Global Genomic Studies of Influenza A Viruses: Identification of Conservative Sequences for Development of Wide Spectrum Vaccines and Gene Therapy Drugs

Kiril Kirilov[1], Vera Maximova[1], Stoyan Markov[2], Ivan Ivanov[1]
[1] Institute of Molecular Biology, Bulgarian Academy of Sciences, "Acad. G. Bonchev" Str., Bld. 21, 1113 Sofia, Bulgaria
[2] National Center for Supercomputer Applications, "Acad. G. Bonchev" Str., Bld. 25A, 1113 Sofia, Bulgaria

Keywords: influenza virus A, influenza virus genome, flu virus changeability

Influenza virus is highly infectious and causes worldwide flu epidemics and pandemics. The annual flu epidemics result in between three to five million cases of severe illness and between 250,000 and 500,000 deaths around the world (WHO EB111/10). There are three genera of influenza viruses designated as type A, B and C. Type A causes epidemics and pandemics, type B causes epidemics and type C only minor upper respiratory illness. Many subtypes of Influenza A virus are known according to the structure of their major surface antigens hemagglutinin (H) and neuraminidase (N). There are 18 different types of H (H1 to H18) and 11 of the N (N1 to N11) antigens. All combinations between the two are possible. Influenza viruses H17N10 and H18N11 were discovered in 2013 [1, 2]. Besides the great variety in H and N antigens in nature, only H1, H2 and H3, and N 1 and N2 are found in humans [3, 4].

The genome of influenza A virus is represented by 8 different segments of single-stranded RNA composed of 13588 ribonucleotides (nt) total. The 8 subgenomic RNAs encode 11 viral proteins (HA, NA, NP, M1, M2, NS1, NEP, PA, PB1, PB1-F2, PB2). The segmented nature of the viral genome allows whole subgenomic RNA segments to be swapped between different viral strains thus giving rise to novel combinations of subgenomic RNAs and therefore to new strains with increased virulence and lethality. Such viral strains caused the deadly pandemics in 1889, 1918, 1935, 1957, 1968, 1978 as well as the two recent "chicken" (2007) and "swine" (2009) flu pandemics.

The progress in whole genome sequencing after the year 2000 paved the way for comprehensive investigation of the flu virus variations under the *Influenza genome sequencing project* (IGRP). It is initiated in 2004 by the

National Institute of Allergy and Infectious Diseases (NIAID) and presently operates under The Venter Institute, USA [5].

The present study aims to shed more light on the changeability, evolution and adaptation of the influenza virus A genome in order to design new antiflu vaccines and antiflu drugs. To this end we employed the publicly available NCBI GenBank database [6] and the ClustalW_MPI software package adapted to the Bulgarian BlueGene/P supercomputer for identification of conservative domains in the influenza virus A genome that code for immunogenic peptides or representing unique siRNA target sites.

References

[1] Tong S. et al., Proc Nat Acad Sci USA 109 (2012) 4269–4274.

[2] Tong S. et al., PLoS Pathog 9 (2013) e1003657.
doi: 10.1371/journal.ppat.1003657

[3] Cox N. J. et al., Annu Rev Med 51 (2000) 407–421.

[4] Bouvier N., Palese P., Vaccine 26 (2008) Suppl.

[5] US Department of Health and Human Services NIH News, NIAID, November 15, 2004.

[6] http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html