



# Molecular Signature of Influenza A

Raghavendra Krishnappa<sup>1,2</sup>, Rajendran Paramasivam<sup>1,2</sup>

<sup>1</sup>Department of Microbiology, Karpagam University, Coimbatore, India

<sup>2</sup>Department of Microbiology, Sri Ramachandra Medical College & Research Institute Sri Ramachandra University, Chennai, India

Email: [raghavendrakbt@gmail.com](mailto:raghavendrakbt@gmail.com)

Received 29 July 2014; revised 20 September 2014; accepted 22 October 2014

Copyright © 2014 by authors and OALib.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

This research work provides comprehensive meaningful valuable findings from NCBI database for researches to analyze entries of H1N1 strains aiming to design genomic region-specific markers and also for screening the region-specific subtypes. Influenza A viruses are divided into subtypes, based on the nature of their surface glycoproteins, HA and NA. There are 16 different HAs and nine NAs which are distinguishable serologically, and we have sincerely dug NCBI database. We have found 7651 entries of H1N1 strains starting from the year 1918 till 2011 in the NCBI database, most recent entries from past few years includes, 1720 entries in the year 2008, 9112 entries in 2009, 1132 entries in 2010, 733 entries in 2011. Recent past few years has seen a rapid growth in high-throughput technology which has resulted in more entries accounting for 71.93% of the total entries in NCBI H1N1 Genome Database, in that, year 2009 which resulted in pandemic holds 51.62% share of the total H1N1 genome entries. 2009 entries were separated based on region of occurrence which resulted in 89 different regions. The above separation helps the researches for designing region-specific markers for screening the region-specific subtypes. Apart from that researches also get other key information like year of pandemic, subtypes responsible for pandemic, genes and virulence activity, genes functionality which can be used for genotyping.

## Keywords

NCBI, Hemagglutinin (HA), Neuraminidase (NA), Subtypes, Primers, PCR and Molecular Screening

**Subject Areas:** Genetics, Microbiology

## 1. Introduction

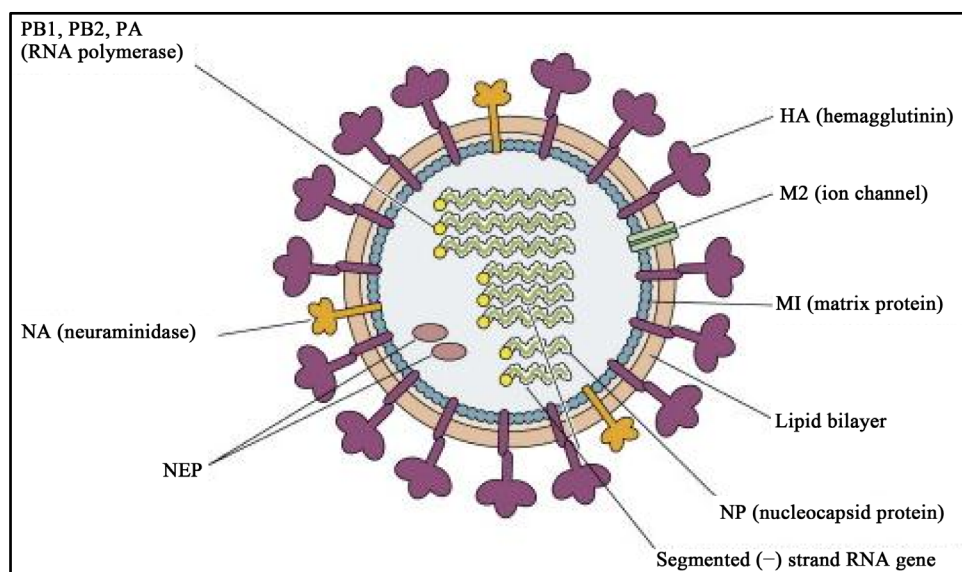
Influenza A viruses are a major cause of morbidity and mortality worldwide and affect large segments of the population every year. The nature of their genome, formed by eight segments of single-stranded RNA, shown in **Figure 2**, favors the constant evolution of the virus by two main mechanisms: the accumulation of single nuc-

leotide mutations in the viral genes introduced by an error-prone viral RNA polymerase and the re-assortment of genes between two strains of different origin. The viral genome encodes 11 proteins shown in **Figure 2**. Most have been shown to play a role in shaping the virulence scenario of influenza A viruses, including the adaptation of infection and transmission into new host species, the ability to modulate the host immune response and the capacity to replicate efficiently at low temperature. Shown in **Figure 1** on the surface of the virus particles there are two principal polypeptides, the hemagglutinin (HA) and the neuraminidase (NA), which are the target for the neutralizing antibodies immune response. There are 16 HA and 9 NA different subtypes in the influenza A virus that circulate in humans and animals. When a virus strain with a new HA or NA subtype appears in the human population by genetic re-assortment, it usually causes a pandemic because there is no preexisting immunity against the new virus. This was the case for the three pandemics that occurred during the last century (1918, 1957, and 1968) and also for the first pandemic of the 21<sup>st</sup> century, caused by the currently circulating A (H1N1) 2009 virus, which was generated by gene re-assortment between a virus present in pigs of North America and a virus that circulates in the swine population of Euroasia [1]. The Spanish influenza pandemic of 1918-1919 caused acute illness in 25% - 30% of the world's population and resulted in the death of 40 million people [2]. The type A viruses are the most virulent human pathogens among the three influenza types and cause the most severe disease, H1N1, which caused Spanish Flu in 1918 and Swine Flu in 2009. For both the 1918 and H5N1 viruses (which were responsible for pandemic), the hemagglutinin and the polymerase complex contribute to high virulence. Non-structural proteins NS1 and PB1-F2, which block host antiviral responses, also influence pathogenesis [3].

## 2. Summarized Information on Virulent Genes

### 2.1. Structure of Influenza Virus

As shown in **Figure 1** the influenza virion (as the infectious particle is called) is roughly spherical. It is an enveloped virus—that is, the outer layer is a lipid membrane which is taken from the host cell in which the virus multiplies. Inserted into the lipid membrane are “spikes”, which are proteins—actually glycoproteins, because they consist of protein linked to sugars—known as HA (hemagglutinin) and NA (neuraminidase). These are the proteins that determine the subtype of influenza virus (A/H1N1, for example). The HA and NA are important in the immune response against the virus; antibodies (proteins made by us to combat infection) against these spikes may protect against infection. The NA protein is the target of the antiviral drugs Relenza and Tamiflu. Also embedded in the lipid membrane is the M2 protein, which is the target of the antiviral adamantanes—amantadine and rimantadine. Beneath the lipid membrane is a viral protein called M1, or matrix protein. This protein, which forms a shell, gives strength and rigidity to the lipid envelope.



**Figure 1.** Structure of influenza virus [4].

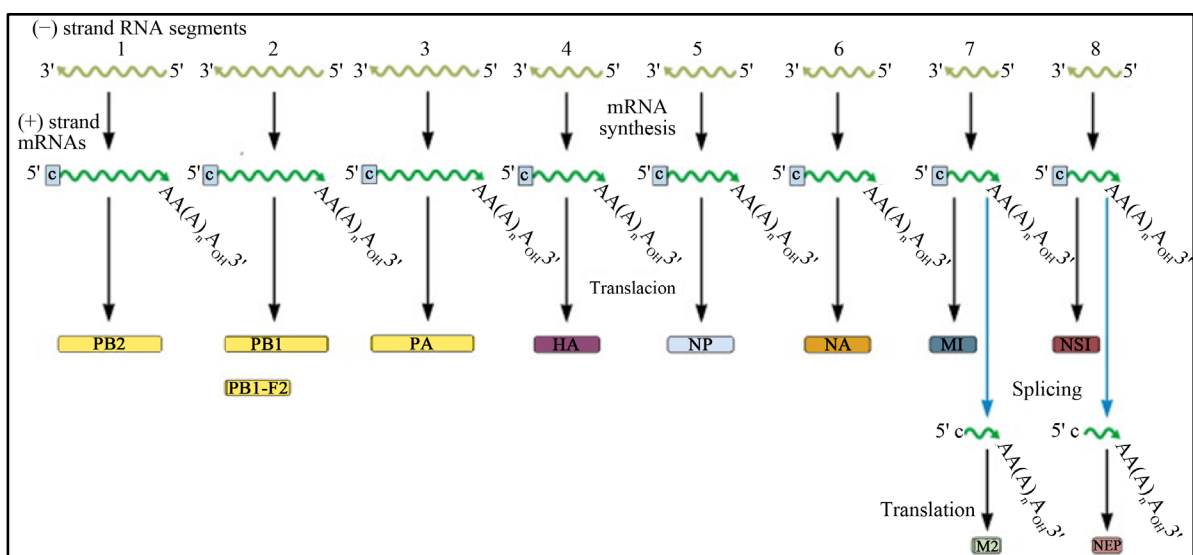
## 2.2. Influenza Virus RNA Genome

Within the influenza A virion are eight segments of viral RNA shown in **Figure 2**. These molecules carry the all the information needed to make new influenza virus particles. These eight RNAs are shown schematically as olive green lines at the top of the illustration. RNAs are chains of four different nucleotides, A, C, G, U. In the case of influenza virus, the eight RNAs are a total of about 14,000 nucleotides in length. The nucleotides make up the genetic code—it is read by the cell's translation machinery in groups of three, with each triplet specifying an amino acid.

Influenza viral RNAs are called (–), or negative strand RNAs, because they are the opposite polarity of the RNA which is translated to make protein. The RNA molecules that are templates for the synthesis of proteins are defined as having (+), or positive polarity. Upon entering the cell, the (–) strand influenza viral RNAs must be copied into complementary (+) strands, so that they can serve as templates for proteins. The viral RNAs are copied by an enzyme-called RNA polymerase—that is carried into the cell with the virus. Once the virion enters the cell, these 8 RNAs are copied into (+) strand mRNAs. Finally, the mRNAs can serve as templates for the synthesis of proteins. Both influenza A and B viruses have 8 RNA segments, while the influenza C viruses have 7.

Virulence: Influenza A viruses replicate in airway epithelium and are transmitted by the respiratory route [6]. They bind host cells via their HA, which interacts with sialic acid residues on host cell surface molecules. Human and avian influenza virus HA differs with regard to their specificity for host cell receptors. Human virus HAs preferentially bind receptors with oligosaccharides ending in sialic-acid  $\alpha$ -2,6-galactose ( $Sa\alpha$ -2,6gal) whereas avian viruses preferentially recognize  $Sa\alpha$ -2,3gal [7] [8]. Time taken for symptoms to happen of viremia usually began about 2 days after exposure to the virus. The potential impact of pandemic influenza on the safety of the blood supply should be small because of the limited viremia and the nature of respiratory tract infection of influenza viruses [9].

Recently, evidence has emerged indicating that receptor specificity influences the anatomical site of replication of influenza viruses and could therefore affect the transmissibility of influenza viruses [10]. Given the fact that transmissibility is an essential characteristic of a pandemic virus, the acquisition of  $Sa\alpha$ -2,6gal specificity would appear to be an important event in the generation of an influenza virus with pandemic potential. Interestingly,  $Sa\alpha$ -2,6gal predominates on epithelial cells in the human nasal mucosa, paranasal sinuses, the pharynx, trachea and bronchi [10]. In contrast,  $Sa\alpha$ -2,3gal is hardly detected in these cells, but it is found in the non-ciliated cuboidal bronchiolar cells at the junction between the respiratory bronchiole, alveolus and alveolar Type II cells. Cumulatively, these studies suggest that receptor specificity contributes to inefficient transmission of H5N1 influenza viruses between humans. In human strains, such as the 1918 virus, receptor specificity has been experimentally demonstrated to influence transmission [11]. However, it must be recognized that other viral factors will also contribute to transmission, but their role remains to be defined.



**Figure 2.** Influenza viral RNA segments [5].

High virulence has been associated with increased levels of influenza A virus replication *in vivo*. Viral replication rates could obviously be influenced by the rate at which a virus can replicate its genes and express its mRNAs. Thus, viral polymerase complexes that display increased rates of function could contribute to increased virulence. Influenza virus RNA synthesis is carried out by a heterotrimeric RNA-dependent RNA polymerase, composed of the PA, PB1 and PB2 proteins, and the viral nucleoprotein (NP), which is always found being associated with the template RNAs. The two viral non-structural proteins, NS1 and PB1-F2, appear to facilitate evasion of host defenses [12]-[14]. The ability of NS1 to block host-cell gene expression contributes to the suppression of IFN $\alpha/\beta$  responses and is important for virulence [12] [13].

Influenza A viruses are divided into subtypes, based on the nature of their surface glycoproteins, HA and NA. There are 16 different HAs and nine NAs which are distinguishable serologically, *i.e.* antibodies to one virus subtype do not react with another [15]. All virus subtypes appear to circulate in aquatic birds [16]. The immune system normally responds to influenza virus by making neutralizing antibodies to regions of the viral spike, the hemagglutinin, that vary year to year. This natural response protects against circulating subtypes but necessitates production of new vaccines annually [15].

Influenza A viruses are known to also infect a variety of other mammals, including non-human primates, pigs, horses, cats, seals, whales and mink (shown in **Table 1**) [15]. The B virus primarily infects humans, although it has also been isolated from seals.

The **Table 1** indicates the subtypes of haemagglutinin (HA) and neuraminidase (NA), and the hosts in which they have been identified.

### 2.3. Influenza entries in NCBI

**Table 2** shows there are totally 17651 entries of H1N1 strains starting from the year 1918 in the NCBI database, most recent entries from past few years includes 1720 entries in the year 2008, 9112 entries in 2009, 1132

**Table 1.** Natural hosts of influenza A viruses.

Natural hosts of influenza A viruses			
Haemagglutinin		Neuraminidase	
Subtype	Predominant hosts	Subtype	Predominant hosts
H1	Human, pig, birds	N1	Human, pig, birds
H2	Human, pig, birds	N2	Human, pig, birds
H3	Birds, human, pig, horse	N3	Birds
H4	Birds	N4	Birds
H5	Birds, (human)	N5	Birds
H6	Birds	N6	Birds
H7	Birds, horse, (human)	N7	Horse, birds
H8	Birds	N8	Horse, birds
H9	Birds, (human)	N9	Birds
H10	Birds		
H11	Birds		
H12	Birds		
H13	Birds		
H14	Birds		
H15	Birds		
H16	Birds		

**Table 2.** Total number of H1N1 entries in NCBI database from the year 1918 to 2011.

Reported in the	Total No.
Yr 1918-2011	17651
Yr 2008	1720
Yr 2009	9112
Yr 2010	1132
Yr 2011	733

entries in 2010, 733 entries in 2011. NCBI H1N1 Genome Database (<http://www.ncbi.nlm.nih.gov/genomes/FLU/SwineFlu.html>) year 2009 which resulted in pandemic holds 51.62% share of the total H1N1 genome entries. NCBI database as seen a rapid increase in H1N1 sequence entries in the recent past years, last four years as contributed about 71.93% of the total entries for H1N1 starting from 1918. New DNA sequencing technologies can sequence up to one billion bases in a single day at low cost, putting large-scale sequencing within the reach of many scientists. Many researchers are forging ahead with projects to sequence a range of species using the new technologies [17].

From 89 different regions H1N1 been reported, the above **Table 3** helps us in detecting region-specific infection. Markers are specific to the above specified strains pertaining to region ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Record from NCBI database mentioning the origin of Influenza A virus from New York (A/New York/DW09D5R/2011 (H1N1)). The current nomenclature system for influenza A viruses includes the host of origin, geographic location of first isolation, strain number and year of isolation. WHO Memorandum, a revised system for nomenclature of influenza viruses. Bull World Health Org 58 (1980) (585 - 591). The HA and NA subtype of influenza A viruses is specified in parentheses, e.g. A/Swine/Iowa/15/30 (H1N1). By convention, the host of origin of human strains is omitted, e.g. A/Puerto Rico/8/34 (H1N1). Since there are no subtypes of influenza B virus, no parenthetical specification is given in this case and, as the B viruses primarily infect humans, the host of origin is not mentioned in the influenza B virus nomenclature, e.g. B/Yamagata/16/88.

### 3. Conclusions

Appearance of H1N1 in the human population was observed from the year 1918 (based on the sequence information available from the NCBI database, Influenza A virus (A/AFIP/1/1918(H1N1)), till 2011, total of 17651 sequence entries can be found for H1N1 originated from different places, past few years has seen a rapid growth in high-through put technology which has resulted in more entries accounting for 71.93% of the total entries in NCBI Genome Database (there are totally 17651 entries of H1N1 strains starting from the year 1918 in the NCBI database, most recent entries from past few years includes, 1720 entries in the year 2008, 9112 entries in 2009, 1132 entries in 2010, 733 entries in 2011), in that, year 2009 which resulted in pandemic holds 51.62% share of the total H1N1 genome entries.

H1N1 viral genome in NCBI genome database is in bits rather than in one complete stretch, this is because the influenza virus genome is not one complete stretch. There are eight RNA segments. These are separate molecules. Complete influenza A H1N1 viral sequences are linked to the NCBI genome record:

<http://www.ncbi.nlm.nih.gov/genome/10290>, <http://www.ncbi.nlm.nih.gov/genomes/FLU/SwineFlu.html>.

HA (hemagglutinin) and NA (neuraminidase). These are the proteins that determine the subtype of influenza virus (A/H1N1, for example). Since HA and NA determine the subtype of Influenza A their gene sequence can be used as marker for diagnosis and genotyping subtypes in molecular screening reaction which includes RT-PCR, real-time RT-PCR, nucleic acid sequence-based amplification, microsphere-based duplexed immunoassay, and DNA microarrays. Conserved region obtained after aligning HA or NA sequence help in identifying subtypes, if the sample is of many subtypes, than in that case two primers pair each specific for the conserved region of HA and NA as to be designed to genotype the subtypes accurately. From **Table 1** we have 17 HA and 9 NA possible subtypes, if we have to detect any of the subtypes, we need to have 17 unique primer pairs for each HA (can be obtained by aligning all 17 HA) variants and 9 unique primer pairs for each NA (can be obtained by aligning all 9 NA) variants. This molecular signatures help in detection of flu subtypes isolated

**Table 3.** H1N1 reported from 89 different regions.

Origin	No.	Origin	No.	Origin	No.
Arad	3	Illinois	2	Philippines	7
Archangelsk	2	Izhevsk	1	Prague	1
Arizona	1	Japane	1	Prahova	1
Beijing	6	Jiangsu	3	Rio Grande do Sul	35
Bihor	2	Joshkar	1	Saint-Petersburg	6
Boston	21	Kaliningrad	1	Salaj	1
Brasov	1	Korea	7	Shanghai	5
Bucuresti	5	Kyoto	8	Shenzhen	90
Budapest	4	Maramures	1	Singapore	21
Calarasi	1	Mexico	33	South Carolina	3
California	1	Milan	51	Suita	15
Canada	1	Missouri	1	Suzhou	2
Casablanca	20	Moscow	27	Switzerland	1
Cheboksary	1	Mures	1	Sydney	2
China	1	Murmansk	1	Taiwan	12
Cluj	1	Nagasaki	5	Thailand	2
Columbia	3	Neamt	1	Timis	1
Constanta	1	New Jersey	1	Tomsk	2
Czech Republic	1	New Mexico	1	Toyama	20
Denmark	11	New York	1	Tula	10
Ecuador	1	Niigata	8	Tunisia	2
Finland	2	Nizhnii Novgorod	13	Turin	10
Florida	2	North Carolina	3	Tver	1
Georgia	1	Novosibirsk	8	Ufa	1
Guangdong	20	Ohio	5	Ulaanbaatar	1
Gunma	3	Orenburg	1	Vladimir	1
Hangzhou	15	Osaka	8	Volgograd	1
Hokkaido	4	Palermo	5	Voronezh	1
Hyogo	10	Pavia	82	Washington	1
Iasi	1	Penza	1		

from any host through polymerase reaction, key point for these holds in getting very stable unique conserved region for each variants.

## References

- [1] Arias, C.F., Escalera-Zamudio, M., de los Dolores Soto-Del Río, M., Cobián-Güemes, A.G., Isa, P. and López, S. (2009) Molecular Anatomy of 2009 Influenza Virus A (H1N1). *Archives of Medical Research*, **40**, 643-654.
- [2] Taubenberger, J.K., Reid, A.H., Janczewski, T.A. and Fanning T.G. (2001) Integrating Historical, Clinical and Molecular Genetic Data in Order to Explain the Origin and Virulence of the 1918 Spanish Influenza Virus. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **356**, 1829-1839. <http://dx.doi.org/10.1098/rstb.2001.1020>
- [3] Basler, C.F. and Aguilar, P.V. (2008) Progress in Identifying Virulence Determinants of the 1918 H1N1 and the Southeast Asian H5N1 Influenza A Viruses. *Antiviral Research*, **79**, 166-178. <http://dx.doi.org/10.1016/j.antiviral.2008.04.006>

- [4] <http://www.virology.ws/2009/04/30/structure-of-influenza-virus/>
- [5] <http://www.virology.ws/2009/05/01/influenza-virus-rna-genome/>
- [6] Lowen, A.C., Mubareka, S., Steel, J. and Palese, P. (2007) Influenza Virus Transmission Is Dependent on Relative Humidity and Temperature. *PLoS Pathogens*, **3**, e151. <http://dx.doi.org/10.1371/journal.ppat.0030151>
- [7] Rogers, G.N. and Paulson, J.C. (1983) Receptor Determinants of Human and Animal Influenza Virus Isolates: Differences in Receptor Specificity of the H3 Hemagglutinin Based on Species of Origin. *Virology*, **127**, 361-373. [http://dx.doi.org/10.1016/0042-6822\(83\)90150-2](http://dx.doi.org/10.1016/0042-6822(83)90150-2)
- [8] Kochs, G., García-Sastre, A. and Martínez-Sobrido, L. (2007) Multiple Anti-Interferon Actions of the Influenza A Virus NS1 Protein. *Journal of Virology*, **81**, 7011-7021. <http://dx.doi.org/10.1128/JVI.02581-06>
- [9] Zou, S. (2006) Potential Impact of Pandemic Influenza on Blood Safety and Availability. *Transfusion Medicine Reviews*, **20**, 181-189. <http://dx.doi.org/10.1016/j.tmr.2006.03.001>
- [10] Shinya, K., Ebina, M., Yamada, S., Ono, M., Kasai, N. and Kawaoka, Y. (2006) Avian Flu: Influenza Virus Receptors in the Human Airway. *Nature*, **440**, 435-436. <http://dx.doi.org/10.1038/440435a>
- [11] Tumpey, T.M., Maines, T.R., Van Hoeven, N., Glaser, L., Solórzano, A., Pappas, C., *et al.* (2007) A Two-Amino Acid Change in the Hemagglutinin of the 1918 Influenza Virus Abolishes Transmission. *Science*, **315**, 655-659. <http://dx.doi.org/10.1126/science.1136212>
- [12] Nabel, G.J. and Fauci, A.S. (2010) Induction of Unnatural Immunity: Prospects for a Broadly Protective Universal Influenza Vaccine. *Nature Medicine*, **16**, 1389-1391. <http://dx.doi.org/10.1038/nm1210-1389>
- [13] Chen, Z.Y., Li, Y.Z. and Krug, R.M. (1999) Influenza A Virus NS1 Protein Targets Poly(A)-Binding Protein II of the Cellular 3'-End Processing Machinery. *The EMBO Journal*, **18**, 2273-2283. <http://dx.doi.org/10.1093/emboj/18.8.2273>
- [14] Noah, D.L., Twu, K.Y. and Krug, R.M. (2003) Cellular Antiviral Responses against Influenza A Virus Are Countered at the Posttranscriptional Level by the Viral NS1A Protein via Its Binding to a Cellular Protein Required for the 3' End Processing of Cellular Pre-mRNAs. *Virology*, **307**, 386-395. [http://dx.doi.org/10.1016/S0042-6822\(02\)00127-7](http://dx.doi.org/10.1016/S0042-6822(02)00127-7)
- [15] Lamb, R.A. and Krug, R.M. (2001) Orthomyxoviridae: The Viruses and Their Replication. In: Knipe, D.M., Howley, P.M. and Griffin, D.E., Eds., *Fields Virology*, 4th Edition, Lippincott Williams & Wilkins, Philadelphia, 1487-1531.
- [16] Hinshaw, V.S., Webster, R.G., Bean, W.J. and Sriram, G. (1980) The Ecology of Influenza Viruses in Ducks and Analysis of Influenza Viruses with Monoclonal Antibodies. *Comparative Immunology, Microbiology and Infectious Diseases*, **3**, 155-164. [http://dx.doi.org/10.1016/0147-9571\(80\)90051-X](http://dx.doi.org/10.1016/0147-9571(80)90051-X)
- [17] Pop, M. and Salzberg, S.L. (2008) Bioinformatics Challenges of New Sequencing Technology. *Trends in Genetics*, **24**, 142-149. <http://dx.doi.org/10.1016/j.tig.2007.12.006>