Infection, Genetics and Evolution 18 (2013) 31–41

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

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

Vincent Bourret [a,b,c,*], Guillaume Croville [b,c], Jérôme Mariette [d], Christophe Klopp [d], Olivier Bouchez [e], Laurence Tiley [a], Jean-Luc Guérin [b,c]

[a] Department of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge CB3 0ES, UK
[b] Université de Toulouse, INP, ENVT, 31076 Toulouse, France
[c] INRA, UMR 1225, IHAP, 31076 Toulouse, France
[d] Plateforme bioinformatique Toulouse Midi-Pyrénées, UBIA, INRA Auzeville, 31326 Castanet-Tolosan, France
[e] 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.

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 9×10^{-6} to over 7×10^{-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, fœtal 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, vjrb2@cam.ac.uk (V. Bourret).

1567-1348/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2013.04.034

(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_2 in 25 cm^2 flasks seeded with 2×10^6 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) fœtal calf serum (FCS). On the next day, cells were washed with phosphate-buffered saline and infected with 1 mL of inoculum consisting of 2.4×10^4 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 2.5×10^4 dilution of the viral yield from the previous passage (corresponding to 3.8×10^5 – 5.0×10^6 viral genome copies). The procedure was repeated 10 consecutive times. The resulting viral stock, harvested from the 10^{th} 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 $MgSO_4$, 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 BsaIseg4 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 μL of 5× ligase buffer, 5 units of ligase T4 (Invitrogen), and water up to 10 μL, and incubated at room temperature for 1 h. Ligation products were transformed into DH5α competent cells (Invitrogen) as per manufacturer's instructions with the following modification: heat shock was 37 °C for 30 s, and grown on ampicillin agar Luria–Bertani medium (LB). After an overnight incubation of the plates at 37 °C, a 3 mL liquid LB broth culture with 100 μg/mL of ampicillin was inoculated with a single colony and incubated overnight at 37 °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 (Mariette et al., 2011) considering different criteria such as length (with reads shorter or longer than mean read length ±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 value 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 (NPTr) 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

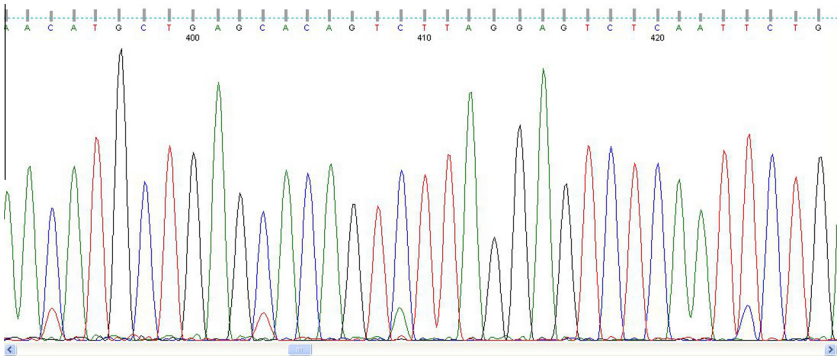

Fig. 1. Section of a chromatogram from the Wt virus' segment 2 illustrating detection of secondary peaks by direct Sanger sequencing (positions 394, 403, 409 and 424 of this read).

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×) 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× 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% (±1.1% standard deviation, SD) in segment 2, 28% (±4.8% SD) in segment 7, and 17% (±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×, while the average coverage depth for the H11 allele was 682×. 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% ± 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

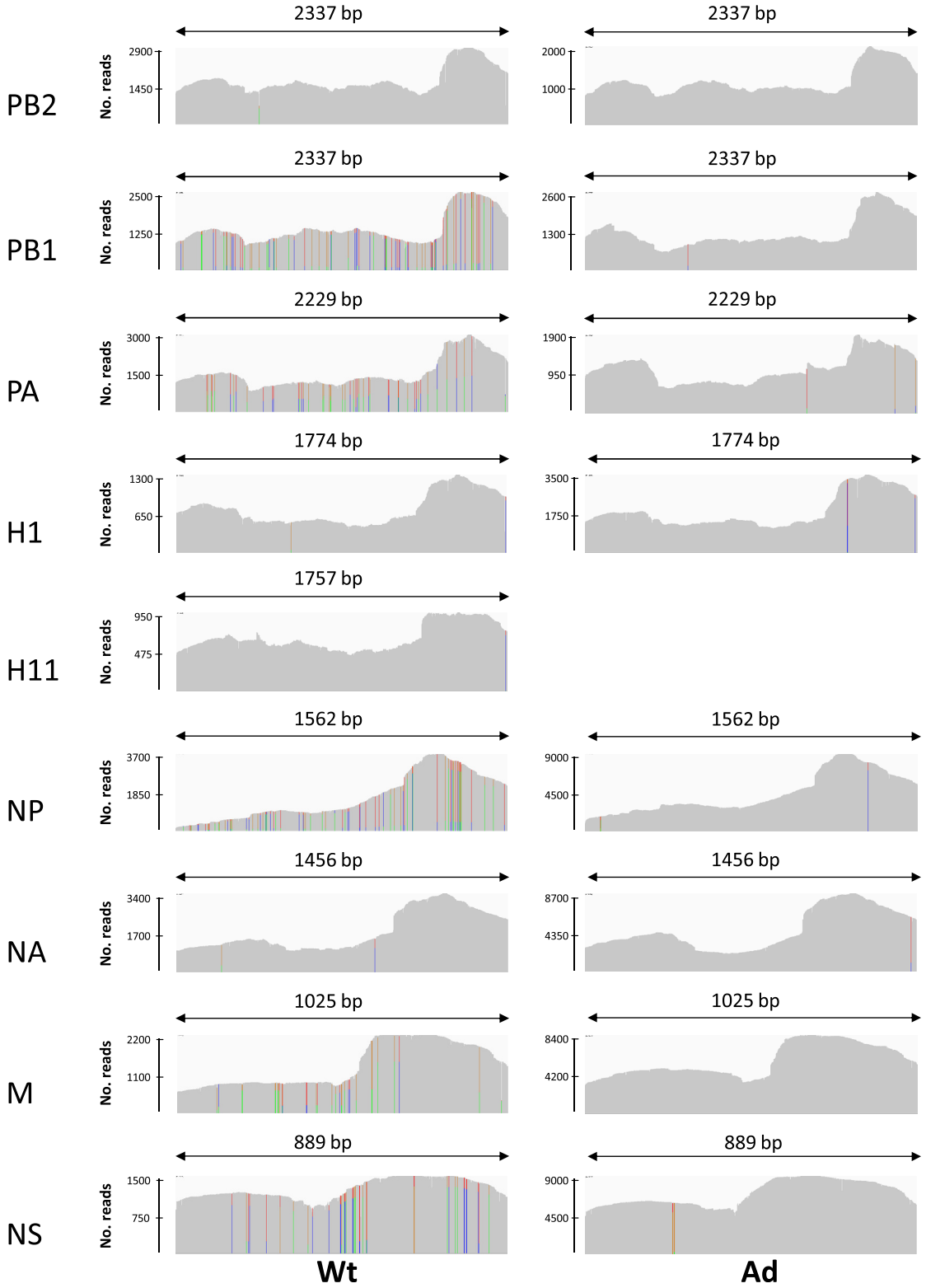

Fig. 2. 454 coverage levels and variability at the 5% allele frequency threshold (a.f.t.) within each gene for viruses Wt and Ad (left and right columns respectively). The H1 and the H11 alleles of Wt are shown (no H11 allele was detected in Ad). Grey indicates no departure from the consensus within 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

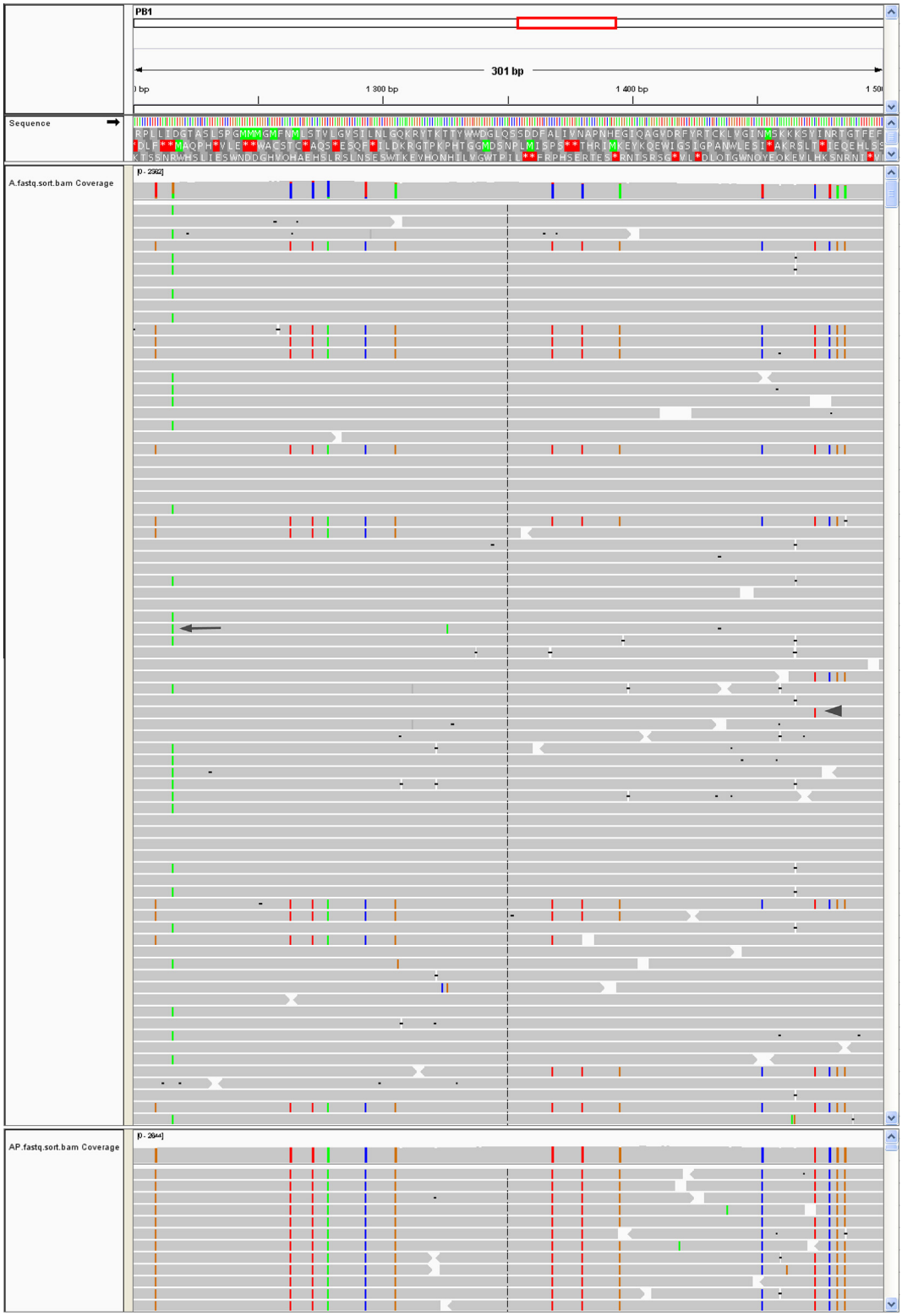

Fig. 3. Image from the 454 analysis of the PB1 genes of viruses Wt and Ad. A 301-bp region of the gene is shown (positions 1200–1500). The display consists of three tracks aligned horizontally below the top panel which indicates the position of the focus region within the whole gene. (i) Track 1 (labelled "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.

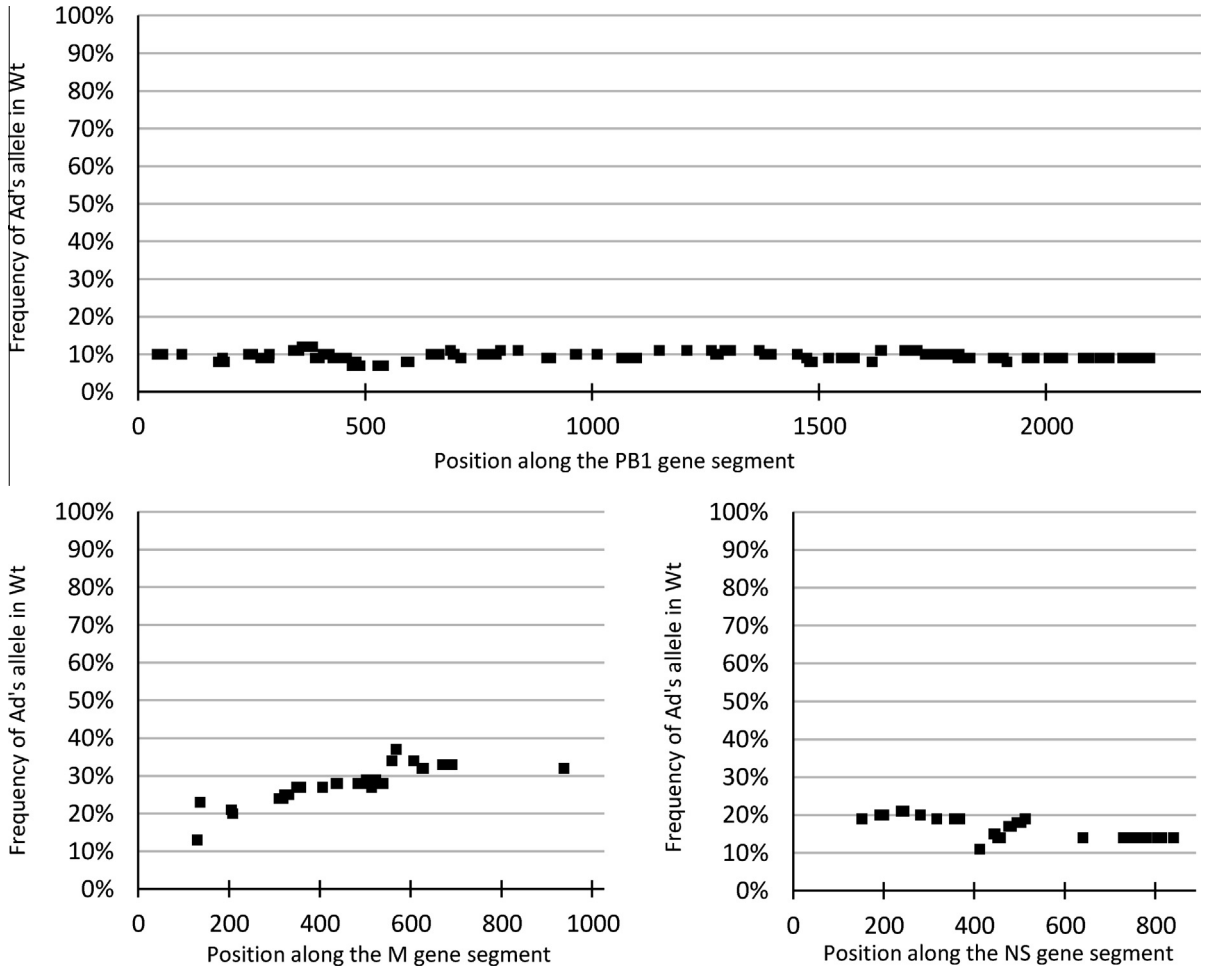

Fig. 4. Frequency of Ad's characteristic alleles in the Wt stock at every position of segments 2 (top), 7 (bottom left) and 8 (bottom right).

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 comprised 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

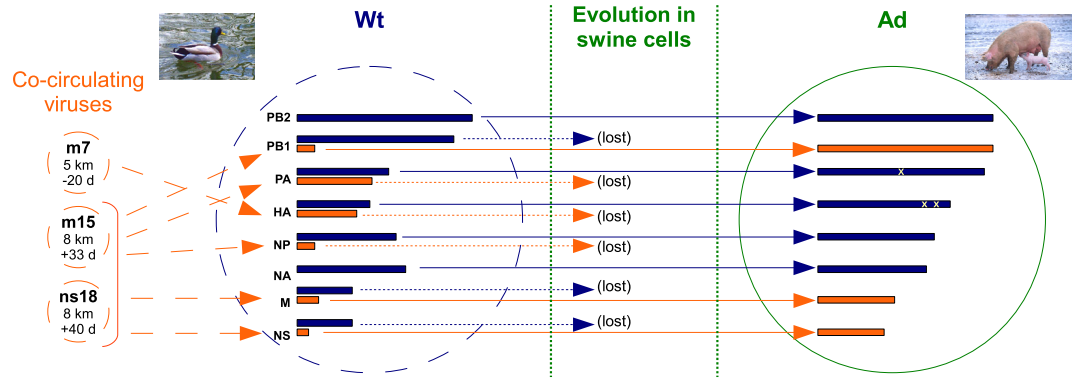

Fig. 5. Reconstructed scenario of genetic evolution of the parental sample upon adaptation to NPTr cells. The genetic make-up of the parental Wt virus shows majority alleles (in blue) co-existing with minority alleles inherited from co-circulating viruses present in varying proportions (in orange). Reassortment 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 4.2×10^{-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×, 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 Lluch, Toulouse Genomics platform, France, who helped with 454 pyrosequencing, and to Etienne Liais, 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.

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., Romanuik, 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., Happ, 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., Milanesi, 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., Meglécz, 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., Engelbrektson, 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., Bailes, 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, . 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., Beerenwinkel, N., 2012. Read length versus depth of coverage for viral quasispecies reconstruction. PLoS ONE 7 (10), e47046.