



**HAL**  
open science

# Dynamics and evolution of infectious diseases

Samuel Alizon

► **To cite this version:**

Samuel Alizon. Dynamics and evolution of infectious diseases. Populations and Evolution [q-bio.PE]. Université Montpellier 2, 2013. tel-01638065

**HAL Id: tel-01638065**

**<https://hal.science/tel-01638065v1>**

Submitted on 26 Nov 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Montpellier 2  
École doctorale SIBAGHE

## **Dynamique et évolution des maladies infectieuses**

## **Dynamics and evolution of infectious diseases**

Samuel ALIZON

Dossier préparé en vue de l'obtention de l'habilitation à diriger des recherches et présenté le 8 novembre 2013.

### **Jury :**

M. Janis A ANTONOVICS, Professeur, Université de Virginie (rapporteur)  
M. Sam P BROWN, Professeur, Université d'Édimbourg (rapporteur)  
M. Jean-Baptiste FERDY, Professeur, Université Paul Sabatier (président du jury)  
M. Yannis MICHALAKIS, Directeur de Recherche, CNRS Montpellier  
Mme Dominique PONTIER, Professeure, Université Claude Bernard (rapporteuse)  
M. François RENAUD, Directeur de Recherche, CNRS Montpellier

# Table des matières

<b>1</b>	<b>Activités d'encadrement</b>	<b>3</b>
<b>2</b>	<b>Research achievements and perspectives</b>	<b>9</b>
2.1	Nested models in evolutionary epidemiology	9
2.2	Why do parasites harm their hosts?	11
2.2.1	The trade-off hypothesis	11
2.2.2	Levels of selection	13
2.2.3	Immunopathology	15
2.2.4	One trade-off to rule them all	15
2.3	Evolutionary epidemiology of multiple infections	24
2.3.1	Multiple infections and virulence evolution: a review	24
2.3.2	Importance of epidemiological feedbacks	25
2.3.3	Co-transmission and the evolution of virulence	28
2.4	Disease spread on weighted networks	31
2.4.1	Weighting for sex	32
2.4.2	Analytical framework to predict spread on weighted networks	34
2.5	Within-host ecology and evolution	37
2.5.1	The problem: what gets in differs from what gets out	37
2.5.2	HIV co-receptor switch	39
2.5.3	Modelling the course of a HIV infection	41
2.5.4	Levels of selection: a signature in HIV evolutionary rates	43
2.6	Phylogenies in evolutionary epidemiology	46
2.6.1	Linking phylogenies and infections traits	46
2.6.2	The case of HCV	49
2.7	Epidemiology and evolution of disease emergence	51
2.7.1	Introducing the outbreak threshold	52
2.7.2	How does host availability affect pathogen emergence?	53
2.8	Perspectives	59
2.8.1	Tolerance, Resistance and Virulence	59
2.8.2	Evolutionary medicine instead of Darwinian medicine	61
2.8.3	More on multiple infections and disease evolution	63
2.8.4	Phylogenies and pathogen evolutionary dynamics	66
2.8.5	Within-host models	71
2.8.6	More weighting for sex	73
2.8.7	Is the HPV vaccine 'evolution-proof'?	73

## 2 Research achievements and perspectives

During my PhD supervised by Minus van Baalen, I developed nested models to study virulence evolution and better understand why parasites harm their hosts. This report presents some of the topics I have been working on since then.

I asked several colleagues what a habilitation dissertation should look like because the official guideline is that it should ‘synthesise my research work’. From the answers I gathered, I think the dissertation can take pretty much any form. I would have loved to write it as a mystery novel or as a poem but I did not have the time so I decided to present some general questions I have been working on. Each section is introduced by a short summary. I tried to use a polemic tone whenever it was possible to stimulate the discussion. If you already know about my work, you can jump to the Perspectives section (page 59), where I discuss some future projects.

### 2.1 Nested models in evolutionary epidemiology

---

I was trained as a biologist but, as often in France, mathematics always had a dominant role in our education. It is only during my third year of undergrad that I found out that not only do mathematics have applications in biology but also that in evolutionary biology this has been going on for over a century. Working with Vincent Jansen, I discovered modelling in evolutionary ecology and decided to try and continue work as a theoretical biologist as long as I could get funding.

During my DEA<sup>1</sup> and my PhD with Minus van Baalen, I worked on the evolution of host-parasite interactions and modelling within-host dynamics. In fact, the main idea was to incorporate within-host dynamics models into epidemiological models. This was partly motivated by some papers from the 90s that added an epidemiological perspective to within-host dynamics (Sasaki and Iwasa, 1991, Antia et al., 1994). Many of such ‘nested models’ were published that linked within-host dynamics and evolutionary epidemiology models were published when I began my PhD (Gilchrist and Sasaki, 2002, André et al., 2003, Ganusov and Antia, 2003) and several others have been published since then (see e.g. Pepin et al., 2010, Handel et al., 2013). My PhD used modelling approaches to study the conditions that lead to a trade-off between virulence and transmission (Alizon, 2006).

Just to finish on nested models, during my first post-doc, Troy Day was asked to write a review on the topic (Mideo et al., 2008). To make a long story short, the most important question we raised in this review has to do with the necessity of building these models.

---

1. The former name of the second year of the MSc in France.

In most cases, the within-host analysis can be done on its own. This way, you know exactly how the parameters that are relevant at the epidemiological level (e.g. transmission rate, virulence, recovery rate) vary as a function of the age of the infection and you can then use classical evolutionary epidemiology frameworks (such as the one by [Day, 2001](#)). In some cases the nesting is essential though. Typically, this occurs when the epidemiological state of the host population feed backs into the within-host dynamics. For example, if multiple infections are allowed, running the within-host analysis separately is not straightforward because one has to take into account the fact that a co-infection can occur at any time during the infection. Another situation we mention is when within-host dynamics depend on the infectious dose, which itself depends on the epidemiological state of the population. Finally, one important case, which is described in details in the following is within-host evolution because then, not only to the epidemiological rates vary through time, but also what gets in the host differs from what gets out.

Even though nesting might not be essential strictly speaking, it can often be useful. To take an example, it is possible to model the effect of an increased death rate of immune cells but it is much easier to first build a within-host model in order to see how this rate affect parasite density dynamics and then incorporate it into an epidemiological model. Therefore, nested models can offer a simple way to increase the biological relevance of a model. One should be careful though because as soon as more than two levels are studied (e.g. combining the cellular, the within-host and the epidemiological level), the meaning of a parameter variation at the first level have little relevance at the third level because of the amount of other processes at play.

## 2.2 Why do parasites harm their hosts?

*In many cases, harming the host may seem detrimental to the parasite: strains that keep their host alive longer have a longer transmission period and therefore achieve a higher fitness. Several hypotheses have been put forward to explain the maintenance of virulence. I review these here and propose a general framework rooted in epidemiological dynamics to capture them all. One of the advantages of this framework is that it could facilitate the linking between theory and experiments, which is always complicated given that virulence is measured in a great variety of ways.*

For many parasites<sup>2</sup>, host death appears to be a costly event because it terminates the infectious period (there are of course exceptions, for instance host death is required for the transmission of parasitoids). Therefore, virulence, which unless specified otherwise is the host mortality due to the infection, has been puzzling researchers for more than a century (Smith, 1904).

The first solution to this apparent paradox was that virulence is only transitory and that, given enough time, parasites will evolve to become benign to their host, whence its name: ‘avirulence theory’ (Méthot, 2012). Some like Ball (1943) did challenge this theory early on. Indeed, as pointed out later on, there is evidence that many parasites have conserved their virulence over centuries. One of the well known cases is tuberculosis, which we know was already deadly in the time of the pharaohs as statues were found that exhibit typical symptoms of vertebral tuberculosis. The analysis of mummies (among which Toutankamon’s) revealed that malaria was already present too in ancient Egypt (Cockburn et al., 1998, Donoghue et al., 2004). Herre (1995) even lean on fossil records to show that nematodes have been infecting fig wasps for a long time. More generally, given that most parasites have low generation times, one expects their virulence to evolve rapidly (on a human time scale). In other words, if parasites are virulent, it is likely no to be just by lack of time to adapt to their host.

Until the 1980s, there was no conceptual framework to challenge the avirulence hypothesis (or, as May and Anderson (1983) called it to discredit it, the ‘common wisdom’). The explanation for the maintenance of parasite virulence has laid the field for intense debates over the years (Lipsitch and Moxon, 1997, Zimmer, 2003). Part of the reason for this is that many people tend to forget about the underlying assumptions made in different models. I first discuss the ideas that have been put forward to explain the evolution and maintenance of parasite virulence (see also Ewald, 1994, Read, 1994, Bull, 1994, Frank, 1996, Ebert and Herre, 1996, Levin, 1996, Schmid-Hempel, 2011). Then, I try to introduce a general framework to explain why parasites harm their host.

### 2.2.1 The trade-off hypothesis

Virulence can be selected if it increases the epidemiological fitness, which is the ability of the pathogen to spread in the host population. This can be captured by linking virulence to other epidemiological parameters.

2. The term ‘parasite’ is used here in its wider acceptation, which includes both micro-parasites (viruses, bacteria, procaryotes) and macro-parasites (such as worms).

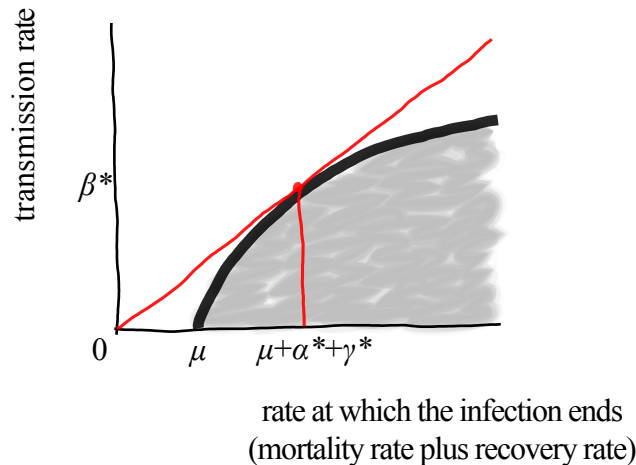


Figure 2.1: **Illustration of the trade-off between the transmission rate and the duration of the infection.** The evolutionary stable level of virulence ( $\alpha^*$ , which maximises parasite epidemiological fitness) is given by the tangent of the curve that goes through the origin. Strains cannot be above the trade-off curve but they can be below it (the gray area) if they are maladapted (with the same level of virulence, they could achieve a higher transmission rate). Figure from [Alizon et al. \(2009\)](#).

This idea was introduced originally by [Anderson and May \(1982\)](#) and [Ewald \(1983\)](#). In their article, [Anderson and May](#) propose a trade-off between the virulence and the recovery rate. By analyzing data on infections of rabbits by myxoma virus, they suggest that more virulent strains also cause longer infections. In other words, the cost of virulence (killing the host) is traded-off against the cost of being cleared by the host's immune response.

Nowadays, the trade-off hypothesis usually refers to the idea that the cost of decreasing the duration of the infection (i.e. being virulent) is compensated by an increase in the transmission rate. In other words, in order to reach high enough number within the host (or to produce specialised transmission forms), the parasite needs to exploit its host. Several experimental approaches have shown such a relationship (reviewed in [Alizon et al., 2009](#)). Amongst these, three stand out because the shape of the relationship they find between virulence and transmission supports the prediction that viruses should evolve towards intermediate levels of virulence. Indeed, it is not only necessary to show a positive correlation between transmission and virulence to conclude that there is an evolutionary stable level of virulence. For instance, if the relationship is linear, infinite levels of virulence are selected for. In order to observe an intermediate level of virulence, the trade-off curve must saturate, which biologically means that at some points the benefits conferred by increased virulence need to decrease ([van Baalen and Sabelis, 1995](#)). This of course assumes that the parasite spreads in the population according to an idealised epidemiological setting with no vertical transmission and only direct transmission (a limitation that will be discussed later). These three examples are myxoma virus in rabbits ([Dwyer et al., 1990](#), [Bolker et al., 2010](#)), which was already mentioned above but for the recovery-virulence trade-off, HIV in

humans (Fraser et al., 2007) and a protozoan parasite of the Monarch butterfly (de Roode et al., 2008).

I mention these examples explicitly because the trade-off hypothesis seems to be one of the scientific hypothesis, which is the easiest to attack and every 10 years or so an article comes out that explains this hypothesis is unsupported by data or unrealistic. As Mike Boots puts it, the (crying) lack of data probably comes from the fact that people did not look for it. However, as we will see below, part of this problem probably comes from the fact that there are not so many systems that allow us to easily test this hypothesis. In particular, one needs to use a correct measure of virulence (Day, 2002a) and closing the life cycle of the parasite can be challenging.

We have mentioned trade-offs between virulence and transmission (deadlier strains transmitted better), between virulence and recovery (it is more difficult to recover from deadlier strains) but a trade-off between recovery and transmission can also have implications for virulence evolution if all these three components are linked to parasite density (Alizon, 2008). Of course in the latter case, the trade-off would then involve three components but the saturating effect, which is what causes the existence of an optimal parasite strategy, is only due to transmission and recovery. Such a trade-off relationship can emerge from within-host models of acute infections that are inspired from predator-prey models. As I showed, the trade-off relationship saturates if the activation of the immune response depends not only on parasite density but also on parasite replication rate. The latter is consistent with observations made for many viral infections, where it is the virus replication inside the cell that triggers the immune response (see Yewdell, 2007, for a review). In fact, ‘dormant’ viruses typically evade the immune response by not replicating. Experimental support for this is scarce (but see Bocharov et al., 2004) because one needs to be able to measure both virus density and virus replication rate to compare explanatory models. One of the uses of this trade-off is that it can apply to viruses that do not kill their host, e.g. many rhinoviruses in humans that only cause colds.

### 2.2.2 Levels of selection

Part of the reason why infectious diseases are such a fascinating object for an evolutionary biologist is because they often illustrate the idea of levels of selection (which refers to the idea of units of selection Lewontin, 1970). At the epidemiological level, there is a competition between the different strains circulating in the population: the most competitive strain is the one that infects the highest number of hosts. However, the occurrence of more than one ‘strain’ (or ‘genotype’) in the same host generates within-host competition: at the within-host level, the strain that reaches higher numbers than the other strain(s) is the most competitive. Two situations can lead to such a within-host diversity.

The first situation concerns rapidly evolving parasites that mutate rapidly, which allows the parasite to evolve over the course of an infection. As originally proposed in the 90s (Bonhoeffer and Nowak, 1994, Bull, 1994, Levin and Bull, 1994), within-host evolution implies that the parasite can adapt to better exploit the host, thus maximizing its competitive ability at the within-host level, but also potentially sacrificing its competitiveness at the between-host level. This is yet another illustration of the idea popularised in evolutionary biology since the 60s that natural selection does not necessarily optimise fitness. One of the examples put forward by Levin and Bull (1994) is that of meningitis. In most infections, the bacterium *Neisseria meningitidis* infects



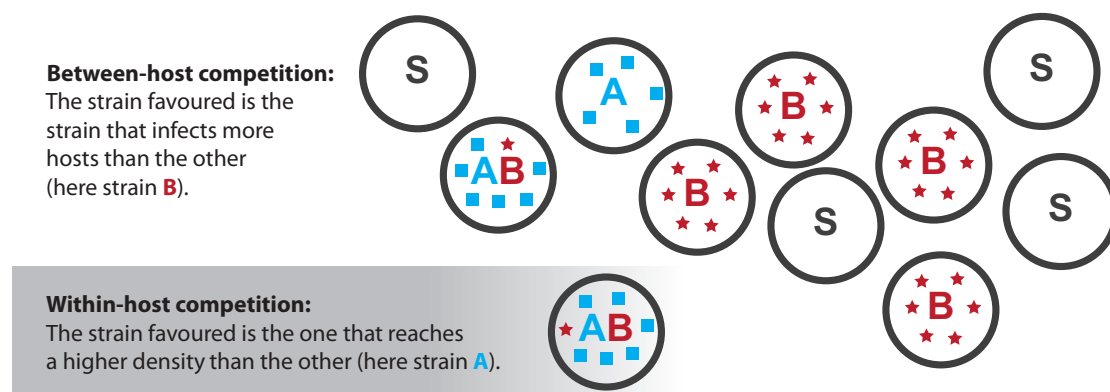


Figure 2.2: **Difference between competitive ability at the within-host and at the between-host level.** The blue strain (A) is more competitive than the red strain (B) at the within-host level. However, the red strain is more efficient at exploiting hosts in single infections and therefore spreads better. Note that the density of susceptible hosts is an important component in shaping the fitness at the epidemiological level. Figure from [Alizon et al. \(2013\)](#).

a host without causing any virulence. In some infections however, the bacterium colonises the brain of the host, thus leading to a lethal infection. This event is very likely to be selected against because there is no transmission once the brain is infected and, as shown experimentally by [Margolis and Levin \(2007\)](#), mutations that lead to virulence are likely to be ‘re-discovered’ every time (and of course lost soon after being rediscovered). Coming back to the problem of virulence evolution, if more virulent strains have a competitive advantage at the within-host level, then within-host evolution could explain why virulence persists at the epidemiological level ([Alizon et al., 2011](#)).

The second situation that requires to invoke levels of selection to understand virulence evolution is multiple infection, which occurs when a host is simultaneously infected by more than one parasite strain. In many ways, this situation is similar to within-host evolution: there is competition between strains at the within-host level and virulence can be favoured if it confers a competitive advantage at this level. One reason for singling it out is that multiple infections cannot be separated from an epidemiological framework. Indeed, the within-host diversity originates from successive infections of the host and not from within-host mutations. Another reason is that (as discussed below), multiple infection can involve unrelated parasites or even parasites from different species, which calls for specific frameworks. For further details about multiple infections and virulence evolution are given in the next section (and in our recent review [Alizon et al., 2013](#)).

As an aside, note that multiple infections or within-host evolution do not require a trade-off at the epidemiological level to explain the maintenance of virulent parasites. Indeed, even in a situation where virulence would always be costly at the between-host level, the fact that virulence provides a competitive advantage at the within-host level is sufficient to maintain virulent

strains in the population. However, the two processes can of course be acting at the same time and one should not see within-host competitiveness as an exclusive alternative to the trade-off hypothesis (Alizon and Michalakis, 2011).

Finally, if we wish to stick to a general definition of levels of selection, it is worth emphasising that one can consider levels above the host levels. For instance, the host population can be structured into sub-population in a meta-population fashion. More generally, any ‘viscosity’ (as introduced by Hamilton, 1964) in the host population will generate a level of selection beyond the host level. Typically, these higher levels tend to select against virulence (Boots and Sasaki, 1999, van Baalen, 2002). The idea is that more virulent strains kill their host too rapidly, which means they tend to deplete their local susceptible population pool before spreading to other parts of the population.

### 2.2.3 Immunopathology

For a long time, the immune response has been seen as the host’s perfect weapon. However, experiments showed that not only does the maintenance of this immune response has a cost but also that its activation can decrease host survival (Moret and Schmid-Hempel, 2000, Schmid-Hempel, 2011). In fact, experiments on mice infected by *Plasmodium chabaudi* showed that tuning down some of the components of the host’s immune response strongly decreased virulence (Long et al., 2008). In other words, part of the disease-induced mortality should actually be referred to immune activation-induced mortality and this part could even be the majority for some infections (Graham et al., 2005).

Some models of virulence evolution have studied the effect of including the immunopathological contribution to virulence (Gilchrist and Sasaki, 2002, Alizon and van Baalen, 2005, Day et al., 2007). The main result is that increased immunopathology favours more virulent strains if the intensity of immunopathology is independent of parasite replication but that different outcomes can occur is immunopathology depends on parasite host exploitation intensity. In this case, if parasites do not become avirulent, it is because it is the immune response that harms the host.

With Eileen Butterfield and Yannis Michalakis, we also explored these ideas by assuming that tolerance and resistance were a property of the parasite instead of being that of the host as most models assume. What evolved in Eileen’s model was the ability of the parasite to be tolerated or to be resisted to. Coming back to the previous paragraph, one of the goals was to find conditions under which parasites evolve to be tolerated, which means less immunopathology (and less virulence).

### 2.2.4 One trade-off to rule them all

From the layman’s eye, there can appear to be a multitude of theories explaining why parasites harm their host. We already mentioned trade-offs involving transmission, virulence and recovery, levels of selection (within-host evolution and multiple infections) and immunopathology. One could also add coincidental selection to the list (Brown et al., 2012). A way to capture all these explanations with a unique framework is to focus on their common denominator, which is

that virulence is somehow adaptive. (If virulence is not adaptive, natural selection will select for less virulent parasites until there is no variation left in the parasite population to select on.)

Thinking about the ‘adaptiveness’ of virulence implicitly calls for an ecological vision because this is the only way to define parasite epidemiological fitness (Metz et al., 1996). One of the advantages of working on parasites is that, at the epidemiological level, fitness is intuitive: it is the ability of the parasite to infect new hosts over the course of an infection. A more accurate expression than epidemiological fitness might be ‘fitness at the highest level of selection’. Usually, this highest level corresponds to the between-host level (unless one only focuses on within-host evolution) but if the population is structured the highest level can be a meta-population level. From hereon, we will refer to this ‘parasite fitness at the highest level of selection’ as ‘parasite fitness’.

Parasite epidemiology can be captured by a system of differential equations, which requires to know the parasite’s life-cycle. Then, one can calculate parasite fitness by deriving this system to obtain what is known as the Jacobian matrix (Diekmann and Heesterbeek, 2000). The dominant eigenvalue of this matrix indicates if the parasite can spread or not in the host population. One can make a parallel with the calculation of the population growth rate in ecology when the population is structured into different classes (Caswell, 1989). Another way to see this is that each of the life-stages of the parasite can contribute differentially to the next generation of the parasite. In fact, the partitioning of the Jacobian matrix into a birth and a death matrix often yields a simpler invasion condition according to the ‘next generation theorem’ (Diekmann and Heesterbeek, 2000, van den Driessche and Watmough, 2002, Hurford et al., 2010). Note that if the parasite population is structured, fitness can be expressed fairly easily in some situations (for instance if there are distinct host classes Diekmann and Heesterbeek, 2000) but nearly impossibly in some others, where approximations are needed (in a lattice structure for instance Lion and van Baalen, 2008).

Being able to express parasite epidemiological fitness as a function of the components of the life-cycle informs us on which quantities to measure. For instance, in the case of horizontally and directly transmitted parasites (in a well-mixed host population), this fitness ( $R$ ) can simply be written as a product of the transmission rate of the parasite and the duration of the infectious period or, more mathematically, as:

$$R = \frac{\beta}{\mu + \alpha + \gamma} \tilde{S} \quad (2.1)$$

where  $\beta$  is the transmission rate,  $\mu$  is the host baseline mortality rate,  $\alpha$  is the disease-induced mortality (or virulence),  $\gamma$  is the recovery rate and  $\tilde{S}$  is the density of susceptible hosts. Equation 2.3 will clearly remind some readers of the baseline reproduction rate ( $R_0$ ), which is the number of secondary infections caused by an infected host in a fully susceptible population (Anderson and May, 1979). In fact, as further discussed at the end of this section, this fitness at the epidemiological level is often identical to  $R_0$ <sup>3</sup>.

By plotting parasite fitness as a function of virulence, we directly see how the harm parasites cause to the host should evolve. In equation 2.3, if none of the terms on the right hand side

---

3. The difference between  $R$  and  $R_0$  comes from the state of the resident population as in the latter case the population is entirely susceptible (there is no resident strain).

depend on  $\alpha$ , the lower  $\alpha$ , the higher the fitness. This is just restating the avirulence theory. In order for non-zero virulence levels to be selected, there needs to be a correlation between  $\alpha$  and other components of the parasite's fitness (here the transmission rate or the recovery rate). This time, we just restated the trade-off hypothesis.

The advantage of writing the fitness function explicitly is that it *de facto* solves the problem of parasites that do not have a direct and horizontal transmission. Let us for example consider the case of spore-transmitted diseases. As shown by Day (2002b), parasite fitness in this case is:

$$R = \frac{\beta}{\delta} \frac{p + (\mu + \alpha) b}{\mu + \alpha + \gamma} \tilde{S} \quad (2.2)$$

where  $\beta$  is the transmission rate of spores to susceptible hosts,  $p$  is the spore production rate of infected hosts,  $b$  is the 'burst size' (the number of spores released by an infected host upon death),  $\delta$  the survival rate of a free-living spore and, as above,  $\mu$  is the host baseline mortality rate,  $\alpha$  is the disease-induced mortality and  $\gamma$  is the host recovery rate. Our goal is still to see if virulence can be adaptive. However, the number of potential trade-offs that can explain such an adaptiveness has exploded as  $\alpha$  can now be traded off against  $\beta$  (if the spores that transmit better are more virulent),  $\delta$  (if the spores that survive better in the environment are more virulent<sup>4</sup>),  $p$ ,  $b$ ,  $\gamma$ ... The problem with all these correlations is that they can go in different directions. For instance, more virulent strains could have at the same time a lower transmission rate ( $\beta$ ) and a better survival ( $\delta$ ). By only investigating one of the trade-offs, and hence considering only part of the parasite life cycle, there is a non negligible risk to make erroneous predictions. The only option therefore seems to directly consider fitness as a function of virulence.

For the sake of completeness, let us apply the same approach to vector-borne parasites. The fitness of these can be written as (Anderson and May, 1991, Mideo and Day, 2008, Froissart et al., 2010):

$$R = \rho b^2 \frac{e^{-\mu_v T} \beta_{v \rightarrow h}}{\mu_v + \alpha_v + \gamma_v} \frac{\beta_{h \rightarrow v}}{\mu_h + \alpha_h + \gamma_h} \quad (2.3)$$

where  $\rho$  is the number of vectors per host,  $b$  is the vector biting rate per day,  $T$  is the incubation time of the parasite in the vector (i.e. the time before the vector becomes infectious),  $\mu_h$  and  $\mu_v$  are the baseline mortality of the host and of the vector,  $\alpha_h$  and  $\alpha_v$  is the virulence (or disease induced mortality) of the parasite to the host and to the vector,  $\gamma_h$  and  $\gamma_v$  are the recovery rates of the the host and of the vector,  $\beta_{h \rightarrow v}$  is the transmission rate from the host to the vector and  $\beta_{v \rightarrow h}$  is the transmission rate from the vector to the host. The virulence in the host ( $\alpha_h$ ), which is usually the variable of interest, can now be traded-off against at least 8 other infection traits!

### Introducing the generalised trade-off approach

Our thesis is that looking for trade-offs between two components of the parasite life-cycle is a lost cause (except maybe for some special cases such as HIV where there really seems to be only

4. This is known as the 'curse of the Pharaoh' hypothesis (Bonhoeffer et al., 1996, Gandon, 1998) following the myth that several archeologists who entered the tomb of pharaoh Toutânkhamon died short after. It is interesting to note that this myth was popularised in the XIX<sup>th</sup> century, that is, several decades before the tomb was actually discovered in 1922.

two relevant components). We argue that a more appropriate approach is to compute parasite fitness as a function of a trait of interest (usually virulence). One can then see whether some trait values are optimal, i.e. maximise parasite fitness. Mathematically, if the trait (usually virulence) is denoted  $x$  and fitness is denoted  $R$ , we are looking for values  $x^*$  that maximise the function  $R(x)$ .

This approach has already been applied to real data. In the case of HIV for instance, [Fraser et al. \(2007\)](#) define the pathogen fitness (they call it ‘transmission potential’), which depends on the viral load ( $V$ ) in the host, as follows:

$$R(V) = D(V) \times \beta(V) \quad (2.4)$$

where  $D(V)$  is the duration of an infection with set point viral load  $V$  and  $\beta(V)$  is its transmission rate. This is of course the classical expression of the  $R_0$  for diseases transmitted per horizontal contact. Note that the authors perform additional fitting to infer the shape of the function  $D$  and  $\beta$  from the data. The authors find that the duration of infection is a decreasing function of  $V$  but that transmission rate is an increasing function of  $V$  and that there exists an intermediate set-point virus load ( $V^*$ ) that maximises fitness ( $R$ ). More strikingly, the observed distribution of set-point viral load values in two populations match the predicted fitness curve (with more hosts exhibiting the set-point viral loads that have the highest fitness).

We saw that there are recipes to derive  $R$  from an epidemiological model. Unfortunately, measuring it experimentally is a different story. Probably the simplest way to proceed is to group components of the fitness into terms that are easier to measure. This is exactly how [de Roode et al. \(2008\)](#) proceeded to show that there is an optimal virulence for a pathogen transmitted through a spore stage. The trait they follow is not the virulence directly but rather the pathogen spore load ( $p$ ). They express the pathogen fitness as:

$$R(p) = E(p) \times M(p) \times T(p) \times I(p) \quad (2.5)$$

where  $E(p)$  is the probability that an adult will emerge from an infected larvae (the inverse of the virulence),  $M(p)$  is the probability of mating,  $T(p)$  is the proportion of eggs that acquire parasite spores and  $I(p)$  is the proportion of offspring that become infected. This pathogen fitness function covers the whole life cycle from the pathogen, as it follows all the steps to go from one infected adult host to another infected adult host. As in the case of HIV, there exists an intermediate value of spore load ( $p^*$ ) that maximises fitness.

More recently, [Chapuis et al. \(2012\)](#) showed that intermediate virulence maximises pathogen fitness in the case of a bacterium that can alternate between two hosts (an insect and a nematode vector). In this model, the fitness is expressed as a function of the number of bacteria per infected host ( $N$ ):

$$R(N) = \frac{\beta(N)}{v(N)} f(N) p(N) \quad (2.6)$$

where  $\beta$  is the capacity of nematodes to infect insects,  $f$  is the number of nematodes released by an infected insect (the ‘burst size’),  $p$  is the fraction of nematodes that are infected and  $v$  is the death rate of the nematodes during dispersal (i.e. during their free-living stage).

In these last two examples, fitness covers the whole life-cycle of the parasite and it is expressed through terms that are easier to measure experimentally. Unfortunately, there is no way to determine *a priori* how the components of the fitness expression should be grouped into terms.

There are not many more examples to cite here. The reason for this is that there are few cases where the authors ‘close the pathogen life-cycle’ and many trade-offs could be contradicted by the rest of the cycle. For instance, [Jensen et al. \(2006\)](#) found that a castrating bacterial parasite of the crustacean *Daphnia magna* maximises the spore load produced from an infected host for intermediate levels of virulence (here castration). This is indeed a very strong hint for the existence of a trade-off but the rest of the pathogen life-cycle might go against this. For instance, if the spores that are produced at an intermediate level of virulence have a poorer quality. Formally, this fitness function can be written by highlighting the castration level induced by the parasite ( $c$ ), as

$$R(c) = \frac{\beta(c)}{\delta(c)} B(c) \quad (2.7)$$

The authors have indeed shown that intermediate values of  $c$  maximise spore load burst size ( $B$ ) but this could be irrelevant if these same values minimise the infectivity of the spores ( $\beta$ ) or maximise their removal rate from the environment ( $\delta$ ). A similar concern is valid for [Doumayrou et al. \(2013\)](#): even though they provide the first experimental support for a transmission-virulence trade-off for plant viruses, their analysis only covers the host part of the life-cycle and it might be that a virus that has an optimal virulence at the host level would be extremely deleterious at the vector level.

### Solving the virulence definition problem

One may then ask: ‘Why study fitness as a function of virulence? Why not use another trait such as host recovery rate?’ This is of course a legitimate concern but in the end the question that usually matters is why do parasites harm their host. In fact, in the three examples cited above ([Fraser et al., 2007](#), [de Roode et al., 2008](#), [Chapuis et al., 2012](#)), the variable through which the pathogen fitness is expressed is positively correlated with virulence expressed through host mortality.

Embracing the whole parasite life-cycle has the advantage that it sort of solves the recurrent problem of choosing an appropriate measure of virulence. As shown in a (quite often mis-cited) article by [Day \(2002a\)](#), measuring virulence as case mortality, death rate or lethal dose has a huge influence on model predictions. Potentially, a lot of the virulence evolution trends measured experimentally should be compared to different models using the same measure. If one measures fitness as a function of any virulence measure, this sort of solves the problem. The fact remains though that all the components (or terms) of the life-cycle still need to be measured in the same way as they are expressed in the fitness expression.

### Applicability of the general approach

Writing the parasite fitness function explicitly has the great advantage that it forces us to bear in mind that virulence can be adaptive in each part of the parasite life-cycle. Consider for in-

stance the case of opportunistic parasites, as recently reviewed by [Brown et al. \(2012\)](#)<sup>5</sup>. These parasites are not able to infect a healthy hosts but they can become deadly if they encounter immunocompromised hosts. In this case, the selection and maintenance of this virulence (for instance virulence factors in bacteria) can appear to be outside the frame of the trade-off hypothesis. Yet, when the whole life-cycle of the parasite is included, it becomes possible to see how virulence in the host of interest can be traded-off against components of the life-cycle that are outside the host (such as growth in another host or survival in the environment).

Of course, it is also possible that virulence is not under any selective pressure. This is the case for instance if the host of interest is a ‘dead-end’ host, which means there is no further transmission from this host. In this case, virulence does not appear in the equation of the parasite epidemiological fitness. An example of such situation for humans is toxoplasmosis, which can infect humans but will never be transmitted out of them.

Overall, the applicability of the general trade-off approach could be restrained if some stages of the parasite life-cycle are difficult to re-create *in vitro* but it does solve the problem that virulence can potentially be traded-off against a legion of traits. Finding that parasite fitness depends on virulence and that it is not maximised for zero virulences is sufficient to show that somehow a trade-off must be acting. The next step is to identify the other trait(s) involved in this trade-off. In fact it could very well be that in some cases only one component of the parasite life-cycle is involved in the trade-off (as in the case of HIV for instance), which would then make parallels with the current theory easier.

One possible limitation of the general approach could reside in incorporating temporal trait variations in the fitness function. In a key article, [Day \(2001\)](#) showed that the fitness of a HDT parasite can be written as a function of the age of the infection ( $a$ ) such that

$$W = \int_0^{\infty} \beta(a) e^{-\int_0^a (\mu + \alpha(s) + \gamma(s)) ds} da \quad (2.8)$$

where the terms inside the main integral are the transmission rate of an individual infected for  $a$  units of time ( $\beta(a)$ ) times the probability that the infection did not end before due to host baseline mortality ( $\mu$ ), or disease-induced mortality ( $\alpha(a)$ ), or recovery ( $\gamma(a)$ ). This explicit tracking of variations in infection life-history traits over the course of an infection can be extremely important (see for the instance the case of HIV where transmission rate more or less follows the variations in viral load [Hollingsworth et al., 2008](#)).

Accounting for variations in rate depending on infection age is easy if there is only one host compartment (equation 2.8 more or less applies). Problems begin when there are several host types of several life stages. One possibility would be to separate scales and develop expressions similar to equation 2.8 for each type of host the parasite infects. Unfortunately, by doing so, we potentially loose temporal effects.

Finally, to be exhaustive, the only situation that does not yet seem to fit perfectly in this comprehensive framework is within-host evolution. To quote Minus van Baalen, the problem with within-host evolution is that what gets in differs from what gets out (which is a major difference with multiple infections for instance). It is not clear to me how this could be captured

---

5. But the idea was introduced and discussed by Bruce Levin in several articles ([Levin and Svanborg Edén, 1990](#), [Levin and Bull, 1994](#), [Levin, 1996](#))

mathematically. Unless the within-host and the between-host scales can be completely separated (which seems oversimplifying), it would require some sort of invasion analysis running at the same time at the within and at the between-host level. Also, a question we have to ask before undertaking such a difficult task is whether it is worth it. Consider the case of HIV for instance. A trade-off has been shown to exist (Fraser et al., 2007). Furthermore, in terms of virulence, there seems to be a link between that of the donor host and that of the recipient host (Alizon et al., 2010, Hollingsworth et al., 2010). Therefore, even though we know within-host evolution is occurring, its strength does not seem sufficient to buffer the trade-off. This is supported by a recent model from Lythgoe et al. (2013), which finds that if given a little bit of importance, within-host evolution will swamp between host selective pressures. In the end, perhaps the diversity that matters in terms of virulence evolution is at another scale than the within-host diversity (sort of different ‘quasi-species’ if one wants to use this proteiform terminology).

### A comment about multiple explanatory hypotheses

Until the publication of an article by smith (2011), I did not realise how prevalent the confusion about different adaptive hypotheses to explain the maintenance of virulence was. Even brilliant researchers like smith consider that these are competing hypotheses and that if multiple infections allow to explain virulence evolution it means that the trade-off hypothesis is not at work. Interestingly, others think that in order to observe an effect of multiple infections on virulence evolution, one needs to assume that there is a trade-off.

To clarify these issues, we wrote a note in which we developed a model tailored to study plasmids infecting bacteria (Alizon and Michalakis, 2011). Since our model is general, I here present it for any type of host-parasite interaction. We used a classical SI epidemiological model with superinfection, in which we allowed for vertical transmission (with a probability  $v$ ). We decomposed superinfection into three processes: the rate at which  $i$  is infectiously transmitted ( $\beta_i$ ), the rate at which it takes hosts over ( $\epsilon_i$ ) and the rate at which the resident parasite strain  $j$  resists superinfection ( $\rho_j$ ). We assumed for simplicity that these three processes were independent but this could be modified to match specific biological situations. Overall, superinfection leading to the replacement of parasite  $j$  by parasite  $i$  occurred at a rate  $\beta_i \epsilon_i (1 - \rho_j)$ .

The Price equation approach (Day and Proulx, 2004) allows us to predict the evolution of the parasite host exploitation strategy (denoted  $x$ ), which is here taken to be the decrease in host fitness (as measured in the experiments). In the following of this subsection, we will refer to  $x$  as virulence. After some calculations based on Day and Gandon (2006), we find that the rate of change of the average value of virulence ( $\bar{x}$ ) in the parasite population is governed by the equation

$$\begin{aligned} \frac{d\bar{x}}{dt} = & \overbrace{-rv \left(1 - \frac{U + I_T}{K}\right) \sigma_{x,s}}^{\text{vert. transmission}} - \overbrace{\sigma_{x,\alpha}}^{\text{host death}} \\ & + \underbrace{U \sigma_{x,\beta}}_{\text{horiz. transmission}} + \underbrace{I_T (1 - \bar{\rho}) \sigma_{x,\beta\epsilon}}_{\text{hosts won}} - \underbrace{I_T \bar{\beta} \epsilon (\bar{x} - \sigma_{x,\rho})}_{\text{hosts lost}} \end{aligned} \quad (2.9)$$

superinfection



where the  $\sigma$  indicate genetic covariances between two parasite traits,  $r$  is the reproduction rate of uninfected hosts,  $K$  is the hosts maximum population size,  $U$  is the density of uninfected hosts and  $I_T$  is the density of infected hosts.

The four terms in the right hand side of equation 2.9 represent how virulence evolution is affected by the relationship between virulence ( $x$ ) and, respectively, parasite effects on host fecundity ( $s$ ), on host survival ( $\alpha$ ), ‘infectious transmission’ ( $\beta$ ), and finally superinfection. Genetic covariances between parasite traits allows us to predict how virulence should evolve in response to perturbations in the host population. More precisely:

- $\sigma_{x,s}$  is the correlation between virulence and the decrease in host fecundity. If the parasite does not affect host fecundity this term is nil. For a host population close to its carrying capacity, this term becomes negligible. For an expanding population, decreasing host reproduction is very costly for the parasite because of its vertical transmission ( $v$ ).
- $\sigma_{x,\alpha}$  is the correlation between virulence and the parasite induced decrease in host survival. If the parasite does not affect host survival this term is nil.
- $\sigma_{x,\beta}$  is the correlation between virulence and horizontal transmission to susceptible hosts (‘infectious transmission’). This is usually what is referred to as the transmission-virulence trade-off. If there is a trade-off, we expect the correlation to be positive; otherwise zero virulence would maximise transmission. Note that the reverse condition is not necessarily true: there can be a positive correlation with transmission being a saturating function of virulence (which is necessary to have an optimal intermediate level of virulence).
- $\sigma_{x,\beta\varepsilon}$  and  $\sigma_{x,\rho}$  are both related to superinfection (i.e. transmission to infected hosts). The former term reflects the correlation between virulence and the ability to take over already infected hosts. The latter is the correlation between virulence and the ability to resist superinfection by another strain. We expect both of them to be positive; otherwise there would be no cost on the evolution of these parameters. Note that this term is scaled by the total number of infected hosts.

Several biological points are captured by equation 2.9. First, superinfections and a transmission virulence trade-off can be acting at the same time and act in the same direction on virulence (if both  $\sigma_{x,\beta}$  and  $\sigma_{x,\beta\varepsilon}$  are positive). Second, virulence can be adaptive even if it does not affect transmission rate ( $\beta$ ), e.g. if  $\sigma_{x,\rho}$  is positive. Third, the way virulence is measured matters as, for instance, a correlation between virulence and castration ( $\sigma_{x,s}$ ) is not relevant if the parasite cannot be transmitted vertically ( $v = 0$ ). Finally, host densities matter. For instance, if most of the hosts are infected ( $I_T \gg U$ ), then the superinfection terms will dominate over the transmission-virulence term.

### A side-comment on $R_0$

Some might be surprised that I would use  $R_0$  to refer to parasite invasion fitness. Over these last years, criticizing the  $R_0$  has become almost as popular as criticising the trade-off hypothesis. Some of the critiques are justified (and sometimes even relevant as we will see below), but several seem out of place.

For instance, a recurring critique stems from the fact that when people think about  $R_0$ , they think  $\beta S/(\mu + \alpha + \gamma)$ , which is the expression for diseases transmitted horizontally and by contact. If the disease has a different transmission mode or a more complex life-cycle, this

expression becomes useless. Therefore (obviously?),  $R_0$  should be derived for each type of host-parasite interaction.

A more valid concern is that the  $R_0$  requires the host population to be fully susceptible. Yet, by definition, this is not the case when a parasite evolves because typically mutants emerge in a host population already infected by one (or several) resident strains. Therefore, what should be measured is  $R$ , i.e. the  $R_0$  of a mutant strain in an already-infected population (a measure that competes the resident and the mutant strain). This is discussed in details by [Dieckmann \(2002\)](#), who shows that unless there are some unusual feedbacks<sup>6</sup>,  $R_0$  and  $R$  yield the same evolutionary stable strategy for the parasite. Arguably, *the* case where feedbacks are crucial is when there are multiple infections.

Finally, one could mention two more recent concerns about  $R_0$ . The first concern is about persistence: the strain that has the highest  $R_0$  might also be the strain that is the less likely to persist in the environment due to stochastic perturbations. As shown by [King et al. \(2009\)](#), other measures can then be used to ensure both high reproductive ratio and high persistence. The second point also has to do with stochasticity: even with an  $R_0 > 1$ , a disease might not invade. In fact, in a totally homogeneous population, stochastic models predict that the probability of emergence is  $1 - 1/R_0$ . In terms of virulence evolution, [André and Day \(2005\)](#) have shown that such stochastic emergence favours less virulent strains. We will come back to this issue when discussing disease emergence. Another point that will be mentioned later on is spatial structure: typically, the existence of a structured contact network between hosts affects disease spread and often makes it complicated to express  $R_0$ .

---

6. One example he considers is when the virulence is a function of the number of infected hosts, which could for instance reflect a scenario where hospitals become crowded with infected people.

## 2.3 Evolutionary epidemiology of multiple infections

*In many ways, multiple infections are the rule rather than the exception as very often hosts are often simultaneously infected by different parasite strains or even different parasite species. I first present a review in which we discuss how multiple infections are expected to affect virulence evolution. Then I discuss two research questions I investigated related to co-infections. The first illustrates the importance of epidemiological feedbacks and the fact that knowing which strain benefits from the within-host competition is not always sufficient to predict virulence evolution. The second study investigates the consequences of co-transmission (the fact that two parasite strains or species can be transmitted simultaneously instead of sequentially) on parasite evolution.*

With the accessibility of massive sequencing, we are witnessing an exponential increase of studies that consider what they call ‘communities’ (or even ‘meta-communities’) of pathogens within hosts. My goal is not to say that massive sequencing is useless but rather to focus on the supposedly simpler case of co-infections by up to two parasites (what one might call ‘micro-communities’) to show that our understanding of these is already quite limited. Therefore, claiming that we can get a detailed understanding of the ecology and evolution of the whole diversity of commensal and parasite micro-organisms within a host is perhaps premature.

### 2.3.1 Multiple infections and virulence evolution: a review

I only outline here some of the more important findings we made while writing our review with Jaap de Roode and Yannis Michalakis. the reader should refer to it for further details and examples ([Alizon et al., 2013](#)).

Our first goal was to highlight that many, if not a majority, of studies that claim to be measuring virulence evolution actually measure what we call the ‘overall virulence’, which is the virulence expressed by a co-infected host. This is summarised in [Figure 2.3](#). Measuring the overall virulence is not uninteresting *per se* (actually this is the most direct consequence co-infections have on host health) but it is not an evolutionary question. In order to know how virulence evolves, one need to be able to assess which parasite (the most or the less virulent) has benefited from the co-infection. This is more difficult to measure than overall virulence because one needs to be able to discriminate between the two parasites and estimate their respective transmission successes.

The second point that had us think more than we anticipated is the difference between co-infections caused by different parasite strains and co-infection by different parasite species. In earlier models of virulence evolution, this question was not really addressed (see [Alizon and van Baalen, 2008](#), for an example of researchers who sort of dodge this question). A step forward was made in the study by [Choisy and de Roode \(2010\)](#). Ironically enough, they write that their model is a model for co-infections by parasites from the same species even though this is arguably the first evolutionary epidemiology model to capture co-infections by different species. Indeed, in their model, there are always two resident strains. Each of these strains can be challenged by a mutant. This differs from other co-infection models, which have only one resident strain and where co-infections occur between either two resident strains or between the

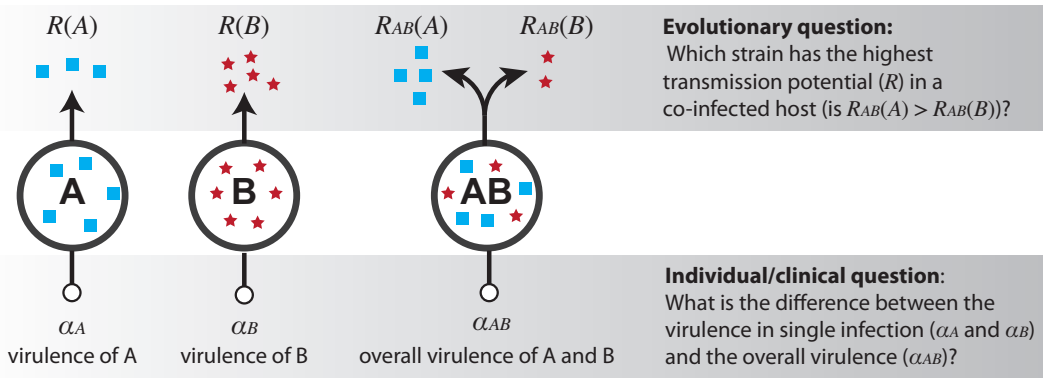


Figure 2.3: **Difference between virulence evolution and overall virulence.** Considering the virulence expressed in co-infected hosts (the ‘overall virulence’) can inform us on how parasites interact inside the host but it is not very helpful to predict how multiple infections affect virulence evolution. Figure from [Alizon et al. \(2013\)](#).

resident strain and a mutant strain ([van Baalen and Sabelis, 1995](#)). In the latter case, since co-infections always occur between more or less similar parasites, it seems impossible to envisage multiple species. These models are summarised in Figure 2.5. Although our reflections were triggered by epidemiology models, they also have biological implications. One of these is that for competition experiments within hosts, the two-species context requires to compete at least 3 parasite strains: one parasite strain of species A and two strains of species B (which should differ in virulence). The consequence of the co-infection on virulence evolution is then assessed by comparing the reproductive success (or the number of transmission events realised over the course of the co-infection) of the two strains of species B. Practically, this can prove to be challenging because not only is it necessary to distinguish between different parasite species, but also between parasites of different strains.

One of the main conclusions from our review is that evolutionary experiments with different species seem particularly appropriate to study the evolutionary consequences of multiple infections. Indeed, the evolutionary experiment solves the issue of having to differentiate between parasite strains of the same species: all one needs to do is measure the virulence of a parasite species before and after the experiment. Note also that working with different species solves another issue, which is that of the control treatment. In an experiment with only one species, it seems difficult to perform the experiment with or without multiple infections. Conversely, if the co-infections occur between different species, it is then straightforward to have one treatment with one species evolving alone and another treatment where the parasite species evolves with another parasite species (one can also think of co-evolving treatments of course).

### 2.3.2 Importance of epidemiological feedbacks

A large proportion of the theoretical literature on co-infection consists in kin selection models because they offer a simple way to capture conflicts of interests between unrelated strains. Put

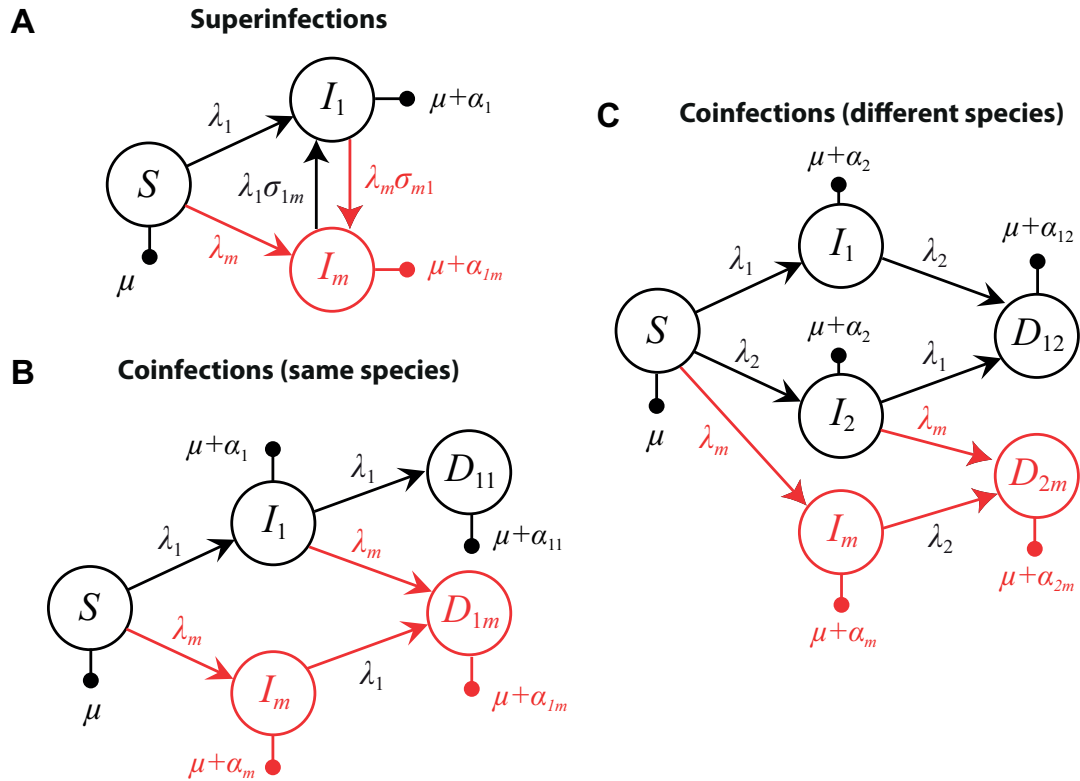


Figure 2.4: **Epidemiological models to study virulence evolution.** A) With super infections, the parasites never coexist inside a host. B) With co-infections by related parasites, coexistence can occur until the end of the infection. Note that hosts co-infected twice by the same strain need to be accounted for for the invasion analysis to be correct. C) Considering co-infections by different species requires a different framework because there needs to be two resident strains (here we only show the mutant strain of resident strain 1). Figure from [Alizon et al. \(2013\)](#).

differently, the average relatedness between parasites infecting a host will depend on the number of strains that co-infect this host. In single infections, relatedness should be close to 1 and in co-infections by  $n$  strains it should be close to  $1/n$  if the strains have similar within-host densities. A major result of these models was to show that depending on how parasites interact within the host, increasing the number of strains per hosts (i.e. decreasing the relatedness) needs not select for more virulent strains. For instance, if parasites produce a costly public goods inside the host, then free-rider strains that do not pay the cost of producing the public goods but are nevertheless able to use it should be fitter. In such a situation, it is likely that the more virulent strain is the strain that produces the public goods (because this is what harms the host) but it is also less competitive in a co-infection setting.

The major limitation of most (if not all) of these co-infection models (reviewed in [Chao et al.](#),

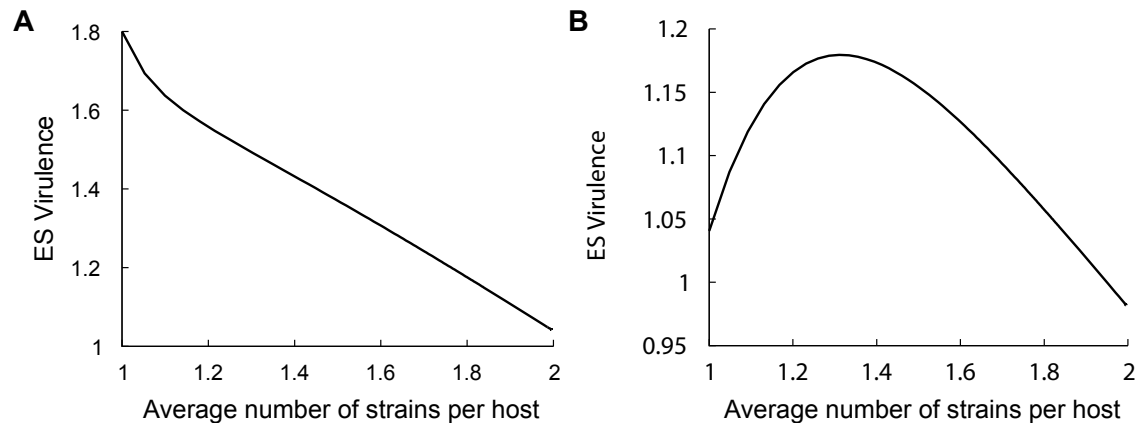


Figure 2.5: **Evolutionary stable level of virulence for bacteria producing siderophores in the host.** In A, virulence is a linear function of parasite within-host density and in B virulence is a linear function of parasite within-host density and of the total amount of siderophores produced. The pattern in A is the one expected from the within-host interactions. The pattern in B occurs due to epidemiological feedbacks. Figure from [Alizon and Lion \(2011\)](#).

2000, [Brown et al., 2002](#), [Buckling and Brockhurst, 2008](#)) is that they lack an epidemiological dimension. They do derive a between-host fitness but this is only based on what happens within a co-infected hosts. In a way, what these models are equivalent to models that would assume that all the hosts in the population are co-infected. We addressed this major limitation in a study with Sébastien Lion ([Alizon and Lion, 2011](#)).

To investigate the importance of epidemiological feedbacks in co-infection kin selection models, we decided to focus on the case of siderophore-producing bacteria. Siderophores are molecular compounds (similar in essence to hemoglobin) that chelate iron, thus allowing the pathogen to divert some of the host's resource. Siderophore production can be seen as a paradox in evolutionary biology because these molecules are costly to produce but free to use. In other words, 'cheater' bacterial genotypes that lost the ability to produce siderophores but still have the capacity to use them should be favoured. This means siderophore production is a typical example of a public goods production scenario. Evolutionary game theory predicts that 'cooperating' (i.e. producing siderophores) should be counter-selected unless the relatedness among co-infecting parasites is non-zero. Earlier models have shown that optimal siderophore production depends on the average level of within-host relatedness, which is often approximated through the inverse of the number of co-infecting strains ([West and Buckling, 2003](#)).

We used a nested model approach (by modelling both within-host and epidemiological dynamics [Mideo et al., 2008](#)) to compare four epidemiological scenarios. The first case corresponds to the assumption made by earlier models that the epidemiology can be inferred by looking only at co-infected hosts ([West and Buckling, 2003](#)). The second model adds an explicit epidemiological co-infection setting (identical to that of [van Baalen and Sabelis, 1995](#)). The

third scenario is based on the second but we also assume a saturating transmission relationship between transmission and virulence, which means that at some points the cost of increasing virulence outweighs the benefit it induces through increase transmission rate. Finally, the last scenario is also based on the second scenario with the additional assumption that virulence of an infection depends not only on the within-host parasite density but also on the total amount of siderophore produced.

Scenario 1 and 2 lead to qualitatively similar results: increasing the average number of strains per hosts always selects for lower levels of siderophore production and hence for less virulent strains (Fig. 2.5A). However, scenario 3 and 4 show that this is not necessarily the case if epidemiological feedbacks differ. In scenario 4, where the virulence is assumed to depend not only on the within-host parasite density but also on the total number of siderophores produced, we find that at first increasing the prevalence of co-infection selects for an increase in siderophore production/virulence and that it is only for higher prevalences of co-infection that siderophore production/virulence starts to decrease (Fig. 2.5B). As shown in the expression of the selection gradient in the Appendix of our article, this two stage effect of the prevalence of co-infections could be due to the fact that decreasing the prevalence of co-infection does increase the relatedness of an infection but it also increases the virulence of the infection. Hence, the benefit reaped by cooperators when relatedness is high is partly consumed by the additional cost of increased virulence.

Our study shows the importance of both epidemiological feedbacks and biological assumptions on the evolution of virulence. As we discuss in our review (Alizon et al., 2013), one of the reasons that make multiple infections by different species such a promising area for research is because they allow for epidemiological feedbacks (while also avoiding several obstacles present when working on co-infections with strains from the same species). Such evolutionary experiments could be used to see if the conclusions drawn at the level of a single host are still valid when the whole parasite life cycle is accounted for and when epidemiological feedbacks are present.

### 2.3.3 Co-transmission and the evolution of virulence

A known limitation of co-infection model is that they tend not to allow simultaneous infection by more than two parasite strains. A less obvious problem is that current models typically assume that only one parasite can be transmitted from a co-infected host upon contact. When looking at the data, this can seem like a strong assumption: several strains can be transmitted at the same time for HIV (Keele et al., 2008, reported that more than 20 % of the 104 infections they considered seem to have been initiated by at least two founder strains) or hepatitis C virus (Bull et al., 2011). There is even indication that co-transmission of HIV and HCV might occur through needle transmission (Ridzon et al., 1997). A large body of evidence supporting co-transmission, i.e. simultaneous transmission of more than one parasite strain upon the same transmission event, can be found among vector-borne diseases. In the case of human malaria, several genotypes of the same *Plasmodium* species or even from different *Plasmodium* species can infect the same mosquito (Taylor et al., 1997). More recent data on the genetic relatedness between co-infecting strains of *Plasmodium falciparum* in 8 individuals seems to be more consistent with simultaneous infection by all the strains rather than sequential infections (Nkhoma

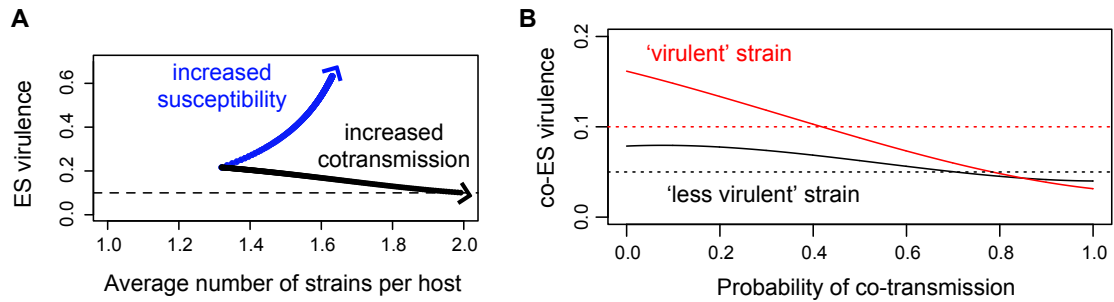


Figure 2.6: **Co-transmission and the evolution of virulence.** In A, the number of strains per host increases either due to increased susceptibility of infected hosts (in blue) or due to co-transmission (in black). The process that generates the co-infections affects the optimal level of virulence. Panel B show the level of virulence at the co-ESS for an asymmetric scenario. The dashed lines indicated the optimal level of virulence in single infections only. Note that for high levels of co-transmission the strain which is most virulent in single infection actually evolves to become the less virulent. Panels from [Alizon \(2013\)](#).

[et al., 2012](#)). In the case of viruses, [Vazeille et al. \(2010\)](#) have shown that *Aedes albopictus* can be co-infected by dengue and chikungunya and are likely to transmit the two parasite species. Infection of mosquitoes by different strains of the same virus species also occur and [Craig et al. \(2003\)](#) even isolated a mosquito infected by two dengue 2 ‘parental’ genomes and their recombinant genome. Finally, co-transmission has been described and studied for a long time in plant viruses. In fact, for some of these viruses, co-transmission is likely to be fairly common. In the *Sequiviridae* family for instance, species from the *sequivirus* genera require the presence of a ‘helper’ molecule produced by viruses from the *waikavirus* genera in order to be transmitted from the plant ([Pirone and Blanc, 1996](#)).

I studied how the probability of co-transmission of the different strains co-infesting a host affects the evolutionary epidemiology of parasite virulence ([Alizon, 2013](#)). I focused on a situation where parasites compete for host resources (such as malaria parasites competing for red blood cells, [de Roode et al., 2005](#)) because models predict that multiple infections should then select for more virulent strains.

The first result I obtained was that the prevalence of co-infections can be a poor predictor of virulence evolution. In a ‘classical’ scenario, I find that more multiple infections selects for more virulent strains (blue curve in Figure 2.6A). However, if the increase in co-infection is due to increased co-transmission, then I find the opposite result with a negative correlation between the prevalence of co-infections and the optimal level of virulence (black curve in Figure 2.6A). This illustrates that knowing host parasites interact within the host is not enough to predict virulence evolution.

This work was ideal to compare cases where co-infections are caused by two parasite strains of the same species or by two strains from different species. The case with two parasite species obviously allows us to capture richer biological scenarios. It also allows us to study three evo-



lutionary outcomes: we can search for the evolutionary stable strategy (ESS) of species 1 when the trait of species 2 is fixed, we can search for that of species 2 when the trait of species 1 is fixed and we can search for the co-ESS, which is the combination of strategies of species 1 and 2 where each ‘plays’ the best strategy in response to the other species’ strategy.

When we look at the co-ESS, results are similar to that obtained in the ‘one species’ scenario if the system is symmetric (i.e. if species 1 and 2 behave similarly). However, this needs not be the case if there is asymmetry. For instance, I consider the situation where one species is more virulent than the other. This is modelled by assuming different transmission-virulence trade-off shapes, which have different optimal virulences (the dashed lines in Figure 2.6B). Due to the asymmetry of the system, the co-ESS values of the two species differ (the less virulent species has a lower virulence at the co-ESS). As the probability of co-transmission increases, the co-ESS virulence of each species decreases. Interestingly, when the probability of co-transmission becomes extremely high, we find that the co-ESS virulence of the less virulent species becomes *higher* than that of the virulent species. Furthermore, the co-ESS virulence of the virulent species in a system with co-infections is then lower than with single infections only. These two effects are due to the fact that the higher the co-transmission probability, the more the interests of the two species are aligned. When co-transmission is extremely likely, it becomes more adaptive for the virulent species to stay on the side and for the less virulent species to take in charge the exploitation of the host because it is more efficient at exploiting host resources.

To summarize, accounting for co-transmission allowed us to show that some widespread ideas might not always be accurate. First, the prevalence of multiple infections is not necessarily a good proxy for selection for higher virulence, even if more virulent strains are more competitive at the within-host level. Second, the optimal level of virulence in a co-infection setting can be lower than that observed in a single infection setting, again even if more virulent strains are more competitive within hosts. Third, co-infections by different parasite species only lead to the same results as co-infection by the same species if there is no asymmetry and if we consider values at the co-ESS.

## 2.4 Disease spread on weighted networks

*Host population structure has long been known to affect disease spread. In particular, many results have been drawn from contact network models. However, studies typically assume that these edges of these networks are unweighted. I discuss here why this assumption might not always be realistic and how to analyse disease spread on weighted networks.*

During my PhD, I avoided to mess up with spatial structure as much as possible because this was the main interest of Sébastien Lion. In return, he refrained from studying virulence evolution and focused on relatedness. This way, each of us exploited half of the expertise of our supervisor (Minus van Baalen). Since we both graduated, I got more involved into issues related to relatedness with Peter Taylor<sup>7</sup> and with spatial structure, while Sébastien started working on virulence evolution (by combining population dynamics feedbacks, spatial structure and virulence evolution [Lion and Boots, 2010](#)). In fact, we still hope to extend our siderophore model to include spatial structure in order to have within-host relatedness (number of strains) and between-host relatedness (governed by spatial structure). In the following, I discuss some of the work I have been doing on predicting disease spread on weighted networks.

Contact structure between hosts is known to have a key influence on disease spread ([Anderson and May, 1991](#)). A striking result is for instance that the more heterogeneous the contact network, i.e. the higher the variance in the number of contacts per individual, the more rapid the initial disease spread.

One way to capture contact structure is to use a network ([Newman, 2002](#)). Such contact networks are typically described by a square binary adjacency matrix, where each term on the  $i^{\text{th}}$  line and  $j^{\text{th}}$  column can take the value 0 or 1 to indicate respectively the absence or the presence of a contact between individuals  $i$  and  $j$ . Contact networks are widely used because they possess several convenient properties, one of which being that the dominant eigenvalue of the adjacency matrix is an indicator of the initial propagation speed on this network (e.g. [Moslonka-Lefebvre et al., 2012b](#)).

The main limitation of contact networks is that their exact shape is often difficult to infer. This is why there is a continuous effort to predict disease spread from network summary statistics that are easier to estimate, such as the distribution of the number of contacts (degrees). For instance, the number of secondary infections generated by a typical infected host in a fully susceptible population, scales with the ratio of the second moment  $\langle k^2 \rangle$  and first moment (mean)  $\langle k \rangle$  of the distribution in the number of contacts  $k$  ([May and Anderson, 1987](#)) such that

$$R_0 = \frac{\beta}{\gamma} \frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{\beta}{\gamma} \left( \frac{\sigma_k^2}{\langle k \rangle} + \langle k \rangle \right) \quad (2.10)$$

where  $\sigma_k^2 = \langle k^2 \rangle - \langle k \rangle^2$  is the variance of the distribution of the number of contacts. Note that this  $R_0$  assumes that individuals update their contacts dynamically in a fully mixed fashion within the population (as approximatively seen in airborne infections). Similar results can be obtained

<sup>7</sup> We analysed a model where altruism can evolve in an island model with empty sites ([Alizon and Taylor, 2008](#)). One of the reasons why I like it is that it draws a parallel between more ecological models ([Lion and van Baalen, 2008](#)) and classical population genetics models ([Rousset and Ronce, 2004](#)).

for networks in which the identity of contacts is fixed, as approximatively seen in sexual contact networks (Durrett, 2007).

The  $R_0$  in equation 2.10 represents upper bound of the basic reproductive ratio for SIR epidemics on random networks if individuals transmit the infection at a rate  $\beta$  and recover from the infection at a rate  $\gamma$  (Volz and Meyers, 2009). The result implies that heterogeneous networks with a large or even diverging variance in the distribution of the number of contacts have a very small or even vanishing epidemic threshold (as  $R_0$  increases with  $\sigma_k^2$ ).

One of the key assumptions network models make to obtain such an elegant expression for  $R_0$  is that the transmission rate is the same between all pairs of individuals. This is materialised by the fact that all the edges of the contact matrix have a weight of 0 or 1. This is known to be a simplifying assumption (Barrat et al., 2004). A well-studied example related to infectious diseases is that of sexual contact networks, where the number of sex acts per unit of time is not constant in all partnerships (Blower and Boe, 1993, Nordvik and Liljeros, 2006, Britton et al., 2007). More generally, the number of interaction events (which correspond to potential transmission events) may vary among contact pairs. An increasing number of studies in epidemiology point to the importance of considering weighted networks. Some examples include respiratory diseases of humans (Stehlé et al., 2011) or general infectious diseases of human spreading on a social contact network (Eames et al., 2009) or on airline connection networks (Colizza et al., 2006). Several more conceptual studies have also been published in the theoretical physics literature (e.g. Newman, 2002, Britton et al., 2011).

Simplifying the reality is commendable but the problem is that tempering with the weighting of the network can affect important epidemiological properties of heterogeneous unweighted networks.

### 2.4.1 Weighting for sex

With Mathieu Moslonka-Lefebvre (who was an MSc student at ETH Zürich) and Sebastian Bonhoeffer<sup>8</sup>, we studied a situation where contact networks are likely to be weighted: the spread of sexually-transmitted infections (Moslonka-Lefebvre et al., 2012a).

Current models of STDs rely on unweighted networks. By doing so, they make the implicit assumption that the number of partnerships does not affect the sexual activity per partner. If we think in terms of sex acts, they assume that the total number of sex acts per unit of time is strictly proportional to the number of sexual partners. Existing data does not support this assumption (Blower and Boe, 1993, Nordvik and Liljeros, 2006, Britton et al., 2007). In fact, in a collaboration with Christian Althaus, we are analysing data from the National Survey of Sexual Attitudes and Lifestyles (NATSAL, Johnson et al., 2001) in the UK to address this question.

In order to study the consequences of relaxing the oversimplifying assumption that the number of sex acts is strictly proportional to the number of partners, we designed a framework where each node of the network (i.e. each individual) is assigned a ‘potential number of sex act’ (one could see this as a ‘libido’ value). Then, each individual shares this potential number of sex acts with his/her contacts. Note that again classical models make an implicit assumption, which is that the sharing is equal (an individual offers the same exact number of sex acts to each partner) but this need not be the case and in fact a random sharing would seem more appropriate. Once

8. Who came up with the title of this subsection.

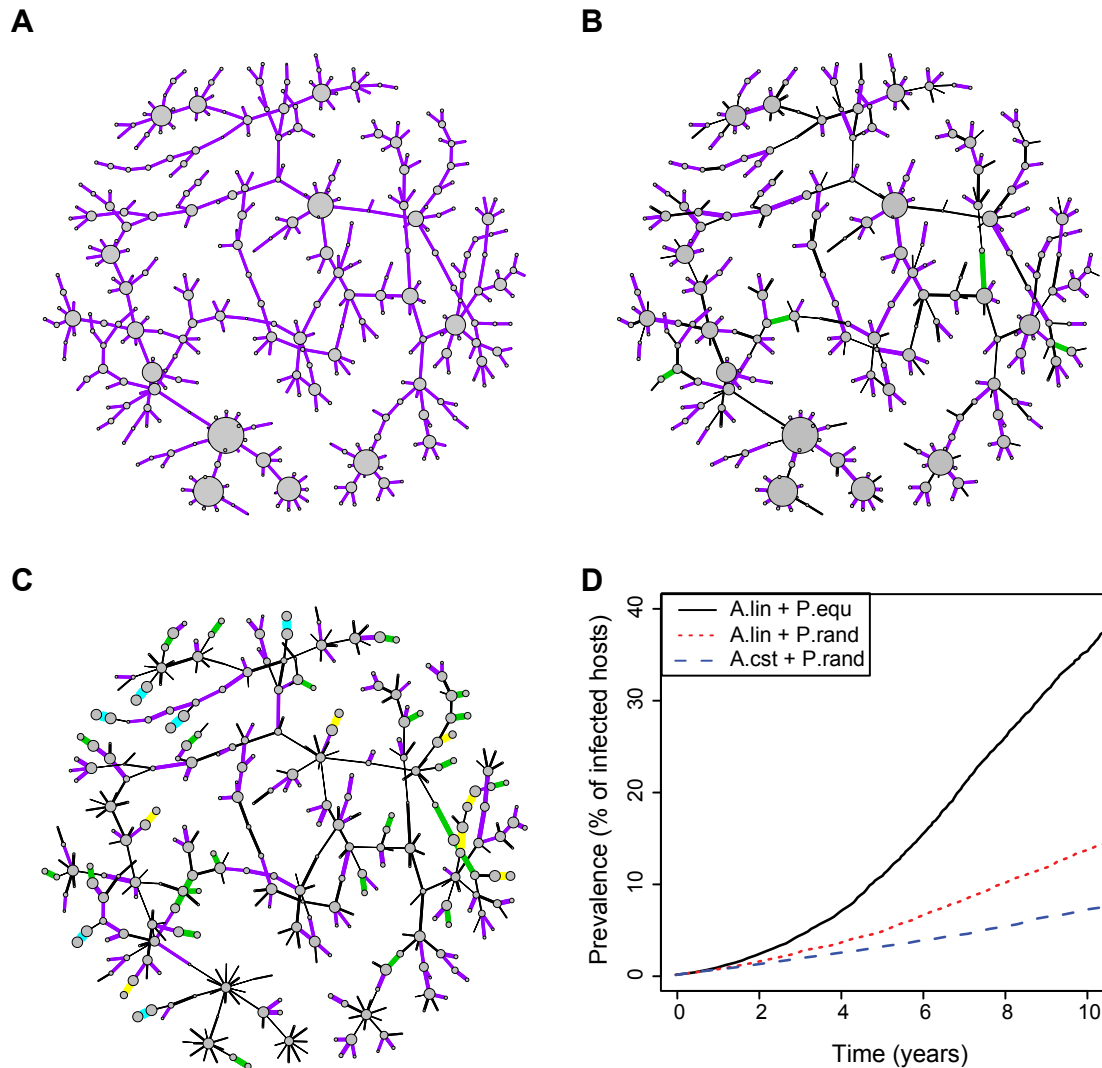


Figure 2.7: **Consequences of weighting an empirical network with realised sex acts.** A) Empirical contact network weighted with the number of realised sex acts implicitly assumed by classical models, B) same as A but with a random partitioning, C) same as A but with a constant allocation and a random partitioning. In panels A, B and C node diameter represents the number of realised sex acts per time step by an individual. Edge colour indicates less than 2 (in black), between 2 and 4 (purple), between 4 and 6 (green), between 6 and 8 (yellow) and more than 8 (cyan) sex acts per week. D) Median STI prevalence as a function of time with the ‘classical’ approach (black), with a random partitioning (red) and with a constant allocation and a random partitioning (blue). Figure from [Moslonka-Lefebvre et al. \(2012a\)](#).

we know how many potential number of sex acts each node of the network has and how these are shared by each node to each contact, we can infer the weighted of the network edges by taking the min of what two partners allocated to each other. The network weights are then normalised to make sure that the total number of sex acts occurring on the network is constant.

Our method allows us to study the effect of individual variations in total number of sex acts and variations in how sex acts are partitioned amongst contacts. We show that, even with the oversimplifying assumption that the total number of sex acts is strictly proportional to the number of partners, assuming that the partitioning of sex acts is random (and not equal) leads to important modifications of the network topology that strongly decrease the speed of disease spread. Further assuming that the total number of sex acts is constant and independent of the number of partners slows the speed of spread even more.

Interestingly, this slowing of disease spread can be analysed in the context of HIV. Several studies have pointed out that given that HIV (and other STI) spread on sexual contact networks and that these are known to be extremely heterogeneous, one could have predicted an even faster spread of HIV. What we suggest in this work is that even though the sexual-contact networks appear to be heterogeneous, they can nevertheless behave as random networks. The reason for this is that they are weighted but that this weighting does not appear in contact tracing studies. In the article, we show that when we use conservative estimates for the number of sex acts per week and the probability of disease transmission, we find a disease doubling time at the epidemiological level which is strongly over-estimated on unsighted networks but much more realistic on weighted networks.

#### 2.4.2 Analytical framework to predict spread on weighted networks

Most of the studies on weighted networks (including ours) have in common that they resort to (heavy) numerical simulations. Indeed, contrary to unweighted networks, we lack analytical frameworks to analyse epidemic spread on weighted networks.

With Christel Kamp, we were able to build a framework to obtain some analytical insights on disease spread on a weighted network (Kamp et al., 2013). These results build on configuration type network epidemic approaches that Christel and others have developed (Volz, 2008, Kamp, 2010). The idea of this approach is to divide the host population into subclasses, each of which is defined by the number of contacts a host has. The model then follows the proportion of each host class. The difficulty of the model is to express and to keep track of the probabilities that an susceptible host interacts with an infected host ( $p_{SI}$ ) and that an infected host interacts with another infected host ( $p_{II}$ ). These are calculated by injecting the probability generating functions of the probability for an individual (susceptible or infected) to have  $k$  contacts (see Kamp, 2010, for further details).

Recently, Christel has managed to extend her framework to include joint distributions of the number of partners and number of sex acts (instead of ‘simply’ using distribution of number of partners). The calculations are much too technical to be discussed here. The general idea is that we can parametrize a model in a general fashion using the joint probability distribution  $P_{kl}$  of an individual to have  $k$  contacts among which (s)he randomly distributes  $l$  interaction events.

One interesting insight we get is that the framework provides us with analytical expressions for the rate of early epidemic expansion ( $r_0$ ) and the basic reproductive ratio ( $R_0$ ) of the infection.

This is particularly interesting because one of the reasons why weighted networks are not very studied is because they usually require numerical simulations. With a mean field approximation, that is without spatial structure, the exponential growth of the epidemics is given by  $r_0^{MF} = \beta \langle l \rangle - \gamma$ , where  $\langle l \rangle$  is the mean number of sex acts in the population. Also,  $R_0^{MF} = \beta \langle l \rangle / \gamma$ . In a weighted network, these quantities can be expressed in the following way:

$$r_0 = \left( \frac{\langle kl \rangle}{\langle l \rangle} - 2 \right) \frac{\langle l \rangle}{\langle k \rangle} \beta - \gamma \quad (2.11a)$$

$$R_0 = \frac{\langle \frac{l}{k} \rangle \beta}{\langle \frac{l}{k} \rangle \beta + \gamma} \left( \frac{\langle kl \rangle}{\langle l \rangle} - 1 \right) \quad (2.11b)$$

The important point is that these expressions scale with the second moment  $\langle kl \rangle$  of the joint probability distribution  $P_{kl}$ . This implies that the number of contacts ( $k$ ) and the interaction events ( $l$ ) an individual maintains both affect epidemic spread and that the correlation between these quantities matters: for epidemic control and targeting individuals with most contacts or interaction events can prove to be much less efficient than targeting those who maximise both.

The framework also allows for a derivation of the full time course of epidemic prevalence and contact behaviour of susceptible, infected and recovered individuals (in terms of the probability generating functions of the degree distributions).

Since this framework relies on summary statistics of the network and does not require knowledge of the exact shape of the network, it can be parametrized using large scale survey data. As shown in Figure 2.8A, we could use the publicly available data from the National Survey of Sexual Attitudes and Lifestyles (NATSAL [Johnson et al., 2001](#)) carried out in the UK in order to obtain information about the distribution of the number of sexual partners over the last 5 years and the distribution of the number of number of sex acts over the last 4 weeks (and of course the joint distributions). In Figure 2.8B, we show the results of simulations performed on weighted networks (in grey), the result for a non-weighted network (in black), the prediction made using our framework (in red) and the prediction made with an additional assumption to correct for assortativity (in orange), which is the fact that nodes tend to be in contact with nodes that are similar (in terms of  $k$  and  $l$ ).

The exponential growth rates of the epidemics are  $r_0 = 0.021$  per year for the unweighted network and  $r_0 = 0.0034$  per year for the weighted network. This confirms that the classical scenario supports faster epidemic expansion. The correction for assortative effects (in orange) underestimates the epidemic prevalence in the network because in the NATSAL network heterogeneity in the number of contacts and interaction events does not lead to a strong network segregation, i.e. individuals with a single or few contacts are not isolated. Although the survey data shown in Fig. 2.8A only provide us with a rough picture of the real transmission network and although relying on the number of partners during 5 years overestimates the number of concurrent partners, the data are sufficient to confirm a remarkable reduction in the speed of epidemic expansion when shifting from a classical unweighted transmission network towards a more realistic weighted transmission network. This finding is in particular consistent with our earlier simulation study on epidemic spreading along a network of homosexual contacts ([Moslonka-Lefebvre et al., 2012a](#)).

Most analytical and numerical models predict disease spread on network using only one sum-

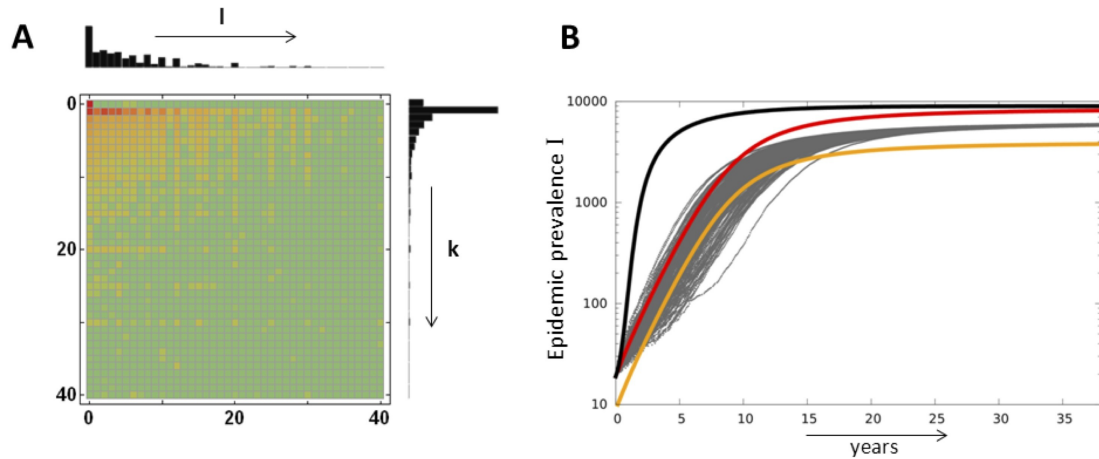


Figure 2.8: **Epidemics on a weighted sexual contact networks inferred from epidemiological data.** A) Characteristics of the heterosexual contact network inferred from the NATSAL contact tracing study (Johnson et al., 2001). Histograms on the top and the right sides show the distributions of number of sex acts ( $P_l$ ) and number of partners ( $P_k$ ) respectively. The graph shows the joint distribution ( $P_{kl}$ ) with higher values in red and lower values in green. B) Dynamics of an SI epidemic spreading on an unweighted (black line) or a weighted sexual contact network. Epidemic prevalence from simulations on the weighted empirical network (grey dotted lines) are enclosed by the lines corresponding to the predictions from  $P_{kl}$  (red line) and predictions with assortativity corrections (orange line). The network has been reduced to nodes with  $k > 0$  and transmission probability per sex act is  $\beta = 0.01$ .

mary statistics, the distribution of the number of partners. We show that additional insights can be gained, while maintaining some analytical results, by including another summary statistics, such as the distribution of the number of sex acts knowing the number of partners. These data are easier to collect than full information of the contact network (especially for a weighted network), which makes our framework widely applicable. We demonstrate this applicability here using data from the NATSAL study conducted in the UK. We note that for some artificial distributions, our results begin to diverge from simulations on real networks. However, the framework has proven to be applicable for empirical distributions and analysis of more empirical data will allow us to further test the robustness of the method using more realistic assumptions.

## 2.5 Within-host ecology and evolution

*It could be tempting to say that within-host evolution and multiple infections are but one process. However, within-host evolution has several particularities. First, the nature of the transmitted strains varies. Second, within-host evolutionary dynamics can be quite complicated and possibly explain clinical symptoms. Third, I show that levels of selection can be detected by comparing within- and between-host evolutionary rates.*

One of the consequences of the emergence of HIV is that it led to the realisation that pathogens evolve over the course of the infection. This can generate complicated within-host evolutionary dynamics, which one can try to link to disease progression in the case of chronic infections. Within-host evolution also grasped the attention of evolutionary biologists because it is a clear case of evolution at two distinct levels (Levin and Bull, 1994). However, in spite of this, within-host evolution is still mainly considered for quantitative traits (such as replication rate) in chronic infections and most of the studies on more categorical traits (such as drug resistance) still consider within-host processes as instantaneous (a host switches from the class ‘infected by a sensitive strain’ to the class ‘infected by a resistant strain’). In Alizon et al. (2011), we argued that the dynamics of strain replacement can give some insights concerning the speed at which drug resistance evolves. Below, I described some of the research projects I have developed that address questions related to within-host evolution.

### 2.5.1 The problem: what gets in differs from what gets out

I met Fabio Luciani at a Jacques Monod conference in 2007 and we were both struggling to get some papers out. He was working on hepatitis C virus (HCV) and at the time I was working on my transmission-recovery trade-off model. We decided to combine our efforts to study the within-host evolutionary dynamics of HCV.

One of the peculiarities of this virus is that some patients clear the infection naturally within a few weeks or months, whereas others remain infected for life. By developing a within-host model capturing the interaction between infected cell and the immune response (Luciani and Alizon, 2009), we managed to capture this pattern. More precisely, in some runs the parasite managed to evade the immune response to establish a chronic infection, whereas in others the failure of escape mutants to emerge led to an acute infection. We studied in further details how variation in model parameters affected this outcome.

One other aspect we focused on was the evolution of the virus replication rate during the infection. Our model incorporated a transmission-recovery trade-off (Alizon, 2008) such that increasing virus replication rate led to increased detection (and clearance) by the immune response. One of the difficulties we faced was that this rate could evolve strongly in chronic infections. In the end, we had one initial trait value (that of the strain that infected the host) but many replication rates for the strains transmitted from this host. This highlighted the difficulty to predict virus evolution at the epidemiological level from the information obtained in a single infection. One possibility would have been to develop a between-host component of the model (which we are now working on now) but the within-host model was already quite complicated. The solution we found was to compute the ‘mean weighted replication rate of the



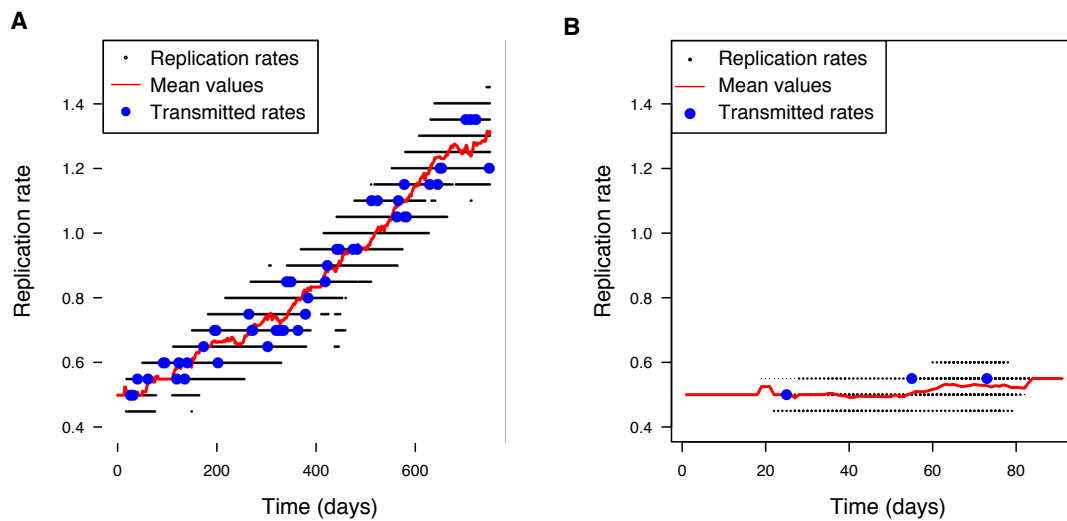


Figure 2.9: **Within-host evolution of replication rate in a chronic (A) or an acute (B) infection.** The replication rate values are in black, the mean trait value is in red and the ‘transmitted strains’ (which are the ones we sample) are in blue. Note that the parameter values are the same in the two panels. The differential outcome of the infection comes from stochastic mutations (whether the virus manages to escape the initial immune response). Figure from [Luciani and Alizon \(2009\)](#).

transmitted strains’. To do so, we calculated the trait value (replication rate) of all the transmitted strains and weighted each of these by the number of transmission events they would generate in an infection. One can make a parallel with an inclusive fitness calculation where the contribution of the offspring of an individual is weighted by the reproductive success of these offspring. By plotting this mean rate, we can graphically determine an ESS value for the trait provided that we make (strong) simplifying assumptions about the epidemiology. This approach allowed us to show that even if the number of transmission events is maximised for low replication rates, selection can favour strains with high replication rates due to within-host competition.

In our Figure 2.10, we show how increasing cross-immunity intensities selects for less rapidly replicating strains. This pattern might seem surprising because one could expect cross-immunity to exacerbate the competition between viral strains. However, we are in a setting where viruses compete for ‘immune free’ space and, since immune activation is a function of replication rate, low replication is the only way to persist in a host with strong immunity. In fact, what happens is that if the level of cross-immunity is strong, strains with high replication rate always lead to acute infections and never get transmitted. That low levels of cross-immunity lead to the highest replication rates can also appear as counter-intuitive because both the number of transmitted strains (Fig. 2.10A) and the duration of the infection (not shown) are maximised for values ranging from 0.2 and 1.0 (depending on the replication rate of the initial strain). What happens is that within-host evolution favours more rapidly replicating strains because, even though they

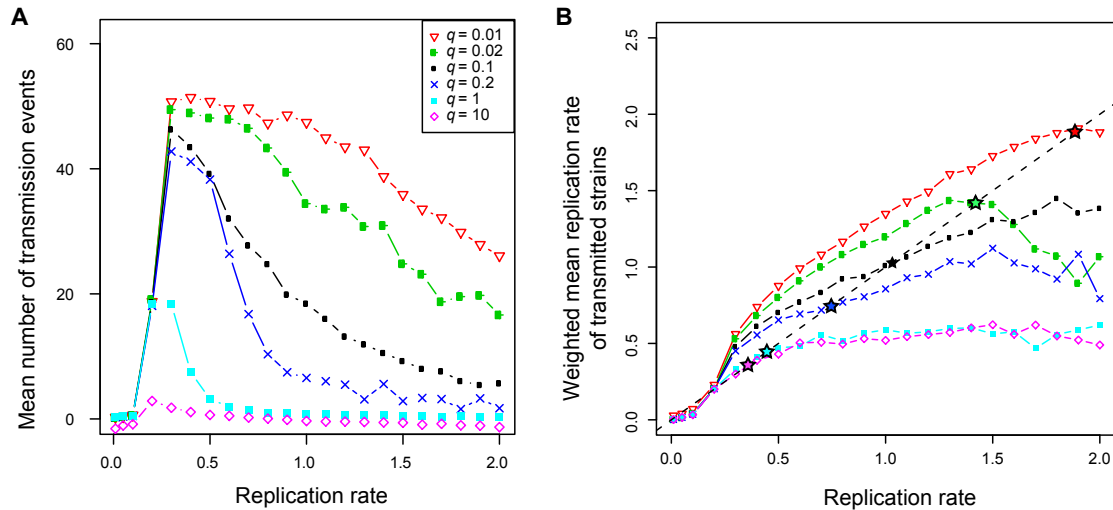


Figure 2.10: **Number of transmission events (A) and Weighted replication rate of transmitted strains (B) as a function of the replication rate of the initial strain infecting the host.** The higher the value of  $q$ , the higher the cross-immunity. In B, the ESS values (denoted with a star) are found by taking the intersection between a curve and the  $y = x$  line. Figure from [Luciani and Alizon \(2009\)](#).

lead to shorter infections, they have a short-term advantage. This is an illustration of the short-sightedness of within-host evolution. This effect is less present with strong cross-immunity because it generates a strong selective pressure in the other direction.

### 2.5.2 HIV co-receptor switch

One of the (many) puzzling features of HIV infections is that in many patients (not all of them though) an important change occurs in the ecology of the virus after a couple of years. This virus infects cells that express CD4 receptors on their surface but, in order to enter the cell, the virus also needs to bind a second receptor (known as the co-receptor). Early in an infection, HIV uses the CCR5 coreceptor but after several years a switch occurs and the virus tends to use the CXCR4 coreceptor. This coreceptor switch has attracted a lot of attention because its timing strongly correlates with a deterioration in the patient's health and the onset of AIDS.

Interestingly, it is possible to tell from the virus genome which coreceptor the virus uses (some viruses can even use both). Even more puzzling, it seems to only take 2 or 3 amino-acid substitutions to change an R5 virus (using CCR5) into an X4 virus (using CXCR4). A recurrent question is 'Why does this switch occur so late in the infection?'. Indeed, even if a host is initially infected by an R5 virus, it should be a matter of days before an X4 mutant virus appears. One explanation has been suggested by [Regoes and Bonhoeffer \(2005\)](#) that only requires evolution. Let us imagine that the X4 virus has a higher fitness than the R5 virus and that it only takes 3 intermediate mutants to get from one to the other. If these intermediate

mutants have a lower fitness and if the fixation event is stochastic, then it can take a while before the X4 mutant invades the population. The more pronounced the fitness valley the virus has to cross, the longer it all takes for the coreceptor switch. Another explanation, more popular amongst virologists, was proposed by [Ribeiro et al. \(2006\)](#). They designed a within-host model capturing the dynamics of naive and memory target cells. In their model, all virus types (R5 and X4) are always present. However, they assume highly non-linear functions to describe the within-host dynamics, which generates an instability if target cells drop below a threshold value. This instability goes along with the replacement of one virus type by the other.

With Barbara Boldin, we wanted to combine the ecological and evolutionary approach to tackle this question ([Alizon and Boldin, 2010](#)). More precisely, our idea was to see whether the co-receptor could be described as an evolutionary branching event using an adaptive dynamics framework. The specificity of adaptive dynamics is that fitness is highly relative and a trait value can be very adaptive or very deleterious depending on the state of the environment. In this particular case, the state of the environment is the density of target cells, which is governed by the trait of the resident virus strain. A behaviour observed in many adaptive dynamics models is that a population can evolve towards an evolutionary singular strategy, which turns out to be evolutionarily unstable. As a consequence, a resident strain with this trait value can be invaded by mutant strains with any trait value (higher or lower), which can lead to evolutionary branching into a dimorphic population ([Geritz et al., 1998](#)).

We developed a within-host model to study the evolution of virus replication rate in a heterogeneous environment (with two types of target cells corresponding to the CCR5 and the CXCR4 receptor). Our model exhibited many features of what was observed *in vivo*, i.e. that it can take a long time before the evolutionary branching occurs, that evolutionary branching might not always occur and that target cell density drops after the switch. We also determined the parameter region where the branching occurred and found that it requires some degree of spatial clustering between CCR5<sup>+</sup> and CXCR4<sup>+</sup> T-cells. Another result is that once the branching has occurred, we have a dimorphic virus population with a rapidly replicating strain and a slowly replicating strain. The former is found more in CCR5<sup>+</sup> T-cells, whereas the latter is found more in CXCR4<sup>+</sup> T-cells. Interestingly, if we seed an infection with a slowly replicating strain, the virus will predominantly be found in CCR5<sup>+</sup> T-cells (because these cells are a better resource and they are accessible if there is no competitor).

What we think is novel about this result is that the R5 or X4 molecular status of a virus is not required to explain how the infection works (similar dynamics are obtained with or without specifying the type of cells a virus can infect). This helps to explain why the co-receptor switch would only occur later in the infection. Even if the infection is seeded by a virus from the X4 molecular type, this virus can rapidly evolve to become an R5 virus, thus accessing a more valuable resource (CCR5<sup>+</sup> T-cells). It is only when the virus population has evolved to the branching point that the branching can occur because both specialist morphs can replace the generalist. Seen this way, the molecular switching is more a consequence of the branching in the phenotypic trait. As a aside, it is worth mentioning that in the beginning of the HIV research, the switch was referred to as the ‘phenotypic switch’ because the early and late viruses led to distinct patterns in cell culture ([Moore et al., 2004](#)).

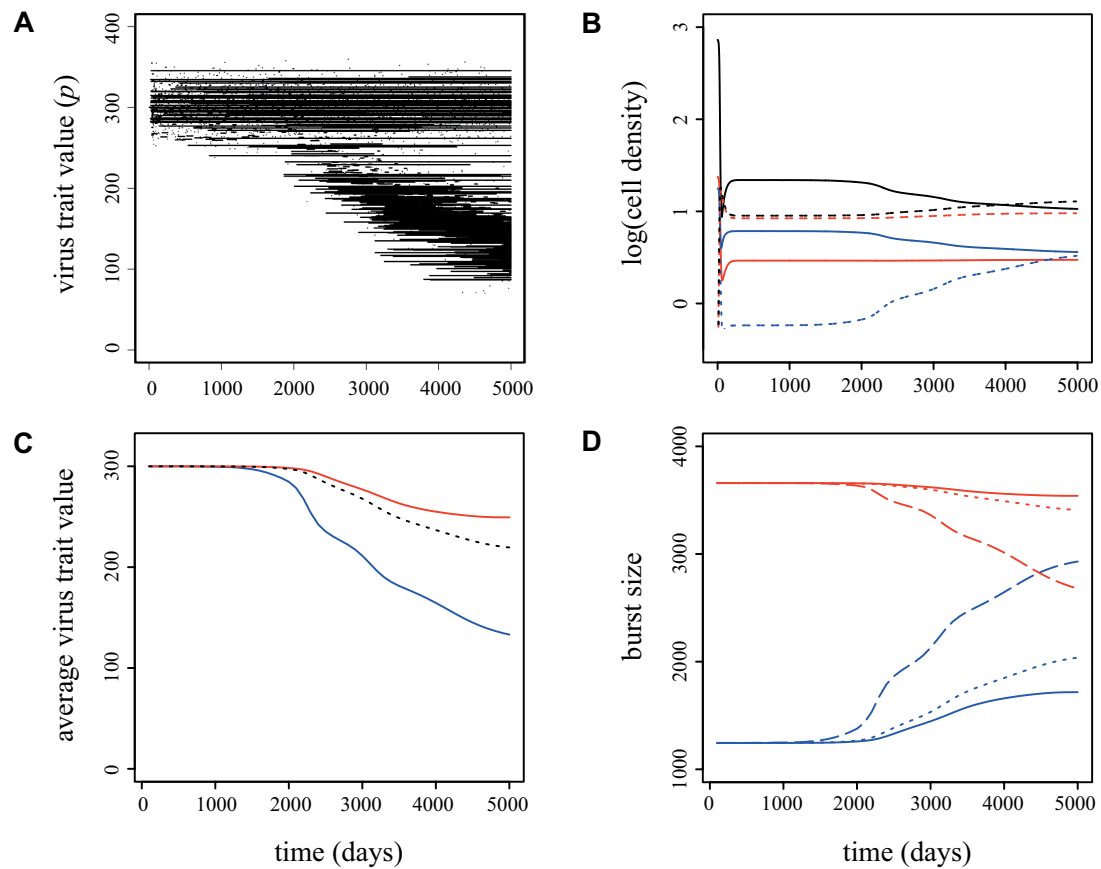


Figure 2.11: **Modelling the HIV co-receptor switch.** A Evolutionary dynamics of the virus within-cell production rate,  $p$ , of the strains present in the host at a given time; B Densities of uninfected (plain lines) and infected cells (dashed lines) for CCR5<sup>+</sup> cells (red), CXCR4<sup>+</sup> cells (blue) and overall (black); C Average exploitation strategy in the host (in black), in CCR5<sup>+</sup> cells (red) and in CXCR4<sup>+</sup> cells (blue); D Burst size in CCR5<sup>+</sup> cells (red) or in CXCR4<sup>+</sup> cells (blue) of viruses with the host average value (dotted line), viruses found in CCR5<sup>+</sup> cells (plain line) and viruses found in CXCR4<sup>+</sup> cells (dashed line). Figure from [Alizon and Boldin \(2010\)](#).

### 2.5.3 Modelling the course of a HIV infection

There are plethora of models of HIV within-host dynamics but relatively few reviews. This is perhaps because authors in this field specialise in writing rather than reading (although, in their defence, it is probably impossible to read all the models published on HIV). With Carsten Magnus, we focused on a specific question, which is how theoreticians have modelled the course of a HIV infection ([Alizon and Magnus, 2012](#)). HIV offers an illustration of a very repeatable

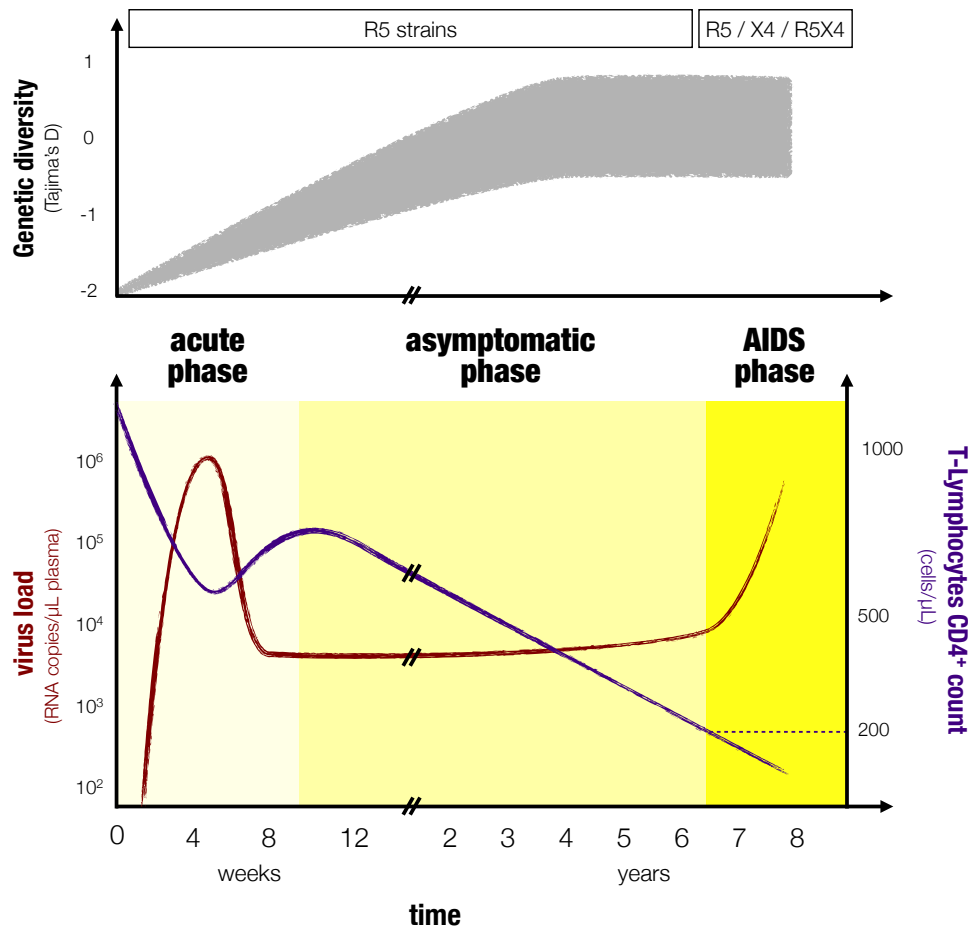


Figure 2.12: **Typical course of a HIV infection.** The top panel shows the diversity along with the type of HIV variant that dominates (as in [Tebit et al., 2007](#)). The diversity measure shown here is Tajima's D, which compares the average pairwise distance of a set of sequences to the number of sites that are polymorphic ([Tajima, 1989](#)). The bottom panel shows the dynamics of the viral load, in red, and the CD4<sup>+</sup> T-cells, in blue ([Levy, 1988](#)) during the three phases of a HIV infection. Figure from [Alizon and Magnus \(2012\)](#).

dynamical process, which invariably ends with the AIDS phase and the death of the host (Figure 2.12). Part of the efforts in the field have been directed towards understanding how such a property of the host level (host death) can result from within-host dynamics.

In our review, we highlight several aspects. One of these is the role of virus evolution in the progression of the infection. Does this evolution, which is undeniable (e.g. with increasing levels of diversity, Fig. 2.12), have a driving effect? Models without evolution were instrumental in understanding several aspects of the biology of HIV. For instance, they allowed to show that immune limitation is key because if there was only target-cell limitation, the virus load

should be unaffected by the presence of drugs (the decrease in new cell infections would be counterbalanced by the increased availability of susceptible cells). Some of these ‘ecological’ models manage to mimic the course of an HIV infection. However, they suffer from several limitations. First, capturing the slow decline of the target cell density in the asymptomatic phase is complicated. Furthermore, it is particularly difficult to end the asymptomatic phase after a variable time period. In fact, most of these ‘population dynamics’ models need to resort to some kind of trick. A classical one (which I find particularly inelegant) is to change the parameter values of the model over given time. Another possibility is to use complicated and non-linear functions, as mentioned above (Ribeiro et al., 2006).

Evolutionary models had a lot of success in the HIV field because they managed to account for the variable duration of the asymptomatic phase and for the AIDS phase (Nowak et al., 1990). They also showed an increase in virus diversity over time, which was consistent with the data. In our review, we stress the change we went between the 1990s, where almost any HIV within-host model would get into at least *PNAS* and the present time where generalist journals tend not to consider a model on HIV if it does not include data analysis. This reversal is probably due to an excessive enthusiasm in the 90s concerning Martin Nowak’s models. As we discuss in the review, his models were criticisable in several ways (one of these being that mutations occurred independently of virus densities) and he never really bothered to test them with data. However the fact remains that these models have had a major influence on evolutionary ecologists because they showed that the combination of population dynamics and evolutionary biology can lead to emergent properties at another level.

Overall, the conclusion is that virus evolution seems to be required to explain the progression to AIDS. However, none of the current models are really satisfying. Given the high hopes that were put into modelling (in the 90s, any model on HIV went at least to *PNAS*), one could interpret this as a failure. To balance this, one could argue that experimental approaches did not either provide us with a satisfying explanation as to why infected individuals progress so slowly to AIDS (and why the variance amongst hosts is so great). Finally, while writing this review, we were surprised by the strong decrease in the number of articles that fitted our selection criterion (mathematical models of the course of an HIV infection) over time. In fact, the majority of the articles we cite were published in the 1990s. Over the last decade, it has become increasingly difficult to publish models in the field of HIV if they do not analyse (preferentially novel) data. That models are backed up with biological observations is of course a necessity, however imposing data analysis also tends to restrict the role of mathematical models to parameter estimation at the expense of conceptual modelling.

#### 2.5.4 Levels of selection: a signature in HIV evolutionary rates

Are levels of selection relevant in the context of HIV? It could be that the within-host selective pressures are the only ones that matter. Evolutionary rates (ER), i.e. the rate at which substitutions are fixed in the HIV genome, seem to suggest that there are indeed two levels to study. More precisely, evidence based on part of the envelope gene (*env*) showed that within-host evolutionary rates (WHER) are higher than between-host evolutionary rates (BHER) for HIV (Lemey et al., 2006, Pybus and Rambaut, 2009).

If transmission is random, by which I mean that all the strains present in a host at a given

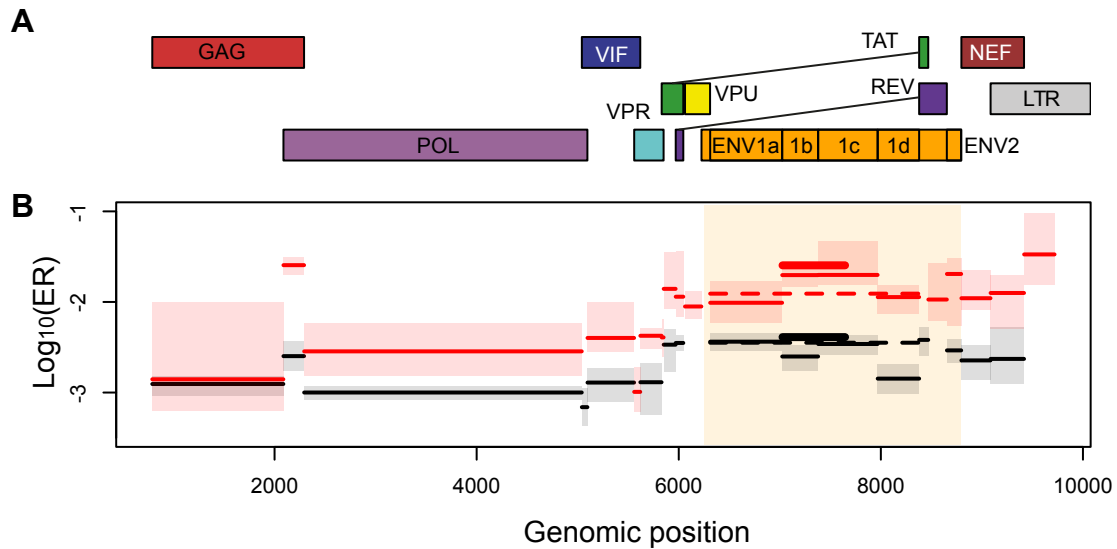


Figure 2.13: **Evolutionary rates at the WH and BH level across the HIV genome.** A) HIV genome and B) Median evolutionary rates for the pooled WH data (in red) and BH data (in black). In panel B, shaded boxes indicate 50% credibility intervals. The thick line shows the C2V5 region (studied by former studies) and the dashed line the ENV1 segment. Figure from [Alizon and Fraser \(2013\)](#).

time have the same probability to be transmitted, we expect WHER and BHER to be identical. At most, there could be a higher variance in the BHER because the transmitted strain can be slightly ‘more evolved’ or ‘less evolved’ than the average. That WHER are higher than BHER by an order of magnitude (approx.  $10^{-2}$  for the former *versus*  $10^{-3}$  for the latter) suggests that something happens when a new infection is generated. [Lythgoe and Fraser \(2012\)](#) narrowed it down to three possibilities: 1) there is backwards evolution and in the initial stages of an infection a virus loses all the mutations it has accumulated during the course of the previous infection, 2) evolutionary rates vary over the course of an infection (the virus evolves faster in the chronic stage when it is sampled but it is more transmitted in the early stages) and 3) ‘more evolved’ strains are less likely to initiate a new infection. They argue that current data seems not to support hypotheses 1 and 2. Furthermore, they develop a model based on virus generations, which suggests that hypothesis 3 can explain the observed mismatch in ER.

One of the limitations of the current data is that it only concerns part of the envelope gene (the C2V5 region sequenced by [Shankarappa et al., 1999](#)). A recent study on HCV showed for instance that what happens in *env* can differ from the rest of the genome ([Gray et al., 2011](#)). In their study, the mismatch between WHER and BHER was only detected in *env* (although one of the strong limitations of their study is that for the WHER, [Gray et al.](#) pooled sequences from several patients who were all infected by the same source after blood transfusion).

We gathered publicly available sequences from the Los Alamos HIV database (<http://www.hiv.lanl.gov/>). Our selection criterion was that we wanted full genome sequences for pa-

tients infected by HIV-1 subtype B and not treated. For the BH dataset, we found several sequences that fitted out criteria. For the WH datasets however, the search was less successful. In the end, we found one chronic infection with a very dense sampling (Liu et al., 2006) and 5 dataset with a lighter sampling in the acute phase of the infection (Herbeck et al., 2011).

As shown in Figure 2.13, we found that the mismatch between WHER and BHER is present throughout the whole genome (Alizon and Fraser, 2013). This mismatch seems to be more pronounced in the *env* region. We found a slight difference in ER between overlapping and non-overlapping genome regions but not as strong as we expected. This could be due to the limited number of WH datasets.

That the difference in ER is not only restricted to *env* but also affects regions evolving less rapidly bring additional support to the hypothesis Lythgoe and Fraser (2012) called ‘store and retrieve’, according to which some viruses are ‘trapped’ in quiescent immune cells and released later in the host seems to be the most parsimonious. This sees more parsimonious than the ‘adapt and revert’ hypothesis, which would require backward evolution throughout the HIV genome. Why viruses that have less adapted to the host would be more prone to being transmitted or to initiate an infection still remains to be explained.



## 2.6 Phylogenies in evolutionary epidemiology

*The presence of phylogenies in articles on host-parasite interactions has increased tremendously. However, in many cases, these phylogenies are very descriptive. I have worked on combining phylogenies with infection trait data in order to better understand links between pathogen genotype and infection phenotype. I applied this to set-point viral load in HIV infections and to the infection outcome in HCV infections.*

Since it has become unthinkable of publishing sequences in the main text of an article, phylogenies have gained in popularity, especially in the medical literature. We now frequently see a tree inferred from pathogen sequences collected from different patients or from the same patient. Leaving aside the robustness of these phylogenies, it is striking to realise that these phylogenies tend to be added for illustrative purposes. At most, there is a graphical interpretation.

Yet, phylogenetic methods have tremendously progressed over the last decade, allowing to address many questions using phylogenies. For instance, as reviewed by [Lemey et al. \(2009\)](#), it is now possible to infer the geographical expansion of a disease on a continuous spatial scale (and even plot it in Google Earth). Phylogenies can also be used to make detailed inferences on disease epidemiology: by using HIV sequences originating from 5700 patients, [Kouyos et al. \(2010\)](#) were able to show that two HIV epidemics coexist in Switzerland: one that circulates in Men having Sex with Men (MSM) and another that circulates in Heterosexuals (HET) and Injecting Drug Users (IDU). They also showed how the magnitude of each epidemics has changed over time, which they argue is linked to public health policies, such as the handling of clean syringes to drug users<sup>9</sup>.

I focused on linking phylogenies to infection traits. This started during my fellowship with Sebastian Bonhoeffer. I was working on the evolution of HIV virulence and asked him ‘*Do we know if this virulence is even heritable from one infection to the next?*’ This question was not really relevant for many parasites because either they evolve slowly enough for the answer to be patent or because they infect animals or plants and experiments can be performed. If one had asked a clinician back then whether the wide variations in survival times we see amongst patients are due to host or viral factors<sup>10</sup>, he/she would have told you that 99.9 % should come from the host. Most clinicians will probably give you the same answer today but in the last few years a lot of evidence has accumulated showing that viral genetic factors matter.

### 2.6.1 Linking phylogenies and infections traits

Some untreated hosts infected by HIV die within a year whereas others survive for more than 25 years ([Buchbinder et al., 1994](#)). There is ample evidence that host environment and genetics matter and genome wide association studies (GWAS) can explain up to approximately 20 % of this variance with a few single nucleotide polymorphisms (SNPs) located in the MHC region of the genome ([Fellay et al., 2007](#)). However, working on the virus side of this question is difficult because, contrary to non-human diseases, we usually do not know the transmission network,

9. En passant, they show a strong spatial structure with little contact between the HIV epidemics spreading on both side of the rösti barrier (i.e. the French-speaking and the German-speaking cantons).

10. I left aside environmental factors because virus genetic effects are puzzling enough for clinicians.

i.e. ‘who infected whom’. Furthermore, viruses evolve rapidly, which makes it more difficult to link variations in infection traits to a given genetic polymorphism. This problem arises not only for the virulence of HIV but also for many viral infection traits.

Sebastian Bonhoeffer’s idea was to circumvent the lack of data on the transmission network by using phylogenies of virus infections. For HIV infections, the virus is routinely sequenced to avoid drug resistance. Rapidly mutating viruses accumulate mutations from one infection to the next, which means that there can be a link between a phylogeny built using virus sequences isolated from infected hosts (which we will refer to as the ‘virus phylogeny’) and the transmission network (Leitner et al., 1996, Hué et al., 2004). In other words, viruses with more similar genomes are likely to originate from hosts close in the transmission network. There is a lot of noise in these virus phylogenies, which arguably explains why these methods have not been pushed forward in the late 90s. However, even though virus phylogenies cannot be used to prove a transmission network with certainty (e.g. in court trials), they can still be used for other types of analyses involving infection traits. In this case, the uncertainty in the phylogeny translates into an error margin in the calculations. This idea has been applied by some studies to test for correlations between phylogenetic distance and virus origin, thus following the ‘phylogeography’ framework (Parker et al., 2008).

We showed that information can be gained by combining virus phylogenies with infection life history traits (Alizon et al., 2010). For this, we applied a classical method in evolutionary biology, the phylogenetic comparative method (Felsenstein, 2004), to estimate the ‘heritability’ from one infection to the next of an infection trait. This heritability, which can be expressed in a fashion similar to the population genetics notion (Housworth et al., 2004, Alizon et al., 2010), is the fraction of the variance in the trait that is governed by the virus genotype (Lynch and Walsh, 1998). In brief, the phylogenetic comparative method (PCM) is a statistical method to estimate the correlation between the proximity in the virus phylogeny and the similarity in trait values. If a trait is heritable, infected hosts close in the phylogeny should have similar trait values. We applied the PCM to data from the Swiss HIV Cohort Study (SHCS). The infection trait we focused on was the set-point viral load (spVL) because it is a predictor of the time to AIDS in an untreated host, i.e. the virulence (Mellors et al., 1996, Fraser et al., 2007, Mellors et al., 2007). We restricted our analysis to untreated individuals infected by HIV-1 subtype B. Our results show that up to 59 % ( $\pm 7$  %) of the variance in spVL is heritable, i.e. can be explained by the virus genotype<sup>11</sup>. Importantly, we also show that this result is robust (by introducing noise in the phylogeny). The same result was reached using two difference methods (Freckleton et al., 2002, Blomberg et al., 2003).

These result came as a surprise to many from the HIV field because it is strongly believed that variations in virulence are almost exclusively controlled by the host<sup>12</sup>. However, this result was strengthened by the publication the same year of three other studies based on known transmission

11. This high heritability value is only obtained when we reduce the data to men having sex with men and while using a strict criterion to define set-point virus load (that is, we remove patients where the variance in virus load is too high). In other datasets, the values are lower (approximately 10 %).

12. When we present these results, MD often attack us by saying in substance ‘We see patients who do control HIV well and we showed they have key mutations in their HLA, *ergo* what you show on the virus effect cannot be true. As discussed earlier in the case of superinfection and trade-off hypothesis, they seem to have difficulties to envisage two simultaneous processes.

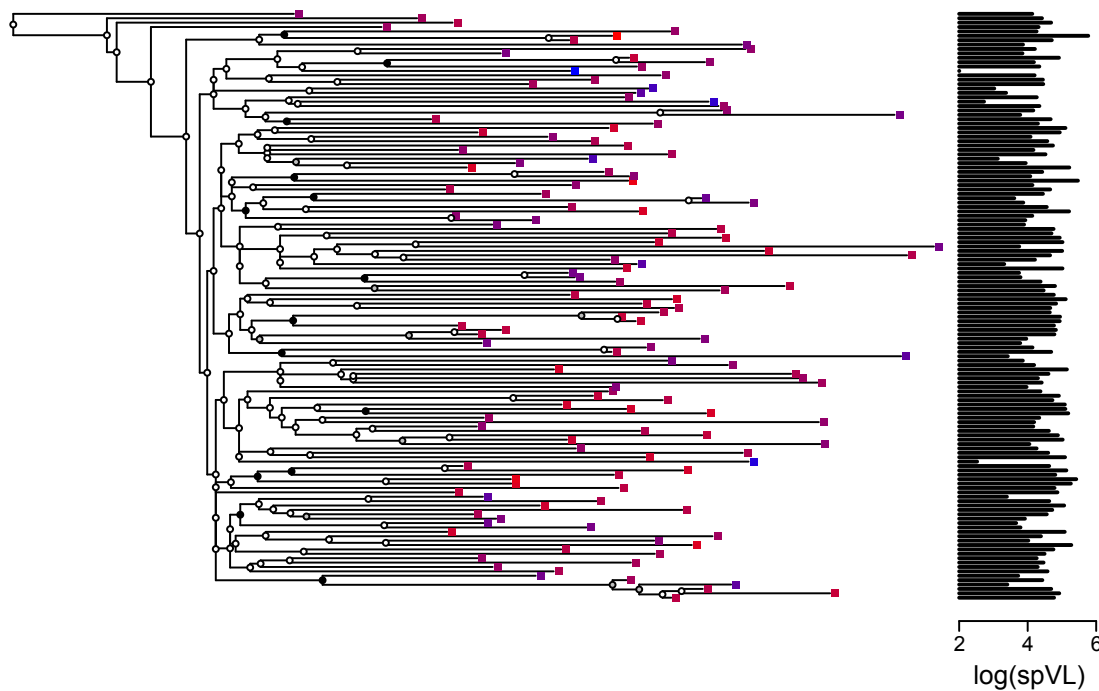


Figure 2.14: **Combining phylogenies and trait values.** A phylogeny based on HIV sequences from the SHCS obtained from patients with known set-point viral loads (spVL). The colour and the graph on the right hand side both indicate the spVL: higher values are in red and lower values are in blue. There are clusters of infected hosts with similar trait value on the phylogeny. Figure from [Alizon et al. \(2010\)](#).

pairs, which have reached similar conclusions ([Hollingsworth et al., 2010](#), [Müller et al., 2011](#)).

Since then, I have been collaborating with George Shirreff, a PhD student of Christophe Fraser, on analysis the accuracy of various phylogeny-based inference methods to estimate heritability. His main result is that most of these methods are only able to detect signal above approx. 40 %. This result has also generated many projects in the field. For instance, Philippe Lemey (University of Louvain) developed a method to estimate phylogenetic signal for spVL, while inferring the phylogeny in the package BEAST. Also, Andrew Leight-Brown and his group are applying pedigree-based methods to estimate heritability from UK datasets<sup>13</sup>.

Together with Christophe Fraser, Katrina Lythgoe and Sebastian Bonhoeffer, we are also working on an Opinion piece to further establish the importance of virus control over virulence in the HIV field.

13. This is quite ironical because I remember Andy being particularly doubtful (to put it mildly) about the relevance of the results when I presented them in Edinburgh in 2009.

### 2.6.2 The case of HCV

Hepatitis C virus (HCV) is estimated to infect approximately 170 million people worldwide and is a major cause of chronic liver disease (Simmonds, 2004). If untreated, acute HCV infections result in one of two infection outcomes. Either the virus is eliminated within six months (a ‘cleared’ infection), or it develops into a ‘chronic’ infection that persists for life unless cured by antiviral therapy (Grebely et al., 2012). Recently, genetic polymorphisms in the promoter region of the host *IL28B* gene that correlate with an increased probability of HCV clearance have been detected (Ge et al., 2009, Rauch et al., 2010). However, a contribution of the virus genome to the infection outcome, although sometimes suspected, remains understudied. Our idea was to adapt the methodology developed to study a quantitative trait in HIV infections (set-point viral load) to a categorical trait (infection outcome) in HCV infections. Again, the virus genotype control value we obtain is analogous to heritability in quantitative genetics (Visscher et al., 2008) but applied to categorical traits.

Our method functions by simulating a set of phylogeny tips (each of which corresponds to an infection outcome) by assuming different virus effects on trait outcomes. We then compare the tip distribution we observe with the real data to the simulated one to determine if there is a virus effect on infection outcome. Note that contrary to the HIV study, we also accounted for confounding effects such as host genetics. This work was carried on by Matthew Hartfield during his postdoctoral fellowship using data collected by members of the team in which my collaborator Fabio Luciani worked as a research fellow. Note that this HITS dataset we used (the Hepatitis C Incidence and Transmission in prisons Study) monitors HCV infections in Australian prisons, where the prevalence is extremely high (approx. 30 %), which explained the high proportion of clearing infections detected.

We analysed genotype 1 and genotype 3 data separately because viruses from each genotype were very distant. Our first analysis did not reveal any significant signal for clustering of infection outcome on phylogeny tips. Knowing that a handful of host SNPs can have a drastic effect on virus clearance, we repeated the analysis but this time replacing infection outcome by host SNP status at two loci of the *IL28B* gene. The implicit assumption we made in our first analysis is that there is no host structure on the virus phylogeny, therefore, if this assumption is correct there shouldn’t be a clustering of host SNP status. This was not the case: for Genotype 1, we found significant correlation estimates of around 0.62 for SNP *IL28B*-917. This suggested that our initial analysis was likely to have been affected by this host clustering.

We re-ran our analysis of the clustering of infection outcome on the virus phylogeny but this time we accounted for the host SNP status (it was added as a co-factor in the analysis run in *BayesTraits*, available from [www.evolution.rdg.ac.uk](http://www.evolution.rdg.ac.uk)). Significant virus control values of 0.653 and 0.569 (the two values correspond to the switching from and to the other infection outcome) were obtained if infection outcome covaried with the *IL28B*-917 SNP. Note that we also detected signal in randomised phylogenies, which means that the estimate we get is unlikely to be the real heritability. If we use the signal obtained in the randomised phylogenies (which we expect to be nil) to rescale these measures, we find that the virus genetics account for 35 and 29 % of the variance in infection outcome.

For Genotype 3, although re-calculated estimates of the virus control over the infection outcome were very high (for example, one estimate gives 0.690), none of these results were signifi-

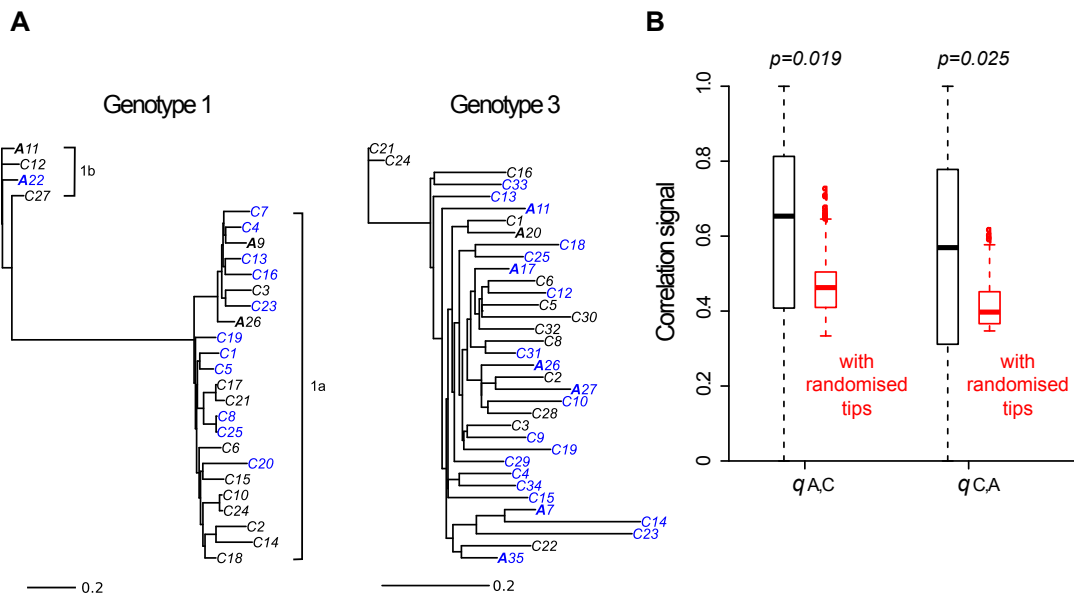


Figure 2.15: **Phylogenetic signal for infection outcome.** A) Phylogenies for Genotype 1 and Genotype 3 sequences. Acute infections are in bold font (and denoted with an ‘A’). Host with a clearing SNP in the *IL28* gene are in blue. B) Estimating phylogenetic signal for Genotype 1 sequences after correcting for host SNP status. In black are the trees inferred from the real data and in red the randomised trees. Note that there is non-zero signal in the randomised trees.

cant. One possible explanation for this is that evolutionary rates seem higher for this genotypes. This alters the structure of the phylogeny (making it more star-like), which is known to decrease the ability to detect phylogenetic signal (Blomberg et al., 2003). Also, there was a higher coalescent effective population size for Genotype 3 than Genotype 1, which could be explained by the fact that Genotype 3 infections responded less well to treatments.

In summary, by creating a method to simulate inheritance of a binary trait along a set of phylogenetic trees, we have found HCV genetics controls 29 to 35 % of the variance in the outcome of acute HCV infection. These correlation estimates apply to HCV genotypes 1a and 1b, and were only found once the host’s *IL28B* status at the 917 SNP was taken into account as a co-factor. Overall, this finding not only suggests that the virus genotype can affect infection outcome, but this is also dependent on the specific subtype of HCV, and host genetics. These findings should motivate further research into host-parasite interactions that affect virus evolution.

As for the results on HIV, publishing this result is not easy. Part of the problem comes from the lack of clearers but I bet that this would not have been a problem if we had been working on the host side. There seems to be a general reluctance in the medical community to admit that genetic diversity in organisms as small as virus matter: differences amongst HIV subtypes or HCV genotypes are acceptable, but variance within subtype/genotype is still taboo.

## 2.7 Epidemiology and evolution of disease emergence

*The initial stages of an epidemics are very stochastic because of the low number of infected hosts. With Matthew Hartfield, we investigated two questions that seemed to have been overlooked (one of them for a surprisingly long time). The first question we asked was how to define an outbreak. Even though this question can seem intuitive, it is seldom discussed. The second project was to study disease emergence in a context where host availability can become limiting.*

Emerging and re-emerging viruses are getting more and more attention, especially since the avian flu outbreak in 1997 in Honk-Kong and the SARS outbreak in south-east Asia in 2003. One of the key pieces of the virus emergence puzzle is stochasticity. Indeed, emerging viruses being rare initially by definition, their spread in the host population does not follow the deterministic dynamics that are typically used in epidemiology models. This is illustrated by an Ebola virus outbreak in 2007 in the Democratic Republic of Congo. As shown by [Leroy et al. \(2009\)](#), during the first three ‘generations’ (in terms of infection time), there was only one infected host. It is only the third infected host, who infected more than one host and led to the outbreak that caused 186 deaths. Had any of these three hosts cleared the disease or died before transmitting it to someone else, there would have been no outbreak. In the following, when referring to ‘virus emergence’ or to the ‘probability of emergence’, we mean the probability that a virus manages to infect a sufficient number of hosts in the population to escape these stochastic dynamics.

Progresses have been made over the last decade in understanding the conditions that favour the emergence of new pathogens. For instance, we know that diversity in the host population, which from the virus’ point of view means that not all hosts are equally good ‘resources’, acts against emergence ([Lloyd-Smith et al., 2005](#)). We also know that disease life-history matters and that viruses causing longer infections are more likely to emerge ([André and Day, 2005](#)). Importantly, evolutionary biology can play a decisive role in disease emergence because an averagely-adapted virus strain that just appeared in a population can evolve and produce a well-adapted virus strain within few mutational steps. The epidemiology of this rare virus strain and its evolution then interact to shape its probability of emergence ([Antia et al., 2003](#), [Yates et al., 2006](#)). Contact networks between hosts, i.e. the fact that hosts only interact with a subset of the host population, also affects emergence ([Alexander and Day, 2010](#)).

One of the limitations common to almost all existing modelling studies on disease emergence is that they implicitly assume that susceptible hosts are so abundant that their availability is never a limiting factor for the parasite. This simplifying assumption is valid if we consider a rare parasite in a population and want to calculate its probability of emergence. However, in some situations it is more problematic.

With Matthew Hartfield, we wanted to study the emergence of a parasite through mutation of a resident strain itself spreading in a population following an SIR (for ‘Susceptible-Infected-Recovered’) epidemiological model. This raised an unexpected question, which is discussed first: ‘How many infected hosts does it take to have an outbreak?’

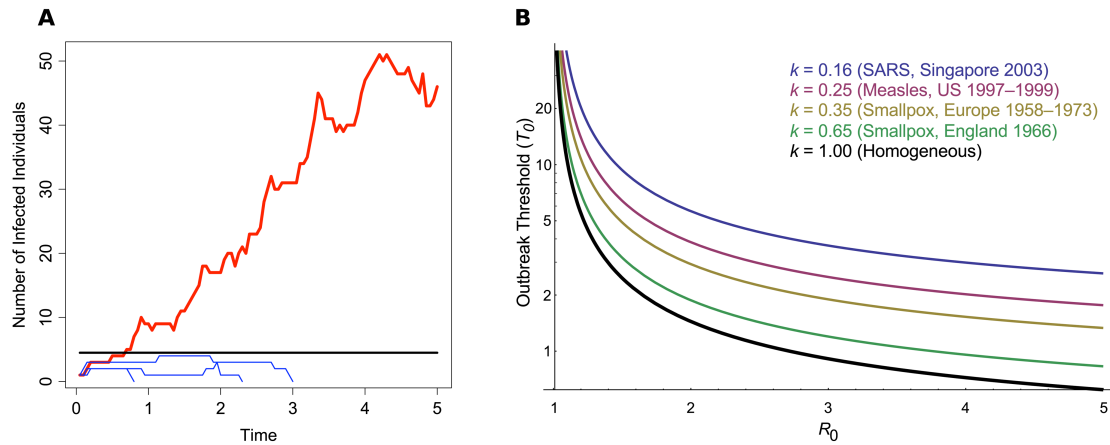


Figure 2.16: **Illustrating the outbreak threshold.** A) Four stochastic epidemics: three do not reach the outbreak threshold (black horizontal line) and go extinct (in blue), whereas one emerges (in red). B) Outbreak threshold as a function of  $R_0$  and of  $k$ . The more heterogeneous the host population, i.e. the lower  $k$ , the higher the threshold. Figure from [Hartfield and Alizon \(2013\)](#).

### 2.7.1 Introducing the outbreak threshold

When there is no limitation of susceptible hosts, the probability of emergence is easy to calculate, it is the probability that the parasite does not go extinct. In an SIR model, this definition fails because, eventually, the parasite will go extinct (at some point there are not enough susceptible hosts left to ensure parasite spread). How then should we define an outbreak? Intuitively, it seems that there should be some sort of threshold because, for instance, one case of flu does not seem enough to qualify as an outbreak.

Typically, authors tend to dodge this question. Most of the time, the outbreak threshold is set to an extreme value. Interestingly, this value is very low in studies that focus on analysing public health data (to make sure an emergence even is not missed) and very high in modelling studies (because running the epidemics many time is not a problem with a computer). We also even found some studies that determine the threshold “*de visu*”<sup>14</sup>. The WHO has more precise definitions in the case of meningitis. If the population is less than 30,000 individuals, 5 cases in a week or a doubling in the number of cases over 3 weeks means the outbreak threshold is reached. If the population is greater than 30,000 individuals, the threshold is set to 0.015 % of the population (or 0.01 % if the risk is known to be greater).

Our suggestion, to formalise the outbreak threshold (which we denote  $T_0$ ), was to define it as the number of infected hosts required for the parasite to escape stochastic dynamics ([Hartfield and Alizon, 2013](#)). Of course, mathematically speaking, the extinction risk due to stochastic dynamics never reached 0 unless the population size is infinite. Therefore, the outbreak threshold

14. A nice informal definition I got later on from Sebastian Bonhoeffer was that one way to define the outbreak threshold is when it's too late.

will always be a function of a cutoff (or error risk). Using branching process models, we show that in the order of  $1/\log(R_0)$  infected individuals are required to escape stochastic dynamics. More precisely, the exact number of infected hosts required so that the probability of extinction of the pathogen is lower than  $c$  (i.e. its probability of emergence is greater than  $1 - c$ ) is

$$T_0 = \frac{-\log(c)}{\log(R_0)} \quad (2.12)$$

Let us consider the meningitis case mentioned above. The  $R_0$  of the disease is approximately 1.36. The WHO sets the threshold to 5 infected individuals. This means the risk of emergence is still approximately equal to 22 %.

In reality though, the threshold could be higher. This is because in order to derive our equation 2.12, we assumed that the host population was completely homogeneous. As shown for instance by Lloyd-Smith et al. (2005), increased host heterogeneity in the number of secondary infections they cause have a strong effect on disease emergence. This can be captured by using a negative binomial distribution to approximate the distribution of the individual  $R_0$  in the population. This distribution has a mean  $R_0$  (the value that is used in deterministic models) and a dispersion parameter  $k$ . If  $k = 1$ , we have a homogenous outbreak (as assumed for eq. 2.12). The more  $k$  drops below 1, the more the majority of infected hosts become ‘dead-end’ (i.e. do not transmit the disease) and a minority become ‘super-spreaders’. Matthew showed that the expression of the outbreak threshold in this case can be approximated by:

$$T_0 = \frac{-\log(c)}{\log(R_0)} \left( 0.334 + \frac{0.689}{k} + \frac{0.408}{R_0} - \frac{0.507}{kR_0} - \frac{0.356}{R_0^2} + \frac{0.467}{kR_0^2} \right) \quad (2.13)$$

Note that if  $k = 1$ , the term in the parenthesis should be equal to 1 for consistency and this occurs if the polynomial  $0.023R_0^2 - 0.099R_0 + 0.111$  is zero. This never occurs for real values of  $R_0$  but the value is lower than 0.05 if  $R_0 \leq 3$ , which suggests that our approximation works best for values of  $R_0$  that are not too high.

As shown in Figure 2.16, the lower  $k$ , the higher the threshold. This makes sense since lower values of  $k$  correspond to more heterogeneous populations and classical epidemiology results show that this heterogeneity acts against emergence (Diekmann and Heesterbeek, 2000, Lloyd-Smith et al., 2005).

In conclusion, let us first stress that similar analyses could be performed for the critical community size (CSS), which is classically defined as the minimal number of hosts required to allow disease persistence in an endemic phase (Anderson and May, 1991). The concept has never really been formalised and, for instance, the fact that a cut-off value ( $c$ ) is required to define this threshold is usually not mentioned. Second, the threshold cannot replace public health policies. A 20 % risk of meningitis outbreak is not the same as 20 % risk of an ebola outbreak and in the end the virulence of the disease will usually have the final word.

### 2.7.2 How does host availability affect pathogen emergence?

In a deterministic model, the spread of a pathogen strain in a host population is governed solely by the number of secondary infection it generates, whence the threshold of 1 for the basic



reproductive ratio  $R_0$ , as shown in equation 2.3 (Anderson and May, 1991). In stochastic models, it is as if each infected host had its own individual  $R_0$  value, which is drawn from a distribution. For instance, if the duration of infection is exponentially distributed and contacts follow a Poisson process, then the probability of emergence equals  $1 - 1/R_0$  (Diekmann and Heesterbeek, 2000). Additionally, the role of pathogen evolution and migration in causing new outbreaks in novel populations has been recognised and widely studied. Previous models highlighted the important fact that even benign strains can be dangerous, as even though they are due to go extinct, they may persist in large populations long enough to give rise to well-adapted mutants that can subsequently cause an outbreak (Antia et al., 2003, Iwasa et al., 2003, 2004). The effect of pathogen life-history (André and Day, 2005) and population structure (Alexander and Day, 2010, Kubiak et al., 2010) have also been modelled, as well as the effect of host heterogeneity and assortative mating (Yates et al., 2006).

Importantly, previous models have focused on the case where the pandemic (well-adapted) strain, with an  $R_0 > 1$ , emerges through mutation from a maladapted strain, with an  $R_0 < 1$ , which is bound to go extinct rapidly. However, this is not the only scenario that can lead to a pandemic, nor arguably the likeliest. An alternate scenario, which has received less attention, is where an intermediately-adapted strain would first create an initial outbreak, which would be limited in spread as it would have a low  $R_0$  close to one. A mutation could then arise in one individual that would subsequently increase the  $R_0$  of the pathogen to a much higher level, creating a more potent outbreak. This type of emergence, from a limited to a large outbreak, plays an important role in the transferring of zoonotic pathogens into complete human agents (Wolfe et al., 2007). This scenario is also an example of ‘evolutionary rescue’ (Gonzalez et al., 2013), where evolution is needed to prevent quick extinction of the weaker strain.

Although this scenario is more likely to occur for many micro-parasitic infections (because the initial and the mutant strain are likely to be antigenically close), it is also more difficult to capture because it leads to the violation of one of the important simplifying assumptions made by earlier models. When a pathogen mutates from  $R_0 < 1$  to  $R_0 \gg 1$ , it is safe to assume that the number of hosts infected by the initial (maladapted) strain is negligible. However, if  $R_0 > 1$  initially, then the susceptible population would, in general, reduce over time as individuals become immune or die out. This reduces the emergence probability of the second strain, as there are less susceptible individuals available to transmit the pathogen to. Furthermore, the initial strain might continue to spread for a while, hence further depleting the pool of susceptible hosts available to the mutant strain. This process is likely to arise due to infection by the first strain rendering the host immune to the mutant strain (cross-immunity), which is the basis for vaccination, or the pathogen killing infected hosts, which is also not accounted for in previous models.

Therefore, a more complete model is needed to account for cases where an initial strain with  $R_0$  close to one mutates into a stronger strain, as observed in recent outbreaks. For instance, mutations arising in the vector-borne chikungunya virus around the Indian Ocean caused a new infection wave to arise, with a higher dissemination rate, from an initially smaller one where viruses had a lower dissemination rate (Schuffenecker et al., 2006). The first strain, which created a very minor outbreak in May-June 2005 at La Réunion (Renault et al., 2007), seemed to have an  $R_0$  around one, because whilst it did not greatly increase in frequency over time, it was maintained in the population at a low frequency. The second strain, which appears to consistently differ from the first by two point mutations (Schuffenecker et al., 2006), created a

major outbreak in La Réunion over January and February 2006 (Renault et al., 2007). Further analyses reveals that this mutated strain was particularly well adapted to exploit the mosquito vector *Aedes albopictus* (Vazeille et al., 2007), which most likely explains the increase in its  $R_0$ . Additionally, these virus mutations seem to have occurred independently in several distant population (de Lamballerie et al., 2008), which underlines the importance of virus evolution triggering further outbreaks. It is known that recovering from a chikungunya infection generates a strong immunity, which seems to target both the initial and the mutated strain (Renault et al., 2007).

Here, we outline an analytical model to investigate this scenario. We derive the probability that, given an initial weakly-spreading pathogen ( $R_0$  close to, but exceeding 1), it can mutate into a second pathogen with a much greater rate of spread ( $R_0$ ). We form this probability via a two-step model. In the first stage, the initial pathogen spreads deterministically, but can mutate into a new strain at any time before it goes extinct. The second stage calculates the probability that this second strain emerges, instead of going extinct by drift. We find that the ongoing depletion of the susceptible population by the initial strain drastically reduces the emergence probability, compared to classical branching process results that assume that the susceptible population remains fixed as the new strain emerges (Allen, 2008). We produce analytical solutions for the probability of emergence in this scenario, which we show to be accurate when compared to stochastic simulations. Specifically, we derive an equation for the emergence probability for the second strain, whilst accounting for the continual depletion of the susceptible population, and show how this mechanism drastically decreases the emergence probability. Finally, we apply our model to the chikungunya virus outbreak in La Réunion, and show the severe reduction in emergence probability due to susceptible depletion, which could counteract the adaptive ability of the initial strain and limit the possibility for future outbreaks to arise.

About the methods, we cannot explicitly solve an SIR model as a function of time, even if the host population size is constant. However, we can find a solution for the dynamics of the density of infected hosts ( $I$ ) as a function of the density of susceptible hosts ( $S$ ), such that:

$$I_1(S) = (N - S) - \frac{N}{R_0} \log\left(\frac{N}{S}\right) + I_0 \quad (2.14)$$

where  $R_0 = \beta/\gamma$  and  $N$  is the total population size and  $I_0$  is the initial number of infected.

For a specific number of susceptible hosts  $S$ , a current infection can mutate with probability  $\mu$  into a new strain; assuming that  $\mu \ll 1$ , so that it is unlikely for more than one new mutant to appear per generation, then the probability that a new strain appears by mutation is  $\mu I_1(S)$ . The new strain can then fully emerge and overcome stochastic loss with probability  $\Pi(S)$ . In order for this strain to never emerge at all, then at each time step a mutant strain never emerges; the probability of this at a single time point is  $1 - \mu I_1(S) \Pi(S)$ . One minus this probability, multiplied for all time points, is the probability  $P_{\text{emer}}$  that a mutant strain ultimately emerges. That is:

$$P_{\text{emer}} = 1 - \prod_{S=S_{\min}}^{S_0} (1 - \mu I_1(S) \Pi(S)) \quad (2.15)$$

where  $S_{\min}$  is the minimum susceptible population needed to give a non-zero emergence probability of the mutant strain.

About the emergence probability of a mutant strain itself (given that the susceptible population is being depleted by a pre-existing strain), we show using some simplifying assumptions that it can be approximated by:

$$\Pi(S) = \frac{\Phi_0}{I_1(S)R_0 + \Phi_0} \left( 1 - \frac{N}{S} \left( 1 - \frac{\Phi_0 - 1}{\Phi_0} e^{-\frac{\Phi_0 - 1}{\Phi_0}} \right) \right) \quad (2.16)$$

where  $\Phi_0$  is the reproductive ratio of the mutant strain and  $R_0$  that of the first strain. The important thing to note is that not only is part of this solution proportional to  $N/S$  (reflecting the growth rate  $R_0$  decreasing by a fraction  $S/N$  as susceptible individuals are removed), but it is also proportional to  $1/(I_1R_0 + \Phi_0)$ . This highlights the fact that even as the second strain appears by mutation and spreads, the emergence probability is decreasing during this initial phase, as the susceptible population is reduced due to the continuing spread of the first strain. This formulation assumes complete cross-immunity (or host death) is present between the original and mutated strains.

Using numerical simulations, we show that these analytical approximations appear to be accurate as long as the population size is not too small ( $N \leq 1000$ ), and the mutation rate is not extremely high. Considering that most outbreaks arise over large areas, where  $N$  is large, and point mutation estimates (for beneficial, neutral and deleterious mutations) for viruses and bacteriophages seldom exceed  $1 \times 10^{-4}$  (Sniegowski et al., 2000), then these assumptions should not be greatly violated when considering real-world cases.

We also applied our model to the two-wave outbreak of chikungunya virus on La Réunion. Over 2005, the virus was present at a low frequency in the population of La Réunion Island, peaking at 500 reported cases and stabilised to 100 until the end of September. Afterwards, a secondary wave emerged in December 2005, causing nearly a 100-fold increase in the number of cases, peaking at more than 45,000 (Renault et al., 2007). More than 90 % of viruses in the secondary outbreak harboured two novel substitutions in the structural region (Schuffenecker et al., 2006). Viruses harbouring these substitutions also replicated faster, compared to ancestral variants (Vazeille et al., 2007). Therefore, the secondary outbreak was almost surely caused by adaptive mutation increasing the replication rate of the pathogen.

The population size of La Réunion is around 776,000 individuals. We assume that the susceptible population size before infection would cover the whole island. This seems a reasonable assumption since infection did not seem to be located at any one location on La Réunion (Renault et al., 2007). Furthermore, previous analysis on other vector-borne diseases suggest that they tend to exhibit little heterogeneity in host transmission (Lloyd-Smith et al., 2005), so we can assume a well-mixed population. To be conservative, we investigate  $\mu$  varying between  $10^{-5}$  and  $10^{-7}$ . Finally, we estimated the reproductive ratios by comparing epidemic data from Renault et al. (2007) with those produced from analytical results. It appears that a good fit for  $R_0$  is 1.04, whilst that for  $\Phi_0$  is much higher at  $\sim 1.4$  (Supplementary Material S3). We therefore fix  $R_0$  at 1.04 and vary  $\Phi_0$  between 1.05 and 1.5 to determine to what extent  $\Phi_0$  affects emergence probability.

Figure 2.17A shows the probability of emergence according to our model, denoted  $P_S$ , scaled to that expected for a single infected individual in a fully susceptible homogeneous population (that is,  $1 - 1/\Phi_0$ ). We see that, unless the mutation rate is very low, the scaled probability of emergence generally lies between 0.2–0.4. Therefore, the need to mutate from a pre-existing

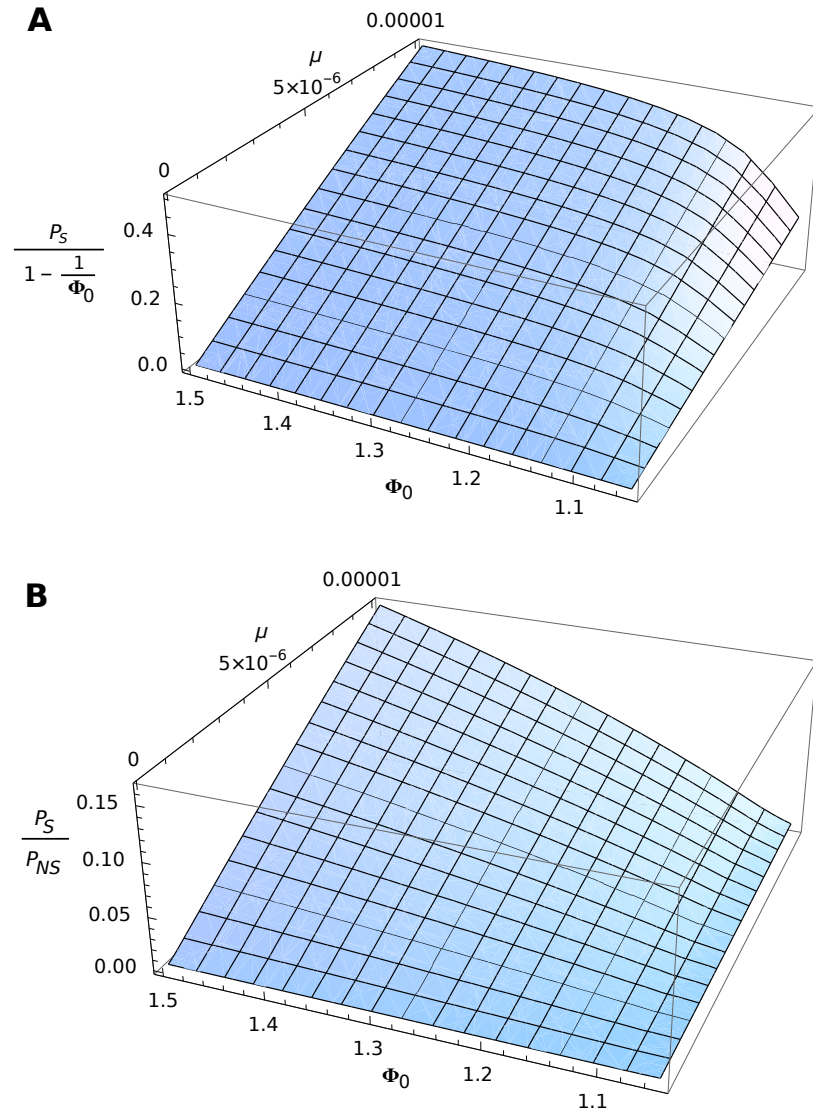


Figure 2.17: **3D plots of emergence probability of a mutated strain of the chikungunya virus in La Réunion.** A) Results predicted by our model ( $P_S$ ), as scaled to the emergence probability for a de-novo strain. B) Results scaling our model to one that assumes that emergence of new pathogens are not affected by limitation of the susceptible population. Values are given as a function of the reproductive ratio of the mutated strain  $\Phi_0$ , and the advantageous mutation rate  $\mu$ .

strain only reduces the probability of emergence by two to five fold, compared to the emergence of a single strain in a fully susceptible population. This is due to the fact that, even though the mutation rate is low, the susceptible population in La Réunion is assumed to be very large, creating multiple opportunities for the first strain to mutate into a faster-spreading strain.

We can also determine how saturation of the susceptible population affected the emergence probability. Figure 2.17B compares  $P_S$  to a model that assumes that susceptible depletion does not affect the emergence probability of new strains ( $P_{NS}$ ). We clearly see that the ongoing depletion of the susceptible population greatly reduced the emergence probability of mutated strains in this epidemic, with a 10–20 fold reduction in probability observed. Similar values are obtained if the pathogen emergence probability is set to  $1 - N/(S\Phi_0)$  in  $P_{NS}$ ; that is, susceptible dilution affects pathogen emergence when it first appears, but not in subsequent generations (result not shown). This result gives a clear indication as to how susceptible depletion strongly affects the emergence of new, mutated strains from pre-existing ones. It could also explain why observing mutated strains arising in the field is uncommon, as the change in the underlying population size would impact their appearance.

In summary, we have produced an accurate model for the probability that a second, stronger outbreak can be caused from an initial weaker one that has an  $R_0$  close to 1. This study highlights how to implement information about changing population sizes (specifically the susceptible population size) into analytical models to determine the probability of a new strain emerging, without relying solely on numerical simulations.

## 2.8 Perspectives

Since my PhD I have moved slightly towards data analysis (although I think I remain at a safe distance from hardcore analyses). There are several general directions to go. I could go back to more conceptual problems. On the contrary I could go towards analysing data, which often makes it easier to get funding. Doing experiments would be completely wild (but why not).

Where the field should go is a different question. A wealth of data is accumulating on host-parasite evolutionary dynamics. As mentioned before, the collection of some (most?) of this data is not really motivated by scientific questions but rather by the fact that we can do it. I would argue that ecology and evolution have long stood aside because researcher in this field were motivated by answering biological question rather than by engineering questions. In fact, molecular biologists are often surprised that relevant answers can be found without first sequencing everything and analysing molecular details (that is of course once the paper is published and after they did all they could as reviewers to reject it).

### 2.8.1 Tolerance, Resistance and Virulence

This will be (very) old news to plant pathologists but host responses to an infection can be categorised as resistance or tolerance (Schafer, 1971). Resistance captures any process that acts to reduce parasite load, thus reducing virulence and transmission. Tolerance means decreasing virulence (leaving parasite transmission rate unaffected). These two responses obviously have different consequences for the pathogen as the former clearly decreases its fitness, while the latter increases it (the infection is equally efficient and lasts longer).

Recently, the study of animal tolerance to parasitic infections has become rather popular (Råberg et al., 2007, 2009, Boots et al., 2009, Ayres and Schneider, 2012). One of the reasons is that disentangling between resistance and tolerance allows to make inferences about disease ecology and evolution without requiring a detailed molecular understanding of the immune response. However, studies have shown that the exact way in which the host resists (Gandon and Michalakis, 2000) or tolerates (Miller et al., 2006) a disease can greatly affect parasite evolution.

#### Tolerance (and resistance) as parasite traits

That infection traits are shared traits between the host and the parasite is an obvious statement to any ecologist (although some people actually manage to publish articles about this<sup>15</sup>). However, although arguably obvious, it is worth applying this to the study of resistance and tolerance because most of the modelling studies consider that the observed variance in these traits comes from host differences (but see Little et al., 2010). An alternative possibility is that these caused by parasite differences. In other words, some strains would be more ‘tolerated’ or ‘resisted to’ than others.

Eileen Butterfield’s MSc project, which we co-supervised with Yannis Michalakis, was to investigate the evolution of such parasite traits. One of her results was that, consistently with the results found by Day et al. (2007) on immunopathology, the way in which tolerance is expressed

15. And let us not get started on the interactions with the environment. . .

matters for the evolution of virulence. She also investigated the consequences of the existence of a trade-off between resistance and tolerance, which is often suspected to occur in studies where resistance and tolerance are considered as host-determined traits (Råberg et al., 2007). More generally, the nested approach she developed can give us more insights into the emergence of the transmission-virulence trade-off. Indeed, most studies that have sought to find the trade-off have assumed that parasite strains only differ in the transmission rate and virulence. If they also differ in their ability to be tolerated or not resisted to, this could affect the results and buffer the trade-off curve.

### Sex differences

An increasing number of studies revolve around the consequences of host heterogeneity on parasite evolution (Regoes et al., 2000, Gandon, 2004, Osnas and Dobson, 2011, Williams, 2012). Incidentally, there are also very few experimental studies on the same topic. One of the reasons for this could be that heterogeneity is described in models in a way that is difficult to link to empirical data. With Susan Cousineau, we applied one of these host heterogeneity models to the more specific case of immune heterogeneity amongst sexes.

Currently, there seems to be a lot of evidence that sexes react differently to infection by the same pathogenic agent. As we show, many of these differences can be argued to be differences in resistance or differences in tolerance. If we just focus on the case of HIV for instance, it has been shown that viral load in men tends to be higher than viral load in women. This viral load is known to be linked to the virulence of the disease in absence of treatment (patients with a lower viral load control the disease better and tend to live longer). Unexpectedly, there HIV virulence is the same in men and in women. This would suggest that men tend to tolerate the infection, while women tend to resist.

Although there are several studies on the consequences of tolerance or resistance on parasite evolution on the one hand, and several studies on the consequences of host heterogeneity on the other hand, we were not aware of any study that combined the two, i.e. looked at whether it makes a difference if heterogeneity is expressed in terms of resistance or in terms of tolerance. We investigated this question by re-deriving a model based on Gandon (2004)'s general framework. This allowed us to disentangle between the effects of increasing heterogeneity or increasing the average level of host tolerance/resistance on virulence evolution.

One of our goals was also to investigate whether heterogeneity amongst sexes can lead to evolutionary branching in the parasite population. This was partly motivated by a contemporary Opinion piece by Duneau and Ebert (2012), who argue that such branching is possible. What I found problematic with their paper was that they showed this hypothesis using a contour plot that looked like the result of a model, except that it was drawn by hand. We showed that obtaining such branching in the parasite population can be very complicated unless there is virtually no transmission of the parasite between the male and female populations. This is consistent with results obtained by Osnas and Dobson (2011). However, further investigation is needed because we assumed a constant host population size and this could affect the possibility for branching (Gandon, 2004).

### 2.8.2 Evolutionary medicine instead of Darwinian medicine

Ever since Darwin's *Origin of species*, there was an elephant in the room, which was man. Darwin himself was very aware of this and cleverly refrained from making claims that could be interpreted in a eugenist way. This is all the more admirable that most of his contemporaries did <sup>16</sup>.

For obvious reasons, the struggle of natural selection theory mostly became a struggle for the acceptance of the origin of humans. It took a while before it became widely acknowledged that humans have evolved from a common ancestor they shared with chimpanzees and that they still evolve. However, we then got faced with a different problem, which can be illustrated by E. O. Wilson's book on sociobiology. His first chapters are insightful and scientifically sound. The problem comes from his last chapter, in which he argues that sociobiology as a science is destined to phagocytose other disciplines such as anthropology and sociology. As stressed by Yudell and DeSalle (2000), it is actually quite striking that the weakest part of his book (that on human sociobiology) is probably what mostly contributed to the success of his ideas. Following Wilson's book, a myriad of studies have emerged that are often referred to as 'human sociobiology' or 'evolutionary psychology'. Contrary to human evolution, which is a field based on solid theories and factual evidences, the vast majority of evolutionary psychology research is not scientific and its proponents thrive on the media coverage that such 'studies' inevitably attract.

It would take a lot of time and energy to dismantle evolutionary psychology (and there are more relevant topics to work on). I will however list a few points to clarify what I mean. First, it is striking that most of these studies rely on natural selection. I write 'striking' because, in theory, similar studies could be conducted using cultural evolution frameworks. Some people like Lewontin et al. (1984) have interpreted this as an ideological bias <sup>17</sup>. This urge to explain human (modern) behaviours through natural selection is almost a guarantee that these studies are wrong. In order for natural selection to act on a trait, three conditions should be fulfilled: i) there should be variation in the trait value in the population (this is the easy bit, even for humans), ii) the trait should have an effect on fitness and iii) the trait should be heritable from parents to offspring. Unless we focus on traits that have a very strong genetic association, genetic heritability is almost indissociable from cultural heritability for humans. Another problem is that humans do not have sex in order to reproduce, which means that any effect on fitness is only likely to be detected if it is massive. In practice, if we are dealing with human behaviour, I doubt that there exist many non-pathological behavioural traits for which it would be possible to show that natural selection is acting on contemporary human populations. Just to be clear, I am not stating that natural selection is not acting on humans. In fact, next generation sequencing techniques reveal that the human genome exhibits many traces of recent natural selection events (Nielsen et al., 2007).

In addition to the natural selection issue, there are also problems with the naivety of these evolutionary psychology theories, which often rely on interactions taking place when humans lived as 'hunter gatherers' during the Pleistocene. First we tend to know extremely little about

---

16. In the preface of the French edition of the *Origin of species* in 1862, Clémence Royer advocates for such applications of Darwin's idea to the human populations (some argue even before Francis Galton).

17. Although unfortunately they tend to largely make the same error by criticising evolutionary psychology because it goes against marxist-leninist theories...



these interactions and second it is not clear at all that behaviours did not evolve before this (Downes, 2009). Evolutionary psychology also tends to use methodologies that are way below the average of other fields (e.g. trying hundreds of factors to explain one behaviour and focusing on the one that turns out significant), which is partly linked with the lack of control variables. Finally, I think it is not a mere coincidence that the outcome of evolutionary psychology studies often follows cultural stereotypes: ‘women are attracted by money and men by breast size’, ‘dancing is important for black people in Jamaica’,... Should they end up with something else than a cliché, reviewers would undoubtedly be more prone to pick up on the methods used.

Why should scientists interested in the evolution of host-parasite interactions be concerned at all by this pseudo-scientific field (besides of course the fact that these ‘studies’ divert some research energy that could be used in a more relevant way)? This is because evolutionary psychology has infiltrated the disease evolution field. As argued by Méthot (2009), Méthot (2011), we should make a distinction between ‘Darwinian medicine’ and ‘evolutionary medicine’. The latter aims at using results from evolutionary biology to understand and fight diseases. Its foundation are old and Bynum (1983) already mentions that medical doctors have learned much from Darwin since the XIX<sup>th</sup> century (cited by Méthot, 2009).

Darwinian medicine, as introduced by Williams and Nesse (1991), takes a different stand by attempting to refound medicine in the light of evolutionary biology (rather than using it to get additional insights). One problem is that by doing so it strongly rests on an adaptationist view, which increases the risk of just-so stories similar to that developed in evolutionary psychology. This problem is of course more obvious for the studies on the human evolution side than on the parasite evolution side. I cannot refrain from citing part of the review by Nesse and Stearns (2008). It is entitled *Evolutionary applications to medicine and public health* and it was not published in a journal without peer-review like *PLoS 1* but in *Evolutionary Applications*:

“Obesity has doubled in the past 40 years in the USA, so that two-thirds of adults are now overweight or obese (Wang and Beydoun 2007). Diabetes and obesity are strongly correlated (Neel et al. 1998a). About 194 million adults worldwide have diabetes, and Type 2 diabetes (late onset) is exploding. [...]

We know what we should do to stay thin. We should eat less and exercise more. So, why don’t we? One answer is that in the past individuals who were thin or who wasted calories in nonproductive exercise tended to have fewer children. Selection favoured those who took advantage of opportunities to eat fat, salt and sugar and who stored some extra calories in good times. Selection has shaped mechanisms that limit weight gain, but they are feeble compared with those that prevent weight loss.”

In this quote, the pseudo-scientific nature of the argument is patent. However, spotting problems is not always that easy. For instance, Byars et al. (2010) analyse a cohort of women to predict future evolutionary changes. They claim to detect natural selection at work on some traits (total cholesterol level, age at first birth) in a North-American population. Since they ignore the exact genetical bases for the trait, they estimate heritability in an essentially flawed way because they cannot separate genotypic and phenotypic variance. Using self-reported variables such as smoking status or level of education is hardly a way to control for social factors. The authors, to some extent, acknowledge this when they write that their methods ‘cannot completely

discriminate between vertical cultural transmission and genetic transmission'. However, they do not mention this (strong) limitation in their conclusion. Decades after the emergence of HIV we are still unable to identify natural selection acting on the CCR5-Δ32 allele that confers resistance to HIV (Sabeti et al., 2005). It seems unlikely that natural selection would be easier to detect on similar time scales when it acts on total cholesterol level.

Again, my message is not that natural selection does not act on contemporary human populations nor that it cannot be detected. Famous examples of sickle-cell anaemia and lactose tolerance clearly show that natural selection has occurred recently (Stearns, 1999). The problem is that detecting it in contemporary human populations is difficult. Genome-wide analyses open new possibilities SabetiEtal2006,NielsenEtal2007,YangEtal2011. In fact, it seems that many of these recent selective events are due to the evolutionary pressure exerted by pathogens (Fumagalli et al., 2011)<sup>18</sup>. However, proving ongoing selection events using these techniques is controversial, partly due to confounding factors introduced by demographic processes, as discussed by Nielsen et al. (2007). Furthermore, recent results show that selective sweeps are unlikely to have occurred during recent human evolution (Hernandez et al., 2011).

Since 1991, approximately 150 papers have been published that refer to evolutionary (or Darwinian) medicine, of which a third are reviews. Yet, evolutionary insights into medicine are not recent and Méthot (2009) even refers to a book by the English medical doctor John Ross on *The Graft Theory of Disease, being an Application of Mr Darwin's Hypothesis of Pangenesis to the Explanation of the Phenomena of the Zymotic Disease* in 1872. This may lead us to question the internal cohesion of 'Darwinian medicine' as a field, but not the relevance of evolutionary thinking for medicine. As many have pointed out, the consequences of evolutionary processes on human health, e.g. antibiotic resistance, are largely under-appreciated in the medical community (Antonovics et al., 2007). Teaching evolutionary biology, instead of Darwinian medicine, to medical doctors is actually the best way to provide them with some of the key tools they need to analyse biological problems of medical significance.

### 2.8.3 More on multiple infections and disease evolution

#### Multiple infections and disease emergence?

Co-infections are likely to affect the probability of disease emergence for two reasons. First because infection by a first pathogen often (but not always) facilitates infection by other pathogens. Second, because infections generate heterogeneity in the host population and host heterogeneity has been shown to act against disease emergence.

One article does mention the effect of HIV infections on disease emergence (Lloyd-Smith et al., 2008), however it only considers the first effect (HIV weakens host defences) and ignores the second (HIV increases host heterogeneity). Furthermore, the authors essentially have no epidemiology in their model: they calculate the  $R_0$  of the disease by averaging that of single infections and that of co-infections instead of inferring it from an epidemiological model (van Baalen and Sabelis, 1995, Choisy and de Roode, 2010). Making the epidemiological model more explicit would have another advantage (besides model correctness) in that it would force

18. Which is consistent with an earlier correlation result from Prugnolle et al. (2005), which showed higher HLA class I genetic diversity in regions of the globe where pathogen diversity is high.

us to consider the various effects co-infections can have on epidemiological parameters. For instance, the ‘overall virulence’ expressed by co-infected hosts is usually assumed to be higher than the virulence of the co-infecting strains (hosts cumulate deleterious effects of each disease). This is true in some cases, e.g. co-infections between HIV and HCV, but it is not always the case and, for instance, infection by the apparently nonpathogenic flavivirus GB virus C has been reported to prolong survival in patients infected by HIV (Xiang et al., 2001).

By allowing different pathogen strains to coexist in the same habitat, co-infections also create opportunities for recombination, which can be involved in disease emergence. So far, many studies point to the importance of recombination but they tend to overlook the epidemiology, i.e. the fact that in order for the molecular process (recombination) to take place, there needs to be a co-infection. One of the application of these co-infection models could be the study of the emergence of bird influenza pathogenic forms capable of being transmitted amongst humans. Recent serial passage experiments (that were fiercely debated in the media) evolved a bird influenza virus into a form that is pathogenic to ferrets, which is usually a sign of pathogenicity to humans (although this is controversial Cohen, 2012). One study (Herfst et al., 2012) showed that the evolved strain was only 5 substitutions away from sequences of wild bird flu viruses and another study (Imai et al., 2012) found that an existing virus needed only four substitutions and reassortment to become transmissible among ferrets<sup>19</sup>. A theoretical model was also built using this data to assess the risk of within-host evolution of such a pathogenic variant but it neglected the epidemiology (Russell et al., 2012), partly because most of the mutations are based on the same genomic segment (HA) and recombination within a segment is known to be rare. However, at least in one of the studies, one of the mutations is on another segment (PB2), which means that reassortment (i.e. exchange of genomic segments between influenza viruses infecting the same host) can be involved in emergence. Furthermore, it is likely that these serial passage experiments only identified one of the few evolutionary paths and that other paths could involve mutations on different segments.

### Recombination and virulence (co-)evolution

Recombination is known to be an important feature for parasites to persist in hosts (Schmid-Hempel, 2008). The vast majority of models have focused on the role of recombination in the evolution of drug resistance and, in the case of HIV for instance, results have shown that the role of recombination can be sensitive to population dynamic feedbacks and stochasticity (see Kouyos et al., 2009, for a model that combines the two effects). Some models have also considered the role of recombination in immune escape (Mostowy et al., 2011).

However, I am not aware of models that have considered the role of recombination on the evolution of virulence. There are two reasons for this. The first reason is that virulence is by essence a multi-factorial trait, which makes it ideal for models based on quantitative traits but problematic for models involving a finite number of loci. Another limitation is the potentially small effect of recombination on virulence evolution. Finally, the third concern is the fact that results are already known: recombination should be advantageous when far from the optimal virulence and disadvantageous when close from the optimum.

19. One cannot help but be amazed that the ‘rediscovery’ of serial passage experiments by molecular pathologists would lead to publications in high-profile magazines.

There are ways to address these concerns however. In order to solve the ‘quantitative trait’ problem, one could for instance use a fitness landscape as realistic as possible. More precisely, what I have in mind is the HIV fitness landscape, which has been inferred by [Hinkley et al. \(2011\)](#) by combining more than 70,000 HIV sequences to their *in-vitro* replicative capacity. At the other extreme, another possibility would be to simplify things as much as possible and consider only two loci, one for virulence and another for transmission. Recombination would then allow to mix the 4 parasite genotype combinations (between low and high virulence with low and high transmission).

Concerning the effect of recombination on virulence evolution, on the long run we do not expect any difference: the parasite genotype selected should be the one closest to the optimum. However, recombination can affect transitory dynamics. Furthermore, unique feedbacks occur when considering recombination and virulence. First, the full epistasis effect (i.e. the fact that mutating each of the two loci can have an overall effect that differs from the sum of the effects of each mutation at easy locus) would be a direct consequence of the epidemiological model instead of being chosen in an *ad hoc* manner. Second, there is an indirect link between recombination and virulence because increased virulence also means shorter infections, which itself implies less opportunities for co-infections. Since multiple infections is a requirement for there to be recombination, we have an indirect effect between the two traits that is mediated by epidemiological feedbacks.

From a practical point of view, one possibility to address this question would be to use the Price equation framework introduced by [Day and Proulx \(2004\)](#) and recently applied to recombination and the evolution of drug resistance ([Day and Gandon, 2012](#)). This framework is particularly well adapted to study the consequences of recombination because it focuses on short-term evolutionary dynamics. In the present case, one limitation is that currently it does not allow to model co-infections ([Day and Gandon](#) use a superinfection assumption), which could affect the results for recombination. This framework could also be used to study the evolution of recombination itself. In the case of viruses of instance, there is a striking dichotomy between viruses such as HIV where recombination is extremely frequent and viruses like HCV and HPV that seldom recombine. Understanding the bases for these differences is something Matthew Hartfield and I are particularly interested in.

### Combining within-host processes

Typically, multiple infection models consider one type of within-host interaction (e.g. competition for host resources, public goods production, spite) but the reality is likely to be a combination of these. Rebecca Schulte-Iserlohe is developing experiments on evolution of the bacterium *Bacillus thuringiensis* in its host *Caenorhabditis elegans* ([Schulte et al., 2010](#)). The life-cycle is as follows: worms ingest free spores, which release a toxin upon being digested. These toxins destroy intestinal cells of the worm, thus freeing nutrients for the bacteria. Importantly, the toxin is a public goods because it is required to initiate the infection but once the spores have germinated (and once the worm’s gut has been destroyed), the toxin is not required anymore. Therefore, a bacterium without the toxin could very well rely on another strain to initiate the infection. Once the host resources are used up, the bacteria produce spores to disperse.

In addition to toxin (public goods) production, these bacteria interact in several other ways.

First they can produce bacteriocins to directly harm unrelated strains. Second, they produce quorum sensing molecules, which allows them to coordinate the sporulation. Finally, they can also exchange plasmids.

We want to use a multiple infection framework to study the co-evolution of toxin production and virulence, which is assumed to be proportional to the within-host replication rate. Using a similar model, we will look at the coevolution between bacteriocin production and virulence. Finally, we will try to see whether it is possible to combine all three processes. One of the goals will be to determine whether the observed patterns match that obtained using evolutionary experiments. In particular, we can compare the virulence we would expect a *B. thuringiensis* to evolve without co-infections to that evolved in presence of a competing strain. Rebecca has also characterised strains that lack the ability to produce the toxin, which should facilitate comparisons between theory and experiments. Finally, coevolutionary dynamics could also be envisaged on the long term.

#### 2.8.4 Phylogenies and pathogen evolutionary dynamics

##### Phylogenetic signal using human phylogenies?

Host genotype is known to affect HIV infections. In addition to striking mutations, such as the CCR5- $\Delta$ 32 deletion, several polymorphisms in the genome have been linked to increased or decreased set-point viral load (spVL) values (Fellay et al., 2007). These genome wide association studies (GWAS) found that at least 20 % of the variability in spVL can be explained by single nucleotide polymorphisms (SNPs) that are mostly located in the HLA class I genomic region. One of the reasons for this somehow low value (e.g. compared to the approx. 50 % explained by the virus genotype) is that (for statistical reasons) only a few genome-wide significant SNPs are included in the estimation of explained variance on the host side, while virus studies use methods that potentially account for all the HIV-1 genomic variation.

Contrary to the GWAS, the phylogenetic comparative method (PCM) does not aim at detecting specific positions associated with higher or lower trait values. Rather, it estimates the fraction of the variance in the infection trait that is due to variations among host (or virus) genotypes. In other words, the PCM will tell us if a genome (as a whole) affects the trait value but not which positions of the genome matter.

Conceptually the idea is simple: we want to apply the PCM to a host phylogeny. By testing the extent to which closer hosts in the phylogeny (i.e. more related hosts) tend to have similar infection trait values, we will estimate the control the host genotype has on the infection trait. If successful, we will then be able to compare it to the virus genotype control over the same trait. This result will be new in itself because we still have an unclear idea of the host overall control over the trait.

The challenging part of course resides in building a phylogeny of hosts infected by HIV. We do not have any ‘pedigree’ information about our hosts, who are all patients enrolled in the Swiss HIV Cohort Study (SHCS). The only information we have comes from the Core genetics projects of the SHCS and consists in small nucleotide polymorphisms (SNPs) with approximately 2 million SNPs per patient spread throughout the genome.

It is not problematic in itself that we do not know the true genealogy (similarly, it was not

problematic that we did not know the true transmission chain and could rely on the virus phylogeny). What matter for the phylogenetic approach is that we have an idea of the degree of relatedness between hosts. One real problem though is that the relatedness amongst hosts varies strongly depending on the genomic region considered. Because of the high recombination rates in the human genome, regions on the same chromosome can evolve almost independently. There are two ways to tackle this problem. One possibility is to only use a very restricted portion of the genome to infer a host phylogeny. Another is to use a method developed by Peter Visscher's lab in order to estimate genome-wide heritability (Yang et al., 2011). According to them, this method can even be used to partition heritability amongst chromosomes (Yang et al., 2011). In the case of HIV virulence, this would allow us to check the consistency of the results because according to earlier GWAS results (Fellay et al., 2007), we would expect a lot of the control to originate from chromosome 6, where the HLA complex is located.

Finally, we would also like to estimate the importance of interactions between host and virus genotypes on the infection outcome. Estimating the host and the virus genotype control over spVL separately using the same method (the PCM) will give us a first insight on their respective effects. A more detailed investigation of these interactions could be achieved by performing the PCM on the virus phylogeny but using residuals from a GWAS analysis instead of trait values. By doing so, we should remove host effects and interaction effects. If we manage to obtain residuals from the PCM, we could perform a similar operation to use residuals in a GWAS. Of course, the GWAS could be replaced by a PCM on host phylogenies if we manage to get these working.

### PEPS project: From population dynamics to phylogenies

Analysing the genome of rapidly mutating pathogens can provide us with insights concerning how a pathogen spreads (i.e. its demography). For instance, the growth rate of an epidemics should affect how mutations are fixated into the pathogen genome.

Inferring phylogenies using a coalescent model requires making assumptions about population sizes. The first model assumed a constant population size (Kingman, 1982) but later implementations of the coalescent managed to allow for population growth (Griffiths and Tavaré, 1994). With such models, it becomes possible to infer values of a parameter for the growth rate of the number of sequences. In the case of infectious diseases, one can make a parallel between this growth rate and the basic reproduction ratio ( $R_0$ ) or the exponential growth rate of the epidemics ( $r_0$ ). Another method that allows to make inferences about epidemiology from sequence data is the skyline plot (Pybus et al., 2000). It discretises time in order to infer a piecewise-constant model of population size. The advantage is that this model does not require to make an assumption about the demography (contrary to the constant or the exponential models for instance). The model was extended by Drummond et al. (2005) to allow Bayesian inference based on a Markov chain Monte Carlo sampling procedure. One of the famous early applications of these methods consisted in inferring the number of hepatitis C virus (HCV) cases in Egypt (Pybus et al., 2003, Drummond et al., 2005). These methods were able to identify an increase in the effective number of infections occurring approximately 75 years in the past, which they argued to be caused by viral contamination of an antischistosomiasis treatment that was widely used in Egypt from the 1920s.

More recently, [Stadler et al. \(2012\)](#) made important progress by replacing the coalescent model by a birth-death model to calculate the likelihood of a phylogeny given model parameters. The birth-death model is well suited to describe epidemics because contrary to the coalescent model, transmission and death events are modelled separately instead of being aggregated into a growth rate. This means that this method can infer both a transmission rate and a removal rate instead of an  $R_0$ . One limitation is that the birth-death model requires a good sampling in the population.

Another limitation lies in the derivation of the likelihood function, which imposes to use epidemiological models in their simplest form. For instance, [Stadler et al. \(2012\)](#) assume that the transmission rate parameter is constant over time. In other words, they consider that there is never any depletion in terms of the number of susceptible hosts. To illustrate the simplicity of this model, we can write it in the form of an ODE:

$$\frac{dI}{dt} = \beta I - \gamma I \quad (2.17)$$

This is still better than the exponential growth rate of the coalescent but it still lacks some of the main properties of epidemiological models. Recently, Gabriel Leventhal and Tanja Stadler managed to write down the likelihood function for an SIS model, which can be written in the form:

$$\frac{dI}{dt} = \beta (N - I) I - \gamma I \quad (2.18)$$

where  $N$  is the total host population size, which is assumed to be constant.

What is particularly striking with the method introduced by [Stadler et al. \(2012\)](#) is its accuracy in estimating parameters (at least based on simulated data) from phylogenies. Since Tanja Stadler is able to calculate the likelihood of a tree given parameter values (transmission rate, recovery rate, sampling rate), she uses all the information of the phylogeny. With Olivier Gascuel, we want to determine whether similar results can be inferred from a selection of summary statistics, such as the distribution of branch lengths or the balance of the tree, instead of the whole likelihood.

The first part of this objective required to generate pathogen phylogenies using different underlying population dynamics models. What we needed was an individual based model where a new node is added to the phylogeny at each infection event and a branch is deleted at each event of end of infection. It is very likely that several colleagues have already built such a program but they were not publicly available when we started this project (since then, we learned about the program MASTER that sort of does this, [Vaughan and Drummond, 2013](#)). Matthieu Jung, a bioinformatician, coded a program called `phyloepid` for us. It is implemented in java and based on a multi-agent simulation technique, which offers more flexibility for further improvements (e.g. adding spatial structure). Matthieu is still working on the code in order to generalise the program to any type of epidemiological model and not just SIR.

Figure 2.18 illustrates the type of output a run of `phyloepid` produces. For a given epidemiological model (here an SIR model with density-dependent transmission), we get a time series with the number of infected (on the left) and a phylogeny of infections (on the right). An important idea also illustrated by this figure is that the phylogeny contains more information than

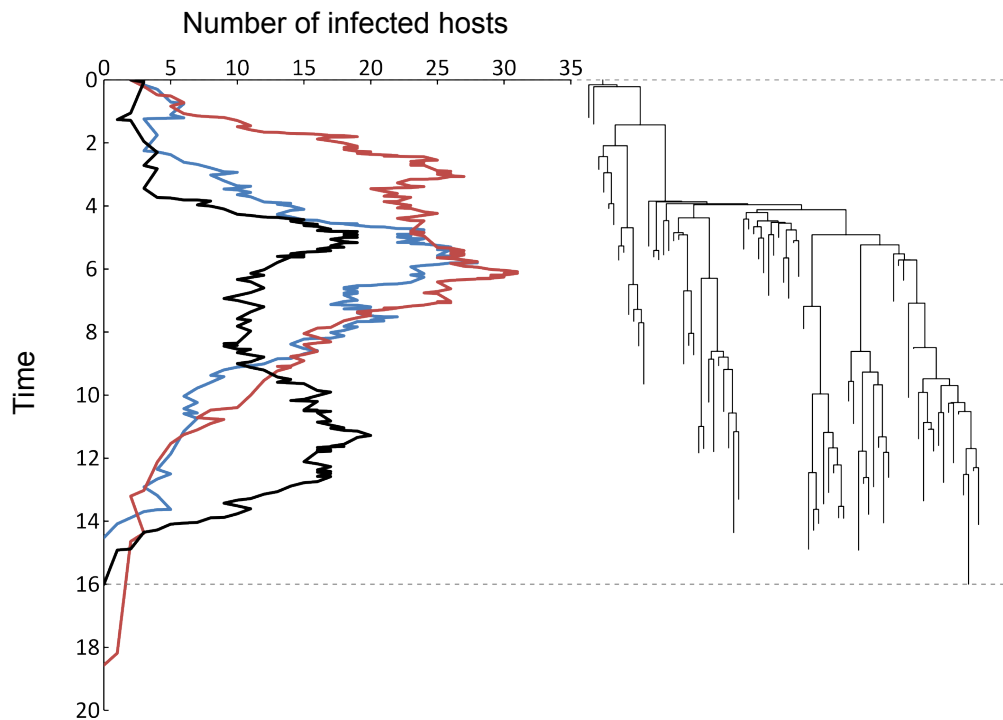


Figure 2.18: **Linking epidemiological dynamics and phylogenies.** On the left, three runs of an individual based SIR model with density-dependent transmission. On the right, the phylogeny built from the black run (assuming perfect sampling). The key point is that the phylogeny contains more information than the time series.

the time series. In fact, the time series could be inferred from the phylogeny by counting the number of branches coexisting at each time point. Some of the additional pieces of information a phylogeny can convey are for instance the time between two events (through the length of internal branches) or the heterogeneity in the structure of the epidemiological model (through the balance of the tree).

The goal of Guilhem Heinrich's 4 months MSc internship was to look for tree summary statistics that allowed to best estimate parameters from a given epidemiological model or, even, to compare the likelihood of different underlying epidemiological models. As shown in Figure 2.19A, when we vary a single parameter, it is possible to find summary statistics that lead to decent credibility intervals. Furthermore, if the  $R_0$  is set to a constant value, we can distinguish between a model with density-dependent transmission and constant transmission (Figure 2.19B). Note that in the later case, the later the sampling, the clearer the difference between the two cases. This is because the effects of host limitation become more pronounced in the later stages of the epidemics.

These results are very preliminary in many ways. First, we only had time to consider two types of SIR models, one with density-dependent transmission and another with constant transmission



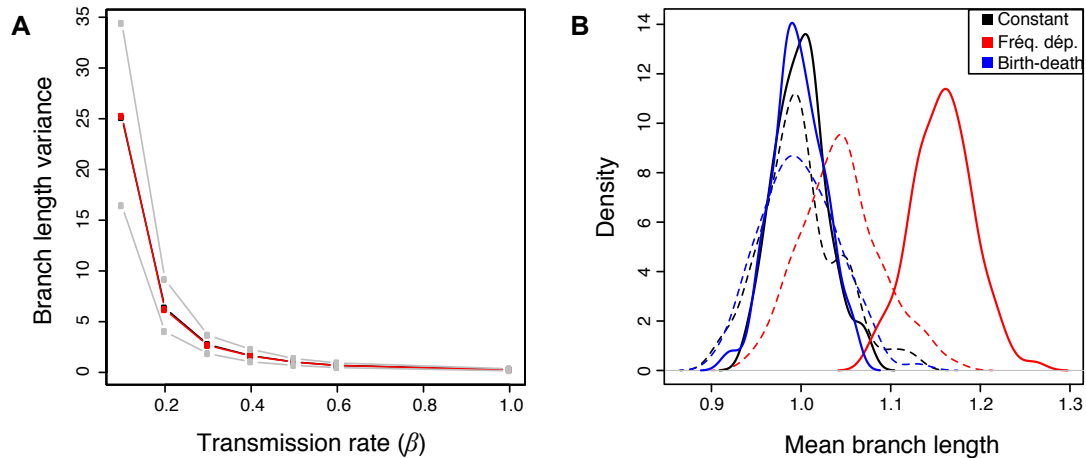


Figure 2.19: **Inferring epidemiological properties from tree summary statistics.** (A) Recovery rate is constant and transmission rate ( $\beta$ ) varies. The observed variance in branch length allows us to infer the value of  $\beta$  with small credibility intervals (in grey). (B) Three models are compared. The constant model (in black) and the birth-death model (in blue) are identical but the second one is coded in R. In all three cases the  $R_0$  is constant. The phylogenies analyse have 200 leaves (A), 800 leaves (B, dashed lines) and 2000 leaves (B, plain lines).

(identical to the birth-death model). Retrospectively, the SIR model clearly is not the simplest because the population size never reaches a steady state (contrary to an SIS model for instance). As a consequence, there is an interaction between the sampling size and the sampling time. Furthermore, for a constant sampling efficiency, we can get the same number of sequences in the ascending or in the descending phase of the epidemics. More generally, sampling is likely to be one of the key processes to define in these models.

Another limitation is that we only varied one parameter at a time. When Guilhem tried to vary both transmission rate and recovery rate, he was unable to distinguish phylogenies with similar  $R_0$  (i.e. high transmission rate with high recovery rate vs. low transmission rate and low recovery rate). Part of the problem is that he did not combine summary statistics. Another problem is also linked with the sampling mode: we assumed that all of our sequences are sampled at the same time point, which means that it is difficult to root our tree and time.

The job of the new PhD student on the project, Emma Saulnier, will be to first determine whether the problems faced by Guilhem can be overcome (either by varying the sampling strategy and/or by combining summary statistics). Another idea is to consider an SIS model, which has the advantage of reaching a steady state. Furthermore, Gabriel Leventhal and Tanja Stadler have developed an approach that calculates the likelihood of the full phylogeny using a birth-death model so we will even be able to compare the power of our summary statistics to the most complete statistics.

If these steps are successful, we will then move to more complicated models. Indeed, the

main motivation of this project was to infer parameter values in cases where the epidemiological model is too complex for a phylogeny likelihood to be calculated analytically. This would be the case of vector-borne models or even for any model that requires keeping track of the dynamics of more than two host compartments (in their unpublished SIS model, Leventhal *et al.* assume a constant population size so they only track one compartment). Such models would quite naturally lead to Approximate Bayesian Computing methods (Sunnåker *et al.*, 2013) in order to infer parameter values or even epidemiological models from pathogen sequence data.

### 2.8.5 Within-host models

#### Immune system modelling

I always tried not to get into the details of how the immune system works. There are so many cellular types and signalling molecules that these can make trophic networks look simple. This is partly the reason why I always try to only include in my models immune system characteristics that I can really understand (B cells *vs.* T-cells, or specific cells *vs.* non-specific cells, etc.). One might ask whether there is any point in doing such a simple modelling of immune dynamics. I think there are several answers to this question.

- **Parasite evolution.** Understanding selective pressures acting on parasites is not a key question to immunologists, who tend to focus on how the immune system works. Considering the specificities of the parasite's life cycle inside the host requires to simplify the complexity of the immune system to keep the analysis tractable. More generally, the role parasite evolution should have in within-host models is not very clear because, as we discussed for the case of HIV (Alizon and Magnus, 2012), it is sometimes difficult to see if evolution is neutral or if it affects the dynamics.
- **Emergence.** The last few years have witnessed a transposition to the within-host level of stochastic models that were initially designed to study emergence at the epidemiological level (Alexander and Bonhoeffer, 2012, Loverdo *et al.*, 2012). This question is particularly interesting with respect to the evolution of drug resistance and combining it with simple immune dynamics could yield interesting results. In fact, deterministic models have already shown that immune dynamics play an important role in the evolution of drug resistance and compensatory mutations (Handel *et al.*, 2009).
- **Self *vs.* non self.** The traditional view in immunology is that elements of the self do not elicit an immune response, whereas non-self elements do (Burnet and Fenner, 1949). This is sometimes a bit tautological because an element of the self is often defined by the fact that it does not elicit an immune response. . . This binary view has become increasingly difficult to reconcile with numerous observations. For instance, cancer cells can be recognised even though they exhibit similar MHC molecules than non-carcinogenic cells. Conversely, the microbiota in our gut clearly does not exhibit self peptides but it is nevertheless tolerated. One of the first challenges to the self/non-self theory came from Matzinger (2002), who introduced the 'danger theory'. This theory stipulates that the immune system is only activated when 'danger' signals are detected. However, as pointed out by Pradeu and Cooper (2012), although in some cases it is possible to define what these signals are, there are cases where the theory is not entirely satisfying. During his PhD, Thomas Pradeu intro-

duced a variation of the danger theory by proposing to replace the self/non-self criterion by a continuity criterion (Pradeu and Carosella, 2006, Pradeu, 2012). His idea is that what the immune system responds to are perturbations of the within-host environment. This view coincides in many aspects with the ecological vision of the immune system that I have adopted and we are working on developing within-host models to capture his theory.

### Within-host evolution challenge

We have already mentioned the fact that HIV seems to evolve more rapidly at the within-host level than at the epidemiological level (Alizon and Fraser, 2013). One implication of this result is that some of the mutations fixed in the virus genome are lost during the transmission bottleneck. In other words, the link between levels of selection is not necessarily trivial: not only does the timing of the transmission even matter a lot (the same infection could transmit different viruses at different time points) but also trade-offs could be involved, such that viruses that are the most frequent in a host are not necessarily the most transmitted. More generally, one of the questions I would like to investigate is the link between within-host evolution, host diversity and epidemiological processes.

Human populations are known to differ for the type and the distribution of Human Leucocyte Antigens (HLA) alleles, which are involved in recognising and fighting pathogens. A seminal study by Poon et al. (2007) considered the acquisition of escape mutations by HIV and HCV in response to immune pressure by cytotoxic T-lymphocytes (CTLs). They complemented their data analysis with a simple population genetics model, which allowed them to explain why selection at the within-host level does not blur all the variability in the virus population at the between-host level.

Poon et al. (2007) make several simplifying assumptions that should be addressed. First, they summarise all the within-host dynamics with a single parameter, when I think an explicit model would be much more appropriate. Second, they only consider a single HLA allele with two types of viruses (wild type and escape) in their model. I would like to introduce a more detailed description of the host-parasite interaction. The idea would be to define the parasite genome as a set of epitopes that can be turned on or off. Then, each HLA allele of the host would be associated with a set of lymphocyte clones that can recognise an epitopes. Turning an epitope off entails a fitness cost for the parasite but it also means that the lymphocytes cannot spot cells infected by the parasite.

With Fabio Luciani, we are working on investigating how the diversity of HLA alleles in the host population affects the evolution of viral escape mutations. The number of alleles present in a population and their distribution vary in human populations. It will be especially interesting to compare data on virus escape mutations obtained in different populations to see if these patterns match those predicted by the model. On a more conceptual level, this study will allow us to address more fundamental questions linked to the evolution of the immune system. For instance, are populations with some HLA allele distributions more resistant to epidemics than other? Or, what is the optimal level of cross-reactivity of the immune response. Again, this conceptual model will allow comparison with the data, as distributions of HLA alleles are well documented for many human populations worldwide.

### 2.8.6 More weighting for sex

With Mathieu Moslonka-Lefebvre and Sebastian Bonhoeffer, we showed that weighting sexual-contact networks with realised number of sex acts between individuals greatly affects epidemic spread (Moslonka-Lefebvre et al., 2012a). However, although we did manage to obtain a real sexual contact network, we did not have any information about the relationship between the number of partners an individual has and his/her number of sex acts. This was not problematic in the context of our study because we could show that any deviation from a purely proportional relationship would affect the results. Furthermore, earlier results have shown that it is clearly unrealistic to assume that the number of sex acts is proportional to the number of partners (Blower and Boe, 1993, Nordvik and Liljeros, 2006, Britton et al., 2007).

In a more recent study, we developed a framework that can be parametrized with a joint probability distribution of number of partners and number of sex acts (Kamp et al., 2013). This is already much more realistic than the assumptions we made earlier (Moslonka-Lefebvre et al., 2012a). However, it is still unsatisfying because these values are self-reported. There are already well-documented biases in self-reporting of the number of sexual partners (Smith, 1992) and these biases are likely to also be present, if not stronger, in self-reported number of sex acts (with the extra complication that sex acts tend to be more difficult to count than partners).

One idea we had with Christian Althaus was to use prevalence data in order to infer the most likely number of sex acts. In the NATSAL data from the UK (Johnson et al., 2001), many individuals enrolled in the cohort are tested for Chlamydia. This sexually-transmitted disease can last for several years with often minimal symptoms and is therefore a good indicator of sexual activity in the population. Our idea is to build an STD transmission model with an explicit description of the relationship between number of partners and number of sex acts in order to see which relationship provide the most parsimonious explanation of the Chlamydia prevalence data.

The model we have in mind with Christian differs from the network-based models because it only considers a heterogeneous host population, where hosts are grouped into classes corresponding to their number of partners. A key parameter in these models is the degree of random vs. assortative mating. Typically, models introduce a parameter  $\varepsilon$  such that when  $\varepsilon = 0$  contacts only occur between hosts from the same class and when  $\varepsilon = 1$  contacts are made with a random host in the population (Hethcote and Yorke, 1984, Garnett et al., 1999). An advantage of our model is that in addition to estimating the most likely relationship between number of partners and number of sex acts, we can also infer the most likely value of  $\varepsilon$ . This would be particularly useful to drive public health policies because a low value of  $\varepsilon$  means that risk groups should be targeted in priority, whereas large values of  $\varepsilon$  suggests that policy measures should target the whole population.

### 2.8.7 Is the HPV vaccine 'evolution-proof'?

The last century has witnessed the development of many drugs, not least antibiotics. The enthusiasm these generated was washed out by the generalisation of drug resistance (and probably more importantly by the emergence of HIV). Indeed, an important variable left out of the equation was that parasites evolve. The study of the evolution of drug resistance has become

one of the trendiest topics to work on these days. But drugs are not our only line of response against infectious diseases. Vaccination is also at the heart of public health policies and there is evidence that vaccination often triggers to parasite evolutionary responses ([Gandon and Day, 2008](#)).

The recently launched Human Papillomavirus (HPV) vaccine is considered to be ‘evolution-proof’. Carmen Murral’s post-doctoral project with me and Ignacio Bravo (from the Catalan institute for cancer in Barcelona) is to investigate the validity of this claim by evaluating the risk that HPV will evolve to give an unwanted response. We envisage three scenarios. First, there could be emergence of cross-immunity escape mutants, which would render the vaccine ineffective. Second, we could observe type replacement. This is because there are many types of HPV and the vaccine only targets the most carcinogenic ones<sup>20</sup>. Vaccination could lead to a shift in type prevalences, which again would make the vaccine inefficient (this has been partly investigated by [Poolman et al., 2008](#)). Finally, a third, and more worrying scenario, is virulence evolution, where the vaccine would lead to the selection of types or mutants that are more virulent. This is justified by the fact that increased virulence after vaccination is an outcome of many evolutionary epidemiology models ([André and Gandon, 2006](#), [Ganusov and Antia, 2006](#)). This is very well documented in the case of Marek disease virus infecting poultry: the release of each new vaccine has led to an increase in the virulence of the virus to its host ([Atkins et al., 2013](#)).

As illustrated by [Bravo et al. \(2010\)](#), the idea that understanding HPV evolution has direct clinical applications is progressing. In addition to the general reluctance of the medical field to add evolutionary biology into the picture, things are made complicated by the fact that this virus seems to evolve slowly due to low mutation rates and extremely low recombination rates. However, we now know that the virus’ genome is not homogeneous with respect to mutation rates and that some regions are evolving more rapidly. In fact, [Bravo and Alonso \(2004\)](#) showed that oncogenes tend to have higher mutation rates. More generally, the HPV genome is highly modular and more recent genes tend to evolve faster than older genes ([García-Vallvé et al., 2005](#)).

This project will involve both data analysis and mathematical modelling. The originality of our approach is that we want to combine a description of the within-host ecology of the virus to an evolutionary epidemiology perspective. Carmen has begun to investigate such within-host models in the case of HPV during her PhD ([Murall et al., 2012](#)). Three aspects will be particularly interesting. First, how does the vaccine affect this within-host ecology? Put differently, can we predict type replacement based on within-host interactions? Second, in case of co-infection by multiple HPV types, which is quite frequent, how do the different viruses interact? Third, what is the result of the co-infection between HPV and another virus, for instance HIV? These questions can be addressed from an ecological perspective but also from an evolutionary perspective. In fact, in the case of co-infections between HPV and HIV, both viruses can evolve in response to the other and it would be useful to determine whether the shape of the virus phylogenies are affected by this co-infection (see the PEPS project above for further details). Note that the modelling of co-infections between HPV and HIV could be a project in itself given the prevalence and the public health importance of these ([Denny et al., 2012](#)).

---

20. The repartition of the types varies across geographical areas, which is why an HPV vaccine that is efficient in western Europe could prove to be completely inefficient in other parts of the world.

Once the within-host ecology (and evolution) of the virus is understood, we will attempt to link it to the between-host level in a context of vaccination. Several techniques can be used for such nested models (Mideo et al., 2008) and the ‘Price equation’ seems like a promising framework to study such short-term evolutionary responses (Day et al., 2011, Mideo et al., 2011).

In terms of data analysis, so far we have initiated three collaborations. First, Ignacio Bravo has access to (and is already analysing) a dataset of HPV in Spanish women. We are also in contact with Gonzague Jourdain in Thailand who has access to a dataset of women coinfecting by HIV and HPV. This will be useful to study the effect of co-infections although the fact that all the patients are treated for HIV is likely to complicate the picture. Finally, we are also collaborating with Eric Leroy and Nicolas Berthet in Gabon. They are initiating a screening campaign of HPV in urban and rural areas (their goal is to see whether different types circulate in different areas). Since the campaign is just beginning, we have an opportunity to influence the data collection.

This data will allow us to estimate the evolutionary potential of the virus at the within-host level and at the between-host level (as in Alizon and Fraser, 2013). One difficulty will of course be HPV’s low evolutionary rates. However, this could be compensated by the fact that recombination is extremely rare. Data from Gabon might also provide us with immunological data. This would allow us to fit within-host dynamics models that include immune cell dynamics and virus loads.

I will conclude by pointing out a coincidence, which is that this latest project combines some of the very first modelling approaches I developed during my PhD (within-host dynamics models and nested models) together with some of the more recent work I did on virus evolution (evolutionary rates and phylodynamics). The project also involves multiple infections and evolutionary responses to treatments so I really think that, although its description is still a bit hasty, it has the potential to grow as an important research theme over the next few years.

# Bibliography

- Alexander, H. K. and S. Bonhoeffer. 2012. Pre-existence and emergence of drug resistance in a generalized model of intra-host viral dynamics. *Epidemics* 4(4):187–202. [→ p. 71]
- Alexander, H. K. and T. Day. 2010. Risk factors for the evolutionary emergence of pathogens. *J R Soc Interface* 7(51):1455–74. [→ pp. 51 and 54]
- Alizon, S. (2006). Parasite virulence evolution: Insights from embedded models, PhD thesis, Université Paris 6. [→ p. 9]
- Alizon, S. 2008. Transmission-recovery trade-offs to study parasite evolution. *Am. Nat.* 172(3):E113–E121. [→ pp. 13 and 37]
- Alizon, S. 2013. Parasite co-transmission and the evolutionary epidemiology of virulence. *Evolution* 67(4):921–933. [→ p. 29]
- Alizon, S., A. Hurford, N. Mideo and M. van Baalen. 2009. Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *J. Evol. Biol.* 22:245–259. [→ pp. 4 and 12]
- Alizon, S. and B. Boldin. 2010. Within-host viral evolution in a heterogeneous environment. Insights into the HIV co-receptor switch. *J. Evol. Biol.* 23:2625–2635. [→ pp. 40 and 41]
- Alizon, S. and C. Fraser. 2013. Within-host and between-host evolutionary rates across the HIV-1 genome. *Retrovirology*. [→ pp. 44, 45, 72, and 75]
- Alizon, S. and C. Magnus. 2012. Modelling the course of an HIV infection: Insights from ecology and evolution. *Viruses* 4(10):1984–2013. [→ pp. 8, 41, 42, and 71]
- Alizon, S. and M. van Baalen. 2005. Emergence of a convex trade-off between transmission and virulence. *Am. Nat.* 165(6):E155–E167. [→ p. 15]
- Alizon, S. and M. van Baalen. 2008. Multiple infections, immune dynamics and virulence evolution. *Am. Nat.* 172(4):E150–E158. [→ p. 24]
- Alizon, S. and P. Taylor. 2008. Empty sites can promote altruistic behaviour. *Evolution* 62(6):1335–1344. [→ p. 31]
- Alizon, S. and S. Lion. 2011. Within-host parasite cooperation and the evolution of virulence. *Proc. R. Soc. Lond. B* 278(1725):3738–3747. [→ p. 27]
- Alizon, S. and Y. Michalakis. 2011. The transmission-virulence trade-off and superinfection: comments to smith. *Evolution* 65(12):3633–8. [→ pp. 15 and 21]
- Alizon, S., F. Luciani and R. R. Regoes. 2011. Epidemiological and clinical consequences of within-host evolution. *Trends Microbiol.* 19(1):24–32. [→ pp. 14 and 37]
- Alizon, S., J. C. de Roode and Y. Michalakis. 2013. Multiple infections and the evolution of virulence. *Ecol. Lett.* 16(4):556–567. [→ pp. 14, 24, 25, 26, and 28]
- Alizon, S., V. von Wyl, T. Stadler, R. D. Kouyos, S. Yerly, B. Hirschel, J. Böni, C. Shah, T. Klimkait, H. Furrer, A. Rauch, P. Vernazza, E. Bernasconi, M. Battegay, P. Bürgisser, A. Telenti, H. F. Günthard, S. Bonhoeffer and the Swiss HIV Cohort Study. 2010. Phylogenetic approach reveals that virus genotype largely determines HIV set-point viral load. *PLoS Pathog.* 6(9):e1001123. [→ pp. 21, 47, and 48]
- Allen, L. 2008. An Introduction to Stochastic Epidemic Models. chapter 3, pp. 81–130 in F. Brauer, P. van den Driessche and J. Wu, eds. *Mathematical Epidemiology*. Vol. 1945 of *Lecture Notes in Mathematics*, Springer, Berlin / Heidelberg. [→ p. 55]
- Anderson, R. M. and R. M. May. 1979. Population biology of infectious diseases. *Nature* 280:361–367. [→ p. 16]
- Anderson, R. M. and R. M. May. 1982. Coevolution of hosts and parasites. *Parasitology* 85:411–426. [→ p. 12]
- Anderson, R. M. and R. M. May. 1991. *Infectious Diseases of Humans. Dynamics and Control*. Oxford University Press, Oxford. [→ pp. 17, 31, 53, and 54]
- André, J.-B. and S. Gandon. 2006. Vaccination, within-host dynamics, and virulence evolution. *Evolution* 60(1):13–23. [→ p. 74]
- André, J.-B. and T. Day. 2005. The effect of disease life-history on the evolutionary emergence of novel pathogens. *Proc. R. Soc. Lond. B* 272(1575):1949–

1956. [→ pp. 23, 51, and 54]
- André, J.-B., J.-B. Ferdy and B. Godelle. 2003. Within-host parasite dynamics, emerging trade-off, and evolution of virulence with immune system. *Evolution* 57(7):1489–1497. [→ p. 9]
- Antia, R., B. R. Levin and R. M. May. 1994. Within-host population dynamics and the evolution and maintenance of microparasite virulence. *Am. Nat.* 144:457–472. [→ p. 9]
- Antia, R., R. R. Regoes, J. Koella and C. T. Bergstrom. 2003. The role of evolution in the emergence of infectious diseases. *Nature* 426:658–661. [→ pp. 51 and 54]
- Antonovics, J., J. L. Abbate, C. H. Baker, D. Daley, M. E. Hood, C. E. Jenkins, L. J. Johnson, J. J. Murray, V. Panjeti, V. H. W. Rudolf, D. Sloan and J. Vondrasek. 2007. Evolution by any other name: antibiotic resistance and avoidance of the E-word. *PLoS Biol.* 5(2):e30. [→ p. 63]
- Atkins, K. E., A. F. Read, N. J. Savill, K. G. Renz, A. F. Islam, S. W. Walkden-Brown and M. E. J. Woolhouse. 2013. Vaccination and reduced cohort duration can drive virulence evolution: Marek’s disease virus and industrialised agriculture. *Evolution* 67(3):851–60. [→ p. 74]
- Ayres, J. S. and D. S. Schneider. 2012. Tolerance of infections. *Annu Rev Immunol* 30:271–94. [→ p. 59]
- Ball, G. H. 1943. Parasitism and evolution. *Am. Nat.* 77(771):345–364. [→ p. 11]
- Barrat, A., M. Barthelemy, R. Pastor-Satorras and A. Vespignani. 2004. The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. USA* 101(11):3747–3752. [→ p. 32]
- Blomberg, S. P., T. J. Garland and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57(4):717–745. [→ pp. 47 and 50]
- Blower, S. M. and C. Boe. 1993. Sex Acts, Sex Partners, and Sex Budgets: Implications for Risk Factor Analysis and Estimation of HIV Transmission Probabilities. *J Acquir. Immune Defic. Syndr.* 6(12):1347–1352. [→ pp. 32 and 73]
- Bocharov, G., B. Ludewig, A. Bertoletti, P. Klenerman, T. Junt, P. Krebs, T. Luzyanina, C. Fraser and R. M. Anderson. 2004. Underwhelming the immune response: effect of slow virus growth on CD8<sup>+</sup>-T-lymphocyte responses. *J. Virol.* 78(5):2247–2254. [→ p. 13]
- Bolker, B. M., A. Nanda and D. Shah. 2010. Transient virulence of emerging pathogens. *J. R. Soc. Interface* 7(46):811–22. [→ p. 12]
- Bonhoeffer, S. and M. A. Nowak. 1994. Intra-host versus inter-host selection: viral strategies of immune function impairment. *Proc. Natl. Acad. Sci. USA* 91:8062–8066. [→ p. 13]
- Bonhoeffer, S., R. E. Lenski and D. Ebert. 1996. The curse of the pharaoh: the evolution of virulence in pathogens with long living propagules. *Proc. R. Soc. Lond. B* 263(1371):715–21. [→ p. 17]
- Boots, M., A. Best, M. R. Miller and A. White. 2009. The role of ecological feedbacks in the evolution of host defence: what does theory tell us? *Philos Trans R Soc Lond B Biol Sci* 364(1513):27–36. [→ p. 59]
- Boots, M. and A. Sasaki. 1999. ‘Small worlds’ and the evolution of virulence: infection occurs locally and at a distance. *Proc. R. Soc. Lond. B* 266:1933–1938. [→ p. 15]
- Bravo, I. G. and A. Alonso. 2004. Mucosal human papillomaviruses encode four different E5 proteins whose chemistry and phylogeny correlate with malignant or benign growth. *J Virol* 78(24):13613–26. [→ p. 74]
- Bravo, I. G., S. de Sanjosé and M. Gottschling. 2010. The clinical importance of understanding the evolution of papillomaviruses. *Trends Microbiol* 18(10):432–8. [→ p. 74]
- Britton, T., M. Deijfen and F. Liljeros. 2011. A weighted configuration model and inhomogeneous epidemics. *J Stat Phys* 145:1368–1384. [→ p. 32]
- Britton, T., M. K. Nordvik and F. Liljeros. 2007. Modelling sexually transmitted infections: the effect of partnership activity and number of partners on  $R_0$ . *Theor. Popul. Biol.* 72(3):389–99. [→ pp. 32 and 73]
- Brown, S. P., D. M. Cornforth and N. Mideo. 2012. Evolution of virulence in opportunistic pathogens: generalism, plasticity, and control. *Trends Microbiol* 20(7):336–42. [→ pp. 15 and 20]
- Brown, S. P., M. E. Hochberg and B. T. Grenfell. 2002. Does multiple infection select for raised virulence? *Trends Microbiol.* 10:401–405. [→ p. 27]
- Buchbinder, S. P., M. H. Katz, N. A. Hessel, P. M. O’Malley and S. D. Holmberg. 1994. Long-term HIV-1 infection without immunologic progression. *AIDS* 8(8):1123–8. [→ p. 46]
- Buckling, A. and M. A. Brockhurst. 2008. Kin selection and the evolution of virulence. *Heredity* 100(5):484–



488. [→ p. 27]
- Bull, J. J. 1994. Virulence. *Evolution* 48(5):1423–1437. [→ pp. 11 and 13]
- Bull, R. A., F. Luciani, K. McElroy, S. Gaudieri, S. T. Pham, A. Chopra, B. Cameron, L. Maher, G. J. Dore, P. A. White and A. R. Lloyd. 2011. Sequential bottlenecks drive viral evolution in early acute hepatitis C virus infection. *PLoS Pathog.* 7(9):e1002243. [→ p. 28]
- Burnet, F. M. and F. Fenner. 1949. The production of antibodies. 2nd Edition, McMillan, Melbourne. [→ p. 71]
- Byars, S. G., D. Ewbank, D. R. Govindaraju and S. C. Stearns. 2010. Colloquium papers: Natural selection in a contemporary human population. *Proc. Natl. Acad. Sci. USA* 107 Suppl 1:1787–92. [→ p. 62]
- Bynum, W. F. 1983. Darwin and the Doctors: Evolution, Diathesis, and Germs in 19th-Century Britain. *Gesnerus* 40(1–2):43–53. [→ p. 62]
- Caswell, H. 1989. Matrix population models: construction, analysis and interpretation. Sinauer Associates, Inc., MA 01375 USA. [→ p. 16]
- Chao, L., K. A. Hanley, C. L. Burch, C. Dahlberg and P. E. Turner. 2000. Kin selection and parasite evolution: higher and lower virulence with hard and soft selection. *Q. Rev. Biol.* 75(3):261–275. [→ p. 26]
- Chapuis, E., A. Arnal and J.-B. Ferdy. 2012. Trade-offs shape the evolution of the vector-borne insect pathogen *Xenorhabdus nematophila*. *Proc. R. Soc. Lond. B* 279(1738):2672–80. [→ pp. 18 and 19]
- Choisy, M. and J. C. de Roode. 2010. Mixed infections and the evolution of virulence: effects of resource competition, parasite plasticity and impaired host immunity. *Am. Nat.* 175(5):E105–E118. [→ pp. 24 and 63]
- Cockburn, T. A., E. Cockburn and T. A. Reyman. 1998. Mummies, Disease and Ancient Cultures. Cambridge University Press. [→ p. 11]
- Cohen, J. 2012. Avian influenza. The limits of avian flu studies in ferrets. *Science* 335(6068):512–3. [→ p. 64]
- Colizza, V., A. Barrat, M. Barthélemy and A. Vespignani. 2006. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc Natl Acad Sci USA* 103(7):2015–20. [→ p. 32]
- Craig, S., H. M. Thu, K. Lowry, X.-f. Wang, E. C. Holmes and J. Aaskov. 2003. Diverse dengue type 2 virus populations contain recombinant and both parental viruses in a single mosquito host. *J. Virol.* 77(7):4463–7. [→ p. 29]
- Day, T. 2001. Parasite transmission modes and the evolution of virulence. *Evolution* 55(12):2389–2400. [→ pp. 10 and 20]
- Day, T. 2002a. On the evolution of virulence and the relationship between various measures of mortality. *Proc. R. Soc. Lond. B* 269:1317–1323. [→ pp. 13 and 19]
- Day, T. 2002b. Virulence evolution via host exploitation and toxin production in spore-producing pathogens. *Ecol. Lett.* 5:471–476. [→ p. 17]
- Day, T., A. Graham and A. Read. 2007. Evolution of parasite virulence when host responses cause disease. *Proc. R. Soc. Lond. B* 274(1626):2685–2692. [→ pp. 4, 15, and 59]
- Day, T. and S. Gandon. (2006). Insights from Price's equation into evolutionary epidemiology, Pages 23–44 in Z. Feng, U. Dieckmann and S. A. Levin, eds. *Disease Evolution: Models, Concepts, and Data Analyses*, Vol. 71 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, American Mathematical Society, Providence, RI (USA), pp. 23–44. [→ p. 21]
- Day, T. and S. Gandon. 2012. The evolutionary epidemiology of multilocus drug resistance. *Evolution* 66(5):1582–97. [→ p. 65]
- Day, T. and S. R. Proulx. 2004. A general theory for the evolutionary dynamics of virulence. *Am. Nat.* 163(4):E40–E63. [→ pp. 21 and 65]
- Day, T., S. Alison and N. Mideo. 2011. Bridging scales in the evolution of infectious disease life histories: theory. *Evolution* 65(12):3448–61. [→ pp. 4 and 75]
- de Lamballerie, X., E. Leroy, R. N. Charrel, K. Tsetsarkin, S. Higgs and E. A. Gould. 2008. Chikungunya virus adapts to tiger mosquito via evolutionary convergence: a sign of things to come? *Virol J* 5:33. [→ p. 55]
- de Roode, J. C., A. J. Yates and S. Altizer. 2008. Virulence-transmission trade-offs and population divergence in virulence in a naturally occurring butterfly parasite. *Proc. Natl. Acad. Sci. USA* 105(21):7489–7494. [→ pp. 13, 18, and 19]
- de Roode, J. C., R. Pansini, S. J. Cheesman, M. E. H. Heilinski, S. Huijben, A. R. Wargo, A. S. Bell, B. H. K. Chan, D. Walliker and A. F. Read. 2005. Virulence and competitive ability in genetically diverse malaria infections. *Proc. Natl. Acad. Sci. USA* 102:7624–

7628. [→ p. 29]
- Denny, L. A., S. Franceschi, S. de Sanjosé, I. Heard, A. B. Moscicki and J. Palefsky. 2012. Human papillomavirus, human immunodeficiency virus and immunosuppression. *Vaccine* 30 Suppl 5:F168–74. [→ p. 74]
- Dieckmann, U. 2002. Adaptive dynamics of pathogen-host interactions. Pages 39–59 in U. Dieckmann, J. A. J. Metz, M. W. Sabelis and K. Sigmund, eds. *Adaptive dynamics of infectious diseases: In pursuit of virulence management*. Cambridge University Press. [→ p. 23]
- Diekmann, O. and J. Heesterbeek. 2000. *Mathematical epidemiology of infectious diseases: model building, analysis, and interpretation*. Wiley, New York. [→ pp. 16, 53, and 54]
- Donoghue, H. D., M. Spigelman, C. L. Greenblatt, G. Lev-Maor, G. K. Bar-Gal, C. Matheson, K. Vernon, A. G. Nerlich and A. R. Zink. 2004. Tuberculosis: from prehistory to Robert Koch, as revealed by ancient DNA. *Lancet. Infect. Dis.* 4(9):584–592. [→ p. 11]
- Doumayrou, J., A. Avellan, R. Froissart and Y. Michalakakis. 2013. An experimental test of the transmission-virulence trade-off hypothesis in a plant virus. *Evolution* 67(2):477–486. [→ p. 19]
- Downes, S. M. 2009. The Basic Components of the Human Mind Were Not Solidified During the Pleistocene Epoch. *Contemporary debates in philosophy of biology* 12:243. [→ p. 62]
- Drummond, A. J., A. Rambaut, B. Shapiro and O. G. Pybus. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22(5):1185–92. [→ p. 67]
- Duneau, D. and D. Ebert. 2012. Host sexual dimorphism and parasite adaptation. *PLoS Biol.* 10:e1001271. [→ p. 60]
- Durrett, R. 2007. *Random graph dynamics*. Cambridge University Press. [→ p. 32]
- Dwyer, G., S. A. Levin and L. Buttel. 1990. A simulation model of the population dynamics and evolution of myxomatosis. *Ecological Monographs* 60(4):423–447. [→ p. 12]
- Eames, K. T. D., J. M. Read and W. J. Edmunds. 2009. Epidemic prediction and control in weighted networks. *Epidemics* 1(1):70–76. [→ p. 32]
- Ebert, D. and E. A. Herre. 1996. The evolution of parasitic diseases. *Parasitol. Today* 12(3):96–101. [→ p. 11]
- Ewald, P. W. 1983. Host-parasite relations, vectors, and the evolution of disease severity. *Annu. Rev. Ecol. Evol. Syst.* 14:465–485. [→ p. 12]
- Ewald, P. W. 1994. *Evolution of Infectious Disease*. Oxford University Press, Oxford. [→ p. 11]
- Fellay, J., K. V. Shianna, D. Ge, S. Colombo, B. Ledgerber, M. Weale, K. Zhang, C. Gumbs, A. Castagna, A. Cossarizza, A. Cozzi-Lepri, A. De Luca, P. Eastbrook, P. Francioli, S. Mallal, J. Martinez-Picado, J. M. Miro, N. Obel, J. P. Smith, J. Wyniger, P. Descombes, S. E. Antonarakis, N. L. Letvin, A. J. McMichael, B. F. Haynes, A. Telenti and D. B. Goldstein. 2007. A whole-genome association study of major determinants for host control of HIV-1. *Science* 317(5840):944–947. [→ pp. 46, 66, and 67]
- Felsenstein, J. 2004. *Inferring phylogenies*. Sinauer Associates, Inc., Sunderland, MA, USA. [→ p. 47]
- Frank, S. A. 1996. Models of parasite virulence. *Q. Rev. Biol.* 71:37–78. [→ p. 11]
- Fraser, C., T. D. Hollingsworth, R. Chapman, F. de Wolf and W. P. Hanage. 2007. Variation in HIV-1 set-point viral load: epidemiological analysis and an evolutionary hypothesis. *Proc. Natl. Acad. Sci. USA* 104(44):17441–17446. [→ pp. 13, 18, 19, 21, and 47]
- Freckleton, R. P., P. H. Harvey and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. *Am. Nat.* 160(6):712–726. [→ p. 47]
- Froissart, R., J. Doumayrou, F. Vuillaume, S. Alizon and Y. Michalakakis. 2010. The virulence-transmission trade-off in vector-borne plant viruses: a review of (non-)existing studies. *Philos. Trans. R. Soc. B* 365(1548):1907–18. [→ p. 17]
- Fumagalli, M., M. Sironi, U. Pozzoli, A. Ferrer-Admetlla, A. Ferrer-Admetlla, L. Pattini and R. Nielsen. 2011. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genet* 7(11):e1002355. [→ p. 63]
- Gandon, S. 1998. The curse of the pharaoh. *Proc. R. Soc. Lond. B* 265:1545–1552. [→ p. 17]
- Gandon, S. 2004. Evolution of multihost parasites. *Evolution* 58(3):455–469. [→ p. 60]
- Gandon, S. and T. Day. 2008. Evidences of parasite evolution after vaccination. *Vaccine* 26S:C4–7. [→ p. 74]

- Gandon, S. and Y. Michalakis. 2000. Evolution of parasite virulence against qualitative or quantitative host resistance. *Proc. R. Soc. Lond. B* 267:985–990. [→ p. 59]
- Ganusov, V. V. and R. Antia. 2003. Trade-offs and the evolution of virulence of microparasites: do details matter? *Theor. Popul. Biol.* 64:211–220. [→ p. 9]
- Ganusov, V. V. and R. Antia. 2006. Imperfect vaccines and the evolution of pathogens causing acute infections in vertebrates. *Evolution* 60(5):957–969. [→ p. 74]
- García-Vallvé, S., A. Alonso and I. G. Bravo. 2005. Papillomaviruses: different genes have different histories. *Trends Microbiol* 13(11):514–21. [→ p. 74]
- Garnett, G. P., K. J. Mertz, L. Finelli, W. C. Levine and M. E. St Louis. 1999. The transmission dynamics of gonorrhoea: modelling the reported behaviour of infected patients from Newark, New Jersey. *Philos. Trans. R. Soc. Lond. B* 354(1384):787–97. [→ p. 73]
- Ge, D., J. Fellay, A. J. Thompson, J. S. Simon, K. V. Shianna, T. J. Urban, E. L. Heinzen, P. Qiu, A. H. Bertelsen, A. J. Muir, M. Sulkowski, J. G. McHutchison and D. B. Goldstein. 2009. Genetic variation in *IL28B* predicts hepatitis C treatment-induced viral clearance. *Nature* 461(7262):399–401. [→ p. 49]
- Geritz, S. A. H., E. Kisdi, G. Meszéna and J. A. J. Metz. 1998. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol. Ecol.* 4:1–79. [→ p. 40]
- Gilchrist, M. A. and A. Sasaki. 2002. Modeling host-parasite coevolution: a nested approach based on mechanistic models. *J. theor. Biol.* 218:289–308. [→ pp. 9 and 15]
- Gonzalez, A., O. Ronce, R. Ferriere and M. E. Hochberg. 2013. Evolutionary rescue: an emerging focus at the intersection between ecology and evolution. *Philos. Trans. R. Soc. Lond. B.* [→ p. 54]
- Graham, A. L., J. E. Allen and A. F. Read. 2005. Evolutionary causes and consequences of immunopathology. *Annu. Rev. Ecol. Evol. Syst.* 36:373–397. [→ p. 15]
- Gray, R. R., J. Parker, P. Lemey, M. Salemi, A. Katzourakis and O. G. Pybus. 2011. The mode and tempo of hepatitis C virus evolution within and among hosts. *BMC Evol Biol* 11:131. [→ p. 44]
- Grebely, J., M. Prins, M. Hellard, A. L. Cox, W. O. Osburn, G. Lauer, K. Page, A. R. Lloyd and G. J. Dore. 2012. Hepatitis C virus clearance, reinfection, and persistence, with insights from studies of injecting drug users: towards a vaccine. *Lancet Infect. Dis.* 12(5):408–414. [→ p. 49]
- Griffiths, R. C. and S. Tavaré. 1994. Sampling theory for neutral alleles in a varying environment. *Philos Trans R Soc Lond B* 344(1310):403–410. [→ p. 67]
- Hamilton, W. 1964. The genetical evolution of social behaviour (I & II). *J. Theor. Biol.* 7(1):1–52. [→ p. 15]
- Handel, A., E. Margolis and B. R. Levin. 2009. Exploring the role of the immune response in preventing antibiotic resistance. *J. theor. Biol.* 256(4):655–62. [→ p. 71]
- Handel, A., J. Brown, D. Stallknecht and P. Rohani. 2013. A multi-scale analysis of influenza A virus fitness trade-offs due to temperature-dependent virus persistence. *PLoS Comput Biol* 9(3):e1002989. [→ p. 9]
- Hartfield, M. and S. Alizon. 2013. Introducing the outbreak threshold in epidemiology. *PLoS Pathog.* 6:e1003277. [→ pp. 7 and 52]
- Herbeck, J. T., M. Rolland, Y. Liu, S. McLaughlin, J. McNevin, H. Zhao, K. Wong, J. N. Stoddard, D. Raugi, S. Sorensen, I. Genowati, B. Birditt, A. McKay, K. Diem, B. S. Maust, W. Deng, A. C. Collier, J. D. Stekler, M. J. McElrath and J. I. Mullins. 2011. Demographic processes affect HIV-1 evolution in primary infection before the onset of selective processes. *J Virol* 85(15):7523–34. [→ p. 45]
- Herfst, S., E. J. A. Schrauwen, M. Linster, S. Chutinimitkul, E. de Wit, V. J. Munster, E. M. Sorrell, T. M. Bestebroer, D. F. Burke, D. J. Smith, G. F. Rimmelzwaan, A. D. M. E. Osterhaus and R. A. M. Fouchier. 2012. Airborne transmission of influenza A/H5N1 virus between ferrets. *Science* 336(6088):1534–41. [→ p. 64]
- Hernandez, R. D., J. L. Kelley, E. Elyashiv, S. C. Melton, A. Auton, G. McVean, 1000 Genomes Project, G. Sella and M. Przeworski. 2011. Classic selective sweeps were rare in recent human evolution. *Science* 331(6019):920–4. [→ p. 63]
- Herre, E. A. 1995. Factors affecting the evolution of virulence: nematode parasites of fig wasps as a case study. *Parasitology* 111(S1):S179–S191. [→ p. 11]
- Hethcote, H. W. and J. A. Yorke. 1984. Gonorrhoea transmission dynamics and control. *Lecture Notes in Biomathematics*. Vol. 56, Springer Berlin. [→ p. 73]
- Hinkley, T., J. Martins, C. Chappey, M. Haddad, E. Stawiski, J. M. Whitcomb, C. J. Petropoulos and

- S. Bonhoeffer. 2011. A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase. *Nat Genet* 43(5):487–9. [→ p. 65]
- Hollingsworth, T. D., O. Laeyendecker, G. Shirreff, C. A. Donnelly, D. Serwadda, M. J. Wawer, N. Kiwanuka, F. Nalugoda, A. Collinson-Streng, V. Ssempijja, W. P. Hanage, T. C. Quinn, R. H. Gray and C. Fraser. 2010. HIV-1 transmitting couples have similar viral load set-points in Rakai, Uganda. *PLoS Pathog* 6(5):e1000876. [→ pp. 21 and 48]
- Hollingsworth, T. D., R. M. Anderson and C. Fraser. 2008. HIV-1 transmission, by stage of infection. *J. Infect. Dis.* 198(5):687–693. [→ p. 20]
- Housworth, E. A., E. P. Martins and M. Lynch. 2004. The phylogenetic mixed model. *Am. Nat.* 163(1):84–96. [→ p. 47]
- Hué, S., J. P. Clewley, P. A. Cane and D. Pillay. 2004. HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. *AIDS* 18(5):719–728. [→ p. 47]
- Hurford, A., D. Cownden and T. Day. 2010. Next-generation tools for evolutionary invasion analyses. *J. R. Soc. Interface* 7(45):561–71. [→ p. 16]
- Imai, M., T. Watanabe, M. Hatta, S. C. Das, M. Ozawa, K. Shinya, G. Zhong, A. Hanson, H. Katsura, S. Watanabe, C. Li, E. Kawakami, S. Yamada, M. Kiso, Y. Suzuki, E. A. Maher, G. Neumann and Y. Kawaoka. 2012. Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature* 486(7403):420–8. [→ p. 64]
- Iwasa, Y., F. Michor and M. A. Nowak. 2003. Evolutionary dynamics of escape from biomedical intervention. *Proc. R. Soc. Lond. B* 270(1533):2573–2578. [→ p. 54]
- Iwasa, Y., F. Michor and M. A. Nowak. 2004. Evolutionary dynamics of invasion and escape. *J. Theor. Biol.* 226:205–214. [→ p. 54]
- Jensen, K. H., T. Little, A. Skorping and D. V. Ebert. 2006. Empirical support for optimal virulence in a castrating parasite. *PLoS Biol.* 4(7):e197. [→ p. 19]
- Johnson, A. M., C. H. Mercer, B. Erens, A. J. Copas, S. McManus, K. Wellings, K. A. Fenton, C. Korovessis, W. Macdowall, K. Nanchahal, S. Purdon and J. Field. 2001. Sexual behaviour in Britain: partnerships, practices, and HIV risk behaviours. *Lancet* 358(9296):1835–42. [→ pp. 32, 35, 36, and 73]
- Kamp, C. 2010. Untangling the Interplay between Epidemic Spread and Transmission Network Dynamics. *PLoS Comput Biol* 6(11):e1000984. [→ p. 34]
- Kamp, C., M. Moslonka-Lefebvre and S. Alizon. 2013. Approximating epidemic spread on weighted networks. *PLoS Comput. Biol.* [→ pp. 5, 34, and 73]
- Keele, B. F., E. E. Giorgi, J. F. Salazar-Gonzalez, J. M. Decker, K. T. Pham, M. G. Salazar, C. Sun, T. Grayson, S. Wang, H. Li, X. Wei, C. Jiang, J. L. Kirchherr, F. Gao, J. A. Anderson, L.-H. Ping, R. Swanstrom, G. D. Tomaras, W. A. Blattner, P. A. Goepfert, J. M. Kilby, M. S. Saag, E. L. Delwart, M. P. Busch, M. S. Cohen, D. C. Montefiori, B. F. Haynes, B. Gaschen, G. S. Athreya, H. Y. Lee, N. Wood, C. Seoighe, A. S. Perelson, T. Bhattacharya, B. T. Korber, B. H. Hahn and G. M. Shaw. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. USA* 105(21):7552–7557. [→ p. 28]
- King, A. A., S. Shrestha, E. T. Harvill and O. N. Bjørnstad. 2009. Evolution of acute infections and the invasion-persistence trade-off. *Am Nat* 173(4):446–55. [→ p. 23]
- Kingman, J. F. C. 1982. The coalescent. *Stochastic processes and their applications* 13(3):235–248. [→ p. 67]
- Kouyos, R. D., D. Fouchet and S. Bonhoeffer. 2009. Recombination and drug resistance in HIV: population dynamics and stochasticity. *Epidemics* 1(1):58–69. [→ p. 64]
- Kouyos, R. D., V. von Wyl, S. Yerly, J. Böni, P. Taffé, C. Shah, P. Bürgisser, T. Klimkait, R. Weber, B. Hirschel, M. Cavassini, H. Furrer, M. Battegay, P. L. Vernazza, E. Bernasconi, M. Rickenbach, B. Ledergerber, S. Bonhoeffer and H. F. Günthard. 2010. Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. *J. Infect. Dis.* 201(10):1488–97. [→ p. 46]
- Kubiak, R. J., N. Arinaminpathy and A. R. McLean. 2010. Insights into the evolution and emergence of a novel infectious disease. *PLoS Comput Biol* 6(9):e1000947. [→ p. 54]
- Leitner, T., D. Escanilla, C. Franzen, M. Uhlen and J. Albert. 1996. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. *Proc. Natl. Acad. Sci. USA* 93(20):10864–10869. [→ p. 47]
- Lemey, P., A. Rambaut and O. G. Pybus. 2006. HIV evolutionary dynamics within and among hosts. *AIDS*

- Rev. 8(3):125–140. [→ p. 43]
- Lemey, P., A. Rambaut, A. J. Drummond and M. A. Suchard. 2009. Bayesian phylogeography finds its roots. *PLoS Comput Biol* 5(9):e1000520. [→ p. 46]
- Leroy, E. M., A. Epelboin, V. Mondonge, X. Pourrut, J.-P. Gonzalez, J.-J. Muyembe-Tamfum and P. Formenty. 2009. Human Ebola outbreak resulting from direct exposure to fruit bats in Luebo, Democratic Republic of Congo, 2007. *Vector Borne Zoonotic Dis* 9(6):723–8. [→ p. 51]
- Leventhal, G. E., R. Kouyos, T. Stadler, V. v. Wyl, S. Yerly, J. Böni, C. Cellera, T. Klimkait, H. F. Günthard and S. Bonhoeffer. 2012. Inferring epidemic contact structure from phylogenetic trees. *PLoS Comput Biol* 8(3):e1002413. [→ p. 4]
- Levin, B. R. 1996. The evolution and maintenance of virulence in microparasites. *Emerg Infect Dis* 2(2):93–102. [→ pp. 11 and 20]
- Levin, B. R. and C. Svanborg Edén. 1990. Selection and evolution of virulence in bacteria: an ecumenical excursion and modest suggestion. *Parasitology* 100:S103–15. [→ p. 20]
- Levin, B. R. and J. J. Bull. 1994. Short-sighted evolution and the virulence of pathogenic microorganisms. *Trends Microbiol.* 2(3):76–81. [→ pp. 13, 20, and 37]
- Levy, J. A. 1988. Mysteries of HIV: challenges for therapy and prevention. *Nature* 333(6173):519–22. [→ p. 42]
- Lewontin, R. C. 1970. The Units of Selection. *Annu. Rev. Ecol. Syst.* 1(1):1–18. [→ p. 13]
- Lewontin, R. C., S. Rose and L. J. Kamin. 1984. Not in our genes: biology, ideology, and human nature. Pantheon Books. [→ p. 61]
- Lion, S. and M. Boots. 2010. Are parasites ‘prudent’ in space? *Ecol. Lett.* 13(10):1245–1255. [→ p. 31]
- Lion, S. and M. van Baalen. 2008. Self-structuring in spatial evolutionary ecology. *Ecol. Lett.* 11(3):277–295. [→ pp. 16 and 31]
- Lipsitch, M. and E. R. Moxon. 1997. Virulence and transmissibility of pathogens: what is the relationship? *Trends Microbiol.* 5:31–37. [→ p. 11]
- Little, T. J., D. M. Shuker, N. Colegrave, T. Day and A. L. Graham. 2010. The coevolution of virulence: tolerance in perspective. *PLoS Pathog.* 6(9):e1001006. [→ p. 59]
- Liu, Y., J. McNevin, J. Cao, H. Zhao, I. Genowati, K. Wong, S. McLaughlin, M. D. McSweyn, K. Diem, C. E. Stevens, J. Maenza, H. He, D. C. Nickle, D. Shriner, S. E. Holte, A. C. Collier, L. Corey, M. J. McElrath and J. I. Mullins. 2006. Selection on the human immunodeficiency virus type 1 proteome following primary infection. *J Virol* 80(19):9519–29. [→ p. 45]
- Lloyd-Smith, J. O., M. Poss and B. T. Grenfell. 2008. HIV-1/parasite co-infection and the emergence of new parasite strains. *Parasitology* 135(7):795–806. [→ p. 63]
- Lloyd-Smith, J. O., S. J. Schreiber, P. E. Kopp and W. M. Getz. 2005. Superspreading and the effect of individual variation on disease emergence. *Nature* 438(7066):355–9. [→ pp. 51, 53, and 56]
- Long, G. H., B. H. K. Chan, J. E. Allen, A. F. Read and A. L. Graham. 2008. Experimental manipulation of immune-mediated disease and its fitness costs for rodent malaria parasites. *BMC Evol. Biol.* 8:128. [→ p. 15]
- Loverdo, C., M. Park, S. J. Schreiber and J. O. Lloyd-Smith. 2012. Influence of viral replication mechanisms on within-host evolutionary dynamics. *Evolution* 66(11):3462–71. [→ p. 71]
- Luciani, F. and S. Alizon. 2009. The evolutionary dynamics of a rapidly mutating virus within and between hosts: the case of hepatitis C virus. *PLoS Comput. Biol.* 5(11):e1000565. [→ pp. 37, 38, and 39]
- Lynch, M. and B. Walsh. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Inc., Sunderland, MA, USA. [→ p. 47]
- Lythgoe, K. and C. Fraser. 2012. New insights into the evolutionary rate of HIV-1 at the within-host and epidemiological levels. *Proc. R. Soc. Lond. B* 279(1741):3367–3375. [→ pp. 44 and 45]
- Lythgoe, K., L. Pellis and C. Fraser. 2013. Is HIV short-sighted? Insights from a multi-strain nested model. *Evolution*. [→ p. 21]
- Margolis, E. and B. R. Levin. 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(7):1068–1075. [→ p. 14]
- Matzinger, P. 2002. The danger model: a renewed sense of self. *Science* 296(5566):301–5. [→ p. 71]
- May, R. M. and R. M. Anderson. 1983. Parasite-host coevolution. Pages 186–206 in D. J. Futuyama and M. Slatkin, eds. *Coevolution*. Sinauer, Sunderland,

- MA. [→ p. 11]
- May, R. M. and R. M. Anderson. 1987. Transmission dynamics of HIV infection. *Nature* 326(6109):137–42. [→ p. 31]
- Mellors, J. W., C. R. J. Rinaldo, P. Gupta, R. M. White, J. A. Todd and L. A. Kingsley. 1996. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science* 272(5265):1167–1170. [→ p. 47]
- Mellors, J. W., J. B. Margolick, J. P. Phair, C. R. Rinaldo, R. Detels, L. P. Jacobson and A. Muñoz. 2007. Prognostic value of HIV-1 RNA, CD4 cell count, and CD4 Cell count slope for progression to AIDS and death in untreated HIV-1 infection. *JAMA* 297(21):2349–50. [→ p. 47]
- Méthot, P.-O. 2009. Darwin et la médecine: Intérêt et limites des explications évolutionnaires en médecine. Pages 657–684 in T. Heams, P. Huneman, G. Lecointre and M. Silberstein, eds. *Les mondes darwiniens: L'évolution de l'évolution*. Syllepses, Paris. [→ pp. 62 and 63]
- Méthot, P.-O. 2011. Research traditions and evolutionary explanations in medicine. *Theor Med Bioeth* 32(1):75–90. [→ p. 62]
- Méthot, P.-O. 2012. Why do Parasites Harm Their Host? On the Origin and Legacy of Theobald Smith's "Law of Declining Virulence" — 1900-1980. *Hist. Phil. Life Sci.* 34:561–601. [→ p. 11]
- Méthot, P.-O. and S. Alizon. 2013. Emerging diseases and the evolution of virulence: the case of the Spanish influenza pandemic. in G. Lambert, P. Huneman and M. Silberstein, eds. *Disease, Classification, and Evidence*. *New Essays in the Philosophy of Medicine*. Vol. in press, Springer. [→ p. 8]
- Metz, J. A. J., S. D. Mylius and O. Diekmann. 1996. When does evolution optimise? On the relation between types of density-dependence and evolutionarily stable life-history parameters. *International Institute for Applied Systems Analysis*. [→ p. 16]
- Mideo, N. and T. Day. 2008. The evolution of reproductive restraint in malaria. *Proc. R. Soc. Lond. B* 275(1639):1217–24. [→ p. 17]
- Mideo, N., S. Alizon and T. Day. 2008. Linking within and between-host dynamics in the evolutionary epidemiology of infectious diseases. *Trends Ecol. Evol.* 23(9):511–517. [→ pp. 4, 9, 27, and 75]
- Mideo, N., W. A. Nelson, S. E. Reece, A. S. Bell, A. F. Read and T. Day. 2011. Bridging scales in the evolution of infectious disease life histories: application. *Evolution* 65(11):3298–310. [→ p. 75]
- Miller, M. R., A. White and M. Boots. 2006. The evolution of parasites in response to tolerance in their hosts: the good, the bad, and apparent commensalism. *Evolution* 60(5):945–56. [→ p. 59]
- Moore, J. P., S. G. Kitchen, P. Pugach and J. A. Zack. 2004. The CCR5 and CXCR4 coreceptors—central to understanding the transmission and pathogenesis of Human Immunodeficiency Virus type 1 infection. *AIDS Res. Hum. Retroviruses* 20(1):111–26. [→ p. 40]
- Moret, Y. and P. Schmid-Hempel. 2000. Survival for immunity: The price of immune system activation for bumblebee workers. *Science* 290:1166–1168. [→ p. 15]
- Moslonka-Lefebvre, M., S. Bonhoeffer and S. Alizon. 2012a. Weighting for sex acts to understand the spread of STI on networks. *J. theor. Biol.* 311:46–53. [→ pp. 4, 32, 33, 35, and 73]
- Moslonka-Lefebvre, M., T. Harwood, M. J. Jeger and M. Pautasso. 2012b. SIS along a continuum (SISc) epidemiological modelling and control of diseases on directed trade networks. *Math. Biosci.* 236(1):44–52. [→ p. 31]
- Mostowy, R., R. D. Kouyos, D. Fouchet and S. Bonhoeffer. 2011. The role of recombination for the coevolutionary dynamics of HIV and the immune response. *PLoS One* 6(2):e16052. [→ p. 64]
- Müller, V., C. Fraser and J. T. Herbeck. 2011. A Strong Case for Viral Genetic Factors in HIV Virulence. *Viruses* 3:204–216. [→ p. 48]
- Murall, C. L., K. S. McCann and C. T. Bauch. 2012. Food webs in the human body: linking ecological theory to viral dynamics. *PLoS One* 7(11):e48812. [→ p. 74]
- Nesse, R. M. and S. C. Stearns. 2008. The great opportunity: evolutionary applications to medicine and public health. *Evolutionary Applications* 1(1):28–48. [→ p. 62]
- Newman, M. E. J. 2002. Spread of epidemic disease on networks. *Phys. Rev. E* 66(1):16128. [→ pp. 31 and 32]
- Nielsen, R., I. Hellmann, M. Hubisz, C. Bustamante and A. G. Clark. 2007. Recent and ongoing selection in the human genome. *Nat Rev Genet* 8(11):857–68. [→ pp. 61 and 63]
- Nkhoma, S. C., S. Nair, I. H. Cheeseman, C. Rohr-Allegrini, S. Singlam, F. Nosten and T. J. C. Anderson.

2012. Close kinship within multiple-genotype malaria parasite infections. *Proc Biol Sci* 279(1738):2589–98. [→ p. 28]
- Nordvik, M. K. and F. Liljeros. 2006. Number of Sexual Encounters Involving Intercourse and the Transmission of Sexually Transmitted Infections. *Sex. Transm. Dis.* 33(6):342–349. [→ pp. 32 and 73]
- Nowak, M. A., R. M. May and R. M. Anderson. 1990. The evolutionary dynamics of HIV-1 quasispecies and the development of immunodeficiency disease. *AIDS* 4:1095–1103. [→ p. 43]
- Osnas, E. E. and A. P. Dobson. 2011. Evolution of virulence in heterogeneous host communities under multiple trade-offs. *Evolution* 66(2):391–401. [→ p. 60]
- Parker, J., A. Rambaut and O. G. Pybus. 2008. Correlating viral phenotypes with phylogeny: Accounting for phylogenetic uncertainty. *Infect. Genet. Evol.* 8:239–246. [→ p. 47]
- Pepin, K. M., I. Volkov, J. R. Banavar, C. O. Wilke and B. T. Grenfell. 2010. Phenotypic differences in viral immune escape explained by linking within-host dynamics to host-population immunity. *J. theor. Biol.* 265:501–510. [→ p. 9]
- Pirone, T. P. and S. Blanc. 1996. Helper-dependent vector transmission of plant viruses. *Annu Rev Phytopathol* 34:227–47. [→ p. 29]
- Poolman, E. M., E. H. Elbasha and A. P. Galvani. 2008. Vaccination and the evolutionary ecology of human papillomavirus. *Vaccine* 26 Suppl 3:C25–30. [→ p. 74]
- Poon, A. F. Y., S. L. Kosakovsky Pond, P. Bennett, D. D. Richman, A. J. Leigh Brown and S. D. W. Frost. 2007. Adaptation to human populations is revealed by within-host polymorphisms in HIV-1 and hepatitis C virus. *PLoS Pathog.* 3(3):e45. [→ p. 72]
- Pradeu, T. 2012. *The Limits of the Self: Immunology and Biological Identity*. Oxford University Press, New York, USA. [→ p. 72]
- Pradeu, T. and E. Carosella. 2006. The self model and the conception of biological identity in immunology. *Biol. Philos.* 21(2):235–252. [→ p. 72]
- Pradeu, T. and E. L. Cooper. 2012. The danger theory: 20 years later. *Front Immunol* 3:287. [→ p. 71]
- Prugnolle, F., A. Manica, M. Charpentier, J. F. Guegan, V. Guernier and F. Balloux. 2005. Pathogen-driven selection and worldwide HLA class I diversity. *Curr. Biol.* 15(11):1022–7. [→ p. 63]
- Pybus, O. G., A. J. Drummond, T. Nakano, B. H. Robertson and A. Rambaut. 2003. The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a Bayesian coalescent approach. *Mol Biol Evol* 20(3):381–7. [→ p. 67]
- Pybus, O. G., A. Rambaut and P. H. Harvey. 2000. An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* 155(3):1429–37. [→ p. 67]
- Pybus, O. G. and A. Rambaut. 2009. Evolutionary analysis of the dynamics of viral infectious disease. *Nat. Rev. Genet.* 10(8):540–550. [→ p. 43]
- Rauch, A., Z. Kutalik, P. Descombes, T. Cai, J. Di Iulio, T. Mueller, M. Bochud, M. Battegay, E. Bernasconi, J. Borovicka, S. Colombo, A. Cerny, J.-F. Dufour, H. Furrer, H. F. Günthard, M. Heim, B. Hirschel, R. Malinverni, D. Moradpour, B. Müllhaupt, A. Witteck, J. S. Beckmann, T. Berg, S. Bergmann, F. Negro, A. Telenti and P.-Y. Bochud. 2010. Genetic Variation in *IL28B* Is Associated With Chronic Hepatitis C and Treatment Failure: A Genome-Wide Association Study. *Gastroenterology* 138(4):1338–1345.e7. [→ p. 49]
- Read, A. 1994. The evolution of virulence. *Trends Microbiol.* 2(3):73–76. [→ p. 11]
- Regoes, R. R. and S. Bonhoeffer. 2005. The HIV coreceptor switch: a population dynamical perspective. *Trends Microbiol.* 13(6):269–277. [→ p. 39]
- Regoes, R. R., M. A. Nowak and S. Bonhoeffer. 2000. Evolution of virulence in a heterogeneous host population. *Evolution* 54(1):64–71. [→ p. 60]
- Renault, P., J.-L. Solet, D. Sissoko, E. Balleydier, S. Larrieu, L. Filleul, C. Lassalle, J. Thiria, E. Rachou, H. de Valk, D. Ilef, M. Ledrans, I. Quatresous, P. Quenel and V. Pierre. 2007. A Major Epidemic of Chikungunya Virus Infection on Réunion Island, France, 2005–2006. *Am. J. Trop. Med. Hyg.* 77(4):727–731. [→ pp. 54, 55, and 56]
- Ribeiro, R. M., M. D. Hazenberg, A. S. Perelson and M. P. Davenport. 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(2):802–9. [→ pp. 40 and 43]
- Ridzon, R., K. Gallagher, C. Ciesielski, M. B. Ginsberg, B. J. Robertson, C. C. Luo and A. DeMaria, Jr. 1997. Simultaneous transmission of human immunodeficiency virus and hepatitis C virus from a needle-stick injury. *N. Engl. J. Med.* 336(13):919–22.

- [→ p. 28]
- Rousset, F. and O. Ronce. 2004. Inclusive fitness for traits affecting metapopulation demography. *Theor. Popul. Biol.* 65(2):127–141. [→ p. 31]
- Råberg, L., A. L. Graham and A. F. Read. 2009. Decomposing health: tolerance and resistance to parasites in animals. *Philos. Trans. R. Soc. B* 364(1513):37–49. [→ p. 59]
- Råberg, L., D. Sim and A. F. Read. 2007. Disentangling genetic variation for resistance and tolerance to infectious diseases in animals. *Science* 318(5851):812–814. [→ pp. 59 and 60]
- Russell, C. A., J. M. Fonville, A. E. X. Brown, D. F. Burke, D. L. Smith, S. L. James, S. Herfst, S. van Boheemen, M. Linster, E. J. Schrauwen, L. Katzelnick, A. Mosterín, T. Kuiken, E. Maher, G. Neumann, A. D. M. E. Osterhaus, Y. Kawaoka, R. A. M. Fouchier and D. J. Smith. 2012. The potential for respiratory droplet-transmissible A/H5N1 influenza virus to evolve in a mammalian host. *Science* 336(6088):1541–7. [→ p. 64]
- Sabeti, P. C., E. Walsh, S. F. Schaffner, P. Varrilly, B. Fry, H. B. Hutcherson, M. Cullen, T. S. Mikkelsen, J. Roy, N. Patterson, R. Cooper, D. Reich, D. Altshuler, S. O'Brien and E. S. Lander. 2005. The case for selection at CCR5-Δ32. *PLoS Biol.* 3(11):e378. [→ p. 63]
- Sasaki, A. and Y. Iwasa. 1991. Optimal growth schedule of pathogens within a host: Switching between lytic and latent cycles. *Theor. Popul. Biol.* 39:201–239. [→ p. 9]
- Schafer, J. F. 1971. Tolerance to plant disease. *Annu Rev Phytopathol* 9(1):235–252. [→ p. 59]
- Schmid-Hempel, P. 2008. Parasite immune evasion: a momentous molecular war. *Trends Ecol. Evol.* 23(6):318–326. [→ p. 64]
- Schmid-Hempel, P. 2011. *Evolutionary Parasitology: The Integrated Study of Infections, Immunology, Ecology, and Genetics*. Oxford Univ Press, Oxford, UK. [→ pp. 11 and 15]
- Schuffenecker, I., I. Itean, A. Michault, S. Murri, L. Frangeul, M.-C. Vaney, R. Lavenir, N. Pardigon, J.-M. Reynes, F. Pettinelli, L. Biscornet, L. Diancourt, S. Michel, S. Duquerroy, G. Guigon, M.-P. Frenkiel, A.-C. Bréhin, N. Cubito, P. Desprès, F. Kunst, F. A. Rey, H. Zeller and S. Brisse. 2006. Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. *PLoS Med* 3(7):e263. [→ pp. 54 and 56]
- Schulte, R. D., C. Makus, B. Hasert, N. K. Michiels and H. Schulenburg. 2010. Multiple reciprocal adaptations and rapid genetic change upon experimental co-evolution of an animal host and its microbial parasite. *Proc Natl Acad Sci USA* 107(16):7359–64. [→ p. 65]
- Shankarappa, R., J. B. Margolick, S. J. Gange, A. G. Rodrigo, D. Upchurch, H. Farzadegan, P. Gupta, C. R. Rinaldo, G. H. Learn, X. He, X. L. Huang and J. I. Mullins. 1999. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J. Virol.* 73(12):10489–502. [→ p. 44]
- Simmonds, P. 2004. Genetic diversity and evolution of hepatitis C virus - 15 years on. *J. Gen. Virol.* 85(11):3173–3188. [→ p. 49]
- smith, j. 2011. Superinfection drives virulence evolution in experimental populations of bacteria and plasmids. *Evolution* 65:831–841. [→ p. 21]
- Smith, T. 1904. Some problems in the life-history of pathogenic microorganism. *Science* 20(520):817–832. [→ p. 11]
- Smith, T. W. 1992. Discrepancies between men and women in reporting number of sexual partners: a summary from four countries. *Soc Biol* 39(3-4):203–11. [→ p. 73]
- Sniegowski, P. D., P. J. Gerrish, T. Johnson and A. Shaver. 2000. The evolution of mutation rates: separating causes from consequences. *BioEssays* 22(12):1057–1066. [→ p. 56]
- Stadler, T., R. Kouyos, V. von Wyl, S. Yerly, J. Böni, P. Bürgisser, T. Klimkait, B. Joos, P. Rieder, D. Xie, H. F. Günthard, A. J. Drummond, S. Bonhoeffer and the Swiss HIV Cohort Study. 2012. Estimating the Basic Reproductive Number from Viral Sequence Data. *Mol Biol Evol* 29(1):347–357. [→ p. 68]
- Stearns, S. C. (1999). *Evolution in health and disease*, Oxford University Press, chapter Introducing evolutionary thinking, pp. 3–15. [→ p. 63]
- Stehlé, J., N. Voirin, A. Barrat, C. Cattuto, V. Colizza, L. Isella, C. Régis, J.-F. Pinton, N. Khanafer, W. Van den Broeck and P. Vanhems. 2011. Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. *BMC Med* 9:87. [→ p. 32]
- Sunnåker, M., A. G. Busetto, E. Numminen, J. Corander, M. Foll and C. Dessimoz. 2013. Approximate Bayesian computation. *PLoS Comput Biol* 9(1):e1002803. [→ p. 71]



- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123(3):585–95. [→ p. 42]
- Taylor, L. H., D. Walliker and A. F. Read. 1997. Mixed-genotype infections of malaria parasites: within-host dynamics and transmission success of competing clones. *Proc. R. Soc. Lond. B* 264(1383):927–35. [→ p. 28]
- Tebit, D. M., I. Nankya, E. J. Arts and Y. Gao. 2007. HIV diversity, recombination and disease progression: how does fitness "fit" into the puzzle? *AIDS Rev.* 9(2):75–87. [→ p. 42]
- van Baalen, M. 2002. Contact networks and the evolution of virulence — Implications for virulence management. Pages 85–103 in U. Dieckmann, J. A. J. Metz, M. W. Sabelis and K. Sigmund, eds. *The Adaptive Dynamics of Infectious Diseases: In Pursuit of Virulence Management*. Cambridge University Press. [→ p. 15]
- van Baalen, M. and M. W. Sabelis. 1995. The dynamics of multiple infection and the evolution of virulence. *Am. Nat.* 146:881–910. [→ pp. 12, 25, 27, and 63]
- van den Driessche, P. and J. Watmough. 2002. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math. Biosci.* 180(1-2):29–48. [→ p. 16]
- Vaughan, T. G. and A. J. Drummond. 2013. A stochastic simulator of birth-death master equations with application to phylodynamics. *Mol Biol Evol* 30(6):1480–93. [→ p. 68]
- Vazeille, M., L. Mousson, E. Martin and A.-B. Failloux. 2010. Orally co-Infected *Aedes albopictus* from La Reunion Island, Indian Ocean, can deliver both dengue and chikungunya infectious viral particles in their saliva. *PLoS Negl. Trop. Dis.* 4(6):e706. [→ p. 29]
- Vazeille, M., S. Moutailler, D. Coudrier, C. Rousseaux, H. Khun, M. Huerre, J. Thiria, J.-S. Dehecq, D. Fontenille, I. Schuffenecker, P. Despres and A.-B. Failloux. 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. [→ pp. 55 and 56]
- Visscher, P. M., W. G. Hill and N. R. Wray. 2008. Heritability in the genomics era – concepts and misconceptions. *Nat. Rev. Genet.* 9(4):255–266. [→ p. 49]
- Volz, E. 2008. SIR dynamics in random networks with heterogeneous connectivity. *J. Math. Biol.* 56(3):293–310. [→ p. 34]
- Volz, E. and L. Meyers. 2009. Epidemic thresholds in dynamic contact networks. *J. R. Soc. Interface* 6(32):233–241. [→ p. 32]
- West, S. A. and A. Buckling. 2003. Cooperation, virulence and siderophore production in bacterial parasites. *Proc. R. Soc. Lond. B* 270:37–44. [→ p. 27]
- Williams, G. C. and R. M. Nesse. 1991. The Dawn of Darwinian Medicine. *Q. Rev. Biol.* 66(1):1–22. [→ p. 62]
- Williams, P. 2012. New Insights into Virulence Evolution in Multigroup Hosts. *Am Nat* 179(2):228–239. [→ p. 60]
- Wilson, E. O. 1975. *Sociobiology: the new synthesis*. Belknap Press of Harvard University Press. [→ p. 61]
- Wolfe, N. D., C. P. Dunavan and J. Diamond. 2007. Origins of major human infectious diseases. *Nature* 447(7142):279–283. [→ p. 54]
- Xiang, J., S. Wunschmann, D. Diekema, D. Klinzman, K. Patrick, S. George and J. Stapleton. 2001. Effect of coinfection with GB virus C on survival among patients with HIV infection. *New England Journal of Medicine* 345(10):707–714. [→ p. 64]
- Yang, J., T. A. Manolio, L. R. Pasquale, E. Boerwinkle, N. Caporaso, J. M. Cunningham, M. de Andrade, B. Feenstra, E. Feingold, M. G. Hayes, W. G. Hill, M. T. Landi, A. Alonso, G. Lettre, P. Lin, H. Ling, W. Lowe, R. A. Mathias, M. Melbye, E. Pugh, M. C. Cornelis, B. S. Weir, M. E. Goddard and P. M. Visscher. 2011. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet* 43(6):519–25. [→ p. 67]
- Yates, A., R. Antia and R. R. Regoes. 2006. How do pathogen evolution and host heterogeneity interact in disease emergence? *Proc. R. Soc. Lond. B* 273(1605):3075–83. [→ pp. 51 and 54]
- Yewdell, J. W. 2007. Plumbing the sources of endogenous MHC class I peptide ligands. *Curr. Opin. Immunol.* 19(1):79–86. [→ p. 13]
- Yudell, M. and R. DeSalle. 2000. Essay Review: Sociobiology: Twenty-five years later. *J. Hist. Biol.* 33(3):577–584. [→ p. 61]
- Zimmer, C. 2003. Infectious diseases. Taming pathogens: an elegant idea, but does it work? *Science* 300(5624):1362–4. [→ p. 11]