

Epidemiological and clinical consequences of within-host evolution

Samuel Alizon¹, Fabio Luciani² and Roland R. Regoes³

¹Laboratoire Génétique et Évolution des Maladies Infectieuses, Unité Mixte de Recherche du Centre national de la Recherche Scientifique (CNRS) et de l'Institut de Recherche pour le Développement (IRD) 2724, 911 avenue Agropolis, BP 64501, 34394 Montpellier CEDEX 5, France

²Infection and Inflammation Research Centre, School of Medical Sciences, Faculty of Medicine, University of New South Wales, Sydney 2052, Australia

³Institut für Integrative Biologie, Eidgenössische Technische Hochschule (ETH), Universitätstrasse 16, 8092 Zürich, Switzerland

Many viruses and bacteria are known to evolve rapidly over the course of an infection. However, epidemiological studies generally assume that within-host evolution is an instantaneous process. We argue that the dynamics of within-host evolution has implications at the within-host and at the between-host levels. We first show that epidemiologists should consider within-host evolution, notably because it affects the genotype of the pathogen that is transmitted. We then present studies that investigate evolution at the within-host level and examine the extent to which these studies can help to understand infection traits involved in the epidemiology (e.g. transmission rate, virulence, recovery rate). Finally, we discuss how new techniques for data acquisition can open new perspectives for empirical and theoretical research.

Evolution is a dynamic process

In epidemiology, evolutionary processes occurring during an infection are usually considered to be instantaneous: the host is assumed to switch from the category 'infected by strain A' to the category 'infected by strain B'. This simplifying view can be inaccurate for two reasons. First, even if a mutant emerges in an infected host, it might not have the time to replace the resident strain before the infection ends. Second, for some rapidly evolving diseases, many mutations can arise, leading to complex dynamics with successive replacements of pathogen strains. Even if epidemiologists and clinicians are aware of the importance of within-host evolution, the fact remains that most models summarise it into a single mutation parameter. In this review we argue that epidemiology can benefit from a better understanding of within-host evolutionary dynamics (WHED), in other words how the size and the genetic composition of the pathogen population change over the course of an infection.

We focus on infectious diseases of humans, including HIV [1], hepatitis C virus (HCV [2]), and microparasites causing persistent infections (e.g. bacteria such as *Helicobacter pylori* [3]). Strikingly, except for intensively studied pathogens such as HIV or HCV, there has been (and largely still is) a clear lack of data on WHED. The rise of new experimental techniques could provide a much more accurate picture of within-host pathogen diversity [4,5] but it

is unlikely that the current lack of data is solely due to technical limitations.

We first discuss some epidemiological consequence of WHED. We then show how earlier work at the within-host level can be used to understand components of the pathogen epidemiological fitness (Box 1). Finally, we outline promising directions for theoretical and experimental research.

Evolution at the between-host level

The implications of evolutionary processes taking place over the course of an infection go beyond the level of a single host and affect the whole host population. The development of the field of evolutionary epidemiology has been based on the observation that trait evolution and disease spread are linked [6]: the epidemiology shapes the selective pressures acting on a pathogen and the resulting evolution of the pathogen modifies the epidemiology (Box 1), which generates a long-term feedback loop.

A first family of models in evolutionary epidemiology, molecular epidemiology models, study how epidemiological processes shape pathogen genomes [7]. In these models pathogen strains differ by their DNA, RNA, or amino acid sequences. A subset of these molecular evolution models, phylodynamics models [8,9], use phylogenies to infer population (or epidemiological) dynamics. We will often refer to these because of the emphasis they put on dynamic processes. A second family of evolutionary epidemiology models tends to leave aside the molecular aspects to focus on phenotypic evolution [6]. In these models strains differ by their phenotypic trait (usually their virulence, transmission or recovery rates). Some of these models, notably nested models, particularly fit the scope of this review because they incorporate within-host dynamics within an evolutionary epidemiology setting [10].

What gets transmitted?

A consequence of within-host evolution is that the pathogen type that enters a host is likely to differ from the type that is transmitted onwards. For instance, during an infection a pathogen can acquire mutations that allow it to better evade the host immune response, but it is not obvious that these escape mutations will be transmitted to the next host. The few nested models that combine within-host evolution and between-host transmission suggest that it is difficult to

Corresponding author: Alizon, S. (samuel.alizon@montp.cnrs.fr).

Box 1. Evolutionary epidemiology

Three conditions are required for natural selection to act on a trait. The trait must be variable in the population, it must affect the fitness of the individual bearing that trait, and it must be heritable. In the case of diseases, we consider that an individual corresponds to an infection by a pathogen of a given genotype. Pathogen fitness is the number of secondary infections generated during the whole infection, usually equivalent to the R_0 [64]. The heritability of a trait is then measured from a donor infection to a receiver infection.

The phenotypic traits of an infection that are usually considered are those that affect the fitness (R_0), in other words the transmission rate (β), the recovery rate (γ) and the virulence (α). The classical formula linking these fitness components is:

$$R_0 = \frac{\beta S}{\mu + \alpha + \gamma} \quad \text{[Equation 1]}$$

where S is the density of susceptible hosts and μ is the baseline mortality of the host. This formula can be understood as the number of transmission events per unit of time (βS) multiplied by the expected duration of an infection [$1/(\mu + \alpha + \gamma)$].

Note that WHED imply that there is within-host genetic diversity. This makes the equivalence between an infection and an individual on which natural selection acts more complicated. Adopting a nested approach is one option to solve this problem by tracking the density of each genotype in each infected host [15].

predict which strain can invade the host population [11,12], a result that also arises from data analysis [13,14].

WHED have implications for traits other than immune escape, for example replication rate [11,15,16]. Studies find that the average replication rate of HIV tends to increase over the course of the infection [17,18]. Figure 1 illustrates the central problem: the nature of the transmitted strain (here its replication rate) depends on when the transmission event occurs.

Considering transmission pairs is a way to analyse the transmission process. One study showed that founder-transmitted HIV viruses, which successfully establish a

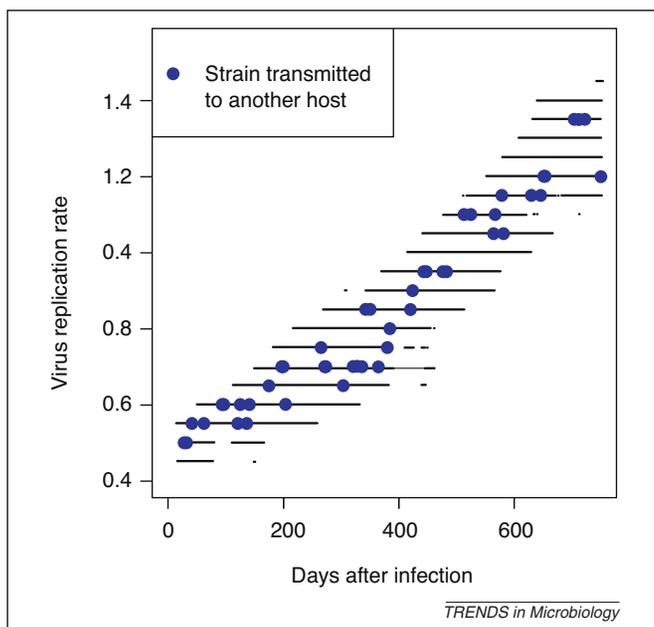


Figure 1. Evolutionary dynamics of virus replication rate within a host. Black dots indicate trait values that are present in the host; blue dots indicate transmitted strains. The replication rate increases over the course of the infection. Depending on when transmission occurs, the trait of the transmitted strain differs more or less from the ancestral trait. Figure modified from [11].

new infection upon transmission, differ from ‘average’ viruses found in donor hosts [19]. In particular, founder viruses tended to display reduced glycosylation of the surface glycoprotein, which facilitates virus binding to CD4 receptors and entry into T cells. This observation shows that there is a bottleneck upon transmission and that the genetic composition of the initial virus population in the next host does not simply reflect that in the donor host (i.e. some strains are proportionally more transmitted).

The focus on transmitted strains is also at the heart of studies on the heritability, from one infection to the next, of set-point viral load (a trait that predicts the virulence of the infection). Data analysis shows that a large fraction of the variance in set-point virus load is heritable – in other words is governed by the genotype of the infecting virus [20,21]. This is important because within-host evolution and host or environmental factors could completely buffer the heritability that is a prerequisite for selection to act on the trait (Box 1). However, including this trait heritability in epidemiological models still remains an open problem.

Conflicts between levels of selection

One way to predict which strain will be transmitted is to consider the conflict that arises because a strategy that is optimal within a host can be poorly adapted at the between-host level. For instance, growing rapidly allows a genotype to win the within-host competition, but this can also lead to rapid host death, thereby reducing the likelihood of infecting a new host [22].

In the case of HIV the conflict between levels of selection is evident from the shape of between-host phylogenies – interior branches (corresponding to between-host evolution) tend to be shorter than terminal branches (corresponding to within-host evolution). Diverse mechanisms have been invoked to explain this difference in evolutionary rates [1,9]. It is conceivable that reverse mutations occur upon infection. Another hypothesis is that only slowly-evolving strains are capable of initiating an infection. More generally, there are likely to be trade-offs associated with levels of selection. We already mentioned that HIV, transmitted viruses tend to have reduced envelope protein glycosylation. In addition to facilitating the infection of host T cells, this also makes the virus more vulnerable to attack by host antibodies [19].

Finally, the conflict between levels of selection does not necessarily have a clear-cut outcome. For instance, a mathematical model of HCV infection finds that the efficiency of the immune response strongly determines whether the dominant circulating virus in the host population is more adapted to the within-host or to the between-host level [11].

Overlap between within- and between-host dynamics

In Box 2 we develop a model to compare cases where a drug-resistant mutant strain takes over treated hosts almost instantaneously with cases where strain replacement takes several days. In the model, replacement time is governed by the within-host fitness advantage of the mutant. We show that taking into account WHED slows the spread of the drug-resistant strain in the population (Figure 1b in Box 2).

Box 2. WHED affect the speed of spread of a mutant

We use a simple epidemiological setting, inspired from [15], where hosts can be infected by two types of strains: a resident, or wild-type, strain (denoted r) and a mutant strain (denoted m). These strains have the same virulence (α) but the mutant strain has a higher transmission rate than the resident strain ($\beta_m > \beta_r$). This is the case, for instance, if the mutant is drug-resistant and if all infected hosts are treated. We therefore know that the mutant strain should ultimately oust the resident. The output of the model is the time taken for this replacement to occur.

Initially, we assume that all the hosts are susceptible to the disease and that one host is infected by the resident strain. The mutant strain evolves from the resident strain by mutation. After a mutation event, a host that was infected by the resident strain becomes doubly infected and can transmit both the resident and mutant strains to a susceptible host (the nature of the transmitted strain depends on the within-host densities of the strains). We consider here that the mutant has a within-host competitive fitness advantage (ΔW), which means that it

will eventually oust the resident strain from doubly-infected hosts. Classical models assume that this advantage is infinite so that the mutant replaces the resident instantaneously (the host switches from I_r to I_m as soon as the mutation event occurs). We study how the time taken for the mutant to take over the host reflects on the time required for the mutant to oust the resident from the host population. The formula and the mathematical details are given in the [supplementary material online](#).

In [Figure 1a](#), within-host dynamics are shown. Increasing the within-host competitive advantage of the mutant (ΔW) leads to a more rapid ousting of the resident strain. [Figure 1b](#) shows the epidemiological dynamics – the proportion of hosts infected by the mutant strain among all the infected hosts (values of ΔW are identical to [Figure 1a](#)). Decreasing the within-host fitness advantage by tenfold can slow the spread by 25%. Therefore, ignoring transient within-host dynamics (i.e. assuming that the within-host fitness advantage of the mutant ΔW is infinite) can lead to overestimation of the speed of spread of a mutant strain.

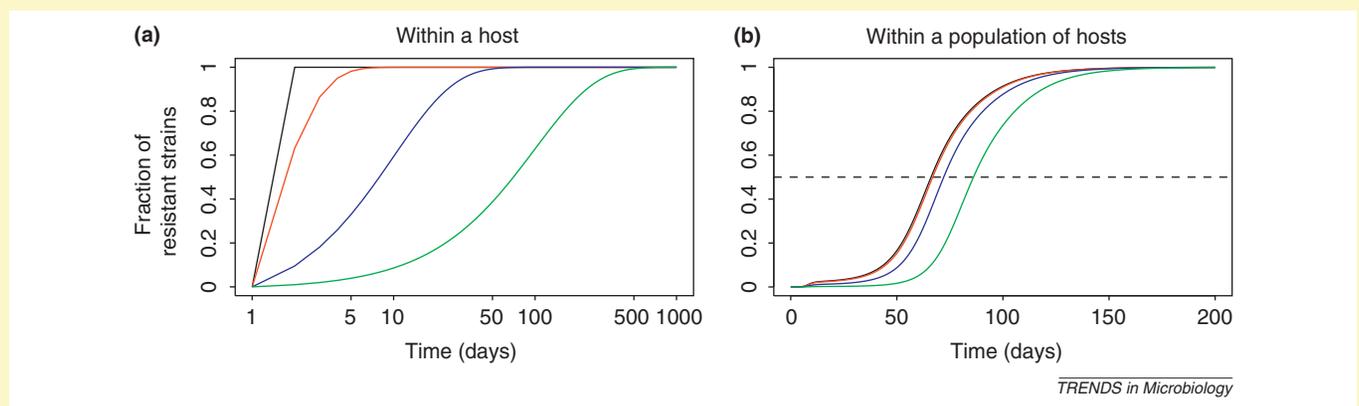


Figure 1. Invasion of a drug-resistant mutant strain. (a) Within a host and (b) in a population of hosts. Different colours correspond to different within-host fitness advantages (ΔW) of the mutant: in black $\Delta W = 10$, in red $\Delta W = 1$, in blue $\Delta W = 0.1$, and in green $\Delta W = 0.01$. In (b) the horizontal dashed line indicates that half of the infected hosts are infected by the mutant strain.

Even when summarising WHED with a single mutation parameter, assuming large within-host mutation rates should also affect epidemiological traits (e.g. virulence) over the course of an infection [23]. This means, such models should consider variations of average infection trait value at the between-host level on short time-scales. Making WHED explicit is a way to include this short-term evolution.

Finally, the overlap between the within-host and the between-host scales is linked to the heterogeneity of selective pressures. In the case of traits such as drug resistance the selective pressure at the between-host level is more or less uniform among treated hosts, and heterogeneity arises because not all the hosts are treated. Consequently, the pathogen evolves at a global scale. By contrast, in the case of immune escape virus evolution is significantly affected by the local level (individual hosts) because the adaptive immune response in each host is influenced by the high polymorphism of the human leukocyte antigen (HLA) alleles in the host population. Data shows that drug resistance can spread rapidly whereas the co-evolution between pathogens and the immune system is an ongoing process [24]. An interesting direction for future work would be to identify the most appropriate measures to fight diseases evolving under local or global selective pressure.

The fact that disease evolution within the host is not instantaneous has implications at the epidemiological

level. First, it affects the timing of events and thus the speed of spread of the disease. Second, it also affects the linkage between the within-host and the between-host level (i.e. the heritability of pathogen traits from one infection to the next), a key point for the understanding of rapidly evolving disease.

Evolution at the within-host level

WHED are more common in studies at the host level ([Box 3](#)), most likely because of their clinical implications. In this section we identify important evolutionary questions that have been addressed at the within-host level and that could have implications at the between-host level.

Natural selection or drift?

Evolution does not necessarily imply natural selection. Discrimination between genetic drift and natural selection generally relies on two summary statistics. First, the effective population size (denoted N_e), which is a population genetics measure that describes the population size that would be required to explain the observed dynamics with an idealised model. Although it is difficult to measure N_e for pathogens such as HIV, existing estimates suggest that it is close to 10^3 [25]. There is thus more room for stochastic processes than the 'real' within-host population size of HIV ($\approx 10^8$ infected cells) might suggest.

Box 3. Processes affecting within-host evolution

Within-host heterogeneity: heterogeneity is a constitutive feature of the within-host environment. First, hosts have different types of cells a pathogen can exploit. Second, these cells tend to be clustered in different organs, and this creates a spatial structure. In the case of HIV the different structures of the spleen and the brain can lead to genetic drift [27,28], and virus replication rates also appear to vary in different regions of the gut [65]. HIV is known to infect several different types of cells, and there have been efforts to link this target cell diversity to clinical observations, especially to the coreceptor switch [66].

Host treatment also adds spatial structure. Drug concentrations vary among organs because treatments are administered to a specific compartment of the body, and this complicates predictions about the evolution of drug resistance [35,44]. Most data concern HIV [60] however one study was based on six *Mycobacterium tuberculosis* (TB)-infected patients suggested that several transmitted strains of *M. tuberculosis* independently evolved drug resistance in the lungs during treatment [58]. This study also showed that there is a diversity of niches within the lungs driven by immunological processes, and that drug resistance is more likely to occur in niches with high bacterial growth.

Co-evolutionary-like dynamics: the immune system acts as a strong selective pressure driving pathogen evolution. This has led to the evolution of many mechanisms of immune evasion such as antigenic variation, or antigen presentation manipulation (see [24] for a review).

The interaction between pathogens and host defences shares many similarities with a co-evolutionary process. Evolution allows pathogens to evade host defences. The host can respond to this escape if its generation time is similar to that of the pathogen, which is the basis of co-evolution models [67]. Within a host, similar processes can occur through adaptive immunity: an initial pathogen strain activates a lymphocyte clone and escape variants will elicit different lympho-

cyte clones. This arms race between host and pathogens can be shown experimentally by challenging viruses isolated at different stages of disease progression with patient sera that were also taken at different stages [61].

A key issue to model this co-evolution is to identify the dimensionality of antigenic space [68]. A study analysing data of influenza A antigens from 1968 to 2003 shows that using two dimensions captures the strains structure of the virus [69]. However, the authors did not consider other dimensions, and the optimal number of dimensions required to best reflect the antigenic space is still an open question.

If assumptions are made concerning the dimension and the topology of the antigenic space, one can then study co-evolutionary dynamics. For instance, a study defining the antigen of a virus strain in an infinite one-dimensional space showed that cross-reactive immunity is a key component to the co-evolutionary dynamics [70]. Less abstractly, it is possible to model the within-host dynamics if we assume that all the particular strains and immune responses that will arise in a system are known [71].

Within-host trade-offs: a growing number of studies allow for pathogen traits to take many values. One of the first theoretical frameworks, known as the 'short-sighted' evolution [22], posits that rapidly replicating strains out-compete slower replicating strains, such that replication rate should increase during an infection [72].

Recent models have shown that within-host trade-offs, for example between the virus replication rate and the lifespan of an infected cell, can alter this evolutionary outcome and favour strains with intermediate replication rates [73,74]. In fact, different outcomes can be observed depending on which parameters are involved in the trade-off [52].

Second, the dN/dS ratio – the ratio of non-synonymous mutations to synonymous mutations. Synonymous mutations can be assumed to be largely neutral because they do not affect the amino-acid sequence. Therefore, if the dN/dS ratio is greater than 1 there is positive selection, and if it is lower than 1 there is purifying selection (if it is equal to 1 evolution is neutral). In the case of HIV, a seminal study showed that during an infection the *nef* gene of viruses sampled in the blood is under positive selection to escape cytotoxic T lymphocytes [26]. However, in other organs, such as the spleen [27] or the brain [28], only drift seems to be at work. These variations in selective pressures could be linked to differences in host compartments, as we discuss below. They also illustrate the gaps in our understanding of the progression to AIDS observed in HIV patients. In fact, the exact role played by virus evolution in HIV infection remains largely unknown [29].

The method based on dN/dS ratios can be used to study more complex traits such as infection duration. HCV infections can be acute and resolve naturally or progress to chronic infection. Results show that infections that will become chronic exhibit more non-synonymous substitutions in one region of the virus genome in the early stages of the infection than in acute infections [30]. This suggests that virus adaptation partly controls the outcome of the infection. Of course, other factors including host factors [31] are involved. A challenge would be to determine which steps of the infection process involve virus evolution.

Longitudinal study of infected patients can provide useful clinical information. In cystic fibrosis patients the comparison of samples of *Pseudomonas aeruginosa* isolated after six months or after 96 months showed important differences due to positive selection [32]. In particular,

genes associated with virulence factors, which are important for acute infections, were selected against over the course of a chronic infection. Insight can also be gained from experiment. In mice, within-host evolution of the bacterium *Haemophilus influenzae* was found to explain part of the outcome of an infection, but stochasticity and host susceptibility also play an important role in explaining why a commensal strain becomes invasive [33].

Arguably, there is a tendency to focus on natural selection because predictions are easier to make. However, studying genetic drift can have public health outcomes. The speed of drift (i.e. the time it takes for an allele to proceed to fixation) is likely to differ between pathogens. Furthermore, drift at the within-host level could have unexpected epidemiological consequences. In the case of *Neisseria meningitidis* a simple stochastic epidemiological model that summarises all the evolution with a single mutation parameter can account for rare meningitis outbreaks with high prevalence and common outbreaks with very low prevalence [34]. Studying links between within-host drift and epidemiology could help in estimating the risk of outbreaks of pathogenic strains.

Selection and within-host dynamics

The type of within-host dynamics is known to affect the suggested epidemiology of a disease [10]. However, these dynamics can also affect within-host evolution, as in the case of drug resistance.

Mathematical models are commonly used to compare and optimise drug therapies [35]. Some of these models account for evolutionary dynamics and show that WHED affect the optimal timing of treatment against HIV [36]. The two main factors that limit pathogen growth in a host

are resource availability and the immune response. A theoretical study shows that the type of limitation can affect the evolution of drug resistance [37]. In particular, the authors predicted that, in absence of immune limitation, the whole bacterial population becomes resistant two to three days after the initiation of drug treatment. Including immunity in their model leads to longer times of emergence of resistance and, for some immune-response functions, the synergy between host immunity and the drug can prevent the emergence of resistance. Finally, their results have implications regarding the optimal dosage of the drug and the timing of administration (both of which are epidemiological parameters).

Specifying the type of within-host dynamics at work for rapidly evolving pathogens is a challenging task. A current methodology consists in fitting within-host dynamics models to data on pathogen and host-cell densities and to identify which assumptions provide the best fit. This approach was used for instance to show that early replication of SIV in macaques is largely controlled by CD8⁺ T cells and not by target cell limitation [38]. However, more work is needed to make sure that this methodology is applicable to rapidly mutating viruses.

Another approach involves dN/dS ratios. The idea is that immune limitation should lead to patterns of positive selection in the viral genome, whereas target-cell limitation should lead to stabilising selection. The problem is that the dN/dS ratio ignores a large amount of genetic information and that it might prove to be limited when analysing single populations [39].

We suggest a third approach inspired from the phylogenetics program in evolutionary epidemiology. This program hypothesises that the shape of the phylogeny also contains valuable information. For example, short infections with partial cross-immunity, such as those caused by influenza A, are under ongoing immune selection, and this is reflected in the shape of some viral gene phylogenies that contain many short branches arising from a central trunk [8].

We think that phylodynamics can be used to infer underlying within-host processes (as illustrated in Box 4), but many caveats remain. For instance, two different models might generate similar phylogenies. Further complexity could arise due to the host immune response. In the case of HCV, for instance, host genes partly control the acute versus chronic outcome of infection [31]. Overall, we still need to determine how within-host processes (e.g. target-cell limitation or not, type of growth function of immune cells) affect the topological and dynamic properties of phylogenies and how robust these properties are to parameter variations (especially the mutation rate). We cannot yet affirm that phylodynamics will provide clinically relevant insights but we think it is a promising method to include WHED in a model-fitting approach.

Host compartments

We have so far implicitly considered the human host as a homogeneous environment. However, in addition to a huge variety of cells, there is also spatial structure. Within-host heterogeneity is known to govern WHED (Box 3) but it can also affect key components of the infection fitness, for

example virulence, transmission rate and recovery rate (Box 1). More precisely, each of these fitness components involves one (or several) specific host compartments. In the case of HCV, sexual organs are a relevant compartment for transmission but the liver is the relevant organ for understanding pathogenesis. Most studies at the within-host level are interested in clinical implications of WHED and therefore tend to concentrate on compartments involved in pathogenesis (or, alternatively, on compartments that are easier to monitor).

Host compartments raise again the question of the role of evolution. Some HIV strains infect the central nervous system, leading to neuropathogenesis. However, there is no clear evidence of tissue-specific adaptation to replication [28]. *N. meningitidis* is a commensal bacterium of the nasopharynx in humans but in some rare cases it can lead to pathogenesis when bacteria colonise the blood and the cerebrospinal fluid. This invasion of new tissues is likely to involve the evolution of specific mechanisms [22], for instance to avoid the immune response. It has been shown that a serogroup of *N. meningitidis* that caused many outbreaks of meningitis has a high prevalence of mutators [40]. These high mutation rates allow the rapid expression or repression of contingency loci in the genome – regions containing virulence factors that could be adaptive in some compartments. In this case pathogen evolution and host structure seem to interact to govern pathogenesis.

Collecting pathogens from several compartments generates rich datasets, as illustrated by a recent clinical study on *Burkholderia pseudomallei* which causes melioidosis in humans [41]. Isolating strains from different sites within the same patient allowed the authors to build a within-host phylogeny; this was used to determine the site of initial infection and also, in some cases, the course of the infection (why some organs were infected). An extension of this study would be to specify the roles of the compartments with respect to transmission, virulence and recovery, and to compare the genetic composition of the pathogen population in these compartments.

Several mathematical models have combined within-host heterogeneity and WHED [42–44] but more insight could be gained from the literature in evolutionary ecology. Some aspects of the evolution of dispersal (such as habitat saturation or specialisation) could be applicable to diseases [45]. Furthermore, much of the kin selection theory deserves to be extended to the within-host level. Ecological studies show that spatial structure is one of the major explanations to why individuals tend to interact with kin. This allows cooperative behaviours to resist the exploitation by cheaters [46]. Within-host structure can thus affect the average level of relatedness of an infection, which is known to be linked to the virulence [47].

Studies on the clinical implications of WHED can provide epidemiological insights. For instance, drift within the host could affect epidemiological processes, such as the periodicity of outbreaks of pathogenic strains. Also, phylodynamic approaches might inform us about how the infection unfolds, and this is necessary in order to include WHED into an epidemiological framework. Finally, more studies are required to analyse WHED in specific compartments and to incorporate them into an epidemiological setting.

Box 4. Within-host dynamics and phylogeny shape

Current techniques used to infer within-host processes from time-series of pathogen densities are probably not extendable to rapidly evolving viruses. We argue that insight could be gained from phylodynamics because the nature of the interactions between the pathogen and the host cells is likely to impact upon the shape of the within-host phylogeny. To illustrate our point, we build four phylogenies using two mathematical models and two clinical datasets. We also present the distribution of the log of the branch lengths to illustrate the summary statistics that can potentially be used to compare phylogenies.

The first model is based on an earlier framework [11] and describes the interaction between a pathogen and immune cells. The second model describes competition between pathogen for host target cells and it is also based on previous models [42,74]. In both models a pathogen trait evolves during the infection. In the first model the pathogen trait is the antigen (protein or peptide). Each antigen has a propensity to activate a repertoire of immune cells, for instance lymphocyte clones. This generates co-evolutionary-like dynamics between pathogen and the immune system (Box 3). In the second model, the trait that evolves is the number of pathogens produced per infected cell. In both models pathogen diversity is generated through stochastic mutation events. Twenty within-host phylogenies are built from the history of mutation events over the first 365 days of infection in each case (Figure 1). Further details about the model are available in the supplementary material online.

The two within-host models generate different patterns that reflect two distinct evolutionary processes. In the immune-limited model

(Figure 1a) the shape of the phylogeny seems to indicate purifying selection. In the resource-competition model, the shape of the tree is closer to stabilising selection (Figure 1b). The distribution of branch lengths differs significantly between the two cases. However, proving that different within-host processes leave different signatures in the pathogen genome will require a demonstration that the differences are robust to parameter variations or to calibrate our models with accurate within-host data.

We also inferred a virus phylogeny using public sequence data [55,30] for three patients infected by HCV and five patients infected by HIV (Figure 1c,d). In both cases the distribution of branch lengths was closer to that expected under immune limitation. In the case of HIV, this is consistent with previous studies [62].

As stressed in the main text, these results are purely illustrative and further work is needed to understand potential biases (host effects, variation in parameter values) and to determine the key properties that are best suited for comparing phylogenies obtained with different within-host processes. In addition, several processes could be at work at the same time [75] and this could blur potential signals. Finally, virus replication rate can evolve during an infection [17,18]. If rapidly replicating viruses are more readily recognised and controlled by host immune responses (see [11] for discussion of possible mechanisms), this will generate shorter inner branches in the within-host phylogeny.

Understanding these within-host mechanisms could open new routes for the design of pathogen-specific therapies [37].

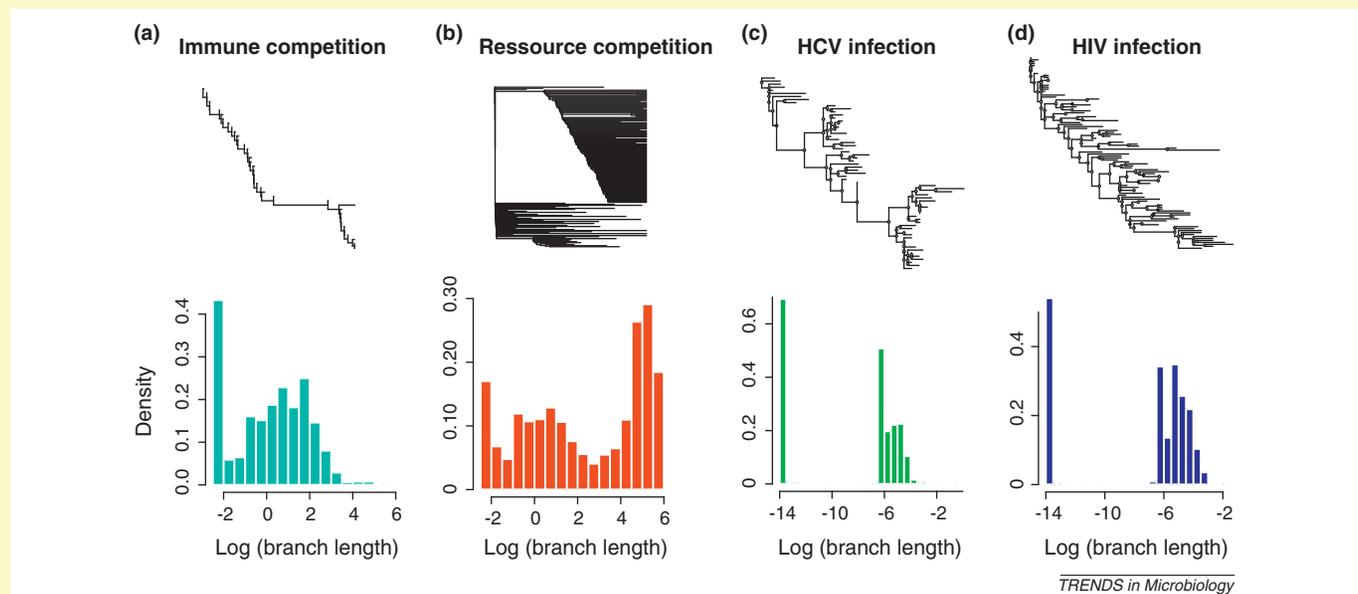


Figure 1. Virus phylogenies and average branch length distributions obtained for different within-host processes. (a) Numerical simulations with immune-mediated competition. (b) Numerical simulations with competition for target cells. (c) HCV infection; data from [30]. (d) HIV infection; data from [55]. Distributions of branch lengths were obtained from 20 phylogenies in A and B, from five phylogenies in C, and from three phylogenies in D. The phylogenies for panel C were obtained from patient 10 in [30] and the phylogeny for D from patient 6 in [55].

Perspectives

What data to collect?

Within-host evolution of many pathogens remains largely unknown. Existing data on WHED predominantly concern viral diseases causing persistent infections (Table 1) and, surprisingly, there seem to be very few (if any) studies on diseases such as human influenza or human malaria that address sequence evolution or trait evolution over the course of an infection. For WHED studies, two types of data can be collected: longitudinal data on the same host, where multiple samples are taken from the host at

different time points, and multi-compartment data, where several organs or tissue of the host are sampled. We argue that in order to link WHED to the evolutionary epidemiology of the disease, compartments linked to transmission need to be investigated, especially if these organs or tissues differ from those associated with pathogenesis.

In the last decade advanced high-throughput nanotechnologies such as next-generation deep sequencing (NGS) have opened new routes to experimental analyses of complex biological systems [4,5,48]. NGS allows us to access 'hidden' diversity and this can be used to determine

Table 1. Empirical data on WHED of human diseases^a

Trait studied	Pathogen	Ref.
DNA/RNA sequence evolution	HIV	[1,55]
	HCV	[30,56]
	<i>Pseudomonas aeruginosa</i>	[32]
	<i>Helicobacter pylori</i>	[3]
	<i>Burkholderia pseudomallei</i>	[41]
	Rhinovirus	[57]
Drug resistance	<i>Escherichia coli</i>	[50]
	<i>Mycobacterium tuberculosis</i>	[58,59]
	HIV	[60,29]
Immune escape	<i>P. aeruginosa</i>	[32]
	HIV	[61,62]
Replication rate	HCV	[63]
	HIV	[17,18]
Virulence factors	<i>P. aeruginosa</i>	[32]
	<i>E. coli</i>	[50]
	<i>H. pylori</i>	[3]

^aOwing to space limitations we were unable to perform an exhaustive review of all studies; instead we have elected to highlight illustrative studies or, where possible, reviews.

why some variants always remain rare and identify weaknesses of the pathogen. NGS thus provides us with a more exhaustive quantitative description of the genetic variations that can exist. Arguably, one of the most interesting perspectives opened by NGS resides in the combination of genetic data with phenotypic data on viral traits such as replication rate (or 'replicative fitness'). Indeed, the speed of replication is inherent to any model of within-host dynamics and/or within-host evolution. Analyses of NGS data can reveal detailed fitness landscapes of pathogens, and these could be the basis of applied evolution models.

However, NGS technologies raise intrinsic difficulties that current bioinformatics and statistics methods can only partly address. A major difficulty lies in the high error rates in the sequences obtained. Standard PCR techniques erroneously generates a mutation at a rate of 1 in 10^6 or 10^7 , whereas for NGS it is currently of 1 in 10^3 [5]. This makes the filtering and statistical analysis of the sequence output a severe burden.

Another challenge presented by NGS is the development of new summary statistics (see also below). Using classical sequencing techniques, only a small number of genotypes were identified and it was possible to identify positions that varied individually and interpret these variations in terms of escape mutations. With NGS, thousands of genotypes need to be analysed simultaneously. Bioinformatics tools could be developed to screen the datasets and identify mutations of importance. The next step would then be to infer evolutionary processes (e.g. drift versus natural selection) from summary statistics.

It would be naive to think that NGS will solve all the current problems. For instance, some parasites display adaptive phenotypic plasticity, as shown in the case of rodent malaria: parasites adjust the sex-ratio of the different sexual forms they produce in response to the presence of unrelated coinfecting strains [49]. Even if there is evolution, processes such as phase-shifting in *N. meningitidis* [40] could also generate specific genetic patterns. Overall, for rapidly evolving pathogens, disentangling evolutionary and plastic adaptive responses and distinguishing between

different types of evolutionary processes could prove challenging.

WHED can be (and have been) studied with current sequencing techniques and the lack of data combining population dynamics and genetics seems to be at least partly due to lack of interest. However, the improvement of sequencing techniques could generate momentum for studies on within-host evolution. A recent study considered the evolution of a specific *Escherichia coli* strain during the colonisation of the bladder of different patients [50]. This study not only shows that evolution takes place on short time-scales but also that this evolution is partly driven by host factors because host reinfection by the same *E. coli* strain generates similar evolutionary trajectories. Analysis of the full genome, of the transcriptome and of the proteome further revealed that these evolutionary changes are due to positive selection and affect processes involved in virulence.

Which models to develop?

Evolutionary theory has already been faced with the problem of increasing data and it is worth noting that a major insight, known as Fisher's fundamental theorem of natural selection, simply states that the rate of increase in fitness of any organism at any time is equal to the genetic variance in fitness in the population at that time [51]. In other words, summary statistics as simple as the genetic variance within a population is sufficient to make important evolutionary predictions. Recent models that combine this classical population genetics theory to population dynamics bring some new insights about the short-term evolution at the within- and between-host level, for instance, the dynamics in the few days after initiating a treatment [51,52]. These models are often referred to as 'Price equation models' because, as the population genetics Price equation, they describe the change in an average trait value from the covariance between a trait value and the fitness of the individual bearing that trait [23]. Such models could prove to be timely because they summarise large amounts of genetic information with variances and covariances, and this could provide a route for incorporating NGS data into a dynamic framework.

Models of molecular epidemiology tend to ignore the within-host level, and often entirely summarise it in a mutation rate term [7]. Comparison with molecular data shows that these models are appropriate for describing the epidemiology of influenza viruses. However, population genetics models indicate that opening the within-host black box could significantly improve our understanding of how pathogens adapt to their host [53]. Different underlying models of immune dynamics lead to different phylogenetic trees at the epidemiological level [54]. It would be interesting to see how incorporating WHED affects these phylogenies.

WHED complicate the epidemiology, largely because of the difficulty to predict which strain(s) will initiate the infection in a new host. In evolutionary ecology, nested models have been developed to include within-host dynamics into an evolutionary epidemiology framework but they tend to leave within-host evolution aside (see [11,12] for exceptions). It is too early to know if including WHED into

epidemiological models will strongly affect the results obtained by current models. WHED are likely to affect the speed of spread of pathogen variants but the open question is to assess if they also have qualitative effects.

Conclusions

The existence of WHED has been known for a long time. Datasets combining population dynamics and pathogen genetics are still lacking, even for major diseases such as malaria or influenza. NGS techniques have substantially increased ability to detect genetic diversity within hosts and could help to fill the gap between theory and data. However, these techniques are still inaccurate and will require theoretical work to solve statistical and bioinformatic problems. In parallel, new mathematical approaches based on nested modelling, phylodynamics and the Price equation could combine epidemiology with evolutionary processes taking place inside a host, and this would undoubtedly improve our understanding of many infectious diseases.

Acknowledgments

We thank P. Schneider and S. Reece for discussion and 3 anonymous reviewers whose comments greatly improved the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.tim.2010.09.005](https://doi.org/10.1016/j.tim.2010.09.005).

References

- Lemey, P. *et al.* (2006) HIV evolutionary dynamics within and among hosts. *AIDS Rev.* 8, 125–140
- Alfonso, V. *et al.* (2005) Intra-host evolutionary dynamics of hepatitis C virus E2 in treated patients. *J. Gen. Virol.* 86, 2781–2786
- Suerbaum, S. and Josenhans, C. (2007) *Helicobacter pylori* evolution and phenotypic diversification in a changing host. *Nat. Rev. Microbiol.* 5, 441–452
- Rokas, A. and Abbot, P. (2009) Harnessing genomics for evolutionary insights. *Trends Ecol. Evol.* 24, 192–200
- Poon, A.F.Y. *et al.* (2010) Phylogenetic analysis of population-based and deep sequencing data to identify coevolving sites in the *nef* gene of HIV-1. *Mol. Biol. Evol.* 27, 819–832
- Restif, O. (2009) Evolutionary epidemiology 20 years on: challenges and prospects. *Infect. Genet. Evol.* 9, 108–123
- Cobey, S. and Koelle, K. (2008) Capturing escape in infectious disease dynamics. *Trends Ecol. Evol.* 23, 572–577
- Grenfell, B.T. *et al.* (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303, 327–332
- Pybus, O.G. and Rambaut, A. (2009) Evolutionary analysis of the dynamics of viral infectious disease. *Nat. Rev. Genet.* 10, 540–550
- Mideo, N. *et al.* (2008) Linking within- and between-host dynamics in the evolutionary epidemiology of infectious diseases. *Trends Ecol. Evol.* 23, 511–517
- Luciani, F. and Alizon, S. (2009) The evolutionary dynamics of a rapidly mutating virus within and between hosts: the case of hepatitis C virus. *PLoS Comput. Biol.* 5, e1000565
- Lange, A. and Ferguson, N.M. (2009) Antigenic diversity, transmission mechanisms, and the evolution of pathogens. *PLoS Comput. Biol.* 5, e1000536
- Leslie, A.J. *et al.* (2004) HIV evolution: CTL escape mutation and reversion after transmission. *Nat. Med.* 10, 282–289
- Park, A.W. *et al.* (2009) Quantifying the impact of immune escape on transmission dynamics of influenza. *Science* 326, 726–728
- Alizon, S. and van Baalen, M. (2008) Multiple infections, immune dynamics and virulence evolution. *Am. Nat.* 172, E150–E158
- Coombs, D. *et al.* (2007) Evaluating the importance of within- and between-host selection pressures on the evolution of chronic pathogens. *Theor. Popul. Biol.* 72, 576–591
- Troyer, R.M. *et al.* (2005) Changes in human immunodeficiency virus type 1 fitness and genetic diversity during disease progression. *J. Virol.* 79, 9006–9018
- Gali, Y. *et al.* (2007) Survey of the temporal changes in HIV-1 replicative fitness in the Amsterdam Cohort. *Virology* 364, 140–146
- Derdeyn, C.A. *et al.* (2004) Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science* 303, 2019–2022
- Alizon, S. *et al.* (2010) Phylogenetic approach reveals that virus genotype largely determines HIV set-point viral load. *PLoS Pathog.* 6, e1001123
- Hollingsworth, T.D. *et al.* (2010) HIV-1 transmitting couples have similar viral load set-points in Rakai, Uganda. *PLoS Pathog.* 6, e1000876
- Levin, B.R. and Bull, J.J. (1994) Short-sighted evolution and the virulence of pathogenic microorganisms. *Trends Microbiol.* 2, 76–81
- Day, T. and Proulx, S.R. (2004) A general theory for the evolutionary dynamics of virulence. *Am. Nat.* 163, E40–E63
- Schmid-Hempel, P. (2008) Parasite immune evasion: a momentous molecular war. *Trends Ecol. Evol.* 23, 318–326
- Kouyos, R.D. *et al.* (2006) Stochastic or deterministic: what is the effective population size of HIV-1? *Trends Microbiol.* 14, 507–511
- Price, D.A. *et al.* (1997) Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. *Proc. Natl. Acad. Sci. U. S. A.* 94, 1890–1895
- Frost, S.D.W. *et al.* (2001) Genetic drift and within-host metapopulation dynamics of HIV-1 infection. *Proc. Natl. Acad. Sci. U. S. A.* 98, 6975–6980
- González-Scarano, F. and Martín-García, J. (2005) The neuropathogenesis of AIDS. *Nat. Rev. Immunol.* 5, 69–81
- Rambaut, A. *et al.* (2004) The causes and consequences of HIV evolution. *Nat. Rev. Genet.* 5, 52–61
- Farci, P. *et al.* (2000) The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science* 288, 339–344
- Thomas, D.L. *et al.* (2009) Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. *Nature* 461, 798–801
- Smith, E.E. *et al.* (2006) Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc. Natl. Acad. Sci. U. S. A.* 103, 8487–8492
- Margolis, E. and Levin, B.R. (2007) Within-host evolution for the invasiveness of commensal bacteria: an experimental study of bacteremias resulting from *Haemophilus influenzae* nasal carriage. *J. Infect. Dis.* 196, 1068–1075
- Stollenwerk, N. *et al.* (2004) Diversity in pathogenicity can cause outbreaks of meningococcal disease. *Proc. Natl. Acad. Sci. U. S. A.* 101, 10229–10234
- Lipsitch, M. *et al.* (1997) The population dynamics of antimicrobial chemotherapy. *Antimicrob. Agents Chemother.* 41, 363–373
- Ribeiro, R.M. and Bonhoeffer, S. (1999) A stochastic model for primary HIV infection: optimal timing of therapy. *AIDS* 13, 351–357
- Handel, A. *et al.* (2009) Exploring the role of the immune response in preventing antibiotic resistance. *J. Theor. Biol.* 256, 655–662
- Regoes, R.R. *et al.* (2004) Roles of target cells and virus-specific cellular immunity in primary simian immunodeficiency virus infection. *J. Virol.* 78, 4866–4875
- Kryazhimskiy, S. and Plotkin, J.B. (2008) The population genetics of dN/dS. *PLoS Genet.* 4, e1000304
- Richardson, A.R. *et al.* (2002) Mutator clones of *Neisseria meningitidis* in epidemic serogroup A disease. *Proc. Natl. Acad. Sci. U. S. A.* 99, 6103–6107
- Price, E.P. *et al.* (2010) Within-host evolution of *Burkholderia pseudomallei* in four cases of acute melioidosis. *PLoS Pathog.* 6, e1000725
- Alizon, S. and Boldin, B. (2010) Within-host viral evolution in a heterogeneous environment. Insights into the HIV co-receptor switch. *J. Evol. Biol.* 23, 2625–2635
- Ribeiro, R.M. *et al.* (2006) Naïve and memory cell turnover as drivers of CCR5-to-CXCR4 tropism switch in human immunodeficiency virus type 1: implications for therapy. *J. Virol.* 80, 802–809
- Holt, R.D. and Barfield, M. (2006) Within-host pathogen dynamics: some ecological and evolutionary consequences of transients, dispersal

- mode, and within-host spatial heterogeneity. In *Disease Evolution: Models, Concepts, and Data Analyses* (Feng Z. *et al.*, eds), (vol. 71, DIMACS Series in Discrete Mathematics and Theoretical Computer Science), pp. 45–66, American Mathematical Society, Providence, RI
- 45 Ravnigné, V. *et al.* (2009) Live where you thrive: joint evolution of habitat choice and local adaptation facilitates specialization and promotes diversity. *Am. Nat.* 174, E141–E169
- 46 Lion, S. and van Baalen, M. (2008) Self-structuring in spatial evolutionary ecology. *Ecol. Lett.* 11, 277–295
- 47 Buckling, A. and Brockhurst, M.A. (2008) Kin selection and the evolution of virulence. *Heredity* 100, 484–488
- 48 Wang, G.P. *et al.* (2010) Hepatitis C virus transmission bottlenecks analyzed by deep sequencing. *J. Virol.* 84, 6218–6228
- 49 Reece, S.E. *et al.* (2008) Sex ratio adjustment and kin discrimination in malaria parasites. *Nature* 453, 609–614
- 50 Zdziarski, J. *et al.* (2010) Host imprints on bacterial genomes – rapid, divergent evolution in individual patients. *PLoS Pathog.* 6, e1001078
- 51 Gandon, S. *et al.* (2009) Evolutionary epidemiology and the dynamics of adaptation. *Evolution* 63, 826–838
- 52 Alizon, S. (2009) The Price equation framework to study disease within-host evolution. *J. Evol. Biol.* 22, 1123–1132
- 53 Poon, A.F.Y. *et al.* (2007) Adaptation to human populations is revealed by within-host polymorphisms in HIV-1 and hepatitis c virus. *PLoS Pathog.* 3, e45
- 54 Ferguson, N.M. *et al.* (2003) Ecological and immunological determinants of influenza evolution. *Nature* 422, 428–433
- 55 Shankarappa, R. *et al.* (1999) Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J. Virol.* 73, 10489–10502
- 56 Stumpf, M.P.H. and Pybus, O.G. (2002) Genetic diversity and models of viral evolution for the hepatitis C virus. *FEMS Microbiol. Lett.* 214, 143–152
- 57 Cordey, S. *et al.* (2010) Rhinovirus genome evolution during experimental human infection. *PLoS One* 5, e10588
- 58 Kaplan, G. *et al.* (2003) *Mycobacterium tuberculosis* growth at the cavity surface: a microenvironment with failed immunity. *Infect. Immun.* 71, 7099–7108
- 59 Meacci, F. *et al.* (2005) Drug resistance evolution of a *Mycobacterium tuberculosis* strain from a noncompliant patient. *J. Clin. Microbiol.* 43, 3114–3120
- 60 Clavel, F. and Hance, A.J. (2004) HIV drug resistance. *N. Engl. J. Med.* 350, 1023–1035
- 61 Richman, D.D. *et al.* (2003) Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc. Natl. Acad. Sci. U. S. A.* 100, 4144–4149
- 62 Goulder, P.J.R. and Watkins, D.I. (2004) HIV and SIV CTL escape: implications for vaccine design. *Nat. Rev. Immunol.* 4, 630–640
- 63 Dustin, L.B. and Rice, C.M. (2007) Flying under the radar: the immunobiology of hepatitis C. *Annu. Rev. Immunol.* 25, 71–99
- 64 Anderson, R.M. and May, R.M. (1991) *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press
- 65 van Marle, G. *et al.* (2007) Compartmentalization of the gut viral reservoir in HIV-1 infected patients. *Retrovirology* 4, 87
- 66 Regoes, R.R. and Bonhoeffer, S. (2005) The HIV coreceptor switch: a population dynamical perspective. *Trends Microbiol.* 13, 269–277
- 67 Woolhouse, M.E.J. *et al.* (2002) Biological and biomedical implications of the coevolution of pathogens and their hosts. *Nat. Genet.* 32, 569–577
- 68 Perelson, A.S. and Oster, G.F. (1979) Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self-non-self discrimination. *J. Theor. Biol.* 81, 645–670
- 69 Smith, D.J. *et al.* (2004) Mapping the antigenic and genetic evolution of influenza virus. *Science* 305, 371–376
- 70 Haraguchi, Y. and Sasaki, A. (1997) Evolutionary pattern of intra-host pathogen antigenic drift: Effect of cross-reactivity in immune response. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 352, 11–20
- 71 Althaus, C.L. and De Boer, R.J. (2008) Dynamics of immune escape during HIV/SIV infection. *PLoS Comput. Biol.* 4, e1000103
- 72 Regoes, R.R. *et al.* (1998) Virus dynamics: the effect of target cell limitation and immune responses on virus evolution. *J. Theor. Biol.* 191, 451–462
- 73 Gilchrist, M.A. *et al.* (2004) Optimizing within-host viral fitness: infected cell lifespan and virion production rate. *J. Theor. Biol.* 229, 281–288
- 74 Ball, C.L. *et al.* (2007) Modeling within-host evolution of HIV: mutation, competition and strain replacement. *Bull. Math. Biol.* 69, 2361–2385
- 75 De Boer, R.J. and Perelson, A.S. (1998) Target cell limited and immune control models of HIV infection: a comparison. *J. Theor. Biol.* 190, 201–214