ont expérimentalement montré qu'un phénomène d'interférence virale existait et dépendait de l'espèce hôte et vraisemblablement de la souche virale [64].

La meilleure compréhension des phénomènes amenant à l'apparition d'un VIAHP ou à la sélection d'un VIAHP qui coexistait avec un VIAFP implique la mise au point de différents modèles, qui vont être détaillés par la suite.

Analyse des phénomènes d'émergence

Plusieurs modèles d'émergence ont été utilisés, dans le but de faire apparaître un VIAHP à partir d'un précurseur VIAFP, en faisant intervenir à chaque fois des passages viraux successifs *in vitro*, *in ovo* et *in vivo*. La multiplication des passages permet d'avoir un très grand nombre de cycles viraux et de sélectionner progressivement les variants avec le meilleur fitness, comme cela peut être le cas lorsqu'une infection progresse dans un élevage. Le passage sur fibroblastes de poulet puis sur poulets d'un virus H5N2 a résulté en l'obtention de la sélection de variants hautement pathogènes [65]. Des passages successifs sur cellules embryonnaires de poulet d'un virus H7N7 isolé sur un mammifère marin a permis l'obtention d'un site de clivage polybasique et d'un phénotype hautement pathogène [66]. Un des inconvénients des modèles cellulaires est l'absence de système immunitaire, donc une pression de sélection très limitée. Les œufs embryonnés permettent de pallier en partie ce problème, bien que le système immunitaire d'un embryon, même proche de l'éclosion, ne soit pas mature [67]. Ainsi, des VIAHP ont été obtenus par passages successifs sur œufs embryonnés de 14 jours de différents sous-types H5 de phénotype faiblement pathogène [68]. Les modèles *in vivo* ont l'avantage de faire intervenir un organisme entier, avec un système immunitaire mature, dès lors que les animaux sont suffisamment âgés. La pression sélective au sein de la quasi-espèce virale permet de reproduire ce qui se passe lorsqu'un VIAFP est naturellement transmis dans un élevage de gallinacées [69]. Des VIAFP H5N2 et H5N3 sont devenus VIAHP par passages successifs chez des poulets [70, 71]. L'augmentation de la pathogénicité virale par passages successifs chez le poulet a également été observée pour d'autres virus, comme le virus de la maladie de Newcastle [72].

La quasi-totalité des émergences de VIAHP à partir d'un précurseur VIAFP rapportées ces dernières années a eu lieu dans des élevages commerciaux de volailles. Seules deux émergences ont été décrites dans des basses-cours [73] et à ce jour, une seule émergence dans la faune sauvage a été rapportée : celle d'un H5N3, en 1961, chez des sternes sauvages, proche des côtes de l'Afrique du Sud [74]. Les animaux élevés dans les élevages commerciaux étant abattus relativement jeunes, ils n'ont généralement pas le temps de mettre en place une immunité adaptative et d'être infectés une seconde fois, ce qui n'est pas le cas des volailles élevées dans les basses-cours.

Les modèles faisant intervenir les passages successifs ne prennent toutefois pas en compte d'autres limites.

Copyright © 2022 John Libbey Eurotext. Téléchargé par M DAMIEN VITOUR le 28/11/2022.

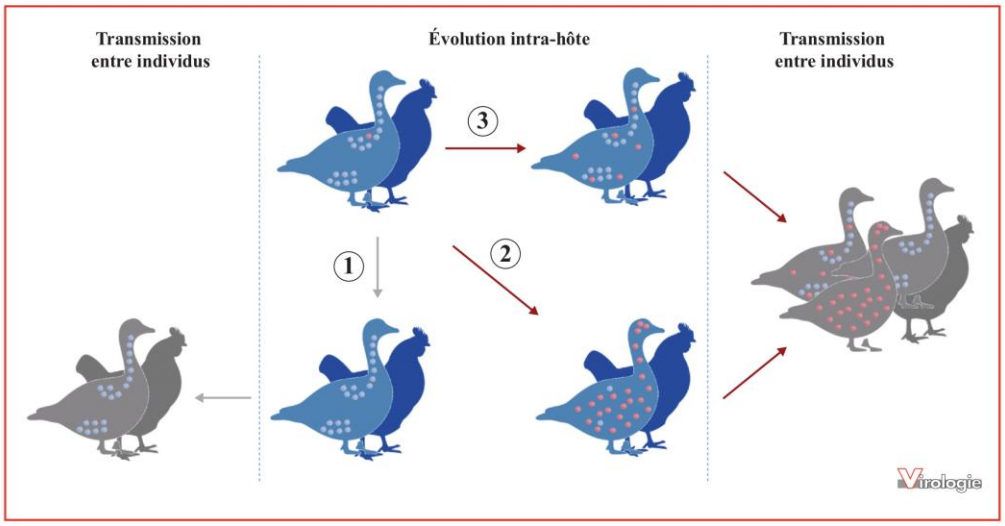


Figure 5. Scénarios possibles suite à l'émergence d'un VIAHP au sein d'un individu. Lorsqu'un VIAHP (rouge) émerge à partir d'un VIAFP précurseur (bleu) au sein d'un hôte, il pourrait disparaître (1), du fait de facteurs dépendant de l'hôte comme de la souche virale, et ne jamais être transmis, ou bien parvenir à devenir un variant (plus ou moins majoritaire) de la quasi-espèce (2 et 3) et être transmis à d'autres individus.

Premièrement, il est possible que l'inoculum de départ, n'étant supposé contenir que le VIAFP, contienne quelques VIAHP, ce qui fait intervenir le concept de quasi-espèce virale. Des sous-populations virales, très différentes en termes de pathogénicité, pourraient coexister et être sélectionnées au cours des passages, sans qu'il y ait d'émergence réelle par l'introduction de mutations dans la séquence du site de clivage. Grâce au séquençage haut débit, la présence de VIAHP en quantité infime a par exemple été suspectée dans certains isolats de VIAFP obtenus lors des épizooties à H7N3 au Canada en 2004 [34], à H7N1 en 1999 en Italie [75] et à H7N7 en Allemagne en 2015 [76]. Deuxièmement, ils ne prennent pas en compte la possibilité qu'un VIAHP puisse apparaître, mais ne pas parvenir à être transmis aux autres animaux, par perte de la compétition avec le VIAFP parental.

S'il existe une claire association entre la clivabilité de HA *in vitro* et la pathogénicité du virus *in vivo*, la présence d'un site de clivage polybasique ne fait pas tout. Comme dit précédemment, certains virus sans site de clivage polybasique sont fortement pathogènes et certains virus avec un site de clivage polybasique ne le sont pas. De nombreuses autres mutations sont impliquées dans la pathogénicité, qu'elles soient localisées autour du site de clivage ou sur d'autres segments que HA. Par exemple, un facteur de pathogénicité bien décrit est la présence d'une NA tronquée. La délétion d'une vingtaine d'acides aminés sur la tige de NA est fréquemment observée chez les souches virales adaptées aux poulets [77] et est associée à une augmentation de la réplication dans l'appareil respiratoire et de la mortalité [78, 79]. Suite à cette délétion, l'équilibre entre les fonctions de HA et NA, respectivement la liaison aux acides sialiques et le clivage de ces derniers, est perturbé : l'activité sialidase de NA est diminuée, la flexibilité de la protéine et sa capacité à atteindre les acides sialiques étant moindres. Ainsi, la probabilité que HA se lie à ces derniers est augmentée [80, 81]. Toutefois, les raisons pour lesquelles l'émergence de virus influenza aviaires avec une activité sialidase diminuée est favorisée chez les poulets restent globalement méconnues [82].

Mesures de prévention

Le rôle des oiseaux sauvages aquatiques, le réservoir des VIA, dans l'introduction de VIAFP ou de VIAHP dans les élevages avicoles est la plupart du temps méconnu. Toutefois, à plusieurs occasions, des études épidémiologiques ont montré qu'ils étaient probablement à l'origine de l'épizootie [83-85]. La claustration des oiseaux d'élevage, notamment lorsque la circulation est forte dans la faune sauvage, paraît donc être une mesure efficace. Toutefois,

Copyright © 2022 John Libbey Eurotext. Téléchargé par M DAMIEN VITOUR le 28/11/2022.

Copyright © 2022 John Libbey Eurotext. Téléchargé par M DAMIEN VITOUR le 28/11/2022.

priver d'accès à l'extérieur les animaux n'est pas une solution applicable au long terme, à la fois d'un point de vue du bien-être animal et d'un point de vue économique. De nombreux producteurs se tournent vers l'élevage extensif afin de répondre à une demande sociétale. Les labels « Agriculture biologique » et « Label rouge » imposent dans leur cahier des charges aux producteurs de fournir à leurs animaux l'accès à un parcours extérieur [86, 87]. En revanche, l'accès aux points d'eau tels que des lacs ou des étangs devrait être limité au maximum, pour limiter les contacts avec le réservoir sauvage. Les mouvements d'animaux, notamment au sein de la filière foie gras, qui fait intervenir plusieurs types d'élevages, ainsi que les flux de matériel, ont déjà été incriminés dans la transmission virale. Pour y remédier, des mesures strictes de biosécurité doivent être mises en place : la quarantaine des animaux, la désinfection du matériel, la présence de sas, le principe de la marche en avant etc. [88-90]. Bien que cela concerne principalement l'Asie, les marchés d'oiseaux vivants sont une source de transmission des VIA et doivent de ce fait être encadrés de manière stricte, voire abrogés [91].

La circulation des virus dans les élevages est également étroitement surveillée par les vétérinaires et les laboratoires vétérinaires départementaux. Dès qu'un VIAFP ou un VIAHP de sous-type H5 ou H7 est détecté dans un élevage, des mesures drastiques sont mises en place, passant par l'abattage des animaux [92]. Ces mesures, adoptées par l'Union européenne, ont un impact économique et sociétal fort, et la législation dans certains pays en voie de développement est souvent différente, la vaccination pouvant être privilégiée, selon que le pays ait un statut indemne ou non vis-à-vis de l'influenza aviaire hautement pathogène.

La vaccination des oiseaux d'élevage contre les VIAHP est pour le moment interdite dans les pays de l'Union européenne, à quelques exceptions près [92]. Elle réduit la mortalité et la morbidité, mais ne permet pas d'empêcher l'infection : les oiseaux infectés et vaccinés peuvent quand même transmettre le virus, lui permettant de circuler et de continuer à évoluer sans être facilement détecté, ce qui, à terme, peut représenter une menace pour la santé publique et la santé animale. De plus, du fait de la dérive antigénique, un vaccin peut vite devenir obsolète [93]. Si lors de l'épizootie italienne à H7N1 en 1999-2000, l'utilisation d'un vaccin inactivé a été autorisée et a contribué au contrôle de la crise [94], la vaccination ne porte pas toujours ses fruits. Par exemple, malgré la campagne de vaccination suite à une épizootie à VIAHP H7N3 au Pakistan en 1995, le virus a continué à circuler dans les années qui ont suivi [95]. Au sein de l'Union européenne, lorsque la vaccination est autorisée, les vaccins doivent être des vaccins DIVA *(differentiating infected from vaccinated animal)* de telle sorte à ce que la vaccination n'interfère pas avec les tests sérologiques utilisés dans le cadre du dépistage de la grippe aviaire [92]. En France, deux vaccins ont été testés au printemps 2022 chez des canards élevés pour la production de foie gras. Si les animaux sont vaccinés dans les élevages, les infections expérimentales (encore non réalisées au moment de l'écriture de cet article) auront lieu en animalerie de niveau de biosécurité 3. Les résultats de ces essais seront disponibles d'ici fin 2022.

Conclusion

Ce n'est qu'une question de temps avant que la prochaine épizootie se déclare sur notre territoire. Les VIAFP circulent dans la faune sauvage depuis potentiellement des millénaires et disposent d'un réservoir considérable, rendant illusoire la possibilité, un jour, de les éradiquer. Certains d'entre eux évolueront vers des formes hautement pathogènes et seront responsables de nouvelles épizooties. D'ici là, la meilleure connaissance et la meilleure maîtrise des VIA restent des enjeux majeurs, par bien des aspects : santé publique, économie, sécurité alimentaire, bien-être animal et bien d'autres encore.

Liens d'intérêt : les auteurs déclarent ne pas avoir de lien d'intérêt en rapport avec cet article.

Références

1. Long JS, Mistry B, Haslam SM, Barclay WS. Host and viral determinants of influenza A virus species specificity. *Nat Rev Microbiol* 2019 ; 17 : 67-81.

2. Trovão NS, Nelson MI. When pigs fly: pandemic influenza enters the 21st century. *PLoS Pathog* 2020 ; 16 : e1008259.

3. Alexander DJ. A review of avian influenza in different bird species. *Vet Microbiol* 2000 ; 74(1–2) : 3-13.

4. Adlhoch C, Fusaro A, Gonzales JL, *et al*. Avian influenza overview-update on 19 November 2020, EU/EEA and the UK. *EFSA Journal* 2020 ; 18 : e06341.

5. Xu X, Subbarao null X, Cox NJ, Guo Y. Genetic characterization of the pathogenic influenza A/Goose/Guangdong/1/96 (H5N1) virus: similarity of its hemagglutinin gene to those of H5N1 viruses from the 1997 outbreaks in Hong Kong. *Virology* 1999 ; 261 : 15-9.

6. Guinat C, Nicolas G, Vergne T, *et al*. Spatio-temporal patterns of highly pathogenic avian influenza virus subtype H5N8 spread, France, 2016 to 2017. *Euro Surveill* 2018 ; 23(26) : 1700791.

7. Shi W, Gao GF. Emerging H5N8 avian influenza viruses. *Science* 2021 ; 372 : 784-6.

8. Vergne T, Gubbins S, Guinat C, Bauzile B, Delpont M, Chakraborty D, *et al*. Inferring within-flock transmission dynamics of highly pathogenic avian influenza (H5N8) in France, 2020. *Transbound Emerg Dis* 2020 ; 68(6) : 3151-5.

9. Bi Y, Chen Q, Wang Q, *et al*. Genesis, evolution and prevalence of H5N6 avian influenza viruses in China. *Cell Host Microbe* 2016 ; 20 : 810-21.

Copyright © 2022 John Libbey Eurotext. Téléchargé par M DAMIEN VITOUR le 28/11/2022.

10. Yuen KY, Chan PK, Peiris M, *et al*. Clinical features and rapid viral diagnosis of human disease associated with avian influenza A H5N1 virus. *Lancet* 1998 ; 351 : 467-71.

11. Pyankova OG, Susloparov IM, Moiseeva AA, *et al*. Isolation of clade 2.3.4.4b A(H5N8), a highly pathogenic avian influenza virus, from a worker during an outbreak on a poultry farm, Russia, December 2020. *Euro Surveill* 2021 ; 26 : 2100439.

12. Stokstad E. *Wrestling with bird flu, Europe considers once-taboo vaccines*. Science, 2022. Available from: https://www.science.org/content/article/wrestling-bird-flu-europe-considers-once-taboo-vaccines (accessed on 31/05/2022).

13. OIE. *OIE manual of diagnostic tests and vaccines for terrestrial animals. Chapter 3.3.4 - Avian influenza (infection with avian influenza viruses)*. 2019.

14. Bertram S, Glowacka I, Steffen I, Kühl A, Pöhlmann S. Novel insights into proteolytic cleavage of influenza virus hemagglutinin. *Rev Med Virol* 2010 ; 20 : 298-310.

15. Tashiro M, Ciborowski P, Klenk HD, Pulverer G, Rott R. Role of Staphylococcus protease in the development of influenza pneumonia. *Nature* 1987 ; 325 : 536-7.

16. Klenk HD, Garten W. Host cell proteases controlling virus pathogenicity. *Trends Microbiol* 1994 ; 2 : 39-43.

17. Webster RG, Yakhno M, Hinshaw VS, Bean WJ, Murti KG. Intestinal influenza: replication and characterization of influenza viruses in ducks. *Virology* 1978 ; 84 : 268-78.

18. van Gils JA, Munster VJ, Radersma R, Liefhebber D, Fouchier RAM, Klaassen M. Hampered foraging and migratory performance in swans infected with low-pathogenic avian influenza A virus. *PLoS ONE* 2007 ; 2 : e184.

19. Kida H, Yanagawa R, Matsuoka Y. Duck influenza lacking evidence of disease signs and immune response. *Infect Immun* 1980 ; 30 : 547-53.

20. Klenk HD, Rott R, Orlich M, Blödorn J. Activation of influenza A viruses by trypsin treatment. *Virology* 1975 ; 68 : 426-39.

21. Slemons RD, Swayne DE. Replication of a waterfowl-origin influenza virus in the kidney and intestine of chickens. *Avian Dis* 1990 ; 34 : 277-84.

22. de Wit JJ, Fabri THF, Molenaar RJ, Dijkman R, de Bruijn N, Bouwstra R. Major difference in clinical outcome and replication of a H3N1 avian influenza strain in young pullets and adult layers. *Avian Pathol* 2020 ; 49 : 286-95.

23. Stieneke-Gröber A, Vey M, Angliker H, *et al*. Influenza virus hemagglutinin with multibasic cleavage site is activated by furin, a subtilisin-like endoprotease. *EMBO J* 1992 ; 11 : 2407-14.

24. Walker JA, Molloy SS, Thomas G, *et al*. Sequence specificity of furin, a proprotein-processing endoprotease, for the hemagglutinin of a virulent avian influenza virus. *J Virol* 1994 ; 68 : 1213-8.

25. Klenk HD, Wöllert W, Rott R, Scholtissek C. Association of influenza virus proteins with cytoplasmic fractions. *Virology* 1974 ; 57 : 28-41.

26. Steinhauer DA. Role of hemagglutinin cleavage for the pathogenicity of influenza virus. *Virology* 1999 ; 258 : 1-20.

27. Figueroa T, Bessière P, Coggon A, *et al*. The microbiota contributes to the control of highly pathogenic H5n9 influenza virus replication in ducks. *J Virol* 2020 ; 94(10) : e00289-320.

28. Grund C, Hoffmann D, Ulrich R, *et al*. A novel European H5N8 influenza A virus has increased virulence in ducks but low zoonotic potential. *Emerg Microbes Infect* 2018 ; 7 : 132.

29. Kleyheeg E, Slaterus R, Bodewes R, *et al*. Deaths among Wild Birds during Highly Pathogenic Avian Influenza A(H5N8) Virus Outbreak, the Netherlands. *Emerg Infect Dis* 2017 ; 23 : 2050-4.

30. Wood GW, Banks J, Strong I, Parsons G, Alexander DJ. An avian influenza virus of H10 subtype that is highly pathogenic for chickens, but lacks multiple basic amino acids at the haemagglutinin cleavage site. *Avian Pathol* 1996 ; 25 : 799-806.

31. Wong S-S, Yoon S-W, Zanin M, *et al*. Characterization of an H4N2 influenza virus from quails with a multibasic motif in the hemagglutinin cleavage site. *Virology* 2014 ; 0 : 72-80.

32. Londt BZ, Banks J, Alexander DJ. Highly pathogenic avian influenza viruses with low virulence for chickens in *in vivo* tests. *Avian Pathol* 2007 ; 36 : 347-50.

33. Suarez DL, Senne DA, Banks J, *et al*. Recombination resulting in virulence shift in avian influenza outbreak, Chile. *Emerging Infect Dis* 2004 ; 10 : 693-9.

34. Pasick J, Handel K, Robinson J, *et al*. Intersegmental recombination between the haemagglutinin and matrix genes was responsible for the emergence of a highly pathogenic H7N3 avian influenza virus in British Columbia. *J Gen Virol* 2005 ; 86(Pt 3) : 727-31.

35. Kwon H, Kim Y-I, Park S-J, *et al*. A novel neuraminidase-dependent hemagglutinin cleavage mechanism enables the systemic spread of an H7N6 avian influenza virus. *mBio* 2019 ; 10 : e02369-2419.

36. Banks J, Speidel EC, McCauley JW, Alexander DJ. Phylogenetic analysis of H7 haemagglutinin subtype influenza A viruses. *Arch Virol* 2000 ; 145 : 1047-58.

37. Monne I, Fusaro A, Nelson MI, *et al*. Emergence of a highly pathogenic avian influenza virus from a low-pathogenic progenitor. *J Virol* 2014 ; 88 : 4375-88.

38. Röhm C, Süss J, Pohle V, Webster RG. Different hemagglutinin cleavage site variants of H7N7 in an influenza outbreak in chickens in Leipzig, Germany. *Virology* 1996 ; 218 : 253-7.

39. Abdelwhab E-SM, Veits J, Mettenleiter TC. Genetic changes that accompanied shifts of low pathogenic avian influenza viruses toward higher pathogenicity in poultry. *Virulence* 2013 ; 4 : 441-52.

40. Horimoto T, Rivera E, Pearson J, *et al*. Origin and molecular changes associated with emergence of a highly pathogenic H5N2 influenza virus in Mexico. *Virology* 1995 ; 213 : 223-30.

41. Naeem K, Siddique N, Ayaz M, Jalalee MA. Avian influenza in Pakistan: outbreaks of low- and high-pathogenicity avian influenza in Pakistan during 2003-2006. *Avian Dis* 2007 ; 51(1 Suppl) : 189-93.

42. Banks J, Speidel ES, Moore E, *et al*. Changes in the haemagglutinin and the neuraminidase genes prior to the emergence of highly pathogenic H7N1 avian influenza viruses in Italy. *Arch Virol* 2001 ; 146 : 963-73.

43. Richard M, Fouchier R, Monne I, Kuiken T. Mechanisms and risk factors for mutation from low to highly pathogenic avian influenza virus. *EFSA J* 2017 ; 14(10) : 1287E.

44. Alexander DJ, Spackman D. Characterisation of influenza A viruses isolated from turkeys in England during March-May 1979. *Avian Pathol* 1981 ; 10 : 281-93.

45. Abolnik C. Evolution of H5 highly pathogenic avian influenza: sequence data indicate stepwise changes in the cleavage site. *Arch Virol* 2017 ; 162 : 2219-30.

46. Lowen AC. Constraints, Drivers, and implications of influenza A virus reassortment. *Annu Rev Virol* 2017 ; 4 : 105-21.

47. Nao N, Yamagishi J, Miyamoto H, *et al*. Genetic predisposition to acquire a polybasic cleavage site for highly pathogenic avian influenza virus hemagglutinin. *mBio* 2017 ; 8 : e02298-2316.

48. Kolakofsky D. Paramyxovirus RNA synthesis, mRNA editing, and genome hexamer phase: a review. *Virology* 2016 ; 498 : 94-8.

Copyright © 2022 John Libbey Eurotext. Téléchargé par M DAMIEN VITOUR le 28/11/2022.

49. Gultyaev AP, Richard M, Spronken MI, Olsthoorn RCL, Fouchier RAM. Conserved structural RNA domains in regions coding for cleavage site motifs in hemagglutinin genes of influenza viruses. *Virus Evol* 2019 ; 5(2) : vez034.

50. García M, Crawford JM, Latimer JW, Rivera-Cruz E, Perdue ML. Heterogeneity in the haemagglutinin gene and emergence of the highly pathogenic phenotype among recent H5N2 avian influenza viruses from Mexico. *J Gen Virol* 1996 ; 77 : 1493-504.

51. Perdue ML, Garcíia M, Senne D, Fraire M. Virulence-associated sequence duplication at the hemagglutinin cleavage site of avian influenza viruses. *Virus Res* 1997 ; 49 : 173-86.

52. Bashiruddin JB, Gould AR, Westbury HA. Molecular pathotyping of two avian influenza viruses isolated during the Victoria 1976 outbreak. *Aust Vet J* 1992 ; 69 : 140-2.

53. Abbas MA, Spackman E, Swayne DE, *et al*. Sequence and phylogenetic analysis of H7N3 avian influenza viruses isolated from poultry in Pakistan 1995-2004. *Virol J* 2010 ; 7 : 137.

54. Gultyaev AP, Spronken MI, Funk M, Fouchier RAM, Richard M. Insertions of codons encoding basic amino acids in H7 hemagglutinins of influenza A viruses occur by recombination with RNA at hotspots near snoRNA binding sites. *RNA* 2021 ; 27 : 123-32.

55. Moumen A, Polomack L, Unge T, Véron M, Buc H, Negroni M. Evidence for a mechanism of recombination during reverse transcription dependent on the structure of the acceptor RNA. *J Biol Chem* 2003 ; 278 : 15973-82.

56. Simon-Loriere E, Martin DP, Weeks KM, Negroni M. RNA structures facilitate recombination-mediated gene swapping in HIV-1. *J Virol* 2010 ; 84 : 12675-82.

57. Wandzik JM, Kouba T, Cusack S. Structure and Function of Influenza Polymerase. *Cold Spring Harb Perspect Med* 2021 ; 11(9) : a038372.

58. Dupré G, Hoede C, Figueroa T, Bessière P, Bertagnoli S, Ducatez M, *et al*. Phylodynamic study of the conserved RNA structure encompassing the hemagglutinin cleavage site encoding region of H5 and H7 low pathogenic avian influenza viruses. *Virus Evol* 2021 ; 7 : veab093.

59. Escalera-Zamudio M, Golden M, Gutiérrez B, *et al*. Parallel evolution in the emergence of highly pathogenic avian influenza A viruses. *Nat Commun* 2020 ; 11 : 5511.

60. Eigen M. Viral quasispecies. *Sci Am* 1993 ; 269 : 42-9.

61. Nowak MA, May RM, Anderson RM. The evolutionary dynamics of HIV-1 quasispecies and the development of immunodeficiency disease. *AIDS* 1990 ; 4 : 1095-103.

62. Tsukiyama-Kohara K, Kohara M. Hepatitis C virus: viral quasispecies and genotypes. *Int J Mol Sci* 2017 ; 19(1) : 23.

63. Domingo E, Perales C. Viral quasispecies. *PLoS Genet* 2019 ; 15 : e1008271.

64. Bessière P, Figueroa T, Coggon A, *et al*. Opposite outcomes of the within-host competition between high- and low-pathogenic H5N8 avian influenza viruses in chickens compared to ducks. *J Virol* 2022 ; 96 : e01366-1421.

65. Brugh M. Highly pathogenic virus recovered from chickens infected with mildly pathogenic 1986 isolates of H5N2 avian influenza virus. *Avian Dis* 1988 ; 32 : 695-703.

66. Li SQ, Orlich M, Rott R. Generation of seal influenza virus variants pathogenic for chickens, because of hemagglutinin cleavage site changes. *J Virol* 1990 ; 64 : 3297-303.

67. Alkie TN, Yitbarek A, Hodgins DC, Kulkarni RR, Taha-Abdelaziz K, Sharif S. Development of innate immunity in chicken embryos and newly hatched chicks: a disease control perspective. *Avian Pathol* 2019 ; 48 : 288-310.

68. Horimoto T, Kawaoka Y. Molecular changes in virulent mutants arising from avirulent avian influenza viruses during replication in 14-day-old embryonated egg. *s. Virology* 1995 ; 206 : 755-9.

69. Dlugolenski D, Jones L, Saavedra G, Tompkins SM, Tripp RA, Mundt E. Passage of low-pathogenic avian influenza (LPAI) viruses mediates rapid genetic adaptation of a wild-bird isolate in poultry. *Arch Virol* 2011 ; 156 : 565-76.

70. Ito T, Goto H, Yamamoto E, *et al*. Generation of a highly pathogenic avian influenza A virus from an avirulent field isolate by passaging in chickens. *J Virol* 2001 ; 75 : 4439-43.

71. Soda K, Cheng M-C, Yoshida H, *et al*. A low pathogenic H5N2 influenza virus isolated in taiwan acquired high pathogenicity by consecutive passages in chickens. *J Vet Med Sci* 2011 ; 73(6) : 767-72.

72. Shengqing Y, Kishida N, Ito H, *et al*. Generation of velogenic newcastle disease viruses from a nonpathogenic waterfowl isolate by passaging in chickens. *Virology* 2002 ; 301 : 206-11.

73. Dhingra MS, Artois J, Dellicour S, *et al*. Geographical and Historical Patterns in the Emergences of Novel Highly Pathogenic Avian Influenza (HPAI) H5 and H7 Viruses in Poultry. *Front Vet Sci* 2018 ; 5 : 84.

74. Becker WB. The isolation and classification of Tern virus: influenza Virus A/Tern/South Africa/1961. *J Hyg (Lond)* 1966 ; 64 : 309-20.

75. Iqbal M, Reddy KB, Brookes SM, Essen SC, Brown IH, McCauley JW. Virus pathotype and deep sequencing of the HA gene of a low pathogenicity H7N1 avian influenza virus causing mortality in turkeys. *PLoS ONE* 2014 ; 9 : e87076.

76. Dietze K, Graaf A, Homeier-Bachmann T, *et al*. From low to high pathogenicity – Characterization of H7N7 avian influenza viruses in two epidemiologically linked outbreaks. *Transbound Emerg Dis* 2017 ; 0 : 1576-87.

77. Li J, zu Dohna H, Cardona CJ, Miller J, Carpenter TE. Emergence and genetic variation of neuraminidase stalk deletions in avian influenza viruses. *PLoS One* 2011 : 6.

78. Hoffmann TW, Munier S, Larcher T, *et al*. Length variations in the NA stalk of an H7N1 influenza virus have opposite effects on viral excretion in chickens and ducks. *J Virol* 2012 ; 86 : 584-8.

79. Sun Y, Tan Y, Wei K, *et al*. Amino acid 316 of hemagglutinin and the neuraminidase stalk length influence virulence of H9N2 influenza virus in chickens and mice. *J Virol* 2013 ; 87 : 2963-8.

80. Baigent SJ, McCauley JW. Glycosylation of haemagglutinin and stalk-length of neuraminidase combine to regulate the growth of avian influenza viruses in tissue culture. *Virus Res* 2001 ; 79 : 177-85.

81. Durrant JD, Bush RM, Amaro RE. Microsecond molecular dynamics simulations of influenza neuraminidase suggest a mechanism for the increased virulence of stalk-deletion mutants. *J Phys Chem B* 2016 ; 120 : 8590-9.

82. Mo J, Youk S, Pantin-Jackwood MJ, *et al*. The pathogenicity and transmission of live bird market H2N2 avian influenza viruses in chickens, Pekin ducks, and guinea fowl. *Vet Microbiol* 2021 ; 260 : 109180.

83. Hanna A, Banks J, Marston DA, Ellis RJ, Brookes SM, Brown IH. Genetic characterization of highly pathogenic avian influenza (H5N8) virus from domestic ducks, England, November 2014. *Emerg Infect Dis* 2015 ; 21 : 879-82.

84. Keawcharoen J, van den Broek J, Bouma A, Tiensin T, Osterhaus ADME, Heesterbeek H. Wild birds and increased transmission of highly pathogenic avian influenza (H5N1) among Poultry, Thailand. *Emerg Infect Dis* 2011 ; 17 : 1016-22.

85. Verhagen JH, van der Jeugd HP, Nolet BA, *et al*. Wild bird surveillance around outbreaks of highly pathogenic avian influenza A(H5N8) virus in the Netherlands, 2014, within the context of global flyways. *Euro Surveill* 2015 ; 20 : 21069.

86. EU. *Règlement (CE) N° 889/2008 de la Commission du 5 septembre 2008*. 2008.

87. MAA. *Cahier des charges du label rouge n° LA 08/87 "Poulet blanc fermier élevé en plein air, entier et découpes, frais*." 2017.

88. Delpont M, Blondel V, Robertet L, *et al*. Biosecurity practices on foie gras duck farms, Southwest France. *Prev Vet Med* 2018 ; 158 : 78-88.

89. Delpont M, Racicot M, Durivage A, *et al*. Determinants of biosecurity practices in French duck farms after a H5N8 Highly Pathogenic Avian Influenza epidemic: the effect of farmer knowledge, attitudes and personality traits. *Transbound Emerg Dis* 2021 ; 68(1) : 51-61.

90. Guinat C, Comin A, Kratzer G, *et al*. Biosecurity risk factors for highly pathogenic avian influenza (H5N8) virus infection in duck farms, France. *Transbound Emerg Dis* 2020 ; 67 : 2961-70.

91. Zhou P, Ma J, Lai A, Gray GC, Su S, Li S. Avian influenza A(H7N9) virus and mixed live poultry-animal markets in Guangdong province: a perfect storm in the making? *Emerg Microbes Infect* 2015 ; 4 : e63.

92. EU. *COUNCIL DIRECTIVE 2005/94/EC of 20 December 2005 on Community measures for the control of avian influenza and repealing Directive 92/40/EEC*. 2006. Available from: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32005L0094&from=en (consulté le 10/09/2022)..

93. Swayne DE. Principles for vaccine protection in chickens and domestic waterfowl against avian influenza: emphasis on Asian H5N1 high pathogenicity avian influenza. *Ann N Y Acad Sci* 2006 ; 1081 : 174-81.

94. Capua I, Marangon S. The use of vaccination to combat multiple introductions of Notifiable Avian Influenza viruses of the H5 and H7 subtypes between 2000 and 2006 in Italy. *Vaccine* 2007 ; 25 : 4987-95.

95. Naeem K, Siddique N. Use of strategic vaccination for the control of avian influenza in Pakistan. *Dev Biol (Basel)* 2006 ; 124 : 145-50.

96. Krammer F, García-Sastre A, Palese P. Is it possible to develop a "Universal" influenza virus vaccine?: potential target antigens and critical aspects for a universal influenza vaccine. *Cold Spring Harb Perspect Biol* 2018 ; 10 : a028845.

Copyright © 2022 John Libbey Eurotext. Téléchargé par M DAMIEN VITOUR le 28/11/2022.

Références bibliographiques

Abbas, Muhammad A, Erica Spackman, David E Swayne, Zaheer Ahmed, Luciana Sarmento, Naila Siddique, Khalid Naeem, Abdul Hameed, and Shafqat Rehmani. 2010. “Sequence and Phylogenetic Analysis of H7N3 Avian Influenza Viruses Isolated from Poultry in Pakistan 1995-2004.” *Virology Journal* 7 (1): 137. https://doi.org/10.1186/1743-422X-7-137.

Abdel-Ghafar, Abdel-Nasser, Tawee Chotpitayasunondh, Zhancheng Gao, Frederick G Hayden, Duc Hien, Menno D De Jong, Azim Naghdaliyev, et al. 2008. “Update on Avian Influenza A (H5N1) Virus Infection in Humans.” *New England Journal of Medicine* 358 (3): 261–73. https://doi.org/10.1056/NEJMRA0707279/SUPPL_FILE/NEJM_HAYDEN_261SA1.PDF.

Abdelwhab, El-Sayed M, Jutta Veits, and Thomas C Mettenleiter. 2013. “Genetic Changes That Accompanied Shifts of Low Pathogenic Avian Influenza Viruses toward Higher Pathogenicity in Poultry.” *Virulence* 4 (6): 441–52. https://doi.org/10.4161/viru.25710.

Abolnik, Celia. 2017. “Evolution of H5 Highly Pathogenic Avian Influenza: Sequence Data Indicate Stepwise Changes in the Cleavage Site.” *Archives of Virology* 162 (8): 2219–30. https://doi.org/10.1007/s00705-017-3337-x.

Agüero, Montserrat, Isabella Monne, Azucena Sánchez, Bianca Zecchin, Alice Fusaro, María José Ruano, Manuel del Valle Arrojo, et al. 2023. “Highly Pathogenic Avian Influenza A(H5N1) Virus Infection in Farmed Minks, Spain, October 2022.” *Eurosurveillance* 28 (3): 2300001. https://doi.org/10.2807/1560-7917.ES.2023.28.3.2300001.

Ayllon, Juan, and Adolfo García-Sastre. 2015. “The NS1 Protein: A Multitasking Virulence Factor.” In *Influenza Pathogenesis and Control - Volume II*, edited by Michael B A Oldstone and Richard W Compans, 73–107. Current {Topics} in {Microbiology} and {Immunology}. Cham: Springer International Publishing. https://doi.org/10.1007/82_2014_400.

Barbezange, Cyril, Louis Jones, Hervé Blanc, Ofer Isakov, Gershon Celniker, Vincent Enouf, Noam Shomron, Marco Vignuzzi, and Sylvie van der Werf. 2018. “Seasonal Genetic Drift of Human Influenza A Virus Quasispecies Revealed by Deep Sequencing.” *Frontiers in Microbiology* 9 (September). https://doi.org/10.3389/fmicb.2018.02596.

Bashiruddin, JB, AR Gould, and HA Westbury. 1992. “Molecular Pathotyping of Two Avian Influenza Viruses Isolated during the Victoria 1976 Outbreak.” *Australian Veterinary Journal* 69 (6): 140–42. https://doi.org/10.1111/J.1751-0813.1992.TB07485.X.

Bean, W J, M Schell, J Katz, Y Kawaoka, C Naeve, O Gorman, and R G Webster. 1992. “Evolution of the H3 Influenza Virus Hemagglutinin from Human and Nonhuman Hosts.” *Journal of Virology* 66 (2): 1129. https://doi.org/10.1128/JVI.66.2.1129-1138.1992.

Beerens, Nancy, Rene Heutink, Frank Harders, Alex Bossers, Guus Koch, and Ben Peeters. 2020. “Emergence and Selection of a Highly Pathogenic Avian Influenza H7N3 Virus.” *Journal of Virology*, January. https://doi.org/10.1128/JVI.01818-19.

Berhane, Yohannes, Tamiko Hisanaga, Helen Kehler, James Neufeld, Lisa Manning, Connie Argue, Katherine Handel, et al. 2009. “Highly Pathogenic Avian Influenza Virus A (H7N3) in Domestic Poultry, Saskatchewan, Canada, 2007.” *Emerging Infectious Diseases* 15 (9): 1492. https://doi.org/10.3201/eid1509.080231.

Bessière, Pierre, Thomas Figueroa, Amelia Coggon, Charlotte Foret-Lucas, Alexandre Houffschmitt, Maxime Fusade-Boyer, Gabriel Dupré, Jean-Luc Guérin, Maxence Delverdier, and Romain Volmer. 2022. “Opposite Outcomes of the Within-Host Competition between High- and Low-Pathogenic H5N8 Avian Influenza Viruses in Chickens Compared to Ducks.” *Journal of Virology* 96 (1). https://doi.org/10.1128/JVI.01366-21.

Bhatt, Samir, Edward C. Holmes, and Oliver G. Pybus. 2011. “The Genomic Rate of Molecular Adaptation of the Human Influenza A Virus.” *Molecular Biology and Evolution* 28 (9): 2443–51. https://doi.org/10.1093/molbev/msr044.

Bjørnstad, Ottar N, and Cecile Viboud. 2016. “Timing and Periodicity of Influenza Epidemics.” *PNAS* 113 (46): 12899–901. https://doi.org/10.1073/pnas.1616052113.

Boni, M. F., Y. Zhou, J. K. Taubenberger, and E. C. Holmes. 2008. “Homologous Recombination Is

Very Rare or Absent in Human Influenza A Virus." *Journal of Virology* 82 (10): 4807–11. https://doi.org/10.1128/jvi.02683-07.
Bragstad, Karoline, Lars P Nielsen, and Anders Fomsgaard. 2008. "The Evolution of Human Influenza A Viruses from 1999 to 2006: A Complete Genome Study." *Virology Journal* 5 (September): 40. https://doi.org/10.1186/1743-422X-5-40.
Brown, Justin D., Ginger Goekjian, Rebecca Poulson, Steve Valeika, and David E. Stallknecht. 2009. "Avian Influenza Virus in Water: Infectivity Is Dependent on PH, Salinity and Temperature." *Veterinary Microbiology* 136 (1–2): 20–26. https://doi.org/10.1016/j.vetmic.2008.10.027.
Brunotte, Linda, Joe Flies, Hardin Bolte, Peter Reuther, Frank Vreede, and Martin Schwemmle. 2014. "The Nuclear Export Protein of H5N1 Influenza A Viruses Recruits Matrix 1 (M1) Protein to the Viral Ribonucleoprotein to Mediate Nuclear Export." *Journal of Biological Chemistry* 289 (29): 20067–77. https://doi.org/10.1074/jbc.M114.569178.
Castro Martin, Isabel Fernández De, Guillaume Fournier, Martin Sachse, Javier Pizarro-Cerda, Cristina Risco, and Nadia Naffakh. 2017. "Influenza Virus Genome Reaches the Plasma Membrane via a Modified Endoplasmic Reticulum and Rab11-Dependent Vesicles." *Nature Communications* 8 (1). https://doi.org/10.1038/s41467-017-01557-6.
Chrisman, Brianna Sierra, Kelley Paskov, Nate. Stockham, Kevin Tabatabaei, Jae-Yoon Jung, Peter Washington, Maya Varma, Min Woo Sun, Sepideh Maleki, and Dennis P. Wall. 2021. "Indels in SARS-CoV-2 Occur at Template-Switching Hotspots." *BioData Mining 2021 14:1* 14 (1): 1–16. https://doi.org/10.1186/S13040-021-00251-0.
Chung, Claire, Bert M. Verheijen, Xinmin Zhang, Biao Huang, Aeowynn Coakley, Eric McGann, Emily Wade, et al. 2023. "The Fidelity of Transcription in Human Cells." *Proceedings of the National Academy of Sciences* 120 (5): e2210038120. https://doi.org/10.1073/PNAS.2210038120.
Claas, Eric C.J., Albert D.M.E. Osterhaus, Ruud Van Beek, Jan C. De Jong, Guus F. Rimmelzwaan, Dennis A. Senne, Scott Krauss, Kennedy F. Shortridge, and Robert G. Webster. 1998. "Human Influenza A H5N1 Virus Related to a Highly Pathogenic Avian Influenza Virus." *Lancet* 351 (9101): 472–77. https://doi.org/10.1016/S0140-6736(97)11212-0.
Cros, Jerome F., and Peter Palese. 2003. "Trafficking of Viral Genomic RNA into and out of the Nucleus: Influenza, Thogoto and Borna Disease Viruses." *Virus Research* 95 (1–2): 3–12. https://doi.org/10.1016/S0168-1702(03)00159-X.
Dadonaite, Bernadeta, Brad Gilbertson, Michael L. Knight, Sanja Trifkovic, Steven Rockman, Alain Laederach, Lorena E. Brown, Ervin Fodor, and David L. V. Bauer. 2019. "The Structure of the Influenza A Virus Genome." *Nature Microbiology*, July, 1–9. https://doi.org/10.1038/s41564-019-0513-7.
Dawson, Wayne K, Michal Lazniewski, and Dariusz Plewczynski. 2017. "RNA Structure Interactions and Ribonucleoprotein Processes of the Influenza A Virus." *Briefings in Functional Genomics* 17 (6): 402–14. https://doi.org/10.1093/bfgp/elx028.
Dietze, Klaas, Annika Graaf, Timo Homeier-Bachmann, Christian Grund, Leonie Forth, Anne Pohlmann, Christa Jeske, et al. 2018. "From Low to High Pathogenicity—Characterization of H7N7 Avian Influenza Viruses in Two Epidemiologically Linked Outbreaks." *Transboundary and Emerging Diseases* 65 (6): 1576–87. https://doi.org/10.1111/tbed.12906.
Dou, Dan, Rebecca Revol, Henrik Östbye, Hao Wang, and Robert Daniels. 2018. "Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement." *Frontiers in Immunology* 9 (July): 1581. https://doi.org/10.3389/fimmu.2018.01581.
Dulin, David, Jamie J Arnold, Theo van Laar, Hyung-Suk Oh, Cheri Lee, Angela L Perkins, Daniel A Harki, Martin Depken, Craig E Cameron, and Nynke H Dekker. 2017. "Signatures of Nucleotide Analog Incorporation by an RNA-Dependent RNA Polymerase Revealed Using High-Throughput Magnetic Tweezers." *Cell Reports* 21 (4): 1063–76. https://doi.org/10.1016/j.celrep.2017.10.005.
Dulin, David, Tao Ju Cui, Jelmer Cnossen, Margreet W. Docter, Jan Lipfert, and Nynke H. Dekker. 2015. "High Spatiotemporal-Resolution Magnetic Tweezers: Calibration and Applications for DNA Dynamics." *Biophysical Journal* 109 (10): 2113–25. https://doi.org/10.1016/j.bpj.2015.10.018.
Dupré, Gabriel, Claire Hoede, Thomas Figueroa, Pierre Bessière, Stéphane Bertagnoli, Mariette Ducatez, Christine Gaspin, and Romain Volmer. 2021. "Phylodynamic Study of the Conserved RNA Structure Encompassing the Hemagglutinin Cleavage Site Encoding Region of H5 and H7

Low Pathogenic Avian Influenza Viruses." *Virus Evolution* 7 (2): 1–16. https://doi.org/10.1093/VE/VEAB093.
Dupré, Gabriel, and Romain Volmer. 2022. "Influence of Viral Genome Properties on Polymerase Fidelity." *Trends in Genetics*, November. https://doi.org/10.1016/J.TIG.2022.10.008.
Edinger, Thomas O., Marie O. Pohl, and Silke Stertz. 2014. "Entry of Influenza A Virus: Host Factors and Antiviral Targets." *Journal of General Virology*. https://doi.org/10.1099/vir.0.059477-0.
Elena, Santiago F., Ricard V. Solé, and Josep Sardanyés. 2010. "Simple Genomes, Complex Interactions: Epistasis in RNA Virus." *Chaos: An Interdisciplinary Journal of Nonlinear Science* 20 (2): 026106. https://doi.org/10.1063/1.3449300.
Escalera-Zamudio, Marina, Michael Golden, Bernardo Gutiérrez, Julien Thézé, Jeremy Russell Keown, Loic Carrique, Thomas A Bowden, and Oliver G Pybus. 2020. "Parallel Evolution in the Emergence of Highly Pathogenic Avian Influenza A Viruses." *Nature Communications* 11 (1): 5511. https://doi.org/10.1038/s41467-020-19364-x.
Evseev, Danyel, and Katharine Magor. 2019. "Innate Immune Responses to Avian Influenza Viruses in Ducks and Chickens." *Veterinary Sciences* 6 (1): 5. https://doi.org/10.3390/vetsci6010005.
Ferhadian, Damien, Maud Contrant, Anne Printz-Schweigert, Redmond P Smyth, Jean-Christophe Paillart, and Roland Marquet. 2018. "Structural and Functional Motifs in Influenza Virus RNAs." *Frontiers in Microbiology* 9: 559. https://doi.org/10.3389/fmicb.2018.00559.
Figlerowicz, Marek. 2000. "Role of RNA Structure in Non-Homologous Recombination between Genomic Molecules of Brome Mosaic Virus." *Nucleic Acids Research* 28 (8): 1714. https://doi.org/10.1093/NAR/28.8.1714.
Fitch, Walter M., Robin M. Bush, Catherine A. Bender, and Nancy J. Cox. 1997. "Long Term Trends in the Evolution of H(3) HA1 Human Influenza Type A." *Proceedings of the National Academy of Sciences of the United States of America* 94 (15): 7712–18. https://doi.org/10.1073/pnas.94.15.7712.
Flynn, Ryan A., Qiangfeng Cliff Zhang, Robert C. Spitale, Byron Lee, Maxwell R. Mumbach, and Howard Y. Chang. 2016. "Transcriptome-Wide Interrogation of RNA Secondary Structure in Living Cells with IcSHAPE." *Nature Protocols* 11 (2): 273–90. https://doi.org/10.1038/nprot.2016.011.
Fodor, E. 2013. "The RNA Polymerase of Influenza A Virus: Mechanisms of Viral Transcription and Replication." *Acta Virologica* 57 (2): 113–22. https://doi.org/10.4149/av_2013_02_113.
França, M., D. E. Stallknecht, and E. W. Howerth. 2013. "Expression and Distribution of Sialic Acid Influenza Virus Receptors in Wild Birds." *Avian Pathology* 42 (1): 60–71. https://doi.org/10.1080/03079457.2012.759176.
French, Hollie, Emmanuelle Pitré, Michael S. Oade, Elizaveta Elshina, Karishma Bisht, Alannah King, David L.V. Bauer, and Aartjan J.W. te Velthuis. 2022. "Transient RNA Structures Cause Aberrant Influenza Virus Replication and Innate Immune Activation." *Science Advances* 8 (36): 8655. https://doi.org/10.1126/SCIADV.ABP8655.
Funk, Mathis, Anja C.M. de Bruin, Monique I. Spronken, Alexander P. Gultyaev, and Mathilde Richard. 2022. "In Silico Analyses of the Role of Codon Usage at the Hemagglutinin Cleavage Site in Highly Pathogenic Avian Influenza Genesis." *Viruses* 14 (7). https://doi.org/10.3390/V14071352.
Gamblin, Steven J., and John J. Skehel. 2010. "Influenza Hemagglutinin and Neuraminidase Membrane Glycoproteins." *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology Inc. https://doi.org/10.1074/jbc.R110.129809.
Garcia, M., J. M. Crawford, J. W. Latimer, E. Rivera-Cruz, and M. L. Perdue. 1996. "Heterogeneity in the Haemagglutinin Gene and Emergence of the Highly Pathogenic Phenotype among Recent H5N2 Avian Influenza Viruses from Mexico." *Journal of General Virology* 77 (7): 1493–1504. https://doi.org/10.1099/0022-1317-77-7-1493.
García, Maricarmen, John M. Crawford, John W. Latimer, Eduardo Rivera-Cruz, and Michael L. Perdue. 1996. "Heterogeneity in the Haemagglutinin Gene and Emergence of the Highly Pathogenic Phenotype among Recent H5N2 Avian Influenza Viruses from Mexico." *Journal of General Virology* 77 (7): 1493–1504. https://doi.org/10.1099/0022-1317-77-7-1493.
Garten, W., and H. D. Klenk. 1999. "Understanding Influenza Virus Pathogenicity." *Trends in Microbiology*. Elsevier Ltd. https://doi.org/10.1016/S0966-842X(99)01460-2.
Gavazzi, Cyrille, Catherine Isel, Emilie Fournier, Vincent Moules, Annie Cavalier, Daniel Thomas, Bruno Lina, and Roland Marquet. 2013. "An in Vitro Network of Intermolecular Interactions

between Viral RNA Segments of an Avian H5N2 Influenza A Virus: Comparison with a Human H3N2 Virus." *Nucleic Acids Research* 41 (2): 1241–54. https://doi.org/10.1093/nar/gks1181.
Geller, Ron, Pilar Domingo-Calap, José M. Cuevas, Paola Rossolillo, Matteo Negroni, and Rafael Sanjuán. 2015. "The External Domains of the HIV-1 Envelope Are a Mutational Cold Spot." *Nature Communications* 6 (1): 1–9. https://doi.org/10.1038/ncomms9571.
Geller, Ron, Úrsula Estada, Joan B. Peris, Iván Andreu, Juan Vicente Bou, Raquel Garijo, José M. Cuevas, Rosario Sabariegos, Antonio Mas, and Rafael Sanjuán. 2016. "Highly Heterogeneous Mutation Rates in the Hepatitis C Virus Genome." *Nature Microbiology* 1 (7): 1–7. https://doi.org/10.1038/nmicrobiol.2016.45.
Gilmer, Orian, Erwan Quignon, Anne Caroline Jousset, Jean Christophe Paillart, Roland Marquet, and Valérie Vivet-Boudou. 2021. "Chemical and Enzymatic Probing of Viral RNAs: From Infancy to Maturity and Beyond." *Viruses 2021, Vol. 13, Page 1894* 13 (10): 1894. https://doi.org/10.3390/V13101894.
Gischke, Marcel, Reiner Ulrich, Olanrewaju I. Fatola, David Scheibner, Ahmed H. Salaheldin, Beate Crossley, Eva Böttcher-Friebertshäuser, Jutta Veits, Thomas C. Mettenleiter, and Elsayed M. Abdelwhab. 2020. "Insertion of Basic Amino Acids in the Hemagglutinin Cleavage Site of H4N2 Avian Influenza Virus (AIV)—Reduced Virus Fitness in Chickens Is Restored by Reassortment with Highly Pathogenic H5N1 AIV." *International Journal of Molecular Sciences* 21 (7): 2353. https://doi.org/10.3390/ijms21072353.
Gorman, O T, R O Donis, Y Kawaoka, and R G Webster. 1990. "Evolution of Influenza A Virus PB2 Genes: Implications for Evolution of the Ribonucleoprotein Complex and Origin of Human Influenza A Virus." *Journal of Virology* 64 (10): 4893. https://doi.org/10.1128/JVI.64.10.4893-4902.1990.
Gorman, Owen T, William J Bean,', Yoshihiro Kawaoka, Isabella Donatelli, Yuanji Guo, and Robert G Webster'. 1991. "Evolution of Influenza A Virus Nucleoprotein Genes: Implications for the Origins of H1N1 Human and Classical Swine Viruses." *Journal of Virology* 65 (7): 3704. https://doi.org/10.1128/JVI.65.7.3704-3714.1991.
Gorman, Owen T, William J Bean, Yoshihiro Kawaoka, and Robert G Webster. 1990. "Evolution of the Nucleoprotein Gene of Influenza A Virus." *Journal of Virology* 64 (4): 1487–97. https://doi.org/10.1128/JVI.64.4.1487-1497.1990.
Gout, Jean-Francois, Weiyi Li, Clark Fritsch, Annie Li, Suraiya Haroon, Larry Singh, Ding Hua, et al. 2017. "The Landscape of Transcription Errors in Eukaryotic Cells." *Science Advances* 3 (10). https://doi.org/10.1126/sciadv.1701484.
Guan, Yi, and Gavin J D Smith. 2013. "The Emergence and Diversification of Panzootic H5N1 Influenza Viruses." *Virus Research*, H5N1, 178 (1): 35–43. https://doi.org/10.1016/j.virusres.2013.05.012.
Gultyaev, Alexander P., Monique I. Spronken, Mathis Funk, Ron A.M. Fouchier, and Mathilde Richard. 2021. "Insertions of Codons Encoding Basic Amino Acids in H7 Hemagglutinins of Influenza A Viruses Occur by Recombination with RNA at Hotspots near SnoRNA Binding Sites." *RNA* 27 (2): 123–32. https://doi.org/10.1261/rna.077495.120.
Gultyaev, Alexander P, Mathilde Richard, Monique I Spronken, René C L Olsthoorn, and Ron A M Fouchier. 2019. "Conserved Structural RNA Domains in Regions Coding for Cleavage Site Motifs in Hemagglutinin Genes of Influenza Viruses." *Virus Evolution* 5 (2): 1–11. https://doi.org/10.1093/ve/vez034.
Gultyaev, Alexander P, Monique I Spronken, Mathilde Richard, Eefje J A Schrauwen, René C L Olsthoorn, and Ron A M Fouchier. 2016. "Subtype-Specific Structural Constraints in the Evolution of Influenza A Virus Hemagglutinin Genes." *Scientific Reports* 6: 38892. https://doi.org/10.1038/srep38892.
Gultyaev, Alexander P, Anton Tsyganov-Bodounov, Monique I J Spronken, Sander van der Kooij, Ron A M Fouchier, and René C L Olsthoorn. 2014. "RNA Structural Constraints in the Evolution of the Influenza A Virus Genome NP Segment." *RNA Biology* 11 (7): 942–52. https://doi.org/10.4161/rna.29730.
Gutierrez, Bernardo, Marina Escalera-Zamudio, and Oliver G. Pybus. 2019. "Parallel Molecular Evolution and Adaptation in Viruses." *Current Opinion in Virology* 34 (February): 90–96. https://doi.org/10.1016/J.COVIRO.2018.12.006.
Han, Guan Zhu, and Michael Worobey. 2011. "Homologous Recombination in Negative Sense RNA Viruses." *Viruses*. Multidisciplinary Digital Publishing Institute (MDPI).

https://doi.org/10.3390/v3081358.
Harris, Audray, Giovanni Cardone, Dennis C. Winkler, J. Bernard Heymann, Matthew Brecher, Judith M. White, and Alasdair C. Steven. 2006. “Influenza Virus Pleiomorphy Characterized by Cryoelectron Tomography.” *Proceedings of the National Academy of Sciences* 103 (50): 19123–27. https://doi.org/10.1073/PNAS.0607614103.
Heaton, Nicholas S., David Sachs, Chi Jene Chen, Rong Hai, and Peter Palese. 2013. “Genome-Wide Mutagenesis of Influenza Virus Reveals Unique Plasticity of the Hemagglutinin and NS1 Proteins.” *Proceedings of the National Academy of Sciences of the United States of America* 110 (50): 20248–53. https://doi.org/10.1073/pnas.1320524110.
Herfst, Sander, Jie Zhang, Mathilde Richard, Ryan McBride, Pascal Lexmond, Theo M. Bestebroer, Monique I.J. Spronken, et al. 2020. “Hemagglutinin Traits Determine Transmission of Avian A/H10N7 Influenza Virus between Mammals.” *Cell Host & Microbe* 28 (4): 602-613.e7. https://doi.org/10.1016/J.CHOM.2020.08.011.
Hirst, Martin, Caroline R. Astell, Malachi Griffith, Shaun M. Coughlin, Michelle Moksa, Thomas Zeng, Duane E. Smailus, et al. 2004. “Novel Avian Influenza H7N3 Strain Outbreak, British Columbia.” *Emerging Infectious Diseases* 10 (12): 2192–95. https://doi.org/10.3201/eid1012.040743.
Horimoto, T, and Y Kawaoka. 1994. “Reverse Genetics Provides Direct Evidence for a Correlation of Hemagglutinin Cleavability and Virulence of an Avian Influenza A Virus.” *Journal of Virology* 68 (5): 3120–28. https://doi.org/10.1128/jvi.68.5.3120-3128.1994.
Horimoto, T, K Nakayama, S P Smeekens, and Y Kawaoka. 1994. “Proprotein-Processing Endoproteases PC6 and Furin Both Activate Hemagglutinin of Virulent Avian Influenza Viruses.” *Journal of Virology* 68 (9): 6074–78. https://doi.org/10.1128/jvi.68.9.6074-6078.1994.
Hu, Jiao, Chunxi Ma, and Xiufan Liu. 2018. “PA-X: A Key Regulator of Influenza A Virus Pathogenicity and Host Immune Responses.” *Medical Microbiology and Immunology* 207 (5–6): 255–69. https://doi.org/10.1007/s00430-018-0548-z.
Iglesias, I., M. Martínez, M. J. Muñoz, A. De La Torre, and J. M. Sánchez-Vizcaíno. 2010. “First Case of Highly Pathogenic Avian Influenza in Poultry in Spain.” *Transboundary and Emerging Diseases* 57 (4): 282–85. https://doi.org/10.1111/j.1865-1682.2010.01145.x.
Ito, T, O T Gorman, Y Kawaoka, W J Bean, and R G Webster. 1991. “Evolutionary Analysis of the Influenza A Virus M Gene with Comparison of the M1 and M2 Proteins.” *Journal of Virology* 65 (10): 5491–98. https://doi.org/10.1128/JVI.65.10.5491-5498.1991.
Jiang, Tian, Aitor Nogales, Steven F Baker, Luis Martinez-Sobrido, and Douglas H Turner. 2016. “Mutations Designed by Ensemble Defect to Misfold Conserved RNA Structures of Influenza A Segments 7 and 8 Affect Splicing and Attenuate Viral Replication in Cell Culture.” Edited by Stefan Pöhlmann. *PLOS ONE* 11 (6): e0156906. https://doi.org/10.1371/journal.pone.0156906.
Kamal, Ram P, Irina V Alymova, and Ian A York. 2017. “Evolution and Virulence of Influenza A Virus Protein PB1-F2.” *International Journal of Molecular Sciences* 19 (1). https://doi.org/10.3390/ijms19010096.
Kawaoka, Yoshihiro, and Robert G. Webster. 1985. “Evolution of the a/Chicken/Pennsylvania/83 (H5N2) Influenza Virus.” *Virology* 146 (1): 130–37. https://doi.org/10.1016/0042-6822(85)90059-5.
Kobayashi, Yuki, Bernadeta Dadonaite, Neeltje van Doremalen, Yoshiyuki Suzuki, Wendy S. Barclay, and Oliver G. Pybus. 2016. “Computational and Molecular Analysis of Conserved Influenza A Virus RNA Secondary Structures Involved in Infectious Virion Production.” *RNA Biology* 13 (9): 883–94. https://doi.org/10.1080/15476286.2016.1208331.
Koelle, Katia, Sarah Cobey, Bryan Grenfell, and Mercedes Pascual. 2006. “Epochal Evolution Shapes the Phylodynamics of Interpandemic Influenza a (H3N2) in Humans.” *Science* 314 (5807): 1898–1903. https://doi.org/10.1126/science.1132745.
Kolakofsky, Daniel. 2016. “Paramyxovirus RNA Synthesis, MRNA Editing, and Genome Hexamer Phase: A Review.” *Virology* 498 (November): 94–98. https://doi.org/10.1016/j.virol.2016.08.018.
Kosik, Ivan, and Jonathan W. Yewdell. 2019. “Influenza Hemagglutinin and Neuraminidase: Yin–Yang Proteins Coevolving to Thwart Immunity.” *Viruses*. MDPI AG. https://doi.org/10.3390/v11040346.
Krammer, Florian, Gavin J. D. Smith, Ron A. M. Fouchier, Malik Peiris, Katherine Kedzierska, Peter C. Doherty, Peter Palese, et al. 2018. “Influenza.” *Nature Reviews Disease Primers 2018 4:1* 4

(1): 1–21. https://doi.org/10.1038/s41572-018-0002-y.
Kuribayashi, Saya, Yoshihiro Sakoda, Takeshi Kawasaki, Tomohisa Tanaka, Naoki Yamamoto, Masatoshi Okamatsu, Norikazu Isoda, et al. 2013. "Excessive Cytokine Response to Rapid Proliferation of Highly Pathogenic Avian Influenza Viruses Leads to Fatal Systemic Capillary Leakage in Chickens." Edited by Elankumaran Subbiah. *PLoS ONE* 8 (7): e68375. https://doi.org/10.1371/journal.pone.0068375.
Laleye, Agnes Tinuke, and Celia Abolnik. 2020. "Emergence of Highly Pathogenic H5N2 and H7N1 Influenza A Viruses from Low Pathogenic Precursors by Serial Passage in Ovo." Edited by Camille Lebarbenchon. *PLOS ONE* 15 (10): e0240290. https://doi.org/10.1371/journal.pone.0240290.
Leahy, M. B., H. C. Dobbyn, and G. G. Brownlee. 2001. "Hairpin Loop Structure in the 3' Arm of the Influenza A Virus Virion RNA Promoter Is Required for Endonuclease Activity." *Journal of Virology* 75 (15): 7042–49. https://doi.org/10.1128/jvi.75.15.7042-7049.2001.
Leahy, M B, D C Pritlove, L L Poon, and G G Brownlee. 2001. "Mutagenic Analysis of the 5' Arm of the Influenza A Virus Virion RNA Promoter Defines the Sequence Requirements for Endonuclease Activity." *Journal of Virology* 75 (1): 134–42. https://doi.org/10.1128/JVI.75.1.134-142.2001.
Lee, Dong-Hun, Miria Ferreira Criado, and David E. Swayne. 2020. "Pathobiological Origins and Evolutionary History of Highly Pathogenic Avian Influenza Viruses." *Cold Spring Harbor Perspectives in Medicine*, January, a038679. https://doi.org/10.1101/cshperspect.a038679.
Lee, Dong-Hun, Mia K Torchetti, Mary Lea Killian, Yohannes Berhane, and David E Swayne. 2017. "Highly Pathogenic Avian Influenza A(H7N9) Virus, Tennessee, USA, March 2017." *Emerging Infectious Diseases* 23 (11): 1860–63. https://doi.org/10.3201/eid2311.171013.
Lee, Dong Hun, Kateri Bertran, Jung Hoon Kwon, and David E. Swayne. 2017. "Evolution, Global Spread, and Pathogenicity of Highly Pathogenic Avian Influenza H5Nx Clade 2.3.4.4." *Journal of Veterinary Science*. Korean Society of Veterinary Science. https://doi.org/10.4142/jvs.2017.18.S1.269.
Lee, Juhye M., John Huddleston, Michael B. Doud, Kathryn A. Hooper, Nicholas C. Wu, Trevor Bedford, and Jesse D. Bloom. 2018. "Deep Mutational Scanning of Hemagglutinin Helps Predict Evolutionary Fates of Human H3N2 Influenza Variants." *Proceedings of the National Academy of Sciences of the United States of America* 115 (35): E8276–85. https://doi.org/10.1073/pnas.1806133115.
Lee, Mi-Kyung, Hee-Eun Kim, Eun-Byeol Park, Janghyun Lee, Ki-Hun Kim, Kyungeun Lim, Seoyun Yum, et al. 2016. "Structural Features of Influenza A Virus Panhandle RNA Enabling the Activation of RIG-I Independently of 5′-Triphosphate." *Nucleic Acids Research* 44 (17): 8407–16. https://doi.org/10.1093/nar/gkw525.
Li, Jing, Meng Yu, Weinan Zheng, and Wenjun Liu. 2015. "Nucleocytoplasmic Shuttling of Influenza a Virus Proteins." *Viruses*. MDPI AG. https://doi.org/10.3390/v7052668.
Li, S Q, M Orlich, and R Rott. 1990. "Generation of Seal Influenza Virus Variants Pathogenic for Chickens, Because of Hemagglutinin Cleavage Site Changes." *Journal of Virology* 64 (7): 3297–3303. http://www.ncbi.nlm.nih.gov/pubmed/2191148.
Lim, Chun Shen, and Chris M. Brown. 2018. "Know Your Enemy: Successful Bioinformatic Approaches to Predict Functional RNA Structures in Viral RNAs." *Frontiers in Microbiology* 8 (January): 2582. https://doi.org/10.3389/fmicb.2017.02582.
Lin, Chenyu, and Wayne O Miles. 2019. "Beyond CLIP: Advances and Opportunities to Measure RBP-RNA and RNA-RNA Interactions." *Nucleic Acids Research* 47 (11): 5490–5501. https://doi.org/10.1093/nar/gkz295.
Liu, Dong, Zhujun Zhang, Lihong He, Zhao Gao, Juan Li, Min Gu, Jiao Hu, Xiaoquan Wang, Xiaowen Liu, and Xiufan Liu. 2018. "Characteristics of the Emerging Chicken-Origin Highly Pathogenic H7N9 Viruses: A New Threat to Public Health and Poultry Industry." *Journal of Infection* 76 (2): 217–20. https://doi.org/10.1016/j.jinf.2017.09.005.
Lo, Chun-Yeung, Yun-Sang Tang, and Pang-Chui Shaw. 2018. "Structure and Function of Influenza Virus Ribonucleoprotein." In , 95–128. Springer, Singapore. https://doi.org/10.1007/978-981-10-8456-0_5.
Lowen, Anice C. 2017. "Constraints, Drivers, and Implications of Influenza A Virus Reassortment." *Annual Review of Virology* 4 (1): 105–21. https://doi.org/10.1146/annurev-virology-101416-041726.

Luo, G X, W Luytjes, M Enami, and P Palese. 1991. “The Polyadenylation Signal of Influenza Virus RNA Involves a Stretch of Uridines Followed by the RNA Duplex of the Panhandle Structure.” *Journal of Virology* 65 (6): 2861–67. http://www.ncbi.nlm.nih.gov/pubmed/2033659.

Lyons, Daniel M, and Adam S Lauring. 2018. “Mutation and Epistasis in Influenza Virus Evolution.” *Viruses* 10 (8). https://doi.org/10.3390/v10080407.

Mailler, Elodie, Jean-Christophe Paillart, Roland Marquet, Redmond P. Smyth, and Valerie Vivet-Boudou. 2019. “The Evolution of RNA Structural Probing Methods: From Gels to next-Generation Sequencing.” *Wiley Interdisciplinary Reviews: RNA* 10 (2): e1518. https://doi.org/10.1002/wrna.1518.

Mair, Caroline M., Kai Ludwig, Andreas Herrmann, and Christian Sieben. 2014. “Receptor Binding and PH Stability — How Influenza A Virus Hemagglutinin Affects Host-Specific Virus Infection.” *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1838 (4): 1153–68. https://doi.org/10.1016/J.BBAMEM.2013.10.004.

Mani, Ramamurthy, Robert P. St. Onge, John L. Hartman IV, Guri Giaever, and Frederick P. Roth. 2008. “Defining Genetic Interaction.” *Proceedings of the National Academy of Sciences of the United States of America* 105 (9): 3461–66. https://doi.org/10.1073/pnas.0712255105.

Martínez-Alonso, Mónica, Narin Hengrung, and Ervin Fodor. 2016. “RNA-Free and Ribonucleoprotein-Associated Influenza Virus Polymerases Directly Bind the Serine-5-Phosphorylated Carboxyl-Terminal Domain of Host RNA Polymerase II.” *Journal of Virology* 90 (13): 6014–21. https://doi.org/10.1128/jvi.00494-16.

Maurer-Stroh, Sebastian, Raphael T C Lee, Vithiagaran Gunalan, and Frank Eisenhaber. 2013. “The Highly Pathogenic H7N3 Avian Influenza Strain from July 2012 in Mexico Acquired an Extended Cleavage Site through Recombination with Host 28S RRNA.” *Virology Journal* 10 (1): 139. https://doi.org/10.1186/1743-422X-10-139.

Mehedi, Masfique, Thomas Hoenen, Shelly Robertson, Stacy Ricklefs, Michael A Dolan, Travis Taylor, Darryl Falzarano, Hideki Ebihara, Stephen F Porcella, and Heinz Feldmann. 2013. “Ebola Virus RNA Editing Depends on the Primary Editing Site Sequence and an Upstream Secondary Structure.” *PLOS Pathogens* 9 (10): e1003677. https://doi.org/10.1371/journal.ppat.1003677.

Mercier, Philippe Le, and Daniel Kolakofsky. 2019. “Bipartite Promoters and RNA Editing of Paramyxoviruses and Filoviruses.” *Rna* 25 (3): 279–85. https://doi.org/10.1261/rna.068825.118.

Merino, Edward J., Kevin A. Wilkinson, Jennifer L. Coughlan, and Kevin M. Weeks. 2005. “RNA Structure Analysis at Single Nucleotide Resolution by Selective 2‘-Hydroxyl Acylation and Primer Extension (SHAPE).” *Journal of the American Chemical Society* 127 (12): 4223–31. https://doi.org/10.1021/ja043822v.

Monne, Isabella, Alice Fusaro, Martha I Nelson, Lebana Bonfanti, Paolo Mulatti, Joseph Hughes, Pablo R Murcia, et al. 2014. “Emergence of a Highly Pathogenic Avian Influenza Virus from a Low-Pathogenic Progenitor.” *Journal of Virology* 88 (8): 4375–88. https://doi.org/10.1128/JVI.03181-13.

Mostafa, Ahmed, Elsayed M Abdelwhab, Thomas C Mettenleiter, and Stephan Pleschka. 2018a. “Zoonotic Potential of Influenza A Viruses: A Comprehensive Overview.” *Viruses* 10 (9). https://doi.org/10.3390/v10090497.

Mostafa, Ahmed, Elsayed Abdelwhab, Thomas Mettenleiter, and Stephan Pleschka. 2018b. “Zoonotic Potential of Influenza A Viruses: A Comprehensive Overview.” *Viruses* 10 (9): 497. https://doi.org/10.3390/v10090497.

Moumen, Abdeladim, Lucette Polomack, Torsten Unge, Michel Véron, Henri Buc, and Matteo Negroni. 2003. “Evidence for a Mechanism of Recombination during Reverse Transcription Dependent on the Structure of the Acceptor RNA.” *Journal of Biological Chemistry* 278 (18): 15973–82. https://doi.org/10.1074/jbc.M212306200.

Munster, V. J., and R. A.M. Fouchier. 2009. “Avian Influenza Virus: Of Virus and Bird Ecology.” *Vaccine*. https://doi.org/10.1016/j.vaccine.2009.02.082.

Muramoto, Y., T. Noda, E. Kawakami, R. Akkina, and Y. Kawaoka. 2013. “Identification of Novel Influenza A Virus Proteins Translated from PA MRNA.” *Journal of Virology* 87 (5): 2455–62. https://doi.org/10.1128/jvi.02656-12.

Nao N, Yamagishi J, Miyamoto H, Tsuda Y Igarashi M, Manzoor R, Ohnuma A, Kishida N Furuyama W, Shigeno A, Kajihara M, and Takada A. Yoshida R. 2017. “Genetic Predisposition To Acquire a Polybasic Cleavage Site for Highly.” *MBio* 8 (1): 1–15. https://doi.org/10.1128/mBio.02298-

16.
Nao, Naganori, Junya Yamagishi, Hiroko Miyamoto, Manabu Igarashi, Rashid Manzoor, Aiko Ohnuma, Yoshimi Tsuda, et al. 2017. “Genetic Predisposition To Acquire a Polybasic Cleavage Site for Highly Pathogenic Avian Influenza Virus Hemagglutinin.” *MBio* 8 (1). https://doi.org/10.1128/mBio.02298-16.
Nayak, Debi P., Eric Ka Wai Hui, and Subrata Barman. 2004. “Assembly and Budding of Influenza Virus.” *Virus Research* 106 (2 SPEC.ISS.): 147–65. https://doi.org/10.1016/j.virusres.2004.08.012.
Nelson, Martha I., and Edward C. Holmes. 2007. “The Evolution of Epidemic Influenza.” *Nature Reviews Genetics*. https://doi.org/10.1038/nrg2053.
Neumann, G, M R Castrucci, and Y Kawaoka. 1997. “Nuclear Import and Export of Influenza Virus Nucleoprotein.” *Journal of Virology* 71 (12): 9690–9700. http://www.ncbi.nlm.nih.gov/pubmed/9371635.
Noda, Takeshi. 2011. “Native Morphology of Influenza Virions.” *Frontiers in Microbiology* 2: 269. https://doi.org/10.3389/fmicb.2011.00269.
Nogales, Aitor, Luis Martinez-Sobrido, David J Topham, and Marta L DeDiego. 2018. “Modulation of Innate Immune Responses by the Influenza A NS1 and PA-X Proteins.” *Viruses* 10 (12). https://doi.org/10.3390/v10120708.
Nshogozabahizi, Jean Claude, Jonathan Dench, and Stéphane Aris-Brosou. 2017. “Widespread Historical Contingency in Influenza Viruses.” *Genetics* 205 (1): 409–20. https://doi.org/10.1534/genetics.116.193979.
Ohuchi, Masanobu, Michaela Orlich, Reiko Ohuchi, Barry E.J. Simpson, Wolfgang Garten, Hans-Dieter Klenk, and Rudolf Rott. 1989a. “Mutations at the Cleavage Site of the Hemagglutinin Alter the Pathogenicity of Influenza Virus a/Chick/Penn/83 (H5N2).” *Virology* 168 (2): 274–80. https://doi.org/10.1016/0042-6822(89)90267-5.
Ohuchi, Masanobu, Michaela Orlich, Reiko Ohuchi, Barry E.J. Simpson, Wolfgang Garten, Hans Dieter Klenk, and Rudolf Rott. 1989b. “Mutations at the Cleavage Site of the Hemagglutinin Alter the Pathogenicity of Influenza Virus a/Chick/Penn/83 (H5N2).” *Virology* 168 (2): 274–80. https://doi.org/10.1016/0042-6822(89)90267-5.
Olsen, Björn, Vincent J Munster, Anders Wallensten, Jonas Waldenström, Albert D M E Osterhaus, and Ron A M Fouchier. 2006. “Global Patterns of Influenza a Virus in Wild Birds.” *Science (New York, N.Y.)* 312 (5772): 384–88. https://doi.org/10.1126/science.1122438.
Olspert, Allan, Betty Y-W Chung, John F Atkins, John P Carr, and Andrew E Firth. 2015. “Transcriptional Slippage in the Positive-sense RNA Virus Family *Potyviridae*.” *EMBO Reports* 16 (8): 995–1004. https://doi.org/10.15252/embr.201540509.
Parsch, J, J M Braverman, and W Stephan. 2000. “Comparative Sequence Analysis and Patterns of Covariation in {RNA} Secondary Structures.” *Genetics* 154 (2): 909–21.
Pasick, John, Katherine Handel, John Robinson, John Copps, Deidre Ridd, Kevin Hills, Helen Kehler, et al. 2005. “Intersegmental Recombination between the Haemagglutinin and Matrix Genes Was Responsible for the Emergence of a Highly Pathogenic H7N3 Avian Influenza Virus in British Columbia.” *The Journal of General Virology* 86 (Pt 3): 727–31. https://doi.org/10.1099/vir.0.80478-0.
Pauly, Matthew D, Megan C Procario, and Adam S Lauring. 2017. “A Novel Twelve Class Fluctuation Test Reveals Higher than Expected Mutation Rates for Influenza A Viruses.” Edited by Karla Kirkegaard. *ELife* 6 (May): e26437. https://doi.org/10.7554/eLife.26437.
Perdue, M L, M Garcia, J Beck, M Brugh, and D E Swayne. 1996. “An Arg-Lys Insertion at the Hemagglutinin Cleavage Site of an H5N2 Avian Influenza Isolate.” *Virus Genes* 12 (1): 77–84. http://www.ncbi.nlm.nih.gov/pubmed/8879123.
Perdue, M L, M García, D Senne, and M Fraire. 1997. “Virulence-Associated Sequence Duplication at the Hemagglutinin Cleavage Site of Avian Influenza Viruses.” *Virus Res.* 49 (2): 173–86.
Perdue, Michael L., Maricarmen García, Dennis Senne, and M. Fraire. 1997. “Virulence-Associated Sequence Duplication at the Hemagglutinin Cleavage Site of Avian Influenza Viruses.” *Virus Research* 49 (2): 173–86. https://doi.org/10.1016/S0168-1702(97)01468-8.
Perdue, Michael L., John W. Latimer, and John M. Crawford. 1995. “A Novel Carbohydrate Addition Site on the Hemagglutinin Protein of a Highly Pathogenic H7 Subtype Avian Influenza Virus.” *Virology* 213 (1): 276–81. https://doi.org/10.1006/VIRO.1995.1571.
Perdue, Michael L, Maricarmen García, Dennis Senne, and M Fraire. 1997. “Virulence-Associated

Sequence Duplication at the Hemagglutinin Cleavage Site of Avian Influenza Viruses." *Virus Research* 49 (2): 173–86. https://doi.org/10.1016/S0168-1702(97)01468-8.
Pflug, Alexander, Delphine Guilligay, Stefan Reich, and Stephen Cusack. 2014. "Structure of Influenza A Polymerase Bound to the Viral RNA Promoter." *Nature* 516 (7531): 355–60. https://doi.org/10.1038/nature14008.
Pflug, Alexander, Maria Lukarska, Patricia Resa-Infante, Stefan Reich, and Stephen Cusack. 2017. "Structural Insights into RNA Synthesis by the Influenza Virus Transcription-Replication Machine." *Virus Research* 234 (April): 103–17. https://doi.org/10.1016/J.VIRUSRES.2017.01.013.
Pielak, Rafal M., and James J. Chou. 2011. "Influenza M2 Proton Channels." *Biochimica et Biophysica Acta - Biomembranes*. https://doi.org/10.1016/j.bbamem.2010.04.015.
Plotkin, Joshua B., and Jonathan Dushoff. 2003. "Codon Bias and Frequency-Dependent Selection on the Hemagglutinin Epitopes of Influenza A Virus." *Proceedings of the National Academy of Sciences of the United States of America* 100 (12): 7152–57. https://doi.org/10.1073/pnas.1132114100.
Qi, Wenbao, Weixin Jia, Di Liu, Jing Li, Yuhai Bi, Shumin Xie, Bo Li, et al. 2017. "Emergence and Adaptation of a Novel Highly Pathogenic H7N9 Influenza Virus in Birds and Humans from a 2013 Human-Infecting Low-Pathogenic Ancestor." *Journal of Virology* 92 (2). https://doi.org/10.1128/jvi.00921-17.
Richard, Mathilde, Ron Fouchier, Isabella Monne, Thijs Kuiken, Isabella Monne, and Thijs Kuiken. 2017. "Mechanisms and Risk Factors for Mutation from Low to Highly Pathogenic Avian Influenza Virus." *EFSA Supporting Publications* 14 (10). https://doi.org/10.2903/sp.efsa.2017.EN-1287.
Riel, Debby Van, Vincent J. Munster, Emmie De Wit, Guus F. Rimmelzwaan, Ron A.M. Fouchier, Ab D.M.E. Osterhaus, and Thijs Kuiken. 2006. "H5N1 Virus Attachment to Lower Respiratory Tract." *Science* 312 (5772): 399. https://doi.org/10.1126/science.1125548.
Rossman, J. S., G. P. Leser, and R. A. Lamb. 2012. "Filamentous Influenza Virus Enters Cells via Macropinocytosis." *Journal of Virology* 86 (20): 10950–60. https://doi.org/10.1128/jvi.05992-11.
Rossman, Jeremy S., and Robert A. Lamb. 2011. "Influenza Virus Assembly and Budding." *Virology*. https://doi.org/10.1016/j.virol.2010.12.003.
Ruigrok, Rob WH, Thibaut Crépin, Darren J. Hart, and Stephen Cusack. 2010. "Towards an Atomic Resolution Understanding of the Influenza Virus Replication Machinery." *Current Opinion in Structural Biology*. https://doi.org/10.1016/j.sbi.2009.12.007.
Sanjuan, R., M. R. Nebot, N. Chirico, L. M. Mansky, and R. Belshaw. 2010. "Viral Mutation Rates." *Journal of Virology* 84 (19): 9733–48. https://doi.org/10.1128/jvi.00694-10.
Sanjuán, Rafael, and Pilar Domingo-Calap. 2016. "Mechanisms of Viral Mutation." *Cellular and Molecular Life Sciences : CMLS* 73 (23): 4433–48. https://doi.org/10.1007/S00018-016-2299-6.
Sanjuán, Rafael, and Santiago F. Elena. 2006. "Epistasis Correlates to Genomic Complexity." *Proceedings of the National Academy of Sciences of the United States of America* 103 (39): 14402–5. https://doi.org/10.1073/pnas.0604543103.
SantaLucia, J. 1998. "A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-Neighbor Thermodynamics." *Proceedings of the National Academy of Sciences of the United States of America* 95 (4): 1460–65. https://doi.org/10.1073/pnas.95.4.1460.
Seekings, Amanda H., Wendy A. Howard, Alejandro Nuñéz, Marek J. Slomka, Ashley C. Banyard, Daniel Hicks, Richard J. Ellis, et al. 2020. "The Emergence of H7N7 Highly Pathogenic Avian Influenza Virus from Low Pathogenicity Avian Influenza Virus Using an in Ovo Embryo Culture Model." *Viruses* 12 (9). https://doi.org/10.3390/v12090920.
Shabman, Reed S., Omar J. Jabado, Chad E. Mire, Timothy B. Stockwell, Megan Edwards, Milind Mahajan, Thomas W. Geisbert, and Christopher F. Basler. 2014. "Deep Sequencing Identifies Noncanonical Editing of Ebola and Marburg Virus RNAs in Infected Cells." *MBio* 5 (6). https://doi.org/10.1128/mbio.02011-14.
Simon-Loriere, Etienne, Roman Galetto, Meriem Hamoudi, John Archer, Pierre Lefeuvre, Darren P Martin, David L Robertson, and Matteo Negroni. 2009. "Molecular Mechanisms of Recombination Restriction in the Envelope Gene of the Human Immunodeficiency Virus." *PLOS Pathogens* 5 (5): e1000418. https://doi.org/10.1371/journal.ppat.1000418.
Simon-Loriere, Etienne, Darren P Martin, Kevin M Weeks, and Matteo Negroni. 2010. "RNA Structures Facilitate Recombination-Mediated Gene Swapping in HIV-1." *Journal of Virology*

84 (24): 12675–82. https://doi.org/10.1128/JVI.01302-10.
Spackman, E. 2009. “The Ecology of Avian Influenza Virus in Wild Birds: What Does This Mean for Poultry?” *Poultry Science* 88 (4): 847–50. https://doi.org/10.3382/ps.2008-00336.
Stallknecht, D. E., M. T. Kearney, S. M. Shane, and P. J. Zwank. 1990. “Effects of PH, Temperature, and Salinity on Persistence of Avian Influenza Viruses in Water.” *Avian Diseases* 34 (2): 412. https://doi.org/10.2307/1591429.
Steinhauer, David A. 1999. “Role of Hemagglutinin Cleavage for the Pathogenicity of Influenza Virus.” *Virology* 258 (1): 1–20. https://doi.org/10.1006/viro.1999.9716.
Suarez, David Lee, Dennis A. Senne, Jill Banks, Ian H. Brown, Steve C. Essen, Chang Won Lee, Ruth J. Manvell, et al. 2004. “Recombination Resulting in Virulence Shift in Avian Influenza Outbreak, Chile.” *Emerging Infectious Diseases* 10 (4): 693–99. https://doi.org/10.3201/eid1004.030396.
Subbarao, Kanta, Alexander Klimov, Jacqueline Katz, Helen Regnery, Wilina Lim, Henrietta Hall, Michael Perdue, et al. 1998. “Characterization of an Avian Influenza A (H5N1) Virus Isolated from a Child with a Fatal Respiratory Illness.” *Science* 279 (5349): 393–96. https://doi.org/10.1126/science.279.5349.393.
Sun, Xiangjie, Jessica A Belser, and Terrence M Tumpey. 2016. “A Novel Eight Amino Acid Insertion Contributes to the Hemagglutinin Cleavability and the Virulence of a Highly Pathogenic Avian Influenza A (H7N3) Virus in Mice.” *Virology* 488 (April): 120–28. https://doi.org/10.1016/j.virol.2015.10.032.
Swayne, D E, and M Pantin-Jackwood. 2006. “Pathogenicity of Avian Influenza Viruses in Poultry.” *Developments in Biologicals* 124: 61–67. http://www.ncbi.nlm.nih.gov/pubmed/16447495.
Swayne, David E. 2007. “Understanding the Complex Pathobiology of High Pathogenicity Avian Influenza Viruses in Birds.” *Avian Diseases* 51 (s1): 242–49. https://doi.org/10.1637/7763-110706-regr.1.
Tomezsko, Phillip J., Vincent D.A. Corbin, Paromita Gupta, Harish Swaminathan, Margalit Glasgow, Sitara Persad, Matthew D. Edwards, et al. 2020. “Determination of RNA Structural Diversity and Its Role in HIV-1 RNA Splicing.” *Nature 2020 582:7812* 582 (7812): 438–42. https://doi.org/10.1038/s41586-020-2253-5.
Tria, F., S. Pompei, and V. Loreto. 2013. “Dynamically Correlated Mutations Drive Human Influenza A Evolution.” *Scientific Reports* 3: 2705. https://doi.org/10.1038/srep02705.
Vandegrift, Kurt J., Susanne H. Sokolow, Peter Daszak, and A. Marm Kilpatrick. 2010. “Ecology of Avian Influenza Viruses in a Changing World.” *Annals of the New York Academy of Sciences*. Blackwell Publishing Inc. https://doi.org/10.1111/j.1749-6632.2010.05451.x.
Veits, Jutta, Siegfried Weber, Olga Stech, Angele Breithaupt, Marcus Gräber, Sandra Gohrbandt, Jessica Bogs, et al. 2012. “Avian Influenza Virus Hemagglutinins H2, H4, H8, and H14 Support a Highly Pathogenic Phenotype.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (7): 2579–84. https://doi.org/10.1073/pnas.1109397109.
Velthuis, Aartjan J. W. te, and Ervin Fodor. 2016. “Influenza Virus RNA Polymerase: Insights into the Mechanisms of Viral RNA Synthesis.” *Nature Reviews Microbiology* 14 (8): 479–93. https://doi.org/10.1038/nrmicro.2016.87.
Velthuis, Aartjan J W Te, and Ervin Fodor. 2016. “Influenza Virus RNA Polymerase: Insights into the Mechanisms of Viral RNA Synthesis.” *Nature Reviews. Microbiology* 14 (8): 479–93. https://doi.org/10.1038/nrmicro.2016.87.
Visher, Elisa, Shawn E. Whitefield, John T. McCrone, William Fitzsimmons, and Adam S. Lauring. 2016. “The Mutational Robustness of Influenza A Virus.” *PLoS Pathogens* 12 (8). https://doi.org/10.1371/journal.ppat.1005856.
Vlugt, Corey De, Dorota Sikora, and Martin Pelchat. 2018. “Insight into Influenza: A Virus Cap-Snatching.” *Viruses* 10 (11): 641. https://doi.org/10.3390/v10110641.
Wahlgren, John. 2011. “Influenza A Viruses: An Ecology Review.” *Infection Ecology & Epidemiology* 1 (1): 6004. https://doi.org/10.3402/iee.v1i0.6004.
Wang, Jingfeng, Jinghua Li, Lili Zhao, Mengmeng Cao, and Tao Deng. 2017. “Dual Roles of the Hemagglutinin Segment-Specific Noncoding Nucleotides in the Extended Duplex Region of the Influenza A Virus RNA Promoter.” *Journal of Virology* 91 (1). https://doi.org/10.1128/jvi.01931-16.
Wang, Qiao, Ranran Liu, Qinghe Li, Fei Wang, Bo Zhu, Maiqing Zheng, Huanxian Cui, Jie Wen, and Guiping Zhao. 2019. “Host Cell Interactome of PB1 N40 Protein of H5N1 Influenza A Virus in

Chicken Cells." *Journal of Proteomics* 197 (April): 34–41. https://doi.org/10.1016/j.jprot.2019.02.011.
Weber, Friedemann, Georg Kochs, Simone Gruber, and Otto Haller. 1998. "A Classical Bipartite Nuclear Localization Signal on Thogoto and Influenza A Virus Nucleoproteins." *Virology* 250 (1): 9–18. https://doi.org/10.1006/viro.1998.9329.
Webster, R. G., W. J. Bean, O. T. Gorman, T. M. Chambers, and Y. Kawaoka. 1992. "Evolution and Ecology of Influenza A Viruses." *Microbiological Reviews*. https://doi.org/10.1111/j.1751-0813.1992.tb09940.x.
Williams, Graham D., Dana Townsend, Kristine M. Wylie, Preston J. Kim, Gaya K. Amarasinghe, Sebla B. Kutluay, and Adrianus C. M. Boon. 2018. "Nucleotide Resolution Mapping of Influenza A Virus Nucleoprotein-RNA Interactions Reveals RNA Features Required for Replication." *Nature Communications* 9 (1): 465. https://doi.org/10.1038/s41467-018-02886-w.
Wise, Helen M., Edward C. Hutchinson, Brett W. Jagger, Amanda D. Stuart, Zi H. Kang, Nicole Robb, Louis M. Schwartzman, et al. 2012. "Identification of a Novel Splice Variant Form of the Influenza A Virus M2 Ion Channel with an Antigenically Distinct Ectodomain." Edited by Andrew Pekosz. *PLoS Pathogens* 8 (11): e1002998. https://doi.org/10.1371/journal.ppat.1002998.
Wong, Sook San, Sun Woo Yoon, Mark Zanin, Min Suk Song, Christine Oshansky, Hassan Zaraket, Stephanie Sonnberg, et al. 2014. "Characterization of an H4N2 Influenza Virus from Quails with a Multibasic Motif in the Hemagglutinin Cleavage Site." *Virology* 468–470 (November): 72–80. https://doi.org/10.1016/j.virol.2014.07.048.
Xue, Katherine S., Louise H. Moncla, Trevor Bedford, and Jesse D. Bloom. 2018. "Within-Host Evolution of Human Influenza Virus." *Trends in Microbiology*. Elsevier Ltd. https://doi.org/10.1016/j.tim.2018.02.007.
Yamayoshi, Seiya, Mariko Watanabe, Hideo Goto, and Yoshihiro Kawaoka. 2015. "Identification of a Novel Viral Protein Expressed from the PB2 Segment of Influenza A Virus." *Journal of Virology* 90 (1): 444–56. https://doi.org/10.1128/JVI.02175-15.
Yang, Yang, Gary Wong, Liuqing Yang, Shuguang Tan, Jianming Li, Bing Bai, Zhixiang Xu, et al. 2019. "Comparison between Human Infections Caused by Highly and Low Pathogenic H7N9 Avian Influenza Viruses in Wave Five: Clinical and Virological Findings." *Journal of Infection* 78 (3): 241–48. https://doi.org/10.1016/J.JINF.2019.01.005.
Zhang, Fuchun, Yuhai Bi, Jian Wang, Gary Wong, Weifeng Shi, Fengyu Hu, Yang Yang, et al. 2017. "Human Infections with Recently-Emerging Highly Pathogenic H7N9 Avian Influenza Virus in China." *Journal of Infection* 75 (1): 71–75. https://doi.org/10.1016/j.jinf.2017.04.001.
Zhang, Jiahao, Kaixiong Ma, Bo Li, Yiqun Chen, Ziwen Qiu, Jinchao Xing, Jinyu Huang, et al. 2021. "A Risk Marker of Tribasic Hemagglutinin Cleavage Site in Influenza A (H9N2) Virus." *Communications Biology* 4 (1). https://doi.org/10.1038/S42003-020-01589-7.
Zhang, Jinwei, and Robert Landick. 2016. "A Two-Way Street: Regulatory Interplay between RNA Polymerase and Nascent RNA Structure." *Trends in Biochemical Sciences* 41 (4): 293–310. https://doi.org/10.1016/j.tibs.2015.12.009.
Zheng, Hongyong, Hye Annie Lee, Peter Palese, and Adolfo García-Sastre. 1999. "Influenza A Virus RNA Polymerase Has the Ability To Stutter at the Polyadenylation Site of a Viral RNA Template during RNA Replication." *Journal of Virology* 73 (6): 5240. https://doi.org/10.1128/JVI.73.6.5240-5243.1999.

Résumé :

Les virus de l'influenza aviaire (AIV) constituent une menace importante pour la santé animale et humaine. Ces virus peuvent être classés en deux groupes en fonction de leur pathogénicité observée chez les volailles. Certains virus influenza aviaires faiblement pathogènes (LPAIV) de sous-types H5 ou H7 ont la capacité d'évoluer et d'acquérir une séquence polybasique dans le site de clivage de l'hémagglutinine (HA), ce qui a pour effet d'augmenter considérablement leur pathogénicité, permettant ainsi l'émergence de virus influenza aviaire hautement pathogène (HPAIV). Le mécanisme génétique de cette évolution est mal compris et certaines questions restent sans réponse. Est-ce que tous les LPAIV ont-ils la même probabilité d'évoluer vers un phénotype hautement pathogène ? Si la probabilité d'émergence d'HPAIV est différente parmi les LPAIV, existe-t-il des caractéristiques spécifiques sur le segment génomique HA qui peuvent influencer l'évolution LPAIV/HPAIV ? Plusieurs études ont suggéré qu'une structure secondaire d'ARN conservée de type tige-boucle d'ARN, située dans la région codant le site de clivage HA, pourrait être impliquée dans l'évolution LPAIV/HPAIV. Dans cette thèse, nous avons étudié la présence possible d'éléments génomiques viraux pouvant influencer les taux de mutation de la polymérase virale, ce qui pourrait potentiellement conduire à l'émergence de HPAIV.

Tout d'abord, une analyse phylogénétique des séquences nucléotidiques HA du sous-type H5/7 du virus de l'influenza aviaire, combinée à des prédictions de structures secondaires de l'ARN, a été réalisée pour révéler les motifs structurels potentiels impliqués dans la transition LPAIV/HPAIV. Grâce à cette approche bioinformatique, l'analyse par clustering des structures d'ARN prédites, associée aux données phylogénétiques, a révélé que la plupart des ancêtres ayant conduit à l'émergence d'HPAIV par recombinaison partagent la même structure prédite d'ARN au niveau de la région HA1/HA2. Cette structure d'ARN pourrait favoriser la recombinaison génétique, conduisant à l'acquisition d'un motif de clivage HA polybasique. De plus, aucune corrélation n'a été identifiée entre la stabilité thermodynamique de la structure ARN conservée et l'émergence du virus HPAIV.

Ensuite, nous avons mis au point un système de génétique inverse permettant de générer des virus H7N1 contenant une autre séquence HA complète sous la forme d'un transgène situé à l'extrémité du segment codant pour la protéine PA virale. Dans ce modèle, la séquence transgénique HA insérée ne peut être traduite et est libérée de toute pression de sélection sur la fonction protéique, permettant ainsi d'étudier l'influence de tous les types d'environnement nucléotidique sur les erreurs de la polymérase virale. Nous avons réussi à produire et à caractériser plusieurs virus contenant des transgènes avec motifs différents au niveau de la région codant le site de clivage HA. En utilisant ce système lors d'infections expérimentales, suivi de séquençage profond du transgène, nous avons étudié comment différents environnements nucléotidiques pouvaient influencer le taux d'erreur de la polymérase virale. Nous avons pu identifier un motif nucléotidique minimal favorable aux événements d'insertion et de substitution pouvant conduire à l'acquisition d'un site de clivage HA polybasique, et donc à l'émergence HPAIV. De plus, plusieurs événements de recombinaison avec l'ARN ribosomal et des segments viraux ont été observés avec notre système. Ce travail permet de mieux comprendre les éléments génétiques viraux qui influencent les erreurs de la polymérase virale et l'émergence de l'HPAIV.

Abstract:

Avian influenza viruses (AIV) constitute a significant threat to both animal and human health. These viruses can be distinguished into two groups in terms of their observed pathogenicity in poultry. Some low pathogenic influenza viruses (LPAIV) of H5 or H7 subtypes have the ability to evolve and acquire a polybasic sequence in the hemagglutinin (HA) cleavage site, which will enhance their pathogenicity and be responsible for the emergence of highly pathogenic avian influenza viruses (HPAIV). The genetic mechanism of this evolution is poorly understood, and some questions remain unanswered. Within the H5 and H7 subtypes, have all the LPAIV the same probability to evolve towards a highly pathogenic phenotype? If the probability of HPAIV emergence among LPAIV is significantly different, are there specific features on the HA genomic segment that can influence the evolution to HPAIV? Several studies suggested that specific determinants of a conserved RNA stem-loop located at the HA cleavage site-encoding region could influence the LPAIV-to-HPAIV evolution. In this thesis, we investigated the possible presence of viral genomic elements that can influence the viral mutation rates, which could potentially lead to HPAIV emergence.

First, a phylogenetic analysis of HA nucleotide sequences of H5/7 subtype-AIV from avian hosts combined with RNA-secondary structures predictions was performed to reveal potential structural motifs involved in LPAIV/HPAIV transition. Thanks to this bioinformatics approach, RNA structure clustering analysis coupled with phylogenetic data revealed that most of the ancestors leading to H7-HPAIV emergences via recombination shared the same viral RNA structure at the HA1/HA2 boundary region. The identified shared predicted RNA structure present in the HA of H7 viruses could promote genetic recombination, leading to the acquisition of HA polybasic cleavage site motif. Moreover, no correlation was identified between the thermodynamic stability of the conserved RNA structure and HPAIV emergence.

We developed a reverse-genetic system that generates H7N1 influenza viruses harbouring a full HA sequence as a transgene located at the end of the PA encoding segment. In this model, the inserted HA sequence cannot be translated and is released from any selection pressure at the level of protein function. We successfully rescued and characterized infectious viruses containing transgene with various HA cleavage site encoding motifs. Using this system and deep-sequencing, we tested how different nucleotide environments could influence the viral polymerase error rate. We were able to identify a minimal nucleotide motif that is favourable to insertion and substitution events that could potentially lead to HPAIV emergence. Moreover, recombination events with both ribosomal RNA and viral segments have been observed in our system. This work permits us to better comprehend the viral genetic elements that influence viral polymerase errors and HPAIV emergence.