Journal of General Virology (2015), **96**, 2982–2993 **DOI** 10.1099/jgv.0.000258

Genome plasticity of triple-reassortant H1N1 influenza A virus during infection of vaccinated pigs

Andres Diaz,[1] Shinichiro Enomoto,[2] Anna Romagosa,†[1] Srinand Sreevatsan,[1] Martha Nelson,[3] Marie Culhane[1] and Montserrat Torremorell[1]

Correspondence
Montserrat Torremorell
torr003@umn.edu

[1]College of Veterinary Medicine, University of Minnesota Saint Paul, Minnesota, USA
[2]Department of Biology University of Utah, South Lake City, Utah, USA
[3]Fogarty International Center, National Institutes of Health, Bethesda, Maryland, USA

To gain insight into the evolution of influenza A viruses (IAVs) during infection of vaccinated pigs, we experimentally infected a 3-week-old naive pig with a triple-reassortant H1N1 IAV and placed the seeder pig in direct contact with a group of age-matched vaccinated pigs (n=10). We indexed the genetic diversity and evolution of the virus at an intra-host level by deep sequencing the entire genome directly from nasal swabs collected at two separate samplings during infection. We obtained 13 IAV metagenomes from 13 samples, which included the virus inoculum and two samples from each of the six pigs that tested positive for IAV during the study. The infection produced a population of heterogeneous alleles (sequence variants) that was dynamic over time. Overall, 794 polymorphisms were identified amongst all samples, which yielded 327 alleles, 214 of which were unique sequences. A total of 43 distinct haemagglutinin proteins were translated, two of which were observed in multiple pigs, whereas the neuraminidase (NA) was conserved and only one dominant NA was found throughout the study. The genetic diversity of IAVs changed dynamically within and between pigs. However, most of the substitutions observed in the internal gene segments were synonymous. Our results demonstrated remarkable IAV diversity, and the complex, rapid and dynamic evolution of IAV during infection of vaccinated pigs that can only be appreciated with repeated sampling of individual animals and deep sequence analysis.

Received 26 May 2015
Accepted 4 August 2015

INTRODUCTION

Influenza A viruses (IAVs) are distributed globally and can infect a wide range of host species, including humans (Russell *et al.*, 2008), birds (Chen & Holmes, 2010), pigs (Vincent *et al.*, 2008), horses (Murcia *et al.*, 2010), dogs (Hoelzer *et al.*, 2010), cats (Ali *et al.*, 2011) and seals (Blanc *et al.*, 2009). Wild waterfowl are considered the natural IAV reservoir (Taubenberger & Kash, 2010) and a genetically distinct lineage of viruses has also been identified in bats (Tong *et al.*, 2013). A swine-origin H1N1 IAV was responsible for the first pandemic of the twenty-first century (Rambaut & Holmes, 2009) and was associated with >200 000 human deaths (Simonsen *et al.*, 2013). In recent decades, the genetic diversity of swine IAVs in North America has increased significantly due to the emergence of triple-reassortant H3N2 viruses in the late 1990s (Kitikoon *et al.*, 2013), the numerous introductions of human-origin viruses in pigs including the 2009 pandemic virus (Nelson *et al.*, 2012) and the large-scale movement of pigs between different US regions (Nelson *et al.*, 2011). It is estimated that >90 % of swine herds in the mid-western US are infected with IAVs (Corzo *et al.*, 2013) and that pigs can be exposed to different IAVs during their lifetime (Diaz *et al.*, 2015; Ducatez *et al.*, 2011).

IAVs belongs to the family *Orthomyxoviridae* and have a segmented genome composed of eight negative-sense ssRNA segments that encode at least 12 proteins: polymerase basic 2 (PB2), polymerase basic 1 (PB1), polymerase acid (PA), haemagglutinin (HA), nucleoprotein (NP), neuraminidase (NA), matrix (M) and non-structural protein (NS). RNA viruses have a high mutation rate that increases their genetic diversity over time (Belshaw *et al.*, 2007; Domingo *et al.*, 2012; Lemey *et al.*, 2006) and the segmented nature of the IAV genome allows the virus to

†Present address: Genus/PIC Europe, Sant Cugat del Vallés, Barcelona, Spain.

The GenBank/EMBL/DDBJ accession numbers for the influenza A virus new sequencing data are KM198687–KM198694.

Three supplementary tables are available with the online Supplementary Material.

exchange (reassort) gene segments with other IAVs, thus contributing to the overall genetic diversity of IAVs.

The main antigenic proteins of the virus, HA and NA, determine IAV subtype. In pigs, H1N1, H1N2 and H3N2 are the most prevalent IAV subtypes (Torremorell *et al.*, 2012). In North American swine there are six antigenically and phylogenetically distinct H1 groups (α, β, γ1, γ2, δ1 and δ2) (Anderson *et al.*, 2015; Lorusso *et al.*, 2011) and four H3 groups (I, II, III and IV) (Kitikoon *et al.*, 2013). Multiple IAV subtypes can co-circulate in swine herds and persist at the population level for prolonged periods of time (Corzo *et al.*, 2013; Diaz *et al.*, 2015). Additionally, multiple alleles (sequence variants of the same virus) can coexist during IAV infection of pigs (Murcia *et al.*, 2012), and the same virus can evolve differently in the upper and lower respiratory tract of pigs (Wei *et al.*, 2014). Furthermore, nucleotide substitutions within the HA antigenic sites can occur shortly after infection of pigs with no significant differences noticed between pigs with or without immunity to the virus (Diaz *et al.*, 2013; Murcia *et al.*, 2012).

In the USA, most pigs may be exposed to one or more IAVs during their lifetime. Hence, the majority of pigs may have immunity to one or several IAVs strains when they are challenged with currently circulating IAVs. However, there is a lack of knowledge about how the virus evolves in swine populations that are seropositive to different IAVs. Therefore, the objective of this study was to explore the genetic diversity of the complete genome of a triple-reassortant H1N1 IAV population during experimental infection of vaccinated pigs. A vaccine with multiple IAV strains was used to mimic field conditions where pigs are usually exposed to different IAVs. We identified several polymorphisms using next-generation sequencing (NGS) technologies directly from nasal swabs and reconstructed 13 complete genomes of the within-host viral populations (metagenomes) using polymorphism overlapping sequence fragment analysis, demonstrating the complex, rapid and dynamic evolution of IAV during infection of vaccinated pigs.

RESULTS

Animal IAV infection status before and after contact with the seeder pig

To study the genomic plasticity of a triple-reassortant H1N1 IAV during infection of vaccinated pigs, we obtained 11 pigs free of IAV, vaccinated 10 and infected one to serve as a seeder. All pigs (n=11) were IAV-negative by real-time reverse transcription-PCR (RRT-PCR) and seronegative by NP ELISA prior to vaccination. Two weeks after the booster vaccination, and before contact with the seeder pig, nine pigs tested ELISA-positive to IAV [sample-to-negative ratio (S/N)$<$0.6] and one was a suspect (S/N=0.803). Three pigs were negative [HA inhibition (HI) titre $<$1 : 20] and seven had HI titres $\leq$1 : 40 to the challenge virus. All vaccinated pigs had HI titres $\geq$1 : 20 to the vaccine viruses (Table 1).

The seeder pig remained negative before challenge and tested IAV-positive by RRT-PCR 48 h after challenge. After introduction of the seeder pig into the isolation unit with the remaining pigs (n=10), five animals tested IAV-positive and five tested negative by RRT-PCR during the study period (Table 2). Prior to the introduction of the seeder pig, the mean S/N ELISA titre was not statistically different (P=0.75) between pigs that tested positive and negative by RRT-PCR. However, 14 days after the introduction of the seeder pig the S/N ELISA titre was lower (P=0.04) in pigs that tested RRT-PCR-positive compared with pigs that tested RRT-PCR-negative (Table 1). Additionally, we also found statistically significant differences ($P<0.05$) in the HI titres before and after contact to the seeder pig, between pigs that tested positive and pigs that tested negative by RRT-PCR (Table 1).

Extensive allelic variation was identified by sequence analysis during infection

The complete genome of IAV was amplified and sequenced from 13 samples, which included the inoculum virus before challenge, two samples from the seeder pig at days 2 and 4 (SD2 and SD4), and two samples from each of the five infected pigs after contact (A1D5, A1D7, A2D4, A2D8, A3D5, A3D6, A4D3, A4D6, A5D4 and A5D6, where 'A' refers to the animal number and 'D' refers to the day of study). The most frequent allele in the inoculum virus for segments 1, 4, 5, 7 and 8 was also the most frequent allele in the pig samples analysed. In contrast, the most frequent allele for segments 2, 3 and 6 was different in pig samples compared with the inoculum virus. Many nucleotide polymorphisms (n=794) were found in all samples throughout the course of the study and were distributed in all gene segments. However, there was great variability in the number of polymorphisms between samples and gene segments (Table S1, available in the online Supplementary Material). The overlapping sequence fragments analysis estimated a total of 327 alleles, of which 214 were unique sequences (Table 3).

Three of the 41 original polymorphisms present in the inoculum virus were not identified in any of the pig samples analysed and not all alleles in the inoculum were identified in the pig samples sequenced. Moreover, only four emergent alleles (defined as alleles not present in the inoculum virus) were found in multiple pigs (two in segment 2, and one each in segments 3 and 4). Finally, for all but the HA segment, the crude ratio of synonymous to non-synonymous substitutions (dS/dN) was $>$1 (Table 3).

Although there was small or no variation in the number of alleles detected between the two samples sequenced of most pigs that tested positive, the allele frequency changed significantly within animals 1 and 4 (Table 3). We noticed that prior to contact with the seeder pig, these two pigs (animals 1 and 4) were negative by HI to the challenge virus and had the lowest HI titre (1 : 80) to A/Swine/Iowa/110600/

Table 1. IAV serology results by ELISA and HI tests for pigs prior to the start of the study (before vaccination), after vaccination and after infection

For simplicity, animals that tested IAV RRT-PCR-positive after exposure to the seeder pig (n=5) are renamed A1–A5 and animals that tested negative are renamed A6–A10. ELISA results are expressed as S/N and considered positive, suspect or negative when S/N<0.6, 0.6<S/N<0.9 or S/N>0.9, respectively. HI titres are expressed as the reciprocal dilutions and considered positive at >1 : 20. The reciprocal mean titre was compared between pigs that tested RRT-PCR-positive and -negative after exposure to the seeder pig and considered statistically significant if the P value for the Kruskal–Wallis test was <0.05.

Animal	PCR	Before vaccination ELISA	Two weeks after second vaccination (prior contact with the seeder pig)					Two weeks after infection			
			ELISA	HI test*			ELISA	HI test*			
				00239	031	110600	MO069		00239	031	110600	MO069
Seeder	Positive	0.967	0.869	Neg†	Neg	Neg	Neg	0.32	80	160	0	40
A1	Positive	0.963	0.503	Neg	20	80	160	0.194	320	320	80	160
A2	Positive	1.700	0.803	20	20	160	320	0.129	640	640	80	320
A3	Positive	0.979	0.331	Neg	40	160	320	0.135	160	160	80	320
A4	Positive	0.925	0.332	Neg	20	80	320	0.202	160	320	80	320
A5	Positive	1.009	0.362	20	40	320	640	0.164	640	640	320	640
Mean		*1.115*	*0.466*	*10*	*28*	*160*	*352*	*0.165*	*384*	*416*	*128*	*352*
A6	Negative	0.958	0.546	40	40	320	640	0.434	20	80	320	640
A7	Negative	0.968	0.187	20	40	320	640	0.219	40	80	320	640
A8	Negative	1.018	0.545	20	20	160	320	0.247	20	40	80	320
A9	Negative	0.968	0.430	40	20	640	640	0.275	40	80	160	640
A10	Negative	0.981	0.193	20	40	320	640	0.162	20	80	80	320
Mean		*0.979*	*0.380*	*28*	*32*	*352*	*576*	*0.2674*	*28*	*72*	*192*	*512*
P value (Kruskal–Wallis test)		0.91	0.75	0.03	0.55	0.048	0.06	0.04	0.03	0.007	0.28	0.16

*IAV isolates used in the HI test: 00239 (challenge virus), A/Swine/IA/00239/2004(H1N1); 031 (vaccine virus), A/Swine/NorthCarolina/031/2005(H1N1); 110600 (vaccine virus), A/Swine/Iowa/110600/2000(H1N1); MO069 (vaccine virus), A/Swine/Missouri/069/2005(H3N2). †: Neg=Negative.

Table 2. IAV RRT-PCR results and samples selected for deep genome sequencing

Nasal swabs were collected and tested for 14 days post-contact. No pig tested positive after 8 days post-contact. Results from five pigs that did not test positive at any point during this study are not shown. The days when pigs tested positive are indicated by '+' and the RRT-PCR C_t value for each sample is shown. A total of 12 pig samples were selected for complete genome sequencing: seeder pig, day 2 (SD2, dark blue) and day 4 (SD4, light blue); animal A1, day 5 (A1D5, dark green) and day 7 (A1D7, light green), animal A2, day 4 (A2D4, dark purple) and day 8 (A2D8, light purple); animal A3, day 5 (A3D5, dark brown) and day 6 (A3D6, light brown); animal A4, day 3 (A4D3, dark grey), and day 6 (A4D6, light grey); animal A5, day 4 (A5D4, dark pink) and day 6 (A5D6, light pink). Sequences obtained from each sample are colour coded in Fig. 1 according to the colours indicated in this table.

Animal	Time tested (days post-contact)									
	1	2	3	4	5	6	7	8	9	10
Seeder	+27.2	+26.0	+27.9	+33.4						
A1				+30.3	+27.3	+30.2	+33.6			
A2			+31.0	+33.1	+29.1	+31.2		+31.8		
A3				+27.6	+27.8	+29.9	+29.5	+34.3		
A4			+32.9	+30.8	+27.6	+29.2	+33.4			
A5			+30.6	+30.7	+28.2	+29.1	+34.2			

2000(H1N1), which was the vaccine virus closest to the challenge virus at the nucleotide level (Table 1).

HA and NA antigenic proteins diverge independently during infection

Whilst 44 different alleles were found in HA, only three alleles were found in NA during this study (Fig. 1). The 44 HA alleles identified (Fig. 1a) yielded 43 different predicted peptides (Table 4). The starting virus inoculum contained two HA alleles that only differed in 1 nt within the HA2 region. In the seeder pig on day 2 (SD2), we found eight HA alleles (Fig. 1a) with variations in the amino acid sequence within both the HA1 and HA2 regions (Table 4). However, the emergent alleles of SD2 were not found in any other pig samples analysed. All the other HA emergent alleles were unique to an animal except for one that was detected in two pigs (A1D5 and A2D8); however, the latter emergent variant (A1D5/A2D8, highlighted in Fig. 1a and Table 4) contained a HA1 region identical to the inoculum alleles. Additionally, we found 32 HA alleles in A1D7; at the nucleotide level, half of these 32 alleles were closer to an allele identified in A4D6, whilst the other half were closer to an inoculum allele (Fig. 1a). At the protein level, these 32 variants contained polymorphisms in all three regions of the HA, signal peptide, HA1, and HA2. Overall, amino acid substitutions within HA1 were only found in alleles obtained from SD2 and A1D7 (Table 4); four of these substitutions happened within antigen sites previously described for HA subtype H1 (Table 4), and their location and nature are illustrated in Fig. 2.

Table 3. Number of alleles distributed by sample and gene segment

The last column of the table indicates the ratio between synonymous mutations (d*S*) and non-synonymous mutations (d*N*) for each gene segment. In total, 214 out of 327 alleles found were unique sequences.

NA, No d*N* mutations were found in segment 6, hence no d*S*/d*N* ratio was estimated.

Segment	Inoculum	SD2	SD4	A1D5	A1D7	A2D4	A2D8	A3D5	A3D6	A4D3	A4D6	A5D4	A5D6	Unique sequences	dS/dN
1 (PB2)	4	1	2	2	0	1	2	1	1	4	2	1	1	11	15.7
2 (PB1)	2	2	1	2	1	2	2	2	2	1	1	1	1	5	3.7
3 (PA)	8	2	2	4	0	2	1	2	0	64	2	2	2	77	12.5
4 (HA)	2	8	2	2	32	2	2	1	1	1	2	2	1	44	0.8
5 (NP)	16	2	2	2	8	2	2	2	8	4	2	2	2	31	10.9
6 (NA)	2	1	1	1	2	1	1	1	1	1	2	1	1	3	NA
7 (M)	2	2	1	1	4	1	1	1	1	1	1	1	1	7	41.0
8 (NS)	2	1	1	1	32	1	1	1	1	4	1	1	1	36	6.1
Total	38	19	12	15	79	12	12	11	15	80	13	11	10	214	

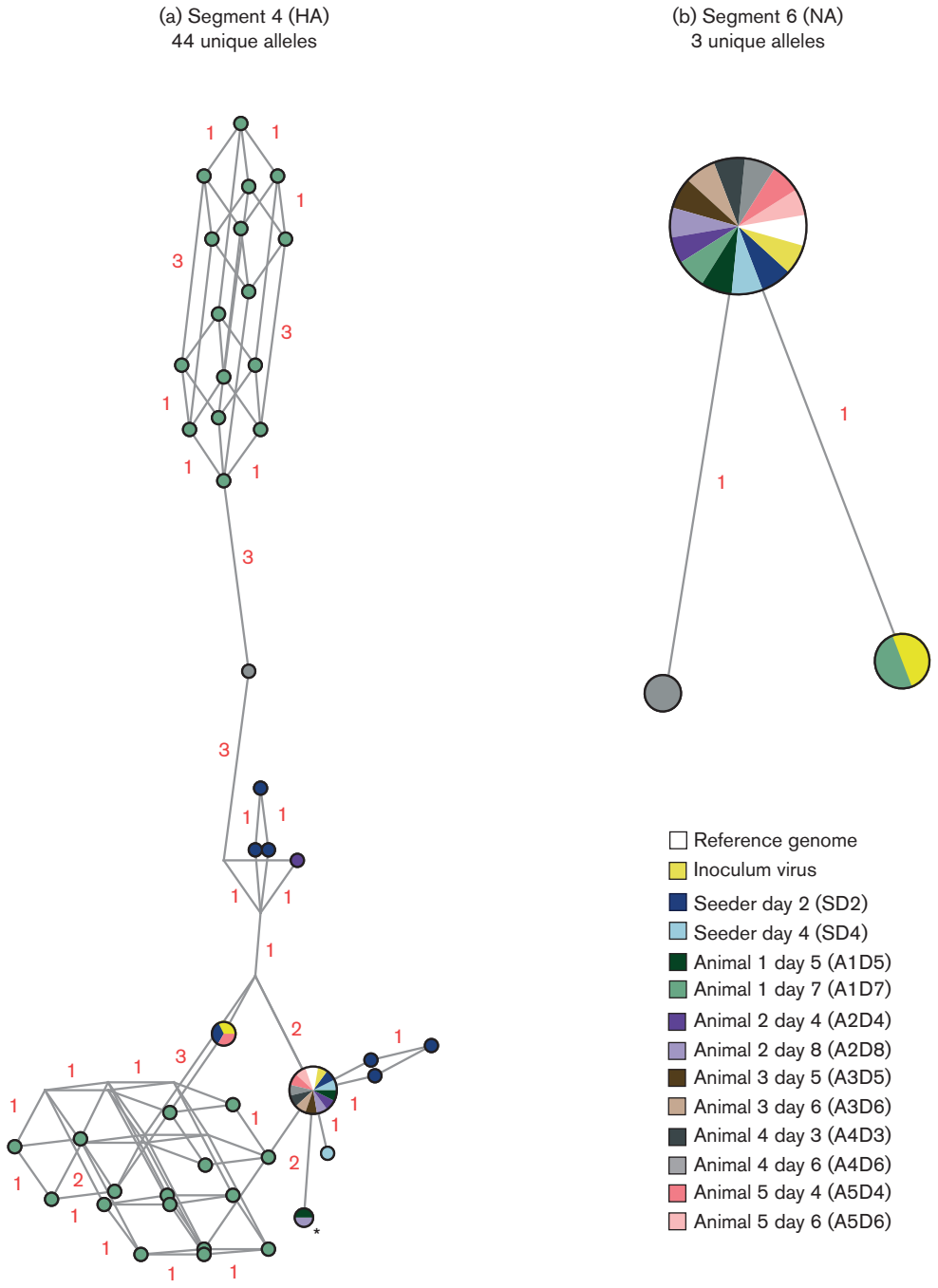

Fig. 1. Median-joining networks of HA and NA alleles found during experimental IAV infection of vaccinated pigs. Each circle represents a sequence variant (allele) and each colour represents the sample where that sequence was found. Red numbers indicate the number of nucleotide differences between sequences (not all numbers are included for brevity). Within each network, the branch length is proportional to the number of differences between alleles. An asterisk indicates an emergent HA allele (not present in the inoculum virus) that was found in more than one pig. The colour code of this figure coincides with the colour code of Table 2.

In contrast, at the NA level only three alleles were found during this study and all of them translated the same NA protein. Two of these alleles were present in the starting inoculum virus. The majority inoculum allele was not detected in most samples sequenced (except in A1D7) and the minority allele became fixed in most of the pig

Table 4. CLUSTAL_X alignment of the complete hypothetical HA proteins found by sample

Only polymorphic sites amongst alleles identified in this study are shown. Superscripts A, B and D indicate the antigenic site where changes in HA1 were observed. The first two rows indicate amino acids found in the vaccine viruses at the polymorphic sites identified in the samples sequenced. The reference amino acid for each polymorphic position is shown in inoculum allele 1. Non-highlighted proteins are unique variants and proteins highlighted with the same shading are 100 % identical amongst them. A total of 58 functional HA sequences were identified amongst all samples. These sequences represented 44 unique alleles and translated 43 different hypothetical HAs. SP, signal peptide; IAV031, A/Swine/NorthCarolina/031/2005(H1N1); IAV110600, A/Swine/Iowa/110600/2000(H1N1). NA, Not applicable [the complete sequence for A/Swine/NorthCarolina/031/2005(H1N1) is not available].

		Protein region and amino acid position											
		Sig_pep	HA1						HA2				
Sample ID	Allele	17	25	29	159^{A}	180^{D}	200^{B}	252^{D}	369	370	371	372	373
IAV031	NA	A	H	S	S	F	N	P	NA	NA	NA	NA	NA
IAV110600	NA	A	H	S	N	K	S	E	H	H	Q	N	E
Inoculum	1	A	H	S	N	K	P	E	H	H	Q	N	E
	2	.	.	.	.	.	.	.	.	.	.	-	K
SD2	1	.	.	.	.	.	.	.	.	.	.	.	.
	2	.	.	.	.	.	.	.	.	.	.	-	K
	3	.	.	.	.	.	.	K	.	.	.	.	.
	4	.	.	.	.	.	.	K	.	.	.	-	K
	5	.	.	.	.	.	L	.	.	.	.	.	.
	6	.	.	.	.	.	L	.	.	.	.	-	K
	7	.	.	.	.	.	L	K	.	.	.	.	.
	8	.	.	.	.	.	L	K	.	.	.	-	K
SD4	1*	.	.	.	.	.	.	.	.	.	.	.	.
	2*	.	.	.	.	.	.	.	.	.	.	.	.
A1D5	1	.	.	.	.	.	.	.	.	.	.	.	.
	2	.	.	.	.	.	.	.	.	.	.	Q	K
A1D7	1	.	R	P	.	.	.	.	.	.	.	.	.
	2	.	R	P	.	.	.	.	.	.	-	Q	K
	3	.	R	P	.	.	.	.	Q	Q	-	.	.
	4	.	R	P	.	.	.	.	Q	Q	-	-	K
	5	.	R	P	K	.	.	.	.	.	.	.	.
	6	.	R	P	K	.	.	.	.	.	.	-	K
	7	.	R	P	K	.	.	.	Q	Q	-	-	K
	8	.	R	P	.	E	.	.	Q	Q	-	.	.
	9	.	R	P	K	E	.	.	.	.	.	.	.
	10	.	R	P	.	E	.	.	.	.	.	.	.
	11	.	R	P	K	.	.	.	Q	Q	-	.	.
	12	.	R	P	K	E	.	.	Q	Q	-	.	.
	13	.	R	P	.	E	.	.	.	.	-	Q	K
	14	.	R	P	K	E	.	.	.	.	-	Q	K
	15	.	R	P	K	E	.	.	Q	-	-	Q	K
	16	.	R	P	.	E	.	.	Q	-	-	Q	K
	17	G	R	P	.	E	.	.	.	.	-	Q	K
	18	G	R	P	.	E	.	.	Q	-	-	Q	K
	19	G	R	P	K	E	.	.	Q	-	-	Q	K
	20	G	R	P	K	E	.	.	.	.	-	Q	K
	21	G	R	P	K	E	.	.	.	.	.	.	.
	22	G	R	P	.	E	.	.	.	.	.	.	.
	23	G	R	P	.	E	.	.	Q	Q	-	.	.
	24	G	R	P	K	E	.	.	Q	Q	-	.	.
	25	G	R	P	.	.	.	.	Q	Q	-	.	.
	26	G	R	P	K	.	.	.	Q	Q	-	.	.
	27	G	R	P	K	.	.	.	Q	Q	-	-	K
	28	G	R	P	.	.	.	.	Q	Q	-	-	K
	29	G	R	P	K	.	.	.	.	.	.	-	K
	30	G	R	P	K	.	.	.	.	.	.	.	.
	31	G	R	P	.	.	.	.	.	.	.	-	K
	32	G	R	P	.	.	.	.	.	.	.	.	.
A2D4	1	.	.	.	.	.	.	.	.	.	.	.	.
	2	.	.	.	.	.	.	.	.	Q	.	-	K
A2D8	1	.	.	.	.	.	.	.	.	.	.	.	.
	2	.	.	.	.	.	.	.	.	.	.	Q	K
A3D5	1	.	.	.	.	.	.	.	.	.	.	.	.
A3D6	1	.	.	.	.	.	.	.	.	.	.	.	.
A4D3	1	.	.	.	.	.	.	.	.	.	.	.	.
A4D6	1	.	.	.	.	.	.	.	.	.	.	.	.
	2	.	.	.	.	.	.	.	Q	Q	-	-	K
A5D4	1	.	.	.	.	.	.	.	.	.	.	.	.
	2	.	.	.	.	.	.	.	.	.	.	-	K
A5D6	1	.	.	.	.	.	.	.	.	.	.	.	.

*Identical proteins within a sample translated from two different alleles.

samples except for A4D6, in which a third emergent NA allele was found (Fig. 1b).

DISCUSSION

To better understand the evolution of IAVs during infection of vaccinated pigs we used deep genome sequencing to compare the viral genetic diversity at two separate sampling points during infection. We demonstrated that the genetic makeup of the virus changed in all gene segments as the virus replicated within the group of animals, yielding a complex collection of viral genomes with similar and distinct variants. The infection produced a population of heterogeneous alleles by gene segment (usually two or more) that was dynamic over time. Therefore, our results indicate that the genetic heterogeneity of IAVs during infection of partially immune pigs is significant and it might have been underestimated. Under this scenario, controlling the transmission of IAVs in pigs is challenging because under natural conditions a large proportion of pigs have maternal or active immunity to different IAV strains (Corzo *et al.*, 2013; Torremorell *et al.*, 2012), IAV are endemic in swine populations (Vincent *et al.*, 2008), multiple genetic lineages of the virus can co-circulate in pigs (Diaz *et al.*, 2015; Ducatez *et al.*, 2011), and pigs are moved and mixed in large batches of animals during their production stage (Knauer & Hostetler, 2013; Oh & Whitley, 2011).

In our study, the genetic diversity of IAVs changed dynamically throughout the course of infection. Two samples that corresponded to two different pigs (A1D7 and A4D3) had a higher number of alleles compared with the rest of the samples. As we did not sequence all samples from all pigs and our sample size was limited, we cannot be certain that high numbers of alleles were not present in all pigs at some point during infection. However, our results proved that the diversity of IAVs could change within a vaccinated pig throughout the course of infection. Interestingly, these two pigs (animals 1 and 4) were negative by HI to the challenge virus (titre $<1:20$) before exposure to the seeder pig and had the lowest HI against A/Swine/Iowa/110600/2000(H1N1), which is the closest vaccine virus compared with the challenge virus. These findings suggest that in pigs the level of antibodies against IAVs might influence the overall diversity of the virus during infection. However, this observation needs to be corroborated in future studies, particularly in the context of heterologous vaccination to infection. Most vaccines are heterologous to circulating viruses and only provide partial protection to infection; therefore, the variability in the immune response to IAV vaccination may influence virus evolution. Individual host factors such as response to social stress (de Groot *et al.*, 2001), individual host genetics and animal behaviour may also affect the immune response to viral infections. In addition, it remains unclear to what extent different alleles are selected for or whether rapid changes in the viral population are primarily stochastic.

In the samples sequenced in this study, the HA segment was more likely to undergo non-synonymous mutations compared with the remaining segments, including NA. Only one HA emergent allele was found in more than one pig and this allele was identical in the HA1 region to the inoculum virus. In agreement with our results, other studies in pigs have shown that nucleotide substitutions can occur in the HA segment very early after infection in pigs with immunity to IAV and that allele fixation can occur amongst infected animals (Diaz *et al.*, 2013; Murcia *et al.*, 2012). However, the time required for these substitutions to become fixed at a population level is still unknown. Our results are also consistent with a previous study of the HA1 region of the HA indicating that IAVs in pigs are not being transmitted as a single genotype, but rather as a population of viruses that may be closely related to each other (Murcia *et al.*, 2012). However, our results also showed that nucleotide substitutions can occur in the signal peptide and the HA2 region of the HA, which were not evaluated by Murcia *et al.* (2012). In addition, other factors in our study, such as the group housing conditions, which facilitated greater interaction between pigs compared with previous studies where pairs of individuals were used to measure intra-host diversity of IAV (Hensley *et al.*, 2009; Murcia *et al.*, 2010, 2012), may have had an effect on the increased overall genetic diversity.

In contrast, we did not find non-synonymous mutations in NA, and there was therefore no evidence of coevolution of HA and NA or epistatic interactions. Although the dominant NA allele in the inoculum was not observed amongst the majority of samples sequenced, including the seeder, the allele was observed in sample A1D7, indicating that it had likely persisted at low levels during transmission. This genotype 'recovery' has been described during replication for other RNA viruses, such as polioviruses (Domingo *et al.*, 1985). It is possible that unique alleles found in pigs were present in the inoculum at a low prevalence and that we were not able to identify them in the inoculum itself. Additionally, unique alleles could have been present in the pig samples that were not sequenced.

Multiple studies have evaluated the intra-host diversity of RNA viruses overtime (Debbink *et al.*, 2014; Salemi, 2013; Tu *et al.*, 2013). In IAVs, this research has focused on HA (Hoelzer *et al.*, 2010; Murcia *et al.*, 2010, 2013). To the best of our knowledge, our study is the first to evaluate the intra-host diversity of the complete genome of IAV during infection of pigs using NGS, and our results are comparable with a recent study in children (Bourret *et al.*, 2015). We uncovered an additional layer of complexity in the evolution of IAVs during infection of immune pigs by demonstrating that all IAV gene segments replicate as a population of alleles that may or may not be transmitted. Although our methods were not able to capture reassortment events within hosts, the intra-host diversity observed here certainly provides opportunity for novel reassortant viruses to emerge. In other species, the evaluation of intra-host reassortment of IAV has shown that

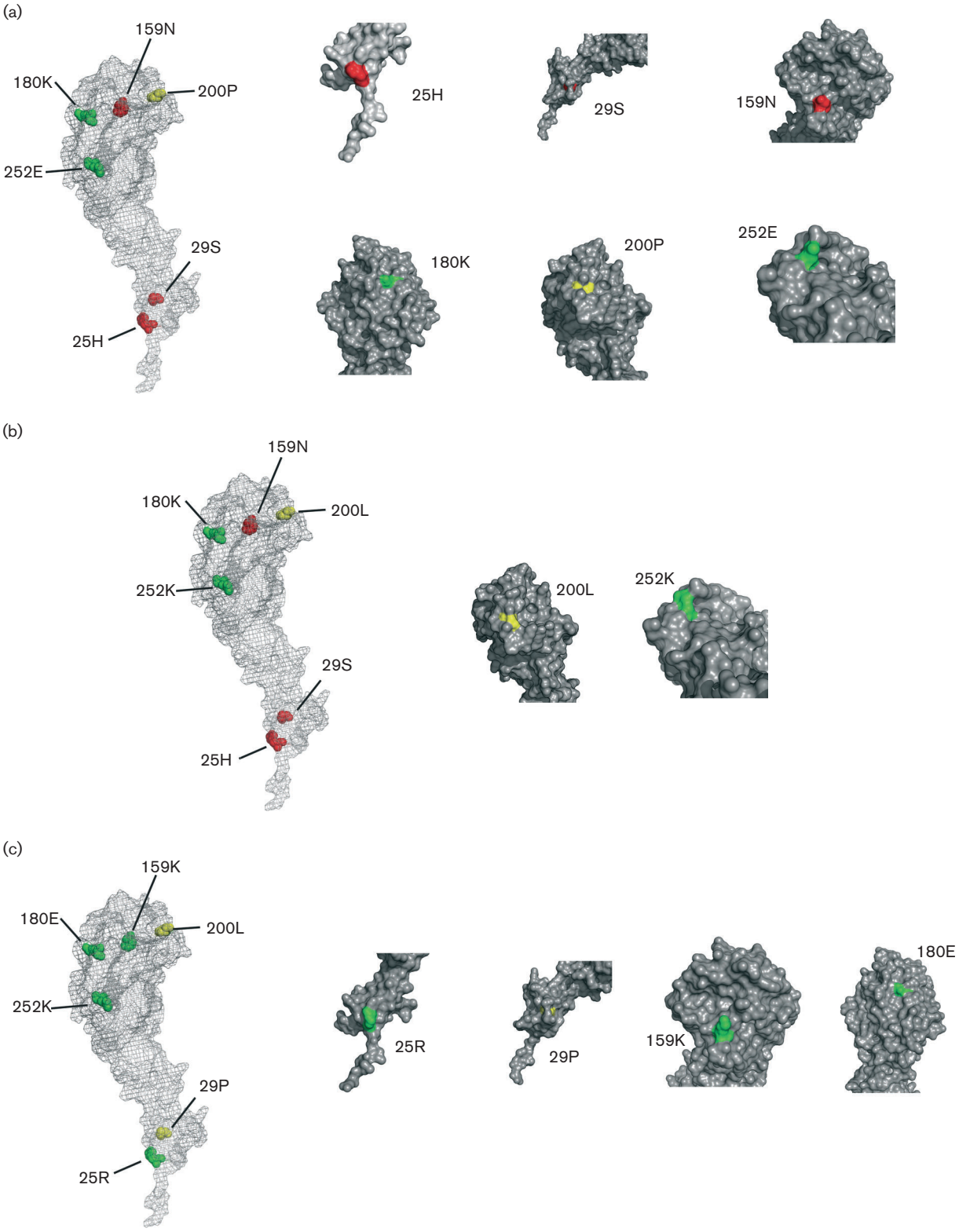

Fig. 2. Three-dimensional models illustrating the HA1 region of the HA and the polymorphic amino acids found during the study. Polymorphic amino acids [histidine (H), serine (S), asparagine (N), lysine (K), glutamate (E), leucine (L), arginine (R) and proline (P)] are indicated and coloured according to their physical properties: polar (red), charged (green) and hydrophobic (yellow). (a) Reference amino acids at polymorphic sites in HA1 for A/Swine/IA/00239/2004(H1N1) (challenge virus). (b) Amino acid residues predicted from the sample sequenced from the seeder pig at day 2 (SD2). (c) Amino acid residues predicted from the sample sequenced from animal 1 at day 7 (A1D7).

two viruses that are closely related to each other reassort at different rates depending on their co-infecting dose (Tao *et al.*, 2014). Genome reassortment should be further investigated as it is a potential source of genetic diversity to swine IAV. It is not clear to what extent the diversity observed in our study is deleterious and not likely to be transmitted onward over longer time periods in pigs. Further understanding of the intra-host dynamics of co-infection and reassortment remains an important outstanding question in IAV evolution.

The dynamic nature of polymorphisms found in our study highlights that IAV genetic diversity ought to be studied directly in the original biological sample (i.e. nasal swab). The study of genetic diversity and evolution of IAV populations through cell culture might be misleading. The cell culture of IAV leads to loss of diversity as cell culture selects for the fastest growing virus in a new environment. As an example, the frequently used Madin–Darby canine kidney (MDCK) and Vero cell lines have differential preferences for IAV variants, which has led to selective rescue of specific alleles during serial passages (Roedig *et al.*, 2011). As IAV diversity and population dynamics are complex and shaped by many factors, including viral fitness, mutation rate, host factors and stochastic events that may produce bottlenecks, a better estimate of IAV diversity at the population level can be obtained directly from the original sample. However, amplifying the complete genome from IAV isolates, where the viral concentration is exponentially higher than the original sample, might be easier for certain studies.

It is important to mention different external sources of potential variation and bias, including sequencing (Ross *et al.*, 2013), depth of coverage (Bidzhieva *et al.*, 2014), PCR (Archer *et al.*, 2012; Cummings *et al.*, 2010) and sampling or intra-assay bias or error. The platform we used, i.e. 454, is mature and errors due to sequencing are considered non-issues as we avoid reads with a Phred score <20, polymorphisms in homonucleotide runs, not represented in both strands and not represented in unique independent sequence runs (Archer *et al.*, 2012; Chen-Harris *et al.*, 2013). The high confidence differences HCDiff file that we used takes these three precautions into account. The 454 has proven to accurately detect human immunodeficiency virus mutants at a prevalence as low as 0.1 % (Shao *et al.*, 2013). Nevertheless, the variability on NGS reads mapped and depth of coverage throughout the complete genome of IAVs remains an issue to better estimate the genetic diversity of the viral populations (Bidzhieva *et al.*, 2014; Bourret *et al.*, 2013). Additionally, PCR (especially when the polymerase is stalled) generates *in vitro* recombinants that inflate and distort the estimates of the number and structure of the true alleles (Cummings *et al.*, 2010; S. Enomoto, unpublished data). To avoid PCR errors, we used a high-fidelity PCR system that uses a blend of DNA polymerases including one isolated from *Pyrococcus furiosus* (*Pfu*) which has $3'\rightarrow5'$ exonuclease (proofreading) activity, with a 10-fold improved error rate compared with *Taq* DNA polymerase (André *et al.*, 1997; Lundberg *et al.*, 1991). High-fidelity polymerases have been shown to enhance the PCR and sequencing conditions (Hedman *et al.*, 2009; Wu *et al.*, 2010), improving the accuracy to estimate microbial diversity.

In conclusion, the swine IAV population in an experimental setting was complex. Although we recognize our sampling bias to estimate the complete genetic composition of the viral population during transmission, our findings demonstrate that the diversity of IAV can change dynamically during infection of vaccinated pigs. New sequencing technologies and bioinformatics algorithms might provide more precise estimates in future studies. In this study, the polymorphisms were abundant, dynamic and not limited to HA and NA. Some variants were maintained whilst others were not identified amongst the samples sequenced. Direct sampling and deep sequencing allowed us to investigate the dynamic plasticity of the IAV population during IAV infection in a small group of pigs. We envision that the plasticity of the IAV genome under field conditions is no less complex as different IAV subtypes can coexist and susceptible animals are continuously introduced into infected populations. Our study emphasizes the need to study IAV evolution directly from the infected host using new-generation sequencing approaches, which will help design better strategies to control influenza in animals and people. More studies are needed in order to evaluate whether the changes observed in this study are due to vaccination or whether they are also found in non-immune pigs.

METHODS

Study design. Eleven 3-week-old specific-pathogen-free piglets were selected from a serologically IAV-negative swine herd and moved to the University of Minnesota animal research units. The IAV-negative status was confirmed by testing individual nasal swabs with RRT-PCR targeting the M gene (Slomka *et al.*, 2010; Spackman & Suarez, 2008) and serum samples by ELISA (Influenza Ab Test kit; IDEXX Laboratories) for antibodies against NP (Ciacci-Zanella *et al.*, 2010).

Viral RNA was eluted using 50 µl of each sample into 50 µl elution buffer using an Ambion MagMax virus RNA isolation kit (Life Technologies). An AgPath-ID One-Step RT-PCR reagent kit (Life Technologies) was used to detect IAV. PCR mix containing 5 µl RNA, 12.5 µl $2\times$ buffer, 1.0 µl $25\times$ enzyme mix, 1.67 µl detection enhancer, 5 pmol each primer and 1.5 pmol probe was run on a LightCycler 480 system (Hoffmann-La Roche) at 45 °C for 10 min, followed by 95 °C for 10 min, and 45 cycles at 94 °C for 1 s and 60 °C for 30 s. Fluorescence was recorded at 60 °C and a sample was considered positive if the C_t was <40. This PCR protocol can detect IAVs in samples containing ≥200 copies of the target amplicon, and has 100 and 95 % diagnostic sensitivity and specificity, respectively (Slomka *et al.*, 2010).

Ten pigs were vaccinated 1 day after arrival and 2 weeks later with 2 ml of a licensed inactivated trivalent IAV vaccine (FluSure; Zoetis Animal Health), containing the δ and γ clusters of H1N1 [A/Swine/North-Carolina/031/2005(H1N1) and A/Swine/Iowa/110600/2000(H1N1), respectively] and one H3N2 [A/Swine/Missouri/069/2005(H3N2)]. Two

weeks after the second vaccination, nasal swabs and blood samples were collected from all pigs and tested for IAV by RRT-PCR (Spackman & Suarez, 2008) and ELISA, respectively. Additionally, blood samples were tested by HI tests against the challenge and vaccine viruses before and after infection as described previously (Direksin *et al.*, 2002). The mean ELISA and HI titres were compared between vaccinated pigs that tested RRT-PCR-positive or -negative during this study, and considered statistically significant if the P value for the non-parametric one-way ANOVA Kruskal–Wallis test was <0.05.

One unvaccinated pig was inoculated with IAV in a separate room to serve as a seeder pig to infect the other pigs. An aliquot of 2 ml 1×10^6 $TCID_{50}$ ml^{-1} A/Swine/IA/00239/2004(H1N1) IAV (GenBank accession number EU139832.1) grown in MDCK cells (Meguro *et al.*, 1979) was used to challenge the seeder pig intranasally and intratracheally. The A/Swine/IA/00239/2004(H1N1) clusters within the β H1 swine IAVs (Lorusso *et al.*, 2011). This virus was selected because it has been fully characterized, genetically and antigenically (Anderson *et al.*, 2015), and it has been used in several pathogenesis (Vincent *et al.*, 2007) and transmission studies (Allerson *et al.*, 2013; Diaz *et al.*, 2013; Romagosa *et al.*, 2011). The challenge virus was 91.5 and 73.7 % identical at the nucleotide level to the H1 γ and δ vaccine virus strains, respectively. The infection was confirmed 48 h later by RRT-PCR and the seeder pig was placed in contact with the rest of the pigs. Nasal swabs were collected from all pigs daily for 14 days into 1.8 ml viral transport medium (minimum essential medium plus 2 % BSA and 1 % penicillin/streptomycin) and an aliquot of the transport medium was used for RRT-PCR testing. All pigs were euthanized on day 14 and all procedures for this study were approved by the University of Minnesota Institutional Animal Care and Use Committee (protocol number 0908A71965).

Sample selection, genome amplification and sequence identification. To explore the within host variability of IAV during infection, two IAV-positive samples from each pig were conveniently selected for complete genome amplification and sequencing using NGS technologies (Table 2). Samples with the lowest C_t value and best genome amplification were targeted for sequencing. The IAV genome was amplified using a modified protocol of Zhou *et al.* (2009). Briefly, the viral RNA was purified from the swabs using a QIAamp Viral RNA Mini kit (Qiagen). IAV cDNA was created from viral RNA using primer MBtuni12(M) (ACGCGTGATCAGCRAAAGCAGG) and Superscript III First Strand Synthesis SuperMix (Invitrogen) cDNA was amplified in a PCR (five cycles of 94 °C 15 s, 45 °C 30 s, 68 °C 180 s and 31 cycles of 94 °C 15 s, 57 °C 30 s, 68 °C 180 s) consisting of PicoMax High Fidelity DNA Polymerase (Agilent), MBtuni12(M) and MBtuni13 (ACGCGTGATCAGTAGAAACAAGG). PCR products were verified by gel electrophoresis and purified using a QIAquick Spin kit (Qiagen). Purified cDNAs from the virus inoculum and 12 pig samples (Table 2) were submitted to the Genomics Center at the University of Minnesota for library preparation and 454 sequencing (454 GS-FLX; Roche Diagnostics) as described in detail by Ramakrishnan *et al.* (2009).

The 454 inoculum reads were assembled with Newbler 2.6 (Roche Diagnostics) using a reference template obtained from GenBank (Table S2) and the inoculum consensus sequence was used as the reference genome (Table S3) to assemble the 454 reads from each pig sample. The polymorphisms present in each sample were extracted from the 454 HCDiff.txt files created during each assembly in Newbler 2.6. These file include only highly confident differences which are defined as variants identified in at least three unique reads, and present in forward and reverse reads.

Allele identification and overlapping reading test. Alleles (sequence variants) were defined as complete functional gene segments identified by aligning overlapping sequence fragments. The Newbler output, 454 HCDiff.txt, is a file of sequence alignments surrounding all the high confidence polymorphic loci. A Ruby (Goto *et al.*, 2010) script was written to test the linkage of two adjacent loci by enumerating the occurrence of the four sequence combinations: consensus–consensus, consensus–variant, variant–consensus and variant–variant. If >80 % of the sequences occurred only as two sequence combinations, the two loci were considered linked. Presence or absence of polymorphisms at each locus was encoded as 1 or 0, respectively. The alleles were deduced by linking together the adjacent intervals between the two polymorphic loci and its functionality verified using the National Center for Biotechnology Information FLu ANnotation tool (FLAN; http://www.ncbi.nlm.nih.gov/genomes/FLU/Database/annotation.cgi) (Bao *et al.*, 2007). Additionally, if the distance that separated two polymorphisms was longer than the length of the reads obtained, then those two polymorphisms were considered not linked. For example, if two adjacent polymorphic loci were linked and recovered as 00 and 11, the segment contained two alleles rather than four alleles. The raw 454 reads, the allele sequences obtained, and the Ruby scripts for overlapping sequence fragments analysis and allele extraction are available upon request.

Sequence analysis. To illustrate the phylogenetic relationship between sequences, alleles were aligned to the reference genome using DNA-Alignment and median-joining networks were estimated using Network (Bandelt *et al.*, 1999). Each network was annotated with Network Publisher (Fluxus Technology) and Adobe Illustrator CC (Adobe Systems). Additionally, for the first ORF we estimated the mean number of synonymous (dS) and non-synonymous (dN) mutations and their ratio (dS/dN) amongst sequences (Korber, 2000; Nei & Gojobori, 1986) using the Synonymous and Non-synonymous Analysis Program (SNAP; www.hiv.lanl.gov).

HA and NA protein analysis. For HA and NA, hypothetical proteins were inferred from nucleotide sequence, aligned using CLUSTAL_X (Larkin *et al.*, 2007) and compared. The amino acid differences amongst HA sequences were mapped to the known H1 antigenic sites (Caton *et al.*, 1982; Deem & Pan, 2009), modelled using the tools available at http://swissmodel.expasy.org/ (Arnold *et al.*, 2006) and illustrated using PyMOL version 1.5.0.4 (https://www.pymol.org/). The HA1 IAV template used for our protein model was A/Swine/Iowa/15/30(H1N1) (Protein Data Bank ID: 1RUY). This template was used because this virus is from swine origin, the HA has been crystallized and it is available for public use.

ACKNOWLEDGEMENTS

This work was supported in whole or in part by federal funds from the National Institutes of Health (contract HHSN266200700007C), the US Department of Agriculture Discretionary GAR Funds (AES0060014) as part of the Signature Program Funding at the College of Veterinary Medicine University of Minnesota and COLCIENCIAS: Departamento Administrativo de Ciencia, Tecnología e Innovación. The authors would like to acknowledge Dr Douglas Marthaler from the University of Minnesota Veterinary Diagnostic Laboratory for his technical assistance, and the Minnesota Supercomputing Institute at the University of Minnesota.

REFERENCES

Ali, A., Daniels, J. B., Zhang, Y., Rodriguez-Palacios, A., Hayes-Ozello, K., Mathes, L. & Lee, C. W. (2011). Pandemic and seasonal human influenza virus infections in domestic cats: prevalence, association with respiratory disease, and seasonality patterns. *J Clin Microbiol* **49**, 4101–4105.

Allerson, M., Deen, J., Detmer, S. E., Gramer, M. R., Joo, H. S., Romagosa, A. & Torremorell, M. (2013). The impact of maternally derived immunity on influenza A virus transmission in neonatal pig populations. *Vaccine* **31**, 500–505.

Anderson, T. K., Campbell, B. A., Nelson, M. I., Lewis, N. S., Janas-Martindale, A., Killian, M. L. & Vincent, A. L. (2015). Characterization of co-circulating swine influenza A viruses in North America and the identification of a novel H1 genetic clade with antigenic significance. *Virus Res* **201**, 24–31.

André, P., Kim, A., Khrapko, K. & Thilly, W. G. (1997). Fidelity and mutational spectrum of *Pfu* DNA polymerase on a human mitochondrial DNA sequence. *Genome Res* **7**, 843–852.

Archer, J., Baillie, G., Watson, S. J., Kellam, P., Rambaut, A. & Robertson, D. L. (2012). Analysis of high-depth sequence data for studying viral diversity: a comparison of next generation sequencing platforms using Segminator II. *BMC Bioinformatics* **13**, 47.

Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201.

Bandelt, H. J., Forster, P. & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**, 37–48.

Bao, Y., Bolotov, P., Dernovoy, D., Kiryutin, B. & Tatusova, T. (2007). FLAN: a web server for influenza virus genome annotation. *Nucleic Acids Res* **35** (Web Server), W280–W284.

Belshaw, R., Pybus, O. G. & Rambaut, A. (2007). The evolution of genome compression and genomic novelty in RNA viruses. *Genome Res* **17**, 1496–1504.

Bidzhieva, B., Zagorodnyaya, T., Karagiannis, K., Simonyan, V., Laassri, M. & Chumakov, K. (2014). Deep sequencing approach for genetic stability evaluation of influenza A viruses. *J Virol Methods* **199**, 68–75.

Blanc, A., Ruchansky, D., Clara, M., Achaval, F., Le Bas, A. & Arbiza, J. (2009). Serologic evidence of influenza A and B viruses in South American fur seals (*Arctocephalus australis*). *J Wildl Dis* **45**, 519–521.

Bourret, V., Croville, G., Mariette, J., Klopp, C., Bouchez, O., Tiley, L. & Guérin, J. L. (2013). Whole-genome, deep pyrosequencing analysis of a duck influenza A virus evolution in swine cells. *Infect Genet Evol* **18**, 31–41.

Bourret, V., Croville, G., Mansuy, J. M., Mengelle, C., Mariette, J., Klopp, C., Genthon, C., Izopet, J. & Guérin, J. L. (2015). Intra-host viral variability in children clinically infected with H1N1(2009) pandemic influenza. *Infect Genet Evol* **33**, 47–54.

Caton, A. J., Brownlee, G. G., Yewdell, J. W. & Gerhard, W. (1982). The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype). *Cell* **31**, 417–427.

Chen, R. & Holmes, E. C. (2010). Hitchhiking and the population genetic structure of avian influenza virus. *J Mol Evol* **70**, 98–105.

Chen-Harris, H., Borucki, M. K., Torres, C., Slezak, T. R. & Allen, J. E. (2013). Ultra-deep mutant spectrum profiling: improving sequencing accuracy using overlapping read pairs. *BMC Genomics* **14**, 96.

Ciacci-Zanella, J. R., Vincent, A. L., Prickett, J. R., Zimmerman, S. M. & Zimmerman, J. J. (2010). Detection of anti-influenza A nucleoprotein antibodies in pigs using a commercial influenza epitope-blocking enzyme-linked immunosorbent assay developed for avian species. *J Vet Diagn Invest* **22**, 3–9.

Corzo, C. A., Culhane, M., Juleen, K., Stigger-Rosser, E., Ducatez, M. F., Webby, R. J. & Lowe, J. F. (2013). Active surveillance for influenza A virus among swine, midwestern United States, 2009-2011. *Emerg Infect Dis* **19**, 954–960.

Cummings, S. M., McMullan, M., Joyce, D. A. & van Oosterhout, C. (2010). Solutions for PCR, cloning and sequencing errors in population genetic analysis. *Conserv Genet* **11**, 1095–1097.

de Groot, J., Ruis, M. A., Scholten, J. W., Koolhaas, J. M. & Boersma, W. J. (2001). Long-term effects of social stress on antiviral immunity in pigs. *Physiol Behav* **73**, 145–158.

Debbink, K., Lindesmith, L. C., Ferris, M. T., Swanstrom, J., Beltramello, M., Corti, D., Lanzavecchia, A. & Baric, R. S. (2014). Within-host evolution results in antigenically distinct GII.4 noroviruses. *J Virol* 7244–7255.

Deem, M. W. & Pan, K. (2009). The epitope regions of H1-subtype influenza A, with application to vaccine efficacy. *Protein Eng Des Sel* **22**, 543–546.

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 Respir Viruses* **7** (Suppl 4), 52–60.

Diaz, A., Perez, A., Sreevatsan, S., Davies, P., Culhane, M. & Torremorell, M. (2015). Association between influenza A virus infection and pigs subpopulations in endemically infected breeding herds. *PLoS One* **10**, e0129213.

Direksin, K., Joo, H. & Goyal, S. M. (2002). An immunoperoxidase monolayer assay for the detection of antibodies against swine influenza virus. *J Vet Diagn Invest* **14**, 169–171.

Domingo, E., Martínez-Salas, E., Sobrino, F., de la Torre, J. C., Portela, A., Ortín, J., López-Galindez, C., Pérez-Breña, P., Villanueva, N. & other authors (1985). The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance - a review. *Gene* **40**, 1–8.

Domingo, E., Sheldon, J. & Perales, C. (2012). Viral quasispecies evolution. *Microbiol Mol Biol Rev* **76**, 159–216.

Ducatez, M. F., Hause, B., Stigger-Rosser, E., Darnell, D., Corzo, C., Juleen, K., Simonson, R., Brockwell-Staats, C., Rubrum, A. & other authors (2011). Multiple reassortment between pandemic (H1N1) 2009 and endemic influenza viruses in pigs. United States. *Emerg Infect Dis* **17**, 1624–1629.

Goto, N., Prins, P., Nakao, M., Bonnal, R., Aerts, J. & Katayama, T. (2010). BioRuby: bioinformatics software for the Ruby programming language. *Bioinformatics* **26**, 2617–2619.

Hedman, J., Nordgaard, A., Rasmusson, B., Ansell, R. & Rådström, P. (2009). Improved forensic DNA analysis through the use of alternative DNA polymerases and statistical modeling of DNA profiles. *Biotechniques* **47**, 951–958.

Hensley, S. E., Das, S. R., Bailey, A. L., Schmidt, L. M., Hickman, H. D., Jayaraman, A., Viswanathan, K., Raman, R., Sasisekharan, R. & other authors (2009). Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. *Science* **326**, 734–736.

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**, 5329–5335.

Kitikoon, P., Nelson, M. I., Killian, M. L., Anderson, T. K., Koster, L., Culhane, M. R. & Vincent, A. L. (2013). Genotype patterns of contemporary reassorted H3N2 virus in US swine. *J Gen Virol* **94**, 1236–1241.

Knauer, M. T. & Hostetler, C. E. (2013). US swine industry productivity analysis, 2005 to 2010. *J Swine Health Prod* **21**, 248–252.

Korber, B. (2000). HIV signature and sequence variation analysis. In *Computational Analysis of HIV Molecular Sequences*, pp. 55–72. Edited by A. G. Rodrigo & G. H. Learn. Dordrecht: Kluwer.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A. & other authors (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948.

Lemey, P., Rambaut, A. & Pybus, O. G. (2006). HIV evolutionary dynamics within and among hosts. *AIDS Rev* **8**, 125–140.

Lorusso, A., Vincent, A. L., Harland, M. L., Alt, D., Bayles, D. O., Swenson, S. L., Gramer, M. R., Russell, C. A., Smith, D. J. & other authors (2011). Genetic and antigenic characterization of H1 influenza viruses from United States swine from 2008. *J Gen Virol* **92**, 919–930.

Lundberg, K. S., Shoemaker, D. D., Adams, M. W. W., Short, J. M., Sorge, J. A. & Mathur, E. J. (1991). High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene* **108**, 1–6.

Meguro, H., Bryant, J. D., Torrence, A. E. & Wright, P. F. (1979). Canine kidney cell line for isolation of respiratory viruses. *J Clin Microbiol* **9**, 175–179.

Murcia, P. R., Baillie, G. J., Daly, J., Elton, D., Jervis, C., Mumford, J. A., Newton, R., Parrish, C. R., Hoelzer, K. & other authors (2010). Intra- and interhost evolutionary dynamics of equine influenza virus. *J Virol* **84**, 6943–6954.

Murcia, P. R., Hughes, J., Battista, P., Lloyd, L., Baillie, G. J., Ramirez-Gonzalez, R. H., Ormond, D., Oliver, K., Elton, D. & other authors (2012). Evolution of an Eurasian avian-like influenza virus in naïve and vaccinated pigs. *PLoS Pathog* **8**, e1002730.

Murcia, P. R., Baillie, G. J., Stack, J. C., Jervis, C., Elton, D., Mumford, J. A., Daly, J., Kellam, P., Grenfell, B. T. & other authors (2013). Evolution of equine influenza virus in vaccinated horses. *J Virol* **87**, 4768–4771.

Nei, M. & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**, 418–426.

Nelson, M. I., Lemey, P., Tan, Y., Vincent, A., Lam, T. T. Y., Detmer, S., Viboud, C., Suchard, M. A., Rambaut, A. & other authors (2011). Spatial dynamics of human-origin H1 influenza A virus in North American swine. *PLoS Pathog* **7**, e1002077.

Nelson, M. I., Gramer, M. R., Vincent, A. L. & Holmes, E. C. (2012). Global transmission of influenza viruses from humans to swine. *J Gen Virol* **93**, 2195–2203.

Oh, S. H. & Whitley, N. C. (2011). Pork production in China, Japan and South Korea. *Asian-Australas J Anim Sci* **24**, 1629–1636.

Ramakrishnan, M. A., Tu, Z. J., Singh, S., Chockalingam, A. K., Gramer, M. R., Wang, P., Goyal, S. M., Yang, M., Halvorson, D. A. & Sreevatsan, S. (2009). The feasibility of using high resolution genome sequencing of influenza A viruses to detect mixed infections and quasispecies. *PLoS One* **4**, e7105.

Rambaut, A. & Holmes, E. (2009). The early molecular epidemiology of the swine-origin A/H1N1 human influenza pandemic. *PLoS Curr* **1**, RRN1003.

Roedig, J. V., Rapp, E., Höper, D., Genzel, Y. & Reichl, U. (2011). Impact of host cell line adaptation on quasispecies composition and glycosylation of influenza A virus hemagglutinin. *PLoS One* **6**, e27989.

Romagosa, A., Allerson, M., Gramer, M., Joo, H. S., Deen, J., Detmer, S. & Torremorell, M. (2011). Vaccination of influenza A virus decreases transmission rates in pigs. *Vet Res* **42**, 120.

Ross, M. G., Russ, C., Costello, M., Hollinger, A., Lennon, N. J., Hegarty, R., Nusbaum, C. & Jaffe, D. B. (2013). Characterizing and measuring bias in sequence data. *Genome Biol* **14**, R51.

Russell, C. A., Jones, T. C., Barr, I. G., Cox, N. J., Garten, R. J., Gregory, V., Gust, I. D., Hampson, A. W., Hay, A. J. & other authors (2008). The global circulation of seasonal influenza A (H3N2) viruses. *Science* **320**, 340–346.

Salemi, M. (2013). The intra-host evolutionary and population dynamics of human immunodeficiency virus type 1: a phylogenetic perspective. *Infect Dis Rep* **5** (Suppl 1), e3.

Shao, W., Boltz, V. F., Spindler, J. E., Kearney, M. F., Maldarelli, F., Mellors, J. W., Stewart, C., Volfovsky, N., Levitsky, A. & other authors (2013). Analysis of 454 sequencing error rate, error sources, and artifact recombination for detection of low-frequency drug resistance mutations in HIV-1 DNA. *Retrovirology* **10**, 18.

Simonsen, L., Spreeuwenberg, P., Lustig, R., Taylor, R. J., Fleming, D. M., Kroneman, M., Van Kerkhove, M. D., Mounts, A. W., Paget, W. J. & GLaMOR Collaborating Teams (2013). Global mortality estimates for the 2009 influenza pandemic from the GLaMOR project: a modeling study. *PLoS Med* **10**, e1001558.

Slomka, M. J., Densham, A. L. E., Coward, V. J., Essen, S., Brookes, S. M., Irvine, R. M., Spackman, E., Ridgeon, J., Gardner, R. & other authors (2010). Real time reverse transcription (RRT)-polymerase chain reaction (PCR) methods for detection of pandemic (H1N1) 2009 influenza virus and European swine influenza A virus infections in pigs. *Influenza Other Respir Viruses* **4**, 277–293.

Spackman, E. & Suarez, D. L. (2008). Type A influenza virus detection and quantitation by real-time RT-PCR. *Methods Mol Biol* **436**, 19–26.

Tao, H., Steel, J. & Lowen, A. C. (2014). Intrahost dynamics of influenza virus reassortment. *J Virol* **88**, 7485–7492.

Taubenberger, J. K. & Kash, J. C. (2010). Influenza virus evolution, host adaptation, and pandemic formation. *Cell Host Microbe* **7**, 440–451.

Tong, S., Zhu, X., Li, Y., Shi, M., Zhang, J., Bourgeois, M., Yang, H., Chen, X., Recuenco, S. & other authors (2013). New World bats harbor diverse influenza A viruses. *PLoS Pathog* **9**, e1003657.

Torremorell, M., Allerson, M., Corzo, C., Diaz, A. & Gramer, M. (2012). Transmission of influenza A virus in pigs. *Transbound Emerg Dis* **59** (Suppl 1), 68–84.

Tu, Z., He, Y. L., Lu, H., Xu, L., Yang, Z. B., Yang, C. & Chen, W. J. (2013). Mutant spectrum of Dengue type 1 virus in the plasma of patients from the 2006 epidemic in South China. *Int J Infect Dis* **17**, e1080–e1081.

Vincent, A. L., Ma, W., Lager, K. M., Janke, B. H., Webby, R. J., García-Sastre, A. & Richt, J. A. (2007). Efficacy of intranasal administration of a truncated NS1 modified live influenza virus vaccine in swine. *Vaccine* **25**, 7999–8009.

Vincent, A. L., Ma, W. J., Lager, K. M., Janke, B. H. & Richt, J. A. (2008). Swine influenza viruses: a North American perspective. *Adv Virus Res* **72**, 127–154.

Wei, K., Sun, H., Sun, Z., Sun, Y., Kong, W., Pu, J., Ma, G., Yin, Y., Yang, H. & other authors (2014). Influenza A virus acquires enhanced pathogenicity and transmissibility after serial passages in swine. *J Virol* **88**, 11981–11994.

Wu, J. Y., Jiang, X. T., Jiang, Y. X., Lu, S. Y., Zou, F. & Zhou, H. W. (2010). Effects of polymerase, template dilution and cycle number on PCR based 16S rRNA diversity analysis using the deep sequencing method. *BMC Microbiol* **10**, 255.

Zhou, B., Donnelly, M. E., Scholes, D. T., St George, K., Hatta, M., Kawaoka, Y. & Wentworth, D. E. (2009). Single-reaction genomic amplification accelerates sequencing and vaccine production for classical and swine origin human influenza A viruses. *J Virol* **83**, 10309–10313.