SciForum
Mol2Net

Interdependence of Influenza HA and NA and Possibilities of New Reassortments

Ashesh Nandy [1,*] and Subhas Basak [2]

[1] Centre for Interdisciplinary Research and Education, 404B Jodhpur Park, Kolkata 700068, INDIA

[2] University of Minnesota Duluth-Natural Resources Research Institute (UMD-NRRI) and Department of Chemistry & Biochemistry, University of Minnesota Duluth, Duluth, MN 55811, USA

* Author to whom correspondence should be addressed; E-Mail: anandy43@yahoo.com; Tel.: +91-94335-79452.

Received: 28 October 2015 / Accepted: 30 October 2015 / Published: 2 December 2015

Abstract: The influenza virion is characterized by two surface proteins, hemagglutinin (HA) and neuraminidase (NA). The changes in their surface antigenic sites have given rise to several subtypes – H1 to H16 for the hemagglutinin and N1 to N9 for the neuraminidase, and each influenza strain is identified with these subtypes such as the H5N1, H7N9, etc. Of the 16 x 9 combinations possible, only certain combinations are observed to proliferate in the wild, such as the H1N1, H3N2, H5N1, etc. This interdependence of the HA and NA on certain subtypes have been noticed, and experimentally demonstrated, but the underlying cause or its systematics have been unknown. We have hypothesized that the base distribution characteristics of the HA and NA constitute a coupling between them. We estimate the coupling strength by measuring the distance in graph radii between the two genes in a graphical representation scheme. We found that this distance was characteristic of each subtype combination and forced combinations with a different HA or NA subtype led to widely different values, which by our hypothesis, and the experimental findings of Zhang et al, implied unstable combinations. This hypothesis implies that given a stable subtype of pathogenic influenza, we can estimate using the coupling constants which other subtype combinations could emerge through reassortment. Thus in the case of the H5N2 strain which had an epidemic form in North America in 2015, we have calculated the consequences of altering the NA component. We found that only H5N4, H5N6 and H5N9 combinations could match the coupling strength of the H5N2, thus implying that the next epidemics could arise from these combinations rather than other subtypes of H5. This allows for more focused monitoring of emerging flu strains for epidemic potential.

.

Keywords: influenza HA-NA coupling strengths; bird flu, neuraminidase; interdependence; new subtypes; HA-NA hemagglutinin

Mol2Net YouTube channel*:*

http://bit.do/mol2net-tube

YouTube link:

https://www.youtube.com/watch?v=vOyKJxv-IVo

1. Introduction

Influenza is a widely prevalent seasonal viral infection that afflicts several million people annually with fatalities that range into tens of thousands. The influenza virus is of three main types, Influenza A, B and C, of which Influenza A is the most prevalent and affects birds, mammals and humans. The influenza A virus genome is a segmented negative strand RNA virus with eight distinct pieces of the nucleic acid coding for 11 proteins. Two of these - hemagglutinin (HA) and neuraminidase (NA) - are surface situated glycoproteins responsible for virus endocytosis and progeny elution. On the basis of their antigenic segments several subtypes of these proteins have been identified – 16 for the HA and 9 for the NA, enabling the viral strains to be characterized as H1N1, H3N2, H7N9, etc.

It has been observed that while 16 x 9 combinations of the H*x*N*y* variety of viral strains are possible, only certain combinations seem to predominate in the wild [1], as can be seen from the database population of the various flu strains [2]. The reason for this interdependence of one subtype of HA for a specific subtype of NA and vice versa is not known, but that such an interdependence exists has been verified by the experiments of Zhang et al [3] who swapped different subtypes to show that outside the preferred subtype combinations the strains were not stable or adequately infective; this has also been observed in other subtype combinations [4].

This issue is quite critical. The evolution of different strains of influenza leads to the possibilities of emergence of epidemic and pandemic strains that can affect very large number of host species as had happened with the Spanish Flu (subsequently identified as a H1N1 subtype) of 1918 where the human death toll exceeded 20 million, the Swine Flu (H1N1) of 2009 where several million people were infected and over 25,000 died within one year, the H5N1 bird flu that apparently surfaced around 1997 and the H7N9 flu of 2013 whose containment necessitated culling of millions of chicken, and the recent H5N2 avian flu epidemic in North America (2015) that has led to culling of millions of poultry and farm birds. Such strains arise through genetic shift and drift, of which reassortments among the various genes, especially the glycoproteins, which can occur when two flu strains infect a single cell, are among the most important processes [5,6]. Such possibilities require incessant monitoring of evolving subtypes and combinations worldwide, a rather daunting, but inescapable task. Taking interdependence of HA and NA into account can help to reduce this task to more manageable proportions since monitoring one of them, say HA, automatically accounts for the associated NA that can retain the infectivity. Quantifying the interdependence of HA and NA therefore can be instrumental in this enterprise.

Our study of this HA-NA interdependence [2] led us to hypothesize a coupling between the two glycoproteins arising out of their base distribution and composition characteristics which are best visualized in a graphical representation. Our research showed that the major influenza subtype combinations had very distinct coupling strengths and interchange between different subtype components compromised such strengths, an outcome in keeping with Zhang et al's experimental results [3]. As a consequence of this model, we considered the current H5N2 avian epidemic in

the USA and could forecast two possible reassortment products that could conceivably fuel new epidemics [7]. This report very briefly summarizes our methodology and these results and observations.

.

2. Results and Discussion

Graphical representation of the HA and NA sequences of H1N1, as described in the Methods section and shown in Fig.1, depicts the base distributions of the two genes. Our hypothesis of a coupling between the two genes is predicated upon the assumption that the base distributions are mutually dependent; indeed, Hu [8] has shown that mutational changes in one lead to a coupled mutational change in the other. We quantify this inter-dependence by the distance between the end-points of the graph radii of the two plots as defined in the Methods section.

The point is that if the coupling between the HA and the NA were to be characteristic of the specific subtype, implying co-ordinated change of some kind, then irrespective of the genetic shifts in the two gene sequences, the coupling factor as measured by the distances of the graph radii should remain constant within a reasonable limit. As shown in Table 1, taking a large sample of the H1N1 strains we found that the coupling factor worked out to 39.49±8.19. Similar analyses for other strains of recent interest showed similar trends. Table 1 lists the results of our analysis with all viral strains in our database showing the coupling actors for each individual viral strain and also an average for each flu subtype where adequate number of strains was available. We notice that this average is different for each subtype with a reasonably low standard deviation implying that each flu subtype has a specific coupling strength, which we may refer to as its characteristic value.

For the coupling to be characteristic of each flu subtype, replacing one gene with another variety should produce quite different result for the coupling factor. In fact, as we replaced the H1 sequence of A/South Dakota/01/2011(H1N1) with a H5 sequence from A/duck/France/05066b/2005(H5N1), the coupling factor changed from 33.97 to 12.03; exchanging only the NA between the two strains changed the coupling factor to 69.72 implying gross lack of compatibility between the HAs and NAs of the two strains. This effect we found in a wide variety of samples tested as shown in Table 2. We note that the HA exchange produced less dramatic or insignificant effects than NA since the HA sequences are more homologous across all HA subtypes compared to the NA subtypes: taking typical examples each of all subtypes of HA and NA, we find that in terms of the composition of the four bases a, c, g, t the standard deviation from the average composition values is <3% for a and t (a: 2.6%, t: 2.8%) and <4% for g and c (g:3.7%, c:3.8%) for the HA, whereas these figures are >3% for a,t (a:3.7%, t:9.2%) and >6% (g:6.7%, c:6.2%) for the NA; the wide variation of the NA and the comparatively lesser variation of the HA subtypes is evident too in the graphical representations shown in Ref 2.

These results of our analyses, summarized in Tables 1 and 2, show that forced exchanges between the HA and NA of the flu strains often lead to coupling factors widely different from the characteristic values. Our observations tie in neatly with the experimental results of Zhang et al [3] who found from HA, NA exchanges within a set of 1918 pandemic H1N1, a 2009 pandemic H1N1 and a HPAI H5N1 that the NA exchanges led to significant decline in influenza infectivity whereas the effect was comparatively much less when the HAs were exchanged. This he attributed to lack of “matching patterns” between the NA of the H5N1 with the HAs of the H1N1s

in the experiment. From the observation noted earlier that the availability in the wild of only a few wild subtypes of flu may imply low stability of other subtypes, and the observations of wide variation in the computed coupling value and Zhang et al's results, we may infer that such "forced" subtypes will not yield stable or efficient infective strains.

This leads us to an interesting prognosis. This year has witnessed a sudden epidemic of highly pathogenic avian influenza (HPAI) H5N2 infection among poultry and farm raised birds like chicken and turkeys in the North American west and mid-west leading to death through infection or culling of millions of birds [9]. While the virus has not affected humans yet, strict monitoring is being done to ensure adequate warning in case the virus develops human-to-human transmission ability [10]. At the same time a watch has to be kept on the possibility that the virus could undergo reassortments and give rise to new subtypes, though one does not know which of the possible subtypes could be highly pathogenic too.

Our analysis provides a guideline here. Once we know that the H5N2 is a HPAI virus, we can forecast which of the possible reassortants have the potential to be stable and possess pathogenic ability [7]; it is pertinent to note here that according to the USDA, the current H5N2 subtype is itself a combination of Asian HPAI H5 and North American N2 [11,12]. We accessed all North American H5N2 gene sequences available at the time, i.e., around mid-May 2015, and determined that the magnitude of the coupling for these strains as measured through the delta-g_R was 38.58±1.46. To assess possible reassortments from these strains we are mindful of the fact that Asian H5 is a highly pathogenic virus that in its H5N1 bird flu form had caused high level of human fatalities at a mortality rate of 1 per 2 infections [13], and a continuing fear that the virus may mutate to a form causing human-to-human transmission leading to a new pandemic [14]. Taking such a HA as one component of the possible reassortants, we tried combinations with all subtypes of the NA available from typical flu sequences (Table 3). Taking a cue from the results given in Table 1 that the standard deviations of the coupling values between the various strains of the flu subtypes is 18.85% (range: 7.65% – 29.12%), we can look for those HA-NA combinations that lie within this range. The results as shown in Table 3 indicate that only combinations of the H5 with a possible N4, N6 or N9 fall within this coupling value range and therefore could be the new HPAI to evolve from reassortments of the H5N2 with other flu subtypes. Our research showed that flu subtypes with these varieties of the neuraminidase have been reported in various places in North America, indicating that it is possible for reassortments of the HPAI H5N2 with these subtypes to take place. While monitoring for genetic shifts and drifts of the H5N2 in North America, close attention, therefore, may be given to development of H5N4, H5N6 and H5N9, if any, of which H5N9 may bear extra scrutiny since a hitherto benign to human H7N9 strain in China suddenly developed a mutation in 2013 that led to human fatalities. Such focused scrutiny might reduce the monitoring overhead to some extent to concentrate on the more potent possibilities. The same exercise can be done with other HA subtypes, but as we have seen, the flu subtypes are more sensitive to the changes in NA.

Table 1. Coupling factors of HA-NA interdependence. The last two columns provide summary data for each major flu type indicated in the first column

Flu Type	Locus ID		Virus Strain Description	Year	gR values		Coupling factor η	Group	
	HA	NA			NA	HA		Averages	Std Dev
H1N1	CY016699	CY016701	A/South Australia/58/2005(H1N1)	2005	113.6024	100.9997	25.37991		
	GQ150342	CY039988	A/Nonthaburi/102/2009(H1N1)	2009	95.58367	118.4997	32.45575		
	KC881952	KC881951	A/South Dakota/01/2011(H1N1)	2011	95.53349	120.0206	33.97239		
	CY039893	CY039895	A/New York/1669/2009(H1N1)	2009	95.73236	120.473	35.51551		
	FJ998208	FJ998214	A/Mexico/InDRE4487/2009(H1N1)	2009	94.97363	120.473	36.33877		
	CY039999	CY040001	A/New York/3008/2009(H1N1)	2009	95.49054	120.9794	36.44679		
	CY039901	CY039903	A/New York/1682/2009(H1N1)	2009	95.45042	120.8896	36.68635		
	KF648252	KF648260	A/Washington/05/2013(H1N1)	2013	93.2853	118.2515	36.95131		
	CY040007	CY040009	A/New York/3012/2009(H1N1)	2009	95.49054	121.9571	37.39781		
	GQ338364	GQ117071	A/Minnesota/02/2009(H1N1)	2009	94.78558	122.4494	38.14657		
	CY148235	CY039528	A/Netherlands/602/2009(H1N1)	2009	92.11382	120.1784	38.49428		
	FJ966952	FJ966956	A/California/05/2009(H1N1)	2009	93.60996	121.7123	38.42935		
	KC781785	GQ377078	A/California/07/2009(H1N1)	2009	94.87404	121.6822	38.93651		
	CY134351	CY134353	A/green-winged teal/California/123/2012(H1N1)	2012	91.47518	83.26524	40.19993		
	CY134359	CY134361	A/northern shoveler/California/138/2012(H1N1)	2012	93.76871	77.66515	33.03013		
	CY134367	CY134369	A/northern pintail/California/183/2012(H1N1)	2012	94.10496	76.34555	34.85523		
	CY077076	CY077078	A/mallard/Sanjiang/390/2007(H1N1)	2007	85.39725	96.98963	55.31523		
	EU026037	EU026039	A/mallard/Maryland/170/2002(H1N1)	2002	89.38879	76.97491	36.1104		
	EU743306	EU743308	A/blue winged teal/LA/B228/1986(H1N1)	1986	87.14094	83.45928	33.5418		
	AB546149	AB546151	A/pintail/Aomori/422/2007(H1N1)	2007	80.85941	88.08622	47.88056		
	AB670330	AB472014	A/duck/Tsukuba/718/2005(H1N1	2005	83.23725	91.84633	54.27835		
	HM193551	HM193628	A/mallard/Alaska/44430-088/2008(H1N1)	2008	91.17319	78.89723	41.29855		
	FJ432778	FJ432780	A/goose/Italy/296426/2003(H1N1)	2003	84.48628	95.62911	54.02458		
	CY014627	CY005686	A/duck/AUS/749/1980(H1N1)	1980	81.72422	92.90199	21.36802		
	FJ536818	FJ536824	A/swine/Shandong/443/2008(H1N1)	2008	88.04054	113.9492	50.04761		
	AB741039	AB741041	A/swine/Narita/aq21/2011(H1N1)	2011	95.94873	122.2932	36.65832		
	GQ229357	GQ229362	A/swine/Hong Kong/9656/2001(H1N1)	2001	109.2822	128.8602	42.5942		
	EU004444	EU004442	A/swine/Tianjin/01/04(H1N1)	2004	105.3392	106.7129	46.18326		
	CY085774	CY085776	A/swine/Hong Kong/434/2006(H1N1)	2006	99.01507	119.5231	52.64454	39.48903	8.187229
H1N2	JX069105	JX069107	A/ostrich/South Africa/AI2887/2011(H1N2)	2011	76.81041	86.06502	17.21649		
	AB741007	AB741009	A/swine/Tochigi/2/2011(H1N2)	2011	91.94878	113.8709	22.75411		
	CY133290	CY133292	A/mallard/Mississippi/10OS4593/2010(H1N2)	2010	82.69537	79.35445	33.5879		
H3N1	CY005943	CY004711	A/mallard duck/ALB/26/1976(H3N1)	1976	87.08971	76.95153	63.14924		
H3N2	CY006467	CY006469	A/New York/516/1997(H3N2)	1997	86.58321	91.76747	21.82011		
	CY091261	CY091263	A/Singapore/NHRC0007/2003(H3N2)	2003	83.2928	96.42491	22.33844		
	AB295605	AB295606	A/Aichi/2/1968(H3N2)	1968	77.86263	63.87713	27.37842		
	CY092233	CY092235	A/Australia/NHRC0010/2005(H3N2)	2005	85.46052	101.1324	27.38328		
	CY116638	CY116639	A/Tbilisi/GNCDC0557/2012(H3N2)	2012	88.55144	103.2946	28.25711		
	CY105870	CY105872	A/KhanhHoa/KH475/2008(H3N2)	2008	84.56058	106.1538	30.82751		

	CY105886	CY105888	A/HaNoi/311/2005(H3N2)	2005	87.52267	105.8014	31.34143		
	AB741031	AB741033	A/swine/Yokohama/aq138/2011(H3N2)	2011	90.78071	97.3364	13.3909		
	CY131029	CY131031	A/swine/Ohio/10SW136/2010(H3N2)	2010	84.88113	98.66828	16.12007		
	AB277754	AB277755	A/duck/Hokkaido/5/1977(H3N2)	1977	79.99952	67.8499	31.90895		
	KC422461	KC422462	A/feline/Guangdong/1/2011(H3N2)	2011	81.79576	94.96184	37.39615		
	KF155145	KF155147	A/canine/Korea/MV1/2012(H3N2)	2012	81.43785	95.31512	37.58191		
	KC422456	KC422458	A/feline/Korea/01/2010(H3N2)	2010	83.5886	92.22999	41.06222	28.21588	8.21681
H5N1	AB212054	AB212056	A/Hong Kong/213/2003(H5N1)	2003	82.66458	110.4352	48.87837		
	GU052518	GU052520	A/chicken/Scotland/1959(H5N1)	1959	86.35189	92.52505	33.31591		
	CY053325	JF758823	A/wood duck/Ohio/623/2004(H5N1)	2001	94.39676	112.5891	36.4933		
	AJ971297	AJ972921	A/duck/France/05066b/2005(H5N1)	2005	94.821	85.81665	43.58987		
	AF144305	AF144304	A/goose/Guangdong/1/1996(H5N1)	1996	76.56076	102.1669	44.23744		
	EF607855	EF607897	A/mute swan/MI/451072-2/2006(H5N1)	2006	87.95686	115.3977	45.63799		
	AF364334	AF364335	A/Goose/Guangdong/3/97(H5N1)	1997	74.52595	101.9541	47.47952		
	KC815853	KC815851	A/mallard/Italy/3401/2005(H5N1)	2005	83.11546	99.35829	52.03776		
	AY585373	AY585404	A/duck/Guangdong/07/2000(H5N1)	2000	82.8014	105.5828	52.64266		
	GU052073	GU052075	A/Goose/Hong Kong/3014.5/2000(H5N1)	2000	92.07489	107.368	53.95241		
	AY747617	AY747618	A/swine/Fujian/F1/2001(H5N1)	2001	86.5242	103.6867	56.02981	46.7541	7.134044
H5N8	CY134101	CY134103	A/mallard/California/2559P/2011(H5N8)	2011	76.93981	112.062	56.45905		
H6N1	AB298279	AB298280	A/duck/Hokkaido/W159/2006(H6N1)	2006	87.5451	102.213	24.3545		
	HM144392	HM144562	A/duck/Hunan/177/2005(H6N1)	2005	85.58079	90.42478	24.70176		
	HM144388	HM144558	A/mallard/Jiangxi/227/2003(H6N1)	2003	83.2108	90.20036	30.04321		
	EF681878	EF681880	A/chicken/Taiwan/2838V/00(H6N1)	2000	89.26956	92.06014	31.88554		
	AB294215	AB294216	A/duck/Hong Kong/716/1979(H6N1)	1979	78.10185	95.91201	33.8106		
	GQ414872	GQ414903	A/spot-billed duck/Korea/545/2008(H6N1)	2008	89.70904	92.77493	40.21473		
	GQ414864	GQ414904	A/mallard/Korea/L08-8/2008(H6N1)	2008	89.54615	94.47328	42.40703	32.4882	6.987585
H7N3	CY125730	CY125728	A/Mexico/InDRE7218/2012(H7N3)	2012	86.18209	73.54771	33.07789		
H7N9	KC853228	KC853231	A/Shanghai/4664T/2013(H7N9)	2013	103.7236	91.82757	38.80746		
	KC885956	KC885958	A/Zhejiang/DTID-ZJU01/2013(H7N9)	2013	106.4451	93.09385	40.50441		
	KC853766	KC853765	A/Hangzhou/1/2013(H7N9)	2013	105.9556	93.48328	41.09839		
	KC994453	KC994454	A/Fujian/1/2013(H7N9)	2013	108.0189	93.09385	41.31705		
	KF420298	KF420296	A/Changsha/1/2013(H7N9)	2013	106.6576	93.71526	42.77973		
	KF278746	KF226113	A/Jiangsu/1/2013(H7N9)	2013	106.8237	92.96191	43.35736		
	KF469231	KF261988	A/Nanchang/1/2013(H7N9)	2013	106.6445	93.71526	42.52465		
	KC899669	KC899671	A/chicken/Zhejiang/DTID-ZJU01/2013(H7N9)	2013	106.2791	85.76458	38.11029		
	AY999981	KF695256	A/Mallard/Sweden/91/02(H7N9)	2002	104.6166	72.43222	40.03544		
	GU060482	GU060484	A/goose/Czech Republic/1848-K9/2009(H7N9)	2009	98.69629	74.68026	43.06948		
	CY133649	CY133651	A/northern shoverl/Mississippi/11OS145/2011(H7N9)	2011	111.8206	69.69097	63.20246		
	CY067678	CY067680	A/blue-winged teal/Guatemala/CIP049-02/2008(H7N9)	2008	113.4246	72.32552	64.78273		
	AB481213	AB481212	A/duck/Mongolia/119/2008(H7N9)	2008	120.3336	79.60081	50.0517		
	KJ508892	KJ508890	A/tree sparrow/Shanghai/01/2013(H7N9)	2013	106.631	91.28917	41.58104	45.0873	8.488588

H9N2	JQ440373	JQ916910	A/chicken/Egypt/114940v/2011(H9N2)	2011	80.29504	113.7064	40.8703		
mixed	CY103245	CY103247	A/mallard/Alberta/115/2007(mixed)	2007	89.43659	82.58275	21.97967		
H5N1 short stalk strains:									
H5N1	AB239125	AB239126	A/Hanoi/30408/2005(H5N1)	2005	68.38075	112.7598	59.9053		
	DQ371928	EU128239	A/Anhui/1/2005(H5N1)	2005	75.77015	112.8584	69.38676		
	CY098681	CY098683	A/Anhui/1/2007(H5N1)	2007	73.62005	112.5678	66.89482		
	DQ835313	DQ835315	A/China/GD01/2006(H5N1)	2006	73.4652	115.0801	72.79483		
	AB478025	AB478033	A/quail/Thanatpin/2283/2007(H5N1)	2007	73.76582	113.7522	63.76267		
	AB780494	AB780496	A/duck/Vietnam/OIE-2212/2012(H5N1)	2012	80.98051	118.7173	70.41153		
	KC261463	KC261473	A/duck/Eastern China/057/2007(H5N1)	2007	76.04691	116.5277	71.06516		
	FR687258	FR687259	A/chicken/Egypt/Q1182/2010(H5N1)	2010	69.47943	113.5217	76.67715	68.86228	5.266735

(Reproduced by permission from Current Comput Aided Drug Design, Ref 2)

Table 2. Coupling factors for forced matches between HA and NA for different strains and subtype combinations

Flu SubTypes	Virus strain description	Natural coupling factor*	Computed coupling factor for first strain		Notes
			HA2+NA1	HA1+NA2	
H1N1 H3N2	A/Mexico/InDRE4487/2009(H1N1) A/Aichi/2/1968(H3N2)	36.338769 27.378422	39.4250615	49.7303551	
H1N1 H5N1	A/New York/1669/2009(H1N1) A/chicken/Egypt/Q1182/2010(H5N1)	30.403674 76.677153	31.0935325	82.399158	1
H1N1 H5N1	A/California/05/2009(H1N1) A/Anhui/1/2005(H5N1)	38.429353 69.386759	34.7546642	73.6664097	1 3
H1N1 H5N1	A/South Dakota/01/2011(H1N1) A/duck/France/05066b/2005(H5N1)	33.972387 43.589874	12.0340407	69.723639	2
H1N1 H6N1	A/Netherlands/602/2009(H1N1) A/mallard/Jiangxi/227/2003(H6N1)	38.494277 30.043214	2.74553914	67.1634074	
H1N1 H7N9	(A/New York/3008/2009(H1N1)) A/Fujian/1/2013(H7N9)	36.446792 41.317053	18.7330372	40.9006368	
H3N2 H5N1	A/Tbilisi/GNCDC0557/2012(H3N2) A/duck/France/05066b/2005(H5N1)	28.257107 43.589874	8.37696138	62.7612626	2
H3N2 H6N1	A/Australia/NHRC0010/2005(H3N2) A/duck/Hunan/177/2005(H6N1)	27.383275 24.701758	11.0170769	59.7319878	
H3N2 H7N9	A/Singapore/NHRC0007/2003(H3N2) A/blue-winged teal/Guatemala/CIP049-02/2008(H7N9)	22.338436 64.782725	12.262058	34.451943	
H5N1 H6N1	A/goose/Guangdong/1/1996(H5N1) A/chicken/Taiwan/2838V/00(H6N1)	44.237442 31.885536	20.1662511	58.710981	2
H5N1 H7N9	A/swine/Fujian/F1/2001(H5N1) A/Shanghai/4664T/2013(H7N9)	56.029805 38.807458	43.2469843	30.6110415	2
H6N1 H7N9	A/duck/Hunan/177/2005(H6N1) A/Zhejiang/DTID-ZJU01/2013(H7N9)	24.701758 40.504412	44.7498079	61.0380129	

Notes: * Values as per Table 1
1 H5N1 with short stalk neuraminidase
2 H5N1 with long stalk neuraminidase
3 Strains used in Zhang et al's (2010) experiments

(Reproduced by permission from Current Comput Aided Drug Design, Ref 2)

Table 3. Hypothetical combinations from 2015 North American H5N2: Coupling a H5 with different NA subtypes.

Gene	Locus ID – typical examples	Typical subtype / averages for avian 'flus	Average Seq Length	Average Mu X	Average Mu Y	Graph radius g_R	Delta g_R for sample HA with other NAs	Diff from Average in Table 1 (in %)
Sample HA of H5N2	KP739389	A/domestic duck/Washington/61-16/2014(H5N2)	1704	-97.9695	-29.4883	102.31118		
NA N1 H1N1	EU743308	Average for H1N1 subtype	1410	-54.4152	-65.8155	85.39725	56.71546	47.02
NA N1 H6N1	HM144562	Average for H6N1 subtype	1410	-55.4681	-65.6722	85.962449	55.81799	44.69
NA N1 ls H5N1	KC815851	Average for H5N1 long stalk subtype	1410	-58.4532	-60.6701	84.247486	50.33731	30.48
NA N1 ss H5N1	KC261473	Average for H5N1 short stalk subtype	1350	-45.6463	-59.3848	74.900875	60.26209	56.21
NA N2 H3N2	AB277755	Average for N2 data from H1N2 and H3N2	1413	-63.7702	-47.3914	79.451709	38.60201	0.06
NA N3 H7N3	CY125730	A/Mexico/InDRE7218/2012(H7N3)	1410	-86.1333	-2.89858	86.182091	29.10508	24.55
NA N4 H4N4	AY207533	A/gray teal/Australia/2/79(H4N4)	1413	-56.1826	-30.2852	63.825366	41.79449	8.34
NA N5 H6N5	FLASHEAU	A/shearwater/Australia/1/1972(H6N5)	1407	-56.2971	-57.3163	80.340011	50.10975	29.89
NA N6 H4N6	JX454731	A/wild bird/Korea/GS26/2006(H4N6)	1413	-87.0035	2.248408	87.032586	33.57779	12.96
NA N7 H10N7	EU747332	A/quail/California/1022/1999(H10N7)	1347	-78.66	-15.3846	80.150346	23.91174	38.02
NA N8 H2N8	KC899750	A/wild duck/SH38-26/2010(H2N8)	1413	-42.2151	-48.0602	63.967938	58.76616	52.33
NA N9 H7N9	KC899669	Average for H7N9 subtype	1409	-108.696	3.491663	108.75254	34.6806	10.10

Note: ss - short stalk; ls - long stalk; Delta-g_R is the coupling factor

(Reproduced by permission from Current Comput Aided Drug Design, Ref 7)

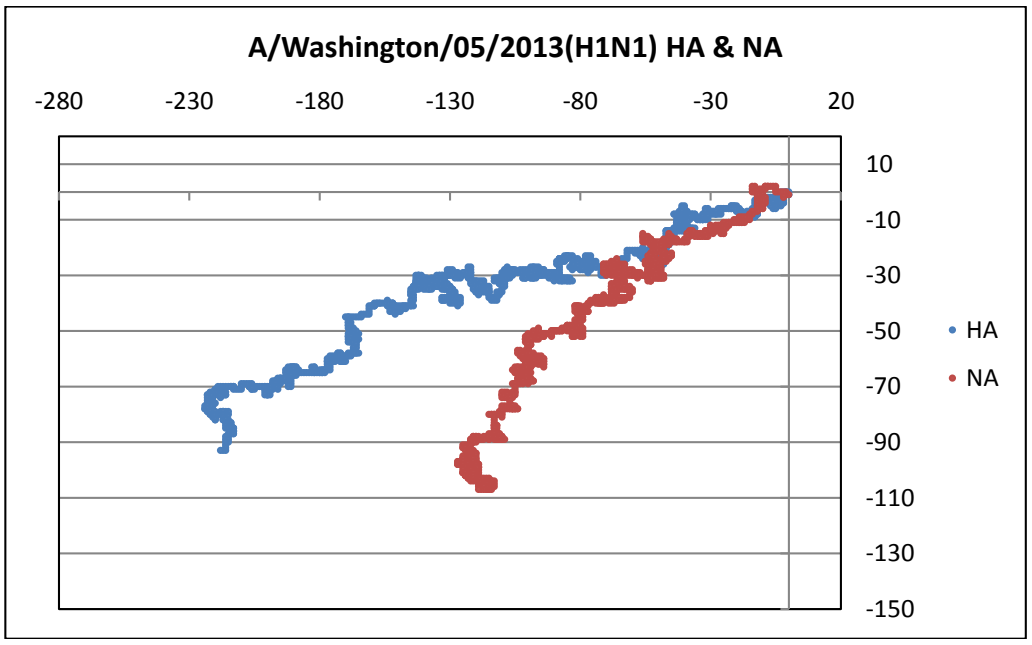

Figure 1. 2D graphical representation of HA (blue) and NA (red) sequences (cds) of A/Washington/05/2013(H1N1)

3. Materials and Methods

All sequences were downloaded from the NCBI GenBank database within the past year [15]. The IDs of the various sequences are given in the tables in Refs 2 and 7 and the details can be accessed from the NCBI website, Ref.11.

The 2D graphical representation method used here [16] is a simple device to visualize the base distribution of any DNA/RNA sequence. On a 2-dimensioanl Cartesian cor-ordinate system, the four cardinal directions are identified with the four bases which preferentially are: adenine with the –ve x-direction, cytosine with the positive y-direction, guanine with the +ve x-direction and thymine with the –ve y-direction. The query sequence is plotted starting from the origin and moving one step in the direction indicated for the base sequentially. This traces out a curve that reflects the distribution of bases along the sequence. Fig.1 is an example of two gene sequences, of the HA and NA, on the same graph.

Quantitative assessments of the different sequences, e.g., descriptors of the two sequences in Fig.1, can be made as a first approximation by defining weighted centre of mass (μ_x,μ_y) of a sequence as

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

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

where (x_i,y_i) represent the co=ordinates of the ith base, N is the total number of bases and we define the distance from the origin to the centre of mass as g_R. Then the difference between two sequences can be represented by the distance between the end points of the graph radii of the two sequences as

$$\Delta g_R = \sqrt{(\mu_x + \mu_{x'})^2 + (\mu_y + \mu_{y'})^2}$$

We use the Δg_R as an indicator of the coupling between the two sequences as explained earlier. More discussions on the properties of g_R and Δg_R can be found in the earlier papers and related documents.

.
.
.
.
.
.

4. Conclusions

In this brief report we have discussed the observation of interdependence of hemagglutinin and neuraminidase in influenza A subtypes that appear to restrict the proliferation of influenza subtypes to a few combinations, although theoretically a much larger number should be possible. We believe the origins of this phenomenon must lie in the base distribution and composition of the related sequences; observations of Hu [8] on HA, NA mutations show that mutations in one sequence appear to regulate mutational changes in the other. To quantify this phenomenon we have hypothesized a coupling of the two sequences with a coupling factor that is characteristic of the related subtypes. Our investigations into real sequences of several different subtypes using graphical representation techniques yielded specific numbers for each flu subtype within a reasonable tolerance level; forced replacements of one gene with another subtype led to different coupling strengths, which was more dramatic in the case of NA exchanges [2].

Since the influenza genome is known to undergo genetic drift to new reassortants quite frequently due to its inherent segmented structure, this interdependence of the HA and NA serves to restrict such reassortants to a reduced subset of possible stable pathogenic varieties. Our methodology described

above, and reported previously in Ref 2, allows us to compute possible such subtypes of the influenza. In response to the recent epidemic of H5N2 influenza among North American poultry, we have made a prognosis on this basis of possible new pathogenic reassortants that may arise out of the current epidemic [7]. This has important consequences on monitoring of influenza strains and mutations that allows opportunity to focus on possible more pathogenic subtypes. On a larger scale, our approach provides an opportunity worldwide to compute and monitor evolution of highly pathogenic influenza viruses.

.

Acknowledgments

The authors would like to acknowledge with thanks the opportunity provided by the organizers of the MOL2NET-1 conference to present our findings to a general audience.

Author Contributions

AN conceived the problem and hypothesis and wrote the paper; SB contributed several suggestions, reviewed the manuscript and made critical comments to improve the presentation.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1 Baigent, S.J.; McCauley, J.W. Glycosylation of haemagglutinin and stalk-length of neuraminidase combine to regulate the growth of avian influenza viruses in tissue culture. Virus Res., 2001, 79, 177-185.

2 Nandy, A; Sarkar, T; Basak, S C; Nandy, P; Das, S. Characteristics of Influenza HA-NA Interdependence Determined Through a Graphical Technique. Curr Comput Aided Drug Design 2014, 10 (4), 285-302.

3 Zhang, Y.; Lin, X.; Wang, G.; Zhou, J.; Lu, J.; Zhao, H.; Zhang, F.; Wu, J.; Xu, C.; Du, N.; Li, Z.; Zhang, Y.; Wang, X.; Bi, S.; Shu, Y.; Zhou, H.; Tan, W.; Wu, X.; Chen, Z.; Wang, Y. Neuraminidase and hemagglutinin matching patterns of a highly pathogenic avain and two pandemic H1N1 influenza A Virus, PLoS One, 2010, 5, e9167

4 Xu, R.; Zhu, X.; McBride, R.; Nycholat, C.M.; Yu, W.; Paulson, J.C.; Wilson, I.A. Functional balance of the hemagglutinin and neuraminidase activities accompanies the emergence of the 2009 H1N1 influenza pandemic. J. Virol., 2012, 86, 9221-9232.

5 Schnitzler, S.U.; Schnitzler, P. An update on swine-origin influenza virus A/H1N1: a review. Virus Genes, 2009, 39, 279-292.

6 CDC Report: Avian influenza A (H7N9) virus. http://www.cdc.gov/flu/avianflu/h7n9-virus.htm. (Accessed April 30, 2014).

7 Nandy, A; Basak, S C. Prognosis of possible reassortments in recent H5N2 epidemic influenza in USA: Implications for computer-assisted surveillance as well as drug/vaccine design. Curr Comput Aided Drug Design 2015, 11(2), 110-116.

8 Wei, Hu. The interaction between the 2009 H1N1 influenza a hemagglutinin and neuraminidase: mutations, co-mutations and the NA stalk motif. J. Biomed. Sc. Engg., 2010, 3, 1-12.
9
http://www.aphis.usda.gov/wps/portal/aphis/newsroom/news/sa_stakeholder_announcements/sa_by_date/sa_2015/sa_04/ct_hpai_sd_m n_wi (Accessed 12 May 2015)
10 Huffstutter, P.J.; Steenhuysen, J. (2015). Increased human protections offered as H5N2 outbreak spreads. Reuters 27 April 2015. http://www.reuters.com/article/2015/04/27/us-health-birdflubiosecurity-insight-idUSKBN0NI0AU20150427 (Accessed 26 May 2015)
11 http://www.aphis.usda.gov/stakeholders/downloads/2015/saees_hpai_minnesota.pdf (Accessed 12 May 2015)
12 Hall, J.S.; Dusek, R.J.; Spackman, E. Rapidly expanding range of highly pathogenic avian influenza viruses. Emerg Infect Dis. 2015 Jul. http://dx.doi.org/10.3201/eid2107.150403 [Accessed 22 May 2015]
13
http://www.who.int/influenza/human_animal_interface/EN_GIP_201503031cumulativeNumberH5N1cases.pdf?ua=1 (Accessed 16 May 2015)
14 World Health Organization. "Avian Influenza: Assessing the Pandemic Threat." January, 2005. WHO/CDS/2005.29
http://apps.who.int/iris/bitstream/10665/68985/1/WHO_CDS_2005.29_eng.pdf?ua=1
(Accessed 22 May 2015)
15 http://www.ncbi.nlm.nih.gov/
16 Nandy, A. A new graphical representation and analysis of DNA sequence structure: I. Methodology and application to Globin genes. Curr. Sci., 1994, 66, 309-314.