

Chapter 7

Viral Genomics: Implications for the Understanding and Control of Emerging Viral Diseases

Christine V.F. Carrington

Introduction

In recent decades, many infectious diseases have significantly increased in incidence and/or geographic range, in some cases impacting heavily on human, animal or plant populations. Some of these ‘emerging infectious diseases’ are associated with pathogens that have appeared in populations for the first time as a result of cross-species transmission (e.g. human immunodeficiency virus—acquired immunodeficiency syndrome (HIV-AIDS), severe acute respiratory syndrome (SARS)), while others were previously known but are rapidly increasing in incidence or geographic range as a result of underlying epidemiological changes (e.g. multi-drug resistant *Staphylococcus aureus* (MRSA) infection, dengue, West Nile encephalitis, foot and mouth disease, cassava mosaic disease). The latter include prominent diseases such as tuberculosis, malaria and yellow fever that were once on the decline but are now re-emerging.

Factors underlying emergence may be broadly grouped into (1) ‘ecological’ changes (such as environmental, agricultural, socio-economic, demographic and behavioural changes) that increase the probability of exposure of susceptible individuals/populations to infected reservoir hosts or vectors, (2) evolutionary changes that lead to increased pathogen virulence, drug resistance, host range or transmissibility and (3) changes in host population susceptibility (e.g. due to malnutrition and HIV-associated immunodeficiency in human populations). In human populations, the majority of disease emergence is driven by ecological factors (Jones et al. 2008; Morens et al. 2004; Taylor et al. 2001; Weiss and McMichael 2004; Woolhouse and Gaunt 2007; Woolhouse 2002; Woolhouse et al. 2005). In particular, anthropogenic

C.V.F. Carrington (✉)
Department of Preclinical Sciences, Faculty of Medical Sciences,
The University of the West Indies, St. Augustine, Republic of Trinidad and Tobago
e-mail: christine.carrington@sta.uwi.edu

factors such as deforestation, habitat fragmentation, urbanisation and modern agricultural practices provide increased opportunities for human interaction with infected reservoirs and vectors, and the existence of rapid global transport networks, and high-density human and animal populations facilitate the spread of pathogens at an unprecedented rate, often over very large distances (Jones et al. 2008; Morens et al. 2004; Taylor et al. 2001; Weiss and McMichael 2004; Woolhouse and Gaunt 2007; Woolhouse 2002; Woolhouse et al. 2005).

Although emerging and re-emerging diseases are associated with all types of microbes, viruses (and in particular RNA viruses) predominate (Taylor et al. 2001). This is considered a consequence of their large population sizes and capacity for rapid evolutionary change (Woolhouse 2002) which together can produce large pools of phenotypic variants including viruses with altered virulence, transmissibility or host range that have increased epidemic potential in their original hosts or are able to jump species boundaries and establish themselves in new hosts. Emergence that leads to successful host switching may be classified into three stages: (1) initial single infection of a new host with no onward transmission (i.e. spillovers into ‘dead-end’ hosts), (2) spillovers that go on to cause local chains of transmission in the new host population before epidemic fade-out (i.e. outbreaks) and (3) epidemic or sustained endemic host-to-host disease transmission in the new host population (Parrish et al. 2008).

In human populations, the majority of emerging diseases are caused by viruses that originate in wildlife populations and spill over into humans either directly or via domestic animals (Taylor et al. 2001). Wolfe et al. (2007) defined five stages through which animal only (i.e. stage 1) viruses progress to become human only (stage 5) viruses such as measles, smallpox and mumps (see Table 7.1). For the majority of emerging viruses, humans apparently represent dead-end hosts and only a few proceed to stage 3 and beyond to achieve human—human transmission and cause epidemics or sustained endemic transmission (Parrish et al. 2008; Wolfe et al. 2007). Nonetheless when this does occur, the impact in terms of morbidity, mortality and economic costs can be immense, as has been well demonstrated by the emergence of HIV, SARS coronavirus and H1N1 influenza.

In terms of understanding and eventually controlling viral disease emergence, the challenge lies in identifying and quantifying the factors that determine which viruses may make the species jump and whether a new disease will progress to epidemic stage or not. At the other end of the spectrum, there is the ever-present challenge of developing effective therapies and vaccines against rapidly evolving viral pathogens. In this regard, emerging viruses, the nature and extent of their diversity, their evolutionary processes and disease mechanisms need to be fully characterised and understood.

Viruses were the first organisms to have their genomes completely sequenced (Fiers et al. 1976), and because of their small size, this could be done relatively quickly and cheaply even prior to the advent of ‘next-generation’ sequencing technologies. There is no doubt, however, that the latter has opened the floodgates since viral genomes can now be generated at lower cost and much more rapidly than was possible using conventional sequencing approaches. The number of viral genomes

Table 7.1 Stages of viral emergence into human populations (Wolfe et al. 2007)

Stage	Description	Examples
1	Animal only viruses	
2	Viruses that are maintained in animal populations and occasionally infect humans, but do not go on to being transmitted human to human	Rabies and West Nile viruses
3	Viruses that are able to maintain a few cycles of transmission between humans such that they cause occasional outbreaks that die out	Ebola and Marburg viruses
4	Viruses that are maintained in human populations for extended periods without involvement of animal hosts but are still maintained in animal populations and have a natural sylvatic cycles that to varying extents involve primary transmission of virus to humans from animal hosts	Yellow fever, dengue and influenza A viruses
5	Human only viruses	Mumps, measles and smallpox viruses

available in public databases continues to increase exponentially. This wealth of data has led to significant progress in terms of rapid identification and characterisation of emerging viruses, as well as knowledge about their biodiversity and evolution. In terms of evolutionary biology, the beauty of working in viral genomics lies in the ability to study evolutionary changes on the same time scales as the events that shape them. For several viruses, historical samples are available for retrospective study, and their analysis has contributed to our understanding of viral evolutionary and epidemiological factors/events accompanying their emergence, maintenance and spatial diffusion (reviewed in (Pybus and Rambaut 2009)).

In addition to enabling an exploitation of existing virus collections, the new sequencing technologies and accompanying bioinformatic tools provide the potential for comprehensive tracking of viral evolution and population dynamics in real time. Unfortunately, much less progress has been made in areas that impact directly on virus control and treatment (Holmes 2009). This is largely a consequence of the lack of appropriate clinical and epidemiological data to accompany the wealth of sequences (Holmes 2009). The other challenge that cannot be ignored is the ability of current computational approaches to deal with the huge volume of sequence data being generated.

In this chapter, I discuss how viral genomics has contributed to our understanding of each of the stages of viral emergence and how it might contribute to disease prevention and control in the future. Although disease emergence in other species can be of equal importance and ultimately impacts on human development, for the purpose of brevity, I concentrate primarily on diseases that have emerged in human populations and draw examples from those that have most deeply affected the developing world. While there is no apparent relationship between the tendency for new human pathogens to be reported and a country's geographic location or level of development (Woolhouse and Gaunt 2007), inadequate public health surveillance and response systems in developing countries coupled with the existence of

underlying disease conditions have meant that disease burden is usually greater in developing than developed countries (as well illustrated by the recent H1N1 pandemic, (Archer et al. 2009)). Additionally, prevention and control strategies that are effective in more developed countries often fall short in resource-limited settings, which can then act as pockets of refuge where pathogens persist, and may serve as future source populations for outbreaks in other regions.

Investigating the Cross-Species Transmission Interface

Identification and Characterisation of Potentially Emergent Viruses in Animal Populations

It has been suggested that preventing viral disease emergence in human populations begins with a systematic survey of viral diversity in animal populations (Wolfe et al. 2007). Such knowledge would enable identification of animal populations harbouring viruses that have previously infected humans or that are likely to do so by virtue of their relatedness to known human pathogens, or perhaps their ability to infect human cell lines (Holmes and Rambaut 2004). Zoonotic viruses generally cause little or no apparent disease in their original hosts; thus animal reservoirs are often not obvious. Since it is clearly impossible to survey all animal species, the focus of animal surveillance should be on species that are more likely to harbour potentially emergent viruses, for example species with large and/or dense populations, and in particular those that live in close proximity to and are more closely related to humans and their domestic mammals, such as rodents, bats and birds (Holmes and Rambaut 2004). Non-human primate populations (regardless of their size or population density) are also worth surveying because of their close evolutionary relationship with humans and the fact that a number of important human pathogens have emerged from them (e.g. dengue virus (DENV), chikungunya virus (CHIKV), yellow fever virus (YFV), human T-cell leukaemia virus (HTLV) and HIV). Finally, any other species having direct (e.g. bushmeat, livestock) or indirect contact (e.g. vector-mediated contact) with humans that could have led to human infections in the past should also be included (Wolfe et al. 2007).

Traditional approaches to virus discovery such as electron microscopy, cell culture, animal inoculation studies and serology (Storch 2007) have a number of limitations, the most important being that not all viruses can be cultured in the laboratory (Amann et al. 1995). There are now a range of sensitive molecular approaches to virus discovery that circumvent this problem by relying on detection and characterisation of viral genomes rather than targeting viral particles, antigens or their cytopathic effects (reviewed in Bexfield 2011). These include hybridisation-, PCR- and sequence-based approaches that have varying levels of reliance on sequence information from known pathogens and thus differ in terms of the range of pathogens

they would be expected to detect. For example, hybridisation-based techniques (such as microarray (Wang et al. 2002) and subtractive hybridisation (Lisitsyn et al. 1993)) require sequence information from known pathogens to detect related pathogens and are unable to detect completely novel virus families. Likewise, PCR-based approaches using degenerate primers are limited to amplification and detection of related viruses. However, there are also sequence independent PCR approaches that facilitate detection of completely novel pathogens. These include sequence-independent single primer amplification (SISPA), degenerate oligonucleotide primed PCR, random PCR and rolling circle amplification (reviewed in Bexfield 2011). When these approaches are coupled with 'next-generation' sequencing technology (Margulies et al. 2005) such as 454 pyrosequencing (Roche), Illumina (Solexa) and SoLiD™ (Applied Biosystems) for definitive identification of amplified fragments, they very efficiently generate large amounts of sequence data that can then be analysed using bioinformatic tools.

Next-generation sequencing also obviates the need for amplification prior to sequencing and has opened the field of metagenomics, i.e. the culture-independent study of microbial communities in environmental or biological samples by analysing the sample's nucleotide content. First applied to environmental samples such as sea water (Angly et al. 2006; Breitbart et al. 2002; Williamson et al. 2008), fresh water (Breitbart et al. 2009; Djikeng et al. 2009), soil (Fierer et al. 2007) and marine sediments (Breitbart et al. 2004), this approach has now been used to define the 'microbiomes' of a range of biological samples including human nasopharyngeal swabs (Bogaert et al. 2011), termite gut (Hongoh 2011) and cow rumen (Hess et al. 2011). It has also been adapted to specifically target viral metagenomes or 'viromes', by enriching samples for intact virions and then treating with nucleases to remove non-virion particle protected (naked) DNA and RNA (Djikeng et al. 2008). In terms of targeting potential reservoirs or vectors for emerging diseases, studies have been performed on faecal, oral, urine and tissue samples from bats (Donaldson et al. 2010; Li et al. 2010a), insect pools (Victoria et al. 2008), chimpanzee and farm animals (Li et al. 2010b).

The metagenomic approach has also been used for the identification and characterization of 2009 pandemic H1N1 influenza A virus from nasopharyngeal swabs (Greninger et al. 2010), to study previously 'uncharacterisable' viruses that have been isolated through culture (Victoria et al. 2008), to explore within host diversity of HIV and SIV (Bimber et al. 2010) and in comparative studies to identify viruses found in diseased versus healthy tissues from a variety of species (Blomström et al. 2010; Ng et al. 2009a; Ng et al. 2009b; Willner et al. 2009). However, one important limitation of this approach to detecting novel viruses is that the protocol currently used to enrich samples for viruses prior to sequencing includes a filtration step designed to exclude cells, cell debris and bacteria, which may also exclude very large viruses ('giant viruses') such as mimiviruses. Also, nuclease treatment eliminates the genomes of any viruses whose integrity has been disrupted by the enrichment process, and depending on the titre of remaining intact virions, these may not be efficiently sequenced (Djikeng et al. 2008).

One intriguing new approach to virus discovery that is worth noting in terms of its ability to characterise viral diversity in insects (which can be important viral vectors) is ‘virus discovery in invertebrates by deep sequencing and assembly of total small RNAs’ or vdSAR (Kreuze et al. 2009; Wu et al. 2010). This approach involves deep sequencing of viral small interfering RNAs (vsiRNA) produced by host immune machinery in response to infection. vsiRNAs are produced by cutting up viral genomes, so piecing their sequences together recovers the virus sequence. In addition to being a sequence independent approach, the process is expected to be more efficient since only a small proportion of host small RNAs need to be sequenced and data-mined (Wu et al. 2010). Additionally, since vdSAR assembles viral genomes from the products of an active host immune response to infection, only replicating and infectious viruses that induce the immune response are identified by this approach (Wu et al. 2010).

Identification and Characterisation of Newly Emerged Viruses

In addition to facilitating surveys of animal reservoirs and vectors, all of the techniques described above (with the exception of vdSAR) may be used to rapidly detect and characterise newly emerged viruses in human populations. This is usually the primary research focus when an apparently new infectious disease first appears, as it facilitates the development of screening tests for early detection and epidemiological investigations aimed at identifying risk groups, reservoirs and possible transmission routes. Such information can then be used to inform control and prevention strategies, including the development of vaccines and antiviral therapies.

The role that viral genomics can play in this regard was well demonstrated during the emergence of SARS, the first cases of which appeared in November 2002 in southern China. In March 2003, traditional cell culture resulted in the isolation of a novel virus from patient specimens (Drosten et al. 2003; Ksiazek et al. 2003; Peiris et al. 2003). Within days of this, the virus was identified as a coronavirus through the use of a pan viral microarray and confirmed by sequencing using two parallel approaches. The first involved designing primers based on known coronaviruses and amplifying regions of the novel virus, and in the second, viral sequences were directly recovered from the surface of the microarray to which they were hybridised, cloned and sequenced without the need to design specific primers (Wang et al. 2003). Comparison with previously characterised coronavirus strains demonstrated that the virus identified was distinct from all known human pathogens (Wang et al. 2003). Thus within 24 h, an unknown virus was identified as a coronavirus and within days partial genome sequences had been generated. Comparative genomics and evolutionary analyses also played the major role in pinpointing bats as the source of the precursor to the SARS virus and the primary reservoirs for SARS-like coronaviruses (Dominguez et al. 2007;

Gloza-Rausch et al. 2008; Lau et al. 2005; Poon et al. 2005; Tang et al. 2006; Tong 2009; Woo et al. 2006; Carrington et al. 2008).

As sequencing costs continue to fall and computing capacity improves, metagenomic approaches to virus detection and characterisation will no doubt become more and more routine aspects of public health activities. Researchers have demonstrated the potential utility of high-throughput pyrosequencing for the detection of viruses in human clinical specimens such as stool (Nakamura et al. 2009), nasopharyngeal swabs (Bogaert et al. 2011; Nakamura et al. 2009), autopsy-derived liver and kidney tissues (Palacios et al. 2008) and serum (Briese et al. 2009). This includes identification of novel viruses associated with high mortality outbreaks of unknown aetiology (Briese et al. 2009) and in tissues from individuals who died following organ transplantation from the same donor (Palacios et al. 2008). Others have demonstrated the potential usefulness of metagenomic sequencing in field surveillance for arboviruses by applying the technique to mosquitoes experimentally infected with dengue virus (Bishop-Lilly et al. 2010). It has even been suggested that metagenomic sequencing may be used for continual surveillance of large human populations for known and unknown viral pathogens (Anderson et al. 2003). The suggestion is that large pooled samples of human serum and plasma (possibly discarded specimens from diagnostic laboratories) could be enriched for viral particles and then subjected to metagenomic sequencing on a routine basis. Such large-scale continual surveillance could allow identification of viruses that have entered the human population even before the usual detection thresholds (which would normally depend on several people being infected) have been reached. According to the authors, this approach could be used to ‘monitor the levels of known viruses, rapidly detect outbreaks and systematically discover novel or variant human viruses’ (Anderson et al. 2003).

Understanding Factors Involved in Cross-Species Transmission and Adaptation to New Hosts

Evidence suggests that transmission of viruses from animal reservoirs to humans is not uncommon (Hahn et al. 2000; Wolfe et al. 2005; Wolfe et al. 2004). However, in the majority of cases, humans are dead-end hosts or even when they are not, the zoonotic virus cannot be sustained in prolonged transmission chains such that outbreaks are small and die out quickly. The barriers to onward transmission are primarily biological (Woolhouse and Gaunt 2007). For example, tissue tropism or viral titres achieved might not allow for efficient human-to-human transmission, or transmission might be restricted by reliance on a vector that does not commonly interact with humans or in which the virus does not achieve high enough titres to efficiently infect humans. In an apparent minority of cases, viruses surmount these barriers and can be maintained in the human population and may even lose their ability to replicate in the animal species they originated from.

The evolutionary events that enable cross-species transmission and subsequent adaptation to the new host are poorly understood. However, they are more likely to be the result of viral rather than human evolutionary changes since the time scale of human evolution is so much longer than the time frame implied by the frequency with which these events occur (Holmes and Rambaut 2004; Schliekelman et al. 2001). Studying viral evolution and comparative genomics applied to viruses before and after a transition, or to phylogenetically related human–animal pathogen pairs, can help us to understand the changes involved in adaptation to humans and other aspects of successful emergence.

This type of approach, coupled with *In vitro* and *in vivo* studies, was used to identify a single amino acid change in the envelope glycoprotein that is responsible for enzootic strains of Venezuelan encephalitis virus (VEEV) gaining the ability to cause epidemics of neurological and potentially fatal disease in horses, with humans as spill-over hosts (Anishchenko et al. 2006). VEEV, an arbovirus belonging to the genus *Alphavirus*, is usually maintained in an enzootic rodent-mosquito-rodent cycle. An amino acid change (Thr → Arg) at position 213 in the E2 glycoprotein confers the ability to cause high titre viraemia in horses, whereas the wild type is either unable to replicate in horses or does so at very low titres (Anishchenko et al. 2006). Likewise, the dramatic emergence of the CHIKV (another mosquito-borne alphavirus) in Asia has been linked to a single amino acid change in the envelope 1 glycoprotein (E1-A226V) of the Indian Ocean lineage responsible (de Lamballerie et al. 2008; Hapuarachchi et al. 2010; Kumar et al. 2008; Ng et al. 2009c; Sam et al. 2009; Schuffenecker et al. 2006). This change results in increased infectivity and transmissibility by *Aedes albopictus* (Tsetsarkin et al. 2007; Vazeille et al. 2007), previously considered as only a secondary vector in human-mosquito-human cycles (urban epidemic cycles), which typically involve *Ae. aegypti*.

While these findings in VEEV and CHIKV provide proof of concept, they are both unusual in that only one amino acid change resulted in adaptation to a new host/vector. This may be because in both cases, the viruses already had the ability to infect the ‘new’ host, albeit inefficiently. In the case of viruses entering a new species for the first time, the scenario is expected to be much more complicated. This may be why mutations associated with emergence remain unknown for other zoonoses including intensely studied viruses like HIV. Also, more recent work on CHIKV has shown that the effect of the E1-A226V mutation is lineage specific, working only in the IOL genomic background, with endemic Asian CHIKV strains requiring a second mutation (E1-98T) to become *Ae. albopictus* adapted (Tsetsarkin et al. 2011).

Next-generation sequencing technology allows for rapid and comprehensive surveys of the extent and nature of viral diversity within and amongst animal reservoir hosts, vectors and human populations. This would provide a basis for investigating the fitness distribution and relevance of mutations produced. The latter, coupled with good ecological, epidemiological, immunological and experimental data from *In vitro* and *in vivo* systems, is crucial if we are to understand the mechanisms involved in adaptation.

Understanding the Spatiotemporal Dynamics of Emerging Viruses

Phylogenetic inference may be used to reconstruct the demographic history of a population from molecular sequences sampled from the population (Drummond et al. 2005). The approach is based on a population genetic model known as the coalescent which describes the relationship between the shape of the genealogical tree of sampled sequences and the demographic history of the population from which they were sampled (i.e. rates of population growth and decline, extent of population subdivision and patterns of migration) (Kingman 1982; Griffiths and Tavaré 1994). In the case of RNA viruses, exploiting this link between population dynamics and molecular evolution, i.e. exploring their ‘phylodynamics’, (Holmes 2009; Grenfell et al. 2004) is particularly attractive since their high mutation rates, short generation times and large population sizes can result in significant genetic differences between sequences sampled within years, months or even days of each other. Additionally, the relatively short time frames involved mean that evolutionary and demographic events may be temporally aligned with the immunological, transmission and ecological events that shaped them. Given date and location stamped sequences, and depending on the nature and spatiotemporal resolution of the sampling, it is then possible to estimate when and where a given epidemic began or particular lineages arose, the order and timing of transmission events, the timing of changes in population growth rates, and the pattern and rate of virus movement between geographic regions, epidemiological risk groups, individuals and even tissues within an individual (reviewed in (Pybus and Rambaut 2009)). All very pertinent given that ecological and immunological rather than genetic factors are thought to be the main determinants of viral emergence (Holmes 2006).

One of the potential pitfalls of this approach is that inferences are based on estimated genealogies that have been derived with a level of uncertainty as the reconstructed genealogy is in fact only one of many that can be derived from the data. While it may be the best estimate, the true genealogy is rarely, if ever, known with absolute certainty. One solution is to account for this uncertainty by using probabilistic models to estimate parameters over many, many plausible genealogies, thereby providing a more rigorous statistical framework. The most commonly used model is the Bayesian skyline plot (Drummond et al. 2005) incorporated into the BEAST software package (Drummond and Rambaut 2007). This approach uses a Markov chain Monte Carlo (MCMC) sampling procedure to derive a distribution of trees from which a distribution of population size estimates is determined at intervals going back to the most recent common ancestor of the gene sequences (Drummond and Rambaut 2007; Drummond et al. 2002). The result is a plot of the estimated effective population size over time with credibility intervals that represent both phylogenetic and coalescent uncertainty (see Boxes 7.1 and 7.2). BEAST also jointly estimates substitution rates and divergence times (i.e. times to the most recent common ancestors of individual lineages and the genealogy as a whole) with credibility intervals and provides the option of using relaxed molecular clock models that allow

Box 7.1 Bayesian coalescent reconstruction of the demographic histories of invading strains of DENV in the Americas

All four DENV serotypes currently circulate in the Americas. DENV-4 was first reported in 1981 and identified as subtype II originating from Asia (Carrington et al. 2005; Lanciotti et al. 1997). In the same year, an Asian strain of DENV-2, distinct from the previously existing American subtype, was also reported (Deubel et al. 1986; Lewis et al. 1993; Twiddy et al. 2002). Figure 7.1 shows molecular clock phylogenies (*top panel*) and skyline plots (*centre panel*) estimated for the invading strains using sequences derived from DENV isolated from several countries in the Caribbean, South and Central America over about 20 years. Both skyline plots describe rapid exponential growth then maintenance of genetic diversity across epidemic peaks and troughs as estimated by the number of countries reporting DENV-2/-4 each year (*bottom panel*). This is likely a result of population subdivision, which is reflected, for example, in the clustering of sequences from mainland (*red*) and island (*blue*) countries. Therefore, in this case, genetic diversity cannot be reliably interpreted as proportional to population size. The faster initial increase in genetic diversity for DENV-4 compared to DENV-2 may reflect the immunological landscape, in that there was no herd immunity to DENV-4 in 1981, but another subtype of DENV-2 had already been circulating for many decades. The dates that the most recent common ancestors of each subtype existed are indicated by arrows along the x-axis of the skyline plot. They pre-date the first epidemiological reports of each virus by about a year. This suggests that viruses remained undetected until the number of infections or disease incidence reached a detection threshold, which might be quite high given the inadequate surveillance in many countries.

Box 7.1 (continued)

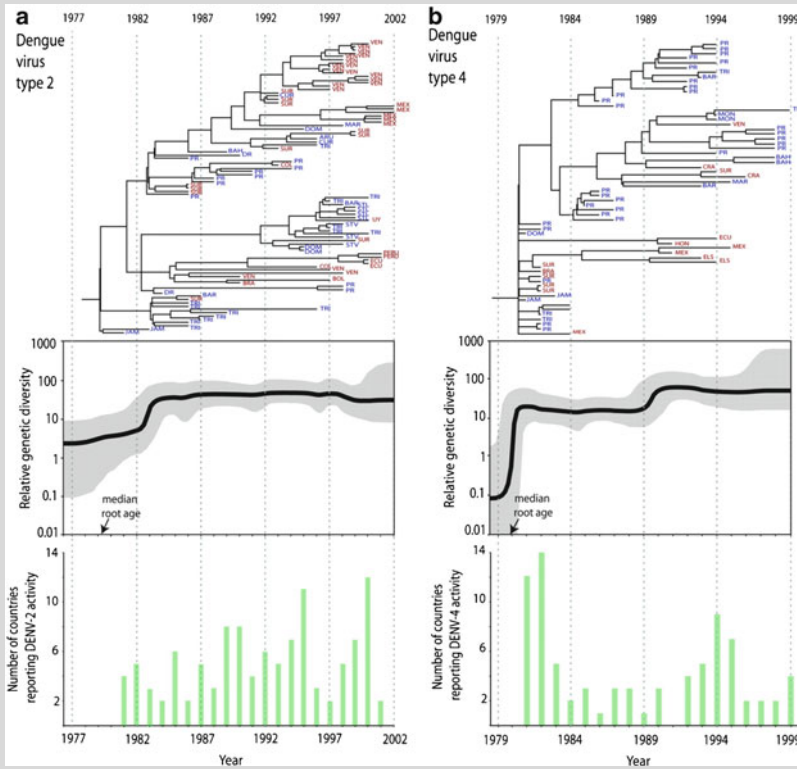


Fig. 7.1 Genealogies and corresponding Bayesian skyline plots showing the transmission histories of (a) DENV-2 subtype III and (b) DENV-4 subtype II, drawn on the same time scale. The y axes of the skyline plots represent relative genetic diversity, which is equal to the product of effective population size and generation length in the absence of population structure. For both viruses, the maximum a posteriori tree is presented on the same time scale as the skyline plot, with tip times corresponding to sampling times. The thick black lines are the median estimates, and the areas between the 95% CIs are shaded grey. Isolates on the trees are identified by their country of origin; mainland countries are labelled in red and islands in blue, and the tips of the phylogenies correspond to their sampling times. The numbers of countries reporting DENV-2 and DENV-4 activity in each year are summarised in the histograms shown. In the case of DENV-2, this represents the activities of both subtype III and V, which are not distinguished in epidemiological reports (Figure reproduced with permission from Carrington et al. (2005) *J Virol* 79(23):14680–14687, doi:10.1128/JVI.79.23.14680–14687.2005. ©2005 American Society for Microbiology)

for substitution rate variation across lineages in a tree (i.e. it does not assume a molecular clock) (Drummond et al. 2006). Several models that assume a particular pattern of population growth (e.g. exponential growth, constant population size) are also available for comparison (Drummond and Rambaut 2007). Boxes 7.1 and 7.2 describe results from two studies in which the demographic histories of dengue viruses were reconstructed from molecular sequences using the skyline plot in BEAST (Bennett et al. 2010; Carrington et al. 2005).

The BEAST programme was also recently extended to allow for inference, visualisation and hypothesis testing of phylogeographic history (Lemey et al. 2009). In the first model implemented, the geographic locations from which sequences were derived are considered as discrete states (Lemey et al. 2009). The spatial diffusion of the virus is then reconstructed using the coalescent approach to infer when and where direct ancestors of the sampled sequences existed. Different scenarios and models of spatial diffusion can be investigated and compared by specifying different prior distributions for the diffusion rates amongst the sampling locations (Lemey et al. 2009; Auguste et al. 2010; Talbi et al. 2010; Allicock et al. 2012). Phylogeographic inferences may be summarised using virtual globe software (Google Earth) such that spread over time may be visualised as an interactive animation. Examples of virtual globe projections demonstrating the diffusion dynamics through time are available online at <http://www.phylogeography.org>.

The above-mentioned discrete model, however, requires the assumption that at any point along the phylogeny, the samples existed in one of the sampled locations. To address this limitation, a more realistic ‘continuous trait’ model that allows for diffusion over a continuous landscape was recently implemented (Lemey et al. 2010). Box 7.3 illustrates the spatial spread of rabies virus amongst racoons in North America reconstructed using this model (Lemey et al. 2010).

Box 7.2 Bayesian coalescent reconstruction of the demographic history of DENV4 in Puerto Rico

All four DENV serotypes have existed in Puerto Rico (PR) since the 1970s causing regular and increasingly severe outbreaks (Bennett et al. 2010). Figure 7.2 shows (A) a skyline plot inferred from DENV4 sequences (~4,000 nucleotides) derived from viruses isolated in PR between 1981 and 1998. The pattern of cyclic epidemics described by the overlaid census data (number of confirmed dengue cases) strongly correlates with the estimates of effective population size (derived using the Bayesian coalescent framework in the software package BEAST), with a 7-month lag (increases in population size precede outbreaks) that is adjusted for in Figure B. The reason for the observed time lag is unclear. It may be that increased diversity provides the variation from which more fit, epidemic-causing strains are more likely to arise. Then during epidemics, diversity might be lost due to selection. Alternatively the discrepancy may be due to non-random and biased sampling in case counts and isolations (e.g. epidemiologic surveillance) (Bennett et al. 2010).

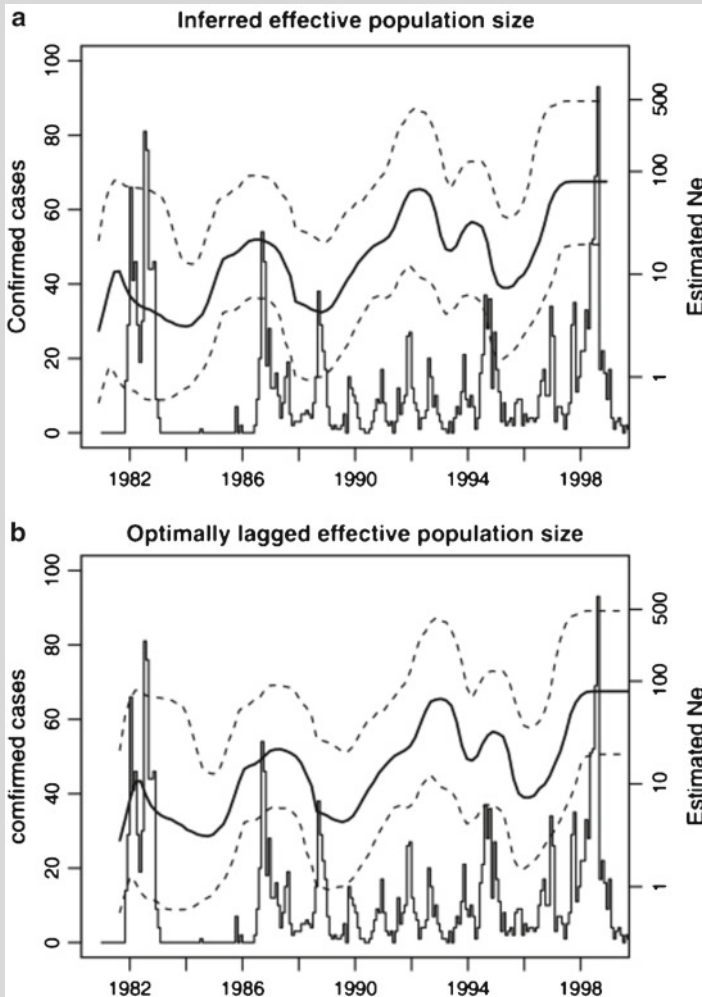
Box 7.2 (continued)

Fig. 7.2 Effective population size estimates in terms of effective numbers of infections per month based on viral genetic diversity and coalescent patterns, superimposed on the number of DENV-4 isolates by month (a) directly and (b) with a 7-month upwards adjustment of N_e (e.g. ahead in time) (Bennett et al. (2010) Epidemic dynamics revealed in dengue evolution. *Mol Biol Evol* 7(4):811–818 by permission of Oxford University Press)

In addition to those shown in Boxes 7.1–7.3, there are numerous other examples where a ‘phylogenetic’ approach to viral evolutionary analysis has been successfully applied. They include, for example, the reconstruction of the origin and global dissemination of HIV-1 (Gilbert et al. 2007; Korber et al. 2000; Vidal et al. 2000; Zhu et al. 1998), reconstruction of the spread of rabies virus in North Africa with an investigation of factors underlying the patterns observed (Talbi et al. 2010), inference of YFV and DENV spatial diffusion in the Americas (Auguste et al. 2010; Allcock et al. 2012) and investigation of the mechanism by which the YFV is maintained between epidemics (Auguste et al. 2010) and elucidation of the role of natural selection and global migration in influenza A epidemic patterns (Nelson et al. 2007; Rambaut et al. 2008; Russell et al. 2008). Although this approach cannot replace good epidemiological data, it complements traditional epidemiological approaches and provides insights into the evolutionary dynamics underlying epidemic behaviour. The ability of the approach to recover information not available in census data (e.g. in the analysis described in Box 7.3, the geographic area where the raccoon rabies virus is estimated to have spread by 1973 includes the location where the first raccoon rabies case was reported in 1977 even though the data did not include a sequence for this case (Lemey et al. 2010)) may be particularly useful in regions that have flawed monitoring and surveillance systems, such as in the developing world.

Box 7.3 Bayesian coalescent reconstruction of the spatial diffusion of rabies virus among North American raccoons

The spatiotemporal dynamics of rabies virus in North America was reconstructed from rabies virus nucleotide sequences using the Bayesian phylogeographic framework in the BEAST software package (Drummond and Rambaut 2007; Lemey et al. 2009; Lemey et al. 2010). The data set consisted of 47 rabies virus genomic fragments (of 2,811 nucleotides in length) sampled from a 30-year epidemic (Biek et al. 2007). Figure 7.3 shows snapshots of the dispersal pattern at different time points as illustrated by a projection of the inferred rabies virus phylogeny onto a map. The shaded regions represent the uncertainty about the locations of the rabies. Interestingly the area contained in 1973 diffusion pattern includes the location of the first raccoon rabies case reported in 1977 (green circle) even though the data do not include a sequence for this case. The changing tempo of the diffusion over time can be observed using the interactive animated visualisation available at www.phylogeography.org.

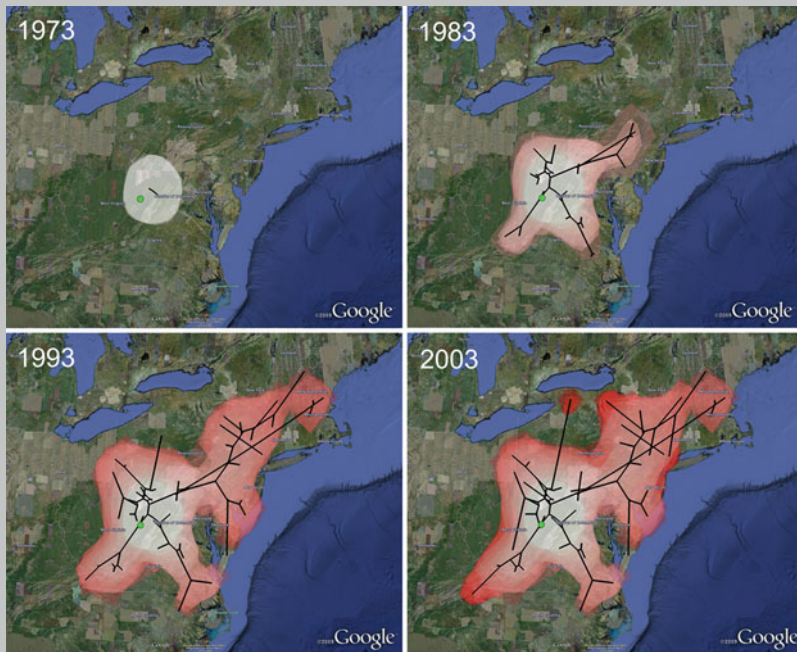
Box 7.3 (continued)

Fig. 7.3 Spatiotemporal dynamics of the rabies epidemic amongst North American raccoons. We provide snapshots of the dispersal pattern for August 1973, 1983, 1993 and 2003. Lines represent MCC phylogeny branches projected on the surface. The uncertainty on the location of raccoon rabies is represented by transparent polygons. These 80% highest posterior density regions are obtained by contouring a time-slice of the posterior phylogeny distribution and imputing the location on each branch in each phylogeny using the precision matrix parameters for the respective sample. The white–red colour gradient informs the relative age of the dispersal pattern (older recent). A *green circle* marks Pendleton County, WV, where the epizootic’s first case was reported in 1977. The maps are based on satellite pictures made available in Google Earth (<http://earth.google.com>). A dynamic visualisation of the spatiotemporal reconstruction can be explored at <http://www.phylogeography.org/> (Lemey et al. (2010) Phylogeography takes a relaxed random walk in continuous space and time. *Mol Biol Evol* 28(8):1877–1885 by permission of Oxford University Press

Antiviral Therapies, Prognostic Markers and Vaccines

The rapid viral evolution that facilitates species jumps and emergence also underlies viruses' ability to escape our immune systems and often presents a challenge in terms of developing effective vaccines and antiviral therapies. As described above, analysis of genomic data can provide valuable insights into virus evolution and epidemiology. The plethora of genomic sequence data being generated and the availability of rapid high-throughput sequencing technology therefore represent valuable resources that have already impacted on the way vaccine and therapeutic development is approached. In particular, they present the opportunity to understand the scope and distribution of the genomic diversity that must be tackled for a given virus and facilitate monitoring of the spatiotemporal dynamics of this biodiversity, thereby underpinning reverse vaccinology, pan-genomic and comparative genomic approaches to identifying vaccine and/or drug targets (Seib et al. 2009) (see Box 7.4). Genomic approaches are also expected to accelerate the identification of genetic and other molecular markers of prognostic and therapeutic relevance, such as markers of disease severity and drug resistance. Access to genomic data also enables researchers to go beyond genomics to transcriptomics, proteomics and other 'omics' approaches to studying emerging viruses.

However, despite the immense potential, with a few exceptions such as the use of pyrosequencing to screen for mutations associated with antiviral resistance in influenza (Deyde et al. 2009; Deyde et al. 2010; Bright et al. 2005; Deng et al. 2011; Dharan et al. 2009; Duwe and Schweiger 2008; Hurt et al. 2009; Lackenby et al. 2008) and resources such as the Stanford HIV drug resistance database (<http://hivdb.stanford.edu/index.html>), genomic developments of direct relevance to clinical care have been slow in coming (Holmes 2009). This is likely to be a consequence of the fact that genomic data are not often associated with data on clinical manifestations and host immunological responses that would enable them to be fully exploited (Holmes 2009). Notable exceptions are the aforementioned Stanford

Box 7.4 Genomic Approaches to Identifying Vaccine and Drug Targets

Reverse vaccinology: Bioinformatic approaches are used to screen whole pathogen genomes for genes that are potential vaccine and drug targets by virtue of the predicted function and other attributes of the proteins they encode.

Pan-genomics: Multiple genomes from a given pathogen are analysed in order to identify conserved antigens/targets that would ensure that vaccines or therapies based on them are effective against the full spectrum of pathogen diversity.

Comparative genomics: Genomes from pathogenic and non-pathogenic strains of a given pathogen are analysed in order to identify antigens/targets associated with disease.

HIV drug resistance database and the Los Alamos HIV databases (<http://www.hiv.lanl.gov/content/index>), and more recently, large-scale whole genome sequencing projects such as the Broad Institute's Genome Resources in Dengue Consortium (GRID) project (<http://www.broadinstitute.org/annotation/viral/Dengue/>) and the influenza genome sequencing projects (IGSP) by The Institute for Genomic Research (TIGR) (<http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html>) have sought to incorporate these and other metadata.

The BROAD dengue sequencing initiative, for example, aims to sequence over 3,500 dengue genomes tagged with information on geographic origin and disease severity (i.e. whether the disease outcome is dengue fever (DF) or the more severe, life-threatening dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS)) in an attempt to determine the impact of introduced strains versus indigenous evolution on disease outcomes, understand genomic correlates of disease severity and provide a map of genomic distributions with reference to DF, DHF and DSS (<http://www.broadinstitute.org/annotation/viral/Dengue/projects.html>). Dengue sequence diversity within individual patients with well-characterised disease outcomes, and for whom time courses for viraemia and status as primary or secondary infections are available, will also be investigated, in order to determine how intra-host diversity drives viraemia and disease and how it correlates with disease severity and primary versus secondary infection.

At the time of writing, the GRID project, which was initiated in 2005, had sequenced 2,372 dengue genomes, IGSP (also launched in the same year) had generated over 3,400 of approximately 7,400 planned genome sequences and there were tens of thousands of HIV sequences available in the Los Alamos database of which 2,788 were HIV1 complete genomes. The current and potential impact of these and other dengue, HIV and influenza sequencing initiatives is well reviewed in Holmes 2009. For dengue, in addition to the previously detailed insights into evolution and epidemiology, analyses suggest that some genotypes differ in virulence and/or fitness (Armstrong and Rico-Hesse 2001; Bennett et al. 2003; Cologna et al. 2005; Cologna and Rico-Hesse 2003; Klungthong et al. 2004; Leitmeyer et al. 1999; Rico-Hesse et al. 1997; Sittisombut et al. 1997; Thu et al. 2004; Wittke et al. 2002; Zhang et al. 2005) and that immune-mediated natural selection may determine which genotypes survive (Adams et al. 2006). Thus the fitness of a given genotype may vary with the changing immunological landscape, which has major implications for vaccine development since tetravalent vaccines designed to induce immunity to all four DENV serotypes are unlikely to provide complete cross-protection (Whitehead et al. 2007). For influenza virus, analysis of IGSP data has already altered basic concepts of influenza virus evolution and shed light on the evolution of drug resistance, identified important source and sink populations and provided data on genomic diversity that will improve and accelerate the process of choosing which strains to incorporate into annual vaccines (reviewed in (Holmes 2009)). HIV is perhaps the greatest disappointment in terms of our inability to arrive at a vaccine despite a wealth of genomic data on the virus. In this regard, the major lesson learned from viral genomics is that HIV is immensely diverse both within and between individual hosts (Rambaut et al. 2004) and vaccines are likely to have to be location/population specific and require regular updating (Holmes 2009).

Conclusions

The ability to generate viral genomes increasingly, rapidly and cheaply and the development of bioinformatic tools for analysing these data have transformed the study of emerging viruses. Metagenomic sequencing and evolutionary analyses will soon become routine diagnostic and surveillance tools, allowing us to detect and visualise viral emergence and spatiotemporal dynamics in real time. In addition to enabling rapid responses in terms of development of pathogen-specific screening tests, identification of source populations and disease tracking, this will facilitate generation of hypotheses about evolutionary mechanisms and ecological factors underlying the patterns observed. However, despite the immense potential, addressing prevention and control issues of more direct clinical relevance such as the development of vaccines and therapeutics will only be possible if genomic data are accompanied by relevant clinical, immunological, phenotypic, host genomic and epidemiological data, with biological measures from *In vitro* and *in vivo* experimental studies incorporated as they arise. The development and maintenance of widely accessible and flexible genomic databases is therefore key in this regard. Furthermore, if we are to avoid the limitations of past efforts, it is essential that data from across the clinical spectrum be included so that the all too common bias towards symptomatic and/or severe cases is avoided. An ideal database would also include viral genomic and corresponding meta-data from animal reservoir and/or vector populations, particularly if our goal is to predict future viral emergence. In addition to traditional sources, these data might be derived from programmes and early warning systems such as the Global Viral Forecasting Initiative (<http://www.gvfi.org/>), USAID PREDICT (<http://www.vetmed.ucdavis.edu/ohi/predict/index.cfm>) and the WHO/FAO/OIE Global Early Warning and Response System (GLEWS; <http://www.glews.net/>), which focus on identification and control of potentially emergent pathogens through surveillance at the animal–human interface.

This is a tall order in terms of the level of coordination and collaboration required to bring all of these data together—public health practitioners, field epidemiologists, clinicians, veterinarians and researchers would all have to work together. More important, however, is the computational challenge. There is no shortage of good ideas, but many of the analyses involved are very computationally intensive, and this is already a limiting factor. Bioinformatic and computational tools will therefore have to further evolve to handle the amounts of genomic and other meta-data generated.

Given our level of globalisation and population mobility (which is only going to increase), it is also essential that all affected geographic regions and populations be represented in these efforts. In addition to providing a complete picture of viral biodiversity against the full span of existing host genomic backgrounds, this will ensure that needs are addressed where the burden of disease is often greatest. It will also reduce the number of surveillance and control ‘blind spots’ where viruses might take refuge and eventually re-emerge. It is therefore essential that developing countries be fully integrated into the genomic age, through collaboration, technology transfer and in-country capacity building. The availability of open source databases, computational tools and scientific literature also goes a long way in this regard.

References

- Adams B, Holmes EC, Zhang C, Mammen MP Jr, Nimmannitya S, Kalayanarooj S et al (2006) Cross-protective immunity can account for the alternating epidemic pattern of dengue virus serotypes circulating in Bangkok. *Proc Natl Acad Sci USA* 103(38):14234–14239
- Allcock OM, Lemey P, Tatem AJ, Pybus OG, Bennett SN, Mueller BA, Suchard MA, Foster JE, Rambaut A, Carrington CV (2012) Phylogeography and population dynamics of Dengue Viruses in the Americas. *Mol Biol Evol* [Epub ahead of print]
- Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 59(1):143–169
- Anderson NG, Gerin JL, Anderson NL (2003) Global screening for human viral pathogens. *Emerg Infect Dis* 9(7):768–774
- Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C et al (2006) The marine viromes of four oceanic regions. *PLoS Biol* 4(11):e368
- Anishchenko M, Bowen RA, Paessler S, Austgen L, Greene IP, Weaver SC (2006) Venezuelan encephalitis emergence mediated by a phylogenetically predicted viral mutation. *Proc Natl Acad Sci USA* 103(13):4994–4999
- Archer BN, Cohen C, Naidoo D, Thomas J, Makunga C, Blumberg L et al (2009) Interim report on pandemic H1N1 influenza virus infections in South Africa. *Epidemiology and Factors associated with fatal cases. Euro Surveill* 14(42):11–19369. <http://www.eurosurveillance.org/Viewarticle.aspx?ArticleId=19369>
- Armstrong PM, Rico-Hesse R (2001) Differential susceptibility of *Aedes aegypti* to infection by the American and Southeast Asian genotypes of dengue type 2 virus. *Vector Borne Zoonotic Dis* 1(2):159–168
- Auguste AJ, Lemey P, Pybus OG, Suchard MA, Salas RA, Adesiyun AA et al (2010) Yellow fever virus maintenance in Trinidad and Its dispersal throughout the Americas †. *J Virol* 84:9967–9977
- Bennett SN, Holmes EC, Chirivella M, Rodriguez DM, Beltran M, Vorndam V et al (2003) Selection-driven evolution of emergent dengue virus. *Mol Biol Evol* 20(10):1650–1658
- Bennett SN, Drummond AJ, Kapan DD, Suchard MA, Munoz-Jordan JL, Pybus OG et al (2010) Epidemic dynamics revealed in dengue evolution. *Mol Biol Evol* 27(4):811–818
- Bexfield N, Kellam P (2011) Metagenomics and the molecular identification of novel viruses. *Vet J* 190(2):191–8
- Biek R, Henderson JC, Waller LA, Rupprecht CE, Real LA (2007) A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. *Proc Natl Acad Sci USA* 104(19):7993–7998
- Bimber BN, Dudley DM, Lauck M, Becker EA, Chin EN, Lank SM et al (2010) Whole-genome characterization of human and simian immunodeficiency virus intrahost diversity by ultra-deep pyrosequencing. *J Virol* 84(22):12087–12092
- Bishop-Lilly KA, Turell MJ, Willner KM, Butani A, Nolan NM, Lentz SM et al (2010) Arbovirus detection in insect vectors by rapid, high-throughput pyrosequencing. *PLoS Negl Trop Dis* 4(11):e878
- Blomström A-L, Widén F, Hammer A-S, Belák S, Berg M (2010) Detection of a novel astrovirus in brain tissue of mink suffering from shaking mink syndrome using viral metagenomics. *J Clin Microbiol* 48:4392–4396
- Bogaert D, Keijsers B, Huse S, Rossen J, Veenhoven R, van Gils E et al (2011) Variability and diversity of nasopharyngeal microbiota in children: a metagenomic analysis. *PLoS One* 6(2):e17035
- Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D et al (2002) Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci USA* 99(22):14250–14255
- Breitbart M, Felts B, Kelley S, Mahaffy JM, Nulton J, Salamon P et al (2004) Diversity and population structure of a near-shore marine-sediment viral community. *Proc Biol Sci* 271(1539):565–574

- Breitbart M, Hoare A, Nitti A, Siefert J, Haynes M, Dinsdale E et al (2009) Metagenomic and stable isotopic analyses of modern freshwater microbialites in Cuatro Ciénegas, Mexico. *Environ Microbiol* 11(1):16–34
- Briese T, Paweska JT, McMullan LK, Hutchison SK, Street C, Palacios G et al (2009) Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. *PLoS Pathog* 5(5):e1000455
- Bright RA, Medina MJ, Xu X, Perez-Oronoz G, Wallis TR, Davis XM et al (2005) Incidence of adamantane resistance among influenza A (H3N2) viruses isolated worldwide from 1994 to 2005: a cause for concern. *Lancet* 366(9492):1175–1181
- Carrington CV, Foster JE, Pybus OG, Bennett SN, Holmes EC (2005) Invasion and maintenance of dengue virus type 2 and type 4 in the Americas. *J Virol* 79(23):14680–14687
- Carrington CV, Foster JE, Zhu HC, Zhang JX, Smith GJ, Thompson N et al (2008) Detection and phylogenetic analysis of group 1 coronaviruses in South American bats. *Emerg Infect Dis* 14(12):1890–1893
- Cologna R, Rico-Hesse R (2003) American genotype structures decrease dengue virus output from human monocytes and dendritic cells. *J Virol* 77(7):3929–3938
- Cologna R, Armstrong PM, Rico-Hesse R (2005) Selection for virulent dengue viruses occurs in humans and mosquitoes. *J Virol* 79(2):853–859
- de Lamballerie X, Leroy E, Charrel RN, Tsetsarkin K, Higgs S, Gould EA (2008) Chikungunya virus adapts to tiger mosquito via evolutionary convergence: a sign of things to come? *Virology* 377(1):1–10
- Deng YM, Caldwell N, Hurt A, Shaw T, Kelso A, Chidlow G et al (2011) A comparison of pyrosequencing and neuraminidase inhibition assays for the detection of oseltamivir-resistant pandemic influenza A(H1N1) 2009 viruses. *Antiviral Res* 90(1):87–91
- Deubel V, Kinney RM, Trent DW (1986) Nucleotide sequence and deduced amino acid sequence of the structural proteins of dengue type 2 virus, Jamaica genotype. *Virology* 155(2):365–377
- Deyde VM, Okomo-Adhiambo M, Sheu TG, Wallis TR, Fry A, Dharan N et al (2009) Pyrosequencing as a tool to detect molecular markers of resistance to neuraminidase inhibitors in seasonal influenza A viruses. *Antiviral Res* 81(1):16–24
- Deyde VM, Sheu TG, Trujillo AA, Okomo-Adhiambo M, Garten R, Klimov AI et al (2010) Detection of molecular markers of drug resistance in 2009 pandemic influenza A (H1N1) viruses by pyrosequencing. *Antimicrob Agents Chemother* 54:1102–1110
- Dharan NJ, Patton M, Siston AM, Morita J, Ramirez E, Wallis TR et al (2009) Outbreak of antiviral drug-resistant influenza A in long-term care facility, Illinois, USA, 2008. *Emerg Infect Dis* 15(12):1973–1976
- Djickeng A, Halpin R, Kuzmickas R, Depasse J, Feldblyum J, Sengamalay N et al (2008) Viral genome sequencing by random priming methods. *BMC Genomics* 9:5
- Djickeng A, Kuzmickas R, Anderson NG, Spiro DJ (2009) Metagenomic analysis of RNA viruses in a fresh water lake. *PLoS One* 4(9):e7264
- Dominguez SR, O'Shea TJ, Oko LM, Holmes KV (2007) Detection of group 1 coronaviruses in bats in North America. *Emerg Infect Dis* 13:1295–1300
- Donaldson EF, Haskew AN, Gates JE, Huynh J, Moore CJ, Frieman MB (2010) Metagenomic analysis of the viromes of three North American bat species: viral diversity among different bat species that share a common habitat. *J Virol* 84(24):13004–13018
- Drosten C, Günther S, Preiser W, van der Werf S, Brodt H-R, Becker S et al (2003) Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med* 348:1967–1976
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214
- Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161(3):1307–1320

- Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22(5):1185–1192
- Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4(5):e88
- Duwe S, Schweiger B (2008) A new and rapid genotypic assay for the detection of neuraminidase inhibitor resistant influenza A viruses of subtype H1N1, H3N2, and H5N1. *J Virol Methods* 153(2):134–141
- Fierer N, Breitbart M, Nulton J, Salamon P, Lozupone C, Jones R et al (2007) Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Appl Environ Microbiol* 73(21):7059–7066
- Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J et al (1976) Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature* 260:500–507
- Gilbert MT, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE, Worobey M (2007) The emergence of HIV/AIDS in the Americas and beyond. *Proc Natl Acad Sci USA* 104(47):18566–18570
- Gloza-Rausch F, Ipsen A, Seebens A, Göttsche M, Panning M, Felix Drexler J et al (2008) Detection and prevalence patterns of group I coronaviruses in bats, northern Germany. *Emerg Infect Dis* 14:626–631
- Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA et al (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303(5656):327–332
- Greninger AL, Chen EC, Sittler T, Scheinerman A, Roubinian N, Yu G et al (2010) A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America. *PLoS One* 5(10):e13381
- Griffiths RC, Tavaré S (1994) Sampling theory for neutral alleles in a varying environment. *Philos Trans R Soc Lond B Biol Sci* 344(1310):403–410
- Hahn BH, Shaw GM, De Cock KM, Sharp PM (2000) AIDS as a zoonosis: scientific and public health implications. *Science* 287(5453):607–614
- Hapuarachchi HC, Bandara KB, Sumanadasa SD, Hapugoda MD, Lai YL, Lee KS et al (2010) Re-emergence of chikungunya virus in South-east Asia: virological evidence from Sri Lanka and Singapore. *J Gen Virol* 91(Pt 4):1067–1076
- Hess M, Sczyrba A, Egan R, Kim T-W, Chokhawala H, Schroth G et al (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* 331:463–467
- Holmes EC (2006) The evolution of viral emergence. *Proc Natl Acad Sci USA* 103(13):4803–4804
- Holmes EC (2009) RNA virus genomics: a world of possibilities. *J Clin Invest* 119:2488
- Holmes EC, Rambaut A (2004) Viral evolution and the emergence of SARS coronavirus. *Philos Trans R Soc Lond B Biol Sci* 359:1059–1065
- Hongoh Y (2011) Toward the functional analysis of uncultivable, symbiotic microorganisms in the termite gut. *Cell Mol Life Sci* 68(8):1311–1325
- Hurt AC, Ernest J, Deng YM, Iannello P, Besselaar TG, Birch C et al (2009) Emergence and spread of oseltamivir-resistant A(H1N1) influenza viruses in Oceania, South East Asia and South Africa. *Antiviral Res* 83(1):90–93
- Jones KE, Patel NG, Levy Ma, Storeygard A, Balk D, Gittleman JL et al (2008) Global trends in emerging infectious diseases. *Nature* 451:990–993
- Kingman J (1982) The coalescent. *Stoch Proc Appl* 13:235–248
- Klunghong C, Zhang C, Mammen MP Jr, Ubol S, Holmes EC (2004) The molecular epidemiology of dengue virus serotype 4 in Bangkok, Thailand. *Virology* 329(1):168–179
- Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A et al (2000) Timing the ancestor of the HIV-1 pandemic strains. *Science* 288(5472):1789–1796
- Kreuze JF, Perez A, Untiveros M, Quispe D, Fuentes S, Barker I et al (2009) Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology* 388(1):1–7
- Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S et al (2003) A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* 348(20):1953–1966

- Kumar NP, Joseph R, Kamaraj T, Jambulingam P (2008) A226V mutation in virus during the 2007 chikungunya outbreak in Kerala, India. *J Gen Virol* 89(Pt 8):1945–1948
- Lackenby A, Democratis J, Siqueira MM, Zambon MC (2008) Rapid quantitation of neuraminidase inhibitor drug resistance in influenza virus quasisppecies. *Antivir Ther* 13(6):809–820
- Lanciotti RS, Gubler DJ, Trent DW (1997) Molecular evolution and phylogeny of dengue-4 viruses. *J Gen Virol* 78(Pt 9):2279–2284
- Lau SKP, Woo PCY, Li KSM, Huang Y, Tsoi H-W, Wong BHL et al (2005) Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc Natl Acad Sci USA* 102:14040–14045
- Leitmeyer KC, Vaughn DW, Watts DM, Salas R, Villalobos I, De C et al (1999) Dengue virus structural differences that correlate with pathogenesis. *J Virol* 73(6):4738–4747
- Lemey P, Rambaut A, Drummond AJ, Suchard Ma (2009) Bayesian phylogeography finds its roots. *PLoS Comput Biol* 5:e1000520
- Lemey P, Rambaut A, Welch JJ, Suchard MA (2010) Phylogeography takes a relaxed random walk in continuous space and time. *Mol Biol Evol* 27:1877–1885
- Lewis JA, Chang GJ, Lanciotti RS, Kinney RM, Mayer LW, Trent DW (1993) Phylogenetic relationships of dengue-2 viruses. *Virology* 197(1):216–224
- Li L, Victoria JG, Wang C, Jones M, Fellers GM, Kunz TH et al (2010a) Bat Guano Virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses. *Society* 84:6955–6965
- Li Y, Ge X, Zhang H, Zhou P, Zhu Y, Zhang Y et al (2010b) Host range, prevalence, and genetic diversity of adenoviruses in bats. *J Virol* 84(8):3889–3897
- Lisitsyn N, Lisitsyn N, Wigler M (1993) Cloning the differences between two complex genomes. *Science* 259(5097):946–951
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380
- Morens DM, Folkers GK, Fauci AS (2004) The challenge of emerging and re-emerging infectious diseases. *Nature* 430(6996):242–249
- Nakamura S, Yang CS, Sakon N, Ueda M, Tougan T, Yamashita A et al (2009) Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PLoS One* 4(1):e4219
- Nelson MI, Simonsen L, Viboud C, Miller MA, Holmes EC (2007) Phylogenetic analysis reveals the global migration of seasonal influenza A viruses. *PLoS Pathog* 3(9):1220–1228
- Ng TF, Manire C, Borrowman K, Langer T, Ehrhart L, Breitbart M (2009a) Discovery of a novel single-stranded DNA virus from a sea turtle fibropapilloma by using viral metagenomics. *J Virol* 83(6):2500–2509
- Ng TF, Suedmeyer WK, Wheeler E, Gulland F, Breitbart M (2009b) Novel anellovirus discovered from a mortality event of captive California sea lions. *J Gen Virol* 90(Pt 5):1256–1261
- Ng LC, Tan LK, Tan CH, Tan SS, Hapuarachchi HC, Pok KY et al (2009c) Entomologic and virologic investigation of chikungunya, Singapore. *Emerg Infect Dis* 15(8):1243–1249
- Palacios G, Druce J, Du L, Tran T, Birch C, Briese T et al (2008) A new arenavirus in a cluster of fatal transplant-associated diseases. *N Engl J Med* 358:991–998
- Parrish CR, Holmes EC, Morens DM, Park E-C, Burke DS, Calisher CH et al (2008) Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol Mol Biol Rev* 72:457–470
- Peiris JSM, Lai ST, Poon LLM, Guan Y, Yam LYC, Lim W et al (2003) Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* 361:1319–1325
- Poon LLM, Chu DKW, Chan KH, Wong OK, Ellis TM, Leung YHC et al (2005) Identification of a novel coronavirus in bats. *J Virol* 79:2001–2009
- Pybus OG, Rambaut A (2009) Evolutionary analysis of the dynamics of viral infectious disease. *Nat Rev Genet* 10:540–550
- Rambaut A, Posada D, Crandall KA, Holmes EC (2004) The causes and consequences of HIV evolution. *Nat Rev Genet* 5(1):52–61

- Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC (2008) The genomic and epidemiological dynamics of human influenza A virus. *Nature* 453(7195):615–619
- Rico-Hesse R, Harrison LM, Salas RA, Tovar D, Nisalak A, Ramos C et al (1997) Origins of dengue type 2 viruses associated with increased pathogenicity in the Americas. *Virology* 230(2):244–251
- Russell CA, Jones TC, Barr IG, Cox NJ, Garten RJ, Gregory V et al (2008) The global circulation of seasonal influenza A (H3N2) viruses. *Science* 320(5874):340–346
- Sam IC, Chan YF, Chan SY, Loong SK, Chin HK, Hooi PS et al (2009) Chikungunya virus of Asian and Central/East African genotypes in Malaysia. *J Clin Virol* 46(2):180–183
- Schliekelman P, Garner C, Slatkin M (2001) Natural selection and resistance to HIV. *Nature* 411:545–546
- Schuffenecker I, Itean I, Michault A, Murri S, Frangeul L, Vaney MC et al (2006) Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. *PLoS Med* 3(7):e263
- Seib KL, Dougan G, Rappuoli R (2009) The key role of genomics in modern vaccine and drug design for emerging infectious diseases. *PLoS Genet* 5:e1000612
- Sittisombut N, Sistayanarain A, Cardoso MJ, Salminen M, Damrongdachakul S, Kalayanaroj S et al (1997) Possible occurrence of a genetic bottleneck in dengue serotype 2 viruses between the 1980 and 1987 epidemic seasons in Bangkok, Thailand. *Am J Trop Med Hyg* 57(1):100–108
- Storch G (2007) Diagnostic Virology. In: Knipe D, Howley P (eds) *Field's Virology*. Lippincott/Williams & Wilkins, Philadelphia, pp 565–604
- Talbi C, Lemey P, Suchard Ma, Abdelatif E, Elharrak M, Jalal N et al (2010) Phylodynamics and human-mediated dispersal of a zoonotic virus. *PLoS Pathog* 6:e1001166
- Tang XC, Zhang JX, Zhang SY, Wang P, Fan XH, Li LF et al (2006) Prevalence and genetic diversity of coronaviruses in bats from China. *Society* 80:7481–7490
- Taylor LH, Latham SM, Woolhouse ME (2001) Risk factors for human disease emergence. *Philos Trans R Soc Lond B Biol Sci* 356(1411):983–989
- Thu HM, Lowry K, Myint TT, Shwe TN, Han AM, Khin KK et al (2004) Myanmar dengue outbreak associated with displacement of serotypes 2, 3, and 4 by dengue 1. *Emerg Infect Dis* 10(4):593–597
- Tong S (2009) Detection of novel SARS-like and other coronaviruses in bats from Kenya. *Emerg Infect Dis* 15:482–485
- Tsetsarkin KA, Vanlandingham DL, McGee CE, Higgs S (2007) A single mutation in chikungunya virus affects vector specificity and epidemic potential. *PLoS Pathog* 3(12):e201
- Tsetsarkin Ka, Chen R, Leal G, Forrester N, Higgs S, Huang J et al (2011) Chikungunya virus emergence is constrained in Asia by lineage-specific adaptive landscapes. *Proc Natl Acad Sci USA* 108(19):7872–7877
- Twiddy SS, Farrar JJ, Vinh Chau N, Wills B, Gould EA, Gritsun T et al (2002) Phylogenetic relationships and differential selection pressures among genotypes of dengue-2 virus. *Virology* 298(1):63–72
- Vazeille M, Moutailler S, Coudrier D, Rousseaux C, Khun H, Huerre M et al (2007) Two Chikungunya isolates from the outbreak of La Reunion (Indian Ocean) exhibit different patterns of infection in the mosquito, *Aedes albopictus*. *PLoS One* 2(11):e1168
- Victoria JG, Kapoor A, Dupuis K, Schnurr DP, Delwart EL (2008) Rapid identification of known and new RNA viruses from animal tissues. *PLoS Pathog* 4(9):e1000163
- Vidal N, Peeters M, Mulanga-Kabeya C, Nzilambi N, Robertson D, Ilunga W et al (2000) Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. *J Virol* 74(22):10498–10507
- Wang D, Coscoy L, Zylberberg M, Avila PC, Boushey HA, Ganem D et al (2002) Microarray-based detection and genotyping of viral pathogens. *Proc Natl Acad Sci USA* 99(24):15687–15692

- Wang D, Urisman A, Liu Y-T, Springer M, Ksiazek TG, Erdman DD et al (2003) Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol* 1:E2
- Weiss Ra, McMichael AJ (2004) Social and environmental risk factors in the emergence of infectious diseases. *Nat Med* 10:S70–S76
- Whitehead SS, Blaney JE, Durbin AP, Murphy BR (2007) Prospects for a dengue virus vaccine. *Nat Rev Microbiol* 5(7):518–528
- Williamson HR, Benbow ME, Nguyen KD, Beachboard DC, Kimbirauskas RK, McIntosh MD et al (2008) Distribution of *Mycobacterium ulcerans* in buruli ulcer endemic and non-endemic aquatic sites in Ghana. *PLoS Negl Trop Dis* 2(3):e205
- Willner D, Furlan M, Haynes M, Schmieder R, Angly FE, Silva J et al (2009) Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLoS One* 4(10):e7370
- Witke V, Robb TE, Thu HM, Nisalak A, Nimmannitya S, Kalayanrooj S et al (2002) Extinction and rapid emergence of strains of dengue 3 virus during an interepidemic period. *Virology* 301(1):148–156
- Wolfe ND, Switzer WM, Carr JK, Bhullar VB, Shanmugam V, Tamoufe U et al (2004) Naturally acquired simian retrovirus infections in central African hunters. *Lancet* 363(9413):932–937
- Wolfe ND, Daszak P, Kilpatrick AM, Burke DS (2005) Bushmeat hunting, deforestation, and prediction of zoonoses emergence. *Emerg Infect Dis* 11(12):1822–1827
- Wolfe ND, Dunavan CP, Diamond J (2007) Origins of major human infectious diseases. *Nature* 447(7142):279–283
- Woo PCY, Lau SKP, Li KSM, Poon RWS, Wong BHL, Tsoi H-w et al (2006) Molecular diversity of coronaviruses in bats. *Virology* 351:180–187
- Woolhouse ME (2002) Population biology of emerging and re-emerging pathogens. *Trends Microbiol* 10(10 Suppl):S3–S7
- Woolhouse M, Gaunt E (2007) Ecological origins of novel human pathogens. *Crit Rev Microbiol* 33(4):231–242
- Woolhouse ME, Haydon DT, Antia R (2005) Emerging pathogens: the epidemiology and evolution of species jumps. *Trends Ecol Evol* 20(5):238–244
- Wu Q, Luo Y, Lu R, Lau N, Lai EC, Li WX et al (2010) Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proc Natl Acad Sci USA* 107(4):1606–1611
- Zhang C, Mammen MP Jr, Chinnawirotpisan P, Klungthong C, Rodpradit P, Monkongdee P et al (2005) Clade replacements in dengue virus serotypes 1 and 3 are associated with changing serotype prevalence. *J Virol* 79(24):15123–15130
- Zhu T, Korber BT, Nahmias AJ, Hooper E, Sharp PM, Ho DD (1998) An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* 391(6667):594–597