An Insight to Segment Based Genetic Exchange in Influenza A virus: an *in silico* Study

Antara De [1,*], **Ashesh Nandy** [2]

[1]Centre for Interdisciplinary Research and Education, 404B Jodhpur Park, Kolkata 700068, India;
E-Mail: antaradnet@gmail.com
[2]Centre for Interdisciplinary Research and Education, 404B Jodhpur Park, Kolkata 700068, India;
Email-: anandy43@yahoo.com
*Author to whom correspondence should be addressed; Antara De E-Mail: antaradnet@gmail.com
Tel.: ++91-33-9038639153

Published: 4 December 2015

Abstract: Influenza virus is well recognized for high level of mutations that lead to development of new strains and subtypes, primarily through genetic shifts and drifts. Another mechanism of genetic change is through recombination, often observed in mammalian genes but controversial in viral genomes, which involves exchanges of short nucleotide sequences between two strains that coinfect the same cell. While evidence of such recombinations are rare to disputed in influenza genomes, we have observed that well-defined segments of influenza genes such as the hemagglutinin and neuraminidase have shown identical sequences between various strains that is best explained by segment exchange. Thus we had in our earlier study of the spread and proliferation of H5N1 bird flu observed that the neuraminidase with three segments – the transmembrane, stalk and body – shows evidence of exchange of one or other segments between different strains. Extending our work to the hemagglutinin of various subtypes, we noticed the same phenomena: Hemagglutinin has two segments, HA1 and HA2, where we found several instances where the segments seem to have been exchanged. Our analyses was based on RNA descriptors calculated in a 2D graphical representation scheme which have been proved to easily identify identical sequences. In this paper we discuss some of the details of this phenomenon in influenza genes which could be important in monitoring development of new highly pathogenic strains..

.

Keywords: Influenza; recombination; graphical representation

Mol2Net YouTube channel*: http://bit.do/mol2net-tube*

1. Introduction

Influenza A virus (IAV) has caused pandemic in human population since antiquity. It belongs to family Orthomyxiviridae and is a negative sense RNA virus with the genome divided into 8 segments which code for 11 proteins. The virus has been classified into different subtypes based on their cell-surface proteins hemagluttin (HA) and neuraminidase (NA). There are there are 18 HA and 11 NA subtypes identified [1]; however only few subtypes docking discrete combinations are found in nature [2]. These subtypes are further identified into different strains depending upon host type, geographical origin, strain number and year of isolation. Influenza exhibits remarkable degree of variability. Two well known mechanisms for the cause of variability are antigenic drift due to lack of proof-reading activity of RNA polymerase and antigenic shift by means of reassortment among the segmented genes of the genome [3]. There is also the question of whether recombination can cause variability, a phenomenon highly debated in negative sense RNA viruses [4]. Theoretically it can happen by non-homologous recombination between different genes or homologous recombination between same genes of different strains co-infecting the same cell. Our aim is to study the possibility of occurrence of homologous recombination in the HA and NA genes.

  The choice of the two surface proteins, HA and NA, for our study is guided by the fact while HA is responsible for viral entry into the host cell, NA is responsible for exit of the progeny virions from the host cell. NA gene has three subunits consisting of transmembrane (TM), stalk (ST) and body (BD) (Fig 1); The HA gene consists of two subunits, HA1 and HA2 (Fig 2). Since recombination relates to exchange of whole sections of the sequence, we want to study whether homologous recombination can happen at the cleavage points of the segmented HA and NA genes. We hypothesize that during replication the RNA polymerase swaps its template from one strain to the other at the cleavage point of the subunits, thus producing unique recombinant having genetic segments of both strains undergoing recombination (Fig 3)**.**

   To determine whether recombination could have taken place by exchange of segments as we hypothesized, we analyzed the major human infecting influenza HA gene sequences, viz., the H1N1, H5N1, H3N2 and H7N9 subtypes over the period 2010 to 2014 in Asia, and the H5N1 bird flu NA from 1997-2009 reported previously [5]. We report here the results of our study that yielded several instances where sequence identities between segments of various strains could be interpreted as homologous recombination via segment exchange, albeit as a small fraction of subtypes tested, but implying enlarged possibility of evolution of new strains of such negative sense RNA viruses.

    .

2. Results and Discussions

Our previous analyses of the neuraminidase sequences of the H5N1 bird flu epidemic of 1997-2009 had shown as an aside from the main thrust of the paper that there were worldwide evidences of duplications in part and whole of the sequences [5]. In particular, considering that the NA had three well-identified segments – transmembrane, stalk and body, there were evidences of sequences that had duplicates of one or two segments also. E.g., the transmembrane segment of A/treesparrow/Henan/4/2004(H5N1) was found to be identical to sequences from Hong Kong in 1999 and 2001

(A/Environment/HongKong/437-6/99(H5N1), A/Goose/HongKong/76.1/01 (H5N1)). The identity search among the rather large number of sequences was facilitated by computation of descriptors, g_R, of each segment whose equality implies sequence identity in the 2D graphical representation system as described in the Materials and Methods section.

This led to a number of sequence identities within each individual segment, which we had interpreted as being the result of RNA polymerase jump during replication in cases of co-infection in the same host cell, i.e. a homologous recombination. Although this was a preliminary study [5], the large number of such duplicated segmental sequences implied a new observation of homologous recombinations in negative sense viral genes.

These observations led us to replicate the analyses for possible such recombinations in hemagglutinin sequences of the Influenza genome. Our analysis of over a thousand hemagglutinin sequences using the 2D graphical representation descriptor, g_R, again revealed segmental duplication (more complete report in Ref [6]); though this time at a much lesser proportion than we had observed for the H5N1 neuraminidase. Of the 1200+ HA sequences analyzed, we found evidence for recombination in 73 instances of a daughter strain having its two segments from two parents under a strict protocol of considering only those daughter strains whose collection dates are marked later than the putative parents, and preferably from the same locality or country; Table 1 displays a short list of some selected sequences that display such connections indicated by the g_R values.. We expect such rigor if applied at the time would have reduced the number of recombination hits observed in the case of the neuraminidase counts.

However, some parent-daughter triplets have been identified with long time lag between their appearances, or the daughter sequence identified appears to have collection date before the parent. Based on previous studies that have shown that influenza strains can survive in the wild for years under suitable benign conditions depending on salinity, pH, temperature and other factors [7] it is possible that the apparent incongruent strain had taken part in the recombination in due course, but only discovered much later.

Segment exchange of the type we have hypothesized here do not fall into the "breakpoint" analysis favoured by researchers into these aspects. We note that such recombinations, if they are to happen, will take place during replication events when the RNA polymerase during its replication could jump from one template to another. For such jumps to take place accurately to create useful replicates, there has to be trigger points like the "breakpoints". This we hypothesize to be the cleavage point between the HA1 and HA2 which is well conserved. Table 2 illustrates this motif by listing 10 bases, 5 on either side, around the cleavage junction. Our hypothesis is that since there are two segments on the HA, the RNA polymerase can jump from one template to another at the cleavage point and continue replicating till the end.

This should be slightly more difficult for the neuraminidase with three segments where replication with jumps from the transmembrane or the stalk will require the polymerase to jump back to the remainder of the first template. The fidelity of the transit points on the neuraminidase sequences are yet to be studied in detail.

We note too that not all segmental exchanges take place at the same frequency. The percentage of recombination observed amongst the total number of strains investigated in the

neuraminidase genes varied significantly between segments. E.g., 2/3rds of all transmembranes showed high degree of similarity as shown in Table 4 of Ref [5]; taking a maximum of 1/3 of these being recombinants, that works out to 2/9ths of all H5N1 strains investigated, admittedly an upper bound. Similarly, it works out to <1/5th for the stalk region and <2% for the body segment. For the hemagglutinin, the recombinations varied by subtype: from 3.74% for H1 of H1N1 to 5.71% for H5 of H5N1, 7.35% of H3N2 and 7.87% of the H7N9 in our database.

 It is noteworthy to watch the apparent disparity among four HA subtypes exhibiting different recombination frequency. However the phenomena can be explained by the fact that sometimes they have different evolutionary history as exemplified by H1 and H3 subtypes [8, 9] or a low pathogenic avian influenza virus (LPAIV) being suddenly mutated to a highly pathogenic avian influenza virus (HPAIV) as

exemplified by H5 and H7 subtypes of H5N1 and H7N9 Influenza A virus which created bird flu and China Flu, 2013[10]. Often these HPAIV are adapted to human host; though possibility of human to human transmission remains debatable. More details can be found in our comprehensive report [Ref 6]. Interestingly, all valid recombinant possibilities noted above appears to have been restricted to parents and daughters from the same hemagglutinin subtype, i.e. there were no examples of mixed subtype marriages. It appears probable that the distinct antigenic sites in the different subtypes play a role to restrict the possible recombinations; the different base compositions, apart from the base distributions, thus may play a regulatory role.

Table 1. Few representative recombinants, identified on the basis of numerical characterization algorithm

Sub Type	Locus id	Description	g_R(HA1)	g_R(HA2)
H1N1	KM029055 (Parent 1)	A/swine/Hong Kong/NS4846/2011(H1N1)	66.18646	49.33498
	KM029103 (Parent2)	A/swine/Hong Kong/4902/2011(H1N1)	65.75885	49.06597
	KM029063 (Daughter)	A/swine/Hong Kong/NS4848/2011(H1N1)	66.18646	49.06597
H1N1	KM028335 (Parent 1)	A/swine/Guangxi/3614/2011(H1N1)	67.93622	49.80615
	KM028423 (Parent2)	A/swine/Guangxi/3880/2011(H1N1)	69.77442	48.80638
	KM028271 (Daughter)	A/swine/Guangxi/NS3248/2011(H1N1)	67.93622	48.80638
H3N2	KM069501 (Parent 1)	A/Singapore/C2010.307/2010(H3N2)	57.06303	36.58159
	JX437712 (Parent2)	A/Singapore/C2010.036V/2010(H3N2)	59.02475	37.64934
	JX437832 (Daughter)	A/Singapore/H2010.370C/2010(H3N2)	57.06303	37.64934

H3N2	JX437832 (Parent 1)	A/Singapore/H2010.370C/2010(H3N2)	57.06303	37.64934
	KM069503 (Parent2	A/Singapore/H2010.310/2010(H3N2)	56.53434	36.58159
	KM069501 (Daughter)	A/Singapore/C2010.307/2010(H3N2)	57.06303	36.58159
H5N1	KF369229 (Parent 1)	A/chicken/Cambodia/W0530391/2012(H5N1	61.16050	37.63436
	KF369214 (Parent2)	A/Cambodia/W0526301/2012(H5N1)	60.92035	37.15533
	KF369222 (Daughter)	A/chicken/Cambodia/W0530389/2012(H5N1)	61.16050	37.15533
H5N1	KF001474 (Parent 1)	A/civet cat/Cambodia/X0313306/2013(H5N1)	61.85977	36.95179
	KF001497 (Parent2)	A/duck/Cambodia/X0220302/2013(H5N1)	60.53024	39.06615
	KF001478 (Daughter)	A/civet cat/Cambodia/X0313307/2013(H5N1)	61.85977	39.06615
H7N9	CY147148 (Parent 1)	A/environment/Shanghai/S1438/2013(H7N9)	50.61318	44.19611
	CY147188 (Parent 2)	A/pigeon/Shanghai/S1423/2013(H7N9)	51.11281	43.71347
	CY147132 (Daughter)	A/environment/Shanghai/S1436/2013(H7N9)	50.61318	43.71347
H7N9	KF542876 (Parent 1)	A/chicken/Shanghai/017/2013(H7N9)	52.00389	43.08626
	KF667751 (Parent 2)	A/environment/Guangdong/30/2013(H7N9)	52.27365	43.67379
	KF667746 (Daughter)	A/environment/Guangdong/25/2013(H7N9)	52.00389	43.67379

Table 2: Conserved region at the cleavage point of HA1 and HA2 segments.

Locus ID	Description	HA1 end	HA2 begin
JF275925	A/swine/Nanchang/F9/2010(H1N1)	ct**aga**	**gg**cct
KJ955515	A/Ho Chi Minh/459.6/2010(H3N2)	ct**aga**	**gg**cat
KM276899	A/swine/Taiwan/NPUST0004/2013(H3N2)	at**aga**	**gg**cat
CY091837	A/Guangdong/322/2010(H3N2)	ct**aga**	**gg**cat
CY124183	A/Singapore/GP1147/2011(H3N2)	ct**aga**	**gg**cat
CY116636	A/Tbilisi/GNCDC0485/2012(H3N2)	ct**aga**	**gg**cat
CY124187	A/Singapore/GP1188/2011(H3N2)	ct**aga**	**gg**cat
AB569348	A/whooper swan/Mongolia/1/2010(H5N1)	aa**aga**	**gg**act
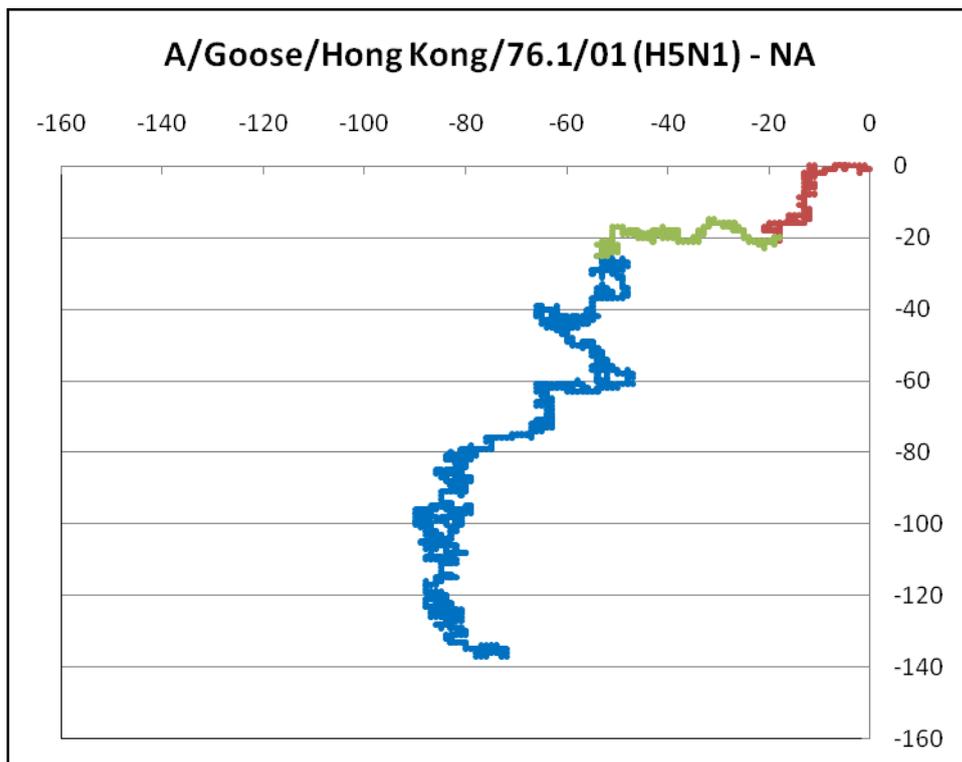
Figure 1: 2D graphical representation of a H5N1 strain neuraminidase sequence
(Segments: Brown - Transmembrane; Green - Stalk; Blue - Body)
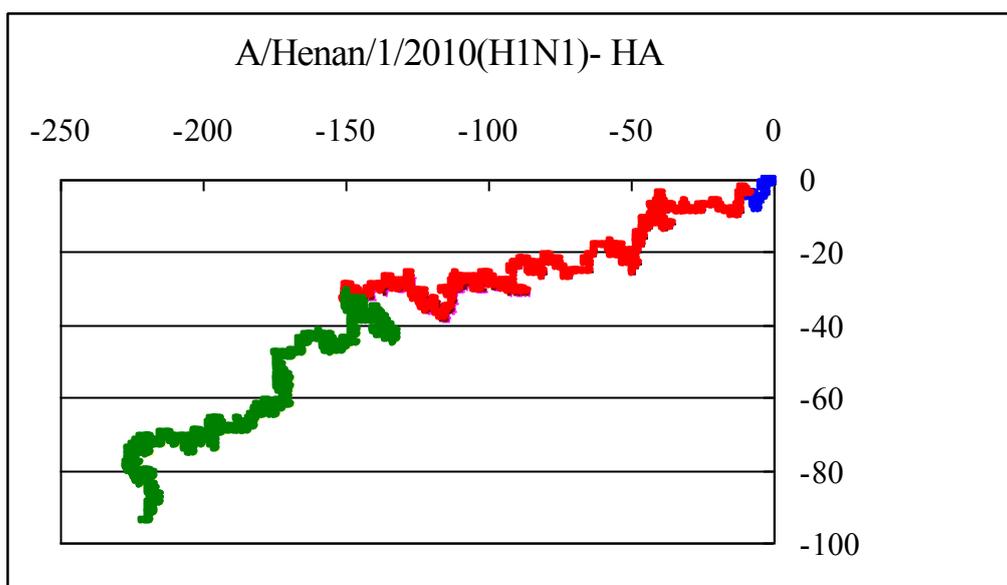
Fig 2: 2D graphical representation of a H1N1 strain hemagglutinin sequence.
(Segments: Blue – Signal peptide; Orange- HA1; Green – HA2)
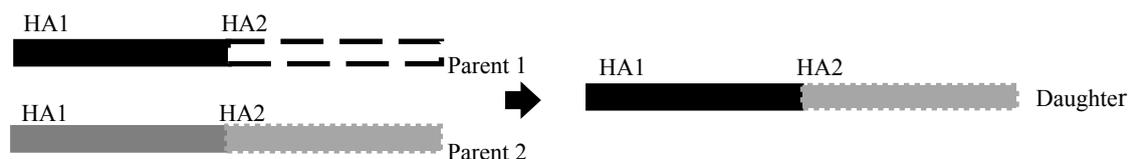
Figure 3. Recombination in HA gene of Influenza virus by genetic exchange at the cleavage point of HA1 and HA2 segments

3. Materials and Methods

Database: We downloaded 1274 full length HA sequences available in Genbank which were major human infecting subtypes i.e H1N1, H5N1, H3N2 and H7N9. All sequences were from Asia during time period of 2010 to 2014. Similarly we analyzed 682 NA of H5N1 subtype available in Genbank till March 2009.

Calculation of sequence similarity: We analyzed the sequences based on a 2D graphical representation system and determined sequence identity from the numerical characterization algorithm [11, 12]. In this method starting from the origin on Cartesian axes a 2D graphical plot is generated for HA and NA sequence of each strain by moving one step in the negative x-direction for an adenine, one step in the positive y-direction for a cytosine, one step in the positive x-direction for a guanine, and one step in the negative y-direction for a thymine. The 2D graphical plot so generated provides a visual representation of the base distribution pattern of the sequence. To numerically characterize the sequence we define a weighted centre of mass of the plot and a graph radius g_R as follows [12] :

$$\mu_x = \frac{\sum_{i=1}^{N} x_i}{N} \ , \ \ \mu_y = \frac{\sum_{i=1}^{N} y_i}{N}$$

$$g_R = \sqrt{\mu_x^2 + \mu_y^2}$$

where x_i, y_i represents the coordinates of the ith base and N is the total number of nucleotides in the sequence under consideration. The graph radius g_R is a base distribution index of the nucleotide sequences. g_R is found to be sensitive to any changes in the base distribution such that sequences having same value of g_R imply sequence identity [13].

.

4. Conclusions

Our recombination study is based on novel concept of segment exchange at the cleavage points of the well defined segments of a gene that codes for different subunits of the associated protein. Previous recombination studies on the influenza virus had been based on breakpoint analysis [14], where the polymerase recognizes a breakpoint sequence and creates a recombinant strain by switching back and forth between two parents, thus producing a daughter having copies of parts of sequences from both the parents. However, breakpoint recombination is a highly debated topic being postulated by some scientists while refuted by others [15, 16]. It is known that RNA polymerase remains loosely attached with few nucleotides of the template strand [17]. Our hypothesis is that such a loose attachment can facilitate a jump between two templates at consensus sequences, which, in the case of the HA and NA could be the cleavage points of their

intrinsic segments. Our analyses revealed many examples where inter-segmental exchanges have led to new daughter strains being developed, thus demonstrating this kind of recombination as a valid evolutionary process. The fact that the exchanges we have observed all appear to occur within same subtypes implies more mechanisms at work than merely the consensus sequences, which though we are yet to clearly understand. However, this new evolutionary process, albeit restricted by segmental recombinations within same subtypes, opens up possibilities of development of many more subtype strains, some of which could turn out to be more pathogenic than the original subtype itself.

Acknowledgments

Author Contributions

AN suggested the problem, AD did the calculations and wrote the paper, which AN critically reviewed and edited.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Tong, S; Zhu, X; Li, Y; Shi, M, Zhang, J; Bourgeois, M; Yang, H; Chen, X; Recuenco, S; Gomez, J; Chen, L.M; Johnson, A; Tao, Y; Dreyfus, C; Yu, W; McBride, R; Carney, P. J; Gilbert, A. T; Chang, J; Guo, Z; Davis, C.T; Paulson, J.C; Stevens, J; Rupprecht, C.E; Holmes, E.C; Wilson, I.A; Donis, R.O. New World Bats Harbor Diverse Influenza A Viruses. PLoS Pathogens 2013, 9(10),e1003657
2. Nandy, A; Sarkar, T; Basak, S.C; Nandy, P; Das, S. Characteristics of influenza HA-N interdependence determined through a graphical technique. Curr Comput Aided Drug Des. 2014,10(4), 285-302.
3. Kageyama, T; Fujisaki, S; Takashita, E; Xu, H; Yamada, S; Uchida, Y; Neumann, G; Saito, T; Kawaoka, Y; Tashiro, M; Genetic analysis of novel avian A(H7N9) influenza viruses isolated from patients in China February to April 2013. Euro Surveill. 2013, 8(15), 20453
4. Worobey, M; Rambaut, A; Pybus, O.G; Robertson, D.L. Questioning the evidence for genetic recombination in the 1918 "Spanish flu" virus. Science 2002, 296(5566), 211
5. Ghosh, A; Nandy A; Nandy, P; Gute, B.D; Basak, S.C. Computational Study of Dispersion and Extent of Mutated and Duplicated Sequences of the H5N1 Influenza Neuraminidase over the Period 1997-2008. J. Chem. Inf. Model. 2009, 49(11), 2627–2638.
6. De, A; Sarkar, T; Nandy, A. Bioinformatics studies of Influenza A hemagglutinin sequence data indicate recombination-like events leading to segment exchanges. (Communicated)
7. Stallknecht, D.E.; Brown, J.D. Tenacity of avian influenza viruses. Rev. Sci. Tech. 2009, 28 (1), 59-67

8. Lindstrom, S.E.; Hiromoto, Y ; Nerome, R; Omoe, K; Sugita, S; Yamazaki, Y; Takahashi, T; Nerome, K. Phylogenetic Analysis of the Entire Genome of Influenza A (H3N2) Viruses from Japan: Evidence for Genetic Reassortment of the Six Internal Genes. J Virol. 1998 ,72, 8021-8031.

9. Zhong, J; Liang, L; Huang, P; Zhu, X; Zou, L; Yu, S; Zhang, X; Zhang, Y; Ni, H; Yan, J. Genetic mutations in influenza H3N2 viruses from a 2012 epidemic in Southern China. Virol J 2013, 10,345.

10. Sarkar, T; Das, S; De, A; Nandy, P; Chattopadhyay, C; Chawla-Sarkar M; Nandy, A. H7N9 influenza outbreak in China 2013: *In silico* analyses of conserved segments of the hemagglutinin as a basis for the selection of peptide vaccine targets. Computational Biology and Chemistry 2015, 59, 8 - 15.

11. Nandy, A. A new graphical representation and analysis of DNA sequence structure: I. Methodology and Application to Globin Genes. Current Sc 1994, 66(4), 309-314.

12. Raychaudhury, C.; Nandy, A. Indexing Scheme and Similarity Measures for Macromolecular Sequences. J Chem Infor Comput Sc. 1999, 39, 243-247.

13. Nandy, A.; Nandy, P. On the uniqueness of quantitative DNA difference descriptors in 2D graphical representation models. Chem Phys Letters 2003, 368,102-107.

14. Boni, M.F; Zhou, Y; Taubenberger, J.K; Holmes, E.C. Homologous recombination is very rare or absent in human influenza A virus. J Virol 2008, 82(10), 4807-4811.

15. Hao, W. Evidence of intra-segmental homologous recombination in influenza A virus Gene 2011, 481(2), 57-64.

16. Boni, M.F; Smith, G.J; Holmes, E.C; Vijaykrishna D. No evidence for intra-segment recombination of 2009 H1N1 influenza virus in swine. Gene. 2012, 494(2), 242-245.

17. Fleischmann, Jr., W.R. Medical Microbiology, 4th edition; Chapter 43 Viral Genetics. The University of Texas Medical Branch, Galveston, Texas, USA. 1996 (Online access at http://www.ncbi.nlm.nih.gov/books/NBK8439/