LETTER

doi:10.1038/nature12515

The genesis and source of the H7N9 influenza viruses causing human infections in China

Tommy Tsan-Yuk Lam[1,2,3]*, Jia Wang[1,3]*, Yongyi Shen[1,3,4]*, Boping Zhou[2], Lian Duan[2,3], Chung-Lam Cheung[3], Chi Ma[1,3], Samantha J. Lycett[5], Connie Yin-Hung Leung[3], Xinchun Chen[2], Lifeng Li[1,2,3], Wenshan Hong[1], Yujuan Chai[2,3], Linlin Zhou[3], Huyi Liang[1,2,3], Zhihua Ou[1,2,3], Yongmei Liu[1,3], Amber Farooqui[6], David J. Kelvin[6], Leo L. M. Poon[2,3], David K. Smith[1,3], Oliver G. Pybus[7,8], Gabriel M. Leung[1,3], Yuelong Shu[9], Robert G. Webster[10], Richard J. Webby[10], Joseph S. M. Peiris[2,3], Andrew Rambaut[5,11], Huachen Zhu[1,2,3] & Yi Guan[1,2,3]

A novel H7N9 influenza A virus first detected in March 2013 has since caused more than 130 human infections in China, resulting in 40 deaths[1,2]. Preliminary analyses suggest that the virus is a reassortant of H7, N9 and H9N2 avian influenza viruses, and carries some amino acids associated with mammalian receptor binding, raising concerns of a new pandemic[1,3,4]. However, neither the source populations of the H7N9 outbreak lineage nor the conditions for its genesis are fully known[5]. Using a combination of active surveillance, screening of virus archives, and evolutionary analyses, here we show that H7 viruses probably transferred from domestic duck to chicken populations in China on at least two independent occasions. We show that the H7 viruses subsequently reassorted with enzootic H9N2 viruses to generate the H7N9 outbreak lineage, and a related previously unrecognized H7N7 lineage. The H7N9 outbreak lineage has spread over a large geographic region and is prevalent in chickens at live poultry markets, which are thought to be the immediate source of human infections. Whether the H7N9 outbreak lineage has, or will, become enzootic in China and neighbouring regions requires further investigation. The discovery here of a related H7N7 influenza virus in chickens that has the ability to infect mammals experimentally, suggests that H7 viruses may pose threats beyond the current outbreak. The continuing prevalence of H7 viruses in poultry could lead to the generation of highly pathogenic variants and further sporadic human infections, with a continued risk of the virus acquiring human-to-human transmissibility.

After the initial reports of H7N9 influenza infection in humans, field surveillance was conducted during 4–18 April 2013 in Wenzhou (Zhejiang province, 500 km south of Shanghai) and Rizhao (Shandong province, 600 km north of Shanghai), which both border the main outbreak region, and in Shenzhen (Guangdong province, 1,200 km south of Shanghai), an area that has not reported human cases (Supplementary Fig. 1). A total of 1,341 pairs of oropharyngeal and cloacal samples were collected from chickens, ducks, geese, pigeons, partridges and quail. A further 1,006 faecal and water samples from live poultry markets (LPMs), farms and wetlands were also collected (Supplementary Table 1). A total of 388 haemagglutin-positive agents were isolated (10.5% of samples), of which 60 and 85 represented H7 and H9 influenza A viruses. The remaining positive isolates represented other subtypes of influenza A virus or avian paramyxovirus (Supplementary Table 1).

H7 influenza A viruses were only detected in Wenzhou and Rizhao, and only in LPMs. All H7 isolates from Rizhao were H7N9 viruses, whereas those from Wenzhou were all H7N7 viruses, except for two duck isolates that were H7N2 and H7N3 viruses. All H9 isolates were H9N2 viruses (80 from LPMs, 5 from farms). At LPMs in Wenzhou, the H7 virus was at its highest prevalence in chickens (10.1%; 46 out of 457), followed by ducks (2.4%; 3 out of 125) and pigeons (1.6%; 3 out of 188). In Rizhao, LPM H7N9 viruses were only found in chickens (0.7%; 8 out of 1,113). Of the chicken isolates, 100% of H7N9, 65.3% of H7N7 and 94.8% of H9N2 viruses were from oropharyngeal swabs (Supplementary Table 1), suggesting that these H7N9 and H7N7 viruses might replicate in the upper respiratory tract of terrestrial poultry, similar to the enzootic H9N2 viruses[6].

These samples were sequenced to investigate the evolutionary history of avian influenza viruses implicated in the current outbreak of H7N9 infections of humans and poultry. Full genome sequences were obtained for 34 H7N7, 4 H7N9 and 19 H9N2 isolates. The H7 and N7/9 genes of 16 mixed H7/H9 infections were sequenced (Supplementary Table 1), as were 3 H7N9 and 3 H7N7 samples that had multiple H9N2-like internal gene segments. The H7 haemagglutinin gene sequences of the H7N9 viruses isolated from chickens in Rizhao formed a tight monophyletic group (Fig. 1a, lineage 'b') with previously reported human and avian viruses from the current H7N9 outbreak. This was most closely related to a group comprising mainly H7N7 viruses obtained from Wenzhou chickens, ducks and pigeons (Fig. 1a, lineage 'c'). All viruses isolated from chickens in these two groups had internal gene complexes that were closely related to those present in co-circulating H9N2 viruses.

To examine the genesis of these H7N9 and H7N7 viruses, we sequenced 197 archived isolates of H7, N9, N7 and H9N2 viruses, obtained during previous influenza surveillance between 2000 and 2013 in southern China (Supplementary Fig. 1). These sequences were analysed together with those obtained in our post-outbreak surveillance, plus all closely related sequences from public databases (see Methods). H7 influenza viruses from East Asian migratory waterfowl were introduced into domestic ducks in China on several occasions during the past decade (Fig. 1a and Supplementary Fig. 2). In 2009–2010, H7 viruses with five NA subtypes were found in duck farms and LPMs in Jiangxi, suggesting an epidemiological bridge from migratory birds to sentinel farm ducks and then to market birds[7] (Supplementary Fig. 2). These introductions to domestic birds persisted for less than two years, except for the introduction of H7N3 viruses (initially isolated in Fujian and Zhejiang in 2010–2011) that led to the 2013 H7N9 outbreak lineage viruses and the H7N7 viruses from Wenzhou (Fig. 1a, lineage 'a').

Previous analyses have suggested that the N9 gene of the H7N9 outbreak lineage was derived from wild bird viruses in Europe and Korea[3,4].

[1]Joint Influenza Research Centre (SUMC/HKU), Shantou University Medical College, Shantou 515041, China. [2]State Key Laboratory of Emerging Infectious Diseases (HKU-Shenzhen Branch), Shenzhen Third People's Hospital, Shenzhen 518112, China. [3]State Key Laboratory of Emerging Infectious Diseases/Centre of Influenza Research, School of Public Health, The University of Hong Kong, Hong Kong SAR, China. [4]State Key Laboratory of Cellular Stress Biology, School of Life Sciences, Xiamen University, Xiamen 361102, China. [5]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK. [6]Joint Vaccine Research Centre (SUMC/UHN), Shantou University Medical College, Shantou 515041, China. [7]Department of Zoology, University of Oxford, Oxford OX1 3PS, UK. [8]Metabiota, San Francisco, California 94104, USA. [9]National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Key Laboratory for Medical Virology, National Health and Family Planning Commission, Beijing 102206, China. [10]Division of Virology, Department of Infectious Diseases, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. [11]Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892-2220, USA.

*These authors contributed equally to this work.

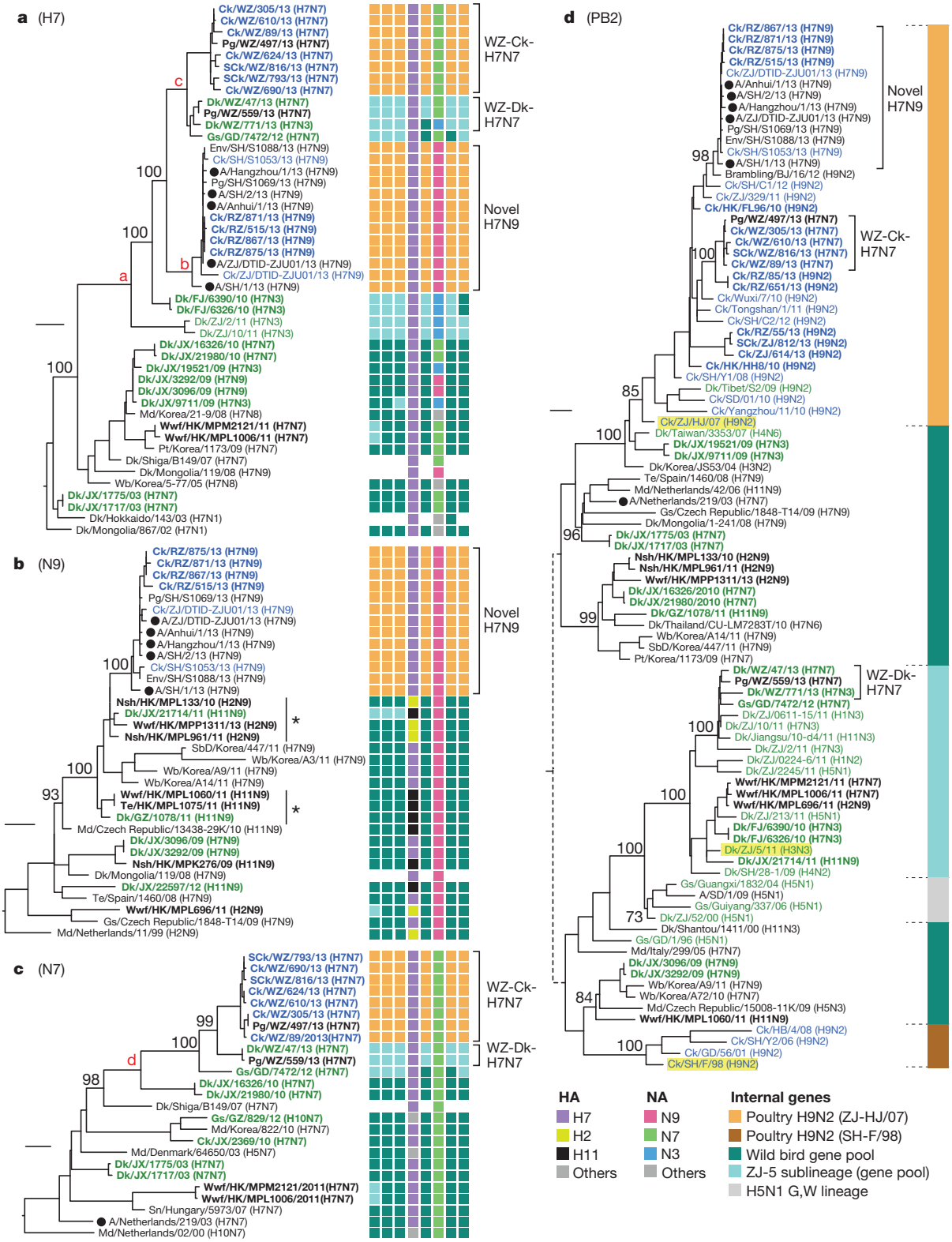

Figure 1 | Phylogenies of haemagglutinin, neuraminidase and PB2 genes. **a**–**d**, Phylogenies of H7 haemagglutinin ($n = 46$) (**a**), N9 neuraminidase ($n = 34$) (**b**), N7 neuraminidase ($n = 25$) (**c**) and PB2 ($n = 93$) (**d**) genes. Sequences reported in this study have their taxa names shown in bold. Genotypes of the influenza viruses are shown on the right (**a**–**c**) as eight coloured blocks representing each gene segment (from left to right: PB2, PB1, polymerase acidic, haemagglutinin, nucleoprotein, neuraminidase, matrix and non-structural; absent if the sequence is unavailable) with the colour indicating the subtype (for haemagglutinin, neuraminidase) or lineage (internal genes; indicated by the solid vertical line in **d**) of that segment. Bootstrap support values (%) from 1,000 pseudo-replicates are shown for selected lineages. Support values for lineages 'a'–'d' were all 100%. The scale bar to the left of each tree represents 0.01 substitutions per site. Asterisks in **b** denote N9 sub-lineages linking the viruses of domestic ducks and wild birds. Host species are: Ck (chicken), Dk (duck), Gs (goose), Md (mallard), Nsh (Northern shoveler), Pg (pigeon), Pt (pintail), SCk (silkie chicken), SbD (spot-billed duck), Te (common teal), Wb (wild bird) and Wwf (wild waterfowl). Geographic locations are: FJ (Fujian), GD (Guangdong), GZ (Guizhou), HK (Hong Kong), JX (Jiangxi), RZ (Rizhao), SD (Shandong), SH (Shanghai), WZ (Wenzhou) and ZJ (Zhejiang). Viruses from different hosts are indicated by: humans, circles; chickens, blue names; domestic ducks or geese, green names.

However, our data show that, for this gene, more closely related H11N9 and H2N9 viruses are found in migratory wild birds in Hong Kong in 2010–2011 (Fig. 1b, asterisk). Phylogenetic linkage between the viruses of these birds and those of domestic ducks in China can be observed at least twice (Fig. 1b, asterisk) before the emergence of the N9 gene in the 2013 outbreak (Fig. 1b and Supplementary Fig. 3).

These data also allow us to reconstruct the genesis of the novel H7N7 lineage closely related to the H7N9 outbreak strain. The N7 of this lineage (Fig. 1c, lineage 'd') arose from H7N7 viruses present in the domestic waterfowl of China since at least 2010 (Fig. 1c and Supplementary Fig. 4). After entering domestic ducks, these early H7N7 viruses and H7N3 viruses (which had internal genes from a sub-lineage, ZJ-5, of the wild bird viral gene pool) co-circulated in these birds in eastern China during 2010–2011 (Fig. 1) and gave rise to H7N7 viruses with ZJ-5 internal genes (WZ-Dk-H7N7; Fig. 1a). H7N7 viruses were also introduced to chickens and by reassortment obtained internal genes from co-circulating H9N2 viruses, generating the H7N7 viruses found in Wenzhou (WZ-Ck-H7N7; Figs 1 and 2).

The H9N2 viruses that contributed internal genes to the H7N9 and H7N7 lineages were formed by reassortment of the major H9N2 lineage in China (SH-F/98) with a Eurasian wild bird virus, from which the former acquired a polymerase basic 2 (PB2) segment, creating a separate sub-lineage (ZJ-HJ/07; Fig. 2 and Supplementary Figs 5–10). The internal genes of the ZJ-HJ/07 sub-lineage form two distinct subgroups (Supplementary Figs 5–10, marked 'α' and 'β'), based on the PB2 segment phylogeny. Reassortment between α and β led to the internal gene cassettes of the H7N9 and H7N7 viruses having an NS gene from α and other internal genes from β (Fig. 2).

Even though the H7N9 and H7N7 chicken isolates obtained their internal gene cassettes from similar H9N2 viruses, all of their six internal genes formed distinct sublineages by subtype (Fig. 1d and Supplementary Figs 5–10). This indicates that the H7N9 and H7N7 viruses resulted from two independent interspecies transmissions to chickens, most likely from domestic ducks, and subsequent reassortment events. Some H7N9 and H7N7 viruses had one or two internal genes that fell outside these sublineages and instead clustered with other co-circulating H9N2 chicken viruses (Supplementary Figs 5–10). Thus, further reassortments with other H9N2 viruses continued to occur in chickens, increasing the genetic diversity of the H7N9 and H7N7 viruses.

Both the H7N9 and H7N7 lineages have three major sources of their eight gene segments (Fig. 2). Although the H7N7 virus suggests a path of two consecutive reassortment events (Fig. 2, yellow arrows), the evidence is not as clear for H7N9 viruses (Fig. 2, purple arrows). Two models are supported for the generation of the H7N9 outbreak lineage by molecular clock dating of divergence times: either a single reassortment event between mid-2011 and mid-2012, or a reassortment between H7 and N9 viruses in 2011 to generate an H7N9 ancestor, probably in ducks, followed by an interspecies transmission and a second reassortment with chicken H9N2 viruses from mid-2011 to late 2012 (Supplementary Fig. 11; substitution rates are shown in Supplementary Table 2). However, alternative reassortant pathways and hosts cannot be definitively excluded due to the limited numbers of H7 and N9 viruses identified during outbreak surveillance, and it is possible for either the H7 or N9 viruses to have first reassorted with H9N2 viruses.

The H7N9 and H7N7 viruses have similar, but independent evolutionary origins. Their haemagglutinin genes originated from H7 viruses that have been introduced to and established among the domestic ducks in China since 2010. At that time, these duck H7 viruses had internal genes from the ZJ-5 sub-lineage, as did H3N3, H1N2 and H11N3 duck viruses found in southern and eastern China during 2009–2013 (Fig. 1). An N9 virus that was a related precursor of the H7N9 viruses (Dk/JX/21714/11; H11N9), had polymerase genes from the ZJ-5 sub-lineage (Fig. 1b, asterisk; Fig. 1d), linking duck N9 viruses to the duck H7N3 precursors of the H7N9 lineage.

The haemagglutinin of the H7N9 outbreak lineage (including viruses isolated from Rizhao, but excluding Shanghai/1/2013 and Ck/ZJ/DTID/ZJU01/2013) had the amino acid substitution Gln235Leu/Ile (position 226 in H3 numbering) indicating that this substitution, which favours α2,6 sialic acid binding, seems to have originated in chickens. The equivalent substitution in chicken H9N2 viruses has been observed since the late 1990s and was linked to virus replication in the upper respiratory tract of birds[6]. The haemagglutinins of the H7N7 lineage from Wenzhou, and their closely related duck viruses (Fig. 1a), do not have this substitution. Of the other amino acid substitutions linked to α2,6 sialic acid binding seen in the H7N9 viruses (Gly195Val, Ala146Ser; 186, 138 in H3 numbering)[4], only Gly195Val is seen in one H7N7 virus (Supplementary Fig. 12). However, two recent studies have shown that, despite these mutations, the H7 haemagglutinin has limited binding to human receptors[8,9]. Both the neuraminidase proteins of the H7N7 and H7N9 chicken viruses have overlapping deletions (20 amino acids in N7, positions 53–72; five amino acids in N9, positions 69–73) in the stalk region, which are often observed in influenza virus lineages established in terrestrial poultry[10]. Duck H7 viruses (N7, N3 and N2) from Wenzhou did not have a stalk deletion. The Glu627Lys and Asp701Asn substitutions in PB2 normally seen in mammalian viruses[11] are only found in human H7N9 isolates (but not in all) and not in the chicken H7N9 and H7N7 viruses obtained here. Although it cannot be excluded, there is no evidence from these data to suggest that

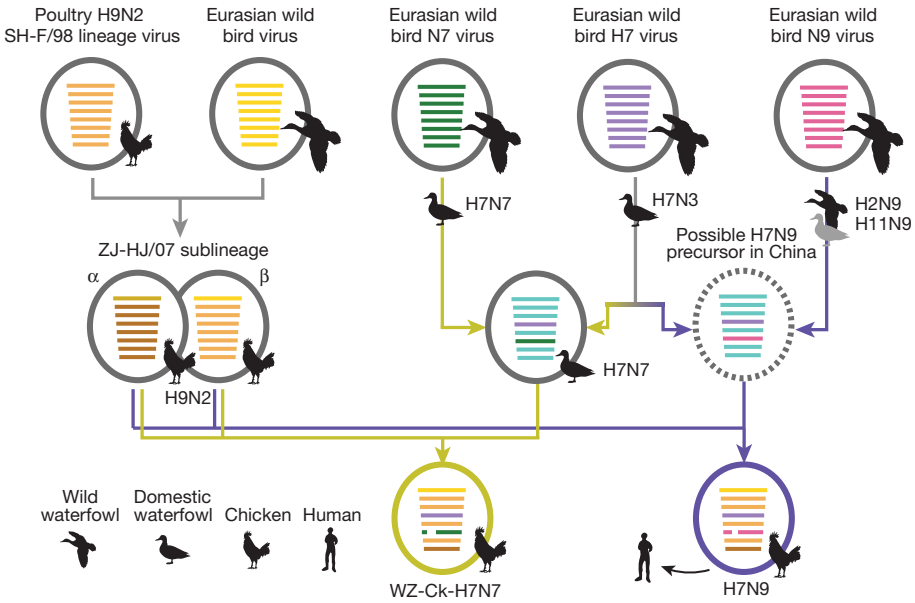

Figure 2 | Evolutionary pathways of the H7N9 and H7N7 viruses. Virus particles are represented by coloured ovals containing horizontal bars that represent the eight gene segments (from top to bottom: PB2, PB1, polymerase acidic, haemagglutinin, nucleoprotein, neuraminidase, matrix and non-structural). Segments in descendent viruses are coloured according to their corresponding source viruses (top) to illustrate gene ancestry through reassortment events. Source viruses for a reassortment are adjacent to arrow tails; arrowheads point to the resulting reassortants. Bars coloured cyan indicate gene segments of the ZJ-5 sub-lineage of wild bird viruses. A broken bar in segment 6 (neuraminidase) indicates a stalk region deletion. The virus indicated by a broken oval represents a hypothetical reassortant.

these viruses are mammalian adapted or that a mammalian intermediate host was involved in the human H7N9 infections.

The chicken H7N7 viruses carry only some of the molecular markers seen in the human H7N9 isolates (Supplementary Fig. 12), but they may still have the potential to infect humans or mammals. To assess infectivity in mammals, two groups ($n = 6$ each) of ferrets were inoculated with Chicken/Wenzhou/610/2013 (H7N7) at $10^{4.5}$ or $10^{6.5}$ plaque-forming units (p.f.u.). Both groups shed virus from 2 days post-inoculation (d.p.i.) (high dose) and 3 d.p.i. (low dose), and virus could be detected in rectal swabs at 4 d.p.i. in both groups (Supplementary Fig. 13). Infectious virus and positive nucleoprotein-stained cells could be detected in the nasal turbinate, and also in the trachea, lungs and hilar lymph nodes, with marginally higher levels from the high dose group at 3 and 5 d.p.i. (Supplementary Figs 13 and 14). This shows that the H7N7 lineage viruses can cause significant infection in mammals under experimental conditions, although virus shedding is lower than those of the 2009 pandemic H1N1 and 2013 human H7N9 viruses[12].

These findings provide a comprehensive picture of the creation and establishment of the H7N9 viruses that have infected humans. Domestic ducks seem to act as key intermediate hosts by acquiring and maintaining diverse influenza viruses from migratory birds, by facilitating the generation of different combinations of H7 and N9 or N7 subtype viruses, and by transmitting these viruses to chickens. After transmission, reassortment with enzootic H9N2 viruses formed the current H7N9 or H7N7 viruses seen in chickens. This probably led to outbreaks in chickens, resulting in the rapid spread of the novel reassortant H7N9 lineage through LPMs, which then became the source of human infections. The cessation of human infections after the closure of LPMs[13], after a precedent set during the Hong Kong H5N1 'bird flu' incident in 1997 (ref. 14), strongly supports this proposition.

The detection of H7N7 chicken viruses in Wenzhou that, like H7N9 viruses, have the potential to infect mammals suggests the current pandemic threat extends beyond the H7N9 virus. Even though human infections with the H7N9 virus seem to be under control, it is too early to know whether this virus has been eradicated from chickens over a larger geographic region. It is possible that H7N9 or H7N7 viruses are still present and may become enzootic in poultry. To control H7N9 and related viruses ultimately, it is necessary to reconsider the management of LPMs in urban areas. Long-term influenza surveillance remains essential for early warning of novel reassortant viruses and interspecies transmission events.

METHODS SUMMARY

Oropharyngeal, cloacal or faecal samples were taken from poultry at LPMs, farms and from wild birds at wetlands in regions flanking the recent H7N9 influenza outbreak. These samples were assessed for the presence of influenza viruses and, with archived H7 and N9 viruses from surveillance in southern China, subjected to full genome sequencing using Sanger or next-generation sequencing methods. After alignment of these sequences with available published sequence data (Supplementary Tables 3–5), maximum likelihood phylogenetic trees were constructed using the GTR+Γ model and topological robustness was assessed[15–17]. Molecular dating was undertaken in a Bayesian Markov chain Monte Carlo framework to calculate divergence times[18]. Ancestral sequences were reconstructed using maximum likelihood methods and codons undergoing positive selection were identified[19–21]. Ferrets were intranasally inoculated with Chicken/Wenzhou/610/2013 (H7N7) at high and low doses. Measurement of virus shedding, titration of virus and examination of pathology in major organs were undertaken as previously described[12].

Full Methods and any associated references are available in the online version of the paper.

Received 22 May; accepted 1 August 2013.
Published online 21 August 2013.

1. Gao, R. *et al.* Human infection with a novel avian-origin influenza A (H7N9) virus. *N. Engl. J. Med.* **368,** 1888–1897 (2013).
2. World Health Organization. Number of confirmed human cases of avian influenza A(H7N9) reported to World Health Organization; http://www.who.int/influenza/human_animal_interface/influenza_h7n9/08_ReportWebH7N9Number.pdf (2013).
3. Liu, D. *et al.* Origin and diversity of novel avian influenza A H7N9 viruses causing human infection: phylogenetic, structural, and coalescent analyses. *Lancet* **381,** 1926–1932 (2013).
4. Kageyama, T. *et al.* Genetic analysis of novel avian A(H7N9) influenza viruses isolated from patients in China, February to April 2013. *Euro Surveill.* **18,** 20453 (2013).
5. Hvistendahl, M., Normile, D. & Cohen, J. Influenza. Despite large research effort, H7N9 continues to baffle. *Science* **340,** 414–415 (2013).
6. Guo, Y. J. *et al.* Characterization of the pathogenicity of members of the newly established H9N2 influenza virus lineages in Asia. *Virology* **267,** 279–288 (2000).
7. Duan, L. *et al.* Influenza virus surveillance in migratory ducks and sentinel ducks at Poyang Lake, China. *Influenza Other Respi. Viruses* **5** (suppl. 1), 65–68 (2011).
8. Tharakaraman, K. *et al.* Glycan receptor binding of the influenza A virus H7N9 hemagglutinin. *Cell* **153,** 1486–1493 (2013).
9. Xiong, X. *et al.* Receptor binding by an H7N9 influenza virus from humans. *Nature* **499,** 496–499 (2013).
10. Cheung, C. L. *et al.* Establishment of influenza A virus (H6N1) in minor poultry species in southern China. *J. Virol.* **81,** 10402–10412 (2007).
11. Yamada, S. *et al.* Biological and structural characterization of a host-adapting amino acid in influenza virus. *PLoS Pathog.* **6,** e1001034 (2010).
12. Zhu, H. *et al.* Infectivity, transmission, and pathology of human-isolated H7N9 influenza virus in ferrets and pigs. *Science* **341,** 183–186 (2013).
13. Xu, J., Lu, S., Wang, H. & Chen, C. Reducing exposure to avian influenza H7N9. *Lancet* **381,** 1815–1816 (2013).
14. Shortridge, K. F. *et al.* Interspecies transmission of influenza viruses: H5N1 virus and a Hong Kong SAR perspective. *Vet. Microbiol.* **74,** 141–147 (2000).
15. Anisimova, M., Gil, M., Dufayard, J. F., Dessimoz, C. & Gascuel, O. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst. Biol.* **60,** 685–699 (2011).
16. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59,** 307–321 (2010).
17. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22,** 2688–2690 (2006).
18. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29,** 1969–1973 (2012).
19. Murrell, B. *et al.* Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **8,** e1002764 (2012).
20. Pond, S. L., Frost, S. D. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21,** 676–679 (2005).
21. Pupko, T., Pe'er, I., Shamir, R. & Graur, D. A fast algorithm for joint reconstruction of ancestral amino acid sequences. *Mol. Biol. Evol.* **17,** 890–896 (2000).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank our colleagues from the Joint Influenza Research Centre (SUMC/HKU) and the State Key Laboratory of Emerging Infectious Diseases for their technical assistance. This study was supported by the National Institutes of Health (National Institute of Allergy and Infectious Diseases contract HSN266200700005C), Li Ka Shing Foundation, the Area of Excellence Scheme of the University Grants Committee of the Hong Kong SAR (grant AoE/M-12/06), Shenzhen Peacock Plan High-End Talents Program (KQTD201203), the University Development Fund (HKU) and the Innovation and Technology Commission of the Hong Kong Government. T.T.-Y.L. was supported in part by a Newton International Fellowship of the Royal Society. Metabiota's involvement was supported by the US Agency for International Development (USAID) Emerging Pandemic Threats Program, PREDICT project, under the terms of Cooperative Agreement Number GHN-A-00-09-00010-00. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no. 278433-PREDEMICS, ERC Grant agreement no. 260864 and the Wellcome Trust (grant 092807) to A.R. and S.J.L.

Author Contributions Y.G., H.Z. and T.T.-Y.L. conceived the study; J.W., Y. Shen, B.Z., L.D., C.Y.-H.L., W.H., Z.O. and X.C. conducted surveillance; H.Z., J.W., C.-L.C., C.M., L.L., Y.C., L.Z., H.L., Y.L., A.F. and D.J.K. performed virus isolation, sequencing and animal experiments; T.T.-Y.L., A.R., O.G.P., H.Z., D.K.S., S.J.L., L.L.M.P., J.S.M.P., G.M.L., Y. Shu, R.G.W, R.J.W. and Y.G. contributed to the analysis; D.K.S. and T.T.-Y.L. wrote the manuscript; Y.G., H.Z., O.G.P. and A.R. edited the manuscript.

Author Information All sequences generated by this study have been deposited in GenBank under accession numbers KF258943–KF260956 and KF297287–KF297322 (Supplementary Table 3). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.Z. (zhuhch@hku.hk) or Y.G. (yguan@hku.hk).

METHODS

Influenza virus surveillance in Zhejiang, Shandong and Guangdong provinces. LPMs, poultry farms and a wetland were surveyed for influenza viruses. If accessible, paired oropharyngeal and cloacal samples were taken from birds, otherwise, samples from isolated individual faecal droppings were collected. Some drinking water samples were also collected. Two cities that flanked the initial outbreak around Shanghai and a more distant city were surveyed (Supplementary Fig. 1), as these locations might have viruses related to the outbreak and/or viruses similar to those that led to the genesis of the outbreak virus. Sampling was conducted immediately before the closure of markets and/or culling of the birds.

Sampling was conducted at LPMs, farms and a wetland in Wenzhou, Zhejiang, a city approximately 500 km south of Shanghai, from 7 to 10 April 2013. Poultry from LPMs and farms were sampled at Rizhao, Shandong, a city approximately 600 km north of Shanghai, on 17 and 18 April 2013. Sampling was also carried out at LPMs in Shenzhen, a major urban centre adjoining Hong Kong and approximately 1,200 km south of Shanghai, on 4, 6 and 7 April 2013. Details of the numbers and types of poultry sampled are given in Supplementary Table 1. After collection, each swab was placed in transport medium (M199) with antibiotics and kept in a cool box before and during shipping to the analysis laboratory. Swab materials were inoculated into 9–10-day-old embryonated chicken eggs and incubated for 48 h at 37 °C. Haemagglutinin-positive isolates were collected and further subtyped by haemagglutination and neuraminidase inhibition (HAI and NAI) assays using a panel of reference antisera as described previously[22]. Standard precautions were taken to avoid cross contamination of samples. Original samples positive for H7 were confirmed with rapid diagnostic quantitative PCR with reverse transcription (qRT–PCR) for the presence of H7 and N9 gene segments (see below). All H7 and H9 influenza isolates from the Wenzhou markets ($n = 78$, Supplementary Table 1), and all H7 isolates ($n = 8$) and 18 of the 51 H9 isolates from the Rizhao markets and farms were selected for full genome sequencing.

Selection of archived samples collected during 2000–2013. Archived H7 ($n = 293$, from 45 sampling occasions) and N9 ($n = 202$, from 27 sampling occasions) viruses isolated during 2000–2013 from our avian influenza surveillance conducted in Fujian, Guangdong, Guangxi, Guizhou, Jiangxi, Yunnan and Hong Kong (Supplementary Fig. 1) were isolated using embryonated eggs and identified by standard HAI and NAI assays as described above. A total of 42 N9 isolates (1 H1, 5 H2, 1 H3, 1 H4, 11 H7, 4 H10 and 19 H11) and a further 57 H7 (1 N1, 26 N3, 4 N6, 23 N7 and 3 N8) isolates were selected for full genome sequencing. For 63 sampling occasions, all H7 and N9 isolates were sequenced. For two occasions from Fujian, 2 out of 5 and 3 out of 6 isolates were sequenced. For six occasions from Jiangxi, 3 out of 5, 3 out of 7, 4 out of 10, 6 out of 11, 3 out of 33 and 10 out of 171 were sequenced. This was based on at least one virus from each HAI and NAI titre group (that is, those that had similar values from these assays). A further 98 H9N2 viruses isolated from Hong Kong retail poultry markets were sequenced.

Genomic sequencing. All selected isolates had all eight segments sequenced either using Sanger sequencing on an ABI 3730 genetic analyser (Applied Biosystems) or high-throughput next-generation sequencing on a Genome Sequencer Junior (Roche). Where a sample contained more than one copy of any segment (indicating a mixed infection of two viruses), only sequences of the haemagglutinin and neuraminidase sequences that could be explicitly defined or have been confirmed by single clonal Sanger sequencing were used in subsequent phylogenetic analyses.

Rapid diagnostic qRT–PCR. For rapid determination of the presence of any H7 and N9 viruses, primers specific for these subtypes were designed, based on the alignment of all H7 and N9 gene sequences available from GenBank on 1 April 2013. Viral RNA was extracted using the QIAamp Viral RNA Minikit (QIAGEN), and complementary DNAs were synthesized with primer Uni12 using the Transcriptor High Fidelity cDNA Synthesis Kit (Roche). Copy numbers of the H7 haemagglutinin or the N9 neuraminidase segment were determined on a LightCycler 480 Real-Time PCR System (Roche) using LightCycler FastStart DNA MasterPLUS SYBR Green I Kit (Roche). Primers were H7-765 (5′-GTTTCAA TGGGGCHTTCATAGC-3′), H7-995R (5′-ACATTCTTCATCCCTGTWGC-3′), N9-160 (5′-CAAGCCAAACAATAATAAACAA-3′) and N9-293R (5′-CATGAAT TTATAGTACAGAGYCCT-3′).

Sequence collection and alignment. All previously published influenza A virus sequences were collated from GenBank on 5 April 2013. Seven full genome sequences of H7N9 influenza viruses from the outbreak in China were downloaded from GISAID (http://gisaid.org) and two from GenBank (http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html). Acknowledgement of the sources of the GISAID sequences is given in Supplementary Table 4. These database sequences were combined with the sequences generated in this study (Supplementary Table 3) for further analysis. Sequences of each gene segment were aligned using MUSCLE v3.5 (ref. 23) with manual adjustment.

Phylogenetic analysis. Preliminary global phylogenetic analysis of all sequences was done using RAxML v7.6.8 (ref. 17). This identified the major Eurasian lineage that contains the H7N9 human isolates and the H7N7 and H7N9 isolates sequenced here. The major Eurasian lineage was reduced to approximately 1,000 sequences. Care was taken to ensure all sequences closely related to the H7N9 outbreak isolates and the novel H7N7 viruses were retained (Supplementary Table 5). Maximum likelihood phylogenies were reconstructed using the GTR+Γ nucleotide substitution model in PhyML v3.0 (ref. 16). Topological robustness was assessed by a Shimodaira–Hasegawa approximate likelihood ratio (SH-like) branch test[15]. The phylogenies for H7, N9, N7, PB2, PB1, polymerase acidic, nucleoprotein, matrix and non-structural segments are shown in Supplementary Figs 2–10. Smaller trees, with only essential taxa are presented in Fig. 1 with bootstrap support values from 1,000 pseudo-replicates to indicate the confidence of the topology.

Molecular dating. The evolutionary timescales of the H7N7 and H7N9 phylogenies were inferred by a molecular clock model that was calibrated using the virus sampling dates. Gene sequences of the H7N9 and H7N7 viruses were analysed with a selected set of closely related viruses. A relaxed clock model with uncorrelated lognormal distribution[24], tree topologies and other evolutionary parameters were jointly estimated using the Bayesian Markov chain Monte Carlo method implemented in BEAST v1.7.5 (ref. 18). The SRD06 nucleotide substitution model was used[25]. A Bayesian skyride model with time-aware smoothing was used[26] and a uniform prior and random walk operator was assigned to virus taxa whose sampling dates were known only to the nearest year. Multiple MCMC trajectories were computed and combined, giving from 2×10^8 to 10×10^8 total steps for each data set, with sampling every 1,000 steps. Convergence, at effective sample sizes >200, of relevant parameters was assessed in Tracer v1.5 (http://tree.bio.ed.ac.uk/software/tracer/). Maximum clade credibility trees with branches scaled by time were summarized, and the time of divergence and time of most recent common ancestor of the H7N9 and H7N7 viruses were obtained as illustrated in Supplementary Fig. 11. All data sets were screened to exclude mosaic sequences, as previously described[27], before molecular clock and further evolutionary analyses were undertaken.

Genome substitutions. The ancestral nucleotide sequence at each internal node of the maximum likelihood tree was reconstructed by the maximum likelihood joint method[21] implemented in HYPHY v2.0 (ref. 20). The sequences at both ends of a tree branch were compared, and the differences represent the substitutions that occurred along that branch. Substitutions in the positively selected codons identified by the mixed effects model of evolution[19] in Datamonkey[28] are highlighted in Supplementary Fig. 12.

Animal infection experiments. Four- to five-month-old female Angora ferrets (*Mustela putorius furo*) were obtained through a laboratory ferret breeding program at the Wuxi Sangosho and confirmed to be influenza virus free by virus isolation in Madin–Darby canine kidney cells (ATCC) from nasal washes and rectal swabs and sero-negative by HAI assay against contemporary swine and human influenza viruses (swine: H1 and H3; human: seasonal H1N1, H3N2 and influenza B), and avian H5N1 and H9N2 that are enzootic in China, and the chicken H7N7 virus to be tested in this study. A minimum number of animals necessary to obtain reproducible results were used according to the ethical guidelines. Ferrets were randomly allocated across the treatment groups, to which investigators were not blinded.

A group of six ferrets were intranasally inoculated with a dose of $10^{6.5}$ p.f.u. (or $10^{6.8}$ 50% tissue culture infectious dose ($TCID_{50}$)) of the Chicken/Wenzhou/610/2013 (H7N7) virus (high dose group) and a second group were inoculated with $10^{4.5}$ p.f.u. (or $10^{4.8}$ $TCID_{50}$) of viruses (low dose group). Nasal and rectal washes were taken daily from each individual and titrated using standard $TCID_{50}$ assays. At 3 and 5 d.p.i., three ferrets from each group were euthanized and tissues from each major organ (pulmonary lobes, nasal turbinate, upper and lower trachea and hilar lymph nodes) were collected for virus titration ($TCID_{50}$ assays), RNA extraction and qRT–PCR, and studies of tissue pathology (haematoxylin and eosin and viral nucleoprotein staining), following protocols described previously[12]. All animal experiment protocols were reviewed and approved by the Institutional Ethical Review Board (IERB) of Shantou University Medical College (ref no. SUMC2013-111) and The University of Hong Kong Committee on the Use of Live Animals in Teaching & Research. All experiments with H7N7 and H7N9 viruses were conducted in biosafety level 3 (BSL3) facilities, using enhanced BSL3 practices for the animal work and following practices in the approved institutional guidelines.

22. Huang, K. *et al.* Establishment and lineage replacement of H6 influenza viruses in domestic ducks in southern China. *J. Virol.* **86,** 6075–6083 (2012).
23. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).
24. Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4,** e88 (2006).
25. Shapiro, B., Rambaut, A. & Drummond, A. J. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol. Biol. Evol.* **23,** 7–9 (2006).
26. Minin, V. N., Bloomquist, E. W. & Suchard, M. A. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol. Biol. Evol.* **25,** 1459–1471 (2008).

27. Lam, T. T. *et al.* Systematic phylogenetic analysis of influenza A virus reveals many novel mosaic genome segments. *Infect. Genet. Evol.* **18,** 367–378 (2013).

28. Delport, W., Poon, A. F., Frost, S. D. & Kosakovsky Pond, S. L. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **26,** 2455–2457 (2010).