DOI:10.1111/irv.12117
www.influenzajournal.com

Original Article

A statistical strategy to identify recombinant viral ribonucleoprotein of avian, human, and swine influenza A viruses with elevated polymerase activity

Alex W. H. Chin,[a] Benjamin D. Greenbaum,[b] Olive T. W. Li,[a] Richard J. Webby,[c] Leo L. M. Poon[a]

[a]Centre of Influenza Research & School of Public Health, The University of Hong Kong, Hong Kong. [b]The Simons Center for Systems Biology, Institute for Advanced Study, Princeton, NJ, USA. [c]Department of Infectious Diseases, Division Virology, St. Jude Children's Research Hospital, Memphis, TN, USA.

Correspondence: Leo L. M. Poon, School of Public Health, LKS Faculty of Medicine, The University of Hong Kong, 21 Sassoon Road, Hong Kong City, Hong Kong. E-mail: llmpoon@hkucc.hku.hk

Accepted 30 March 2013. Published Online 02 May 2013.

Objectives Reassortment of influenza A viruses can give rise to viral ribonucleoproteins (vRNPs) with elevated polymerase activity and the previous three pandemic influenza viruses contained reassorted vRNPs of different origins. These suggest that reassorted vRNP may be one of the factors leading to a pandemic virus. In this study, we reconstituted chimeric vRNPs with three different viral strains isolated from avian, human and swine hosts. We applied a statistical strategy to identify the effect that the origin of a single vRNP protein subunit or the interactions between these subunits on polymerase activity.

Design Eighty one chimeric vRNPs were reconstituted in 293T cells at different temperatures. Polymerase activity was determined by luciferase reporter assay and the results were analysed by multiway ANOVA and other statistical methods.

Results It was found that PB2, PB1, NP, PB2-PB1 interaction, PB2-PA interaction and PB1-NP interaction had significant effect on polymerase activity at 37°C and several single subunits and interactions were identified to lead to elevation of polymerase activity. Furthermore, we studied 27 out of these 81 different chimieric vRNPs in different combinations via fractional factorial design approach. Our results suggested that the approach can identify the major single subunit or interaction factors that affect the polymerase activity without the need to experimentally reproduce all possible vRNP combinations.

Conclusions Statistical approach and fractional factorial design are useful to identify the major single subunit or interaction factors that can modulate viral polymerase activity.

Keywords Fractional factorial design, influenza virus, polymerase, reassortment.

Please cite this paper as: Chin *et al.* (2013) A statistical strategy to identify recombinant viral ribonucleoprotein of avian, human, and swine influenza A viruses with elevated polymerase activity. Influenza and Other Respiratory Viruses 7(6), 969–978.

Introduction

Influenza A virus is a negative-stranded RNA virus, which belongs to the family Orthomyxoviridae. Its genome consists of eight viral RNA segments. Three of them encode for the three subunits of the viral RNA-dependent RNA polymerase (RdRP), namely polymerase basic 2 (PB2), polymerase basic 1 (PB1), and polymerase acid (PA). The three subunits of the polymerase form a heterotrimer, which associates with nucleoprotein (NP) and a segment of viral RNA to form a viral ribonucleoprotein (vRNP).[1–3]

The segmented genome of influenza virus makes it possible for reassortment of gene segments when coinfection of different virus strains occurs in the same host. Swine have been described as a mixing vessel for the influenza viruses.[4] Therefore, reassortment of avian, human, and swine viruses in swine is possible. The reassortments may result in novel virus strains that can infect human, and some of these zoonotic viruses may occasionally develop into a pandemic strain.[5–7]

Apart from the 1918 pandemic H1N1 virus, where the origin is controversial,[8–11] the subsequent three influenza pandemic strains contain polymerase subunits reassorted from strains of various origins. Both the 1957 H2N2 and 1968 H3N2 pandemic viruses contain PB2 and PA of human origin reassorted with PB1 of avian origin.[7] The pandemic H1N1/2009 strain contains polymerase genes originated from the North American triple reassortant swine lineage, of which the PB2 and PA genes are of avian origin and the PB1 gene is of human origin.[12–14] It is possible that, in addition to acquiring a novel HA or/and NA surface protein, a virus with polymerase subunits of different origins may have an increased pandemic potential. Thus, it is necessary to

develop a more systematic strategy to analyze chimeric polymerase complexes with genes from multiple sources.

Previously, we used a 2-level full factorial experimental design with four factors to determine the polymerase activities of all possible reassorted vRNPs of two different strains.[15] However, if more than two strains were allowed to reassort in the same host population, more combinations of reassorted vRNPs would result. This increases the difficulty in comprehending the data generated from a large number of chimeric vRNPs. Hence, we would like to apply statistical methods to analyze the polymerase activities of reassorted vRNPs originated from three strains of different origins. In particular, we aim to test the feasibility of using this strategy to determine whether temperature, the origin of single subunit of vRNP, and/or interactions between subunits of vRNP affect the polymerase activity.

Methods

Plasmids

The gene segments of PB2, PB1, PA, and NP of three influenza viruses of different origins (avian, human, and swine) were cloned into pcDNA vector. The avian, human, and swine viruses used were A/Indonesia/5/2005 (H5N1), A/Puerto Rico/8/1934 (H1N1), and A/swine/Texas/4199-2/1998 (H3N2), respectively. Plasmid expressing vRNA-like luciferase reporter (pPolI-Luc-NS) and GFP (pGFP) was prepared as described.[15]

Reconstitution of chimeric vRNPs

Chimeric vRNPs were reconstituted in human embryonic kidney (293T) cells obtained from ATCC. The 293T cells were maintained in minimum essential medium (MEM) supplemented with 10% fetal bovine serum (Life Technologies, Carlsbad, CA, USA).

Different combinations of plasmid mixtures were prepared by mixing pcDNA plasmids expressing PB2, PB1, PA, and NP proteins from either avian, human, or swine strain, together with pPolI-Luc-NS and pGFP. A total of 81 (3^4) different combinations of vRNP subunits were generated. Only pPolI-Luc-NS and pGFP plasmids were used as the negative control reactions. The plasmid mixtures were transfected into 293T cells on 96-well culture plate and were incubated at 33, 37, or 40°C for 48 hours before determination of polymerase activity by luciferase reporter assay.

Luciferase reporter assay

Transfected cells were lysed by Steady-Glo luciferase assay substrate solution (Promega, Madison, WI, USA) for 5 minutes. The luciferase activities were determined by the luminescence measured using a luminometer (PerkinElmer, Waltham, MA, USA), and the data were normalized with expression level of GFP as described.[15]

Statistical analysis

All the data obtained in this study were expressed as means of three individual experiments. Effect of temperature on distributions of the polymerase activities was determined by two-sample Kolmogorov–Smirnov test performed with MATLAB (The MathWorks Inc., Natick, MA, USA). Single vRNP subunit effect and multifactor interaction of 81-combination full factorial design or 27-run fractional factorial designs were determined by multiway analysis of variance (ANOVA) performed with IBM SPSS Statistics Premium (SPSS Inc., Armonk, NY, USA). Significant difference between two groups of data was determined by *t*-test. Unless otherwise stated, effect was considered significant with $P < 0{\cdot}05$.

Full factorial and fractional factorial designs

Full factorial design was adopted to investigate all the possible vRNP combinations of the three viral strains of different origins. This full factorial design consisted of four factors, which were the subunits of vRNP including PB2, PB1, PA, and NP. Each factor consisted of three levels that were the three different origins of the subunits. This resulted in 81 (3^4) different combinations in the full factorial design. In addition, a 27-run (3^{4-1}) fractional factorial design was applied to a subset of vRNP combinations from the full factorial design. There are 24 different ways in selecting these vRNPs (see Results section).[16,17] Among the selected chimeric vRNPs ($n = 27$) in a subset, nine of them must have a subunit from the same strain (e.g. human PB2), and three of them must have a specific subunit–subunit combination (e.g. human PB2–swine PB1).

Results

Determination of polymerase activity of chimeric vRNPs

Full factorial design was applied in reconstituting chimeric vRNPs of three different strains with different origins. Eighty-one different chimeric vRNPs were reconstituted in 293T cells at 33, 37, or 40°C. The polymerase activities of these chimeric vRNPs were determined by the luciferase reporter assay. The normalized data obtained were the means of three independent experiments ($n = 3$) and were expressed as relative polymerase activity in relative to the polymerase activity of wild-type vRNP complex of human origin (HHHH) at corresponding temperatures (Figures 1–3).

Effect of temperature on polymerase activity

To investigate whether incubation temperature has any effect on the polymerase activity, the polymerase activities of wild-type vRNPs of avian, human, and swine origin at 33, 37, and 40°C were compared. At 33 and 37°C, the polymerase activity of wild-type vRNP of human origin was the highest among the three vRNPs, while the polymerase activity of wild-type vRNP of avian origin was the lowest. However, at

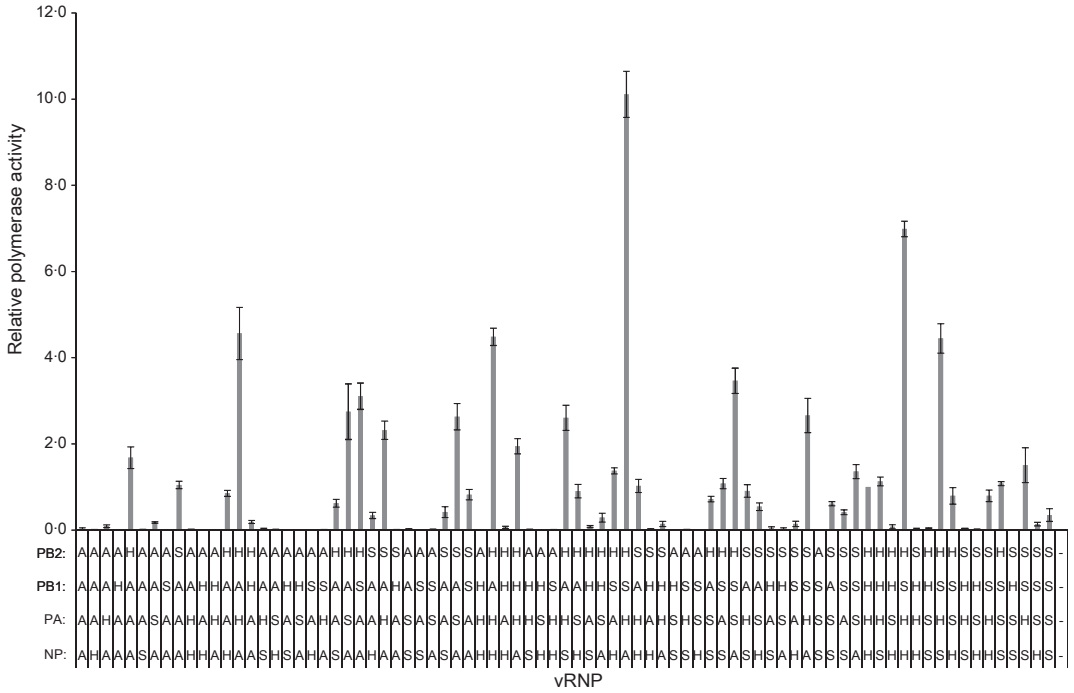

Figure 1. Relative polymerase activity of 81 recombinant viral ribonucleoprotein (vRNP) at 33°C. Data are expressed as mean relative polymerase activity in relative to the polymerase activity of wild-type human origin vRNP complex (HHHH). Error bars represent one standard deviation (n = 3).

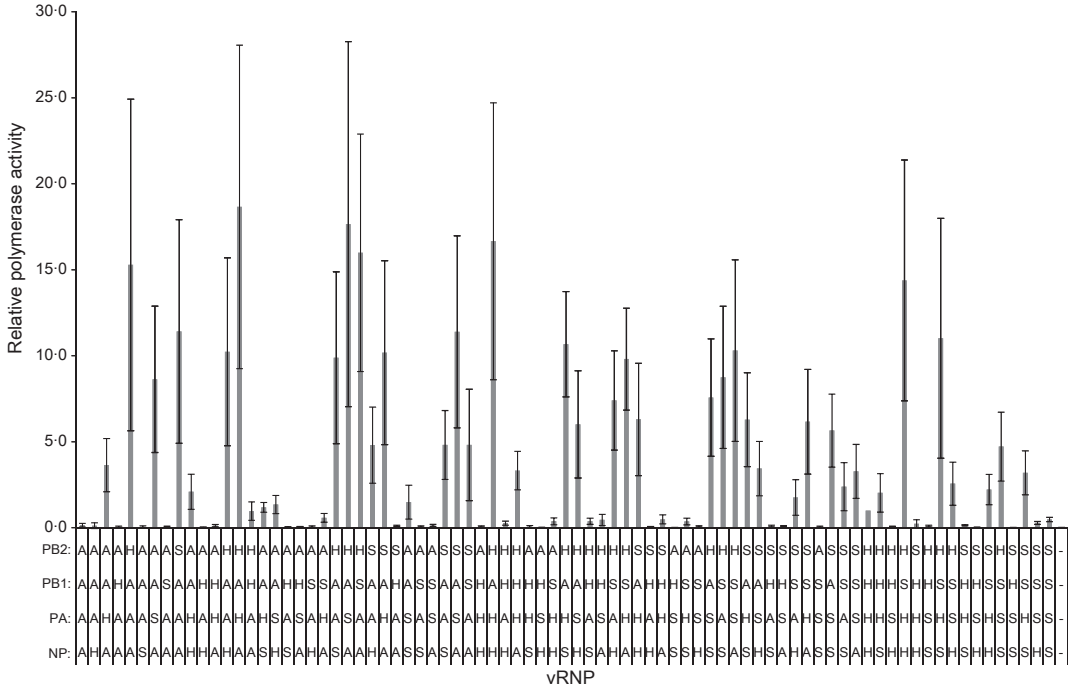

Figure 2. Relative polymerase activity of 81 recombinant viral ribonucleoproteins (vRNPs) at 37°C. Data are expressed as mean relative polymerase activity in relative to the polymerase activity of wild-type human origin vRNP complex (HHHH). Error bars represent one standard deviation (n = 3).

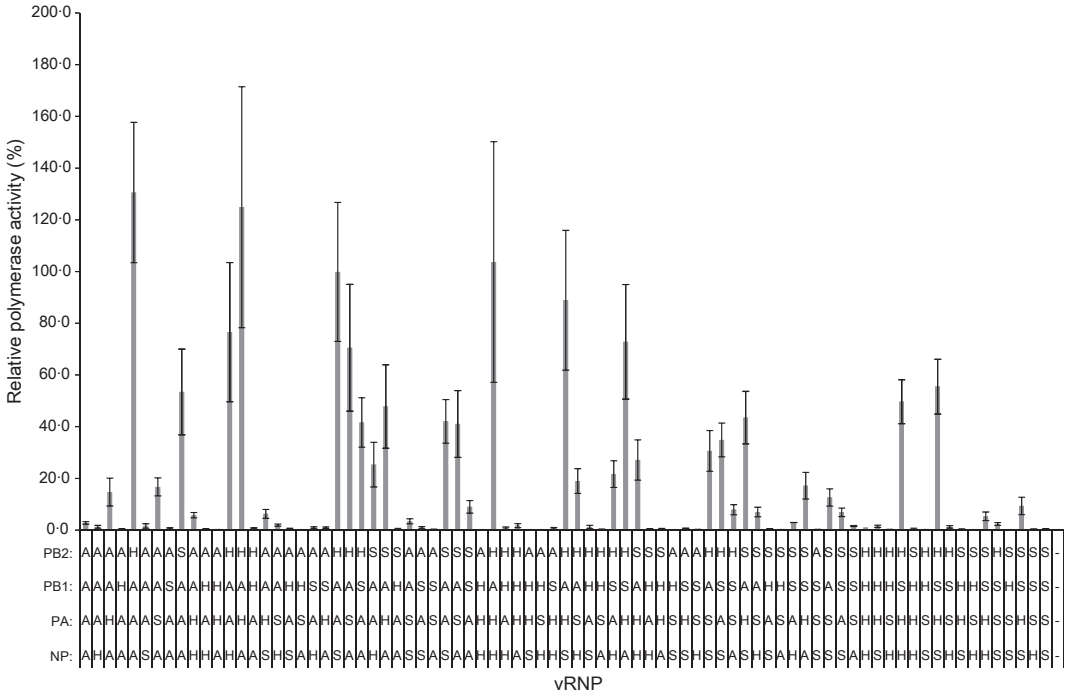

Figure 3. Relative polymerase activity of 81 recombinant viral ribonucleoprotein (vRNPs) at 40°C. Data are expressed as mean relative polymerase activity in relative to the polymerase activity of wild-type human origin vRNP complex (HHHH). Error bars represent one standard deviation (n = 3).

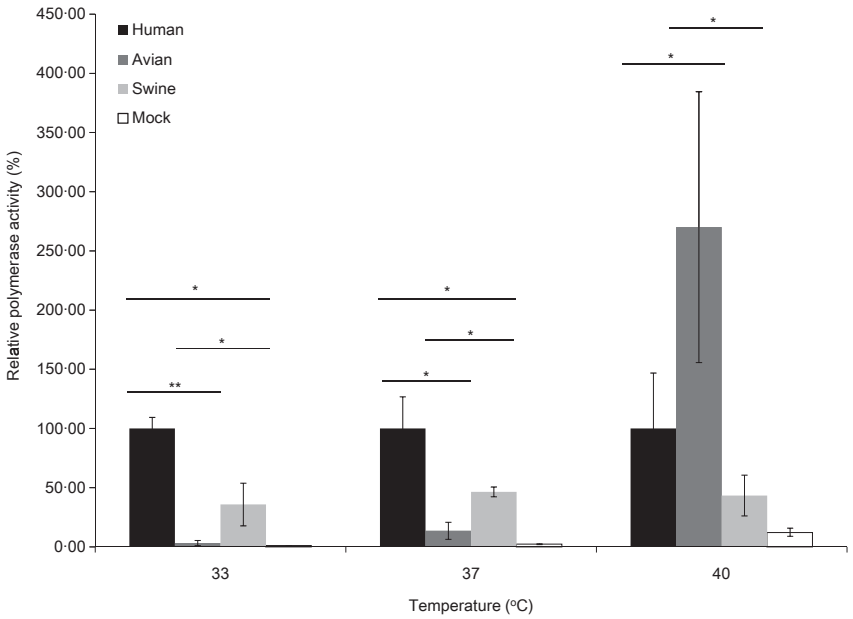

Figure 4. Mean relative polymerase activity of wild-type vRNP complexes of three different origins at different temperatures. Polymerase activity of vRNPs of wild-type human origin (black), avian origin (dark gray), swine origin (light gray), or no vRNP (mock, unfilled) reconstituted in human 293T cells at 33, 37, or 40°C was compared. The normalized data are expressed as mean relative polymerase activity in relative to wild-type human origin polymerase activity at the respective temperatures. Error bars represent one standard deviation (n = 3; *P < 0·05; **P < 0·001, by t-test).

40°C, the polymerase activity of wild-type vRNP of avian origin became the highest (Figure 4).

Additionally, the two-sample Kolmogorov–Smirnov test was carried out to analyze whether the distributions of the polymerase activities of the 81 different chimeric vRNPs obtained differed significantly at the three temperatures. The result showed that there was significant difference between the two distributions of relative polymerase activity pairs for

all pairs (33 and 37°C, 33 and 40°C, or 37 and 40°C; Table 1). Overall, these results agreed with previous findings that different vRNPs have different temperatures in which they work best.[15,18,19]

Effect of the origin of a single vRNP subunit on polymerase activity at 37°C

Next, we investigated whether the origin of a single subunit of a chimeric vRNP has significant effect on polymerase activity at different temperatures. A multiway analysis of variance (ANOVA) was carried out to compare whether, at a given temperature, the polymerase activity significantly varied with the origins of vRNP subunits. It was found that the origin of PB2 or PB1 played a significant role on polymerase activity at all three temperatures ($P < 0{\cdot}05/4 = 0{\cdot}0125$, accounting for multiple hypotheses because there were four segments whose *P*-values were being calculated). However, the origin of PA had significant effect on polymerase activity only at 33 or 40°C, but not at 37°C, while the origin of NP had significant effect on polymerase activity only at 37°C (Table 2). This suggested PB2 and PB1 from the studied strains had more stringent temperature requirements. Nonetheless, these results reiterated that the origins of PA and NP might have different effects on polymerase activity at different temperatures, again demonstrating that temperature had a prominent effect on the polymerase activity of these chimeric vRNPs.

The normal temperature of human lower respiratory tract is 37°C. A virus that can replicate efficiently in human lungs increases the risk to result in more severe illness.[20] Therefore, we conducted a more in-depth analysis on the data generated at 37°C. As demonstrated above, the origin of PB2, PB1, or NP had a significant effect on polymerase activity at this temperature. To test whether there is any significant change in polymerase activity when human origin PB2 is replaced by avian or swine origin PB2, we compared the mean polymerase activity of chimeric vRNPs with PB2 of these three origins. The significance of the differences in polymerase activity was determined by *t*-test. It was demonstrated that chimeric vRNP with PB2 of avian or swine origin had significantly lower polymerase activity than chimeric vRNP with PB2 of human origin. Similar analyses were also carried out on PB1 and NP. It was found that both avian and swine origin PB1 would lead to significant increase in the polymerase activity of chimeric vRNPs when they replaced PB1 of human origin. For NP, only avian origin NP would increase the polymerase activity of chimeric vRNPs significantly when they replaced NP of human origin (Tables 2 and 3). Swine PB2 and PB1 were found to have significant effects on the polymerase activity, but it should be noted that the swine influenza virus used in this study is the North American triple reassortant swine virus, which is a reassortant of avian, swine, and human viruses. Although this virus

Table 1. Comparison on the distributions of relative polymerase activities of all chimeric viral ribonucleoprotein complexes between different temperature pairs by Kolmogorov–Smirnov test

Distribution of relative polymerase activity at temperatures	Significance
33°C versus 37°C	8·21E-05*
33°C versus 40°C	1·66E-08*
37°C versus 40°C	0·0022*

*$P < 0{\cdot}05$.

Table 2. Effect of the origin of single viral ribonucleoprotein (vRNP) subunit on relative polymerase activity at different temperatures determined by multiway ANOVA

	Significance		
vRNP subunit	33°C	37°C	40°C
PB2	0 (A, S)*	0 (A, S)*	0 (A, S)*
PB1	0·0004 (A, S)*	0 (A, S)*	0 (A, S)*
PA	0·0002 (A, S)*	0·0674	0·0055 (S)*
NP	0·0194	0·0006 (A)*	0·0943

NP, nucleoprotein; PA, polymerase acid; PB1, polymerase basic 1; PB2, polymerase basic 2.
Origins (A: avian; S: swine) of vRNP subunit that have significant difference from the subunit of human origin are indicated in brackets.
*$P < 0{\cdot}0125$.

Table 3. Comparisons of mean relative polymerase activities of chimeric viral ribonucleoprotein (vRNP) with PB2, PB1, or NP of avian, human, or swine origin at 37°C

	Mean relative polymerase activity** at 37°C		
vRNP subunit	Avian	Human	Swine
PB2	0·7828*	7·6389	3·3435*
PB1	7·2488*	0·3891	4·1273*
NP	5·7014*	3·0386	3·0251

NP, nucleoprotein; PB1, polymerase basic 1; PB2, polymerase basic 2.
**The value represents the average polymerase activity from vRNPs containing a subunit derived from a specified host. The activities are expressed in relative to the polymerase activity of wild-type vRNP of human origin (HHHH).
*$P < 0{\cdot}05$.

Table 4. Effect of the interactions between the origins of two or three viral ribonucleoprotein (vRNP) subunits on relative polymerase activity at 37°C determined by multiway ANOVA

Interaction	Significance at 37°C
PB2*PB1	0*
PB2*PA	0·0002*
PB2*NP	0·0204
PB1*PA	0·0281
PB1*NP	0·0001*
PA*NP	0·0695
PB2*PB1*PA	0·3858
PB2*PBA*NP	0·7335
PB2*PA*NP	0·4089
PB1*PA*NP	0·1557

NP, nucleoprotein; PA, polymerase acid; PB1, polymerase basic 1; PB2, polymerase basic 2.
*$P < 0·00833$.

may have well adapted in swine, PB2 and PA genes of this swine virus were originated from avian virus, while PB1 was originated from human virus.[12]

Effect of interactions between different subunits of chimeric vRNP on polymerase activity at 37°C

Interactions between different subunits of a vRNP may also play a role in affecting polymerase activity. Multiway ANOVA was carried out at each temperature to test for the significance of origins of the six possible segment pairs PB2–PB1, PB2–PA, PB2–NP, PB1–PA, PB1–NP, and PA–NP. For each segment pair to be studied, like PB2–PB1 pair, the 81 chimeric vRNP combinations were divided into nine different groups according to their PB2–PB1 combinations (e.g. avian PB2–avian PB1, avian PB2–human PB1, etc.). The polymerase activities of these groups of chimeric vRNPs were analyzed by multiway ANOVA to determine whether change in origins of the studied segment pair would lead to a significant change in polymerase activity. It was found that at 37°C, PB2–PB1, PB2–PA, and PB1–NP interactions were significant in affecting the polymerase activity ($P < 0·05/6 = 0·00833$, accounting for multiple hypotheses because there were six interactions whose *P*-values were being calculated; Table 4). Of these three interactions, we performed *t*-tests to identify the combinations of the origins of interacting subunits that would significantly increase the polymerase activity as compared to the interaction between the two subunits of human origin. The analysis revealed that for the PB2–PB1 interaction, there were four combinations of PB2–PB1 having enhanced polymerase activity as compared with human origin PB2–PB1 (HHxx). These combinations were human origin PB2 with avian/swine origin PB1 (HAxx and HSxx) and swine origin PB2 with avian/swine origin PB1 (SAxx and SSxx). For the PB2–PA interaction, all of the other PB2–PA combinations had lower polymerase activity than human origin PB2–PA combination (HxHx). For the PB1–NP interaction, all combinations except human origin PB1 with avian/swine origin NP (xHxA and xHxS) had enhanced polymerase activity when compared with human origin PB1–NP combination (xHxH; Table 5).

Furthermore, the effect of interactions among three vRNP subunits on polymerase activity was also analyzed. However, there was no significant interaction among three vRNP subunits found at 37°C (Table 4).

Factors and interactions that enhance polymerase activity at both 33 and 37°C

The normal temperature of human upper respiratory tract is 33°C. If a virus has more robust polymerase activities at both 33 and 37°C, it is likely that the virus might have a better

Table 5. Comparisons of mean relative polymerase activities** of chimeric viral ribonucleoprotein (vRNP) with different PB2–PB1, PB2–PA, or PB1–NP combinations of avian, human, or swine origin at 37°C

PB2–PB1 interaction	Avian PB2	Human PB2	Swine PB2
Avian PB1	2·0848	12·5160*	7·1454*
Human PB1	0·0593	*0·9600*	0·1481
Swine PB1	0·2043	9·4405*	2·7369*
PB2–PA interaction	**Avian PB2**	**Human PB2**	**Swine PB2**
Avian PA	0·0779	7·6878	3·3679
Human PA	0·9492	*9·7304*	3·9167
Swine PA	1·3213	5·4984	2·7458
PB1–NP interaction	**Avian PB1**	**Human PB1**	**Swine PB1**
Avian NP	10·7811*	0·6370	5·6861*
Human NP	5·6714*	*0·2030*	3·2413*
Swine NP	5·2938*	0·3273	3·4543*

NP, nucleoprotein; PA, polymerase acid; PB1, polymerase basic 1; PB2, polymerase basic 2.
Mean relative polymerase activity with PB2–PB1, PB2–PA, or PB1–NP combinations of both human origins is indicated in italics.
Combinations with mean relative polymerase activity significantly higher than combination of both human origin are indicated with asterisks ($P < 0·05$).
**The value represents the average polymerase activity from vRNPs containing a subunit derived from a specified host. The activities are expressed in relative to the polymerase activity of wild-type vRNP of human origin (HHHH).

Table 6. Design table for obtaining 24 different subsets of chimeric viral ribonucleoprotein (vRNP) in 27-run fractional factorial designs

vRNP combinations	PB2	PB1	PA	Fractional factorial design																							
				1 NP	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	A	A	A	A	A	H	H	S	S	A	A	H	H	S	S	A	A	H	H	S	S	A	A	H	H	S	S
2	A	A	H	H	H	S	S	A	A	H	H	S	S	A	A	S	S	A	A	H	H	S	S	A	A	H	H
3	A	A	S	S	S	A	A	H	H	S	S	A	A	H	H	H	H	S	S	A	A	H	H	S	S	A	A
4	A	H	A	H	H	S	S	A	A	S	S	A	A	H	H	H	H	S	S	A	A	S	S	A	A	H	H
5	A	H	H	S	S	A	A	H	H	A	A	H	H	S	S	A	A	H	H	S	S	H	H	S	S	A	A
6	A	H	S	A	A	H	H	S	S	H	H	S	S	A	A	S	S	A	A	H	H	A	A	H	H	S	S
7	A	S	A	S	S	A	A	H	H	H	H	S	S	A	A	S	S	A	A	H	H	H	H	S	S	A	A
8	A	S	H	A	A	H	H	S	S	S	S	A	A	H	H	H	H	S	S	A	A	A	A	H	H	S	S
9	A	S	S	H	H	S	S	A	A	A	A	H	H	S	S	A	A	H	H	S	S	S	S	A	A	H	H
10	H	A	A	H	S	A	S	A	H	H	S	A	S	A	H	H	S	A	S	A	H	H	S	A	S	A	H
11	H	A	H	S	A	H	A	H	S	S	A	H	A	H	S	A	H	S	H	S	A	A	H	S	H	S	A
12	H	A	S	A	H	S	H	S	A	A	H	S	H	S	A	S	A	H	A	H	S	S	A	H	A	H	S
13	H	H	A	S	A	H	A	H	S	A	H	S	H	S	A	S	A	H	A	H	S	A	H	S	H	S	A
14	H	H	H	A	H	S	H	S	A	H	S	A	S	A	H	H	S	A	S	A	H	S	A	H	A	H	S
15	H	H	S	H	S	A	S	A	H	S	A	H	A	H	S	A	H	S	H	S	A	H	S	A	S	A	H
16	H	S	A	A	H	S	H	S	A	S	A	H	A	H	S	A	H	S	H	S	A	S	A	H	A	H	S
17	H	S	H	H	S	A	S	A	H	A	H	S	H	S	A	S	A	H	A	H	S	H	S	A	S	A	H
18	H	S	S	S	A	H	A	H	S	H	S	A	S	A	H	H	S	A	S	A	H	A	H	S	H	S	A
19	S	A	A	S	H	S	A	H	A	S	H	S	A	H	A	S	H	S	A	H	A	S	H	S	A	H	A
20	S	A	H	A	S	A	H	S	H	A	S	A	H	S	H	H	A	H	S	A	S	H	A	H	S	A	S
21	S	A	S	H	A	H	S	A	S	H	A	H	S	A	S	A	S	A	H	S	H	A	S	A	H	S	H
22	S	H	A	A	S	A	H	S	H	H	A	H	S	A	S	A	S	A	H	S	H	H	A	H	S	A	S
23	S	H	H	H	A	H	S	A	S	S	H	S	A	H	A	S	H	S	A	H	A	A	S	A	H	S	H
24	S	H	S	S	H	S	A	H	A	A	S	A	H	S	H	H	A	H	S	A	S	S	H	S	A	H	A
25	S	S	A	H	A	H	S	A	S	A	S	A	H	S	H	H	A	H	S	A	S	A	S	A	H	S	H
26	S	S	H	S	H	S	A	H	A	H	A	H	S	A	S	A	S	A	H	S	H	S	H	S	A	H	A
27	S	S	S	A	S	A	H	S	H	S	H	S	A	H	A	S	H	S	A	H	A	H	A	H	S	A	S

NP, nucleoprotein; PA, polymerase acid; PB1, polymerase basic 1; PB2, polymerase basic 2.
Each fractional factorial design contains 27 chimeric vRNPs (vRNP combinations 1–27) comprised of the columns PB2, PB1, PA, and one of the column of NP (fractional factorial design 1–24).

chance to adapt in human. Thus, we tried to identify the factors and interactions that could enhance viral polymerase activity at 33°C (data not shown). We observed that chimeric vRNPs with swine PB1 or avian PB1 subunit could significantly enhance viral polymerase at both 33 and 37°C. In addition, of all the interactions identified at this dataset obtained from 33°C, only viral polymerases with the combination of human PB2–avian PB1 (HAxx), human PB2–swine PB1 (HSxx), or swine PB2–avian PB1 (SAxx) could enhance viral polymerase at both 33 and 37°C.

Possibility of using a fractional factorial design to detect the single subunit effect or two-factor interactions

Next, we investigated the possibility of using a fraction of the 81 chimeric vRNPs dataset obtained at 37°C to reflect the single subunit effect and effect of interaction between subunits on polymerase activity. A 27-run (3^{4-1}) fractional factorial design was applied as a model to select a subset of 27 chimeric vRNPs of the 81 combinations.[16,17] Total 24 different subsets of data could be selected (Table 6). Multiway ANOVA was applied on these 24 subsets of data as well as the 81-combination full factorial design. The values of *F*-statistics, which reflect the significance, for all single subunit effects and two-factor interactions were compared and ranked in each subset and the 81-combination full factorial design. The average rank of the 24 subsets was calculated for each of the single subunit effect and two-factor interaction. It was found that the average rank on the significance of single subunit effect and two-factor interactions of the 24 subsets of fractional factorial designs was similar to the rank in the 81-combination full factorial design. The results obtained showed that PB2 and PB1 were the most significant factors affecting polymerase activity determined by both full and fractional factorial designs, while PA–NP interaction, PB1–PA interaction, and PB2–NP

Table 7. *F*-statistics of multiway ANOVA for single subunit effect and two-factor interactions of the 81-combination full factorial design and the 27-run fractional factorial designs

	Value of *F*-statistics (rank)									
	PB2	**PB1**	**PA**	**NP**	**PA*NP**	**PB1*NP**	**PB2*NP**	**PB1*PA**	**PB2*PA**	**PB2*PB1**
Full factorial	100·088 (1)	98·363 (2)	6·14 (5)	19·81 (3)	1·773 (10)	5·769 (6)	2·11 (8)	1·901 (9)	4·199 (7)	19·774 (4)
Fractional factorial design 1	42·254 (1)	25·531 (2)	1·651 (8)	5·467 (4)	1·009 (9)	2·321 (6)	5·107 (5)	2·041 (7)	0·773 (10)	11·179 (3)
Fractional factorial design 2	31·499 (1)	28·766 (2)	2·033 (7)	5·193 (3)	0·51 (10)	1·725 (8)	3·359 (4)	1·277 (9)	2·283 (6)	2·449 (5)
Fractional factorial design 3	25·535 (2)	36·911 (1)	2·913 (6)	5·388 (4)	2·433 (7)	1·643 (8)	0·185 (10)	0·985 (9)	3·348 (5)	6·16 (3)
Fractional factorial design 4	23·126 (2)	32·596 (1)	1·788 (6)	10·004 (3)	0·585 (7)	3·114 (5)	0·213 (10)	0·239 (9)	0·459 (8)	7·546 (4)
Fractional factorial design 5	29·366 (2)	35·891 (1)	2·735 (5)	4·385 (4)	1·423 (8)	0·604 (9)	0·49 (10)	1·926 (7)	2·364 (6)	8·559 (3)
Fractional factorial design 6	36·299 (1)	27·962 (2)	1·661 (7)	8·726 (3)	0·464 (10)	2·539 (5)	0·644 (8)	1·712 (6)	0·571 (9)	2·795 (4)
Fractional factorial design 7	32·959 (2)	33·468 (1)	1·138 (8)	5·347 (4)	1·929 (6)	2·412 (5)	1·068 (9)	1·314 (7)	0·787 (10)	7·212 (3)
Fractional factorial design 8	37·281 (1)	27·991 (2)	5·456 (5)	6·923 (4)	0·761 (8)	0·237 (10)	1·261 (7)	0·404 (9)	3·587 (6)	14·37 (3)
Fractional factorial design 9	28·461 (2)	28·465 (1)	1·756 (7)	6·164 (3)	1·285 (8)	2·569 (6)	0·148 (10)	0·202 (9)	3·613 (5)	5·141 (4)
Fractional factorial design 10	28·745 (2)	31·105 (1)	2·415 (6)	9·828 (3)	0·925 (8)	1·626 (7)	0·824 (9)	0·326 (10)	3·271 (4)	3·039 (5)
Fractional factorial design 11	32·08 (1)	28·333 (2)	2·756 (5)	4·234 (3)	0·637 (9)	2·744 (6)	1·576 (7)	0·154 (10)	1·322 (8)	3·516 (4)
Fractional factorial design 12	28·863 (2)	38·094 (1)	0·211 (10)	5·75 (4)	0·539 (8)	5·034 (5)	1·244 (6)	0·887 (7)	0·308 (9)	9·032 (3)
Fractional factorial design 13	33·012 (1)	28·309 (2)	3·486 (5)	8·607 (4)	0·152 (10)	0·401 (9)	1·789 (7)	0·74 (8)	2·735 (6)	12·453 (3)
Fractional factorial design 14	35·828 (1)	33·535 (2)	2·367 (5)	2·571 (4)	0·182 (8)	0·43 (7)	0·026 (9)	0·975 (6)	0·017 (10)	8·527 (3)
Fractional factorial design 15	34·056 (1)	27·228 (2)	2·375 (5)	5·681 (3)	0·252 (10)	3·882 (4)	0·791 (9)	0·894 (8)	1·744 (7)	1·937 (6)
Fractional factorial design 16	32·189 (2)	35·604 (1)	1·065 (8)	3·739 (4)	0·608 (9)	2·196 (5)	1·936 (7)	0·025 (10)	1·995 (6)	8·281 (3)
Fractional factorial design 17	28·131 (2)	29·188 (1)	1·646 (7)	6·701 (4)	1·503 (8)	3·896 (5)	0·3 (9)	0·182 (10)	2·566 (6)	8·357 (3)

Table 7. (*Continued*)

	Value of *F*-statistics (rank)									
	PB2	**PB1**	**PA**	**NP**	**PA*NP**	**PB1*NP**	**PB2*NP**	**PB1*PA**	**PB2*PA**	**PB2*PB1**
Fractional factorial design 18	24·367 (2)	32·075 (1)	2·381 (5)	12·193 (3)	1·347 (8)	1·456 (7)	0·635 (9)	0·287 (10)	2·198 (6)	3·663 (4)
Fractional factorial design 19	38·657 (1)	27·507 (2)	2·846 (4)	3·322 (3)	0·159 (10)	1·258 (9)	2·55 (7)	1·35 (8)	2·729 (6)	2·763 (5)
Fractional factorial design 20	36·786 (1)	26·228 (2)	1·563 (8)	4·271 (4)	0·813 (10)	2·613 (6)	1·856 (7)	3·362 (5)	1·002 (9)	9·938 (3)
Fractional factorial design 21	28·558 (2)	37·731 (1)	1·536 (6)	8·019 (4)	1·055 (8)	2·447 (5)	0·791 (9)	0·589 (10)	1·291 (7)	9·999 (3)
Fractional factorial design 22	32·407 (1)	26·929 (2)	2·968 (5)	7·952 (3)	1·893 (7)	2·656 (6)	1·23 (9)	1·788 (8)	1·218 (10)	4·7 (4)
Fractional factorial design 23	22·869 (2)	42·971 (1)	0·762 (7)	8·771 (3)	0·549 (9)	3·891 (4)	0·652 (8)	0·218 (10)	2·362 (6)	3·374 (5)
Fractional factorial design 24	29·169 (2)	31·043 (1)	3·32 (5)	7·68 (4)	0·057 (10)	2·495 (6)	0·154 (9)	0·333 (8)	0·445 (7)	7·733 (3)
Average rank of fractional factorial design 1–24	1·542	1·458	6·250	3·542	8·542	6·375	8·083	8·333	7·167	3·708

NP, nucleoprotein; PA, polymerase acid; PB1, polymerase basic 1; PB2, polymerase basic 2.

interaction were the least significant in both designs (Table 7). This suggested that, by running an experiment with a fractional factorial design, it is possible to capture the major factors or interaction effects that would be observed from the results generated from the full factorial design.

Discussion

Previously, evidence has shown that the polymerase activity of an influenza virus is correlated to its rate of replication and host adaptability. Viruses with higher polymerase activity have been shown to have higher rate of replication, indicating that the rate of replication is at least partly influenced by polymerase activity.[19,21,22] In addition, polymerase activity is also believed to have a role to play in allowing the virus to overcome a species barrier and infect a new host.[23] With faster replication and better ability in host adaptation, a virus containing a recombinant vRNP with enhanced polymerase activity may have an enhanced potential to infect and adapt to humans.

Our study has demonstrated a statistical approach analyzing the polymerase activity of recombinant vRNPs of three different strains. We identified several factors including single protein subunits and interactions between two subunits of a chimeric vRNP that would lead to an increase in polymerase activity. This may give us some insights into the correlation between polymerase activity and pandemic potential of a virus. It is important to stress that the critical factors and/or interactions identified in the analysis are highly depended on the strains used in the work. If other avian, human, or swine viruses were used, the results obtained may be different. As the primary objective of this work is to develop a strategy to analysis a large number of chimeric vRNPs, the mechanism of how the interactions between different subunits affect polymerase activity and how the polymerase activity correlates with pandemic potential needs further investigation.

To a further extent, our statistical analysis of polymerase activity of chimeric vRNPs could be applied to vRNP reassortment of more than three strains. For example, if vRNPs from four different strains were allowed to reassort, 256 combinations of reassorted vRNPs could be generated. However, manipulation of 256 or more combinations is resource intensive. For testing chimeric vRNPs with subunits derived from four different strains in a factional factorial designed manner, results from 64 specific combinations of

vRNPs might provide sufficient information for identifying the major single vRNP subunit effect or two-factor interactions. Alternatively, this approach can at least eliminate the least significant factors. As a result, laboratory resources can be focused on studying those subunits or interactions with strong significance.

In nature, multiple virus lineages can coexist in the same host population. It is highly possible that different strains coinfect and reassort in a host, for example, swine.[24] Given the fact the vRNP of the pandemic H1N1/2009 virus is composed of subunits derived from avian, swine, and human viruses, developing analytic tools of this kind would be an important component of pandemic preparedness and risk assessment. As the influenza gene pool in animals is huge, it is not possible to study each of these viruses via this approach. It is more practical to apply our statistical approach to study recombinant vRNPs formed from various prevailing strains in the environment, thus providing a possible way to identify potential reassorted vRNP with enhanced polymerase activity. For example, our strategy can be used to characterize recombinant vRNPs with subunits derived from viruses that are endemic in human, swine, or poultry populations. Besides, our approach can be used to assess the risk of viruses that are known to cause zoonotic transmissions (e.g. swine to human and avian to swine). In parallel with animal influenza virus surveillance, the findings could give hints of emergence of any reassorted virus with high polymerase activity, which has a potential to infect and become established in humans. Nonetheless, the potential uses and limitations of this approach are needed to be fully revealed by further studies.

Acknowledgements

This project is supported by National Institutes of Health (NIAID contract HSN266200700005C), Research Grant Council of Hong Kong (HKU 776611M), Area of Excellence Scheme of the University Grants Committee (Grant AoE/M-12/06). We would like to thank Raul Rabadan for many helpful discussions on the topic. B.D.G. is the Eric and Wendy Schmidt, Member in Biology at the Institute for Advanced Study, and would like to thank them for their support.

References

1 Palese P, Shaw ML. Orthomyxoviridae: the viruses and their replication; in Knipe DM, Howley PM (eds): Fields Virology, 5th edn. Philadelphia, PA: Lippincott Williams & Wilkins, 2007; 1647–1689.
2 Palese P. The genes of influenza virus. Cell 1977; 10:1–10.
3 Biswas SK, Boutz PL, Nayak DP. Influenza virus nucleoprotein interacts with influenza virus polymerase proteins. J Virol 1998; 72:5493–5501.
4 Ma W, Kahn RE, Richt JA. The pig as a mixing vessel for influenza viruses: human and veterinary implications. J Mol Genet Med 2008; 3:158–166.
5 Nelson MI, Viboud C, Simonsen L *et al.* Multiple reassortment events in the evolutionary history of H1N1 influenza A virus since 1918. PLoS Pathog 2008; 4:e1000012.
6 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; 328:101–119.
7 Kawaoka Y, Krauss S, Webster RG. Avian-to-human transmission of the PB1 gene of influenza A viruses in the 1957 and 1968 pandemics. J Virol 1989; 63:4603–4608.
8 Taubenberger JK, Reid AH, Lourens RM, Wang R, Jin G, Fanning TG. Characterization of the 1918 influenza virus polymerase genes. Nature 2005; 437:889–893.
9 Antonovics J, Hood ME, Baker CH. Molecular virology: was the 1918 flu avian in origin? Nature 2006;440:E9; discussion E-10.
10 Gibbs MJ, Gibbs AJ. Molecular virology: was the 1918 pandemic caused by a bird flu? Nature 2006;440:E8; discussion E9–10.
11 Vana G, Westover KM. Origin of the 1918 Spanish influenza virus: a comparative genomic analysis. Mol Phylogenet Evol 2008; 47:1100–1110.
12 Smith GJ, Vijaykrishna D, Bahl J *et al.* Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature 2009; 459:1122–1125.
13 Kingsford C, Nagarajan N, Salzberg SL. 2009 Swine-origin influenza A (H1N1) resembles previous influenza isolates. PLoS ONE 2009; 4: e6402.
14 Dawood FS, Jain S, Finelli L *et al.* Emergence of a novel swine-origin influenza A (H1N1) virus in humans. N Engl J Med 2009; 360:2605–2615.
15 Li OT, Chan MC, Leung CS *et al.* Full factorial analysis of mammalian and avian influenza polymerase subunits suggests a role of an efficient polymerase for virus adaptation. PLoS ONE 2009; 4:e5658.
16 Chen J, Sun DX, Wu CFJ. A catalogue of two-level and three-level fractional factorial designs with small runs. Int Stat Rev 1993; 61:131–145.
17 Xu H. A catalogue of three-level regular fractional factorial designs. Metrika 2005; 62:259–281.
18 Bradel-Tretheway BG, Kelley Z, Chakraborty-Sett S, Takimoto T, Kim B, Dewhurst S. The human H5N1 influenza A virus polymerase complex is active *in vitro* over a broad range of temperatures, in contrast to the WSN complex, and this property can be attributed to the PB2 subunit. J Gen Virol 2008; 89:2923–2932.
19 Lang V, Marjuki H, Krauss SL, Webby RJ, Webster RG. Different incubation temperatures affect viral polymerase activity and yields of low-pathogenic avian influenza viruses in embryonated chicken eggs. Arch Virol 2011; 156:987–994.
20 Uiprasertkul M, Puthavathana P, Sangsiriwut K *et al.* Influenza A H5N1 replication sites in humans. Emerg Infect Dis 2005; 11:1036–1041.
21 Xu C, Hu WB, Xu K *et al.* Amino acids 473V and 598P of PB1 from an avian-origin influenza A virus contribute to polymerase activity, especially in mammalian cells. J Gen Virol 2012; 93:531–540.
22 Zhu W, Zhu Y, Qin K *et al.* Mutations in polymerase genes enhanced the virulence of 2009 pandemic H1N1 influenza virus in mice. PLoS ONE 2012; 7:e33383.
23 Mehle A, Dugan VG, Taubenberger JK, Doudna JA. Reassortment and mutation of the avian influenza virus polymerase PA subunit overcome species barriers. J Virol 2012; 86:1750–1757.
24 Vijaykrishna D, Smith GJ, Pybus OG *et al.* Long-term evolution and transmission dynamics of swine influenza A virus. Nature 2011; 473:519–522.