



Use of Shannon entropy as a tool in evaluating dynamics of Influenza A virus evolution

Igor Balaz^{a*}

^a University of Novi Sad, Faculty of Agriculture, Department of Field and Vegetable Crops, Novi Sad, Serbia

* Corresponding author: ibalaz@polj.uns.ac.rs

ABSTRACT

Influenza A viruses (IAV) are RNA viruses that can infect a range of animal species. Natural reservoirs for a large variety of IAV are wild aquatic birds. From them, IAV can transmit to other species causing costly poultry or swine outbreaks. IAV are highly evolvable due to the continuous interplay of reassortment and genetic drift. Their unpredictable evolution is one of the main reasons why it is very hard to efficiently predict occurrence of the next outbreak. Additional problem is the lack of simple and straightforward approach that could be used to analyze available high-dimensional genomic and epidemiological data. In this paper we performed a large scale analysis of Shannon entropy of avian and swine Influenza A (H1N1, H2N5 and H3N2) nucleotide sequences in order to (i) analyze evolutionary Influenza dynamics and (ii) test to what extent data dimensionality could be reduced while keeping intact the information on the temporal dynamics. We show that numerical characterization of each RNA sequence opened a possibility for straightforward analysis of their evolution. We also demonstrate that correlated clustering in investigated Influenza viruses emerge at the genetic level and sometimes lasts for several years.

KEY WORDS

Influenza, Shannon entropy, evolution.

Introduction

Influenza A virus is a negative-sense segmented RNA virus, with a genome consisting of eight gene segments, that can encode up to 16 proteins (Schrauwen et al., 2014). The subtype of influenza A virus is determined by the antigenicity of two surface glycoproteins, hemagglutinin (HA) and neuraminidase (NA). In nature, aquatic birds host a vast reservoir of Influenza A viruses, while all of the current mammalian influenza A viruses are probably derived from that reservoir (Webster et al. 1992).

Avian Influenza A viruses can be divided into two groups (Alexander, 2007). The first one are the very virulent viruses causing highly pathogenic avian influenza, with flock mortality as high as 100%. These viruses have been restricted to subtypes H5 and H7. All other viruses cause a milder, primarily respiratory, disease. Influenza viruses may infect all types of domestic or captive birds in all areas of the world. Until recently high pathogenic viruses were rarely isolated from wild birds, but low pathogenic viruses are commonly found, with overall figures of about 11% for ducks and geese and around 2% for all other species (Alexander, 2007).

Compared with the cases reported for the past 40 years, the number of outbreaks of Influenza A in poultry has increased sharply since the early 2000s (Capua and Marangon 2006, FAO, 2012). The number of birds involved in Avian Influenza (AI) outbreaks has increased 100-fold, from 23 million from 1959 through 1998 to >200 million from 1999 through 2005 (Capua and Alexander, 2004). As Capua and Alexander (2004) reported, the reasons for this are not clear. One possibility, albeit unlikely, could be the increase in awareness and better diagnostic tools. Another factor could be the climate change and consequent variations in wild bird migratory patterns or populations. There have also been relevant changes in poultry production in recent years with an increase in densely populated poultry areas coupled with a move towards rearing on open range. Whatever the reason is, increase of influenza outbreaks on poultry farms causes a great damage. For example, in 2003 in The Netherlands, Belgium and Germany, high pathogenic avian influenza resulted in culling of over 30 million birds. In that particular case, the reason was the extremely high density and close proximity of poultry farms, which resulted in difficulties in containment (Capua and Alexander, 2004). Although in most countries farmers are compensated for culled poultry, the compensation was often far below market price (Otte et al., 2008, McLeod, 2010). In addition to high economic cost, outbreaks at poultry farms can also pose a direct threat to humans. In 1997, a large outbreak of highly pathogenic avian influenza (HPAI) H5N1 virus in poultry in Hong Kong resulted in the first documented cases of the direct transmission of HPAI H5N1 virus from poultry to humans, with a fatal outcome in 6 out of 18

cases (de Jong et al., 1997).

Economic impact is also significant in pig farms where for example, influenza A H1N2 infections lead to 2.9% more wean-to-finish mortality, as compared to a similar, uninfected population (Donovan, 2005). In addition, influenza infection increases medication usage in order to control secondary bacterial infections (Loving, et al., 2010). Finally, there is an increased risk for public health since transmission of influenza viruses has been detected in both ways: from pigs to human and vice versa (Forgie, et al., 2011; Lantos, et al., 2016; Sikkema, et al., 2016)

Given the recent increase in outbreaks, it could be of great importance to analyze evolutionary trends of influenza viruses in order to better understand, emergence of outbreaks. Therefore we performed a large scale analysis of Shannon entropy of all publicly available Influenza A nucleotide sequences (subtypes H1N1, H5N2 and H3N2) from avian and swine hosts that are stored at the Influenza Research Database. There are two main reasons why using Shannon entropy in this analysis. First, since the Shannon entropy measures the effective number of symbols in a given sequence, it is one of the most basic quantities to characterize RNA sequences. Second, because in each year there are up to tens of thousands available RNA sequences, our goal is to numerically characterize them so we can remove high-dimensionality problems and allow more easily comparison between them.

Material and methods

Influenza RNA segments

Using the Influenza Research Database (<https://www.fludb.org/>) we obtained all available RNA segments of Influenza A viruses (subtypes H1N1 and H3N2) isolated from avian and swine host species, available up to January 2017. From them, we removed all incomplete and duplicate segments, while keeping only one copy if multiple nucleotide segments are exactly identical. Since the entire Influenza genome contains eight RNA segments, when referring to them, in the text we used following abbreviation: Segment 1: PB2; Segment 2: PB1; Segment 3: PA; Segment 4: HA; Segment 5: NP; Segment 6: NA; Segment 7: M; and Segment 8: NS. To form continuous time series, we analysed only the longest time interval where for each year at least five unique and complete segments are available. As a result, for avian H1N1 viruses our analysis cover interval 2005-2016, for H5N1 2000-2017, while for H3N2 covered years are 2004 – 2014. For swine hosts intervals are 1998 – 2016 (H1N1) and 2002 – 2016 (H3N2).

Evolutionary correlation of RNA segment entropies

For each year in the intervals defined above we calculated mean and median of all entropy values for each of eight groups of RNA segments. Entropy is calculated as Shannon entropy $H = -\sum_i p_i \log_2 p_i$

where p_i is the probability of nucleic acid i to appear in the nucleic acid segment. To determine level of correlation of evolution among RNA segments we defined 10-year sliding-time window. Within that window, for each pair of eight available segments (28 pairs in total), first we calculated yearly mean and median values of entropies. In that way, for each pair of segments we obtained two datasets: $\{x_1, x_2, \dots, x_{10}\}$ and $\{y_1, y_2, \dots, y_{10}\}$ for which we then calculated correlation coefficient as Pearson Product-Moment Correlation:

$$r = \frac{\sum_{i=1}^{10} (x - \bar{x})(y - \bar{y})}{\sqrt{\sum_{i=1}^{10} (x - \bar{x})^2} \sqrt{\sum_{i=1}^{10} (y - \bar{y})^2}} \quad (1)$$

where x and y are median entropy values for corresponding years. Statistical significance was calculated as p value. Further significance tests using a permutation test and a multiple testing correction will be performed in future work.

Distribution of entropies

For each year when number of available RNA segments was larger than 50, we analyzed distribution of obtained entropy values, by calculating the adjusted Fisher-Pearson coefficient of skewness as

$$G = \frac{\sqrt{N(N-1)}}{N-2} \frac{\sum_{i=1}^N (Y_i - \bar{Y})^3 / N}{s^3} \quad (2)$$

and kurtosis as

$$k = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^4 / N}{s^4}, \quad (3)$$

where \bar{Y} is the entropy mean, s is the standard deviation of entropies in the given year and N is the number of available sequences.

Principal components

For each 10-year sliding time window we derived eigenvalues of principal components (PC) based on their correlation coefficients r . Since there are eight RNA segments, a maximum of eight PCs are formed for each time window using the following standard procedure: (i) the first PC is derived so that it accounts for the greatest possible variance in the correlation matrix. Using the elements of the data set x that have maximum variance, we created a linear function $\alpha_1^T x$ where α_1 is a vector of p constants $\alpha_{11}x_1, \alpha_{12}x_2, \dots, \alpha_{1p}x_p$ and T denotes matrix transpose, so that

$$\alpha_1^T x = \alpha_{11}x_1 + \alpha_{12}x_2 + \dots + \alpha_{1p}x_p = \sum_{j=1}^p \alpha_{1j}x_j; \quad (4)$$

(ii) Each subsequent k th principal component is derived in the same way by finding linear function $\alpha_k^T x$ such that it has maximum variance and is uncorrelated with all previously defined linear functions $\alpha_1^T x, \alpha_2^T x, \dots, \alpha_{k-1}^T x$ (Jolliffe, 2002). Finally, the eigenvalues for each PC are calculated as the variance of that PC and then normalized to represent the percentage of total variation accounted for by each PC. As an indicator of uniformity of variations during a time interval we calculated the sum of the first k principal components as

$$t_m^k = \frac{100}{p} \sum_{k=1}^k l_k, \quad (5)$$

where l_k is the variance of the k th PC and p is the total number of PCs.

Results and discussion

As expected, temporal evolution of correlation among sequences from different hosts shows no significant correlation in time development of sequence's entropy (data not shown). However within the same host, different subtypes show unique evolution of correlation patterns. In swine H1N1 (Fig. 1), notable is emergence of clearly separated clusters of sequences with strongly positive or negative correlation during the decade 2000-2009. Then, two clusters of statistically significant ($p < 0,05$) very high positive correlations among members appeared: HA-NP-PB1 ($r_{\min} = 0,8034$) and NA-NS-PA-PB2 ($r_{\min} = 0,6405$), while the segment M remains both weakly correlated and not statistically significant (Fig. 2). At the same time members of opposite clusters have highly negative correlation with each other. This period of high correlations overlaps with the emergence of swine influenza lineage that in 2009 transmitted to humans leading to pandemic (Smith et al., 2009). Principal component analysis also shows unusually high evolutionary correlations of RNA segments during the 2000-2009 decade (Fig. 3). Partition of the total variation demonstrate that the sum of the first two principal components accounts for the 88,97% of total variation. For later years it rapidly drops down to the level of approximately 60-65%, where it remains for the rest of the available period.

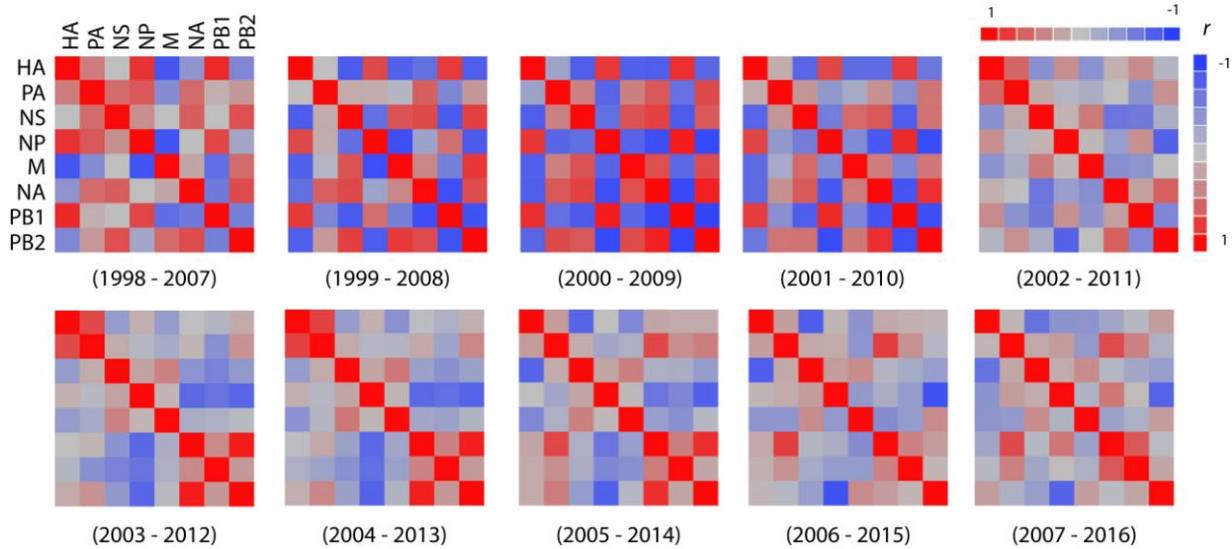


Figure 1. Cell plots of the correlations r between median H values for all sequences of Swine H1N1, on a scale from red (+1) to blue (-1). Correlations are calculated for 10 years periods which are indicated below each cell plot. Order of RNA segments is the same for all cell plots and is indicated in the upper left corner

Slika 1. Prikaz korelacije r između median vrednosti izračunatih entropija svih sekvenci svinjskog virusa gripa H1N1, na skali od crvene (+1) do plave (-1). Korelacije su računane za periode od 10 godina što je naglašeno ispod svakog pod-grafikona. Red RNA sekvenci (gornji levi ugao) je isti za svaki vremenski interval

On the other hand, swine H3N2 show different, but no less curious pattern. Although there were no emergence of genome-wide correlations as in H1N1, and the correlations among RNA segments are much weaker, there is a set of pairs (NS-HA, NS-NA and NS-PA) that remains highly negatively correlated ($p < 0,05$) on each 10-year window during the 2002-2013 period. Only in the last 3 available years (2014-2016) correlations among NS-NA and NS-PA waned, but the NS-HA remains highly negatively correlated, with r values always around -0,8. Why that is the case is not clear. One possibility could be the parallel host-specific evolution of virus proteins. Three out of four of these segments encode for proteins that are important determinants of virulence and play a crucial part in invasion of hosts and spread through their organism. Hemagglutinin (HA) covers most of the virus surface and initiates entry into the target cells (Gamblin and Skehel, 2010).

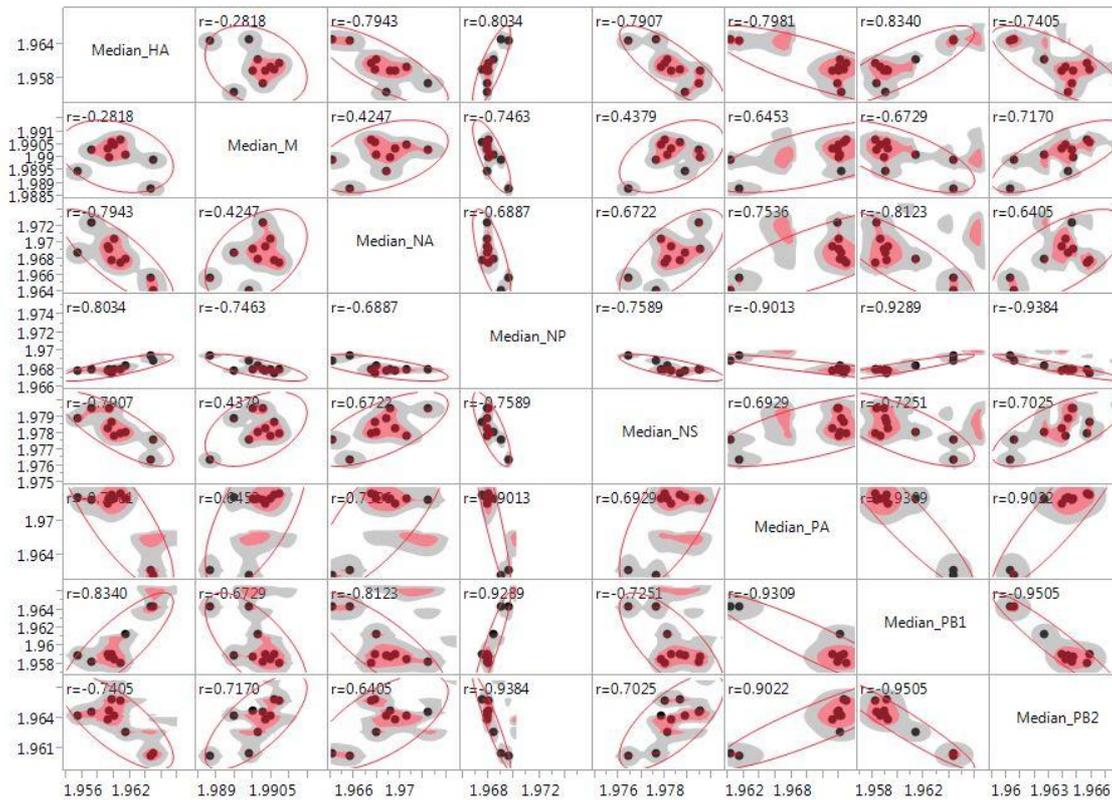


Figure 2. Matrix representation of correlations r between median H values (indicated at both axes) for each pair of segments of swine Influenza H1N1 subtype, for the 10-year period 2000-2009. Red ellipses marks 95% bivariate normal density ellipse. Red and grey shaded areas indicate the density of data points (black dots): grey contour includes 90% of the smoothed density while the red ones include 50%

Slika 2. Matrični prikaz korelacije r između median vrednosti izračunatih entropija svih sekvenci (čiji su nazivi dati u dijagonali matrice) svinjskog gripa H1N1, za desetogodišnji period od 2000-2009. godine. Crvene elipse označavaju 95% bivarijantnu gustinu korelacija. Crvene i sive zone označavaju gustinu datih podataka (koji su označeni crnim tačkama): siva zona uključuje 90% podataka, dok crvena zona uključuje 50%

NA is essential for release and spread of virions from the host cells (Shtyrya, 2009). NS1 protein is important in suppressing the virus-induced host immune response (Wolff and Ludwig, 2009; Hale et al., 2010). Therefore, it is possible that changes in one of these proteins could lead to necessary complementary changes in other ones in order to keep high fitness. Also, it has been shown that functions of the HA and NA proteins are tightly balanced to each other because they both play an important role in acquisition of efficient and sustained host-to-host transmissibility (Wagner et al., 2002, Yen et al., 2011).

For avian Influenza the range of available data for both H3N2 and H1N1 is much shorter so we cannot draw any conclusion on evolutionary trends. The only exception is Influenza strain H5N1 which is enzootic in many bird populations and is much better monitored because it can pose a significant threat for both avian and human populations (Li et al., 2004). For the whole duration of uninterrupted time series of 17 years all RNA segments show positive correlations, either high or moderate (Table 1). Also, during the entire period, the sum of the first two principal components accounts for the 85-95% of total variation, for each 10-year sliding window. It is interesting to note that avian H5N1 is a highly pathogenic virus that caused several severe disease outbreaks in poultry. Due to the enzootic nature of H5N1 in poultry, the brood host range (over 20 different mammalian species) and the accumulation of mammalian adaptation mutations, this virus is currently considered to represent a significant pandemic threat (Schrauwen et al., 2014) In a similar manner, appearance of highly pathogenic swine H1N1 influenza strain corresponds with surge in high correlations among RNA segments. However, to draw any conclusions whether high correlation of RNA segments is in any way connected with higher virulence, additional epidemiological and evolutionary analysis should be performed, which is out of scope of this paper.

Table 1

Correlation coefficients of median values of calculated Shannon entropy for all segments of avian H5N1 Influenza virus, for the period 2002-2016. Values higher than 0,7 are in blue color

Tabela 1

Koeficijenti korelacije između medijan vrednosti Šenonovih entropija za sve segmente ptičjeg virusa gripa H5N1, za period 2002-2016. godina. Vrednosti koje su veće od 0,7 su naglašene plavom bojom

	HA	M	NA	NP	NS	PA	PB1	PB2
HA	1.0000	0.9656	0.6494	0.4840	0.7919	0.7666	0.8460	0.8405
M	0.9656	1.0000	0.7304	0.5538	0.8565	0.8131	0.8920	0.8383
NA	0.6494	0.7304	1.0000	0.7935	0.8186	0.6998	0.8231	0.7190
NP	0.4840	0.5538	0.7935	1.0000	0.5551	0.3921	0.5641	0.6606
NS	0.7919	0.8565	0.8186	0.5551	1.0000	0.9455	0.9307	0.7853
PA	0.7666	0.8131	0.6998	0.3921	0.9455	1.0000	0.9116	0.7111
PB1	0.8460	0.8920	0.8231	0.5641	0.9307	0.9116	1.0000	0.8564
PB2	0.8405	0.8383	0.7190	0.6606	0.7853	0.7111	0.8564	1.0000

Distribution analysis of obtained Shannon entropy values for each year when number of samples was high enough show that evolution of most of the sequences share the similar characteristics. For all subtypes that we investigated, positive skewness and kurtosis values are usually highly correlated (Fig. 4 and Fig. 5). Compared to swine influenza, avian strains are slightly more compact with lower average kurtosis. It could indicate that swine influenza strains are under higher selective pressure which leads to evolutionary favoring of outliers. If that is the case, then our results are in agreement with proposal that avian influenza viruses have reached an “evolutionary stasis” characterized by low rates of evolutionary change, particularly at amino acid-changing level (Webster et al. 1992).

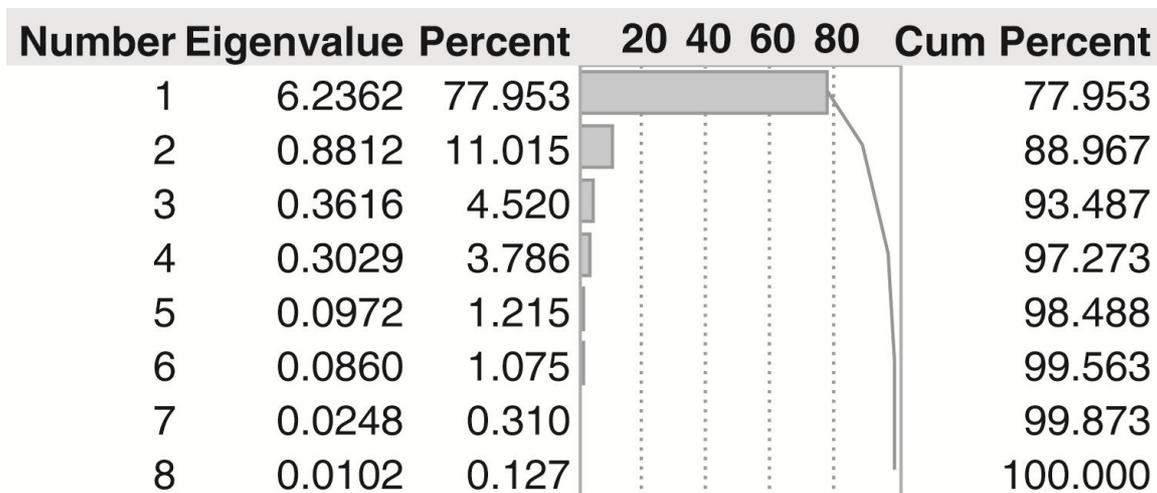


Figure 3. Eigenvalues Pareto plot derived from median H values for swine Influenza H1N1 subtype, for years 2000-2009. All eight eigenvalues that corresponds to each PC, from largest to smallest are listed in the second column. The percent, cumulative percent and horizontal bar visualization of the variation accounted for by each PC are given in the remaining columns. The eigenvalues represent a partition of the total variation in the analyzed sample

Slika 3. Prikaz ajgen vrednosti izvedenih iz median vrednosti entropija za svinjski grip H1N1 za period 2000-2009. Svih osam ajgen vrednosti koje korespondiraju svakoj osnovnoj komponenti (PC), dati su u drugoj koloni, poređani od najvećeg ka najmanjem. Procenat, kumulativni procenat i histogram su dati u narednim kolonama. Ajgen vrednosti ovde predstavljaju udeo u ukupnoj varijaciji analiziranog uzorka

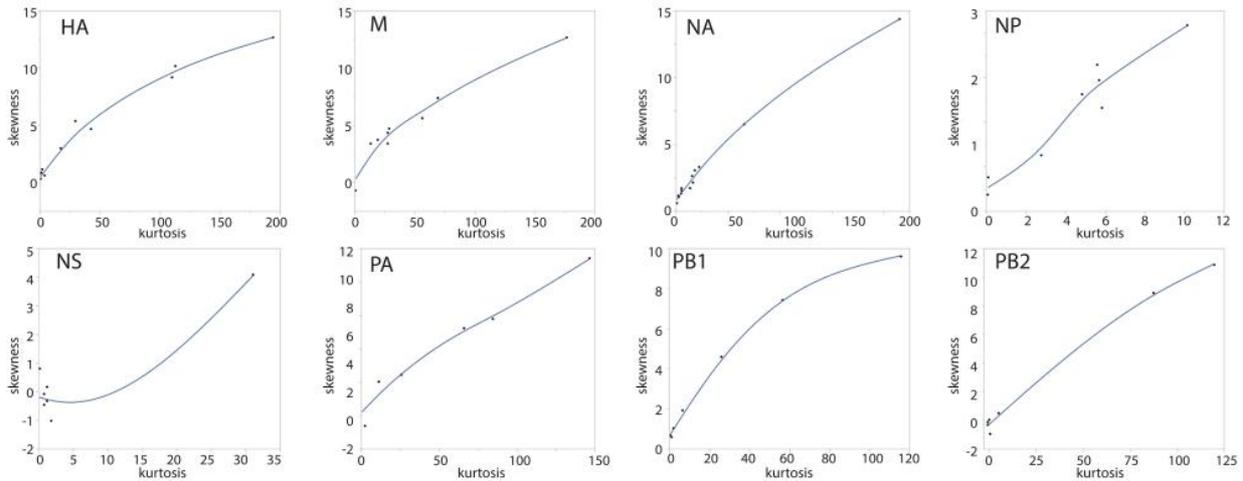


Figure 4. Skewness vs. kurtosis for all segments of swine Influenza H1N1 subtype. Values are calculated for each year when number of samples was higher than 50. Segment HA, NA show data for the period 2003-2016, Segment M, NP and NS period 2008-2016, while Segments PA, PB1 and PB2 are for the period 2009-2016

Slika 4. Odnos koeficijenta asimetričnosti i kurtozisa za sve segmente svinjskog gripa H1N1. Vrednosti su računane za svaku godinu kada je broj dostupnih uzoraka bio veći od 50. Segmenti HA i NA su za period 2003-2016; segmenti M, NP i NS za period 2008-2016; dok su segmenti PA, PB1 i PB2 za period 2009-2016

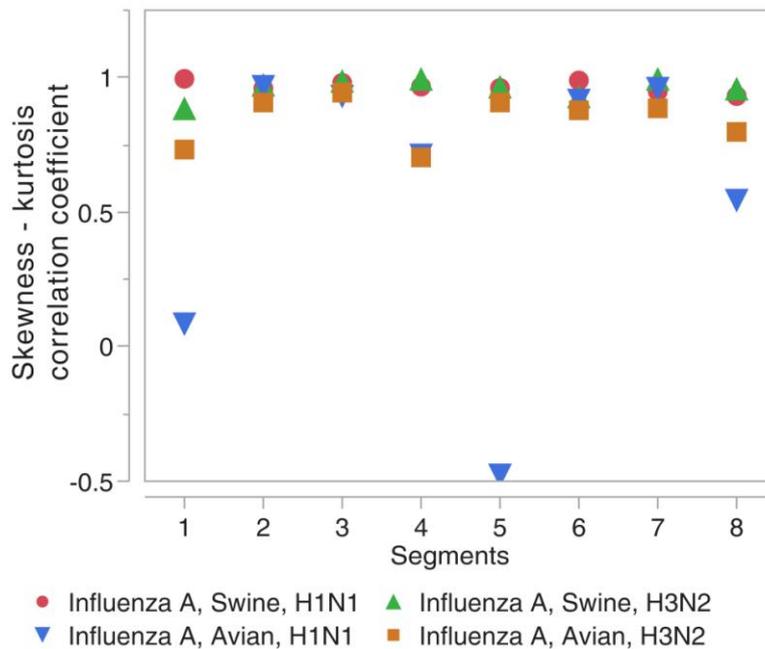


Figure 5. Correlation between skewness and kurtosis for each RNA segment, for avian and swine Influenza A H1N1 and H3N2. Analyzed time span for each segment is given in methods

Slika 5. Korelacija između koeficijenta asimetričnosti i kurtozisa za svaki RNA segment svinjskog i ptičjeg virusa gripa, H1N1 i H3N2. Vremenski interval za svaki segment je dat u sekciji Materijal i metode

Conclusions

The first goal of this paper was to demonstrate whether analysis of entropy of RNA sequences could add a new perspective to the evolution of Influenza viruses. Numerical characterization of each RNA sequence opened a possibility for straightforward analysis of their evolution. Our results, where we demonstrated correspondence of emergence of genome-wide correlations with occurrence of more

pathogenic strains, raises a lot of important questions like: To what extent all Influenza virus genes are functionally integrated? If they are integrated, and our data clearly point out in that direction, how that affects evolution of influenza viruses and the dynamics of their epidemics? What sets of environmental pressures favors such integration? Whether the methodology applied in this paper or some similar genome-wide approach will provide answers to these questions, remains to be seen.

Finally, we demonstrate that correlated clustering in investigated Influenza viruses emerge at the genetic level and sometimes lasts for several years. It is at odds with finding that, as a result of frequent mutations and reassortment, Influenza viruses have very complicated phylogenetic patterns (Holmes et al., 2005, Nelson et al., 2007). Whether it can be explained by parallel host-specific evolution or some other mechanism will be a topic for further investigation.

Acknowledgments

This work is supported by the Ministry of Education and Science of the Republic of Serbia (Grant III 43007: Studying climate change and its influence on the environment: impacts, adaptation and mitigation).

References

- Alexander, D.J. 2007. An overview of the epidemiology of avian influenza. *Vaccine* 25: 5637-5644.
- Capua, I., Alexander, D.J. 2004. Avian influenza: recent developments. *Avian Pathol.* 33:393-404.
- Capua, I., Marangon, S. 2006. Control of Avian Influenza in Poultry. *Emerg Infect Dis.* 12: 1319–1324.
- de Jong, J.C., Claas, E.C., Osterhaus, A.D., Webster, R.G., Lim, W.L. 1997. A pandemic warning? *Nature* 389:554.
- Donovan T. 2005. The role of influenza on growing pig performance. Proc. A.D. Lemna Swine Conference conference, 97-98, August, USA
- FAO. 2012. H5N1 HPAI Global Overview, January–March 2012. Food and Agriculture Organization, Rome, Italy. <http://www.fao.org/docrep/015/an388e/an388e.pdf>
- Forgie, S.E., Keenliside, J., Wilkinson, C., Webby, R., Lu, P., Sorensen, O., Fonseca, K., Barman, S., Rubrum, A., Stigger, E., Marrie, T.J., Marshall, F., Spady, D.W., Hu, J., Loeb, M., Russell, M.L., Babiuk, L.A. 2011. Swine outbreak of pandemic influenza A virus on a Canadian research farm supports human-to-swine transmission. *Clin Infect Dis.* 52(1):10-8.
- Gamblin, S.J., Skehel, J.J. 2010. Influenza Hemagglutinin and Neuraminidase Membrane Glycoproteins. *J Biol Chem.* 285: 28403–28409.
- Hale, B.G., Albrecht, R.A., Garcia-Sastre, A. 2010. Innate immune evasion strategies of influenza viruses. *Future Microbiol.* 5: 23–41.
- Holmes, E.C., Ghedin, E., Miller, N., Taylor, J., Bao, Y., St George, K. 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.
- Jolliffe, I. T. 2002. *Principal Component Analysis*, 2nd ed. Springer-Verlag, New York.
- Lantos, P.M., Hoffman, K., Höhle, M., Anderson, B., Gray, G.C. 2016. Are People Living Near Modern Swine Production Facilities at Increased Risk of Influenza Virus Infection? *Clin Infect Dis.* 63(12): 1558-1563.
- Li, K.S., Wang, G.J., Smith G.J.D., Xu, K.M., Duan, L., Rahardjo, A.P., Puthavathana, P., Buranathai, C., Nguyen, T.D., Estoepongstie, A.T.S., Chaisingh, A., Auewarakul, P., Long, H.T., Hanh, N.T.H., Webby, R.J., Poon, L.L.M., Chen, H., Shortridge, K.F., Yuen, K.Y., Webster, R.G., Peiris, J.S.M. 2004. Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature* 430: 209–213.
- Loving, C.L., Brockmeier, S., Vincent, A.L. 2010. Prior Swine Influenza Virus Infection Enhances Pulmonary Responses to Secondary *Haemophilus parasuis* Infection. In: *Proceedings of the Pig Veterinary Society International Congress*, July 18-21, 2010, Vancouver, Canada. p. 264.
- McLeod, A. 2010. Economics of avian influenza management and control in a world with competing agendas. *Avian Dis.* 54:374–379.
- Nelson, M.I., Simonsen, L., Viboud, C., Miller, M.A., Holmes, E.C. 2007. Phylogenetic analysis reveals the global migration of seasonal influenza A viruses. *PLoS Pathog.* 3: 1220–1228.
- Otte, J., Hinrichs, J., Rushton, J., Roland-Holst, D., Zilberman, D. 2008. Impacts of avian influenza virus on animal production in developing countries. *CAB Rev. Perspect. Agric. Vet. Sci. Nutrition Nat. Res.* 3:080.
- Schrauwen, E.J., de Graaf, M., Herfst, S., Rimmelzwaan, G.F., Osterhaus, A.D., Fouchier, R.A. 2014. Determinants of virulence of influenza A virus. *Eur J Clin Microbiol Infect Dis.* 33:479-90.
- Shtyrya, Y.A., Mochalova, L.V., Bovin, N.V. 2009. Influenza Virus Neuraminidase: Structure and Function. *Acta Naturae.* 1: 26–32.
- Sikkema, R.S., Freidl, G.S., de Bruin, E., Koopmans, M. 2016. Weighing serological evidence of human exposure to animal influenza viruses - a literature review. *Euro Surveill.* 21(44): 30388.
- Smith, G.J., Vijaykrishna, D., Bahl, J., Lycett, S.J., Worobey, M., Pybus, O.G. et al. 2009. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* 459: 1122-5.

- Wagner, R., Matrosovich, M., Klenk, H.D. 2002. Functional balance between haemagglutinin and neuraminidase in influenza virus infections. *Rev Med Virol.* 12: 159–166.
- Webster, R.G., Bean, W.J., Gorman, O.T., Chambers, T.M., Kawaoka, Y. 1992. Evolution and ecology of influenza A viruses. *Microbiol. Mol. Biol. Rev.* 56: 152-179.
- Wolff, T., Ludwig, S. 2009. Influenza viruses control the vertebrate type I interferon system: factors, mechanisms, and consequences. *J Interferon Cytokine Res.* 29: 549–557.
- Yen, H.L., Liang, C.H., Wu, C.Y., Forrest, H.L., Ferguson, A., Choy, K.T. et al. 2011. Hemagglutinin-neuraminidase balance confers respiratory-droplet transmissibility of the pandemic H1N1 influenza virus in ferrets. *Proc Natl Acad Sci U S A* 108: 14264–14269.

Upotreba Šenonove entropije za analizu evolutivne dinamike virusa gripa, tip A

Igor Balaž^{a*}

^a Univerzitet u Novom Sadu, Poljoprivredni fakultet, Departman za ratarstvo i povrtarstvo, Novi Sad, Srbija

* Autor za kontakt: ibalaz@polj.uns.ac.rs

Sažetak

Virus gripe pripada grupi RNA virusa i može da inficira širok spektar različitih vrsta životinja. Ptice vodenih staništa čine njihov prirodan rezervoar, kao i izvor velike varijabilnosti virusa. Kao posledica toga, različiti tipovi virusa gripa mogu da pređu na druge vrste i tako izazovu ekonomski veoma značajne epidemije kod domaće živine i svinja. Usled stalnog genetskog drifta i reasortimana RNA segmenata, stopa evolucije virusa gripa je veoma visoka i generalno je nemoguće predvideti u gde i u kom obliku će se pojaviti sledeća epidemija. Dodatni problem je i nepostojanje jednostavnog i brzog metoda za analizu dostupnih visoko-dimenzionalnih genetskih i epidemioloških podataka. U ovom radu je primenjena analiza Šenonove entropije nukleotidnih sekvenci ptičjeg i svinjskog gripa (tipovi H1N1, H2N5 i H3N2) da bi se: (i) analizirala evolutivna dinamika virusa gripa, (ii) testiralo do koje mere je moguće smanjiti dimenzionalnost dostupnih podataka a da se istovremeno zadrži njihov informacioni sadržaj vezan za vremensku dinamiku evolucije virusa. Pokazano je da ovakva numerička karakterizacija RNA sekvenci otvara mogućnost za jasnu i brzu analizu njihove evolucije. Takođe je pokazano da postoji periodično pojavljivanje klastera korelacija između različitih sekvenci koji u nekim slučajevima traju i do nekoliko godina.

KLJUČNE REČI

Grip, Šenonova entropija, evolucija

Primljen: 13.11.2017.

Prihvaćen: 14.12.2017.