



**HAL**  
open science

# Analyse de la variabilité génétique des virus influenza aviaires par séquençage à très haut débit.

Guillaume Croville

► **To cite this version:**

Guillaume Croville. Analyse de la variabilité génétique des virus influenza aviaires par séquençage à très haut débit.. Sciences du Vivant [q-bio]. 2013. hal-01502556

**HAL Id: hal-01502556**

**<https://ephe.hal.science/hal-01502556>**

Submitted on 5 Apr 2017

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

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



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE

ÉCOLE PRATIQUE DES HAUTES ÉTUDES

Sciences de la Vie et de la Terre

MÉMOIRE

présenté

par

Croville Guillaume

pour l'obtention du Diplôme de l'École Pratique des Hautes Études

## Analyse de la variabilité génétique des virus influenza aviaires par séquençage à très haut débit

soutenu le (jour, mois, année) devant le jury suivant :

N. (nom et prénom) – Président

N. (nom et prénom) – Rapporteur

N. (nom et prénom) – Examineur

Dr. Guérin Jean-Luc – Tuteur scientifique

Dr. Dupressoir Thierry – Tuteur pédagogique

Mémoire préparé sous la direction de :

Dr. Guérin Jean-Luc

UMR 1225 INRA-ENVY IHAP – Directeur : Pr. Schelcher François

et de

Dr. Dupressoir Thierry

Laboratoire de pathologie comparée des invertébrés, EPHE (Sciences de la Vie et de la Terre)

UMR 1333 INRA-UM2 DGIMI – Directeur : Dr. Tailliez Patrick

# Table des matières

Table des matières .....	1
Liste des abréviations .....	5
Liste des figures .....	7
Liste des tableaux .....	8
Introduction bibliographique .....	9
1 Les virus influenza A .....	10
1.1 Un peu d'histoire.....	10
1.2 Qu'est-ce que l'influenza A ?.....	11
1.2.1 Classification .....	11
1.2.2 Structure des virions.....	11
1.2.3 Organisation et structure du génome .....	12
1.2.4 Glycoprotéines de surface.....	12
1.2.5 Étapes de la réplication virale.....	14
1.2.5.1 Attachement du virus .....	14
1.2.5.2 Entrée du virus et endocytose .....	15
1.2.5.3 Import nucléaire des RNP.....	15
1.2.5.4 Synthèse des ARNm viraux .....	15
1.2.5.5 Synthèse des ARNc : futur génome viral .....	16
1.2.5.6 Export des RNP.....	16
1.2.5.7 Assemblage et bourgeonnement des virions.....	16
1.3 Écologie des virus influenza A.....	16
1.3.1 L'influenza aviaire chez les volailles.....	17
1.3.1.1 Signes cliniques et pathogénicité .....	18
1.3.1.2 Transmission du virus.....	19

1.3.1.3	Virulence.....	19
1.3.2	L'influenza aviaire chez les canards.....	20
1.4	L'influenza aviaire .....	20
1.5	Évolution des virus influenza A.....	23
1.6	Conséquences de la variabilité des virus influenza .....	24
2	Analyse des sous-populations virales.....	25
2.1	Nouvelles générations de séquençage.....	25
2.2	Pyroséquençage 454 .....	27
3	Projets scientifiques.....	29
Première partie expérimentale.....		30
1	Introduction .....	31
2	Résultats.....	32
2.1	Historique et premières analyses .....	32
2.2	Analyse des séquences virales des prélèvements de canards et de dindes.....	32
2.3	Analyse des neuraminidases par PCR quantitative en temps réel.....	33
2.4	Séquençage à très haut débit du génome complet des virus de canards et de dindes .....	33
3	Discussion.....	35
Deuxième partie expérimentale .....		38
1	Introduction .....	39
2	Résultats.....	39
2.1	Analyse des séquences « Sanger ».....	40
2.2	Présentation des résultats de pyroséquençage 454 .....	41
2.3	Redistribution des populations minoritaires initialement présentes au sein de la population Wt.....	41
2.3.1	Précurseurs des gènes 2, 7 et 8 du virus Ad dans la population Wt.....	41
2.3.1	Disparition d'allèles minoritaires de la population virale Wt au fil de l'évolution .....	42

2.4	Identification de mutations ponctuelles .....	43
2.4.1	Mutations ponctuelles des séquences consensus.....	43
2.4.2	Évolution des SNP des séquences consensus de Wt et Ad .....	44
2.5	Identification de potentiels virus « donneurs » de segments 2, 3, 4, 5, 7 et 8... .....	44
2.6	Proposition de scénario d'adaptation virale.....	45
3	Discussion.....	46
3.1	Séquençage total des virus parental et adapté .....	46
3.2	Analyse de sous-populations .....	46
3.3	Analyse de SNP minoritaires.....	47
3.4	Infections mixtes chez les oiseaux sauvages .....	48
	Matériels et méthodes .....	50
1	Première partie expérimentale.....	51
1.1	Observations cliniques et prélèvements biologiques.....	51
1.2	Extraction d'ARN, rétrotranscription et PCR .....	51
1.3	Séquençage des produits d'amplification par la technique de Sanger.....	52
1.4	Analyse des neuraminidases par PCR conventionnelle.....	52
1.5	Analyse des neuraminidases par PCR quantitative en temps réel .....	53
1.6	Pyroséquençage des produits d'amplification .....	54
1.7	Analyse de données par bioinformatique .....	54
1.7.1	Numéros d'accès des séquences nucléotidiques .....	55
2	Deuxième partie expérimentale.....	55
2.1	Culture de VIA sur cellules porcines.....	55
2.2	Séquençage par la technique de Sanger.....	56
2.2.1	Amplification des génomes totaux.....	56
2.2.2	Purification des produits de PCR.....	56
2.2.3	Séquençage .....	57

2.3	Pyroséquençage des produits d'amplification .....	57
2.4	Analyse de données par bioinformatique .....	57
	Conclusion générale.....	58
	Références bibliographiques .....	61
	Annexes .....	71

## Liste des abréviations

A : adénosine

aa : acide aminé

ACP : amplification en chaîne par polymérase

ADN : acide désoxyribonucléique

ADNc : acide désoxyribonucléique complémentaire

ARN : acide ribonucléique

ARNc : acide ribonucléique complémentaire

ARNm : acide ribonucléique messenger

ARNv : acide ribonucléique viral

ATP : adénosine triphosphate

C : cytidine

dNTP : désoxyribonucléotide triphosphate

EC : écart-type

emPCR : *emulsion polymerase chain reaction*

G : guanosine

HA : hémagglutinine

MDCK : *Madin-Darby canine kidney*

NA : neuraminidase

NGS : nouvelles générations de séquençage

NP : nucléoprotéine

NPTr : *newborn pig trachea*

OMS : organisation mondiale de la santé

PA : polymérase acide

pb : paire de bases

PB1 : polymérase basique 1

PB2 : polymérase basique 2

PBS : *phosphate buffered saline*

PCR : *polymerase chain reaction*

pH : potentiel hydrogène

RBS : *receptor binding site*

SNP : *single nucleotide polymorphism*

SVF : sérum de veau foetal

T : thymidine

TBE : tris, borate, EDTA

UFP : unité formant plage de lyse

VIA : virus influenza aviaire

VIAFP : virus influenza aviaire faiblement pathogène

VIAHP : virus influenza aviaire hautement pathogène



## Liste des figures

<b>Figure 1</b> Morphologie et structure des virions et des RNP. ....	face p.11
<b>Figure 2</b> L'hémagglutinine des VIA .....	p.12
<b>Figure 3</b> Spectre d'hôtes des VIA. ....	face p.14
<b>Figure 4</b> Les principales étapes du cycle viral d'un VIA.....	face p.15
<b>Figure 5</b> Bourgeonnement des VIA. ....	face p.16
<b>Figure 6</b> Schématisation de la transmission des VIA. ....	p.17
<b>Figure 7</b> Évolution d'un virus influenza pandémique.. ....	p.23
<b>Figure 8</b> Réplication d'un virus avec introduction de mutations ponctuelles. ....	p.24
<b>Figure 9</b> Chromatogramme de séquençage « Sanger ». ....	p.26
<b>Figure 10</b> Étapes de pyroséquençage 454.....	face p.27
<b>Figure 11</b> Électrophorèse sur gel d'agarose.....	face p.33
<b>Figure 12</b> Résultats de pyroséquençage 454 de la première partie.....	face p.33
<b>Figure 13</b> Résultats de pyroséquençage 454 de la première partie, tige de la NA. ....	face p.35
<b>Figure 14</b> Résultats de pyroséquençage 454 de la seconde partie.....	face p.41
<b>Figure 15</b> Résultats de pyroséquençage 454 de la seconde partie, PB1.....	face p.42
<b>Figure 16</b> Scénario d'évolution génétique. ....	45
<b>Figure 17</b> Sondes de PCR quantitative. ....	53

## Liste des tableaux

<b>Tableau I</b> Les protéines codées par le génome des VIA. ....	13
<b>Tableau II</b> SNP non synonymes des gènes PB2 et HA .....	34
<b>Tableau III</b> Séquences nucléotidiques des amorces et sondes de PCRq. ....	53

## **Introduction bibliographique**

# 1 Les virus influenza A

## 1.1 Un peu d'histoire

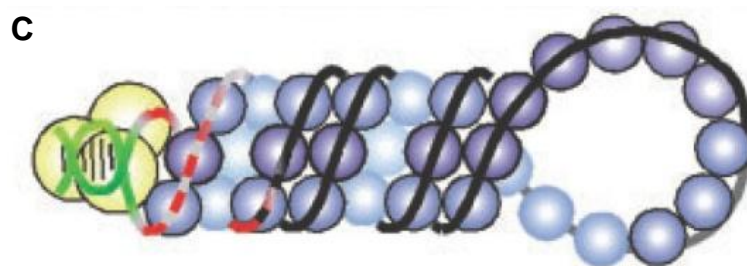
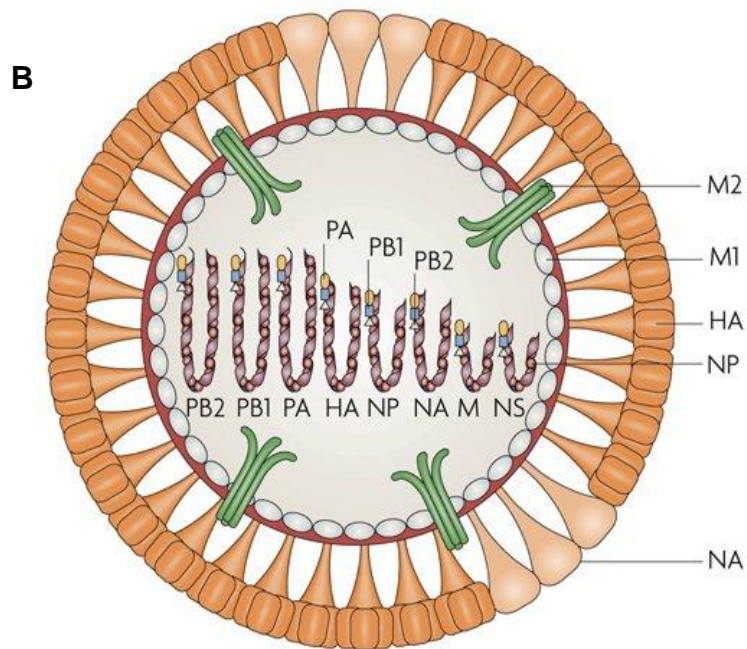
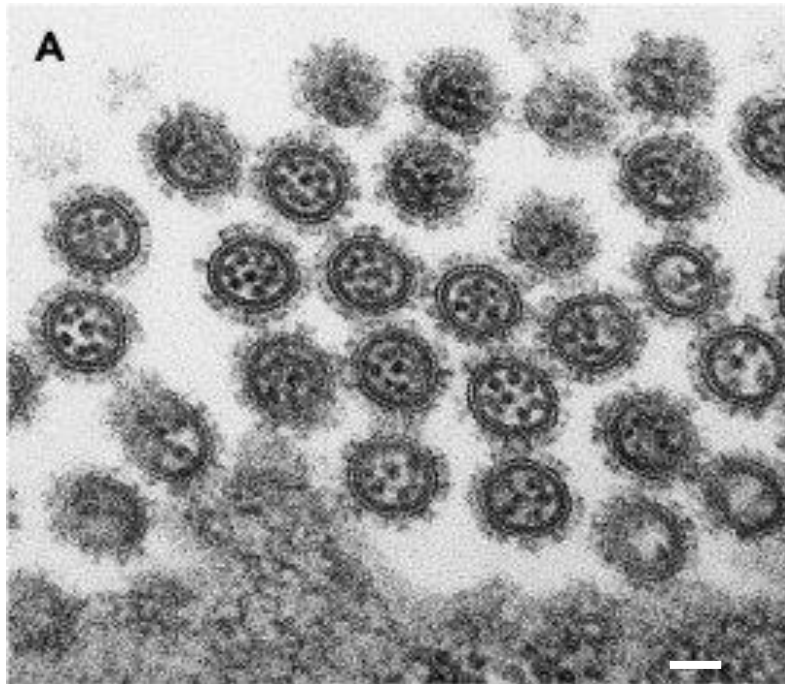
Les virus influenza sont responsables d'une infection virale aiguë nommée grippe. Cette maladie a angoissé les populations à plusieurs reprises au cours des siècles passés et continuera probablement ainsi au fil des siècles à venir. Cette infection est à l'origine de décès innombrables, tant dans la population humaine que chez les animaux. En 412 avant Jésus-Christ, Hippocrate décrit pour la première fois la maladie provoquée par les virus influenza. Depuis 1509, on a dénombré 14 pandémies de grippe soit environ une tous les 36 ans (Taubenberger and Morens, 2010).

La célèbre grippe espagnole (1918-1919), une des pandémies les plus dévastatrices de l'histoire de l'humanité, fut à l'origine de près de 38 millions de décès d'après les données de Johnson and Mueller, 2002. L'agent responsable était un virus influenza A de sous-type H1N1 et d'origine aviaire, il est considéré comme l'ascendant de tous les virus influenza A circulant aujourd'hui chez l'homme (Taubenberger and Morens, 2006; Taubenberger et al., 2007). D'autres pandémies ont aussi marqué l'histoire : la grippe asiatique (1957-1958) causée par un virus de sous-type H2N2, la grippe de Hong Kong (1968-1969) causée par un virus de sous-type H3N2 et enfin, la récente grippe dite porcine en 2009, causée par un virus de sous-type H1N1.

Des épisodes de mortalité massive chez les oiseaux sauvages et domestiques sont décrits depuis l'antiquité (Blancou, 2000) mais nous ne pouvons malheureusement pas affirmer s'il s'agissait de grippe, c'est pourquoi les historiens qualifient ces maladies de « pestes ». Certaines descriptions d'épidémies sont cependant singulières, en 1366 Short rapporte une bataille d'oiseaux précédant une épidémie humaine :

*Cette année, survint aussi une grande querelle entre moineaux, qui tourna en une bataille rangée au cours de laquelle d'innombrables combattants perdirent la vie. Il s'en suivit une importante mortalité chez les êtres humains, dont beaucoup furent trouvés morts le matin, alors qu'ils s'étaient couchés en bonne santé la veille (Flemming, 1871).*

C'est en 1997 que la première infection d'un homme par un virus influenza aviaire de sous-type H5N1 hautement pathogène a été observée. Ce virus provenait de volailles de Hong Kong (région administrative spéciale de Chine) et a de nouveau émergé en 2003-2004



**Figure 1** A) Visualisation au microscope électronique à transmission de virions influenza sphériques libérés d'une cellule infectée, barre d'échelle : 50 nm (dans Knipe and Howley, 2007). B) Représentation schématique d'un virion sphérique influenza A (d'après Subbarao and Joseph, 2007). C) Représentation schématique d'une RNP. Les NP et les polymérases sont respectivement représentées sous forme de sphères bleues ou jaunes. L'ARNv est représenté par le brin coloré en vert, rouge et noir (d'après Hutchinson et al., 2010).

avant de se propager depuis l'Asie vers l'Europe et l'Afrique, il persiste alors à l'état endémique chez les volailles dans certaines régions du monde ([www.who.int](http://www.who.int)).

## 1.2 Qu'est-ce que l'influenza A ?

### 1.2.1 Classification

La nomenclature de ces virus est organisée selon le **genre** de la souche (A, B ou C), l'**espèce** chez laquelle la souche a été isolée pour la première fois (non précisée s'il s'agit de l'espèce humaine), la **localisation** géographique de l'isolat, la **référence** et l'**année** de l'isolement et, exclusivement pour le type A : les **sous-types** de l'hémagglutinine HA et de la neuraminidase NA entre parenthèses et sous la forme HxNy.

Un virus influenza aviaire (VIA) de sous-type H6N1 isolé chez le canard de Barbarie, en France, en 2009 avec pour référence le numéro 09010 sera classifié de la sorte : A/Muscovy duck/France/09010/2009 (H6N1).

Les virus influenza A appartiennent à la famille des *Orthomyxoviridae* qui comprend les trois genres de virus influenza, les *Thogotovirus* et les *Isavirus* (Kawaoka et al., 2004).

### 1.2.2 Structure des virions

Les virions peuvent être observés sous forme de sphères (figure 1-A) dont la taille varie entre 80 et 120 nm de diamètre mais aussi sous forme filamenteuse avec un diamètre de 100 nm pour une longueur de 20 µm. Bien que les souches de laboratoire produisent uniquement des virions sphériques, les virus issus de prélèvements humains produiraient davantage de virions d'aspect filamenteux (dans Rossman and Lamb, 2011). Leur membrane lipidique provient de la cellule hôte à partir de laquelle les virions se sont échappés par bourgeonnement. À la surface de l'enveloppe virale, on retrouve 3 des 16 protéines codées par le génome, à savoir les deux principaux déterminants antigéniques du virion la HA et la NA (environ quatre HA pour une NA), ainsi que la protéine transmembranaire M2. La protéine de matrice M1 tapisse la face interne de l'enveloppe et le cœur de la particule virale est constitué de ribonucléoprotéines (RNP) (figure 1-B).

Les RNP sont constituées d'un segment d'acide ribonucléique (ARN) couplé aux polyméras PB2, PB1, PA et à la NP (figure 1-C).

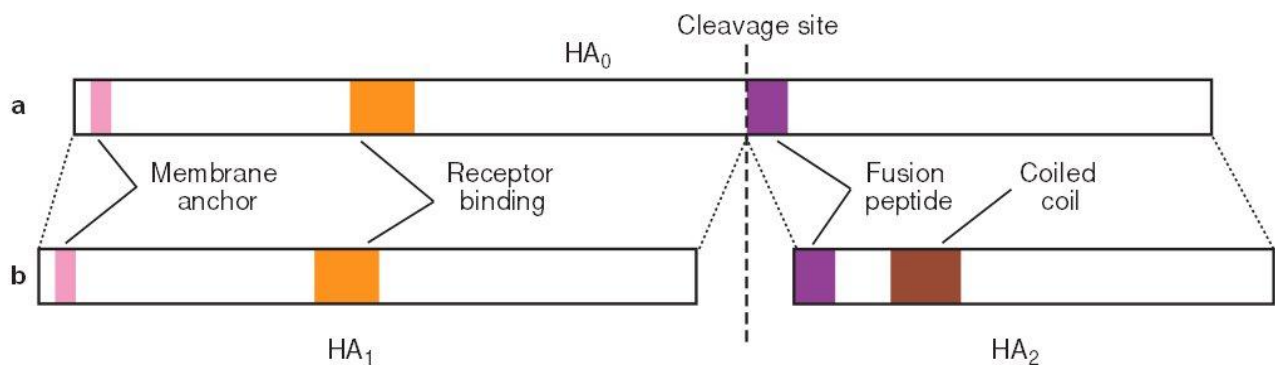
Récemment, un 17<sup>ème</sup> sous-type de HA a été identifié (Tong et al., 2012), pour 10 sous-types de NA.

### 1.2.3 Organisation et structure du génome

Les virus influenza A possèdent huit fragments d'ARN simple brin, segmentés et de polarité négative pour un total de 13 500 paires de bases. Les 13 nucléotides de l'extrémité 5' et les 12 nucléotides de l'extrémité 3' sont conservés entre les segments d'ARNv puis suivis d'une région codante spécifique du segment d'ARNv. À ce jour 16 protéines ont été identifiées, codées par les huit segments d'ARNv, chacun d'entre eux codant pour une à quatre protéines (tableau I).

### 1.2.4 Glycoprotéines de surface

L'hémagglutinine et la neuraminidase (tableau I) sont les deux glycoprotéines de surface et les principaux déterminants antigéniques des VIA. La HA est synthétisée sous forme d'un précurseur nommé HA<sub>0</sub>, ce précurseur est clivé par des protéases extracellulaires en un dimère HA<sub>1</sub>/HA<sub>2</sub>. HA<sub>1</sub> supporte le site de liaison aux récepteurs cellulaires (*receptor binding site* en anglais, RBS) et HA<sub>2</sub> supporte le peptide de fusion (figure 2).

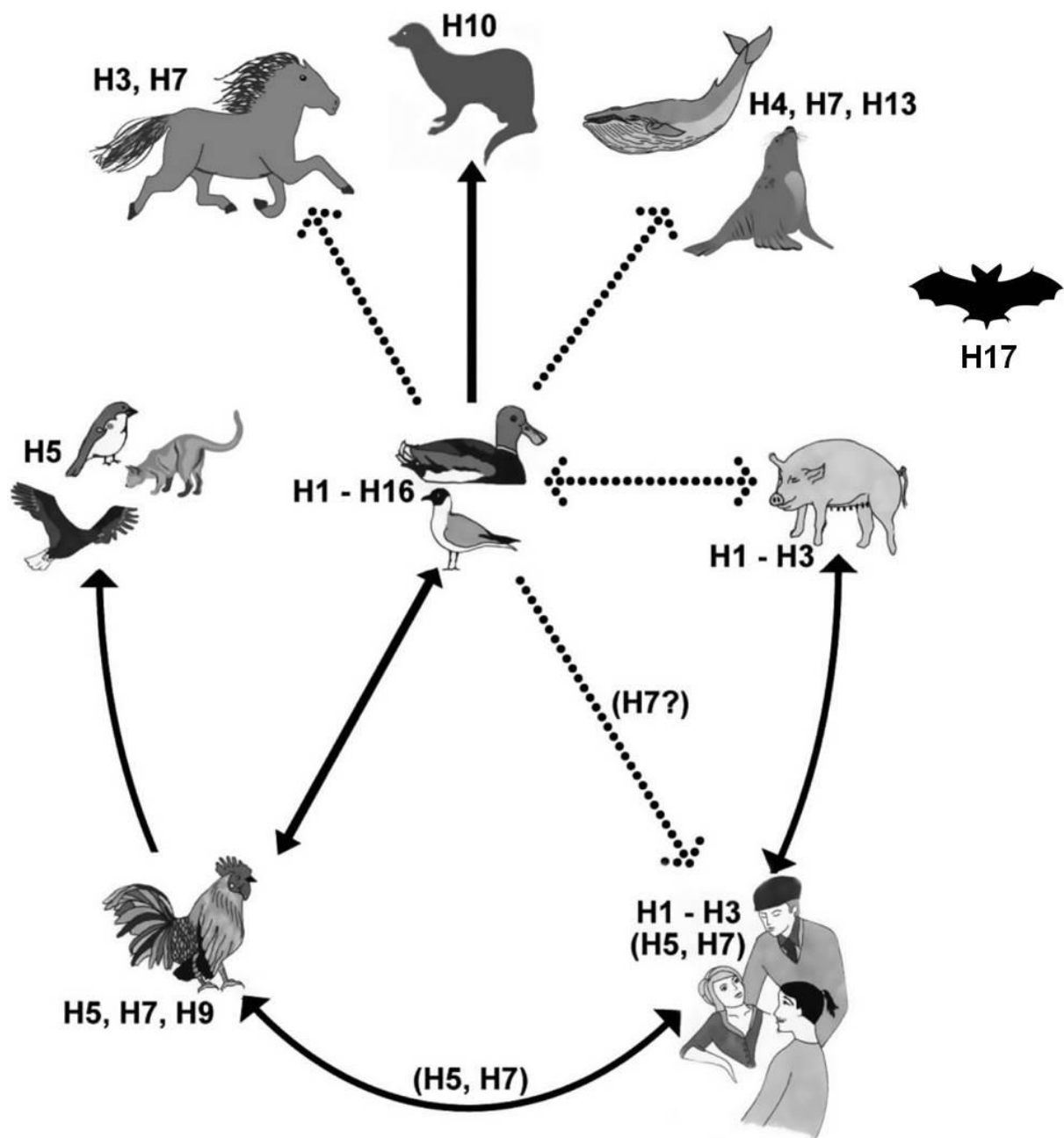


**Figure 2** Clivage de l'hémagglutinine des virus influenza. A) Hémagglutinine sous forme de précurseur HA<sub>0</sub>. B) HA<sub>0</sub> est clivé par des protéases en deux sous-unités : HA<sub>1</sub> et HA<sub>2</sub>. HA<sub>1</sub> supporte le domaine d'accroche aux récepteurs cellulaires (zone orange), ancre la protéine dans la membrane virale (zone rose) et est reliée à HA<sub>2</sub> par des ponts disulfure. HA<sub>2</sub> supporte le peptide de fusion (zone violette) qui est activé suite à une modification conformationnelle de HA<sub>2</sub> (zone marron) dépendante d'un pH acide (Whittaker, 2001).

**Tableau I** Les protéines codées par le génome des virus influenza A et leurs fonctions.

Segment d'ARNv	Taille du segment d'ARNv (pb)	Protéine(s)	Fonction(s) générale(s) de la ou des protéine(s)
1	2341	PB2	Sous-unité du complexe polymérase, en charge de la reconnaissance de la coiffe 5' des ARNm de l'hôte (dans Cheung and Poon, 2007).
2	2341	PB1 PB1-F2 PB1-N40	Sous-unité du complexe polymérase (dans Cheung and Poon, 2007). Facteur de virulence (Zamarin et al., 2006). Protéine non essentielle mais aide à la réplication virale de certaines souches (Wise et al., 2009).
3	2233	PA PA-X PA-N155 PA-N182	Responsable du clivage de la coiffe de l'ARNm hôte (dans Cheung and Poon, 2007). Régule la réponse immunitaire de l'hôte et rôle dans la répression de gènes cellulaires de l'hôte (Jagger et al., 2012). Fonctions de réplication (Muramoto et al., 2013). Fonctions de réplication (Muramoto et al., 2013).
4	1778	HA	Glycoprotéine de surface. Impliquée dans l'attachement des particules virales aux récepteurs cellulaires. Impliquée dans la fusion des membranes virale et cellulaire. Support du caractère « hautement pathogène » des VIA. Cible d'anticorps neutralisants et de traitements antiviraux (dans Cheung and Poon, 2007).
5	1565	NP	Liaison aux ARN. Régulation de la transcription des ARNv. Transport nucléaire des ARNv (dans Cheung and Poon, 2007).
6	1413	NA	Glycoprotéine de surface. Rôle dans la libération des particules de la progénie virale depuis la cellule infectée. Cible d'anticorps neutralisants et de traitements antiviraux (dans Cheung and Poon, 2007).
7	1027	M1 M2 M42	Protéine de matrice. Impliquée dans l'export nucléaire de RNPv. Protéine membranaire formant des canaux ioniques. Cible d'anticorps neutralisants et de traitements antiviraux (dans Cheung and Poon, 2007). Forme alternative de M2 (Wise et al., 2012).
8	890	NS1 NEP	Intervient dans la régulation de l'expression de protéines virales et cellulaires et notamment de l'interféron. Impliquée dans l'export nucléaire des RNPv (dans Cheung and Poon, 2007).





**Figure 3** Illustration du spectre d'hôtes des VIA avec le réservoir naturel des VIA au centre, les hôtes accidentels et les sous-types qui ont été identifiés dans chaque groupe d'hôtes apparaissent en périphérie. (Modifié d'après Wahlgren, 2011).

L'hémagglutinine détermine également le spectre d'hôte des VIA, chaque sous-type de HA est propre à certaines espèces animales (figure 3).

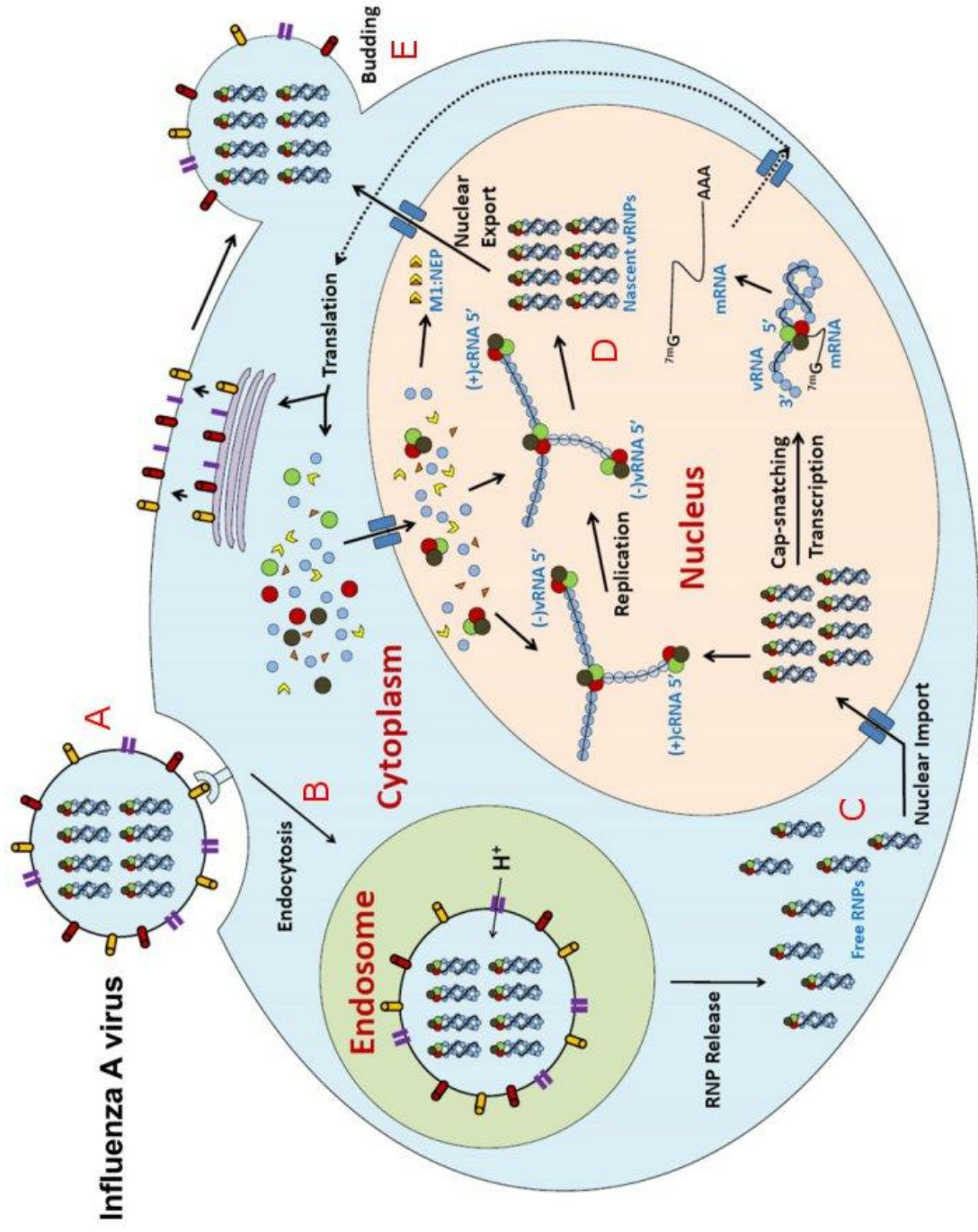
La neuraminidase, quant à elle, est composée d'une tête globulaire, d'une étroite tige, d'un domaine transmembranaire et d'un domaine cytoplasmique. La glycosylation de la NA peut être un déterminant important de la neurovirulence des VIA. La NA peut exister sous forme d'un homotétramère pouvant détruire les récepteurs cellulaires, c'est-à-dire qu'elle peut cliver la liaison  $\alpha$ -kétosidique existant entre les résidus d'acide sialique et un résidu de D-galactose ou D-galactosamine. La NA joue un rôle primordial dans le relargage de la progénie virale depuis la cellule infectée (dans Cheung and Poon, 2007).

## 1.2.5 Étapes de la réplication virale

### 1.2.5.1 Attachement du virus

Les virus se lient aux cellules par l'interaction entre l'hémagglutinine et les résidus d'acide sialique portés par des glycoprotéines ou glycolipides présent(e)s à la surface des cellules (figure 4-A). Il existe une spécificité de liaison selon l'espèce d'origine du virus : les virus aviaires se lient préférentiellement aux résidus d'acide sialique liés par une liaison de type  $\alpha_{2,3}$  aux résidus galactose alors que les virus d'origine humaine présentent une meilleure affinité pour les résidus d'acides sialiques reliés aux résidus galactose par une liaison  $\alpha_{2,6}$  (Rogers and Paulson, 1983). Il faut cependant noter que cette spécificité de liaison n'est pas absolue et que les cellules aviaires comme les cellules humaines peuvent arborer les deux types de conjugués. Chez l'homme, les liaisons de type  $\alpha_{2,6}$  sont dominantes tout au long de l'arbre respiratoire mais des liaisons de type  $\alpha_{2,3}$  sont retrouvées au niveau de l'arbre respiratoire profond (Shinya et al., 2006).

Chez les canards et les poulets, les cellules intestinales ne présentent que des conjugués de type  $\alpha_{2,3}$ . Au niveau de la trachée, les cellules de poulet présentent majoritairement des conjugués de type  $\alpha_{2,6}$  alors que ce sont des conjugués de type  $\alpha_{2,3}$  qui sont majoritaires chez le canard. Quant aux porcs, les deux types de conjugués sont retrouvés au niveau des appareils respiratoire et digestif (Nelli et al., 2010).



**Figure 4** Schématisation des principales étapes du cycle viral d'un virus influenza. **A)** Attachement du virus à la surface de la cellule. **B)** Endocytose du virus. **C)** Transfert des RNP vers le noyau. **D)** Formation des nouvelles RNP. **E)** Bourgeonnement du virion (modifié d'après Zheng and Tao, 2013).

#### 1.2.5.2 Entrée du virus et endocytose

La voie d'entrée des virus influenza la plus décrite est la voie des clathrines (Matlin et al., 1981). Ces dernières recouvrent la face cytosolique de la cellule qui deviendra la vésicule d'endocytose (Rust et al., 2004), l'internalisation du virus a ensuite lieu (figure 4-B). L'endosome formé, deux phénomènes ont lieu : (i) la diminution du potentiel hydrogène (pH) dans l'endosome entraîne une modification conformationnelle de la HA qui résulte en l'exposition du peptide de fusion localisé dans la région N-terminale de HA<sub>2</sub>. Le peptide de fusion va fusionner avec la membrane de l'endosome.

Cet évènement ayant lieu simultanément pour chacune des HA liées et clivées, un pore est formé ; (ii) la protéine M2 s'arrange en tétramère, le canal transmembranaire ainsi formé entraîne l'afflux de proton H<sup>+</sup> dans la particule virale, le pH chute et les interactions entre les RNP et M1 sont déstabilisées.

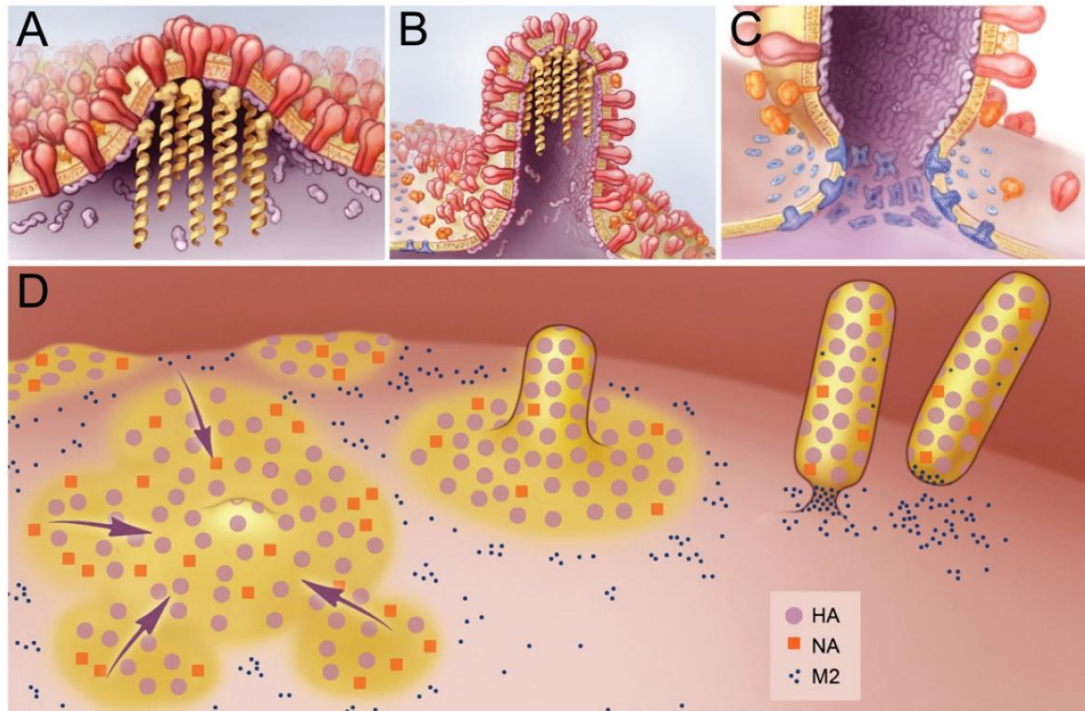
L'addition de ces deux évènements résulte en la libération des RNP dans le cytoplasme, en direction du noyau (figure 4-C).

#### 1.2.5.3 Import nucléaire des RNP

Les virus influenza se distinguent des autres virus à ARN par leur dépendance aux fonctions nucléaires, en effet la totalité de la synthèse des ARN a lieu dans le noyau. Les RNP étant de taille trop importante pour diffuser de façon passive dans le noyau, chacune des protéines les constituant possède un signal de localisation nucléaire (NLS). Les NLS leur permettent d'interagir avec la machinerie d'import nucléaire pour leur entrée dans le noyau.

#### 1.2.5.4 Synthèse des ARNm viraux

L'initiation de la transcription débute par une liaison de la sous-unité de PB1 à l'extrémité 5' du brin d'ARNv, ceci entraîne une modification conformationnelle de PB1, PB2 peut alors se lier au complexe grâce à une coiffe d'ARN pré-messager volée à la cellule hôte. L'extrémité 3' du brin d'ARNv se lie à PB1, la coiffe est clivée puis la transcription est initiée suite à l'ajout d'une guanosine en 3' de la coiffe.



**Figure 5** Bourgeoisement de virus influenza. **A)** Formation des radeaux lipidiques. **B)** Élongation du virion en bourgeoisement suite à la polymérisation de M1. **C)** Membrane en cours de scission. **D)** Vue globale du processus de bourgeoisement des virus influenza. (d'après Rossman and Lamb, 2011).

#### 1.2.5.5 Synthèse des ARNc : futur génome viral

Les ARNc sont des copies complètes de l'ARNv, de polarité positive, ils sont également générés par la polymérase virale. Une fois produits, ces ARNc servent de matrice pour la synthèse d'ARNv, brin complémentaire de l'ARNc. Les ARNv nouvellement produits peuvent former de nouvelles RNP qui seront empaquetées dans les futurs virions (figure 4-D).

#### 1.2.5.6 Export des RNP

Les RNP sont formées et assemblées dans le noyau, la protéine M1 permet l'association des NP avec les brins d'ARNv (Huang et al., 2001). NEP intervient ensuite en se liant au complexe RNP/M1 et recrute la machinerie nécessaire à leur exportation dans le cytoplasme. Avant d'être empaquetées dans les virions, les huit RNP s'assemblent dans le noyau ou le cytoplasme tout en s'arrangeant autour d'une structure bien définie grâce à des interactions spécifiques entre les segments d'ARNv (Noda et al., 2012).

#### 1.2.5.7 Assemblage et bourgeonnement des virions

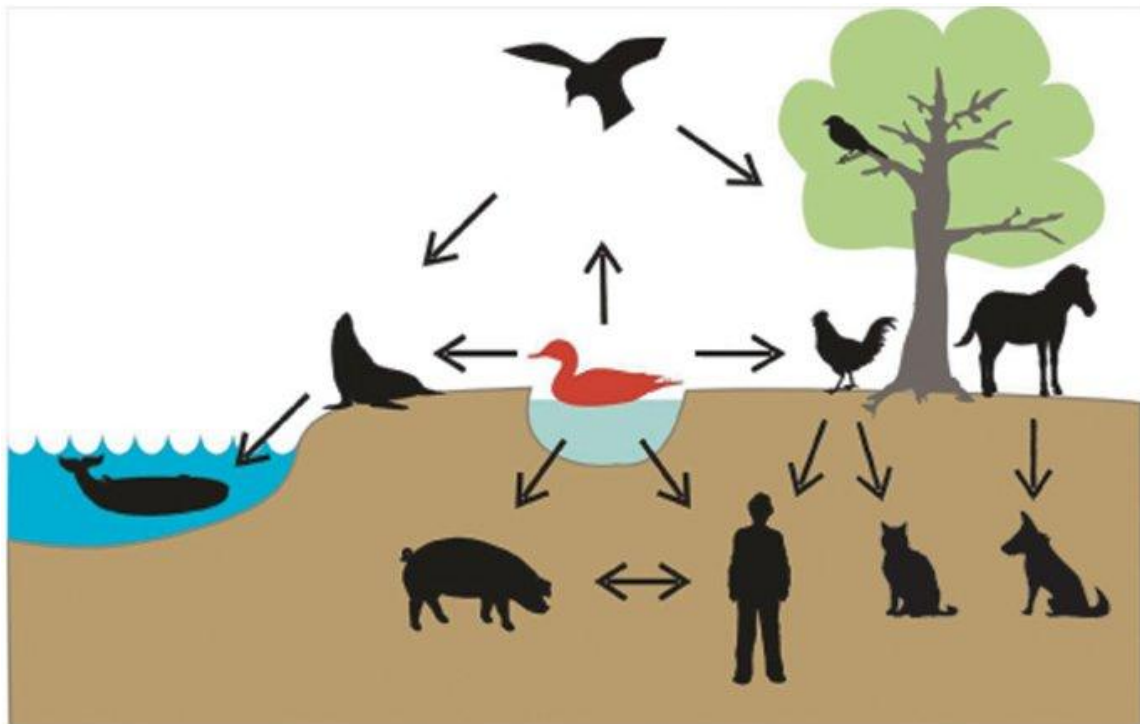
La dernière étape, le bourgeonnement, est initiée par HA et NA, leur accumulation au niveau de la membrane plasmique des cellules infectées entraîne la formation de radeaux lipidiques, sites d'assemblage de virions en cours de formation. Les radeaux provoquent une altération de la courbe de la membrane plasmique, ce qui induit le processus de bourgeonnement (figure 5 A-D). Les extrémités intra-cytoplasmiques de HA et NA recrutent M1, cela offre une structure à la nouvelle membrane virale, cet arrangement déterminerait la future morphologie des virions, filamenteux ou sphériques (Calder et al., 2010). La polarisation de M1 (figure 5-B) permet l'élongation du virion puis les RNP sont recrutées pour être empaquetées. Enfin, une accumulation de M2 a lieu à la frontière entre le virion bourgeonnant et la membrane plasmique, elle exerce une force supplémentaire pour courber davantage la membrane, ce qui se traduit en une scission de cette dernière (figure 5-C) (Rossman and Lamb, 2011). NA termine le travail en clivant les acides sialiques de la surface cellulaire de sorte à ce que les HA ne s'y lient pas, les virions sont libres.

### 1.3 Écologie des virus influenza A

Les oiseaux sauvages aquatiques constituent le réservoir principal des virus influenza A, on y a isolé la majorité des sous-types de HA et de NA (Olsen et al., 2006). Les plans d'eau

sont un lieu idéal pour la transmission de virus, ce qui se résulte en un brassage génétique intense. L'infection de ces espèces, dont beaucoup sont migratrices, entraîne une dissémination des virus le long des couloirs de migration au niveau des continents eurasien ou américain. L'infection asymptomatique ainsi que la dissémination lors des migrations favorise la circulation des virus depuis l'avifaune sauvage vers les volailles domestiques.

Les VIA peuvent diffuser depuis les oiseaux sauvages aquatiques vers les humains, les porcs, les mammifères marins ou les volailles (figure 6). Les chauves-souris sont également un réservoir de virus influenza à prendre en compte (Tong et al., 2012).



**Figure 6** Schématisation de la transmission de virus influenza A des oiseaux sauvages aquatiques aux humains, porcs, mammifères marins et volailles. La transmission est aussi observée entre les porcs et les humains et de la volaille aux humains ou aux chats. Les virus influenza équin peuvent aussi être transmis aux chiens (d'après Knipe and Howley, 2007).

### 1.3.1 L'influenza aviaire chez les volailles

L'influenza aviaire provoque l'une des maladies les plus importantes rencontrées chez les volailles. Il existe aussi la maladie de Newcastle qui est causée par un virus de la famille des *Paramyxoviridae*. Il s'agit d'une maladie très contagieuse touchant de nombreuses espèces aviaires, les signes cliniques peuvent être sévères avec un taux de mortalité très élevé, quelque soit l'espèce touchée (Alexander D. J. et al., 1999).

Un large éventail de sous-types de VIA peuvent infecter les volailles et causer des infections asymptomatiques comme létales (Pantin-Jackwood and Swayne, 2009). Les sous-types H5, H6, H7 et H9 sont les plus fréquents en Eurasie (Alexander, 2007; Brown, 2010).

Les poulets, les dindes, les canards, les oies et d'autres volailles présentent des profils d'infection distincts. Les dindes et les cailles sont particulièrement sensibles aux infections par des VIA (Makarova et al., 2003; Pillai et al., 2010). D'autres virus d'origine aviaire peuvent infecter des volailles, un des exemples est le virus A/H1N1/pdm09 qui est capable d'infecter la dinde (Mathieu et al., 2010).

#### 1.3.1.1 Signes cliniques et pathogénicité

Chez la volaille, l'infection par un VIAFP débute par une réplication au niveau de l'arbre respiratoire puis se disperse au niveau de l'appareil digestif. Classiquement, les infections à VIAFP causent une mortalité inférieure à 5 % mais induisent une morbidité d'environ 50 % (Pantin-Jackwood and Swayne, 2009). Une diminution de la consommation d'eau et d'aliment, de la prostration ainsi qu'une légère augmentation de la mortalité quotidienne sont des signes d'infection à VIAFP. Les oiseaux présentent les signes cliniques suivants : éternuements, toux, râles respiratoires, sécrétions oculaires excessives et parfois de la diarrhée.

Les VIAFP de sous-type H5 ou H7 originaires d'oiseaux sauvages ont un potentiel hautement pathogène dépendant de mutations localisées au niveau du site de clivage HA0. Ces mutations peuvent survenir après la circulation du virus chez des volailles (Ito et al., 2001; Pantin-Jackwood and Swayne, 2009). L'introduction de résidus basiques (Arginine ou Lysine) au niveau du site de clivage HA0 permet aux protéases ubiquitaires comme la furine de cliver HA0, il s'ensuit une réplication systémique et des complications pathologiques (Alexander, 2000; Rott, 1992; Stieneke-Gröber et al., 1992). L'infection des cellules endothéliales permet la diffusion du virus par le système sanguin ou lymphatique, la virémie entraîne une diffusion systémique du virus, notamment jusqu'au cerveau. Les dommages causés aux cellules endothéliales entraînent une augmentation de la perméabilité vasculaire qui mène à des hémorragies, des œdèmes, des nécroses et des dysfonctionnements de certains organes. Les signes cliniques peuvent être une cyanose de la crête et de la caroncule tout comme un gonflement péri-orbital. Les signes pathologiques causés par les VIA sont dus à la réplication virale, la production de cytokines et une ischémie (Pantin-Jackwood and Swayne, 2009).



### 1.3.1.2 Transmission du virus

L'infection de volailles par des VIA d'oiseaux sauvages s'effectue par contact direct ou indirect à des équipements, des hommes, de l'eau ou de l'air contaminé(e)s par le virus (Alexander, 2000; Hinshaw et al., 1979). Le risque majeur de diffusion du virus entre élevages est principalement le déplacement d'oiseaux infectés ou de matières inertes contaminées. Le virus peut diffuser jusqu'à 45 mètres depuis un bâtiment d'élevage par l'intermédiaire de poussières, d'aérosols ou de plumes transporté(e)s par le vent ([www.who.int](http://www.who.int)), cependant ces modes de contamination ne sont pas les plus fréquents.

Les VIA peuvent persister jusqu'à un an dans des eaux présentant des conditions idéales en termes de pH, de température et de salinité ainsi que dans des sédiments de lacs qui sont supposés être importants pour la transmission de virus (Lebarbenchon et al., 2011; Nazir et al., 2011). Il se peut que les VIA se maintiennent à l'état congelé dans les sols durant l'hiver (Ito et al., 1995) ou bien que le virus persiste à faible niveau chez les oiseaux migrateurs qui les ramènent dans les zones de reproduction au printemps (Krauss et al., 2004).

### 1.3.1.3 Virulence

Selon le « code sanitaire pour les animaux terrestres, chapitre 10.4 : influenza aviaire » de l'OIE, la virulence des VIA est définie comme telle :

*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 moyen quotidien individuel observé à l'issue une période de 10 jours après inoculation intraveineuse de 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.

D'un point de vue pratique, le caractère hautement ou faiblement pathogène des VIA est déterminé en première intention par détermination du sous-type (H5, H7 ou autre) et pour les virus H5 ou H7, par séquençage du site de clivage de HA. La détermination de l'indice de pathogénicité intraveineuse s'effectue en parallèle, mais son résultat intervient bien plus tard, souvent trop tard pour servir de base à la gestion d'un foyer d'influenza aviaire.

### **1.3.2 L'influenza aviaire chez les canards**

Au sein des Ansériformes, les VIA sont le plus fréquemment détectés chez les *Anatidae*. Les canards et les canards colvert (*Anas Platyrhynchos*) en particulier sont considérés comme les hôtes primaires de VIA (Olsen et al., 2006). Chez les canards, les VIA se répliquent dans les cellules épithéliales du tractus intestinal bas et des bourses de Fabricius sans causer de lésions morphologiques (Daoust et al., 2011). De grandes quantités de virus sont retrouvées dans les fèces, c'est donc la voie féco-orale qui permet la transmission du virus lors, par exemple, de consommation d'eau de surface contaminée (VanDalen et al., 2010; Webster et al., 1978).

La prévalence des VIA chez les oiseaux aquatiques varie selon la saison, la localisation géographique, l'espèce et l'âge (Munster et al., 2007). En Europe du nord et en Amérique du nord, les VIA sont le plus fréquemment détectés chez les canards à la fin de l'été et à l'automne (Krauss et al., 2004; Wallensten et al., 2007) soit à une période durant laquelle les jeunes canards se rassemblent avant la migration (Hinshaw et al., 1980).

## **1.4 L'influenza aviaire**

La première référence à l'influenza aviaire date de 1878 avec la description, faite par Perroncito, d'une maladie aviaire associée à une forte mortalité. Baptisée « peste aviaire »,

cette maladie était initialement confondue avec le choléra aviaire (infection à *Pasteurella multocida*). Néanmoins il ne s'écoula que quelques décennies avant que l'on puisse distinguer les deux maladies. En 1955, l'agent responsable est identifié comme étant un virus influenza de type A (dans Lupiani and Reddy, 2009) et c'est en 1981 que le terme « influenza aviaire hautement pathogène » est adopté, lors du premier symposium international sur l'influenza aviaire.

Le spectre d'hôte des virus influenza aviaires (VIA) est relativement large puisque, en plus des oiseaux aquatiques, on les retrouve chez les humains, les porcs, les chiens, les chats ou encore les volailles. On souligne là leur importante capacité de transmission inter-espèces. Chez l'homme, la diffusion de VIA reste cependant limitée, en raison de l'affinité plus marquée des VIA pour des récepteurs situés dans l'arbre respiratoire profond, un contact rapproché est donc nécessaire à la contamination. Des cas humains de contamination à VIAHP de sous-type H5N1 restent toutefois d'actualité, en effet, le bulletin épidémiologique de l'OMS datant du 1<sup>er</sup> février 2013 reporte le décès de quatre patients cambodgiens : « Tous les cas ont présenté de la fièvre, de la toux et d'autres signes de syndrome grippal. Quatre sont morts, celui ayant survécu étant le petit garçon de 8 mois qui récupère après avoir souffert d'un syndrome grippal bénin. Les données préliminaires n'étaient pas une transmission interhumaine et l'on sait pour quatre de ces cas qu'ils ont été en contact rapproché avec des volailles malades ou mortes. » (<http://www.who.int>)

Les VIA sont classés sur la base de la virulence qu'ils induisent chez le poulet (Suarez, 2010). Les VIA hautement pathogènes (VIAHP) causent une infection systémique létale pouvant provoquer 100 % de mortalité au cours des 48 heures suivant l'infection (Swayne and Suarez, 2000). Les VIA faiblement pathogènes (VIAFP), qui représentent la majeure partie des VIA, causent une mortalité de moindre importance et sont rarement à l'origine d'épisodes de forte mortalité chez la volaille (Webster et al., 1992). Les VIAHP sont descendants des VIAFP de sous-type H5 ou H7 qui ont été transmis des oiseaux aquatiques sauvages aux volailles domestiques (Suarez, 2010). D'un point de vue législatif, en plus des VIAHP, les sous-types H5 et H7 de VIAFP sont donc soumis à déclaration obligatoire auprès des services de santé (OIE, Code sanitaire pour les animaux terrestres, chapitre 10.4).

Dans les années 1960, des anticorps anti-VIA ont été mis en évidence chez des oiseaux aquatiques migrateurs (Easterday et al., 1968). Les premiers VIAFP isolés chez les oiseaux sauvages, en 1972, l'ont été à partir de canards sauvages en Californie et d'un oiseau de mer en Australie (Downie and Laver, 1973; Slemons et al., 1974). Des programmes de

surveillance conduits en Amérique du nord ont démontré que les VIA étaient plus fréquemment isolés chez des canards colvert (*Anas Platyrhynchos*), ce qui a permis d'identifier les canards comme un réservoir de VIA qui pourraient potentiellement infecter d'autres espèces comme les mammifères et les oiseaux domestiques (Hinshaw et al., 1985; Webster et al., 1992).

Le virus parental du VIAHP H5N1 responsable de l'épidémie de 2003 a été isolé chez une oie domestique malade dans la province de Guangdong en Chine, en 1996 (Sims et al., 2005). À Hong Kong, en 1997, 6 des 18 cas d'infections humaines ont été fatals. Le virus a été éradiqué d'Hong Kong par l'élimination de toutes les volailles d'élevage et de marché (Sims et al., 2003). En 2003, des infections humaines à VIAHP H5N1 sont de nouveau survenues à Hong Kong et des infections de volailles ont eu lieu en Corée, en Thaïlande et au Vietnam ([www.who.int](http://www.who.int)). Deux tigres et deux léopards d'un zoo thaïlandais ont aussi été infectés et en sont morts (Keawcharoen et al., 2004).

Dès 2003, le VIAHP H5N1 a diffusé dans toute l'Asie, infectant des volailles ([www.who.int](http://www.who.int)). Au printemps 2005, des milliers d'oiseaux sauvages en sont morts, au lac Qinghai en Chine centrale (Chen et al., 2005; Liu et al., 2005). La crainte que le virus ne diffuse par le biais d'oiseaux migrateurs a motivé de nombreux pays à mettre en place des programmes de surveillance. Depuis 2005, plus de 250 millions d'oiseaux ont été infectés par le virus ou abattus dans 60 pays d'Asie, d'Europe et d'Afrique (FAO). Depuis 2003 jusqu'au 4 juin 2013, l'organisation mondiale de la santé (OMS) a recensé 630 cas confirmés de VIAHP H5N1 chez l'homme parmi lesquels 375 ont été mortels ([www.who.int](http://www.who.int)).

En 2009, au Mexique, un nouveau virus H1N1 (A/H1N1/2009pdm) a déclenché une épidémie chez l'homme, elle a été qualifiée de « grippe porcine » (Neumann et al., 2009). Ce virus a beaucoup préoccupé, d'autant plus qu'il a été démontré expérimentalement qu'un virus contenant le gène H5 d'un VIAHP H5N1 avec quatre mutations et les sept autres segments d'ARNv du virus A/H1N1/2009pdm était transmissible par voie aérosol chez le furet mais sans causer de pathologie grave ni de mortalité (Imai et al., 2012).

Dans la plupart des pays développés, rares sont les émergences de VIAHP bien que celles de VIAFP soient sporadiques. De nos jours, le sous-type H5N1 demeure à l'état endémique dans certaines parties de l'Asie et de l'Afrique (Beato et al., 2013; Daniels et al., 2012; Kayali et al., 2011). Certains virus H5N1 circulant en Europe, dans le Moyen-Orient et en Afrique peuvent reconnaître les récepteurs cellulaires humains (Yamada et al., 2006). De plus, les VIA H9N2 sont endémiques chez les volailles dans certaines parties de l'Asie et du

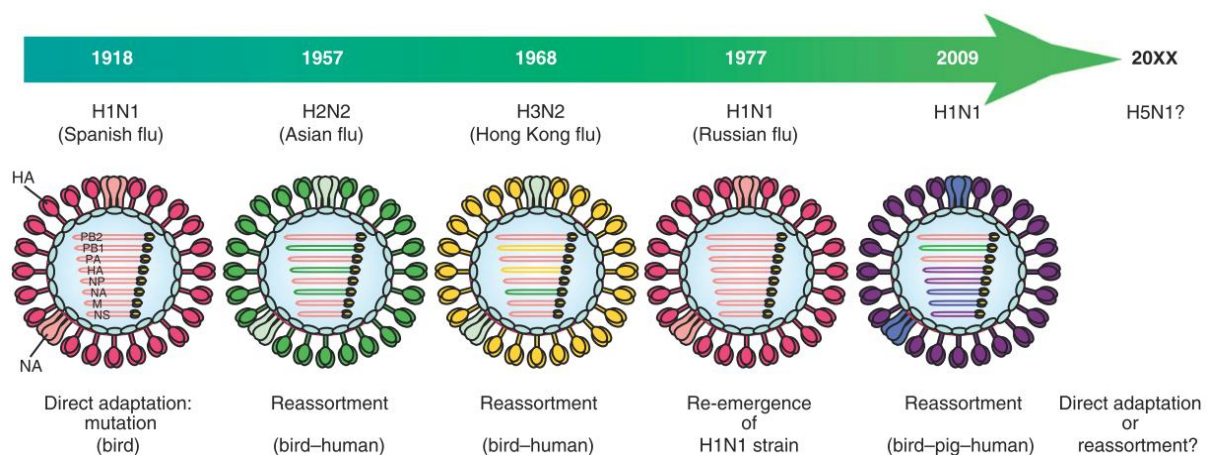
Moyen-Orient et plusieurs de ces virus ont aussi acquis une spécificité de liaison aux récepteurs cellulaires humains (Matrosovich et al., 2001). Il est donc supposé que la prochaine pandémie sera causée par un virus de sous-type H5N1 ou H9N2 adapté à l'homme (Imai and Kawaoka, 2012).

Les émergences de VIAHP H5N1 en Asie du sud-est, tout comme la récente épidémie de 2009 à virus influenza H1N1, ont souligné l'importance de l'étude des dynamiques de VIA en relation avec l'écologie de leurs hôtes naturels (Munster et al., 2007). Les programmes de surveillance du VIAHP H5N1 eurasiens qui ont été appliqués chez les oiseaux sauvages ont permis d'accroître les connaissances sur les VIAFP, en particulier chez le canard (Wallensten et al., 2007).

### 1.5 Évolution des virus influenza A

Les virus influenza A possèdent plusieurs caractéristiques leur conférant une grande capacité d'évolution. La polymérase virale, en charge de la réplication du génome, peut être à l'origine de mutations ponctuelles, il s'ensuit le phénomène de « glissement antigénique » puisque la nature génétique va évoluer au fur et à mesure des mutations.

Lors de l'infection d'une même cellule par deux virus génétiquement différents, le phénomène de « réassortiment » peut avoir lieu, il s'agit d'un mélange de segments d'ARN. Le matériel génétique de la progénie virale sera composé d'un mélange des génomes parentaux. Dans ce cas, la conséquence peut être le « saut antigénique » puisque la nature génétique des nouveaux virions se trouve radicalement modifiée et ce nouveau statut antigénique peut conférer aux virions un potentiel pandémique (figure 7).

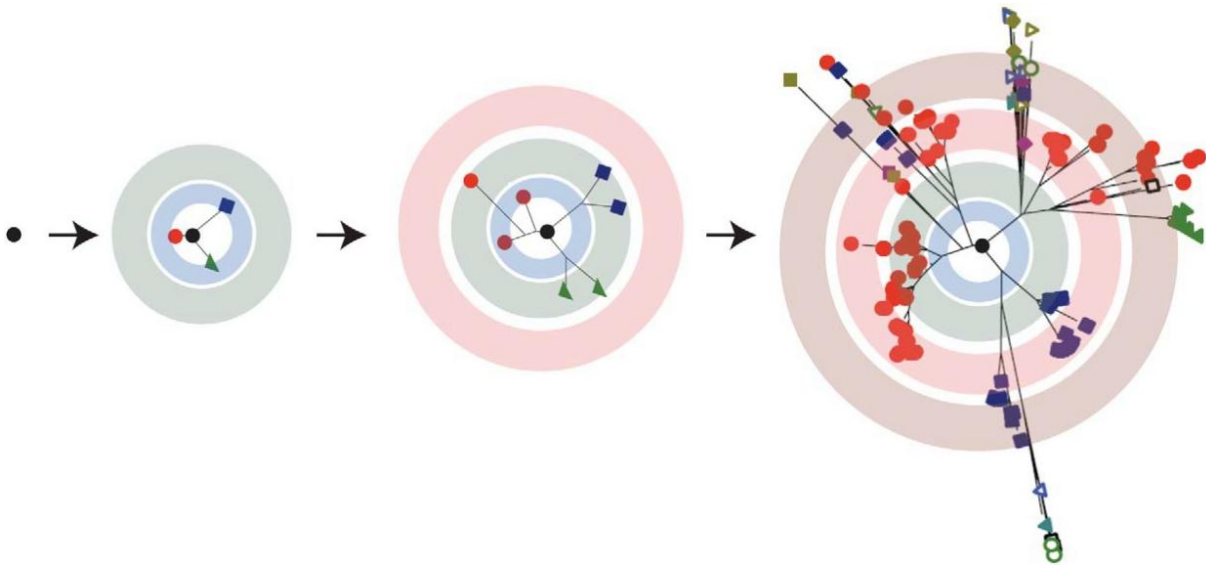


**Figure 7** Évolution d'un virus pandémique. Les pandémies du 20<sup>ème</sup> siècle ont été causées par un virus d'origine aviaire (1918), un réassortant aviaire-humain (1957 puis 1968) puis par un VIA réémergent de 1918 (1977). La pandémie de 2009 a été causée par un réassortant aviaire-porcine-humain (d'après Watanabe et al., 2012).

## 1.6 Conséquences de la variabilité des virus influenza

Qu'elle soit spontanée ou inhérente à un changement d'hôte voire à une modification de pression de sélection, la plasticité génétique des virus influenza leur confère un avantage adaptatif certain dans la mesure où cette diversité peut assurer la survie de leur espèce. On peut imaginer une population de virus influenza comme un nuage de mutants, ils sont suffisamment semblables pour évoluer dans le même environnement mais suffisamment différents pour qu'une des sous-populations puisse échapper à la pression de sélection que représentent le système immunitaire, un traitement antiviral ou un changement d'hôte.

Lorsque l'on aborde un tel sujet, la notion de quasi-espèces est incontournable. Un génome à ARN, lors d'un cycle de réplication, va générer des centaines de descendants mutés sur une position. Les cycles suivants entraineront une distribution plus complexe puisque les génomes obtenus seront de plus en plus éloignés de la matrice d'origine (figure 8). La fréquence d'un variant donné est théoriquement liée à sa capacité de réplication et de pérennité dans son environnement.



**Figure 8** Représentation schématique d'un virus se répliquant avec un fort taux de mutation. Une collection de mutants, au sein de la même cellule, est produite au fur et à mesure des cycles de réplication. Chaque branche relie deux variants différant d'une mutation. Les cercles concentriques représentent les cycles de réplication. Ce schéma bidimensionnel est une représentation extrêmement simpliste des connexions existant entre quasi-espèces (d'après Lauring and Andino, 2010).

Dans le cas des quasi-espèces, quand le taux de mutation est élevé au sein d'une population de virus se répliquant rapidement, un variant peu adapté peut être maintenu à une fréquence plus élevée que prévu grâce à sa cohabitation avec une sous-population bien adaptée au milieu (Lauring and Andino, 2010).

Les études détaillées ensuite dans ce manuscrit ne contiennent pas les éléments nécessaires à l'emploi du terme quasi-espèces ; les interactions entre les populations et les effets que les unes peuvent avoir sur les autres n'ont pas été approfondis. Il est donc plus adapté de définir les groupes de variants comme des sous-populations.

## **2 Analyse des sous-populations virales**

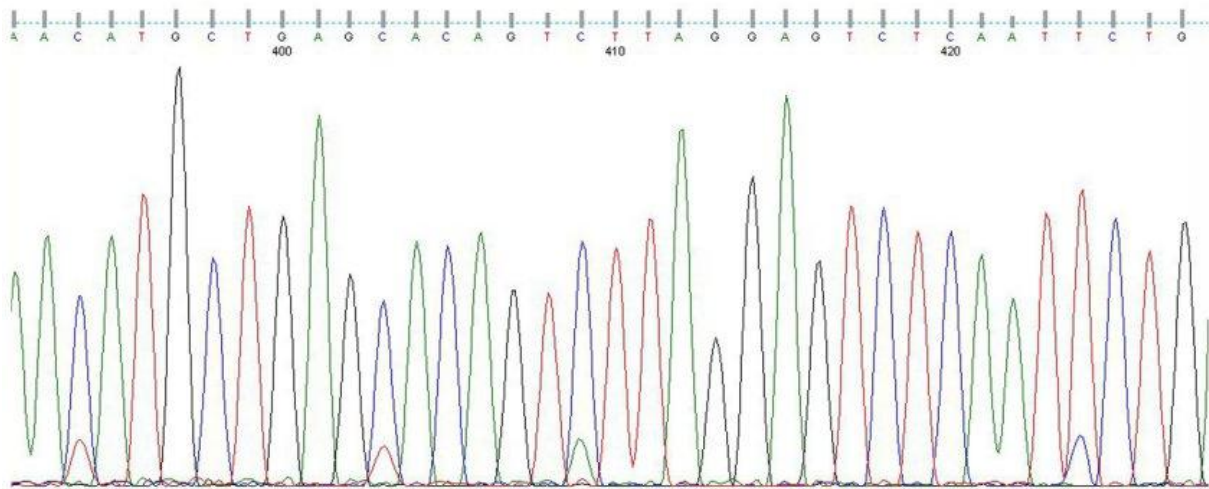
Comme évoqué précédemment (*cf.* §1.5, §1.6), la variabilité génétique des virus, et notamment des virus influenza représente une menace. Des sous-populations « dormantes » à caractère pandémique ou à potentiel hautement pathogène peuvent subsister chez n'importe quel type de sujet infecté. Un pourcentage élevé de maladies émergentes chez l'homme sont des zoonoses ; en raison de la propension humaine à étendre de plus en plus son « territoire », le contact avec des espèces animales sauvages réservoir est favorisé, cela accroît les probabilités d'émergence de nouvelles maladies chez les animaux domestiques ou chez les hommes (Blomström, 2011).

La meilleure façon d'anticiper ou de contrôler les maladies émergentes est d'élargir nos connaissances grâce à une analyse fine des populations virales circulant chez les animaux sauvages ou domestiques. L'utilisation d'outils permettant d'explorer la flore virale est indispensable. Le développement de ces outils a été initié pour plusieurs raisons : (i) une grande partie des micro-organismes est difficile voire impossible à cultiver ; (ii) la variabilité génétique au sein d'une même famille de virus contraint le développement de réaction d'amplification en chaîne par polymérase (ACP ou PCR en anglais) ciblant tous les virus d'une même famille ; (iii) le séquençage par la technique de Sanger (Sanger et al., 1977) fournit une séquence d'acide désoxyribonucléique (ADN) consensus qui représente uniquement la population majoritaire de l'échantillon.

### **2.1 Nouvelles générations de séquençage**

Les nouvelles générations de séquençage (NGS) bouleversent le monde de la biologie qui était dominé depuis les années 1980 par le séquençage suivant la technique de Sanger.

Cette dernière permet difficilement d'obtenir une vision exhaustive des micro-organismes présents dans un environnement donné. La séquence lue est présentée sous forme de chromatogramme (figure 9); des polymorphismes peuvent être observés mais il est impossible d'apprécier la quantité de variants présents ; les pics secondaires peuvent être dus à des mutations ponctuelles présentes sur des brins d'ADN distincts. Néanmoins, le clonage est une alternative adaptée à l'analyse des sous-populations, chaque clone bactérien représente une séquence d'ADN issue d'une PCR donc un variant précis, mais cette technique reste d'une efficacité limitée, fastidieuse, coûteuse et chronophage.

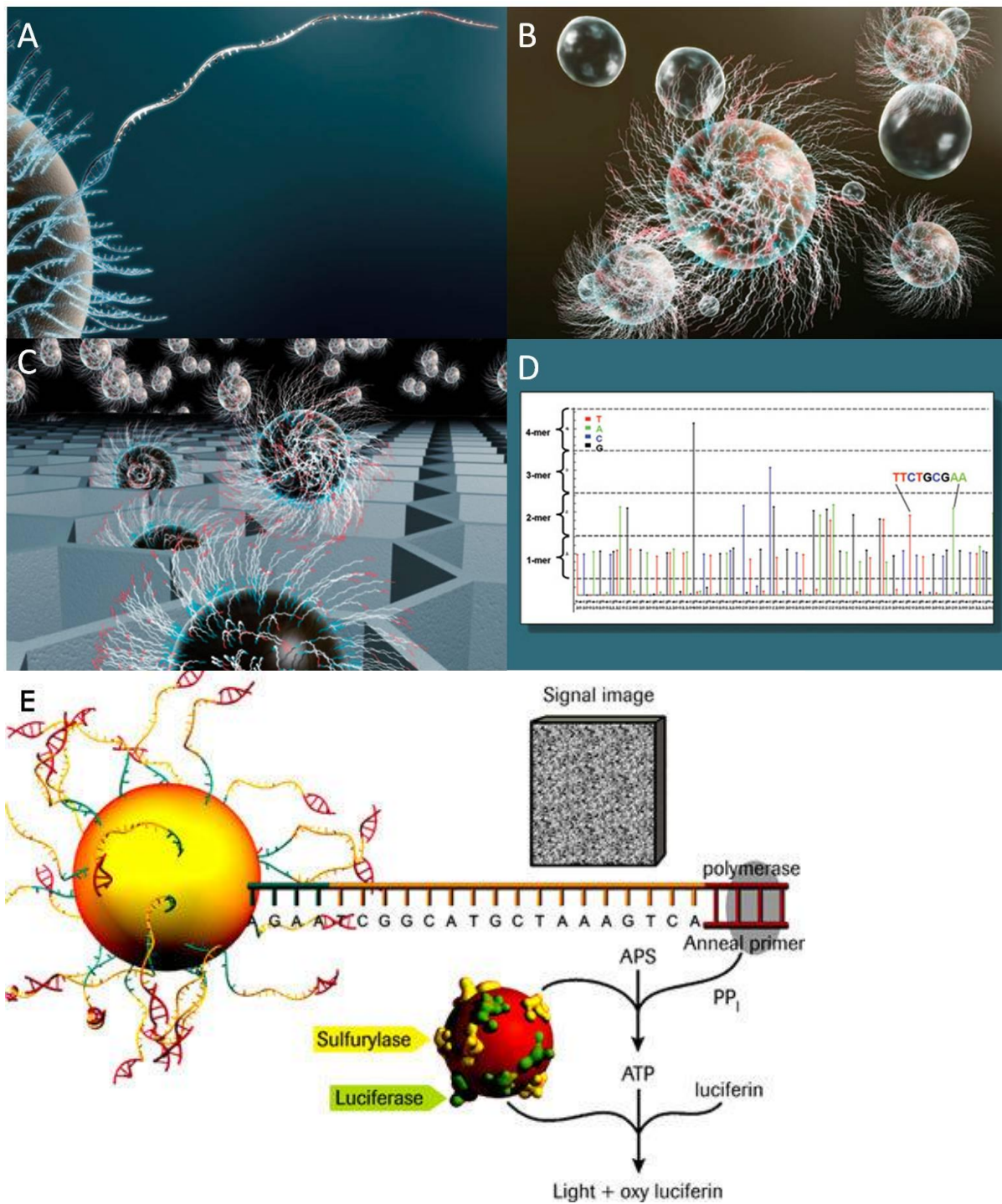


**Figure 9** Extrait d'un chromatogramme issu du séquençage Sanger d'un fragment d'ADN correspondant au gène de la PB2 d'un VIA. Quatre positions présentent des pics secondaires soit des SNP (chromatogramme affiché avec le logiciel *FinchTV* version 1.4.0) (modifié d'après Bourret et al., 2013).

Le séquençage à très haut débit permet la détection de mutations présentes à très faible fréquence et donne une idée de la structure de la population virale. Ces méthodes permettent de séquencer massivement et de façon consécutive des milliers de courts fragments d'ADN d'une taille variant entre 250 et 1000 paires de bases (pb). Le clonage n'est pas nécessaire à ces techniques, en effet, les échantillons sont directement séquencés après liaison ou immobilisation du fragment d'ADN sur un support particulier. Différentes techniques sont disponibles et se distinguent par la préparation des échantillons, la technologie de séquençage, l'acquisition des signaux d'incorporation de bases, la taille des séquences obtenues, l'analyse bioinformatique des données obtenues ou encore leur coût.

Les travaux décrits ci-après font exclusivement usage de la technologie 454 car les séquences obtenues sont de grande taille (environ 400 à 500 pb, taille maximale disponible à l'époque de l'analyse), cela est plus approprié à l'analyse de sous-populations et facilite l'assemblage du génome. Le paragraphe suivant est donc dédié à cette technologie.





**Figure 10** Quelques étapes de pyroséquençage 454. **A)** Hybridation d'un brin d'ADN à une microbille. **B)** Microbilles après amplification clonale du brin d'ADN. **C)** Distribution des microbilles en microplaque. **D)** Pyrogramme (modifié d'après [www.454.com](http://www.454.com)). **E)** Cascade de réactions du pyroséquençage 454. Lorsqu'un dNTP est incorporé, le pyrophosphate libéré entraîne la production d'ATP qui est transformé en signal lumineux par la luciférase (modifié d'après [www.454.com](http://www.454.com)).

## 2.2 Pyroséquençage 454

Dès 2005, le pyroséquençage 454 a été la technique de séquençage à très haut débit la plus couramment employée (Margulies et al., 2005), elle est commercialisée par Roche/454.

Cette approche innovante repose sur le principe de séquençage par synthèse : la composition en bases est déterminée au travers de la détection d'une activité chimioluminescente. Cette dernière est entraînée par l'incorporation d'un désoxyribonucléotide triphosphate (dNTP) pendant la synthèse d'un brin d'ADN complémentaire par une ADN polymérase.

De façon résumée, les étapes de la méthode sont les suivantes :

L'ADN à séquencer est au préalable fragmenté mécaniquement, cette étape est appelée « nébulisation » car l'ADN est soumis à 2,1 bar d'azote gazeux. Les fragments obtenus ont alors une taille moyenne de 400 à 500 pb, taille correspondant aux performances de séquençage du 454.

Les fragments sont liés à des séquences adaptatrices leur permettant un attachement à des microbilles (figure 10-A) pour une PCR en émulsion (emPCR) qui est une amplification clonale car chaque bille arbore un seul fragment d'ADN. À la suite de l'emPCR, les billes sont recouvertes sur toute leur surface du même brin d'ADN amplifié (figure 10-B), cela permet d'accroître le signal lumineux qui surviendra lors du séquençage.

Les microbilles sont distribuées sur une microplaque (figure 10-C), dans chaque puits se trouve une microbille qui est incubée en présence d'ADN polymérase, d'ATP sulfurylase, de luciférase et d'apyrase.

Lors de la synthèse du brin complémentaire, quand l'ADN polymérase incorpore un dNTP complémentaire au brin matrice, un pyrophosphate est libéré, ce dernier est converti en adénosine triphosphate (ATP). L'ATP interagit avec la luciférase pour catalyser la conversion de luciférine en oxyluciférine, ce qui libère de l'énergie sous forme de photons en quantités proportionnelles à l'ATP produit par l'incorporation du ou des dNTP(s) (figure 10-E).

Les molécules d'ATP et les dNTPs inutilisés persistants dans le milieu sont dégradés par l'apyrase puis des lavages ont lieu. Cette procédure est répétée tout au long de la synthèse du brin d'ADN complémentaire.

La lumière produite par l'incorporation des dNTPs est détectée par une caméra qui traduit le signal en un pyrogramme (figure 10-D).

Ce processus est répété séquentiellement, les bases sont introduites dans le milieu de réaction tour à tour : des adénines (A), thymines (T), guanines (G), cytosines (C) avec des lavages entre chaque incorporation de bases. Au moment de l'introduction des A dans le milieu, si trois bases sont incorporées dans le brin d'ADN complémentaire en cours de synthèse, l'intensité lumineuse produite sera traduite informatiquement par un pic trois fois plus important que pour une seule base. Ce signal lumineux étant concomitant à la présence des A, la séquence déduite sera AAA.

Les performances de la technologie 454, à l'époque de l'étude, permettaient d'obtenir 640 000 séquences de 400 à 500 pb soit près de 300 millions de bases lues.

Le taux moyen d'erreur a été estimé à 1.07 % ; les facteurs de ce taux d'erreur seraient : (i) un défaut d'acquisition du signal lumineux ; (ii) un lavage incomplet du puits donc la persistance de bases à l'origine d'une incorporation prématurée ; (iii) un échec d'incorporation de base donc l'élongation du brin d'ADN s'arrête jusqu'à la prochaine introduction de la base complémentaire dans le milieu (Gilles et al., 2011).

De plus, il a été démontré que 39 % des erreurs observées sont dues à l'effet homopolymère (Huse et al., 2007). En effet, lorsque la même base est répétée quatre fois ou plus, des substitutions ou des bases indéterminées (N) apparaissent ; ce phénomène est davantage marqué dans le cas d'homopolymères de A ou de T et cela pourrait être dû au fait que lorsqu'une adénine est incorporée, le ratio [signal lumineux / bruit de fond] n'est pas aussi puissant que pour les autres bases (dans Becker et al., 2012).

### 3 Projets scientifiques

Les deux études décrites dans ce manuscrit ont été explorées par pyroséquençage 454 pour (i) rechercher des populations virales ultra minoritaires ayant émergé chez un nouvel hôte et étant responsables d'une importante mortalité et (ii) pour suivre au niveau génétique l'adaptation d'un virus aviaire sur cellules de mammifère.

La nécessité de séquencer des génomes grippaux totaux, de rechercher des populations virales ultra minoritaires et de reconstituer un scénario d'évolution virale a naturellement orienté les travaux vers l'utilisation de nouvelles technologies de séquençage.

La masse de données générées par le séquençage à très haut débit permet de déduire avec fiabilité les liens épidémiologiques existants entre les isolats à l'échelle spatiale et chronologique, de plus les chaînes de transmissions inter-hôtes peuvent être établies. L'évolution des virus est une caractéristique biologique ayant un fort impact en virologie médicale, avec des conséquences sur la pathogénicité, le franchissement de la barrière d'espèce ainsi que l'échappement aux antiviraux et aux vaccins.

Les deux cas approfondis ci-dessous ont, au cours de leur analyse, induit les problématiques suivantes :

Un virus émergent suite à un changement de pression de sélection émerge-t-il *de novo* ou s'agit-il d'une bascule de la représentation des populations virales avec une population initialement ultra-minoritaire qui devient majoritaire ?

Quel type de scénario d'adaptation est adopté par les virus influenza aviaires lorsqu'ils sont cultivés sur cellules porcines ?

Les résultats obtenus ont permis de proposer des réponses et ont fait l'objet de publications scientifiques (Annexe 1 et Annexe 2).

## Première partie expérimentale

Suivi de VIA de terrain : séquençage du génome total et suivi de l'évolution de la neuraminidase par pyroséquençage 454

*Field monitoring of avian influenza viruses: whole-genome sequencing and tracking of neuraminidase evolution using 454 pyrosequencing*

(Croville et al., 2012)

# 1 Introduction

Les infections à VIAFP chez les volailles ne présentent généralement pas de signes cliniques et sont par conséquent très peu étudiées malgré le potentiel qu'ont ces virus à devenir hautement pathogènes. La signature des VIAHP se trouve au niveau de la HA, lorsque celle-ci comporte une série d'acides aminés (aa) basiques à proximité du site de clivage, ce dernier peut être opéré par des protéases ubiquitaires, le spectre tissulaire de la réplication virale devient alors élargi.

Des facteurs d'adaptation des VIAFP aux volailles sont également connus, l'un d'entre eux est caractérisé par des délétions d'aa au niveau de la tige de la NA, cela survient lors de la circulation de VIA chez les volailles. Ces délétions permettent l'adaptation du virus aux volailles (Li et al., 2011; Matrosovich et al., 1999) et entraînent une modification du tropisme du virus depuis l'appareil digestif vers l'arbre respiratoire des oiseaux (Sorrell et al., 2010). Des études de génétique inverse ont récemment confirmé que des délétions au niveau de la NA sont associées à une augmentation de la virulence des VIA chez la volaille (Matsuoka et al., 2009; Munier et al., 2010; Zhou et al., 2009). Ces délétions ont régulièrement été observées chez des virus infectant des volailles, d'un point de vue phylogénétique ces virus sont proches de ceux circulant chez les oiseaux sauvages (Giannecchini et al., 2006; Di Trani et al., 2004). Néanmoins, les virus parentaux et délétés n'ont jamais été suivis jusqu'alors dans des conditions de terrain.

Cette étude consiste en un suivi d'épisode infectieux à VIA de sous-type H6N1 au sein d'un élevage de canards de Barbarie, l'infection s'est ensuite propagée à l'élevage de dindes présent sur le même site. L'évolution génétique a été suivie sur les huit segments d'ARNv et la NA fut au centre de l'investigation. En effet, grâce à l'usage de séquençage à très haut débit, des variants arborant une NA tronquée ont été détectés chez le canard à une fréquence ultra-minoritaire alors qu'ils représentaient 100 % de la population virale chez la dinde. Dans des temps plus tardifs, les populations virales étudiées chez la dinde ont présenté deux types de variants majoritaires avec deux profils de délétion de la NA.

## **2 Résultats**

### **2.1 Historique et premières analyses**

Trois jours après le départ des canards pour l'abattoir, les dindes ont présenté des signes cliniques respiratoires et de la prostration, ce qui s'est soldé par une mortalité d'environ 10 % de l'effectif total des animaux. Des prélèvements ont été effectués sur 10 dindes affectées et, au même moment, du lisier de canards a été échantillonné. Les 10 animaux ont été autopsiés et chacun d'entre eux était atteint de trachéite exsudative sévère, d'aérosacculite et de pneumonie. Une antibiothérapie de routine a été mise en place pour prévenir une éventuelle infection bactérienne secondaire mais cela n'a pas limité la mortalité. Des autopsies ont donc de nouveau été réalisées sur 10 autres oiseaux, huit jours plus tard. Des écouvillons trachéaux ont été exécutés et des organes prélevés sur les 10 volailles. Tous les prélèvements ont été réalisés par des vétérinaires et ont été mis à disposition du laboratoire.

Des réactions de PCR quantitative en temps réel se sont révélées négatives pour *Mycoplasma gallisepticum* et *Mycoplasma synoviae* ou encore pour le metapneumovirus aviaire (Guionie et al., 2007) au niveau trachéal mais positives en VIA (Fouchier et al., 2000).

Un immunomarquage ciblant l'influenza A a révélé des cellules infectées au niveau de l'épithélium trachéal et du parenchyme pulmonaire.

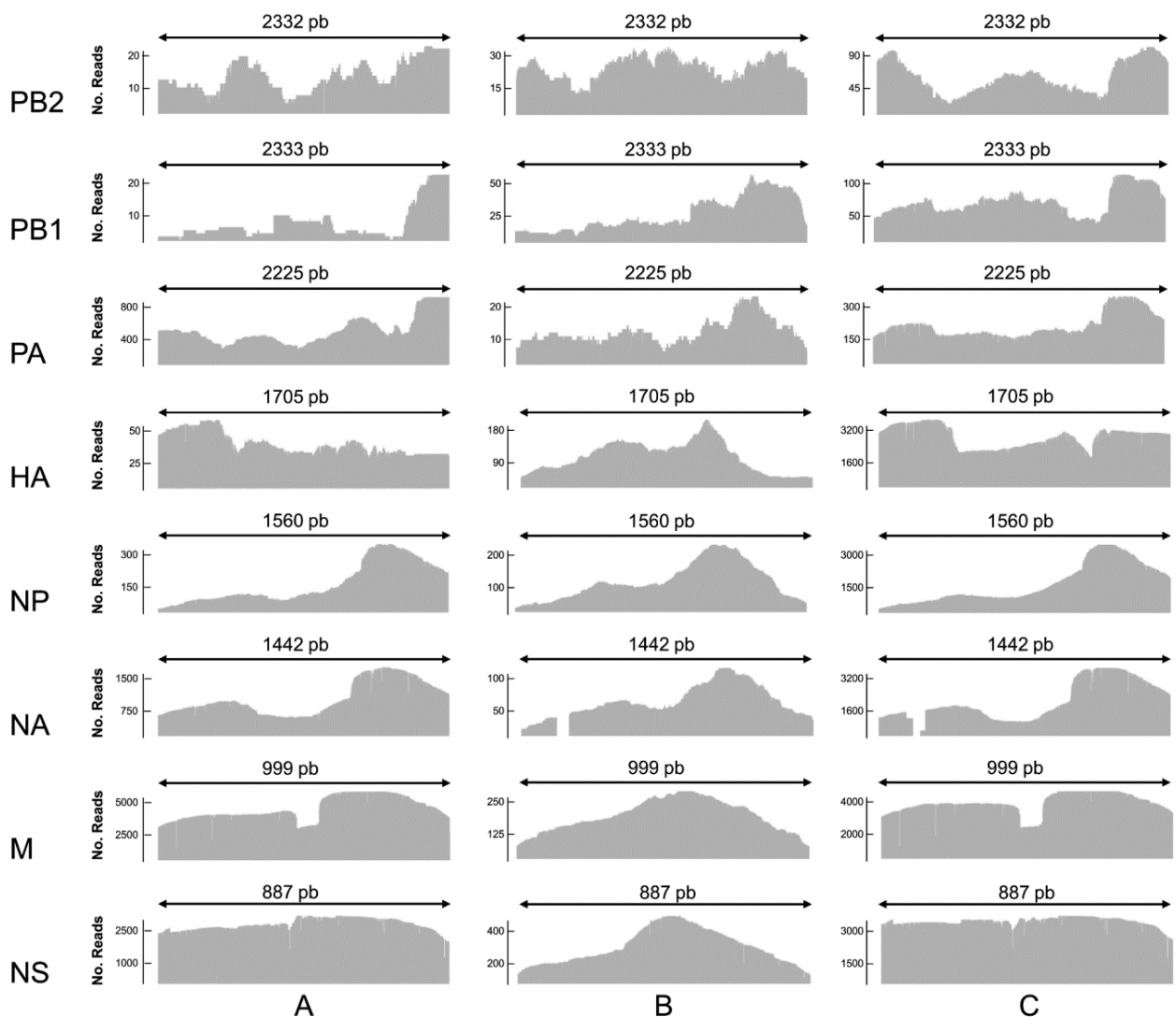
### **2.2 Analyse des séquences virales des prélèvements de canards et de dindes**

Des RT-PCR couvrant le génome viral total ont été réalisées sur les prélèvements de lisier de canards et sur les écouvillons de trachées de dindes en employant un jeu d'amorces spécifiques des

huit segments d'ARN des virus influenza (Hoffmann et al., 2001). Le séquençage par la technique de Sanger appliqué aux produits d'amplification obtenus précédemment a permis l'identification d'un virus influenza A de sous-type H6N1. La comparaison des séquences de virus de lisier de canards avec celles de virus de trachées de dindes a révélé plus de 99,8 % d'identité nucléotidique.



**Figure 11** Électrophorèse sur gel d'agarose à 1% des produits d'amplification de la NA par PCR spécifique de la zone de délétion. Dépôt n°1 : lisier de canard. Dépôts n° 2 à 5 : prélèvements de dindes en début d'infection, les bandes correspondent à la délétion de 60 nucléotides. Dépôts n°6 à 9 : prélèvements de dindes en fin d'infection, les bandes correspondent à la délétion de 69 nucléotides et/ou celle de 60 nucléotides. M : marqueur de poids moléculaire. \* : NA complète ;  $\Delta 60$  et  $\Delta 69$  : délétions de 60 et 69 nucléotides.



**Figure 12** Représentation graphique de la couverture et de la profondeur de séquençage des génomes totaux obtenus à partir du lisier de canards (A), des prélèvements précoces de dindes (B) et des prélèvements tardifs de dindes (C). PB1 : polymérase basique 1 ; PB2 : polymérase basique 2 ; PA : polymérase acide ; HA : hémagglutinine ; NP : nucléoprotéine ; NA : neuraminidase ; M : protéine de matrice ; NS : protéine non structurale.



Deux profils de délétions ont été observés : (i) une délétion interne de 60 nucléotides du gène de la NA se traduisant par une délétion de 20 aa au niveau de la tige de la NA dans chacun des échantillons de dindes et (ii) un assortiment comprenant la délétion de 60 nucléotides du gène de la NA citée précédemment et une nouvelle délétion de 69 nucléotides du gène de la NA dans certains prélèvements de dindes effectués huit jours après l'infection (figure 11). Quatre des 10 dindes écuvillonnées au stade avancé de l'infection présentaient le profil mixte de délétions.

### **2.3 Analyse des neuraminidases par PCR quantitative en temps réel**

Afin de déterminer si les formes de neuraminidase tronquée étaient présentes, même à un faible niveau, dans les échantillons de lisier de canards, une PCR quantitative en temps réel a été mise au point pour discriminer les formes complètes des formes tronquées de la neuraminidase.

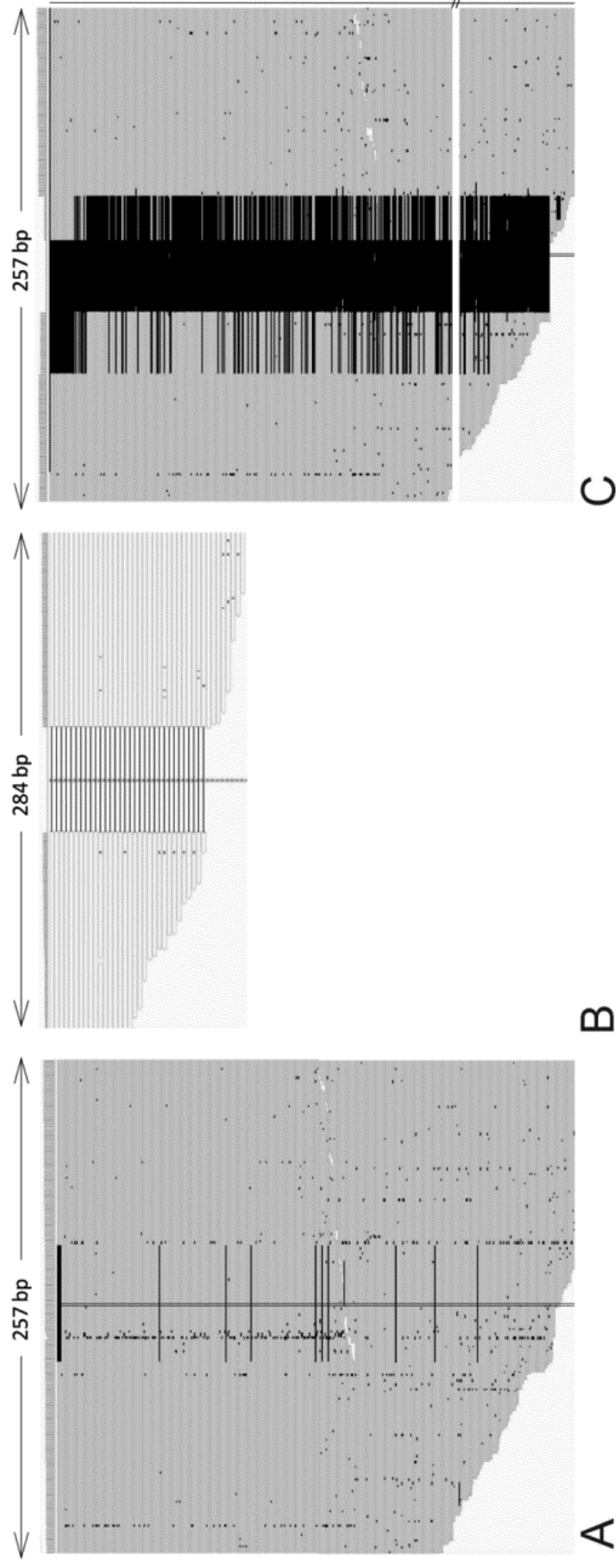
L'estimation de la quantité de copies de génomes a été déduite d'une gamme plasmidique (de  $10$  à  $10^6$  copies avec un pas de dilution de 10) contenant la séquence de NA complète ou tronquée. Le seuil de détection de la réaction était de 10 copies, chaque échantillon a été analysé en triplicat. Cette analyse a permis de détecter exclusivement la forme complète de NA dans les échantillons de canards à hauteur de  $787 \text{ copies} \pm 311 \text{ EC}$  (Ecart-type) et uniquement les formes tronquées de NA dans les prélèvements trachéaux de dindes soit  $5109 \text{ copies} \pm 1585 \text{ EC}$ .

### **2.4 Séquençage à très haut débit du génome complet des virus de canards et de dindes**

Les génomes complets ont été obtenus par séquençage à très haut débit 454 pour les trois types d'échantillon (lisier de canards ainsi que trachées de dindes aux stades précoce et tardif de l'infection). En moyenne, 21 600 séquences d'ADN ont été obtenues pour chaque type d'échantillon, avec une profondeur de lecture du génome de 680 fois (figure 12). Pour chacun des gènes, les polymorphismes nucléotidiques uniques (ou SNP : *single nucleotide polymorphism*) ont été répertoriés avec un seuil de détection arbitrairement fixé à cinq pourcent. Le tableau II regroupe les SNP non synonymes et leur représentation dans les trois types d'échantillons pour PB2 et HA. Huit SNP au niveau de PB2 qui étaient absents des échantillons de canards sont devenus majoritaires à 100 % dans les prélèvements de dindes. Des mutants minoritaires présents dans le lisier de canards n'ont pas été sélectionnés et ont

**Tableau II** SNP non synonymes répertoriés pour les gènes PB2 et HA

Protein	Nucleotide			Protein			Population ratio (%)			
	Base(s)	Position	Substitution	Amino acid	Position	Substitution	A	B	C	
PB2	AGA	181	AAG	R	61	K	0	95	95	
	C	300	A	N	100	K	0	0	8	
	A	346	C	K	116	Q	0	0	5	
	A	361	G	K	121	E	0	7	10	
	C	440	T	T	147	I	0	100	100	
	A	563	G	E	188	G	25	0	0	
	A	569	G	K	190	R	23	0	0	
	G	571	A	E	191	K	21	0	0	
	G	874	A	V	292	I	0	100	100	
	GTG	883	ATT	V	295	I	0	100	100	
	GTG	1189	GCA	V	397	A	0	100	100	
	G	1261	A	V	421	I	0	0	23	
	ATT	1531	GTA	I	511	V	0	100	100	
	CCT	1735	ACA	P	579	T	0	14	0	
	ACT	1753	CCC	T	585	P	25	100	100	
	AAG	1756	CAA	K	586	Q	0	7	14	
	GAT	2038	GGC	D	680	G	0	100	100	
	HA	A	353	G	K	118	R	0	100	62
		C	358	T	L	120	F	6	0	0
		A	509	G	K	170	R	0	0	6
G		534	T	K	178	N	9	0	0	
A		560	T	Q	187	L	0	0	38	
G		685	A	E	229	K	0	0	25	
C		692	T	A	231	V	0	0	25	
G		722	C	G	241	A	0	0	14	
T		1004	C	L	335	P	0	0	20	
C		1041	A	F	347	L	0	0	6	
T		1262	C	I	421	T	0	0	23	



**Figure 13** Alignement des séquences nucléotidiques correspondant à la tige de la NA et laissant apparaître les profils de délétion. La troncation de 60 pb (positions 161 à 220) est retrouvée dans 13 des 670 séquences du prélèvement de canards (A) et dans la totalité des séquences des prélèvements effectués en début d'infection chez la dinde (B). Un assortiment de délétions de 60 et 69 nucléotides (positions 129 à 197) apparaît huit jours plus tard dans les prélèvements de dindes (C).

totalemment disparu des échantillons de dindes. Au niveau de l'hémagglutinine, un seul SNP (K118R) a été identifié comme étant apparu chez la dinde et est resté majoritaire chez ces dernières. Huit autres SNP minoritaires, représentant 6 à 38 % des séquences des prélèvements tardifs ont été répertoriés entre les positions 170 et 241 soit la région du récepteur d'attachement de la HA. Les données des six autres segments sont disponibles en Annexe 3, des SNP ont été observés chez les dindes au niveau de PB1 (sept positions), PB1-F2 (14 positions), PA (deux positions) et HA (une position).

Une analyse ciblée a été réalisée sur le gène de la NA ; comme attendu, un variant présentant une délétion de 60 nucléotides représentait 100 % de la population virale des prélèvements de dindes effectués en début d'infection et un mélange de variants arborant les délétions de 60 ou 69 nucléotides constituait les prélèvements de dindes effectués en fin d'infection.

L'alignement des séquences du gène de la NA des trois types d'échantillons met clairement en évidence les différents profils de délétion (figure 13).

De façon intéressante, les analyses menées sur les prélèvements de lisier de canard ont mis en évidence une population ultra-minoritaire de variants possédant la délétion de 60 nucléotides du gène de la NA sur 13 séquences sur les 670 obtenues, soit à hauteur de 1,94 % des lectures pour cette zone ; il est important de noter que les autres moyens mis en œuvre n'ont pas permis la détection de cette population. Ces résultats sont la preuve qu'une infime quantité de virions à neuraminidase mutée était présente dans le nuage de mutants et a émergé dans le nouvel environnement que représentait l'appareil respiratoire des dindes.

### **3 Discussion**

Cette étude a consisté en un suivi de l'évolution de virus influenza aviaires faiblement pathogènes par pyroséquençage d'échantillons cliniques. Les génomes complets de trois lots d'échantillons ont été obtenus par séquençage profond à très haut débit permettant une analyse des sous-populations virales. La couverture de séquençage était suffisante pour observer les variations génétiques des populations très minoritaires au potentiel émergent en cas de modification de la pression de sélection. Le choix méthodologique s'est porté sur l'usage de la technologie de pyroséquençage 454 qui produisait des séquences de 400 à 500 nucléotides, ceci facilitant les travaux d'assemblage des séquences. L'observation la plus notable a été celle de l'émergence de virus mutés au niveau de la NA suite à l'infection de

dindes par un VIA de canard. L'intérêt de ce cas a été le suivi de terrain à l'échelle d'un élevage avec un lien direct entre les virus d'origine et ceux ayant subi des mutations leur conférant vraisemblablement des capacités d'adaptation au nouvel hôte.

Les données présentées ici vont en la faveur de l'émergence d'une population virale concomitante à la contamination de l'élevage de dindes. De plus, aucune forme de virus à la NA non mutée n'a été détectée dans les prélèvements de dindes ; cela laisse suggérer un fort avantage adaptatif des formes tronquées aux dépens de la forme complète. Après huit jours de circulation virale chez les dindes, certains oiseaux présentaient un second variant arborant une délétion plus importante qui ne recouvrait pas totalement la troncature de 60 pb précédemment observée. Cet allèle tronqué de 69 pb n'a pas été détecté dans les prélèvements de dindes effectués en début d'infection, probablement à cause de la profondeur de lecture qui n'aurait pas été suffisante. Ces données suggèrent une importante plasticité de la NA.

Les événements de troncature de la NA, réduisant l'activité enzymatique de la protéine, sont associés à des modifications du profil de glycosylation de la tête globulaire de la HA, ce qui se traduit par une modification d'affinité de la HA pour les récepteurs cellulaires (Matsuoka et al., 2009; Mitnaul et al., 2000). Le suivi de l'évolution de la HA a permis de mettre en évidence l'émergence de mutants minoritaires présentant des substitutions nucléotidiques principalement au niveau du site d'attachement aux récepteurs, dans la tête globulaire de la glycoprotéine. Ce type de SNP a été observé sur 38 % des séquences virales des prélèvements de fin d'infection. Aucune de ces mutations n'est décrite comme néfaste pour les sites de glycosylation de la HA, elles sont supposées contribuer à l'amélioration des capacités d'attachement aux récepteurs cellulaires de l'arbre respiratoire supérieur des dindes et/ou contourner la réponse immunitaire de l'hôte (McHardy and Adams, 2009).

Il faut noter que la mortalité spontanée des dindes est contemporaine de la détection des délétions de la NA alors que les canards n'ont présenté aucun signe clinique dû à l'infection. Bien que des délétions de la tige de la NA soient associées à une augmentation de la virulence des VIAFP chez des poulets infectés par des virus produits par génétique inverse (Hoffmann et al., 2012; Munier et al., 2010), il reste à démontrer la corrélation entre les délétions décrites dans ce cas et une augmentation de la virulence chez la dinde.

Au sein de cet élevage, les deux espèces d'oiseaux étaient élevées séparément dans deux bâtiments distincts, il se peut que la contamination des dindes soit liée à une faille de biosécurité ayant eu lieu lors du transfert des canards de l'élevage à l'abattoir. Ce fait

confirme la sensibilité particulière des dindes vis-à-vis des VIA (Pillai et al., 2010), de plus, des données récentes en épidémiologie démontrent une importante capacité de transmission des VIAFP au sein d'élevages de dindes (Comin et al., 2011). Cet épisode infectieux met l'accent sur les risques d'élevages aviaires mixtes de volailles et d'oiseaux aquatiques, situation fréquente en Asie pouvant engendrer l'émergence rapide de variants davantage pathogènes.

Dans d'autres études, les techniques de séquençage à très haut débit ont fait leurs preuves pour le suivi de l'évolution génétique des virus en ciblant des positions bien précises soumises à la pression de traitements antiviraux, notamment chez des patients infectés par le virus de l'immunodéficience humaine (VIH) (Hedskog et al., 2010) ou chez des patients traités à l'oseltamivir lors d'épisodes de grippe (Inoue et al., 2010). Une approche plus large du suivi de l'évolution des génomes viraux devrait permettre une détection précoce de mutants sans cibler de génotypes particuliers et de suivre les événements de mutations ayant lieu le long des génomes.

L'équipement de pyroséquençage 454 a l'avantage de produire des séquences longues (400 pb en 2011, 800 à 1000 pb en 2013), ce qui facilite l'assemblage des séquences dans une approche de séquençage de génomes complets. D'autres techniques sont disponibles, chacune d'entre elles étant en constant développement avec un accroissement spectaculaire de performances au cours du temps.

Bien que les données présentées dans ce document n'expliquent pas comment la délétion de la NA est survenue chez le canard, elles indiquent que ce mutant peut exister à une échelle ultra-minoritaire chez les oiseaux aquatiques et être amplifié lors d'un changement d'hôte.

## Deuxième partie expérimentale

Analyse par séquençage à très haut débit de l'adaptation d'un virus influenza A de canard sur cellules porcines et séquençage de génomes complets

*Whole-genome, deep pyrosequencing analysis of a duck influenza A virus evolution in swine cells*

(Bourret et al., 2013)

## 1 Introduction

Les oiseaux sauvages aquatiques sont considérés comme le réservoir primaire et naturel des virus influenza A (Alexander, 2000; Hinshaw et al., 1982; Horimoto and Kawaoka, 2001; Webster et al., 1992). Certains virus peuvent être transmis des oiseaux aquatiques vers les mammifères que sont les hommes, les porcs, les chevaux, les mammifères marins, les furets ou encore les visons (Forrest and Webster, 2010) chez qui ils peuvent s'adapter. Lors de ce type de changement d'hôte les infections à virus influenza peuvent avoir d'importantes conséquences avec notamment de la mortalité. Les virus influenza font partie des agents pathogènes respiratoires majeurs chez l'homme ([www.who.int](http://www.who.int)) et le porc (Kuntz-Simon et al., 2010). De plus, les porcs sont suspectés de faire office d'hôte intermédiaire pour l'adaptation des VIA aux humains et de potentiellement permettre un réassortiment entre virus d'origines aviaire et humaine (Castrucci et al., 1993; Horimoto and Kawaoka, 2005; Scholtissek et al., 1985). L'étude de l'évolution de VIA chez le porc est donc d'un intérêt majeur dans l'épidémiologie des virus influenza.

Dans cette étude, le pyroséquençage 454 a permis de suivre l'évolution du génome d'un virus influenza de canard adapté à des cellules de trachée de porc.

## 2 Résultats

Le virus Wt a été soumis à 10 passages successifs sur cellules de lignée de trachée de porcelet (*newborn pig trachea* ou NPTr) pour obtenir le virus descendant adapté Ad. Les virus Wt et Ad ont d'abord été analysés par séquençage Sanger de sorte à obtenir les séquences consensus des divers segments d'ARNv, la technique de Sanger permet également de détecter qualitativement des variants minoritaires par l'observation de pics secondaires sur les chromatogrammes. L'analyse par pyroséquençage 454 réalisée permet, quant à elle, d'obtenir des informations émanant de séquences nucléotidiques individuelles. La détection de sous-populations minoritaires, de polymorphismes associés et l'estimation de leurs fréquences est alors possible de sorte à établir le scénario d'évolution virale qui a eu lieu lors de l'adaptation du virus Wt sur cellules NPTr.



## 2.1 Analyse des séquences « Sanger »

Les virus Wt et Ad ont été en première intention analysés par séquençage « Sanger » dans le laboratoire d'origine de ces prélèvements (Bourret, 2012).

Au niveau des séquences consensus il n'a pas été observé de différences entre les virus Wt et Ad pour les segments 1 (PB2), 5 (NP) et 6 (NA).

Sur le segment 2, 114 différences au niveau nucléotidique ont été rapportées, ces dernières se traduisant par 7 et 11 modifications respectivement au niveau protéique pour PB1 et PB1-F2. Sur le segment 4 (HA), qui est représenté par un mélange d'allèles H1 et H11, deux mutations non synonymes ont été détectées sur l'allèle H1. Sur le segment 7 (M), 32 différences nucléotidiques synonymes ont été observées. Sur le segment 8 (NS), 35 différences nucléotidiques ont été relevées pour deux mutations au niveau protéique pour NS1 et une mutation protéique pour NEP.

Dans le cas des segments 2, 7 et 8, les précurseurs du virus Ad ont été détectés dans les séquences du virus Wt sous forme de pics secondaires sur les chromatogrammes.

Ces données suggèrent que, pour ces gènes, les allèles représentant la population Ad étaient déjà présents dans la population de virus Wt et sont devenus majoritaires après des passages successifs sur cellules porcines. L'observation des chromatogrammes du virus Ad indique que la diversité des génomes viraux a diminué sous la contrainte imposée par le protocole de sélection.

Des questions d'ordre quantitatif et qualitatif se sont posées au regard de ces résultats :

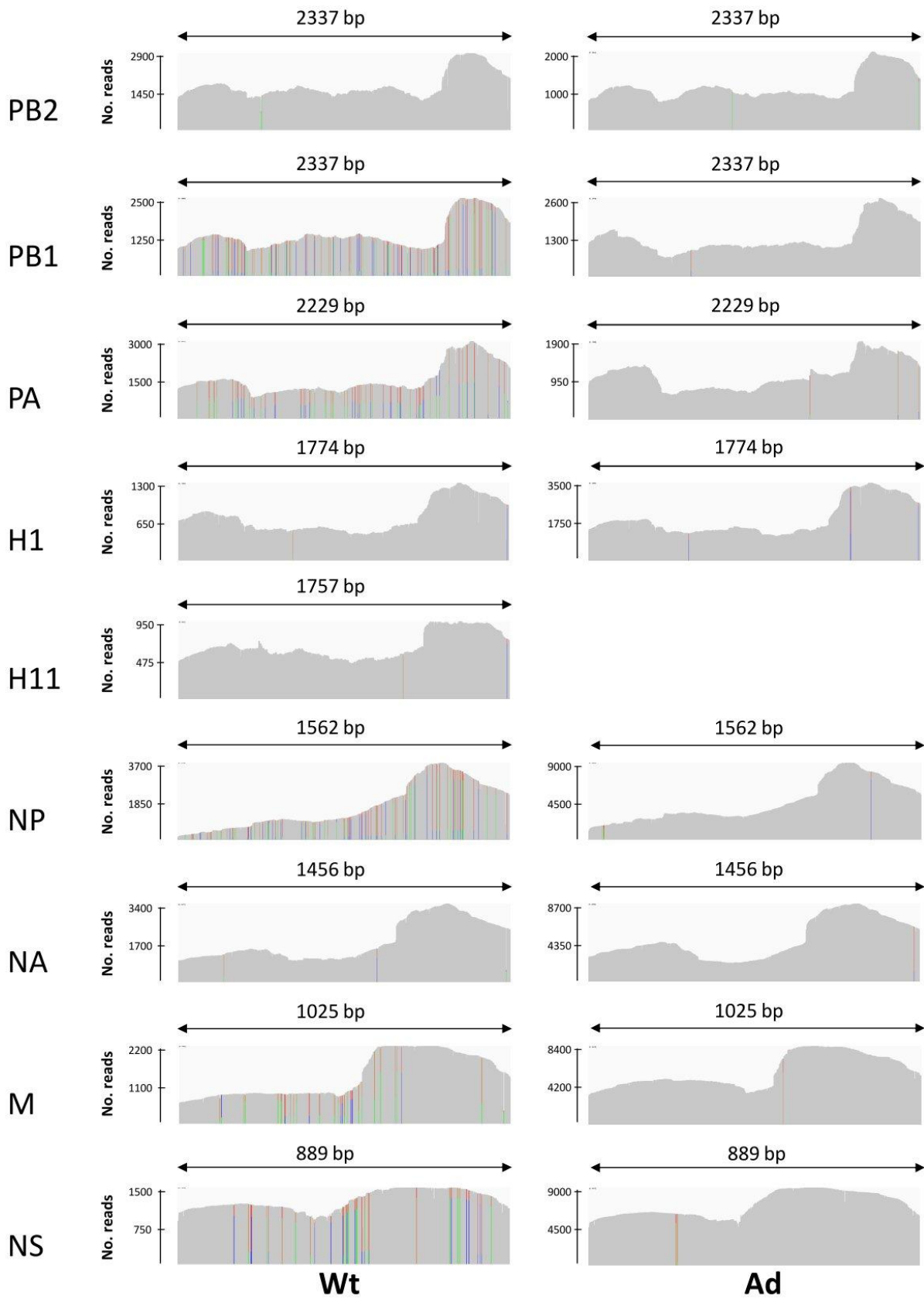
Dans quelles proportions les segments 2, 7 et 8 du virus Ad étaient-ils présents dans le stock de virus Wt ?

Ces segments ont-ils muté ou s'agissait-il de réassortiments ?

Les mutations caractéristiques des segments 3 et 4 du virus Ad étaient-elles présentes sous forme de population ultra minoritaire au sein de la population Wt ou sont-elles apparues *de novo* ?

La population de virus Ad était-elle pure ou existait-il des populations ultra minoritaires non détectées par le séquençage Sanger ?

L'allèle H11 a-t-il totalement disparu du stock de virus Ad ?



**Figure 14** Niveaux de couverture et de variabilité avec un seuil de détection des allèles mineurs fixé à 5 % pour les virus Wt et Ad. Les barres colorées indiquent qu'un polymorphisme nucléotidique est détecté à cette position avec une fréquence supérieure au seuil de 5 %. L'échelle horizontale indique la longueur du gène alors que l'échelle verticale indique la profondeur de lecture.

Quelle était la variabilité, s'il devait y en avoir une, des segments 1, 5 et 6 du virus Wt ?

Toutes ces questions ont pu être abordées grâce au pyroséquençage 454 des génomes complets des virus sauvage et adapté.

## **2.2 Présentation des résultats de pyroséquençage 454**

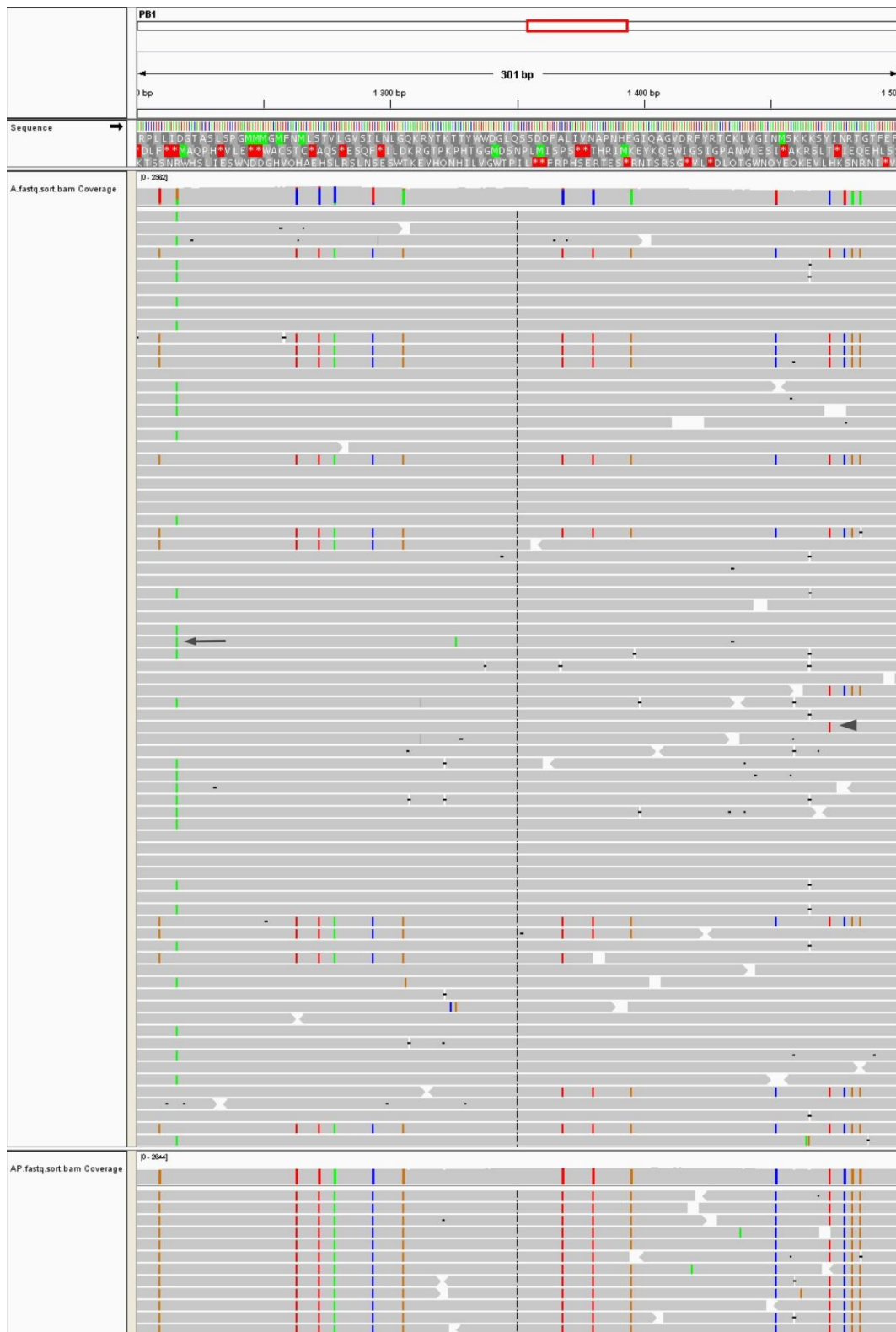
Concernant le virus Wt, 52 346 séquences ont été obtenues avec une longueur moyenne de 413 pb. Cela s'est traduit par une profondeur de lecture moyenne de 1589 (notée 1589 x) par position. Concernant le virus Ad, 93 905 séquences ont été obtenues avec une longueur moyenne de 416 pb soit 2 868 x de profondeur de lecture moyenne. La couverture du génome ainsi que la profondeur de lecture obtenues pour chacun des gènes des virus Wt et Ad sont décrites en figure 14. La profondeur de lecture a varié de quelques centaines à quelques milliers de lectures par position. Le seuil de détection de polymorphismes a arbitrairement été fixé à 5 %, de nombreux polymorphismes sont observables pour les segments 2, 3, 5, 7 et 8 du virus Wt par rapport aux segments 1, 4, 6 et tous les segments du virus Ad qui ne montrent que très peu de polymorphismes.

Ces données indiquent que le virus Ad est composé d'une population davantage homogène que le virus Wt, avec seulement 17 SNP minoritaires détectables (en tenant en compte du seuil de détection de 5 %).

## **2.3 Redistribution des populations minoritaires initialement présentes au sein de la population Wt**

### **2.3.1 Précurseurs des gènes 2, 7 et 8 du virus Ad dans la population Wt**

Pour chacune des positions différant entre la séquence consensus de Wt et Ad pour les segments 2, 7 et 8, un précurseur de l'allèle Ad existait dans la population de virus Wt. Ces précurseurs représentaient en moyenne 9,5 % (écart-type de  $\pm 1,1$  %, EC) de la population Wt pour le segment 2, 28 % (EC=4,8 %) pour le segment 7 et 17 % (EC=2,8 %) pour le segment 8. Ces fréquences de gènes de Ad relativement homogènes chez Wt suggèrent que ces allèles de Ad pourraient provenir des mêmes molécules d'ARN. Cette suggestion est confirmée par l'observation des séquences obtenues par pyroséquençage 454, en effet les mutations minoritaires de Wt qui deviennent majoritaires chez Ad étaient généralement présentes sur



**Figure 15** Séquences du gène PB1 obtenues par pyroséquençage 454 pour les virus Wt et Ad. Région de 301 pb de la position 1200 à 1500. L’affichage consiste en l’alignement horizontal de trois pistes. La première piste nommée « Séquence » correspond à la séquence consensus du virus Wt. La deuxième piste nommée « A.fastq.sort.bam Coverage » est divisée en deux parties, la partie du haut correspond à la profondeur de lecture de Wt et celle du bas correspond aux séquences nucléotidiques. La troisième piste nommée « Ap.fastq.sort.bam Coverage » est l’homologue de Wt mais pour le virus Ad. Les quantités de séquences ne sont pas identiques pour des raisons d’affichage. Cette figure représente l’émergence des séquences minoritaires de Wt chez Ad et permet d’observer que les polymorphismes apparaissent sur les mêmes séquences.

les mêmes séquences nucléotidiques donc provenaient de la même molécule d'ARNv, l'exemple du segment 2 est visible sur la figure 15.

Cependant, les différences de fréquence entre les segments suggèrent qu'ils aient entrepris des réassortiments indépendamment les uns des autres, les brins d'ARNv minoritaires n'étaient pas systématiquement assemblés dans les mêmes virions. Il en est de même pour les segments 3, 4 et 5.

Les SNP présents sur les segments 2,7 et 8 du virus Wt et qui ont été perdus au fil des passages (qui sont donc absents du virus Ad) étaient en règle générale localisés sur des séquences nucléotidiques différentes de celles contenant les allèles minoritaires devenus majoritaire chez la population Ad. Cependant il y a des exceptions car certaines séquences arboraient seulement certains polymorphismes caractéristiques du virus Ad. Alors que les fréquences d'allèles Ad au sein de la population Wt étaient homogènes le long du segment PB1, dans le cas des segments M et NS une certaine hétérogénéité est observée.

Le séquençage à très haut débit a démontré que la plupart des segments du virus Wt qui étaient absents de la séquence consensus de Ad présentaient une fréquence moyenne inférieure à 0,1 % dans la population virale Ad dans laquelle les segments 2, 7 et 8 étaient très purs. Le segment 2 ne présentait aucune fréquence d'allèle atteignant les 10 %, 15 allèles minoritaires étaient présents à moins de 1%. Au niveau du segment 7, un allèle minoritaire dépassait les 10 % (21 %) et huit allèles minoritaires étaient présents à une fréquence de 1 % ou plus. Dans le segment 8, un allèle dépassait les 10 % (18 %) et cinq allèles minoritaires étaient présents à une fréquence de 1 % ou plus.

### **2.3.1 Disparition d'allèles minoritaires de la population virale Wt au fil de l'évolution**

Le segment 3 du virus Wt présentait 66 positions pour lesquelles des allèles minoritaires avaient une fréquence comprise entre 30 et 50 %. Comme pour les segments 2, 7 et 8 ces polymorphismes se situaient généralement sur les mêmes séquences.

Cependant, contrairement à ce qui est arrivé pour les segments 2, 7 et 8, ces populations minoritaires plutôt bien établies n'ont pas été sélectionnées et ont disparu de la population Ad.

La profondeur de lecture moyenne de l'allèle H1 du virus Wt était de 796x alors que celle de l'allèle H11 était de 682x. Les proportions approximatives étaient alors de 54 % pour H1 et de 46 % pour H11 dans cette population virale. Toutefois, l'allèle H11 a totalement disparu du stock Ad. Cette suppression radicale suggère que ce segment a été éliminé lors de la sélection exercée par le système cellulaire porcin.

L'analyse du segment 5 révèle qu'avec un seuil de détection de polymorphismes situé à 10 %, 97 positions ont été mutées dans la population virale Wt à une fréquence de 14 %  $\pm$  1,5 % EC. Cette faible répartition suggère que ces allèles minoritaires auraient été présents sur les mêmes molécules d'ARN et représentaient une sous-population minoritaire comme pour les segments 2, 3, 4, 7 et 8. Après l'observation des séquences nucléotidiques c'est en effet le même phénomène que pour les segments 3 et 4 qui est survenu, ces polymorphismes ont disparu du stock viral Ad.

## **2.4 Identification de mutations ponctuelles**

### **2.4.1 Mutations ponctuelles des séquences consensus**

Segments 3 et 4 : les SNP caractéristiques de la population virale Ad existaient déjà dans le stock Wt à une fréquence de 10 %. Il ne s'agit donc pas de mutations ponctuelles mais de polymorphismes qui ont émergé lors du processus de sélection et sont devenus majoritaires chez Ad (les allèles qui étaient majoritaires chez Wt sont descendus à 1 % chez Ad). Néanmoins, les deux SNP caractéristiques du segment 4 étaient apparemment absents de la population Wt. Ces SNP seraient apparus *de novo* ou bien étaient présent chez Wt à une fréquence inférieure au seuil de détection.

Segments 2, 7 et 8 : aucune mutation apparue *de novo* n'a été détectée au niveau du segment 2. Un SNP minoritaire du segment 7 du virus Ad présent à une fréquence de 21% semblerait provenir d'une mutation *de novo* puisqu'il n'a pas été détecté dans la population virale Wt. De la même façon, le gène NS du virus Ad présentait un SNP à une fréquence de 18% ; absent du virus Wt, ce SNP est donc supposé être apparu *de novo*.

L'échec d'émergence de sous-populations, même après dix passages successifs, suggère que leurs mutations spontanées ne leur conféraient pas de fort avantage adaptatif ou bien qu'elles sont apparues tardivement au fil du protocole de sélection.

#### 2.4.2 Évolution des SNP des séquences consensus de Wt et Ad

Segment 1 : avec un seuil de détection de polymorphismes fixé à 10 %, un seul SNP apparaissait au niveau de la séquence consensus de Wt avec une fréquence de 14 %. Ce SNP était descendu à 3% chez Ad. Ad présentait un seul allèle minoritaire à plus de 10 % (19 %) qui aurait pu être le résultat d'une mutation spontanée puisqu'il n'a pas été détecté chez Wt. Au seuil de détection des polymorphismes de 1 %, 13 positions polymorphes apparaissaient chez Wt et 21 chez Ad.

Segment 5 : au seuil de détection de 1 %, 107 positions polymorphes ont été observées chez Wt tandis que seulement 19 sites mutés subsistaient chez Ad.

Segment 6 : au seuil de 10 %, deux positions polymorphes étaient observables chez Wt. Ces polymorphismes étaient trop distants sur le génome pour déterminer s'ils appartenaient aux mêmes molécules. Ces SNP n'ont pas été détectés chez Ad. Au seuil de 1 % il y avait deux positions polymorphiques chez Wt et, encore, aucune association ne peut être établie entre ces sites mutés.

#### 2.5 Identification de potentiels virus « donneurs » de segments 2, 3, 4, 5, 7 et 8

Comme vu précédemment, les segments 2, 3, 4, 5, 7 et 8 du virus Wt présentaient des profils de sous-populations relativement établis. Il se peut que ces gènes aient été hérités d'un ou plusieurs virus présents dans l'échantillon de départ. De sorte à identifier ces potentiels virus donneurs, les séquences des segments 2, 7 et 8 ont été comparées à la base de données BLAST (Altschul et al., 1990).

Segment 2 : les résultats de BLAST les plus probants ont été A/mallard/Netherlands/15/1999 (H11N9) et A/northern shoveler/Netherlands/18/1999 (H11N9) qui partageaient le même segment 2. Seule une position différait entre ces segments et le segment 2 de Ad.

Segment 7 : les isolats A/mallard/Netherlands/15/1999, A/northern shoveler/Netherlands/18/1999 et Ad présentaient 100 % d'identité au niveau du segment 7.

Segment 8 : comme pour le segment 7, une identité de 100 % a été retrouvée entre les trois isolats.

Dans le cas des segments 3, 4 et 5 il n'était pas possible de se fier aux gènes de Ad pour déterminer l'origine des segments minoritaires de Wt puisqu'ils ont tous disparu de Ad. Une autre stratégie a donc été adoptée.

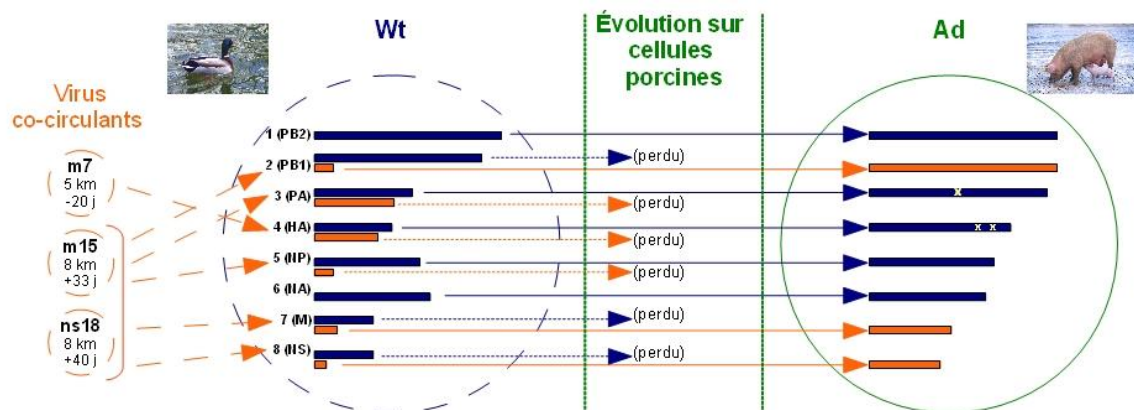
Segment 4 : la séquence de H11 de Wt a été comparée à la base de données BLAST. Le meilleur résultat fut A/mallard/Netherlands/7/1999 (H11N9), à la différence d'une position. Les autres résultats les plus proches ont été A/mallard/Netherlands/15/1999 (avec trois différences) et A/northern shoveler/Netherlands/18/1999 (avec trois différences).

Segment 3 : l'allèle minoritaire de Wt présente 100 % d'identité avec le segment 3 de A/mallard/Netherlands/15/1999 et A/northern shoveler/Netherlands/18/1999.

Segment 5 : les allèles minoritaires du segment 5 de Wt avaient deux différences avec A/mallard/Netherlands/15/1999 et A/northern shoveler/Netherlands/18/1999.

## 2.6 Proposition de scénario d'adaptation virale

Un scénario probable d'évolution du virus Wt vers le virus Ad basé sur les résultats de séquençage est proposé en figure 16.



**Figure 16** Scénario d'évolution génétique de l'échantillon parental Wt lors de son adaptation aux cellules NPTc. Le profil génétique du virus parental Wt présente des allèles majoritaires (en bleu) co-existants avec des allèles minoritaires glanés d'autres virus co-circulants (représentés en orange). Le réassortiment entre des segments minoritaires et majoritaires présents dans Wt est à l'origine du lot de gènes du virus Ad qui est adapté aux cellules porcines. Les mutations ponctuelles sont représentées par des x. Les virus potentiellement donneurs de gènes sont ns18 A/northern shoveler/Netherlands/18/1999 (H11N9) prélevé 40 jours après Wt à 8 kilomètres de distance; m7 A/mallard/Netherlands/7/99 (H11N9) prélevé 20 jours avant Wt à 5 kilomètres de distance et m15 A/mallard/Netherlands/15/1999 (H11N9) prélevé 33 jours après Wt à 8 kilomètres de distance (modifié d'après Bourret, 2013).



La population virale Wt était un mélange de populations minoritaires de segments 2, 3, 4, 5, 7 et 8 probablement hérités de souches de virus influenza co-circulantes dans des proportions décrites précédemment. Les allèles minoritaires des segments 3, 4 (H11) et 5 du virus Wt ont été perdus lors de l'adaptation de la population virale aux cellules NPTr avec un maintien des segments majoritaires uniquement (avec des mutations ponctuelles sur les segments 3 et 4). Inversement, ce sont les allèles minoritaires des segments 2, 7 et 8 qui ont été sélectionnés et ont remplacé les versions majoritaires chez Ad. Les segments 1 et 6 ne présentaient pas de sous-populations bien définies chez Wt ni d'évolution au niveau des séquences consensus lors de leur adaptation aux cellules NPTr.

### **3 Discussion**

#### **3.1 Séquençage total des virus parental et adapté**

Les génomes complets du virus parental Wt et du virus adapté Ad ont été analysés par pyroséquençage avec une profondeur de lecture comprise entre quelques centaines et quelques milliers de lectures par position sur le génome. Cette couverture du génome n'était cependant pas totalement homogène car elle était plus élevée aux extrémités des brins d'ADNc. La longueur moyenne des séquences nucléotidiques obtenues était d'environ 400 pb, ces séquences ont été obtenues à partir de fragments d'ADNc liés à des micro-billes. Le creux de profondeur de lecture observable au milieu des séquences peut être dû à une fragmentation inefficace des brins d'ADNc. En effet, si le brin d'ADNc est d'une longueur supérieure au double de la capacité de séquençage 454, seules les extrémités des brins seront séquencées (Bainbridge et al., 2006). Cependant, cette suggestion n'est pas applicable au segment 5 (NP) puisque l'extrémité 3' est davantage représentée que l'extrémité 5'.

#### **3.2 Analyse de sous-populations**

Une approche par pyroséquençage a permis d'identifier des sous-populations virales au sein de l'échantillon Wt. Cette méthode présente de nombreux avantages en comparaison avec le séquençage direct par la technique de Sanger. Dans un premier temps, l'analyse de séquences a permis d'identifier des polymorphismes minoritaires de façon plus catégorique que par l'observation de pics secondaires apparaissant sur les chromatogrammes de séquençage Sanger. De plus, un lien entre les SNP a pu être établi lorsqu'ils étaient situés sur

les mêmes séquences d'ADN, d'où la possibilité d'affirmer la présence de sous-populations minoritaires. Enfin, la fréquence des SNP a permis d'estimer le ratio relatif entre les populations minoritaires et les populations majoritaires dans l'échantillon de virus.

Les estimations quantitatives de fréquences de sous-populations réalisées peuvent être biaisées pour plusieurs raisons. En premier lieu, s'il est nécessaire de cultiver l'isolat viral sur œufs de sorte à l'amplifier, les œufs embryonnés de poule sont considérés comme un système permissif pour la culture de la plupart des VIA, l'évolution des proportions de variants pendant la culture est donc inévitable. Ensuite, le processus de traitement des échantillons composé d'extraction d'ARN, de rétro-transcription et de PCR est également un facteur d'erreurs et de biais (*cf.* §3.3).

### **3.3 Analyse de SNP minoritaires**

En plus des sous-populations virales minoritaires, le pyroséquençage a permis la détection de mutations ponctuelles présentes à très bas bruit. Ces dernières peuvent cependant être des artefacts dus aux étapes d'amplification et de séquençage. L'étape de rétro-transcription a été conduite avec une enzyme dérivée d'une transcriptase inverse du virus de la leucémie murine (revertaid h minus reverse transcriptase, Fermentas) qui introduirait une erreur toutes les 37000 bases (Ji and Loeb, 1992), ce qui représente statistiquement 0,37 erreur par génome influenza. L'étape ultérieure de PCR a été réalisée avec l'enzyme phusion (Finnzymes), cette enzyme génèrerait  $4,2 \times 10^{-7}$  erreur par base et par cycle (Li et al., 2006). De plus, la réaction de pyroséquençage est réputée pour introduire par erreur des insertions/délétions (indel) au sein d'homopolymères (succession de plusieurs nucléotides identiques), cela représente la principale source d'erreurs du pyroséquençage 454 (Kunin et al., 2010). En se basant sur un taux d'erreur estimé à 1.07 % (Gilles et al., 2011), lorsque la profondeur de lecture est de 2000 x environ, 20 lectures devraient présenter un polymorphisme. Or, il est fréquent d'observer des positions ne présentant aucun polymorphisme. Cela suggère que les erreurs de pyroséquençage ne sont pas réparties de façon homogène, la détermination d'un seuil discriminant les artefacts des SNP ayant un sens biologique est alors complexe. L'interprétation de séquences uniques devient délicate puisqu'elles devraient être écartées de l'analyse.

D'autres approches, tel le séquençage Sanger de banques de clones bactériens, permettent également d'aborder la variation génétique de populations de virus. Cependant, une telle approche, lorsqu'appliquée au séquençage de génomes totaux, deviendrait bien trop

lourde. Le pyroséquençage 454 a donc permis une caractérisation fine de sous-populations de VIA au sein des échantillons Wt et Ad ainsi que l'établissement d'un scénario d'évolution virale expérimentale.

### **3.4 Infections mixtes chez les oiseaux sauvages**

L'objectif de cette étude était de caractériser l'évolution génétique d'un VIA de canard lors de sa culture sur cellules porcines. Le virus sélectionné a initialement été identifié comme étant A/mallard/Netherlands/10/1999 (H1N8) mais le séquençage réalisé a permis de constater que le stock viral était majoritairement constitué de H1N1 sans N8 et avec une considérable proportion de H11.

Les recherches en bases de données (BLAST) ont suggéré que le virus Wt contenait des segments minoritaires probablement issus des isolats A/northern shoveler/Netherlands/18/1999 (H11N9), A/mallard/Netherlands/15/1999 (H11N9) et A/mallard/Netherlands/7/99 (H11N9). Ces trois virus ont été isolés à partir d'espèces phylogénétiquement proches ou bien à partir de la même espèce d'oiseaux que A/mallard/Netherlands/10/99 (H1N8) dans un secteur géographique restreint et dans un pas de temps de quelques semaines. Il se peut donc que des canards aient été co-infectés par deux voire davantage de ces virus compte tenu de l'hétérogénéité observée dans l'échantillon.

De tels profils d'infection sont classiquement observés dans l'étude des virus influenza chez les oiseaux sauvages. Une étude de 167 échantillons de virus influenza d'espèces ansériformes provenant de diverses origines géographiques aux États-Unis a démontré que 44 de ces isolats (soit 26 %) présentaient des profils d'infections mixtes sur la base de l'analyse des segments 4 et 6 et sans faire usage de séquençage à très haut débit (Dugan et al., 2008).

Ce type d'infections mixtes peut intégrer le concept de la théorie des quasi-espèces, c'est-à-dire qu'un ensemble de souches virales toutes différentes mais tout de même semblables d'un point de vue génétique pourraient agir comme un ensemble face à la pression de sélection qui leur est imposée (Domingo and Holland, 1997). En plus des SNP, les quasi-espèces peuvent arborer des segments d'ARNv nettement différents formant des sous-populations distinctes. Des versions de segments d'ARNv peuvent être maintenues à l'état de sous-populations dominées par des sous-populations majoritaires, elles peuvent alors demeurer indétectables tant qu'elles ne sont pas soumises à un goulet d'étranglement, une

pression de sélection positive ou alors une analyse par séquençage à très haut débit. Si l'on devait ne pas tenir compte de la théorie des quasi-espèces (Holmes, 2010), la mixité des infections jouerait tout de même son rôle dans l'écologie des virus influenza. Les co-infections suivies de réassortiments seraient une stratégie d'évolution de ces virus en milieu naturel, amplifiant leur pouvoir d'adaptation à de nouvelles niches écologiques.

## **Matériels et méthodes**

# 1 Première partie expérimentale

## 1.1 Observations cliniques et prélèvements biologiques

L'élevage de volailles concerné dans cette étude, situé en Pays de la Loire, était composé de deux bâtiments qui, lors de l'épisode infectieux en mars 2009, abritaient 8 700 canards de barbarie et 8 000 dindes de chair. Les deux bâtiments étaient séparés, hermétiques aux oiseaux de l'extérieur et traités indépendamment par le même personnel. Lors du premier jour de l'épisode clinique, des prélèvements biologiques et des autopsies ont été réalisés sur 10 dindes et du lisier de canards a été prélevé dans le bâtiment réservé aux canards. Une seconde autopsie a été effectuée sur 10 dindes huit jours plus tard. Des écouvillons trachéaux et des prélèvements d'organes ont été opérés sur chacun des animaux étudiés puis congelés à  $-80^{\circ}\text{C}$  pour des analyses ultérieures. Les procédures de routine ont été appliquées à l'analyse histologique des trachées et des poumons de dindes.

## 1.2 Extraction d'ARN, rétrotranscription et PCR

Chacun des 20 écouvillons réalisés sur les trachées de dindes a été repris dans 500  $\mu\text{l}$  de tampon phosphate salin (PBS) puis vortexé durant 15 secondes. L'extraction des ARN viraux a été effectuée à partir de 150  $\mu\text{l}$  de PBS en utilisant le kit d'extraction d'ARNv RNA virus kit (fourni par Macherey-Nagel, Düren, Allemagne) en respectant les consignes du fournisseur. Le volume d'élution des ARNv a été fixé à 50  $\mu\text{l}$  d'eau traitée sans ribonucléases.

Chaque segment d'ARNv a été amplifié en deux étapes par rétrotranscription puis PCR (RT-PCR).

La réaction de rétrotranscription a consisté en une incubation de 20  $\mu\text{l}$  d'ARNv avec 40 pmol de l'amorce UN112 (Hoffmann et al., 2001) pendant cinq minutes à  $65^{\circ}\text{C}$ . Ensuite, 8  $\mu\text{l}$  de tampon concentré cinq fois (fourni par Fermentas, Glen Burnie, USA) ont été ajoutés ainsi que 40 unités d'inhibiteur de RNase ribolock (Fermentas), 40 nmol de dNTPS (fourni par Finnzymes, Espoo, Finlande) et 200 unités de transcriptase inverse revertaid h minus reverse transcriptase (Fermentas). Le volume total de la réaction a été complété à 40  $\mu\text{l}$  par de l'eau distillée. La réaction s'est déroulée à  $42^{\circ}\text{C}$  pendant 60 minutes puis achevée par chauffage à  $70^{\circ}\text{C}$  durant 10 minutes.

Les ADN complémentaires (ADNc) obtenus ont été amplifiés par PCR en utilisant l'ADN polymérase phusion high-fidelity (Finnzymes) et un jeu d'amorces spécifiques de chaque gène (Hoffmann et al., 2001).

Les réactions de PCR, au nombre de huit (une réaction correspond à un gène), étaient composées de 29,5 µl d'eau distillée, 10 µl de tampon concentré cinq fois, 10 nmol de dNTP (Finnzymes), 25 pmol de l'amorce spécifique, 4 µl d'ADNc et une unité d'ADN polymérase. Le programme d'amplification débutait par 30 secondes à 98°C suivies de 35 cycles respectant les étapes suivantes : 10 secondes à 98°C, 20 secondes à 61°C et 90 secondes à 72°C. Une dernière étape de 10 minutes à 72°C terminait la réaction.

### **1.3 Séquençage des produits d'amplification par la technique de Sanger**

Les produits d'amplification obtenus lors de la PCR ont été séparés par électrophorèse en les faisant migrer sur un gel d'agarose concentré à 1 % en tampon tris, borate, EDTA (TBE), sous un champ électrique de 80 volts pendant une heure. Les bandes d'intérêt ont été extraites du gel puis purifiées à l'aide du kit nucleospin extract II (Macherey-Nagel) en respectant les consignes du fournisseur. Les produits d'amplification ont été élués dans 50 µl d'eau distillée. Le séquençage par la technique de Sanger (réalisé par la société Millegen, Labège, France) a permis d'obtenir la séquence consensus du génome total de chacun des trois types de prélèvement : le lisier de canards et les écouvillonnages de trachées de dindes réalisés en début et en fin d'épisode infectieux.

### **1.4 Analyse des neuraminidases par PCR conventionnelle**

De sorte à pouvoir discriminer les différentes formes de NA par PCR conventionnelle, un couple d'amorces encadrant la zone de délétion a été dessiné : NAs 5'-TGGTAATTGGAATAGTCAGTTTGA-3' et NAr 5'-ATCCCCTTTGGAACCAATTC-3'. Les réactions de PCR ont été réalisées à l'aide du kit phusion high-fidelity (Finnzymes); elles étaient composées de 15,75 µl d'eau distillée, 2,5 µl de tampon concentré cinq fois, 5 nmol de dNTP (Finnzymes), 12,5 pmol de chaque amorce, 2,5 µl d'ADNc et 0,25 unité d'ADN polymérase. Le programme d'amplification débutait par 30 secondes à 98°C suivies de 35 cycles respectant les étapes suivantes : 10 secondes à 98°C, 20 secondes à 61°C et 12 secondes à 72°C. Une dernière étape de 10 minutes à 72°C terminait la réaction.

### 1.5 Analyse des neuraminidases par PCR quantitative en temps réel

Afin de quantifier les formes complètes et tronquées de la NA dans les échantillons de lisier de canards ou de trachées de dindes, une PCR quantitative en temps réel a été mise au point. D'après les séquences obtenues avec le séquençage par la technique de Sanger, un couple d'amorces ainsi que deux sondes pour distinguer les formes complètes de NA des formes tronquées ont été dessinés (tableau III). Les ARNv ont été rétrotranscrits comme décrit précédemment (*cf.* §1.2) puis la PCR quantitative en temps réel pour l'amplification et la quantification des ADNc obtenus a été réalisée à l'aide d'un thermocycleur de type 7000 sequence detection system (fourni par PE Applied biosystems). Le kit de PCR utilisé était platinum quantitative PCR supermix-UDG (fourni par Invitrogen SRL, Milan, Italie). Les réactions de PCR étaient composées de 5 pmol de chaque amorce et 50 pmol de sonde taqman. La programmation du thermocycleur débutait par deux étapes à 50°C et 95°C pendant deux minutes pour chacune des températures puis 40 cycles de 95°C pendant 15 secondes et 60°C pendant 30 secondes.

**Tableau III** Séquences nucléotidiques des amorces sens, antisens et des sondes taqman employées lors de la PCR quantitative en temps réel pour la détection et la quantification des formes complètes et tronquées de NA dans les prélèvements de lisier de canards ou de trachées de dindes. FAM et JOE sont les fluorophores accepteurs couplés aux sondes taqman et BHQ est un accepteur non fluorescent.

	Séquence nucléotidique (5' vers 3')
amorce sens NA-F	AGACAGGGAATCAGCACCAG
amorce antisens delta-R	TTGGGCAAAGAGATGCATTA
sonde « NA complète »	FAM-TGAAAACAACACTTGGGTAAATCAGACC-BHQ
sonde « NA tronquée »	JOE-AAAAGCATCATTCTTACTGAGCAAGCTGT-BHQ



**Figure 17** A) Hybridation de la sonde « NA complète » au niveau de la séquence nucléotidique susceptible à la délétion. B) Hybridation de la sonde « NA tronquée » après troncation de la NA.



La sonde « NA complète » est complémentaire de la séquence nucléotidique absente des formes de NA tronquée (figure 17-A). Quant à la sonde « NA tronquée », elle s'hybride sur les zones flaquant la délétion (figure 17-B).

## 1.6 Pyroséquençage des produits d'amplification

Les produits d'amplification obtenus lors de la PCR ont été quantifiés en utilisant le kit quant-iT picogreen (Molecular probes, Eugene, USA). Les librairies d'ADN ont été préparées après une « nébulisation » par l'usage du kit GS DNA rapid library preparation (fourni par Roche, Basel, Suisse) selon les instructions du fournisseur. Trois librairies d'ADN ont été préparées soit une librairie par type de prélèvement, chacune contenant 500 ng d'ADN ; les huit segments génétiques ont été répartis au sein des librairies de façon équimolaire. Après une étape d'ajout de séquences d'identification aux extrémités des brins d'ADN, les librairies ont été quantifiées en utilisant le kit slingshot (fourni par Fluidigm, San Fransisco, USA) en respectant les consignes du fournisseur. Le pyroséquençage, exécuté selon la chimie GS FLX titanium de 454/Roche, a été effectué avec les kits de PCR et de séquençage GS titanium SV emulsion (fournis par 454/Roche) d'après les consignes du fournisseur. Les trois librairies d'ADN ont été séquencées sur une région de plaque représentant  $1/8^{\text{ème}}$  d'une plaque picotitrée.

## 1.7 Analyse de données par bioinformatique

Les séquences issues du pyroséquençage 454 ont d'abord été attribuées à leur librairie d'origine en employant l'outil *SFF file* (fourni par 454/Roche). Les lectures d'ADN ont ensuite été « nettoyées » avec le programme *pyrocleaner* (Mariette et al., 2011) en prenant en compte différents critères comme la longueur des séquences, par exemple. Les lectures d'ADN d'une longueur différant de deux écarts type par rapport à la longueur moyenne des séquences étaient éliminées. Les lectures d'ADN ont également été sélectionnées d'après des critères de qualité c'est-à-dire (i) en fonction de leur complexité, (ii) selon l'indice de qualité phred (Ewing et al., 1998) qui devait être supérieur à 20, ou encore (iii) suivant le taux de bases indéterminées qui devait être inférieur à quatre pourcent. Ensuite, pour rechercher les polymorphismes nucléotidiques répartis tout au long des séquences d'ADN, les lectures subsistantes ont été assemblées par rapport à un génome de référence avec le programme *brwa* (Li and Durbin, 2009) puis une séquence consensus a été créée en utilisant l'ensemble de programmes *SAMtools* (Li et al., 2009). Les séquences nucléotidiques ont de nouveau été alignées par rapport à la séquence consensus précédemment obtenue. Pour la détection de

délétions de taille supérieure à quelques nucléotides, c'est l'algorithme *BLAT* (Kent, 2002) qui a été appliqué aux séquences. Les données obtenues ont été observées avec le logiciel de visualisation *IGV* version 2.2.4 (Robinson et al., 2011).

### 1.7.1 Numéros d'accès des séquences nucléotidiques

Les séquences nucléotidiques complètes des virus A/muscovy duck/France/09010/2009, A/turkey/France/09010-1/2009 et A/turkey/France/09010-2/2009 (H6N1) ont été déposées sur la base de données genbank sous les numéros d'accès [JN860169](#) à [JN860193](#).

## 2 Deuxième partie expérimentale

### 2.1 Culture de VIA sur cellules porcines

Le virus A/mallard/Netherlands/10-Nmkt/1999 est un stock de virus influenza aviaire faiblement pathogène (nommé « Wt » tout au long de ce document) qui a été soumis à trois séries d'amplification en œufs embryonnés de poule. Ce stock de virus a ensuite subi des séries d'amplification sur une lignée cellulaire provenant de trachée de porcelet (NPTr) (Ferrari et al., 2003). Les passages sur cellules ont été réalisés à 37°C sous 5 % de CO<sub>2</sub> dans des flasques de 25 cm<sup>2</sup> dans lesquelles 2x10<sup>6</sup> cellules étaient en culture en milieu minimum essentiel de Eagle (EMEM) contenant de la pénicilline à 100 U/mL, de la streptomycine à 50 µg/mL, de l'amphotéricine à 2,6 µg/mL, 10 mM de L-glutamine, des aa non essentiels à 1 % et 10 % (volume à volume) de sérum de veau foetal (SVF). Vingt-quatre heures plus tard, les cellules ont été lavées en PBS puis infectées avec 1 mL d'inoculum contenant 2,4x10<sup>4</sup> copies de génome viral diluées en EMEM contenant 0,3 % (poids à volume) d'albumine de sérum bovin (en remplacement de SVF) et 0,25 µg/mL de trypsine (Worthington).

Cet inoculum était équivalent à 50 unités formant plaque de lyse (UFP) sur cellules de rein de chien de Madin-Darby (MDCK). Une heure après l'inoculation, 4 mL du milieu ayant servi à diluer l'inoculum sont ajoutés à la culture. Quarante-huit heures plus tard, de nouvelles cellules sont infectées avec le surnageant de culture virale obtenu dilué au 1/25 000<sup>ème</sup> (ce qui correspond, après titrage par PCR, à 3.8x10<sup>5</sup> à 5x10<sup>6</sup> copies de génome viral). Cette procédure est répétée dix fois. Le stock viral final obtenu après les dix passages est nommé « Ad ».

## **2.2 Séquençage par la technique de Sanger**

### **2.2.1 Amplification des génomes totaux**

Les ARNV ont été extraits à partir des liquides allantoïques et des surnageants de culture cellulaire à l'aide du kit qiaamp viral mini kit (Qiagen) selon les recommandations du fournisseur. La transcription inverse des ARNV en ADNc a été réalisée avec la transcriptase inverse revertaid premium (Fermentas) en respectant les instructions du fournisseur ainsi que l'amorce UNI12 (Hoffmann et al., 2001) complémentaire des 12 nucléotides conservés au niveau de l'extrémité 3' des huit segments d'ARNV. De sorte à prendre compte la position n°4 variable de l'extrémité 3' des brins d'ARNV, la base de l'amorce correspondante était A ou G (soit R) ce qui donne l'amorce vRNART : 5'-AGCRAAAGCAGG-3'.

Les réactions de PCR ont été réalisées avec la polymérase platinum pfx (Invitrogen) dans un volume final de 50 µl contenant du tampon d'amplification pfx, des dNTP à 0,3 mM, 0,9 mM de MgSO<sub>4</sub>, 1,25 unité de polymérase, 300 nM de chaque amorce (Annexe 2) et 1 µL d'ADNc. Le programme d'amplification débutait par cinq minutes à 94°C suivies de 35 cycles respectant une étape de dénaturation initiale de 15 secondes à 94°C puis une hybridation pendant 30 secondes et une étape d'élongation à 68°C. Les températures d'hybridation et les temps d'élongation dépendaient du segment d'ARNV et sont détaillées en (Annexe 2). Un programme de PCR dite « touchdown » a été employé pour l'amplification des gènes codant pour les polymérases, lors de ce programme la température d'hybridation est abaissée d'un degré Celsius à chaque cycle lors des 12 premiers cycles puis reste constante pour les cycles suivants.

Étant donné que le stock de virus Wt contient un mélange de sous-types de HA, les amorces utilisées ne ciblent que les extrémités 3' et 5' des segments d'ARNV qui sont conservées entre tous les sous-types de HA connus. De la sorte, il n'y a pas de sous-type de HA privilégié par rapport à un autre. Lorsque l'amplification spécifique de la HA de sous-type H1 était nécessaire, le couple d'amorces spécifique de la H1 était employé. Les différents couples d'amorces amplifiant la HA sont détaillés en Annexe 2.

### **2.2.2 Purification des produits de PCR**

Les produits de PCR ont été séparés par électrophorèse sur gel d'agarose concentré à 1 % en tampon TBE sous un champ électrique de 80 volts pendant une heure. Les bandes

d'intérêt ont ensuite été purifiées avec le kit qiaquick PCR purification kit (Qiagen). Les concentrations et quantités d'ADN amplifié ont été mesurées par l'usage d'un nanodrop (Thermo scientific).

### **2.2.3 Séquençage**

Un premier séquençage des produits de PCR par la technique de Sanger (réalisé par la société Millegen, Labège, France) a permis de déterminer les séquences consensus de chaque gène pour Wt et Ad. Les chromatogrammes peuvent également permettre de détecter d'éventuels allèles minoritaires lorsque des pics secondaires apparaissent.

Pour séquencer le sous-type d'hémagglutinine H1 ce sont des amorces spécifiques qui ont été utilisées (Annexe 3). Concernant le séquençage du sous-type H11, des clones bactériens ont été sélectionnés sur la base de leurs profils de restriction puis le séquençage de la HA a été réalisé à l'aide d'amorces internes à la séquence ainsi qu'un couple d'amorces ciblant les régions promotrices du vecteur flanquant la séquence de la HA.

## **2.3 Pyroséquençage des produits d'amplification**

Le matériel génétique ayant servi au pyroséquençage 454 des échantillons Wt et Ad a été respectivement obtenu à partir de liquide allantoïque ou de surnageant de culture cellulaire. La méthode utilisée respecte les étapes décrites en paragraphe 1.6 de la première partie expérimentale. Pour cette analyse ce sont deux librairies d'ADN qui ont été constituées, chacune contenant 500 ng d'ADN représentant les huit segments d'ARNv et correspondant aux échantillons Wt et Ad.

## **2.4 Analyse de données par bioinformatique**

Les analyses de séquences obtenues par pyroséquençage ont été réalisées en respectant les étapes détaillées en paragraphe 1.7 de la première partie expérimentale.

## **Conclusion générale**

D'un point de vue évolutif, les virus se caractérisent par un nombre élevé d'entités produites par génération, un cycle de réplication court et un fort taux de mutations introduites dans le génome lors des cycles réplicatifs. Ces caractéristiques sont à l'origine d'une diversité des populations faisant des virus un modèle de choix pour étudier l'évolution du vivant.

En virologie médicale, l'impact de l'évolution du vivant s'observe au niveau de la pathogénicité, du franchissement de la barrière d'espèce et de l'échappement aux antiviraux et aux vaccins.

Le niveau de diversité des populations virales dépend de la nature du virus étudié et des mécanismes mis en jeu. Les virus à ARN sont les plus variables avec un taux de substitution par site et par an de  $10^{-3}$  à  $10^{-5}$  soit 4 à 6  $\log_{10}$  au-dessus des taux de substitution observés chez les organismes pluricellulaires (Duffy et al., 2008).

Les dynamiques d'évolution des pathogènes et la diversité des populations virales étaient auparavant explorées par les techniques lourdes de clonage et de séquençage par la technique de Sanger. Les NGS permettent dorénavant d'étudier la diversité virale *in situ* en prenant garde toutefois de prendre en compte les erreurs d'amplification, de séquençage et d'assemblage des séquences.

La quantité d'informations générées par ces méthodes de nouvelle génération permet de caractériser les dynamiques d'évolution des virus chez un hôte (éventuellement à différents stades de l'infection) ou entre différents hôtes.

Les travaux de métagénomique virale ou d'exploration de la diversité génétique par les NGS sont confrontés à deux problématiques majeures : la première est celle de l'analyse bioinformatique et la seconde, celle de l'interprétation des données et la capacité à leur donner un sens biologique.

Les deux projets scientifiques détaillés ici ont confirmé l'atout que représente la plasticité génétique des VIA dans un contexte original infectieux en élevage et dans un contexte expérimental lors de l'adaptation d'une population virale à un système cellulaire. Dans les deux cas nous avons observé à quel point la diversité virale est une richesse puisqu'elle est à l'origine de l'adaptation des virus lors d'une modification de l'environnement. En effet, les deux études traitent de l'émergence de populations initialement présentes à bas bruit ; ces populations au demeurant dormantes ont finalement émergé pour devenir ultra-majoritaires en raison de leurs profils génétiques adaptés aux

nouveaux environnements. Les NGS ont apporté la masse d'information nécessaire à l'exploration de la diversité génétique, c'est ainsi qu'il fut possible de proposer des *scénarii* d'adaptation de populations virales.

Les nouvelles générations de séquençage sont en constante évolution et tendent à être supplantées par les « nouvelles nouvelles » générations de séquençage ou NNGS qui promettent toujours plus d'informations générées, une meilleure qualité des données, des séquences plus longues et des quantités de matrice de plus en plus faibles, tout comme les tarifs. Idéalement, le séquenceur de demain sera capable de séquencer le matériel génétique d'un seul virion ou d'une seule cellule avec un maximum de fiabilité, en quelques heures et pour quelques euros.

Les séquenceurs aujourd'hui accessibles fournissent tout de même des performances toujours meilleures et toujours plus intéressantes en tous points. De plus, la diversité des appareils disponibles sur le marché permet d'avoir une technique adaptée à chaque problématique, selon que l'on souhaite séquencer quelques génomes totaux de 13 kilobases ou déterminer le métagénome viral de 100 prélèvements.

Les études décrites dans ce document ont permis une familiarisation avec les NGS. Dès lors, des expérimentations deviennent réalisables afin d'explorer encore plus les mécanismes de la diversité génétique virale. La technologie de génétique inverse va nous permettre de créer des mélanges de populations virales dont les ratios et la composition génétique sont contrôlés. Ensuite, des infections expérimentales à partir de ces différents mélanges suivies d'une analyse par NGS permettra d'établir les conditions (environnementales et populationnelles) nécessaires à l'émergence de certains allèles viraux.

## **Références bibliographiques**



Alexander, D.J. (2000). A review of avian influenza in different bird species. *Vet. Microbiol.* 74, 3–13.

Alexander, D.J. (2007). Summary of avian influenza activity in Europe, Asia, Africa, and Australasia, 2002–2006. *Avian Dis.* 51, 161–166.

Alexander D. J., Manvell R. J., Banks J., Collins M. S., Parsons G., Cox B., Frost K. M., Speidel E. C., Ashman S., and Aldous E. W. (1999). Experimental assessment of the pathogenicity of the Newcastle disease viruses from outbreaks in Great Britain in 1997 for chickens and turkeys, and the protection afforded by vaccination. *Avian Pathology* 28, 501–511.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

Bainbridge, M.N., Warren, R.L., Hirst, M., Romanuik, T., Zeng, T., Go, A., Delaney, A., Griffith, M., Hickenbotham, M., Magrini, V., et al. (2006). Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. *BMC Genomics* 7, 246.

Beato, M.S., Mancin, M., Yang, J., Buratin, A., Ruffa, M., Maniero, S., Fusaro, A., Terregino, C., Wan, X.-F., and Capua, I. (2013). Antigenic characterization of recent H5N1 highly pathogenic avian influenza viruses circulating in Egyptian poultry. *Virology* 435, 350–356.

Becker, E.A., Burns, C.M., León, E.J., Rajabojan, S., Friedman, R., Friedrich, T.C., O'Connor, S.L., and Hughes, A.L. (2012). Experimental Analysis of Sources of Error in Evolutionary Studies Based on Roche/454 Pyrosequencing of Viral Genomes. *Genome Biol Evol* 4, 457–465.

Blancou, J. (2000). Histoire de la surveillance et du contrôle des maladies animales transmissibles (Paris).

Blomström, A.-L. (2011). Viral metagenomics as an emerging and powerful tool in veterinary medicine. *Veterinary Quarterly* 31, 107–114.

Bourret, V. (2012). Genetic adaptation of an avian influenza A virus to swine cells. University of Cambridge.

Bourret, V., Croville, G., Mariette, J., Klopp, C., Bouchez, O., Tiley, L., and Guérin, J.-L. (2013). Whole-genome, deep pyrosequencing analysis of a duck influenza A virus evolution in swine cells. *Infect. Genet. Evol.* 18C, 31–41.

Brown, I.H. (2010). Summary of avian influenza activity in Europe, Asia, and Africa, 2006–2009. *Avian Dis.* 54, 187–193.

Brown, J.D., Goekjian, G., Poulson, R., Valeika, S., and Stallknecht, D.E. (2009). Avian influenza virus in water: infectivity is dependent on pH, salinity and temperature. *Vet. Microbiol.* 136, 20–26.

Calder, L.J., Wasilewski, S., Berriman, J.A., and Rosenthal, P.B. (2010). Structural organization of a filamentous influenza A virus. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10685–10690.

Castrucci, M.R., Donatelli, I., Sidoli, L., Barigazzi, G., Kawaoka, Y., and Webster, R.G. (1993). Genetic reassortment between avian and human influenza A viruses in Italian pigs. *Virology* 193, 503–506.

Chen, H., Smith, G.J.D., Zhang, S.Y., Qin, K., Wang, J., Li, K.S., Webster, R.G., Peiris, J.S.M., and Guan, Y. (2005). Avian flu: H5N1 virus outbreak in migratory waterfowl. *Nature* 436, 191–192.

Cheung, T.K.W., and Poon, L.L.M. (2007). Biology of influenza a virus. *Ann. N. Y. Acad. Sci.* 1102, 1–25.

Comin, A., Klinkenberg, D., Marangon, S., Toffan, A., and Stegeman, A. (2011). Transmission dynamics of low pathogenicity avian influenza infections in Turkey flocks. *PLoS ONE* 6, e26935.

Croville, G., Soubies, S.M., Barbieri, J., Klopp, C., Mariette, J., Bouchez, O., Camus-Bouclainville, C., and Guérin, J.-L. (2012). Field monitoring of avian influenza viruses: whole-genome sequencing and tracking of neuraminidase evolution using 454 pyrosequencing. *J. Clin. Microbiol.* 50, 2881–2887.

Daniels, P., Wiyono, A., Sawitri, E., Poermadjaja, B., and Sims, L.D. (2012). H5N1 Highly Pathogenic Avian Influenza in Indonesia: Retrospective Considerations. *Curr. Top. Microbiol. Immunol.*

Daoust, P.-Y., Kibenge, F.S.B., Fouchier, R.A.M., van de Bildt, M.W.G., van Riel, D., and Kuiken, T. (2011). Replication of low pathogenic avian influenza virus in naturally infected Mallard ducks (*Anas platyrhynchos*) causes no morphologic lesions. *J. Wildl. Dis.* 47, 401–409.

Domingo, E., and Holland, J.J. (1997). RNA virus mutations and fitness for survival. *Annu. Rev. Microbiol.* 51, 151–178.

Downie, J.C., and Laver, W.G. (1973). Isolation of a type A influenza virus from an Australian pelagic bird. *Virology* 51, 259–269.

Duffy, S., Shackelton, L.A., and Holmes, E.C. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet.* 9, 267–276.

Dugan, V.G., Chen, R., Spiro, D.J., Sengamalay, N., Zaborsky, J., Ghedin, E., Nolting, J., Swayne, D.E., Runstadler, J.A., Happ, G.M., et al. (2008). The evolutionary genetics and emergence of avian influenza viruses in wild birds. *PLoS Pathog.* 4, e1000076.

Easterday, B.C., Trainer, D.O., Tůmová, B., and Pereira, H.G. (1968). Evidence of infection with influenza viruses in migratory waterfowl. *Nature* 219, 523–524.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8, 175–185.

FAO Task Force on Avian Influenza Broadened to Cover Wider Disease Risks in Wildlife.

Ferrari, M., Scalvini, A., Losio, M.N., Corradi, A., Soncini, M., Bignotti, E., Milanese, E., Ajmone-Marsan, P., Barlati, S., Bellotti, D., et al. (2003). Establishment and characterization of two new pig cell lines for use in virological diagnostic laboratories. *J. Virol. Methods* *107*, 205–212.

Flemming, G. (1871). *Animal plagues: their history, nature and prevention.* (London).

Forrest, H.L., and Webster, R.G. (2010). Perspectives on influenza evolution and the role of research. *Anim Health Res Rev* *11*, 3–18.

Fouchier, R.A., Bestebroer, T.M., Herfst, S., Van Der Kemp, L., Rimmelzwaan, G.F., and Osterhaus, A.D. (2000). Detection of influenza A viruses from different species by PCR amplification of conserved sequences in the matrix gene. *J. Clin. Microbiol.* *38*, 4096–4101.

Giannecchini, S., Campitelli, L., Calzoletti, L., De Marco, M.A., Azzi, A., and Donatelli, I. (2006). Comparison of in vitro replication features of H7N3 influenza viruses from wild ducks and turkeys: potential implications for interspecies transmission. *J. Gen. Virol.* *87*, 171–175.

Gilles, A., Megléc, E., Pech, N., Ferreira, S., Malausa, T., and Martin, J.-F. (2011). Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics* *12*, 245.

Guionie, O., Toquin, D., Sellal, E., Bouley, S., Zwingelstein, F., Allée, C., Bougeard, S., Lemièrre, S., and Etteradossi, N. (2007). Laboratory evaluation of a quantitative real-time reverse transcription PCR assay for the detection and identification of the four subgroups of avian metapneumovirus. *J. Virol. Methods* *139*, 150–158.

Hedskog, C., Mild, M., Jernberg, J., Sherwood, E., Bratt, G., Leitner, T., Lundeberg, J., Andersson, B., and Albert, J. (2010). Dynamics of HIV-1 quasispecies during antiviral treatment dissected using ultra-deep pyrosequencing. *PLoS ONE* *5*, e11345.

Hinshaw, V.S., Webster, R.G., and Turner, B. (1979). Water-borne transmission of influenza A viruses? *Intervirology* *11*, 66–68.

Hinshaw, V.S., Webster, R.G., and Turner, B. (1980). The perpetuation of orthomyxoviruses and paramyxoviruses in Canadian waterfowl. *Can. J. Microbiol.* *26*, 622–629.

Hinshaw, V.S., Air, G.M., Gibbs, A.J., Graves, L., Prescott, B., and Karunakaran, D. (1982). Antigenic and genetic characterization of a novel hemagglutinin subtype of influenza A viruses from gulls. *J. Virol.* *42*, 865–872.

Hinshaw, V.S., Wood, J.M., Webster, R.G., Deibel, R., and Turner, B. (1985). Circulation of influenza viruses and paramyxoviruses in waterfowl originating from two different areas of North America. *Bull. World Health Organ.* *63*, 711–719.

Hoffmann, E., Stech, J., Guan, Y., Webster, R.G., and Perez, D.R. (2001). Universal primer set for the full-length amplification of all influenza A viruses. *Arch. Virol.* *146*, 2275–2289.

Hoffmann, T.W., Munier, S., Larcher, T., Soubieux, D., Ledevin, M., Esnault, E., Tourdes, A., Croville, G., Guérin, J.-L., Quéré, P., et al. (2012). Length variations in the NA stalk of an H7N1 influenza virus have opposite effects on viral excretion in chickens and ducks. *J. Virol.* *86*, 584–588.

Holmes, E.C. (2010). The RNA virus quasispecies: fact or fiction? *J. Mol. Biol.* *400*, 271–273.

Horimoto, T., and Kawaoka, Y. (2001). Pandemic threat posed by avian influenza A viruses. *Clin. Microbiol. Rev.* *14*, 129–149.

Horimoto, T., and Kawaoka, Y. (2005). Influenza: lessons from past pandemics, warnings from current incidents. *Nat. Rev. Microbiol.* *3*, 591–600.

Huang, X., Liu, T., Muller, J., Levandowski, R.A., and Ye, Z. (2001). Effect of influenza virus matrix protein and viral RNA on ribonucleoprotein formation and nuclear export. *Virology* *287*, 405–416.

Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., and Welch, D.M. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.* *8*, R143.

Imai, M., and Kawaoka, Y. (2012). The role of receptor binding specificity in interspecies transmission of influenza viruses. *Curr Opin Virol* *2*, 160–167.

Imai, M., Watanabe, T., Hatta, M., Das, S.C., Ozawa, M., Shinya, K., Zhong, G., Hanson, A., Katsura, H., Watanabe, S., et al. (2012). Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature* *486*, 420–428.

Inoue, M., Barkham, T., Leo, Y.S., Chan, K.P., Chow, A., Wong, C.W., Tze Chuen Lee, R., Maurer-Stroh, S., Lin, R., and Lin, C. (2010). Emergence of oseltamivir-resistant pandemic (H1N1) 2009 virus within 48 hours. *Emerging Infect. Dis.* *16*, 1633–1636.

Ito, T., Okazaki, K., Kawaoka, Y., Takada, A., Webster, R.G., and Kida, H. (1995). Perpetuation of influenza A viruses in Alaskan waterfowl reservoirs. *Arch. Virol.* *140*, 1163–1172.

Ito, T., Goto, H., Yamamoto, E., Tanaka, H., Takeuchi, M., Kuwayama, M., Kawaoka, Y., and Otsuki, K. (2001). Generation of a highly pathogenic avian influenza A virus from an avirulent field isolate by passaging in chickens. *J. Virol.* *75*, 4439–4443.

Jagger, B.W., Wise, H.M., Kash, J.C., Walters, K.-A., Wills, N.M., Xiao, Y.-L., Dunfee, R.L., Schwartzman, L.M., Ozinsky, A., Bell, G.L., et al. (2012). An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. *Science* *337*, 199–204.

Ji, J.P., and Loeb, L.A. (1992). Fidelity of HIV-1 reverse transcriptase copying RNA in vitro. *Biochemistry* *31*, 954–958.

Johnson, N.P.A.S., and Mueller, J. (2002). Updating the accounts: global mortality of the 1918-1920 “Spanish” influenza pandemic. *Bull Hist Med* 76, 105–115.

Kawaoka, Y., Cox, N., and Haller, O. (2004). *Virus Taxonomy: VIIIth Report of the International Committee on Taxonomy of Viruses.* (Fauquet C, Mayo MA, Maniloff J, et al.).

Kayali, G., El-Shesheny, R., Kutkat, M.A., Kandeil, A.M., Mostafa, A., Ducatez, M.F., McKenzie, P.P., Govorkova, E.A., Nasraa, M.H., Webster, R.G., et al. (2011). Continuing threat of influenza (H5N1) virus circulation in Egypt. *Emerging Infect. Dis.* 17, 2306–2308.

Keawcharoen, J., Oraveerakul, K., Kuiken, T., Fouchier, R.A.M., Amonsin, A., Payungporn, S., Noppornpanth, S., Wattanodorn, S., Theambooniers, A., Tantilertcharoen, R., et al. (2004). Avian influenza H5N1 in tigers and leopards. *Emerging Infect. Dis.* 10, 2189–2191.

Kent, W.J. (2002). BLAT--the BLAST-like alignment tool. *Genome Res.* 12, 656–664.

Krauss, S., Walker, D., Pryor, S.P., Niles, L., Chenghong, L., Hinshaw, V.S., and Webster, R.G. (2004). Influenza A viruses of migrating wild aquatic birds in North America. *Vector Borne Zoonotic Dis.* 4, 177–189.

Kunin, V., Engelbrekton, A., Ochman, H., and Hugenholtz, P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ. Microbiol.* 12, 118–123.

Kuntz-Simon, G., Kaiser, C., and Madec, F. (2010). Swine influenza. In *Infectious and Parasitic Diseases of Livestock*, (Paris: Lavoisier), pp. 273–285.

Lauring, A.S., and Andino, R. (2010). Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog.* 6, e1001005.

Lebarbenchon, C., Yang, M., Keeler, S.P., Ramakrishnan, M.A., Brown, J.D., Stallknecht, D.E., and Sreevatsan, S. (2011). Viral replication, persistence in water and genetic characterization of two influenza A viruses isolated from surface lake water. *PLoS ONE* 6, e26566.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.

Li, J., Zu Dohna, H., Cardona, C.J., Miller, J., and Carpenter, T.E. (2011). Emergence and genetic variation of neuraminidase stalk deletions in avian influenza viruses. *PLoS ONE* 6, e14722.

Li, M., Diehl, F., Dressman, D., Vogelstein, B., and Kinzler, K.W. (2006). BEAMing up for detection and quantification of rare sequence variants. *Nat. Methods* 3, 95–97.

Liu, J., Xiao, H., Lei, F., Zhu, Q., Qin, K., Zhang, X.-W., Zhang, X.-L., Zhao, D., Wang, G., Feng, Y., et al. (2005). Highly pathogenic H5N1 influenza virus infection in migratory birds. *Science* 309, 1206.

Lupiani, B., and Reddy, S.M. (2009). The history of avian influenza. *Comparative Immunology, Microbiology and Infectious Diseases* 32, 311–323.

Makarova, N.V., Ozaki, H., Kida, H., Webster, R.G., and Perez, D.R. (2003). Replication and transmission of influenza viruses in Japanese quail. *Virology* 310, 8–15.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380.

Mariette, J., Noirot, C., and Klopp, C. (2011). Assessment of replicate bias in 454 pyrosequencing and a multi-purpose read-filtering tool. *BMC Res Notes* 4, 149.

Mathieu, C., Moreno, V., Retamal, P., Gonzalez, A., Rivera, A., Fuller, J., Jara, C., Lecocq, C., Rojas, M., Garcia, A., et al. (2010). Pandemic (H1N1) 2009 in breeding turkeys, Valparaiso, Chile. *Emerging Infect. Dis.* 16, 709–711.

Matlin, K.S., Reggio, H., Helenius, A., and Simons, K. (1981). Infectious entry pathway of influenza virus in a canine kidney cell line. *J. Cell Biol.* 91, 601–613.

Matrosovich, M., Zhou, N., Kawaoka, Y., and Webster, R. (1999). The surface glycoproteins of H5 influenza viruses isolated from humans, chickens, and wild aquatic birds have distinguishable properties. *J. Virol.* 73, 1146–1155.

Matrosovich, M.N., Krauss, S., and Webster, R.G. (2001). H9N2 influenza A viruses from poultry in Asia have human virus-like receptor specificity. *Virology* 281, 156–162.

Matsuoka, Y., Swayne, D.E., Thomas, C., Rameix-Welti, M.-A., Naffakh, N., Warnes, C., Altholtz, M., Donis, R., and Subbarao, K. (2009). Neuraminidase stalk length and additional glycosylation of the hemagglutinin influence the virulence of influenza H5N1 viruses for mice. *J. Virol.* 83, 4704–4708.

McHardy, A.C., and Adams, B. (2009). The role of genomics in tracking the evolution of influenza A virus. *PLoS Pathog.* 5, e1000566.

Mitnaul, L.J., Matrosovich, M.N., Castrucci, M.R., Tuzikov, A.B., Bovin, N.V., Kobasa, D., and Kawaoka, Y. (2000). Balanced hemagglutinin and neuraminidase activities are critical for efficient replication of influenza A virus. *J. Virol.* 74, 6015–6020.

Munier, S., Larcher, T., Cormier-Aline, F., Soubieux, D., Su, B., Guigand, L., Labrosse, B., Cherel, Y., Quéré, P., Marc, D., et al. (2010). A genetically engineered waterfowl influenza virus with a deletion in the stalk of the neuraminidase has increased virulence for chickens. *J. Virol.* 84, 940–952.

Munster, V.J., Baas, C., Lexmond, P., Waldenström, J., Wallensten, A., Fransson, T., Rimmelzwaan, G.F., Beyer, W.E.P., Schutten, M., Olsen, B., et al. (2007). Spatial, temporal, and species variation in prevalence of influenza A viruses in wild migratory birds. *PLoS Pathog.* *3*, e61.

Muramoto, Y., Noda, T., Kawakami, E., Akkina, R., and Kawaoka, Y. (2013). Identification of Novel Influenza A Virus Proteins Translated from PA mRNA. *J. Virol.* *87*, 2455–2462.

Nazir, J., Haumacher, R., Ike, A.C., and Marschang, R.E. (2011). Persistence of avian influenza viruses in lake sediment, duck feces, and duck meat. *Appl. Environ. Microbiol.* *77*, 4981–4985.

Nelli, R.K., Kuchipudi, S.V., White, G.A., Perez, B.B., Dunham, S.P., and Chang, K.-C. (2010). Comparative distribution of human and avian type sialic acid influenza receptors in the pig. *BMC Veterinary Research* *6*, 4.

Neumann, G., Noda, T., and Kawaoka, Y. (2009). Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature* *459*, 931–939.

Noda, T., Sugita, Y., Aoyama, K., Hirase, A., Kawakami, E., Miyazawa, A., Sagara, H., and Kawaoka, Y. (2012). Three-dimensional analysis of ribonucleoprotein complexes in influenza A virus. *Nat Commun* *3*, 639.

Olsen, B., Munster, V.J., Wallensten, A., Waldenström, J., Osterhaus, A.D.M.E., and Fouchier, R.A.M. (2006). Global patterns of influenza a virus in wild birds. *Science* *312*, 384–388.

Pantin-Jackwood, M.J., and Swayne, D.E. (2009). Pathogenesis and pathobiology of avian influenza virus infection in birds. *Rev. - Off. Int. Epizoot.* *28*, 113–136.

Pillai, S.P.S., Pantin-Jackwood, M., Yassine, H.M., Saif, Y.M., and Lee, C.W. (2010). The high susceptibility of turkeys to influenza viruses of different origins implies their importance as potential intermediate hosts. *Avian Dis.* *54*, 522–526.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* *29*, 24–26.

Rogers, G.N., and Paulson, J.C. (1983). Receptor determinants of human and animal influenza virus isolates: differences in receptor specificity of the H3 hemagglutinin based on species of origin. *Virology* *127*, 361–373.

Rossman, J.S., and Lamb, R.A. (2011). Influenza virus assembly and budding. *Virology* *411*, 229–236.

Rott, R. (1992). The pathogenic determinant of influenza virus. *Vet. Microbiol.* *33*, 303–310.

Rust, M.J., Lakadamyali, M., Zhang, F., and Zhuang, X. (2004). Assembly of endocytic machinery around individual influenza viruses during viral entry. *Nat. Struct. Mol. Biol.* *11*, 567–573.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* *74*, 5463.

Scholtissek, C., Bürger, H., Kistner, O., and Shortridge, K.F. (1985). The nucleoprotein as a possible major factor in determining host specificity of influenza H3N2 viruses. *Virology* *147*, 287–294.

Shinya, K., Ebina, M., Yamada, S., Ono, M., Kasai, N., and Kawaoka, Y. (2006). Avian flu: Influenza virus receptors in the human airway. *Nature* *440*, 435–436.

Sims, L.D., Ellis, T.M., Liu, K.K., Dyrting, K., Wong, H., Peiris, M., Guan, Y., and Shortridge, K.F. (2003). Avian influenza in Hong Kong 1997–2002. *Avian Dis.* *47*, 832–838.

Sims, L.D., Domenech, J., Benigno, C., Kahn, S., Kamata, A., Lubroth, J., Martin, V., and Roeder, P. (2005). Origin and evolution of highly pathogenic H5N1 avian influenza in Asia. *Vet. Rec.* *157*, 159–164.

Slemons, R.D., Johnson, D.C., Osborn, J.S., and Hayes, F. (1974). Type-A influenza viruses isolated from wild free-flying ducks in California. *Avian Dis.* *18*, 119–124.

Sorrell, E.M., Song, H., Pena, L., and Perez, D.R. (2010). A 27-amino-acid deletion in the neuraminidase stalk supports replication of an avian H2N2 influenza A virus in the respiratory tract of chickens. *J. Virol.* *84*, 11831–11840.

Stieneke-Gröber, A., Vey, M., Angliker, H., Shaw, E., Thomas, G., Roberts, C., Klenk, H.D., and Garten, W. (1992). Influenza virus hemagglutinin with multibasic cleavage site is activated by furin, a subtilisin-like endoprotease. *EMBO J.* *11*, 2407–2414.

Suarez, D.L. (2010). Avian influenza: our current understanding. *Anim Health Res Rev* *11*, 19–33.

Swayne, D.E., and Suarez, D.L. (2000). Highly pathogenic avian influenza. *Rev. - Off. Int. Epizoot.* *19*, 463–482.

Taubenberger, J.K., and Morens, D.M. (2006). 1918 Influenza: the mother of all pandemics. *Emerging Infect. Dis.* *12*, 15–22.

Taubenberger, J.K., and Morens, D.M. (2010). Influenza: the once and future pandemic. *Public Health Reports* *125*, 16.

Taubenberger, J.K., Hultin, J.V., and Morens, D.M. (2007). Discovery and characterization of the 1918 pandemic influenza virus in historical context. *Antivir. Ther. (Lond.)* *12*, 581–591.

Tong, S., Li, Y., Rivaller, P., Conrardy, C., Castillo, D.A.A., Chen, L.-M., Recuenco, S., Ellison, J.A., Davis, C.T., York, I.A., et al. (2012). A distinct lineage of influenza A virus from bats. *Proc. Natl. Acad. Sci. U.S.A.* *109*, 4269–4274.

Di Trani, L., Bedini, B., Cordioli, P., Muscillo, M., Vignolo, E., Moreno, A., and Tollis, M. (2004). Molecular characterization of low pathogenicity H7N3 avian influenza viruses isolated in Italy. *Avian Dis.* *48*, 376–383.



VanDalen, K.K., Franklin, A.B., Mooers, N.L., Sullivan, H.J., and Shriner, S.A. (2010). Shedding light on avian influenza H4N6 infection in mallards: modes of transmission and implications for surveillance. *PLoS ONE* 5, e12851.

Wahlgren, J. (2011). Influenza A viruses: an ecology review. *Infection Ecology & Epidemiology* 1.

Wallensten, A., Munster, V.J., Latorre-Margalef, N., Brytting, M., Elmberg, J., Fouchier, R.A.M., Fransson, T., Haemig, P.D., Karlsson, M., Lundkvist, Å., et al. (2007). Surveillance of Influenza Virus A in Migratory Waterfowl in Northern Europe. *Emerg Infect Dis* 13, 404–411.

Webster, R.G., Yakhno, M., Hinshaw, V.S., Bean, W.J., and Murti, K.G. (1978). Intestinal influenza: replication and characterization of influenza viruses in ducks. *Virology* 84, 268–278.

Webster, R.G., Bean, W.J., Gorman, O.T., Chambers, T.M., and Kawaoka, Y. (1992). Evolution and ecology of influenza A viruses. *Microbiol. Rev.* 56, 152–179.

Whittaker, G.R. (2001). Intracellular trafficking of influenza virus: clinical implications for molecular medicine. *Expert Reviews in Molecular Medicine* 3, 1–13.

Wise, H.M., Foeglein, A., Sun, J., Dalton, R.M., Patel, S., Howard, W., Anderson, E.C., Barclay, W.S., and Digard, P. (2009). A Complicated Message: Identification of a Novel PB1-Related Protein Translated from Influenza A Virus Segment 2 mRNA. *J. Virol.* 83, 8021–8031.

Wise, H.M., Hutchinson, E.C., Jagger, B.W., Stuart, A.D., Kang, Z.H., Robb, N., Schwartzman, L.M., Kash, J.C., Fodor, E., Firth, A.E., et al. (2012). Identification of a novel splice variant form of the influenza A virus M2 ion channel with an antigenically distinct ectodomain. *PLoS Pathog.* 8, e1002998.

Yamada, S., Suzuki, Y., Suzuki, T., Le, M.Q., Nidom, C.A., Sakai-Tagawa, Y., Muramoto, Y., Ito, M., Kiso, M., Horimoto, T., et al. (2006). Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors. *Nature* 444, 378–382.

Zamarin, D., Ortigoza, M.B., and Palese, P. (2006). Influenza A virus PB1-F2 protein contributes to viral pathogenesis in mice. *J. Virol.* 80, 7976–7983.

Zhou, H., Yu, Z., Hu, Y., Tu, J., Zou, W., Peng, Y., Zhu, J., Li, Y., Zhang, A., Yu, Z., et al. (2009). The special neuraminidase stalk-motif responsible for increased virulence and pathogenesis of H5N1 influenza A virus. *PLoS ONE* 4, e6277.

## **Annexes**

# Field Monitoring of Avian Influenza Viruses: Whole-Genome Sequencing and Tracking of Neuraminidase Evolution Using 454 Pyrosequencing

Guillaume Croville,<sup>a,b</sup> Sébastien Mathieu Soubies,<sup>a,b</sup> Johanna Barbieri,<sup>a,b</sup> Christophe Klopp,<sup>c</sup> Jérôme Mariette,<sup>c</sup> Olivier Bouchez,<sup>d,e</sup> Christelle Camus-Bouclainville,<sup>a,b</sup> and Jean-Luc Guérin<sup>a,b</sup>

Université de Toulouse, INP, ENVT, Toulouse, France<sup>a</sup>; INRA, UMR 1225, Toulouse, France<sup>b</sup>; Plateforme Bioinformatique Toulouse Midi-Pyrénées, UBIA, INRA, Auzeville Castanet-Tolosan, France<sup>c</sup>; INRA, UMR 444 Laboratoire de Génétique Cellulaire, INRA Auzeville, Castanet-Tolosan, France<sup>d</sup>; and GeT-PlaGe, Genotoul, INRA Auzeville, Castanet-Tolosan, France<sup>e</sup>

Adaptation of avian influenza viruses (AIVs) from waterfowl to domestic poultry with a deletion in the neuraminidase (NA) stalk has already been reported. The way the virus undergoes this evolution, however, is thus far unclear. We address this question using pyrosequencing of duck and turkey low-pathogenicity AIVs. Ducks and turkeys were sampled at the very beginning of an H6N1 outbreak, and turkeys were swabbed again 8 days later. NA stalk deletions were evidenced in turkeys by Sanger sequencing. To further investigate viral evolution, 454 pyrosequencing was performed: for each set of samples, up to 41,500 reads of ca. 400 bp were generated and aligned. Genetic polymorphisms between duck and turkey viruses were tracked on the whole genome. NA deletion was detected in less than 2% of reads in duck feces but in 100% of reads in turkey tracheal specimens collected at the same time. Further variations in length were observed in NA from turkeys 8 days later. Similarly, minority mutants emerged on the hemagglutinin (HA) gene, with substitutions mostly in the receptor binding site on the globular head. These critical changes suggest a strong evolutionary pressure in turkeys. The increasing performances of next-generation sequencing technologies should enable us to monitor the genomic diversity of avian influenza viruses and early emergence of potentially pathogenic variants within bird flocks. The present study, based on 454 pyrosequencing, suggests that NA deletion, an example of AIV adaptation from waterfowl to domestic poultry, occurs by selection rather than *de novo* emergence of viral mutants.

A critical biological feature of influenza viruses, shared with all RNA viruses, is their high genetic variability generated by poor proofreading activity. As a consequence, influenza viruses appear as clouds of mutants, referred to as quasispecies, which can hardly be assessed by a consensus sequence (15). Deep sequencing is therefore an alternative approach to dissecting this genetic complexity by detecting minority mutants in complex viral populations.

Low-pathogenicity avian influenza virus (LPAIV) infections in poultry are mostly nonclinical and therefore remain poorly investigated despite the role of precursors of highly pathogenic (HP) strains.

The presence of a polybasic cleavage site in the hemagglutinin (HA) sequence is known as the main support of pathogenicity in avian hosts. However, determinants involved in the adaptation to gallinaceous birds have also been identified. Examples of such determinants are represented by deletions in the stalk of the neuraminidase (NA), which occur during viral circulation in poultry. These deletions appear to contribute to viral adaptation to these species (13, 16) and to a change in viral tropism from the intestine to the respiratory tract of birds (23). Reverse genetic studies recently confirmed that NA stalk deletions are associated with an increased virulence in poultry (17, 20, 24). These deletion events have been recurrently observed in viral isolates infecting gallinaceous birds and are phylogenetically close to strains circulating in wild birds (2, 4). Nevertheless, parental and truncated viruses have not been tracked thus far in field conditions.

In this study, we investigate a case of H6N1 infection in Muscovy ducks, which resulted in in-farm contamination of turkeys, and monitored genetic changes on the 8 viral segments. A partic-

ular interest was focused on NA: using deep sequencing, we were able to detect ultraminority deleted NA alleles in ducks, which were found in 100% of NA sequences in turkeys, and then evolved further with cocirculation of 2 forms of NA deletions in the same birds. Altogether, these results illustrate the invaluable contribution of deep sequencing to the fine dissection of genetic species of influenza viruses.

## MATERIALS AND METHODS

**Clinical examination and sampling of birds.** The poultry farm, localized in Pays de la Loire, France, included 2 sheds, which housed, at the time of the infectious episode in March 2009, 8,700 Muscovy ducks and 8,000 meat turkey poults, respectively. The 2 sheds were totally closed, bird-proofed, and independently cared for by the same farmer. On the first day of a clinical episode, a sampling of 10 diseased turkeys was performed, and at the same time, several samples of fecal material in the duck house were collected and pooled. A second necropsy was performed on 10 birds 8 days after the first one. Tracheal swabs and organ samples were taken from all dead animals. Swabs were immediately stored at  $-80^{\circ}\text{C}$  until further pro-

Received 1 May 2012 Returned for modification 3 June 2012

Accepted 10 June 2012

Published ahead of print 20 June 2012

Address correspondence to Jean-Luc Guérin, [jl.guerin@envt.fr](mailto:jl.guerin@envt.fr).

G.C. and S.M.S. contributed equally to this article.

Supplemental material for this article may be found at <http://jcm.asm.org/>.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JCM.01142-12

**TABLE 1** Comparison of nucleotide and amino acid sequences of A/Muscovy duck/France/09010/09(H6N1) with A/turkey/France/09010-1/09(H6N1) and A/turkey/France/09010-2/09(H6N1) viruses

Segment	Protein	A/Turkey/France/09010-1/09(H6N1)			A/Turkey/France/09010-2/09(H6N1)		
		Nucleotide identity (%)	Amino acid identity (%)	Amino acid changes	Nucleotide identity (%)	Amino acid identity (%)	Amino acid changes
1	PB2	2,005/2,280 (87.9)	751/759 (98.9)	R61K, T147I, V292I, V295I V397A, V397A, I511V, T585P, D680G	2,005/2,280 (87.9)	751/759 (98.9)	R61K, T147I, V292I, V295I V397A, I511V, T585P, D680G
2	PB1	2,190/2,274 (96.3)	750/757 (99)	K52N, R54K, A56T, M195I, L550I, R584H, N694S	2190/2274 (96.3)	750/757 (99)	K52N, R54K, A56T, M195I, L550I, R584H, N694S
	PB1-F2	258/273 (94.5)	76/90 (84.4)	E4G, R21 M, G22E, G23S, Q26R, L30P, T46 M, Q48R, A49V, K53R, Q54R, K65R, Q79R, S82L	258/273 (94.5)	76/90 (84.4)	E4G, R21 M, G22E, G23S, Q26R, L30P, T46 M, Q48R, A49V, K53R, Q54R, K65R, Q79R, S82L
3	PA	2,059/2,151 (95.7)	714/716 (99.7)	A70V, E538G	2,059/2,151 (95.7)	714/716 (99.7)	A70V, E538G
4	HA	1,698/1,701 (99.8)	565/566 (99.8)	K118R	1,698/1,701 (99.8)	565/566 (99.8)	K118R
5	NP	1,497/1,497 (100)	498/498 (100)		1,497/1,497 (100)	498/498 (100)	
6	NA	1,350/1,350 (excluding deletion)	449/449 (100)	N.B.: amino acids 55-74 are absent in the turkey isolate (deleted region)	1,350/1,350 (excluding deletion)	449/449 (100)	N.B.: amino acids 55-74 are absent in the turkey isolate (deleted region)
7	M1	759/759 (100)	252/252 (100)		1,341/1,341 (excluding deletion)	446/446 (100)	N.B.: amino acids 43-66 are absent in the turkey isolate (deleted region)
	M2	294/294 (100)	97/97 (100)				
	NS1	693/693 (100)	230/230 (100)				
8	NEP	366/366 (100)	121/121 (100)				

cessing. Histological examination of the trachea and lung of turkeys was performed using routine procedures.

**RNA extraction and RT-PCR.** All the swabs received at the National Veterinary School of Toulouse were first vortexed in a 500- $\mu$ l PBS solution. Viral RNA was extracted from 150  $\mu$ l of sample by use of the RNA Virus kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's protocol. The RNA was eluted into 50  $\mu$ l RNase-free water. Each of the viral gene segments was amplified by 2-step reverse transcription-PCR (RT-PCR). For reverse transcription, 20  $\mu$ l of total RNA was mixed with 40 pmol of Uni12 primer (7) and incubated for 5 min at 65°C. Then, 8  $\mu$ l of 5 $\times$  reaction buffer (Fermentas, Glen Burnie, MD) was added to this mixture at the same time as 1  $\mu$ l of RiboLock RNase inhibitor (Fermentas), 4  $\mu$ l of 10 mmol/liter deoxynucleoside triphosphate (dNTP) solution (Finnzymes, Espoo, Finland), and 2  $\mu$ l of RevertAid reverse transcriptase (Fermentas). The reaction volume was completed to 40  $\mu$ l with distilled water. The RT reaction was performed at 42°C for 60 min and terminated by heating at 70°C for 10 min.

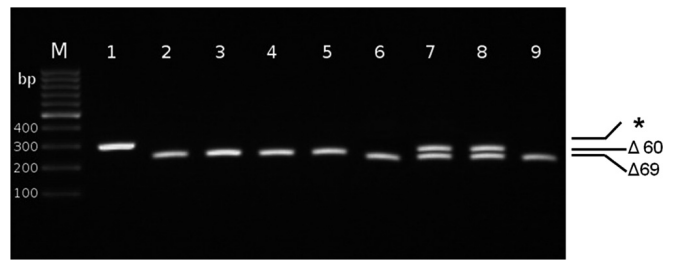
cDNAs were amplified by using Phusion High-Fidelity DNA polymerase (Finnzymes) and a set of gene-specific primers (7).

Eight separated PCR mixtures, corresponding to the 8 genes, were prepared and contained 29.5  $\mu$ l of distilled water, 10  $\mu$ l of 5 $\times$  buffer, 1  $\mu$ l of 10 mmol/liter dNTP solution (Finnzymes), 2.5  $\mu$ l of 10  $\mu$ mol/liter primers, 4  $\mu$ l of the cDNA, and 1 unit Phusion High-Fidelity DNA polymerase. The amplification program consisted of a first 30-s step at 98°C followed by 35 cycles with the following conditions: 98°C for 10 s, 61°C for 20 s, and 72°C for 1 min 30 s. The program ended with 1 step at 72°C for 10 min.

**Sanger sequencing of the amplification products.** The amplification products were separated by electrophoresis using 1% agarose gels. Observed bands were excised from the gel and purified using the Nucleospin Extract II kit (Macherey-Nagel) according to the manufacturer's protocol. These PCR products were eluted into 50  $\mu$ l of distilled water and subsequently used in Sanger sequencing and pyrosequencing. By Sanger sequencing, we determined consensus whole-genome sequences for each of the 3 types of sampling.

**Differential real-time PCR on neuraminidase alleles.** Quantitative real-time PCR was also used to characterize the neuraminidase allele(s) present in duck feces or turkey tracheas. A set of primers and probes was designed for full-length neuraminidase or truncated neuraminidase detection, according to the Sanger sequencing results. The forward primer was NA-F (5'-AGACAGGGAATCAGCACCAG-3'), and the reverse primer was Delta-R (5'-TTGGGCAAAGAGATGCATTA-3'). The TaqMan probe sequences targeting the complete or the truncated NA were 5'-FAM-TGAAAACAACACTTGGGTAAATCAGACC-BHQ-3' and 5'-JOE-AAAAGCATCATTCTTACTGAGCAAGCTGT-BHQ-3', respectively. Both were coupled with a Black Hole Quencher dye at the 3' end. Viral RNA was retrotranscribed as described before, and cDNA was amplified and detected in a 7000 Sequence Detection System (PE Applied Biosystems) by using the Platinum Quantitative PCR SuperMix-UDG (Invitrogen SRL, Milan, Italy). The 25- $\mu$ l PCR mixture contained 5 pmol of each primer and 50 pmol probe. After 2 thermal steps at 50°C for 2 min and 95°C for 2 min, 40 cycles of PCR at 95°C for 15 s and at 60°C for 30 s were performed.

**454 pyrosequencing of the amplification products.** The previously purified PCR products were quantified using a Quant-iT PicoGreen kit (Molecular Probes, Eugene, OR). Libraries were prepared after PCR product nebulization using the GS DNA Rapid Library Preparation kit according to the manufacturer's recommendations (Roche, Basel, Switzerland). Three pools (corresponding to the 3 different sets of samplings), each containing 500 ng DNA, were used to generate the libraries; the 8 segments were pooled in equimolar amounts. Each DNA pool was ligated to an adapter containing a multiplex identifier sequence (MID) for emulsion PCR and sequencing. Final libraries were quantified using the SlingShot kit, using the Fluidigm Digital Array for sample quantification, following the manufacturer's instructions. Pyrosequencing using the 454/Roche GS



**FIG 1** Typical agarose gel electrophoresis pattern of NA gene-specific RT-PCR products from duck and turkey clinical specimens, using internal primers (NA-F: 5'-TGGTAATTGGAATAGTCAGTTGA-3'; NA-R: 5'-ATCCCCTTGGAAACCAATTC-3'). Lane 1: duck feces; lanes 2 to 5: representative early turkey swabs, showing a 60-nucleotide deletion; lanes 6 to 9: representative late turkey swabs, showing various, mixed patterns of deletion. M: molecular mass marker; \*: full-length NA.  $\Delta$ 60 and  $\Delta$ 69: 60-nt and 69-nt deletion, respectively.

FLX Titanium chemistry was carried out with the GS Titanium SV emulsion PCR kit and sequencing kit according to the manufacturer's instructions. The 3 samples were sequenced on 1 region of an 8-region picotiter plate.

**Bioinformatic analysis.** Sequences were first demultiplexed using Roche's tool SFF file without allowing any error per MID. Reads were then filtered using pyrocleaner (14), considering different criteria such as length. Reads shorter or longer than the mean length of the reads  $\pm$  2 standard deviations were discarded. Reads were also filtered based on their complexity, which is computed using the compressed string length (library zip) on several subsequences generated using a sliding window approach. Moreover, the pyrocleaner script removes sequences for which all base pairs have a phred quality value under 20 or if the rate of under-terminated bases is higher than 4%, which has been proven to correlate with poor quality. Next, to carry out the single nucleotide polymorphism (SNP) calling, the remaining sequences were assembled on the reference to produce a consensus sequence for each sample. To do so, sequences were aligned against the reference genome using bwa (11), and the consensus was computed with the SAMtools (12) software package. Reads were then realigned against this consensus sequence before SNP calling using other modules of the same package. For large deletion discovery, the BLAT (10) aligner was used on the same reads for each sample. The resulting files were then displayed within the IGV (22) browser.

**Nucleotide sequence accession numbers.** The complete sequences of A/Muscovy duck/France/09010/09, A/turkey/France/09010-1/09, and A/turkey/France/09010-2/09 (H6N1) viruses have been deposited in GenBank under accession numbers JN860169 to JN860193.

## RESULTS

**Case history and virological findings.** Three days after the removal of ducks to the slaughterhouse, the turkey flock started to show prostration and respiratory signs, resulting in approximately 10% mortality. A sampling of 10 diseased turkeys was performed, and at the same time, several samples of fecal material in the duck house were collected and pooled. Necropsies were performed on the 10 birds, and all showed severe exudative tracheitis, airsaccu-litis, and pneumonia. A routine antibiotic treatment was performed to prevent secondary bacterial infections, but it did not stop mortality. A second necropsy was therefore performed on 10 other birds 8 days later. Tracheal swabs and organ samples were taken from all dead animals.

Real-time PCR assays were negative for *Mycoplasma gallisepticum* and *Mycoplasma synoviae*. Real-time RT-PCR analyses of the tracheal samples were negative for avian metapneumovirus (5) but positive for avian influenza virus (AIV) (3).

TABLE 2 Quantitative RT-PCR distinguishing between wild-type and truncated NA<sup>a</sup>

Specimen type	Probe sequence	Mean no. of copies ± SD	
		Wild-type NA	Truncated NA
Duck feces	TGAAAACAACACTTGG GTAAATCAGACC	787 ± 311	ND
Turkey swabs	AAAAGCATCATTCTTA CTGAGCAAGCTGT	ND	5,109 ± 1,585

<sup>a</sup> ND, not detected. 5' primer sequence AGACAGGGAATCAGACCAG; 3' primer sequence, TTGGGCAAAGAGATGCATTA.

Immunostaining for influenza virus A revealed infected cells in the tracheal epithelium and in the pulmonary parenchyma at the time of the first sampling (data not shown).

**Sequence analysis of duck and turkey isolates.** RT-PCR am-

plifications encompassing all the viral genomic segments' coding sequences were performed on both turkey tracheal swabs and duck feces using a set of gene-specific primers for influenza viruses (7). This procedure, followed by Sanger sequencing, led to the identification of the H6N1 subtype. Comparison of complete sequences of duck and turkey viruses (Table 1) revealed more than 99.8% nucleotide (nt) identity genome-wide, except for (i) one internal deletion of 60 nt in segment 6 resulting in a putative 20-aa deletion in the neuraminidase stalk in all of the first turkey isolates collected and (ii) a singular pattern of NA stalk deletions consisting of a mix of the 60-nt deleted allele and a new, second 69-nt deleted one in some turkeys swabbed after 8 days of infection, as confirmed by cloning and sequencing. Four of 10 birds sampled at this later stage of infection in the turkey barn showed this mixed pattern (representative specimens shown in Fig. 1).

**Differential real-time PCR on Neuraminidase alleles.** To deter-

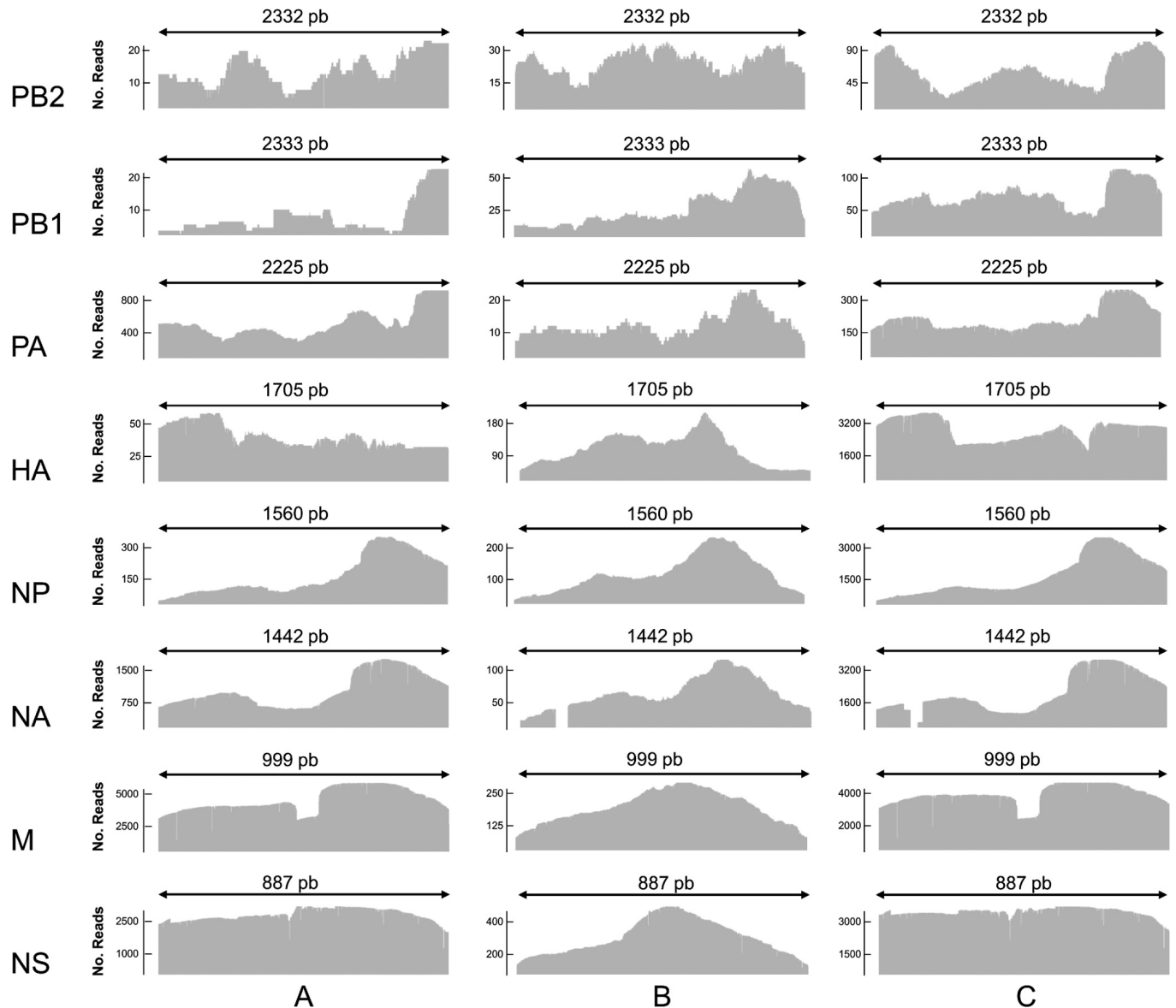


FIG 2 Map of sequence depth of the whole influenza genome for duck (A), early turkey (B), and late turkey virus (C). For each segment, the coverage depth is plotted as a function of nucleotide position. PA: PA polymerase; HA: hemagglutinin; NP: nucleoprotein; NA: neuraminidase; M: matrix; NS: nonstructural.

TABLE 3 Nonsynonymous single nucleotide polymorphisms (SNPs) detected on PB2 and HA<sup>a</sup>

Protein	Nucleotide			Protein			Population ratio (%)		
	Base(s)	Position	Substitution	Amino acid	Position	Substitution	A	B	C
PB2	AGA	181	AAG	R	61	K	0	95	95
	C	300	A	N	100	K	0	0	8
	A	346	C	K	116	Q	0	0	5
	A	361	G	K	121	E	0	7	10
	C	440	T	T	147	I	0	100	100
	A	563	G	E	188	G	25	0	0
	A	569	G	K	190	R	23	0	0
	G	571	A	E	191	K	21	0	0
	G	874	A	V	292	I	0	100	100
	GTG	883	ATT	V	295	I	0	100	100
	GTG	1189	GCA	V	397	A	0	100	100
	G	1261	A	V	421	I	0	0	23
	ATT	1531	GTA	I	511	V	0	100	100
	CCT	1735	ACA	P	579	T	0	14	0
	ACT	1753	CCC	T	585	P	25	100	100
	AAG	1756	CAA	K	586	Q	0	7	14
	GAT	2038	GGC	D	680	G	0	100	100
	HA	A	353	G	K	118	R	0	100
C		358	T	L	120	F	6	0	0
A		509	G	K	170	R	0	0	6
G		534	T	K	178	N	9	0	0
A		560	T	Q	187	L	0	0	38
G		685	A	E	229	K	0	0	25
C		692	T	A	231	V	0	0	25
G		722	C	G	241	A	0	0	14
T		1004	C	L	335	P	0	0	20
C		1041	A	F	347	L	0	0	6
T		1262	C	I	421	T	0	0	23

<sup>a</sup> For each library, corresponding to duck feces (A), early samples in turkeys (B), and late samples in turkeys (C), reads were aligned against the consensus sequence before SNP calling. The percentage of reads sharing the SNP was determined with a threshold of 5%.

mine if the deleted forms were present, even at low copy number, in duck feces samplings, a real-time PCR assay was designed to discriminate between undeleted and deleted NA, using a unique couple of primers encompassing the deletion region of NA and 2 sequence-specific internal probes (Table 2). The copy number was deduced from a plasmidic standard dilution (ranging from 10 to 10<sup>6</sup> copies) containing the cognate sequences. This allele-specific real-time PCR showed a very low threshold of less than 10 copies per reaction (data not shown) and was performed in triplicate on duck and turkey samples. Using this sensitive method, we readily detected the full-length NA but not the truncated segment in duck samples. The same analysis, when applied to early turkey tracheal swabs, only detected the truncated form of the NA segment (Table 2).

**Deep sequencing of complete viral genome of duck and turkey viruses.** Complete genomes were obtained for the 3 pools of samples, with a mean number of reads of 21,600 per pool and an average coverage depth of 680 for all segments. A comprehensive coverage map is shown in Fig. 2.

For each viral gene, SNPs were tracked on the basis of a 5% threshold. Table 3 summarizes the nonsynonymous SNPs and the ratio of mutated viral sequences for PB2 and HA. Briefly, 8 SNPs on PB2 appeared within a few days in 100% of the turkey viral reads. Some minority mutants in the ducks disappeared as readily as in the turkeys. On HA, one SNP (K118R) was identified and remained preponderant in turkeys in the late samples. Eight other minority SNPs, concerning 6% to 32% of the reads, were evi-

denced in late samples on positions 170 to 241, corresponding to the receptor binding region of HA. Data concerning the 6 other segments are shown in Table SA1 in the supplemental material; SNPs were observed in turkeys on PB1 (7 positions), PB1-F2 (14 positions), and PA (3 positions).

A particular analysis was performed on the NA gene; as expected, we identified a 100% 60-nt deleted population in early turkey samples and a mixed 60- and 69-nt deletion pattern in late turkey samples. Alignments of the NA genes could readily evidence a comprehensive pattern of deletions (Fig. 3). More surprising is that the same analysis on duck fecal material evidenced an ultraminority population of deleted 60-nt NA alleles (13/670 reads, i.e., 1.94%), which was not detected by other means (Fig. 3). This critical result demonstrates that minute amounts of deleted NA alleles were present in the duck viral mutant cloud and “emerged” in the turkeys.

## DISCUSSION

In this study, we monitored the evolution of LPAIVs by pyrosequencing performed directly on clinical samples. The whole genomes of 3 pooled samples were subjected to deep sequencing and analysis of viral subpopulations. The global sequence coverage was sufficient to observe low population variation, which may possibly represent quasispecies, and to track the evolution of minority mutants. For this purpose, a critical advantage of 454 pyrosequencing, when compared with other next-generation sequencing platforms, relies on its ability to generate quite long

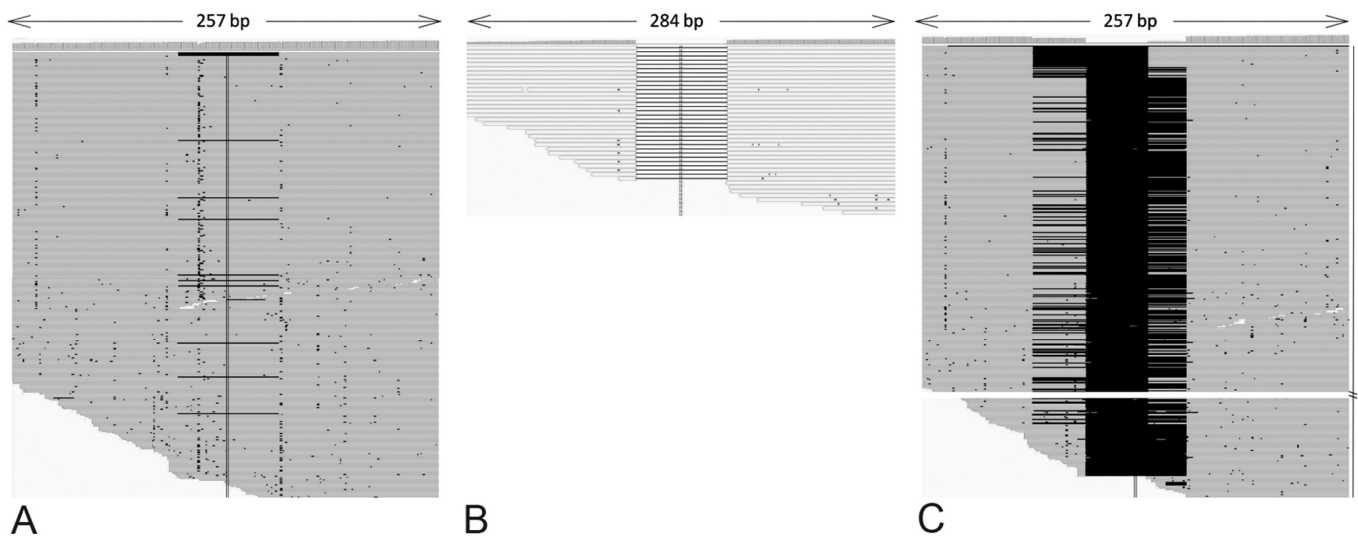


FIG 3 Read alignments on the NA stalk region, focusing on deletion patterns; 454 reads were contiged and aligned. Deletion in the stalk region (positions 161 to 220) was found in only 13 of 670 reads in duck feces (A) but in 100% of reads in early samples from turkeys (B). A mix of 60- and 69-nt (positions 129 to 197) deletions was observed in turkeys 8 days later (C).

reads, of roughly 400 nt (at the time of the experiment), facilitating subsequent sequence reconstruction.

One of the most significant changes observed was the emergence of 2 avian influenza NA stalk deletion variants after virus transmission from the ducks to the turkeys. In this case study, the NA stalk deletions could be monitored in the field, at the scale of a single farm, with a direct link from the parental to the deletion variants.

Our data favor the idea that NA stalk truncation emerged early after turkey contamination, as a minute number of reads of deleted NA could be detected in duck feces. Besides, we did not detect any full-length NA form in turkeys, even in early samples, suggesting a strong evolutionary advantage of the truncated form at the expense of the full-length one in this poultry species.

Very surprising is that after 8 more days of viral circulation in turkeys, some birds appeared to be coinfecting with a second variant harboring a larger deleted region, which did not totally overlap with the 60-nt deletion previously observed. This later  $\Delta 69$  allele could not be detected in early turkey samples, maybe because the depth coverage of NA was quite low for this early sample. This finding suggests a higher plasticity of neuraminidase than previously thought.

NA stalk truncation events, which reduce the protein enzymatic activity, have been shown to be associated with changes in the glycosylation pattern of the HA globular head, resulting in a modification of HA affinity for cellular receptors (5, 17, 19).

We tracked the evolution of HA and could evidence the emergence of minority mutants, with substitutions mostly in the receptor binding site on the globular head: these SNPs concerned up to 38% of viral species after 8 days of clinical evolution in the flock. None of these mutations altered glycosylation sites of HA. They may have contributed to the refinement of binding to cellular receptors of the turkeys' upper respiratory tract and/or to circumvent the host's immune response (18).

It is of interest to note that the observed mortality in turkeys was contemporary to NA deletion variant detection, whereas ducks did not show any clinical sign of infection. Even though NA

stalk deletion was recently associated with an increased virulence in chickens infected by reverse-genetics-engineered viruses (8, 20), whether the deletions described in this field case are responsible for the observed virulence in turkey remains to be demonstrated.

In this farm, the 2 species were bred separately in 2 different houses. The contamination of turkeys can likely be explained by a biosecurity failure, which probably occurred when ducks were transported to the slaughterhouse. This case suggests further that the turkey is a very peculiar host among poultry species and shows very high receptivity and sensitivity to influenza A viruses (21). Recent epidemiological modeling data suggest a very high within-flock transmissibility of LPAIVs in turkeys (1). This case also emphasizes the risks of breeding together different avian species, i.e., terrestrial birds (chickens or turkeys) and waterfowl. Such a situation, which seems relatively common in Asia, can result in the quick emergence of variants with altered phenotypes and increased pathogenicity. Moreover, it illustrates the genetic heterogeneity of avian influenza viruses within a single flock. In a first attempt, our approach was intentionally based on the poultry population considered as a whole, irrespective of the individual level. The next step will be to investigate quasispecies variations between birds in a flock and then between different tissues (i.e., respiratory versus digestive tracts) in given birds.

In previous studies, next-generation sequencing approaches have already proven to be of interest for monitoring viral evolution at precise viral genome positions following evolutionary constraints such as antiretroviral treatment during HIV infection (6) or oseltamivir treatment during influenza infection (9). A broader approach of viral genome evolution may allow us to monitor the early apparition of mutants without *a priori* consideration and to follow complex chains of mutational events occurring on several positions of viral genomes. The 454 sequencing platform used here has the critical advantage of producing long reads—from 400 nt to more than 700 nt in recent updates—facilitating sequence reconstruction in a whole viral genome approach. Alternative technologies are in constant development and will likely provide



improved sequencing capacities in the near future. The development of next-generation sequencing approaches should enable us to monitor both viral genome diversity within bird flocks and the early apparition of potentially pathogenic variants.

Although our data do not explain how the NA stalk deletion appeared in the duck samples in the first place, this present study suggests that this deletion can exist as a minority mutant in waterfowl, awaiting a change of host to be selected.

#### ACKNOWLEDGMENTS

We thank Mariette Ducatez for helpful comments on the manuscript. We acknowledge the assistance of Jerome Lluch, Genomics platform of Toulouse, France, with the 454 pyrosequencing.

This project was partly supported by French Region Midi-Pyrénées.

#### REFERENCES

1. Comin A, Klinkenberg D, Marangon S, Toffan A, Stegeman A. 2011. Transmission dynamics of low pathogenicity avian influenza infections in Turkey flocks. *PLoS One* 6:e26935.
2. Di Trani L, et al. 2004. Molecular characterization of low pathogenicity H7N3 avian influenza viruses isolated in Italy. *Avian Dis.* 48:376–383.
3. Fouchier RA, et al. 2000. Detection of influenza A viruses from different species by PCR amplification of conserved sequences in the matrix gene. *J. Clin. Microbiol.* 38:4096–4101.
4. Giannecchini S, et al. 2006. Comparison of in vitro replication features of H7N3 influenza viruses from wild ducks and turkeys: potential implications for interspecies transmission. *J. Gen. Virol.* 87:171–175.
5. Guionie O, et al. 2007. Laboratory evaluation of a quantitative real-time reverse transcription PCR assay for the detection and identification of the four subgroups of avian metapneumovirus. *J. Virol. Methods* 139:150–158.
6. Hedskog C, et al. 2010. Dynamics of HIV-1 quasispecies during antiviral treatment dissected using ultra-deep pyrosequencing. *PLoS One* 5:e11345.
7. Hoffmann E, Stech J, Guan Y, Webster RG, Perez DR. 2001. Universal primer set for the full-length amplification of all influenza A viruses. *Arch. Virol.* 146:2275–2289.
8. Hoffmann TW, et al. 2012. Length variations in the NA stalk of an H7N1 influenza virus have opposite effects on viral excretion in chickens and ducks. *J. Virol.* 86:584–588.
9. Inoue M, et al. 2010. Emergence of oseltamivir-resistant pandemic (H1N1) 2009 virus within 48 hours. *Emerg. Infect. Dis.* 16:1633–1636.
10. Kent WJ. 2002. BLAT: the BLAST-like alignment tool. *Genome Res.* 12:656–664.
11. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
12. Li H, et al. 2009. The Sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
13. Li J, Zu Dohna H, Cardona CJ, Miller J, Carpenter TE. 2011. Emergence and genetic variation of neuraminidase stalk deletions in avian influenza viruses. *PLoS One* 6:e14722.
14. Mariette J, Noirot C, Klopp C. 2011. Assessment of replicate bias in 454 pyrosequencing and a multi-purpose read-filtering tool. *BMC Res. Notes* 4:149.
15. Mathieu C, et al. 2010. Pandemic (H1N1) 2009 in breeding turkeys, Valparaiso, Chile. *Emerg. Infect. Dis.* 16:709–711.
16. Matrosovich M, Zhou N, Kawaoka Y, Webster R. 1999. The surface glycoproteins of H5 influenza viruses isolated from humans, chickens, and wild aquatic birds have distinguishable properties. *J. Virol.* 73:1146–1155.
17. Matsuoka Y, et al. 2009. Neuraminidase stalk length and additional glycosylation of the hemagglutinin influence the virulence of influenza H5N1 viruses for mice. *J. Virol.* 83:4704–4708.
18. McHardy AC, Ben Adams B. 2009. The role of genomics in tracking the evolution of influenza A virus. *PLoS Pathog.* 5:e1000566.
19. Mitnaul LJ, et al. 2000. Balanced hemagglutinin and neuraminidase activities are critical for efficient replication of influenza A virus. *J. Virol.* 74:6015–6020.
20. Munier S, et al. 2010. A genetically engineered waterfowl influenza virus with a deletion in the stalk of the neuraminidase has increased virulence for chickens. *J. Virol.* 84:940–952.
21. Pillai SPS, Pantin-Jackwood M, Yassine HM, Saif YM, Lee CW. 2010. The high susceptibility of turkeys to influenza viruses of different origins implies their importance as potential intermediate hosts. *Avian Dis.* 54:522–526.
22. Robinson JT, et al. 2011. Integrative genomics viewer. *Nat. Biotechnol.* 29:24–26.
23. Sorrell EM, Song H, Pena L, Perez DR. 2010. A 27-amino-acid deletion in the neuraminidase stalk supports replication of an avian H2N2 influenza A virus in the respiratory tract of chickens. *J. Virol.* 84:11831–11840.
24. Zhou H, et al. 2009. The special neuraminidase stalk-motif responsible for increased virulence and pathogenesis of H5N1 influenza A virus. *PLoS One* 4:e6277.



## Whole-genome, deep pyrosequencing analysis of a duck influenza A virus evolution in swine cells



Vincent Bourret<sup>a,b,c,\*</sup>, Guillaume Croville<sup>b,c</sup>, Jérôme Mariette<sup>d</sup>, Christophe Klopp<sup>d</sup>, Olivier Bouchez<sup>e</sup>, Laurence Tiley<sup>a</sup>, Jean-Luc Guérin<sup>b,c</sup>

<sup>a</sup> Department of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge CB3 0ES, UK

<sup>b</sup> Université de Toulouse, INP, ENVT, 31076 Toulouse, France

<sup>c</sup> INRA, UMR 1225, IHAP, 31076 Toulouse, France

<sup>d</sup> Plateforme bioinformatique Toulouse Midi-Pyrénées, UBIA, INRA Auzeville, 31326 Castanet-Tolosan, France

<sup>e</sup> GeT-PlaGe, Genotoul, INRA Auzeville, 31326 Castanet-Tolosan, France

### ARTICLE INFO

#### Article history:

Received 30 November 2012

Received in revised form 14 March 2013

Accepted 29 April 2013

Available online 7 May 2013

#### Keywords:

Influenza A virus

Pyrosequencing

Viral evolution

Within-host diversity

Mixed infection

Reassortment

### ABSTRACT

We studied the sub-population level evolution of a duck influenza A virus isolate during passage in swine tracheal cells. The complete genomes of the A/mallard/Netherlands/10-Nmkt/1999 strain and its swine cell-passaged descendent were analysed by 454 pyrosequencing with coverage depth ranging from several hundred to several thousand reads at any point. This allowed characterization of defined minority sub-populations of gene segments 2, 3, 4, 5, 7, and 8 present in the original isolate. These minority sub-populations ranged between 9.5% (for segment 2) and 46% (for segment 4) of their respective gene segments in the parental stock. They were likely contributed by one or more viruses circulating within the same area, at the same period and in the same or a sympatric host species. The minority sub-populations of segments 3, 4, and 5 became extinct upon viral passage in swine cells, whereas the minority sub-populations of segments 2, 7 and 8 completely replaced their majority counterparts. The swine cell-passaged virus was therefore a three-segment reassortant and also harboured point mutations in segments 3 and 4. The passaged virus was more homogenous than the parental stock, with only 17 minority single nucleotide polymorphisms present above 5% frequency across the whole genome. Though limited here to one sample, this deep sequencing approach highlights the evolutionary versatility of influenza viruses whereby they exploit their genetic diversity, predilection for mixed infection and reassortment to adapt to a new host environmental niche.

© 2013 Elsevier B.V. All rights reserved.

### 1. Introduction

The genome of influenza A viruses consists of eight negative-sense single stranded RNA segments (Palese, 1977). The viral RNA polymerase lacks proofreading function, and error rates of  $9 \times 10^{-6}$  to over  $7 \times 10^{-5}$  mutation per nucleotide per RNA replication cycle have been proposed (Drake, 1993; Suárez-López and Ortín, 1994; Drake and Holland, 1999; Sanjuán et al., 2010). Influenza viruses also evolve through genetic reassortment, a highly efficient form of modular recombination. Distinct viruses, infecting

the same cell, can exchange gene segments giving rise to progeny with new gene combinations. The high mutation rate of the viral RNA polymerase together with segment exchange make influenza a rapidly evolving and highly adaptable virus considered to exist naturally within individual hosts as swarms of many variant viruses that closely conform to the population consensus sequence (e.g. Iqbal et al., 2009; Murcia et al., 2010, 2012; Croville et al., 2012).

Wild aquatic birds are considered to be the primary natural reservoir for influenza A viruses (Hinshaw et al., 1982 and references therein; reviewed in Webster et al., 1992; Alexander, 2000; Horimoto and Kawaoka, 2001). Some viruses can be transmitted from aquatic birds to mammals including humans, pigs, horses, sea mammals, ferrets and mink (reviewed in Forrest and Webster, 2010) where they may adapt to become established in such new host species. Influenza infections can have serious consequences in these spill-over hosts, including substantial mortality. Notably, influenza is among the primary respiratory pathogens in humans

*Abbreviations:* a.f.t., allele frequency threshold; bp, base pairs; dNTP, deoxyribonucleotides triphosphate; EMEM, Eagle's Minimum Essential Medium; FCS, foetal calf serum; LB, Luria-Bertani; MID, multiplex identifier sequence; NPTr, Newborn Pig Trachea; SD, standard deviation; SNP, single nucleotide polymorphism.

\* Corresponding author. Address: Unité de maladies contagieuses, Ecole Nationale Vétérinaire de Toulouse, 23, chemin des Capelles, 31076 Toulouse, France. Tel.: +33 561 193 233; fax: +33 561 193 974.

E-mail addresses: [v.bourret@envt.fr](mailto:v.bourret@envt.fr), [vjrb2@cam.ac.uk](mailto:vjrb2@cam.ac.uk) (V. Bourret).

(reviewed in WHO, 2009) and swine (reviewed in Kuntz-Simon et al., 2010). Further, swine has been suspected to act as an intermediate host for adaptation of avian influenza strains to humans, and potentially enable reassortment between strains of avian and human origins (Scholtissek et al., 1985; Castrucci et al., 1993; reviewed in Horimoto and Kawaoka, 2005). Avian influenza evolution in swine is therefore a topic of prominent interest in influenza epidemiology.

The 454 pyrosequencing approach has the power to study minority variants and mixed infections in field-derived influenza samples (Ramakrishnan et al., 2009), and is a method of choice for reconstruction of full-length viral sub-population haplotypes (Zagordi et al., 2012). Here we extended this approach to a deep, whole-genome analysis of the genetic evolution of a duck influenza strain in swine tracheal cells at the sub-population level.

## 2. Materials and methods

### 2.1. Passaging in swine cells

A/mallard/Netherlands/10-Nmkt/1999 (mixed subtype H1N1 and H11N1, *cf. infra*) is a low pathogenicity avian influenza virus stock (hereafter referred to as “Wt”) which has undergone relatively little passage since field isolation (a total of three amplification rounds in embryonated hen’s eggs). This virus stock was then serially passaged in the Newborn Pig Trachea (NPTr) cell line (Ferrari et al., 2003). Passaging was done at 37 °C, 5% CO<sub>2</sub> in 25 cm<sup>2</sup> flasks seeded with 2 × 10<sup>6</sup> cells in Eagle’s Minimum Essential Medium (EMEM) supplemented with penicillin (100 U/mL), streptomycin (50 µg/mL), amphotericin (2.6 µg/mL), L-glutamine (10 mM), 1× non-essential amino acids, and 10% (v/v) foetal calf serum (FCS). On the next day, cells were washed with phosphate-buffered saline and infected with 1 mL of inoculum consisting of 2.4 × 10<sup>4</sup> viral genome copies diluted in EMEM supplemented with 0.3% (w/v) bovine serum albumin instead of FCS and 0.25 µg/mL trypsin (Worthington). This inoculum was equivalent to 50 plaque forming units on Madin–Darby Canine Kidney cells. After one hour, an additional 4 mL of the same medium was added to the culture and the virus was left to propagate for 2 days. Fresh cells seeded the day before were then infected with a 1 in 2.5 × 10<sup>4</sup> dilution of the viral yield from the previous passage (corresponding to 3.8 × 10<sup>5</sup> – 5.0 × 10<sup>6</sup> viral genome copies). The procedure was repeated 10 consecutive times. The resulting viral stock, harvested from the 10<sup>th</sup> passage, is referred to as “Ad”.

### 2.2. Sanger sequencing

#### 2.2.1. Amplification of full-length influenza segments

The QiaAmp Viral Mini Kit (Qiagen) was used for RNA extraction from allantoic fluid and cell culture supernatant according to manufacturer’s instructions. Reverse transcription of viral RNA into cDNA was carried out using Fermentas RevertAid Premium reverse transcriptase as per manufacturer’s instructions, with a primer targeted at the 3′ 12-nucleotide conserved region of the eight vRNA segments. To accommodate the known variation at position 4 of the 3′ end of vRNA, the corresponding base in the primer was A or G (denoted R) (primer vRNART: AGCRAAAGCAGG).

Polymerase chain reactions were carried out using Platinum Pfx DNA polymerase (Invitrogen) in 50 µL reactions each containing 1× Pfx Amplification Buffer, 0.3 mM dNTP, 0.9 mM MgSO<sub>4</sub>, 1.25 units of Pfx DNA polymerase, 300 nM forward and reverse primers (detailed in Supplementary Table S1) and 1 µL of template cDNA. Cycling consisted of an initial enzyme activation step at 94 °C for 5 min, followed by 35 cycles of denaturation at 94 °C for 15 s, annealing for 30 s, and extension at 68 °C. Optimised

annealing temperatures and extension times depended on the segment and are detailed in Table S1. For the polymerase genes, a “touchdown” programme was used where the annealing temperature decreased by 1 °C every cycle for the first 12 cycles and then was maintained over the remaining cycles.

The Wt stock was found to be a mixture containing different haemagglutinin (HA) subtypes. For segment 4 of this virus, we therefore used a primer pair targeted only at the 3′ and 5′ regions conserved between all known HA subtypes (*i.e.*, non subtype-specific). This allowed amplification of all haemagglutinins present in the sample. In order to amplify only H1 haemagglutinins out of the mixture, we then used a pair of primers extending into the H1 subtype-specific region of the gene. These different HA primer pairs are detailed in Table S1.

#### 2.2.2. PCR products purification

Amplification products were run on a 1% agarose gel in Tris–Borate–EDTA. Amplification conditions were optimised to achieve single specific amplification products that were then purified using the QIAquick PCR Purification Kit (Qiagen). Where multiple products could not be avoided, the target band was excised from the gel and purified using the GeneJET Gel Extraction Kit (Fermentas). Concentration and purity of amplified DNA was assessed by NanoDrop (Thermo Scientific).

#### 2.2.3. Sequencing

We first carried out direct Sanger sequencing of PCR products, allowing determination of the consensus sequence and giving an opportunity to detect some minority alleles as secondary peaks on sequence chromatograms. Purified PCR products were diluted to the appropriate concentration and sent for sequencing to an external company (GeneService) using internal plus-sense sequencing primers detailed in Supplementary Table S2. Sequencing of the 5′ end of full-length genes was achieved by using one anti-sense primer for each segment.

To sequence the H1 population amplified using specific primers (Table S1), we used the H1 sequencing primers detailed in Table S2. To sequence the H11 population, we selected H11 clones based on restriction profiles, and then sequenced them using two H11 internal primers and two primers targeted at promoters from the cloning vector flanking the HA sequence.

Sequence contig assembly, translation and alignments were carried out using a web interface to the European Molecular Biology Open Software Suite (EMBOSS, Rice et al., 2000). These Sanger sequences were deposited in GenBank under accession numbers KC209503–KC209519.

#### 2.2.4. Molecular cloning

Cloning was carried out in the RF483 plasmid, known as pDual, which was modified in Prof. Fouchier’s laboratory (Erasmus Medical Centre, Rotterdam) from the pHW2000 plasmid (Hoffmann et al., 2000). This plasmid bears a resistance gene to ampicillin and a cassette comprising two BsmBI cloning sites to receive the viral cDNA. The plasmid was digested twice sequentially using BsmBI (New England Biolabs) in a 100 µL reaction in water containing 10 µL of a 10× buffer, 4 µg of plasmid, and 20 units of enzyme. The reaction was covered with mineral oil and incubated at 55 °C for 3 h. The plasmid was then purified and eluted using the QIAquick PCR Purification Kit (Qiagen) before the process was repeated. The viral segments amplified with the Bsalseg4 primers (Supplementary Table S1) were digested once using a similar protocol with BsaI (New England Biolabs), adjusting the buffer and reaction temperature as per manufacturer’s instruction. Cut segments were purified using the QIAquick PCR Purification Kit. The overhangs were directionally compatible with the BsmBI sites of RF483. Ligations were done in 10 µL reactions with 50 ng of

plasmid, three times as much insert as plasmid in molar ratio, 2  $\mu$ L of 5 $\times$  ligase buffer, 5 units of ligase T4 (Invitrogen), and water up to 10  $\mu$ L, and incubated at room temperature for 1 h. Ligation products were transformed into DH5 $\alpha$  competent cells (Invitrogen) as per manufacturer's instructions with the following modification: heat shock was 37  $^{\circ}$ C for 30 s, and grown on ampicillin agar Luria–Bertani medium (LB). After an overnight incubation of the plates at 37  $^{\circ}$ C, a 3 mL liquid LB broth culture with 100  $\mu$ g/mL of ampicillin was inoculated with a single colony and incubated overnight at 37  $^{\circ}$ C with a 225 rotation per minute agitation. The plasmids were then purified using the QIAprep Spin Miniprep Kit (Qiagen) as per manufacturer's instructions.

### 2.3. Pyrosequencing

#### 2.3.1. Laboratory protocols

For pyrosequencing analyses, the complete genomes of the Wt and Ad samples were amplified and purified from the liquid samples (allantoic fluid or cell culture supernatant for Wt and Ad, respectively) following the same protocol as in Croville et al. (2012). This allowed purification of sufficient quantities of each segment from each virus.

The purified PCR products were then quantified using Quant-iT PicoGreen kit (Molecular Probes). Two libraries were generated, corresponding to the Wt and the Ad virus respectively and each containing 500 ng DNA comprising the eight influenza gene segments. Libraries were fragmented and prepared using the GS DNA Rapid Library Preparation Kit (Roche) according to manufacturer's recommendations. Each DNA pool was ligated to an adapter containing a multiplex identifier sequence for emulsion PCR and sequencing. Final libraries were quantified using the SlingShot™ kit with the Fluidigm® Digital Array for sample quantification following manufacturer's instructions. Pyrosequencing using the 454/Roche GS FLX Titanium chemistry was carried out with the GS Titanium SV emulsion PCR kit and sequencing kit according to manufacturer's instructions. Each sample was sequenced on a separate region of an 8-region Picotiterplate.

#### 2.3.2. Bioinformatic analysis

Sequences were first demultiplexed using Roche's tool SFF file without allowing any error per MID. Reads were then filtered using Pyrocleaner (Marette et al., 2011) considering different criteria such as length (with reads shorter or longer than mean read length  $\pm 2$  standard deviations being discarded). Reads were also filtered based on their complexity which is computed using the compressed string length (library zip) on several sub-sequences generated using a sliding window approach. The pyrocleaner script removed sequences for which all base pairs had a Phred quality va-

lue under 20, or if the rate of undetermined bases was higher than 4% (which has been correlated with poor quality).

The remaining sequences were next aligned on the reference sequences obtained from Sanger sequencing using the Burrows–Wheeler Alignment package (BWA; Li and Durbin, 2009), and nucleotide polymorphisms were then identified using the SAMtools software package (Li et al., 2009).

Gene by gene analysis of pyrosequencing alignments was carried out using the Integrative Genomics Viewer (IGV) graphic interface (Robinson et al., 2011). Searches into an online database for sequences most similar to a given query sequence were done using the Basic Local Alignment Search Tool (BLAST; Altschul et al., 1990).

### 3. Results

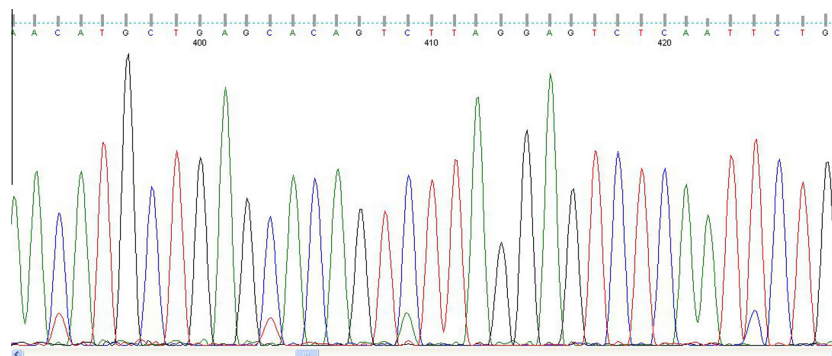
The A/mallard/Netherlands/10-Nmkt/1999 low pathogenicity avian influenza virus stock (referred to as “Wt”) was serially passaged in the Newborn Pig Trachea (NPT<sub>r</sub>) cell line (Ferrari et al., 2003). The descendent virus is referred to as Ad. The Wt and Ad stocks were first analysed by direct Sanger sequencing, allowing determination of the consensus sequence for the various segments. This technique can also allow, to some extent, qualitative detection of minority alleles in a sample. The Wt and Ad viruses were then pyrosequenced using the 454 technique (Roche). Because it sequences individual nucleic acid molecules from the population, this technique has the advantages of allowing detection of minority sub-populations, quantitative estimates of their respective frequencies, and detection of polymorphism associations at the molecular level.

#### 3.1. Sanger sequences analysis

At the consensus level, no difference was detected between the Wt and the Ad viruses on segments 1 (PB2), 5 (NP) and 6 (NA).

On segment 2, 114 nucleotide differences resulted in 7 amino acid differences on the PB1 protein and 11 amino acid differences on the PB1-F2 protein. On segment 3 (PA), one synonymous nucleotide difference was detected. On segment 4 (HA), which originally contained a mixture of an H1 and H11 alleles, the H11 allele was lost and two non-synonymous differences appeared on the H1 allele. On segment 7 (M), 32 nucleotide differences were detected, resulting in no amino acid difference in M1 or M2 but 2 amino acid differences on M42 (Wise et al., 2012). On segment 8 (NS), 35 nucleotide differences were detected, resulting in 2 amino acid differences on NS1 and one amino acid difference on NS2.

In the case of segments 2, 7 and 8, precursors to the Ad virus' alleles were detected in the Wt virus population due to secondary



**Fig. 1.** Section of a chromatogram from the Wt virus' segment 2 illustrating detection of secondary peaks by direct Sanger sequencing (positions 394, 403, 409 and 424 of this read).

peaks on chromatograms (Fig. 1). This suggested that, for those genes, the alleles that went on to define the Ad population pre-existed as minority alleles in the Wt population and became dominant upon passage in swine cells. By contrast, chromatograms from the Ad virus indicated the virus diversity had decreased as a result of the selection protocol.

These results raised a number of quantitative and qualitative questions: at what proportion were the Ad virus' segments 2, 7, and 8 already present in the Wt stock? Did these segments mutate as well as undergoing reassortment? Were the point mutations on segments 3 and 4 characteristic of Ad present at a low level in the Wt population, or did they appear *de novo*? Is Ad a completely pure population, or are there sub-populations undetected by Sanger sequencing? Notably, has the H11 allele disappeared completely? What is the variability, if any, within Wt's segments 1, 5, 6? Insights into these and other issues were provided by 454 pyrosequencing of the complete wild type and adapted viruses.

### 3.2. Pyrosequencing results overview

For the Wt virus, 52,346 reads were obtained for a total of 21,613,370 base pairs (bp) (average read length: 412.9 bp). This resulted in an average depth of 1589 reads (noted 1589 $\times$ ) at each position. For the Ad virus, 93,905 reads were obtained for a total of 39,033,643 bp (average read length: 415.7 bp). This resulted in an average depth of 2868 $\times$  at each position.

Fig. 2 sums up the coverage and variability at the 5% allele frequency threshold (a.f.t.) for each gene of both viruses. The coverage depth varied from several hundred to several thousand reads at any position, with, as a general trend, the extremities of the genes (especially towards the 3' end of the cDNA sequences) tending to be read more often than the central parts.

There was substantial polymorphism at the 5% a.f.t. within Wt's segments 2, 3, 4 (with the H1 and H11 alleles present in nearly equal proportions), 5, 7 and 8, with numerous positions showing polymorphism within each of these genes. By contrast, Wt's segments 1 and 6, and all of Ad's segments, showed very few polymorphic positions at this threshold. Overall, Ad was a much purer population than Wt, with only 17 minority SNP present at or above 5% across the whole genome.

### 3.3. Reassortment of defined minority populations of gene segments initially present within virus Wt

#### 3.3.1. Precursors for Ad's gene segments 2, 7 and 8 within Wt

At every position differing between the respective consensus of Wt and Ad for segments 2, 7 and 8, a precursor for Ad's allele existed within Wt. These precursors were present in Wt at an average frequency of 9.5% ( $\pm 1.1\%$  standard deviation, SD) in segment 2, 28% ( $\pm 4.8\%$  SD) in segment 7, and 17% ( $\pm 2.8\%$  SD) in segment 8. These relatively homogeneous frequencies within each Wt gene, reflected by low to moderate SD values, suggested that these Ad alleles could tend to be borne by the same RNA molecules within the Wt pool. This was confirmed by visual inspection of individual reads: the minority alleles from Wt which became dominant in Ad tended to be present on the same reads (which originated from the same vRNA molecules), as exemplified for segment 2 on Fig. 3. However, the frequency differences between segments (9.5% vs 28% vs 17%) suggest that each was present in the population at a different level and they were already undergoing reassortment in the original mix and not simply present as a sub-population of a single co-infecting virus. This is also evident with sub-populations of segments 3, 4 and 5 (see below) constituting 14–46% of their respective total.

The SNP within Wt's segments 2, 7 and 8 that were lost upon passaging (*i.e.* that are absent from Ad) were generally located on

different reads to the ones bearing the minority alleles that became dominant in Ad (see for instance the minority adenine indicated by an arrow towards the left of Fig. 3).

There were exceptions to this general pattern with some individual sequence reads bearing some, but not all polymorphisms characteristic of virus Ad (see for instance one lone T residue indicated by an arrowhead towards the right of Fig. 3). While the frequencies of Ad's alleles in the Wt stock along the various positions of the PB1 gene segment were very homogeneous (Fig. 4, top), more heterogeneity was observed in the cases of the M and NS gene segments (Fig. 4, bottom). Deviations from a flat line suggested that recombination events may have occurred within the Wt population between consensus and minority segments upon co-infection of single cells, prior to passaging in swine cells. This interpretation however needs to be balanced in the view of possible PCR artefacts resulting in chimeric segments (see Discussion).

Deep sequencing showed that the majority segments from Wt that were absent from the consensus sequence of Ad had their frequency reduced to less than 0.1% on average in Ad, whose segments 2, 7 and 8 were generally very pure. In segment 2, no minority allele attained 10% in frequency and 15 minority alleles were present at or above 1%. In segment 7, one minority allele exceeded 10% (21%) and 8 minority alleles were present at or above 1%. In segment 8, one minority allele exceeded 10% (18%) and five minority alleles were present at or above 1%.

#### 3.3.2. Suppression of strong, defined minority populations from the Wt mix upon evolution

**Segment 3.** Wt's segment 3 showed 66 positions along the gene where a minority allele was present at a frequency comprised between 30% and 50%. As with segments 2, 7 and 8, these minority polymorphisms tended to be present on the same reads. However, opposite to what happened in segments 2, 7, and 8, this strong minority population was selected out and almost completely disappeared in Ad.

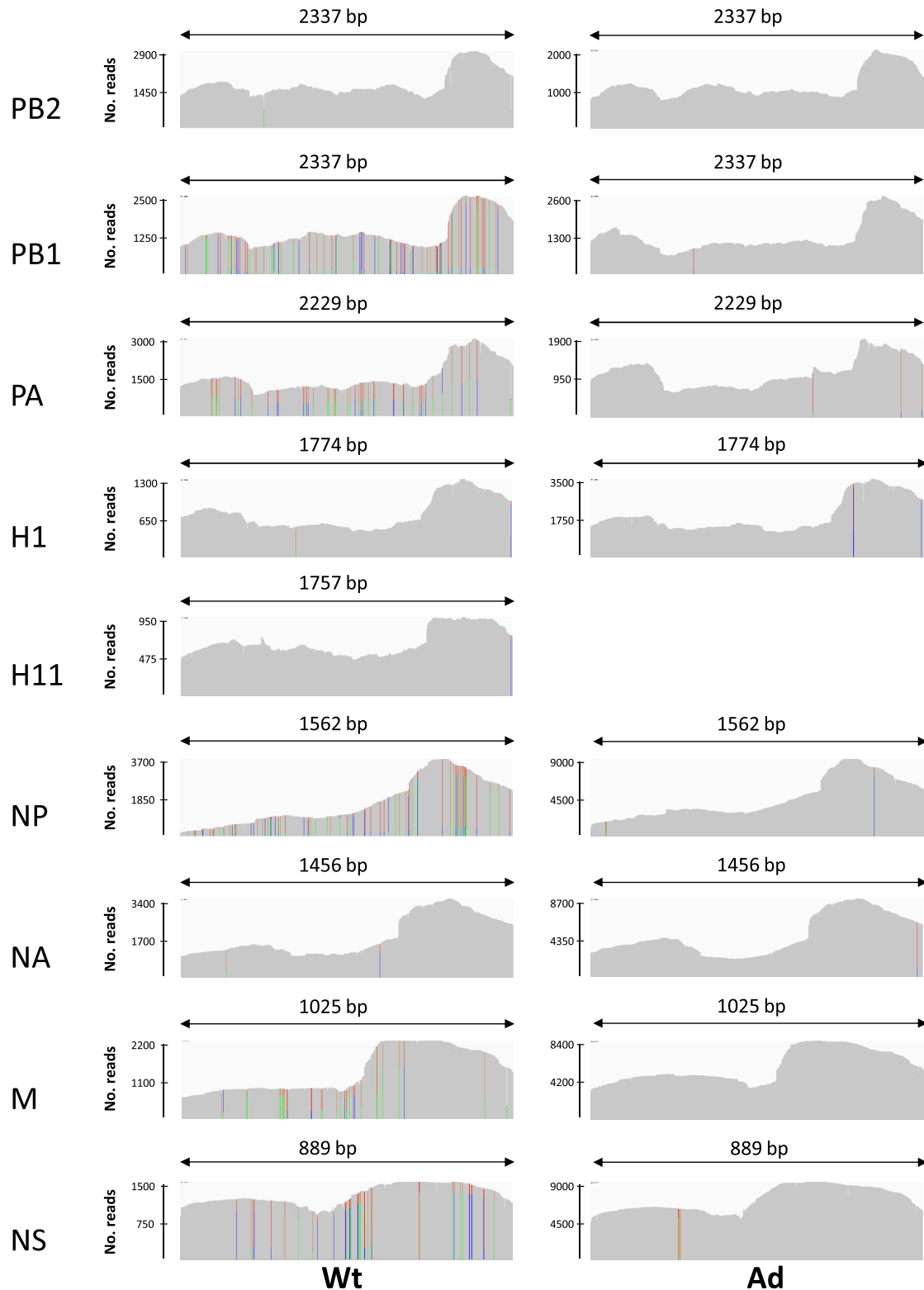
**Segment 4.** The average coverage depth for the H1 allele in Wt was 796 $\times$ , while the average coverage depth for the H11 allele was 682 $\times$ . This suggested proportions of approximately 54% for H1 and 46% for H11 in this virus. The H11 allele however went on to be completely undetected in Ad. Its complete absence suggests that this segment was strongly selected against in the swine cells.

**Segment 5.** At the 10% a.f.t., there were 97 polymorphic sites in Wt with minority alleles at an average frequency of 14%  $\pm$  1.5% SD. This low dispersion suggested that those minority alleles could tend to be present on the same RNA molecules and make up a defined minority sub-population as for segments 2, 3, 4, 7, 8. This was confirmed by visual inspection of individual sequence reads. As with segments 3 and 4, these polymorphisms largely disappeared in Ad.

### 3.4. Frequency and location of point mutations

#### 3.4.1. Point mutations observed at the consensus level

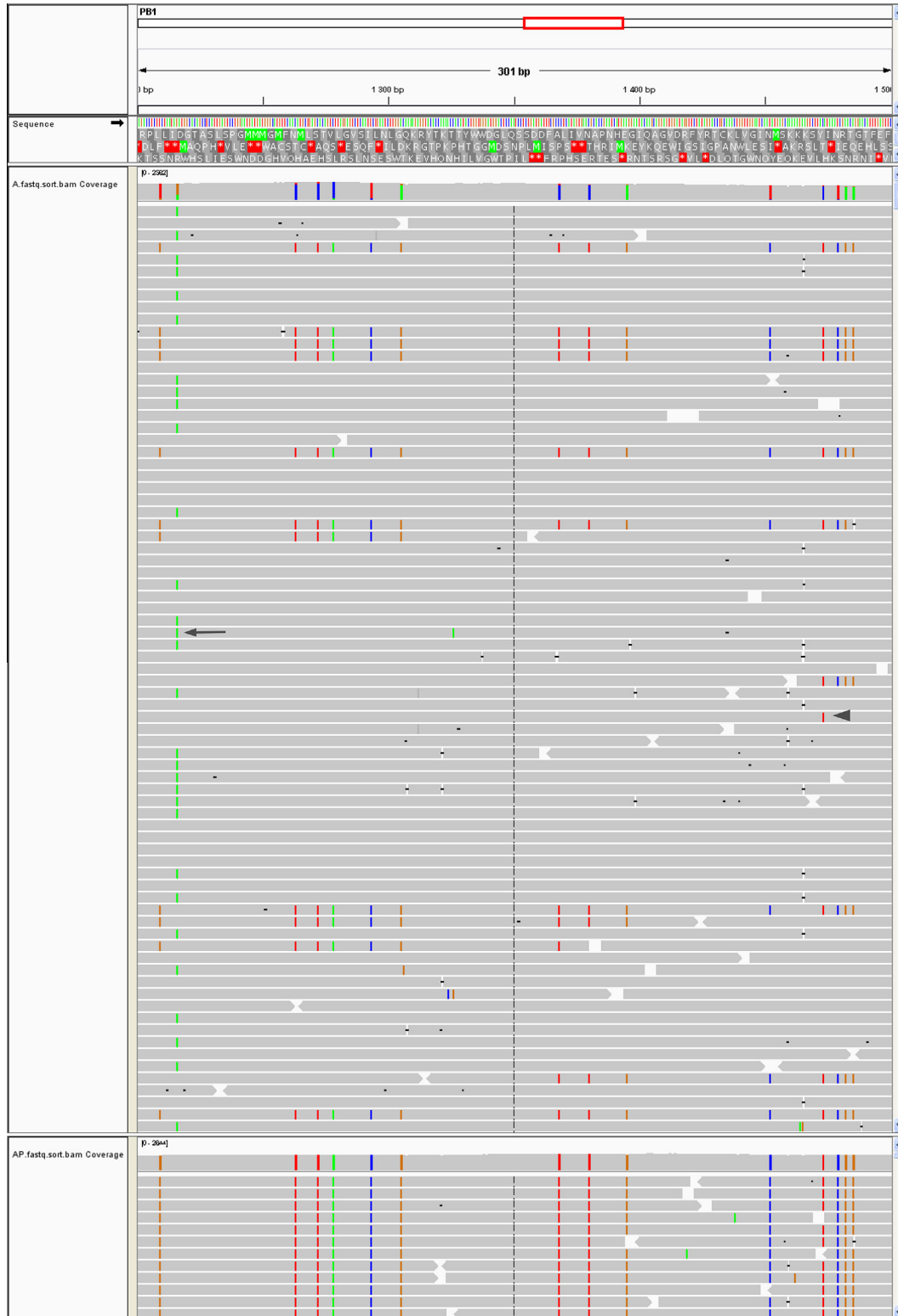
**Segments 3 and 4.** Ad's characteristic segment 3 consensus single nucleotide polymorphism (SNP) was found to pre-exist in Wt at a frequency of 10% when analysed by deep sequencing. Therefore it was not a spontaneous mutation that occurred during the selection process but it became 99% dominant in Ad after selection (Wt's majority allele came down to 1% in Ad). In contrast, the two segment 4 SNP characteristic of Ad were completely undetected in Wt (0/656 and 0/1238 counts respectively). It is therefore likely that they appeared through *de novo* mutation, or that they were present below our detection threshold. Wt's allele at one position



**Fig. 2.** 454 coverage levels and variability at the 5% allele frequency threshold (a.f.t.) within each gene for viruses Wt and Ad (left and right columns respectively). The H1 and the H11 alleles of Wt are shown (no H11 allele was detected in Ad). Grey indicates no departure from the consensus within the viral population at the 5% a.f.t., while coloured bars indicate different alleles present at or above that threshold following the code: green, A; red, T; brown, G; blue, C; coloured bars are proportional in height to the number of reads bearing a given nucleotide. A vertical scale by each gene indicates the coverage depth. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

was almost totally lost in Ad (where it remained at a frequency of less than 0.1%) while Wt's allele at the other position persisted in Ad at a frequency of 37%.

*Segments 2, 7 and 8.* No evidence of *de novo* point mutation was detected for segment 2. A minority SNP of Ad's segment 7 present at a 21% frequency likely resulted from a *de novo* mutation as it



**Fig. 3.** Image from the 454 analysis of the PB1 genes of viruses Wt and Ad. A 301-bp region of the gene is shown (positions 1200–1500). The display consists of three tracks aligned horizontally below the top panel which indicates the position of the focus region within the whole gene. (i) Track 1 (labelled “Sequence”) indicates the consensus nucleotide for Wt at every position (colour-coded as in Fig. 2) together with the amino acid sequence for each open reading frame. (ii) Track 2 (labelled “A.fastq.sort.bam Coverage”) is divided vertically into two panels. The top panel shows the coverage depth for the Wt virus at every position, with polymorphisms at the 5% a.f.t. indicated by coloured bars (grey indicates no departure from the consensus at the chosen a.f.t.). The bottom panel shows individual Wt reads represented by horizontal grey stripes (a depth of 77 reads is visible on this shot). Departures from the consensus are indicated by coloured bars. (iii) Track 3 (labelled “AP.fastq.sort.bam Coverage”) is Ad’s analogous to Wt’s track 2, with only the first 12 reads displayed here for this virus. Note how minority polymorphisms in Wt that went on to make the Ad consensus tended to be present on the same reads. The arrow and arrowhead refer to some particular patterns described in the text.

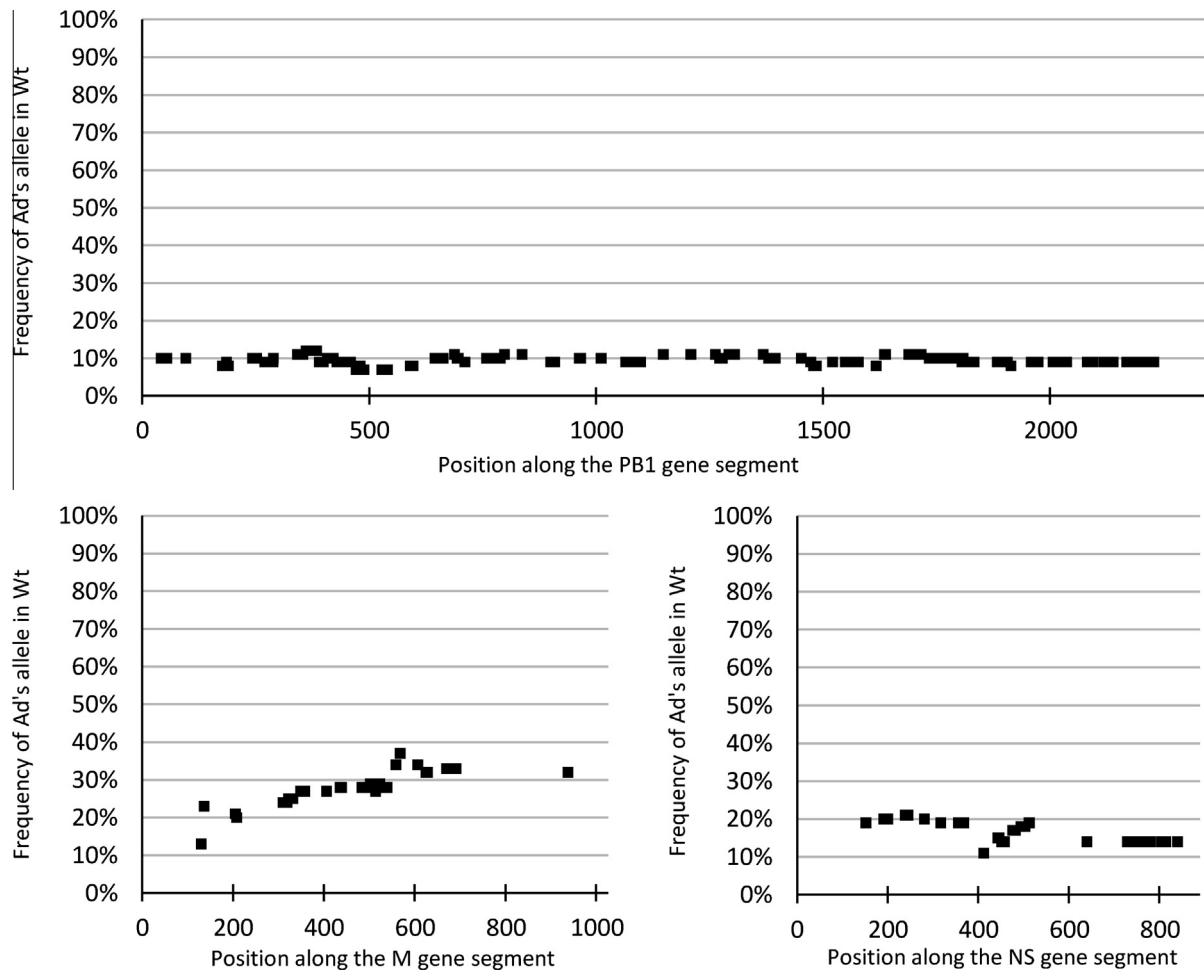


Fig. 4. Frequency of Ad's characteristic alleles in the Wt stock at every position of segments 2 (top), 7 (bottom left) and 8 (bottom right).

was undetected in Wt (0/1975 counts at that position). Likewise, the Ad NS gene had one minority SNP present at an 18% frequency which was absent from Wt (0/1164 counts) and therefore assumed to have appeared through *de novo* mutation. Their failure to become dominant even after 10 passages suggests that these spontaneous mutations did not confer a strong selective advantage, or that they appeared very late in the passaging process.

#### 3.4.2. Evolution of SNP in segments identical at the consensus level

**Segment 1.** At the 10% allele frequency threshold (a.f.t.), only one departure from the consensus was detected in Wt, with an allele present at 14%. This allele went down to 3% in Ad. Ad had only one minority allele present over 10% (19%), which may have been the result of a spontaneous mutation as it was undetected in Wt (0/1445 counts at this position). At the 1% a.f.t., there were 13 polymorphic sites in Wt and 21 polymorphic sites in Ad. These were not identified as defined minority alleles from known viruses and this diversity is therefore suspected to have originated from spontaneous mutation.

**Segment 5.** At the 1% a.f.t., there were 107 polymorphic sites in Wt (the majority of them at or above 10% making up the defined segment 5 minority sub-population described above), and there were 19 polymorphic sites in Ad at this threshold.

**Segment 6.** At the 10% a.f.t., there were two polymorphic sites in Wt. Those polymorphisms were located too far apart to tell reliably whether they were present on the same molecules, and went completely undetected in Ad. At the 1% a.f.t., there were 6 polymorphic

sites in Wt. We did not detect a clear pattern of association between these either.

#### 3.5. Identification of possible donor viruses for segments 2, 3, 4, 5, 7 and 8

As seen above, segments 2, 3, 4, 5, 7 and 8 of the Wt virus showed defined sub-populations. We therefore suspected that genes inherited from another virus or several other viruses might be present in the sample. In order to identify such donor viruses, Ad's alleles of segments 2, 7, and 8, suspected to be reassorted from such donor viruses, were searched against the BLAST database (Altschul et al., 1990).

**Segment 2.** The closest matches were A/mallard/Netherlands/15/1999 (H11N9) and A/northern shoveler/Netherlands/18/1999 (H11N9) which shared an identical segment 2. There was only one position differing between these and Ad's segment 2.

**Segment 7.** The closest matches (100% identity) were A/mallard/Netherlands/15/1999 and A/northern shoveler/Netherlands/18/1999 which shared an identical segment 7.

**Segment 8.** The closest matches (100% identity) were A/mallard/Netherlands/15/1999 and A/northern shoveler/Netherlands/18/1999 which shared an identical segment 8.

In the case of segments 3, 4 and 5, it was not possible to use Ad's genes to infer the origin of Wt's minority segments as these had disappeared in Ad. We therefore used the following strategies:

**Segment 4.** For that segment, we searched the consensus from three H11 clones from Wt against the BLAST database. The best



match was A/mallard/Netherlands/7/1999 (H11N9), with two differences including one position with a G + T mixture in that virus instead of a pure G in Wt's H11. Other close matches were A/mallard/Netherlands/15/1999 (3 differences including a mixture) and A/northern shoveler/Netherlands/18/1999 (3 differences).

In order to infer full-length minority haplotypes for Wt's segments 3 and 5, we assembled overlapping reads from Wt bearing multiple linked minority SNP, and searched these constructs against the BLAST database.

**Segment 3.** Wt's minority segment 3 had 100% identity with segment 3 of A/mallard/Netherlands/15/1999 and A/northern shoveler/Netherlands/18/1999.

**Segment 5.** Wt's minority segment 5 had two differences from segment 5 of A/northern shoveler/Netherlands/18/1999 and A/mallard/Netherlands/15/1999 (which are recorded as having a guanine at position 4 from the 5' end of their cDNA).

### 3.6. Inference of a scenario of viral evolution

Fig. 5 shows the inferred scenario of viral evolution from Wt to Ad upon passaging based on the interpretation of sequencing results. The "Wt" viral population was a mixture containing defined minority populations of segments 2, 3, 4, 5, 7 and 8 possibly inherited from co-circulating influenza strains as identified and present in different proportions as described above. Wt's minority versions of segments 3, 4 (H11) and 5 were lost upon NPTr adaptation, with only the majority segments remaining (and undergoing point mutation on segments 3 and 4). On the contrary, the minority segments 2, 7 and 8 were selected for and entirely replaced their majority counterparts in Ad. Segments 1 and 6 did not show well-defined sub-populations in Wt nor evolution at the consensus level during NPTr adaptation.

## 4. Discussion

### 4.1. Complete 454 coverage of the parental and adapted viruses

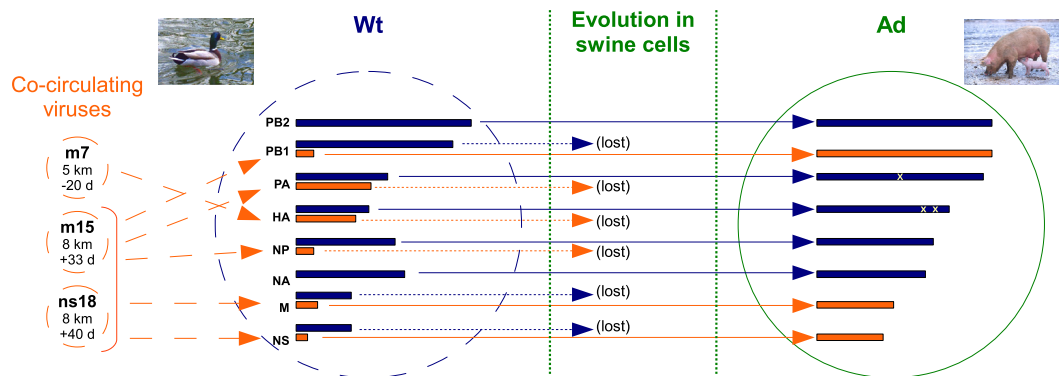
The complete genomes of both the parental (Wt) and passaged (Ad) virus stocks were analysed by pyrosequencing, with a coverage depth at any position generally comprising between several hundred and several thousand reads (Fig. 2). That coverage was however not completely uniform, with a tendency towards deeper coverage of a ~400 to ~500 bp region at the extremity of each

gene's cDNA. An average read is slightly over 400 bp in length, and reads originate from beads on to which gene fragments are attached before emulsion PCR. Therefore, it is suspected that fragmentation of genes prior to their attachment on to the beads may not have been optimal. This would have resulted in longer fragments or complete genes being attached to the beads and then read; as reads start close to the bead and stop after a few hundred bases, this could account for the observed greater coverage of the extremities of the segments. That deeper coverage could be noticed at both cDNA ends for most genes, and analogous conclusions have been drawn independently by others (e.g. Bainbridge et al., 2006). There was also a tendency towards even deeper coverage of the 3' than the 5' end of the cDNA, especially in the case of the NP genes which showed relatively shallow coverage of their 5' cDNA ends. The mechanism accounting for this difference between the two cDNA ends is uncertain, and might be related to sequence composition.

### 4.2. Analysis of well-defined subpopulations

The pyrosequencing approach allowed reliable identification of well-defined viral sub-populations within the Wt sample. There are several advantages to this approach compared with direct Sanger sequencing only. First, analysis of individual reads allows identification of minority polymorphisms with much less ambiguity than analysis of secondary peaks on Sanger chromatograms. Second, analysis of individual reads also allows visualisation of physical linkage between minority polymorphisms, enabling their assignment to defined minority sub-populations. Finally, the relative abundance of reads bearing minority polymorphisms is used as a quantitative estimate of the relative frequency of these polymorphisms in the viral pool.

Potential sources of bias can be considered with respect to these quantitative estimates of sub-population frequencies. First, although embryonated chicken eggs are generally considered a permissive system for amplification of most avian influenza viruses, some evolution of the respective variant proportions during amplification in eggs is inevitable. Furthermore, the downstream sample processing entails RNA extraction, reverse transcription, and two subsequent PCR (including the emulsion PCR) before the sequencing reaction itself. At each of these preliminary steps, there is a statistical chance of some of the rarest variants being lost if the efficiency is not 100%. As examples, the RNA extraction kit manufacturer mentions typical recovery rates over



**Fig. 5.** Reconstructed scenario of genetic evolution of the parental sample upon adaptation to NPTr cells. The genetic make-up of the parental Wt virus shows majority alleles (in blue) co-existing with minority alleles inherited from co-circulating viruses present in varying proportions (in orange). Reassortment between some minority and majority segments present within Wt, as well as point mutations, contributed to the final constellation of genes of the purer swine cell-tropic Ad virus as shown. Point mutations are indicated by ×. Potential donor viruses were: ns18, A/northern shoveler/Netherlands/18/1999 (H11N9); m7, A/mallard/Netherlands/7/1999 (H11N9); m15, A/mallard/Netherlands/15/1999 (H11N9). The distance and time difference between the sampling location and date of each of these viruses and those of A/mallard/Netherlands/10/1999 are given. Possible recombinations within Wt's segments 7 and 8 are not depicted here. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

90%. Also, agarose gel electrophoresis of subsequent PCR products suggests that the full-length PCR efficiency is not 100% for all genes, with, as a typical tendency, the shorter genes being more efficiently amplified than the longest ones. A previous study has sought to validate the correlation between relative 454 read abundance and taxa abundance within a mixed sample of fungi (Amend et al., 2010). A maximal 10-fold discrepancy suggested that read abundance was not a precise estimate of taxa abundance in that setting. However, that study was conducted on fungal cells from different genera that may well differ in their target rDNA gene copy number, with close to 10-fold variations in rDNA copy number reported within yeast species (Liti et al., 2009). This issue should not affect estimates of viral sub-population abundance in our context (unless some influenza virions bore multiple copies of some genes).

Quantitative frequency estimates also allowed suspicion of possible recombination events between different versions of Wt's segments 7 and 8 (Fig. 4, bottom, with deviations from a flat line indicating abrupt shifts in SNP frequency suggestive of recombination). However, artefactual chimeras originating during PCR could be a confounding factor (Görzer et al., 2010). Such chimeras are believed to appear mostly when incomplete extension events result in partially elongated segments that can act as mismatching primers for a subsequent amplification round (reviewed in Judo et al., 1998; Smyth et al., 2010). When the DNA polymerase and extension time are identical across segments, as is the case here, longer segments would therefore be expected to be more prone to incomplete extension, resulting in less efficient amplification and more chimera formation. Gel electrophoresis confirmed less efficient PCR amplification of the longer segment 2 compared to segments 7 and 8 (data not shown). However, the short segments 7 and 8 showed more SNP frequency shifts than the longer segment 2 (Fig. 4) despite benefiting from over twice as much relative elongation time, a factor known to reduce chimera formation (Meyerhans et al., 1990). On the other hand, the more efficient amplification of the shorter products could cause an amplification plateau to be reached earlier with primers occurring in limiting quantities, favouring non-specific template annealing to potential incompletely extended products (reviewed in Smyth et al., 2010). This would be consistent with Judo et al. (1998)'s views that chimeras mostly occurs late in PCR when products have accumulated to high concentrations with most chimeras produced by independent recombination events rather than reamplification. However it is unclear why most artefactual recombination events should then occur independently at only a few defined positions as suggested by Fig. 4.

It is possible that the recombinants observed here be genuine, however the natural occurrence of homologous recombination in influenza viruses has been debated (see for instance Gibbs et al., 2001 and Worobey et al., 2002), and we cannot definitely rule out our observations to be artefactual.

#### 4.3. Analysis of minority SNP

Aside from well-defined minority sub-populations, pyrosequencing also allows detection of point variants present at a very low level, with sensitivity being a function of sequencing depth.

Various limitations exist in such minority SNP analyses, with potential sources of artefactual mutations. The reverse transcription step, for instance, is carried out with an enzyme derived from the Moloney murine leukemia virus reverse transcriptase, which has been estimated to introduce one error every 37,000 bases on average (Ji and Loeb, 1992); that is statistically 0.37 error per complete influenza genome. The subsequent Phusion (Finnzymes)-mediated PCR, used to build the libraries prior to the 454 sequencing proper, is estimated to generate about  $4.2 \times 10^{-7}$  errors per

base pair per cycle (Li et al., 2006). Further, the pyrosequencing reaction is known to introduce frequent artefactual insertions or deletions (indels) in homopolymers (stretches of several identical nucleotides) where they represent the major source of errors in pyrosequencing data (reviewed in Kunin et al., 2009). Overall, the 454 sequencing reaction has been estimated to call an average of 0.53% erroneous bases over the first 101 bases of a read, and 1.07% erroneous bases on full-length reads 500–572 bp long (Gilles et al., 2011). Since our reads are on average 412.9–415.7 bp long, an average error rate close to 1% is expected in our dataset. At a position where the sample is in fact completely homogeneous and the coverage depth is, as an example, 2000 $\times$ , about 20 reads should therefore display polymorphism. However that was not the case in our dataset, with many completely homogeneous positions. This confirms suspicions of others (Gilles et al., 2011) that pyrosequencing errors are probably not homogeneously distributed, and makes it difficult to decide on a threshold above which minority SNP can be considered biologically significant. Caution is required when interpreting pyrosequencing singletons (*i.e.* polymorphisms detected only once) as it is suspected that most of them may be artefactual (Tedersoo et al., 2010) and these should therefore probably be disregarded in conservative interpretations.

There are other approaches to assess within-host genetic variation when looking at influenza virus populations, such as Sanger sequencing of a large number of cDNA clones (*e.g.* Iqbal et al., 2009; Murcia et al., 2010). This method can allow coverage of genome regions over 1000 bp in length with a depth of up to several hundred reads. However, such approaches need to be targeted as applying them to the full influenza genome would be extremely labour intensive. The pyrosequencing approach used here allowed a genome-wide, sensitive characterisation of sub-populations within the Wt and Ad samples, enabling a relatively fine reconstruction of the viral evolution scenario.

#### 4.4. Mixed infection in wild birds

The intention of this study was to characterise the genetic evolution of a duck influenza virus in porcine cells. The virus chosen had initially been identified as A/mallard/Netherlands/10/1999 (H1N8) but sequencing determined that our working stock (A/mallard/Netherlands/10-Nmkt/1999) was predominantly an H1N1 with no N8 but with a substantial proportion of H11. Using deep, full genome pyrosequencing we analysed the process at the sub-population level and could infer the underlying population changes that occurred. While this exposes the complexity of working with field isolates that have not been cloned by limiting dilution, it also highlights the versatility and opportunity for evolution that segmented genomes confer on influenza viruses.

Database (BLAST) searches suggested that Wt contained minority segments probably originating from close relatives of A/northern shoveler/Netherlands/18/1999 (H11N9), A/mallard/Netherlands/15/1999 (H11N9), and A/mallard/Netherlands/7/1999 (H11N9). These strains were sampled in a sympatric species or the same species as A/mallard/Netherlands/10/1999 (H1N8), within a few kilometres, and within a few weeks of each other. Ducks could therefore have been co-infected with two or more of those viruses. Due to reassortment events in the duck, the respective proportions of different minority segments in the Wt mix could have evolved differently from one another, accounting for the heterogeneity within the sample.

Though this report describes only one sample, such mixed infections are likely to be a common feature of influenza ecology in wild birds. This is exemplified by a study of 167 influenza samples from wild Anseriformes from various locations across the USA, which showed that at least 44 of these isolates (26%) represented mixed infections based on analysis of segments 4 and 6 alone

and without the sensitivity of deep sequencing (Dugan et al., 2008). Another study found co-infection of a bald eagle with both an H1 and an H2 as well as two different neuraminidase segments, and an H1, an H3 and an H4 all combined in a mallard cloacal swab (Ramakrishnan et al., 2009).

Such mixed infections can be considered in the context of viral quasispecies theory, where viral strains comprising many different closely related genomes respond as an ensemble to selective forces acting on them (reviewed in Domingo and Holland, 1997). In addition to SNP variants, quasispecies could also harbour substantially different gene segments forming distinct sub-populations. Subject to the constraints imposed by protein:protein compatibility and packaging specificity, segments may nevertheless operate individually as the units of evolution within individual hosts (and possibly even between hosts, if infective pressure or exposure is high). Distinct segments could be maintained as a sub-population within a complementing dominant population, undetected until subjected to bottle-necking, selective pressure – or deep sequencing analysis. Irrespective of whether or not quasispecies theory applies to influenza viruses in nature (Holmes, 2010), mixed infections undoubtedly play a key role in the ecology of influenza viruses. In the case of the Wt virus sample described here, co-infection most likely occurred within the natural reservoir. This provided the genetic diversity that was ultimately exploited to enable the adaptation to a new host system. Co-infection followed by reassortment may be a standard evolutionary strategy for these viruses in nature, enhancing their ability to conquer new ecological niches.

## Acknowledgements

This work was supported by a grant from DEFRA and HEFCE under the Veterinary Training and Research Initiative to the Cambridge Infectious Diseases Consortium (VB, LT), BBSRC grants BB/H014306/1 and BB/G00479X/1 (LT), and the French Ministry of Agriculture, INRA and the French Région Midi-Pyrénées (GC, J-LG, VB).

We are grateful to Jon Lyall for advice at the early stages of the project. We also thank Ron Fouchier and his team at the Erasmus Medical Centre, Rotterdam, The Netherlands, who provided the parental influenza virus strain, Pascal Lexmond and Vincent Munster who sent the samples, and Josanne Verhagen who facilitated information on sampling of the Netherlands/1999 viruses. We are also grateful to Adam Rash, Carley Jervis and Debra Elton at the Animal Health Trust, Newmarket, UK, who helped grow the isolate in eggs, to Helen Wise, Department of Pathology, University of Cambridge who donated the RF483 plasmid, to Sophie Ruzafa and Jérôme Luch, Toulouse Genomics platform, France, who helped with 454 pyrosequencing, and to Etienne Liats, Ecole Nationale Vétérinaire de Toulouse, for help with PCR. The duck picture in Fig. 5 was modified from an original courtesy of Mathilde Bourret; the swine picture is by Scott Bauer (public domain) from [http://en.wikipedia.org/wiki/File:Sow\\_with\\_piglet.jpg](http://en.wikipedia.org/wiki/File:Sow_with_piglet.jpg).

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.meegid.2013.04.034>.

## References

- Alexander, D.J., 2000. A review of avian influenza in different bird species. *Vet. Microbiol.* 74 (1–2), 3–13.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215 (3), 403–410.
- Amend, A.S., Seifert, K.A., Bruns, T.D., 2010. Quantifying microbial communities with 454 pyrosequencing: does read abundance count? *Mol. Ecol.* 19 (24), 5555–5565.
- Bainbridge, M.N., Warren, R.L., Hirst, M., Romanuk, T., Zeng, T., Go, A., Delaney, A., Griffith, M., Hickenbotham, M., Magrini, V., Mardis, E.R., Sadar, M.D., Siddiqui, A.S., Marra, M.A., Jones, S.J., 2006. Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. *BMC Genomics* 7, 246.
- Castrucci, M.R., Donatelli, I., Sidoli, L., Barigazzi, G., Kawaoka, Y., Webster, R.G., 1993. Genetic reassortment between avian and human influenza A viruses in Italian pigs. *Virology* 193 (1), 503–506.
- Croville, G., Soubies, S.M., Barbieri, J., Klopp, C., Mariette, J., Bouchez, O., Camus-Bouclainville, C., Guérin, J.-L., 2012. Field monitoring of avian influenza viruses: whole-genome sequencing and tracking of neuraminidase evolution using 454 pyrosequencing. *J. Clin. Microbiol.* 50 (9), 2881–2887.
- Domingo, E., Holland, J.J., 1997. RNA virus mutations and fitness for survival. *Ann. Rev. Microbiol.* 51 (1), 151–178.
- Drake, J.W., Holland, J.J., 1999. Mutation rates among RNA viruses. *Proc. Natl. Acad. Sci. U.S.A.* 96 (24), 13910–13913.
- Drake, J.W., 1993. Rates of spontaneous mutation among RNA viruses. *Proc. Natl. Acad. Sci. U.S.A.* 90 (9), 4171–4175.
- Dugan, V.G., Chen, R., Spiro, D.J., Sengamalay, N., Zaborsky, J., Ghedin, E., Nolting, J., Swayne, D.E., Runstadler, J.A., Hupp, G.M., Senne, D.A., Wang, R., Slemons, R.D., Holmes, E.C., Taubenberger, J.K., 2008. The evolutionary genetics and emergence of avian influenza viruses in wild birds. *PLoS Pathog.* 4 (5).
- Ferrari, M., Scalvini, A., Losio, M.N., Corradi, A., Soncini, M., Bignotti, E., Milanese, E., Ajmone-Marsan, P., Barlati, S., Bellotti, D., Tonelli, M., 2003. Establishment and characterization of two new pig cell lines for use in virological diagnostic laboratories. *J. Virol. Methods* 107 (2), 205–212.
- Forrest, H.L., Webster, R.G., 2010. Perspectives on influenza evolution and the role of research. *Anim. Health Res. Rev.* 11 (Special Issue 01), 3–18.
- Gibbs, M.J., Armstrong, J.S., Gibbs, A.J., 2001. Recombination in the hemagglutinin gene of the 1918 « Spanish Flu ». *Science* 293 (5536), 1842–1845.
- Gilles, A., Megléc, E., Pech, N., Ferreira, S., Malausa, T., Martin, J.-F., 2011. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics* 12 (1), 245.
- Görzer, I., Guelly, C., Trajanoski, S., Puchhammer-Stöckl, E., 2010. The impact of PCR-generated recombination on diversity estimation of mixed viral populations by deep sequencing. *J. Virol. Methods* 169 (1), 248–252.
- Hinshaw, V.S., Air, G.M., Gibbs, A.J., Graves, L., Prescott, B., Karunakaran, D., 1982. Antigenic and genetic characterization of a novel hemagglutinin subtype of influenza A viruses from gulls. *J. Virol.* 42 (3), 865–872.
- Hoffmann, E., Neumann, G., Kawaoka, Y., Hobom, G., Webster, R.G., 2000. A DNA transfection system for generation of influenza A virus from eight plasmids. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6108–6113.
- Holmes, E.C., 2010. The RNA virus quasispecies: fact or fiction? *J. Mol. Biol.* 400 (3), 271–273.
- Horimoto, T., Kawaoka, Y., 2001. Pandemic threat posed by Avian influenza A viruses. *Clin. Microbiol. Rev.* 14 (1), 129–149.
- Horimoto, T., Kawaoka, Y., 2005. Influenza: lessons from past pandemics, warnings from current incidents. *Nat. Rev. Microbiol.* 3 (8), 591–600.
- Iqbal, M., Xiao, H., Baillie, G., Warry, A., Essen, S.C., Londt, B., Brookes, S.M., Brown, I.H., McCauley, J.W., 2009. Within-host variation of avian influenza viruses. *Phil. Trans. R. Soc. B.* 364 (1530), 2739–2747.
- Ji, J., Loeb, L.A., 1992. Fidelity of HIV-1 reverse transcriptase copying RNA in vitro. *Biochemistry* 31 (4), 954–958.
- Judo, M.S., Wedel, A.B., Wilson, C., 1998. Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Res.* 26 (7), 1819–1825.
- Kunin, V., Engelbrekton, A., Ochman, H., Hugenholtz, P., 2009. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ. Microbiol.* 12 (1), 118–123.
- Kuntz-Simon, G., Kaiser, C., Madec, F., 2010. Swine influenza. In: Lefèvre, P.-C., Blancou, J., Chermette, R., Uilenberg, G. (Eds.), *Infectious and Parasitic Diseases of Livestock*. Lavoisier, Paris, pp. 273–285.
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25 (14), 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079.
- Li, M., Diehl, F., Dressman, D., Vogelstein, B., Kinzler, K.W., 2006. BEAMing up for detection and quantification of rare sequence variants. *Nat. Methods* 3 (2), 95–97.
- Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., Davey, R.P., Roberts, I.N., Burt, A., Koufopanou, V., Tsai, I.J., Bergman, C.M., Bensasson, D., O’Kelly, M.J.T., van Oudenaarden, A., Barton, D.B.H., Bales, E., Nguyen, A.N., Jones, M., Quail, M.A., Goodhead, I., Sims, S., Smith, F., Blomberg, A., Durbin, R., Louis, E.J., 2009. Population genomics of domestic and wild yeasts. *Nature* 458 (7236), 337–341.
- Mariette, J., Noirot, C., Klopp, C., 2011. Assessment of replicate bias in 454 pyrosequencing and a multi-purpose read-filtering tool. *BMC Res. Notes* 4 (1), 149.
- Meyerhans, A., Vartanian, J., Wainhobson, S., 1990. DNA recombination during PCR. *Nucleic Acids Res.* 18 (7), 1687–1691.
- Murcia, P.R., Baillie, G.J., Daly, J., Elton, D., Jervis, C., Mumford, J.A., Newton, R., Parrish, C.R., Hoelzer, K., Dougan, G., Parkhill, J., Lennard, N., Ormond, D., Moule, S., Whitwham, A., McCauley, J.W., McKinley, T.J., Holmes, E.C., Grenfell, B.T.,

- Wood, J., Intra- and interhost evolutionary dynamics of equine influenza virus. *J. Virol.* 84 (14), 6943–6954.
- Murcia, P.R., Hughes, J., Battista, P., Lloyd, L., Baillie, G.J., Ramirez-Gonzalez, R.H., Ormond, D., Oliver, K., Elton, D., Mumford, J.A., Caccamo, M., Kellam, P., Grenfell, B.T., Holmes, E.C., Wood, J.L.N., 2012. Evolution of an eurasian avian-like influenza virus in naïve and vaccinated pigs. *PLoS Pathog.* 8 (5), e1002730.
- Palese, P., 1977. The genes of influenza virus. *Cell* 10 (1), 1–10.
- Ramakrishnan, M.A., Tu, Z.J., Singh, S., Chockalingam, A.K., Gramer, M.R., Wang, P., Goyal, S.M., Yang, M., Halvorson, D.A., Sreevatsan, S., 2009. The feasibility of using high resolution genome sequencing of influenza A viruses to detect mixed infections and quasispecies. *PLoS ONE* 4 (9), e7105.
- Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 16 (6), 276–277.
- Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. *Nat. Biotech.* 29 (1), 24–26.
- Sanjuán, R., Nebot, M.R., Chirico, N., Mansky, L.M., Belshaw, R., 2010. Viral mutation rates. *J. Virol.* 84 (19), 9733–9748.
- Scholtissek, C., Bürger, H., Kistner, O., Shortridge, K.F., 1985. The nucleoprotein as a possible major factor in determining host specificity of influenza H3N2 viruses. *Virology* 147 (2), 287–294.
- Smyth, R.P., Schlub, T.E., Grimm, A., Venturi, V., Chopra, A., Mallal, S., Davenport, M.P., Mak, J., 2010. Reducing chimera formation during PCR amplification to ensure accurate genotyping. *Gene* 469 (1–2), 45–51.
- Suárez-López, P., Ortín, J., 1994. An estimation of the nucleotide substitution rate at defined positions in the influenza virus haemagglutinin gene. *J. Gen. Virol.* 75 (Pt. 2), 389–393.
- Tedersoo, L., Nilsson, R.H., Abarenkov, K., Jairus, T., Sadam, A., Saar, I., Bahram, M., Bechem, E., Chuyong, G., Kõljalg, U., 2010. 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytol.* 188 (1), 291–301.
- Webster, R.G., Bean, W.J., Gorman, O.T., Chambers, T.M., Kawaoka, Y., 1992. Evolution and ecology of influenza A viruses. *Microbiol. Rev.* 56 (1), 152–179.
- WHO, 2009. WHO|Influenza (Seasonal). <<http://www.who.int/mediacentre/factsheets/fs211/en/>> (accessed 13.07.11.).
- Wise, H.M., Hutchinson, E.C., Jagger, B.W., Stuart, A.D., Kang, Z.H., Robb, N., Schwartzman, L.M., Kash, J.C., Fodor, E., Firth, A.E., Gog, J.R., Taubenberger, J.K., Digard, P., 2012. Identification of a novel splice variant form of the influenza A virus M2 ion channel with an antigenically distinct ectodomain. *PLoS Pathog.* 8, e1002998.
- Worobey, M., Rambaut, A., Pybus, O.G., Robertson, D.L., 2002. Questioning the evidence for genetic recombination in the 1918 « Spanish Flu » virus. *Science* 296 (5566), 211–211.
- Zagordi, O., Däumer, M., Beisel, C., Beerwinkler, N., 2012. Read length versus depth of coverage for viral quasispecies reconstruction. *PLoS ONE* 7 (10), e47046.

**Annexe 1** Comparaison des séquences nucléotidiques et d'acides aminés entre le virus A/Muscovy duck/France/ 09010/09(H6N1) et les virus

Segments	Protéines	A/turkey/France/09010-1/09(H6N1)			A/turkey/France/09010-2/09(H6N1)		
		Identité nucleotidique (%)	Modifications d'acides aminés (aa) et pourcentages d'identité	aa modifiés	Identité nucleotidique (%)	Modifications d'acides aminés (aa) et pourcentages d'identité	aa modifiés
1	PB2	2005/2280 (87,9%)	751/759 (98.9%)	R61K, T147I, V292I, V295I V397A, I511V, T585P, D680G	2005/2280 (87,9%)	751/759 (98.9%)	R61K, T147I, V292I, V295I V397A, I511V, T585P, D680G
2	PB1	2190/2274 (96.3%)	750/757 (99%)	K52N, R54K, A56T, M195I, L550I, R584H, N694S	2190/2274 (96.3%)	750/757 (99%)	K52N, R54K, A56T, M195I, L550I, R584H, N694S
	PB1-F2	258/273 (94.5%)	76/90 (84.4%)	E4G, R21M, G22E, G23S, Q26R, L30P, T46M, Q48R, A49V, K53R, Q54R, K65R, Q79R, S82L	258/273 (94.5%)	76/90 (84.4%)	E4G, R21M, G22E, G23S, Q26R, L30P, T46M, Q48R, A49V, K53R, Q54R, K65R, Q79R, S82L
3	PA	2059/2151 (95.7%)	714/716 (99.7%)	A70V, E538G	2059/2151 (95.7%)	714/716 (99.7%)	A70V, E538G
4	HA	1698/1701 (99.8%)	565/566 (99,8%)	K118R	1698/1701 (99.8%)	565/566 (99,8%)	K118R
5	NP	1497/1497 (100%)	498/498 (100%)	-	1497/1497 (100%)	498/498 (100%)	-
6	NA	1350/1350 (excluding deletion)	449/449 (100%)	N.B. : Les aa 55 à 74 sont absents des prélèvements de dinde (région délétee)	1350/1350 (région délétee non prise en compte)	449/449 (100%)	N.B.: Les aa 55 à 74 sont absents des prélèvements de dinde (région délétee)
					1341/1341 ( région délétee non prise en compte )	446/446 (100%)	N.B.: Les aa 43 à 66 sont absents des prélèvements de dinde (région délétee)
7	M1	759/759 (100%)	252/252 (100%)	-	759/759 (100%)	252/252 (100%)	-
	M2	294/294 (100%)	97/97 (100%)	-	294/294 (100%)	97/97 (100%)	-
8	NS1	693/693 (100%)	230/230 (100%)	-	693/693 (100%)	230/230 (100%)	-
	NEP	366/366 (100%)	121/121 (100%)	-	366/366 (100%)	121/121 (100%)	-

**Annexe 2** Conditions de PCR pour l'amplification de chaque segment d'ARNv.

Segment	Amorces de PCR	Température d'hybridation	Durée d'élongation
<b>PB2</b>	<b>PB2F:</b> AGCGAAAGCAGGTCAAATA <b>PB2R:</b> AGTAGAAACAAGGTCGTTTTTAAAC	Programme « Touchdown » : 67 °C à 55 °C	2.5 min
<b>PB1</b>	<b>PB1F:</b> AGCGAAAGCAGGCAAAC <b>PB1R:</b> AGTAGAAACAAGGCATTTTTTCA	Programme « Touchdown » : 67 °C à 55 °C	2.5 min
<b>PA</b>	<b>PAF:</b> AGCGAAAGCAGGTACTGAT <b>PAR:</b> AGTAGAAACAAGGTACTTTTTTGGGA	Programme « Touchdown » : 67 °C à 55 °C	2.5 min
<b>HA</b> (Tous sous-types)	<b>BsaIseg4F:</b> CACACACAGGTCTCAGGGGAGCAAAAGCAGGGGAAAA <b>BsaIseg4R:</b> CACACACAGGTCTCATATTAGTAGAAACAAGGGTGTTTTTCC	68 °C	2 min
<b>HA</b> (H1)	<b>HAIF:</b> AGCAAAAGCAGGGGAA <b>HAIR:</b> AGTAGAAACAAGGGTGTTTTTC	57 °C	2 min
<b>HA</b> (Ad virus)	<b>HAF:</b> AGCAAAAGCAGGGGAAAA <b>HAR:</b> AGTAGAAACAAGGGTGTTTTTCC	59 °C	2 min
<b>NP</b>	<b>NPF:</b> AGCAAAAGCAGGGTAGATAATC <b>NPR:</b> AGTAGAAACAAGGGTATTTTTCTTTA	54 °C	2 min
<b>NA</b>	<b>NAF:</b> AGCAAAAGCAGGAGTTTA <b>NAR:</b> AGTAGAAACAAGGAGTTTTTT	52 °C	1.5 min
<b>M</b>	<b>MF:</b> AGCAAAAGCAGGTAGATATTGAA <b>MR:</b> AGTAGAAACAAGGTAGTTTTTTACTCCA	56 °C	1.5 min
<b>NS</b>	<b>NSF:</b> AGCAAAAGCAGGGTGACA <b>NSR:</b> AGTAGAAACAAGGGTGTTTTTTATCA	57 °C	1 min

**Annexe 3** Amorces internes utilisées pour le séquençage des génomes totaux du virus parental A/mallard/Netherlands/10/99 et du virus Ad adapté aux cellules NPTr. \* : amorce anti-sens.

Segment	Internal primers
<b>PB2</b>	<b>PB21</b> : AGGAATGGACCAACGACAAG
	<b>PB22b</b> : TAGCACTGAGGTCTGCATGG*
	<b>PB23</b> : GGGGAGGTGAGAAATGATGA
	<b>PB24</b> : ATGCACCAACTCCTGAGACAC
	<b>PB25</b> : TTGAGCCTTTCCAATCCTTG
	<b>PB26</b> : CAGTCAGACAGCGACCAAAA
<b>PB1</b>	<b>PB11</b> : GCCATGGAACAGGAACAGGAT
	<b>PB12</b> : GCTTGGCTCGTTATCCTCAG*
	<b>PB13</b> : CCAATGAATCGGGGAGACTA
	<b>PB14</b> : CGTTGGAGGGAATGAGAAGA
	<b>PB15</b> : GGGCAAAAGAGGTACACCAA
	<b>PB16</b> : CCATTCAAAGGCAGGACTGT
	<b>PB17</b> : CCGTAAACAATGCTGTGGTG
<b>PA</b>	<b>PA1</b> : GCTGCGATATGCACACACTT
	<b>PA2</b> : CGGGGAGAAATTTAGGCTTC*
	<b>PA3</b> : TGGGATTCCTTTCGTCAGTC
	<b>PA4</b> : GAAGACATTTTTTCGGCTGGA
	<b>PA5</b> : ACAGCGGAAGTATCCCATTG
	<b>PA6</b> : AGGCCGAGTCTTCTGTCAAA
<b>HA (H1 subtype)</b>	<b>AHA1</b> : CTCAGTTAATTTACTCGAAGACA
	<b>AHA2</b> : GGTAGGATGAATTATTATTGGAC
	<b>AHA3</b> : GCTGATCAGAAGAGCACACAAAATGC
	<b>AHA4</b> : TCTGGGTTACCAAGGAGCCA*
<b>HA (H11 subtype)</b>	<b>H111</b> : GCAAGTTCGAAGCATTCAACA
	<b>H112</b> : GCTCTACCAAATGCCAAACA
<b>NP</b>	<b>NP1</b> : GGGGAAGGATCCGAAGAAAAC
	<b>NP2</b> : AGCAGTTGCGTCTTCTCCAT*
	<b>NP3</b> : TTCCAAACAGCAGCACAAAG
	<b>NP4</b> : AACCAGGAGTGGAGGAAACA
<b>NA</b>	<b>NA1b</b> : TGAGCAAGCTGTGGTTTCAG
	<b>NA2b</b> : ATCCCCTTTGGAACCAATTC*
	<b>NA3b</b> : GCAAGTGCTTGTGATGATGG
	<b>NA4b</b> : GTTGCGGTCCAGTGTCTTCT
<b>M</b>	<b>M1</b> : CTTTGTCCAAAATGCCCTA
	<b>M2</b> : CCAAAGCCACTTCTGTGGT*
	<b>M3</b> : GTGCCGGTCTGAAAGATGAT
<b>NS</b>	<b>NS1</b> : CAGCACTCTTGGTCTGGACA
	<b>NS2</b> : CGGCACTGAAGCAATAGTCA*
	<b>NS3</b> : GTCCTCATCGGAGGACTTGA

ÉCOLE PRATIQUE DES HAUTES ÉTUDES  
SCIENCES DE LA VIE ET DE LA TERRE

Analyse de la variabilité génétique des virus influenza aviaires par séquençage à très haut débit

Croville Guillaume

Date de soutenance

RÉSUMÉ

Les virus influenza A sont des virus à ARN segmenté et présentent une importante variabilité génétique. Les virus influenza aviaires (VIA) peuvent être transmis depuis leur réservoir (les oiseaux aquatiques sauvages) vers les volailles domestiques ou les mammifères chez qui ils peuvent causer des épizooties avec un fort potentiel de pertes économiques et une menace pour la santé humaine.

Le suivi virologique des VIA lors d'épisodes infectieux est primordial pour la détection précoce de modifications génétiques. Le développement de nouvelles générations de séquençage (NGS) permet d'effectuer un suivi de l'évolution génétique virale de façon pratique et accessible.

Ce document décrit l'application du pyroséquençage 454 de Roche (i) à l'analyse de l'adaptation d'un VIA de sous-type H6N1 d'un élevage de canards vers un élevage de dindes lors d'un épisode infectieux de terrain et (ii) au suivi de l'adaptation d'un VIA de sous-type H1N1 à des cellules porcines.

Dans le contexte de l'évolution du VIA de sous-type H6N1, une troncation du gène de la neuraminidase (NA) a été mise en évidence par pyroséquençage dans moins de 2 % des échantillons de canards alors qu'elle était présente dans 100 % des échantillons de dindes. Cela suggère que la délétion de la NA permet l'adaptation des VIA aux volailles et peut émerger suite à une modification de la pression de sélection.

Lors du suivi de l'adaptation d'un VIA de sous-type H1N1 à des cellules de trachée de porcelet, les génomes complets des virus parental et adapté ont été analysés par pyroséquençage. La couverture et la profondeur de séquençage ont permis la caractérisation de populations virales minoritaires dans l'échantillon parental qui ont émergé après l'adaptation de l'échantillon viral sur cellules porcines.

Ces deux études démontrent à quel point le séquençage à très haut débit met en évidence les fortes capacités d'évolution des VIA qui leur permettent de s'adapter à de nouvelles niches environnementales.

**MOTS-CLÉS** : virus influenza, pyroséquençage, aviaire, évolution génétique, adaptation, émergence.