

# Role of genomic and proteomic tools in the study of host–virus interactions and virus evolution

Soumen Bhattacharjee

Received: 28 February 2013 / Accepted: 24 July 2013 / Published online: 8 August 2013  
© Indian Virological Society 2013

**Abstract** Viruses have short replication cycles and produce genomic variants within a host, a process that seems to adapt to their specific host and also enable them to infect new hosts. The recent emergence of viral genomic variants from the circulating pool within the host population and re-emergence of the old ones are posing serious threat to agriculture, animal husbandry and humanity as a whole. This review assesses the potential role of genomic and proteomic tools that can monitor not only the course of infection and pathogenesis, but also predict the pandemic or zoonotic epidemic potential of a virus in a previously exposed or immunologically naive biological population.

**Keywords** Viral molecular epidemiology · Viral genomics · Viral proteomics · Codon-usage · Virus-host adaptation

## Introduction

In the recent times, new pandemics are mostly viral in nature. Viruses represent a serious problem faced by humanity, livestock and agriculture. New viruses constantly emerge/re-emerge while old ones evolve to challenge the latest advances in antiviral pharmaceuticals and thus generate tremendous social alarm, sanitary problems, and economical losses in the modern world. Viral

evolution involves rapid mutation, recombination, and reassortments within their genomes, and combinations of these events produce complex and phenotypically diverse populations of viruses, that constitute a raw material on which natural selection acts [12]. Owing to this intrinsic capacity for genetic changes, viruses evolve their fitness level rapidly, and therefore are able to parasitize alternative host species. The recent growth of genome/proteome-based data from both viral pathogens and their hosts can now complement classical epidemiological, clinical, and pathological studies all together to predict virus evolution, host–virus interactions (viral adaptation), interspecies transmissions, and possible emergence of a virulent/pandemic strain [21].

## Viral epidemiology and pandemics

Viral pathogens that are causing serious pandemics or posing to cause one are the *Human immunodeficiency virus 1* (HIV-1) causing AIDS, hepatitis viruses (HBV and HCV) causing chronic hepatitis and hepatocellular carcinoma, *Dengue virus* (DENV) causing Dengue hemorrhagic fever, *SARS coronavirus* (SARS-CoV) causing respiratory disease and Influenza A virus (subtypes H5N1, H1N1, swine origin-Influenza A Virus/SO-IAV) causing flu. Researches into the origin and future of viral pandemics are possibly largely inadequate owing to their complex aetiologies, especially in the case of RNA viruses. For example, flu viruses show seasonal and sudden emergence, explosive growth, and sudden remission [10]. Although we still do not have an effective vaccine, the prospects of controlling HIV/AIDS pandemic appear brighter with the accumulation of genomic data, phylogenetic delineation of its origin and studies on its molecular epidemiology [27].

---

S. Bhattacharjee (✉)  
Cell and Molecular Biology Laboratory, Department of Zoology,  
University of North Bengal, Raja Rammohunpur,  
P.O. North Bengal University, Siliguri 734 013,  
District Darjeeling, West Bengal, India  
e-mail: soumenb123@rediffmail.com

### **Viral epidemiology: from early to current age**

Historically, viral epidemiology has been performed with substantial success by virus isolation through inoculation of tissue cultures, serological methods (complement-fixation, haemagglutination-inhibition and antibody detections) employed both in cross-sectional (covering diverse geographical localities) and longitudinal (within families) surveys. Virus isolation is both costly and a time consuming method which needs special laboratory facilities and expertise. Serological techniques, on the other hand, are sometimes misleading owing to pathogen's cross-species reactivity and ubiquitous presence of antibodies as against co-evolving viruses. Transmission electron microscope (TEM) had been instrumental in morphological characterization and identification of many clinically important viruses in the 1970s and 1980s [48]. However, developments in immunofluorescence imaging techniques, enzyme-linked immunosorbent assays (ELISA), reverse-transcriptase polymerase chain reaction (RT-PCR) have gradually replaced TEM in virus diagnosis in veterinary and human medicine [48]. In the last three decades, detection of viral nucleic acid and phylogenetic analyses based on nucleotide and amino acid sequence divergences have substantially improved the conventional epidemiology and the tracking of viral transmission/evolution of a disease in the human and animal populations. These studies on viral phylogeography revealed the trends of viral evolution, shed light on the most important topics in disease epidemiology: predicting from what reservoir species and locations a new viral infection will emerge and spread in human populations in the future [32].

### **Viral molecular epidemiology**

Molecular phylogenetic studies and virus-host co-divergence analyses have been used to predict the origin, emergence and/or the re-emergence of a viral disease. The evolution of viruses is determined mainly by their relatively small genomes, enormous population sizes, and usually short generation times. Additionally, at least in the RNA viruses, large selection coefficients, antagonistic epistasis and high mutation rates govern the course of their evolution. Virus population ("quasi-species") seems to exist as a dynamic collection of a large genomic variants maintained in equilibrium within the host which is subjected to intense selective forces, a strategy that would give rise to a progeny fitter to infect another host or to adapt to a host more effectively [13].

The study of origin, evolution, molecular epidemiology, phylogenetic/phylogeographic and codivergence analyses in many important human viruses like HIV-1 [19, 29],

*Human Papillomavirus* [40, 42], *Human Polyomavirus* [5, 47, 57], *Influenza A virus* subtype H5N1 [6, 59], *Influenza A virus* subtype H1N1 [43], *Hepatitis B virus* [37, 45], *Hepatitis C virus* [36, 49], *Dengue Virus* [11, 50], *SARS Coronavirus* [34] have been possible due to the development of sensitive and robust molecular detection methods like polymerase chain reaction (PCR), quantitative-PCR (Q-PCR), high-throughput DNA sequencing methods and bioinformatic tools in the past 25 years.

### **Phylogenetic analyses of viral genes and proteins**

The most active and productive area of research in modern evolutionary biology is the study of the origins, emergence, and spread of viral infections in human populations [26]. These phylogenetic analyses, based on nucleotide data, have been more successful in RNA viruses owing to their high mutation rates in their genomes. Although rather less evolutionary work has been undertaken on the human DNA viruses, a great wealth of gene and genome sequence data from DNA viruses is gradually growing that might provide a powerful means by which their phylogenetic history and the mechanisms of evolution can be elucidated [15]. Molecular phylogenetic analyses and dating of origin and divergence of viruses are often inconsistent, being largely dependent on the assumption of the models of nucleotide/amino acid substitution [56]. For elucidating the evolutionary relationships the use of nucleotide sequences especially DNA sequences, is well documented in many works [7, 8, 58]. Virologists require precise estimation of emergence/divergence of viruses and their subtypes to correlate these important events with historical demographic changes, geographical invasions, zoonoses, viral transmission pattern and its adaption to its host, all enabling the development of an effective antiviral strategy. At least for influenza viral subtype divergence there appears to be strong evidence that the process of sub-speciation is associated not just with large genomic changes but also with an accelerated finite process of its adaption to host [44]. Moreover, presently it is also difficult to understand the precise mechanistic basis of evolutionary strategy of RNA viruses, which maintain much higher mutation rates as opposed to the DNA viruses.

There are three general theories of origin of viruses: regressive, cellular and independent origin. Molecular phylogenetic analyses of viral nucleic acid and co-divergence analysis with their respective hosts are probably the only way of testing the contesting hypotheses at present [53]. While relatively ample fossil data of vertebrate hosts exists, co-divergence analysis of vertebrate hosts with that of the nucleotide divergence of their cognate viruses can throw some light on the time of origin of a virus when done

on the basis of a constant molecular clock (constant mutation incorporation by DNA replication machinery). Although co-evolutionary studies have been successful in few small DNA viruses (parvoviruses, polyomaviruses and papillomaviruses), sequence analysis of RNA viruses pose some problem owing to high mutation rates in their genes and rapid emergence of reassortants. Still, predicting the future course of viral epidemiology or emergence of a new virus is difficult because evolution is probabilistic and the issues are complicated by several factors like host ecology, environmental parameters, human migration and land use patterns. Emergence of a new epidemic depends on several factors like, the ability of a virus to infect new host through mutation/reassortment of its genome, extent of contact between the host population and the virus reservoir and on host immunity. However, the genetic mutations that facilitate virus adaptation to new host are likely to reduce its fitness in the donor host [46]. Therefore the mechanisms that determine these “fitness trade-offs” may largely depend on host–virus interaction both at the genomic and proteomic levels.

With the rapid accumulation of genomic sequence data of both hosts and their viruses, and concomitant developments in transcriptomics, microarray-based SNP analysis, the current “omic” age provides a new vista for studying viral evolution, viral epidemiology, host immune responses and disease management, at least in some human RNA viruses (Influenza A, HIV, Dengue and SARS-CoV) [26, 32].

#### **Application of genomics and proteomics in the study of virus evolution and host–virus interactions**

In the age of genomics and proteomics, important questions that demand answers from the virologists perhaps centres around interactions among viruses, intra and inter-host evolutionary trends, genome wide interactions within the viral genomes, ecological interference among different virus groups and virus–host adaptation dynamics [25, 39]. Molecular phylogenetic analyses based on key coding sequences or whole genomes, at least in small DNA tumour viruses are indicating interactions between similar viral genomes [4]. Application of genomic and proteomic tools thus is expected to improve our understanding of viral evolution, to improve prediction of disease outbreaks and subsequent managements, and to develop better vaccines and therapeutics [51].

There can be a spectrum of host–virus interactions, while some viruses infect a broad range of species others infect only a specific host. Three steps are important in viral epidemic or pandemic- (1) introduction of a viral pathogen into a new host species, (2) establishment of the

pathogen in the new host, and (3) dissemination of the pathogen among a large number of individuals of the new host species. All of these steps require adaptation of the virus to its new host in addition to several other modifying factors that allow the virus to get into a stable dynamic relationship with the host [2, 14].

It has been postulated as early as 1961 that nucleotide composition of the genome strongly influences gene sequence evolution [55] and also that organisms evolve unique genetic signature shaped by specific genomic pressures giving rise to specific codon bias. Use of selective genetic codons or differential use of synonymous codons in gene expression is characteristic of a genome in a species [33, 54]. Therefore, analyses of the extent and causes of these unique codon choices and nucleotide composition are essential to understand viral evolution and virus–host interactions [52]. Recently, researches in codon-usage and amino acid preferences in viruses have kindled interesting proposition with regard to fine-tuning of existing virus phylogenies and viral adaptation towards their hosts [3, 17, 31, 52, 60]. It is known at least in *E. coli*, *Saccharomyces* and *Drosophila* [17], that codon-usage determines gene expression at the level of translation [24]. Recent studies on virus codon-usage patterns indicate that mutation pressure may be more important factor than translational selection in determining codon-usage bias [17]. The study of codon-usage patterns in viruses can be rewarding with respect to understanding not only viral gene expression and vaccine development but also virus adaptation, evolution and host–virus interactions [9, 17]. Adaptation to host codon-usage pattern can also help explain both tissue and species specificities of these viruses, and thus cross-species transmissibility [9].

Codon-usage statistics have been used in identifying highly expressed genes in prokaryotes and yeast [16, 18] and have been shown to correlate with evolutionary distances between organisms, while atypical codon-usages seem to indicate horizontal transmission of genes between populations [41]. Translational selection and mutation bias are thought to have jointly shaped the favoured codon-usage in many organisms [23]. It has been shown in many viruses that viral codon-usage profile is similar to their specific hosts or to abundantly expressed host gene, a viral mechanism to adapt to host-specific requirements [22, 35]. Amino acid distribution, distance matrices for similar amino acid distribution, host-versus-viral codon-usage comparisons have been used to delineate the pattern of virus gene expression, evolution of viruses and host–virus interactions. These studies are expected to explain emergence, adaptation and evolution of at least few important viruses [28].

Virus adaptation to host can involve only few non-synonymous mutations within key viral proteins that

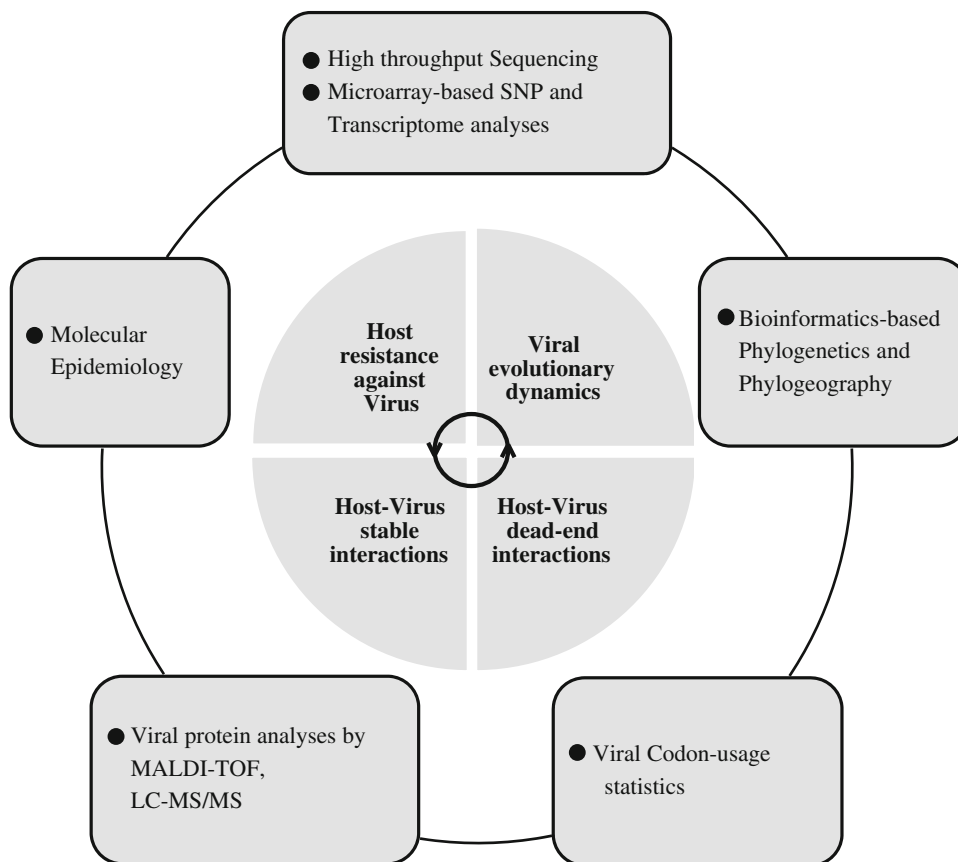
accompany substantial changes in host-gene expression patterns. It has been documented through in vivo experiments and DNA microarray-based host transcriptome analysis [1]. Evolutionary changes within the viral genome and interaction of viral genes with the host transcriptome can reveal the degree of adaptation of the pathogen to its host and to the environmental conditions. This revelation can predict the outcome of an infection of naive or experienced host population with an emergent virus strain. Prevalence of specific amino acid appears to be virus specific. Moreover, there is a preferential usage of only few codons for most prevalent amino acids, e.g., CGC in human viruses and AGA in bacteriophages for Arginine [3] for example. It has been documented that single amino acid change in a key position of H5N1 NS1 protein can markedly increase its pathogenicity through immune modulation in host [30]. Proteome-based studies have indicated that there is an enhanced diversity in amino acid usage among viruses, those that infect humans are not only adapted to human hosts but also show similarities to non-human mammals and non-mammalian vertebrates [3].

Viruses also have a strong pattern of correlation with their host GC content; especially the bacteria, human and rat viruses share strong resemblances with their hosts in their codon-usage profiles [3]. Human viruses rather have

conserved and unique codon-usages. Codon-usage adaptation of the viruses may be determined from the structural proteins; and proteins required for new host recognition show wide variation in codon-usage profiles with that of their hosts. Codon-usage preferences correlate with the abundance of the respective tRNAs, which is expected to affect translational efficiency of the gene product [3]. Thus, codon-usage profile may be correlated to viral gene expression patterns. A host infectivity-shift may occur in a well-adapted virus through a genetic adaptation process within their structural proteins, that are involved in host recognition. Therefore, the knowledge of viral codon-usage pattern and amino-acid preferences can enrich our understanding of viral evolutionary trends, host-switch mechanisms and disease transmissibility.

Proteomic methods like matrix assisted laser desorption ionization-time of flight-mass spectrometry (MALDI-TOF-MS) and liquid chromatography linked tandem mass spectrometry (LC-MS/MS) are now being used to catalogue and characterize viral proteins, viral protein expression pattern in the host cell and compare it with host protein expression patterns, mostly in RNA and in few DNA viruses [20, 38]. In addition to the characterization of viral proteins these studies reported few host protein incorporations within viral architecture, significance of which is still to be understood.

**Fig. 1** Roles of genomic and proteomic tools in the study of host–virus interaction dynamics



Effect of viral infection on host gene expression has been studied by DNA microarray-based transcriptome analysis, which has given us important insights with regard to this complex interplay of viral proteins and host metabolic networks (Fig. 1). If gene-disease association is identified, target human/animal groups can be chosen for either vaccination or prophylaxis [21].

## Conclusion

It is gradually being realized that virus proteome interacts in a complex manner with that of the host proteome and understanding the nature and quantifying the effects of this interactions will be the cornerstone of future viral disease management and vaccine development. Genomic, proteomic and microarray-based transcriptome analysis can elucidate the gene-interacting networks in the virus–host interplay. Thus host-dependent selection pressure and mutation-dependent codon-usage preferences of the viruses will be instrumental in defining proper prognosis and health-care initiatives. Genomic, proteomic and bioinformatic tools, if applied to complement traditional epidemiology, can substantially improve viral epidemiology, trace geographic diffusion of a pathogenic and potentially hazardous virus and can play an important role in monitoring not only the course of an infection and pathogenesis but also in predicting the zoonotic epidemic potential of a virus in a previously exposed or immunologically naive biological population.

**Acknowledgments** I thank Dr. Ranjan Ghosh, Assistant Professor, Department of English, University of North Bengal for extensive reviewing and editing the text.

## References

- Agudelo-Romero P, Carbonell P, Perez-Amador MA, Elena SF. Virus adaptation by manipulation of host's gene expression. *PLoS One*. 2008. doi: [10.1371/journal.pone.0002397](https://doi.org/10.1371/journal.pone.0002397).
- Antia R, Regoes RR, Koella JC, Bergstrom CT. The role of evolution in the emergence of infectious diseases. *Nature*. 2003;426:658–61.
- Bahir I, Fromer M, Prat Y, Linial M. Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol Syst Biol*. 2009;5:311. doi: [10.1038/msb.2009.71](https://doi.org/10.1038/msb.2009.71).
- Bhattacharjee S. Evolutionary interrelationships among polyomaviruses based on nucleotide and amino acid variations. *Indian J Biotech*. 2010;9:252–64.
- Bhattacharjee S, Chakraborty T. High reactivation of BK virus variants in Asian Indians with renal disorders and during pregnancy. *Virus Genes*. 2004;28:157–68.
- Buchy P, Fourment M, Mardy S, Sorn S, Holl D, Ly S, Vong S, Enouf V, Peiris JSM, van der Werf S. Molecular epidemiology of clade 1 influenza A viruses (H5N1), southern Indochina peninsula, 2004–2007. *Emerg Infect Dis*. 2009;15:1641–4.
- Cann RL, Stoneking M, Wilson AC. Mitochondrial DNA and human evolution. *Nature*. 1987;325:31–6.
- Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duval RR, Price R, Hillis HG, Qui Y-L, Kron KA, Rettig JH, Conti E, Palmer JD, Manhart JR, Sytsma KJ, Michaels HJ, Kress WJ, Karol KG, Clark WD, Hedren M, Gaut BS, Jansen RK, Kim K-J, Wimpee CF, Smith JF, Furnier GR, Strauss SH, Xiang Q-Y, Plunkett GM, Soltis PS, Williams SE, Gadek PA, Quinn CJ, Eguiarte LE, Golenberg E, Learn GH, Graham S, Barrett SCH, Dayanandan S, Albert VA. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann Missouri Bot Garden*. 1993;80:528–80.
- Cladel NM, Bertotto A, Christensen ND. Human alpha and beta papillomaviruses use different synonymous codon profiles. *Virus Genes*. 2010;40:329–40.
- De Jong JC, Claas EC, Osterhaus ADME, Webster RG, Lim WL. A pandemic warning? *Nature*. 1997;389:554.
- De Paula SO, Pires Neto RJ, Correa JA, Assumpcao SR, Costa ML, Lima DM, Fonseca BA. The use of reverse transcription-polymerase chain reaction (RT-PCR) for the rapid detection and identification of dengue virus in an endemic region: a validation study. *Trans R Soc Trop Med Hyg*. 2002;96:266–9.
- Domingo E. Mechanisms of viral emergence. *Vet Res*. 2010; 41(3):8. doi: [10.1051/vetres/201001](https://doi.org/10.1051/vetres/201001).
- Eigen M, McCaskill J, Schuster P. The molecular quasi-species. *Adv Chem Phys*. 1989;75:149–263.
- Elena SF, Sanjuan R. Virus evolution: insights from an experimental approach. *Annu Rev Ecol Evol Syst*. 2007;38:27–52.
- Esposito JJ, Sammons SA, Frace AM, Osborne JD, Olsen-Rasmussen M, Zhang M, Govil D, Damon IK, Kline R, Laker M, Li Y, Smith GL, Meyer H, Leduc JW, Wohlhueter RM. Genome sequence diversity and clues to the evolution of variola (smallpox) virus. *Science*. 2006;313:807–12.
- Freire-Picos MA, Gonzalez-Siso MI, Rodriguez-Belmonte E, Rodriguez-Torres AM, Ramil E, Cerdan ME. Codon usage in *Kluyveromyces lactis* and in yeast cytochrome *c*-encoding genes. *Gene*. 1994;139:43–9.
- Fu M. Codon usage bias in herpesvirus. *Arch Virol*. 2010;155:391–6.
- Gharbia SE, Williams JC, Andrews DMA, Shah HN. Genomic clusters and codon usage in relation to gene-expression in oral Gram-negative anaerobes. *Anaerobe*. 1995;1:239–62.
- Gifford RJ, de Oliveira T, Rambaut A, Pybus OG, Dunn D, Vandamme A-M, Kellam P, Pillay D. Phylogenetic surveillance of viral genetic diversity and the evolving molecular epidemiology of human immunodeficiency virus type 1. *J Virol*. 2007;81: 13050–6.
- Go EP, Wikoff WR, Shen Z, O'Maille G, Morita H, Conrads TP, Nordstrom A, Trauger SA, Uritboonthai W, Lucas DA, Chan KC, Veenstra TD, Lewicki H, Oldstone MB, Schneemann A, Siuzdak G. Mass spectrometry reveals specific and global molecular transformations during viral infection. *J Proteome Res*. 2006;5: 2405–16.
- Haagmans BL, Andeweg AC, Osterhaus ADME. The application of genomics to emerging zoonotic viral diseases. *PLoS Pathog*. 2009;5:e1000557. doi: [10.1371/journal.ppat.1000557](https://doi.org/10.1371/journal.ppat.1000557).
- Haas J, Park E-C, Seed B. Codon usage limitation in the expression of HIV-1 envelope glycoprotein. *Curr Biol*. 1996;6: 315–24.
- Hershberg R, Petrov DA. General rules for optimal codon choice. *PLoS Genet*. 2009. doi: [10.1371/journal.pgen.1000556](https://doi.org/10.1371/journal.pgen.1000556).
- Holm L. Codon usage and gene expression. *Nucleic Acids Res*. 1986;14:3075–87.
- Holmes EC. Viral evolution in the genomic age. *PLoS Biol*. 2007;5:e278.
- Holmes EC. Evolutionary history and phylogeography of human viruses. *Annu Rev Microbiol*. 2008;62:307–28.

27. Holmes EC. RNA virus genomics: a world of possibilities. *J Clin Invest*. 2009;119:2488–95.
28. Holmes EC. The comparative genomics of viral emergence. *Proc Natl Acad Sci USA*. 2010;107:1742–6.
29. Hughes GJ, Fearnhill E, Dunn D, Lycett SJ, Rambaut A, Brown AJL. Molecular phylodynamics of the heterosexual HIV epidemic in the United Kingdom. *PLoS Pathog*. 2009;5:e1000590. doi:10.1371/journal.ppat.1000590.
30. Jiao P, Tian G, Li Y, Deng G, Jiang Y, Liu C, Liu W, Bu Z, Kawaoka Y, Chen H. A single-amino-acid substitution in the NS1 protein changes the pathogenicity of H5N1 avian influenza viruses in mice. *J Virol*. 2008;82:1146–54.
31. Karlin S, Blaisdell BE, Schachtel GA. Contrasts in codon usage of latent versus productive genes of Epstein-Barr virus: data and hypotheses. *J Virol*. 1990;64:4264–73.
32. Kilpatrick M, Daszak P, Goodman SJ, Rogg H, Kramer LD, Cenedo V, Cunningham AA. Predicting pathogen introduction: West Nile Virus spread to Galapagos. *Conserv Biol*. 2006;20:1224–31.
33. Kurland CG. Codon bias and gene expression. *FEBS Lett*. 1991;285:165–9.
34. Lipsitch M, Cohen T, Cooper B, Robins JM, Ma S, James L, Gopalakrishna G, Chew SK, Tan CC, Samore MH, Fisman D, Murray M. Transmission dynamics and control of severe acute respiratory syndrome. *Science*. 2003;300:1966–70.
35. Lobo FP, Mota BEF, Pena SDJ, Azevedo V, Macedo AM, Tauch A, Machado CR, Gloria R, Franco GR. Virus–host coevolution: common patterns of nucleotide motif usage in Flaviviridae and their hosts. *PLoS One*. 2009. doi: 10.1371/journal.pone.0006282.
36. Magiorkinis G, Magiorkinis E, Paraskevis D, Ho SYW, Shapiro B, Pybus OG, Allain J-P, Hatzakis A. The global dynamics and phylogeography of hepatitis C virus 1a and 1b. *PLoS Med*. 2009;6:e1000198. doi:10.1371/journal.pmed.1000198.
37. Magnus LO, Norder H. Subtypes, genotypes and molecular epidemiology of the hepatitis B virus as reflected by sequence variability of the S-gene. *Intervirology*. 1995;38:24–34.
38. Maxwell KL, Frappier L. Viral proteomics. *Microbiol Mol Biol Rev*. 2007;71:398–411.
39. McHardy AC, Adams B. The role of genomics in tracking the evolution of influenza A virus. *PLoS Pathog*. 2009;5:e1000566. doi:10.1371/journal.ppat.1000566.
40. McKaig RG, Baric RS, Olshan AF. Human papillomavirus and head and neck cancer: epidemiology and molecular biology. *Head Neck*. 1998;20:250–65.
41. Medigue C, Rouxel T, Vigier P, Henaut A, Danchin A. Evidence for horizontal gene transfer in *Escherichia coli* speciation. *J Mol Biol*. 1991;222:851–6.
42. Menton JF, Cremin SM, Canier L, Horgan M, Fanning LJ. Molecular epidemiology of sexually transmitted human papillomavirus in a self-referred group of women in Ireland. *Viol J*. 2009;6:112. doi:10.1186/1743-422X-6-112.
43. Novel Swine-Origin Influenza A (H1N1) Virus Investigating Team. Emergence of a novel swine-origin influenza A (H1N1) virus in humans. *N Engl J Med*. 2009;360:2605–15.
44. O'Brien JD, She Z-S, Suchard MA. Dating the time of viral subtype divergence. *BMC Evol Biol*. 2008;8:172. doi:10.1186/1471-2148-8-172.
45. Okamoto H, Tsuda F, Sakugawa H, Sastrosoewignjo RI, Imai M, Miyakawa Y, Mayumi M. Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. *J Gen Virol*. 1988;69:2575–83.
46. Parrish CR, Holmes EC, Morens DM, Park E-C, Burke DS, Calisher CH, Laughlin CA, Saif LJ, Daszak P. Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol Mol Biol Rev*. 2008;72:457–70.
47. Perez-Losada M, Christensen RG, McClellan DA, Adams BJ, Viscidi RP, Demma JC, Crandall KA. Comparing phylogenetic codivergence between polyomaviruses and their hosts. *J Virol*. 2006;80:5663–9.
48. Roingard P. Viral detection by electron microscopy: past, present and future. *Biol Cell*. 2008;100:491–501.
49. Romano M, de Carvalho-Mello IMVG, Jamal LF, de Melo FL, Iamarino A, Motoki M, Pinho JRR, Holmes EC, de Andrade Zanolto PM, VGDN Consortium. Social networks shape the transmission dynamics of hepatitis C virus. *PLoS One*. 2010;5:e11170. doi:10.1371/journal.pone.0011170.
50. Schreiber MJ, Holmes EC, Ong SH, Soh HSH, Liu W, Tanner L, Aw PPK, Tan HC, Ng LC, Leo YS, Low JGH, Ong A, Ooi EE, Vasudevan SG, Hibberd ML. Genomic epidemiology of a dengue virus epidemic in urban Singapore. *J Virol*. 2009;83:4163–73.
51. Seib KL, Dougan G, Rappuoli R. The key role of genomics in modern vaccine and drug design for emerging infectious diseases. *PLoS Genet*. 2009;5:e1000612. doi:10.1371/journal.pgen.1000612.
52. Shackelton LA, Parrish CR, Holmes EC. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *J Mol Evol*. 2006;62:551–63.
53. Shadan FF, Villarreal LP. Coevolution of persistently infecting small DNA viruses and their hosts linked to host-interacting regulatory domains. *Proc Natl Acad Sci USA*. 1993;90:4117–21.
54. Shields DC, Sharp PM. Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational biases. *Nucleic Acids Res*. 1987;15:8023–40.
55. Sueoka N. Compositional correlation between deoxyribonucleic acid and protein. *CSH Symp Quant Biol*. 1961;26:35–43.
56. Susko E, Inagaki Y, Roger AJ. On inconsistency of the neighbor-joining method, least squares and minimum evolution estimation when distances are incorrectly specified. *Mol Biol Evol*. 2004;21:1629–42.
57. Whitley DM, Mackay IM, Sloots TP. Detection and differentiation of human polyomaviruses JC and BK by light cycler PCR. *J Clin Microbiol*. 2001;39:4357–61.
58. Woese CR. Bacterial evolution. *Microbiol Rev*. 1987;51:221–71.
59. Xu X, Subbarao K, Cox NJ, Guo Y. Genetic characterization of the pathogenic influenza A/Goose/Guangdong/1/96 (H5N1) virus: similarity of its hemagglutinin gene to those of H5N1 viruses from the 1997 outbreaks in Hong Kong. *Virology*. 1999;261:15–9.
60. Zhao K-N, Liu WJ, Frazer IH. Codon usage bias and A+T content variation in human papillomavirus genomes. *Virus Res*. 2003;98:95–104.