



## Evidence of intra-segmental homologous recombination in influenza A virus

Weilong Hao\*

Department of Laboratory Medicine and Pathobiology, University of Toronto, ON, Canada  
Mount Sinai Hospital, Toronto, ON, Canada

### ARTICLE INFO

#### Article history:

Accepted 25 April 2011

Available online xxx

Received by A.J. van Wijnen

#### Keywords:

Recombination

Reassortment

Influenza

Genome evolution

Circulating evolution

### ABSTRACT

The evolution of influenza viruses is remarkably dynamic. Influenza viruses evolve rapidly in sequence and undergo frequent reassortment of different gene segments. Homologous recombination, although commonly seen as an important component of dynamic genome evolution in many other organisms, is believed to be rare in influenza. In this study, 256 gene segments from 32 influenza A genomes were examined for homologous recombination, three recombinant H1N1 strains were detected and they most likely resulted from one recombination event between two closely related parental sequences. These findings suggest that homologous recombination in influenza viruses tends to take place between strains sharing high sequence similarity. The three recombinant strains were isolated at different time periods and they form a clade, indicating that recombinant strains could circulate. In addition, the simulation results showed that many recombinant sequences might not be detectable by currently existing recombinant detection programs when the parental sequences are of high sequence similarity. Finally, possible ways were discussed to improve the accuracy of the detection for recombinant sequences in influenza.

© 2011 Elsevier B.V. All rights reserved.

### 1. Introduction

Influenza A is a rapidly evolving single-stranded negative-sense RNA virus (Webster et al., 1992; Rambaut et al., 2008). The propensity of the rapid evolution of the influenza A genome features high mutation rates due to the notoriously error-prone RNA polymerase (Drake, 1993) and frequent reassortment of different RNA segments in multiple host types (Holmes et al., 2005; Schweiger et al., 2006; Nelson et al., 2008; Smith et al., 2009; Vijaykrishna et al., 2010). However, homologous recombination, a rather important evolutionary contributor to sequence diversity, has rarely been reported in influenza A virus (Boni et al., 2008), while homologous recombination has been well recognized across a broad spectrum of life forms such as, bacteria (Nesbo et al., 2006), archaea (Papke et al., 2004), yeasts (Ruderfer et al., 2006), plant mitochondria (Hao and Palmer, 2009), human mitochondria (White and Gemmell, 2009), and various forms of viruses [e.g. foot-and-mouth disease virus (FMDV) (Heath et al., 2006; Lewis-Rogers et al., 2008) and severe acute respiratory syndrome (SARS) virus (Stavriniades and Guttman, 2004)]. The relatively low observed frequency of recombination in both influenza and negative-strand RNA viruses in general could be explained by the

fact that the genomic RNA generated during replication is rapidly packaged with ribonucleoprotein, which acts to prevent the occurrence of template-switching that is essential to RNA recombination (Chare et al., 2003). To date, there are only three peer-reviewed articles that have proposed the occurrence of intra-segmental homologous recombination in influenza A virus (Gibbs et al., 2001; He et al., 2008, 2009) and each study has been criticized and challenged (Worobey et al., 2002; Boni et al., 2010).

In the first published recombination study, Gibbs et al. (2001) suggested that the hemagglutinin (HA) segment of the deadly 1918 H1N1 “Spanish flu” virus was a recombinant between a swine virus and a human virus, but this claim was challenged by the lack of phylogenetic support (Worobey et al., 2002). Recently, He and coworkers (He et al., 2008, 2009) reported several cases of homologous recombination in influenza, their proposed recombination cases, however, suffer from the criticisms of 1) no phylogenetic support, 2) requiring parental sequences from substantially different geographic locations and time periods, 3) lone recombinant sequences and no circulating clades of recombinant viruses, and 4) unknown quality in sample-handling and sequencing (Boni et al., 2010). On the other hand, the search for evidence of intra-segmental homologous recombination in influenza is still legitimate, since evidence of homologous recombination has been found in other segmented negative-strand RNA viruses (Sibold et al., 1999; Charrel et al., 2001). To convincingly demonstrate the occurrence of intra-segmental homologous recombination in influenza, sequences need to be placed under rigorous quality control [e.g., following the standard of the Influenza Genome Sequencing Project (IGSP) (Boni et al., 2010)].

*Abbreviations:* AU test, approximately unbiased test; FMDV, foot-and-mouth disease virus; HA, hemagglutinin; IGSP, Influenza Genome Sequencing Project; NIAID, National Institute of Allergy and Infectious Diseases; PB2, polymerase basic 2 protein; SARS, severe acute respiratory syndrome virus.

\* Tel.: +1 416 649 4475.

E-mail address: [haoweilong@gmail.com](mailto:haoweilong@gmail.com).

Moreover, identification of a circulating clade of recombinant viruses instead of individual lone recombinant sequences would provide further compelling evidence that intra-segmental homologous recombination occurs among influenza viruses.

In this study, a thorough examination for intra-segmental homologous recombination was conducted in 256 gene segments from 32 swine influenza A genomes generated by Vijaykrishna et al. (2010). Vijaykrishna et al. (2010) have already shown multiple reassortment events in these 32 genomes via a sophisticated phylogenetic analysis on each gene segment. The goal of this study is to address whether there is evidence of intra-segmental homologous recombination in these strains that have undergone multiple reassortment events, and when homologous recombination did occur, whether we can effectively detect the recombinant sequences.

## 2. Methods

Nucleotide sequences of 256 gene segments (GenBank accessions CY061626–CY061881) from 32 swine influenza A genomes in (Vijaykrishna et al., 2010) were downloaded from the GenBank website (<http://www.ncbi.nlm.nih.gov/genbank/>). The 32 swine viruses were collected from June 2009 to February 2010 in Hong Kong as part of the IGSP project funded by the National Institute of Allergy and Infectious Diseases (NIAID) (Vijaykrishna et al., 2010). Sequences were aligned individually for each gene segment using the MUSCLE program (Edgar, 2004). Recombination analysis was performed on the aligned sequences using RDP (Martin and Rybicki, 2000), SEQ (Boni et al., 2007), OnePop (Hao, 2010) and Max  $\chi^2$  (Smith, 1992). Detected recombinant sequences were further assessed by classical statistical tests using R (R Development Core Team, 2009).

Phylogenetic trees were initially constructed using a maximum likelihood method via the RAxML program (Stamatakis, 2006) under a GTR+ $\Gamma$ +I substitution model. To assess the robustness of the obtained phylogenetic trees, neighbor-joining trees were constructed using neighbor of the PHYLIP package (Felsenstein, 1989) version 3.6. Furthermore, two other maximum likelihood programs PhyML (Guindon and Gascuel, 2003) and PUZZLE (Strimmer and von Haeseler, 1996) were applied on the same dataset using a variety of substitution models (including GTR+ $\Gamma$ +I). Phylogenetic incongruence between sequence regions defined by the break point identified by the recombination detection programs was examined by the approximately unbiased (AU) test (Shimodaira, 2002). In brief, the site-by-site likelihoods for the trees were calculated with PUZZLE (Strimmer and von Haeseler, 1996), the AU-test was then implemented using CONSEL (Shimodaira and Hasegawa, 2001) to assign the tree probabilities.

Simulation studies were performed using the Seq-Gen program (Rambaut and Grassly, 1997) to generate phylogenetically related 2000-nucleotide sequences using the empirical base composition and transition/transversion ratio measured from the polymerase basic 2 protein (PB2) gene alignment using PUZZLE (Strimmer and von Haeseler, 1996). The empirical base composition is 34.0% (A), 18.9% (C), 25.4% (G), 21.7% (T), and the transition/transversion ratio is 5.29. Recombinant sequences were artificially constructed between simulated nucleotide sequences. For simplicity, the recombination break point was assumed to be the middle of the sequences for all simulated recombinant sequences. Each constructed recombinant sequence and its parental sequences were then tested for recombination using the RDP program (Martin and Rybicki, 2000) as a representative of the recombination detection programs.

## 3. Results and discussion

### 3.1. Evidence of intra-segmental homologous recombination

In this study, 256 gene segments from 32 influenza A genomes published in Vijaykrishna et al. (2010) were carefully examined.

Vijaykrishna et al. (2010) have already shown multiple reassortment events in these 32 genomes via a sophisticated phylogenetic analysis on each gene segment, in this study, the main focus was on the detection of homologous recombination within each gene segment. Table 1 shows that there are significant signals for homologous recombination within the polymerase basic 2 protein (PB2) gene supported by RDP, SEQ, Max  $\chi^2$  and OnePop. RDP, Max  $\chi^2$ , and OnePop identified the breakpoint at nucleotide site 1197 in the alignment, while 3SEQ suggested a breakpoint ranging from sites 1197–1212. For convenience, the breakpoint was assumed to be at site 1197 in this study, since no nucleotide differences were observed from sites 1198–1212 between recombinant and parental sequences (Fig. S1), and different breakpoint locations in this range will not alter the results of statistical tests.

Fig. 1 shows that the 32 PB2 gene sequences form two very distinct clusters and the intra-segmental homologous recombination event involves three lineages containing 10 highly closely related strains. As shown in Fig. 2, the PB2 sequences in Sw/HK/NS1809/2009 and Sw/HK/NS1810/2009 are likely to be recombinants between the clade that Sw/HK/NS1583/2009 belongs to (or the P1 clade) and the clade that Sw/HK/2886/2009 belongs to (the P2 clade). The regions 1–1197 and 1198–2232 show significantly different tree topologies. The clade of Sw/HK/189/2010, Sw/HK/NS1809/2009 and Sw/HK/NS1810/2009 is most closely related to the P2 clade in region 1–1197 supported by node A, while it is most closely related to the P1 clade in region 1198–2232 supported by node B (Fig. 3). The two key nodes (A and B) are well supported by RAxML and neighbor as shown in Fig. 3. The two nodes are also well supported by PhyML and PUZZLE, regardless of nucleotide substitution models, rate heterogeneity or the proportion of invariable sites (Table 2). The bootstrap values for node A range from 73 to 88, while the bootstrap values for node B range from 85 to 98. The quartet puzzle support values obtained from PUZZLE are even higher and they are all close to 100 for both nodes (Table 2). These results suggest different evolutionary relationships between region 1–1197 and region 1198–2232.

The difference between the two tree topologies is also supported by the AU test for phylogenetic incongruence. The *P*-values in the AU test are 0.029 and 0.031, respectively, for the two regions (Table 1). In both regions, Sw/HK/189/2010 is consistently the nearest neighbor to Sw/HK/NS1809/2009 and Sw/HK/NS1810/2009. Furthermore, there are 19 distinct nucleotides between the P1 and P2 clades (Fig. 2); Sw/HK/189/2010, Sw/HK/NS1809/2009 and Sw/HK/NS1810/2009 share the exactly same set of nucleotides in all these 19 sites. That is, of these 19 nucleotide sites, 9 sites are identical with the P2 clade and 10 sites are identical with the P1 clade (Fig. 2). For these reasons and more reasons elaborated below, Sw/HK/189/2010 was considered to belong to the recombinant clade together with Sw/HK/NS1809/2009 and Sw/HK/NS1810/2009. It is important to note that the findings of homologous recombination are not artifacts

**Table 1**

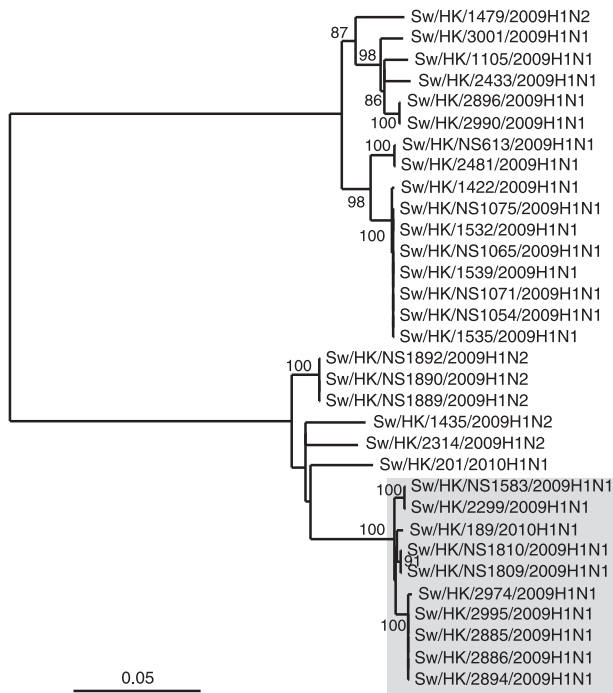
*P*-values obtained from different recombination detection programs on the PB2 gene<sup>a</sup>.

Recombinant sequence <sup>b</sup>	OnePop	3SEQ	RDP	Max $\chi^2$	AU-test <sup>c</sup>	
Sw/HK/NS1810/2009	$3.66 \times 10^{-6}$	0.042	0.042	0.048	0.029	0.031
Sw/HK/189/2010	–	0.156	–	0.082		

<sup>a</sup> *P*-values were computed based on the recombinant sequence and two parental type sequences Sw/HK/NS1583/2009 and Sw/HK/2886/2009 using OnePop, 3SEQ, RDP and Max  $\chi^2$ . *P*-values in the AU-test were computed based on the sequences and break point shown in Fig. 3.

<sup>b</sup> Sw/HK/NS1809/2009 has an otherwise identical sequence with Sw/HK/NS1810/2009 in the alignment except for a slight difference in available length (Fig. S1) and produces the same *P*-values as Sw/HK/NS1810/2009. Sw/HK/189/2010 is shown, because its PB2 gene has likely undergone the same recombination event as Sw/HK/NS1809/2009 and Sw/HK/NS1810/2009 (see Fig. 2 and main text for details), even though the *P*-values are not significant.

<sup>c</sup> The first *P*-value was from the test of the sequence alignment of region 1 against the topology of region 2, while the second *P*-value was from the test of the sequence alignment of region 2 against the topology of region 1.



**Fig. 1.** Maximum likelihood tree of the PB2 nucleotide sequences from 32 influenza genomes in Vijaykrishna et al. (2010). The tree is mid-point rooted for purposes of clarity. Ten strains that were detected to be involved in homologous recombination in this study are shaded. One hundred bootstrap replicates were generated and bootstrap values, when >75, were shown on the phylogeny.

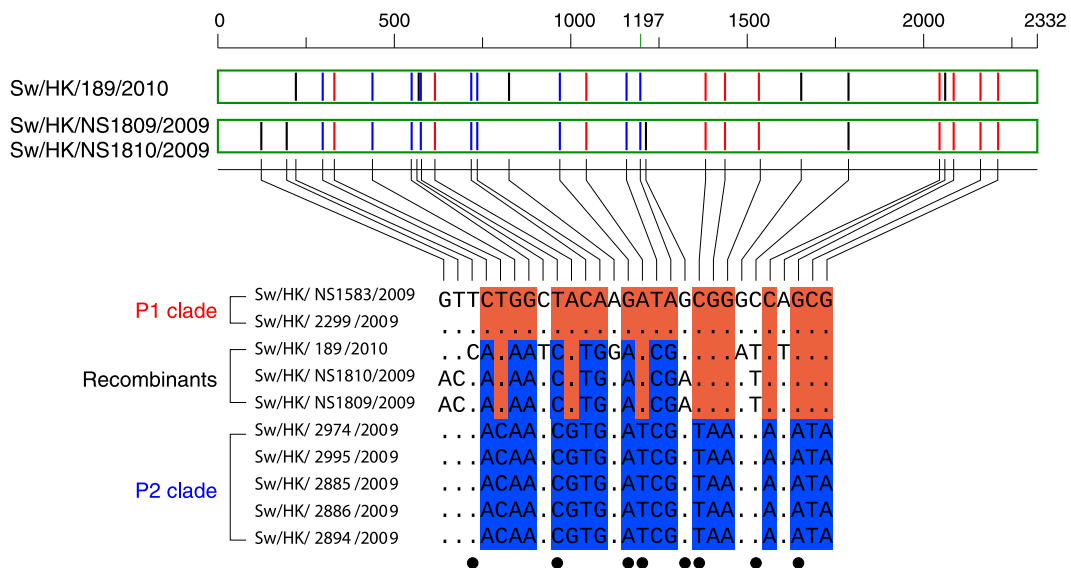
of sequencing/handling errors. In fact, Prof. Yi Guan and colleagues at the University of Hong Kong have kindly confirmed that the PB2 sequences in Sw/HK/189/2010, Sw/HK/NS1809/2009 and Sw/HK/NS1810/2009 were sequenced, as per IGSP standard, multiple times on different plates and dates, the trace signal for each sequence

was very clean with no ambiguity, and the assembly was done properly (Zhu and Guan personal communication).

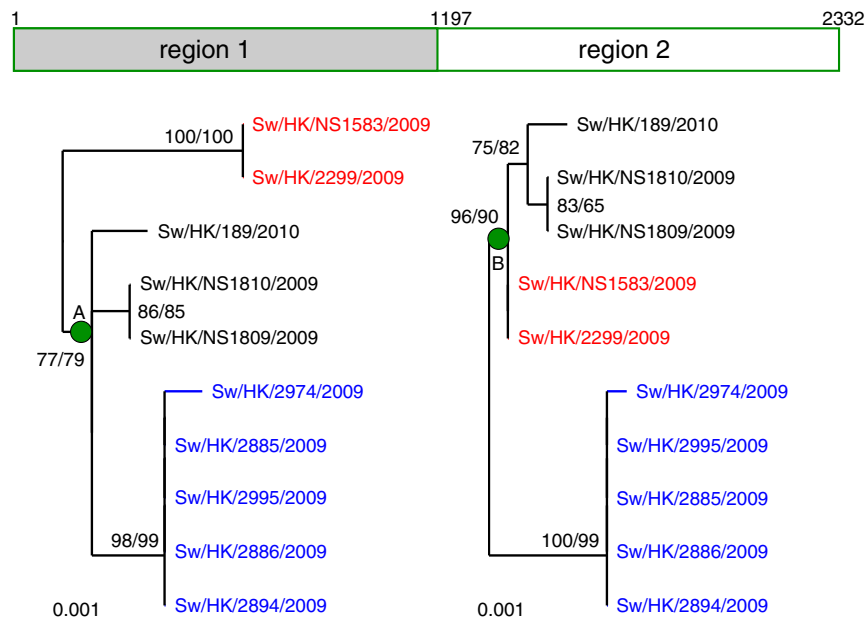
The occurrence of recombination in the PB2 gene could potentially have functional consequences. The PB2 gene encodes a subunit of the trimeric viral RNA polymerase complex containing PA, PB1, and PB2 (Engelhardt and Fodor, 2006). The PB2 protein has been shown to interact with the antiviral signaling protein (Graef et al., 2010), affect host range (Subbarao et al., 1993; Labadie et al., 2007) and virulence of influenza viruses (Hatta et al., 2001; Shinya et al., 2004). The recombinant sequences contain eight nonsynonymous nucleotide changes, and five of them are identical with only one of the two parental clades but not the other (Fig. 2). Such unique amino acid combination generated by recombination could possibly result in different pathogenicity of the recombinant viruses than that of each parental clade.

### 3.2. Parallel convergent evolution is unlikely to explain the mosaic sequence pattern

Homologous recombination in Sw/HK/NS1809/2009 and Sw/HK/NS1810/2009 was widely supported by a variety of programs, but it is also notable that because of the little difference among sequences, the *P*-values all fall between 0.01 and 0.05 with one exception of  $3.66 \times 10^{-6}$  from OnePop (Table 1). The increased recombination detection power among closely related sequences in OnePop is due to the removal of the examined region from the calculation of sequence divergence [see (Hao, 2010) for more details], even though OnePop and RDP are based on essentially the same methodology. Since the sequences involved in recombination share high sequence similarity and only a small number of nucleotide changes are involved, it becomes crucial to address whether the observed nucleotide distribution pattern in Fig. 2 could be explained by parallel convergent evolution in different lineages, as opposed to by recombination. Evidence of parallel evolution has been previously documented in viral genome evolution and it is almost always associated with adaptive selection (Crandall et al., 1999; Keleta et al., 2008). However, adaptive evolution is not very likely the explanation



**Fig. 2.** Chimeric PB2 genes in H1N1. Positions with nucleotide changes are shown (for a complete sequence alignment, see Fig. S1). Three recombinant strains are suggested to be the result of homologous recombination between two parental clades (or P1 and P2, in red and blue, respectively). In the nucleotide alignment, dots indicate identities relative to the strain Sw/HK/NS1583/2009, while letters show nucleotide differences. Nucleotide differences between the P1 and P2 clades are colored in blue in the P2 clade. Nucleotides in the recombinant strains that are identical to P2 are colored in blue, while those identical to P1 are colored in red. Eight nucleotide changes that have resulted in amino acid changes are labeled with filled black circles. The positions of these nucleotides are also shown in scale as two sequence types for the three recombinant strains at the top part of the figure, with red vertical lines representing the red nucleotides in the nucleotide alignment, blue vertical lines representing the blue nucleotides, black vertical lines representing the lineage-specific changes. Nucleotide position 1197 is the break point used for the phylogenetic incongruence test in Fig. 3 and some other statistical tests. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 3.** Phylogenetic support for homologous recombination in PB2. The phylogenetic trees (with associated bootstrap values) supporting the contrasting phylogenetic positions between regions 1–1197 and 1198–2332 in the three recombinant strains are shown. The shown phylogenetic trees were obtained using RAxML, while bootstrap values from both RAxML (first) and neighbor (second) are shown for a given branch. The trees are rooted using Sw/HK/201/2010H1N1 (see Fig. 1) as the outgroup. The two parental clades, P1 and P2, are colored in red and blue, respectively as per Fig. 2. The two key nodes that indicate different evolutionary origins between the two regions are labeled as A and B respectively. The two phylogenetic trees are significantly different in the AU-test (*P*-values shown in Table 1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 2**

Bootstrap values for the two key nodes (A and B as shown in Fig. 3) that support different evolutionary relationships between the two regions. Bootstrap values were obtained based on a variety of substitution models.

Phylogenetic programs	Substitution models	Bootstrap values for	
		A node	B node
RAxML	GTR+ $\Gamma$ +I	77	96
PhyML	JC79	83	92
PhyML	JC79+ $\Gamma$	84	97
PhyML	JC79+ $\Gamma$ +I	88	95
PhyML	K2P	84	90
PhyML	K2P+ $\Gamma$	87	91
PhyML	K2P+ $\Gamma$ +I	84	96
PhyML	F81	82	98
PhyML	F81+ $\Gamma$	79	95
PhyML	F81+ $\Gamma$ +I	87	88
PhyML	F84	87	87
PhyML	F84+ $\Gamma$	87	88
PhyML	F84+ $\Gamma$ +I	83	85
PhyML	HKY	82	95
PhyML	HKY+ $\Gamma$	74	94
PhyML	HKY+ $\Gamma$ +I	79	92
PhyML	TN93	80	92
PhyML	TN93+ $\Gamma$	73	92
PhyML	TN93+ $\Gamma$ +I	78	90
PhyML	GTR	84	93
PhyML	GTR+ $\Gamma$	82	91
PhyML	GTR+ $\Gamma$ +I	87	88
Puzzle <sup>a</sup>	HKY	99 <sup>b</sup>	100
Puzzle	HKY+ $\Gamma$	100	100
Puzzle	HKY+ $\Gamma$ +I	99	100
Puzzle	TN93	99	100
Puzzle	TN93+ $\Gamma$	99	100
Puzzle	TN93+ $\Gamma$ +I	99	100

<sup>a</sup> The GTR model was not included in the PUZZLE analysis, since PUZZLE does not directly estimate the GTR relative rate parameters. PUZZLE does not provide options for models JC79, K2P, F81 and F84.

<sup>b</sup> Values are quartet puzzling support values.

for the mosaic sequence pattern in this study. First, of the 19 distinct nucleotides between the P1 and P2 clades, 14 sites are synonymous changes (Figs. 2, S1 and S2). Second, the two parental clades are highly similar (with <1% of nucleotide divergence), and have diverged very recently. In fact, if we assume that the PB2 gene has a similar rate of mutation as the nonstructural protein NS gene whose mutation rate ranges from  $1.8 \times 10^{-3}$ – $2.2 \times 10^{-3}$  nucleotide substitutions/site/year (Ludwig et al., 1991; Kawaoka et al., 1998), the P1 and P2 clades have just diverged for 3–5 years. Third, despite a growing body of literature focused on identifying mutations associated with adaptive evolution [e.g., amino acid sites 114, 491, 560, 684 (Furuse et al., 2010), 158 (Zhou et al., 2011), 271 (Bussey et al., 2010), 504 (Rolling et al., 2009), 627, and 701 (Ozawa et al., 2011), and synonymous changes in amino acid sites 741 and 742 ((Liang et al., 2008), corresponding to nucleotide sites from 2245–2250 in Fig. S1)], none of the reported adaptive amino acid or nucleotide changes were observed in this study (Fig. S2). If the observed nucleotide pattern were solely due to independently derived substitutions, most of the substitutions would have to be selectively neutral or nearly neutral. Since the shared nucleotide changes in each of the three clades (P1, P2 and recombinant) are clear and of nearly no ambiguity, the substitutions would have been derived either in the detected recombinant lineage or in the two parental lineages. We can test each substitution scenario by calculating its *P*-value using classical statistical tests.

The first scenario is that a number of substitutions had been accumulated in the recombinant clade and by chance some derived nucleotides in the recombinant clade are identical to some distinct clade. In Fig. 2, the recombinant clade shares 9 distinct nucleotides with the P2 clade and 10 with the P1 clade. Since the recombinant clade is slightly more similar to the P1 clade, we could assume that the detected recombinants were derived from the P1 clade by random substitution. Then we can calculate the *P*-value for the shared nucleotides between the recombinant and P2 clades being derived by random substitution. The probability that two independent substitutions occur at the same site could be calculated from 1 over the “effective site number”, which is ideally the number of nucleotide

sites free to change. In the homoplasy test (Maynard Smith and Smith, 1998), the number of synonymous changes from an outgroup species was assumed to be the effective site number. In this study, there are 271 synonymous changes and 102 non-synonymous changes between Sw/HK/NS1054/2009 and Sw/HK/2886/2009, and 273 synonymous changes and 100 non-synonymous changes between Sw/HK/1105/2009 and Sw/HK/2886/2009. To be conservative, I assumed that the effective site number is 271 in the PB2 gene. Thus, when a substitution occurred in a PB2 sequence derived from the P1 clade, the probability that this substitution is one of the 19 distinct nucleotides identical with the P2 clade is equal to or less than  $19/271 = 0.07$  (since, for each site, there are 3 possible nucleotide substitution changes instead of just one). For instance, the Sw/HK/189/2010 strain has 15 nucleotides different from the P1 clade with 6 specific to Sw/HK/189/2010 or the recombinant clade and 9 shared with the P2 clade. Then, a *P*-value can be calculated for 9 shared nucleotides out of 15 nucleotide changes with the probability of each substitution identical with the P2 clade being 0.07 under a binomial distribution with  $P = 1.37 \times 10^{-7}$ . This suggests that the PB2 gene in Sw/HK/189/2010 is likely a recombinant even though it could not be detected by any programs used in this study (Table 1). It is worth to note that the effective site number 271 should be considered conservative. In fact, the number of synonymous sites in Sw/HK/2886/2009 was estimated as 631 using the method developed by Yang and Nielsen (2000) and some nonsynonymous sites are believed to be selectively neutral or nearly neutral. Together, this suggests a much greater effective site number than 271, and the calculated *P*-value will become much smaller and more significant with a larger effective site number. Therefore, the observed mosaic pattern is unlikely due to accumulated substitutions in the detected recombinant clade.

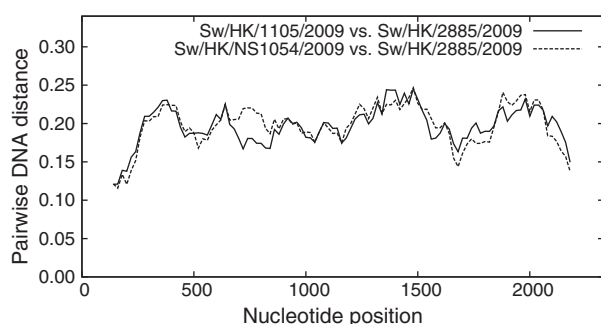
Second, substitution changes would have been accumulated in the P1 and P2 lineages separately with the recombinant lineage as the ancestral sequence type. If this were true, one should expect that substitutions arise in a somewhat independent, random manner throughout the gene in both the P1 and P2 lineages. According to the break point shown in Fig. 3, the P1 clade would have 9 nucleotide changes in the first half of the gene and 1 nucleotide change (because of the T at position 1787 in all three recombinants) in the second half of the gene from the ancestral sequence, and the P2 clade would have 3 nucleotide changes in the first half and 8 nucleotide changes in the second half. The Fisher's Exact Test was used to test whether the P1 and P2 lineages have significantly different substitution patterns in these two regions. The *P*-value of the Fisher's Exact Test is 0.0075. If we assume that the ancestral

nucleotide at position 1787 is C, as in both the P1 and P2 clades, instead of T, the *P*-value would be 0.0031. This suggests that, to achieve the observed nucleotide pattern from an ancestral sequence similar to the recombinant strains, the P1 and P2 lineages would have to have substantially different substitution pattern in each of the two regions. Fig. 4 shows the sequence divergence between two pairs of distantly related strains Sw/HK/NS1054/2009 vs. Sw/HK/2886/2009, and Sw/HK/1105/2009 vs. Sw/HK/2886/2009. It shows that there is a nearly constant rate of substitution across the PB2 gene in each of the two gene pairs and the two regions do not show significantly different evolutionary rates in either gene pair (statistical data not shown). It is therefore unlikely that the observed nucleotide pattern in Fig. 2 is due to derived substitution changes in the P1 and P2 lineages.

In the above analyses, significant signals for recombination were obtained based on the assumption of a single breakpoint at nucleotide site 1197, which could potentially be over-simplified. As shown in Fig. 2, there are 3 P1-like nucleotide sites embedded in P2-like region 1 of the putative recombinant sequences. Applying the same approach using effective site number described above, we can assess the probability that these 3 P1-like nucleotides arose via random mutation. In region 1, we assumed the effective site number is 139 based on the synonymous changes between Sw/HK/NS1054/2009 and Sw/HK/2886/2009. By assuming region 1 in the detected recombinants derived from the P2 clade, the probability to have three nucleotides identical to the P1 clade is 0.006 for Sw/HK/NS1809/2009 and Sw/HK/NS1810/2009, or 0.011 for Sw/HK/189/2010. It is, therefore, possible that these 3 P1-like nucleotides resulted from intricate fine-scale recombination. A similar phenomenon was found by Baldo et al. (2005) that there have been extensive recombination and shuffling of a relatively conserved set of amino acid motifs within each of the four hypervariable regions in the Wolbachia surface protein. Chan and colleagues further showed that the physical position for recombination to occur is not restricted by any functional units (e.g., motif, domain or gene [Chan et al., 2009a,b]). Recently, Hao et al. (2010) have demonstrated that intricate recombination events could lead to reduced or even not significant recombination signals. In fact, analyses excluding the 3 P1-like nucleotides revealed that these 3 nucleotide sites are responsible for the relatively long branch leading to the P2 clade in region 1–1197 (Fig. 3) and slightly lower bootstrap values for node A than those for node B (Table 2 and Fig. 3). It is important to note that significant recombination signals were detected in the PB2 gene without considering the 3 P1-like nucleotides as the result of fine-scale recombination (Table 1). If these 3 nucleotide sites were removed from the analyses, much more significant *P*-values for recombination would be expected. Furthermore, the analyses conducted in this study lie in the assumption of independent and random mutation, which has been widely used in various evolutionary analyses to make inferences (Jukes and Cantor, 1969; Nei and Gojobori, 1986) but not necessarily realistic (Clegg et al., 1994; Miyamoto and Fitch, 1995). For instance, some substitutions in influenza could have compensatory roles with one another (Simon et al., 2011), substitutions could sometimes be dependent (e.g., the dinucleotide bias [Greenbaum et al., 2008]), and the substitution rate of sites in a gene can vary over time (known as heterotachy [Lopez et al., 2002]). All of these circumstances could confound the detection of recombination.

### 3.3. Recombination tends to occur between closely related sequences

In this study, three strains were found to be the result of homologous recombination between two parental sequences of very low sequence divergence. Given the fact that homologous recombination occurs rarely in influenza A virus (Boni et al., 2008), the findings suggest that homologous recombination in influenza viruses, when occurs, tends to occur between closely related strains. There are two



**Fig. 4.** Nucleotide conservation across the PB2 gene. The plots used sliding windows of 200 nucleotides, slid 20 nucleotides at a time. The y axis corresponds to the estimated DNA distance (number of substitutions/changes per site) measured using the F84 matrix. These plots are based on the comparisons between Sw/HK/1105/2009 vs. Sw/HK/2886/2009 and between Sw/HK/NS1054/2009 vs. Sw/HK/2885/2009 (for a complete sequence alignment, see Fig. S3) as they represent the most divergence among the PB2 sequences in the Vijaykrishna et al., 2010 study.

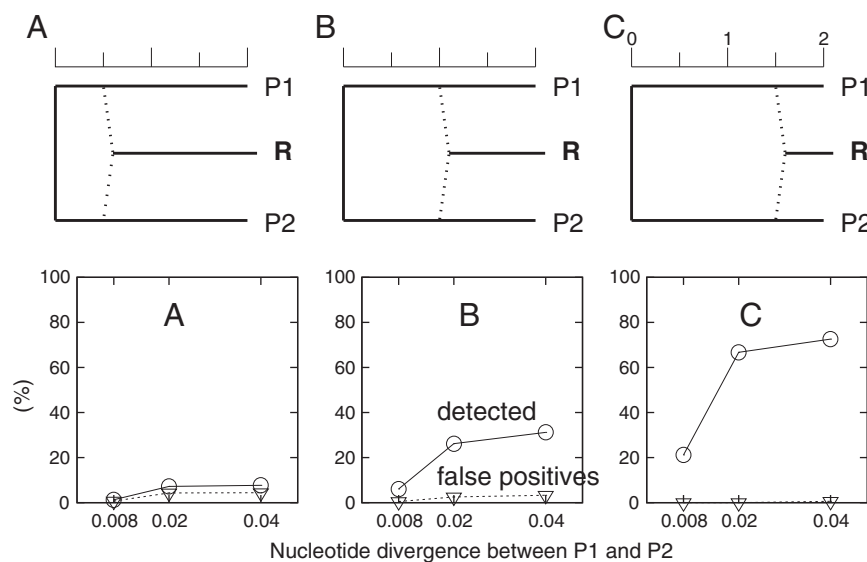
obvious reasons. First, the frequency of homologous recombination correlates positively and tightly with nucleotide sequence similarity and decreases sharply with the level of relatedness between the two parental sequences (Stratz et al., 1996; Majewski and Cohan, 1999). Second, recombination between highly conserved sequences will introduce relatively few changes. Such events are therefore more likely to be relatively neutral (or even beneficial) and thus be fixed. If the occurred homologous recombination events are predominantly between closely related strains, the relatively small number of substitution changes between the closely related sequences could potentially prevent intra-segmental homologous recombination from being detected. On the other hand, recombination events between distantly related strains, though occur at a much lower frequency, are the ones likely to be detected. In fact, both empirical and simulation studies have shown that all existing recombination detection programs are not sensitive at very low sequence diversity (Drouin et al., 1999; Posada and Crandall, 2001; Wiuf et al., 2001; Posada, 2002). Furthermore, the predictive accuracy of recombination detection programs could be further diminished by post-recombination substitutions (Chan et al., 2006). It is therefore not surprising that little evidence of homologous recombination has been previously discovered in influenza virus, and the only few reported recombination cases largely involve divergent parental lineages (He et al., 2008, 2009).

Given the different perspectives of sequence diversity between occurrence and detection of homologous recombination, it is interesting to examine and compare sequence diversity among the eight influenza gene segments. To do so, two measurements were obtained from the 10 closely related strains in Fig. 2. 1) Nucleotide diversity ( $\pi$ ), which is the average number of nucleotide differences per site between any two DNA sequences from the sample population, was calculated using DnaSP 4.0 (Rozas et al., 2003). 2) DNA distance of each sequence pair was also calculated using dnadist of the PHYLIP package (Felsenstein, 1989) version 3.6, and the maximum pairwise DNA distance from each gene segment was used as another indicator for sequence divergence. The results revealed that the PB2 gene has the highest  $\pi$  value among the 8 gene segments (Fig. S4), suggesting the highest diversity in the PB2 gene. In the presence of recombination, the high  $\pi$  value in the PB2 gene could possibly be, at least in part, due to recombination. Importantly, the PB2 gene also has the highest

maximum pairwise DNA distance among the 8 gene segments (Fig. S4). It is therefore likely that the high nucleotide diversity in the PB2 gene has facilitated the detection of recombination.

#### 3.4. Many recombinant sequences are undetectable

It has been well accepted that existing recombination detection programs perform poorly in sequences of low diversity (Drouin et al., 1999; Posada and Crandall, 2001; Wiuf et al., 2001; Posada, 2002). To assess how well homologous recombination between closely related strains could be detected at a level of sequence diversity comparable to the sequences in this study, I conducted a series of simulation studies as illustrated in Fig. 5. Three different sequence divergences were simulated between the P1 and P2 lineages, they are 0.008, 0.02 and 0.04 nucleotide divergence (substitutions per site), respectively, between the P1 and P2 lineages. As a reference, the sequence divergence of members between the two parental clades in Fig. 2 ranges from 0.0084 to 0.0105 (Fig. S4). Thereafter, recombination events were introduced at three different time points between the P1 and P2 lineages with the recombination break point at the middle of the sequence (Fig. 5). Then the constructed recombinant sequence (R) and the two parental sequences (P1 and P2) were examined by the RDP program (Martin and Rybicki, 2000). As suggested previously, the accuracy of recombination detection is low when the parental sequences have high sequence similarity and the accuracy increases when the parental sequences become more divergent. It is consistently shown that recombination events, when occurred recently, are more likely to be detected. The accuracy of recombination detection could range from nearly 0% for ancient recombination with 0.008 nucleotide divergence between P1 and P2 to about 75% for recent recombination with 0.04 nucleotide divergence between P1 and P2. It is noteworthy that Sw/HK/NS1809/2009 and Sw/HK/NS1810/2009 were detected with a parental divergence close to 0.008, at which level of parental divergence the accuracy of recombination detection is only about 20%. Since the accuracy of recombination detection is already very low, it is not surprising that Sw/HK/189/2010 with some additionally accumulated substitutions has become insignificant to the recombination detection programs (Table 1). The failure of recombination detection in highly similar sequences raises the



**Fig. 5.** Recombination detection power and rate of false positives using the RDP program on a series of simulations. In the simulations, two parental taxa have been diverging for 2 simulation units. A recombination event was introduced in each simulation, but could be introduced at three different evolutionary time (0.5, 1, and 1.5 simulation units respectively). In all simulations, the recombination break point was the middle of a 2000-nucleotide sequence. The results are based on the analysis of the recombinant sequence and the two parental sequences given three different levels of nucleotide divergence.

demand for developing more sophisticated and sensitive methodologies specially for closely related sequences.

### 3.5. The choice of taxa matters in phylogenetic analyses

Homologous recombination could also be detected by phylogenetic methods. A routine analysis of recombination detection in influenza is to test for phylogenetic incongruence (e.g., using the AU-test) between regions identified by some recombination detection programs. It is often seen that a large set of diverse sequences or even all available homologs were tested for phylogenetic incongruence. Since homologous recombination would likely occur between closely related strains, it is important to know whether the presence of distantly related sequences affects the performance of the incongruence test. Specifically, I sought to address whether the choice of taxa that are not involved in recombination affects the successful detection of homologous recombination. Three lineages (P1, P2, and R) were simulated in 100 iterations as per Fig. 5C and, additionally, two outgroups were introduced with three different scales (1×, 2×, and 5×) for the branch length between the two outgroups and the P1, R, P2 lineages. The AU-test was then performed in each condition using the middle of the sequence as the break point, and the number of iterations that failed in recombination detection was counted (Fig. 6). When the branch length between the outgroups and the P1, R, P2 lineages was 1×, 2 out of 100 iterations failed to show significant phylogenetic incongruence. When the branch length between the outgroups and the P1, R, P2 lineages was increased to 5×, 19 iterations failed to show significant phylogenetic incongruence. In other words, the results reveal that, with the exactly same set of sequences involved in recombination, using distantly related outgroup sequences could lead to a poorer performance of recombination detection using phylogenetic methods. Unfortunately, many published homologous recombination studies in influenza searched for incongruent phylogenetic signals from data sets of very diverse sequences. It is possible that some genuine recombinant sequences might have been overlooked by analyzing diverse sequences. Hence, the best way to detect homologous recombination using phylogenetic

methods is to analyze the putative recombinant, parental sequences, and their close relatives and to avoid distantly related sequences that are not involved in recombination. It is important to note that, in the simulation, the AU-test was performed on the actual simulated breakpoint, which is usually unknown. The accuracy of recombination detection of the AU-test and other similar phylogenetic methods is therefore dependent on the successful detection of recombinant sequences and the successful identification of recombination breakpoints.

## 4. Conclusion

Evidence of intra-segmental homologous recombination was found in three strains between very closely related parental sequences. Given the very low rate of homologous recombination, homologous recombination might be more likely to occur among closely related strains. Furthermore, the three recombinant strains were isolated at different time periods. The finding of a clade of recombinant influenza viruses suggests that recombinant strains could persist in a host population and have the potential to circulate. The results of simulation studies suggest that homologous recombination between divergent strains have a better chance to be detected than between very closely related strains. When homologous recombination occurs between sequences at the same level of divergence, it would be increasingly difficult to detect when the recombination events are relatively old. Finally, the use of distantly related taxa that are not involved in recombination could also result in poor performance in recombination detection. Therefore, when we search for evidence of homologous recombination among closely related strains using phylogenetic methods, it would be the best to avoid analyzing distantly related strains. Successful detection of intra-segmental homologous recombination events in influenza genomes demands much more careful sequence analysis and more sophisticated and sensitive recombination detection algorithms.

## Acknowledgments

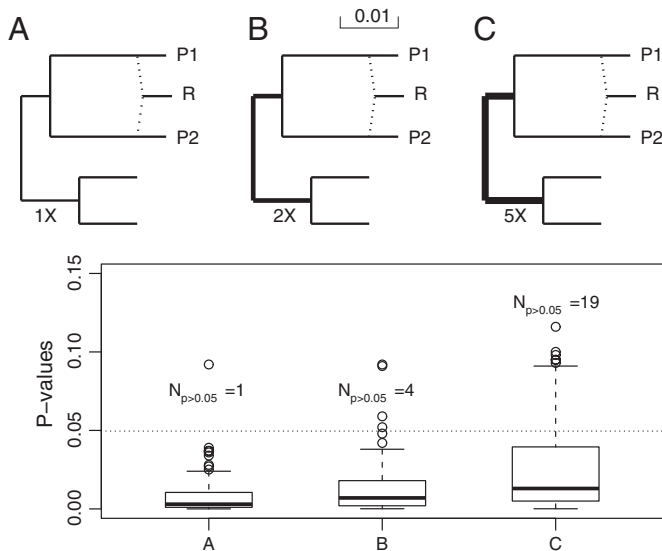
The author thanks the four anonymous reviewers for their thoughtful suggestions. The author would also like to thank Brian Golding, David Alexander, Wilfried Haerty and Yi Guan for comments on the manuscript; Jonathan Gubbay and Reza Eshaghi for discussion. The author is also grateful to Prof. Yi Guan, Dr. Huachen Zhu and colleagues at the University of Hong Kong for verification of the quality of the sequences they generated. The author was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) fellowship.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.gene.2011.04.012.

## References

- Baldo, L., Lo, N., Werren, J.H., 2005. Mosaic nature of the *Wolbachia* surface protein. *J. Bacteriol.* 187, 5406–5418.
- Boni, M.F., Posada, D., Feldman, M.W., 2007. An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* 176, 1035–1047.
- Boni, M.F., Zhou, Y., Taubenberger, J.K., Holmes, E.C., 2008. Homologous recombination is very rare or absent in human influenza A virus. *J. Virol.* 82, 4807–4811.
- Boni, M.F., de Jong, M.D., van Doorn, H.R., Holmes, E.C., 2010. Guidelines for identifying homologous recombination events in influenza A virus. *PLoS One* 5, e10434.
- Bussey, K.A., Bousse, T.L., Desmet, E.A., Kim, B., Takimoto, T., 2010. PB2 residue 271 plays a key role in enhanced polymerase activity of influenza A viruses in mammalian host cells. *J. Virol.* 84, 4395–4406.
- Chan, C.X., Beiko, R.G., Ragan, M.A., 2006. Detecting recombination in evolving nucleotide sequences. *BMC Bioinformatics* 7, 412.
- Chan, C.X., Beiko, R.G., Darling, A.E., Ragan, M.A., 2009a. Lateral transfer of genes and gene fragments in prokaryotes. *Genome Biol. Evol.* 2009, 429–438.



**Fig. 6.** The detecting power of phylogenetic methods on homologous recombination could be significantly affected by the high divergence of outgroup taxa. In each simulation, the parental taxa and the recombinant remain the same while different divergence was simulated from the two outgroup taxa. As per Fig. 5, the recombination break point was the middle of a 2000-nucleotide sequence. *P*-values were measured in the AU-test by testing the first half sequence against the topology of the second half of the sequence. 100 iterations were conducted, and the number of iterations that are not significant are shown.

- Chan, C.X., Darling, A.E., Beiko, R.G., Ragan, M.A., 2009b. Are protein domains modules of lateral genetic transfer? *PLoS One* 4, e4524.
- Chare, E.R., Gould, E.A., Holmes, E.C., 2003. Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. *J. Gen. Virol.* 84, 2691–2703.
- Charrel, R.N., de Lamballerie, X., Fulhorst, C.F., 2001. The Whitewater Arroyo virus: natural evidence for genetic recombination among Tacaribe serocomplex viruses (family Arenaviridae). *Virology* 283, 161–166.
- Clegg, M.T., Gaut, B.S., Learn Jr., G.H., Morton, B.R., 1994. Rates and patterns of chloroplast DNA evolution. *Proc. Natl. Acad. Sci. U. S. A.* 91, 6795–6801.
- Crandall, K.A., Kelsey, C.R., Imamichi, H., Lane, H.C., Salzman, N.P., 1999. Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection. *Mol. Biol. Evol.* 16, 372–382.
- Drake, J.W., 1993. Rates of spontaneous mutation among RNA viruses. *Proc. Natl. Acad. Sci. U. S. A.* 90, 4171–4175.
- Drouin, G., Prat, F., Ell, M., Clarke, G.D., 1999. Detecting and characterizing gene conversions between multigene family members. *Mol. Biol. Evol.* 16, 1369–1390.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
- Engelhardt, O.G., Fodor, E., 2006. Functional association between viral and cellular transcription during influenza virus infection. *Rev. Med. Virol.* 16, 329–345.
- Felsenstein, J., 1989. PHYLIP (phylogeny inference package). Version 3.2. *Cladistics* 5, 164–166.
- Furuse, Y., Suzuki, A., Oshitani, H., 2010. Reassortment between swine influenza A viruses increased their adaptation to humans in pandemic H1N1/09. *Infect. Genet. Evol.* 10, 569–574.
- Gibbs, M.J., Armstrong, J.S., Gibbs, A.J., 2001. Recombination in the hemagglutinin gene of the 1918 “Spanish flu”. *Science* 293, 1842–1845.
- Graef, K.M., et al., 2010. The PB2 subunit of the influenza virus RNA polymerase affects virulence by interacting with the mitochondrial antiviral signaling protein and inhibiting expression of beta interferon. *J. Virol.* 84, 8433–8445.
- Greenbaum, B.D., Levine, A.J., Bhanot, G., Rabadan, R., 2008. Patterns of evolution and host gene mimicry in influenza and other RNA viruses. *PLoS Pathog.* 4, e1000079.
- Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704.
- Hao, W., 2010. OrgConv: detection of gene conversion using consensus sequences and its application in plant mitochondrial and chloroplast homologs. *BMC Bioinformatics* 11, 114.
- Hao, W., Palmer, J.D., 2009. Fine-scale mergers of chloroplast and mitochondrial genes create functional, transcompartimentally chimeric mitochondrial genes. *Proc. Natl. Acad. Sci. U. S. A.* 106, 16728–16733.
- Hao, W., Richardson, A.O., Zheng, Y., Palmer, J.D., 2010. Gorgeous mosaic of mitochondrial genes created by horizontal transfer and gene conversion. *Proc. Natl. Acad. Sci. U. S. A.* 107, 21576–21581.
- Hatta, M., Gao, P., Halfmann, P., Kawaoka, Y., 2001. Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. *Science* 293, 1840–1842.
- He, C.Q., et al., 2008. Homologous recombination evidence in human and swine influenza A viruses. *Virology* 380, 12–20.
- He, C.Q., et al., 2009. Homologous recombination as an evolutionary force in the avian influenza A virus. *Mol. Biol. Evol.* 26, 177–187.
- Heath, L., van der Walt, E., Varsani, A., Martin, D.P., 2006. Recombination patterns in aphthoviruses mirror those found in other picornaviruses. *J. Virol.* 80, 11827–11832.
- Holmes, E.C., et al., 2005. Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. *PLoS Biol.* 3, e300.
- Jukes, T.H., Cantor, C.R., 1969. Evolution of protein molecules. In: Munro, H.N. (Ed.), *Mammalian Protein Metabolism*. Academic Press, New York, pp. 21–132.
- Kawaoka, Y., et al., 1998. Influence of host species on the evolution of the nonstructural (NS) gene of influenza A viruses. *Virus Res.* 55, 143–156.
- Keleta, L., Ibricevic, A., Bovin, N.V., Brody, S.L., Brown, E.G., 2008. Experimental evolution of human influenza virus H3 hemagglutinin in the mouse lung identifies adaptive regions in HA1 and HA2. *J. Virol.* 82, 11599–11608.
- Labadie, K., Dos Santos, Afonso E., Rameix-Welti, M.A., van der Werf, S., Naffakh, N., 2007. Host-range determinants on the PB2 protein of influenza A viruses control the interaction between the viral polymerase and nucleoprotein in human cells. *Virology* 362, 271–282.
- Lewis-Rogers, N., McClellan, D.A., Crandall, K.A., 2008. The evolution of foot-and-mouth disease virus: impacts of recombination and selection. *Infect. Genet. Evol.* 8, 786–798.
- Liang, Y., Huang, T., Ly, H., Parslow, T.G., Liang, Y., 2008. Mutational analyses of packaging signals in influenza virus PA, PB1, and PB2 genomic RNA segments. *J. Virol.* 82, 229–236.
- Lopez, P., Casane, D., Philippe, H., 2002. Heterotachy, an important process of protein evolution. *Mol. Biol. Evol.* 19, 1–7.
- Ludwig, S., Schultz, U., Mandler, J., Fitch, W.M., Scholtissek, C., 1991. Phylogenetic relationship of the nonstructural (NS) genes of influenza A viruses. *Virology* 183, 566–577.
- Majewski, J., Cohan, F.M., 1999. DNA sequence similarity requirements for interspecific recombination in *Bacillus*. *Genetics* 153, 1525–1533.
- Martin, D., Rybicki, E., 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16, 562–563.
- Maynard Smith, J., Smith, N.H., 1998. Detecting recombination from gene trees. *Mol. Biol. Evol.* 15, 590–599.
- Miyamoto, M.M., Fitch, W.M., 1995. Testing the covarion hypothesis of molecular evolution. *Mol. Biol. Evol.* 12, 503–513.
- 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., et al., 2008. Multiple reassortment events in the evolutionary history of H1N1 influenza A virus since 1918. *PLoS Pathog.* 4, e1000012.
- Nesbo, C.L., Dlutek, M., Doolittle, W.F., 2006. Recombination in thermotoga: implications for species concepts and biogeography. *Genetics* 172, 759–769.
- Ozawa, M., Basnet, S., Burley, L.M., Neumann, G., Hatta, M., Kawaoka, Y., 2011. Impact of amino acid mutations in pb2, pb1-f2, and ns1 on the replication and pathogenicity of pandemic (h1n1) 2009 influenza viruses. *J. Virol.* 85. doi:10.1128/JVI.00029–11.
- Papke, R.T., Koenig, J.E., Rodriguez-Valera, F., Doolittle, W.F., 2004. Frequent recombination in a saltern population of *Halorubrum*. *Science* 306, 1928–1929.
- Posada, D., 2002. Evaluation of methods for detecting recombination from DNA sequences: empirical data. *Mol. Biol. Evol.* 19, 708–717.
- Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* 98, 13757–13762.
- R Development Core Team, 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria 3–900051–07–0.
- Rambaut, A., Grassly, N.C., 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.* 13, 235–238.
- Rambaut, A., Pybus, O.G., Nelson, M.I., Viboud, C., Taubenberger, J.K., Holmes, E.C., 2008. The genomic and epidemiological dynamics of human influenza A virus. *Nature* 453, 615–619.
- Rolling, T., et al., 2009. Adaptive mutations resulting in enhanced polymerase activity contribute to high virulence of influenza A virus in mice. *J. Virol.* 83, 6673–6680.
- Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X., Rozas, R., 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19, 2496–2497.
- Ruderfer, D.M., Pratt, S.C., Seidel, H.S., Kruglyak, L., 2006. Population genetic analysis of outcrossing and recombination in yeast. *Nat. Genet.* 38, 1077–1081.
- Schweiger, B., Bruns, L., Meixnerberger, K., 2006. Reassortment between human A (H3N2) viruses is an important evolutionary mechanism. *Vaccine* 24, 6683–6690.
- Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* 51, 492–508.
- Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17, 1246–1247.
- Shinya, K., Hamm, S., Hatta, M., Ito, H., Ito, T., Kawaoka, Y., 2004. PB2 amino acid at position 627 affects replicative efficiency, but not cell tropism, of Hong Kong H5N1 influenza A viruses in mice. *Virology* 320, 258–266.
- Sibold, C., et al., 1999. Recombination in Tula hantavirus evolution: analysis of genetic lineages from Slovakia. *J. Virol.* 73, 667–675.
- Simon, P., Holder, B.P., Bouhy, X., Abed, Y., Beauchemin, C.A., Boivin, G., 2011. The I222V neuraminidase mutation has a compensatory role in replication of an oseltamivir-resistant influenza virus A/H3N2 E119V mutant. *J. Clin. Microbiol.* 49, 715–717.
- Smith, J.M., 1992. Analyzing the mosaic structure of genes. *J. Mol. Evol.* 34, 126–129.
- Smith, G.J., et al., 2009. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* 459, 1122–1125.
- Stamatakis, A., 2006. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690.
- Stavriniades, J., Guttman, D.S., 2004. Mosaic evolution of the severe acute respiratory syndrome coronavirus. *J. Virol.* 78, 76–82.
- Stratz, M., Mau, M., Timmis, K.N., 1996. System to study horizontal gene exchange among microorganisms without cultivation of recipients. *Mol. Microbiol.* 22, 207–215.
- Strimmer, K., von Haeseler, A., 1996. Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* 13, 964–969.
- Subbarao, E.K., London, W., Murphy, B.R., 1993. A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. *J. Virol.* 67, 1761–1764.
- Vijaykrishna, D., et al., 2010. Reassortment of pandemic H1N1/2009 influenza A virus in swine. *Science* 328, 1529.
- Webster, R.G., Bean, W.J., Gorman, O.T., Chambers, T.M., Kawaoka, Y., 1992. Evolution and ecology of influenza A viruses. *Microbiol. Rev.* 56, 152–179.
- White, D.J., Gemmell, N.J., 2009. Can indirect tests detect a known recombination event in human mtDNA? *Mol. Biol. Evol.* 26, 1435–1439.
- Wiuf, C., Christensen, T., Hein, J., 2001. A simulation study of the reliability of recombination detection methods. *Mol. Biol. Evol.* 18, 1929–1939.
- Worobey, M., Rambaut, A., Pybus, O.G., Robertson, D.L., 2002. Questioning the evidence for genetic recombination in the 1918 “Spanish flu” virus. *Science* 296, 211.
- Yang, Z., Nielsen, R., 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* 17, 32–43.
- Zhou, B., Li, Y., Halpin, R., Hine, E., Spiro, D.J., Wentworth, D.E., 2011. PB2 residue 158 is a pathogenic determinant of pandemic H1N1 and H5 influenza A viruses in mice. *J. Virol.* 85, 357–365.