Exploring the use of influenza virus sequence diversity

for the identification and characterization of transmission events

Marcel Jonges

Exploring the use of influenza virus sequence diversity

for the identification and characterization of transmission events

Marcel Jonges

ISBN: 978-90-6464-903-5

The research described in this thesis was conducted at the National Institute for Public Health and the Environment (RIVM), the Netherlands, and was funded by the Dutch Ministry of Economic Affairs (Castellum Project and Impulse Veterinary Avian Influenza Research in the Netherlands program of the Dutch Economic Structure Enhancement Fund)

Cover design and layout: M. Jonges
Printed by GVO drukkers & vormgevers B.V. | Ponsen & Looijen, Ede

Exploring the Use of Influenza Virus Sequence Diversity

for the identification and characterization of transmission events

Verkenning van het gebruik van influenza virus sequentie diversiteit

ten behoeve van de identificatie en karakterisering van transmissie gevallen

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
donderdag 8 oktober 2015 om 13:30 uur

door

Marcel Jonges

geboren te Voorburg

Erasmus University Rotterdam

Promotiecommissie

Promotor:	Prof.dr. M.P.G. Koopmans
Overige leden:	Prof.dr. R.A.M. Fouchier Prof.dr. A.D.M.E. Osterhaus Prof.dr. J.H. Richardus
Copromotor:	Dr. A. Meijer

Table of contents

Chapter 1

Introduction and scope of this thesis

Background on influenza viruses

Influenza A, B and C viruses are enveloped RNA viruses, classified as members of the family of *Orthomyxoviridae*. Influenza A viruses can be further divided into subtypes based on the surface glycoproteins haemagglutinin (HA) and neuraminidase (NA). Currently, 18 HA (H1-H18) and 11 NA (N1-N11) surface glycoproteins are known. All influenza A virus subtypes, except for the newly discovered bat influenza A viruses subtypes H17N10 and H18N11 [1,2], circulate in their natural reservoir of wild aquatic birds [3-5]. Although avian influenza A viruses typically cause mild and unspecific symptoms in aquatic birds, they are classified as either low pathogenic avian influenza (LPAI) or highly pathogenic avian influenza (HPAI) depending on their pathogenicity in chickens [6]. LPAI viruses can contain any HA, while HPAI viruses (with >75% mortality rates in naïve chickens) contain so far H5 or H7 glycoproteins [7,8]. Avian influenza viruses may cross species barriers and establish stable lineages in mammalian species, including humans, through adaptive evolutionary processes [9]. Introduction of LPAI viruses from birds into mammals has resulted in the establishment of equine and multiple swine influenza virus lineages [10]. Fortunately, avian influenza viruses of subtypes H5 and H7 have not adapted to mammalian species [11,12].

Seasonal influenza viruses

Co-circulating influenza A viruses of subtypes H1N1, H3N2 and influenza B viruses are accountable for annual epidemics that affect 7-20% of the human population. Most seasonal influenza virus infected individuals recover within two weeks without medical treatment, but infection may also lead to severe illness and death, especially in elderly people and very young children [13]. Although immunological memory is induced during influenza virus infection, reinfection may still cause disease again 3-5 years later [14]. Among host factors, this is caused by antigenic drift of the virus, i.e. the accumulation of amino acid substitutions in the antibody bindingsites of the HA and NA which cause evasion from antibodies. This antigenic drift necessitates the annual update of vaccines, which are based on both HA and NA.

Vaccines

The continued monitoring of seasonal influenza viruses by the WHO Global Influenza Surveillance and Response System, by analyzing influenza virus specimens obtained at National Influenza Centers and Reference Laboratories, provides both phenotypic and genotypic data on spreading and circulating influenza viruses. Twice a year, the surveillance data and candidate vaccine viruses are reviewed by a WHO committee in anticipation of the next epidemic [15], and vaccine manufacturers can start their 6-month production period to deliver vaccines to health care providers in time. Using this vaccine production process, in which vaccine viruses are grown in the allantoic cavity of embryonated eggs, it is impossible to respond timely to sudden antigenic drift

variants or the introduction of a non-human influenza virus in the human population [16]. Although vaccines are generally provided to the people who are at high risk of serious complications from influenza virus infection, transmission on population level still is responsible for substantial morbidity and mortality, and associated with an enormous economic burden [13].

Antivirals

Although vaccination is the primary means to protect against influenza virus infection, antiviral drugs can offer a valuable addition to vaccines. Two classes of influenza antiviral drugs are currently available, targeting the matrix (M2) ion-channel or the NA enzyme of the virus. Adamantane derivatives amantadine and rimantadine, both M2 inhibitors, bind and irreversibly block the function of the viral M2 ion-channel protein. While adamantanes have been used for decades with <1% adamantane-resistance reported, adamantanes are currently rarely used because of their side effects combined with naturally occurring resistance of circulating influenza A viruses [17-19]. The conserved nature of the NA active site among different influenza viruses, responsible for removal of sialic acids from the infected cell surface, led to the development of two neuraminidase inhibitor (NAI) antivirals: oral oseltamivir and inhaled zanamivir. Contrary to adamantanes, influenza viruses were thought to be less susceptible to develop NAI resistance due to the conserved function of the NA enzyme [20-22]. Therefore, oseltamivir resistant viruses that emerged and spread during the 2007/2008 influenza season came as a surprise [23,24]. As monitoring for emergence of antiviral resistance is vital for both the patient as well as public health, several genotypic assays have been described for the rapid detection of antiviral resistance markers in clinical laboratories [25-27]. In addition to the use of NAI for prophylaxis in high-risk groups and therapy of influenza in nursing homes and hospitals, NAI are stockpiled as part of national programs of pandemic preparedness plans.

Pandemic influenza viruses

The introduction of a novel influenza virus subtype into the human population, to which no or limited immunity exists, may cause an influenza virus pandemic if such a virus adapts to both efficient infection of humans and sustained transmission between humans. Influenza A viruses of subtypes H1N1, H2N2 and H3N2 caused pandemics during the past century [28]. These include the 1918 Spanish flu, caused by the introduction of subtype H1N1 and killing 40 million people worldwide. In 1957, the emergence of subtype H2N2 caused the Asian flu, and in 1968 the subtype H3N2 caused the Hong Kong flu. In 2009, influenza A(H1N1)pdm09 jumped species and caused the first pandemic of the 21st century [29]. Because the genome of influenza A viruses comprises eight negative-strand RNA segments, new subtypes can emerge due to antigenic shift, i.e. reassortment of the RNA segment encoding the HA glycoprotein

when a single cell is co-infected by two different viral strains. This process enabled the 1957 A(H2N2) and 1968 A(H3N2) pandemic strains to acquire avian surface glycoproteins in a human A(H1N1) virus background, sparking pandemics following the human adaptation of these avian glycoproteins [28]. In addition to RNA segment reassortment, non-human influenza viruses may evolve towards more human-tropic variants by adaptive mutations that facilitate the virus' ability to infect and efficiently replicate in humans and efficiently transmit between humans [30-32]. Critical mutations required for air-borne transmission of influenza avian influenza A(H5N1) between ferrets affect the receptor specificity of the virus, the glycosylation pattern, HA stability and polymerase activity [33,34]. Assessment of avian influenza A(H5N1) surveillance data for these mutations identified circulating A(H5N1) viruses that are only four nucleotide mutations away from becoming air-borne transmissible [35]. In parallel, the avian influenza A(H7N9) outbreak that emerged in China 2013 already possessed several mutations associated with human adaptation [36-38]. Although both influenza A(H5N1) and A(H7N9) have yet failed to spread efficient among humans, the widespread presence of non-human influenza viruses that are able to infect humans constitutes a continuous pandemic threat.

The origin of sequence diversity

Influenza virus infection of a single host is characterized by presence of genetic diversity in the viral population, termed virus quasi species, which is generated and maintained during error-prone influenza genome replication and transcription by the polymerase complex. The viral advantage of such diverse quasi species population, is that variants may be present or generated that are more fit in a new environment or under selection pressure [39]. This allows adaptation to changing environments without losing fitness and is essential for virus survival. The spectrum of all influenza virus mutants covered by the quasi species population in sequence space is defined as the quasi species cloud size. The quasi species cloud size is influenced by the error rate of the RNA polymerase but also by host factors. Using models, it has been illustrated that the level of genetic diversity within the quasi species population varies depending on host environment. Either directed by interactions of host components and replicase complex or selection pressure, different hosts may affect the quasi species cloud size [40,41]. The larger the quasi species cloud size, the larger the spectrum of influenza virus mutants covered by the quasi species population, the higher the ability to adapt to changing environments [42]. As an essential characteristic of the within-host quasi species population is its high connectivity, sequencing of influenza viruses obtained from consecutive infections might illustrate the genetic virus population moving through its genetic space.

Utilizing influenza sequence data

Two decades ago molecular techniques such as PCR amplification of viral nucleic acid in clinical samples became available for virus detection, that today have become routine in hospital and public health laboratories. Although the amplification of specific viral nucleic acid allows virus diagnostics, additional molecular approaches can be applied on the same specimen to provide virus specific information. This enables early detection of strains with increased public health risk by surveillance for vaccine escape, antiviral susceptibility, human adaptation and virulence markers of influenza viruses.

Compared with the average evolutionary rate of mammalian genomes, RNA viruses mutate 500.000 times faster [43,44]. Specifically, influenza virus evolutionary rates range from 1.8 to 8.4 × 10^{-3} substitutions/site/year for avian influenza viruses in wild birds and 5.7 × 10^{-3} substitutions/site/year for influenza A (H3) viruses in humans [45,46]. For outbreaks of HPAI virus subtypes H5 and H7 in poultry, evolutionary rates of above 1 × 10^{-2} substitution/site/year have been demonstrated [47,48]. Previous studies have already demonstrated the use of influenza virus sequence data for molecular characterization of outbreak strains and identification strain relatedness between outbreaks [49-53]. Given the high evolutionary rates observed for influenza viruses, characterization of the influenza virus sequence diversity between infected hosts might provide sufficient resolution to identify transmission events during influenza virus outbreaks and epidemics, in addition to the detection of markers associated with a phenotypic trait.

Scope of this thesis

In this thesis we evaluate the use of influenza sequence diversity to support outbreak control measures. Specifically, we investigated the possibility of identifying clustered influenza virus cases as well as chains of influenza virus transmission, and thereby gain information on transmission events to inform public health decisions.

The possibility to identify clustered seasonal influenza virus cases using molecular surveillance as a background dataset was evaluated (Chapter 2). Initially, influenza virus sequence data was used to supplement epidemiological data on a nosocomial patient cluster (Chapter 2.1). Secondly, influenza virus sequence diversity observed during surveillance was characterized and compared with virus sequences from patients and staff with influenza in one hospital, to evaluate the use of sequence data to identify potential nosocomial patient clusters in the absence of epidemiological data (Chapter 2.2). Next, the possibility to identify influenza virus transmission events using virus sequence data from an influenza virus outbreak was evaluated (Chapter 3). Chapter 3.1 exploited the sequence diversity observed within a densely sampled HPAI A(H7N7) outbreak, to study whether influenza virus mutations with increased public health risk emerged during infection of humans, or were already present in viruses obtained from poultry. When comparative analysis of HPAI A(H7N7) sequence data and meteorological data demonstrated statistical evidence that the direction of farm-to-farm spread correlated with the wind direction at the date of infection [54], field experiments were performed to explore the possibility of wind-mediated influenza virus spread during outbreaks in poultry (Chapter 3.2). For the early detection of markers associated with increased public health risk, the discriminatory power of Next Generation Sequencing (NGS) to robustly characterize the within-host influenza virus sequence diversity was evaluated (Chapter 4). Therefore, the capacity to the distinguish biological from artificial variation in NGS datasets was assessed (Chapter 4.1). An optimized NGS analysis pipeline was applied to study the emergence of an influenza virus virulence marker using longitudinal samples obtained from a single farm-to-human HPAI A(H7N7) virus transmission event (Chapter 4.2). In Chapter 5 we investigated whether the obtained knowledge on the use of influenza virus sequence diversity within outbreaks and epidemics can be applied to prospectively assess the public health risk of an outbreak, using initial virological findings from the A(H7N9) outbreak in China 2013. Lastly, the main findings of this thesis are summarized and the use of influenza sequence data to support outbreak control measures is discussed in relation to the current knowledge in the field (Chapter 6).

References

1. Tong S, Li Y, Rivailler P, Conrardy C, Castillo DA, et al. (2012) A distinct lineage of influenza A virus from bats. Proc Natl Acad Sci U S A 109: 4269-4274.
2. Tong S, Zhu X, Li Y, Shi M, Zhang J, et al. (2013) New world bats harbor diverse influenza a viruses. PLoS Pathog 9: e1003657.
3. Causey D, Edwards SV (2008) Ecology of avian influenza virus in birds. J Infect Dis 197 Suppl 1: S29-33.
4. Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y (1992) Evolution and ecology of influenza A viruses. Microbiol Rev 56: 152-179.
5. Fouchier RA, Munster V, Wallensten A, Bestebroer TM, Herfst S, et al. (2005) Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls. J Virol 79: 2814-2822.
6. Cardona CJ, Xing Z, Sandrock CE, Davis CE (2009) Avian influenza in birds and mammals. Comp Immunol Microbiol Infect Dis 32: 255-273.
7. OIE (2009) High Pathogenicity Avian Influenza. Centre for food security and public health.
8. Swayne DE (2007) Understanding the complex pathobiology of high pathogenicity avian influenza viruses in birds. Avian Dis 51: 242-249.
9. Webby R, Hoffmann E, Webster R (2004) Molecular constraints to interspecies transmission of viral pathogens. Nat Med 10: S77-81.
10. Reperant LA, Rimmelzwaan GF, Kuiken T (2009) Avian influenza viruses in mammals. Rev Sci Tech 28: 137-159.
11. Wang H, Feng Z, Shu Y, Yu H, Zhou L, et al. (2008) Probable limited person-to-person transmission of highly pathogenic avian influenza A (H5N1) virus in China. Lancet 371: 1427-1434.
12. Ungchusak K, Auewarakul P, Dowell SF, Kitphati R, Auwanit W, et al. (2005) Probable person-to-person transmission of avian influenza A (H5N1). N Engl J Med 352: 333-340.
13. Molinari NA, Ortega-Sanchez IR, Messonnier ML, Thompson WW, Wortley PM, et al. (2007) The annual impact of seasonal influenza in the US: measuring disease burden and costs. Vaccine 25: 5086-5096.
14. Palese P, Garcia-Sastre A (2002) Influenza vaccines: present and future. J Clin Invest 110: 9-13.
15. Klimov AI, Garten R, Russell C, Barr IG, Besselaar TG, et al. (2012) WHO recommendations for the viruses to be used in the 2012 Southern Hemisphere Influenza Vaccine: epidemiology, antigenic and genetic characteristics of influenza A(H1N1)pdm09, A(H3N2) and B influenza viruses collected from February to September 2011. Vaccine 30: 6461-6471.

16. Gerdil C (2003) The annual production cycle for influenza vaccine. Vaccine 21: 1776-1779.
17. Deyde VM, Xu X, Bright RA, Shaw M, Smith CB, et al. (2007) Surveillance of resistance to adamantanes among influenza A(H3N2) and A(H1N1) viruses isolated worldwide. J Infect Dis 196: 249-257.
18. Bright RA, Medina MJ, Xu X, Perez-Oronoz G, Wallis TR, et al. (2005) Incidence of adamantane resistance among influenza A (H3N2) viruses isolated worldwide from 1994 to 2005: a cause for concern. Lancet 366: 1175-1181.
19. Ziegler T, Hemphill ML, Ziegler ML, Perez-Oronoz G, Klimov AI, et al. (1999) Low incidence of rimantadine resistance in field isolates of influenza A viruses. J Infect Dis 180: 935-939.
20. Yen HL, Hoffmann E, Taylor G, Scholtissek C, Monto AS, et al. (2006) Importance of neuraminidase active-site residues to the neuraminidase inhibitor resistance of influenza viruses. J Virol 80: 8787-8795.
21. Aoki FY, Boivin G, Roberts N (2007) Influenza virus susceptibility and resistance to oseltamivir. Antivir Ther 12: 603-616.
22. Carr J, Ives J, Kelly L, Lambkin R, Oxford J, et al. (2002) Influenza virus carrying neuraminidase with reduced sensitivity to oseltamivir carboxylate has altered properties in vitro and is compromised for infectivity and replicative ability in vivo. Antiviral Res 54: 79-88.
23. Hauge SH, Dudman S, Borgen K, Lackenby A, Hungnes O (2009) Oseltamivir-resistant influenza viruses A (H1N1), Norway, 2007-08. Emerg Infect Dis 15: 155-162.
24. Meijer A, Lackenby A, Hungnes O, Lina B, van-der-Werf S, et al. (2009) Oseltamivir-resistant influenza virus A (H1N1), Europe, 2007-08 Season. Emerg Infect Dis 15: 552-560.
25. van der Vries E, Jonges M, Herfst S, Maaskant J, Van der Linden A, et al. (2010) Evaluation of a rapid molecular algorithm for detection of pandemic influenza A (H1N1) 2009 virus and screening for a key oseltamivir resistance (H275Y) substitution in neuraminidase. J Clin Virol 47: 34-37.
26. Deyde VM, Okomo-Adhiambo M, Sheu TG, Wallis TR, Fry A, et al. (2008) Pyrosequencing as a tool to detect molecular markers of resistance to neuraminidase inhibitors in seasonal influenza A viruses. Antiviral Res.
27. Carr MJ, Sayre N, Duffy M, Connell J, Hall WW (2008) Rapid molecular detection of the H275Y oseltamivir resistance gene mutation in circulating influenza A (H1N1) viruses. J Virol Methods 153: 257-262.
28. Horimoto T, Kawaoka Y (2005) Influenza: lessons from past pandemics, warnings from current incidents. Nat Rev Microbiol 3: 591-600.

29. Dawood FS, Jain S, Finelli L, Shaw MW, Lindstrom S, et al. (2009) Emergence of a novel swine-origin influenza A (H1N1) virus in humans. N Engl J Med 360: 2605-2615.
30. Miotto O, Heiny AT, Albrecht R, Garcia-Sastre A, Tan TW, et al. (2010) Complete-proteome mapping of human influenza A adaptive mutations: implications for human transmissibility of zoonotic strains. PLoS One 5: e9025.
31. Steel J, Lowen AC, Mubareka S, Palese P (2009) Transmission of influenza virus in a mammalian host is increased by PB2 amino acids 627K or 627E/701N. PLoS Pathog 5: e1000252.
32. Palese P (2004) Influenza: old and new threats. Nat Med 10: S82-87.
33. Imai M, Watanabe T, Hatta M, Das SC, Ozawa M, 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.
34. Herfst S, Schrauwen EJ, Linster M, Chutinimitkul S, de Wit E, et al. (2012) Airborne transmission of influenza A/H5N1 virus between ferrets. Science 336: 1534-1541.
35. Russell CA, Fonville JM, Brown AE, Burke DF, Smith DL, et al. (2012) The potential for respiratory droplet-transmissible A/H5N1 influenza virus to evolve in a mammalian host. Science 336: 1541-1547.
36. Gao R, Cao B, Hu Y, Feng Z, Wang D, et al. (2013) Human Infection with a Novel Avian-Origin Influenza A (H7N9) Virus. N Engl J Med.
37. Kageyama T FS, Takashita E, Xu H, Yamada S, Uchida Y, Neumann G, Saito T, Kawaoka Y, Tashiro M. (2013) Genetic analysis of novel avian A(H7N9) influenza viruses isolated from patients in China, February to April 2013. . Euro Surveilance 18(15):pii=20453.
38. Jonges M, Meijer A, Fouchier R, Koch G, Li J, et al. (2013) Guiding outbreak management by the use of influenza A(H7Nx) virus sequence analysis. Euro Surveill 18.
39. Domingo E, Sheldon J, Perales C (2012) Viral Quasispecies Evolution. Microbiology and Molecular Biology Reviews 76: 159-216.
40. Schneider WL, Roossinck MJ (2000) Evolutionarily Related Sindbis-Like Plant Viruses Maintain Different Levels of Population Diversity in a Common Host. Journal of Virology 74: 3130-3134.
41. Schneider WL, Roossinck MJ (2001) Genetic Diversity in RNA Virus Quasispecies Is Controlled by Host-Virus Interactions. Journal of Virology 75: 6566-6571.
42. Li H, Roossinck MJ (2004) Genetic bottlenecks reduce population variation in an experimental RNA virus population. J Virol 78: 10582-10587.
43. Jenkins GM, Rambaut A, Pybus OG, Holmes EC (2002) Rates of Molecular Evolution in RNA Viruses: A Quantitative Phylogenetic Analysis. Journal of Molecular Evolution 54: 156-165.

44. Kumar S, Subramanian S (2002) Mutation rates in mammalian genomes. Proc Natl Acad Sci U S A 99: 803-808.
45. Fitch WM, Bush RM, Bender CA, Cox NJ (1997) Long term trends in the evolution of H(3) HA1 human influenza type A. Proceedings of the National Academy of Sciences 94: 7712-7718.
46. Chen R, Holmes EC (2006) Avian Influenza Virus Exhibits Rapid Evolutionary Dynamics. Molecular Biology and Evolution 23: 2336-2341.
47. Lee C-W, Senne DA, Suarez DL (2004) Effect of Vaccine Use in the Evolution of Mexican Lineage H5N2 Avian Influenza Virus. Journal of Virology 78: 8372-8381.
48. Bataille A, van der Meer F, Stegeman A, Koch G (2011) Evolutionary analysis of inter-farm transmission dynamics in a highly pathogenic avian influenza epidemic. PLoS Pathog 7: e1002094.
49. Kim SH, Kim HJ, Jin YH, Yeoul JJ, Lee KK, et al. (2013) Isolation of influenza A(H3N2)v virus from pigs and characterization of its biological properties in pigs and mice. Arch Virol 158: 2351-2357.
50. Bountouri M, Fragkiadaki E, Ntafis V, Kanellos T, Xylouri E (2011) Phylogenetic and molecular characterization of equine H3N8 influenza viruses from Greece (2003 and 2007): evidence for reassortment between evolutionary lineages. Virol J 8: 350.
51. Forgie SE, Keenliside J, Wilkinson C, Webby R, Lu P, et al. (2011) Swine outbreak of pandemic influenza A virus on a Canadian research farm supports human-to-swine transmission. Clin Infect Dis 52: 10-18.
52. Kapczynski DR, Pantin-Jackwood M, Guzman SG, Ricardez Y, Spackman E, et al. (2013) Characterization of the 2012 highly pathogenic avian influenza H7N3 virus isolated from poultry in an outbreak in Mexico: pathobiology and vaccine protection. J Virol 87: 9086-9096.
53. Monne I, Hussein HA, Fusaro A, Valastro V, Hamoud MM, et al. (2013) H9N2 influenza A virus circulates in H5N1 endemically infected poultry population in Egypt. Influenza Other Respir Viruses 7: 240-243.
54. Ypma RJ, Jonges M, Bataille A, Stegeman A, Koch G, et al. (2013) Genetic data provide evidence for wind-mediated transmission of highly pathogenic avian influenza. J Infect Dis 207: 730-735.

Chapter 2

Identification of clustered influenza virus cases

Chapter 2.1

Morbidity and Mortality Associated with Nosocomial Transmission of Oseltamivir-Resistant Influenza A(H1N1) Virus

Jairo Gooskens[1], Marcel Jonges[2], Eric C.J. Claas[1], Adam Meijer[2], Peterhans J. van den Broek[3], Aloys C.M. Kroes[1]

JAMA. 2009;301(10):1042-1046

[1]Department of Medical Microbiology, Center of Infectious Diseases, Leiden University Medical Center, Leiden, The Netherlands

[2]Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, The Netherlands

[3]Department of Infectious Diseases, Center of Infectious Diseases, Leiden University Medical Center, Leiden, The Netherlands

Abstract

Context The sudden emergence and rapid spread of oseltamivir-resistant influenza A(H1N1) viruses with neuraminidase (NA) gene H274Y amino acid substitution is the hallmark of global seasonal influenza since January 2008. Viruses carrying this mutation are widely presumed to exhibit attenuated pathogenicity, compromised transmission, and reduced lethality.

Objective To investigate nosocomial viral transmission in a cluster of patients with influenza A(H1N1) virus infection.

Design, Setting, and Patients Descriptive outbreak investigation of 2 hematopoietic stem cell transplant recipients and an elderly patient who developed hospital-acquired influenza A virus infection following exposure to an index patient with community-acquired H274Y-mutated influenza A(H1N1) virus infection in a medical ward at a Dutch university hospital in February 2008. The investigation included a review of the medical records, influenza virus polymerase chain reaction and culture, phenotypic oseltamivir and zanamivir susceptibility determination, and hemagglutinin chain 1 (HA_1) gene and NA gene sequence analysis.

Main Outcome Measure Phylogenetic relationship of patient cluster influenza A(H1N1) viruses and other 2007-2008 seasonal influenza A(H1N1) viruses.

Results Viral HA_1 and NA gene sequence analysis from the 4 patients revealed indistinguishable nucleotide sequences and phylogenetic clustering of H274Y-mutated, oseltamivir-resistant influenza A(H1N1) virus, confirming nosocomial transmission. Influenza virus pneumonia (3 patients) and attributable mortality (2 patients) during active infection was observed in patients with lymphocytopenia at onset.

Conclusion Seasonal oseltamivir-resistant influenza A(H1N1) viruses with NA gene H274Y mutation are transmitted and retain significant pathogenicity and lethality in high-risk patients.

Introduction

A global emergence and rapid spread of oseltamivir-resistant influenza A(H1N1) viruses carrying a neuraminidase (NA) gene with an H274Y (N2 numbering; H275Y in N1 numbering) amino acid substitution has been observed since January 2008 [1-3]. Viruses carrying this mutation are presumed to exhibit attenuated pathogenicity [4], compromised transmission [5], and reduced lethality [6]. However, current widespread circulation of oseltamivir-resistant influenza A(H1N1) viruses associated with typical influenza illnesses and viral pneumonia suggest that these viruses retain significant transmissibility and pathogenicity [2,3,7,8]. While these resistant variants may cause significant mortality and retain efficient transmission, these properties have not yet been firmly established.

Methods

In February 2008, an outbreak of influenza A(H1N1) virus occurred in a medical ward at a Dutch university hospital. Clinical specimens from symptomatic contact patients of the presumed index patient were tested by influenza polymerase chain reaction (PCR) and sequences were further analyzed. Medical records of contact patients with related influenza virus infection were reviewed for underlying disease, clinical findings, and outcome. Screening specimens were obtained from contacts of the last outbreak patient (patient 4) to rule out further spread of the virus.

Influenza virus diagnostics

Influenza virus detection was performed on clinical respiratory specimens using rapid antigen testing and PCR as described previously [9] along with viral culture. Antigenic characterization (hemagglutination inhibition testing) and phenotypic oseltamivir and zanamivir susceptibility (IC_{50}, concentration of drug needed to inhibit enzyme activity by 50%) were determined as described [10,11]. Viral RNA extracted from clinical specimens was further transcribed and amplified. Hemagglutinin chain 1 (HA_1) and neuraminidase gene sequences from patients with confirmed influenza virus infection were analyzed and phylogenetically related to other 2007-2008 seasonal influenza viruses obtained at the hospital or sentinel surveillance isolates collected nationwide using BioNumerics version 5.1 (Applied Maths, Sint-Martens-Latem, Belgium).

Ethical Considerations

The investigation of this outbreak did not involve any planned activity that could have been reviewed prospectively by an institutional review board or ethics committee. Nevertheless, all necessary precautions were taken to prevent identification of the

patients and health care workers involved. The physicians involved all gave signed permission to use clinical data and were informed on the outcome of the investigation. All of them agreed with the intention to publication. Details including the age and role of the health care workers were omitted or described in a nonspecific way, while we also took care to preserve all clinically meaningful details. The chairman of the ethics committee at the Leiden University Medical Center was consulted retrospectively and agreed to the approach as described for reporting the clinical information obtained during the investigation and included herein.

Results

Nosocomial influenza A(H1N1) virus outbreak

The clinical characteristics and timeline of the outbreak of influenza A(H1N1) virus are shown in the TABLE and in FIGURE 1, respectively. Community-acquired oseltamivir-resistant influenza A(H1N1) virus with NA gene H274Y mutation, isolated from the presumed index case, was detected in 3 additional patients (mean oseltamivir IC_{50}, 484 nM; mean zanamivir IC_{50},1.1 nM). The presumed index case (patient 1), who was vaccinated for 2007-2008 seasonal influenza and received high-dose (cumulative) prednisolone therapy for systemic lupus erythematosus, was admitted to the hospital on January 29, 2008, with fever, cough, dyspnea, and lymphocytopenia. Mechanical ventilation and broad-spectrum empirical antibacterial treatment were initiated for acute respiratory failure and apparent pulmonary consolidations by chest radiography. Oseltamivir administration was initiated following influenza A virus detection using rapid antigen testing and PCR along with contact and droplet isolation. No other viral and bacterial respiratory pathogens were detected and blood cultures remained negative. The patient was transferred to a medical ward following clinical improvement on February 1 and isolation precautions were continued for the duration of symptoms until February 3. Viral clearance was confirmed by PCR upon lymphocyte reconstitution in an outpatient setting on March 4, 2008.

Two hematopoietic stem cell transplant recipients (patient 2 and patient 4) and an 89-year-old elderly patient (patient 3) developed hospital-acquired influenza A virus infection (Table). These patients were present at the same time as the index patient at the medical ward without isolation procedures (Figure 1), but the 4 patients never shared rooms. Patient 2 developed mild influenza symptoms with rapid viral clearance, whereas both patients 3 and 4 developed pulmonary consolidations and fatal respiratory failure with viral excretion under broad-spectrum antibacterial therapy. The influenza vaccination status of these patients is not known. The attribution of mortality to influenza was supported by detection of influenza A(H1N1) viral RNA from

postmortem pulmonary tissue and histopathological pulmonary findings consistent with viral pneumonia in patient 4, with the exclusion of other pathogens. Further nosocomial spread to other contacts within the wards was excluded by PCR.

Table. Clinical Characteristics of 4 Hospitalized Patients With Oseltamivir-Resistant Influenza A(H1N1) Virus Infection

	Patient 1 (Index)	Patient 2	Patient 3	Patient 4
Sex	Male	Female	Male	Male
Age, y	47	57	89	66
Medical history	Systemic lupus erythematosus, sarcoidosis, diabetes	Multiple myeloma, allogeneic stem cell transplantation following reduced-intensity conditioning engraftment	Diabetes, cardiac ischemia	Acute myelogenous leukemia, allogeneic stem cell transplantation following myeloablative conditioning, chronic graft-vs-host disease
Immunosuppressive therapy	Prednisolone	None	None	Mycophenolate mofetil, prednisone, recent cyclosporine
Reason for admission	Fever, dyspnea	Fever, mild cough	Fever	Fever
Confirmed diagnosis	Influenza lower respiratory tract infection	Bacteremia, source unknown	Bacteremia, urosepsis	Bacteremia, abdominal sepsis, active graft-vs-host disease
First positive test result for influenza A[a]	January 29, 2008	February 11, 2008	February 15, 2008	February 14, 2008
No. of days following admission in hospital	0	16	17	17
Antiviral therapy	Oseltamivir	Oseltamivir	Oseltamivir	Oseltamivir
Absolute lymphocyte count within 48 h of symptom onset, cells/μL	699 (lymphocytopenia)	3195 (normal count)	726 (lymphocytopenia)	217 (lymphocytopenia)
Respiratory symptoms	Dyspnea, oxygen dependent	Cough	Dyspnea, oxygen dependent	Dyspnea, oxygen dependent
Chest radiograph result	Left lower-lobe consolidation	No pulmonary abnormalities	Bilateral lower-lobe consolidations	Perihilar consolidations
Absolute lymphocyte count during follow-up, cells/μL[b]	1134 (normal count)	4739 (normal count)	Not available	129 (lymphocytopenia)
Influenza outcome	Viral clearance	Viral clearance	Fatal influenza pneumonia[c]	Fatal influenza pneumonia[c, d]

[a]Determined by polymerase chain reaction (PCR).
[b]Lymphocyte count determined 2 to 6 weeks after influenza confirmation.
[c]Clinical diagnosis (dyspnea, PCR detection of influenza A in respiratory specimens).
[d]Pathological diagnosis (histological features, PCR detection of influenza A in pulmonary tissue).

Phylogenetic relationship

HA_1 gene and NA gene sequence analysis of viruses from the 4 outbreak patients revealed indistinguishable nucleotide sequences and phylogenetic clustering (Figure 2). In addition to NA gene H274Y substitution, a rare T284A substitution was sequenced (NA gene sequence Genbank accession numbers AB476754, AB476755, AB476756, AB462370) and could therefore be recognized as a marker for transmission. An unlinked national surveillance isolate (A/NL/088/08) with identical HA_1 gene nucleotide sequence revealed no NA gene phylogenetic clustering and lacked the specific T284A substitution. The NA gene T284A mutation was not observed in other 2007-2008 seasonal influenza A(H1N1) viruses or in worldwide sequences submitted to public databases, reinforcing the unique genetic relatedness of this influenza A(H1N1) virus patient cluster.

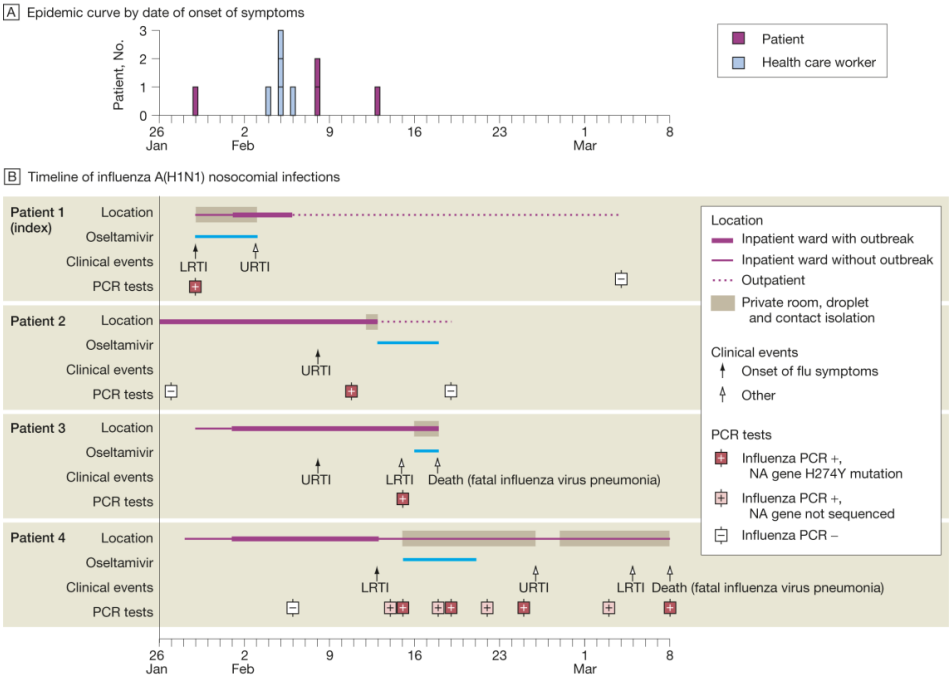


Figure 1. Epidemic Curve by Date of Onset of Symptoms and Timeline of Influenza A(H1N1) Nosocomial Infections. During influenza A(H1N1) virus outbreak, all 4 infected patients were admitted in the same department and never shared rooms. LRTI indicates lower respiratory tract infection; NA, neuraminidase; PCR, polymerase chain reaction; and URTI, upper respiratory tract infection.

Healthcare workers

Five health care workers developed influenza-like illness (onset February 4–6, 2008) during admission of the presumed index patient. One health care worker, vaccinated for the 2007-2008 seasonal influenza, developed influenza-like illness following established contact with the index patient and continued working. Four other health care workers took sick leave within 24 to 48 hours of symptom onset. However, no samples for influenza testing were obtained from any of these 5 health care workers. Thus, their role in possibly contributing to this apparent nosocomial spread of influenza could not be confirmed.

Viruses cultured from the patient cluster revealed a poor antigenic match with the 2007-2008 vaccine reference strain A/Solomon Islands/3/06 (approximately 16-fold difference by duplicate hemagglutination inhibition testing). This may in part explain the sustained susceptibility and infectivity of the vaccinated index patient and suspected health care worker resulting in the hospital outbreak.

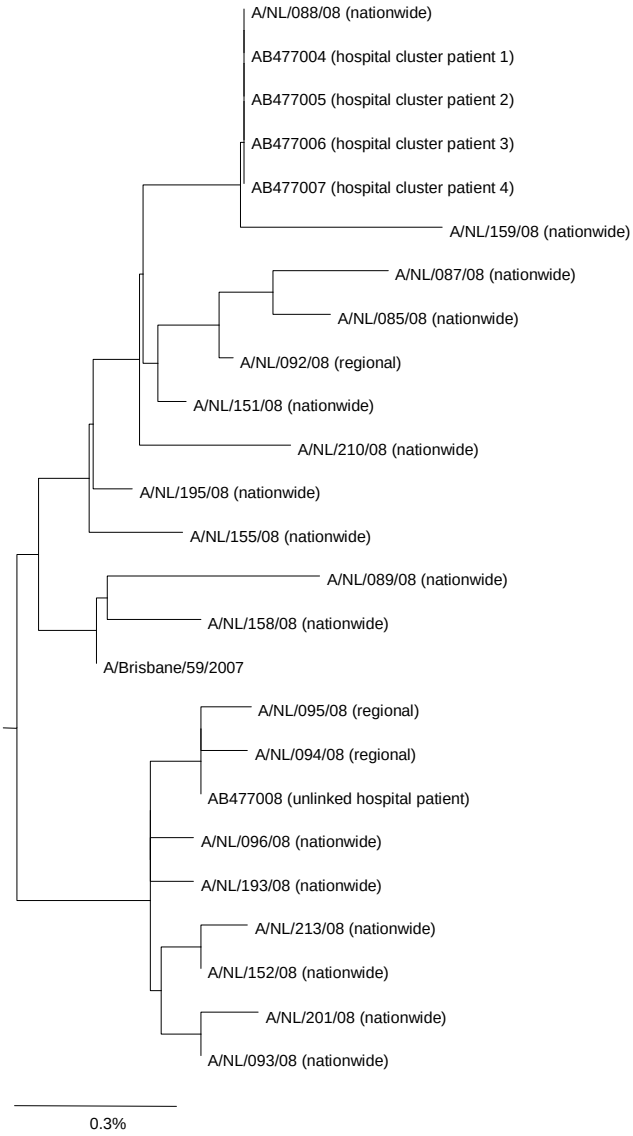


Figure 2. Phylogenetic Relationship of Nosocomial Patient Cluster Influenza A(H1N1) Viruses and Other 2007-2008 Seasonal Influenza A Viruses.

Influenza A(H1N1) virus HA_1 gene sequences obtained from the patient cluster (n=4, marked as hospitalized cluster patient 1, 2, 3, 4) were related to available unlinked 2007-2008 seasonal influenza A(H1N1) viral sequences obtained at the hospital (n=1, marked as unlinked hospitalized patient), surveillance isolates collected within a 10-km regional zone from the hospital (n=3, marked as regional), nationwide collected surveillance isolates (n=17, marked as nationwide), and vaccine strain A/Brisbane/59/2007. The HA_1 gene (nucleotides 1-1071) neighbor-joining tree was rooted on vaccine strain A/Solomon Islands/3/2006. Viral sequence Genbank accession numbers are depicted for hospitalized patients.

Comment

This outbreak provided evidence that circulating oseltamivir-resistant influenza A(H1N1) viruses with NA gene H274Y mutation are transmitted between humans. Limitations of this observational study include the small number of patients, therefore the findings require careful interpretation and do not allow conclusions on the frequency of this complication in hospital settings. The vaccination status of secondarily infected cases (patients 2, 3, and 4) remained unclarified. Information on the mechanism of spread was limited by the circumstances in this study. Data obtained from clinical specimens suggest different routes of transmission; however, this could not be further explored because the sampling of symptomatic health care workers and testing of fomites are not routinely performed. However, analysis of data obtained from clinical specimens provided some insight to different routes of transmission and suggested a limited viral spread.

Early identification and prolonged isolation precautions appear prudent in the care for infected immunocompromised patients to prevent nosocomial influenza virus outbreaks. This study confirmed that circulating H274Y-mutated A(H1N1) viruses can retain significant pathogenicity and lethality, as shown in these elderly or immunocompromised patients with lymphocytopenia, underlining the urgency for the introduction of new effective antiviral agents and therapeutic strategies [12].

Acknowledgment

The authors thank the Dutch sentinel GPs for providing respiratory specimens for virological surveillance of influenza-like illness from which viral sequences were included in this analysis for comparison, and Dr. G. Donker, NIVEL Netherlands institute for health services research, coordinator of the Dutch GP sentinel surveillance network. M. Jonges reports that he was funded by the Impulse Veterinary Avian Influenza Research in the Netherlands program of the Economic Structure Enhancement Fund (FES). The other authors declare not to have any disclosures of financial or potential conflicts of interest.

References

1. Lackenby A, Hungnes O, Dudman SG, Meijer A, Paget WJ, et al. (2008) Emergence of resistance to oseltamivir among influenza A(H1N1) viruses in Europe. Euro Surveill 13.
2. WHO (2009) Epidemic and pandemic alert response: influenza A(H1N1) virus resistance to oseltamivir.
3. Prevention CfDCa (2009) CDC issues interim recommendations for the use of influenza antiviral medications in the setting of oseltamivir resistance among circulating influenza A (H1N1) viruses, 2008-09 influenza season.
4. Ives JA, Carr JA, Mendel DB, Tai CY, Lambkin R, et al. (2002) The H274Y mutation in the influenza A/H1N1 neuraminidase active site following oseltamivir phosphate treatment leave virus severely compromised both in vitro and in vivo. Antiviral Res 55: 307-317.
5. Herlocher ML, Truscon R, Elias S, Yen HL, Roberts NA, et al. (2004) Influenza viruses resistant to the antiviral drug oseltamivir: transmission studies in ferrets. J Infect Dis 190: 1627-1630.
6. Yen HL, Ilyushina NA, Salomon R, Hoffmann E, Webster RG, et al. (2007) Neuraminidase inhibitor-resistant recombinant A/Vietnam/1203/04 (H5N1) influenza viruses retain their replication efficiency and pathogenicity in vitro and in vivo. J Virol 81: 12418-12426.
7. Nicoll A, Ciancio B, Kramarz P, Influenza Project T (2008) Observed oseltamivir resistance in seasonal influenza viruses in Europe interpretation and potential implications. Euro Surveill 13.
8. van der Vries E, van den Berg B, Schutten M (2008) Fatal oseltamivir-resistant influenza virus infection. N Engl J Med 359: 1074-1076.
9. Templeton KE, Scheltinga SA, Beersma MF, Kroes AC, Claas EC (2004) Rapid and sensitive method using multiplex real-time PCR for diagnosis of infections by influenza a and influenza B viruses, respiratory syncytial virus, and parainfluenza viruses 1, 2, 3, and 4. J Clin Microbiol 42: 1564-1569.
10. Gooskens J, Kuiken T, Claas EC, Harinck HI, Thijssen JC, et al. (2007) Severe influenza resembling hemorrhagic shock and encephalopathy syndrome. J Clin Virol 39: 136-140.
11. Potier M, Mameli L, Belisle M, Dallaire L, Melancon SB (1979) Fluorometric assay of neuraminidase with a sodium (4-methylumbelliferyl-alpha-D-N-acetylneuraminate) substrate. Anal Biochem 94: 287-296.
12. Hayden F (2009) Developing new antiviral agents for influenza treatment: what does the future hold? Clin Infect Dis 48 Suppl 1: S3-13.

Chapter 2.2

Sequence-Based Identification and Characterization of Nosocomial Influenza A(H1N1)pdm09 Virus Infections

Marcel Jonges[1,2], Janette Rahamat-Langendoen[3], Adam Meijer[1], Hubert Niesters[3], Marion Koopmans[1,2]

J Hosp Infect. 2012 Nov;82(3):187-93

[1]Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, The Netherlands

[2]Department of Viroscience, Erasmus MC, Rotterdam, The Netherlands

[3]University Medical Center Groningen, the University of Groningen, Department of Medical Microbiology, Groningen, the Netherlands

Summary

Background

Highly transmissible viruses such as influenza are a potential source of nosocomial infections and thereby cause increased patient morbidity and mortality.

Aim

To assess whether influenza virus sequence data can be used to link nosocomial influenza transmission between individuals.

Methods

Dutch A(H1N1)pdm09-positive specimens from one hospital (N = 107) were compared with samples from community cases (N = 685). Gene fragments of haemagglutinin, neuraminidase and PB2 were sequenced and subsequently clustered to detect patients infected with identical influenza viruses. The probability of detecting a second patient was calculated for each hospital cluster against the background diversity observed in hospital and community strains. All clusters were further analysed for possible links between patients.

Findings

Seventeen A(H1N1)pdm09 hospital clusters were detected of which eight had a low probability of occurrence compared with background diversity ($P < 0.01$). Epidemiological analysis confirmed a total of eight nosocomial infections in four of these eight clusters, and a mother–child combination in a fifth cluster. The nine clusters with a high probability of occurrence involved community cases of influenza without a known epidemiological link.

Conclusion

If a background sequence dataset is available, the detection of hospital sequence clusters that differ from dominant community strains can be used to select clusters requiring further investigation by hospital hygienists before a nosocomial influenza outbreak is epidemiologically suspected.

Introduction

Despite its overall mild appearance, the 2009 influenza A(H1N1) pandemic (A(H1N1)pdm09) resulted in a significantly increased demand on health services [1,2]. Outbreaks of nosocomial influenza can occur, facilitated by transmission from HCWs to patients and colleagues [3,4]. Ward closure is frequently performed when preventive measures fail to control virus spread, illustrating the problems in containing nosocomial outbreaks of viruses that are highly transmissible and require a low infectious dose such as influenza viruses [5-7]. Consequently, identification of sources of nosocomial virus infection can direct outbreak control measures, leading to reduced patient morbidity and mortality, and preserving a functioning healthcare system [7].
The aim of this study was to assess whether influenza sequence data can be used to detect nosocomial influenza transmission, and to identify its sources and risk factors associated with nosocomial spread. Since April 2009, clinical samples from A(H1N1)pdm09-positive cases have been systematically sequenced for mutations previously associated with increased virulence and reduced antiviral susceptibility to monitor the emergence of the new influenza virus. The data have been used to explore the potential for patient cluster detection. The sequencing protocol was implemented by several healthcare institutes such as the University Medical Center Groningen (UMCG). The national and UMCG sequence dataset are analysed for clusters of nosocomial influenza. A sensible sequencing strategy targeting antiviral resistance markers is discussed for retrospective as well as real-time characterization of nosocomial influenza infections in healthcare institutes during an epidemic.

Methods

Cases in the UMCG hospital

Between weeks 40 and 53 in 2009, throat or nose swabs from patients as well as healthcare workers (HCWs) with symptoms of acute respiratory illness were tested by reverse transcriptase–polymerase chain reaction (RT–PCR) for the presence of influenza A(H1N1)pdm09 virus [8,9]. Further characterization was done by sequence analysis as described below. Clinical and epidemiological data from hospitalized patients with influenza A(H1N1)pdm09 were gathered using a standardized case report form. Based on the previously reported two-day (median) incubation period for A(H1N1)pdm09 virus infections, nosocomial infection was defined as development of respiratory illness two or more days after hospital admission [10]. The influenza viruses in this UMCG hospital dataset are referred to in the text as ‘hospital influenza viruses’.

Cases from national surveillance

A background dataset consisting of all (N = 685) sequenced A(H1N1)pdm09-positive samples obtained from the Dutch influenza GP sentinel surveillance system (N = 103) and 582 non-sentinel samples that were collected between October 2009 and April 2010 were compared with data for hospital influenza viruses [11]. The influenza viruses in this background dataset are referred to in the text as 'community influenza viruses'.

Sequencing

Throat or nose swabs from A(H1N1)pdm09-positive cases were sequenced directly for surveillance of virulence and antiviral resistance markers, as described previously [11]. Briefly, RNA was extracted from the clinical specimen and transcribed into cDNA using ThermoScript ™ reverse transcriptase (Invitrogen, Carlsbad, CA, USA) following target amplification using HotstarTaq Master Mix (Qiagen, Hilden, Germany). PCR primers targeted the receptor binding site within haemagglutinin (HA) (nt 8–789), the neuraminidase inhibitor resistance markers in neuraminidase (NA) (nt 669–1323) and known virulence markers for influenza A viruses in the polymerase binding protein 2 (PB2) (nt 1684–2223). A viral load cut-off below Ct 25 in the Matrix RT-PCR assay was used for sequencing specimens obtained from HCWs. DNA sequences were analysed using BioNumerics V6.5 (Applied Maths, Sint-Martens-Latem, Belgium).

Molecular data analysis

The character data from a concatenated HA (nt 58–713), NA (nt 702–1323) and PB2 (nt 1684–2218) nucleotide alignment was used to build a Maximum Parsimony network in BioNumerics [12]. This method was chosen for its ability to link A(H1N1)pdm09-positive cases to their suspected source of infection using a minimum number of 'evolutionary events' based on the simplest, most parsimonious, explanation of an observation and free of specific evolutionary assumptions. Patient clusters were defined by 100% sequence identity of the sequenced A(H1N1)pdm09 gene fragments. As antiviral resistant influenza viruses can emerge during oseltamivir therapy of immunocompromised patients [11,13], nucleotide variation at NA codon 275 encoding resistance marker H275Y was excluded from the cluster analysis. Hospital and community influenza virus datasets were compared for their Maximum Parsimony using character data and the evolutionary divergence over all sequence pairs using the Kimura 2-parameter model conducted in MEGA5 [14,15]. Standard error estimates were obtained by a bootstrap procedure (1000 replicates). After sequence clustering, the resolution of the hospital sequence data was analysed by calculating the probability of observing a second patient with identical virus sequence for each virus strain that defined a hospital cluster ($n \geq 2$). This was done per gene fragment using the combined community and hospital virus sequence data sets as a measure for the

total sequence diversity observed during the 2009 influenza epidemic, assuming that the proportions of sequence clusters in our dataset correspond with the proportions of clusters within circulating viruses in the Dutch population.

Epidemiological analysis

Influenza sequence clusters were analysed further by reviewing the date of hospitalization, reported onset of illness and ward(s) where the person was hospitalized to find epidemiological links between patients. For HCWs data were collected on reported onset of illness, absenteeism from work, presence of patient contacts during regular work and the ward or specialty where HCWs were working before illness.

Results

Between October and December 2009, 288 of 1470 hospital respiratory samples (20%) tested positive for influenza A(H1N1)pdm09 virus. They were obtained from 121 patients (75 hospitalized patients and 46 outpatients) and 71 HCWs. Most prevalent symptoms were fever, cough and shortness of breath. Almost 30% of the patients had to be admitted to the Intensive Care Unit. In seven patients (9%) the onset of symptoms occurred two or more days after admission to the hospital, suggesting hospital-acquired infection. A total of 107 influenza A(H1N1)pdm09-positive hospital samples were sequenced: 43% of the outpatients (N = 20), 83% of the hospitalized patients (N = 62) and 35% of the HCWs (N = 25), yielding sequence information on 101 NA fragments, 102 PB2 fragments and 69 HA fragments.

Molecular resolution in the hospital sequence dataset

To distinguish clusters of hospitalized patients, the hospital sequence dataset was compared with the national dataset to assess whether influenza viruses obtained from one hospital are closely related or whether they reflect community influenza virus diversity. This was achieved by using Maximum Parsimony networks built with influenza HA, NA or PB2 gene fragments, supplemented by a sequence diversity comparison using a character and substitution based method (Figures 1 and 2). Both methods demonstrate that influenza viruses obtained from the hospital show diversity similar to that of the national dataset. This could allow the detection of patient clusters using hospital sequence data.

Molecular clustering

Initially, hospital sequence data were used without community data to detect patient clusters. Maximum Parsimony networks were constructed using 101 NA fragments, 96

PB2 fragments, 65 HA fragments, 96 combined NA and PB2 fragments, and 65 combined HA, NA and PB2 fragments to detect clusters (N ≥ 2) of 100% identical influenza A(H1N1)pdm09 sequences. Clustering of hospital sequence data based on HA, NA or PB2 fragments identified 7, 14 and 7 sequence clusters, respectively. When combining the clustering results obtained using individual gene fragments, 17 molecularly distinct clusters were identified (Figure 3). Six of the seven patients with onset illness two or more days after hospitalization were present in four clusters. The seventh patient did not cluster with hospital and community sequences due to a silent mutation in NA codon 413. A proportion of the hospital clusters was characterized by unique sequences compared with community sequence data. To enumerate the uniqueness of all detected hospital sequence clusters, a probability was assigned to each sequence cluster based on the frequency with which these variants were observed. Eight patient clusters were identified with a low probability ($P < 0.01$) that these were expected based on random selection in a background diversity of hospital and community influenza strains (Figure 3).

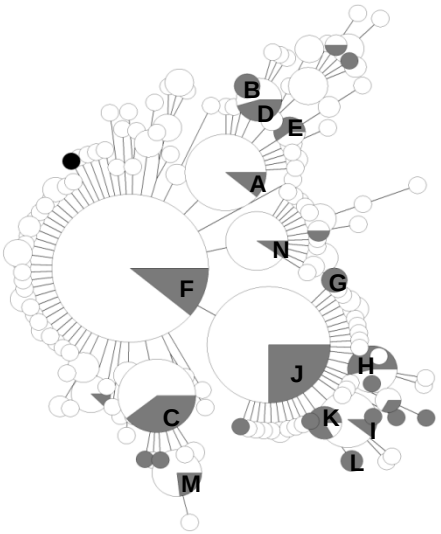


Figure 1. Maximum Parsimony network of influenza A(H1N1)pdm09 NA (neuraminidase) sequences obtained from the hospital (N = 101; grey) and the community (N = 601; white), supplemented by vaccine strain A/California/07/09 (black), showing that hospital A(H1N1)pdm09 viruses are dispersed similarly to community viruses. Genotypes (A–N) were assigned to clusters (N ≥ 2) of 100% identical hospital influenza sequences.

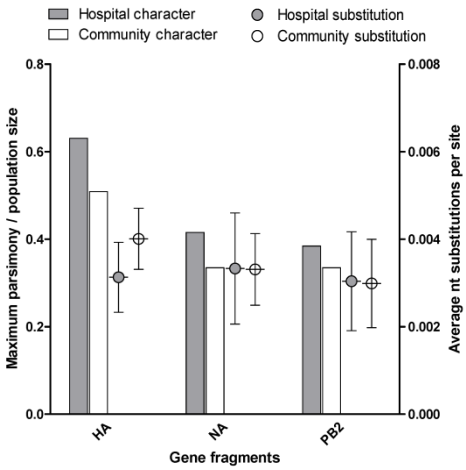


Figure 2. Influenza virus sequence diversities of the hospital and community datasets were compared using a character and substitution based method and expressed as the relative Maximum Parsimony per gene fragment (left Y-axis) and the evolutionary divergence over all sequence pairs per gene fragment (right Y-axis), respectively. Both methods demonstrate similar sequence diversity for each gene segment present in the hospital sequence dataset compared with the observed national diversity.

		Sequence clusters						Probability					
Cluster information		**HA**	**(genotype)**	**NA**	**(genotype)**	**PB2**	**(genotype)**	**p (HA)**	**p (NA)**	**p (PB2)**	**p (NA, PB2)**	**p (HA, NA, PB2)**	**Presence of p < 0.01**
community acquired	patient	+	A	+	A	+	A	0.0199	0.0598	0.1029	0.0313	0.0099	YES
	patient	+		+		+							
	patient			+		+							
	patient			+									
nosocomial cluster 1	patient A	+	B	+	B	+		0.0066	0.0050	0.1029	0.0067	0.0066	YES
	patient C	+		+		+							
	patient B *			+		+							
community acquired	patient	+	C	+	C	+	B	0.2185	0.0548	0.0224	0.0179	0.0199	NO
	patient	+		+		+							
	patient	+		+		+							
	HCW	+		+		+							
	patient			+		+							
	HCW			+		+							
	patient			+									
	HCW			+									
	HCW			+									
community acquired	patient	+	D	+		+	C	0.0265	0.0548	0.2662	0.0403	0.0331	NO
	patient	+		+		+							
	HCW	+		+		+							
	patient			+		+							
community acquired	patient			+	D	+		-	0.0183	0.2662	0.0179	-	NO
	patient			+		+							
	patient			+		+							
	HCW			+		+							
	patient			+									
nosocomial cluster 3	patient			+	E	+		-	0.0083	0.2662	0.0089	-	YES
	patient A *	+		+		+							
	patient B *			+		+							
community acquired	child	+	E	+	F	+	D	0.0497	0.2259	0.0067	0.0067	0.0099	YES
	mother	+		+		+							
	patient	+		+		+							
nosocomial cluster 2	patient A	+	C	+		+	C	0.2185	0.2259	0.2662	0.0559	0.0099	YES
	patient C *	+		+		+							
	patient B *	+		+		+							
	HCW			+		+							
	HCW			+		+							
	patient			+		+							
	HCW			+		+							
	patient			+		+							
	patient			+		+							
	patient			+		+							
	patient			+									
	patient			+									
community acquired	patient	+		+	G	+	E	0.2185	0.0050	0.2416	0.0067	0.0099	YES
	patient	+		+		+							
	HCW	+		+		+							
community acquired	child	+		+	H	+		0.2185	0.0216	0.2416	0.0201	0.0232	NO
	patient	+		+		+							
	HCW	+		+		+							
	patient			+		+							
	mother			+		+							
	patient			+									
	patient			+									

community acquired	patient patient			+ I +		+ +		0.2185	0.0316	0.2416	0.0179	-	NO
community acquired	HCW HCW HCW patient	+ +		+ + + + J		+ + +		0.2185	0.1395	0.2416	0.1074	0.0265	NO
community acquired	patient patient HCW HCW patient patient patient patient patient patient patient patient patient HCW HCW	+ + + + + F		+ + + + + + + + + + + + + + +		+ + + + + + + + +		0.0762	0.1395	0.2416	0.1074	0.0331	NO
community acquired	patient HCW patient patient patient	+ + G		+ + + + + K		+ + + +		0.0232	0.0100	0.2416	0.0089	0.0066	YES
nosocomial cluster 4	patient A * patient B	 +		+ L +		+ +		0.0232	0.0033	0.2416	0.0045	-	YES
community acquired	patient patient HCW			+ + + M		+ + F		-	0.0216	0.0201	0.0179	-	NO
HCW	HCW HCW	 +		+ N +		+ + G		-	0.0349	0.0224	0.0291	-	NO

Figure 3. Sequence clustering of hospital A(H1N1)pdm09 viruses identified a total of 17 molecular clusters based on the combined results obtained from individual gene fragments. Sequence analysis of haemagglutinin (HA) fragments identified seven molecularly distinct clusters (genotypes A–G), of neuraminidase (NA) 14 clusters (NA genotypes A–N) and of PB2 seven clusters (PB2 genotypes A–G). By combining these results, NA genotype C and F viruses are subdivided by addition of HA and/or PB2 sequence data, and NA genotype J viruses are subdivided by addition of HA data, yielding the 17 molecular distinct patient clusters. The probability of observing a second patient with an identical influenza gene fragment sequence is provided for all sequence clusters per separate and combined gene fragment(s). The left column shows the result of the epidemiological cluster investigation. Patients with epidemiological links are marked A, B or C, corresponding with the analysis of patient clusters in Figure 4.

+ Gene fragment sequence data present. ∗Patients with onset illness ≥2 days after hospitalization. HCW, healthcare worker.

Epidemiological clustering

Epidemiological analysis of nine sequence clusters that had a clustering probability P > 0.01 did not identify any epidemiological associations between the patients and/or HCWs within these clusters, except for the detection of a mother–child link within one cluster. Based on first day of illness, all patients within these nine clusters were infected with community-acquired influenza virus infection.

For the eight clusters having a low probability (P < 0.01) that these were expected based on random selection, epidemiological analysis identified no associations between the patients and/or HCWs within three out of eight clusters. One cluster included a mother and child, and epidemiological analysis of the remaining four clusters was suggestive for transmission of influenza virus between patients (Figures 3 and 4). These four clusters each contained one or more patients with a presumed hospital-acquired influenza virus infection based on the first day of illness two or more days after admission.

Hospital cluster 1 contained three A(H1N1)pdm09 virus-infected patients. Patient A, infected with community-acquired influenza, was probably the source of infection for patients B and C. All three were on the same ward at the time as patient A tested positive for influenza A(H1N1)pdm09 virus infection. Patient C was temporarily sent home, but was readmitted with influenza-like illness four days after leaving the hospital for the first time. Based on onset of symptoms compared to date of admission, patient C initially was considered as having a community-acquired infection.

Hospital cluster 2 contained three patients with A(H1N1)pdm09 virus infection including two patients that had a first day of illness respectively 2 and 7 days after admission to the hospital. Patient A, infected with community-acquired influenza, is regarded as the index patient for the other two patients.

Hospital cluster 3 consisted of three patients including two patients with hospital-acquired influenza virus infection. As the third patient stayed at a different ward and was discharged before patient B was admitted, the actual source of infection is missing. Both nosocomial patients were on the same ward before becoming ill, but were already transferred to other wards when tested positive for influenza A(H1N1)pdm09 virus. Most likely the source of these two hospital-acquired infections occurred in ward J.

Hospital cluster 4 consisted of two patients including one patient with a hospital-acquired influenza virus infection. Both patients were on the same ward at the time patient A developed a respiratory illness caused by influenza A(H1N1)pdm09 virus.

Patient B was discharged from the hospital but became ill four days after returning home, eight days after the first day of illness of patient A. Patient B was readmitted because of respiratory illness and subsequently tested positive for influenza A(H1N1)pdm09 virus infection. This readmitted patient was not considered a patient with a hospital-acquired infection in the initial comparison of first day of illness with date of admission. As in cluster 3, the source of infection is missing and most likely was a patient, HCW or visitor from ward N.

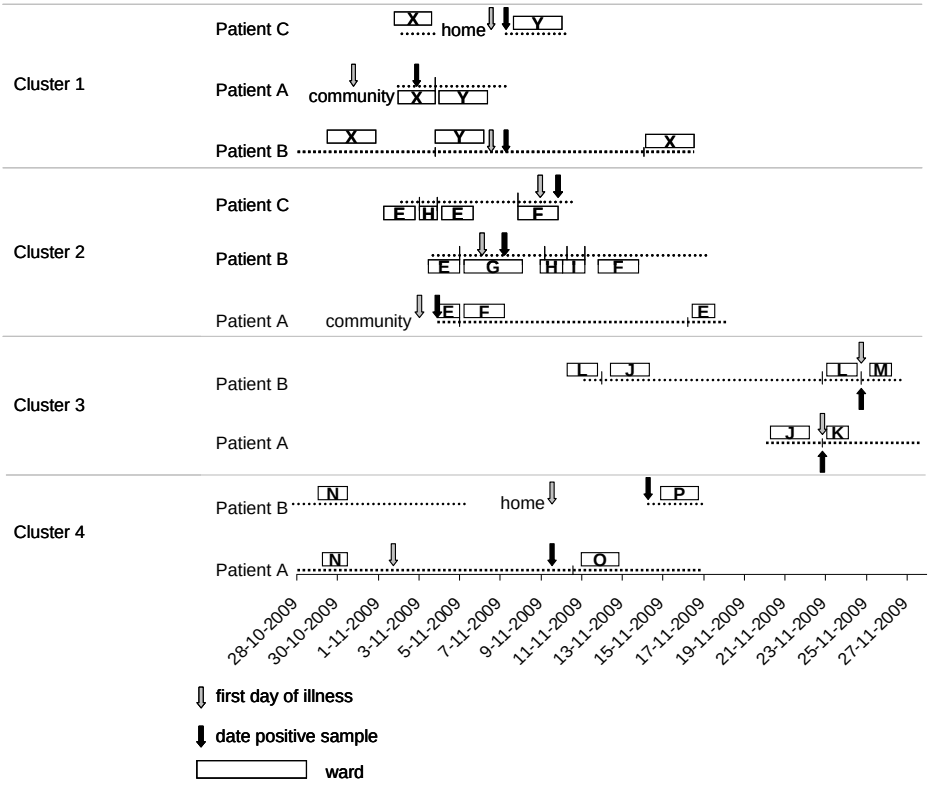


Figure 4. Epidemiological analysis of four patient clusters each containing ≥1 patients with presumed hospital-acquired A(H1N1)pdm09 virus infection suggests that the source of nosocomial influenza virus infections within clusters 1 and 2 was a patient with community-acquired influenza. This analysis identified two additional nosocomial influenza patients (cluster 1, patient C; cluster 4, patient B) that initially were considered as having a community-acquired infection. An index case of nosocomial influenza virus infections within clusters 3 and 4 is missing, but might be associated with a specific ward. Separate wards have been assigned different letters

Discussion

Retrospective analysis of sequence data combined with epidemiological data was performed to evaluate the usefulness of sequencing influenza gene fragments for the characterization of nosocomial influenza virus infections. Although characterization of nosocomial A(H1N1)pdm09 cases by a combined molecular and epidemiological approach has been reported, the evaluation of this approach to guide the identification of nosocomial influenza clusters to our knowledge has not previously been reported [16,17].

The present analysis identified nine nosocomial influenza infections during the 2009 influenza epidemic in a tertiary referral hospital, which comprises 15% of all the diagnoses during the study period. Although transmission from an index case to a hospitalized patient was observed four times, no onward transmissions could be demonstrated. In addition, the molecular and epidemiological analysis did not provide evidence for transmission between patients and staff. Our results highlight some important observations. First, patient transfer from a ward with influenza-positive patients to another ward should be done with great care and follow-up of the transferred patient. Second, follow-up of recently discharged patients could identify possible spread of infection, which is highly relevant if antiviral resistant or virulent influenza virus variants are identified in the hospital. Finally, our analysis was able to link patients across wards in two clusters that would be difficult to detect using epidemiological data only.

Sequencing of HA fragments was less often successful than sequencing of the NA and PB2 fragments due to reduced sensitivity for the HA fragment amplification, compared with NA and PB2 amplification (data not shown). This affected HA-based clustering as these sequences were missing for 52% of the hospital patients. Consequently, an unbiased comparison of HA with NA sequence data for the detection of patient clusters was not possible. The lower substitution rate for A(H1N1)pdm09 PB2 gene segments compared with NA segments explains the discrepancy between the number of molecular clusters detected using NA (N = 14) and PB2 (N = 7) sequence data [18].

Nevertheless, the calculation of the probabilities that A(H1N1)pdm09-positive patients were infected by viruses with 100% identical gene fragments proved to be a useful selection. When comparing the NA sequences with GenBank (consulted on 6 February 2012) data, the five strains that defined NA clusters with $P < 0.01$ all remained unique, whereas the strains that defined clusters with $P > 0.01$ demonstrated on average 57 identical strains. This suggests that our observations did not result from the lack of

resolution in the background data, and implies that, during an influenza epidemic, commonly circulating influenza strains are detected at all geographical areas and can be used to express the uniqueness of a local patient cluster. Although sequence-based typing is possible in advanced clinical laboratories, routine application is relatively costly when considering its use for the sole purpose of identification of nosocomial infections. Therefore, we looked at the potential for using this approach on the basis of data collected for a different purpose, i.e. sequencing to detect markers of antiviral resistance. As antiviral treatment of immunocompromised patients is associated with the emergence of antiviral resistance, monitoring for emergence of resistance is vital for both the patient as well as public health, and is increasingly common practice [13]. Hospital molecular diagnostics commonly utilize a targeted approach for the rapid identification of the primary oseltamivir resistance substitution H275Y in NA subtype 1, but our results demonstrate the possible added value of Sanger sequencing a larger region (>600 bp)[19-21]. Currently, sequence capacities and application of routine sequence analysis in clinical laboratories are increasing, especially in academic hospitals. Although the capacity to perform sequence analysis is not present in all hospitals, these hospitals could benefit from the knowledge generated on the characterization of influenza virus transmission chains in hospitals that are able to incorporate sequence analysis into routine clinical use. Both retrospective and prospective approaches produce valuable insights into nosocomial virus spread that can direct outbreak control measures, leading to reduced patient morbidity and mortality.

Acknowledgements

We would like to thank R. van der Heide and F. Geeraedts for their help in performing the sequence analysis of UMCG influenza viruses. This study was funded by the Impulse Veterinary Avian Influenza Research in The Netherlands programme of the Economic Structure Enhancement Fund.

References

1. Dominguez-Cherit G, Lapinsky SE, Macias AE, Pinto R, Espinosa-Perez L, et al. (2009) Critically Ill patients with 2009 influenza A(H1N1) in Mexico. JAMA 302: 1880-1887.
2. Kumar A, Zarychanski R, Pinto R, Cook DJ, Marshall J, et al. (2009) Critically ill patients with 2009 influenza A(H1N1) infection in Canada. JAMA 302: 1872-1881.
3. Blachere FM, Cao G, Lindsley WG, Noti JD, Beezhold DH (2009) Enhanced detection of infectious airborne influenza virus. J Virol Methods 176: 120-124.
4. Mermel LA (2009) Preventing the spread of influenza A H1N1 2009 to health-care workers. Lancet Infect Dis 9: 723-724.
5. Tellier R (2006) Review of aerosol transmission of influenza A virus. Emerg Infect Dis 12: 1657-1662.
6. Salgado CD, Farr BM, Hall KK, Hayden FG (2002) Influenza in the acute hospital setting. Lancet Infect Dis 2: 145-155.
7. Hansen S, Stamm-Balderjahn S, Zuschneid I, Behnke M, Ruden H, et al. (2007) Closure of medical departments during nosocomial outbreaks: data from a systematic analysis of the literature. J Hosp Infect 65: 348-353.
8. Jonges M, Liu WM, van der Vries E, Jacobi R, Pronk I, et al. (2010) Influenza virus inactivation for studies of antigenicity and phenotypic neuraminidase inhibitor resistance profiling. J Clin Microbiol 48: 928-940.
9. Meijer A, Beerens A, Claas E, Hermans M, de Jong A, et al. (2009) Preparing the outbreak assistance laboratory network in the Netherlands for the detection of the influenza virus A(H1N1) variant. J Clin Virol 45: 179-184.
10. Lagueux M, Perrodou E, Levashina EA, Capovilla M, Hoffmann JA (2000) Constitutive expression of a complement-like protein in toll and JAK gain-of-function mutants of Drosophila. Proc Natl Acad Sci U S A 97: 11427-11432.
11. Meijer A, Jonges M, Abbink F, Ang W, van Beek J, et al. (2011) Oseltamivir-resistant pandemic A(H1N1) 2009 influenza viruses detected through enhanced surveillance in the Netherlands, 2009-2010. Antiviral Res 92: 81-89.
12. Fitch WM (1971) Toward Defining the Course of Evolution: Minimum Change for a Specific Tree Topology. Systematic Zoology 20: 406-416.
13. Gooskens J, Jonges M, Claas EC, Meijer A, Kroes AC (2009) Prolonged influenza virus infection during lymphocytopenia and frequent detection of drug-resistant viruses. J Infect Dis 199: 1435-1441.
14. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood,

evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731-2739.

15. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16: 111-120.
16. Rodriguez-Sanchez B, Alonso M, Catalan P, Sanchez Conde M, Gonzalez-Candelas F, et al. (2011) Genotyping of a nosocomial outbreak of pandemic influenza A/H1N1 2009. J Clin Virol 52: 129-132.
17. Chen LF, Dailey NJ, Rao AK, Fleischauer AT, Greenwald I, et al. (2011) Cluster of oseltamivir-resistant 2009 pandemic influenza A (H1N1) virus infections on a hospital ward among immunocompromised patients--North Carolina, 2009. J Infect Dis 203: 838-846.
18. Mengelle C, Mansuy JM, Saune K, Barthe C, Boineau J, et al. (2012) A new highly automated extraction system for quantitative real-time PCRs from whole blood samples: routine monitoring of opportunistic infections in immunosuppressed patients. J Clin Virol 53: 314-319.
19. Carr MJ, Sayre N, Duffy M, Connell J, Hall WW (2008) Rapid molecular detection of the H275Y oseltamivir resistance gene mutation in circulating influenza A (H1N1) viruses. J Virol Methods 153: 257-262.
20. van der Vries E, Jonges M, Herfst S, Maaskant J, Van der Linden A, et al. (2010) Evaluation of a rapid molecular algorithm for detection of pandemic influenza A (H1N1) 2009 virus and screening for a key oseltamivir resistance (H275Y) substitution in neuraminidase. J Clin Virol 47: 34-37.
21. Gooskens J, Jonges M, Claas EC, Meijer A, van den Broek PJ, et al. (2009) Morbidity and mortality associated with nosocomial transmission of oseltamivir-resistant influenza A(H1N1) virus. JAMA 301: 1042-1046.

Chapter 3

Characterizing transmission events during influenza virus outbreaks

Chapter 3.1

Comparative Analysis of Avian Influenza Virus Diversity in Poultry and Humans During a Highly Pathogenic Avian Influenza A(H7N7) Virus Outbreak

Marcel Jonges[1,2], Arnaud Bataille[3,4], Remko Enserink[1], Adam Meijer[1], Ron A.M. Fouchier[2], Arjan Stegeman[3], Guus Koch[4], Marion Koopmans[1,2]

J Virol. 2011 Oct;85(20):10598-604

[1]Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, The Netherlands

[2]Department of Viroscience, Erasmus MC, Rotterdam, The Netherlands

[3]Department of Farm Animal Health, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands

[4]Central Veterinary Institute, Wageningen University & Research Center, Lelystad, The Netherlands

Abstract

Although increasing data have become available that link human adaptation with specific molecular changes in nonhuman influenza viruses, the molecular changes of these viruses during a large highly pathogenic avian influenza virus (HPAI) outbreak in poultry along with avian-to-human transmission have never been documented. By comprehensive virologic analysis of combined veterinary and human samples obtained during a large HPAI A(H7N7) outbreak in the Netherlands in 2003, we mapped the acquisition of human adaptation markers to identify the public health risk associated with an HPAI outbreak in poultry. Full-length hemagglutinin (HA), neuraminidase (NA), and PB2 sequencing of A(H7N7) viruses obtained from 45 human cases showed amino acid variations at different codons in HA (n=20), NA (n=23), and PB2 (n=23). Identification of the avian sources of human virus infections based on 232 farm sequences demonstrated that for each gene about 50% of the variation was already present in poultry. Polygenic accumulation and farm-to-farm spread of known virulence and human adaptation markers in A(H7N7) virus-infected poultry occurred prior to farm-to-human transmission. These include the independent emergence of HA A143T mutants, accumulation of four NA mutations, and farm-to-farm spread of virus variants harboring mammalian host determinants D701N and S714I in PB2. This implies that HPAI viruses with pandemic potential can emerge directly from poultry. Since the public health risk of an avian influenza virus outbreak in poultry can rapidly change, we recommend virologic monitoring for human adaptation markers among poultry as well as among humans during the course of an outbreak in poultry.

Introduction

Avian influenza viruses are typically restricted to the intestinal tract of aquatic birds, in which they cause mild or subclinical and unspecific symptoms. Occasionally, however, they acquire the properties to directly infect the respiratory tracts of terrestrial birds and mammals, including humans [1]. Introduction of avian influenza in humans could result in a pandemic when the viruses (I) possess a hemagglutinin (HA) subtype antigenically distinct from the currently circulating human influenza virus subtypes, (II) have the ability to infect and efficiently replicate in humans, and (III) transmit efficiently from human to human [2-4]. Infection of humans in direct contact with wild or domestic poultry has been observed for influenza A viruses with HA subtypes H5, H7, H9, and H10 over the past 2 decades. Low-pathogenic avian influenza (LPAI) A(H9N2) virus caused human infections in China in 1998 and Hong Kong in 1999 and 2003 [5-7]. LPAI A(H7N2) virus infected humans in the United States in 2002 and 2003 and the United Kingdom in 2007, and LPAI A(H7N3) virus infected humans in the United Kingdom in 2006 [8-10]. Additionally, LPAI A(H10N7) virus caused human infections in Egypt in 2004 [11]. Symptoms of these LPAI virus infections were mild, and no fatalities were reported. In contrast, avian influenza viruses of subtypes H5 and H7 are notorious for their abilities to mutate from LPAI viruses to viruses with increased virulence for poultry. These highly pathogenic avian influenza (HPAI) viruses typically emerge after introduction of an LPAI virus from wild birds into large commercial poultry flocks. This was exemplified by LPAI A(H7N1) viruses in Italy in 1999, LPAI A(H7N3) in Canada in 2004 and Chile in 2002, and LPAI A(H7N7) in Spain in 2009 that initiated outbreaks in poultry before mutating into the HPAI phenotype [12-15]. The emergence of a lineage of HPAI A(H5N1) virus in poultry and wild birds in Southeast Asia and its dissemination over a wide geographic region have been of particular concern. This virus now has evolved through drift and reassortment into distinct lineages, with reports of transmission to > 500 laboratory-confirmed infected humans and a case fatality rate of 60% [16]. The widespread presence of an influenza virus that can be transmitted to humans in regions with limited surveillance constitutes a permanent pandemic threat. So far, however, the prerequisite of efficient human-to-human transmission has not been met. Similarly, the unprecedented number of human infections during the HPAI A(H7N7) outbreak in The Netherlands in 2003 was cause for concern. This outbreak struck a large number of Dutch poultry farms and resulted in the culling of 30 million chickens. In addition, 89 virologically confirmed human cases, including 1 death and limited human-to-human transmission, were reported [17-19].

Avian influenza viruses must undergo molecular changes to adapt to a mammalian host [20]. This is a complex process that involves adaptation to human host cell factors to facilitate replication and transmission [21-25]. In addition to emergence of the HPAI phenotype by alteration of the HA cleavage site, LPAI virus replication in poultry can induce polygenic molecular changes. Examples are a deletion in the neuraminidase (NA) stalk region, which is associated with a shift in tissue tropism in chickens [26] and receptor binding site (RBS) mutations in HA, which affect host cell binding affinity [24,27]. Collectively, poultry can serve as an intermediate host and facilitate mammalian host adaptation of avian influenza viruses. Although increasing data that link human adaptation with specific molecular changes in the virus have become available, the molecular changes of influenza viruses during a large HPAI outbreak in poultry, along with cross-species transmission events, have never been documented. Such information could help to improve surveillance aimed at the timely detection of emergence of human adaptation markers to direct public health measures in the future. In this study, the 2003 HPAI A(H7N7) outbreak in the Netherlands was used as a model [18]. We combined virologic and epidemiologic data collected during the veterinary and medical outbreak control activities in order to identify the most likely source of infection for each person. We then compared viruses from humans and poultry to identify possible human adaptation markers. Precursors of potential pandemic variants with increased public health risk were detected in poultry prior to avian-to-human transmission events.

Materials and Methods

HPAI A(H7N7) outbreak

From individuals in contact with HPAI A(H7N7) and suffering from conjunctivitis and/or influenza-like illness, throat, nose, and conjunctiva swabs were collected into virus transport medium; the patients' written consent was obtained and a questionnaire was completed. After diagnostic testing [18], leftovers of clinical specimens were stored in a biobank at −80°C. For all human clinical specimens, the date of sampling, the first day of illness, demographic data, symptoms, and exposure data were stored in the human case finding database [18]. The veterinary database contained information on type of poultry, geographical location, date of sampling, date of culling, daily mortality data, and a unique farm identification number [28]. For each farm, the probable day of introduction of infection had been estimated previously from daily mortality data based on a back-calculation method [28].

Sequencing of viruses in human clinical samples

All A(H7N7) influenza virus-positive cases for which clinical samples were still available in a biobank (n=53) were used. Prior to direct sequencing of the virus in clinical

specimens, the viral load of the specimens was determined by quantitative influenza virus reverse transcription-PCR targeting a 94-bp matrix gene fragment [29]. Due to the low viral load in the 53 eye swabs, a direct sequencing experiment was designed that covered the HA receptor binding site and previously reported host range markers PB2 A199S, E627K, D701N, and K702R [19,30,31]. Therefore, three PCRs targeting HA fragment amino acids (aa) 160 to 239 and PB2 fragments aa 156 to 225 and aa 585 to 720 were developed by using the PSQ Assay Design package v1.0.6 (Biotage AB) for conventional Sanger sequencing. In addition, pyrosequencing targeting of the four PB2 host range markers was performed. Total RNA was isolated using a High Pure RNA Isolation kit (Roche) followed by reverse transcription using SuperScript III (Invitrogen) and cDNA amplification using HotstarTaq Mastermix (Qiagen). All PCR products were Sanger sequenced using an ABI 3730 sequencer, and PB2 host determinant markers were pyrosequenced using the PSQ MA96 platform (Biotage AB). Contig assembly and alignment were done using BioNumerics Package 6.5. Primers and PCR protocols are available upon request.

Sequencing of human virus isolates

Due to insufficient viral load in some clinical specimens, influenza virus culture supernatants of human A(H7N7) viruses were used to complement the data set. Routine culture of clinical specimens obtained from virologic confirmed A(H7N7) cases (n=89) using MDCK-I cells yielded 47 A(H7N7) virus isolates corresponding to 45 human cases [19]. Forty-three A(H7N7) virus isolates were obtained from eye swabs, three from throat swabs, and one from postmortem lung tissue from the fatal case. As initial screening of poultry viruses using whole-genome analysis did not reveal substantial differences when comparing nucleoprotein (NP), matrix, NS, PA, and PB1 genes from A(H7N7) viruses sampled early and late in the outbreak (data not shown), we focused on the more-diverse HA, NA, and PB2 genes. Hence, the full-length sequences of the HA, NA, and PB2 gene segments of all 47 human A(H7N7) virus isolates were determined. Total RNA was isolated using a High Pure RNA Isolation kit, followed by reverse transcription using ThermoScript (Invitrogen) and HA, NA, and PB2 gene amplification with the LongRange kit (Kapa Biosystems). After sequencing of the amplicons, trace files were assembled and aligned.

Data analysis

Human A(H7N7) sequences were compared with HA, NA, and PB2 gene sequences obtained from 184 infected poultry farms that were available from a separate study [32] supplemented by sequences obtained from 47 other poultry farms infected during the outbreak (combined sequences were obtained from 91% of the infected farms). Each A(H7N7) farm sequence is the consensus sequence of RNA extracted from a pool of five A(H7N7) tracheal tissue samples of poultry. The character data from a

concatenated full-length HA, NA, and PB2 gene nucleotide alignment were used to build a Maximum Parsimony network in BioNumerics using the Fitch method [33] with a greedy tree construction algorithm, followed by random branch swapping to find the optimal network topology. This method was chosen for its ability to link human A(H7N7) viruses to their suspected avian source of infection by using a minimum number of "evolutionary events" based on the simplest, most parsimonious explanation of an observation and free of specific evolutionary assumptions. The faithfulness of the entire network was determined by calculation of the global cophenetic correlation coefficient [34]. The significance of branches linking human viruses with poultry viruses was assessed by bootstrap analysis (1,000 iterations) and permutation resampling (200 resampling cycles) [35]. Finally, the constructed transmission network was validated by comparing the first day of human illness with the estimated day of introduction of A(H7N7) virus into a poultry farm for all farm-human links.

After validation of the transmission network, potential human adaptation markers were explored using the combined avian and human A(H7N7) sequence data set following categorization based on the following characteristics: first, the observed mutations were labeled poultry induced or human induced, depending on the host in which they were first identified. Subsequently, positively selected sites within the human data set were detected using the single-likelihood ancestor counting (SLAC), the random effect likelihood (REL), and the fixed effect likelihood (FEL) methods of the Datamonkey website [36]. Finally, HA and PB2 mutations were categorized based on the location in the gene segments and literature, while NA mutations were analyzed phenotypically for their effects on susceptibility to the antiviral drug oseltamivir carboxylate as described previously [37,38].

Nucleotide sequence accession numbers

GISAID accession numbers assigned to our sequences included the following: Human A(H7N7) sequence, EPI_ISL_90869-90912; poultry A(H7N7) sequences, EPI_ISL_68268-68352 and EPI_ISL_82373-82472 [32], supplemented by EPI_ISL_83984-84031.

Results

Direct sequencing of A(H7N7) virus fragments from clinical specimens

Direct sequencing of virus in the 53 influenza A(H7N7) virus eye swabs, which had an average virus load of ~3,000 genomic viral RNA copies/ml, was successful for 44 human clinical samples, yielding HA receptor binding site fragments (n=43) and sequences covering previously reported host determinant markers in PB2 at residue 199 (n=43) and the residues 627, 701, and 702 (n=39). Sequence comparison with influenza A/Netherlands/33/03 (H7N7) virus obtained from the first human

conjunctivitis case demonstrated the absence of PB2 host determinant markers A199S, E627K, D701N, and K702R but the presence of an amino acid variation in the HA receptor binding site (HA V223I [n=1] and HA subpopulation V223A [n=1]) (table 1).

Sequencing of A(H7N7) virus isolates

To expand the search for human adaptation markers beyond these known positions, influenza virus isolates were utilized to obtain sequences for the entire HA, NA, and PB2 genes corresponding with 45 human cases. Compared with A/Netherlands/33/03 (H7N7) virus, amino acid variations were observed at 20 different HA codons, 23 NA codons, and 23 PB2 codons.

Table 1. Amino acid substitutions found in human A(H7N7) viruses compared to their avian source of infection[a]

		HA gene	NA gene	PB2 gene
Virus isolates (n=48) from human conjunctivitis cases	Fragment sequenced	aa 1-562	aa 1-471	aa 1-759
	Mutations detected	R65K, A156S, V223A, V223I, D264G, E279K, E287G, R504K, E505K, D515N	A14T, V19A, Q44K, N57S, N67S, V115I, M174V, V263I, E267V, I275T, N284I, N338S, V426I, V456I	V89M, T117P, V139I, V167I, C409R, M467L, I562T, D567N, K586R, P620Q, N711K
Clinical specimen (n=44) from human conjunctivitis cases	Fragment sequenced	aa 160-239		aa 156-225, 585-720
	Mutations detected	V223A, V223I		N711K
Virus isolate (n=1) from human fatal case [24]	Fragment sequenced	aa 1-562	aa 1-471	aa 1-759
	Mutations detected	K416R		E627K

[a] Underlined mutations are associated with human adaptation based on literature or selection pressure analysis

Detection of sources of human A(H7N7) virus infection

To discriminate mutations that emerged in poultry from mutations that were detected in viruses from humans only, the data set was supplemented with 232 veterinary A(H7N7) HA, NA, and PB2 sequences, representing 231 poultry farms, to allow identification of sources of human A(H7N7) virus infection. The high level of genetic diversity observed in avian influenza A(H7N7) viruses isolated during the 2003 HPAI outbreak in the Netherlands [32] was used to build a maximum parsimony transmission network to identify chains of transmission and probable sources of infection of the sequenced human A(H7N7) cases. Since initial analysis with 231 poultry farms did not identify clustering of human A(H7N7) viruses with avian A(H7N7) viruses containing a deletion in the NA stalk region (n=17), these were removed from the data set. Removal of these sequences from the data set did not affect the results

of the study. Moreover, as a result of observed nucleotide variations in the NA stalk region, the final transmission network was constructed with concatenated HA, NA, and PB2 sequences from avian A(H7N7) viruses containing full-length NA sequences, representing 214 poultry farms, to maximize the network resolution (Fig. 1). The global cophenetic correlation of the transmission network was 99%, and the branches connecting human cases (n=32) with farm sequences were supported through a bootstrap analysis and permutation resampling by 100%.

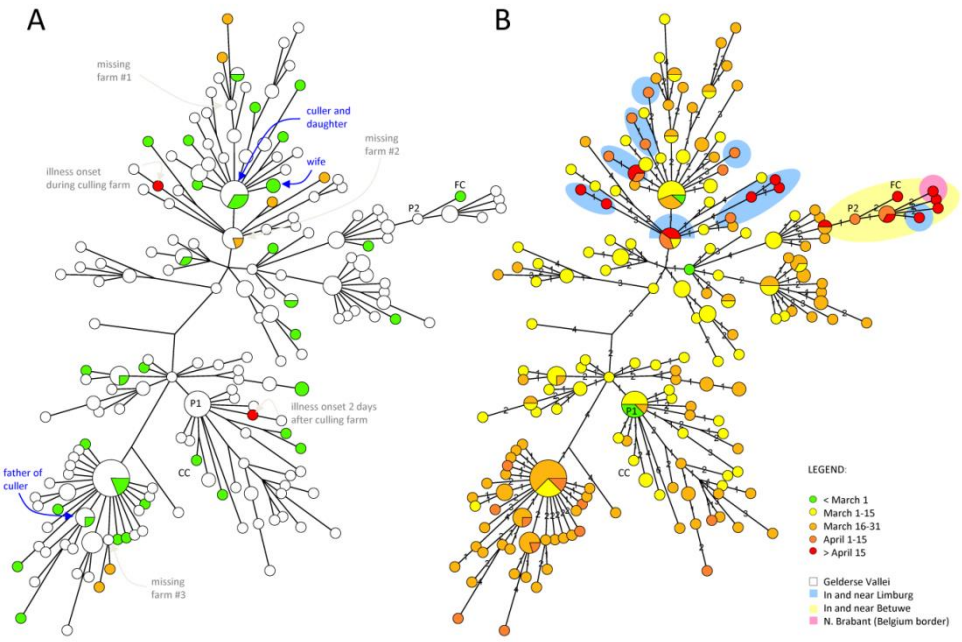


Figure 1. Transmission network built with concatenated full-length HA, NA, and PB2 gene segments of the A(H7N7) virus outbreak, illustrating the position of 45 human A(H7N7) virus infections (colored) in relation to 215 A(H7N7) virus-infected farms (white) (A) and the nucleotide differences between the A(H7N7) sequences supplemented by information on sample period and farm location (B). (A) Human A(H7N7) virus infections in agreement with epidemiologic data are shown in green. The positions of the three secondary human cases within the network are indicated in blue. Human infections (n=7) with onset of illness before their avian sources were estimated to be infected are indicators for missing poultry data (yellow). Two human viruses within the transmission network could not have caused human-to-avian transmissions based on infection dates (red). (B) After notification of the A(H7N7) outbreak on 28 February 2003, multiple poultry A(H7N7) virus clusters and chains of transmission emerged that spread over time (colored circles) and location (colored areas). The locations of human cases refer to the locations of their suspected avian source of infection. The nucleotide changes between specific A(H7N7) virus strains are shown on the branches. CC, A/Netherlands/033/03, FC, A/Netherlands/219/03; P1, A/chicken/Netherlands/01/03; P2, A/chicken/Netherlands/03010132/03.

The genetic diversity observed at 214 poultry farms, represented by 149 distinct avian A(H7N7) virus sequences in the transmission network, illustrates the emergence of multiple avian A(H7N7) virus clusters and chains of transmission during the outbreak. Human A(H7N7) viruses are interspersed with avian viruses, reflecting the virus diversity in poultry, allowing identification of probable sources of human A(H7N7) virus infection. In addition to the three human secondary cases, 22% (10/45) of the human A(H7N7) viruses were identical to their presumed avian source of infection, while 71% (32/45) of the human A(H7N7) viruses displayed genetic variation from their avian source. Thirty of these 32 human A(H7N7) viruses are at the tips of the network, illustrating that the majority of human cases were dead-end A(H7N7) virus infections. One A(H7N7)-positive culler infected both his wife and daughter, and another culler infected his father (Fig. 1A). Of the probable avian sources of the remaining 42 human A(H7N7) virus infections, 35 (83%) persons had their first symptoms after the estimated date of introduction of virus on the farm (back-calculated from mortality data [39]. The avian source of infection of three different clusters of human cases (n=7) did fit molecularly, but infection chronology was reversed (Fig. 1A, yellow dots). As 9% (n=24) of the infected poultry farms were missing in our analysis, the farms of infection origin of these human clusters could be missed. Furthermore, two human cases with unique virus sequences were found within the transmission network (Fig. 1A, red dots). Although these could represent human-to-avian transmissions, the infection chronology demonstrated avian-to-human transmission. A phylogenetic tree constructed with Neighbor Joining and representative A(H7N7) virus isolates demonstrated identical A(H7N7) virus relations (see Fig. S1 in the supplemental material).

Acquisition of human adaptation markers in humans

The observed A(H7N7) virus diversity among viruses in poultry could mask the occurrence of possible human adaptation during infection of humans when comparing grouped human and avian sequences. Therefore, we assessed which differences were observed between A(H7N7) sequences from human influenza isolates with those from their most likely source, based on the transmission network. A total of 37 amino acid substitutions were identified between human A(H7N7) viruses and their avian source of infection (HA [n=11], NA [n=14], and PB2 [n=12]) (table 1). Three of these mutations (HA V223A, HA V223I, and PB2 N711K) were detected in both the virus isolate and the corresponding human clinical specimen, while the presence of the remaining 34 mutations could not be confirmed by direct sequence analysis because there were no leftover clinical specimens in our biobank, the viral load was too low, or the site was not covered by direct sequencing.

Selection pressure analysis, using the combined avian and human sequence data set to determine whether the potential human adaptation markers presented in table 1 emerged under (host) selection pressure, only recognized HA codon 223 as a positive selection site (REL method; Bayes Factor, 1.72E+08). Comparison of human A(H7N7) virus sequences with previously identified host-specific residues [23,40,41], detected PB2 mutation D567N in addition to the human adaptation marker PB2 E627K, which was identified in the fatal human case.

During the A(H7N7) virus outbreak, the neuraminidase inhibitor oseltamivir was used for antiviral treatment and prophylaxis of A(H7N7) virus infection. Human influenza virus isolates (n=15) with and without NA mutations compared to their avian source of infection were screened for emergence of oseltamivir resistance, but no resistance or reduced susceptibility was observed (mean 50% inhibitory concentration [IC_{50}], 0.46 nM; range, 0.12 to 0.70 nM). Moreover, one human A(H7N7) isolate with NA mutations N67S, V115I, and I275T displayed 4-fold-enhanced sensitivity against oseltamivir compared with the mean (IC50, 0.12 nM).

Detection of potential human adaptation markers in poultry

Based on a selection of positions that are potentially relevant for human adaptation of the avian virus, we detected variations in PB2 codons 292, 567, 701, and 714 of poultry A(H7N7) viruses that were previously identified to be host-specific residues [23,40,41]. Mutation I292V is not associated with host adaptation but represents natural variation within avian sequences [41]. Mutation D567Y is not a known human adaptation marker but is present in A/swine/Shizuoka/120/97(H3N2) (GenBank). Mutation D701N was previously associated with efficient influenza virus replication and transmission in mammalian species, similar to PB2 E627K [42,43]. This mutation did emerge and subsequently spread to four poultry farms. PB2 mutation S714I is associated with increased pathogenicity for mammals and emerged in a chicken farm before it spread to one of the cullers and another chicken farm [42].

We also tracked the accumulation of mutations leading up to the fatal human infection, previously described by de Wit and colleagues [24]. Their study identified substitutions HA A143T and PB2 E627K, attributable to increased pathogenicity in the mouse model, in addition to four replication-enhancing NA substitutions. Mapping of markers in farm sequence data shows the emergence and spread of these mutations during the A(H7N7) virus outbreak in poultry (Fig. 2). HA substitution A143T emerged during the outbreak at three different chicken farms and spread to subsequent farms. For two of these farms, this mutation emerged in the first week after the outbreak was identified and authorities were notified. In addition to the virus isolated from the fatal A(H7N7) case, a 49-year-old owner of a layer farm who developed conjunctivitis

without fever appeared to be infected with virus variant HA A143T. He was infected by an A(H7N7) virus identical to the virus variant HA A143T obtained from his 17,500-layer hen farm, except for one silent PB2 nucleotide substitution. The accumulation of NA mutations T442A and P458S (73 farms) supplemented by NA A346V (30 farms) and NA N308S (11 farms) in poultry eventually gave rise to an A(H7N7) variant with enhanced NA activity that was transmitted to the fatal human case [24]. The PB2 E627K mutation detected in the virus isolated from the fatal A(H7N7) case was not detected in poultry.

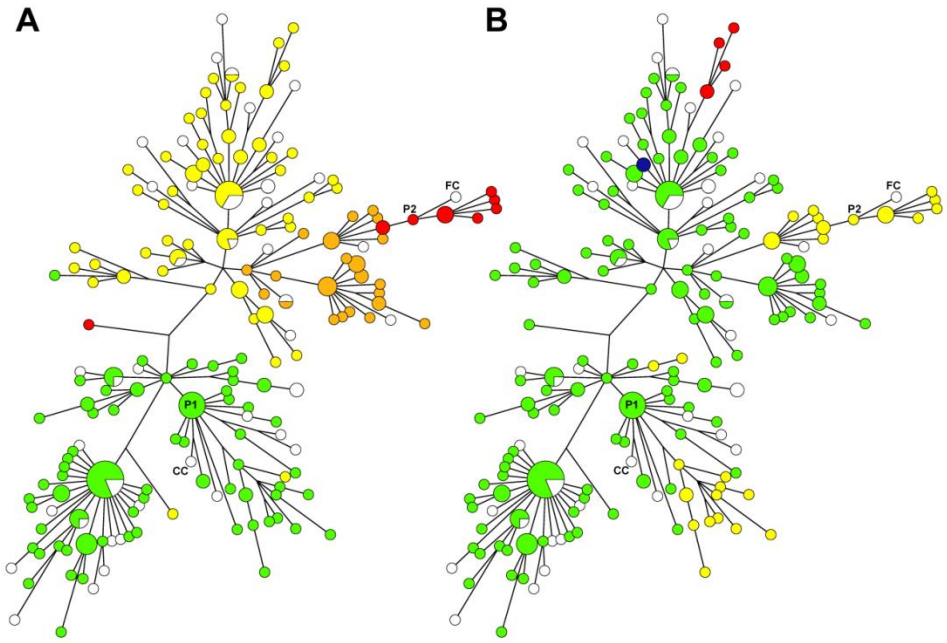


Figure 2. The accumulation of virulence and human adaptation markers in poultry farms (green) per gene segment, illustrating the accumulation of NA mutations T442A and P458S (yellow), T442A, P458S plus A346V (orange) and T442A, P458S, A346V plus N308S (red) (A) and independent emergence of HA A143T variants (yellow) and emergence of PB2 D701N (red) and PB2 S714I (blue) (B). Human viruses are represented by open circles. CC, A/Netherlands/033/03; FC, A/Netherlands/219/03; P1, A/chicken/Netherlands/01/03; P2, A/chicken/ Netherlands/03010132/03.

Discussion

In contrast to the ongoing outbreak of HPAI A(H5N1), in which human infections have occurred as isolated cases or in a small cluster (n < 8), the HPAI A(H7N7) outbreak that struck the Netherlands in 2003 infected at least 89 humans, including three secondary

cases and one death [19,44]. The culling of >30 million birds combined with control measures did not prevent A(H7N7) spread to a total of 255 poultry farms in The Netherlands and 9 farms in the neighboring countries of Belgium (n=8) and Germany (n=1) [45,46]. By comprehensive virologic analysis of combined veterinary and human A(H7N7) samples obtained during this large avian influenza outbreak, this study provides unique knowledge on the acquisition of human adaptation markers and the public health risk associated with an HPAI virus outbreak in poultry. We demonstrated the polygenic accumulation and farm-to-farm spread of known virulence and human adaptation markers in A(H7N7)-infected poultry farms following farm-to-human transmission. These include the independent emergence of HA mutants with increased replication kinetics, accumulation of NA mutations facilitating efficient release of virus particles from the host cell, and farm-to-farm spread of virus variants harboring mammalian host determinant D701N and S714I in PB2 [3,24]. The emergence of PB2 E627K detected in the virus obtained from the fatal case could not be assigned to the avian source of infection and might have emerged in the human host. Similarly, PB2 mutation D567N and HA V223I/A in the RBS were detected in human A(H7N7) viruses only.

Sequence analysis of A(H7N7) viruses obtained from human cases was performed on all available clinical samples and virus isolates. For 26 cases, both a clinical sample and virus isolate were present, allowing comparison of direct sequencing results with full-length sequences obtained from virus isolates (table 1). This demonstrated an absence of virus adaptation to MDCK-I cells within the regions sequenced. Moreover, HA RBS sequences were preserved during virus isolation using MDCK-I cells. Since the RBS is considered to be one of the most variable regions of the influenza virus and prone to host cell adaptation, we reasoned that molecular variation shown in table 1 is the result of human adaptation. We cannot, however, fully exclude that some variation was MDCK culture induced. Nonetheless, human adaptation markers PB2 S714I and HA V223I/A were confirmed by direct sequencing.

Although popular phylogenetic analysis uses parametric techniques of maximum likelihood and Bayesian Markov chain Monte Carlo methods, such results can become inconsistent when evolutionary rates vary. Especially when using short time scales, e.g., during local virus outbreaks, the combination of rapidly evolving viruses with non-evolving variants challenges the molecular clock model, resulting in statistically inconsistent results [47]. Because maximum parsimony analysis does not assume a specific distribution and is best suited for analyzing sequences that are quite similar, this simple approach is the method of choice for analyzing defined virus outbreaks, such as the HPAI A(H7N7) virus outbreak. By addition of statistical parameters, like the global cophenetic correlation coefficient, bootstrap, and permutation resampling

support, the potential avian sources of human A(H7N7) virus infection could be identified with high significance [34,35].

Genetic variation was not associated with a diagnostic delay, as the difference between the first day of illness and day of consultation was, on average, 1.3 days for patients infected with either identical viruses (n=14), infected with viruses with synonymous mutations (n=8), or infected with viruses with nonsynonymous mutations (n=23). Furthermore, one A(H7N7) virus obtained from a culler was detected in his daughter (100% identical) and wife (1 synonymous substitution, 2 days later). From the wife and a culler, influenza A(H7N7) virus isolates from eye and throat swabs were available for sequence analysis. Sequence comparison showed no variation between viruses obtained from eye or throat. The latter can be explained by replication-independent spread of virus from the eye to the throat [48,49]. The above illustrates limited A(H7N7) virus adaptation during human infection, probably hampered by suboptimal virus replication in the human eye combined with oseltamivir treatment. However, the prolonged 24-day A(H7N7) course of illness of the veterinarian that eventually died could have facilitated virus adaptation to humans by means of the mutation PB2 E627K. The phenotype of virus variant PB2 E627K underlines the importance of the viral RNA replication complex (NP, PA, PB1, and PB2) in human adaptation of avian influenza viruses [24,30]. Available PA sequences from the A(H7N7) outbreak revealed amino acid variation of PA codon 666 [24]. Although NP and PB1 amino acid sequences were identical, variation (L62P) was observed within the full-length PB1-F2 proteins. Interestingly, this PB1-F2 variant was present in the A(H7N7) virus obtained from the culler and it subsequently spread to both his wife and daughter. However, experiments are needed to characterize the effect of PB1-F2 sequence variation on viral pathogenesis.

The HPAI A(H7N7) virus outbreak was most probably initiated by the introduction of an LPAI virus in poultry, following multiple weeks of silent replication and spread until it was identified and authorities were notified at the first poultry farm on 28 February 2003 [19,32]. The HPAI A(H7N7) virus subsequently infected 255 poultry farms in the Netherlands and 8 poultry farms in Belgium [46]. Of interest is the long farm-to-farm transmission distance of the A(H7N7) variant harboring the four NA substitutions N308S, A346V, T442A, and P458S, which spread on average >10 km (Fig. 1B). These variants initiated an A(H7N7) outbreak in Belgium before reemerging in a hobby farm in the southern part of the Netherlands [46]. A similar observation was made by Shi and colleagues, who associated NA mutations with rapid geographical spread of avian influenza A(H9N2) virus in China [50]. The exceptional level of human exposure resulting from active culling and screening activities over a wide geographical region during 2.5 months resulted in 89 human A(H7N7) cases. Our results demonstrate that

viruses obtained from human A(H7N7) conjunctivitis cases reflect the virus diversity generated in poultry. Of greater concern, we observed the polygenic accumulation and spread of known virulence and human adaptation markers in A(H7N7) virus-infected poultry farms following farm-to-human transmission. This implies that HPAI viruses with pandemic potential can emerge directly from poultry without the need to adapt in the human host. The dynamics of emerging virus variants in poultry with increased virulence and enhanced transmission characteristics provide a challenge during culling activities. Since the public health risk of an avian influenza virus outbreak in poultry can rapidly change, we recommend virologic monitoring of poultry in addition to humans during the course of an outbreak in poultry.

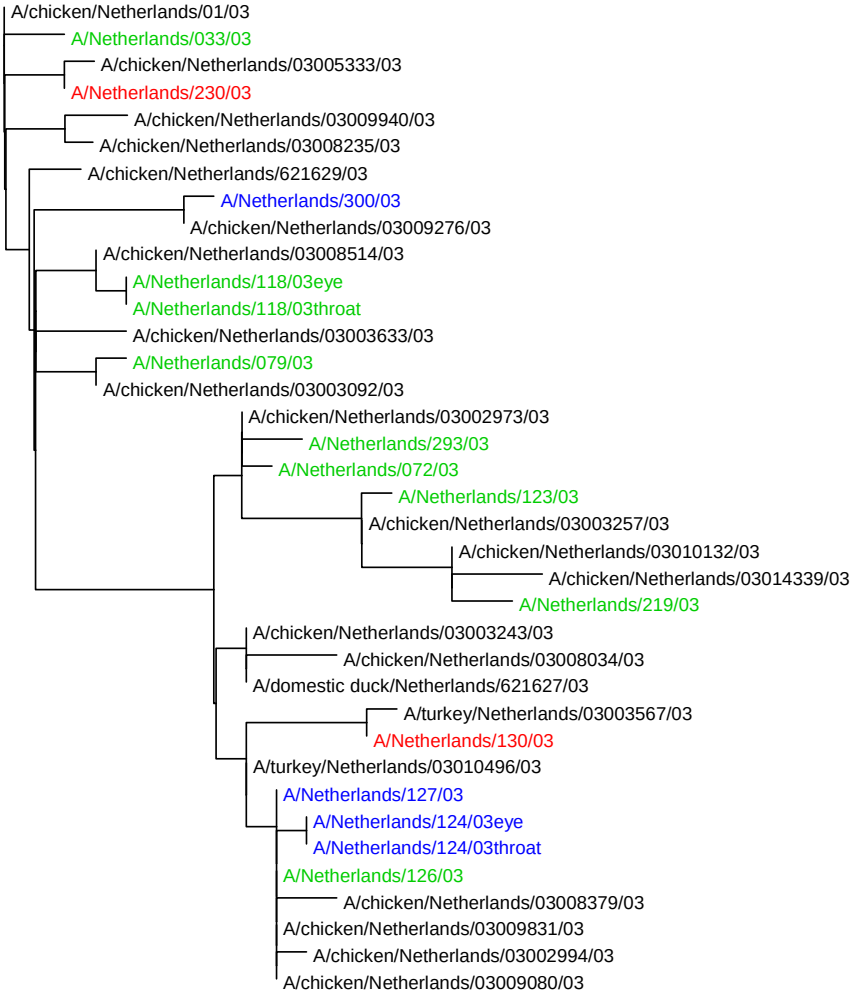
Although the emergence and spread of viruses with multiple potential human adaptation mutations were detected in poultry, the impact of detecting specific mutations for public health risk assessment is uncertain. Nonetheless, this study suggests that the effective control of influenza virus in poultry can prevent avian influenza viruses from acquiring all their prerequisites to become human tropic, in addition to the reduction of human exposure [51]. Increased awareness by veterinary and medical authorities is needed to design more efficient surveillance, diagnostic algorithms, and prepandemic planning strategies.

Acknowledgements

This work was funded by the Centers for Disease Control and Prevention (Avian Influenza Cooperative Research Centers: Studies at the Human-Animal Interface project; grant1U19CI000404-02) and the Dutch Ministry of Economic Affairs (Impulse Veterinary Avian Influenza Research in the Netherlands program). We are grateful to Gert-Jan Godeke, Ilse Zutt and Shireen Jenny for virologic technical assistance.

Supplemental material

Figure S1. Neighbor Joining tree of concatenated HA, NA and PB2 sequences, rooted on A/chicken/Netherlands/01/03, illustrates the molecular relation of representative influenza A(H7N7) strains during the A(H7N7) outbreak similar to the transmission network.



References

1. Webby R, Hoffmann E, Webster R (2004) Molecular constraints to interspecies transmission of viral pathogens. Nat Med 10: S77-81.
2. Miotto O, Heiny AT, Albrecht R, Garcia-Sastre A, Tan TW, et al. (2010) Complete-proteome mapping of human influenza A adaptive mutations: implications for human transmissibility of zoonotic strains. PLoS One 5: e9025.
3. Steel J, Lowen AC, Mubareka S, Palese P (2009) Transmission of influenza virus in a mammalian host is increased by PB2 amino acids 627K or 627E/701N. PLoS Pathog 5: e1000252.
4. Palese P (2004) Influenza: old and new threats. Nat Med 10: S82-87.
5. Peiris M, Yam WC, Chan KH, Ghose P, Shortridge KF (1999) Influenza A H9N2: aspects of laboratory diagnosis. J Clin Microbiol 37: 3426-3427.
6. Butt KM, Smith GJ, Chen H, Zhang LJ, Leung YH, et al. (2005) Human infection with an avian H9N2 influenza A virus in Hong Kong in 2003. J Clin Microbiol 43: 5760-5767.
7. Guo Y, Li J, Cheng X (1999) [Discovery of men infected by avian influenza A (H9N2) virus]. Zhonghua Shi Yan He Lin Chuang Bing Du Xue Za Zhi 13: 105-108.
8. Eames KT, Webb C, Thomas K, Smith J, Salmon R, et al. Assessing the role of contact tracing in a suspected H7N2 influenza A outbreak in humans in Wales. BMC Infect Dis 10: 141.
9. CDC (2004) Update: influenza activity--United States and worldwide, 2003-04 season, and composition of the 2004-05 influenza vaccine. MMWR Morb Mortal Wkly Rep 53: 547-552.
10. Nguyen-Van-Tam JS, Nair P, Acheson P, Baker A, Barker M, et al. (2006) Outbreak of low pathogenicity H7N3 avian influenza in UK, including associated case of human conjunctivitis. Euro Surveill 11: E060504 060502.
11. WHO (2004) Avian Influenza Virus A (H10N7) Circulating among Humans in Egypt. EID Weekly Updates: Emerging and Reemerging Infectious Diseases, Region of the Americas Vol. 2.
12. Capua I, Mutinelli F, Pozza MD, Donatelli I, Puzelli S, et al. (2002) The 1999-2000 avian influenza (H7N1) epidemic in Italy: veterinary and human health implications. Acta Trop 83: 7-11.
13. Hirst M, Astell CR, Griffith M, Coughlin SM, Moksa M, et al. (2004) Novel avian influenza H7N3 strain outbreak, British Columbia. Emerg Infect Dis 10: 2192-2195.
14. Rojas H, Moreira R, Avalos P, Capua I, Marangon S (2002) Avian influenza in poultry in Chile. Veterinary Record 151: 188.

15. Iglesias I, Martinez M, Munoz MJ, De La Torre A, Sanchez-Vizcaino JM (2010) First case of highly pathogenic avian influenza in poultry in Spain. Transboundary and Emerging Diseases 57: 282-285.
16. WHO (2011) Cumulative Number of Confirmed Human Cases of Avian Influenza A/(H5N1) Reported to WHO. Confirmed Human Cases of Avian Influenza A(H5N1).
17. Du Ry van Beest Holle M, Meijer A, Koopmans M, de Jager CM (2005) Human-to-human transmission of avian influenza A/H7N7, The Netherlands, 2003. Euro Surveill 10: 264-268.
18. Koopmans M, Wilbrink B, Conyn M, Natrop G, van der Nat H, et al. (2004) Transmission of H7N7 avian influenza A virus to human beings during a large outbreak in commercial poultry farms in the Netherlands. Lancet 363: 587-593.
19. Fouchier RA, Schneeberger PM, Rozendaal FW, Broekman JM, Kemink SA, et al. (2004) Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome. Proc Natl Acad Sci U S A 101: 1356-1361.
20. Beare AS, Webster RG (1991) Replication of avian influenza viruses in humans. Arch Virol 119: 37-42.
21. Stevens J, Blixt O, Tumpey TM, Taubenberger JK, Paulson JC, et al. (2006) Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. Science 312: 404-410.
22. Parrish CR, Holmes EC, Morens DM, Park EC, Burke DS, et al. (2008) Cross-species virus transmission and the emergence of new epidemic diseases. Microbiol Mol Biol Rev 72: 457-470.
23. Naffakh N, Massin P, Escriou N, Crescenzo-Chaigne B, van der Werf S (2000) Genetic analysis of the compatibility between polymerase proteins from human and avian strains of influenza A viruses. J Gen Virol 81: 1283-1291.
24. de Wit E, Munster VJ, van Riel D, Beyer WE, Rimmelzwaan GF, et al. (2010) Molecular determinants of adaptation of highly pathogenic avian influenza H7N7 viruses to efficient replication in the human host. J Virol 84: 1597-1606.
25. Neumann G, Kawaoka Y (2006) Host range restriction and pathogenicity in the context of influenza pandemic. Emerg Infect Dis 12: 881-886.
26. 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.
27. Wan H, Perez DR (2007) Amino acid 226 in the hemagglutinin of H9N2 influenza viruses determines cell tropism and replication in human airway epithelial cells. J Virol 81: 5181-5191.

28. Bos ME, Van Boven M, Nielen M, Bouma A, Elbers AR, et al. (2007) Estimating the day of highly pathogenic avian influenza (H7N7) virus introduction into a poultry flock based on mortality data. Vet Res 38: 493-504.
29. Jonges M, Liu WM, van der Vries E, Jacobi R, Pronk I, et al. (2010) Influenza virus inactivation for studies of antigenicity and phenotypic neuraminidase inhibitor resistance profiling. J Clin Microbiol 48: 928-940.
30. Finkelstein DB, Mukatira S, Mehta PK, Obenauer JC, Su X, et al. (2007) Persistent host markers in pandemic and H5N1 influenza viruses. J Virol 81: 10292-10299.
31. Li Z, Chen H, Jiao P, Deng G, Tian G, et al. (2005) Molecular basis of replication of duck H5N1 influenza viruses in a mammalian mouse model. J Virol 79: 12058-12064.
32. Bataille A, van der Meer F, Stegeman A, Koch G (2011) Evolutionary analysis of inter-farm transmission dynamics in a highly pathogenic avian influenza epidemic. PLoS Pathog 7: e1002094.
33. Fitch WM (1971) Toward Defining the Course of Evolution: Minimum Change for a Specific Tree Topology. Systematic Zoology 20: 406-416.
34. Sokal RR, Rohlf FJ (1962) The comparison of dendrograms by objective methods. Taxon 11: 33-40.
35. Pouseele H, Vauterin P, Vauterin L (2011) A resampling strategy for reliable network construction. Mol Phylogenet Evol 60: 273-286.
36. Pond SL, Frost SD (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21: 2531-2533.
37. Jonges M, van der Lubben IM, Dijkstra F, Verhoef L, Koopmans M, et al. (2009) Dynamics of antiviral-resistant influenza viruses in the Netherlands, 2005-2008. Antiviral Res 83: 290-297.
38. Potier M, Mameli L, Belisle M, Dallaire L, Melancon SB (1979) Fluorometric assay of neuraminidase with a sodium (4-methylumbelliferyl-alpha-D-N-acetylneuraminate) substrate. Anal Biochem 94: 287-296.
39. Bos ME, Nielen M, Koch G, Bouma A, De Jong MC, et al. (2009) Back-calculation method shows that within-flock transmission of highly pathogenic avian influenza (H7N7) virus in the Netherlands is not influenced by housing risk factors. Prev Vet Med 88: 278-285.
40. Chen GW, Chang SC, Mok CK, Lo YL, Kung YN, et al. (2006) Genomic signatures of human versus avian influenza A viruses. Emerg Infect Dis 12: 1353-1360.
41. Miotto O, Heiny A, Tan TW, August JT, Brusic V (2008) Identification of human-to-human transmissibility factors in PB2 proteins of influenza A by large-scale mutual information analysis. BMC Bioinformatics 9 Suppl 1: S18.
42. Gabriel G, Dauber B, Wolff T, Planz O, Klenk HD, et al. (2005) The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. Proc Natl Acad Sci U S A 102: 18590-18595.

43. de Jong MD, Simmons CP, Thanh TT, Hien VM, Smith GJ, et al. (2006) Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia. Nat Med 12: 1203-1207.
44. Tweed SA, Skowronski DM, David ST, Larder A, Petric M, et al. (2004) Human illness from avian influenza H7N3, British Columbia. Emerg Infect Dis 10: 2196-2199.
45. Capua I, Alexander DJ (2004) Avian influenza: recent developments. Avian Pathol 33: 393-404.
46. van den Berg T, Houdart P (2008) Avian influenza outbreak management: action at time of confirmation, depopulation and disposal methods; the 'Belgian experience' during the H7N7 highly pathogenic avian influenza epidemic in 2003. Zoonoses Public Health 55: 54-64.
47. Kolaczkowski B, Thornton JW (2004) Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature 431: 980-984.
48. Belser JA, Wadford DA, Xu J, Katz JM, Tumpey TM (2009) Ocular infection of mice with influenza A (H7) viruses: a site of primary replication and spread to the respiratory tract. J Virol 83: 7075-7084.
49. Bitko V, Musiyenko A, Barik S (2007) Viral infection of the lungs through the eye. J Virol 81: 783-790.
50. Shi HY, Liu XF (2006) [Molecular mechanism affecting route of transmission for H9N2 subtype AIV]. Wei Sheng Wu Xue Bao 46: 48-54.
51. Rabinowitz P, Perdue M, Mumford E (2010) Contact variables for exposure to avian influenza H5N1 virus at the human-animal interface. Zoonoses Public Health 57: 227-238.

Chapter 3.2

Wind-Mediated Spread of Low-Pathogenic Avian Influenza Virus into the Environment During Outbreaks at Commercial Poultry Farms

Marcel Jonges[1,2], Jeroen van Leuken[1,3], Inge Wouters[3], Guus Koch[4], Adam Meijer[1], Marion Koopmans[1,2]

PLoS One. 2015 May 6;10(5):e0125401

[1] Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, The Netherlands

[2]Department of Viroscience, Erasmus MC, Rotterdam, The Netherlands

[3] Institute for Risk Assessment Sciences, Faculty of Veterinary Sciences, Utrecht University, Utrecht, The Netherlands

[4] Central Veterinary Institute, Wageningen University & Research Center, Lelystad, The Netherlands

Abstract

Avian influenza virus-infected poultry can release a large amount of virus-contaminated droppings that serve as sources of infection for susceptible birds. Much research so far has focused on virus spread within flocks. However, as fecal material or manure is a major constituent of airborne poultry dust, virus-contaminated particulate matter from infected flocks may be dispersed into the environment. We collected samples of suspended particulate matter, or the inhalable dust fraction, inside, upwind and downwind of buildings holding poultry infected with low-pathogenic avian influenza virus, and tested them for the presence of endotoxins and influenza virus to characterize the potential impact of airborne influenza virus transmission during outbreaks at commercial poultry farms. Influenza viruses were detected by RT-PCR in filter-rinse fluids collected up to 60 meters downwind from the barns, but virus isolation did not yield any isolates. Viral loads in the air samples were low and beyond the limit of RT-PCR quantification except for one in-barn measurement showing a virus concentration of 8.48×10^4 genome copies/m^3. Air samples taken outside poultry barns had endotoxin concentrations of ~50 EU/m^3 that declined with increasing distance from the barn. Atmospheric dispersion modeling of particulate matter, using location-specific meteorological data for the sampling days, demonstrated a positive correlation between endotoxin measurements and modeled particulate matter concentrations, with an R^2 varying from 0.59 to 0.88. Our data suggest that areas at high risk for human or animal exposure to airborne influenza viruses can be modeled during an outbreak to allow directed interventions following targeted surveillance.

Introduction

Avian influenza A viruses are highly heterogeneous, with varying pathogenicity across different species. They are classified into subtypes based on the surface glycoproteins haemagglutinin (HA) and neuraminidase (NA). Pathogenicity of the virus in chickens is related to the pathotype: low-pathogenic avian influenza (LPAI) viruses can contain any type of HA, while highly pathogenic avian influenza (HPAI) viruses invariably contain H5 or H7 [1]. Mortality is a prominent sign of HPAI-infected flocks, whereas LPAI-infected flocks show milder or even subclinical signs that can wax and wane, making LPAI more difficult to detect.

In wild birds, avian influenza virus is primarily transmitted through fecally contaminated surface water in shared aquatic habitats. In these habitats, the viruses can persist for extended periods, depending on water temperature and physico-chemical characteristics [2]. In domesticated birds, or poultry, HPAI viruses are typically found in both feces and respiratory secretions, while LPAI viruses are mainly shed through the enteric route [3]. Virus-contaminated droppings serve as source of infection for susceptible birds, and influenza viruses can remain infectious for many days in poultry litter [4,5]. Dispersal of infectious material into the environment may occur through ventilation of virus-contaminated dust. In commercial poultry operations, concentrations of airborne dust are high and include a large component of fecal material along with food, dander (skin material), feather material, and micro-organisms [6]. Detection of influenza A virus in air measurements collected within farms suggest that particulate matter from infected poultry may play a role in avian influenza virus transmission to humans and birds, and other animals [7,8].

One of the routes for pathogen transmission is through dispersal into outdoor air. Viruses may be dispersed as single particles or by using other particles (particulate matter) as a vehicle [9-11]. Recently, Ypma et al. estimated that wind direction could explain about 18% of the total transmission of avian influenza between farms during an outbreak of influenza A(H7N7) virus in 2003 [12]. However, wind direction alone does not quantify the amount of pathogen transmitted to a certain distance, as wind speed is another important factor [13]. Atmospheric dispersion models (ADMs) take these and other factors into account and have been applied to analyse the correlation between airborne pathogen transmission and the incidence of disease in the nearby surroundings for *Legionella pneumophila* [14], foot-and-mouth disease [15], *Coxiella burnetii* [16], and avian influenza virus [17]. Although the ADM data are suggestive of airborne pathogen dispersion, laboratory data have not yet confirmed that airborne avian influenza viruses are indeed detectable in the air downwind of a source.

Previously, we demonstrated that farm-to-farm spread of avian influenza viruses was associated with accumulated mutations that increase the public health risk of HPAI

A(H7N7) viruses [18]. In addition, LPAI virus replication in poultry may trigger the emergence of an HPAI variant by alteration of the HA cleavage site, facilitating systemic infections. The consequent importance of early control of outbreaks became very clear with the 2013 emergence of avian influenza A(H7N9) viruses in China. Despite causing severe illness in humans, these viruses have the LPAI phenotype, making it hard to identify the avian sources and rendering humans as sentinels [19-21]. Gaining more insight into the transmission routes of avian influenza will help provide a more solid basis for current outbreak response strategies, and thereby could eventually reduce the public health risk associated with outbreaks.

In this study, we collected samples of suspended particulate matter, or inhalable dust fraction, inside, upwind and at several distances downwind of buildings holding poultry infected with LPAI. The samples were tested for the presence of influenza virus and for endotoxins, a marker for microbial exposure of poultry and livestock, since they have a high presence in commercial farms and can be quantified in the adjacent outdoor air [22]. We hypothesized that particulate matter may be used as a substitute for dispersion monitoring of avian influenza transmitted into the environment during outbreaks [10]. Consequently, airborne microbial exposure was determined by measuring endotoxin concentrations at different distances from farms and compared with an ADM to test the applicability of this model for the rapid characterization of a geographical region exposed during future outbreaks of avian influenza.

Materials and Methods

Farm description

At the following five LPAI-infected farms and one control farm, air samples were taken at multiple distances from the farms.

Farm 1 was a naturally ventilated organic chicken farm composed of an indoor-housed flock and a free-range flock. An LPAI A(H7N7) virus infection was detected by targeted investigation following signs of reduced food consumption, diarrhea, and limited growth of 8900 22-week old chickens. All chickens were culled on the day of confirmation of virus presence, in accordance with European guidelines for avian influenza virus subtypes H5 and H7 in commercial poultry (EU directive 92/40/EEC). Outdoor air sampling was initiated approximately six hours after the culling, at locations upwind and downwind of the farm.

Farm 2 was a mechanically ventilated turkey farm with 20,600 one-month old turkey chicks. It was tested for the presence of influenza virus because of negative health reports, showing infection with LPAI A(H9N2) virus. Nine days after sampling of the birds, air sampling was performed at locations upwind and downwind of the farm. No

control measures were applied following outbreak confirmation, in accordance with EU guidelines.

Farm 3 was a bird-breeding farm that also housed various mammals and reptiles. Air was sampled upwind and downwind of 83 healthy-appearing wild swans that had been captured and were destined for export to a foreign zoo. The swans were quarantined following a positive screen for LPAI A(H5N2) virus performed as part of export guidelines. The air sampling was performed eleven days after A(H5N2) virus-positive cloaca swabs were collected. Twenty-four days after the initial A(H5N2) virus was detected, cloaca swabs indicated a continuing infection.

Farm 4 was a mechanically ventilated turkey farm housing three flocks: two with a total of 4000 21-week-old hens and one with 18,000 one-week-old chicks. In the hens, an LPAI A(H10N9) virus infection was detected following reports of reduced food consumption, respiratory signs including coughing, and malaise. Air sampling was performed at downwind locations nine days after the A(H10N9) virus-positive cloaca and trachea swabs were collected. No control measures were applied following outbreak confirmation.

Farm 5 was a mechanically ventilated mixed farm composed of two turkey flocks and a number of pigs. One turkey flock included 4000 20-week-old turkey cocks; the other included an unknown number of chicks. An LPAI A(H10N9) virus infection was detected following reports of increased mortality, nasal discharge, and respiratory signs in the 20-week-old turkeys. Air sampling inside and at downwind locations of the barn was performed three days after the A(H10N9) virus-positive cloaca and trachea swabs were collected. No control measures were applied following outbreak confirmation.

Farm 6, included as a control, was a naturally ventilated turkey farm that housed 16,500 one-month-old chicks. It was chosen because it was relatively isolated, with no commercial turkey farms (closest at ± 10 km) or chicken farms (closest at ± 4 km) in the immediate surroundings. In addition, the distance to nearest other livestock farms was >1 km. Air sampling was performed at downwind locations.

Environmental samples

Environmental sampling was performed either on private land with permission of the owner or on public roads requiring no permissions. The experiment did not involve endangered or protected species.

Airborne inhalable dust samples were initially captured on a 37-mm diameter Teflon filter with a pore-size of 2.0 µm (SKC, PA, USA) using the GSP personal sampler (JS Holdings, Stevenage, UK) equipped with a conical inlet with an 8-mm diameter orifice at the front. The sampler meets the CEN/ISO/ACGIH criterion for inhalable dust when operated at 3.5 L per minute, which was achieved with a constant-flow pump (Gill air 5, Gillian, UK). Using a tripod, sampling was performed 1.5 m above ground for a six-hour period, resulting in a filtered-air volume of 1.3 m^3. Multiple GSP samplers were

used for simultaneous collection of inhalable dust samples at several distances from a farm. Immediately after sampling, the GSP sampling heads were wrapped in plastic before transport from field to laboratory, where they were stored at −20°C until further use.

In addition to the six-hour GSP air-sampling strategy, a short-term strategy using a portable air sampler was incorporated halfway into this study. The short-term air samples were obtained with an MD8-AirPort Air Sampler (Sartorius, Göttingen, Germany) equipped with cellulose nitrate filters having a pore size of 8 µm. This sampler was operated at 50 L per minute, with a sampling time of 20 minutes, resulting in a filtered-air volume of 1.0 m^3. One MD8-AirPort sampler was used for consecutive collection of air samples at several distances from a farm. Immediately after sampling, each cellulose nitrate filter was transferred to a sterile Petri dish before transport from field to laboratory and storage at −20°C [23,24].

Influenza virus recovery

Our procedure for the detection of airborne influenza viruses was adopted from knowledge gained during a Q fever outbreak. We therefore evaluated whether the method for detecting *Coxiella burnetii* DNA in inhalable airborne dust collected on Teflon filters could be used for recovery of influenza virus by reverse transcriptase PCR (RT-PCR), using cell-culture grown influenza virus as a control [25]. We used three additional filter extraction procedures to determine which allowed the most sensitive RT-PCR detection of influenza viruses on filters.

To each filter we applied 20 individual 5-µL drops of heat-inactivated LPAI virus A/Mallard/NL/12/2000 (H7N3) in Dulbecco's modified Eagle medium (Gibco, NY, USA), corresponding with 1.3 x 10^5 influenza genome copies [26]. The filters were air-dried and shaken for one hour in 4 mL pyrogen-free water with 0.05% Tween 20 (Calbiochem, CA, USA), with (method A) or without (method B) subsequent enzyme treatment intended to free bacterial DNA. Enzyme treatment consisted of adding 100 µL of 1 mg/mL lysostaphin (Sigma, MO, USA) and 20 µL of 20 mg/mL lysozyme (Sigma) followed by incubation for 35 minutes at 37 °C, after which 400 µL of 20 mg/mL proteinase K (Roche diagnostics, Rotkreuz, Switzerland) was added and incubated for 10 minutes at 55 °C. Enzymes were then heat-inactivated at 95 °C for 10 minutes. This step was followed by DNA/RNA extraction using the NucliSens Magnetic Extraction Kit (bioMérieux, Marcy-l'Etoile, France) according to the manufacturer's instructions [25].

Alternatively, spiked filters transferred to 2mL Eppendorf tubes containing 1 mL PBS and 1% Triton X-100 (BDH Chemicals, Poole, UK) were vortexed 3x10 seconds (method C), were mixed using a bench rocker for 30 minutes (method D) or sonicated for 30 minutes (method E), followed by RNA extraction as described below. For use as reference material, 100 µL influenza A/Mallard/NL/12/2000 (H7N3) virus was directly

resuspended in 900 µL PBS and 1% Triton X-100. All procedures were performed *in triplicate*.
The presence of influenza virus was measured by a real-time influenza virus-specific RT-PCR as described below [26]. All obtained cycle-threshold (Ct) values were within the linear part (R^2 = 0.9995) of the RT-PCR amplification. Influenza virus recovery was calculated by comparing the averaged Ct values per extraction procedure with the averaged Ct values of the reference material. The method with highest recovery was used in subsequent experiments.

Environmental air sample processing
The Teflon filters (37mm diameter) collected from the GSP after air sampling were cut in half and transferred to 2 mL Eppendorf tubes prefilled with either 1 mL PBS containing 1% Triton X-100 or 1.5 mL infection medium consisting of Modified Eagle Medium with Hanks' BSS (BioWhittaker, Verviers, Belgium) supplemented with 10% PGR-albumin, penicillin, streptomycin, nystatin, L-glutamine, HEPES, and trypsin. The larger cellulose nitrate filters (80mm diameter) were likewise cut in half and transferred to 15-mL Greiner tubes prefilled with either 1.5 mL PBS containing 1% Triton X-100 for molecular testing, or 5 mL infection medium. Filters in infection medium were vortexed for 10 seconds followed by 0.22 µm filtration, and 300 µL (Teflon) or 2 mL (cellulose nitrate) of the flow-through was subsequently used for virus isolation.

Detection of influenza virus
RNA was extracted from 600 µL of the recovered fluids from Triton X-100-treated Teflon and cellulose nitrate filters using the High Pure RNA isolation Kit (Roche), and influenza virus real- time RT-PCR was used to detect the matrix gene of the influenza virus [26,27]. The influenza virus RT-PCR had a linear amplification range up to Ct value 31.15, corresponding with a limit of quantification of 1.1 x 10^4 genome copies per ml or 3.0 x 10^2 50% egg infectious dose (EID50) per ml. The detection limit of the influenza virus RT-PCR was 320 genome copies per ml or 8.9 EID50 per ml.

Influenza virus isolation
Filter-rinse fluids in infection medium were cultured on tertiary cynomolgus monkey kidney cells [28] and maintained in culture for a maximum of 2 weeks, or until cytopathic effect was observed. Presence of influenza virus in the culture supernatants was verified by RT-PCR as described above. The foregoing is standard procedure for human influenza virus isolation in our laboratory, which was proven effective for the isolation of avian influenza viruses during the influenza A(H7N7) virus outbreak in the Netherlands in 2003 [29]. Consequently, we hypothesized that it could be used to isolate avian influenza virus from filter fluids.

Endotoxin measurement

In addition to influenza virus RNA, air samples obtained from farms 4, 5 and 6 were assayed for endotoxins, which can serve as a generic proxy for airborne poultry and livestock associated microbial exposure [22]. Endotoxin content of fluids from Triton X-100-rinsed filters was analyzed by the quantitative kinetic chromogenic Limulus amebocyte lysate (LAL) assay, described previously [30]. Inhibition or enhancement of the LAL-assay by application of 1% Triton X-100 was verified in dilution series, but was not observed when samples were diluted at least 1:50 in the assay. Consequently, fluids from Triton X-100-rinsed filters were tested in a dilution of 1:50 or higher. Results were expressed as endotoxin units (EU) per m^3 (18 EU=1 ng). The limit of detection was 1 and 2 EU per m^3 of filtered air for the GSP and MD8-AirPort measurements, respectively.

Detection of turkey cells

In addition to the quantification of endotoxins as a generic proxy for airborne microbial exposure, PCR detection of the *Meleagris gallopavo* (turkey) gene for mitochondria cytochrome oxidase 1 (CO1) was performed as a turkey farm-specific proxy for airborne exposure. Of fluids from Triton X-100-rinsed filters, 200 µL was used for automated total nucleic acid isolation on a MagNA Pure 96 extraction robot (Roche) with the MagNA Pure 96 DNA and Viral NA Small Volume Kit. Next, a real-time PCR assay was performed, targeting a 90-nucleotide fragment of the CO1 gene, using LightCycler® 480 DNA SYBR Green I Master and the primers TurkeyCOI-F (5'-ACAACCATATTCTTATCATTAACC-3') and TurkeyCOI-R (5'-GTTGCATTAAGTATAGGTGTTT-3').

Atmospheric dispersion model (ADM)

We compared the endotoxin measurements to relative spatial particulate matter concentrations calculated by the atmospheric dispersion model OPS-ST (Operational Priority Substances, Short Term) model, version 4. This ADM was developed by the Dutch National Institute for Public Health and the Environment (RIVM) for the dispersion modeling of chemical pollutants, i.e. particulate matter and ammonia/nitrogen oxides generated by traffic, industries, agriculture, and natural sources [31-33]. The OPS-ST model downloaded hourly-averaged meteorological data from the Royal Netherlands Meteorological Institute webserver, including wind speed, wind direction, solar radiation, temperature, precipitation amounts, and precipitation duration [33,34]. We used coarse particulate matter as a proxy for endotoxin and assumed an environmental roughness length of 20 cm.

We defined the farms under investigation as point sources in the OPS-ST model, with arbitrary PM_{10} emission amounts per hour per source from 10:00AM to 16:00PM on sampling days. We calculated the concentrations at an above-ground height of 1.5 m

on a grid of 2 x 2 km, with a grid-cell size of 10 m. Since we converted the modeled concentration levels relative to the concentration near the source, we were able to compare these modeled data to measured endotoxin concentrations by performing a linear regression analysis.

Results

Influenza virus recovery

Influenza virus recovery measured by RT-PCR was 10% when Teflon filters were processed by using method A, but recovery increased to 43% in the absence of enzyme treatment (method B). Rinsing of filters with PBS containing 1% Triton X-100 using the alternatives of vortexing (method C), a bench rocker (method D), or sonication (method E) resulted in influenza virus recovery of 60%, 26% and 57%, respectively. Based on these results, all filters were processed using method C prior to RNA extraction and influenza virus detection by RT-PCR.

Influenza virus detection

As the applied influenza A virus RT-PCR demonstrated a limit of detection of approximately 6 genome copies per reaction, the processing of filter-rinse fluids combined with the available virus recovery data led to a theoretical limit of detection of approximately 300 and 500 influenza genome copies per m^3 of filtered air for the GSP and MD8-AirPort measurements, respectively. Influenza A viruses could be detected by RT-PCR in outdoor-air samples obtained up to 60 meters downwind of commercial turkey farms 2, 4 and 5, which had ongoing LPAI infection (Table 1). However, the corresponding Ct values were high and beyond the linear RT-PCR amplification range (Ct value > 31.15), hampering virus quantification except for an indoor-air sample from farm 5 that had an influenza virus concentration of 8.48×10^4 genome copies per m^3. The air measurements initiated approximately six hours after the culling of farm 1 yielded no influenza virus. In addition, none were detected in the filter fluids from the air measurements obtained near the LPAI-positive swans at farm 3 or near control farm 6. Despite the fact that some of the air filters tested positive for influenza virus RNA, the GSP and MD8-AirPort filters yielded no virus isolates.

Endotoxin measurement

For farms 4, 5 and 6, fluids from Triton X-100-rinsed filters were likewise tested for the amount of air-suspended endotoxins (Table 1, Fig 1). As expected, endotoxin concentrations decreased as function of distance from the source, suggesting a reduction in microbial exposure with increasing distance. Air samples taken outside poultry barns had endotoxin concentrations of ~50 EU/m^3 at distances up to 50 meters

Table 1. Combined dataset depicting the farms, air-sample type, and location relative to the barn, and corresponding laboratory results including influenza virus RT-PCR detection, turkey COI RT-PCR detection, endotoxin quantification, and modeled relative particulate matter concentrations of all GSP and MD8-AirPort filters assayed.

Farm No.	species	virus subtype	air sample type	distance from barn	bearing from barn	type of measurement	influenza virus (Ct value)	turkey COI (Ct value)	Endotoxin EU/m3	OPS conc. (log)
1	chickens	H7N7	GSP: Teflon	42m	93°	downwind	neg	nd	nd	nd
			GSP: Teflon	53m	61°	downwind	neg	nd	nd	nd
			GSP: Teflon	150m	95°	downwind	neg	nd	nd	nd
			GSP: Teflon	200m	86°	downwind	neg	nd	nd	nd
			GSP: Teflon	190m	232°	upwind control	neg	nd	nd	nd
2	turkeys	H9N2	GSP: Teflon	40m	81°	downwind	36.59	nd	nd	nd
			GSP: Teflon	44m	129°	downwind	38.91	nd	nd	nd
			GSP: Teflon	58m	44°	downwind	37.76	nd	nd	nd
			GSP: Teflon	560m	91°	downwind	neg	nd	nd	nd
			GSP: Teflon	320m	143°	upwind control	neg	nd	nd	nd
3	swans	H5N2	GSP: Teflon	4m	30°	downwind	neg	nd	nd	nd
			GSP: Teflon	20m	77°	downwind	neg	nd	nd	nd
			GSP: Teflon	20m	348°	downwind	neg	nd	nd	nd
			GSP: Teflon	60m	70°	downwind	neg	nd	nd	nd
			GSP: Teflon	98m	54°	downwind	neg	nd	nd	nd
			GSP: Teflon	290m	177°	upwind control	neg	nd	nd	nd
4	turkeys	H10N9	GSP: Teflon	7m	56°	downwind	34.00	32.10	56.99	-0.42
			GSP: Teflon	8m	109°	downwind	35.81	33.83	26.14	-0.42
			GSP: Teflon	54m	59°	downwind	34.97	neg	8.60	-1.17
			GSP: Teflon	100m	53°	downwind	neg	neg	6.44	-1.56
5	turkeys	H10N9	MD8: Cellulose nitrate	0m		inside turkey stable	30.54	31.05	339.99	0.00
			GSP: Teflon	15m	104°	downwind	32.96	40.00	31.93	-0.76
			GSP: Teflon	28m	130°	downwind	36.3	neg	7.10	nd#
			MD8: Cellulose nitrate	44m	64°	downwind	35.94	neg	111.42	-1.01
			GSP: Teflon	47m	71°	downwind	34.31	40.00	48.31	-1.01
			MD8: Cellulose nitrate	54m	36°	downwind	36.38	neg	71.89	-1.18
			GSP: Teflon	59m	30°	downwind	34.56	neg	13.71	-1.37
			MD8: Cellulose nitrate	walkabout		downwind	neg	neg	96.72	nd
			MD8: Cellulose nitrate	0m		inside pig stable	neg	neg	98990.06	nd
			MD8: Cellulose nitrate	0m		inside turkey chick stable	neg	37.29	86.26	nd
6	turkeys	None	MD8: Cellulose nitrate	37m	15°	downwind	neg	neg	6.97	-2.21
			MD8: Cellulose nitrate	38m	24°	downwind	neg	neg	31.59	-1.09
			MD8: Cellulose nitrate	44m	42°	downwind	neg	neg	15.21	-1.00
			MD8: Cellulose nitrate	100m	47°	downwind	neg	neg	7.47	-1.89
			MD8: Cellulose nitrate	110m	59°	downwind; rain	neg	neg	2.60	-1.59
			MD8: Cellulose nitrate	110m	50°	downwind	neg	neg	7.07	-1.81
			MD8: Cellulose nitrate	160m	85°	downwind	neg	neg	5.00	-1.45
			MD8: Cellulose nitrate	190m	40°	downwind	neg	neg	*	-2.35
			MD8: Cellulose nitrate	200m	55°	downwind	neg	neg	*	-2.18
			MD8: Cellulose nitrate	410m	71°	downwind	neg	neg	3.82	-2.37

nd) not determined

*) Below detection limit

#) outside plume

from the farm. At 100 meters and further from the farm, endotoxin concentrations decreased to <10 EU/m^3. The highest endotoxin concentrations detected in outdoor-air samples corresponded with two MD8-AirPort measurements at poultry farm 5, probably influenced by emissions from a nearby upwind-located pig shed with an indoor endotoxin level of 99.000 EU/m^3.

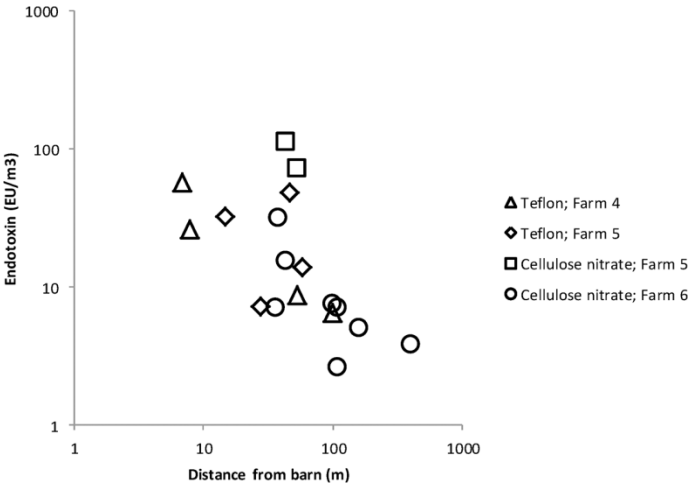


Figure 1. Endotoxin concentrations in air samples outside poultry barns are depicted in relation to the distance from the poultry barn, illustrating a reduction of airborne endotoxin with increasing distance from the source.

Detection of turkey cells

The presence of air-suspended turkey cells was confirmed in samples obtained from farms 4 and 5 (Table 1). The three downwind dust samples collected closest to the farm tested positive for the turkey cells. However, at distances ≥ 10 m, turkey cell concentrations reached the limit of CO1 detection.

Atmospheric dispersion modeling of particulate matter

Fig 2A shows the modeled relative concentrations of particulate matter at farms 4, 5 and 6. We next determined the modeled concentrations at the locations of air sampling (Table 1) and plotted them against the corresponding endotoxin concentration of each air filter location (Fig 2B). Air filters exposed to rain, located outside the dust plume, or showing an endotoxin concentration below the limit of detection were excluded. In general, endotoxin measurements and modeled relative concentrations of particulate matter showed a good correlation: linear regression analysis for farms 4 and 5 resulted in slopes of 0.72 (95% CI: 0.09 – 1.55) and 0.78 (95% CI: 0.08 – 1.64), respectively, and an R^2 of 0.88 and 0.61, respectively. Analysis of farm 6 resulted in a slope of 0.45 (95% CI: 0.02 – 0.88) and an R^2 of 0.59. Combining the data of farms 4, 5 and 6 resulted in an overall slope of 0.69 (95% CI: 0.42 – 0.97) and an R^2 of 0.65.

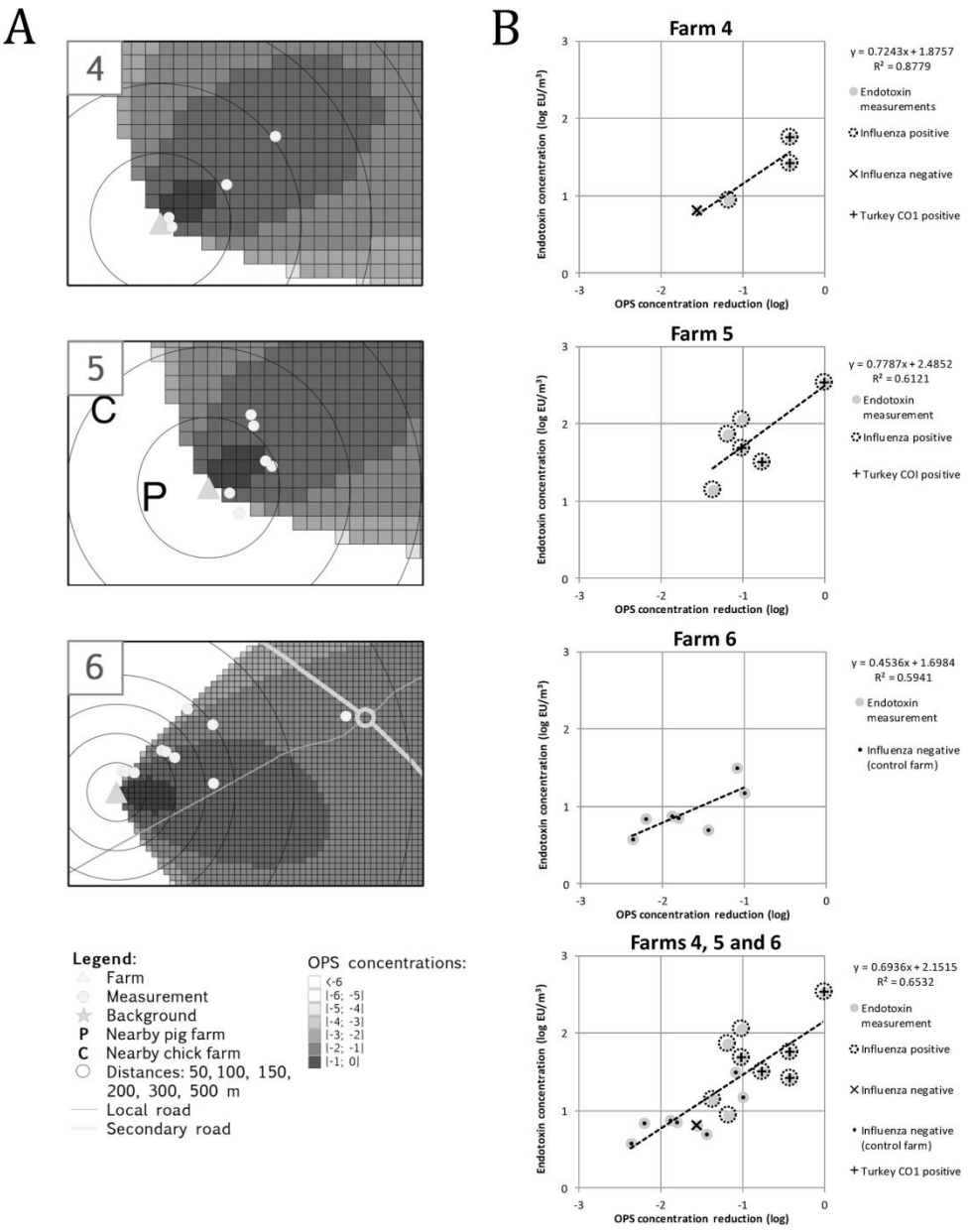


Figure 2. Dispersion of particulate matter around poultry farms, based on field measurements of endotoxin concentrations in air samples and OPS-ST particulate matter modeling. A) Maps illustrating the air sampling locations together with the atmospheric dispersion of particulate matter (relative to the source) that was modeled using meteorological data corresponding with the day and timeframe (10:00AM—16:00PM) of air sampling. B) Scatterplot of modeled dispersion and measured endotoxin concentration. Qualitative results of influenza virus RNA and turkey cell DNA detection are depicted as well.

Discussion

We demonstrate the wind-mediated spread of influenza virus-contaminated poultry dust into the environment during influenza outbreaks in commercial poultry farms based on detection of the air-suspended virus downwind of farms. The observed influenza virus concentration of 8.5 x 10^4 genome copies per m^3 air inside turkey farm 5 is in agreement with previous reports of 6.9 x 10^4 influenza virus particles per m^3 detected in chicken houses and 3.7 ×10^4 particles per m^3 in a chicken pen, respectively [7,8]. As the mechanical ventilation rates of commercial poultry housing range from a minimum of 0.5 m^3/kg/hr up to 4.0 m^3/kg/hr, the amount of air that is forced into the environment was at least 40.000 m^3 per hour for farm 5, corresponding with the emission of over 3 x 10^9 influenza virus genome copies per hour. A very large volume of virus particles could be shed into the environment during an outbreak, particularly with a source of prolonged emission like farm 4, where the infection started two weeks before the positive air samples were obtained. Such geographic dispersal of airborne virus may explain the detection of influenza virus RNA in, for example, dust swabs and samples of soil and mud puddles taken in areas surrounding farms positive for influenza A(H5N1) virus [35].

A crucial question when using this data for risk assessment is whether the viruses remain infectious during dust-mediated dispersal. We did not detect any infectious virus in our study, perhaps due to the sampling procedure. Mandal and Brandl (2011) have shown that dehydration stress caused by filtration reduces survival rates of bacteria [36]. Likewise for influenza viruses, filtration was found to reduce viability [37,38]. Even when air is not passed through a filter, infectious influenza virus can decay with a 1-2 log reduction after drying at room temperature [39,40]. Filtration and the subsequent processing of samples is therefore not optimal for the detection of live microorganisms, although Spekreijse et al. demonstrated that using a less dehydrating filter medium like gelatin allows detection of infectious influenza viruses in air-filter samples [41]. In our study, environmental conditions could also have led to influenza virus inactivation. Finally, virus recovery may have been reduced by 0.22 µm syringe filtration of the inoculum prior to cell culture inoculation, to minimize bacterial contamination. Our results are in agreement with the RT-PCR detection of swine influenza at locations downwind from swine farms using liquid cyclonic collectors, for which virus infectivity could not be confirmed, possibly due to low virus loads combined with physical disruption of viruses during air sampling [42].

In addition to virus detection by RT-PCR, the presence of turkey-specific CO1 gene was confirmed by PCR. Although this proxy for airborne microbial exposure is specific for turkey (farms), it appeared to be less sensitive than the generic endotoxin proxy. Nevertheless, the association between detectable virus and host nucleic acids

illustrated the potential application of CO1 gene sequencing to assess sources of zoonotic agents in environmental samples.

Data from the six-hour measurements with GSP personal samplers confirmed the presence of influenza virus in the inhalable dust fraction of outdoor air near infected poultry farms. In addition to the approximately 10^6 microbial cells present in 10 m^3 air that humans inhale during the course of a day [36], our data suggests that humans living downwind of influenza virus-positive farms are possibly exposed to these virus particles. Unfortunately, our use of low-volume air samplers resulted in the detection of unquantifiable influenza virus RNA in the outdoor air. The use of high-volume air samples will presumably provide data that are more robust and allow outdoor air characterization at larger distances from an infected barn. In addition to the characterization of wind-borne influenza virus exposure, determining the infectivity of wind-borne influenza virus is challenging. Due to the loss of virus viability by air sampling methods, placement of naïve sentinel animals at a grid downwind of infected poultry farms is the method of choice to obtain relevant and conclusive data on the public health impact of wind-borne influenza virus spread. Such data can also be gained by sero-surveillance studies of humans and animals previously exposed to the virus.

As the average size of single influenza virus particles is 80 – 120 nm, our capture of viruses with 2.0 µm and 8.0 µm pore-sized filters suggests that these virus particles are indeed dispersed using other particles (particulate matter) as a vehicle [9-11,43]. However, the mesh of fibers in aerosol filters (including Teflon and cellulose nitrate filters) enable efficient collection of much smaller particles than their pore-size would indicate [44]. Consequently, more research is needed to characterize the particle size distribution in relation to the detection of influenza virus RNA. Although the amount of influenza virus RNA in outdoor air near barns was unquantifiable, airborne poultry and livestock associated microbial exposure was determined by measuring concentrations of endotoxin. Endotoxin concentrations measured at different distances from farms were compared with modeled concentrations of particulate matter. Despite the many variables that potentially influenced the relationship between emission of particles and concentration, a good correlation between field measurements and modeled particulate concentrations was observed (Fig 2B). Although the 95% confidence intervals of the slopes corresponding with individual farms were large, combining all farms reduced the confidence interval, possibly by averaging out errors.

Our results suggest that an ADM like the OPS-ST model could be used to model the dispersion of outdoor airborne pathogens prospectively. However, the quantitative model outcomes should be regarded as indicative, given the number of unknown uncertainty factors. For example, the model predicts dispersion of PM_{10} while it is yet unknown what proportion of influenza viruses are associated with the PM_{10} fraction of airborne particulate matter. Nor is it known how the pathogens are distributed over

the different size fractions within PM_{10}. Although an association was found between modeled and measured concentrations, the slopes in Fig 2B did deviate from 1, suggesting that dispersion is slightly different for endotoxins than for modeled dust. The difference should be clarified in future studies. Moreover, it should be noted that virus inactivation (e.g., as a result of UV-radiation or dehydration) is not included in the model. Since our measurements were performed at relatively short distances from the infected farms, absence of an inactivation rate will not lead to high biases. At larger distances, however, the inactivation rate will be more important and must be included when this type of information becomes available.

In accordance with EU directive 92/40/EEC, controlling outbreaks of HPAI viruses relies on movement restrictions for farms within a radius of at least 10 km of an infected farm, along with culling of infected poultry. Depending on the outbreak severity, additional control measures can be taken including preventive ring-culling of farms within a (1-5 km) radius of an infected farm and extended (nationwide) standstill for the transport of live poultry. Despite such measures, avian influenza virus outbreaks in areas with high concentrations of poultry have been difficult to control, resulting in large-scale culling in the Netherlands, Canada and Mexico [45-47]. More directed interventions can potentially limit the duration of the outbreak and the number of culled farms [48]. Gaining insight on farm-specific virus spread allows directed interventions following targeted surveillance.

In this study, we demonstrated the presence of airborne influenza virus RNA downwind from buildings holding LPAI-infected birds, and observed correlation between field data on airborne poultry and livestock associated microbial exposure and the OPS-ST model. These findings suggest that geographical estimates of areas at high risk for human and animal exposure to airborne influenza virus can be modeled during an outbreak, although additional field measurements are needed to validate this proposition. In addition, the outdoor detection of influenza virus-contaminated airborne dust during outbreaks in poultry suggests that practical measures can assist in the control of future influenza outbreaks.

In general, exposure to airborne influenza virus on commercial poultry farms could be reduced both by minimizing the initial generation of airborne particles and implementing methods for abatement of particles once generated [6,49]. As an example, emergency mass culling of poultry using a foam blanket over the birds instead of labor-intensive whole-house gassing followed by ventilation reduces both exposure of cullers and dispersion of contaminated dust into the environment, contributing to the control of influenza outbreaks [50,51].

Acknowledgements

This work was supported by the Dutch Ministry of Economic Affairs, Castellum Project (MJ MK). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Swayne DE (2007) Understanding the complex pathobiology of high pathogenicity avian influenza viruses in birds. Avian Dis 51: 242-249.
2. Stallknecht DE, Brown JD (2009) Tenacity of avian influenza viruses. Rev Sci Tech 28: 59-67.
3. Spickler AR, Trampel DW, Roth JA (2008) The onset of virus shedding and clinical signs in chickens infected with high-pathogenicity and low-pathogenicity avian influenza viruses. Avian Pathol 37: 555-577.
4. Lu H, Castro AE, Pennick K, Liu J, Yang Q, et al. (2003) Survival of avian influenza virus H7N2 in SPF chickens and their environments. Avian Dis 47: 1015-1021.
5. Reis A, Stallknecht D, Ritz C, Garcia M (2012) Tenacity of low-pathogenic avian influenza viruses in different types of poultry litter. Poult Sci 91: 1745-1750.
6. Cambra-Lopez M, Aarnink AJ, Zhao Y, Calvet S, Torres AG (2010) Airborne particulate matter from livestock production systems: a review of an air pollution problem. Environ Pollut 158: 1-17.
7. Chen P-S, Lin CK, Tsai FT, Yang C-Y, Lee C-H, et al. (2009) Quantification of Airborne Influenza and Avian Influenza Virus in a Wet Poultry Market using a Filter/Real-time qPCR Method. Aerosol Science and Technology 43: 290-297.
8. Lv J, Wei B, Chai T, Xia X, Miao Z, et al. (2011) Development of a real-time RT-PCR method for rapid detection of H9 avian influenza virus in the air. Arch Virol 156: 1795-1801.
9. Angelakis E, Raoult D (2010) Q Fever. Vet Microbiol 140: 297-309.
10. Zhao Y, Aarnink AJA, De Jong MCM, Groot Koerkamp PWG (2013) Airborne Microorganisms from Livestock Production Systems and Their Relation to Dust. Critical Reviews in Environmental Science and Technology: null-null.
11. Cafarchia C, Camarda A, Iatta R, Danesi P, Favuzzi V, et al. (2014) Environmental contamination by Aspergillus spp. in laying hen farms and associated health risks for farm workers. J Med Microbiol 63: 464-470.
12. Ypma RJ, Jonges M, Bataille A, Stegeman A, Koch G, et al. (2013) Genetic data provide evidence for wind-mediated transmission of highly pathogenic avian influenza. J Infect Dis 207: 730-735.
13. Stull RB (2000) Meteorology for Scientists and Engineers. Pacific Grove: California Brooks/Cole.
14. Nguyen TM, Ilef D, Jarraud S, Rouil L, Campese C, et al. (2006) A community-wide outbreak of legionnaires disease linked to industrial cooling towers--how far can contaminated aerosols spread? J Infect Dis 193: 102-111.
15. Sørensen JH, Jensen CØ, Mikkelsen T, Mackay DKJ, Donaldson AI (2001) Modelling the atmospheric dispersion of foot-and-mouth disease virus for emergency

preparedness. Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere 26: 93-97.

16. Wallensten A, Moore P, Webster H, Johnson C, van der Burgt G, et al. (2010) Q fever outbreak in Cheltenham, United Kingdom, in 2007 and the use of dispersion modelling to investigate the possibility of airborne spread. Euro Surveill 15.
17. Ssematimba A, Hagenaars TJ, de Jong MC (2012) Modelling the wind-borne spread of highly pathogenic avian influenza virus between farms. PLoS One 7: e31114.
18. Jonges M, Bataille A, Enserink R, Meijer A, Fouchier RA, et al. (2011) Comparative analysis of avian influenza virus diversity in poultry and humans during a highly pathogenic avian influenza A (H7N7) virus outbreak. J Virol 85: 10598-10604.
19. Wang X, Fang S, Lu X, Xu C, Cowling BJ, et al. (2014) Seroprevalence to Avian Influenza A(H7N9) Virus Among Poultry Workers and the General Population in Southern China: A Longitudinal Study. Clinical Infectious Diseases.
20. Chen Y, Liang W, Yang S, Wu N, Gao H, et al. (2013) Human infections with the emerging avian influenza A H7N9 virus from wet market poultry: clinical analysis and characterisation of viral genome. The Lancet 381: 1916-1925.
21. Pantin-Jackwood MJ, Miller PJ, Spackman E, Swayne DE, Susta L, et al. (2014) Role of Poultry in the Spread of Novel H7N9 Influenza Virus in China. Journal of Virology 88: 5381-5390.
22. Heederik DJJI, C. J. (eds.) (2011) Mogelijke effecten van intensieve-veehouderij op de gezondheid van omwonenden: onderzoek naar potentiële blootstelling en gezondheidsproblemen.
23. Engelhart S, Glasmacher A, Simon A, Exner M (2007) Air sampling of Aspergillus fumigatus and other thermotolerant fungi: comparative performance of the Sartorius MD8 airport and the Merck MAS-100 portable bioaerosol sampler. Int J Hyg Environ Health 210: 733-739.
24. de Bruin A, van der Plaats RQ, de Heer L, Paauwe R, Schimmer B, et al. (2012) Detection of Coxiella burnetii DNA on small-ruminant farms during a Q fever outbreak in the Netherlands. Appl Environ Microbiol 78: 1652-1657.
25. Hogerwerf L, Borlee F, Still K, Heederik D, van Rotterdam B, et al. (2012) Detection of Coxiella burnetii DNA in inhalable airborne dust samples from goat farms after mandatory culling. Appl Environ Microbiol 78: 5410-5412.
26. Jonges M, Liu WM, van der Vries E, Jacobi R, Pronk I, et al. (2010) Influenza virus inactivation for studies of antigenicity and phenotypic neuraminidase inhibitor resistance profiling. J Clin Microbiol 48: 928-940.
27. van der Goot JA, van Boven M, Stegeman A, van de Water SGP, de Jong MCM, et al. (2008) Transmission of highly pathogenic avian influenza H5N1 virus in Pekin ducks is significantly reduced by a genetically distant H5N2 vaccine. Virology 382: 91-97.

28. van Wezel AL, van Steenis G, van der Marel P, Osterhaus ADME (1984) Inactivated Poliovirus Vaccine: Current Production Methods and New Developments. Review of Infectious Diseases 6: S335-S340.
29. Fouchier RA, Schneeberger PM, Rozendaal FW, Broekman JM, Kemink SA, et al. (2004) Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome. Proc Natl Acad Sci U S A 101: 1356-1361.
30. Douwes J, Versloot P, Hollander A, Heederik D, Doekes G (1995) Influence of various dust sampling and extraction methods on the measurement of airborne endotoxin. Appl Environ Microbiol 61: 1763-1769.
31. van der Swaluw E, Asman WAH, van Jaarsveld H, Hoogerbrugge R (2011) Wet deposition of ammonium, nitrate and sulfate in the Netherlands over the period 1992–2008. Atmospheric Environment 45: 3819-3826.
32. Jaarsveld JAV, Klimov D (2011) Modelling the impact of sea-salt particles on the exceedances of daily PM10 air quality standards in the Netherlands. International Journal of Environment and Pollution 44: 217-225.
33. Jaarsveld JAV (2004) The operational priority substances model. Bilthoven: National Institute for Public Health and the Environment.
34. KNMI Uurgegevens van het weer in Nederland. De Bilt: Royal Netherlands Meteorological Institute.
35. Horm SV, Gutierrez RA, Sorn S, Buchy P (2012) Environment: a potential source of animal and human infection with influenza A (H5N1) virus. Influenza Other Respir Viruses 6: 442-448.
36. Mandal JB, H. (2011) Bioaerosols in indoor environment—a review with special reference to residential and occupational locations. Open Environmental and Biological Monitoring Journal 4: 83–96.
37. Cao G, Noti JD, Blachere FM, Lindsley WG, Beezhold DH (2011) Development of an improved methodology to detect infectious airborne influenza virus using the NIOSH bioaerosol sampler. J Environ Monit 13: 3321-3328.
38. Fabian P, McDevitt JJ, Houseman EA, Milton DK (2009) Airborne influenza virus detection with four aerosol samplers using molecular and infectivity assays: considerations for a new infectious virus aerosol sampler. Indoor Air 19: 433-441.
39. Tuladhar E, Terpstra P, Koopmans M, Duizer E (2012) Virucidal efficacy of hydrogen peroxide vapour disinfection. J Hosp Infect 80: 110-115.
40. Tuladhar E, de Koning MC, Fundeanu I, Beumer R, Duizer E (2012) Different virucidal activities of hyperbranched quaternary ammonium coatings on poliovirus and influenza virus. Appl Environ Microbiol 78: 2456-2458.

41. Spekreijse D, Bouma A, Koch G, Stegeman A (2013) Quantification of dust-borne transmission of highly pathogenic avian influenza virus between chickens. Influenza and Other Respiratory Viruses 7: 132-138.
42. Corzo CA, Culhane M, Dee S, Morrison RB, Torremorell M (2013) Airborne detection and quantification of swine influenza a virus in air samples collected inside, outside and downwind from swine barns. PLoS One 8: e71444.
43. Lamb RA, Choppin PW (1983) The Gene Structure and Replication of Influenza Virus. Annual Review of Biochemistry 52: 467-506.
44. Burton NC, Grinshpun SA, Reponen T (2007) Physical collection efficiency of filter materials for bacteria and viruses. Ann Occup Hyg 51: 143-151.
45. Stegeman A, Bouma A, Elbers AR, de Jong MC, Nodelijk G, et al. (2004) Avian influenza A virus (H7N7) epidemic in The Netherlands in 2003: course of the epidemic and effectiveness of control measures. J Infect Dis 190: 2088-2095.
46. Bowes VA (2007) After the outbreak: how the British Columbia commercial poultry industry recovered after H7N3 HPAI. Avian Dis 51: 313-316.
47. FAO (2012) Highly Pathogenic Avian Influenza in Mexico (H7N3); A significant threat to poultry production not to be underestimated. Empress Watch http://www.fao.org/docrep/016/an395e/an395e.pdf.
48. te Beest DE, Hagenaars TJ, Stegeman JA, Koopmans MP, van Boven M (2011) Risk based culling for highly infectious diseases of livestock. Vet Res 42: 81.
49. Aarnink A, Mosquera J, Winkel A, Cambra-Lopez M, Van Harn J, et al. Options for dust reduction from poultry houses; 2009. Engineers Australia. pp. 47.
50. Gerritzen MA, Lambooij E, Stegeman JA, Spruijt BM (2006) Slaughter of poultry during the epidemic of avian influenza in the Netherlands in 2003. Veterinary Record 159: 39-42.
51. Benson E, Malone GW, Alphin RL, Dawson MD, Pope CR, et al. (2007) Foam-Based Mass Emergency Depopulation of Floor-Reared Meat-Type Poultry Operations. Poultry Science 86: 219-224.

Chapter 4

Characterizing the within-host influenza virus sequence diversity

Chapter 4.1

Improved Detection of Artifactual Viral Minority Variants in High-Throughput Sequencing Data

Matthijs R.A. Welkers[1], Marcel Jonges[2,3], Rienk E. Jeeninga[1], Marion Koopmans[2,3] and Menno D. de Jong[1]

Front Microbiol. 2015 Jan 22;5:804.

[1]Dept. of Medical Microbiology, Academic Medical Centre, Amsterdam, The Netherlands

[2]Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, The Netherlands

[3]Department of Viroscience, Erasmus MC, Rotterdam, The Netherlands

Abstract

High-throughput sequencing (HTS) of viral samples provides important information on the presence of viral minority variants. However, detection and accurate quantification is limited by the capacity to distinguish biological from artificial variation. In this study, errors related to the Illumina Hiseq2000 library generation and HTS process were investigated by determining minority variant frequencies in an influenza A/WSN/1933(H1N1) virus reverse-genetics plasmid pool. Errors related to amplification and sequencing were determined using the same plasmid pool, by generation of infectious virus using reverse genetics followed by in duplo reverse-transcriptase PCR (RT-PCR) amplification and HTS in the same sequence run.

Results showed that after "best practice" quality control (QC), within the plasmid pool, one minority variant with a frequency >0.5% was identified, while 84 and 139 were identified in the RT-PCR amplified samples, indicating RT-PCR amplification artificially increased variation. Detailed analysis showed that artifactual minority variants could be identified by two major technical characteristics: their predominant presence in a single read orientation and uneven distribution of mismatches over the length of the reads. We demonstrate that by addition of two QC steps 95% of the artifactual minority variants could be identified. When our analysis approach was applied to three clinical samples 68% of the initially identified minority variants were identified as artifacts.

Our study clearly demonstrated that, without additional QC steps, overestimation of viral minority variants is very likely to occur, mainly as a consequence of the required RT-PCR amplification step. The improved ability to detect and correct for artifactual minority variants, increases data resolution and could aid both past and future studies incorporating HTS. The source code has been made available through Sourceforge (https://sourceforge.net/projects/mva-ngs).

Introduction

RNA viruses are present within hosts as a cloud of closely related non-identical genomes, often referred to as the viral quasispecies [1, 2]. The development of high-throughput sequencing (HTS) technologies has provided the opportunity to investigate the viral quasispecies composition and detection of viral minority variants at frequencies well below 10–20%, generally accepted as the detection limit of Sanger population sequencing [3, 4]. However, it is essential to distinguish true biological variation from artificial variation introduced by the laboratory and sequencing methods used. Over time, several approaches to error-correction have been developed incorporating complicated probabilistic methods. These methods use the read sequence information but ignore the corresponding phred quality scores [5–7]. In addition, most error-correction algorithms were designed for invariable genomes like the human genome, and generally assumed that errors were not only random and infrequent, but can be corrected using the majority of the reads that have the correct base [6, 7]. For usage in viral population sequencing this assumption does not apply as each variant in theory could represent a unique viral variant, and different approaches for viral quasispecies reconstruction have been designed, like local haplotype reconstruction incorporated in Shorah [5, 8, 9]. However, these methods still rely solely on read sequence information requiring prior QC steps. In addition, while Shorah performs well using long read sequences and relatively small sized datasets, typical for Roche 454 data, it does not scale well to the shorter reads and much larger Illumina datasets as it can only handle up to one million reads, limiting the usability with Illumina sequencers. As different HTS platforms utilize different sequencing methods we showed in a previous study on serially passaged influenza A(H1N1)pdm09 virus, that while identical minority variants were detected using the Roche 454 and Illumina GAIIx HTS platforms, the frequencies of these variants could vary depending on the platform used [10]. Sequencing chemistries of these HTS platforms have been upgraded to produce longer read lengths with improved error profiles which might lead to more sensitive and reliable detection of viral minority variants [11].

The aim of the current study was to identify and characterize the type of technical errors that were introduced due to the RT-PCR amplification and HTS process itself. Therefore, HTS datasets were generated directly from a pool of eight influenza A virus reverse genetics plasmids as well as from RT-PCR amplified viral RNA produced by reverse genetics using the same plasmid pool. After current “best practice” QC, mismatch frequencies per reference genome position were reduced in all three samples, but a large number of minority variants were still identified and assumed to be the result of the RT-PCR amplification and sequencing procedures. We analyzed these minority variants in more detail and show that by using two additional

bioinformatic analysis steps, applicable to large datasets, 95% of erroneous bases were detected and could be corrected for. Additionally, we then demonstrate the usability of our approach using three influenza A positive clinical samples. Finally, we discuss the flexibility of our methods for the application to HTS datasets generated from other platforms.

Materials and methods

Plasmid DNA and reverse genetics

Reverse genetic system plasmids containing the individual full-length segments of influenza A/WSN/33 (H1N1) virus PB2, PB1, PA, HA, NA, NP, MP, and NS in a pHW2000 backbone were kindly provided by Webster. Plasmid DNA was isolated using Maxi prep column purification according to the manufacturer's protocol (Macherey-Nagel). An 80% confluent monolayer of epithelial human embryonic kidney (HEK) 293T/17 cells (ATCC no. CRL-11268) in a T25 tissue culture flask was transfected with an equimolar mixture of the eight WSN33 reverse genetic plasmids (1 µg DNA each) and 20 µl Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. Medium was replaced 16 h after transfection with Dulbecco's modified eagle medium (DMEM; Invitrogen) containing 1% fetal calf serum (FCS). Virus-containing supernatant was harvested 48 h after transfection, and cells and cell debris were removed by centrifugation [4 min at 400 relative centrifugation force (RCF)] and filtration through a 0.2-µm filter (FP 030/3; Schleicher and Schuell). For RNA extraction and consecutive full-genome amplification, 200 µl supernatant was added to 400 µl lysis binding buffer (High Pure RNA isolation kit; Roche). The remaining supernatant was aliquoted and stored at −80°C.

Clinical samples

From three patients with laboratory-confirmed influenza A(H1N1)pdm09 virus infection, viral RNA was extracted from throat swabs using the High Pure RNA isolation kit (Roche) with an on-column DNase treatment, according to the manufacturer's protocol. The clinical samples were obtained as part of routine diagnostic procedures in the Academic Medical Centre (Amsterdam, The Netherlands) in accordance with national and institutional regulations concerning the procurement and usage of patient-derived materials.

Full influenza genome RT-PCR amplification

RT-PCR amplification was performed as described previously [12]. In short, for each sample, two separate RT-PCR reactions were performed using primers common-uni12R (5'-GCCGGAGCTCTGCAGATATCAGCRAAAGCAGG-3'), common-uni12G (5'-GCCGGAGCTCTGCAGATATCAGCGAAAGCAGG-3'), and common-uni13 (5'-

CAGGAAACAGCTATGACAGTAGAAACAAGG-3'). The first RT-PCR contained the primers common-uni12R and common-uni13. The second RT-PCR contained the primers common-uni12G and common-uni13 that greatly improved the amplification of the PB2, PB1, and PA segments. Reactions were performed using the One-Step RT-PCR kit High Fidelity (Invitrogen) in a volume of 50 µl, and contained 5.0 µl eluted RNA, and final concentrations of 1X SuperScript™ III One-Step RT-PCR reaction buffer, 0.2 µM of each primer and 1.0 µL SuperScript™ III RT/PlatinumTM Taq High Fidelity Enzyme Mix (Invitrogen). Thermal cycling conditions were: reverse transcription at 42°C for 15 min, 55°C for 15 min, 60°C for 5 min; initial denaturation/enzyme activation of 94°C for 2 min; five cycles of 94°C for 30 s, 45°C for 30 s, slow ramp (0.5°C/s) to 68°C, 68°C for 3 min; 30 cycles 94°C for 30 s, 57°C for 30 s, 68°C for 3 min; and final extension of 68°C for 5 min. After the RT-PCR equal volumes of both reactions were combined to produce a well distributed mixture of all eight influenza segments.

Illumina HiSeq2000 sequencing

Each sample was diluted to a DNA concentration of 50 ng/µL and sheared by nebulization. Samples were then subjected to end-repair, A-overhang, and adaptor ligation with MID-tags using the Illumina TruSeq DNA sample preparation kit. The libraries were multiplexed, clustered, and sequenced on an Illumina HiSeq2000 (TruSeq v3 chemistry) with a paired-end 100 cycles sequencing protocol with indexing according to manufacturer's protocol [10, 12]. After analysis with the Illumina CASAVA pipeline version 1.8.2, resulting datafiles were split into separate sample specific fastq files based on the used multiplex identifier (MID) sequence using the readset parser function of QUASR version 7.0.1 [10, 13].

Pre-mapping quality control

Quality control (QC) consisted of several bioinformatic steps which were divided in pre-mapping and post-mapping QC. All steps were performed using in-house generated "PHP: Hypertext Preprocessor" (PHP) scripts which are available upon request. Pre-mapping QC steps included the removal of reads from datasets containing uncalled bases (N's) followed by the removal of read parts that exactly matched any of the primers used. Next, the removal of low quality nucleotides from read ends (defined as nucleotides with phred score < 30) and removal of low quality read pairs using the paired end QC function from QUASR version 7.0.1 with median phred cutoff of 30 and minimal resulting read length of 50 nucleotides (settings -m30 and -l50) [10]. Phred quality scores are logarithmically linked to the base call accuracy and a phred score of 30 resembles a base call accuracy of 99.9% while a phred score of 40 represent an accuracy of 99.99% [15]. The selection of a median phred value of 30 as a cutoff value was based on the observation that the largest fraction of reads was

maintained in the datasets while still providing a strong reduction in low quality sequence reads (supplementary Figure 1).

Read mapping

Datasets from the plasmid pool and both RT-PCR amplified rgWSN33 samples were mapped to the influenza A/WSN/1933(H1N1) reference sequence (taxonomy ID: 382835) using the program Burrows-Wheeler Aligner (BWA) version 0.5.9-r16 with the paired-end mapping option (sampe) and default settings [16]. The three influenza A(H1N1)pdm09 clinical samples were mapped to the A/California/04/2009 H1N1pdm09 virus (taxonomy ID: 641501; Genbank accession numbers FJ966079-FJ966086).

Post-mapping quality control

Post-mapping QC consisted of the removal of unmapped and unpaired reads from the Sequence Alignment Map (SAM) files as well as removal of reads containing clipped ends (i.e., reads with read sequence parts automatically removed by BWA) and/or insertions/deletions (indels) as identified in the read CIGAR string [17]. Lastly, orphan reads (i.e., read sequences present without a mate) were removed to generate the final cleaned datasets.

Mismatch frequency (MMF) per sequence read position

For each mapped sequence read, the read sequence, most left-based starting position and mapping information summarized in the CIGAR string were extracted from the SAM file and nucleotides at each mapped read position were compared to the corresponding reference sequence position [17]. The MMF per sequence read position was defined as the percentage of all sequenced nucleotides not matching the reference nucleotide at that read position.

Influenza genome coverage and MMF per reference sequence position

The coverage per influenza genome position was defined as the total of all mapped nucleotides per reference sequence position with the consensus nucleotide being the mapped nucleotide with the highest frequency. The MMF per reference position was defined as the percentage of all mapped nucleotides excluding the consensus nucleotide frequency.

Sequence specific error (SSE) analysis

Artificial variation which is introduced by the sequencing process will lead to differences in the MMF in a particular sequencing orientation. Therefore, the MMF at each reference sequence position was determined for reads mapping to the forward and the reverse strand of the reference sequence based on the bitwise flag value [17].

An SSE was defined as a position in which the difference in MMF of nucleotides mapping to the forward and/or reverse strand exceeded the average MMF of all mapped nucleotides at that position, with a minimal sequencing depth of 50 for both mapping orientations.

Non-uniform spread of mismatches and frequency recalculation

Biological variation will be present in the viral quasispecies in multiple different RNA templates which are separately amplified by RT-PCR, hence DNA fragmentation during library preparation will lead to an even distribution of mutations over the length of the sequence reads (Figure 1A). In contrast, artificial variation by for instance either residual primers/adapters or overamplification of a (subset) of libraries during library amplification will lead to mismatches overrepresented at particular sequence read positions producing a spike-like pattern when visualized in a “hairy caterpillar plot” (HCPP). Therefore, reads covering influenza genome positions with an MMF above 0.5% were extracted from the cleaned SAM file and the position of each mutation within the read was determined. Artificial variation was arbitrarily defined as a minimal three-fold higher number of mutations at a particular read position above expected and MMF's were recalculated by excluding the excess of mutations.

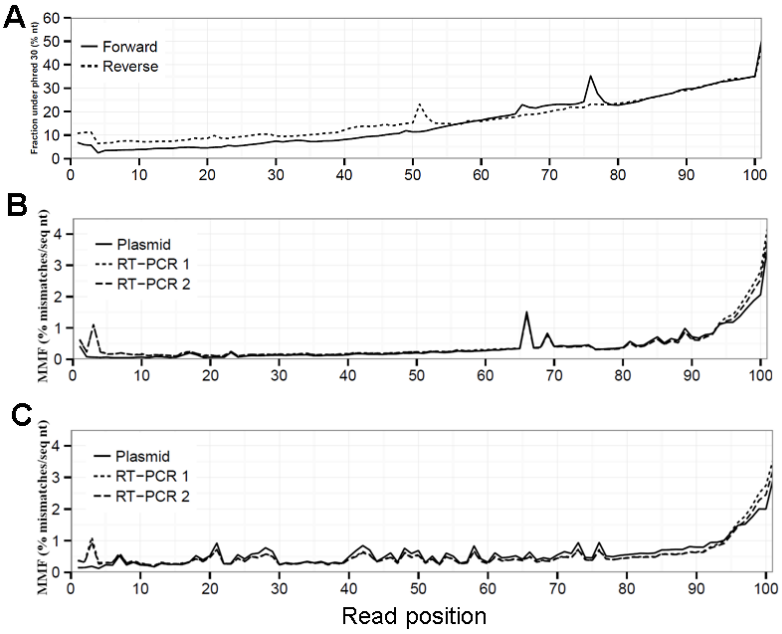


Figure 1. Sequencing quality decreased toward the 3’ read end displayed by an increased fraction of sequenced nucleotides with a phred ≤30 (A). This reduced sequencing quality led to an increased MMF in both the forward (B) and reverse (C) run for the plasmid, RT-PCR 1 and RT-PCR 2 datasets.

Results

Mismatches introduced by Illumina HiSeq2000 sequencing and RT-PCR

Illumina HiSeq2000 sequencing generated 14.2 million read pairs for the reverse genetic plasmid pool sample, and 7.5 million and 4.0 million read pairs for RT-PCR samples 1 and 2, respectively (Table 1). All three samples were sequenced in the same run, hence variation in the plasmid pool sample was assumed to be due to the sample preparation and sequencing process and differences in variation between both RT-PCR samples and the plasmid pool sample were assumed to be due to the RT-PCR amplification step.

Table 1. Overview sequence data.

		Plasmid	RT-PCR 1	RT-PCR 2	CS1	CS2	CS3
RT-PCR amplified		No	Yes	Yes	Yes	Yes	Yes
Number of read pairs		14,192,571	7,455,869	4,040,876	3,438,654	3,577,913	5,480,898
Number of sequenced nucleotides		2,866,899,342	1,506,085,538	816,256,952	687,730,800	715,582,600	1,096,179,600
Read pairs removed containing N's (% of total)	Forward	0.3	0.3	0.3	0.1	0.1	0.1
	Reverse	0.8	0.8	0.8	0.2	0.2	0.3
Nucleotides removed due to primer/adapter removal (% of total)	Forward	1.4	7.2	7.2	9.2	9.3	9.3
	Reverse	1.2	7.2	7.1	4.2	4.2	11.4
Removed due to median phred read quality <30 (% of total)	Read pairs	6.0	4.4	4.3	4.9	4.5	1.4
	Nucleotides	6.0	4.8	4.7	5.4	5.2	2.4
Total read pairs removed after pre-mapping QC		1,001,818 (7.1%)	399,154 (5.4%)	216,962 (5.4%)	166,215 (5.4%)	159,772 (5.0%)	75,711 (1.8%)
Total nucleotides removed after pre-mapping QC		336,369,593 (11.7%)	228,338,456 (15.2%)	123,364,266 (15.1%)	96,126,836 (14.0%)	98,606,211 (13.8%)	158,272,488 (14.4%)

Read sequence quality is known to decrease toward the 3' read end [11], which was also observed in our datasets demonstrated by the up to seven-fold increase in the fraction of sequenced nucleotides with a phred value ≤ 30 in both the forward (from 6.8% at read position 1 to 50.5% at read position 101) and reverse run (from 10.8 to 47.3%, Figure 1A). Without any QC on the sequence reads, this resulted in an average MMF of 0.49% per read position which increased up to 1.58% in the last 10 read positions for both the forward and reverse run (Figures 1B,C, supplementary Table 1). This increase was only partially due to the before mentioned reduction in sequencing quality, as after applying a phred cutoff of 30 the MMF in the last 10 positions only decreased to 0.36% (supplementary Table 1). Further investigation of the 3' sequence read ends with more than two mismatches showed the presence of partial and/or mutated Illumina adapters (data not shown). Furthermore, while the average MMF per read position for both RT-PCR samples were not markedly different when compared to the plasmid pool sample the average MMF in the first 10 nucleotides of the forward run of both RT-PCR samples were increased nearly three-fold compared to the plasmid pool sample (Figures 1B,C). Further investigation showed that this was due to presence

of RT-PCR primer sequences (data not shown). If not removed, these Illumina adapter and primer read sequences would lead to erroneous interpretation of biological sequence diversity as these sequences were successfully mapped to the reference sequence. We also investigated whether the library preparation, sequencing process and RT-PCR amplification influenced the resulting coverage at specific influenza genome positions. In order to compare between samples, the coverage was expressed as a percentage of the maximum obtained coverage within that sample (relative coverage).

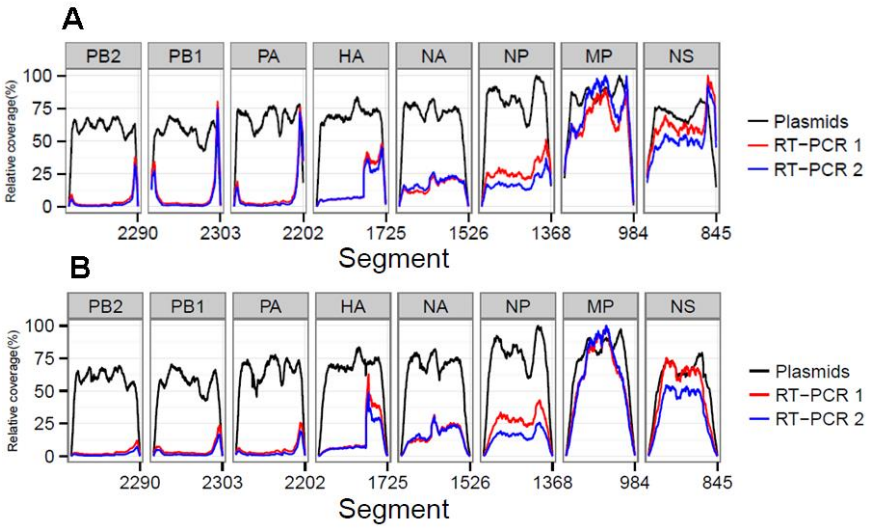


Figure 2. The sequencing process and quality control steps do not lead to a selective bias for particular influenza segments as comparable relative coverages for the plasmid dataset before **(A)** and after **(B)** QC were obtained. In both RT-PCR samples 1 and 2 preferential amplification of the smaller segments was observed.

Results showed that in the plasmid pool sample similar relative coverages were obtained for the different influenza segments indicating that the library preparation and sequencing steps did not create a bias (Figures 2A,B). However, for the RT-PCR amplified samples, we observed increased coverages per position for the smaller influenza segments (Figures 2A,B). In addition, both RT-PCR amplified samples displayed a marked four-fold increase in coverage in the HA segment starting at reference position 1189. At reference position 1176, a Uni13-like primer sequence was identified with only two mismatches compared to the Uni13 primer used for amplification, highly suggestive for non-specific primer binding and subsequent amplification.

In summary, preferential amplification of the smaller segments in the RT-PCR combined with the presence of primer and Illumina adapter sequences led to

increased mismatch frequencies per sequence read position, which could impact the accurate detection and identification of viral minority variants.

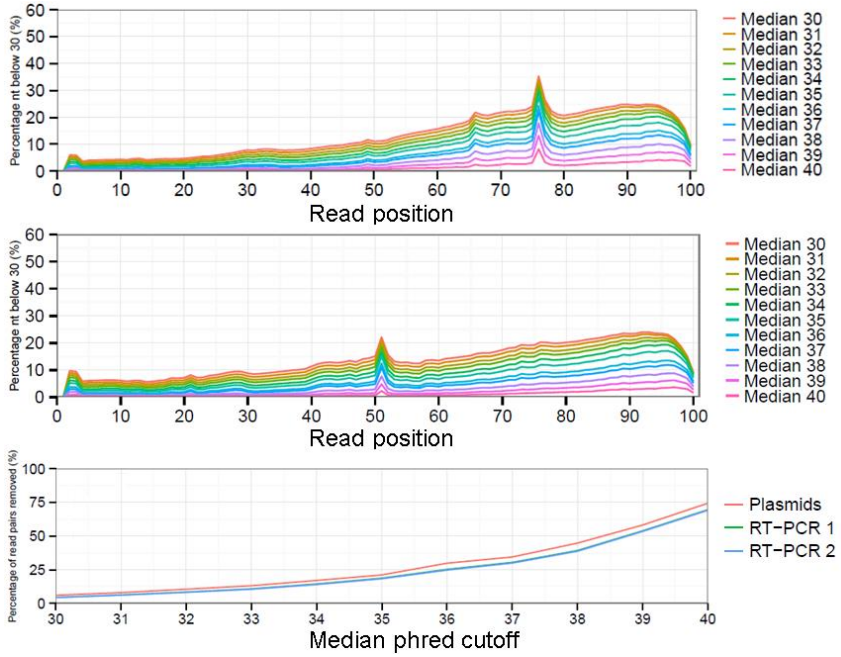


Figure 3. The impact of selecting different median phred quality cutoff values on the remaining reads as displayed by fraction of sequenced nucleotides with a phred ≤30 for the forward **(A)** and reverse **(B)** run. Increasing the minimal median phred read quality will result in increased numbers of read pairs being removed from the analysis **(C).**

Quality control of sequence reads

While there is no gold standard for QC, it is generally divided in pre-mapping QC and post-mapping QC steps (see Materials and methods). In pre-mapping QC read pairs were removed containing uncalled bases, primer sequences as well as low quality nucleotides from read ends. In the final pre-mapping QC step all remaining read pairs were required to have a minimal length of 50 and minimal median phred value of 30. While the selection of a particular median phred cutoff is rather arbitrary, increasing the median phred cutoff value removed increasing numbers of read pairs from the datasets to nearly up to 75% for the plasmid pool dataset with a median phred cutoff of 40 (Figure 3C). With a median phred cutoff of 30, pre-mapping QC led to a total removal of 7.1% of read pairs (11.7% of sequenced nucleotides) from the plasmid pool dataset and 5.4% of read pairs (15.1-15.2% of sequenced nucleotides) from the RT-PCR datasets (Table 1). This lead to a decrease in the fraction of nucleotides with a phred < 30 in both the forward and reverse sequence run (Figures 3A,B). The increased

percentage of removed nucleotides in the RT-PCR samples was due to removal of primer sequences from the sequence reads.

Subsequently, datasets were mapped to influenza A/WSN/1933(H1N1) consensus sequence and 93.7% of the reads that passed pre-mapping QC were successfully mapped for the plasmid pool dataset, while this was 70.5-70.8% for the RT-PCR datasets, indicating the presence of sequence reads which differed from the used reference sequence (Table 2).

Table 2. Overview of mapping data.

Post mapping QC	Plasmid	RT-PCR 1	RT-PCR 2	CS1	CS2	CS3
Total sequenced reads	28,385,142	14,911,738	8,081,752	6,877,308	7,155,826	10,961,796
Reads passing pre-mapping QC (% of total)	26,381,506 (92.9)	14,113,430 (94.6)	7,647,828 (94.6)	6,504,799 (94.6)	6,795,790 (95.0)	10,766,776 (98.2)
Mapped reads (% of pre-mapping QC passed reads)	24,710,168 (93.7)	9,986,294 (70.8)	5,389,605 (70.5)	6,452,417 (99.2)	6,603,449 (97.2)	10,109,171 (93.9)
Unmapped reads (% of pre-mapping QC passed reads)	1,671,338 (6.3)	4,127,136 (29.2)	2,258,223 (29.5)	52,382 (0.8)	192,341 (2.8)	657,605 (6.1)
Unpaired reads (% of mapped)	982,417 (4.0)	1,168,572 (11.7)	606,869 (11.3)	526,237 (8.2)	513,868 (7.6)	491,450 (4.9)
Clipped reads (% of mapped)	498,217 (2.1)	825,102 (9.4)	420,532 (8.8)	61,272 (0.9)	53,788 (0.8)	936,993 (9.3)
Orphan reads (% of mapped)	649,810 (2.8)	1,523,536 (19.1)	837,872 (19.2)	64,346 (1.0)	56,685 (0.9)	939,518 (9.3)
Total sequenced reads removed by pre- and post-mapping QC (% of total)	5,805,418 (20.5)	8,442,654 (56.6)	4,557,420 (56.4)	1,076,746 (15.7)	1,176,718 (16.4)	3,220,586 (29.4)
Total sequenced nucleotides removed by pre- and post-mapping QC (% of total)	683,255,156 (23.8)	881,452,097 (58.5)	475,561,924 (58.3)	166,794,256 (24.0)	179,258,081 (24.8)	424,048,489 (28.3)

Post-mapping QC steps showed that in RT-PCR datasets 1 and 2, clipped reads (i.e., reads with parts of the read sequence automatically removed by the mapping software) and unpaired reads were nearly four- and three-fold more frequent compared to the plasmid pool dataset respectively. In more detail, these reads predominantly mapped to segment ends and consisted mainly of residual (mutated) primer and adapters sequences (data not shown). Both clipped and unpaired reads were removed from the datasets as well as their mate reads (orphan reads) during further analysis. The percentages of removed reads in the combined QC steps were 20.5% for the plasmid pool dataset and 56.6-56.4% for RT-PCR 1 and RT-PCR 2 respectively, corresponding to 23.8, 58.5, and 58.3% of all sequenced nucleotides respectively (Table 2).

The average MMF per sequence read position was recalculated using the QC-controlled datasets and in the plasmid pool datasets the 20.5% removal of sequence reads led to a 19-fold reduction to an average of 0.02 and 0.03% for the forward and reverse run respectively. Within the RT-PCR amplified samples an average removal of 56.5% of reads only led to a nine-fold reduction to an average of 0.06% (0.05-0.07%,

supplementary Table 1). These results indicate, that not only RT-PCR amplification greatly increased the sample diversity by primarily the introduction of different (mutated) primer sequences, but also that this artificial variation is not adequately removed using current “best practice” QC steps.

Analysis of unmapped reads for false negative detection of minority variants

While concerns mainly exist for the presence erroneous minority variants, overly stringent QC and removal of reads could also lead to a false negative detection of minority variants. Therefore, sequence reads that passed pre- and post-mapping QC steps but were not mapped to the reference sequence were analyzed for the presence of influenza-like sequences. As unmapped reads were most often unpaired, the reads from the forward and reverse run were analyzed separately. For the plasmid pool sample 17.9 and 15.2% of unmapped reads from the forward and reverse run mapped to reference sequence *Escherichia coli* strain K-12 (Table 3). The source of the *E. coli* was most likely the plasmid isolation, as the plasmids were propagated in *E. coli*. In the RT-PCR samples only a very low percentage of unmapped reads mapped to *E. coli* (0.002 and 0.001%). Human DNA was detected at an average frequency of 1.5% (1.3-1.6%) in the plasmid pool sample while this was only 0.2% (0.1-0.2%) for the RT-PCR samples (Table 3). As the majority of unmapped reads remained of unknown origin, all remaining sequence reads were split in half and both the first and last half of each sequence read were separately mapped to the influenza A/WSN/1933(H1N1) reference sequence. Of the first half of the reads nearly 50% (48.2-50.5%) of the plasmid sample and 57% (55.8-58.0%) of the RT-PCR samples were successfully mapped. For the last half of the sequence read, these percentages were 26% (25.8 and 27.3%) and 37% (35.5-37.9%). Of these mapped reads, nearly 48.8% in the plasmid sample and 0.002% in the RT-PCR samples mapped to the plasmid pHW2000 backbone sequence, while the rest of the mapped read parts were spread over the influenza segment ends. This indicated that the unmapped reads mainly represented the sections where the influenza sequence connected with the plasmid sequence. Interestingly, in 1.0% of split reads, both read halves mapped to the same segment but at opposite ends, while in only 0.03% of split reads the read parts mapped to separate gene segments. This indicates that the ligation step during library preparation is not prone to inducing ligation artifacts. In all three datasets, the reads halves that successfully mapped were predominantly located at the segment ends, indicating that these initially unmappable reads consisted of a combination of a mutated Uni12/Uni13 primer and influenza sequence. In conclusion, these observations suggest that false negatives may occur due to unsuccessful mapping, but the frequency is very likely to be low and mainly restricted to the segment ends.

Table 3. Analysis of the unmapped reads.

Post mapping QC			Plasmid	RT-PCR 1	RT-PCR 2	CS1	CS2	CS3
Total unmapped reads			1,671,338	4,127,136	2,258,223	52,382	192,341	657,605
	Forward run		789,931	1,972,388	1,079,863	26,481	97,125	332,493
	Reverse run		881,407	2,154,748	1,178,360	25,901	95,216	325,112
Unmapped reads mapped to *E. coli*	Forward (%)		141,391 (17.9)	40 (0.002)	16 (0.001)	0 (0.0)	3 (0.0)	4 (0.0)
	Reverse (%)		134,312 (15.2)	41 (0.002)	11 (0.001)	0 (0.0)	3 (0.0)	3 (0.0)
Remaining unmapped reads mapped to human genome sequence hg19	Forward (%)		12,478 (1.6)	4,378 (0.2)	1,788 (0.2)	7,915 (29.9)	78,479 (80.8)	77,225 (23.8)
	Reverse (%)		11,867 (1.3)	4,012 (0.2)	1,592 (0.1)	7,894 (30.5)	77,976 (81.9)	77,363 (23.3)
Remaining unmapped reads split and remapped to corresponding reference sequence	Forward	Front (%)	399,145 (50.5)	1,143,072 (58.0)	621,760 (57.6)	5,555 (29.9)	4,092 (21.9)	98,904 (39.9)
		Back (%)	236,917 (27.3)	824,898 (37.3)	456,299 (37.9)	8,195 (44.1)	8,690 (46.6)	82,033 (33.1)
		Plasmid (%)	48.8	0.002	0.002	NA	NA	NA
	Reverse	Front (%)	424,536 (48.2)	1,202,129 (55.8)	667,583 (56.7)	4892 (27.2)	3954 (22.9)	106,518 (41.8)
		Back (%)	310,692 (25.8)	948,566 (35.8)	509,174 (35.5)	6713 (37.3)	6305 (36.6)	85,395 (33.5)
		Plasmid (%)	49.0	0.002	0.001	NA	NA	NA

NA, not available.

Detection of viral minority variants

Pre- and post-mapping QC steps reduced the average MMF per reference genome position in both RT-PCR samples to 0.08 and 0.09% respectively and to 0.03% in the plasmid pool sample, with only minor differences between the individual influenza segments (supplementary Table 2). In further detail, the plasmid pool sample had only one position with a MMF greater than 0.5%, while RT-PCR sample 1 had 84 positions and RT-PCR sample 2 had 139 positions above this percentage. Minority variants at these positions would normally be classified as “true” biological minority variants while a fraction actually reflects artificially introduced variation.

Sequence specific error (SSE) analysis and detection of artificial minority variants

At a combined total of 224 positions, in three separate samples, the MMF was above 0.5%. These positions were located at specific sites in the influenza genome (Figure 4). For instance, in all three samples an increased MMF was detected at reference position 666 of PB2. Without a per nucleotide quality cutoff, the MMF's at this position were 18.4% in the plasmid sample and 21.3-23.8% in the RT-PCR samples, while these were 0.6%, and 1.1-1.3% after applying a phred quality cutoff of 30. What was striking was that the mismatches at this position were unevenly distributed between the reads mapping to the forward and reverse strand. In the plasmid sample, the MMF was 8.8% in reads mapping to the forward strand and 0.04% in reads mapping to the reverse strand, indicating a strand bias. The specific T666G mutation was located in a homopolymer region and it has been known that the Illumina platform has increased error rates at homopolymeric regions [11]. After SSE analysis (Figure 5) and removal of positions at which the MMF was associated with a strand bias, no positions with an MMF above 0.5% remained in the plasmid sample while in RT-PCR samples 1 and 2 only 6 (out of 84) and 21 (out of 139) positions remained. Interestingly, 5 of the 6 positions in RT-PCR 1 were also identified in RT-PCR 2, with only 0.1-0.3% difference in

mismatch frequencies, indicating that these minority variants were either present prior to RT-PCR (i.e., introduced by the reverse genetics procedure) or that these particular errors were specifically introduced by the RT-PCR step.

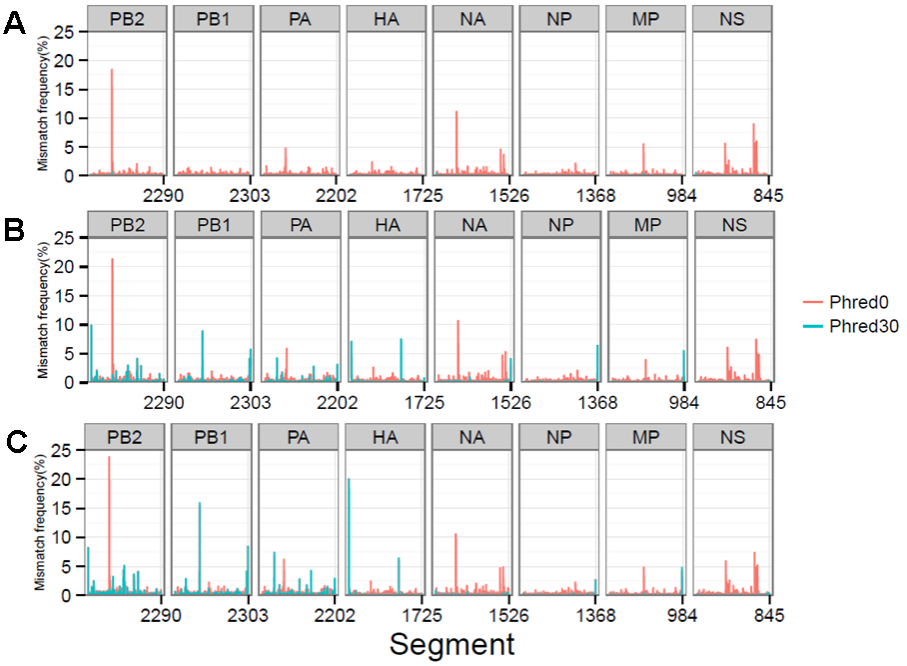


Figure 4. MMF's differ at specific reference genome positions depending on the per nucleotide phred quality cutoff for the plasmid pool (A), RT-PCR 1 (B), and RT-PCR 2 (C) samples.

Detection of artificial variation due to non-random distribution of mismatches

Artificial variation can be identified by the non-random distribution of mismatches over the sequence reads using hairy caterpillar plots (HCPP analysis). For instance, influenza position 1976 of PA had a MMF in RT-PCR samples 1 and 2 of 1.2 and 1.5%, respectively. HCPP analysis showed an even spread of the mismatches over the length of the sequence reads (Figure 6A). This even dispersal of mismatches indicated that the increased MMF was not related to any technical error and the corrected mismatch frequencies were therefore nearly identical at 1.1 and 1.0%. In contrast, influenza position 2189 of PA showed mismatch frequencies of 0.70 and 0.85% in RT-PCR samples 1 and 2 and HCPP analysis showed that the mutations were primarily located within the first and last 10 positions of the reads mapping to the forward and reverse strand respectively (Figure 6B). This indicated that variation is introduced artificially by for instance (over)amplification of the mutation in the library preparation. By the removal of excess mutations the corrected mismatch frequencies decreased for both RT-PCR samples to 0.2%.

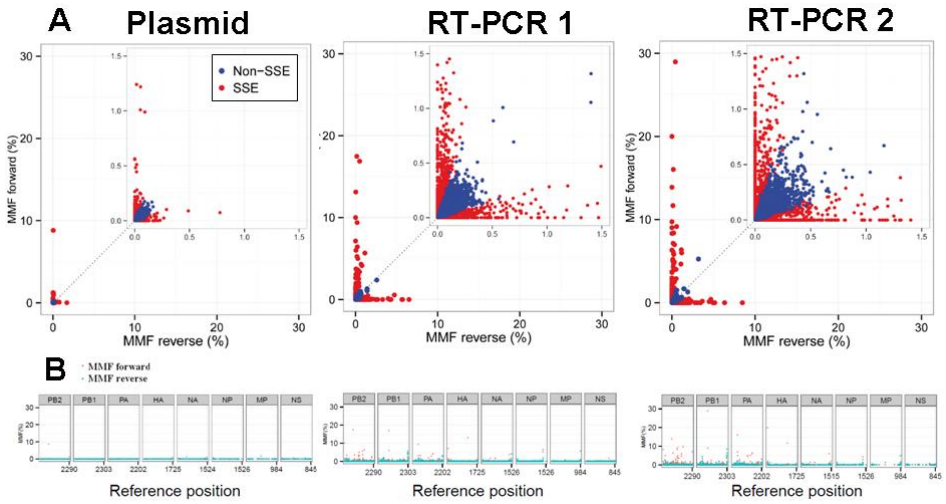


Figure 5. Sequence specific error (SSE) was defined as a position at which the difference in MMF of nucleotides mapping to the forward and the reverse strand exceeded the average MMF. Positions that were classified as SSE are highlighted in red. Both RT-PCR samples showed increased MMF's compared to the plasmid sample indicating introduction of mismatches due to the RT-PCR amplification process **(A).** The distribution of SSE's over the influenza genome was primarily located in the large polymerase segments of the RT-PCR samples and at segment ends **(B).**

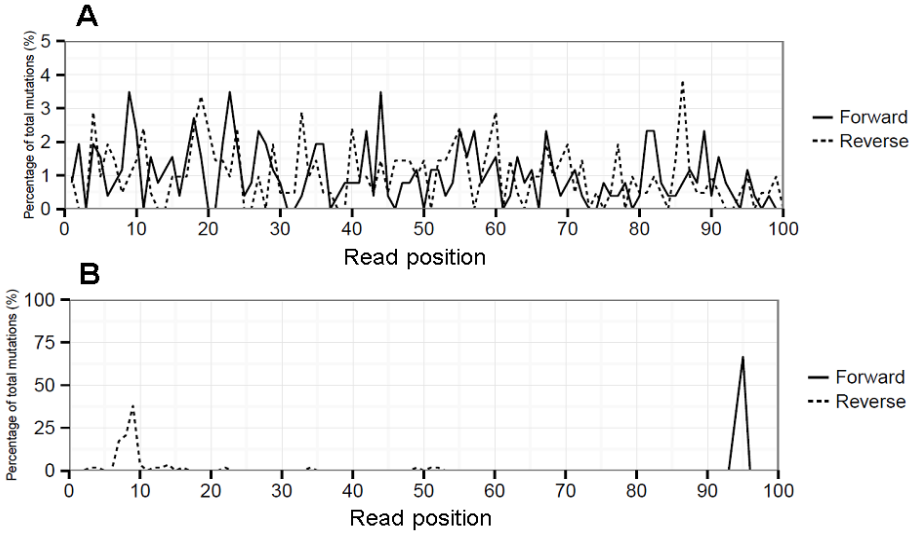


Figure 6. Hairy caterpillar plot (HCPP) of RT-PCR 1 influenza position PA-1976 **(A)** which showed that mismatches at this position were spread evenly over all read positions and strand orientations, indicative for an actual biological variation. At position PA-2189 **(B)** the mismatches were predominantly present at specific read positions in both strand orientations indicative for artificial variation.

After the HCPP analysis, two of the six positions in RT-PCR sample 1 remained with a MMF above 0.5% and 12 of 21 positions in RT-PCR sample 2. Both remaining positions from RT-PCR sample 1 were also identified in RT-PCR sample 2 (positions PA-1976 and PB1-1935) with similar mismatch frequencies. As the HCPP analysis is dependent upon adequate genome coverage it may suggest that reduced coverage is the primary cause of failing to detect the remaining artificially introduced variation. Indeed, the average coverage of the 10 remaining positions was 1.489 (675–3.382), which was nearly 22-fold lower compared the average genome coverage of 33.086 for RT-PCR sample 2.

In conclusion, with the addition of the SSE and HCPP analysis the positions with increased MMF decreased from 86 to 2 positions in RT-PCR 1 and from 139 to 12 positions in RT-PCR 2 corresponding to an average reduction of 95% (91.4-97.7%). In addition, the positions with the highest initial mismatch frequencies in RT-PCR 1 and 2 were reduced from an initial 8.9 and 15.9% to 1.1 and 1.5%, respectively. This 1.5% limit reflects the lowest frequency at which artificial variation remained clearly distinguishable from biological variation after all pre- and post-mapping steps as well as SSE and HCPP analysis.

Applicability to clinical samples

The detection of minority variants in clinical samples could be complicated by additional (unknown) factors, such as low viral load, multiple rounds of viral replication and/or the presence of bacterial, viral or human DNA. Therefore, our approach was applied to three influenza A(H1N1)pdm09 virus positive clinical samples, which generated between 3.4 and 5.5 million read pairs for each clinical sample (Table 1). Less than 5.0% of the read pairs of each dataset were removed in pre-mapping QC steps. The number of unmapped reads per sample were 0.8% for clinical sample CS 1, 2.8% for CS2 and 6.1% for CS3 (Table 2) and predominantly consisted of reads which mapped to the human genome with substantial variation between samples (23.3-81.9%). Additional BLAST analysis also showed the presence of reads of bacterial origin in all three clinical samples consistent with the type of clinical specimen used (data not shown). After splitting the remaining unmapped reads in two, influenza was still detected in all three samples (Table 3). As observed with the RT-PCR datasets, read parts were detected that mapped to opposite ends of gene segments particularly in the large polymerase genes (PB2, PB1, and PA) in all three clinical samples. After pre- and post-mapping QC steps, 15.7, 16.4, and 29.4% of sequenced reads were removed from CS1, CS2, and CS3 datasets, respectively (Table 2).

The average mismatch frequencies per influenza genome position were calculated and a total of 28, 72, and 162 positions with a MMF above 0.5% were identified in CS1, CS2, and CS3, respectively. The highest mismatch frequencies were found in HA for CS1 and CS2 [reference position 590 for CS 1 (12.5%) and reference position 261 for CS2

(10.7%)] while for CS3 the highest MMF was found in two adjacent positions in PB2 [position 275 and 276 (25.0 and 11.9%)]. After SSE analysis, 27 positions remained in CS1, 70 positions in CS2 and 86 positions in CS3. HCPP analysis further reduced the number of positions to 16 in CS 1, 58 positions in CS2 and 66 positions in CS3. As the lowest mismatch percentage at which biological variation was distinguishable from artificial variation was 1.5% this frequency was used as a cutoff value and five positions remained in CS1, 23 in CS2 and 5 in CS3. At these positions, most nucleotide mutations in CS1 and CS3 were silent (four out of five for both samples), 14 out of 23 nucleotide mutations in CS2 lead to amino acid substitutions of which the highest frequencies were observed in the HA gene (Table 4). A BLAST search for the HA-N81D substitution showed that this variant has been identified in Japanese patients. Similarly, several other HA substitutions were also identified in patients, indicating that these variants are likely to be actual biological variants.

Table 4. Minority amino acid variants above 1.5% in clinical samples after QC and SSE/HCPP analysis.

Sample	Identified minority variants (>1.5%)	
	Number of synonimous mutations	Identified non-synonimous mutations (%)
CS1	4	PA E300K (3.1%)
CS2	9	HA N81D (10.5%), HA L547V (2.5%), PA I438L (2.3%), NP D151N (2.2%), HA S123G (2.1%), PB2 D153N (2.0%), PB2 P200H (2.0%), NP A342T (2.0%), NP A33T (1.9%), NA P328H (1.7%), NA I32M (1.7%), PB2 A204E (1.6%), PB1 D638H (1.5%), HA G519E (1.5%)
CS3	4	PB1 T42K = PB1-F2 Q11K (1.7%)

These results showed that a high proportion of observed minority variants were introduced artificially and could be identified and removed efficiently with proper QC steps and additional SSE and HCPP analyses without the need for complicated mathematical haplotype reconstruction algorithms. This ultimately resulted in a reduction of false-positive detection of minority variants hence improving the data resolution and identification of true biological variants. We emphasize that while variation caused by artificial introduction was successfully removed; frequencies of assumed biological minority variants in the clinical samples remained nearly unchanged after the additional SSE and HCPP analysis steps.

Discussion

In this study, the average MMF per read position related to the Illumina Hiseq2000 sequencing process was 0.5% while in the last 10 read positions the MMF increased to 1.6% primarily due to the presence of (mutated) Illumina adapters. The MMF after RT-PCR amplification and sequencing was determined using RNA from virus produced by reverse genetics using cells which do not support productive viral replication. As the virus progeny will have been transcribed from the same plasmids as used to determine the sequencing-related MMF, sequence variation is assumed to only result from mismatches introduced during this process or the subsequent RT-PCR. In comparison to the plasmid control sample, the MMF in the first 10 read positions increased due to the presence of primer sequences used in RT-PCR amplification, while the overall MMF remained nearly identical. The finding of increased MMF's at read ends are in line with previous observations using the Illumina Genome Analyzer platform in the quasispecies analysis of foot-and-mouth disease [18].

After pre- and post-mapping QC steps, the average MMF per read position in the RT-PCR amplified datasets reduced to 0.06% overall and 0.1% in the last 10 read positions. This represents the lowest possible technical MMF, indicating that minority variants with a frequency below 0.1% will remain indistinguishable from technical error. Further increasing the sequencing depth will not improve the identification of these very low frequency (<0.1%) minority variants without additional analysis steps for the detection of artificial variation. In this study, after “best practice” QC, 223 influenza genome positions remained which would normally be considered to be “true” biological variants, while 95% were in fact artifacts, hampering reliable interpretation of the sequence data and hindering efficient follow-up studies. These artifacts were not randomly distributed over the influenza virus genome but were shown to be specifically influenced by the sequence context, detectable by a mutational strand bias. The SSE analysis indicated that nearly 90% of the positions with increased MMF were artifacts. Furthermore, shearing or nebulization of amplified DNA during library preparation will lead to biological variant mutations to be evenly spread over the sequence read. When a particular fragment containing a mutation is preferentially (over)amplified during library preparation (i.e., resequenced due to low copy number) this random spread would be disrupted and the prevalence of this mutation overestimated. This bias has been described before and complicated time-consuming laboratory methods, like primer-ID tagging, have been developed that can identify and correct for this [19]. We designed a more straightforward approach, the HCPP analysis, that was able to simultaneously detect variation associated with non-specific RT-PCR products, residual primer and adapter nucleotides and preferential overamplification, as each would lead to disruption of the random spread. The HCPP analysis not only

provides qualitative information on the presence of artifacts, but also allows for correction of a specific MMF by subtracting a quantifiable artifact. The HCPP analysis can in theory be applied to all shotgun-based deep sequencing datasets as the principle of artifact identification and quantification is not based on the sequencing platform but on the shearing of the DNA template during library preparation. However, as the cutoff value at which variation is defined as artificial is determined by the division of detected mutations by the sequence read length, it lacks sensitivity when either the sequencing depth or the MMF for the genome position in question are low. A solution to this problem would be to perform replicate RT-PCR reactions.

Combined, the SSE and HCPP analysis led to a near 95% reduction of identified positions with increased MMF's in RT-PCR amplified genomes from reverse-genetics generated virus samples, providing a simple and efficient method to reduce artifacts and more reliable identification of true minority variants. While in clinical samples a larger fraction of the positions with increased mismatch frequencies can be assumed to be due to biological variation, the lowest reduction in positions we obtained was still nearly 68%. The impact of HCPP analysis was greatest in clinical sample CS3, which also had the highest percentage of nucleotides removed during pre-mapping QC steps, particularly the primer/adapter removal. This indicates that most of the initial variation was due to remaining mutated primers which were efficiently recognized and removed during the HCPP analysis step. Our study clearly demonstrates that without proper testing for technical errors, in particular for strand bias, overestimation of minority variants (and their frequencies) in biological samples is very likely to occur.

This is the first study that describes in detail the MMF's of full genome influenza RT-PCR amplification combined with Illumina HiSeq2000 sequencing. We present additional analysis steps to remove artifacts leading to a 95% reduction of otherwise erroneously identified minority variants in plasmid controlled datasets and a 68% reduction in clinical samples. Our approach improved the distinction between artifactual and biological variants in high throughput sequencing datasets, hence facilitating studies requiring reliable high resolution identification of viral minority variants.

Acknowledgements

We kindly acknowledge Paul Kellam from the Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK for sequencing the clinical samples and BaseClear BV (Leiden, The Netherlands) for sequencing the clinical samples. This work was supported by funding from the European Community's Seventh Framework Programme [FP7/2007–

2013] under the project EMPERIE, EC grant agreement number 223498; from the Dutch Ministry of Economic Affairs, Agriculture, and Innovation, under the Castellum Project; and from the Academic Medical Centre (AMC) Graduate School [PhD scholarship to MRAW].

Supplementary Material

Supplementary Table 1. Average mismatch frequencies (MMF) per read position before and after QC and phred quality cutoff

Raw data		**Plasmid**		**RT-PCR 1**		**RT-PCR 2**	
	Run	*Forward (% of seq. nt)*	*Reverse (% of seq. nt)*	*Forward (% of seq. nt)*	*Reverse (% of seq. nt)*	*Forward (% of seq. nt)*	*Reverse (% of seq. nt)*
No quality control	*Average*	0.39	0.58	0.46	0.57	0.44	0.53
	First 10	0.09	0.24	0.32	0.43	0.32	0.40
	Last 10	1.56	1.58	1.87	1.82	1.71	1.66
Phred cutoff 30	*Average*	0.06	0.09	0.14	0.17	0.13	0.15
	First 10	0.04	0.03	0.24	0.22	0.24	0.20
	Last 10	0.36	0.39	0.72	0.81	0.66	0.72
Pre-mapping QC							
No quality control	*Average*	0.16	0.23	0.19	0.25	0.18	0.24
	First 10	0.05	0.10	0.11	0.18	0.11	0.18
	Last 10	0.28	0.30	0.24	0.26	0.23	0.25
Phred cutoff 30	*Average*	0.02	0.03	0.05	0.07	0.05	0.06
	First 10	0.03	0.03	0.09	0.10	0.09	0.10
	Last 10	0.01	0.02	0.04	0.05	0.04	0.04

Supplementary Table 2. Average mismatch frequencies per gene segment after applying a phred quality cutoff of 30

Segment	**Plasmid**	**RT-PCR 1**	**RT-PCR 2**	**CS1**	**CS2**	**CS3**
PB2	0.03 (0 – 0.6)	0.09 (0 – 4.1)	0.11 (0 – 5.1)	0.04 (0 - 1.9)	0.03 (0 - 2.5)	0.09 (0 - 25.0)
PB1	0.03 (0 – 0.2)	0.10 (0 – 10)	0.12 (0 – 15.9)	0.04 (0 - 1.2)	0.03 (0 - 6.7)	0.06 (0 - 11.8)
PA	0.03 (0 – 0.1)	0.09 (0 – 4.2)	0.10 (0 – 7.4)	0.04 (0 - 3.4)	0.03 (0 - 2.8)	0.07 (0 - 4.6)
HA	0.03 (0 – 0.2)	0.07 (0 – 7.5)	0.06 (0 – 6.4)	0.05 (0 - 12.5)	0.03 (0 - 10.7)	0.04 (0 - 0.8)
NA	0.03 (0 – 0.6)	0.07 (0 – 6.5)	0.06 (0 – 2.8)	0.04 (0 - 1.0)	0.03 (0 - 2.1)	0.03 (0 - 1.2)
NP	0.03 (0 – 0.6)	0.07 (0 – 4.1)	0.06 (0 – 0.7)	0.04 (0 - 0.7)	0.04 (0 - 2.5)	0.04 (0 - 1.4)
MP	0.03 (0 – 0.4)	0.07 (0 – 5.5)	0.07 (0 – 4.8)	0.04 (0 - 0.6)	0.03 (0 - 1.9)	0.04 (0- 0.5)
NS	0.03 (0 – 0.1)	0.06 (0 – 0.3)	0.06 (0 – 0.4)	0.05 (0 - 4.3)	0.03 (0 - 1.2)	0.04 (0 - 0.6)
Average mmf per genome position (%)	0.03 (0 – 0.6)	0.08 (0 - 10)	0.09 (0 - 20)	0.04 (0 - 12.5)	0.03 (0 - 10.7)	0.06 (0 - 25.0)

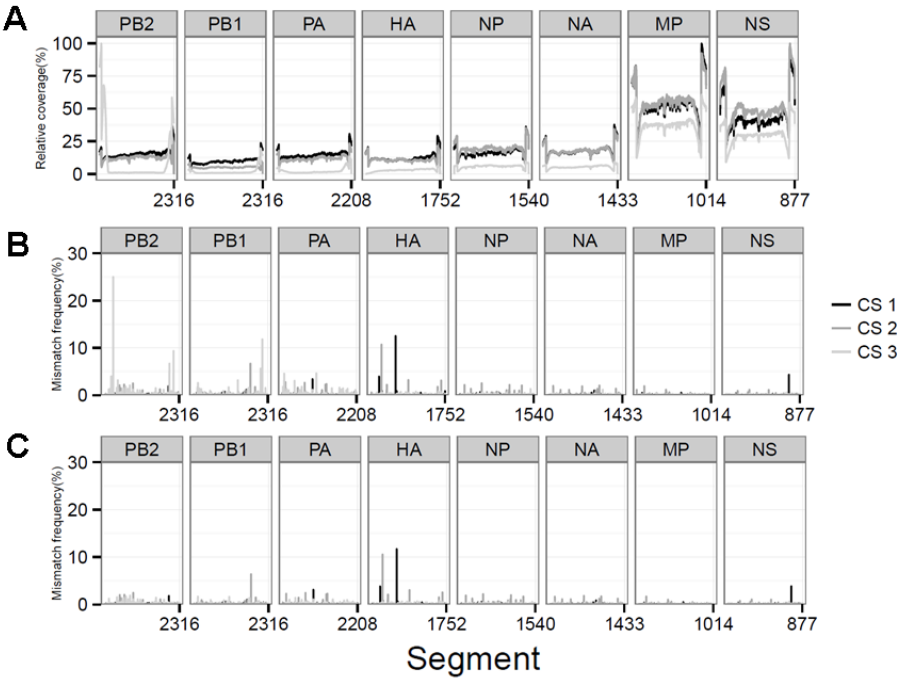


Figure S1. Relative coverage per influenza genome position after pre- and post-mapping QC **(A)** and the MMF's before **(B)** and after **(C)** application of the SSE and HCPP analysis for the clinical samples CS1, CS2 and CS3.

References

1. Domingo E, Sheldon J, Perales C: Viral quasispecies evolution. *Microbiol Mol Biol Rev* 2012, 76:159–216.
2. Lauring AS, Andino R: Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog* 2010, 6:e1001005.
3. Leitner T, Halapi E, Scarlatti G, Rossi P, Albert J, Fenyö EM, Uhlén M: Analysis of heterogeneous viral populations by direct DNA sequencing. *Biotechniques* 1993, 15:120–7.
4. Palmer S, Kearney M, Maldarelli F, Halvas EK, Bixby CJ, Bazmi H, Rock D, Falloon J, Davey RT, Dewar RL, Metcalf JA, Hammer S, Mellors JW, Coffin JM: Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *J Clin Microbiol* 2005, 43:406–13.
5. Zagordi O, Bhattacharya A, Eriksson N, Beerenwinkel N: ShoRAH: estimating the genetic diversity of a mixed sample from next-generation sequencing data. *BMC Bioinformatics* 2011, 12:119.
6. Beerenwinkel N, Günthard HF, Roth V, Metzner KJ: Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. *Front Microbiol* 2012, 3(September):329.
7. Yang X, Chockalingam SP, Aluru S: A survey of error-correction methods for next-generation sequencing. *Brief Bioinform* 2013, 14:56–66.
8. Zagordi O, Klein R, Däumer M, Beerenwinkel N: Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. *Nucleic Acids Res* 2010, 38:7400–9.
9. Zagordi O, Geyrhofer L, Roth V, Beerenwinkel N: Deep sequencing of a genetically heterogeneous sample: local haplotype reconstruction and read error correction. *J Comput Biol* 2010, 17:417–28.
10. Watson SJ, Welkers MRA, Depledge DP, Coulter E, Breuer JM, Kellam P, B PTRS, de Jong MD: Viral population analysis and minority-variant detection using short read next-generation sequencing. *Philos Trans R Soc Lond B Biol Sci* 2013, 368:20120205.
11. Minoche AE, Dohm JC, Himmelbauer H: Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol* 2011, 12:R112.
12. Jonges M, Welkers MR a, Jeeninga RE, Meijer A, Schneeberger P, Fouchier R a M, de Jong MD, Koopmans M: Emergence of the Virulence-Associated PB2 E627K Substitution in a Fatal Human Case of Highly Pathogenic Avian Influenza Virus

A(H7N7) Infection as Determined by Illumina Ultra-Deep Sequencing. *J Virol* 2014, 88:1694–702.

13. Cock PJ a, Fields CJ, Goto N, Heuer ML, Rice PM: The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res* 2010, 38:1767–71.
14. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Edgar R, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Khovayko O, Landsman D, Lipman DJ, Madden TL, Maglott DR, Miller V, Ostell J, Pruitt KD, Schuler GD, Shumway M, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Tatusov RL, et al.: Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2008, 36(Database issue):D13–21.
15. Ewing B, Green P: Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 1998, 8:186–94.
16. Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009, 25:1754–60.
17. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009, 25:2078–9.
18. Wright CF, Morelli MJ, Thébaud G, Knowles NJ, Herzyk P, Paton DJ, Haydon DT, King DP: Beyond the Consensus: Dissecting Within-Host Viral Population Diversity of Foot-and-Mouth Disease Virus by Using Next-Generation Genome Sequencing. *J Virol* 2011, 85:2266–75.
19. Jabara CB, Jones CD, Roach J, Anderson J a, Swanstrom R: Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. *Proc Natl Acad Sci U S A* 2011, 108:20166–71.

Chapter 4.2

Emergence of the Virulence-associated PB2 E627K Substitution in a Fatal Human Case of Highly Pathogenic Avian Influenza Virus A(H7N7) Infection as Determined by Illumina Ultra-deep Sequencing

Marcel Jonges[1,2], Matthijs R.A. Welkers[3], Rienk E. Jeeninga[3], Adam Meijer[1], Peter Schneeberger[4], Ron A.M. Fouchier[2] , Menno D. de Jong[3] and Marion Koopmans[1,2]

J Virol. 2014 Feb;88(3):1694-702

[1]Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, The Netherlands

[2]Department of Viroscience, Erasmus MC, Rotterdam, The Netherlands

[3]Department of Medical Microbiology, Academic Medical Center, Amsterdam, The Netherlands

[4]Department of Medical Microbiology and Infection Control, Jeroen Bosch Hospital, 's-Hertogenbosch, The Netherlands

Abstract

Avian influenza viruses are capable of crossing the species barrier and infecting humans. Although evidence of human-to-human transmission of avian influenza viruses to date is limited, evolution of variants toward more-efficient human-to-human transmission could result in a new influenza virus pandemic. In both the avian influenza A(H5N1) and the recently emerging avian influenza A(H7N9) viruses, the polymerase basic 2 protein (PB2) E627K mutation appears to be of key importance for human adaptation. During a large influenza A(H7N7) virus outbreak in the Netherlands in 2003, the A(H7N7) virus isolated from a fatal human case contained the PB2 E627K mutation as well as a hemagglutinin (HA) K416R mutation. In this study, we aimed to investigate whether these mutations occurred in the avian or the human host by Illumina ultra-deep sequencing of three previously uninvestigated clinical samples obtained from the fatal case. In addition, we investigated three chicken samples, two of which were obtained from the source farm. Results showed that the PB2 E627K mutation was not present in any of the chicken samples tested. Surprisingly, the avian samples were characterized by the presence of influenza virus defective RNA segments, suggestive for the synthesis of defective interfering viruses during infection in poultry. In the human samples, the PB2 E627K mutation was identified with increasing frequency during infection. Our results strongly suggest that human adaptation marker PB2 E627K has emerged during virus infection of a single human host, emphasizing the importance of reducing human exposure to avian influenza viruses to reduce the likelihood of viral adaptation to humans.

Introduction

Influenza viruses have been isolated from a wide range of host species, and they occasionally cross the species barrier to infect humans and vice versa [1]. The persistence of avian influenza viruses in their natural wild waterbird reservoir can trigger avian influenza epidemics in poultry, in which these viruses generally cause mild or subclinical infection [2]. For this reason, they are referred to as low-pathogenic avian influenza (LPAI) viruses. LPAI viruses of subtypes H5 and H7 have the ability to mutate to viruses with increased virulence for poultry, referred to as high-pathogenic avian influenza (HPAI) viruses, which is why they are targeted through mandatory veterinary surveillance programs. The transmission of LPAI virus from birds to mammals has resulted in the establishment of adapted variants of equine, canine, and multiple swine influenza virus lineages [3]. At present, avian influenza viruses of subtypes H5 and H7 have not adapted to mammalian species. However, the emergence and dissemination over a wide geographic region of a lineage of HPAI virus A(H5N1) in wild birds and poultry, with reported transmissions to >600 humans including >300 fatalities, is of concern [4]. Similarly, the recent emergence of human cases of influenza A(H7N9) virus infections in China in 2013 poses a public health risk [5,6]. Although evidence of avian influenza virus transmission between humans has to date been limited, possible evolution toward more-transmissible virus variants carries the specter of a potential pandemic. Fortunately, avian influenza viruses have not yet acquired mutations that facilitate efficient aerosol or respiratory droplet transmission between humans. Assessment of HPAI A(H5N1) surveillance data for mutations required for airborne transmission of A(H5N1) viruses between ferrets identified viruses in clade 2.3.2.1 that are only four nucleotide substitutions away from becoming transmissible by the airborne route [7-9]. Corresponding A(H5N1) viruses have been sampled in Nepal, Mongolia, Japan, and Korea and lack a glutamate-to-lysine change at position 627 of the viral polymerase basic 2 protein (PB2 E627K), one of the four prerequisite substitutions for airborne transmissibility, which has also been associated with increased viral replication and pathogenicity in mammals [10-15]. In addition, viruses obtained from human fatal cases detected during the recent outbreak of A(H7N9) in China are characterized by PB2 E627K, while A(H7N9) viruses obtained from poultry are not [5,16,17]. This mutation allows increased virus replication at the body temperature of mammals, thereby increasing virulence and the potential for transmission [18]. An important question for the assessment of pandemic risk is whether this change occurs in the avian host or emerges during human infection.

Between February 2003 and May 2003, a large HPAI A(H7N7) outbreak occurred in the Netherlands. This outbreak struck 255 Dutch poultry farms in a 9-week period, resulting in the culling of nearly 30 million chickens. A total of 89 humans were infected, primarily persons involved in the direct handling of infected poultry, resulting in one fatal case of a veterinarian [19-21]. Previously, two amino acid mutations had been identified in the A(H7N7) virus obtained from this fatal human case, relative to an avian virus obtained from the source farm [10]. The first was a K416R mutation in the hemagglutinin (HA), which was shown to have no impact on replication kinetics *in vitro*. The second mutation was the well-characterized human adaptation marker PB2 E627K. Previously, investigating the presence of minority variants in clinical samples required either culturing of the clinical sample (disturbing the viral population composition) or using direct Sanger sequencing methods, which are hampered by a high 20% detection limit. With the rise of next-generation sequencing technologies, it is now possible to characterize the composition of the viral quasispecies in great detail. Therefore, we traced previously unanalyzed clinical samples from the fatal A(H7N7) human case as well as original A(H7N7)-positive poultry samples from the farm at which the veterinarian was infected and a control sample from an unrelated farm. Using Illumina ultra-deep sequencing, we were able to investigate the viral quasispecies dynamics associated with a fatal human A(H7N7) virus infection.

Materials and methods

Controls

Plasmid DNA and reverse genetics

Plasmid DNA was isolated using Maxi prep column purification according to the manufacturer's protocol (Macherey-Nagel). The preparations of eight reverse genetic system plasmids containing the individual full-length segments of influenza A/WSN/33 (H1N1) virus PB2, PB1, PA, HA, NA, NP, MP, and NS in a pHW2000 backbone were obtained from R.G. Webster. Plasmid insert sequences were previously confirmed using standard Sanger sequencing [22]. An 80% confluent monolayer of epithelial human embryonic kidney (HEK) 293T/17 cells (ATCC no. CRL-11268) in a T25 tissue culture flask was transfected with a mixture of all eight WSN33 reverse genetic plasmids (1 µg DNA each) and 20 µl Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. Medium was replaced 16 h after transfection with Dulbecco's modified eagle medium (DMEM; Invitrogen) containing 1% fetal calf serum. Virus-containing supernatant was harvested 48 h after transfection, and cells and cell debris were removed by centrifugation (4 minutes at 400 relative centrifugation force [RCF]) and filtration through a 0.2-µm filter (FP 030/3; Schleicher & Schuell). For RNA

extraction and consecutive full-genome amplification, 200 µl supernatant was added to 400 µl lysis-binding buffer (High pure RNA isolation kit; Roche). The remaining supernatant was aliquoted and stored at −80°C.

Samples

(i) Chicken tracheal samples

The source of the fatal human A(H7N7) virus infection was shown to be a chicken egg layer farm that housed around 15,000 chickens in battery cages [10]. From 27 March 2003 onwards, the farmer reported an increasing number of dead chickens. On 2 April 2003, the poultry farm was visited by the veterinarian for screening purposes, and 2 days later, on 4 April 2003, all chickens from this farm were culled as screening samples tested positive for influenza A(H7N7) virus. Two pools of five influenza A(H7N7) virus-positive chicken tracheal samples were obtained from the source farm. One pool was obtained on 2 April, and the second pool was obtained on 4 April 2003. In addition, we analyzed a control pool of five A(H7N7) virus-positive chicken tracheal samples from an unrelated chicken egg layer farm, located 50 km away from the source farm, sampled on 31 March 2003. From each pool of five tracheal samples positive for A(H7N7) virus, 200 µl was used for RNA extraction.

(ii) Human samples

Two days after his visit to the source farm, on 4 April 2003, the veterinarian developed influenza-like illness with high fever and severe headache. Throat and eye swabs were obtained on 10 April 2003 and tested negative for influenza virus by two separate laboratories. He was hospitalized on 11 April 2003, 9 days after his farm visit, with high fever and pneumonia with infiltrates visible on the chest X-ray in the lower right lung lobe. On 13 April 2003, the patient was transferred to the intensive care unit due to respiratory distress requiring mechanical ventilation. A bronchoalveolar lavage (BAL) was performed on the same day with concurrent sputum analysis. On 14 April 2003, kidney function decreased, requiring continuous veno-venous hemofiltration, while blood gas values failed to improve despite mechanical ventilation with 100% oxygen. On 17 April 2003, the patient died as a result of respiratory failure [23]. After extensive laboratory testing for bacterial and viral pathogens, influenza A(H7N7) virus was detected in the BAL sample on 19 April. Virus culture of the BAL sample provided virus isolate A/Netherlands/219/03 on 23 April 2003 [21].

As part of this study, we obtained the original uncultured sputum sample as well as the original uncultured BAL sample from 13 April 2003 and used a total of 50 µl and 25 µl, respectively, for RNA extraction. In addition, RNA was extracted from a freshly cut section of formalin-fixed paraffin-embedded (FFPE) lung tissue obtained during autopsy, performed on 18 April 2003.

RNA extraction and viral load determination

RNA extraction was performed using the High Pure RNA isolation kit (Roche) with an on-column DNase treatment according to the manufacturer's protocol. RNA was extracted from the FFPE lower lung tissue using the RNeasy FFPE Kit (Qiagen) according to the manufacturer's protocol. Total RNA was eluted in a volume of 50 µl. All A(H7N7)-positive RNA extracts were subjected to a semiquantitative RT-PCR targeting the influenza virus matrix gene segment to prevent possible resequencing due to low-copy-number input in the reverse transcription (RT)-PCR [24].

Universal influenza virus full-genome amplification

Viral RNA was RT-PCR amplified using a universal 8-segment PCR method as described previously [25]. In short, two separate RT-PCRs were performed for each sample, using primers common-uni12R (5'-GCCGGAGCTCTGCAGATATCAGCRAAAGCAGG-3'), common-uni12G (5'-GCCGGAGCTCTGCAGATATCAGCGAAAGCAGG-3'), and common-uni13 (5'-CAGGAAACAGCTATGACAGTAGAAACAAGG-3'). The first RT-PCR mixture contained the primers common-uni12R and common-uni13. The second RT-PCR mixture contained the primers common-uni12G and common-uni13, which greatly improved the amplification of the PB2, PB1, and PA segments. Reactions were performed using the One-Step RT-PCR kit High Fidelity (Invitrogen) in a volume of 50 µl containing 5.0 µl eluted RNA with final concentrations of 1× SuperScript III One-Step RT-PCR buffer, 0.2 µM each primer, and 1.0 µl SuperScript III RT/Platinum Taq High Fidelity Enzyme Mix (Invitrogen). Thermal cycling conditions were as follows: reverse transcription at 42°C for 15 min, 55°C for 15 min, and 60°C for 5 min; initial denaturation/enzyme activation of 94°C for 2 min; 5 cycles of 94°C for 30 s, 45°C for 30 s, slow ramp (0.5°C/s) to 68°C, and 68°C for 3 min; 30 cycles of 94°C for 30 s, 57°C for 30 s, and 68°C for 3 min; and a final extension of 68°C for 5 min. After the PCR, equal volumes of the two reaction mixtures were combined to produce a well-distributed mixture of all 8 influenza virus segments. All RT-PCRs were performed in duplicate.

Influenza A(H7N7) virus HA and PB2 fragment amplification

Although the influenza virus matrix RT-PCR demonstrated that the viral load of the human samples was higher than that of the poultry samples, the universal influenza virus full-genome amplification of human samples was unsuccessful. The largest RT-PCR products detectable after the full-genome amplification of the human samples had a length of ~1 kb, illustrating that the amplification of larger influenza virus gene segments had failed due to RNA fragmentation. As mutation PB2 E627K is located on the largest of eight influenza virus gene segments with a length of >2,000 nucleotides, an amplification strategy targeting smaller fragments was needed to allow successful

characterization of minority variants. Therefore, specific amplification of a 410-nucleotide PB2 fragment covering codon 627 and a 362-nucleotide HA fragment covering codon 416 was applied *in duplicate*. A two-step protocol with SuperScript III and 1.0 µM specific primers followed by amplification with HotStarTaq Master Mix (Qiagen) and 0.5 µM each primer was performed [26]. PB2 primers H7-PB2F (5'-GAGCCCTTTCAATCCTTGGT-3') and H7-PB2R (5'-CTTGCCCTATCAATACATTAGCCT-3') and HA primers H7-HAF (5'-ATGTCCGAGATATGTTAA-3') and H7-HAR (5'-GATTCTCCATTGCTACTA-3') were used. Thermal cycling conditions were as follows: reverse transcription at 50°C for 1 h and denaturation at 95°C for 5 min; enzyme activation of 95°C for 15 min; 45 cycles of 94°C for 30 s, 50°C for 30 s, and 72°C for 1 min; and final extension of 72°C for 10 min.

Illumina sequencing

(i) Genome Analyzer IIx (GAIIx)

Products from the full-genome influenza virus RT-PCR of the pooled chicken samples were diluted to a DNA concentration of 50 ng/µl and sheared to a length of 200 to 400 bp using a Covaris AFA (Covaris, Woburn, MA). The sheared fragments were end repaired, A-tailed, and ligated to Illumina sequencing adaptors containing molecular identifier (MID) tags using the NEBNext DNA Library Prep Reagent Set (New England Biolabs, United Kingdom) to allow for multiplex sequencing per lane. Cluster generation was done using the TruSeq PE cluster generation kit version 4 (Illumina) according to the manufacturer's instructions. The three samples were sequenced in the same run on an Illumina GAII sequencer with a paired-end 54-bp run using the Sequencing by Synthesis kit version 5 according to the manufacturer's instructions (Illumina). All sample processing from shearing to sequencing was performed by the Wellcome Trust Sanger Institute (Hinxton, Cambridge, United Kingdom) as part of the European FP7 framework Emperie.

(ii) HiSeq 2000

Products from the HA and PB2 RT-PCR and plasmid pool RT-PCR (PCR control) as well as a plasmid pool sample (no-PCR control) were diluted to a DNA concentration of 50 ng/µl and sheared by nebulization. Samples were subjected to end repair, A-overhang, and adaptor ligation with MID tags using the Illumina TruSeq DNA sample preparation kit. The libraries were multiplexed, clustered, and sequenced on an Illumina HiSeq 2000 (TruSeq v3 chemistry) with a paired-end 100-cycle sequencing protocol with indexing. All sample processing from shearing to sequencing was performed by BaseClear BV (Leiden, The Netherlands).

Data analysis

Demultiplexing

The raw fastq files containing the sequences with corresponding quality scores were split into separate sample-specific files based on the corresponding MID sequence using the readset parser function of the Quality Assessment of Short Reads (QUASR) package version 7.0.1 [25].

Quality control

To correct for the inaccurate calling of nucleotides by the sequencer, an error correction step was performed on all MID-split data files using the paired-end quality control function of the QUASR pipeline version 7.0.1. In short, the median phred quality score of each read was calculated, and if this value was below a (user-defined) threshold, the read was trimmed from the 3' end until either the median quality cutoff requirement was met or the read length reached a user-defined minimal length cutoff and was discarded. In paired-end mode, the quality control step requires both mates to pass quality control. In this study, for the three chicken samples that were sequenced on an Illumina GAIIx, a median read phred quality score cutoff of 30 (corresponding to a 99.9% base call accuracy) was used in the quality control process and a length cutoff of 50. For all HiSeq 2000 data sets, a median read phred quality score cutoff of 33 was used in the quality control process and a length cutoff of 90. This difference in phred cutoff scores was determined empirically and was necessary due to inherent differences in the overall quality produced by the two different types of Illumina sequencers. Full details concerning error correction of deep sequencing data are published elsewhere [27].

Primer removal

The fastq files containing the reads that passed quality control were parsed to remove any sequences with exact matches to any of the primers used in the amplification or sequencing process in both the 5'and 3' orientation using in-house-generated PHP scripts.

Mapping strategy

All MID-split files containing quality-controlled primer-removed reads of the plasmid controls were mapped to the WSN reference sequence (NCBI taxonomy ID 382835) using the Burrows-Wheeler Aligner (BWA) version 0.5.9-r16 with the paired-end mapping option [28]. All data files containing quality-controlled reads from the human clinical samples as well as the pooled chicken samples were mapped to the consensus

sequence of virus isolate A/Nederland/219/03, which was isolated previously from the BAL specimen of the fatal case (NCBI taxonomy ID 680693). A quality-controlled consensus sequence was generated for each of the chicken samples using only nucleotides with a phred score above 30. This consensus sequence was subsequently used to remap all reads. Due to the internal concatenation of reference sequences by BWA, the resulting sequence alignment map (SAM) files were checked for mapping inconsistencies by removing reads with a mismatch rate of >10% for further analysis. In addition, mapped reads with either hard or soft clipping were removed for further analysis. As we mapped to a near-homologous consensus sequence, we removed reads from the SAM file with excessive mutations that might overestimate the viral diversity and limited the per-read mutation frequency to 2% by analyzing the NM tag (genetic distance from consensus sequence) from the SAM file [29]. Reads remaining after the removal of high mutation reads formed the cleaned (RHM) data sets. To characterize the type of influenza virus-specific sequence reads that were removed from the RAW data set to obtain the RHM data set, all quality-controlled but unmapped and clipped reads were split in half, and both parts were remapped against their sample-specific consensus sequence. Defective influenza virus RNA was defined when the difference in starting mapping positions of both read parts was greater than the length of the read [30].

Generation of coverage overview and quality-based consensus sequence

Both the RAW and RHM SAM-files of each sample were analyzed to generate a coverage overview per reference position disregarding nucleotides with a phred score below 30. Indel correction was done by analyzing the read CIGAR string [29], removing insertions and adding gaps to sites of deletions.

Determining SSEs

Sequence-specific errors (SSEs) are associated with a discrepancy in the mismatch frequency of mutations observed in mapped reads in forward and reverse orientation. To investigate the impact of SSEs, we split the SAM files in reads mapped in the forward orientation as well as the reverse orientation based on the bitwise flag value [29]. From the two resulting SAM files, a coverage overview was generated, and the mismatch frequency per position (i.e., the frequency of nucleotides not being the consensus nucleotide) was determined. To determine whether a nucleotide variation was an SSE, the difference in mismatch variation between the two orientations had to exceed the mean mismatch variation.

Mapping position of mutants in sequence reads

In case a particular mutation is due to a technical error, the mutation would be primarily located on a single specific location on all sequenced reads, as it originates from a single source read. A true variant is more evenly distributed over the entire length of the sequence reads, as it originates from more than 1 source read. To characterize the distribution of a particular mutation on the read, SAM files were parsed to extract all reads with a mutation on a predetermined position within the viral genome. The location of the mutation on the read was then determined for all individual reads. Mutational position graphs were generated and visually inspected for presence of a "hairy caterpillar"-like shape that indicates efficient shearing, unbiased amplification, and elimination of variation associated with the terminus of PCR products. When a mutational position graph was characterized by either one peak of >50% or more peaks of >20%, the corresponding minority variant was marked as an artifact and discarded. For the plasmid data set, the 200 nucleotide positions with the highest MMF (range, 0.078 to 0.38%) were analyzed, while all positions with a MMF of >0.5% were analyzed for the RT-PCR and A(H7N7) data sets.

Confirmation of influenza A(H7N7) virus defective RNA

To confirm the presence of defective influenza virus RNA in A(H7N7)-positive material obtained from the 2003 outbreak in the Netherlands, 32 influenza A(H7N7) virus-positive poultry samples and 6 human samples were analyzed [26,31]. Using a two-step protocol with SuperScriptTM III followed by amplification with HotStarTaq Master Mix and 0.5 µM H7-PB2ncrF (5'-AGCAAAAGCAGGTCAAATATATTC-3') and H7-PB2ncrR (5'-GGTCGTTTTTAAACAATTC-3'), the presence of PB2-defective RNA was examined in both human and poultry samples and Sanger sequenced with an ABI 3730 sequencer (see Fig. S1 in the supplemental material). In addition, this RT-PCR was performed on the BAL fluid, sputum, and lung tissue samples from the fatal case.

Results

The entire process from influenza A(H7N7) virus-positive samples to deep sequencing data consists of many steps in which a possible bias in the detection and quantification of minority variants can be introduced. Therefore, we first determined the technical error rates induced by the sequencing platform by directly sequencing a plasmid pool, as well as after the required RT-PCR, and optimized the data analysis in favor of sensitive detection of true virus minority variants.

Quantification of technical errors

A pool of 8 WSN reverse genetics plasmid DNA preparations were sequenced directly to determine the Illumina HiSeq2000-associated technical error. The average coverage per nucleotide position in the WSN genome was 66.100 and displayed a mean mismatch frequency (MMF) of 0.079% per reference position. The duplicate RT-PCR data sets (the average WSN genome coverage was 43.300) displayed a 3-fold increase of the mean MMF compared with the directly sequenced WSN plasmids and were characterized by the presence of nucleotide positions that reached 20% MMF. Although random induction of RT-PCR errors was expected, correlating (R^2 = 0.89) nucleotide positions in both RT-PCR data sets were characterized by a high MMF (Fig. 1A).

Characterization and elimination of artifacts

To identify the source of the high MMF in the RT-PCR data sets, we investigated the impact of unspecific primer binding and observed strong increases of the MMF at WSN genome locations that had up to 85% homology with the uni13-primer, indicative for nonspecific binding. To improve the overall sequence quality, clipped sequence reads, reads with indels, and reads that deviated more than 2% from the sample-specific consensus sequence were removed (RHM data set; Fig. 1A). Next, we examined the occurrence of sequence-specific errors for all nucleotide positions that displayed variation (Fig. 1B). For the RHM plasmid data set, all nucleotide positions with an MMF of >0.16% were characterized by SSE. For the RHM RT-PCR data sets, however, removal of SSE proved insufficient in eliminating the RT-PCR-induced variation. As primer sequences are associated with the 5′ and 3′ termini of PCR products, mutational position graphs were lastly applied to assess the distribution of sequence variation in the sequence reads (Fig. 1C). The optimized data analysis pipeline that allowed detection of minority variants present at ≥2% in RT-PCR products was subsequently applied on the deep sequence data sets of A(H7N7)-positive specimens (Fig. 1D). The full details of the analysis pipeline are described elsewhere [27].

Characterizing the avian source of infection

In a previous study, the consensus sequence of A/Netherlands/219/03 obtained from the fatal case contained two additional amino acid substitutions, PB2 E627K and HA K416R, compared with the consensus sequence of A/chicken/Netherlands/03010132/03 (H7N7) obtained from the source farm of the fatal human infection [10]. Deep sequence analysis of A(H7N7)-positive chicken specimens obtained from the source farm at the day of the visit of the fatal case and the day of culling, with acceptable threshold cycle (CT) values of 30.5 and 28.3 in the

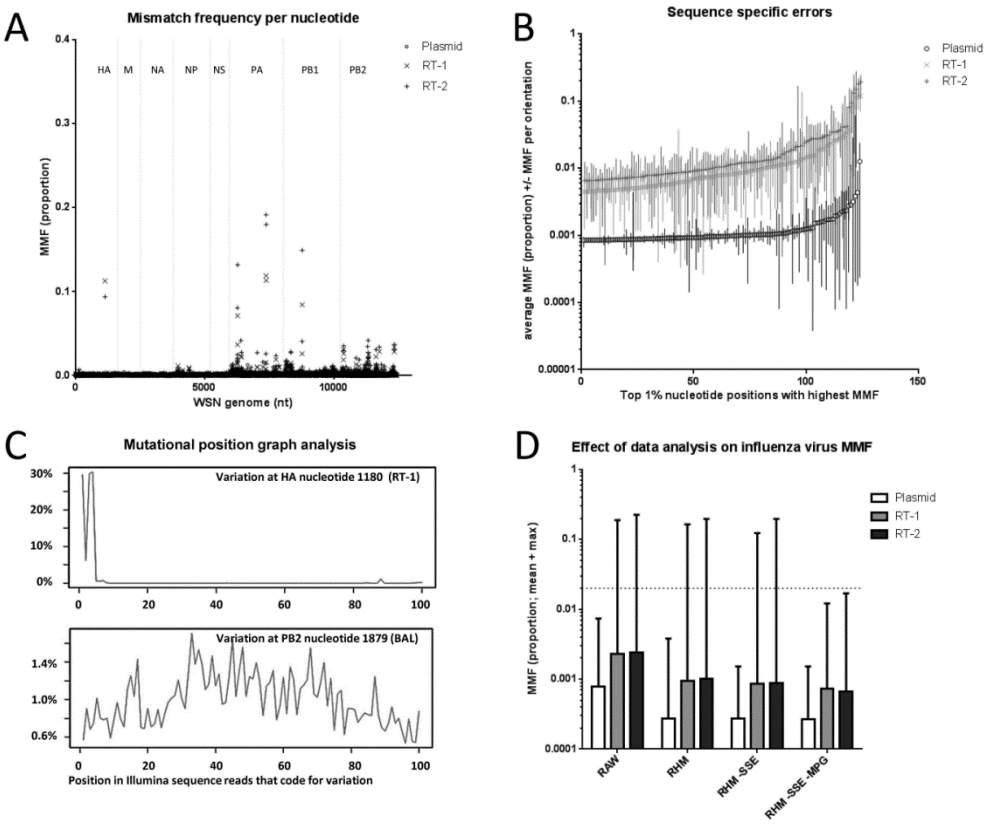


Figure 1. Steps of the data analysis pipeline and their effect on the influenza virus data set. (A) From the three control data sets, the mismatch frequency is expressed for each influenza virus nucleotide. This shows that the enhanced mapping algorithm was insufficient in removing the variation observed at specific nucleotides of the duplicate RT-PCR sets. (B) Focusing on 1% of the nucleotides that displayed the highest mismatch frequency, the sequence-specific errors are expressed by error bars. While plasmid nucleotides with the highest mismatch frequency were associated with sequence-specific errors, RT-PCR nucleotides with the highest mismatch frequency were not. (C) During the final step of the data analysis pipeline, all nucleotide positions displaying variation were assessed individually by mutational position graphs that demonstrate the location within Illumina sequence reads that code for variation. A mutational position graph of variation associated with the 5'-terminus of a PCR product is presented (top) together with a "hairy caterpillar"-shaped plot showing variation regardless of the position in the sequence fragment, thus considered a true variant (bottom). (D) effect of each type of data filtering on the mismatch frequency expressed as mean (+ range) for the plasmid, RT-PCR 1, and RT-PCR 2 data sets. While the enhanced mapping algorithm (RHM) and removal of sequence-specific errors (SSE) successfully reduced the mismatch frequency of nucleotides in the plasmid data set, mutational position graphs (MPG) were required to reduce the maximum mismatch frequency to less than 2% of the RT-PCR data sets.

matrix RT-PCR, failed to detect the presence of human adaptation marker PB2 E627K as well as the HA K416R change (Fig. 2). For each of the three poultry deep sequence data sets, variation from the consensus sequence was observed (Table 1). While the full-genome A(H7N7) consensus sequences of the two samples obtained from the source farm were identical except for one synonymous mutation at codon 72 in the Matrix gene segment, the A(H7N7) deep sequence data displayed variation at 20 nucleotide positions in the April 2 sample and 5 positions in the April 4 sample. Of the observed variation, only the nucleotide variation within codon 72 in the Matrix gene was present in A(H7N7) viruses obtained from both samples. The A(H7N7) viruses obtained from the control farm, with an acceptable ct value of 28.2 in the matrix RT-PCR, displayed 7 positions with variation. Human adaptation markers including PB2 E627K as well as HA K416R substitutions were absent in the A(H7N7) viruses obtained from both the source and control farms.

Table 1. Summary of within-farm influenza virus sequence diversity

Farm and collection date	No. of synonymous substitutions	No. of nonsynonymous substitutions	List and percentage of nonsynonymous substitutions
Source farm, 2 Apr '03	8	12	3% HA I335V, 3% HA M429I, 2% HA V441A, 2% HA I481T, 2% HA Y498H, 2% NA T137I, 3% NP R446I, 2% NP E454G, 2% PA D294Y, 3% PB1 P698S, 2% PB1 Q210P, 5% PB2 M315V
Source farm, 4 Apr '03	4	1	34% HA S102N
Control farm, 31 March '03	3	4	10% HA I211T, 9% NP I116V, 3% PA E181D, 27% PB2 I90M

In addition to the screening for human adaptation markers, sequence diversity can be used to characterize transmission chains. Conceivably, the observed deep sequence variation might harbor information that could enhance the resolution of transmission markers obtained by Sanger sequence data. The previously reported transmission network of the A(H7N7) outbreak demonstrated that our source farm was part of a transmission chain, characterized by the accumulation of mutations in HA, NA, and PB2, that ultimately led to introduction of the A(H7N7) virus in Belgium [26,31,32]. None of the mutations that characterized the A(H7N7)-positive poultry farms in the transmission chain before and after our source farm could be detected as minority variants using deep sequencing. In contrast, our control farm demonstrated minority variant PB2 I90M (27%), which characterized the transmission chain prior to infection of our control farm, providing a direct virological relation between viruses sampled at the control farm and its suggested source of infection (Table 1) [26].

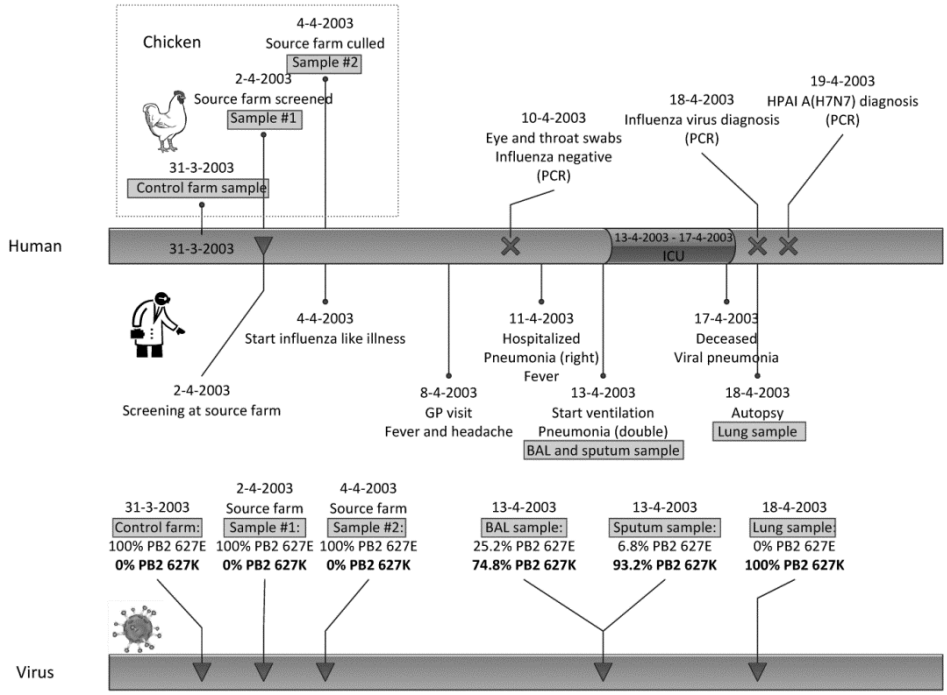


Figure 2. Timeline integrating the data on poultry and the fatal case and the corresponding virological data. It shows the increase of PB2 E627K during human infection in the absence of PB2 E627K detected at the avian source of infection. A quarter of the influenza A(H7N7) virus population in the lower respiratory tract of the veterinarian contained avian PB2 627E virus 11 days postexposure, while the postmortem lower lung tissue contained fully adapted (PB2 E627K) virus 16 days postexposure, suggesting that the PB2 E627K mutation emerged during infection of the human host.

Characterizing the human A(H7N7) virus infection

Three A(H7N7) virus-positive specimens were obtained from the fatal case: BAL fluid and sputum obtained 1 day after hospitalization (11 days postexposure) and postmortem lung tissue (16 days postexposure). The samples demonstrated acceptable ct values of 29.8, 25.1, and 27.5 in the matrix RT-PCR for the BAL fluid, sputum, and lung tissue, respectively. Since the influenza virus RNA in these specimens was too degraded for successful amplification of the larger influenza virus gene segments, specific RT-PCRs targeting small PB2 and HA fragments were performed *in duplicate* to allow characterization by deep sequence analysis. After quality control and data filtering, the only subpopulation that was detectable in the duplicate PB2 and HA fragments was located within PB2 codon 627, verified by the “hairy caterpillar”-shaped mutational position graph (Fig. 1C, bottom). The presence of the avian PB2

627E was detected at 30.0% and 20.5% in the *duplicate* BAL samples and at 7.3% and 6.3% in the *duplicate* sputum samples obtained in parallel. In the postmortem lung tissue obtained 5 days later, PB2 627E was not detectable. This demonstrated that the A(H7N7) virus infection was characterized by a mean 75% and 93% PB2 E627K mutation in the BAL and sputum samples obtained 11 days after exposure, followed by a dominant (100%) PB2 E627K mutation in the sample obtained during autopsy 5 days later (Fig. 2). The sequence variation observed at HA codon 416 was marked as artificial in the data analysis pipeline, based on the absence of a "hairy caterpillar"-shaped mutational position graph. These plots illustrated that >50% of the observed variation at HA codon 416 was associated with position 11 in sequence reads (Fig. 1C, top). Overrepresentation of one specific DNA template followed by resequencing was found to be the cause of this phenomenon. After correcting for the resequencing bias, we were able to demonstrate that the HA mutation K416R was the dominant population (~80%) at both sample days.

Identification of A(H7N7) virus defective RNA

In the data analysis pipeline, on average, 9% of the deep sequence data were removed from the RAW data sets to obtain the RHM data sets. To investigate the type of influenza virus-specific information that was lost by the removal of high mutation reads, these sequence reads were split and remapped against their sample-specific consensus sequence. Influenza A(H7N7) virus-specific sequence reads that demonstrated the presence of internal deletions for all eight influenza gene segments were encountered in the three A(H7N7) chicken samples, consistent with the detection of defective RNA (Fig. 3). The majority of the defective RNAs (80.0%) were derived from the gene segments that form the polymerase complex. Although internal deletions were observed with various sizes, the PB1 and PB2 subgenomic RNA (sgRNA) fragments were characterized by an ~1,900-nucleotide internal deletion and PA by an ~1,800-nucleotide deletion. In addition, 6% of the detected influenza virus defective RNAs comprised two different influenza virus gene segments. To independently confirm the presence of defective RNA in A(H7N7)-positive samples, additional A(H7N7)-positive samples obtained from 32 poultry farms and six human conjunctivitis cases during the 2003 outbreak in the Netherlands were analyzed by conventional RT-PCR and Sanger sequencing. This illustrated the presence of influenza virus PB2-defective RNA in 17/32 (53%) A(H7N7)-positive poultry samples and in 3/6 (n = 50%) A(H7N7)-positive human conjunctivitis samples (see Fig. S1 in the supplemental material). No PB2-defective RNA was detected in the samples obtained from the fatal case.

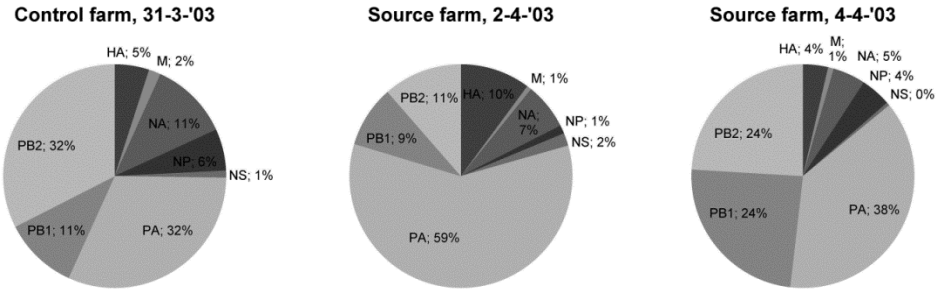


Figure 3. Detection of influenza virus defective RNA by deep sequencing. Characterization of the start and endpoint of internal deletions in A(H7N7) virus gene segments by deep sequence analysis identified a multitude of influenza virus defective RNAs in chicken samples obtained from the source and control farm. Of the total reads that contained an internal deletion, the pie charts illustrate the contribution for each of the influenza A(H7N7) virus gene segments.

Discussion

In this study, next-generation sequencing was applied to investigate the emergence of human adaptation marker PB2 E627K and HA K416R during a fatal case of HPAI A(H7N7) virus infection. Two uncultured samples from the avian source of infection and three uncultured samples obtained from the human fatal case were RT-PCR amplified prior to ultradeep sequencing using the Illumina platform. This platform was chosen for its ability to generate deep sequence data with the lowest reported error rates of all currently available platforms, allowing for ultradeep sequencing of our specimens with high resolution. For usage in next-generation sequencing, a high input DNA quantity is needed, requiring RT-PCR amplification of the influenza viral RNA. Our results illustrate the paradox in using RT-PCR for generating sufficient sequencing depth, as RT-PCR was associated with artifacts that hamper minority variant calling. By extended analysis of the data sets obtained from plasmid DNA and RT-PCR products, we developed a universal approach for distinguishing variation of viral RNA from artifacts and sequence bias that was subsequently applied on A(H7N7)-positive human and chicken samples.

The Illumina HiSeq 2000 platform had a maximum error below 0.2% for all nucleotides, which rose to a maximum of 20% after applying the RT-PCR. The “hairy caterpillar” plot analysis reduced the maximum error to less than 2% for all nucleotides, permitting characterization of viral diversity when present above 2%. Subsequent analysis of A(H7N7)-positive poultry samples illustrated the presence of a dominant influenza virus genotype for each farm supplemented with a multitude of low-frequency

sequence variants that are directly linked to the dominant variant, in agreement with conventional cloning results [31]. Interestingly, one minority variant (PB2 I90M) correlated with mutations from a previously characterized farm-to-farm transmission chain [26], suggesting that ultradeep sequence can be applied for more robust determination of a transmission network than that obtained with consensus sequencing approaches. In addition, the poultry samples were characterized by the presence of defective influenza virus RNA, as reported previously [33]. The formation of virus-derived defective RNA requires a high multiplicity of infection and has been reported under laboratory conditions [34,35]. Similarly, HPAI virus infection of poultry is associated with extremely high titers that might facilitate the emergence of defective influenza virus RNA. The presence of defective RNA corresponding with all eight influenza virus gene segments in the three A(H7N7) poultry samples illustrated the variety of defective RNA segments generated during HPAI virus infection of a chicken farm. It has been illustrated that defective influenza virus RNAs compete with full-length viral gene segments at the step of virion assembly and thereby inhibit the production of infectious progeny virus [36,37]. Subsequent virions that carry such an incomplete genome are known as defective interfering (DI) viruses. It has been demonstrated that DI influenza viruses *in vivo* provide protection against homologous and heterologous influenza virus infection in laboratory animals [38-40]. The presence of DIs in HPAI-infected poultry could attenuate the virulence for exposed humans, possibly aided by enhanced interferon (IFN) host response [41,42]. Additional experiments are needed to confirm a potential inhibitory role of DI influenza viruses during human HPAI virus infection.

For the samples obtained from the fatal human A(H7N7) case, mean frequencies of PB2 E627K ranged from 75% in the BAL fluid to 93% in the sputum and 100% in the postmortem sample (Fig. 2). The HA K416R mutation was present with a frequency of ~80% at both sample days and is in agreement with the previously determined consensus sequence A/Netherlands/219/03 [21]. Additional minority variants could be present, but failure of the universal influenza virus genome RT-PCR did not allow a full-genome analysis. The highest percentage of PB2 E627K was found in the sputum sample, containing mucus and spit from the carina of the lungs and up, while the BAL sample is restricted to the lower part of the airways. The difference in PB2 E627K frequency between BAL fluid and sputum might be due to the preferential replication temperature difference between the human-adapted and avian type PB2, allowing increased virus replication of the PB2 E627K variant at lower temperatures [11,18]. Modeling of the emergence and positive selection of human adaptation mutations, including PB2 E627K, during human influenza A(H5N1) virus infection demonstrated the plausibility of such a variant to become dominant during infection of a single

human host, consistent with our *in vivo* data [8]. The absence of PB2 E627K variants in poultry samples, combined with the detection of initial mixed PB2 627E and E627K variants during human infection, strongly suggests that the PB2 E627K mutation emerged during infection of the human host and correlated with disease severity.

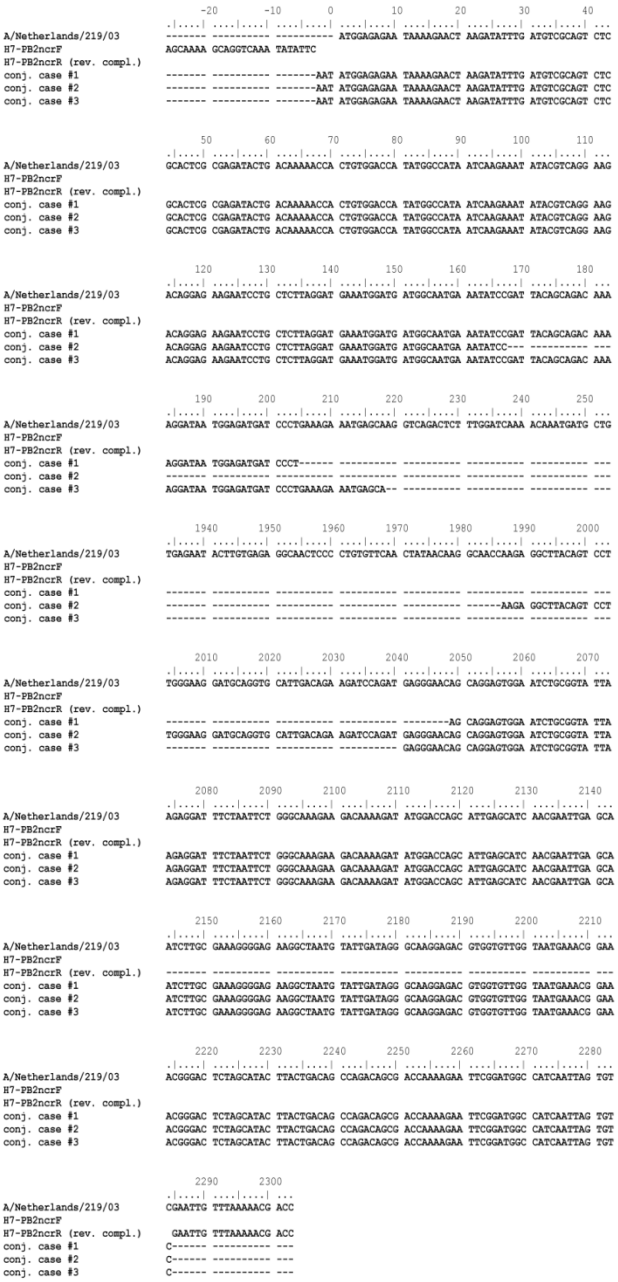
Herfst et al. demonstrated that PB2 E627K was a prerequisite for the development of airborne A(H5N1) virus in ferrets [7]. The accumulation of other human adaptation markers than PB2 E627K observed in avian influenza viruses from poultry and the wild bird population suggests that the introduction of this particular mutation in avian influenza viruses dramatically increases the virus pandemic potential and public health risk [8,26]. Moreover, the fatal human cases observed during the recent LPAI A(H7N9) outbreak in China are characterized by a virus carrying the PB2 E627K substitution, while viruses obtained from poultry are not [5,16,17]. Here, we illustrate that the conversion of an avian influenza virus toward a more human-tropic and potentially transmissible variant may occur during one infection cycle in humans, emphasizing the importance of reducing human exposure to avian influenza viruses on a global scale.

Acknowledgments

The authors would like to acknowledge G. Koch and A. Bataille for obtaining and sharing of influenza A(H7N7) virus-positive RNA from poultry farms. We thank R.G. Webster for providing the WSN33 reverse genetics plasmid set. This work was supported by the Dutch Ministry of Economic Affairs, Agriculture, and Innovation, Castellum Project. M. R. A. Welkers was funded by a personal Ph.D. scholarship from the AMC Graduate School.

Supplemental Figure S1

Nucleotide alignment of PB2 defective RNA segments obtained after RT-PCR amplification and Sanger sequencing of influenza A(H7N7) positive human eye swabs. Both the primers and the defective RNA segments are mapped against reference influenza strain A/Netherlands/219/03 (H7N7), illustrating the presence of a large deletions in the PB2 segment.



References

1. Webby R, Hoffmann E, Webster R (2004) Molecular constraints to interspecies transmission of viral pathogens. Nat Med 10: S77-81.
2. Swayne DE (2007) Understanding the complex pathobiology of high pathogenicity avian influenza viruses in birds. Avian Dis 51: 242-249.
3. Reperant LA, Rimmelzwaan GF, Kuiken T (2009) Avian influenza viruses in mammals. Rev Sci Tech 28: 137-159.
4. WHO (2013) Cumulative number of confirmed human cases of avian influenza A(H5N1) reported to WHO. http://www.who.int/influenza/human_animal_interface/H5N1_cumulative_table_archives/en/.
5. Gao R, Cao B, Hu Y, Feng Z, Wang D, et al. (2013) Human Infection with a Novel Avian-Origin Influenza A (H7N9) Virus. N Engl J Med.
6. Li Q, Zhou L, Zhou M, Chen Z, Li F, et al. (2013) Preliminary Report: Epidemiology of the Avian Influenza A (H7N9) Outbreak in China. N Engl J Med.
7. Herfst S, Schrauwen EJ, Linster M, Chutinimitkul S, de Wit E, et al. (2012) Airborne transmission of influenza A/H5N1 virus between ferrets. Science 336: 1534-1541.
8. Russell CA, Fonville JM, Brown AE, Burke DF, Smith DL, et al. (2012) The potential for respiratory droplet-transmissible A/H5N1 influenza virus to evolve in a mammalian host. Science 336: 1541-1547.
9. Imai M, Watanabe T, Hatta M, Das SC, Ozawa M, 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.
10. de Wit E, Munster VJ, van Riel D, Beyer WE, Rimmelzwaan GF, et al. (2010) Molecular determinants of adaptation of highly pathogenic avian influenza H7N7 viruses to efficient replication in the human host. J Virol 84: 1597-1606.
11. Hatta M, Gao P, Halfmann P, Kawaoka Y (2001) Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. Science 293: 1840-1842.
12. Li Z, Chen H, Jiao P, Deng G, Tian G, et al. (2005) Molecular basis of replication of duck H5N1 influenza viruses in a mammalian mouse model. J Virol 79: 12058-12064.
13. Subbarao EK, London W, Murphy BR (1993) A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. J Virol 67: 1761-1764.
14. Gabriel G, Dauber B, Wolff T, Planz O, Klenk HD, et al. (2005) The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. Proc Natl Acad Sci U S A 102: 18590-18595.

15. Le QM, Sakai-Tagawa Y, Ozawa M, Ito M, Kawaoka Y (2009) Selection of H5N1 influenza virus PB2 during replication in humans. J Virol 83: 5278-5281.
16. Kageyama T FS, Takashita E, Xu H, Yamada S, Uchida Y, Neumann G, Saito T, Kawaoka Y, Tashiro M. (2013) Genetic analysis of novel avian A(H7N9) influenza viruses isolated from patients in China, February to April 2013. . Euro Surveilance 18(15):pii=20453.
17. Jonges M, Meijer A, Fouchier R, Koch G, Li J, et al. (2013) Guiding outbreak management by the use of influenza A(H7Nx) virus sequence analysis. Euro Surveill 18.
18. Massin P, van der Werf S, Naffakh N (2001) Residue 627 of PB2 is a determinant of cold sensitivity in RNA replication of avian influenza viruses. J Virol 75: 5398-5404.
19. Du Ry van Beest Holle M, Meijer A, Koopmans M, de Jager CM (2005) Human-to-human transmission of avian influenza A/H7N7, The Netherlands, 2003. Euro Surveill 10: 264-268.
20. Koopmans M, Wilbrink B, Conyn M, Natrop G, van der Nat H, et al. (2004) Transmission of H7N7 avian influenza A virus to human beings during a large outbreak in commercial poultry farms in the Netherlands. Lancet 363: 587-593.
21. Fouchier RA, Schneeberger PM, Rozendaal FW, Broekman JM, Kemink SA, et al. (2004) Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome. Proc Natl Acad Sci U S A 101: 1356-1361.
22. Hoffmann E, Neumann G, Kawaoka Y, Hobom G, Webster RG (2000) A DNA transfection system for generation of influenza A virus from eight plasmids. Proc Natl Acad Sci U S A 97: 6108-6113.
23. Kemink SA, Fouchier RA, Rozendaal FW, Broekman JM, Koopmans M, et al. (2004) [A fatal infection due to avian influenza-A (H7N7) virus and adjustment of the preventive measures]. Ned Tijdschr Geneeskd 148: 2190-2194.
24. Jonges M, Liu WM, van der Vries E, Jacobi R, Pronk I, et al. (2009) Influenza virus inactivation for studies of antigenicity and phenotypic neuraminidase inhibitor resistance profiling. J Clin Microbiol 48: 928-940.
25. Watson SJ, Welkers MR, Depledge DP, Coulter E, Breuer JM, et al. (2013) Viral population analysis and minority-variant detection using short read next-generation sequencing. Philos Trans R Soc Lond B Biol Sci 368: 20120205.
26. Jonges M, Bataille A, Enserink R, Meijer A, Fouchier RA, et al. (2011) Comparative analysis of avian influenza virus diversity in poultry and humans during a highly pathogenic avian influenza A (H7N7) virus outbreak. J Virol 85: 10598-10604.
27. Welkers MR, Jonges M, Jeeninga RE, Koopmans MP, de Jong MD (2014) Improved detection of artifactual viral minority variants in high-throughput sequencing data. Front Microbiol 5: 804.

28. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754-1760.
29. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079.
30. Saira K, Lin X, DePasse JV, Halpin R, Twaddle A, et al. (2013) Sequence analysis of in vivo defective interfering-like RNA of influenza A H1N1 pandemic virus. J Virol 87: 8064-8074.
31. Bataille A, van der Meer F, Stegeman A, Koch G (2011) Evolutionary analysis of inter-farm transmission dynamics in a highly pathogenic avian influenza epidemic. PLoS Pathog 7: e1002094.
32. Van Borm S, Jonges M, Lambrecht B, Koch G, Houdart P, et al. (2012) Molecular Epidemiological Analysis of the Transboundary Transmission of 2003 Highly Pathogenic Avian Influenza H7N7 Outbreaks Between The Netherlands and Belgium. Transbound Emerg Dis.
33. Bean WJ, Kawaoka Y, Wood JM, Pearson JE, Webster RG (1985) Characterization of virulent and avirulent A/chicken/Pennsylvania/83 influenza A viruses: potential role of defective interfering RNAs in nature. J Virol 54: 151-160.
34. Choppin PW, Pons MW (1970) The RNAs of infective and incomplete influenza virions grown in MDBK and HeLa cells. Virology 42: 603-610.
35. Crumpton WM, Dimmock NJ, Minor PD, Avery RJ (1978) The RNAs of defective-interfering influenza virus. Virology 90: 370-373.
36. Duhaut SD, Dimmock NJ (2002) Defective segment 1 RNAs that interfere with production of infectious influenza A virus require at least 150 nucleotides of 5' sequence: evidence from a plasmid-driven system. J Gen Virol 83: 403-411.
37. Odagiri T, Tashiro M (1997) Segment-specific noncoding sequences of the influenza virus genome RNA are involved in the specific competition between defective interfering RNA and its progenitor RNA segment at the virion assembly step. J Virol 71: 2138-2145.
38. Dimmock NJ, Rainsford EW, Scott PD, Marriott AC (2008) Influenza virus protecting RNA: an effective prophylactic and therapeutic antiviral. J Virol 82: 8570-8578.
39. Mann A, Marriott AC, Balasingam S, Lambkin R, Oxford JS, et al. (2006) Interfering vaccine (defective interfering influenza A virus) protects ferrets from influenza, and allows them to develop solid immunity to reinfection. Vaccine 24: 4290-4296.
40. Noble S, Dimmock NJ (1994) Defective interfering type A equine influenza virus (H3N8) protects mice from morbidity and mortality caused by homologous and heterologous subtypes of influenza A virus. J Gen Virol 75 (Pt 12): 3485-3491.
41. Sekellick MJ, Marcus PI (1982) Interferon induction by viruses. VIII. Vesicular stomatitis virus: [+/-]DI-011 particles induce interferon in the absence of standard virions. Virology 117: 280-285.

42. Strahle L, Garcin D, Kolakofsky D (2006) Sendai virus defective-interfering genomes and the activation of interferon-beta. Virology 351: 101-111.

Chapter 5

Near real-time application of influenza virus sequence data

Guiding Outbreak Management by the Use of Influenza A(H7Nx) Virus Sequence Analysis

Marcel Jonges[1,2], Adam Meijer[1], Ron A.M. Fouchier[2], Guus Koch[3], Jun Li[4], Jin Cao Pan[4], Hualan Chen[5], Yue Long Shu[6], Marion Koopmans[1,2]

Euro Surveill. 2013 Apr 18;18(16):20460

[1]Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, The Netherlands

[2]Department of Viroscience, Erasmus MC, Rotterdam, The Netherlands

[3]Central Veterinary Institute, Wageningen University & Research Center, Lelystad, The Netherlands

[4]Hangzhou Center for Disease Control and Prevention, Hangzhou, China

[5]National Avian Influenza Reference Laboratory, Harbin Veterinary Research Institute, Harbin, China

[6]National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China

Abstract

The recently identified human infections with avian influenza A(H7N9) viruses in China raise important questions regarding possible source and risk to humans. Sequence comparison with an influenza A(H7N7) outbreak in The Netherlands in 2003 and an A(H7N1) epidemic in Italy in 1999–2000 suggests that widespread circulation of A(H7N9) viruses must have occurred in China. The emergence of human adaptation marker PB2 E627K in human A(H7N9) cases parallels that of the fatal A(H7N7) human case in the Netherlands.

Background

Since 31 March 2013, Chinese health authorities have been reporting human cases of avian influenza A(H7N9) virus infections. This novel reassortant influenza virus, carrying six internal gene segments of poultry A(H9N2) viruses, supplemented with a haemagglutinin (HA) subtype 7 and a neuraminidase (NA) subtype 9 originating from wild birds [1,2], has caused infections in at least 82 persons, of whom 17 have died, as of 17 April 2013. The human infections occurred in Eastern China in four provinces (Henan, Anhui, Jiangsu, and Zhejiang) and two municipalities (Shanghai and Beijing). Currently, the source of the human infections is unclear. However, in response to the detection of the influenza A(H7N9) virus among chickens, pigeons, ducks and environmental samples from some animal markets, as reported to the World Organisation for Animal Health (OIE), Chinese authorities have suspended live poultry trade and implemented the immediate closure of poultry markets, launched road inspections for transport of poultry, and have culled birds in an effort to deal with the issue. The outbreak raises important questions regarding possible source and risk to humans, and these will be addressed through case investigations. Here, we compare some findings from the first two weeks of the outbreak with those from a large highly pathogenic avian influenza (HPAI) A(H7N7) virus outbreak in The Netherlands in 2003 and from a low pathogenic avian influenza (LPAI) A(H7N1) epidemic in Italy in 1999–2000 [3-5] and discuss issues related to diagnosis and the use of molecular surveillance to monitor the outbreak.

Influenza A(H7N7) outbreak in The Netherlands in 2003

Exactly 10 years ago, the Netherlands was struck by an HPAI A(H7N7) virus outbreak that resulted in the infection of poultry on 255 farms and the subsequent culling of about 30 million chickens. A total of 453 exposed persons had mild symptoms and were investigated, of whom 89 were laboratory-confirmed as having an A(H7N7) virus infection [6,7].

Diagnosis of influenza A(H7Nx) virus infection

During the HPAI A(H7N7) virus outbreak in the Netherlands, almost all human cases had mild symptoms, particularly conjunctivitis, but one veterinarian died after an episode of severe influenza-like illness complicated by acute respiratory distress syndrome (ARDS)[7]. Diagnosis was based on virus detection by reverse transcription polymerase chain reaction (RT-PCR) from eye swabs, or combined nose and throat swabs. An important observation was that the sensitivity of eye swab-based diagnostics was much higher than that of diagnostics based on combined nose and throat swabs [6,7]. Similarly, in later sporadic infections of humans with H7 influenza A viruses, ocular symptoms were observed, probably caused by a preference of H7 influenza viruses for receptors in the eye [8]. Studies have shown that H7 influenza

viruses may use the ocular mucosa as portal of entry for systemic infection and that this is strain dependent [9,10]. Such symptoms have not been described for the cases of A(H7N9) virus infection in China in 2013, but it may be important to actively monitor for conjunctivitis in the outbreak investigation, as it may increase the success of case finding, particularly for mild cases.

Serological surveillance is important to rule out infection in patients sampled too late for direct virus detection and to assess the extent of transmission. This may be a problem since serological responses in persons with confirmed influenza A(H7Nx) virus infection have been difficult to detect, making assessment of A(H7N9) virus exposure using serosurveys challenging [11,12]. However, determining the kinetics of the antibody response in confirmed cases of influenza A(H7N9) virus infection will provide important information that can inform public health action.

Comparative analysis based on virus sequencing

Detecting the novel virus in animals is challenging as the A(H7N9) virus is a LPAI virus that is expected to cause few or no signs of disease in poultry, allowing silent spread among poultry flocks. The sharing of influenza A(H7N9) virus sequence data by both Chinese veterinary and public health institutes through the Global Initiative on Sharing All Influenza Data (GISAID) allows comparison with the sequences obtained during the Dutch outbreak. We therefore performed a comparative analysis using HA, NA and PB2 (subunit of the influenza virus RNA polymerase complex) fragment sequences from Chinese A(H7N9) viruses in 2013, Dutch A(H7N7) viruses in 2003 and sequences from a well-described LPAI A(H7N1) epidemic in Italy in 1999–2000 [5]. Providers of sequences downloaded from GISAID, listed with accession numbers, are acknowledged in the Table.

Sequence analysis of the Dutch viruses detected in poultry and in humans showed rapid diversification of the outbreak strain into multiple lineages (Figure). On the basis of the combined epidemiological and laboratory analyses, we demonstrated that sequences from humans were positioned mostly at ends of the branches of minimal spanning trees, confirming that humans were probably not involved in onward transmission [5].

In the current study, we compared the sequence diversity observed during the Dutch A(H7N7) outbreak and Italian A(H7N1) epidemic with the initial A(H7N9) virus sequences from the current outbreak in China. The maximum genetic distance generated during the three months of the Dutch HPAI A(H7N7) outbreak in concatenated HA, NA and PB2 segments of A(H7N7) viruses was 25 nucleotide substitutions. For the Italian LPAI A(H7N1) epidemic, the distance generated during a nine-month period was 66 nucleotide substitutions. For the A(H7N9) outbreak strains, this genetic distance is 35 substitutions, or 21 substitutions when the outlier strain A/Shanghai/1/2013 is ignored (Figure).

Table, panel A

Origin of the sequences of influenza A(H7Nx) viruses used for the comparative analysis

Segment ID	Segment	Country	Collection date	Isolate name	Originating laboratory	Submitting laboratory	Authors
EPI439488	PB2	China	2013-Jan-01	A/Shanghai/1/2013		WHO Chinese National Influenza Center	
EPI439486	HA	China	2013-Jan-01	A/Shanghai/1/2013		WHO Chinese National Influenza Center	
EPI439487	NA	China	2013-Jan-01	A/Shanghai/1/2013		WHO Chinese National Influenza Center	
EPI439495	PB2	China	2013-Jan-01	A/Shanghai/2/2013		WHO Chinese National Influenza Center	
EPI439500	NA	China	2013-Jan-01	A/Shanghai/2/2013		WHO Chinese National Influenza Center	
EPI439502	HA	China	2013-Jan-01	A/Shanghai/2/2013		WHO Chinese National Influenza Center	
EPI439504	PB2	China	2013-Jan-01	A/Anhui/1/2013		WHO Chinese National Influenza Center	
EPI439507	HA	China	2013-Jan-01	A/Anhui/1/2013		WHO Chinese National Influenza Center	
EPI439509	NA	China	2013-Jan-01	A/Anhui/1/2013		WHO Chinese National Influenza Center	
EPI441601	PB2	China	2013-Mar-24	A/Hangzhou/1/2013	Hangzhou Center for Disease Control and Prevention	Hangzhou Center for Disease Control and Prevention	Li,J; Pan,JC; Pu,XY; Yu,XF; Kou,Y; Zhou,YY
EPI440095	HA	China	2013-Mar-24	A/Hangzhou/1/2013	Hangzhou Center for Disease Control and Prevention	Hangzhou Center for Disease Control and Prevention	Li,J; Pan,JC; Pu,XY; Yu,XF; Kou,Y; Zhou,YY
EPI440096	NA	China	2013-Mar-24	A/Hangzhou/1/2013	Hangzhou Center for Disease Control and Prevention	Hangzhou Center for Disease Control and Prevention	Li,J; Pan,JC; Pu,XY; Yu,XF; Kou,Y; Zhou,YY
EPI440682	PB2	China	2013-Apr-03	A/Chicken/Shanghai/S1053/2013	Harbin Veterinary Research Institute	Harbin Veterinary Research Institute	
EPI440684	NA	China	2013-Apr-03	A/Chicken/Shanghai/S1053/2013	Harbin Veterinary Research Institute	Harbin Veterinary Research Institute	
EPI440685	HA	China	2013-Apr-03	A/Chicken/Shanghai/S1053/2013	Harbin Veterinary Research Institute	Harbin Veterinary Research Institute	
EPI440690	PB2	China	2013-Apr-03	A/Environment/Shanghai/S1088/2013	Harbin Veterinary Research Institute	Harbin Veterinary Research Institute	
EPI440692	NA	China	2013-Apr-03	A/Environment/Shanghai/S1088/2013	Harbin Veterinary Research Institute	Harbin Veterinary Research Institute	
EPI440693	HA	China	2013-Apr-03	A/Environment/Shanghai/S1088/2013	Harbin Veterinary Research Institute	Harbin Veterinary Research Institute	
EPI440698	PB2	China	2013-Apr-02	A/Pigeon/Shanghai/S1069/2013	Harbin Veterinary Research Institute	Harbin Veterinary Research Institute	
EPI440700	NA	China	2013-Apr-02	A/Pigeon/Shanghai/S1069/2013	Harbin Veterinary Research Institute	Harbin Veterinary Research Institute	
EPI440701	HA	China	2013-Apr-02	A/Pigeon/Shanghai/S1069/2013	Harbin Veterinary Research Institute	Harbin Veterinary Research Institute	

We acknowledge the authors, originating and submitting laboratories of the sequences from the Global Initiative on Sharing All Influenza Data (GISAID)s EpiFlu Database, on which this research is based.

Table, panel B

Origin of the sequences of influenza A(H7Nx) viruses used for the comparative analysis

Segment ID	Segment	Country	Collection date	Isolate name	Submitting laboratory	Authors
EPI238407	HA	Italy	1999-Jan-01	A/chicken/Italy/1067/1999	Other Database Import	Kim,L.M.; Scott,M.A.; Suarez,D.L.; Spackman,E.; Swayne,D.E.; Afonso,C.L.
EPI238409	NA	Italy	1999-Jan-01	A/chicken/Italy/1067/1999	Other Database Import	Kim,L.M.; Scott,M.A.; Suarez,D.L.; Spackman,E.; Swayne,D.E.; Afonso,C.L.
EPI238414	PB2	Italy	1999-Jan-01	A/chicken/Italy/1067/1999	Other Database Import	Kim,L.M.; Scott,M.A.; Suarez,D.L.; Spackman,E.; Swayne,D.E.; Afonso,C.L.
EPI63315	HA	Italy	1999-Jan-01	A/chicken/Italy/1082/1999	Other Database Import	
EPI63320	NA	Italy	1999-Jan-01	A/chicken/Italy/1082/1999	Other Database Import	
EPI63332	PB2	Italy	1999-Jan-01	A/chicken/Italy/1082/1999	Other Database Import	
EPI63258	HA	Italy	1999-Dec-21	A/chicken/Italy/4746/1999	Other Database Import	
EPI63263	NA	Italy	1999-Dec-21	A/chicken/Italy/4746/1999	Other Database Import	
EPI63275	PB2	Italy	1999-Dec-21	A/chicken/Italy/4746/1999	Other Database Import	
EPI69063	HA	Italy	1999-Dec-22	A/chicken/Italy/4789/1999	Other Database Import	
EPI69068	NA	Italy	1999-Dec-22	A/chicken/Italy/4789/1999	Other Database Import	
EPI69080	PB2	Italy	1999-Dec-22	A/chicken/Italy/4789/1999	Other Database Import	
EPI107331	PB2	Italy	1999-Jan-01	A/chicken/Italy/5093/99	Other Database Import	
EPI107338	HA	Italy	1999-Jan-01	A/chicken/Italy/5093/99	Other Database Import	
EPI107342	NA	Italy	1999-Jan-01	A/chicken/Italy/5093/99	Other Database Import	
EPI69044	HA	Italy	1999-Dec-30	A/quail/Italy/4992/1999	Other Database Import	
EPI69049	NA	Italy	1999-Dec-30	A/quail/Italy/4992/1999	Other Database Import	
EPI69061	PB2	Italy	1999-Dec-30	A/quail/Italy/4992/1999	Other Database Import	
EPI69215	HA	Italy	1999-Apr-10	A/turkey/Italy/1265/99	Other Database Import	
EPI69220	NA	Italy	1999-Apr-10	A/turkey/Italy/1265/99	Other Database Import	
EPI69232	PB2	Italy	1999-Apr-10	A/turkey/Italy/1265/99	Other Database Import	
EPI69177	HA	Italy	1999-Jul-12	A/turkey/Italy/2715/99	Other Database Import	
EPI69182	NA	Italy	1999-Jul-12	A/turkey/Italy/2715/99	Other Database Import	
EPI69194	PB2	Italy	1999-Jul-12	A/turkey/Italy/2715/99	Other Database Import	
EPI238618	HA	Italy	1999-Jan-01	A/turkey/Italy/2732/1999	Other Database Import	Kim,L.M.; Scott,M.A.; Suarez,D.L.; Spackman,E.; Swayne,D.E.; Afonso,C.L.
EPI238619	NA	Italy	1999-Jan-01	A/turkey/Italy/2732/1999	Other Database Import	Kim,L.M.; Scott,M.A.; Suarez,D.L.; Spackman,E.; Swayne,D.E.; Afonso,C.L.
EPI238624	PB2	Italy	1999-Jan-01	A/turkey/Italy/2732/1999	Other Database Import	Kim,L.M.; Scott,M.A.; Suarez,D.L.; Spackman,E.; Swayne,D.E.; Afonso,C.L.
EPI68484	HA	Italy	1999-Sep-03	A/turkey/Italy/3185/99	Other Database Import	
EPI68489	NA	Italy	1999-Sep-03	A/turkey/Italy/3185/99	Other Database Import	
EPI68501	PB2	Italy	1999-Sep-03	A/turkey/Italy/3185/99	Other Database Import	
EPI238625	HA	Italy	1999-Jan-01	A/turkey/Italy/3283/1999	Other Database Import	Kim,L.M.; Scott,M.A.; Suarez,D.L.; Spackman,E.; Swayne,D.E.; Afonso,C.L.
EPI238627	NA	Italy	1999-Jan-01	A/turkey/Italy/3283/1999	Other Database Import	Kim,L.M.; Scott,M.A.; Suarez,D.L.; Spackman,E.; Swayne,D.E.; Afonso,C.L.
EPI238632	PB2	Italy	1999-Jan-01	A/turkey/Italy/3283/1999	Other Database Import	Kim,L.M.; Scott,M.A.; Suarez,D.L.; Spackman,E.; Swayne,D.E.; Afonso,C.L.

We acknowledge the authors, originating and submitting laboratories of the sequences from the Global Initiative on Sharing All Influenza Data (GISAID)s EpiFlu Database, on which this research is based.

Table, panel C

Origin of the sequences of influenza A(H7Nx) viruses used for the comparative analysis

Segment ID	Segment	Country	Collection date	Isolate name	Submitting laboratory	Authors
EPI68503	HA	Italy	1999-Sep-23	A/turkey/Italy/3488/1999	Other Database Import	
EPI68508	NA	Italy	1999-Sep-23	A/turkey/Italy/3488/1999	Other Database Import	
EPI68520	PB2	Italy	1999-Sep-23	A/turkey/Italy/3488/1999	Other Database Import	
EPI69139	HA	Italy	1999-Sep-23	A/turkey/Italy/3489/99	Other Database Import	
EPI69144	NA	Italy	1999-Sep-23	A/turkey/Italy/3489/99	Other Database Import	
EPI69156	PB2	Italy	1999-Sep-23	A/turkey/Italy/3489/99	Other Database Import	
EPI69158	HA	Italy	1999-Sep-27	A/turkey/Italy/3560/99	Other Database Import	
EPI69163	NA	Italy	1999-Sep-27	A/turkey/Italy/3560/99	Other Database Import	
EPI69175	PB2	Italy	1999-Sep-27	A/turkey/Italy/3560/99	Other Database Import	
EPI238633	HA	Italy	1999-Jan-01	A/turkey/Italy/3675/1999	Other Database Import	Kim,L.M.; Scott,M.A.; Suarez,D.L.; Spackman,E.; Swayne,D.E.; Afonso,C.L.
EPI238635	NA	Italy	1999-Jan-01	A/turkey/Italy/3675/1999	Other Database Import	Kim,L.M.; Scott,M.A.; Suarez,D.L.; Spackman,E.; Swayne,D.E.; Afonso,C.L.
EPI238640	PB2	Italy	1999-Jan-01	A/turkey/Italy/3675/1999	Other Database Import	Kim,L.M.; Scott,M.A.; Suarez,D.L.; Spackman,E.; Swayne,D.E.; Afonso,C.L.
EPI42171	PB2	Italy	1999-Jan-01	A/turkey/Italy/4169/99	Other Database Import	
EPI89837	NA	Italy	1999-Jan-01	A/turkey/Italy/4169/99	Other Database Import	
EPI90388	HA	Italy	1999-Jan-01	A/turkey/Italy/4169/99	Other Database Import	
EPI69082	HA	Italy	1999-Nov-22	A/turkey/Italy/4294/99	Other Database Import	
EPI69087	NA	Italy	1999-Nov-22	A/turkey/Italy/4294/99	Other Database Import	
EPI69099	PB2	Italy	1999-Nov-22	A/turkey/Italy/4294/99	Other Database Import	
EPI68465	HA	Italy	1999-Nov-22	A/turkey/Italy/4295/1999	Other Database Import	
EPI68470	NA	Italy	1999-Nov-22	A/turkey/Italy/4295/1999	Other Database Import	
EPI68482	PB2	Italy	1999-Nov-22	A/turkey/Italy/4295/1999	Other Database Import	
EPI69125	NA	Italy	1999-Nov-22	A/turkey/Italy/4301/1999	Other Database Import	
EPI69137	PB2	Italy	1999-Nov-22	A/turkey/Italy/4301/1999	Other Database Import	
EPI69120	HA	Italy	1999-Nov-22	A/turkey/Italy/4301/1999	Other Database Import	
EPI238648	HA	Italy	1999-Jan-01	A/turkey/Italy/4482/1999	Other Database Import	Kim,L.M.; Scott,M.A.; Suarez,D.L.; Spackman,E.; Swayne,D.E.; Afonso,C.L.
EPI238650	NA	Italy	1999-Jan-01	A/turkey/Italy/4482/1999	Other Database Import	Kim,L.M.; Scott,M.A.; Suarez,D.L.; Spackman,E.; Swayne,D.E.; Afonso,C.L.
EPI238655	PB2	Italy	1999-Jan-01	A/turkey/Italy/4482/1999	Other Database Import	Kim,L.M.; Scott,M.A.; Suarez,D.L.; Spackman,E.; Swayne,D.E.; Afonso,C.L.
EPI238656	HA	Italy	1999-Jan-01	A/turkey/Italy/4580/1999	Other Database Import	Kim,L.M.; Scott,M.A.; Suarez,D.L.; Spackman,E.; Swayne,D.E.; Afonso,C.L.
EPI238658	NA	Italy	1999-Jan-01	A/turkey/Italy/4580/1999	Other Database Import	Kim,L.M.; Scott,M.A.; Suarez,D.L.; Spackman,E.; Swayne,D.E.; Afonso,C.L.
EPI238663	PB2	Italy	1999-Jan-01	A/turkey/Italy/4580/1999	Other Database Import	Kim,L.M.; Scott,M.A.; Suarez,D.L.; Spackman,E.; Swayne,D.E.; Afonso,C.L.
EPI69101	HA	Italy	1999-Dec-14	A/turkey/Italy/4617/1999	Other Database Import	
EPI69106	NA	Italy	1999-Dec-14	A/turkey/Italy/4617/1999	Other Database Import	
EPI69118	PB2	Italy	1999-Dec-14	A/turkey/Italy/4617/1999	Other Database Import	

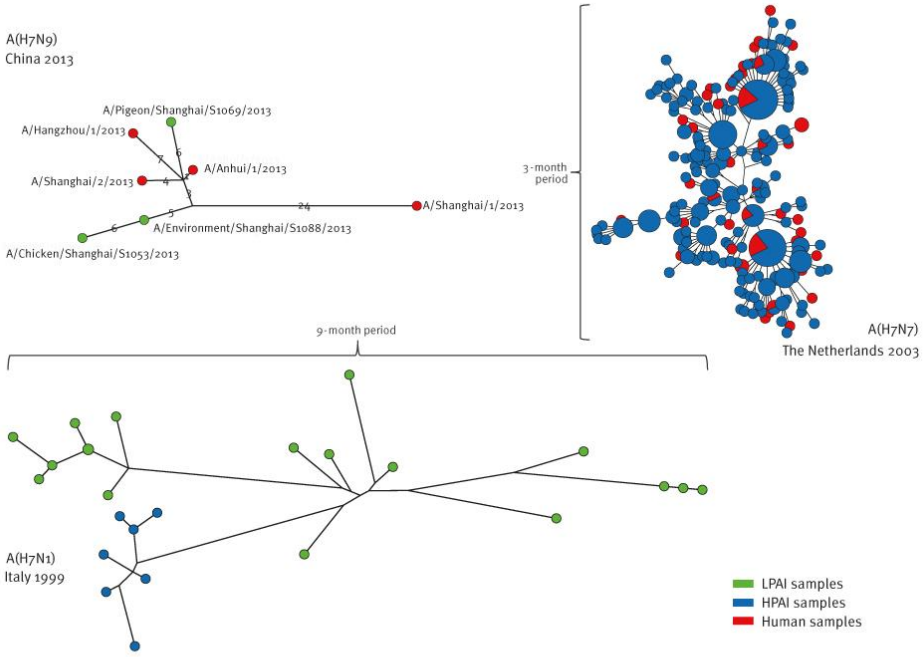
We acknowledge the authors, originating and submitting laboratories of the sequences from the Global Initiative on Sharing All Influenza Data (GISAID)s EpiFlu Database, on which this research is based.

Table, panel D

Origin of the sequences of influenza A(H7Nx) viruses used for the comparative analysis

Segment ID	Segment	Country	Collection date	Isolate name	Submitting laboratory	Authors
EPI69025	HA	Italy	1999-Dec-16	A/turkey/Italy/4644/99	Other Database Import	
EPI69030	NA	Italy	1999-Dec-16	A/turkey/Italy/4644/99	Other Database Import	
EPI69042	PB2	Italy	1999-Dec-16	A/turkey/Italy/4644/99	Other Database Import	
EPI68427	HA	Italy	1999-Dec-21	A/turkey/Italy/4708/1999	Other Database Import	
EPI68432	NA	Italy	1999-Dec-21	A/turkey/Italy/4708/1999	Other Database Import	
EPI68444	PB2	Italy	1999-Dec-21	A/turkey/Italy/4708/1999	Other Database Import	
EPI238664	HA	Italy	1999-Jan-01	A/turkey/Italy/977/1999	Other Database Import	Kim,L.M.; Scott,M.A.; Suarez,D.L.; Spackman,E.; Swayne,D.E.; Afonso,C.L.
EPI238666	NA	Italy	1999-Jan-01	A/turkey/Italy/977/1999	Other Database Import	Kim,L.M.; Scott,M.A.; Suarez,D.L.; Spackman,E.; Swayne,D.E.; Afonso,C.L.
EPI238671	PB2	Italy	1999-Jan-01	A/turkey/Italy/977/1999	Other Database Import	Kim,L.M.; Scott,M.A.; Suarez,D.L.; Spackman,E.; Swayne,D.E.; Afonso,C.L.

We acknowledge the authors, originating and submitting laboratories of the sequences from the Global Initiative on Sharing All Influenza Data (GISAID)s EpiFlu Database, on which this research is based.



HPAI: highly pathogenic avian influenza; LPAI: low pathogenic avian influenza.

The minimum spanning trees were constructed using concatenated haemagglutinin, neuraminidase and PB2 (subunit of the influenza virus RNA polymerase complex) nucleotide sequences in BioNumerics software version 6.6.4. The scaling of the branches, representing nucleotide substitutions, is equal for the three outbreaks.

Figure. Genetic diversity of three influenza A(H7Nx) virus outbreaks expressed by minimum spanning trees

All (n=7) NA sequences of the A(H7N9) viruses are characterised by a deletion in the stalk region, associated with adaptation to gallinaceous hosts [1,2,13]. Similar deletions in the NA stalk were also observed during the A(H7N7) outbreak in the Netherlands and the A(H7N1) epidemic in Italy [3]. Given the degree of sequence diversity present in initial A(H7N9) virus sequences, compared with that of the Dutch HPAI A(H7N7) and Italian LPAI A(H7N1) outbreak strains, and the large geographical area affected, the data are suggestive of (silent) spread and adaptation in domestic animals before the novel A(H7N9) virus was identified in humans.

Human adaptation markers

The majority of the Dutch human cases of A(H7N7) virus infection had mild symptoms, with the exception of one fatal case who was diagnosed with an A(H7N7) virus with the mammalian adaptation marker PB2 E627K. This mutation most probably occurred during infection of this case and was associated with high virulence [14]. Remarkably, the PB2 segments of the four available human virus genome sequences from China all

carry this E627K substitution, which is absent in the virus isolates obtained from birds and the environment [2]. In addition, three of the four infections with the virus with PB2 E627K were fatal. There are two plausible explanations for this observation:

1. The mammalian adaptation markers are selected during replication in humans following exposure to viruses that do not have this mutation, which are circulating in animals;
2. The mammalian adaptation markers result from virus replication in animals from which humans become infected.

The relatively protracted disease course in the current outbreak of A(H7N9) virus infection, with relatively mild symptoms at first, followed by exacerbation in the course of a week or longer, is suggestive of the first hypothesis, similar to the outbreak in the Netherlands. In this scenario, an important difference in the A(H7N7) observations from the Netherlands is the frequency of finding the PB2 E627K mutation in humans (4/4 A(H7N9) sequenced patient strains compared with 1/61 sequenced A(H7N7) patient strains). Therefore, an outstanding question is whether the A(H7N9) viruses are more readily mutating in humans or milder cases are being missed. Contact investigations have found no mild cases and only one asymptomatic case), but in order to address this issue, more enhanced testing of persons exposed to a similar source is needed, using the most sensitive tests available on the optimal clinical specimen type obtained at the right time.

Although human infections with H7 influenza viruses have occurred repeatedly over the last decades without evidence of sustained human-to-human transmission, the absence of sustained human-to-human transmission of A(H7N9) viruses does not come with any guarantee. Five of seven A(H7N9) virus strains obtained from humans (n=2), birds (n=2) and the environment (n=1) have a mutation in HA, Q226L, that is associated with binding to alpha(2,6)-linked sialic acids, the virus receptors in the human upper respiratory tract [2]. This Q226L substitution in combination with G228S has been associated with human receptor preference for influenza viruses that caused the pandemics of 1957 and 1968 and with airborne transmission of A(H5N1) virus [15,16]. For H7 viruses, it has recently been demonstrated that these mutations also increased human receptor-binding affinity [17]. In combination with the PB2 E627K mutation, the A(H7N9) virus thus contains two well-known mammalian adaptation markers.

Conclusion

Comparative analysis of the first virological findings from the current outbreak of influenza A(H7N9) virus infection in China with those from other influenza A(H7Nx) outbreaks suggests that widespread circulation must have occurred, resulting in major

genetic diversification. Such diversification is of concern, given that several markers associated with increased risk for public health are already present. Enhanced monitoring of avian and mammalian animal reservoirs is of utmost importance as the public health risk of these A(H7N9) viruses may change following limited additional modification.

Acknowledgements
This work was supported by the Dutch Ministry of Economic Affairs, Agriculture, and Innovation, Castellum Project. We acknowledge the authors, originating and submitting laboratories of the sequences from the Global Initiative on Sharing All Influenza Data (GISAID)s EpiFlu Database (www.gisaid.org), on which this research is based.

References

1. Gao R, Cao B, Hu Y, Feng Z, Wang D, et al. (2013) Human Infection with a Novel Avian-Origin Influenza A (H7N9) Virus. N Engl J Med.
2. Kageyama T FS, Takashita E, Xu H, Yamada S, Uchida Y, Neumann G, Saito T, Kawaoka Y, Tashiro M. (2013) Genetic analysis of novel avian A(H7N9) influenza viruses isolated from patients in China, February to April 2013. . Euro Surveilance 18(15):pii=20453.
3. Capua I, Mutinelli F, Pozza MD, Donatelli I, Puzelli S, et al. (2002) The 1999-2000 avian influenza (H7N1) epidemic in Italy: veterinary and human health implications. Acta Trop 83: 7-11.
4. Bataille A, van der Meer F, Stegeman A, Koch G (2011) Evolutionary analysis of inter-farm transmission dynamics in a highly pathogenic avian influenza epidemic. PLoS Pathog 7: e1002094.
5. Jonges M, Bataille A, Enserink R, Meijer A, Fouchier RA, et al. (2011) Comparative analysis of avian influenza virus diversity in poultry and humans during a highly pathogenic avian influenza A (H7N7) virus outbreak. J Virol 85: 10598-10604.
6. Koopmans M, Wilbrink B, Conyn M, Natrop G, van der Nat H, et al. (2004) Transmission of H7N7 avian influenza A virus to human beings during a large outbreak in commercial poultry farms in the Netherlands. Lancet 363: 587-593.
7. Fouchier RA, Schneeberger PM, Rozendaal FW, Broekman JM, Kemink SA, et al. (2004) Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome. Proc Natl Acad Sci U S A 101: 1356-1361.
8. Olofsson S, Kumlin U, Dimock K, Arnberg N (2005) Avian influenza and sialic acid receptors: more than meets the eye? Lancet Infect Dis 5: 184-188.
9. Belser JA, Wadford DA, Xu J, Katz JM, Tumpey TM (2009) Ocular infection of mice with influenza A (H7) viruses: a site of primary replication and spread to the respiratory tract. J Virol 83: 7075-7084.
10. Belser JA, Davis CT, Balish A, Edwards LE, Zeng H, et al. (2013) Pathogenesis, transmissibility, and ocular tropism of a highly pathogenic avian influenza A (H7N3) virus associated with human conjunctivitis. J Virol.
11. Meijer A, Bosman A, van de Kamp EE, Wilbrink B, Du Ry van Beest Holle M, et al. (2006) Measurement of antibodies to avian influenza virus A(H7N7) in humans by hemagglutination inhibition test. J Virol Methods 132: 113-120.
12. Min JY, Vogel L, Matsuoka Y, Lu B, Swayne D, et al. (2010) A live attenuated H7N7 candidate vaccine virus induces neutralizing antibody that confers protection from challenge in mice, ferrets, and monkeys. J Virol 84: 11950-11960.

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. de Wit E, Munster VJ, van Riel D, Beyer WE, Rimmelzwaan GF, et al. (2010) Molecular determinants of adaptation of highly pathogenic avian influenza H7N7 viruses to efficient replication in the human host. J Virol 84: 1597-1606.
15. Pappas C, Viswanathan K, Chandrasekaran A, Raman R, Katz JM, et al. (2010) Receptor specificity and transmission of H2N2 subtype viruses isolated from the pandemic of 1957. PLoS One 5: e11158.
16. Herfst S, Schrauwen EJ, Linster M, Chutinimitkul S, de Wit E, et al. (2012) Airborne transmission of influenza A/H5N1 virus between ferrets. Science 336: 1534-1541.
17. Srinivasan K, Raman R, Jayaraman A, Viswanathan K, Sasisekharan R (2013) Quantitative description of glycan-receptor binding of influenza a virus h7 hemagglutinin. PLoS One 8: e49597.

Chapter 6

Summary & Discussion

This thesis provides an overview of my research done on the use of influenza sequence data to support outbreak control measures. Specifically, it enabled the detection of sources of infection and patient clusters based on retrospective analysis of influenza virus sequence data sets obtained during influenza virus outbreaks and epidemics. The knowledge obtained in this thesis can target intervention methods in the future that eventually may reduce virus spread.

The introduction of this thesis, Chapter 1, presents general information on influenza viruses and introduces the engine of virus sequence diversity, generated and maintained by the error-prone viral RNA polymerase. Seasonal influenza viruses obtained during the molecular surveillance provide a representation of the sequence diversity present in circulating viruses and it was hypothesized that they can be used as reference to assess the uniqueness of clustered influenza virus cases (Chapter 2). Initial experiments targeted potential nosocomial transmission of oseltamivir-resistant viruses, using a limited number of national (n=17), regional (n=3) and hospital influenza A(H1N1) viruses (n=5). We were able to support epidemiological data on a suspected cluster of nosocomial resistant virus infections as the influenza A(H1N1) patient cluster was characterized by a unique mutation in the NA enzyme (Chapter 2.1). Besides this patient cluster, none of the NA sequences available in GenBank contain this specific NA sequence motif (access date: March 27, 2015). Following the emergence of pandemic influenza A(H1N1)pdm09, a sequence dataset of 685 community influenza A(H1N1)pdm09 sequences was generated during the 2009/2010 epidemic, providing knowledge on national influenza virus sequence diversity. Sequence comparison of A(H1N1)pdm09 viruses (n=107) obtained from one hospital with community viruses identified eight patient clusters that had a low probability of occurrence ($P < 0.01$) compared with national diversity (Chapter 2.2). Epidemiological analysis demonstrated nosocomial infections in four of these eight clusters, and a mother-child combination in a fifth cluster, implicating that hospital sequence clusters that differ from dominant community strains should requiring further investigation. In contrast, patient clusters (n=9) having a higher probability of occurrence compared with national diversity ($P > 0.01$), all were associated with community acquired infections. The results demonstrated in Chapters 2.1 and 2.2 suggest that transmission chains within a single epidemic are associated with the synthesis of a vast amount of nearly identical influenza viruses, of which the majority of the variants become extinct because of dead-end transmissions.

Next, we evaluated the possibility to identify influenza virus transmission events using virus sequence data from an influenza virus outbreak (Chapter 3). A combined human and veterinary sequence dataset from a densely sampled avian influenza A(H7N7) virus outbreak was exploited to study whether influenza virus mutations with increased public health risk emerged during infection of humans, or were already present in

viruses obtained from poultry (Chapter 3.1). Viruses obtained from human A(H7N7) conjunctivitis cases reflected the virus diversity generated in poultry and appeared dead-end infections. Moreover, the virus transmission network showed that the polygenic accumulation and spread of virulence and human adaptation markers occurred in poultry following farm-to-human transmissions. When comparative analysis of HPAI A(H7N7) sequence data and meteorological data demonstrated statistical evidence that the direction of farm-to-farm spread correlated with the wind direction at the date of infection [1], field experiments were conducted to investigate the possible role of wind-mediated spread of avian influenza viruses during outbreaks in poultry (Chapter 3.2). Samples of suspended particulate matter were collected, inside and on downwind locations from buildings holding poultry infected with avian influenza viruses. In addition to one in-barn measurement showing an influenza virus concentration of $8.48x10^4$ genome copies/m^3, airborne influenza viruses were detected by reverse transcriptase PCR collected up to 60 meters downwind from the barns. As a good correlation between field measurements and modeled particulate concentrations was observed, geographical estimates of areas at high risk for human and animal exposure to airborne influenza virus could be modeled during an outbreak.

It was the A(H7N7) transmission network that stressed the importance of investigating farm-to-farm transmission due to the its potential impact on public health (Chapter 3.1). One possible exception was the detection of virulence marker PB2 E627K in a fatal human A(H7N7) case, that could have emerged in poultry but also during human infection. A NGS approach was attempted to characterize the A(H7N7) within-host diversity in samples obtained from poultry and the human fatal case. As characterization of viral minority variants is limited by the capacity to distinguish biological from artificial variation, the impact of quality control on both technical errors as well as RT-PCR induced errors in NGS datasets was evaluated (Chapter 4.1). Two types of artifacts were characterized and their removal from NGS datasets resulted in a 95% reduction of artificial variation. The optimized NGS data analysis pipeline was subsequently applied on NGS datasets of longitudinal A(H7N7) virus positive samples obtained from both poultry and the human case to investigate the emergence of the virulence marker PB2 E627K (Chapter 4.2). In the human samples, the PB2 E627K mutation was identified with increasing frequency during the course of the infection. Our results strongly suggest that human adaptation marker PB2 E627K has emerged during virus infection of a single human host, emphasizing the importance of reducing human exposure to avian influenza viruses to reduce the likelihood of viral adaptation to humans.

The knowledge gained on the use of influenza virus sequence diversity to characterize patient clusters (Chapter 2), transmission chains (Chapter 3) and to monitor the emergence of human adaptation markers (Chapter 3 and 4) were combined in Chapter 5. From March 31 2013 onward, Chinese health authorities reported human cases of

A(H7N9) virus infections and immediately shared their virological findings. Near real-time comparative sequence analysis of A(H7N9) viruses with those from the A(H7N7) virus outbreak in the Netherlands in 2003 and from an A(H7N1) epidemic in Italy in 1999/2000 suggested that widespread circulation must have occurred, resulting in major genetic diversification. Such diversification was of concern, given that several human adaptation markers were already present.

Using transmission markers; Résumé

When sequence based information on patient clusters and transmission chains is generated in a timely fashion, such data might influence infection control measures and lead to isolation of sources and/or interruption of transmission chains. There are however, specific practical issues that should be considered before drawing conclusions on sequence datasets and provide unsubstantiated recommendations. In the next paragraphs, the issues affecting conclusions drawn on influenza sequence datasets are addressed, comprising the sequence methodology and analysis, and availability of metadata.

Method dynamics

Introduction NGS

In the last decade, the standard in the sequence analysis of viruses has shifted from conventional Sanger sequencing technique towards high throughput NGS approaches, which can be applied to provide complete genome coverage of a virus [2]. Moreover, as a NGS approach generates millions of reads from a single specimen, it also permits characterization of minority variants and thereby provides information on the mutants covered by the quasi species population in sequence space. Sequence data covering the mutant spectrum of each virus infection in theory could provide the basis for highly robust transmission networks. So far, NGS-based studies have been able to identify direct transmission between individual patients infected with Norovirus, Hepatitis C virus and Human immunodeficiency virus by the detection of shared sequence space between infected patients [3-5]. Interestingly, these studies illustrated that the uniqueness of the established phylogenetic relationships were not resulting from the use of all informative sites (whole genome sequences in contrast to gene fragments), but from the sequence depth obtained within genome fragments. By deep amplicon NGS, it was shown that it were often minority variants of the donor that dominated the infection of the recipient host, demonstrating direct phylogenetic relationships between cases in the absence of a conserved consensus sequence. A similar observation but on a significantly larger scale was made by deep amplicon NGS analysis of chicken samples during the A(H7N7) virus outbreak in the Netherlands, where quasi species characterization was able to identify the consensus sequence of a source farm as minority variant in the recipient farm [6]. Although the closeness

between viral populations found in short transmission chains might suggest that a deep NGS approach is the method of choice for the robust characterization of transmission chains, prospective real-time implementation is complicated by the duration of the NGS experiment and the laborious NGS analysis needed to eliminate artificial sequence variation induced during each NGS experiment.

NGS platforms

For prospective NGS implementation during virus outbreaks, in order to be used to target outbreak control interventions, the data on virus transmissions should be available within days. In this thesis we showed that a consensus Sanger sequencing approach can provide data on patient clusters or virus transmissions at the same day the specimen was obtained. In contrast, the HiSeq (Illumina) NGS platforms that dominate the sequencing industry recent years have a runtime of over a week, with a median turnaround time in routine clinical care of 33 days on a HiSeq 2000 and 18 days on a HiSeq 2500 platform [7,8]. More recently, two competitive MiSeq (Illumina) and Proton PI (LifeTechnologies) benchtop NGS platforms were released that enable a combined sample preparation time and sequencing runtime of approximately one day, costing $46-53 (MiSeq) and $14-15 (Proton PI) for each sample (assuming 1 million raw reads), and launching the breakthrough in the public health use of NGS technologies [9]. However, in depth data on platform performance is yet sparse for the Proton PI, and based on the performance of its predecessor Ion Torrent PGM the error rate is significantly higher than the error rate of the MiSeq and datasets are produced where error-free reads are the minority [7]. Although this might suggest that the semi-conductor technology of the Proton PI is not suitable for the robust quasi species characterization, acknowledged difficulties with homopolymer stretches dominate the errors and these are less of a problem when sequencing virus genomes [10]. Therefore, timely peer-reviewed data on platform performance is relevant for all potential future users, in addition to the available information provided by the NGS manufacturers. To gain acceptance of NGS in clinical laboratories, manufacturers should collaborate with independent multidisciplinary researchers to validate their NGS platforms and share their results in contrast to the current situations where the performance evaluation responsibility lies with the end-user. Currently, evidence based implementation of a NGS platform for both rapid and robust quasi species characterization is limited to the MiSeq platform, due to absence of thorough Proton PI platform evaluations.

NGS data handling

Viral genomes have to be amplified several orders of magnitude to yield the required amount of DNA for current NGS platforms. Unavoidable errors are introduced during the reverse transcription of viral RNA by error-prone transcriptases that lack proof-

reading activity [11,12]. During the subsequent PCR, mispriming of utilized primers, nucleotide misincorporation by polymerases and resampling due to low DNA input may introduce a subsequent source of errors that are not completely avoidable [13-15]. And finally, NGS platform specific sequence errors are introduced during each NGS run [16,17]. When errors introduced during sample preparation and sequencing run remain unrecognized during the NGS data analysis, this may lead to an overestimation of the viral diversity.

The initial step in NGS data analysis is the filtering or trimming of low-quality sequence reads, which can result in a 300-fold reduction of the error rate [18]. For quasi species analysis, the next step is the alignment of the remaining sequence reads to a reference sequence, followed by an optional re-alignment to its global consensus sequence [19]. Finally, the relative nucleotide frequencies at all genomic positions are characterized from the aligned sequence reads, taking into account e.g. the individual nucleotide quality scores, reads containing insertions or deletions, read orientation and/or the distribution of mutations over the sequence reads [20]. The reliability of the characterizing nucleotide variation at low frequencies, termed minority variant calling, can be further improved by including a control experiment using a clonal sample, to characterize the experimentally induced error patterns and compare these with the biological samples [20,21]. However, there is yet no uniformity on any of the NGS data analysis steps mentioned above for the robust characterization of minority variants. Current NGS users apply a multitude of existing, updated or novel analytical software tools for individual NGS data analysis steps [22]. Many of these tools have not been published and/or evaluated for their performance, consistency and quality [23]. To facilitate the synchronization of NGS data analysis in the nearby future, a platform for the continuous sharing and evaluation of available NGS software tools would be a valuable resource [24]. A recent example of an open-access directory of classified NGS software tools is OMICtools, where detailed information on available commercial and non-profit tools including published evaluations are provided [25]. However, this continuously updated directory of NGS tools does not yet contain a knowledge-sharing resource for the NGS community to address experimental and computational aspects of NGS tools as has been provided by e.g. SEQanswers [26]. Consequently, an integrated global effort would be favored that allows rapid NGS tool development and validation to allow standardized and robust NGS data interpretation and sharing [27]. Scientific journals can contribute to the global understanding on NGS data analysis, firstly by allowing a detailed description of the analysis into the methods section, and secondly by making it mandatory to provide these details. Almost a decade after the introduction of NGS technologies the time has come to harmonize NGS data analysis, by the listing of specific types of sequence artifact and subsequent development of tools for their identification and removal from NGS datasets. An international team of NGS experts and bioinformaticists can fulfill this task, and give rise to a publicly

available NGS data validation pipeline enabling more standardized NGS data interpretation in the near future.

Sequence analysis

For the analysis of nearly identical consensus sequences obtained using short time scales from densely sampled hosts, the maximum parsimony tree depicted in chapter 3.1 highly correlated with epidemiological data. Although this simple analysis in which the tree is inferred based on the minimal number of necessary substitutions, proved useful in the short time scale of influenza virus outbreaks where the inter-sample diversity is low due to high sampling density, the credibility of the results needs to be verified.

The central problem concerning maximum parsimony analysis is the frequent existence of multiple equally parsimonious solutions that all yield equally valid interpretations of virus transmissions [28,29]. To assess the credibility of an inferred tree, bootstrapping or the use of Bayesian posterior probabilities are commonly applied [30]. To characterize the impact of multiple equally parsimonious solutions on the obtained maximum parsimony tree, we performed a permutation resampling strategy to calculate the support of individual branches, leading to the identification of sources of infection with high significance [6,31]. The maximum parsimony method utilizes character data, where the number of pairwise differences is simply counted (Hamming distance) and substitution models cannot be applied [32]. As a consequence, the number of observed substitutions is in theory smaller than the actual number of substitutions because complicated genuine substitution patterns e.g. back-substitutions are ignored and all substitutions are weighed equally [33]. Novel approaches that combine maximum parsimony with substitution models or defined distributions of pairwise differences are available, but demand details on mutations rate, transmission dynamics and/or substitution model that typically are not known at the start of an emerging disease outbreak [34,35]. Despite these caveats, the maximum parsimony approach appeared accurate: by combining both epidemiological and sequence data containing 260 concatenated influenza virus sequences obtained during one outbreak, only two transmission events were identified that might have been associated with back-mutations [6]. Furthermore, the availability of a high sample density and a short time span, yield influenza virus sequence datasets where the majority of the observed substitutions do not enhance the phenotype. The phenotypic stability provided by mutational robustness forms the basis of virus adaptability and eventually virus evolution, by the accumulation of genetic diversity that allows access to great phenotypic diversity by few substitutions. Although phenotypic differences drive (Darwinian) evolution, it is unclear whether the underlying mutational robustness is merely accidental or selected for. Consequently, we do not favor evolutionary models that take into account substitution models

and/or mutation rates, but we use the minimum number of substitutions to explain the data.

We illustrate that maximum parsimony trees were able to identify sources of infection with high significance (Chapter 3.1), but we specifically did not apply the method to estimate the origin of the corresponding outbreak as branches located at the root of the tree were associated with low resampling support [6]. We did apply the method to identify clusters of identical sequences (Chapter 2.2) or to visualize virus diversification (Chapter 5) without the need for prior knowledge on viral infection dynamics, irrespective of the availability of such data [36,37]. Computationally demanding maximum likelihood and Bayesian methods are considered to be the most accurate for phylogenetic inference. Recent development of Bayesian phylogeographic inference can even combine spatial and temporal dynamics, and expand evolutionary hypothesis testing by including geographical priors [38]. While use of an evolutionary clock is rational for calculating evolutionary trends, the extremely variable accumulation of mutations observed within an defined influenza virus outbreak complicate a method based on average mutation rates [39]. Moreover, as discussed above, mutational robustness observed in short transmission chains dictates a different approach, where the least number of mutations explains the observed genetic diversity. Resulting maximum parsimony trees enabled visualization of hundreds of sequences and demonstrated effective in the transfer of knowledge on transmission patterns or patient clusters to a broad audience. Despite the analysis method used, there are limitations associated with the use of conventional consensus sequence data. Influenza virus infection of a single host is characterized by the complex reality of quasi species dynamics [40]. To designate the progressively and spatially changing mutant virus spectrum within an infected host as a single viral consensus sequence - as typically is done - comes with the expense of loss of detail. Therefore, implementation of quasi species characterization by deep (amplicon) NGS for studying virus transmission, while challenging, will provide the basis for highly credible transmission networks in the near future.

Enhancing the spatial power of international sequence data

In addition to local influenza virus outbreaks, influenza viruses cause annual epidemics and incidentally spark pandemics. To understand international spatial spread of influenza viruses, sequence data can be utilized to connect influenza virus infections through space and time, enabling the identification of sources of infection and viral migration patterns on continents or even across the globe [41,42]. Owing to the portability of sequence data, influenza virus sequences from one laboratory can be compared with others and can be shared in sequence databases, allowing analysis on international sequence datasets [43]. There is, however, a shortcoming of the

currently used nomenclature system for influenza viruses for monitoring spatial spread of viruses [44]. Although the system includes the geographic location of first virus isolation, geographical data on the suspected source of infection e.g. travel history is lacking. In addition, information on a known or suspected source of infection is systematically missing in the metadata of publicly available sequence databases, compromising our knowledge on the spatial and temporal evolutionary dynamics of influenza viruses by the use of geographical locations corresponding with the location of virus isolation. To our knowledge, the database of The European Surveillance System (TESSy) is the only influenza virus database that provides the opportunity to share information on the source of infection. In addition to the presence of a database field, clinicians should request and properly document a patient's travel history in a more systematical way, to support geographical linking of a virus sequence [45]. Given the global burden of infectious diseases like influenza, the addition of a patient's travel history to the reported metadata associated with the virus sequence is a minor effort that will enhance the robustness of international transmissions networks. Moreover, by the sharing of information on sources of infection, returning travelers can be used as sentinels for sustained transmission in regions that lack an influenza virus surveillance system and fill in international surveillance gaps.

Real-time sequencing in outbreak response

Epidemiological data lack resolution to reveal common sources and transmission dynamics during outbreaks when the causative agent lacks a distinctive phenotype e.g. LPAI outbreaks in poultry or nosocomial outbreaks of influenza virus during winter season [46]. As these are crucial data for effective control measures, the discriminatory power of influenza virus sequences could allow evidence based appropriate and effective interventions. The majority of the sequence based transmission networks of outbreaks have been generated retrospectively, but have revealed valuable insights in the spread of viruses as well as bacteria. Real-time sequence applications during outbreaks are yet sparse, although they have offered direct guidance to infection control teams [36,47-50].

The principal step, prior to any outbreak response, is to evaluate whether an observed increase in morbidity or mortality is associated with a related virus strain, deviating from circulating strains [51]. In addition to epidemiological data, patient cluster detection by sequence analysis provides information on the size of a patient cluster or outbreak by inclusion of related cases, as well as information on phenotypic markers associated with increased public health impact [36,52]. This use of sequence data necessitates detailed knowledge on between-host diversity, and significantly benefits from having a background sequence dataset to rule cases in or out belonging to the same cluster, transmission chain or outbreak [53,54].

Following cluster detection, sequence analysis can guide outbreak management. Successful isolation of sources and interruption of transmission chains based on sequence directed interventions have been accomplished for defined healthcare associated infections. As an example, Halachev *et al* prospectively linked unexplained nosocomial *A. baumannii* transmissions to environmental samples using sequence data. Initially a contaminated bed was identified and eventually a burns operating theater, leading to improved decontamination procedures that stopped the outbreak [49]. Near real-time sequencing of Ebola Viruses obtained from human cases during the West Africa outbreak that started early 2014, illustrated that the outbreak most probably was initiated by a single introduction into humans, followed by human-to-human transmission and subsequent geographical spread from Guinea to Sierra Leone [48]. Human-to-human transmission demonstrated to be clustered, stressing out the importance of rapid contact tracing [47].

As the interval between the beginning and the detection of an outbreak has direct impact on the magnitude of the outbreak and the response required to contain it, early recognition facilitates containment of infectious diseases [55]. Likewise, early recognition enhances the impact of real-time sequence based information on infection control measures, due to the feasibility of high-density sampling of transmission chains, allowing credible transmission network reconstruction.

Concluding remarks

In conclusion, the influenza virus genomic diversity arising during outbreaks and epidemics that may result in virus variants with increased public health risk can be utilized to gain insights in transmission chains and patient clusters using a transmission network. For transmission network reconstruction of influenza virus outbreaks, access to specimens obtained from densely sampled populations is the primary requirement, followed by a timely analysis of the sequence data and the availability of associated metadata. Given the host range variation of influenza viruses, this demands close collaboration between veterinarians, public health officials and clinicians, and will be beneficial for both human and animal health owing to reduced virus spread by targeted interventions. The explorations in using influenza sequence data to support outbreak control measures performed in this thesis demonstrate applications in both influenza virus outbreaks and epidemics. Technological advances will increase to power transmission networks in the nearby future and provide novel insights in the contribution of the quasi species population in the transmission dynamics. Investigation of samples from virus outbreaks is explanatory as well as exciting and should be continued as it will further our understanding on the control of influenza.

References

1. Ypma RJ, Jonges M, Bataille A, Stegeman A, Koch G, et al. (2013) Genetic data provide evidence for wind-mediated transmission of highly pathogenic avian influenza. J Infect Dis 207: 730-735.
2. Metzker ML (2010) Sequencing technologies - the next generation. Nat Rev Genet 11: 31-46.
3. Escobar-Gutierrez A, Vazquez-Pichardo M, Cruz-Rivera M, Rivera-Osorio P, Carpio-Pedroza JC, et al. (2012) Identification of hepatitis C virus transmission using a next-generation sequencing approach. J Clin Microbiol 50: 1461-1463.
4. Eshleman SH, Hudelson SE, Redd AD, Wang L, Debes R, et al. (2011) Analysis of genetic linkage of HIV from couples enrolled in the HIV Prevention Trials Network 052 trial. J Infect Dis 204: 1918-1926.
5. Bull RA, Eden JS, Luciani F, McElroy K, Rawlinson WD, et al. (2012) Contribution of intra- and interhost dynamics to norovirus evolution. J Virol 86: 3219-3229.
6. Jonges M, Bataille A, Enserink R, Meijer A, Fouchier RA, et al. (2011) Comparative analysis of avian influenza virus diversity in poultry and humans during a highly pathogenic avian influenza A (H7N7) virus outbreak. J Virol 85: 10598-10604.
7. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, et al. (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13: 341.
8. Hagemann IS, Devarakonda S, Lockwood CM, Spencer DH, Guebert K, et al. (2014) Clinical next-generation sequencing in patients with non-small cell lung cancer. Cancer.
9. Chen S, Li S, Xie W, Li X, Zhang C, et al. (2014) Performance comparison between rapid sequencing platforms for ultra-low coverage sequencing strategy. PLoS One 9: e92192.
10. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, et al. (2012) Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 30: 434-439.
11. Preston BD, Poiesz BJ, Loeb LA (1988) Fidelity of HIV-1 reverse transcriptase. Science 242: 1168-1171.
12. Roberts JD, Bebenek K, Kunkel TA (1988) The accuracy of reverse transcriptase from HIV-1. Science 242: 1171-1173.
13. Eckert KA, Kunkel TA (1991) DNA polymerase fidelity and the polymerase chain reaction. PCR Methods Appl 1: 17-24.
14. Liu SL, Rodrigo AG, Shankarappa R, Learn GH, Hsu L, et al. (1996) HIV quasispecies and resampling. Science 273: 415-416.

15. Kanagawa T (2003) Bias and artifacts in multitemplate polymerase chain reactions (PCR). J Biosci Bioeng 96: 317-323.
16. Gilles A, Meglecz E, Pech N, Ferreira S, Malausa T, et al. (2011) Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. BMC Genomics 12: 245.
17. Archer J, Baillie G, Watson SJ, Kellam P, Rambaut A, et al. (2012) Analysis of high-depth sequence data for studying viral diversity: a comparison of next generation sequencing platforms using Segminator II. BMC Bioinformatics 13: 47.
18. Reumers J, De Rijk P, Zhao H, Liekens A, Smeets D, et al. (2012) Optimized filtering reduces the error rate in detecting genomic variants by short-read sequencing. Nat Biotechnol 30: 61-68.
19. Archer J, Rambaut A, Taillon BE, Harrigan PR, Lewis M, et al. (2010) The evolutionary analysis of emerging low frequency HIV-1 CXCR4 using variants through time--an ultra-deep approach. PLoS Comput Biol 6: e1001022.
20. Jonges M, Welkers MR, Jeeninga RE, Meijer A, Schneeberger P, et al. (2014) Emergence of the virulence-associated PB2 E627K substitution in a fatal human case of highly pathogenic avian influenza virus A(H7N7) infection as determined by Illumina ultra-deep sequencing. J Virol 88: 1694-1702.
21. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, et al. (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22: 568-576.
22. Nekrutenko A, Taylor J (2012) Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. Nat Rev Genet 13: 667-672.
23. Ince DC, Hatton L, Graham-Cumming J (2012) The case for open computer programs. Nature 482: 485-488.
24. Cannata N, Merelli E, Altman RB (2005) Time to organize the bioinformatics resourceome. PLoS Comput Biol 1: e76.
25. Henry VJ, Bandrowski AE, Pepin AS, Gonzalez BJ, Desfeux A (2014) OMICtools: an informative directory for multi-omic data analysis. Database (Oxford) 2014.
26. Li JW, Schmieder R, Ward RM, Delenick J, Olivares EC, et al. (2012) SEQanswers: an open access community for collaboratively decoding genomes. Bioinformatics 28: 1272-1273.
27. Aarestrup FM, Brown EW, Detter C, Gerner-Smidt P, Gilmour MW, et al. (2012) Integrating genome-based informatics to modernize global disease monitoring, information sharing, and response. Emerg Infect Dis 18: e1.
28. Cassens I, Mardulyn P, Milinkovitch MC (2005) Evaluating intraspecific "network" construction methods using simulated sequence data: do existing algorithms outperform the global maximum parsimony approach? Syst Biol 54: 363-372.
29. Mardulyn P (2012) Trees and/or networks to display intraspecific DNA sequence variation? Mol Ecol 21: 3385-3390.

30. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572-1574.
31. Pouseele H, Vauterin P, Vauterin L (2011) A resampling strategy for reliable network construction. Mol Phylogenet Evol 60: 273-286.
32. Steel M, Penny D (2000) Parsimony, likelihood, and the role of models in molecular phylogenetics. Mol Biol Evol 17: 839-850.
33. Xia X (2007) Molecular Phylogenetics: Mathematical Framework and Unsolved Problems. In: Bastolla U, Porto M, Roman HE, Vendruscolo M, editors. Structural Approaches to Sequence Evolution: Springer Berlin Heidelberg. pp. 169-189.
34. Weng JF, Thomas DA, Mareels I (2011) Maximum parsimony, substitution model, and probability phylogenetic trees. J Comput Biol 18: 67-80.
35. Worby CJ, Chang HH, Hanage WP, Lipsitch M (2014) The distribution of pairwise genetic distances: a tool for investigating disease transmission. Genetics 198: 1395-1404.
36. Jonges M, Meijer A, Fouchier R, Koch G, Li J, et al. (2013) Guiding outbreak management by the use of influenza A(H7Nx) virus sequence analysis. Euro Surveill 18.
37. Jonges M, Rahamat-Langendoen J, Meijer A, Niesters HG, Koopmans M (2012) Sequence-based identification and characterization of nosocomial influenza A(H1N1)pdm09 virus infections. J Hosp Infect 82: 187-193.
38. Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian phylogeography finds its roots. PLoS Comput Biol 5: e1000520.
39. Kolaczkowski B, Thornton JW (2004) Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature 431: 980-984.
40. Lauring AS, Andino R (2010) Quasispecies theory and the behavior of RNA viruses. PLoS Pathog 6: e1001005.
41. Lemey P, Suchard M, Rambaut A (2009) Reconstructing the initial global spread of a human influenza pandemic: A Bayesian spatial-temporal model for the global spread of H1N1pdm. PLoS Curr 1: RRN1031.
42. Gog JR, Ballesteros S, Viboud C, Simonsen L, Bjornstad ON, et al. (2014) Spatial Transmission of 2009 Pandemic Influenza in the US. PLoS Comput Biol 10: e1003635.
43. Sharma D, Priyadarshini P, Vrati S (2015) Unraveling the web of viroinformatics: computational tools and databases in virus research. J Virol 89: 1489-1501.
44. (1980) A revision of the system of nomenclature for influenza viruses: a WHO memorandum. Bull World Health Organ 58: 585-591.
45. Cleton N, Reusken C, Murk JL, de Jong M, Reimerink J, et al. (2014) Using routine diagnostic data as a method of surveillance of arboviral infection in travellers: a

comparative analysis with a focus on dengue. Travel Med Infect Dis 12: 159-166.

46. Robinson ER, Walker TM, Pallen MJ (2013) Genomics and outbreak investigation: from sequence to consequence. Genome Med 5: 36.
47. Scarpino SV, Iamarino A, Wells C, Yamin D, Ndeffo-Mbah M, et al. (2015) Epidemiological and viral genomic sequence analysis of the 2014 ebola outbreak reveals clustered transmission. Clin Infect Dis 60: 1079-1082.
48. Gire SK, Goba A, Andersen KG, Sealfon RS, Park DJ, et al. (2014) Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science 345: 1369-1372.
49. Halachev MR, Chan JZ, Constantinidou CI, Cumley N, Bradley C, et al. (2014) Genomic epidemiology of a protracted hospital outbreak caused by multidrug-resistant Acinetobacter baumannii in Birmingham, England. Genome Med 6: 70.
50. Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, et al. (2011) Prospective genomic characterization of the German enterohemorrhagic Escherichia coli O104:H4 outbreak by rapid next generation sequencing technology. PLoS One 6: e22751.
51. Tang P, Gardy JL (2014) Stopping outbreaks with real-time genomic epidemiology. Genome Med 6: 104.
52. Price JR, Golubchik T, Cole K, Wilson DJ, Crook DW, et al. (2014) Whole-genome sequencing shows that patient-to-patient transmission rarely accounts for acquisition of Staphylococcus aureus in an intensive care unit. Clin Infect Dis 58: 609-618.
53. Qi X, Qian YH, Bao CJ, Guo XL, Cui LB, et al. (2013) Probable person to person transmission of novel avian influenza A (H7N9) virus in Eastern China, 2013: epidemiological investigation. BMJ 347: f4752.
54. Jonges M, Meijer A, Koopmans MPG (2013) Letter to the Editor; Response to: "Probable person to person transmission of novel avian influenza A (H7N9) virus in Eastern China, 2013: epidemiological investigation". BMJ: BMJ.
55. Hitchcock P, Chamberlain A, Van Wagoner M, Inglesby TV, O'Toole T (2007) Challenges to global surveillance and response to infectious disease outbreaks of international importance. Biosecur Bioterror 5: 206-227.

Chapter 7

Summary in Dutch /

Nederlandse samenvatting

Influenza is een infectieziekte veroorzaakt door type A-, B- of C- influenzavirussen. Dit zijn RNA virussen geclassificeerd als leden van de *Orthomyxovirus* familie. Influenza A virussen kunnen verder worden onderverdeeld in subtypen op basis van oppervlakte eiwitten hemagglutinine (HA) en neuraminidase (NA). Tot op heden zijn er 18 HA (H1-H18) en 11 NA (N1-N11) eiwitten bekend. Alle influenza A virus subtypen, behalve de recentelijk ontdekte vleermuis influenza A virus subtypen H17N10 en H18N11, circuleren in hun natuurlijke reservoir van watervogels. Tegelijk rondgaande influenza A virussen van subtypen H1N1 en H3N2, en influenza B virussen, zijn verantwoordelijk voor de jaarlijks terugkerende griepepidemie waarbij 7-20% van de wereldbevolking geïnfecteerd raakt. De meeste influenzavirus geïnfecteerde personen herstellen binnen twee weken, maar infecties kunnen ook tot complicaties leiden waaraan patiënten kunnen overlijden.

De introductie van een nieuw niet-humaan influenzavirus in de humane populatie, waartegen geen of beperkte immuniteit aanwezig is, kan uitmonden in een pandemie. Dit is alleen mogelijk wanneer een dergelijk virus zich aanpast aan effectieve infectie van mensen en makkelijk overdraagbaar wordt tussen mensen. Influenzavirussen zijn in staat zich aan te passen aan een nieuwe of veranderende omgeving door uitwisseling van gen-segmenten afkomstig van verschillende influenzavirussen en/of door mutaties. Doordat influenzavirussen van nature muteren om te kunnen overleven, is er een continu gevaar van een pandemie doordat niet-humane influenzavirussen zich zouden kunnen aanpassen aan de mens.

In vergelijking met zoogdieren, muteren RNA virussen ongeveer 500.000 keer sneller. Influenzavirus evolutie vertoont gemiddelde mutatiesnelheden van 2 tot 8 × 10^{-3} substituties per nucleotide per jaar, terwijl bij uitbraken van aviaire influenza in pluimvee zelfs mutatiesnelheden boven de 1 × 10^{-2} substituties per nucleotide per jaar zijn waargenomen. Gegeven deze mutatiesnelheden, zou het in kaart brengen van (accumulerende) mutaties in het RNA van influenzavirussen informatief kunnen zijn om virusoverdracht te identificeren. Binnen dit promotieonderzoek is geëvalueerd in hoeverre influenzavirus RNA diversiteit, bepaalt door middel van sequentie analyse, benut kan worden bij de bestrijding van uitbraken. De mogelijkheid is onderzocht om zowel lokale influenzavirus clusters als overdrachtsketens te identificeren, opdat kennis verkregen wordt ten behoeve van volksgezondheid besluitvorming bij uitbraken.

In hoofdstuk 2 is de mogelijkheid onderzocht om lokale geclusterde influenzavirus infecties op te sporen met behulp van een achtergrond dataset verkregen uit moleculaire surveillance. Hiervoor is beoordeeld of de totale influenzavirus sequentiediversiteit gedetecteerd tijdens nationale influenzavirus surveillance gebruikt

kan worden om de exclusiviteit van lokale geclusterde infecties aan te tonen. Met slechts een beperkt aantal nationale (n=17), regionale (n=3) en ziekenhuis (n=5) influenza A(H1N1) virus sequenties, is in hoofdstuk 2.1 een epidemiologisch verdacht nosocomiaal cluster van oseltamivir-resistente A(H1N1) virusinfecties onderbouwd. Het verantwoordelijke virus bleek gekarakteriseerd door een unieke mutatie in het NA enzym, welke tevens afwezig bleek in beschikbare internationale A(H1N1) sequentiedata (GenBank: maart 2015). Ten gevolge van de influenza A(H1N1)pdm09 pandemie in 2009, is tijdens de moleculaire surveillance van het 2009/2010 influenza seizoen een A(H1N1)pdm09 sequentie dataset gegenereerd bestaande uit sequenties van 685 circulerende virussen. Daarmee is kennis over nationale influenzavirus sequentiediversiteit verkregen. Vergelijking van 107 A(H1N1)pdm09 virus sequenties afkomstig van één ziekenhuis met sequenties van circulerende virussen, leverde acht patiënt clusters op welke gegeven de nationale virus sequentiediversiteit een lage kans op voorkomen hadden ($P < 0.01$) (Hoofdstuk 2.2). Epidemiologische analyse toonde in vier van de acht clusters aanwezigheid van nosocomiale infecties aan, plus een moeder-kind combinatie in een vijfde cluster. Dit impliceert dat clusters van ziekenhuis influenzavirus sequenties welke afwijken van circulerende virus sequenties nader onderzocht dienen te worden. Daarentegen bleken clusters (n=9) van ziekenhuis influenzavirus infecties welke gegeven de nationale virus sequentiediversiteit een hogere kans op voorkomen hadden ($P > 0.01$), allen geassocieerd met in de gemeenschap opgelopen virusinfecties. De resultaten van Hoofdstukken 2.1 en 2.2 suggereren dat overdrachtsketens binnen een influenza epidemie leiden tot een enorme variatie bijna-identieke influenzavirussen, waarvan de meerderheid uiteindelijk uitsterft door doodlopende overdrachten.

In Hoofdstuk 3 is de mogelijkheid onderzocht om influenzavirus overdracht te identificeren met behulp van virus sequentiedata van een influenzavirus uitbraak. Een gecombineerde dataset van humane en veterinaire influenzavirus sequenties van een influenza A(H7N7) virusuitbraak in pluimvee is gebruikt om te onderzoeken of influenzavirus mutaties met een verhoogd risico voor de volksgezondheid ontstonden tijdens infecties van personen, of al aanwezig waren in de virussen afkomstig uit pluimvee (Hoofdstuk 3.1). De virussen van humane A(H7N7) conjunctivitis gevallen weerspiegelden de virusdiversiteit in pluimvee en bleken doodlopende overdrachten. Het virusoverdracht netwerk toonde aan dat een toename en verspreiding van virulentie en humane adaptatie markers plaatsvond in pluimvee, alvorens overdracht naar personen plaatsvond. Nadat er met een gecombineerde analyse van A(H7N7) virus sequentiedata en meteorologische data statistisch bewijs geleverd werd dat de richting van boerderij-naar-boerderij verspreiding correleerde met de windrichting op de dag van infectie, zijn veldstudies uitgevoerd om de mogelijke rol van de wind bij de verspreiding van aviair influenzavirus te onderzoeken tijdens uitbraken in pluimvee

(Hoofdstuk 3.2). Hiervoor is fijnstof in de lucht, in de stal en op benedenwindse locaties van stallen met influenzavirus positief pluimvee verzameld. Naast een gemeten influenzavirus concentratie van $8.48x10^4$ genoom kopieën/m^3 binnenin een stal, bleken virussen met behulp van een reverse transcriptase PCR tot 60 meter benedenwinds van stallen te detecteren. Doordat een goede correlatie tussen veldmetingen en gemodelleerde concentraties werd waargenomen, zou een geografische raming van gebieden met verhoogd risico op blootstelling aan aviaire influenza gemodelleerd kunnen worden tijdens een uitbraak.

In Hoofdstuk 4 is het onderscheidend vermogen van nieuwe generatie sequencers (NGS) voor de robuuste karakterisering van binnen-gastheer influenzavirus sequentiediversiteit geëvalueerd. Dit onderscheidend vermogen is vereist voor de vroege signalering van mutaties met een verhoogd risico voor de volksgezondheid. Doordat het karakteriseren van virale minderheidsvarianten afhankelijk is van het vermogen om biologische variatie te onderscheiden van artefacten, is het effect van kwaliteit controles op zowel technische als reverse transcriptase PCR geïnduceerde artefacten in NGS datasets bestudeerd (Hoofdstuk 4.1). De identificatie en eliminatie van twee verschillende soorten artefacten in NGS datasets resulteerden in een 95% reductie van artificiële variatie. De geoptimaliseerde NGS data analyse is vervolgens toegepast op NGS datasets van influenza A(H7N7) virus positieve materialen afkomstig van pluimvee en van een fatale humane casus, om het ontstaan van virulentie marker PB2 E627K te onderzoeken (Hoofdstuk 4.2). Deze toepassing suggereert dat marker PB2 E627K ontstond tijdens infectie van een enkele humane gastheer en benadrukt het belang van het reduceren van humane blootstelling aan aviaire influenzavirussen om daarmee de kans op virusaanpassing aan mensen te verminderen.

In Hoofdstuk 5 is onderzocht of de verkregen kennis over het gebruik van influenzavirus sequentiediversiteit binnen uitbraken en epidemieën prospectief toegepast kan worden om het volksgezondheidrisico van een uitbraak in te schatten. Vanaf 31 maart 2013 rapporteerden Chinese volksgezondheidautoriteiten humane infecties met influenza A(H7N9) en deelden direct hun virologische gegevens. Een onmiddellijke vergelijking van A(H7N9) sequentiedata met sequenties van de A(H7N7) virusuitbraak in Nederland in 2003 en een A(H7N1) virusuitbraak in Italië in 1999/2000 suggereerde dat wijdverspreide circulatie van A(H7N9) virussen in China had plaatsgevonden, gebaseerd op de geconstateerde diversificatie. Deze diversificatie was verontrustend, omdat reeds enkele humane adaptatie markers in het A(H7N9) virus aanwezig waren.

Tenslotte zijn in Hoofdstuk 6 de belangrijkste bevindingen van dit proefschrift samengevat en is het gebruik van influenzavirus sequentiediversiteit bij de bestrijding van uitbraken bediscussieerd in relatie tot de huidige inzichten.

Samengevat verschaft dit proefschrift een overzicht van mijn onderzoek naar het gebruik van influenzavirus sequentiediversiteit. Het stelt ons in staat om zowel lokale geclusterde influenzavirus infecties als overdrachtsketens te identificeren, opdat kennis verkregen wordt ten behoeve van volksgezondheid besluitvorming bij uitbraken. De verkregen kennis zou toekomstige interventies kunnen sturen, met verminderde virus verspreiding als gevolg. In de nabije toekomst zal de technologische vooruitgang zowel de resolutie als de betrouwbaarheid van virusoverdracht netwerken verbeteren, en nieuwe inzichten opleveren over de bijdrage van binnen-gastheer virusdiversiteit aan de dynamiek van virus overdracht.

Chapter 8

Word of thanks /

Dankwoord

Ik heb de afgelopen jaren mogen genieten van de dynamische en zeer interessante wereld van RNA virussen, in het bijzonder influenza virussen. De onvoorspelbaar muterende virussen gecombineerd met evoluerende moleculaire technieken, creëren een onderzoekswereld waar ik nog lang niet op uitgekeken ben. Naast de bijdrage van virussen en technieken, zou dit proefschrift niet mogelijk geweest zijn zonder velen.

In de eerste plaats wil ik mijn promotor Marion bedanken. Je hebt een kostbare combinatie van kwaliteiten, waarbij jouw blik op het grote geheel voor mij erg waardevol is gebleken. De implicaties van experimentele data, het bewaken van de rode draad van mijn onderzoek, en het “Public Health” denken zijn enkele voorbeelden. Tevens wist je zonder te souffleren, zowel experimenten als publicaties te sturen in richtingen waarbij de weg naar het resultaat te allen tijde uitdagend was. Naast je inhoudelijke bijdragen hebben we, vooral buiten het RIVM, bijzondere en leuke momenten beleefd tijdens o.a. ons eerste bezoek aan Guangdong CDC, congressen en de feesten bij je thuis. Dank voor het creëren van kansen, de nuttige discussies en adviezen, die mij veel geleerd hebben en het werk zeer ten goede zijn gekomen.

Uiteraard wil ik ook dank uitspreken aan mijn copromotor Adam, waarmee ik veelvuldig contact heb gehad. Je bent een ontzettend prettige collega met ongelooflijk veel kennis; de “influenza vraagbaak”. In het bijzonder het diepgaan over mogelijke fundamentele oorzaken van onverwachte resultaten, is waar onze passies volledig overlappen. Je bent nauwkeurig, geduldig, en altijd bereid adviezen te geven. Als directe begeleider heb je de nodige avonturen met me meegemaakt, van de euforie na geslaagde experimenten tot totale stress voor mijn presentatie tijdens de Options VI Meeting in Hong Kong. In ieder geval smaakte het eten na afloop goed… Dank voor de interessante overleggen, je vertrouwen en betrokkenheid.

Ook wil ik de leden van de leescommissie danken: Prof.dr. Ron Fouchier, Prof.dr. Ab Osterhaus en Prof.dr. Jan Hendrik Richardus, dank voor de tijd die jullie gestoken hebben in het kritisch beoordelen van dit proefschrift.

Gedurende mijn promotieonderzoek ben ik in contact gekomen met veel niet-RIVMers waaruit mooie samenwerkingen zijn ontstaan. Enkelen wil ik persoonlijk bedanken. Jairo: Wat hebben wij over het ontstaan van influenza virus antivirale resistentie veel contact gehad! Je bent erg gedreven en geïnteresseerd, en onze samenwerking heeft mooie resultaten opgeleverd. Succes met jouw afronding binnenkort! Janette: Uren hebben we gepuzzeld, met zowel de moleculaire als epidemiologische data om alle mogelijke transmissies boven tafel te krijgen. Dank voor je geduldige speurwerk. Inge: Dank voor het delen van jouw ervaringen met luchtmetingen, waarbij uiteindelijk

succesvolle virusdetecties zijn verricht. Zonder jouw kennis en de inhoudelijke discussies was dit nooit gelukt. Matthijs: Het was een bijzonder leuke en intensieve samenwerking waarbij we elkaar goed aanvulden, vernieuwende ideeën hadden, wat heeft geleid tot mooie publicaties en de basis heeft gelegd voor toekomstige re-sequencing NGS analyses. Dank je, en succes met de laatste loodjes van jouw proefschrift.

Binnen het RIVM heb ik gedurende verschillende fases van mijn onderzoek mogen samenwerken met meerdere afdelingen. Maaike, Evy en Dieneke; dank voor de bereidheid mijn ad hoc sequentie analyses in te zetten. Rozemarijn, Ingmar, Arnout en Marjolijn; dank voor alle support, adviezen en nuttige discussies welke geresulteerd hebben in geslaagde luchtmetingen bij pluimveebedrijven. Last but not least: Paul, Anouk, Imke, Leslie, Madelief, Marleen, Corien, Andre en Aura; het was een bijzonder intensieve maar geslaagde samenwerking met jullie allen n.a.v. het contactonderzoek.

Binnen de afdeling Virologie gaat in eerste instantie mijn dank uit naar de Moleculaire Pool, waarmee ik de laatste jaren nauw heb samengewerkt om de moleculaire diagnostiek van de afdeling efficiënt, betrouwbaar en (grotendeels) geautomatiseerd te krijgen. Jeroen, Anne-Marie, Bas, Sharon, Naomi en Lisa; dank voor al jullie bloed, zweet en tranen. De Moleculaire Pool werkt inmiddels als een geoliede machine en is een schoolvoorbeeld van samenwerking voor heel IDS. Daar mogen we met z'n allen trots op zijn!

Daarnaast wil ik mijn collega-AIO's bedanken met wie ik een kamer heb mogen delen, en die daarmee eerstelijns sparringpartner waren maar ook voor een hoop vertier hebben gezorgd. Joukje, Sanela, Sabine. Van vrijdagmiddag-experimenten op de gang tot zeilwedstrijden welke tot frustratie van de organisatie keer op keer uitliepen op excessen. Dank voor de gezelligheid, zowel binnen als buiten het instituut! Era, Faizel, Sabine, Dominique, Giovanna, Rita, Janko, Gudi en Dennis. Speciale dank dat jullie mijn klankboard waren, de gezellige vrijdagmiddag borrels, jullie verhalen en inspiratie.

Van tijd tot tijd moet mijn brein worden gereset of worden opgeladen. De verbouwing van ons huis was een lekkere uitlaatklep. Met dank aan de loodgieterslessen van Maurice heeft dit veel woongenot opgeleverd. Structurele uitlaatklep blijft de afreageer-zondag waarbij ik met Heren Veteranen C onze tegenstander wekelijks te lijf mag gaan. Daarnaast heeft mijn eerste autootje, een Volvo Amazon uit 1968, voor erg veel oplaadmomenten gezorgd. Zowel erin rijdend als eronder liggend en aan de klus, was een genot. Tijdens klussen bij mij voor de deur vaak een sociale happening, met alle nieuwsgierige huisvaders uit de buurt. Daarbij biedt onze straat meer dan alleen een woning; dank voor de gezellige avonden en BBQ's.

Papa en mama, dank voor het faciliteren van kansen en mogelijkheden, sturen van ideeën, inzichten en gedachtegoed. De onvergetelijke vakanties, de vrijheid; het zorgeloze bestaan. Ontpopt als opa en oma, waarbij kleinkinderen ervoor zorgen dat jullie oude tijden herbeleven en ik nu bewust meemaak hoe vrolijk en zorgzaam mama voor mij is geweest. Dank voor jullie steun al deze jaren. Steven, bepakt aan memorabele hoogtepunten die we tijdens onze vakanties samen beleefden zijn we beiden uitgevlogen. Jij letterlijk. Tof dat je met mij achterop de motor door Zwitserland durfde en met Lieveke op de achterbank van mijn Volvo wilde trouwen. Happy flying broertje!

Ben en Sonja, uiteraard ook dank voor jullie. Of het nou met een fles Soberano in de avondzon aan de Middellandse Zee, of op vrijdagavonden in jullie achtertuin was; er is altijd tijd voor gezelligheid, goede gesprekken en steun. Het is maar goed dat jullie deur altijd voor mij opstaat, anders zal ik samen met Jan weer moeten inbreken…

Lieve Kim, samen kunnen we de wereld aan. Gisteren, vandaag en morgen. Bijzonder hoe wij telkens weer in gezamenlijke avonturen of “missies” belanden, elkaar kansen en ruimte geven, en elkaar blijven uitdagen. Dank voor je steun, het ontzien en het afzien, dat dit proefschrift heeft mogelijk gemaakt. Ik hou van jou! Milou, wat word je snel groot en wijs. Lucas, kleine charmeur! Allebei dank voor de vrolijke afleiding tijdens het schrijven.

Chapter 9

About the author

Curriculum vitae

Marcel Jonges was born on January 14, 1979, in Voorburg, the Netherlands. After having graduated at the Sint-Odulphus Lyceum in Tilburg, he started his study in Molecular Science at Wageningen University (the Netherlands) in 1998, and continued his studies in Biotechnology. Marcel performed an internship at the department of Viroscience at ErasmusMC (Rotterdam, the Netherlands) where he studied the impact of an amino acid substitution in a T-helper cell epitope on the immune response to influenza A viruses. Next, he performed an internship at the Division of Virology at the National Institute for Public Health and the Environment (RIVM, Bilthoven, the Netherlands) where he studied the susceptibility of the influenza viruses for oseltamivir by fluorometric analysis of neuraminidase glycoprotein activity. After finishing his Master of Science degree in June 2006, Marcel worked as a young scientist at RIVM where he incorporated antiviral susceptibility testing in the national influenza virus surveillance program. In July 2007, he started as PhD candidate on the use of influenza virus sequence diversity to identify and characterize transmission events. He was employed at the department of Viroscience at ErasmusMC and was hosted fulltime at the Division of Virology at RIVM under supervision of Prof. dr. M.P.G. Koopmans. Guided by dr. A. Meijer he developed influenza virus sequencing protocols, initially focusing on influenza virus mutations associated with phenotypic traits. Soon all virus sequence diversity was taken into account, providing sufficient resolution to study influenza virus transmission. From 2011 to 2015, Marcel was Project Manager Roche FLOW Implementation at the Division of Virology (RIVM). As of March 1, 2012, he is Coordinator Molecular Diagnostics at the Division of Virology (RIVM) where Marcel is responsible for development and validation of real-time PCR reagents and protocols that allow efficient routine diagnosis of both endemic as well as emerging virus infections. The results obtained at RIVM, and during collaborations with LUMC (Leiden, the Netherlands), UMCG (Groningen, the Netherlands), AMC (Amsterdam, the Netherlands), UU (Utrecht, the Netherlands) and CVI (Lelystad, the Netherlands) are presented in this thesis.

PhD portfolio

Name: Marcel Jonges
Research group: Department of Viroscience
Research school: Post-Graduate Molecular Medicine
PhD period: 2007-2015
Promotor: Prof.dr. M.P.G. Koopmans
Co-promotor: dr. A. Meijer

Experience

2012-present	Coordinator Molecular Diagnostics at the Division of Virology, Centre for Infectious Disease Control, RIVM, Bilthoven, the Netherlands
2011-2015	Project Manager Roche FLOW Implementation at the Division of Virology, Centre for Infectious Disease Control, RIVM, Bilthoven, the Netherlands

Education

2007-2012	PhD position at the department of Viroscience, ErasmusMC, Rotterdam, the Netherlands

In-Depth Courses

2014	Techniques in Clinical Virology, Medical University Vienna, Austria (ESCV)
2012	Next Generation Sequencing (NGS) Data analysis, ErasmusMC (Molmed)
2011	Succesvol gebruik van moleculaire diagnostiek, Boerhaave, Leiden
2010	May 31 - June 4 2010, Course in Virology 2010, ErasmusMC (Molmed)
2008	Workshop on Phylogeny and Genetics in microbiology and virology, Erasmus MC (Molmed)
2008	Course Molecular Diagnostics III, Erasmus MC (Molmed)
2007	Workshop BioNumerics & GelCompar, Sint-Martens-Latem, Belgium (Applied-Maths)
2007	antiviral Workshop, HPA Colindale, London, UK (VIRGIL)
2006	antiviral Course, HPA Colindale/MRC, London, UK (VIRGIL)
2006-present	Internal and external presentations at the Division of Virology (RIVM)

Scientific presentations

2015	3rd International One Health Congress, Amsterdam, the Netherlands (poster)

2013	XV International Symposium on Respiratory Viral Infections, Rotterdam, the Netherlands (presentation)
2012	ESCV Madrid, Spain (presentation and poster)
2011	IMED, Vienna, Austria (presentation)
2011	Spring Meeting NVMM, Papendal, the Netherlands (presentation)
2010	Dutch Annual Virology Symposium, Amsterdam, the Netherlands (presentation)
2010	4th European Congress of Virology, Cernobbio, Italy (poster)
2010	Options for the Control of Influenza VII, Hong Kong, China (presentation)
2008	X International Symposium on Respiratory Viral Infections, Sentosa Island, Singapore (poster)
2007	Third European Congress of Virology, Nürnberg, German (poster)
2006	Spring Meeting NVMM, Papendal, the Netherlands (presentation)

Teaching

2014	WHO training course, Rotterdam, the Netherlands
2013	Metagenomics in Virology meeting, Amsterdam, the Netherlands
2013	Summer / Winter course, Rotterdam, the Netherlands
2012	Summer / Winter course, Rotterdam, the Netherlands
2010	Technical Officer of the WHO was trained in executing an influenza antiviral resistance test (NAI-assay)
2010	Supervising BSc graduate student
2009	Supervising MSc Biology graduate student
2008	Technicians of Guangdong-CDC and Dongguan-CDC (China) were trained in executing an influenza antiviral resistance test (NAI-assay)

Miscellaneous

2008-2011	Board Member of Promovendus Netwerk RIVM (Proneri): organization of workshops, lectures, meetings and a retreat for PhD students
2007-present	Reviewer for Antiviral Res., Euro Surveill., Emerg Infect Dis., Infect Genet Evol., J Clin Virol., J Virol Methods, Lancet, PLOS ONE, Sci. Rep., Vaccine, Virol. J.

List of publications

2015

Jonges M, van Leuken J, Wouters I, Koch G, Meijer A, Koopmans M. Wind-Mediated Spread of Low-Pathogenic Avian Influenza Virus into the Environment during Outbreaks at Commercial Poultry Farms. PLoS One. 2015 May 6;10(5)

Welkers MR, Jonges M, Jeeninga RE, Koopmans MP, de Jong MD, Improved detection of artifactual viral minority variants in high-throughput sequencing data. Front Microbiol. 2015 Jan 22;5:804

Mollers M, Jonges M, Pas SD, van der Eijk AA, Dirksen K, Jansen C, et al. Follow-up of contacts of Middle East respiratory syndrome coronavirus–infected returning travelers, the Netherlands, 2014. Emerg Infect Dis. 2015 Sept

2014

Jonges M, Welkers MR, Jeeninga RE, Meijer A, Schneeberger P, Fouchier RA, de Jong MD, Koopmans M, Emergence of the Virulence-Associated PB2 E627K Substitution in a Fatal Human Case of Highly Pathogenic Avian Influenza Virus A(H7N7) Infection as Determined by Illumina Ultra-Deep Sequencing, J Virol. 2014 Feb;88(3):1694-702

Van Borm S, Jonges M, Lambrecht B, Koch G, Houdart P, van den Berg T. Molecular epidemiological analysis of the transboundary transmission of 2003 highly pathogenic avian influenza H7N7 outbreaks between the Netherlands and Belgium, Transbound Emerg Dis. 2014 Feb;61(1):86-90.

Reusken CB, Farag EA, Jonges M, Godeke GJ, El-Sayed AM, Pas SD, Raj VS, Mohran KA, Moussa HA, Ghobashy H, Alhajri F, Ibrahim AK, Bosch BJ, Pasha SK, Al-Romaihi HE, Al-Thani M, Al-Marri SA, AlHajri MM, Haagmans BL, Koopmans MP. Middle East respiratory syndrome coronavirus (MERS-CoV) RNA and neutralising antibodies in milk collected according to local customs from dromedary camels, Qatar, April 2014. Euro Surveill. 2014 Jun 12;19(23).

Haagmans BL, Al Dhahiry SH, Reusken CB, Raj VS, Galiano M, Myers R, Godeke GJ, Jonges M, Farag E, Diab A, Ghobashy H, Alhajri F, Al-Thani M, Al-Marri SA, Al Romaihi HE, Al Khal A, Bermingham A, Osterhaus AD, Alhajri MM, Koopmans MP. Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation. Lancet Infect Dis. 2014 Feb;14(2):140-5

Kraaij-Dirkzwager M, Timen A, Dirksen K, Gelinck L, Leyten E, Groeneveld P, Jansen C, Jonges M, Raj S, Thurkow I, van Gageldonk-Lafeber R, van der Eijk A, Koopmans M; MERS-CoV outbreak investigation team of the Netherlands. Middle East respiratory syndrome coronavirus (MERS-CoV) infections in two returning travellers in the Netherlands, May 2014. Euro Surveill. 2014 May 29;19(21)

Fanoy EB, van der Sande MA, Kraaij-Dirkzwager M, Dirksen K, Jonges M, van der Hoek W, Koopmans MP, van der Werf D, Sonder G, van der Weijden C, van der Heuvel J, Gelinck L, Bouwhuis JW, van Gageldonk-Lafeber AB. Travel-related MERS-CoV cases: an assessment of exposures and risk factors in a group of Dutch travellers returning from the Kingdom of Saudi Arabia, May 2014. Emerg Themes Epidemiol. 2014 Oct 17;11:16

2013

Jonges M, Meijer A, Fouchier RA, Koch G, Li J, Pan JC, Chen H, Shu YL, Koopmans MP, Guiding outbreak management by the use of influenza A(H7Nx) virus sequence analysis, Euro Surveill. 2013 Apr 18;18(16):20460

Ypma RJ, Jonges M, Bataille A, Stegeman A, Koch G, van Boven M, Koopmans M, van Ballegooijen WM, Wallinga J. Genetic data provide evidence for wind-mediated transmission of highly pathogenic avian influenza, J Infect Dis. 2013 Mar;207(5):730-5

2012

Jonges M, J. Rahamat-Langendoen, A. Meijer, H. Niesters, M. Koopmans, Sequence-based identification and characterization of nosocomial influenza A(H1N1)pdm09 virus infections, J Hosp Infect. 2012 Nov;82(3):187-93.

Meijer A, Jonges M, van Beek P, Swaan C, Osterhaus A, Daniels R, Hurt A, Koopmans M. Oseltamivir-resistant influenza A(H1N1)pdm09 virus in Dutch travellers returning from Spain, August 2012. Euro Surveill. 2012 Sep 6;17(36).

Liu WM, Nahar TE, Jacobi RH, Gijzen K, van Beek J, Hak E, Jonges M, Boog CJ, van der Zeijst BA, Soethout EC. Impaired production of TNF-α by dendritic cells of older adults leads to a lower CD8+ T cell response against influenza. Vaccine. 2012 Feb 21;30(9):1659-66

2011

Jonges M, Bataille A, Enserink R, Meijer A, Fouchier RA, Stegeman A, Koch G, Koopmans M, Comparative analysis of avian influenza virus diversity in poultry and humans during a highly pathogenic avian influenza A (H7N7) virus outbreak, J Virol. 2011 Oct;85(20):10598-604.

Meijer A, Jonges M, Abbink F, Ang W, van Beek J, Beersma M, Bloembergen P, Boucher C, Claas E, Donker G, van Gageldonk-Lafeber R, Isken L, de Jong A, Kroes A, Leenders S, van der Lubben M, Mascini E, Niesters B, Oosterheert JJ, Osterhaus A, Riesmeijer R, Riezebos-Brilman A, Schutten M, Sebens F, Stelma F, Swaan C, Timen A, van 't Veen A, van der Vries E, te Wierik M, Koopmans M, Oseltamivir-resistant pandemic A(H1N1) 2009 influenza viruses detected through enhanced surveillance in the Netherlands, 2009-2010, Antiviral Res. 2011 Oct;92(1):81-9.

Dijkstra F, Jonges M, van Beek R, Donker GA, Schellevis FG, Koopmans M, van der Sande MA, Osterhaus AD, Boucher CA, Rimmelzwaan GF, Meijer A. Influenza A(H1N1) Oseltamivir Resistant Viruses in the Netherlands During the Winter 2007/2008. Open Virol J. 2011;5:154-62

Zhou J, Zou L, Zhang X, Liao J, Ni H, Hou N, Wang Y, Li H, Wu J, Jonges M, Meijer A, Koopmans M, Ke C., Adamantane- and Oseltamivir-Resistant Seasonal A(H1N1) and Pandemic A(H1N1) 2009 Influenza Viruses in Guangdong, China, During 2008 and 2009., J Clin Microbiol. 2011 May 18

2010

Jonges M, Liu WM, van der Vries E, Jacobi R, Pronk I, Boog C, Koopmans M, Meijer A, Soethout E. Influenza virus inactivation for studies of antigenicity and phenotypic neuraminidase inhibitor resistance profiling. J Clin Microbiol. 2010 Mar;48(3):928-40

van der Vries E, Jonges M, Herfst S, Maaskant J, Van der Linden A, Guldemeester J, Aron GI, Bestebroer TM, Koopmans M, Meijer A, Fouchier RA, Osterhaus AD, Boucher CA, Schutten M. Evaluation of a rapid molecular algorithm for detection of pandemic influenza A (H1N1) 2009 virus and screening for a key oseltamivir resistance (H275Y) substitution in neuraminidase, J Clin Virol. 2010 Jan;47(1):34-7

Herfst S, Chutinimitkul S, Ye J, de Wit E, Munster VJ, Schrauwen EJ, Bestebroer TM, Jonges M, Meijer A, Koopmans M, Rimmelzwaan GF, Osterhaus AD, Perez DR, Fouchier RA,. Introduction of virulence markers in PB2 of pandemic swine-origin influenza virus does not result in enhanced virulence or transmission, J Virol. 2010 Apr;84(8):3752-8

Chutinimitkul S, Herfst S, Steel J, Lowen AC, Ye J, van Riel D, Schrauwen EJ, Bestebroer TM, Koel B, Burke DF, Sutherland-Cash KH, Whittleston CS, Russell CA, Wales DJ, Smith DJ, Jonges M, Meijer A, Koopmans M, Rimmelzwaan GF, Kuiken T, Osterhaus AD, García-Sastre A, Perez DR, Fouchier RA., Virulence-associated substitution D222G in

the hemagglutinin of 2009 pandemic influenza A(H1N1) virus affects receptor binding, J Virol. 2010 Nov;84(22):11802-13

2009

Jonges M, van der Lubben IM, Dijkstra F, Verhoef L, Koopmans M, Meijer A, Dynamics of antiviral-resistant influenza viruses in the Netherlands, 2005-2008, Antiviral Res. 2009 Sep;83(3):290-7

Gooskens J ; Jonges M ; Claas ECJ ; Meijer A ; Broek PJ van den ; Kroes ACM, Morbidity and mortality associated with nosocomial transmission of oseltamivir-resistant influenza A (H1N1) virus, JAMA 2009; 301(10):1042-6

Gooskens J, Jonges M, Claas EC, Meijer A, Kroes AC, Prolonged Influenza Virus Infection during Lymphocytopenia and Frequent Detection of Drug-Resistant Viruses, J Infect Dis. 2009 May 15;199(10):1435-41.

Hahné S, Donker T, Meijer A, Timen A, van Steenbergen J, Osterhaus A, van der Sande M, Koopmans M, Wallinga J, Coutinho R; Beaujean D, Beersma T, Boucher C, van Boven M, Brandsema P, de Bruin E, Coutinho R, Deuning C, Dijkstra F, Dittrich S, Donker T, van Eijk A, van Gageldonk R, Hahné S, ten Ham P, van der Have J, van den Hoek A, van der Hoek W, Isken L, Jacobi A, Jacobs P, Jonges M, van den Kerkhof H, van Kessel R, Koopmans M, Kroneman A, van der Lubben M, Meijer A, Monen J, Osterhaus A, Petrignani M, Ruijs H, van der Sande M, ter Schegget R, Schutten M, Siebbeles M, van Steenbergen J, Steens A, Swaan C, Timen A, Vennema H, Verhoef L, Vriend R, Waegemaekers T, Wallinga J., Epidemiology and control of influenza A(H1N1)v in the Netherlands: the first 115 cases, Euro Surveill. 2009 Jul 9;14(27).

TGTTGCACTTGACTCAAGGAACCTGCTGGGAACAAATGTACACACCGGGAGGGGAGGTGAGAAATGATGATGTTGATCAGAGTTTAATTATTGCTGCTAGAAATATTGTTAGGAGAG
GCTTCGCTCTTGGAGATGTGCCATAGTACACAAATTGGCGGAATAAGGATGGTAGACATCCTTAGACAAAACCCAACAGAAGAGCAAGCTGTGGATATATGCAAAGCAGCAATGGGT
TTTTGGAGGTTTCACTTTCAAAAGGACAAGTGGGTCATCTGTCAAAAGAGAAGAGGAAGTGCTCACAGGCAACCTCCAAACATTGAAAATAAAAGTGCATGAAGGATATGAGGAAT
ACAGCCATTCTAAGAAAGGCAACCAGAAGGCTGATCCAATTGATAGTGAGTGGGAGAGACGAGCAGTCAATCGCCGAAGCAATTATAGTGGCAATGGTGTTCTCACAAGAGGATTG
TTGAATTTTGTCAACAGAGCGAACCAGCGGCTAAATCCTATGCATCAGCTCCTGAGGCATTTCCAAAAGGATGCAAAGGTGCTGTTCCAAAACTGGGGAATTGAACCCATTGACAAT

9 789064 649035 >