REVIEW ARTICLE

Annals of Agricultural and Environmental Medicine 2013, Vol 20, No 3, 455–459
www.aaem.pl

Molecular evolution and phylogenetic implications in clinical research

Łukasz Podsiadło, Małgorzata Polz-Dacewicz

Department of Virology, Medical University, Lublin, Poland

Podsiadło Ł, Polz-Dacewicz M. Molecular evolution and phylogenetic implications in clinical research. Ann Agric Environ Med. 2013; 20(3): 455–459.

Abstract

A phylogenetic tree shows graphically the evolutionary relationships among various organisms. The dynamic development of molecular biology and bioinformatics has led to a revolution in our knowledge of biological evolution and the kinships between living organisms and viruses. Nowadays, the available laboratory techniques and computer software allow reconstruction of the actual changes which occurred in the evolutionary process. The derivation of molecular evolution models and several methods for building phylogenetic trees have played a huge role in that enterprise. The emergence of new infectious agents is a problem afflicting mankind since prehistoric times. The study of phylogenetic implications among pathogenic microorganisms allows tracking the process of evolution, the indirect understanding of their biology, and thus facilitates the implementation of treatment.

The presented article demonstrates the basic methods for constructing phylogenetic trees, as well as the benefits of reconstructing the evolution process and kinship with the study of microorganisms; in particular, viruses are considered from the clinical aspect.

Key words

phylogenetic tree, molecular evolution, substitution models, phylogenetic inference, taxonomy, viruses

INTRODUCTION

The concept of evolution is connected with the susceptibility to making errors during DNA replication, which means that duplicates are not always identical to the original. If DNA replication is accurate, there would be no variation on which natural selection could act. Errors are thus the key to evolution.

In many cases, the classification of the species was carried out on the basis of morphological features. Sometimes, molecular studies confirm the findings made on the basis of morphological characteristics, but they are also often contradictory. The main task of molecular phylogenetic analysis is the construction of the tree. At the present time, there are computer programmes that allow for fast and reliable data analysis and the construction of a phylogenetic tree. The software Frequently used for this purpose are MEGA, PHYML and MrBayes [1].

Ernst Haeckel constructed the first phylogenetic tree of living organisms in 1866, based primarily on morphological features. He was inspired by Charles Darwin's theory of evolution [2, 3], but he could not expect that one hundred years later laboratory techniques and computer software would exist which enable the reconstruction of the evolutionary process in an incomparably more credible way. In the past, phylogenetics was considered a very difficult discipline of biology in which only systematists were engaged [1]. Their tasks were primarily the identification, nomenclature and classification of organisms; additionally, they attempted to explain the historical relationships between them. Currently, more and more researchers from various fields of biology are using this type of analysis. Phylogenetic trees help to understand better the biological processes occurring in the world of living organisms and viruses. In the case of microorganisms, for instance, there can be illustrated a kinship of different species of bacteria or track an evolution process and the emergence of new viral strains.

A phylogenetic tree shows the evolutionary relationships between sequences or species of living organisms. A typical phylogenetic tree is composed of branches connected by nodes, and their arrangement is called the topology of the tree. External nodes, so-called Operational Taxonomic Units (OTU), represent sequences of DNA, proteins, or the taxa used in the analysis. Frequently, trees are constructed with an indication of the length of its branches. This shows the time of emergence of new evolutionary lineages and the subsequent degree of sequence divergence, which is expressed by the number of substitutions per site [4].

As mentioned, researchers from various fields of biology increasingly apply the phylogenetic trees to present their results. There are several types and shapes of trees. Figures 1A, 1D and 1E are common phylograms with the branch length and scale provided, which represents the number of substitutions per site. Figures 1B and 1C are cladograms, which represent only the relationship between the OTUs, and the branch length does not reflect the degree of sequence divergence. There are also a few shapes of trees. Figures 1A, 1C, 1D and 1E illustrate rectangular, straight, radial and circular shapes, respectively.

The phylogenetic tree can be rooted or unrooted. An unrooted tree (Fig. 1D) shows only the relationships among the examined taxonomic units and, in this case, no conclusion can be made about the direction of evolution. In a rooted tree (Fig. 1A, 1B, 1C, 1E) the root represents the common ancestor of all analyzed sequences, which allows for inference about the order of sequence inheritance. To place the root properly, an out group (external group) has to be added to the analysis [1, 4]. An outgroup is defined as one or several sequences more distantly related to the sequences of the internal group than to the OTUs from the second group to each other. In

Address for correspondence: Łukasz Podsiadło, Department of Virology, Medical University, Chodźki 1, 20-093 Lublin, Poland
e-mail: lukasz.podsiadlo@op.pl

Received: 20 October 2012; accepted: 16 January 2013

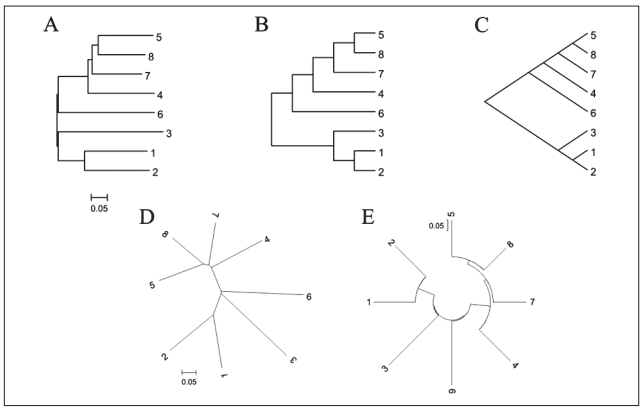

Figure 1. Basic types and shapes of phylograms and cladograms. Types: A, D, E – phylograms; B, C – cladograms. Shapes: A, B – rectangular; C, D, E – straight, radial and circular, respectively. Branch length scale under the A tree represents the number of substitutions per site.

Figure 1 (A, B, C, E), the outgroup includes OTU 1, 2 and 3 in relation to the rest of the OTUs.

Models of molecular evolution. Molecular evolution is a process that occurs at the level of DNA, RNA and proteins. One of the important process in the evolution of living organisms is mutations, called substitutions, which are divided into transitions and transversions that are only at the nucleic acid level. Transition is defined as a point mutation which occurs within the same group of nucleotides (purines or pyrimidines), whereas transversions involve conversion of purine to pyrimidine and vice versa. Although the probability of transversion is twice as large, in nature it is observed at twice the rate of transitions [5, 6].

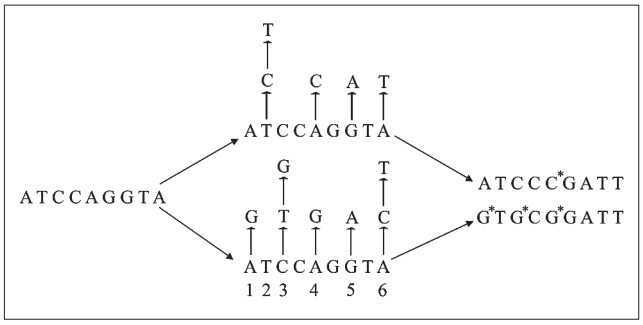

Figure 2. Demonstration of two homologous sequence alignment and types of substitutions that occurred in the course of evolution. Although there were twelve mutations, the differences can be observed only in three positions (*). 1 – single substitution; 2 – back substitution; 3 – multiple substitution; 4 – simultaneous substitution; 5 – parallel substitution; 6 – convergent substitution.

An evolutionary distance between the analyzed sequences can be expressed using a mathematical model which has a biological justification. The simplest way for its determination is to calculate a percentage of different positions in compared sequences. This method, however, does not reflect the exact evolutionary distance; it does not take into account the different rate of mutational changes and multiple substitutions [7]. Mathematical models of evolution can be briefly defined as some assumptions concerning the process of nucleotide/amino acid substitutions in the sequences of DNA/proteins [8]. Reconstruction of the evolution is a very difficult task. Models of nucleotide substitutions allow for a better understanding of the evolution and construction of the tree, which will be more reliably show the relationship between organisms.

The model devised by Jukes and Cantor (JC69) is the simplest, one-parameter model of sequence evolution, which assumes that the probability of any substitution is equal and identical in each sequence position. Furthermore, each of the four types of nucleotides in the sequence occurs with equal frequency [8, 9]. The Jukes-Cantor distance between sequences is the evolutionary distance defined as an estimated number of substitutions per position in the sequence [7]. More complex models of DNA evolution use more parameters to describe the process of substitutions. These parameters reflect differences in the frequency of nucleotides, the transition and transversion rate, and the rate of substitutions in various places in the sequence [8].

The Felsenstein model 81 (F81) [10] is an extension of the JC69 model and assumes a different frequency of nucleotides but model of Kimura (K80) [11] not, similarly to JC69 [12]. Tamura-Nei model (TN93) [13], the Felsenstein 84 (F84), Hasegawa, Kishino and Yano (HKY) models introduce an additional distinction between transitions for purines and pyrimidines. One of the most complex models is the General Reversible Time model (GTR) with six parameters that determine the frequency of each substitution. In the literature there is a lot of information about these models and their comparison [8, 9, 12].

Table 1. Summary of some commonly used substitutions models. a: A↔C, b: A↔G, c: A↔T, d: C↔G, e: C↔T, f: G↔T [8].

Model	No. of parameters	Substitution rate
JC69	1	a=b=c=d=e=f
K80	2	a=c=d=f, b=e
F81	4	Not included
HKY85	6	a=c=d=f, b=e
GTR	10	a, b, c, d, e, f

The discussed models assume the same rate of substitutions, regardless of location in the sequence. In fact, there are positions (e.g. region of the initiation of translation) where the rate is slower. This is due to natural selection, which does not allow for mutations in important places in the sequence. The distribution of substitution rate along the molecule can be described by the gamma distribution (Γ distribution). To apply this, the value of α parameter has to be known, which describes the shape of the gamma distribution. There can also be also introduced into the analysis the I parameter, which determines the proportion of invariant positions in the sequence [7, 12].

Methods for building phylogenetic trees. In the literature, the division of phylogenetic inference methods are divided into two main groups: distance-based and character-based. The unquestionable advantage of distance-based methods is the velocity of tree construction, while character-based methods are characterized by higher credibility of phylogeny reconstruction, but are more time-consuming [14]. Unfortunately, there is no objective way to choose an appropriate method for the construction of trees. It is worth considering which is the important – accuracy, easiness of interpretation, or the time of analysis.

Distance-based methods. Phylogeny estimation by using distance-based methods consists in determining the evolutionary distance between sequences, based on multiple alignments and defining the distance matrix. The multiple alignments of analyzed sequences is achieved using appropriate software (e.g. ClustalX). The distance matrix shows the obtained distances between each pair of sequences and is used to determine the tree topology and calculate the length of branches. There are two crucial distance-based methods: Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and Neighbour Joining (NJ) method [7].

UPGMA is one of the simplest methods of phylogenetic trees construction. It has numerous limitations and wrong assumptions; therefore, this method is not readily used. The UPGMA algorithm assumes that the tree is additive (the distance between any two nodes is equal to the total length of the branches connecting them) and ultrametric (all OTU are at the same distance from the root). This means that it applies the molecular clock hypothesis, whereby the evolution of all different species occurs at the same pace. This assumption is obviously wrong [7, 8, 15]. The tree in the NJ method is not ultrametric, which has the ability to obtain quickly a relatively credible phylogram. For this reason, this method is widely used in current research [14, 16].

Character-based methods. The distance methods discussed above are undoubtedly fast and build only one tree from a given set of data. Otherwise, character-based methods are time-consuming and labour-intensive. Subsequently, the suitable software selects one or a few trees, which reflect the reality in the best way possible. Character-based methods include maximum parsimony, maximum likelihood and Bayesian inference [1].

Maximum parsimony (MP) is one of the earliest methods proposed for the reconstruction of phylogeny. It is based on the main assumption that the best is the tree, which explains the changes in the sequences by the smallest possible number of substitutions. For instance, if cytosine is in the same position of two compared sequences, their common ancestor also has a cytosine at this position [17, 18, 19]. In the case of a very large diversity of 'characters', this method justifies this by the principle of reversion, convergence and parallelism, which are described here using the common term homoplasies. Reversion is a change of the feature and then returns to its initial state. Convergence is a process of independent development of the same traits in unrelated organisms, while parallel is a development of similar traits in related but distinct species [1].

The maximum likelihood (ML) allows for the likelihood estimation of data. A suitable computer algorithm builds the tree with the biggest value of the reliability logarithm, which is the sum of logarithms of all positions in the compared sequences. The phylogenetic tree parameters, for which the reliability is calculated, are the topology, branch length and substitution model. ML method requires a software which analyze different models of DNA evolution and their parameters [20, 21]. In order to choose the appropriate model, several methods are used: hLRTs (hierarchical Likelihood-Ratio Tests), AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) [22, 23].

Bayesian inference, in contrast to ML, works on sets and instead of selecting the single, most reliable tree, it creates a set of trees with large credibility. For the three types of tree parameters (topology, branch length and substitution model) the values of the *a priori* probabilities are calculated. From the set of trees with calculated maximum likelihood value the programme builds one tree in which each node has its *a posteriori* probability [21, 24].

Molecular phylogenetics in clinical research. Nowadays, molecular phylogenetics is a powerful research tool. It is increasingly commonly used not only in biology but also in clinical research, namely, in bacteriology, mycology and virology. Phylogeny is primarily applied in taxonomy, epidemiology and forensic medicine. It enables the tracking of the evolution of pathogens, the study their origin, and identification of new infectious agents.

Taxonomy of viruses. The enormous amount of viral species in nature arouses curiosity about not only their origin, but also forces their naming and organizing them into hierarchically arranged systematic units. To achieve this, the virology section of the International Union of Microbiology Societies (IUMS) has appointed an International Committee on Taxonomy of Viruses (ICTV).

The *Papillomaviridae* family is a good example to demonstrate the importance of taxonomy in clinical research. This is a very large group of viruses which infect both animals and humans, and together with polyomaviruses were classified into the *Papovaviridae* family. The molecular biology techniques, such as PCR or sequencing, revealed important differences between these two groups of viruses according to the ICTV regulations: the *Papovaviridae* family was distinguished into *Papillomaviridae* and *Polyomaviridae* family [25]. Detailed sequences analysis of papillomaviruses resulted in supplementing this group with additional taxonomic units, such as 'types', 'subtypes' and 'variants'. This division was based on the percentage variation in the capsid protein L1 gene sequence. The above-mentioned taxa show differences in the sequence at the level of >10%, 2%-10% and <2%, respectively [25, 26].

Papillomaviruses have tropism for the squamous epithelium, are responsible for the formation of warts, benign skin lesions, and may lead to cancer [27, 28]. The majority of human papillomaviruses belong to *Alpha-*, *Beta-* and *Gammapapillomavirus genera*, and there are many species and types among them. In the literature, there are often found the division of viruses with high, medium and low oncogenic potential. This division is mainly used by clinicians. From a systematic point of view, for instance, viruses with high oncogenic potential belong to the species 7 (HPV-18, -39, -45, -59, -68, -70) and 9 (HPV-16, -31, -33, -35, -52, -58, -67) [26]. Phylogenetic analysis revealed that papillomaviruses detected in similar localized lesions are in distant branches on phylogenetic trees. HPV-1, HPV-2, HPV-4 and HPV-41, which cause benign skin lesions, can be used as such examples. Interestingly, HPV-16 and HPV-18 are more closely related to the types undetectable in the cervical epithelium than with each other [25]. However, correct genotype determination is a very important issue. It enables assessment of the risk scale and implementation of antiviral therapy.

Phylogenetics in epidemiology and criminology. Molecular biology techniques are also used in epidemiology and forensic medicine. The rapid evolution of viruses, in particular RNA

viruses, has led to emergence of many new genotypes of these microorganisms. For instance, hepatitis B virus (HBV) and hepatitis C virus (HCV) have 8 (A-H) and 6 (1–6) genotypes, respectively [29, 30]. From the epidemiological point of view, it is very important to outline the geographic range of viruses with their genotypes. For this, the sequencing of whole genomes, genes or gene fragments is more and more commonly used. These sequences can be subsequently compared with homologous sequences available in the international database GenBank (http://www.ncbi.nlm.nih.gov/genbank/). With appropriate software, conclusions can be drawn about the occurrence of a particular genotype in the population of the region. Moreover, phylogenetic analysis allows determination of the directions of virus migrations from one country or continent to another [31, 32].

Nowadays, the methods such as PCR, sequencing, and DNA fingerprinting are also used successfully in medicine. The great advantage of this type of analysis is its high sensitivity and repeatability. A well-known examples is the case of an American gastroenterologist, who was accused of injecting his former girlfriend with a mixture of blood or blood products, collected from one patient infected with HIV-1 and the second infected with hepatitis C virus. Phylogenetic analysis showed that HIV-1 from the patient and victim were very closely related and located in the common branch on phylogenetic trees, compared with control samples. The Supreme Court of the United States accused the doctor of attempting to commit second-degree murder [33]. A similar case concerned the patients infected by a dentist in Florida who was infected with HIV-1 [34]. Obviously, in such cases the error probability should be minimized. Therefore, researchers have to be very accurate, several methods of tree construction should be applied and analyses should be conducted in at least two separate laboratories.

Origin, evolution and emergence of new viruses. Infectious diseases have accompanied mankind since time immemorial. Everyone realizes the importance of the issue of the origin and emergence of new pathogens. More than half of the infectious disease agents, such as Ebola, yellow fever, influenza A or hepatitis B viruses originate from animals [35]. Different types of African tribes hunt and eat the meat of local fauna, often monkey meat, because the sources of animal protein such as pork or beef are rarely available in this region. The probability that viruses and other microorganisms from primates will transfer to humans is much greater than in the case of antelope, whose relationship with man is more distant. Therefore, not only the microorganisms found in domestic pets should be controlled, but also those developing in animals living in the wild [35, 36].

The SFV virus (Simian Foamy Virus) belongs to *Retroviridae* family and was detected in hunters from the Central Africa region. This virus is present in most primates. In one, particular case, a 45-year-old man contracted the SFV variant which occurs in gorillas. It appeared that the man hunted these animals [37]. In the members of the same community, HTLV-3 and HTLV-4 viruses have also been discovered. Based on the close similarity of HTLV-3 to the STLV-3 counterpart, it can be assumed that the infection occurred during the hunting for monkeys infected with STLV-3 [38]. In the case of HIV, the first sample containing HIV-1 was taken in 1959 from a man in the Democratic Republic of Congo [39]. It is now well known that the HIV-2 virus, which also causes AIDS in humans, is a distant cousin of HIV-1 and is very closely related to SIV virus found in sooty mangabeys, while HIV-1 has a close relationship with viruses isolated from chimpanzees. Furthermore, even before 1959, the region of Central Africa was considered the place of departure for the expansion of the human immunodeficiency virus,. Studies have shown that HIV appeared in the human population long before the defining of the disease known today as AIDS [40].

These examples clearly highlight the role of phylogenetic analysis and reconstruction of the evolutionary process of viruses and other microorganisms. Such tools are used by the scientists of the Global Viral Forecasting Initiative, headed by Nathan Wolfe. The scientists focus on early detection of new viruses and prevention of pandemics [41].

CONCLUSIONS

The dynamic development of the aforementioned methods used in phylogenetic analysis triggered a breakthrough in the perception of the world around us. Charles Darwin during his travel in the *Beagle* ship sketched the first trees of species relationship, solely on the basis of observed morphological differences. Today, it is known that evolution is a fact, and techniques such as PCR, RFLP, hybridization or sequencing enable its reconstruction. Therefore, Darwin was a man with enormous potential for observation and intuition.

Phylogenetics has been increasingly used in biological and medical sciences, and in clinical research it is increasingly more important. It is used, among other things, in the identification and classification of pathogenic microorganisms, epidemiology, forensics, and the study of the origin and evolution of pathogens. Phylogenetic trees reflect the relationship and direction of the evolution of living organisms. Nowadays, algorithms allow for more and more accurate phylogeny reconstruction. Our knowledge of the living world is constantly increasing and has contributed to the derivation of different models of molecular evolution and methods of tree construction. It is not only important but also essential for a proper understanding of the relationships between organisms.

Virology is currently a multidisciplinary science in which molecular biology, genetics or proteomics play an increasingly significant role. In order to understand properly the functioning of viruses, their world should be addressed in terms of not only the diseases they induce, but also the knowledge of biology, and the replication mechanisms or interactions with host cells. All this is aimed at improving the lives of mankind.

REFERENCES

1. Hall BG. Phylogenetic Trees Made Easy: A How-to Manual, Third Edition. Massachusetts: Sinauer Associates, 2008.
2. Darwin C. On the origin of species. London, 1859.
3. Dayrat B. The Roots of Phylogeny: How Did Haeckel Build His Trees? Syst Biol. 2003; 52(4): 515–527.
4. Baldauf SL. Phylogeny for the faint of heart: a tutorial. Trends Genet. 2003; 19: 345–351.
5. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Mol Evol. 2003; 16: 111–120.
6. Takahata N, Kimura M. A model of evolutionary base substitutions and its application with special reference to rapid change of pseudogenes. Genetics 1981; 98: 641–657.

7. Higgs PG, Attwood TK. Bioinformatics and Molecular Evolution. New Jersey: Wiley-Blackwell, 2004.
8. Bos DH, Posada D. Using models of nucleotide evolution to build phylogenetic trees. Dev Comp Immunol. 2005; 29: 211–227.
9. Whelan S, Lio P, Goldman N. Molecular phylogenetics: state-of-theart methods for looking into the past. Trends Genet. 2001; 17: 262–272.
10. Felsenstein J. Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach. J Mol Evol. 1981; 17: 368–376.
11. Kimura M. Estimation of evolutionary distances between homologous nucleotide sequences. Proc NatL Acad Sci. USA 1981; 78(1): 454–458.
12. Lio P, Goldman N. Models of Molecular Evolution and Phylogeny. Genome Res. 1998; 8: 1233–1244.
13. Tamura K, Nei M. Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial DNA in Humans and Chimpanzees. Mol Biol Evol. 1993; 10(3): 512–526.
14. Criscuolo A, Gascuel O. Fast NJ-like algorithms to deal with incomplete distance matrices. BMC Bioinformatics 2008; 9: 166.
15. Nei M, Roychoudhury AK. Evolutionary Relationships of Human Populations on a Global Scale. Mol Biol Evol. 1993; 10: 927–943.
16. Saitou N, Nei M. The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. Mol Biol Evol. 1987; 4: 406–425.
17. Kolaczkowski B, Thornton JW. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature 2004; 431: 980–984.
18. Sober E. Parsimony in Systematics: Philosophical Issues Annual Review of Ecology and Systematics. Annu Rev Ecol Syst. 1983; 14: 335–357.
19. Steel M, Penny D. Parsimony, Likelihood, and the Role of Models in Molecular Phylogenetics. Mol Biol Evol. 2000; 17: 839–850.
20. Guindon S, Gascuel OA. Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Syst Biol. 2003; 52: 696–704.
21. Holder M, Lewis PO. Phylogeny Estimation: Traditional and Bayesian Approaches. Nat Rev Genet. 2003; 4: 275–284.
22. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. Bioinformatics 1998; 14: 817–818.
23. Posada D. jModelTest: Phylogenetic Model Averaging. Mol Biol Evol. 2008; 25: 1253–1256.
24. Lunter G, Miklos I, Drummond A, Jensen JL, Hein J. Bayesian coestimation of phylogeny and sequence alignment. BMC Bioinformatics 2005; 6: 83.
25. Bernard HU. The clinical importance of the nomenclature, evolution and taxonomy of human papillomaviruses. J Clin Virol. 2005; 32: S1-S6.
26. de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H. Classification of papillomaviruses. Virology 2004; 324: 17–27.
27. Bravo IG, de Sanjose S, Gottschling M. The clinical importance of understanding the evolution of papillomaviruses. Trends Microbiol. 2010; 18: 432–438.
28. Chow LT, Broker TR, Steinberg BM. The natural history of human papillomavirus infections of the mucosal epithelia. Acta Path Micro Im C. 2010; 118: 422–449.
29. Cantaloube JF, Gallian P, Attoui H, Biagini P, De Micco P, de Lamballerie X. Genotype Distribution and Molecular Epidemiology of Hepatitis C Virus in Blood Donors from Southeast France. J Clin Microbiol. 2005; 43: 3624–3629.
30. Forbi JC, Vaughan G, Purdy MA, Campo DS, Xia G, Lilia M, Ganova-Raeva LM, Ramachandran S, Thai H, Khudyakov YE. Epidemic History and Evolutionary Dynamics of Hepatitis B Virus Infection in Two Remote Communities in Rural Nigeria. PLoS One. 2010; 5: 1–14.
31. Holmes EC. The phylogeography of human viruses. Mol Ecol. 2004; 13: 745–756.
32. Mild M, Simon M, Albert J, Mirazimi A. Towards an understanding of the migration of Crimean–Congo hemorrhagic fever virus. J Gen Virol. 2010; 91: 199–207.
33. Metzker ML, Mindell DP, Liu XM, Ptak RG, Gibbs RA, Hillis DM. Molecular evidence of HIV-1 transmission in a criminal case. P Natl Acad Sci. USA 2002; 99: 14292–14297.
34. Smith TF, Waterman MS. The Continuing Case of the Florida Dentist. Science 1992; 256: 1155–1156.
35. Wolfe ND, Dunavan CP, Diamond J. Origins of major human infectious diseases. Nature 2007; 447: 279–283.
36. Dobson AP, Carper ER. Infectious Diseases and Human Population History. Bioscience 1996; 46: 115–126.
37. Wolfe ND, Switzer WM, Carr JK, Bhullar VB, Shanmugam V, Tamoufe U, Prosser AT, Torimiro JN, Wright A, Mpoudi-Ngole E, McCutchan FE, Birx DL, Folks TM, Burke DS, Heneine W. Naturally acquired simian retrovirus infections in central African hunters. Lancet 2004; 363: 932–937.
38. Wolfe ND, Heneine W, Carr JK, Garcia AD, Shanmugam V, Tamoufe U, Torimiro JN, Prosser AT, LeBreton M, Mpoudi-Ngole E, McCutchan FE, Birx DL, Folks TM, Burke DS, Switzer WM. Emergence of unique primate T-lymphotropic viruses among central African bushmeat hunters. P Natl Acad Sci. USA 2005; 102: 7994–7999.
39. Zhu T, Korber BT, Nahmias AJ, Hooper E, Sharp PM, Ho DD. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. Nature 1998; 391: 594–597.
40. Korber B, Gaschen B, Yusim K, Thakallapally R, Kesmir C, Detours V. Evolutionary and immunological implications of contemporary HIV-1 variation. Brit Med Bull. 2001; 58: 19–42.
41. Pike BL, Saylors KE, Fair JN, LeBreton M, Tamoufe U, Djoko CF, Rimoin AW, Wolfe ND. The Origin and Prevention of Pandemics. Clin Infect Dis. 2010; 50: 1636–1640.