



Published in final edited form as:

*Science*. 2012 June 22; 336(6088): 1541–1547. doi:10.1126/science.1222526.

## The potential for respiratory droplet transmissible A/H5N1 influenza virus to evolve in a mammalian host

Colin A. Russell<sup>1,2,3</sup>, Judith M. Fonville<sup>1,2</sup>, André E. X. Brown<sup>4</sup>, David F. Burke<sup>1,2</sup>, David L. Smith<sup>3,5,6</sup>, Sarah L. James<sup>1,2</sup>, Sander Herfst<sup>7</sup>, Sander van Boheemen<sup>7</sup>, Martin Linster<sup>7</sup>, Eefje J. Schrauwen<sup>7</sup>, Leah Katzelnick<sup>1,2</sup>, Ana Mosterín<sup>1,2,8</sup>, Thijs Kuiken<sup>7</sup>, Eileen Maher<sup>9</sup>, Gabriele Neumann<sup>9</sup>, Albert D. M. E. Osterhaus<sup>7</sup>, Yoshihiro Kawaoka<sup>9,10,11,12</sup>, Ron A. M. Fouchier<sup>7</sup>, and Derek J. Smith<sup>1,2,3,7,\*</sup>

<sup>1</sup>Department of Zoology, University of Cambridge, Cambridge, CB2 3EJ, United Kingdom <sup>2</sup>WHO Collaborating Center for Modeling, Evolution, and Control of Emerging Infectious Diseases, Cambridge, CB2 3EJ, United Kingdom <sup>3</sup>Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA <sup>4</sup>MRC Laboratory of Molecular Biology, Cambridge, United Kingdom <sup>5</sup>Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA <sup>6</sup>Malaria Research Institute, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, 21205, USA <sup>7</sup>Department of Virology, Erasmus Medical Center, Rotterdam, the Netherlands <sup>8</sup>Center for Research in Health and Economics, Universitat Pompeu Fabra, Barcelona, Spain <sup>9</sup>Department of Pathobiological Sciences, School of Veterinary Medicine, University of Wisconsin, Madison, WI, USA <sup>10</sup>Department of Special Pathogens, International Research Center for Infectious Diseases, Institute of Medical Science, University of Tokyo, Shirokanedai, Minato-ku, Tokyo, Japan <sup>11</sup>Division of Virology, Department of Microbiology and Immunology, Institute of Medical Science, University of Tokyo, Shirokanedai, Minato-ku, Tokyo, Japan <sup>12</sup>ERATO Infection-Induced Host Responses Project, Japan Science and Technology Agency, Saitama, Japan

### Abstract

Avian A/H5N1 influenza viruses pose a pandemic threat. As few as five amino acid substitutions, or four with reassortment, might be sufficient for mammal-to-mammal transmission by respiratory droplets. From surveillance data we find that two of these substitutions are common in A/H5N1 viruses and thus some viruses might require only three additional substitutions to become transmissible via respiratory droplets between mammals. We use a mathematical model of within-host virus evolution to study factors that could increase and decrease the probability of the remaining substitutions evolving after the virus has infected a mammalian host. These factors combined with the presence of some of these substitutions in circulating strains, make a virus evolving in nature a potentially serious threat. These results highlight critical areas where more data are needed for assessing, and potentially averting, this threat.

Recent studies have shown that the A/Indonesia/5/2005 avian A/H5N1 influenza virus may require as few as five amino acid substitutions (1), and the A/Vietnam/1203/2004 A/H5N1 influenza virus requires four substitutions and reassortment (2), to become transmissible between ferrets via respiratory droplets. Here we assess the likelihood that these substitutions could arise in nature. We first analyze A/H5N1 sequence surveillance data to identify if any of these substitutions are already circulating. We then explore the probability of the virus evolving the remaining substitutions after a spill-over event of an avian virus

\*Correspondence to: dsmith@zoo.cam.ac.uk.

into a single mammalian host and in a short chain of transmission between mammalian hosts.

The minimal set of substitutions identified by (1) (the Herfst *et al*-set) contains two receptor binding amino acid substitutions, Q222L and G224S (H5 numbering used throughout) in the hemagglutinin (HA), known to change the virus from the more avian-like alpha-2-3 linked sialic acid specificity to the more human-like alpha-2-6 linked sialic acid (3,4). The remaining three substitutions in the set are T156A in HA which disrupts the N-linked glycosylation sequon spanning positions 154-156, H103Y in the HA trimer-interface, and E627K in the PB2 which is a common mammalian polymerase adaptation (5).

The four amino acid substitutions in HA identified by (2) (the Imai *et al*-set) also contain two receptor binding amino acid substitutions, N220K and Q222L, one of which is in common with the Herfst *et al*-set, and which together are known to change the sialic acid linkage preference to the more human-like alpha-2-6 linkage (2). The remaining two substitutions are N154D which disrupts the same N-linked glycosylation sequon as the T156A substitution in the Herfst *et al*-set, and T315I in the stalk region.

Of the three receptor binding substitutions in the two sets, only N220K in the Imai *et al*-set has been detected by surveillance in consensus sequencing of the HA of A/H5N1 viruses, and only in two of 3,392 sequences (both avian viruses, one from 2007 in Vietnam, one from Egypt in 2010, indicated by arrows in Figs. 1B and 1F). The T315I stalk substitution and H103Y trimer interface substitution have each been detected once in two viruses from China in 2002 (brown arrows Figs. 1A and 1B). T315I has been detected in two pre-1997 H5N1 viruses, four H5N2 viruses, two H5N3 viruses, and two H5N9 viruses. H103Y has been detected in five H5N2 viruses and one H5N3 virus. The remaining substitutions, N154D and T156A in the HA glycosylation sequon, and E627K in PB2 however are common and occur in 942/3,392, 1,803/3,392 and 432/1,612 sequences respectively. Fig. S1 and Tab. S1 show a summary of the substitutions detected in surveillance. For viruses where both HA and PB2 have been sequenced 338/1,533 have lost the 154-156 glycosylation sequon and have E627K in PB2. These viruses have been collected in at least 28 countries in Europe, the Middle East, Africa, and Asia.

The HA glycosylation sequon substitutions, N154D and T156A, have drifted in and out of the avian virus population over time suggesting that they may be under little selective pressure in birds. The other substitutions, which are rare in birds, particularly those that change the sialic acid linkage preference, are likely to be negatively selected in birds.

Fig. 1 shows phylogenetic trees of the A/H5N1 HA color-coded by the number of nucleotide mutations required to obtain the five Herfst *et al*-set (column 1) and four Imai *et al*-set (column 2) of substitutions in HA. Note, obtaining these mutations does not necessarily mean the virus will be transmissible by respiratory droplets between ferrets as the genetic background of each strain is different from the strain used by Herfst *et al* (Fig. 1A blue circle) and the strain used by Imai *et al* (Fig. 1J red circle). Other than for clade 2.3.2.1, the variation in color in Fig. 1 columns 1 and 2 is due to the presence (mostly in E and SE Asia) or absence (mostly outside of E and SE Asia) of the glycosylation sequon at positions 154-156.

The sequenced viruses that are closest to the Herfst *et al*-set are in clade 2.3.2.1 (Fig. 1A, S1A). These HAs have acquired a silent nucleotide mutation that makes the amino acid substitution G224S require only one nucleotide mutation instead of the two mutations for other strains. It is the requirement of these two nucleotide mutations that makes viruses usually farther from the Herfst *et al*-set than the Imai *et al*-set. The viruses in clade 2.3.2.1 have been sampled in Nepal, Mongolia, Japan and Korea from 2009 to 2011. Seventeen out

of 94 of these viruses have been sequenced in PB2 (Fig. 1D) and none have the E627K mutation. Thus the closest known viruses to the Herfst *et al*-set by consensus sequencing are four nucleotide substitutions away.

The majority of H5 viruses in clade 2.2 (and its subclades) are three nucleotide mutations from the Imai *et al*-set in HA (Figs. 1F and 1J). These viruses have been sampled in Europe from 2005 to 2007, in the Middle East (including Egypt) from 2005 to 2011, and from Africa from 2005 to 2007. Viruses sampled in 2010 and 2011 are indicated by the red portion of the vertical line delimiting the clade (Fig. 1, and by the timeseries in Figs. S1F and S1J). Fig. 1 column 3 shows that if it is the loss of glycosylation that is important, rather than any other effect of N154D, that almost all the non-Asian viruses have lost the glycosylation sequon, and thus all these viruses would potentially be functionally three nucleotides from the Imai *et al*-set in HA.

The viruses indicated by the black arrows in Figs. 1B and 1F (one from Vietnam in 2007, one from Egypt in 2010) have the N220K receptor binding substitution and have lost the glycosylation sequon at positions 154–156. Thus these two viruses are two nucleotide substitutions from the Imai *et al*-set in HA, and are the viruses closest to having the full Imai *et al*-set in HA detected to date by consensus sequencing.

Surveillance has detected humans with A/H5N1 viruses four nucleotide mutations from the full Herfst *et al*-set, and three from the Imai *et al*-set in HA. Viruses isolated from human A/H5N1 infections (Fig. 1, bottom row) are generally the same number of mutations in HA away from the Herfst *et al*- and Imai *et al*-sets, by consensus sequencing, as their most closely related avian viruses. The within-host evolution modeling below indicates that any host adaptation substitutions would only reach a small proportion of the total virus population in the first spill-over host, and while potentially critical in the host-adaptation process, would not be detected by consensus sequencing. Thus the absence of evidence of host-adaptation by consensus sequencing is not evidence for the absence of potentially critical adaptation to the mammalian host. See (6) for details of human strains and their most genetically similar avian A/H5N1 viruses.

## Model of within-host evolution

To explore the probability of accumulating the remaining nucleotide mutations after the avian virus has been transmitted to a human (or other mammalian host), we constructed a mathematical model (7–10) of the within-host evolutionary dynamics of the virus. In the model, errors made by the virus polymerase are the source of mutation ( $10^{-5}$  mutations per site, per genome replication), the initial virus population expands exponentially (each infected cell produces  $10^4$  virions (11–12),  $10^{10}$  cells can be infected (13)) until it reaches  $10^{14}$  virions after which the virus population size stays roughly constant, and selection is modeled by differences in expected numbers of progeny [(6) Fig. S2, Tab. S2]. The results of the model are largely insensitive to number of cells that can be infected, maximum virus population size, and whether the virus population remains roughly constant or declines (Figs. S3, S4, S5). Typical infections were simulated out to five days corresponding to the approximate time of peak viral load, and long-duration infections to 14 days (14).

It is not possible to calculate the level of risk precisely because of uncertainties in some aspects of the biology. We use the model to compare the relative effects of factors that could increase or decrease the probability of accumulating mutations, and to identify areas for further investigation that are critical for more accurate risk assessment. We compare and contrast the effects of factors that can increase the probability of accumulating mutations, and thus evolving a respiratory droplet transmissible A/H5N1 influenza virus in a mammalian host, and factors that could decrease the probability of evolving a such a virus.

The factors we consider that can increase the probability are: random mutation, positive selection, long infection, alternate functionally-equivalent substitutions, and transmission of partially adapted viruses as a proportion of the within-host diversity both in the avian-mammal and the mammal-mammal transmission events (10,14–18). The factors we consider that can decrease the probability are: an effective immune response, deleterious mutations, and order dependence in the acquisition of mutations. We considered these factors for starting viruses differing in the number of mutations that separates them from a respiratory droplet transmissible A/H5N1 virus; i.e., viruses that require five, four, three, two, or one mutations at specific positions in the virus HA, reflecting that zero, one, two, three, or four of the mutations are already present in the avian population and thus are present at the start of the infection in mammals. We treat each amino acid substitution as if it can be acquired by a single nucleotide mutation as is the case for the circulating viruses closest to acquiring the Herfst *et al*- or Imai *et al*-sets (see (6) for the general case).

## Random mutation

Even without any positive selection pressure, the random process of mutations introduced by the virus polymerase in the expanding population of viruses will on average produce viruses that contain the required single, double or triple mutations, and even some quadruple mutants. These mutants will arise after a few days of an infection in a host in which the virus replicates efficiently (Fig. 2A), and would be delayed if replication is impaired (Fig. S5). However, the existence of a virus within-host does not mean that it will transmit, as it might exist only as a small proportion of the total virus population and thus have little chance of being excreted (Fig. 2B). The minimum proportion of mutant virus required to make transmission likely is not known, but increased proportion translates into increased probability of transmission, thus we focus on proportion of mutant virus in the total virus population. These proportions (equivalent to the probability of a single virion to be a mutant), both here and below, cannot yet be determined precisely — they are sensitive to some biological parameters that are not yet known accurately, and some which are specific to a particular virus or mutant. For such parameters we test a range of the current best estimates, and focus on the relative, rather than the absolute, effects (6).

## Positive selection

Some of the substitutions identified by (1) and (2) have been shown to increase within-host virus fitness, specifically the loss of glycosylation at positions 154 and 156, and E627K in PB2. However, given the absence of specific information on the within-host selective advantage or disadvantage conferred by each substitution, or combination of substitutions, we considered two cases of positive selection: one in which each individual substitution confers an additive advantage (hill-climb) and one in which only viruses that have acquired all substitutions have an advantage (all-or-nothing). We consider a total advantage of 1.1-, 2- or 10-fold in each genome replication step for the full set of respiratory droplet transmission enabling substitutions (Tab. S2, Fig. S6) (A two-fold advantage at each genome replication step translates into an approximately 100-fold increase in mutant virus titer after 36 hours). In the all-or-nothing scenario, a strong increase in proportion occurs for viruses that have acquired all mutations, due to its substantial fitness advantage over the rest of the population. The rate at which all-or-nothing selection increases the proportion of respiratory droplet transmissible A/H5N1 viruses, compared to the neutral case, is mostly independent of the number of mutations required (Fig. 3A). In contrast, for hill-climb selection, the rate of increase above the neutral case decreases when fewer mutations are required (Fig. 3A). This difference between the all-or-nothing and hill-climb is because the fitness differential from the starting virus to the respiratory droplet transmissible A/H5N1 virus decreases as the number of needed mutations decreases (i.e. if some of the mutations

are already present in the avian host) (Tab. S2, (6)). We consider this hill-climb case to be the most likely situation during the host-adaptation we are modeling here (in the absence of deleterious mutations). However, we have also compared the two selection scenarios when the starting fitness of all-or-nothing and hill-climb are the same independently of the starting number of necessary mutations, and discuss the subtle tradeoff between the fitness advantage of, and clonal-interference among, intermediate mutants (6,19) (Figs. S7, S8).

## Long infection

Because both random mutation and positive selection increase the expected proportion of mutated virions with every viral generation, the longer a host is infected, the greater the proportion of a particular mutant (Fig. 4, (15)). Human A/H5N1 infections lasting 14 days or longer have been reported especially in children, the elderly and the immunocompromised (14), and have been associated with the evolution of oseltamivir resistance (20). It might be that only immunocompromised individuals can typically transmit the virus late in a long-infection. The increasing proportion of mutant virus is only dependent on continued virus production and is independent of whether the virus load stays constant or declines (Fig. S4 (21)). The variance in the proportion of mutant virus (pale regions, Fig. 4) increases with each additional mutation required due to the increased number of combinatorial options and the greater selective advantage of mutant viruses compared to wildtype viruses in the hill-climb scenario. The pale regions only reflect the within-model variance in results as indicated by the different runs of the stochastic model, and not uncertainty as a result of other factors; sensitivity of the outcomes for model parameters such as the error rate and the branching factor are explored in (6) (Fig. S5).

## Functionally-equivalent substitutions

The sets of substitutions required for a respiratory droplet transmissible A/H5N1 virus identified by (1) and (2) are unlikely to be the only combinations of substitutions capable of producing a respiratory droplet transmissible A/H5N1 virus. If particular biological traits could be achieved by other substitutions, this would increase the expected proportions of respiratory droplet transmissible A/H5N1 viruses. This is likely to be the case given that there are multiple substitutions that can cause changes in receptor binding specificity and two sites where substitutions will result in loss of glycosylation: positions 154 and 156 (Tab. S3). If the five mutations could be from any 10 specific positions in the virus genome (or if two already existed in nature, three from any eight), then there would be 252 (or 56) combinations, and this would raise the proportion of respiratory droplet transmissible A/H5N1 virus within a host by  $\sim 10^{2.5}$  (or  $\sim 10^{1.5}$ ) above the case of positive selection alone after five days (Figs. 3B, S9, S10, Tab. S4).

## Avian–mammal transmission of partially adapted mutants

We consider the case in which one of the required mutations exists as a small proportion of the avian within-host viral population, or in the viral populations from the  $>20$  mammalian hosts in which A/H5N1 infections have been observed (22–25), such that they would not be detected by the usual consensus sequencing techniques. If the mutant is one of the 100 virions that seed an infection (16,17), then with positive selection, the probability of acquiring the remaining mutations increases by  $10^3$  after five days of infection above the case of positive selection alone (Fig. 3B). If the proportion of mutant in the seeding population is  $10^{-4}$  however, the increase in proportion of respiratory droplet transmissible A/H5N1 virions in the mammalian host is small (Fig. S11).

## Mammal-mammal transmission of partially adapted viruses

Transmission of viruses between mammals that have some but not all of the substitutions necessary for respiratory droplet transmission potentially increases the risk of evolving a respiratory droplet transmissible A/H5N1 virus but this increase is modulated by the difficulty of transmitting partially adapted strains and the loss of diversity at transmission. Two primary factors strongly modulate the effect of transmission on the accumulation of mutations. First, transmission could decrease the accumulation of mutations by the loss of low-proportion mutants because only a limited portion of the virus population will be transmitted. Second, transmission could increase the accumulation of respiratory droplet transmission enabling mutations by concentrating transmissible virus during excretion from or seeding into a host; for example, if the adapted virus has increased tropism for the mammalian upper respiratory tract and therefore concentrated in the nose and throat. Thus the effect of transmission can range from negligible if mutants are culled by the loss of diversity at transmission, to substantial if selection favors mutants at transmission (Tab. S5). Given that A/H5N1 virus infections have been observed in >20 mammalian species there is a potentially large pool of non-human hosts where short chains of transmission could play a role in the emergence of respiratory droplet transmissible A/H5N1 viruses.

In contrast to these factors that could increase the rate of accumulating substitutions here we discuss factors that could decrease this rate.

## An effective immune response

An effective immune response that substantially shortened an infection would decrease the probability of the accumulation of mutations, however there are many reported cases of infections up to and beyond five days (14,21). Variation in the number of virions produced by each infected cell does not affect the deterministic calculations of the proportion of mutants. However, if this number is substantially lower for the stochastic simulations, for example 25 (6), compared to 10,000 (used for most of the figures), the slower growth and lower total number of viruses could substantially delay the appearance of mutants within host. As the number of required mutations increases, stochastic effects caused by the slower growth decrease the proportion of these mutants (Fig. S5, (6)).

## Deleterious intermediate mutations

The receptor binding, and trimer-interface or stalk substitutions, required by (1) and (2) are, as we have seen, either rare or absent in influenza viruses isolated to date. The receptor-binding substitutions, though deleterious in birds, would be expected to be advantageous in humans. However the details of this host-adaptation are not yet elucidated, and so we also consider the possibility that there are deleterious intermediate mutations, and explore a variety of scenarios (Figs. S12 and S13). When two of the required mutations are individually deleterious (i.e. for these two specific mutations either mutation alone reduces the replicative fitness of the virus to zero) this slows the rate of accumulation of mutations for the three-mutation case by less than the amount that hill-climb positive selection increases the rate above the neutral case (Fig. 3C). When a triple mutation is required (i.e. all single and double mutations reduce the replicative fitness of a mutant virus to zero), this can lower the accumulation rate two logs below the neutral case (Fig. S12). Deleterious (or advantageous) substitutions other than the respiratory droplet transmissible A/H5N1 substitutions can, to a first approximation, be ignored in calculating proportions as such substitutions would on average affect all viruses equally and thus would not specifically affect the accumulation of respiratory droplet transmissible A/H5N1 mutations (6).

## Order dependence in the acquisition of mutations

It is not currently known whether the acquisition of some or all of the respiratory droplet transmission enabling mutations is dependent on the order in which viruses accumulate those mutations. For example, the gain of 2–6-receptor binding might be required before loss of 2–3-receptor binding. If there were any order dependence it would slow down the rate of accumulation of mutations. However, even in the most extreme scenario in which there is a single specific order in which the mutations must be acquired, and any other order results in a virion with a replicative fitness of zero, if fewer than four mutations are required the effect on the rate of accumulation of mutations is less than that of the deleterious scenario described above (Fig. 3C, S14, S15).

In addition to the substitutions in HA, the Imai *et al* virus was a reassortment with an A/H1pdm09 virus. The probability of a reassortment event is difficult to determine given current knowledge. In one study (26) it has been estimated to be more likely than the likelihood of acquiring a single mutation as calculated here.

Highly pathogenic avian A/H5N1 viruses have been infecting humans for over a decade, with ~600 reported cases to date (and possibly many more that have not been reported), but there have yet to be known cases of efficient human-to-human transmission (27,28). One hypothesis for the lack of sustained transmission is that it is not possible for A/H5N1 viruses to become respiratory droplet transmissible in mammals -- (1) and (2) have shown that this may not be the case in ferrets. Another hypothesis is that the number of mutations necessary for respiratory droplet transmissibility might be so great that such a virus would be unlikely to evolve -- we show here that, in biologically plausible scenarios, respiratory droplet transmissible A/H5N1 viruses can evolve during a mammalian infection. Given that respiratory droplet transmission between mammals is possible and that respiratory droplet transmissible A/H5N1 mutants are likely to evolve in infected individuals, the primary impediment to transmission could be whether the respiratory droplet transmissible A/H5N1 viruses comprise a sufficient proportion of the within-host viral population to actually transmit.

The minimum proportion of virus required for transmission is not known, but increased proportion likely translates into increased probability of transmission. There cannot be respiratory droplet transmission if there are no viruses in the air. Given a peak excretion rate  $\sim 10^7$  viruses per day (29,30), a proportion of which are likely to become aerosolized (31), mutants at proportions near or above  $10^{-7}$  might thus be among the particles excreted. Each of the factors analyzed above has a potentially substantial effect on the rate of accumulating mutations (Fig. 3), and the effects of each can be additive. With plausible combinations of these factors, a virus that requires three mutations reaches proportions at which a few respiratory droplet transmissible A/H5N1 viruses are likely to be among the particles excreted. For a virus that requires five mutations, it may only reach such proportions with more extreme combinations of factors, a large number of infected mammals, or if an event occurs that is not encompassed by the model (32). However, it is known that influenza viruses are capable of respiratory droplet transmission in animal models at low infectious doses (33), and that transmission routes other than in respiratory droplets could be important, thus despite the three key current unknowns about transmission (6), even low numbers of excreted respiratory droplet transmissible A/H5N1 virus may be relevant for emergence. The output of the model is a guide to understand the approximate effects of different factors, and should not be interpreted as actual proportions of virus and probabilities of transmission, given the uncertainty inherent in parameter estimates and model structure, the inherent unpredictability of rare events (34).

These results highlight four areas of investigation that are critical to more accurately assess and monitor the risk of a respiratory droplet transmissible A/H5N1 virus emerging, and increasing our understanding of virus emergence in general. Some of this work is already ongoing, planned, or suggested. The work of Herfst *et al* (1) and Imai *et al* (2) and the analyses here help to prioritize particular areas.

First, in the area of surveillance. Additional surveillance in higher risk regions where respiratory droplet transmission enabling mutations are already prevalent (Figs. 1 and S1) (and in regions connected by travel, trade and migratory flyways) is key for monitoring the emergence of a respiratory droplet transmissible A/H5N1 virus. Surveillance of non-human mammalian hosts, especially any that harbor long infections or live in large groups, is important for the early identification of mammalian adaptation. Additionally, studies are needed on the accumulation of mutations within-host and in short chains of transmission in mammals (22–25), even when endemic circulation has not been observed.

Second, deep sequencing of avian and other non-human virus samples is necessary to accurately estimate the prevalence of the respiratory droplet transmission enabling amino acid substitutions in nature. Deep sequencing of human samples, particularly at multiple time points from individuals with long infections, would be useful for evaluating within-host evolution, for estimating selective advantage of substitutions, and for testing the underlying dynamics and assumptions of the model (15). Respiratory droplet transmissible A/H5N1 mutations present in a proportion higher than the polymerase error rate, i.e. exceeding approximately  $10^{-5}$ , but far below the threshold for detection by consensus sequencing and thus not detectable by current surveillance practices, would increase the risk of respiratory droplet transmissible A/H5N1 evolving. Thus, sequencing deeper than that currently routinely achieved for RNA viruses (ideally detecting mutations at 0.1% frequency and lower for detailed studies) is necessary to more accurately assess the risk posed by intra-host variability (15).

Third, experiments are needed to determine which substitutions, besides the already identified receptor binding substitutions by (1) and (2), are capable of producing respiratory droplet transmissible A/H5N1 viruses, including the important case of functionally equivalent substitutions or alternative sets of substitutions that would require fewer nucleotide mutations than the Herfst *et al*- or Imai *et al*-sets. This work will be important for calculating risk and for monitoring in surveillance.

Fourth, further studies are needed to elucidate the changes in within-host fitness and between-host transmissibility associated with each respiratory droplet transmission enabling substitution and combination of substitutions. These studies are necessary for determining the dynamics of within-host selection (including data on, and modeling of, the effects of glycan heterogeneity between the upper and lower respiratory tract (6)) and the potential for transmission of partially adapted viruses. It is important to determine the strength of selection at transmission, as it can increase the proportion of respiratory droplet transmission enabling substitutions. Further work is needed to refine the estimate for virus excretion and the minimum human infectious dose (29).

Numerous avian A/H5N1 viruses have been sampled in the last two years that are four nucleotide mutations from acquiring the Herfst *et al*-set of HA and PB2 substitutions, and three nucleotide mutations from acquiring the Imai *et al*-set in HA (the Imai *et al*-set also requires a reassortment event). Precise estimates of the probability of evolving the remaining mutations for the virus to become a respiratory droplet transmissible A/H5N1 virus cannot be accurately calculated at this time because of gaps in knowledge of the factors described above. However, the analyses here, using current best estimates, indicate that the remaining



mutations could evolve within a single mammalian host, making the possibility of a respiratory droplet transmissible A/H5N1 virus evolving in nature a potentially serious threat.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

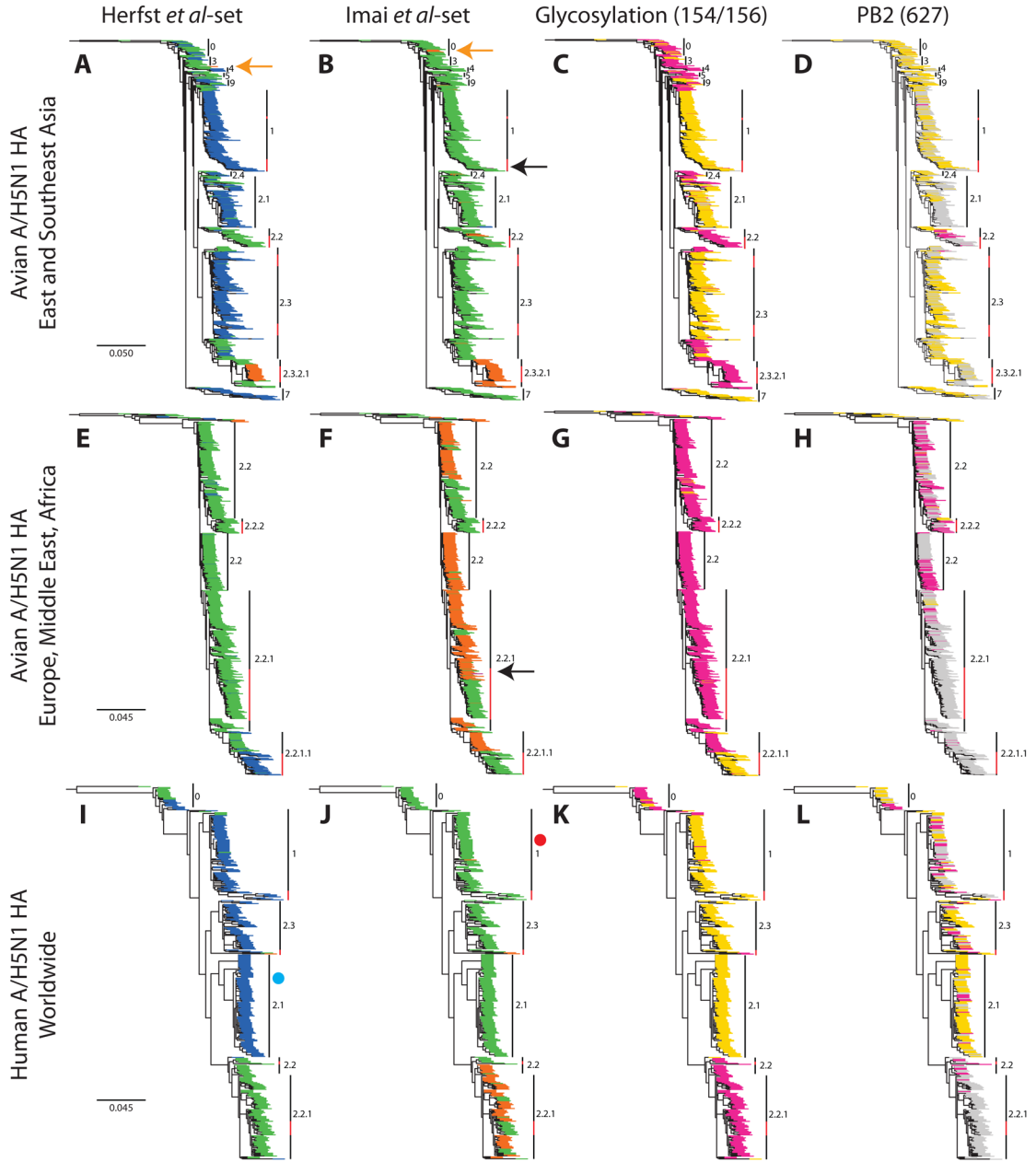
## Acknowledgments

C.A.R. was supported by a University Research Fellowship from the Royal Society. The authors acknowledge an NWO VICI grant, EU FP7 programs EMPIRE (223498) and ANTIGONE (278976), Human Frontier Science Program (HFSP) program grant P0050/2008, Wellcome 087982AIA, and NIH Director's Pioneer Award DP1-OD000490-01. R.A.M.F. was supported by NIAID-NIH contract HHSN266200700010C. A.E.X.B. was supported by a long-term fellowship from the HFSP. E.A.M., G.N. and Y.K. are supported by the Bill and Melinda Gates Foundation (OPPGH5383) and NIAID-NIH grant R01 AI 069274; in addition, Y.K. was supported by a Grant-in-Aid for Specially Promoted Research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, and by ERATO (Japan Science and Technology Agency). Y.K. and Y.K. and G.N. have a financial interest, as founders, of FluGen, and hold a patent on influenza virus reverse genetics. Y.K. and G.N. have a paid consulting relationship with Theraclone, Y.K. also with Crucell. Y.K. has received speaker's honoraria from Chugai Pharmaceuticals, Novartis, Daiichi-Sankyo Pharmaceutical, Toyama Chemical, Wyeth, GlaxoSmithKline and Astellas Inc., and grant support from Chugai Pharmaceuticals, Daiichi Sankyo Pharmaceutical, Toyama Chemical, Otsuka Pharmaceutical Co., Ltd.. A.D.M.E.O. (on behalf of Viroclinics Biosciences BV) has advisory affiliations with GSK, Novartis, and Roche. A.D.M.E.O. is CSO of Viroclinics Biosciences BV. We thank S. Cornell, R. Johnstone and L. Reperant for helpful discussions and the reviewers for their detailed and thoughtful comments.

## References and Notes

- Herfst S, et al. *Science*. :XXX.
- Imai M, et al. *Nature*. 2012;1038/nature10831
- Vines A, et al. *J Virol*. 1998; 72:7626. [PubMed: 9696865]
- Chutinimitkul S, et al. *J Virol*. 2010; 84:6825. [PubMed: 20392847]
- Hatta M, et al. *PLoS Pathog*. 2007; 3:e133.
- Supplementary materials are available on *Science Online*.
- Luria SE, Delbruck M. *Genetics*. 1943; 28:491. [PubMed: 17247100]
- Mode CJ, Raj T, Sleeman CK. *J Prob and Stat*. 2011; 2011:1.
- Coffin JM. *Science*. 1995; 267:483. [PubMed: 7824947]
- Perelson AS, Rong L, Hayden FG. *J Infect Dis*. 2012;1093/infdis/jis265
- Sidorenko Y, Reichl U. *Biotechnol Bioeng*. 2004; 88:1. [PubMed: 15384040]
- Möhler L, Flockerzi D, Sann H, Reichl U. *Biotechnol Bioeng*. 2005; 90:46. [PubMed: 15736163]
- Weibel ER. *Bull Europ Physiopathol Resp*. 1979; 15:999.
- Writing Committee WHO. *N Engl J Med*. 2008; 358:261. [PubMed: 18199865]
- Murcia PR, et al. *J Virol*. 2010; 84:6943. [PubMed: 20444896]
- Kuiken T, et al. *Science*. 2006; 312:394. [PubMed: 16627737]
- Bonhoeffer S, Nowak MA. *Proc R Soc London Ser B*. 1997; 264:631.
- Wain-Hobson S. *Curr Opin Genet Dev*. 1993; 3:878. [PubMed: 7509668]
- Hill WG, Robertson A. *Genet Res*. 1966; 8:269. [PubMed: 5980116]
- Anton A, et al. *Diag Micr Infec Dis*. 2010; 68:214.
- de Jong MD, et al. *Nature Med*. 2006; 12:1203. [PubMed: 16964257]
- Nidom CA, et al. *Emerg Infect Dis*. 2010; 16:1515. [PubMed: 20875275]
- Keawcharoen J, et al. *Emerg Infect Dis*. 2004; 10:2189. [PubMed: 15663858]
- Qi X, et al. *PLoS one*. 2009; 4:e4682. [PubMed: 19270752]
- Reperant L, et al. *Rev Sci Tech*. 2009; 1:137. [PubMed: 19618623]
- Ferguson NM, et al. *Science*. 2004; 304:968. [PubMed: 15143265]

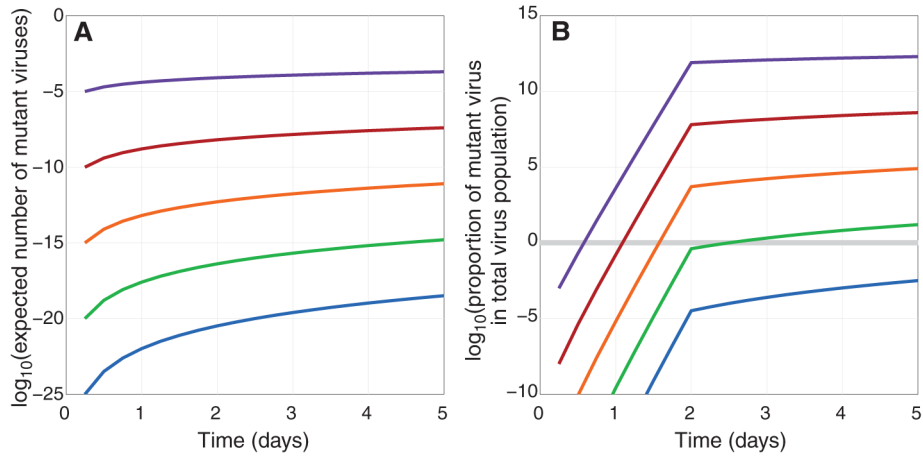
27. Aditama TY, et al. PLoS one. 2012; 7:e29971. [PubMed: 22238686]
28. Yang Y, Halloran ME, Sugimoto JD, Longini IM. Emerg Infect Dis. 2007; 13:1348. [PubMed: 18252106]
29. Atkinson MP, Wein LM. Bull Math Biol. 2008; 70:820. [PubMed: 18278533]
30. Fabian P, et al. PLoS one. 2008; 3:e2691. [PubMed: 18628983]
31. Tellier R. J Roy Soc Interface. 2008; 6:S783. [PubMed: 19773292]
32. Ord T, Hillerbrand R, Sandberg A. J Risk Research. 2010; 13:191.
33. Lednicky JA, et al. Virol J. 2010; 7:231. [PubMed: 20843329]
34. Spiegelhalter DJ, Riesch H. Philos Trans R Soc London Ser A. 2011; 269:4730.
35. Bahl J, et al. Influenza and Other Respiratory Pathogens. 2012; 6:1.
36. Guindon S, et al. Syst Biol. 2010; 59:307. [PubMed: 20525638]
37. Posada D, Crandall KA. Bioinformatics. 1998; 14:817. [PubMed: 9918953]
38. Zwickl, DJ. thesis. University of Texas; 2006.



**Fig. 1.** Phylogenetic trees of the A/H5N1 HA1 nucleotide sequences. The sequences are split into three trees: 2,022 avian H5 sequences from East and South-East (E and SE) Asia (top row); 1,097 avian H5 sequences from Europe, the Middle East, and Africa (middle row); and 385 human H5 sequences (bottom row). Each sequence is color coded by the minimum number of nucleotide mutations required: to obtain the four amino acid substitutions in HA in the Herfst *et al*-set (column 1), to obtain the four amino acid substitutions in the Imai *et al*-set (column 2), to disrupt the N-linked glycosylation sequon spanning positions 154–156 in HA (column 3), and to obtain E627K in the PB2 segment of the corresponding virus in these HA trees (column 4). In columns 1 and 2, blue indicates five nucleotide changes, green four, and orange three. In columns 3 and 4 yellow indicates viruses that require one mutation and pink zero mutations. Gray indicates PB2 not sequenced. Clades as defined by (35) are marked to

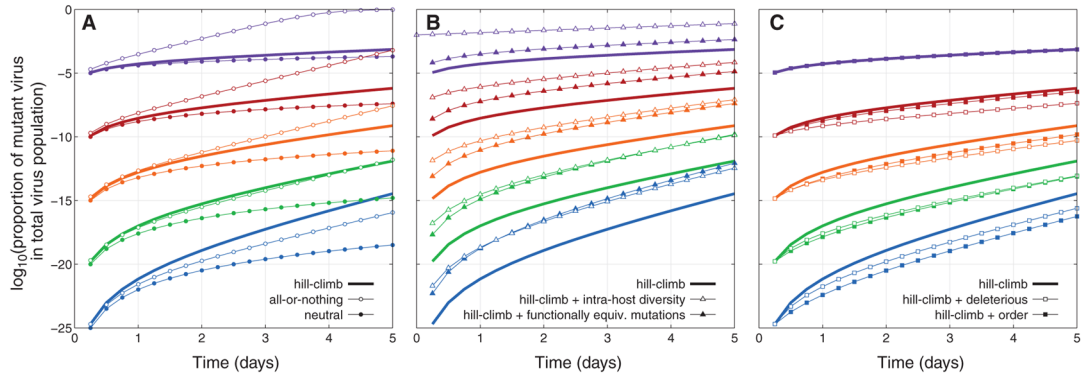
the right of the branches, the red portion of the vertical clade-identification lines indicates strains sampled in 2010 or 2011. The viruses indicated by black arrows are two nucleotides from the Imai *et al*-set. The virus indicated in A by the brown arrow has the H103Y substitution, the virus indicated in B by the brown arrow has the T315I substitution. The blue circle indicates A/Indonesia/5/2005, the red circle indicates A/Vietnam/1203/2004, the starting viruses used by (1) and (2) respectively. The initial trees were constructed with PhyML version 3.0 (36), with A/Chicken/Scotland/1959 as the root, using GTR+I+ $\Gamma$ 4 [determined by Model Test (37)] as the evolutionary model. GARLI version 0.96 (38) was run on the best tree from PhyML for one million generations to optimize tree topology and branch lengths. Fig. S1 shows “zoom-able” versions of these trees to show detail.

Russell, fig2, x28p6



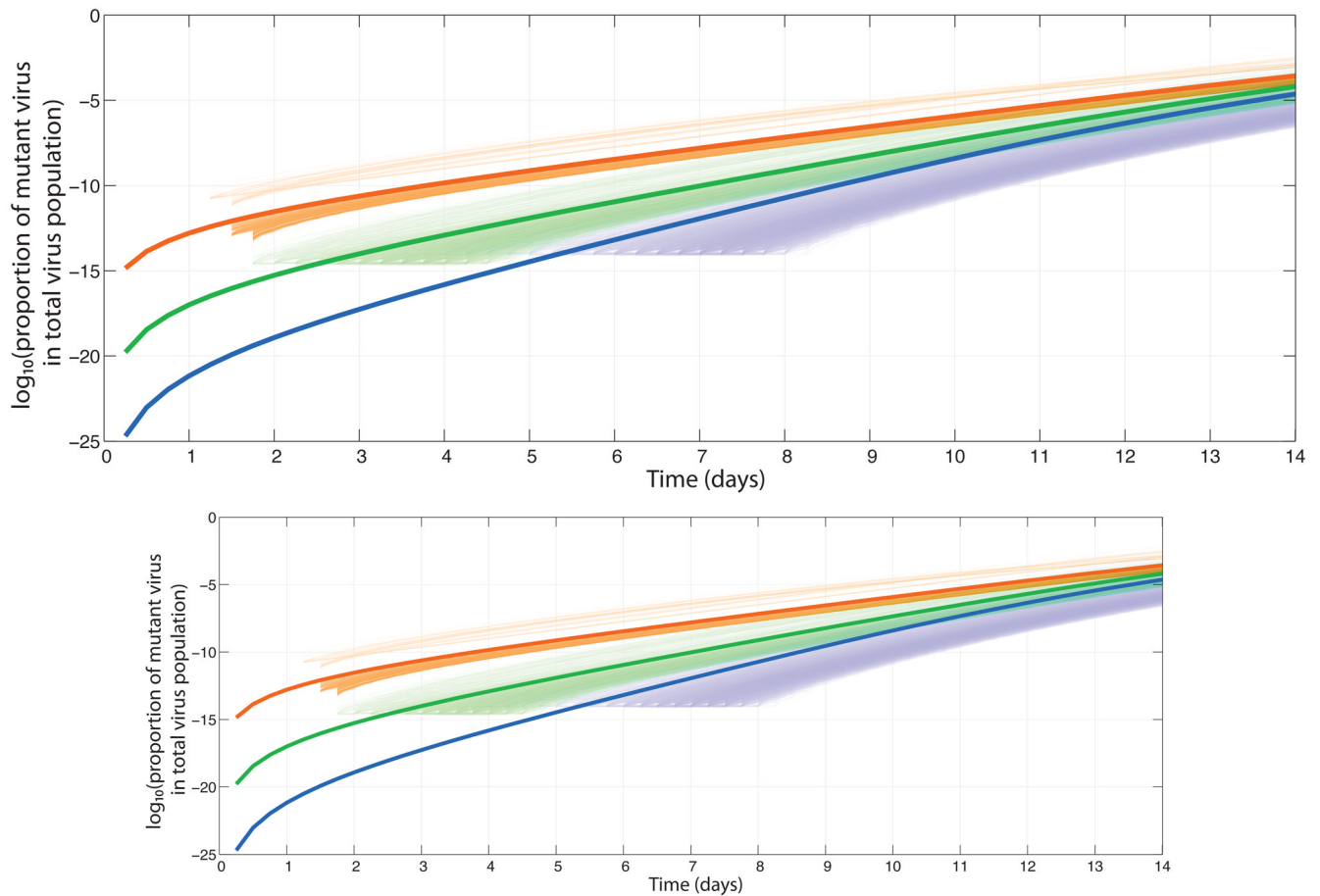
**Fig. 2.** Expected proportions and absolute numbers of respiratory droplet transmissible A/H5N1 virions within a host initially infected by strains that require five (blue), four (green), three (orange), two (red), or one (purple) mutation(s) to become respiratory droplet transmissible, calculated from the deterministic model. **A.** Expected proportion of respiratory droplet transmissible A/H5N1 viruses in the total virus population over time in the random mutation case (when all mutations are fitness neutral). **B.** The absolute number of respiratory droplet transmissible A/H5N1 viruses in a host; the intersections with the grey line indicate the point when at least one virus in each host is expected to have the required mutations. The change in slope is due to the transition in the virus population from exponential expansion to constant size.

Russell, fig3, x43p



**Fig. 3.**

Factors that increase the proportion of respiratory droplet transmissible A/H5N1 virus based on starting viruses that require five (blue), four (green), three (orange), two (red), or one (purple) mutation(s) to become respiratory droplet transmissible. **A.** The effect of hill-climb and all-or-nothing positive selection compared to random mutation alone. **B.** The effect of avian–mammal transmission of partially adapted virus (100 viruses start the infection, one of which has a mutation) and the effect of alternate substitutions with 10 functionally equivalent sites for a virus requiring five mutations (blue), nine requiring four (green), eight requiring three (orange), seven requiring two (red), six requiring one (purple); both with hill-climb selection, as compared to hill-climb selection alone. **C.** The effect of two of the required mutations being individually deleterious (i.e. for these two specific mutations either mutation alone reduces the replicative fitness of the virus to zero) and the effect of complete order dependence of acquiring mutations, both with hill-climb selection as compared to hill-climb selection alone.



**Fig. 4.** Proportion of respiratory droplet transmissible A/H5N1 virus in a long infection with virus replication for 14 days in the presence of hill-climb selection. Bold lines show results from a probability-based deterministic model of virus mutation, the pale region (composed of lines) shows 10,000 stochastic model simulations for each starting virus. Starting viruses require either five (blue), four (green), three (orange) mutations to become respiratory droplet transmissible. For the stochastic simulations the lines start when the first virion that has the required mutations appears.