Review article

Mechanisms of viral emergence

Esteban DOMINGO*

Centro de Biología Molecular Severo Ochoa (CSIC-UAM), C/Nicolas Cabrera, 1, Cantoblanco, Madrid 28049,
Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd),
Barcelona, Spain

Abstract – A number of virologic and environmental factors are involved in the emergence and re-emergence of viral disease. Viruses do not conservatively occupy a single and permanent ecological niche. Rather, due to their intrinsic capacity for genetic change, and to the evolvability of fitness levels, viruses display a potential to parasitize alternative host species. Mutation, recombination and genome segment reassortment, and combination of these molecular events, produce complex and phenotypically diverse populations of viruses, which constitute the raw material on which selection acts. The majority of emerging viral diseases of humans have a zoonotic origin. Sociologic and ecologic factors produce diverse and changing environments in which viral subpopulations have ample opportunities to be selected from intrinsically heterogeneous viral populations, particularly in the case of RNA viruses. In this manner, new human, animal and plant viruses have emerged periodically and, from all evidence, will continue to emerge. This article reviews some of the mechanisms that have been identified in viral emergence, with a focus on the importance of genetic variation of viruses, and on the general concept of biological complexity.

Table of contents

1. INTRODUCTION

Very few topics in Virology relate so closely to the general concept of biological complexity as the emergence and re-emergence of viral disease. In the introduction to their classic book, Solé and Goodwin define the sciences of complexity as "the study of those systems in which there is no simple and predictable relationship between levels, between the properties of parts and of wholes" [78]. The emergence of viral disease involves several levels of complexity. The underlying level stems from the population structure of viral populations as they replicate in their standard hosts. Model studies of plaque-to-plaque transfers (bottleneck

---

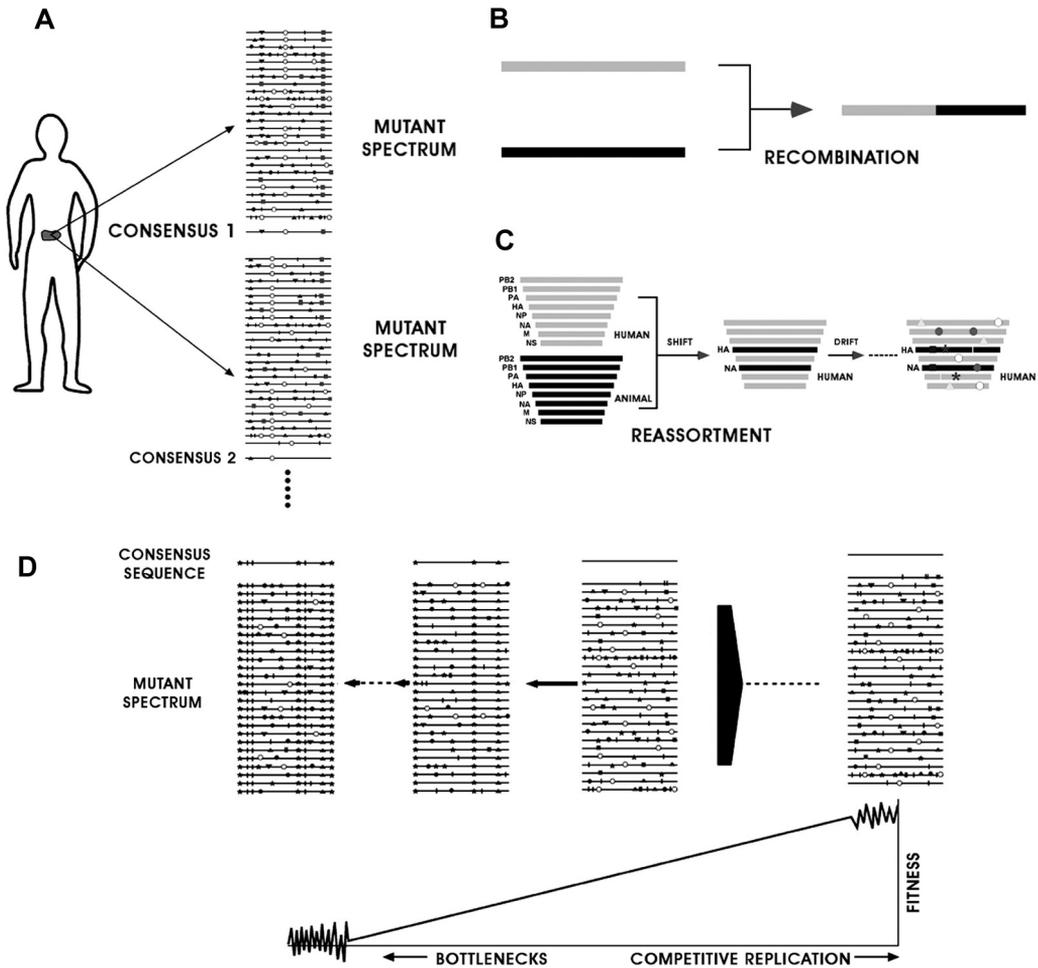* Corresponding author: edomingo@cbm.uam.es

Figure 1. Schematic representation of genomic variations and population dynamics of RNA viruses. (A) In infected hosts RNA viruses replicate as complex distributions of related genomes termed viral quasispecies. Here two out of multiple distributions are represented. Horizontal lines depict individual genomes and symbols on the lines represent different mutation types (transitions, transversions and short insertions or deletions termed indels). A distribution is defined by a consensus sequence and a mutant spectrum with a complexity given by the average pairwise genetic (also termed Hamming) distance among its components or the average mutation frequency. (B) Molecular recombination. (C) Genome segment reassortment, using influenza A virus (eight genomic segments) as an example. Reassortment (in this case the replacement of HA and NA genes) gives rise to an antigenic shift. Continued accumulation of mutations results in gradual antigenic drift. (D) A simplified view of quasispecies dynamics and fitness change. Unrestricted replication (large black arrow-head on the right, with multiple passages indicated by the dotted line) results in fitness gain, as depicted by the triangle at the bottom. Fitness gain can occur without variation of the consensus sequence (top). In contrast, repeated bottleneck transfers (left, with the dotted line representing multiple transfers) result in accumulation of mutations that modify the consensus sequences, and in fitness decrease. At low and high fitness values significant fluctuations of fitness values have been observed. This figure is based on previously published data and concepts [4, 9, 11–14, 16, 17, 19, 21, 25, 29, 31, 35, 41, 57, 59–61, 66, 71, 74].

passes) of foot-and-mouth disease virus (FMDV) in BHK-21 cells (Fig. 1) showed that the pattern of fitness decay of the virus followed a Weibull distribution [42]. This type of statistical distribution suggests that the mutations fixed in the viral genome at each transfer produced a cascade of perturbations in the virus-host interactions that were sensed in the form of a change in virus yield [42]. The unpredictability of the effect of mutations is further reinforced by the increasing evidence that viral proteins are multi-functional so that mutations can alter one or more of the interactions between viral and host components that determine the viral yield per cell [42].

The second level of complexity results from a network of environmental, ecologic, and sociologic influences that affect the probability that a potentially pathogenic virus comes into contact with a new host. A good number of such influences are subjected to indetermination. To give a specific example, a change in temperature and humidity in a large geographical area may alter flora and fauna, and, as a result, the distribution of viral vectors (arthropods, birds, mammals). However, which vectors and how they will be affected by weather or climate conditions are usually difficult to predict since they depend on sets of ecological interactions. Finally, another level of complexity intervenes in the epidemiological outcome of a viral disease, once it has emerged. Severe acute respiratory syndrome (SARS) faded months after its emergence, while Acquired Immune Deficiency Syndrome (AIDS) has become a severe pandemics, to last for many decades or forever. The difference can be attributed to the pattern of virus-host interaction and the capacity of the virus to be transmitted from an infected host individual into a susceptible individual. Again, these processes fall into the domain of biological complexity. Advances in molecular virology and the development of computational tools for viral surveillance have greatly improved our knowledge of the factors that underlie the emergence and re-emergence of viral disease, and, interestingly and paradoxically, we understand some of the reasons behind the unpredictability of emergences.

2. BASIC FEATURES OF VIRUSES AT THE POPULATION LEVEL. FITNESS

Viral disease emergence is a major concern in public health, and the events behind their occurrence have been extensively analyzed [1, 33, 55, 69, 77]. Although in most published accounts on viral disease emergence the role of genetic variation of viruses has been considered, the majority of studies have examined factors that affect viral traffic, and have described many examples of inter-species transmissions. Few treatments have dissected intra-host genetic variation of viruses as a key element in virus adaptability and, therefore, in virus disease emergence.

Viral emergences and re-emergences can be regarded as episodes of virus adaptation to a new environment. Adaptation can be measured as an increase of replicative fitness, defined as the capacity of a virus to produce infectious progeny in a given environment [11, 14, 71]. The term epidemiologic fitness describes in semiquantitative ways (for example, through diagnostic surveys, by sampling nucleotide sequences of viruses that compete in a given geographical area, etc.) the capacity of a virus (a serotype, clade or variant) to become dominant, relative to other serotypes, clades or variants of the same virus [14]. For example, a FMDV of serotype O termed FMDV O PanAsia was first isolated in India in 1990 and then, in the following years, the virus spread to several countries. FMDV PanAsia manifested high epidemiologic fitness, since it displaced other FMDV strains that circulated at that time in the same geographical areas [39]. A more recent example is provided by the new H1N1 human influenza virus that originated in swine, crossed the species barrier into the human population early in 2009, and is currently displacing the H1N1 and H3N1 viruses that were dominant at the time of irruption of the new H1N1 [74]. Replicative fitness is a relevant concept in evolutionary biology in general, and it has become very important to quantitate virus adaptability in numerous scenarios, including clinical and diagnostic virology (as reviews, see [11, 14, 71]) (Fig. 1).

What are the molecular mechanisms underlying fitness changes in viral populations? Many studies with human, animal and plant viruses suggest that fitness variations are associated with genetic change, mediated by mutation, recombination and genome segment reassortment during viral replication (Fig. 1). Mutation is a universal mechanism of genetic variation that affects all viruses, while recombination appears to vary greatly in frequency among different viruses. Segment reassortment is obviously restricted to viruses with segmented genomes. It must be emphasized that fitness values, as well as the fitness effects of genetic variation, are dependent on the sequence context of a viral genome and on the physical and biological environments in which the virus replication capacity is tested. Next I discuss how genetic change underlies fitness variations and how such variations unpredictably affect the adaptation of viruses to new environments.

3. BIOLOGICAL IMPLICATIONS OF MUTATION, RECOMBINATION AND REASSORTMENT. QUASISPECIES AS RAW GENOMIC DISTRIBUTIONS FOR SELECTION OF VIRAL SUBPOPULATIONS

Viruses can be divided into those that have DNA and those that have RNA as genetic material. Relevant with regard to the occurrence of mutations during viral genome replication is the distinction among four replicative schemes that can be schematically described as follows:

1. DNA → DNA → DNA
2. DNA → RNA → DNA
3. RNA → RNA → RNA
4. RNA → DNA → RNA

The nucleic acid written in the first and third place in each of the four schemes is the one found in the parental and progeny viral particles. The nucleic acid in the middle position indicates the type of nucleic acid that acts as replicative intermediate (not the transcripts involved in gene expression that participate in the infectious cycle). The presence or not of RNA in the scheme determines the potential of genetic variation of the virus. RNA viruses and DNA viruses that have RNA as replicative intermediate (that is, viruses that follow replicative schemes 2, 3 and 4) display elevated mutation rates that have been estimated in 10^{-3} to 10^{-5} misincorporations per nucleotide copied [4, 14, 17]. The molecular basis for high mutation rates is the absence of a proofreading-repair activity in the viral RNA-dependent RNA polymerases (RdRp) and RNA-dependent DNA polymerases (RdDp) (also termed reverse transcriptases (RT)), evidenced by functional and structural studies [24, 52, 79]. There appear to be some exceptions to this rule, as suggested by the capacity of the influenza virus polymerase to remove excess GMP residues added to a capped oligonucleotide primer [37], a 3′-end repair mechanism described in a satellite RNA of a plant virus [56], and a 3′-5′ exonuclease activity encoded by some coronaviruses [18, 54]. However, it has not been established whether the decrease in mutation rate promoted by such mechanisms has a major consequence for the biology of the corresponding viruses. In addition to the general absence of proofreading-repair activities in RdRps and RdDps, the cellular post-replicative-repair pathways that act to correct mismatches in double-stranded DNA are not effective in repairing either double-stranded RNA or RNA-DNA hybrids [26]. Lack of proofreading-repair and post-replicative repair activities contributes decisively to high mutation rates of riboviruses, retroviruses and hepadnaviruses, a feature that has a profound influence in their biology [14].

DNA viruses with large genomes (poxviruses, iridoviruses, adenoviruses) encode DNA polymerases endowed with a proofreading 3′-5′ exonuclease activity, similar to those found in cellular, replicative DNA-dependent DNA polymerases (DdDp). Therefore, such DNA viruses are not expected to display the general high mutation rates typical of RNA viruses. However, it would be important to measure mutation rates for DNA viruses in order to compare the values with those available for RNA viruses. Relevant factors that can influence the mutation rates displayed by DNA viruses are

Table I. Some biological implications of the quasispecies nature of RNA viruses.

1. Viral genomes are collections of mutant genomes termed mutant spectra or mutant clouds. Mutant clouds may include phenotypic variants adequate to respond to selective constraints (antibody- and cytotoxic T cell-escape mutants, inhibitor- or mutagen-resistant mutants; cell tropism and host-range mutants, etc.). The phenotypic repertoire of a viral quasispecies can contribute to viral persistence, pathogenesis and to the limited efficacy of treatments designed to limit viral replication.
2. Viral quasispecies can include memory genomes as minority components of their mutant spectra. Memory provides an adaptive advantage to viral populations.
3. Mutant spectra are not merely collections of mutant viruses acting independently. Positive interactions (of complementation) or negative interactions (of interference) can be established within mutant spectra. Thus viral quasispecies act as a unit of selection and cannot be accurately described by classical Wright-Fisher formulations of population genetics.
4. The understanding of quasispecies dynamics has helped defining protocols for preventive and therapeutic designs (vaccines to control viral quasispecies must be multivalent, antiviral agents must be sued in combination) and has impelled new antiviral strategies such as lethal mutagenesis (virus extinction by excess mutations).

Based in references [2, 3, 6, 8, 11–16, 19, 25, 34, 36, 49, 58, 59].

whether the viral genome DNA is replicated by a high or low fidelity DNA polymerase, and whether post-replicative repair pathways reach the viral replication sites, and can cope with the correction of errors that may occur in many nascent DNA genomes in viral replication complexes where viral DNA accumulates. These questions have not been solved [14].

High mutation rates of RNA viruses and some DNA viruses result in dynamic distributions of mutants termed viral quasispecies (Fig. 1). Viral quasispecies constitute reservoirs of genetic and phenotypic variants, with several biological implications for viral evolution and pathogenesis, some of which have become apparent over the last few years (reviews in [13, 14, 16, 25]) (Tab. I). The main conceptual extension that quasispecies has represented for RNA virus genetics is that the "wild type" can no longer be assigned to a specific genomic nucleotide sequence but to a distribution of sequences. Mutation is not an occasional event but a continuously occurring event. When the consensus sequence is seen as invariant it is not because mutations did not occur but because they occurred in components of mutant spectrum without a change in the average or consensus sequence (Fig. 1). This is a key departure from classic genetics, and one of the fuels for viral disease emergences, since viral

quasispecies may provide viral subpopulations ready to be selected in response to an environmental change, for example, an alternative host. As discussed below, the multi-phenotypic potential inherent to viral quasispecies represents a major driving force for viral disease emergence. In some cases, however a mutation reflected in the consensus (epidemic VEEV amplification in horses, expansion of influenza host range, etc.) may also serve as determinant of an emergence.

Recombination occurs both with RNA and DNA viruses, but it is infrequent with most negative-strand RNA viruses. Recombination has been divided in replicative and non-replicative, according to whether viral genome replication is required or not [9, 29]. From the point of view of evolution, recombination has been viewed as a means to rescue fit viral genomes from low fitness parents (a force to eliminate deleterious mutations) or a means to produce highly divergent genomes that provide an opportunity to explore the adaptative potential of rare genomic combinations (a force for large evolutionary transitions). There is good evidence that recombination and reassortment events, followed by fitness adjustments through mutation, have played a role in the emergence and re-emergence of new viral pathogens. In fact, most (but not all) emergences

and re-emergences of human viral diseases, that have amounted to about one per year during the last decades, have been associated with RNA viruses that display active recombination or reassortment. As examples, several recent poliomyelitis outbreaks have been related to viruses generated by recombination between attenuated vaccine polioviruses and other circulating enteroviruses [75]. As an historical event, the coronavirus mouse hepatitis virus appears to have acquired its hemagglutinin-esterase gene by recombination with an influenza type C virus [46]. Also, the alphavirus Western equine encephalitis virus probably resulted from an ancient recombination between a Sindbis-like and a Easter-equine encephalitis-like virus (reviewed in [87]). The emergence of the SARS coronavirus around 2002 probably involved the transfer of the virus from a bat to a carnivore (probably a civet cat); contact with civets may have facilitated the human outbreak. The human ACE2 receptor-binding sequences used by the human SARS coronavirus may have been acquired by recombination with a preexisting group 1 human coronavirus [45]. In this, and in the other cases in which recombination played a role in the generation of a new pathogen, selection of virus variants for a more efficient replication in the new host (for receptor use or other) probably took place [91].

Genome segment reassortment is a major evolutionary force for some viruses with a segmented genome, the most studied case being reassortments in the influenza type A viruses. Reassortants that have acquired the hemagglutinin and/or neuraminidase genes from viruses of some animal reservoirs (notably birds or swine) have been associated with the major human influenza pandemics such as the Spanish influenza of 1918 (H1N1 virus), the Hong Kong influenza of 1968 (H3N2 virus) or the pandemics of 2009 (H1N1 virus) [35, 66, 74]. Such viruses acquired a selective advantage relative to previously circulating influenza viruses because the new hemagglutinin and/or neuraminidase surface antigens were not recognized (or were only poorly recognized) by the neutralizing antibodies present in the human population. The acquisition of new genome segments that produce a change in antigenic specificity is termed antigenic shift. The gradual antigenic modification undergone by the virus as it circulates in the human population is termed antigenic drift. The latter is linked to point mutations that affect the antigenic determinants, and can be regarded as the fitness adjustment by mutation as the virus explores new environments (some of which include anti-influenza antibodies) following a reassortment event. Drastic and gradual versions of antigenic variation are not exclusive of influenza virus, and they have been amply recognized as a major factor in the adaptation of viruses to host organisms endowed with an immune system [48, 67]. Viruses are undoubtedly subjected to selective constraints additional to the several components of an immune response, and some of them probably have not even been identified. Multiple constraints contribute to shape the genomic compositions of the viral populations as they complete successive life cycles. Mutation, recombination and reassortment can produce new viral forms, a minority of which might be competent to replicate in a new environment. There is evidence that a true recombinant hemagglutinin might have been one of the virulence determinants of the influenza pandemics of 1918 [28].

The mechanisms of genetic variation of viruses are blind forces (mutation, recombination and gene transfer events) that have acted historically to build the biosphere we know, under the guidance of natural selection. Eigen and Biebricher adapted the concept of sequence space, originally formulated by Maynard Smith for proteins [50], to describe all the possible nucleotide sequences that a given replicating entity can have [19]. The theoretical sequence space of any organism or any virus is huge (4^n), being n the number of genomic nucleotides. Only a minute portion of the theoretical sequence space is compatible with virus infectivity and survival. Mutation prompts exploration of neighbor sites in sequence space, while recombination offers the possibility to explore distant sites in sequence space. Non-homologous recombination events (that is, those that occur between genomes or genomic sites that do not show nucleotide sequence identity) constitute a means to promote exchanges

of genetic material among organisms that belong to distant phyla, including exchanges between viruses and cells. Recent developments in genomics are unveiling surprising relationships among very distant organisms, giving rise to a tree of life with inter-kingdom connections due to lateral gene transfers, that were unsuspected just a few decades ago [7].

Interestingly, despite the obvious impact of recombination for viral evolution [14, 41, 57], some negative-strand RNA viruses do not seem to display an active recombination, at least among standard infectious genomes. These viruses show, however, an extremely high adaptive potential comparable to that of other RNA viruses that display an active recombination [61]. This suggests that the viruses that have survived to the present time, and probably also their ancestors, have redundant adaptive mechanisms to respond to the multiple selective constraints that their hosts have evolved to limit their replication. All mechanisms of genetic variation of viruses exert a key influence in viral disease emergence, as discussed next.

4. STEPS IN VIRAL DISEASE EMERGENCE AND THE ROLE OF VIRAL VARIATION

A viral emergence is generally defined as the appearance of a new pathogen for a host, such as human immunideficiency virus (HIV)-1 for humans in the twentieth century. Viral re-emergence often refers to the reappearance of a viral pathogen after a period of absence, such as the periodic human influenza epidemics or pandemics. Three major steps have been distinguished in the processes of viral disease emergence or re-emergence: (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 to bring about outbreaks, epidemics or pandemics. Viral genome adaptability has been listed as one of 13 major influences that determine the emergence of infectious disease [1, 55, 69, 77]. The influences cited by Smolinski et al. include modifications of climate, weather and ecosystems,

human demographics and behavior, international travel and commerce, poverty, social inequality, war and famine, and lack of political will. Peters has emphasized the role of viral variation by expressing viral emergence in the following terms: (viral variation) + (changing ecology) + (travel and transport) = (increased opportunities for viral emergence) [69].

It is not easy to specify at which stage of an emergence or re-emergence process virus variation plays a role and to what extent virus variation is essential or accessory. However, the presently available evidence suggests that viral genome variation participates in each of the processes that culminate in dissemination of the viral pathogen among individuals of the newly colonized host. Concerning the introduction step, current knowledge of viral population dynamics and of virus-host interactions suggests that chance encounters between viruses and potential new hosts must occur frequently. Consider the many areas of the world in which humans have regular contacts with animals or vectors that carry viruses (farm animals, mosquitoes, etc.) [89]. Most likely, only a tiny minority of such chance encounters gives rise to an emergence or re-emergence of a viral disease.

Regarding the role of genetic variation, there are several possibilities to explain a successful introduction of a virus into a new host species. The main ones are: (1) the virus is competent to replicate equally well in the donor host species and in the recipient new host species. No genetic variation for the change of host range is needed. The reason why the emergence did not occur before is that either the encounter between the two hosts had not occurred until then, or that a threshold dose of virus needed for the successful introduction had not been attained in past encounters. Such a threshold amount of virus can be imposed by limited virus particle stability, the immune response of the recipient host, and other factors. (2) The virus is not efficient in replicating in the new host, and genetic variation of the virus must intervene to achieve a successful introduction. In this case, again, several factors come into play. They are briefly described here solely to emphasize the complex and unpredictable

nature of a successful introduction step. Mutagenesis is a highly stochastic event subject to the quantum-mechanic uncertainties associated with formation of infrequent tautomeric forms, or other alterations of nucleotides that contribute to misincorporation events [5, 26]. Although in RNA virus genetics it is often considered that mutation rates are high enough not to be the limiting factor for adaptability, there is a caveat in this argument when applied to a disease emergence that requires modification of the host range of a virus. In general, the generation of a specific mutant type will not be a limiting factor provided that: (i) the standard mutation rates that have been measured for RNA viruses operate [4, 14, 17] (discussed in the previous section), (ii) the population size of the replicating virus is sufficiently large, and (iii) that once the required mutation has occurred, the mutant virus maintains a sufficient fitness level, in the context of the surrounding mutant spectrum. However, a natural chance contact between an infected donor host and a potential recipient of another host species is unlikely to involve a large viral population size. Under such conditions, the occurrence of the mutations required for a change in host range may become a limiting factor for the introduction phase of disease emergence. In some designed experiments of adaptation of a virus in a new host species it has been shown than an initial inefficient replication in a recipient host is followed by fitness increase mediated by point mutations that occurred gradually in the mutant spectrum of the viral quasispecies evolving in vivo [64]. The barriers imposed by the host to a new viral pathogen may render establishment unlikely. The required viral genetic change must lie within the range achievable by mutation, recombination or reassortment (or their combinations) in the given environmental context.

Once established, additional adaptive mutations are likely to occur while the virus completes its infectious cycle in the target cells and organs of the newly invaded host. There have been many descriptions that viruses replicating either in cell culture or in vivo, when given the opportunity, increase their replicative fitness [21, 59, 60]. Fitness increase is the expected response when a virus maintains replicative competence in a new environment, or when it escapes from natural (neutralizing antibodies or cytotoxic T cells) or artificial (antiviral drug) selective pressures (reviewed in different chapters of [12, 13, 16]). In vivo an "arms race" is established between the virus and the host: the virus increases its fitness but it must overcome the multiple physiological responses intended to limit virus replication and to clear the virus from the organism. Successful evasion contributes to viral persistence, as documented with different viral pathogens [6, 8, 13, 15, 49] (Tab. I).

Virus dissemination in a new host species occurs when the viral fitness in the recipient host is such that transmission from one infected individual into susceptible individuals of the same species attains a minimum required value. Here, again, several interconnected factors can play a role: the viral load achieved in the infected donor, the amount of infectious particles shed by the donor, the amount of virus in the different secretions and excretions (saliva, blood, semen, feces, etc.), transmission routes (contact, aerosol, blood, sexual, etc.) [22, 53], the stability of the virus particles, etc. Thus, many influences that depend on the genetic make up of the virus and also on the nature of the virus-host relationship affect virus transmissibility. The capacity of expansion in the new host species is often expressed by a reproductive ratio of $R_0 > 1$, that quantitates the average number of infected contacts produced by each infected host [62, 63]. Some viruses, such as the SARS coronavirus, did not reach an average $R_0 > 1$ once introduced in the human population, preventing their continued expansion. Because of its epidemiologic parameters, SARS could be controlled by public health measures. In contrast, other viruses such as HIV-1 or the rhinoviruses are, unfortunately, highly efficient in sustained human-to-human transmission. In the case of HIV-1, asymptomatic donor individuals are capable of transmitting the virus to multiple recipient individuals [44]. In this case, a change of social habits (including sexual promiscuity, and travel) have contributed to the dissemination of HIV-1 in the human population. Obviously R_0 is not
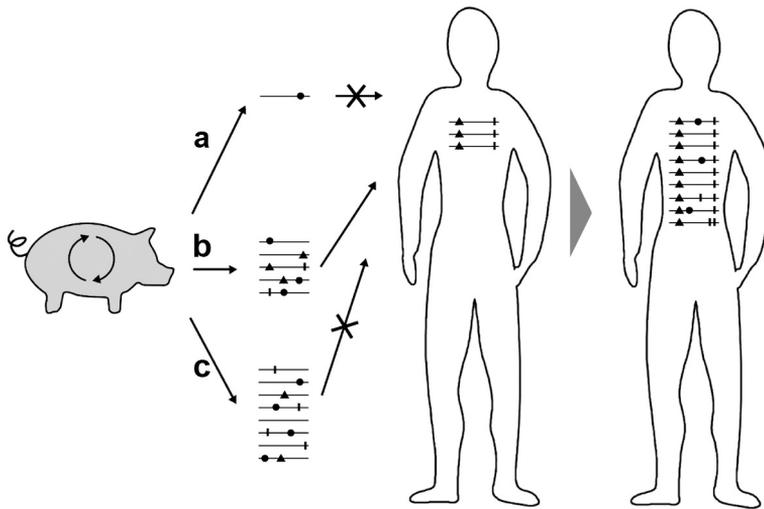
Figure 2. A simplified view of some events involved in RNA virus transmission, that takes into consideration that RNA viruses are quasispecies (mutant distributions). A swine-human transmission is used as illustration. An RNA virus replicates in an infected swine and it can shed different amounts of virus (a, b, c) that can reach a susceptible human. If two specific mutations are needed for the virus to initiate replication in the exposed human host, only the mutant spectrum that includes the required mutations will be established in the human host (in this case distribution b). If viral replication continues in the same individual (human figure on the right, two time points of the same human depicted as separated by the grey arrow), a new mutant distribution will be generated, shaped by the selective constraints imposed by each individual human (immune response and others). Human-to-human transmission (not shown in this scheme) will again involve bottlenecks of different intensities (number of genomes that reach the successive recipient humans). When transmissions occur among individuals of the same host species, it is likely that most mutant distributions will include genomes that are replication-competent in the recipient host. The picture is in reality more complex, as documented in the text. However, the scheme emphasizes the fact that viruses are not defined sequences but mutant distributions, and very frequently this pro-adaptive population structure has not been considered in treatments of emerging viral infections.

invariant, and may be modified by genetic change of the virus.

The number of infectious particles shed by each infected individual is critical not only because the amount of virus that is shed affects the probability of encountering a susceptible host, but also because the larger the number of infectious particles the broader the spectrum of mutants that will reach the recipient host individuals (Fig. 2). Viral population numbers can be largely amplified in immunocompromised individuals, that have been termed "super-shedders" or "super-spreaders", and constitute large reservoirs of variant viruses [27, 38, 65, 68, 72, 76].

A biological feature relevant to emergence is the overlap that has been identified in several viruses between receptor-recognition sites and antigenic sites. The fact that some exposed amino acid residues are part of one or several epitopes for antibody binding, as well as part of the receptor-recognition domain, may favor the coevolution of host cell tropism and antigenicity [2, 3, 36, 58, 83]. Such overlap is, however, double-edged. It opens the possibility of pleiotropic effects of single mutations (that may affect positively or negatively more than one viral function), and of coevolution of phenotypic traits. However, it may also introduce additional evolutionary constraints to the virus,

as will the overlap of domains (in viral proteins or nucleic acids) in general. Structural and biochemical studies have indicated that residues located at the surface of viral particles tend to be subjected to less structural constraints than residues located in the interior of viral capsids or envelopes and, therefore, they may be prone to accept amino acid substitutions. Tolerance to change of surface amino acid residues may facilitate the emergence of new cell receptor specificities mediated by antigenic change with obvious implications for viral disease emergence and re-emergence [10, 34, 43, 90]. Partial immunity to FMDV, evoked by synthetic peptides that mimicked the major antigenic site of the virus, produced the selection of antigenic variants in cattle, and the variants displayed alterations of cell tropism [80, 81]. Escape mutants selected by ineffective vaccines have been reported in other viral systems [85]. In addition, tolerance to amino acid substitutions at surface sites of capsid or envelope proteins may result in antigenic variation independent of immune selection, as documented with several DNA and RNA viruses [10].

Long-term coexistence of viruses and their hosts has given coevolution a good chance to reach some relative equilibrium that has excluded "super-virulent" pathogens from our biosphere, because if they existed they would be eliminated with their hosts. However, transient disequilibria may often occur when viruses cross a species barrier. The great majority of viruses are not pathogenic, although most of the viruses that have been studied in detail are associated with disease, for obvious reasons. Many viruses do not cause disease in their natural reservoir hosts but can be highly pathogenic when transmitted to a new host species. This is amply documented by the zoonotic nature of most of the viruses that have emerged in the human population (influenza virus type A from aquatic birds, HIV-1 from Old World primates, SARS coronavirus from bats probably via palm civets, Nipah and Hendra viruses from fruit bats via pigs, among others) [1, 33, 40, 55, 59, 69, 77, 89].

Population equilibria and disequilibria in viruses depend also on the nature and tempo of genetic change imposed by the surrounding host environments, in ways that only now we are beginning to understand.

5. SOME GENERAL CONCLUSIONS ON VIRAL DISEASE EMERGENCE. ZOONOTIC RESERVOIRS AND FITNESS GAINS

Viruses generally have multiple genes involved in cellular tropism and host range, that encode non-structural or structural viral proteins [2, 51]. Modification of such genes by mutation or their acquisition by recombination or reassortment opens the way to a change in host range and to disease emergence. In the present article I have emphasized a major issue of RNA virus genetics that has been considered only marginally in several studies on the mechanisms of viral disease emergence. The issue is that RNA viruses replicate as complex mutant distributions, termed viral quasispecies, and that the complexity and composition of mutant distributions are key factors in virus adaptation and persistence [13, 14, 16, 23, 70, 73, 84] (Tab. I). Contrary to designed experiments of adaptation of a virus to new host in which the role of mutant spectra in the adaptation process has been observed [64], no such observations have been made in the process of a natural emergence. Therefore, there is no definitive proof that a number of related mutants from a donor host reached a potential new host, or that genetic variation of the colonizing virus was essential for the establishment of the virus in the new host. However, what the molecular studies on viral genomes tell us is that when more than one virus particle reaches a recipient host there is a high probability that the genomes will not be identical, and that they may display fitness differences. It is worth noting that fitness changes can occur in viral quasispecies without being reflected in any modification of the genomic consensus nucleotide sequence [30–32]. Thus, relevant biological information can be lost if surveys of viral genomic sequences (for the purpose of the control of disease emergence or other), are restricted to the determination of consensus sequences. A new era to penetrate into the mutant spectrum composition

has started with the use of the new ultra deep sequencing or microarray-based methodologies by which the characterization of mutant spectra can be based on thousands of sequences [20, 47, 82, 86, 88].

Despite difficulties of implementation, it would be important to survey those viruses that have manifested a potential for emergence in the human population (human and animal influenza viruses, animal coronaviruses, primate lentiviruses, animal picornaviruses, etc.) not only by sampling consensus sequences but also by ultra deep sequencing. However, for obvious reasons covered in this and other articles of this special issue, it is not possible to anticipate which viruses from the multiple animal reservoirs are going to emerge as new human pathogens and when. Survey of the circulating zoonotic viruses, with effective communications networks, the development of new vaccines and antiviral agents, and an increasing understanding of viral population dynamics all should contribute to either delaying viral emergences or reducing their impact once they occur.

REFERENCES

[1] Antia R., Regoes R.R., Koella J.C., Bergstrom C.T., The role of evolution in the emergence of infectious diseases, Nature (2003) 426:658–661.

[2] Baranowski E., Ruiz-Jarabo C.M., Domingo E., Evolution of cell recognition by viruses, Science (2001) 292:1102–1105.

[3] Baranowski E., Ruiz-Jarabo C.M., Pariente N., Verdaguer N., Domingo E., Evolution of cell recognition by viruses: a source of biological novelty with medical implications, Adv. Virus Res. (2003) 62:19–111.

[4] Batschelet E., Domingo E., Weissmann C., The proportion of revertant and mutant phage in a growing population, as a function of mutation and growth rate, Gene (1976) 1:27–32.

[5] Bloomfield V.A., Crothers D.M., Tinoco J., Nucleic acids. Structures, properties, and functions, Section I, University Science Books, Sausalito, CA, 2000.

[6] Bowen D.G., Walker C.M., The origin of quasi-species: cause or consequence of chronic hepatitis C viral infection?, J. Hepatol. (2005) 42:408–417.

[7] Bushman F., Lateral DNA transfer, mechanisms and consequences, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2002.

[8] Ciurea A., Hunziker L., Martinic M.M., Oxenius A., Hengartner H., Zinkernagel R.M., CD4+ T-cell-epitope escape mutant virus selected in vivo, Nat. Med. (2001) 7:795–800.

[9] Chetverin A.B., Kopein D.S., Chetverina H.V., Demidenko A.A., Ugarov V.I., Viral RNA-directed RNA polymerases use diverse mechanisms to promote recombination between RNA molecules, J. Biol. Chem. (2005) 280:8748–8755.

[10] Domingo E., Díez J., Martínez M.A., Hernández J., Holguín A., Borrego B., Mateu M.G., New observations on antigenic diversification of RNA viruses. Antigenic variation is not dependent on immune selection, J. Gen. Virol. (1993) 74:2039–2045.

[11] Domingo E., Holland J.J., RNA virus mutations and fitness for survival, Annu. Rev. Microbiol. (1997) 51:151–178.

[12] Domingo E., Biebricher C., Eigen M., Holland J.J., Quasispecies and RNA virus evolution: principles and consequences, Landes Bioscience, Austin, 2001.

[13] Domingo E., Quasispecies: concepts and implications for virology, Springer Verlag, 2006.

[14] Domingo E., Virus evolution, in: Fields Virology, 5th ed., Lappincott Williams & Wilkins, Philadelphia, 2007.

[15] Domingo E., Gomez J., Quasispecies and its impact on viral hepatitis, Virus Res. (2007) 127:131–150.

[16] Domingo E., Parrish C., Holland J.J.E., Origin and evolution of viruses, 2nd ed., Elsevier, Oxford, 2008.

[17] Drake J.W., Holland J.J., Mutation rates among RNA viruses, Proc. Natl. Acad. Sci. USA (1999) 96:13910–13913.

[18] Eckerle L.D., Lu X., Sperry S.M., Choi L., Denison M.R., High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants, J. Virol. (2007) 81:12135–12144.

[19] Eigen M., Biebricher C.K., Sequence space and quasispecies distribution, CRC Press, Boca Raton, FL, 1988.

[20] Eriksson N., Pachter L., Mitsuya Y., Rhee S.Y., Wang C., Gharizadeh B., et al., Viral population estimation using pyrosequencing, PLoS Comput. Biol. (2008) 4:e1000074.

[21] Escarmís C., Dávila M., Domingo E., Multiple molecular pathways for fitness recovery of an RNA virus debilitated by operation of Muller's ratchet, J. Mol. Biol. (1999) 285:495–505.

[22] Evans A.S., Kaslow R.A., Viral infections of humans. Epidemiology and control, Plenum Medical Book Company, New York and London, 1997.

[23] Farci P., Shimoda A., Coiana A., Diaz G., Peddis G., Melpolder J.C., et al., The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies, Science (2000) 288:339–344.

[24] Ferrer-Orta C., Arias A., Escarmis C., Verdaguer N., A comparison of viral RNA-dependent RNA polymerases, Curr. Opin. Struct. Biol. (2006) 16:27–34.

[25] Figlerowicz M., Alejska M., Kurzynska-Kokorniak A., Figlerowicz M., Genetic variability: the key problem in the prevention and therapy of RNA-based virus infections, Med. Res. Rev. (2003) 23:488–518.

[26] Friedberg E.C., Walker G.C., Siede W., Wood R.D., Schultz R.A., Ellenberger T., DNA repair and mutagenesis, American Society for Microbiology, Washington, DC, 2006.

[27] Gavrilin G.V., Cherkasova E.A., Lipskaya G.Y., Kew O.M., Agol V.I., Evolution of circulating wild poliovirus and of vaccine-derived poliovirus in an immunodeficient patient: a unifying model, J. Virol. (2000) 74:7381–7390.

[28] Gibbs M.J., Armstrong J.S., Gibbs A.J., The haemagglutinin gene, but not the neuraminidase gene, of "Spanish flu" was a recombinant, Philos. Trans. R. Soc. Lond. B Biol. Sci. (2001) 356:1845–1855.

[29] Gmyl A.P., Korshenko S.A., Belousov E.V., Khitrina E.V., Agol V.I., Nonreplicative homologous RNA recombination: promiscuous joining of RNA pieces?, RNA (2003) 9:1221–1231.

[30] González-López C., Arias A., Pariente N., Gómez-Mariano G., Domingo E., Preextinction viral RNA can interfere with infectivity, J. Virol. (2004) 78:3319–3324.

[31] González-López C., Gómez-Mariano G., Escarmís C., Domingo E., Invariant aphthovirus consensus nucleotide sequence in the transition to error catastrophe, Infect. Genet. Evol. (2005) 5:366–374.

[32] Grande-Pérez A., Lazaro E., Lowenstein P., Domingo E., Manrubia S.C., Suppression of viral infectivity through lethal defection, Proc. Natl. Acad. Sci. USA (2005) 102:4448–4452.

[33] Haagmans B.L., Andeweg A.C., Osterhaus A.D., The application of genomics to emerging zoonotic viral diseases, PLoS Pathog. (2009) 5:e1000557.

[34] Haydon D.T., Woolhouse M.E., Immune avoidance strategies in RNA viruses: fitness continuums arising from trade-offs between immunogenicity and antigenic variability, J. Theor. Biol. (1998) 193: 601–612.

[35] He C.Q., Xie Z.X., Han G.Z., Dong J.B., Wang D., Liu J.B., et al., Homologous recombination as an evolutionary force in the avian influenza A virus, Mol. Biol. Evol. (2009) 26:177–187.

[36] Hueffer K., Parrish C.R., Parvovirus host range, cell tropism and evolution, Curr. Opin. Microbiol. (2003) 6:392–398.

[37] Ishihama A., Mizumoto K., Kawakami K., Kato A., Honda A., Proofreading function associated with the RNA-dependent RNA polymerase from influenza virus, J. Biol. Chem. (1986) 261:10417–10421.

[38] Khetsuriani N., Prevots D.R., Quick L., Elder M.E., Pallansch M., Kew O., Sutter R.W., Persistence of vaccine-derived polioviruses among immunodeficient persons with vaccine-associated paralytic poliomyelitis, J. Infect. Dis. (2003) 188:1845–1852.

[39] Knowles N.J., Samuel A.R., Molecular epidemiology of foot-and-mouth disease virus, Virus Res. (2003) 91:65–80.

[40] Krauss H., Weber A., Appel M., Enders B., Isenberg H.D., Schiefer H.G., et al., Zoonoses. Infectious diseases transmissible from animals to humans, ASM Press, Washington, DC, 2003.

[41] Lai M.M.C., Genetic recombination in RNA viruses, Curr. Top. Microbiol. Immunol. (1992) 176:21–32.

[42] Lazaro E., Escarmis C., Perez-Mercader J., Manrubia S.C., Domingo E., Resistance of virus to extinction on bottleneck passages: study of a decaying and fluctuating pattern of fitness loss, Proc. Natl. Acad. Sci. USA (2003) 100:10830–10835.

[43] Lea S., Hernández J., Blakemore W., Brocchi E., Curry S., Domingo E., et al., The structure and antigenicity of a type C foot-and-mouth disease virus, Structure (1994) 2:123–139.

[44] Levy J.A., HIV and the pathogenesis of AIDS, ASM Press, Washington, DC, 2007.

[45] Li W., Wong S.K., Li F., Kuhn J.H., Huang I.C., Choe H., Farzan M., Animal origins of the severe acute respiratory syndrome coronavirus: insight from ACE2-S-protein interactions, J. Virol. (2006) 80: 4211–4219.

[46] Luytjes W., Bredenbeek P.J., Noten A.F., Horzinek M.C., Spaan W.J., Sequence of mouse hepatitis virus A59 mRNA 2: indications for RNA recombination between coronaviruses and influenza C virus, Virology (1988) 166:415–422.

[47] Mardis E.R., The impact of next-generation sequencing technology on genetics, Trends Genet. (2008) 24:133–141.

[48] Martinez M.A., Hernandez J., Piccone M.E., Palma E.L., Domingo E., Knowles N., Mateu M.G., Two mechanisms of antigenic diversification of foot-and-mouth disease virus, Virology (1991) 184: 695–706.

[49] Mascola J.R., The cat and mouse of HIV-1 antibody escape, PLoS Pathog. (2009) 5:e1000592.

[50] Maynard Smith J.M., Natural selection and the concept of a protein space, Nature (1970) 225: 563–564.

[51] McFadden G., Poxvirus tropism, Nat. Rev. Microbiol. (2005) 3:201–213.

[52] Menéndez-Arias L., Molecular basis of fidelity of DNA synthesis and nucleotide specificity of retroviral reverse transcriptases, Prog. Nucleic Acid Res. Mol. Biol. (2002) 71:91–147.

[53] Mims C., Nash A., Stephen J., Mims' pathogenesis of infectious disease, Academic Press, San Diego, 2001.

[54] Minskaia E., Hertzig T., Gorbalenya A.E., Campanacci V., Cambillau C., Canard B., Ziebuhr J., Discovery of an RNA virus $3' \rightarrow 5'$ exoribonuclease that is critically involved in coronavirus RNA synthesis, Proc. Natl. Acad. Sci. USA (2006) 103: 5108–5113.

[55] Morse S.S., The evolutionary biology of viruses, Raven Press, New York, 1994.

[56] Nagy P.D., Carpenter C.D., Simon A.E., A novel $3'$-end repair mechanism in an RNA virus, Proc. Natl. Acad. Sci. USA (1997) 94:1113–1118.

[57] Nagy P.D., Simon A.E., New insights into the mechanisms of RNA recombination, Virology (1997) 235:1–9.

[58] Nemerow G.R., Stewart P.L., Antibody neutralization epitopes and integrin binding sites on nonenveloped viruses, Virology (2001) 288:189–191.

[59] Nijhuis M., van Maarseveen N.M., Boucher C.A., Antiviral resistance and impact on viral replication capacity: evolution of viruses under antiviral pressure occurs in three phases, Handb. Exp. Pharmacol. (2009) 189:299–320.

[60] Novella I.S., Duarte E.A., Elena S.F., Moya A., Domingo E., Holland J.J., Exponential increases of RNA virus fitness during large population transmissions, Proc. Natl. Acad. Sci. USA (1995) 92:5841–5844.

[61] Novella I.S., Contributions of vesicular stomatitis virus to the understanding of RNA virus evolution, Curr. Opin. Microbiol. (2003) 6:399–405.

[62] Nowak M.A., May R.M., Virus dynamics. Mathematical principles of immunology and virology, Oxford University Press Inc., New York, 2000.

[63] Nowak M.A., Evolutionary dynamics, The Belknap Press of Harvard University Press, Cambridge, Massachusetts and London, England, 2006.

[64] Nunez J.I., Molina N., Baranowski E., Domingo E., Clark S., Burman A., et al., Guinea pig-adapted foot-and-mouth disease virus with altered receptor recognition can productively infect a natural host, J. Virol. (2007) 81:8497–8506.

[65] Odoom J.K., Yunus Z., Dunn G., Minor P.D., Martin J., Changes in population dynamics during long-term evolution of sabin type 1 poliovirus in an immunodeficient patient, J. Virol. (2008) 82:9179–9190.

[66] Palese P., Shaw M.L., Orthomyxoviridae: the viruses and their replication, in: Knipe D.M., Howley P.M. (Eds.), Fields Virology, 5th ed., Lappincott Williams & Wilkins, Philadelphia, 2007.

[67] Parrish C.R., Kawaoka Y., The origins of new pandemic viruses: the acquisition of new host ranges by canine parvovirus and influenza A viruses, Annu. Rev. Microbiol. (2005) 59:553–586.

[68] Paunio M., Peltola H., Valle M., Davidkin I., Virtanen M., Heinonen O.P., Explosive school-based measles outbreak: intense exposure may have resulted in high risk, even among revaccinees, Am. J. Epidemiol. (1998) 148:1103–1110.

[69] Peters C.J., Emerging viral diseases, in: Knipe D.M., Howley P.M. (Eds.), Fields Virology, 5th ed., Lappincott Williams & Wilkins, Philadelphia, 2007, pp. 605–625

[70] Pfeiffer J.K., Kirkegaard K., Ribavirin resistance in hepatitis C virus replicon-containing cell lines conferred by changes in the cell line or mutations in the replicon RNA, J. Virol. (2005) 79:2346–2355.

[71] Quiñones-Mateu M.E., Arts E., Virus fitness: concept, quantification, and application to HIV population dynamics, Curr. Top. Microbiol. Immunol. (2006) 299:83–140.

[72] Rocha E., Cox N.J., Black R.A., Harmon M.W., Harrison C.J., Kendal A.P., Antigenic and genetic variation in influenza A (H1N1) virus isolates

recovered from a persistently infected immunodeficient child, J. Virol. (1991) 65:2340–2350.

[73] Sanz-Ramos M., Diaz-San Segundo F., Escarmis C., Domingo E., Sevilla N., Hidden virulence determinants in a viral quasispecies in vivo, J. Virol. (2008) 82:10465–10476.

[74] Schnitzler S.U., Schnitzler P., An update on swine-origin influenza virus A/H1N1: a review, Virus Genes (2009) 39:279–292.

[75] Simmonds P., Welch J., Frequency and dynamics of recombination within different species of human enteroviruses, J. Virol. (2006) 80:483–493.

[76] Small M., Tse C.K., Walker D.M., Super-spreaders and the rate of transmission of the SARS virus, Physica D (2006) 215:146–158.

[77] Smolinski M.S., Hamburg M.A., Lederberg J., Microbial threats to health. Emergence, detection and response, The National Academics Press, Washington, DC, 2003.

[78] Solé R., Goodwin B., Signs of life, how complexity pervades biology, Basic Books, New York, 2000.

[79] Steinhauer D.A., Domingo E., Holland J.J., Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase, Gene (1992) 122:281–288.

[80] Taboga O., Tami C., Carrillo E., Núñez J.I., Rodríguez A., Saíz J.C., et al., A large-scale evaluation of peptide vaccines against foot-and-mouth disease: lack of solid protection in cattle and isolation of escape mutants, J. Virol. (1997) 71:2606–2614.

[81] Tami C., Taboga O., Berinstein A., Nuñez J.I., Palma E.L., Domingo E., et al., Evidence of the coevolution of antigenicity and host cell tropism of foot-and-mouth disease virus in vivo, J. Virol. (2003) 77:1219–1226.

[82] Tsibris A.M., Korber B., Arnaout R., Russ C., Lo C.C., Leitner T., et al., Quantitative deep sequencing reveals dynamic HIV-1 escape and large population shifts during CCR5 antagonist therapy in vivo, PLoS ONE (2009) 4:e5683.

[83] Verdaguer N., Mateu M.G., Andreu D., Giralt E., Domingo E., Fita I., Structure of the major antigenic loop of foot-and-mouth disease virus complexed with a neutralizing antibody: direct involvement of the Arg-Gly-Asp motif in the interaction, EMBO J. (1995) 14:1690–1696.

[84] Vignuzzi M., Stone J.K., Arnold J.J., Cameron C.E., Andino R., Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population, Nature (2006) 439:344–348.

[85] Villefroy P., Letourneur F., Coutsinos Z., Mortara L., Beyer C., Gras-Masse H., et al., SIV escape mutants in rhesus macaques vaccinated with NEF-derived lipopeptides and challenged with pathogenic SIVmac251, Virol. J. (2006) 3:65.

[86] Wang C., Mitsuya Y., Gharizadeh B., Ronaghi M., Shafer R.W., Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance, Genome Res. (2007) 17:1195–1201.

[87] Weaver S.C., Evolutionary influences in arboviral disease, Curr. Top. Microbiol. Immunol. (2006) 299:285–314.

[88] Webster D.R., Hekele A.G., Lauring A.S., Fisher K.F., Li H., Andino R., De Risi J.L., An enhanced single base extension technique for the analysis of complex viral populations, PLoS ONE (2009) 16:e7453.

[89] Woo P.C., Lau S.K., Yuen K.Y., Infectious diseases emerging from Chinese wet-markets: zoonotic origins of severe respiratory viral infections, Curr. Opin. Infect. Dis. (2006) 19:401–407.

[90] Woolhouse M.E., Webster J.P., Domingo E., Charlesworth B., Levin B.R., Biological and biomedical implications of the co-evolution of pathogens and their hosts, Nat. Genet. (2002) 32:569–577.

[91] Zhang C.Y., Wei J.F., He S.H., Adaptive evolution of the spike gene of SARS coronavirus: changes in positively selected sites in different epidemic groups, BMC Microbiol. (2006) 6:88.