Inter-Host Reassortment Patterns in Swine Influenza Viruses

Hossein Khiabanian, Vladimir Trifonov, Raul Rabadan

Department of Biomedical Informatics and Center for Computational Biology and Bioinformatics, Columbia University College of Physicians and Surgeons, 1130 St. Nicholas Ave., New York, NY 10032.

Abstract. Three pandemic influenza strains occurred in the 20th century, in 1918, 1957 and 1968. Influenza pandemic strains are the result of an emerging virus spreading in people which have little or non immunity. At least two of these pandemics strains, in 1957 and in 1968, were the result of reassortments between human and avian viruses. In 1957 three segments, PB1, HA and NA, and in 1968 two segments, PB1 and HA, were of avian origin. Recently new influenza viruses have been isolated in Mexico and the United States. These viruses are found to be a complicated reassortement of swine, avian and human strains. Pigs are frequently found co-infected with human, avian and swine viruses. This observation has led to the conjecture that pigs are the mixing vessel that cause the avian-human reassortments, and hence the pandemics. Understanding the process and the patterns of viral reassortment, especially in pigs, is key to estimate and/or predict the likelihoods and rates of influenza pandemics. In the last few years databases collecting the sequences of influenza viruses in diverse geographical locations since 1918, including swine viruses, have been developed and made publicly available. In this paper, we study the ensemble of the swine influenza viruses to analyze how the reassortments happen in pigs. Inter-host reassortment patterns in pigs confirm similar previous results found in human viruses, both *in vitro* and *in vivo*. More interestingly, we have found that one of the polymerases, PB1 reassorts more often than other segments. In the last two pandemics, 1957 and 1968, PB1 was of avian origin. This observation reinforces the hypothesis that pigs constitute the inter-host mixing vessel responsible for at least two of the three pandemics in the 20th century.

Introduction

Pandemics are epidemics that rapidly spread on a worldwide scale infecting a large part of the human population, with an associated large number of casualties. Influenza pandemics are caused by emerging influenza viruses from a non-human reservoir. Of the three influenza pandemics of the twentieth century, at least two (1957 and 1968) have been caused by reassorted viruses from avian and human origin. In 1957, the human H1N1 strain that had been circulating since 1918 recombined to become a human H2N2 strain with new PB1, HA, and NA segments of avian origin [Lindstrom2004]. Also, in 1968, reassortment of the PB1 and HA segments created a new human H3N2 strain which is currently co-circulating with the human H1N1 strain that reappeared in 1977 [Nakajima1978, Scholtissek1978].

Recently, reassortant viruses of human, avian and swine origin have been isolated in humans in Mexico, the United States, Canada and Europe [CDC2009,WHO2009-1]. The virus present a complicated constellation of segments, from different origin. These complicated sets of multiple reassortments is characteristic of influenza A viruses found in swine. Since 2003, high pathogenic H5N1 avian virus has successfully infected more than 400 humans with a mortality rate of 60% [WHO2009-2]. It is not clear if any of these viruses will be the cause of the next pandemic, but it is vital to our society to understand the mechanisms behind reassortments to estimate the rates and likelihoods of possible pandemic strains.

Influenza A virus can be found in humans and a variety of animals with aquatic birds being considered as its main reservoir. Influenza viruses do not usually spread between different hosts. However, pigs are frequently documented to be infected by avian and human viruses, in addition to the swine strains. Co-infection is common and multiple reassotments are found to happen under natural conditions [CDC2009,WHO2009-1, Webby2000]. Hence, it has been postulated that swine are the mixing vessel for inter-host influenza viruses [Scholtissek1990].

Classical H1N1 swine virus have been circulating in pigs probably since the human influenza A pandemic in 1918. The swine H1N1 virus was the dominant strain in the United States until 1998, when two new swine H3N2 strains were identified in the United States. These new strains were the result of a double reassortment of classical swine H1N1 strain with the PB1, HA, and NA segments from a human H3N2 virus, and a triple reassortment (NP, M, and NS segments of classical swine H1N1 strain, PB1, HA, and NA segments of human H3N2 strain, and PB2 and PA segments of avian lineage). Since 1999, multiple strains of influenza virus have been isolated in pigs, both whole genome adaptations of human and/or avian viruses and inter-host reassortments [Zhou1999,Webby2000].

In this paper, we employ the temporally and geographically diverse information deposited in the Influenza Virus Resource of the National Center for Biotechnology Information [NCBI] to study the inter-host reassortment phenomena in swine influenza A viruses. By integrating all sequences that have been made publicly available, we can study the patterns behind the reasortment phenomenon. We have applied several techniques to identify the differential variability of the genes in the influenza genome. We study the diversity/entropy of each segment and the correlations between different segments. We enumerate the inter-host reassortment events to extract the segmental patterns in this process. We confirm some of the results that have been already reported in human viruses [Lubeck1979, Rabadan2008]: polymerase genes travel together, and HA and NA reassort more frequently than other segments. Surprisingly, we find that one of the polymerase genes, PB1, reassorts more frequently than other segments. Similar results were experimentally observed by Downie [Downie2004].

Methods

To compare the diversity within the segments of swine influenza A virus we use the strains deposited in the Influenza Virus Resource of the NCBI that have all eight segments completely sequenced. These 142 sequences contain 94 H1N1, 24 H1N2, 21 H3N2, and 3 H3N1 viruses. For each segment, we align the sequences using the Smith-Waterman algorithm and calculate the normalized Hamming distances only at the third codon positions, to eliminate the effects of evolutionary pressure due to positive selection.

To measure the diversity of each segment, we calculate D_i, Rao's quadratic entropy [Rao1982], according to

$$D_{i} = \frac{1}{N^{2}} \sum_{a,b=1}^{N} d_{ab}^{i},$$

where, d_{ab}^{i} is the Hamming distance between strains *a* and *b* and *N* is the total number of strains in the dataset. We estimate the confidence intervals for the diversity measurements, via 1000 bootstrap resamplings of the dataset.

To find the possible reassortant strains, we primarily follow the method introduced by Rabadan et al. (2008) [Rabadan2008], which was initially applied to complete sequences of human influenza A strains. Briefly, in this method, the number of differences in nucleotide sequences between the segments of any two strains is calculated. Assuming that RNA segments have proportional substitution rates at the third codon positions, the differences between two segments of two viruses should be proportional if the two segments have a common origin. A violation of this rule indicates that the histories of the two segments are different, i.e. there has been a reassortment event. Therefore, when the distances between two segments of different viruses are plotted against each other, the points corresponding to the possible reassortment events lie off the diagonal.

Assuming similar average substitution rates for the third codon positions in all segments, the probability that the distances between two viruses coming from a common ancestor are equal to d^{i}_{ab} , follows a hypergeometric distribution:

$$P_{ab}(d^i_{ab}, L^i + L^j, d^i_{ab} + d^j_{ab}) = \frac{\binom{L^i}{d^j_{ab}}\binom{L^j}{d^j_{ab}}}{\binom{L^i + L^j}{d^j_{ab} + d^j_{ab}}},$$

where L_i and L_j are the respective lengths of the segments. Hence, the probability of observing a difference of no more than d^i_{ab} will be the cumulative of the hypergeometric distribution. In this way, we are able to model the probability of any two segments to have a common ancestor, i.e. lower probabilities indicate reassortant pairs. As many sequences are compared, one has to correct for multiple hypotheses testing. We generate 100 sets of permutated sequences by randomly permuting the third codon positions of any two segments of each strain. Pairs of sequences lying outside the region covered by the permutated set indicate a reassortment event.



Figure 1: Differences in PB2 vs PB1 in third codon position. Points outside the diagonal line (p-values $< 10^{-6}$) indicate a possible reassortment with the logarithm of the statistical value indicated by the color.

Finally, we generate for each strain a list of strains with which it has a low probability of having a common ancestor, hinting to reassortment events between them. We further investigate the origin of the segments by the NCBI BLAST tool.

Results and Discussion



Figure 2: a) Diversity in swine influenza viruses measured by Rao's quadratic entropy. Segments 6 (NA), 4 (HA) and 2 (PB1) present a higher diversity than the rest. B) Diversity in swine H1N1 influenza viruses measured by Rao's quadratic entropy. When the HA and NA type are fixed, segment 2 (PB1) shows a higher diversity than the rest.

Viruses present an enormous diversity due to their high mutation rates, short replication times and high number of replicates. There are several ways of measuring the diversity of a viral population: richness, evenness, Rao's entropy, Shannon entropy and other Renyi entropies, etc [Rao1982]. All these measures encounter similar problems when applied to actual viral populations: the sampling is biased (for instance, in human influenza most of the samples come from a few studies in New York state and New Zealand

[NCBI]), the exponential growth and bottleneck structures of viral populations, population stratification, etc. Although the interpretation of these measures applied to highly structured populations is not clear, one can use them to compare the variation of diversity in the different genes of a particular organism. Since similar histories imply similar diversity measures, strong differences in these measures along the genome indicate different histories for each gene.

Although the third codon evolutionary rates in influenza A viruses are thought to be similar in all segments, the analysis of the genomic diversity (figure 2) reveals a very inhomogeneous pattern. The Rao's quadratic entropy for the 142 complete sequences indicates a statistically significant difference between HA, NA, PB2, PB1 and PA compared to NP, M, and NS segments. Figure 2 (a) shows the diversities and their 95% bootstrap percentile confidence intervals. Within a particular subtype, PB1 appears as the most diverse gene. Figure 2(b) shows the classic H1N1 strains that were isolated in the 70's, 80's and 90's. This analysis shows that the segments do not have a common history, i.e. reassortments are common in influenza A viruses found in pigs, and segments PB1, HA and NA present a higher level of diversity than the rest.

Perhaps more indicative of what is the cause of the variable diversity, is the pair-wise Pearson correlation between the distances between the eight segments. Correlations, linear or non-linear, or any other measure of dependence, as mutual information, encounter the same problems as the measures of diversity (sampling bias, population stratification, etc), but nonetheless provide interesting information. When considering all swine isolates, the stronger correlations appear to be between the polymerase segments, in accordance with previous results *in vitro* [Lubeck1979] and *in vivo* [Rabadan2008], that indicate that in reassortments polymerase segments are likely to travel together (see figure 3a). More interesting is the fact that, when considering a particular subtype, PB1 has a low correlation relative to the other segments. This is evident, among the classical swine H1N1 sequences isolated in the 70's, 80's and 90's (see figure 3b).



Figure 3: Pearson correlation between third codon position Hamming distances between the different segments. In figure a, we can see that polymerases form a higher correlated group in accordance to previous results in vitro (Palese). When we fix the HA and NA subtype (figure b), we can appreciate how PB1 presents a distinctive role.

The above observations from diversity measures and correlations hint to a distinct evolutionary behavior in the HA, NA, and PB1. To elucidate the role of the process of reassortment in these patterns, we have enumerated the reassortment events that can be found in pigs. Many of these events have been already reported in independent publications (see last column in Table 1). We have used the hypergeometric distribution analysis of [Rabadan2008] to identify reassortment events. A much more complicated question is how to disentangle the reassortment history of each segment. Reassortement events are frequent in swine and sampling is not, creating a set of possible compatible reassortement histories. We have analyzed all the complete sequences from swine influenza A isolates since 1930, identify the possible reassortment events and, when possible, reconstruct the events. We have summarized some of these events in Table 1. The notation is S:swine, A:avian, S/T: swine/turkey and H:human, and the colors are blue for the most common set of segments, then red and then green. Simple inspection of Table 1 reveals the frequent role of PB1, HA and NA in the reassortment process:

Year	Strain	PB2	PB1	PA	HA	NP	NA	MP	NS	Reference
1998	A/Swine/North Carolina/35922/98(H3N2)	S	Н	S	Н	S	Н	S	S	[Zhou1999]
1998	A/Swine/Texas/4199-2/98(H3N2), A/Swine/Iowa/569/99 (H3N2)	А	Н	А	Н	S	Н	S	s	[Zhou1999] [Karasin2000]
2002	A/Swine/Spain/39139/2002 (H3N2)	S1	S1	S1	S2	S1	S2	S1	S1	
2002	A/Swine/Spain/42386/2002 (H3N2)	S1	S3	S1	S2	S1	S2	S1	S1	
2003	A/Swine/Ontario/53518/03 (H1N1)	S2	Н	S1	S1	S1	S1	S1	S1	[Karasin2006]
2004	A/Swine/MI/PU243/04 (H3N1)	S1	S1	S1	S1	S1	S2	S1	S1	[Lekcharoensu k2006]
2004	A/Swine/Zhejiang/1/2004 (H1N2)	S	S	S	S	S	Н	S	S	[Qi2006]
2004	A/Swine/Ontario/11112/04 (H1N1)	S	Н	S	S	S	S	S	S	[Karasin2006]
2004	A/Swine/Korea/CAN01/2004 (H1N1)	S2	Н	S2	S1	S1	S1	S1	S1	
2005	A/Swine/Alberta/14722/2005 (H3N2)	S	S	S	S/T	S	Н	S	S	[Olsen2006]
2005	A/Swine/Doetlingen/IDT4735/2005 (H1N2)	S1	S1	S1	Н	S1	S2	S1	S1	[Zell2008]
2006	A/Swine/Missouri/2124514/2006 (H2N3)	S1	S2	А	А	S1	А	S1	S1	[Ma2007]
2006	A/Swine/Miyazaki/1/2006 (H1N2)	S1	S1	S1	S1	S1	S2	S1	S1	[Saito2008]

Summarizing, we have found from the diversity analysis, the correlations and the enumeration of inter-host reassortment events in swine that not every segment reassorts in the same fashion. As it was already observed in human influenza A viruses, *in vivo* [Rabadan2008] and *in vitro* [Lubeck1979], some segments tend to reassort together (polymerases) and some of the segments present a higher reassortment rate (HA and NA). Perhaps, the most interesting result of our analysis is the characteristic role of one of the polymerases (PB1) that frequently reassorts in inter-host mixing. Interestingly enough, this is the same pattern that occurred in at least the 1957 and 1968 pandemics when the pandemic strain, mostly human, obtained an avian PB1.

It is not clear what are the mechanisms behind the preferential reassortments. Several hypotheses can be advanced: biases in the packaging of the vRNA into the virion [Marsh2007], and compensatory mutations. As the different proteins encoded in the viral genome interact, coding mutations in one segment could be compensated in other segments. It is interesting to note that PB1 is very conserved at the protein level, presenting more than 60% identity between influenza A and B, while the rest of the segments are around 35% identical. Gene exchange with distant viruses (distant avian and human viruses, for instance) would be easier in more conserved structures.

References

[Lindstrom 2004] Lindstrom SE, Cox NJ, Klimov A. Genetic analysis of human H2N2 and early H3N2 influenza viruses, 1957-1972: evidence for genetic divergence and multiple reassortment events. Virology. 2004 Oct 10;328(1):101-19.

[CDC2009] http://www.cdc.gov/swineflu/.

[WHO2009-1] http://www.who.int/csr/disease/swineflu/en/index.html.

[WHO2009-2] Cumulative Number of Confirmed Human Cases of Avian Influenza A/(H5N1) Reported to WHO (April 21st 2009).

[Nakajima1978] Nakajima K, Desselberger U, Palese P. Recent human influenza A (H1N1) viruses are closely related genetically to strains isolated in 1950. *Nature*. 1978 Jul 27;274(5669):334–339.

[Scholtissek1978] Scholtissek C, von Hoyningen V, Rott R. Genetic relatedness between the new 1977 epidemic strains (H1N1) of influenza and human influenza strains isolated between 1947 and 1957 (H1N1). *Virology*. 1978 Sep;89(2):613–617.

[Scholtissek1990] Scholtissek C. 1990. Pigs as the 'mixing vessel' for the creation of

new pandemic influenza A viruses. Med Princip Prac, 2, 65-71.

[Webby2000] Webby RJ, Swenson SL, Krauss SL, Gerrish PJ, Goyal SM, Webster RG. Evolution of swine H3N2 influenza viruses in the United States. J Virol. 2000 Sep;74(18):8243-51.

[Zhou1999] Zhou NN, Senne DA, Landgraf JS, Swenson SL, Erickson G, Rossow K, Liu L, Yoon K, Krauss S, Webster RG. Genetic reassortment of avian, swine, and human influenza A viruses in American pigs. J Virol. 1999 Oct;73(10):8851-6.

[NCBI] http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html.

[Lubeck1979] Lubeck MD, Palese P, Schulman JL. Nonrandom association of parental genes in influenza A virus recombinants. Virology. 1979 May;95(1):269-74.

[Rabadan2008] Rabadan R, Levin AJ, Krasnitz M, Non-random reassortment in human influenza A viruses. 2008 Influenza and Other Respiratory Viruses. Volume 2 Issue 1:9-22.

[Rao1982] Rao CR, Diversity and dissimilarity coefficients: a unified approach. 1982 Theor. Popul. Biol. 21.

[Karasin2000] Genetic characterization of H3N2 influenza viruses isolated from pigs in North America, 1977-1999: evidence for wholly human and reassortant virus genotypes.

Karasin AI, Schutten MM, Cooper LA, Smith CB, Subbarao K, Anderson GA, Carman S, Olsen CW. Virus Res. 2000 Jun;68(1):71-85.

[Olsen2006] Olsen, C.W., Karasin, A.I., Carman, S., Li, Y., Bastien, N., Ojkic, D., Alves, D., Charbonneau, G., Henning, B.M., Low, D.E., Burton, L. and Broukhanski, G.

Emerging Infect. Dis. 12 (7), 1132-1135 (2006)

[Ma2007] Ma,W., Vincent,A.L., Gramer,M.R., Brockwell,C.B., Lager,K.M., Janke,B.H., Gauger,P.C., Patnayak,D.P., Webby,R.J. and Richt,J.A. Proc. Natl. Acad. Sci. U.S.A. 104 (52), 20949-20954 (2007) [Lekcharoensuk2006] Novel swine influenza virus subtype H3N1, United States. Lekcharoensuk P, Lager KM, Vemulapalli R, Woodruff M, Vincent AL, Richt JA.

Emerg Infect Dis. 2006 May;12(5):787-94.

[Zell2008] Zell,R., Motzke,S., Krumbholz,A., Wutzler,P., Herwig,V. and Durrwald,R.

Novel reassortant of swine influenza H1N2 virus in Germany

J. Gen. Virol. 89 (PT 1), 271-276 (2008)

[Qi2006]Genetic characterization of novel reassortant H1N2 influenza A viruses isolated from pigs in southeastern China. Qi X, Lu CP. Arch Virol. 2006 Nov;151(11):2289-99. Epub 2006 Jun 8.

[Saito2008]Molecular characterization of an H1N2 swine influenza virus isolated in Miyazaki, Japan, in 2006. Saito T, Suzuki H, Maeda K, Inai K, Takemae N, Uchida Y, Tsunemitsu H. J Vet Med Sci. 2008 Apr;70(4):423-7.

[Karasin2006]Identification of human H1N2 and human-swine reassortant H1N2 and H1N1 influenza A viruses among pigs in Ontario, Canada (2003 to 2005). Karasin AI, Carman S, Olsen CW. J Clin Microbiol. 2006 Mar;44(3):1123-6.

[Marsh2007] Glenn A. Marsh, Raul Rabadan, Arnold J. Levine, Peter Palese . Highly conserved regions of the influenza A virus polymerase gene segments are critical for efficient vRNA packaging. J Virol. 2007 Dec 19.