Vaccine immune pressure influences viral population complexity of avian influenza virus during infection

Adelaide Milani[a], Alice Fusaro[a], Francesco Bonfante[a], Gianpiero Zamperin[a], Annalisa Salviato[a], Marzia Mancin[a], Eleonora Mastrorilli[a], Joseph Hughes[e], Hussein A. Hussein[b], Magdi Hassan[c], Egbert Mundt[f], Calogero Terregino[a], Giovanni Cattoli[a,d], Isabella Monne[a]

[a]Istituto Zooprofilattico Sperimentale delle Venezie, Legnaro (Padova), Italy; [b]Virology department, Faculty of Veterinary Medicine, Cairo University, Giza, Egypt; [c]IFT Corporation, Cairo, Egypt; [d]Animal Production and Health Laboratory, Joint FAO/IAEA Division, International Atomic Energy Agency, Seibersdorf, Austria; [e]MRC-University of Glasgow Center for Virus Research, Glasgow, United Kingdom; [f]Boehringer Ingelheim Veterinary Research Center, Hannover, 30559, Germany

Corresponding author: Adelaide Milani, Division of Comparative Biomedical Sciences, Istituto Zooprofilattico Sperimentale delle Venezie (IZSVe), Viale dell'Università 10, 35020 Legnaro (PD), Italy

E-mail: amilani@izsvenezie.it

Tel. 049-8084368

Fax. 049-8084360

Abstract

Vaccines are useful tools to control influenza A virus infection in poultry, but they need to be periodically reformulated to guarantee appropriate protection from infection and to limit viral replication and circulation, which could favour the emergence of new variants. In this study, a deep sequencing approach was used to characterize and follow the evolution of the hemagglutinin of the H5N1 highly pathogenic avian influenza viral population in infected animals vaccinated with two

vaccines conferring different protection levels. Results from this preliminary investigation suggested that the evolution of the viral population, as well as the abundance and heterogeneity of minority variants could be influenced by the immune pressure conferred by vaccination.

Keywords

Influenza A virus; H5N1; vaccine; deep sequencing; evolution.

Introduction

Influenza A virus is a zoonotic agent with a significant impact on both public health and poultry industry. Vaccination is a useful tool used worldwide to support intervention strategies, such as stamping out and biosecurity policies, in order to keep the infection under control and prevent the transmission of avian influenza viruses in poultry (Lee and Suarez, 2005). However, as demonstrated in previous studies, the use of a vaccine strain antigenically different from the circulating viruses or application of inadequate vaccine protocols may favour the antigenic drift and cause vaccination failure (Cattoli et al., 2011b; Lee et al., 2004; Swayne, 2012). A more extensive knowledge of the mechanisms underlying intra-host evolution of avian influenza viruses circulating in vaccinated poultry populations could be of help to formulate and adopt more adequate vaccine strategies.

Previous studies conducted in partially immune pigs indicated that the variability in immune response may influence the overall diversity of swine influenza virus during infection (Diaz et al., 2015) and showed that the hemagglutinin gene displayed nucleotide mutations at the very beginning of viral infection (Diaz et al., 2013; Murcia et al., 2012). Still, to date there is no information on the intra-host evolution of highly pathogenic avian influenza (HPAI) viruses circulating in vaccinated poultry populations. The hemagglutinin (HA) is a surface glycoprotein, involved in the induction of a protective humoral and cell mediated immune response, and represents one of the major antigenic determinants of type A influenza viruses.

To provide some preliminary data on the impact of vaccination on the intra-host diversity and evolution of HPAI viruses, in this study we performed a deep sequencing analysis of swabs sampled from H5N1 HPAI experimentally infected chickens, which showed different levels of clinical and virological protection conferred by two different vaccine formulations.

Material and methods

A deep sequencing analysis on the HA gene segment was performed on samples collected in a previously performed vaccination/challenge study for assessing the protective efficacy of two avian influenza vaccines against a HPAI H5N1 virus. Birds were housed into HEPA filtered poultry isolators in BSL3 animal facilities and received food and water *ad libitum*. All animals were handled in strict accordance with the Decree of the Italian Ministry of Health (n. 26 of 4 March 2014) on the protection of animals used for scientific purposes, implementing Directive 2010/63/EU.

Briefly, two groups of ten Specific Pathogen Free (SPF) day-old chicks were vaccinated twice at a 10-day interval by the sub-cutaneous route, using two distinct influenza inactivated experimental vaccines. The first vaccination was performed with 0.3 ml and the second vaccination was performed with 0.5 ml via the subcutaneous route in the back of the neck. Both vaccines contained each a recombinant H5-antigen which was encoded by recombinant baculoviruses. Generation of the recombinant baculoviruses and formulation of the experimental vaccines was performed as described recently (Oliveira et al, 2016) with one exception, no inactivated antigen of Newcastle disease virus was added, and 512 HA units were added as H5 antigen. One recombinant baculovirus encoded for a H5-protein which is included in the VOLVAC® B.E.S.T AI + ND KV (A) belonging to clade 2.3.2. The second recombinant baculovirus (B) encoded for a consensus sequence (Patent WO 2012/054907 A2; Sequence ID 43) which was generated based on 37 amino acid sequences of Egyptian origin viruses from 2010 belonging to either clade 2.2.1 or clade 2.2.1.1.

The birds were challenged with 10^6 50% Embryo Infectious Dose (EID_{50}) of the HPAI H5N1 A/chicken/Egypt/11VIR4453-7/VRLCU/2010 virus (clade 2.2.1) (WHO/OIE/FAO H5N1 Evolution Working Group, 2012) 21 days after the booster vaccination. The virus had been isolated from pooled organs in 9- to 11-day-old SPF embryonated fowls' eggs. Antibody responses were assessed by means of hemagglutination inhibition assay (HI), 10 days from the first vaccination, 21 days from the second vaccination and 2 weeks after the challenge. Tracheal swabs (TS) and cloacal swabs (CS) were collected on days 2, 4, 6, 8 and 10 post challenge (p.c.) to evaluate viral shedding by quantitative real-time RT-PCR (qRRT-PCR) targeting the M gene (Spackman et al., 2002) and calculate the EID_{50} equivalents from the cycle threshold (Ct) values. The Egyptian HPAI H5N1 virus used for the challenge, as well as 20 TS positive by qRRT-PCR, which contain a sufficient amount of RNA (six samples from group A and fourteen samples from group B), were processed as described below. Neither of the samples obtained from CS did result in a sufficient signal from the qRRT-PCR to be included in the analysis. Total RNA was isolated using Nucleospin RNA kit (Macherey-Nagel, Düren, Germany). Viral RNA encoding the HA gene segment was retro-transcribed and amplified using SuperScript III one-step reverse transcription-PCR (RT-PCR) system with PlatinumTaq High Fidelity (Invitrogen, Carlsbad, CA) using H5 specific primers (H5-for: 5'-CRAAAGCAGGGGTYCAATC-3', H5-rev: 5'-GAAACAAGGGTGTTTTTAAC-3') . Sequencing libraries were prepared using Nextera XT DNA Sample preparation kit (Illumina) and processed as described by Monne et al., 2014 on an Illumina Miseq desktop sequencer. Raw data were submitted to the NCBI Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/Traces/sra/) under accession numbers SRR4244068, SRR4244069, SRR4244070, SRR4244071, SRR4244072, SRR4244073, SRR4244074, SRR4244076, SRR4244077, SRR4244078, SRR4244079, SRR4244080, SRR4244081, SRR4244082, SRR4244083, SRR4244084, SRR4244085, SRR4244086, SRR4244087, SRR4244088, SRR4244089.

FASTQC software was used to inspect quality score of raw sequence files and post processing data coming from the high-throughput sequencing pipelines. Fastq files were cleaned with Trimmomatic v0.32 (Bolger et al., 2014), using a 4-base-pair sliding-window algorithm with a quality score cut-off of 20; only reads longer than 80 nucleotides were considered and mapped to the hemagglutinin H5 reference sequence A/chicken/Egypt/11VIR4453-7/VRLCU/2010 (Gisaid accession number EPI348162) using bwa v0.7.5 (Li and Durbin, 2010). The alignment was processed with LoFreq v2.1.2 (Wilm et al., 2012) for the SNP calling; according to LoFreq usage recommendations, the alignment was first processed with Picard-tools v2.1.0 (http://broadinstitute.github.io/picard/) and GATK v3.5 (DePristo et al., 2011; McKenna et al., 2010; Van der Auwera et al., 2013). The frequencies of variants in the final set were computed using the diversiTools program v0.1.19 (http://josephhughes.github.io/btctools/) and confirmed by those found with LoFreq. Variants with sequence coverage less than 500X or frequency less than 1% were discarded. The numbering scheme of the amino acid residues used in this study is based on the mature sequence (signal peptide was cleaved off) of the HA of A/Vietnam/1203/2004 (H5N1) (GenBank accession number HM006759).

To explore the amino acid variability at the resulted polymorphic sites in the Egyptian H5N1 viral population, the full-length hemagglutinin sequences of H5N1 Egyptian viruses (from 2006 to 2016) were downloaded from the public database (GenBank) collapsing identical ones; a total of 345 protein sequences was obtained. The sequences were then aligned using the on-line MAFFT program (http://mafft.cbrc.jp/alignment/server/) and amino acid composition at the selected positions was carefully inspected.

From our NGS data we then calculated the Shannon Entropy (SE). SE is a measure of diversity within a viral population at a single nucleotide position or a whole sequence level. In a specific site, the highest value of SE is reached when the four nucleotides are present at a frequency of 25%, while the lowest value (zero) is obtained when only one nucleotide is present at a frequency of 100%. Low entropy may be observed in case of a significant reduction of the population size, i.e.,

following a bottleneck event, which at a single nucleotide level results in low or no variability. In our study Shannon Entropy was used to quantify average diversity of the HA gene of the viral populations of each sample belonging to group A and B, using the following formula:

$$E = -\frac{1}{N}\sum_{i=1}^{N}(f_{iA}\ln f_{iA} + f_{iG}\ln f_{iG} + f_{iT}\ln f_{iT} + f_{iC}\ln f_{iC})$$

where f_i is the frequency of the nucleotide A, C, G or T at position i and N is the total length of the hemagglutinin gene.

The Wilcoxon Mann–Whitney rank-sum test was used to verify whether the distribution of EID_{50} and Entropy values was the same in both vaccine groups; only six samples were collected at 4 days p.c., five from group B and one from group A, therefore they were excluded from the statistical comparison. The same test was used to assess whether the distributions of Entropy were identical between the challenge virus and each sample of groups A and B. P-value < 0.10 was considered significant. The Pearson product-moment correlation coefficient was used to verify whether the population diversity (represented by SE) and the quantity of virus (represented by EID50) were independent so as to exclude that differences of SE between group A and B were due to EID50 values.

Results

To provide some preliminary data on the impact of vaccination on the intra-host diversity and evolution of HPAI viruses, in this study we performed a deep sequencing analysis of swabs sampled from SPF chickens immunized with two different vaccines that, upon H5N1 HPAI challenge exhibited different levels of clinical and virological protection, expressed as reduction of mortality and shedding, respectively. All of the birds that received vaccine A survived the challenge (100% protection from mortality) and only 2/10 birds showed moderate depression that in one animal was coupled with nervous signs. Vaccine B prevented mortality in 70% of the birds (Tab. 1) and all of the surviving animals showed moderate signs of inactivity and depression. Moreover, the

vaccines differed in terms of suppression of viral shedding, as at each time p.c. fewer birds in group A had shed viral RNA from the trachea (Tab. 1), and the amount of virus shed was significantly lower than in group B on day 2 p.c. (p=0.072) (Fig. 1). Ten days after the first vaccination, animals in both groups showed no detectable levels of HI antibodies against the challenge virus. After the booster vaccination, birds in group A recorded a 2.6 log2 HI geometric mean titre (GMT) against the challenge virus, whereas in group B all birds resulted either negative or recorded HI titres of 1 log2 (GMT of 0.2 log2). After the challenge, seroconversion, expressed as an HI GMT increase equal to or higher than 2 log2, was observed in all of the surviving birds in group B (GMT of 2.2 log2), as opposed to 50% of the animals in group A (HI GMT of 3.0 log2).

A sufficient amount of amplification product (>0.2 ng/µl) for deep sequencing analysis was recovered from TS only on days 2 and 4 p.c.. Data were obtained for a total of twenty-one positive samples, specifically: a) the challenge virus , b) six samples belonging to group A, five of which at day 2 p.c. (34A2, 35A2, 37A2, 47A2, 59A2) and one at day 4 p.c. (34A4), and c) fourteen samples belonging to group B, nine of which at day 2 p.c. (72B2, 73B2, 75B2, 79B2, 80B2, 81B2, 83B2, 86B2, 88B2) and five at day 4 p.c. (73B4, 79B4, 81B4, 86B4, 88B4). Each sample was identified as follows: the first two digits refer to the animal identification code, the alphabetic character (A or B) identifies the group and the last digit indicates the number of days p.c.

To characterize the complexity of the viral population of the 20 swab samples from the vaccinated birds, the per-site Shannon Entropy was calculated, considering the frequencies of nucleotide substitutions across the hemagglutinin gene. The Entropy measures fluctuated considerably: the samples with the lowest values belonged to group A (0 and 2.3×10^{-4}), while the ones with the highest values (7.2×10^{-4} and 7.7×10^{-4}) belonged to group B.

The non-parametric Wilcoxon Mann–Whitney test indicated that the distribution of Entropy values were different in the two vaccination groups (p=0.0041), with group B showing the highest values (fig 1). Compared to the challenge virus, four out of five samples of group A collected at day 2 p.c.

showed significantly different Entropy values (p-values range from 0.0142 to 0.0587). These four samples had a lower Entropy value than the one of the challenge virus (data not shown). All samples belonging to group B recorded Entropy values without significant differences compared to the Entropy of the challenge virus. Pearson test between the within-host virus diversity and the viral shedding (EID_{50}) showed no significant correlation between Entropy and EID_{50} values (r=0.23 with p-value=0.4297), suggesting that the different Entropy values observed in the viruses of group A may have been due to the occurrence of bottleneck events as a consequence of the vaccine selective pressure.

The analysis of the nucleotide sequence diversity of the hemagglutinin gene showed several synonymous and non-synonymous polymorphisms distributed on the HA gene of almost all samples. However, a comparison between the two groups revealed a great variability in the number of polymorphisms among samples. Specifically, the six samples belonging to group A showed from zero to four minority variants per sample (tab. 2), randomly distributed across nine nucleotide positions, with a frequency ranging from 1.12% to 6.88%. None of the identified polymorphic sites were shared among samples. On the other hand, group B displayed a higher number of polymorphisms (tab. 3), from two to thirteen per sample, distributed in fifty-seven positions and showing a frequency ranging from 1.02% to 68.70%. Interestingly, three of these polymorphisms (residues 258, 470 and 1379) were independently acquired by two or more samples of group B, while none of the minority variants was shared between group A and B.

In addition, we found that most of the identified mutations had been newly acquired during the viral replication in host. Indeed, only three minority variants identified at the nucleotide positions 164, 1032 and 1395 of the HA gene of seven group B samples were already present in the challenge strain material, while none of the polymorphisms observed in the challenge virus material were later identified among samples belonging to group A, thus suggesting an even narrower bottleneck.

Non-synonymous substitutions represented 67% of the total polymorphisms both in group A and B and were randomly distributed across the HA gene. Among samples belonging to group A, six

nucleotide positions, across the whole HA gene, showed non-synonymous minority variants with a frequency ranging from 1.12% to 6.88%. None of them were located at the globular head of the HA1 protein near or within the three secondary structural elements of the receptor binding domain (RBD) (Kovácová et al., 2002), formed by the 130-loop, 190-helix, and 220-loop, and/or in previously identified antigenic sites (Kaverin et al., 2007, 2002). On the other hand, group B showed a higher number of nucleotide positions (forty-two) displaying non-synonymous mutations with a frequency ranging from 1.02% to 38.84% (tab. 3). Interestingly, we identified two non-synonymous SNPs at positions 1018 and 1019 responsible of the mutations R325K and R325G situated at the HA cleavage site for two samples, collected respectively at 2 and 4 days p.c. In addition, ten of the polymorphisms acquired by six out of nine samples at 2 days p.c. and by one out of five samples at 4 days p.c. fell within the receptor binding cavity of the HA. In particular, C135F and S142Y/F were positioned respectively within and close to the 130 loop, whereas K218E was located within the 220 loop. Moreover, five of these minority variants fell close to or within antigenic sites previously characterized in the H5 strains: site 1 (S1) and site 2 (S2) (Kaverin et al., 2007, 2002). In particular, three non-synonymous SNPs (C135F, S142Y and S142F), found in one sample at 2 days p.c. and in one sample at 4 days p.c., flanked the site 1 (136-141), which corresponds to antigenic site A in H3 and Ca2 of H1, while the two minority variants T151I and Y157C, detected in sample 79B2 and 80B2, were respectively located within and close to site 2 (residues 151-156), which corresponds to antigenic site B in H3. Considering the distinct results obtained from the two experimental groups, the percentage of amino acid identity between the two vaccines and the challenge virus was calculated. The sequence of the HA protein (except for the signal peptide) of the challenge virus showed a percentage of amino acid identity with vaccine A and B of 96.4% and 95.3%, respectively. Moreover, the identity of the sequence coding for the HA receptor binding pocket (from amino acid 130 to 225, H5 numbering) of the challenge virus with vaccine A and B corresponded to 93.6% and 85.1%, respectively. These findings may suggest that

the suboptimal protection conferred by vaccine B seems to have favoured the occurrence of a higher number of molecular changes in the challenge virus.

Two samples belonging to group B (73B2 and 79B4) displayed stop codons at amino acid positions 122 in sample 73B2 and 149 and 518 in sample 79B4. This observation could suggest the formation of defective virus particles as previously shown (Hoelzer et al., 2010; Murcia et al., 2012).

To explore whether the amino acid polymorphisms identified in this study had emerged also in field strains circulating under vaccination pressure, we analysed all the non-redundant hemagglutinin sequences of H5N1 viruses available in GenBank that had been collected in Egypt between 2006 and 2016. We found at least one sequence displaying the same mutation identified in our samples at the amino acid positions 121, 151, 312, 324, 325, 369, 373, 393 and 528. Positions 151 and 325 showed the highest variability; in particular, amino acid position 151 showed 51% T and 62% I, whereas position 325 showed 52% R, 29% K and 12% G. Furthermore position 324 showed 97.3% E and 2.3% K, whereas position 373 showed 96.5% K, and 3.5% R.

Discussion

For many influenza subtypes, such as HPAI H5N1 virus, vaccination programmes are currently implemented in a number of countries with the goal to control this disease in poultry populations (Swayne, 2012). However, influenza A viruses evolve rapidly in response to selection pressures generated through vaccine protection (Cattoli et al., 2011a), and the emergence of virus strains, for which existing vaccines are not well matched and offer little protection, continuously challenges the effectiveness of vaccines in the field (Connie Leung et al., 2013; Kim et al., 2010). To better understand the dynamics of the emergence of antigenic drifted variants, we performed a deep sequencing analysis of 20 samples collected in the framework of a previous experimental study conducted in our institute to assess the protection efficacy of two distinct vaccines against a HPAI H5N1 virus. In particular, the two experimental challenge groups (A and B) allowed to mimic

different level of immunity and to explore how viruses evolve within hosts that have developed suboptimal immunity with influenza vaccines. All of the birds that had received vaccine A survived the challenge, whereas vaccine B prevented mortality in 70% of the birds. The results obtained applying a deep sequencing approach to samples collected from the groups A and B and to the challenge virus suggests that a suboptimal level of antibody protection may favour the increase of viral population heterogeneity from the early stages of infection and may promote the selection of minority variants, some of which may be involved in antigenic drift.

We identified a total of 9 minority variants (frequency range from 1.12% to 6.88%) in group A and 64 minority variants (frequency range from 1.05% to 38.84%) and 2 fixed mutations (frequencies 54.44% and 68.70%) in group B. Although no marked differences were highlighted between the two groups in the ratio of synonymous to non-synonymous substitutions (67% of SNP identified in both groups were non-synonymous), evidence of non-synonymous mutations in the RBD or in the antigenic sites (S1 and S2) were found only in group B. However, whether these non-synonymous polymorphisms had emerged by chance or as a consequence of strong selection could not be assessed. None of the samples taken at day 4 p.c. showed a fixation of non-synonymous substitutions indicating they may be transient, or else there may have been insufficient time for the minority variant to become fixed in the viral population, since mutations that alter HA antigenicity may engender a cost in the viral fitness, which requires compensatory mutations to restore viral replicative ability. Interestingly, mutation T151I located in the RBD and antigenic site 2 is present in 62% of the Egyptian H5N1 sequences available in GenBank, meaning that this mutation may be naturally selected during circulation of H5N1 viruses in the field. In addition, two samples belonging to group B displayed minority variants R325K and R325G situated within the cleavage site of the hemagglutinin protein. A previous study carried out on Egyptian HPAI H5N1 viruses showed that the amino acid glycine at position 325 had significantly reduced pathogenicity without altering the transmission efficiency (Yoon et al., 2013). Both mutations were observed in the H5N1

viruses collected from poultry in Egypt. Specifically, R325K was observed in the recent viruses from clade 2.2.1.2, while R325G characterized the Egyptian clade 2.2.1.1a (Arafa et al., 2016). Considering that our sample size was rather limited and that no samples from non-vaccinated control birds were suitable for sequencing, and since all of the animals had died soon after the HPAI H5N1 infection, this study should be further confirmed by making an assessment on a greater number of samples collected over a wider range of time. In addition, further studies on the whole influenza A virus genome could provide an overview on the effect of vaccine immune pressure on the evolution of the viral genome and, consequently, on viral proteins. However, the identification of some of the minority variants in field viruses circulating in a partially immune population may indicate that in natural conditions these mutations can arise and be selected, providing a fitness advantage to the viruses.

To our knowledge, this is the first time that we are provided with evidence showing that infection of properly vaccinated birds is dominated by narrow genetic bottlenecks, resulting in a reduction of genetic variation. On the other hand, the infection of partially immune poultry appears to allow the virus to increase its genetic heterogeneity and to accumulate mutations that may alter HA antigenicity. Furthermore, vaccine A seems to have played a bottleneck effect as four out of five animals at day 2 p.c. had showed significantly lower Entropy values compared to the challenge virus. Overall our findings are in accordance with the results obtained in infected and vaccinated pigs (Diaz et al., 2015), where the authors suggested that the viral population diversity and evolution might be influenced by the level of immune response to influenza A virus vaccine. Results from this study provide proof of concept data that pave the way for further investigations on the effect of vaccine immune pressure on the generation of genetic diversity at the subpopulation level.

Conclusion

Deep sequencing analysis proved to be a valid tool to explore and characterize differences among viral populations in vaccinated animals during infection. This technique detected the presence of minority variants that could not have been revealed by a classical sequencing method. This work highlights the potential value of the NGS approach for the assessment of vaccine efficacy and as a suitable method to understand the mechanisms that underpin how viruses escape vaccine protection. The deep sequencing analysis of samples collected from vaccine challenge studies early during the infection might serve as a predictive tool to infer virus population heterogeneity and hence effectiveness of vaccine control programmes.

Acknowledgments

This work was financially supported by the European projects Epi-SEQ (research project supported under the 2nd Joint Call for Transnational Research Projects by EMIDA ERA-NET [FP7 project no. 219235]), by the NoFlu project, Fondazione Cariplo Vaccine Program (grant number 2009-3594) and by Behringer Ingelheim Vetmedica. The authors would like to acknowledge Francesca Ellero for providing help with the language. This study was conducted in the framework of the Doctoral school in Bioscience and Biotechnology at the University of Padua (Adelaide Milani).

References

Arafa, A., El-Masry, I., Kholosy, S., Hassan, M.K., Dauphin, G., Lubroth, J., Makonnen, Y.J., 2016. Phylodynamics of avian influenza clade 2.2.1 H5N1 viruses in Egypt. Virol. J. 13, 49. doi:10.1186/s12985-016-0477-7

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–20. doi:10.1093/bioinformatics/btu170

Cattoli, G., Fusaro, A., Monne, I., Coven, F., Joannis, T., El-Hamid, H.S.A., Hussein, A.A., Cornelius, C., Amarin, N.M., Mancin, M., Holmes, E.C., Capua, I., 2011a. Evidence for differing evolutionary dynamics of A/H5N1 viruses among countries applying or not applying

avian influenza vaccination in poultry. Vaccine 29, 9368–75. doi:10.1016/j.vaccine.2011.09.127

Cattoli, G., Milani, A., Temperton, N., Zecchin, B., Buratin, A., Molesti, E., Aly, M.M., Arafa, A., Capua, I., 2011b. Antigenic drift in H5N1 avian influenza virus in poultry is driven by mutations in major antigenic sites of the hemagglutinin molecule analogous to those for human influenza virus. J. Virol. 85, 8718–24. doi:10.1128/JVI.02403-10

Connie Leung, Y.H., Luk, G., Sia, S.-F., Wu, Y.-O., Ho, C.-K., Chow, K.-C., Tang, S.-C., Guan, Y., Malik Peiris, J.S., 2013. Experimental challenge of chicken vaccinated with commercially available H5 vaccines reveals loss of protection to some highly pathogenic avian influenza H5N1 strains circulating in Hong Kong/China. Vaccine 31, 3536–42. doi:10.1016/j.vaccine.2013.05.076

DePristo, M.A., Banks, E., Poplin, R., Garimella, K. V, Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–8. doi:10.1038/ng.806

Diaz, A., Allerson, M., Culhane, M., Sreevatsan, S., Torremorell, M., 2013. Antigenic drift of H1N1 influenza A virus in pigs with and without passive immunity. Influenza Other Respi. Viruses 7 Suppl 4, 52–60. doi:10.1111/irv.12190

Diaz, A., Enomoto, S., Romagosa, A., Sreevatsan, S., Nelson, M., Culhane, M., Torremorell, M., 2015. Genome plasticity of triple-reassortant H1N1 influenza A virus during infection of vaccinated pigs. J. Gen. Virol. 96, 2982–93. doi:10.1099/jgv.0.000258

Hoelzer, K., Murcia, P.R., Baillie, G.J., Wood, J.L.N., Metzger, S.M., Osterrieder, N., Dubovi, E.J., Holmes, E.C., Parrish, C.R., 2010. Intrahost evolutionary dynamics of canine influenza virus in naive and partially immune dogs. J. Virol. 84, 5329–35. doi:10.1128/JVI.02469-09

Kaverin, N. V, Rudneva, I.A., Govorkova, E.A., Timofeeva, T.A., Shilov, A.A., Kochergin-

Nikitsky, K.S., Krylov, P.S., Webster, R.G., 2007. Epitope mapping of the hemagglutinin molecule of a highly pathogenic H5N1 influenza virus by using monoclonal antibodies. J. Virol. 81, 12911–7. doi:10.1128/JVI.01522-07

Kaverin, N. V, Rudneva, I.A., Ilyushina, N.A., Varich, N.L., Lipatov, A.S., Smirnov, Y.A., Govorkova, E.A., Gitelman, A.K., Lvov, D.K., Webster, R.G., 2002. Structure of antigenic sites on the haemagglutinin molecule of H5 avian influenza virus and phenotypic variation of escape mutants. J. Gen. Virol. 83, 2497–505. doi:10.1099/0022-1317-83-10-2497

Kim, J.-K., Kayali, G., Walker, D., Forrest, H.L., Ellebedy, A.H., Griffin, Y.S., Rubrum, A., Bahgat, M.M., Kutkat, M.A., Ali, M.A.A., Aldridge, J.R., Negovetich, N.J., Krauss, S., Webby, R.J., Webster, R.G., 2010. Puzzling inefficiency of H5N1 influenza vaccines in Egyptian poultry. Proc. Natl. Acad. Sci. U. S. A. 107, 11044–9. doi:10.1073/pnas.1006419107

Kovácová, A., Ruttkay-Nedecký, G., Haverlík, I.K., Janecek, S., 2002. Sequence similarities and evolutionary relationships of influenza virus A hemagglutinins. Virus Genes 24, 57–63.

Lee, C.-W., Senne, D. a, Suarez, D.L., 2004. Effect of vaccine use in the evolution of Mexican lineage H5N2 avian influenza virus. J. Virol. 78, 8372–8381. doi:10.1128/JVI.78.15.8372-8381.2004

Lee, C.-W., Suarez, D.L., 2005. Avian influenza virus: prospects for prevention and control by vaccination. Anim. Health Res. Rev. 6, 1–15.

Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–95. doi:10.1093/bioinformatics/btp698

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–303. doi:10.1101/gr.107524.110

Monne, I., Fusaro, A., Nelson, M.I., Bonfanti, L., Mulatti, P., Hughes, J., Murcia, P.R., Schivo, A., Valastro, V., Moreno, A., Holmes, E.C., Cattoli, G., 2014. Emergence of a highly pathogenic

avian influenza virus from a low-pathogenic progenitor. J. Virol. 88, 4375–88. doi:10.1128/JVI.03181-13

Murcia, P.R., Hughes, J., Battista, P., Lloyd, L., Baillie, G.J., Ramirez-Gonzalez, R.H., Ormond, D., Oliver, K., Elton, D., Mumford, J.A., Caccamo, M., Kellam, P., Grenfell, B.T., Holmes, E.C., Wood, J.L.N., 2012. Evolution of an Eurasian avian-like influenza virus in naïve and vaccinated pigs. PLoS Pathog. 8, e1002730. doi:10.1371/journal.ppat.1002730

Spackman, E., Senne, D.A., Myers, T.J., Bulaga, L.L., Garber, L.P., Perdue, M.L., Lohman, K., Daum, L.T., Suarez, D.L., 2002. Development of a Real-Time Reverse Transcriptase PCR Assay for Type A Influenza Virus and the Avian H5 and H7 Hemagglutinin Subtypes. J. Clin. Microbiol. 40, 3256–3260. doi:10.1128/JCM.40.9.3256-3260.2002

Swayne, D.E., 2012. Impact of vaccines and vaccination on global control of avian influenza. Avian Dis. 56, 818–28. doi:10.1637/10183-041012-Review.1

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V, Altshuler, D., Gabriel, S., DePristo, M.A., 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatics 43, 11.10.1-33. doi:10.1002/0471250953.bi1110s43

WHO/OIE/FAO H5N1 Evolution Working Group, 2012. Continued evolution of highly pathogenic avian influenza A (H5N1): updated nomenclature. Influenza Other Respi. Viruses 6, 1–5. doi:10.1111/j.1750-2659.2011.00298.x

Wilm, A., Aw, P.P.K., Bertrand, D., Yeo, G.H.T., Ong, S.H., Wong, C.H., Khor, C.C., Petric, R., Hibberd, M.L., Nagarajan, N., 2012. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. 40, 11189–201. doi:10.1093/nar/gks918

Yoon, S.-W., Kayali, G., Ali, M.A., Webster, R.G., Webby, R.J., Ducatez, M.F., 2013. A single amino acid at the hemagglutinin cleavage site contributes to the pathogenicity but not the

transmission of Egyptian highly pathogenic H5N1 influenza virus in chickens. J. Virol. 87, 4786–8. doi:10.1128/JVI.03551-12

Oliveira Cavalcanti, M., Vaughn, E., Capua, I., Cattoli, G., Terregino, C., Harder, T., Grund, C., Vega, C., Robles, F., Franco, J., Darji, A., Arafa, A. S., Mundt E., 2017. A genetically engineered H5 protein expressed in insect cells confers protection against different clades of H5N1 highly pathogenic avian influenza viruses in chickens. Avian Pathol. Jan 27:1-10. doi: 10.1080/03079457.2016.1250866.

Figure legend

Figure 1. Box plots of EID50 and Entropy by vaccination group. The boxplots display the distribution of the quantitative values of EID50 (left) and Entropy (right) for samples collected at 2 days p.c. of each vaccination group. The y-axis represents the EID and Entropy values. Each box shows the degree of dispersion in the data. The line inside the box is the median value; the points above the box are the outliers.

Table 1. Number of positive swabs (viral shedding) and survival rates of groups A and B.

	Number of positive swabs										Survival rate %
Days p.c.	2		4		6		8		10		
	TS[b]	CS[c]	TS	CS	TS	CS	TS	CS	TS	CS	
Vaccine A[a]	8/10[d]	0/10	5/10	0/10	1/10	0/10	0/10	0/10	0/10	1/10	100
Vaccine B[a]	10/10	0/10	8/10	1/10	2/8	1/8	1/8	2/8	1/7	0/7	70
Control	NA[e]	NA	NA	NA	NA	NA	NA	NA	NA	NA	0

[a] Vaccine A (B.E.S.T AI) and B (Consensus) were used for a two-fold vaccination at day 1 and 10 post hatch

[b] Tracheal swab

[c] Cloacal swab

[d] Number of chickens positive in qRRT-PCR/total number of chickens

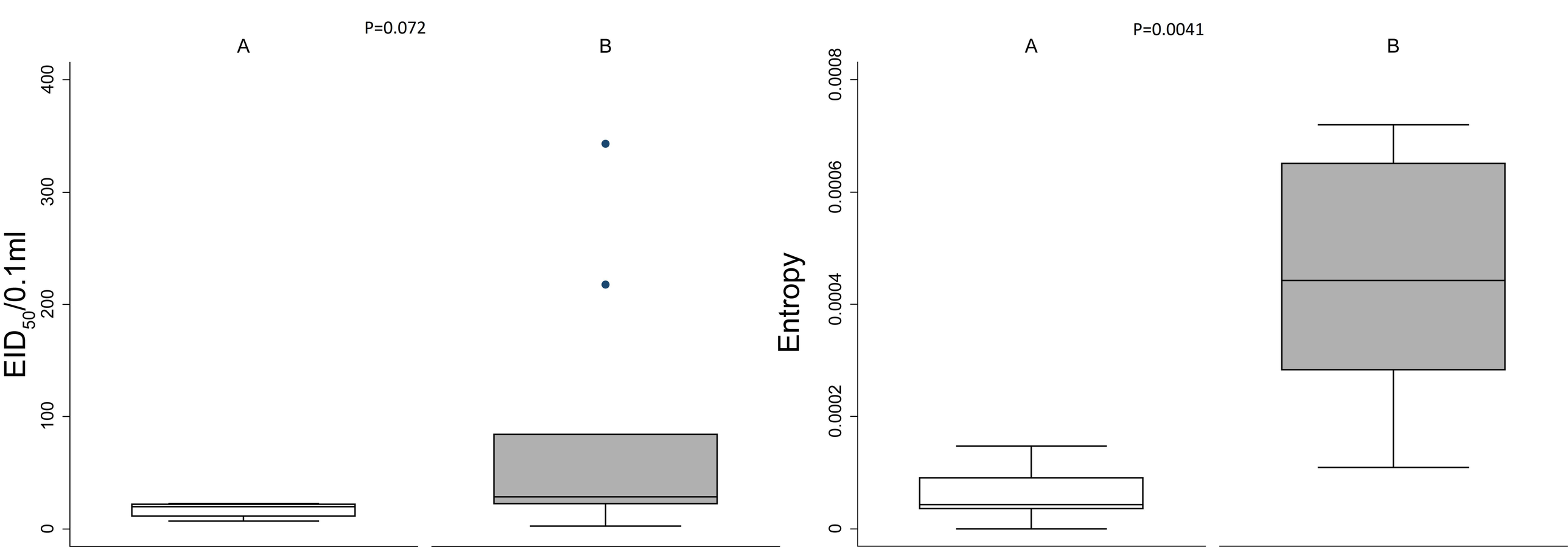
[e] Not applicable since all birds of the control group died on day 1 post challenge infection

Table 2. SNPs identified in samples belonging to group A and virus of challenge; frequency values are in percentage. SIL: silent mutation, CDS: coding sequence, POLYM: polymorphism at aa position (signal peptide is not included in the numbering).

CDS	POLYM	Chall	34A2	35A2	37A2	47A2	59A2	34A4
164	**G39E**	5.06						
229	**L61F**		1.34					
238	**N64D**		1.54					
489	**SIL**							1.13
849	**SIL**	1.26						
1032	**SIL**	1.64						
1190	**V382A**				1.12			
1222	**F393L**					6.88		
1236	**SIL**							2.01
1290	**SIL**	1.64						
1395	**SIL**	1.07						
1400	**V452A**							2.72
1615	**SIL**	2.32						
1625	**L527P**							2.35
1701	**SIL**			1.37				

Table 3. SNPs identified in samples belonging to group B and virus of challenge; frequency values are in percentage. S1: antigenic site S1; S2: antigenic site S2; 130L: 130 loop; 220L: 220 loop. The box represents the residues included in the Receptor Binding Domain; SIL: silent mutation, CDS: coding sequence, POLYM: polymorphism at amino acid position (signal peptide is not included in the numbering).

CDS	POLYM	Site	Chall	72B2	73B2	75B2	79B2	80B2	81B2	83B2	86B2	88B2	73B4	79B4	81B4	86B4	88B4
149	E34G												7.93				
153	***SIL***																2.86
160	N38D								1.52								
164	G39E		5.06							38.84							
213	***SIL***												1.59				
258	***SIL***				2.38							16.68					32.74
268	***SIL***																2.72
357	H103Q				2.30												
410	S121F															3.49	
411	***SIL***				1.44												
414	W122*				1.49												
449	C135F	S1 130L			2.28												
458	***SIL***			1.25													
470	S142Y/F	S1			1.65											3.88	
492	W149*													1.75			
497	T151I	S2					4.46										
515	Y157C	S2						3.00									
536	Y164C										5.35						
555	E170D								5.29								
585	***SIL***															2.89	
682	I213V											1.50					
686	A214D						3.77										
697	K218E	220L														1.05	
804	***SIL***														1.74		
828	***SIL***							1.39									
848	E268G									3.56							
849	***SIL***		1.26														
880	Q279K										3.92						
931	P296S						5.16										
979	V312I				1.30												
985	A314T													1.98			
1015	E324K									2.66							
1018	R325G														1.44		
1019	R325K				1.33												
1032	***SIL***		1.64		1.58	68.70		54.44					6.74	1.25			15.16
1048	A335T				1.27												
1071	***SIL***									1.66							
1107	***SIL***													1.20			
1119	N358D				1.40												
1150	E369K						10.51										
1163	K373R													1.02			
1199	I385T										2.26						
1264	I407L													1.29			
1290	***SIL*** /E415D		1.64		1.42												
1327	L428F				2.13												
1349	E435G					2.92											
1364	F440S														2.71		
1365	**SIL**													1.84			
1379	V445A						5.67		1.54								
1394	D450G									3.96							
1395	***SIL***		1.07										6.62				
1429	E462K													1.61			
1504	Y487H								1.22								
1517	Q491L						2.50										
1522	S493P			2.53													
1575	***SIL***															2.67	
1598	S518*													1.05			
1615	***SIL***		2.32														
1628	A528V											1.25					


P=0.072
A
B
EID50/0.1ml
0
100
200
300
400
P=0.0041
A
B
Entropy
0
0.0002
0.0004
0.0006
0.0008