Mapping of the influenza A hemagglutinin serotypes evolution by the ISSCOR method

Jan P. Radomski\textsuperscript{1,4*}, Piotr P. Słonimski\textsuperscript{2*}, Włodzimierz Zagórski-Ostoja\textsuperscript{3} and Piotr Borowicz\textsuperscript{4}

\textsuperscript{1}Interdisciplinary Center for Mathematical and Computational Modeling, Warsaw University, Warsaw, Poland; \textsuperscript{2}Centre de Génétique Moléculaire du CNRS & Université Pierre-et-Marie Curne, Paris, Gif–sur–Yvette, France; \textsuperscript{3}Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland; \textsuperscript{4}Institute of Biotechnology and Antibiotics, Warsaw, Poland

Analyses and visualizations by the ISSCOR method of influenza virus hemagglutinin genes of different A-subtypes revealed some rather striking temporal relationships between groups of individual gene subsets. Based on these findings we consider application of the ISSCOR-PCA method for analyses of large sets of homologous genes to be a worthwhile addition to a toolbox of genomics — allowing for a rapid diagnostics of trends, and ultimately even aiding an early warning of newly emerging epidemiological threats.

Key words: ISSCOR descriptors; phylogenetic analysis; influenza virus; hemagglutinin; phylogenetic maps

Received: 28 May, 2014; revised: 26 August, 2014; accepted: 11 September, 2014; available on-line: 12 September, 2014

\*Piotr P. Słonimski, passed away on April 25th, 2009, too early, leaving behind a large portfolio of many joint genomic projects and ideas at various stages of development, some of which need long time to be fulfilled. Specifically, in the case of large assemblies of the HA orthologs, which we have started to analyze already back in 2006–2007, the sufficiently detailed data are available only recently, although the first basic results showing the peculiar triangular distributions presented here, were obtained in the 2008.
Mapping of the Influenza A Hemagglutinin Serotypes Evolution by the ISSCOR Method

Jan P. Radomski 1,4*, Piotr P. Slonimski 2, Włodzimierz Zagórski-Ostoja 3, and Piotr Borowicz 4

1 Interdisciplinary Center for Mathematical and Computational Modeling, Warsaw University, Pawińskiego 5A, Bldg. D, PL–02106 Warsaw, Poland; 2 Centre de Génétique Moléculaire du CNRS & Université Pierre-et-Marie Curie (Paris-6), 91190 Gif-sur-Yvette, France; 3 Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawińskiego 5A, Bldg. D, PL–02106 Warsaw, Poland 4 Institute of Biotechnology and Antibiotics, Starościńska 5, PL–02516 Warsaw, Poland

Supplementary on-line materials

Figure S1 – The NJ+ phylogram of the 2009 swine-flu pandemic H1N1 hemagglutinin early 178 sequences, which were isolated in the 2009 before April 30th.

The 2009 pandemic H1N1 hemagglutinin early 178 sequences (light rose, and the enlarged inset on the left). Calculated together with 380 putative precursor sequences of other serotypes (collected during the same period): seasonal H1N1 (yellow – human hosts, and light green – porcine hosts), H3N2 (magenta – human), H5N1 (blue – avian; and light blue – human), H1N2 (olive – porcine). The zoomable Fig_S1_large.pdf file can be used for a better visibility of details. The two most probable swine precursor gene sequences are located at the very bottom on the left: the H1N1 A/swine/Missouri/46519-5/2009 (HQ378741, light green) and the H1N2 A/swine/Hong_Kong/ NS252/2009 (CY085998, olive).

* corresponding author: janr@icm.edu.pl
Figure S2 – The NJ+ phylogram of the 2009 swine-flu pandemic H1N1 hemagglutinin early 178 sequences (light rose), which were isolated in the 2009, before April 30th.

Calculated as that on Fig. S1, but besides the 380 putative precursor sequences of other serotypes, collected during the same period, also additional 569 candidate genes, collected during the whole 2008 year were used. The region corresponding to the left inset of the Fig S1 is enlarged here. The result was that in addition to the two most probable swine precursor gene sequences are located at the bottom: the H1N1 A/swine/Missouri/46519-5/2009 (HQ378741, light green) and the H1N2 A/swine/Hong_Kong/NS252/ 2009 (CY085998, olive), also three other porcine H1N1 strains (light green), preceding in the same, small subclade are present: the A/swine/North_Carolina/3793/2008 (JQ624667), the A/swine/Illinois/ 02064/2008 (CY099095), and the A/swine/Ohio/02026/2008 (CY099159).
Positions of all the 9131 full-length hemagglutinin sequences analyzed (light gray points) superposed with the sequences corresponding to the hemagglutinin early 178 sequences (isolated early in the 2009, before April 30th, orange points; c.f. the bottom-left panel in Fig. 1; the rest of pandemic human isolates is marked yellow). The 380 putative precursor sequences of all the other serotypes, collected during the same period are also mapped (blue). The two most probable swine precursor gene sequences (from the phylogram on Fig. S1) are: the H1N1 A/swine/Missouri/46519-5/2009 (HQ378741, red) and the H1N2 A/swine/Hong_Kong/NS252/2009 (CY085998, black). Also three other porcine H1N1 strains (as in the legend to Fig. S2) are shown: the A/swine/North_Carolina/3793/2008 (JQ624667, cyan), the A/swine/Illinois/02064/2008 (CY099095, light green), and the A/swine/Ohio/02026/2008 (CY099159, magenta). Yet another H1N2 swine strain – the A/swine/Hong_Kong/1435/2009 (CY061653, square) was isolated later, on July 23rd.

Within the pandemic 2009 H1N1 cluster (at approx. PC-1 = -24, and PC-2 = -10) there are also five other porcine-host HA genes, but they are all more recent than May 2009: the H1N2 (circles) – A/swine/Italy/116114/ (CY067662), A/swine/Minnesota/A01076209/2010 (JQ906868), and A/swine/Nebraska/A01203626/2012 (JX444788); and the H1N1 (crosses) – A/swine/Illinois/ A01076179/2009 (JX042553, isolated Dec. 6th, 2009), and A/swine/Shepparton/6/2009 (JQ273542, isolated Aug. 17th, 2009). It is possible that in those cases infection might have followed a reversed transmission human → swine paths.

It is interesting to note that while all human strains of the pandemic 2009 H1N1 isolates are mapped at the very tip of their cluster (orange points on the Fig. S3), the later strains of the 2009 pandemic (yellow points) map more in direction towards the “common origin” location. Similarly like it was for H1N1 strains of the 1918, 1930’, 1940’, 1950’, etc. clusters (c.f. bottom-right panel of the Fig. 1), described in the main text.
Laboratory-induced transition to the droplet-transmissible infectivity

The Table ST1 contains the list of accession numbers for all H5N1 sequences carrying the full pattern of the seven nucleotides (C355, C510, A514, A713, G718, A720, C992) required to be present in order to undergo the transition described by Herfst et al. (2012) from avian transmissible (the “before” state) to mammals airborne-transmissible (the “after” state, i.e. carrying the pattern: T355, A510, G514, T713, A718, C720, T992). The distributions of genes carrying each of the seven individual nucleotides, specific to this transition (c.f. Herfst et al. 2012 and their table S1 on page 15 of the corresponding supplementary materials – pls. note that the numeration of nucleotides in their table differs from the shown here – our numbering starts always from the UAG codon of each gene) are mapped on Fig. S4 for the “before”, and Fig. S5 for the “after” transition nucleotide positions. The distribution of the single/individual “after” positions is quite strikingly different from the distribution of the single/individual “before” positions – too numerous to list here, except for the T355 position (present only in 17 strains; top-left panel on Fig. S5), and the T713 position (present only in 23 sequences; second top-right panel on Fig. S5).

It is also noteworthy, that as this transition can take place only among H5N1 strains, as all other strains are already capable of droplet borne infectivity in mammals, the three mutations: C355T, A713T, and A720C are indeed crucial for the emerging strain – among all the 9131 alleles they are present together in just a single pair of sequences: CY116646 and CY116654, indicative of their artificial, laboratory origin. Indeed, the visualization which of the “before” seven mutated positions were present with just one exception (that is which genes were carrying six out of the seven “before” nucleotides, c.f. Fig. S4) is quite striking.

The Fig. S6 shows the enlarged part of the “before” maps in Fig. S5, corresponding to just H5N1 strains – it is immediately evident that in contrast to the “after” distributions, the “before” one are quite ubiquitous. In contrast to a very broad dispersal for the single “before” positions as seen on Fig. S4 – encompassing each of HA serotypes; all of the six-out-of-seven cases were present exclusively among H5N1 isolates (mostly avian, but also 8 porcine, and 54 human, c.f. Table ST1); ranging in numbers from 220 (positions: C355, G718, C992; and also the all seven positions) to 404 (for the position A514) sequences. Our results differ from these of Russell et al. (2012) who performed a phylogenetic analysis of some H5N1 strains to reveal temporal, and to some extent also spatial, distributions of the between two to five of these seven mutations. On the one hand, it might seem that 220 orthologs, each carrying already all seven of the prerequisite “before” positions indicate quite substantial presence within all H5N1 strains in the whole data set. However, the comparison to just one sequence in the mutated “after” state clearly indicates rather very small probability of such an event occurrence. Finally, the Fig. S7 shows the details of positions of the five strains that are closest to the droplet-infection transmissible ferret #1 of Herfst et al. (2012).
Figure S4 – the map of all sequences carrying the “before” transition for the corresponding nucleotide positions

Figure S5 – the map of all sequences carrying the “after” transition for the corresponding nucleotide positions
<table>
<thead>
<tr>
<th>host</th>
<th>number</th>
<th>accessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>avian</td>
<td>158</td>
<td>EU124091, EU124082, EU124094, EU124158, EU124157, EU124093, EU124083, EU124153, EU124163, EU124162, EU124160, EU124155, EU124151, EU124149, EU930980, EU931004, EU124159, FJ784842, CY091811, CY091812, CY091815, CY091816, CY091819, CY091813, CY091817, CY091820, CY091800, CY091803, CY091874, CY091875, CY091876, CY091877, CY091789, CY091779, CY091782, CY091783, CY091784, CY091290, QG184236, QG184238, QG184239, JN807782, JN807777, JN807783, JN807785, JQ858472, JN807840, JN807843, FJ686831, CY062602, FJ686832, CY062603, CY062605, CY062606, CY062607, CY062608, FJ686833, FJ686836, CY062610, FJ357723, CY095680, CY095686, CY095701, JN055363, EU930892, EU124276, EU124277, CY091951, CY091289, CY014280, CY014465, CY014489, CY014497, CY014529, CY017638, CY017654, CY017662, CY017670, CY017688, CY019376, CY098603, CY098634, CY098641, CY098655, CY098668, CY098681, CY098688, CY098695, CY098702, CY098716, CY098730, CY098737, CY098751, DQ371928, DQ371930, EU146622, EU146640, EU146648, EU146681, EU146688, EU146755, EU146777, FJ432879, FJ432880, FJ432881, FJ492882, FJ492883, FJ492884, FJ492885, FJ492886, GQ466176, HM114545, HM114561, HM114569, HM114577, HM114585, HM114593, HM114609, HM114617, AY651334, AY818135, CY116646, GQ194235</td>
</tr>
<tr>
<td>human</td>
<td>54</td>
<td>CY014280, CY014465, CY014489, CY014497, CY014529, CY017638, CY017654, CY017662, CY017670, CY017688, CY019376, CY098603, CY098634, CY098641, CY098655, CY098668, CY098681, CY098688, CY098695, CY098702, CY098716, CY098730, CY098737, CY098751, DQ371928, DQ371930, EU146622, EU146640, EU146648, EU146681, EU146688, EU146755, EU146777, FJ432879, FJ432880, FJ432881, FJ492882, FJ492883, FJ492884, FJ492885, FJ492886, GQ466176, HM114545, HM114561, HM114569, HM114577, HM114585, HM114593, HM114609, HM114617, AY651334, AY818135, CY116646, GQ194235</td>
</tr>
<tr>
<td>swine</td>
<td>8</td>
<td>AY646424, DQ997253, DQ997262, HM440083, HM440091, HM440123, HM440131, HM440147</td>
</tr>
</tbody>
</table>

Table ST1 - the list of accession numbers for all H5N1 sequences carrying full configuration of nucleotides (C355, C510, A514, A713, G718, A720, C992) required to undergo the transition from avian transmissible to mammals droplet-transmissible (i.e. carrying the configuration: T355, A510, G514, T713, A718, C720, T992).
Figure S6 – the enlarged region of the PC-1 vs. PC-2 map H5N1 hemagglutinin, pinpointing the sequences corresponding to the “before” transitional state of avian transmissible to mammals airborne-transmissible. The sequences carrying all the seven positions (as described in Herfst et al. 2012), as well as carrying combinations of the six positions with an exclusion the one indicated, are color-coded as shown in their panel headings.
Figure S7 – the enlarged fragment of the PC-1 vs. PC-2 scatter-plot of the H5N1 hemagglutinin sequences, displaying the five strains that are closest to the droplet-infection transmissible ferret #1 of Herfst et al. (2012): A/ferret/Indonesia/5-F1/2005, CY116654 strain (c.f. Figs. S4, S5; and the main text Fig. 3B – panels of the H5N1 strains).