



**HAL**  
open science

# Theory and examples of reciprocal influence between hosts and pathogens, from short-term to long term interactions: coevolution, cospeciation and pathogen speciation following host shifts

Aurélien Tellier, Damien M. de Vienne, Tatiana Giraud, Michael Hood, Guislaine Refrégier

## ► To cite this version:

Aurélien Tellier, Damien M. de Vienne, Tatiana Giraud, Michael Hood, Guislaine Refrégier. Theory and examples of reciprocal influence between hosts and pathogens, from short-term to long term interactions: coevolution, cospeciation and pathogen speciation following host shifts. Annette W. Barton. Host-Pathogen Interactions: Genetics, Immunology, and Physiology, Nova Science Publishers, Inc.; 1 edition (September 1, 2010), 2010, Immunology and Immune System Disorders, 978-1608762866. hal-02326400

**HAL Id: hal-02326400**

**<https://hal.science/hal-02326400v1>**

Submitted on 22 Oct 2019

**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.

**Title : Theory and examples of reciprocal influence between hosts and pathogens, from short-term to long term interactions: coevolution, cospeciation and pathogen speciation following host shifts**

Authors : Tellier A (3), de Vienne DM (4), Giraud T (2), Hood ME (5) and Refrégier G (1, 2).

(1) Institut de Génétique et Microbiologie, Université Paris-Sud 11, Bâtiment 400, F-91405 Orsay-Cedex, France ; CNRS -UMR8621, F-91405 Orsay cedex, France.

(2) Ecologie, Systématique et Evolution, Bâtiment 360, Université Paris-Sud, F-91405 Orsay cedex, France ; CNRS UMR 8079, F-91405 Orsay cedex, France.

(3) Evolutionary Biology, LMU Biocenter, University of Munich, Grosshaderner Str. 2, D-82152 Planegg-Martinsried, Germany.

(4) Laboratoire de Recherche en Informatique, Bâtiment 490, Université Paris-Sud, F-91405 Orsay cedex, France ; CNRS UMR 8623, F-91405 Orsay cedex, France.

(5) Department of Biology, Amherst College, Amherst, Massachusetts USA

## Abstract

Hosts and their obligate pathogens exhibit intimate physiological interactions. How such interactions impact upon evolution of each partner varies depending on the time-scale of interest. This chapter reviews theoretical advances and available data on host and pathogen evolution with regard to both short term (coevolution) and long term (cospeciation and speciation following host shifts) dynamics, and then discusses the possible links between processes acting at these two time scales.

The genetic basis of the physiological interactions between host and pathogens has been identified in some systems, showing that as few as a single locus in each species can determine the success or failure of infection. Such simple interacting mechanism of infection and disease resistance has triggered theoretical developments on how allele frequencies in one species should evolve in response to the allele frequencies in the other species. This reciprocal influence in the short term is referred to as “coevolution” (in its narrow sense). The concept states that each species, host or pathogen, acts as a selective pressure on its partner in a frequency-dependent manner. Two simple outcomes for host and pathogen allele frequencies have been suggested. The “arms race” model describes frequency dynamics where advantageous new variants go to fixation. The “trench warfare” model depicts allele frequencies oscillating dynamically in time, several host and pathogen alleles being maintained over the long term.

In this chapter, we detail the theoretical bases underlying arms race and trench warfare dynamics, and review empirical data supporting the different types of models. We also discuss recent theoretical advances that attempt to analyze more complex coevolutionary scenarios and integrate environmental effects into these simple models.

Another field of research on host-pathogen evolution is devoted to elucidating whether the partners in such associations speciate in parallel, i.e. diversify by cospeciation, or by other processes such as host shifts. We outline the approaches to compare the divergence between two species or loci and present insight gained on the long-term evolution of host-parasite associations, whether pathogenic or not. Speciation in parasites seems to regularly occur without tracking host speciation, for instance after host shifts.

We finally examine the relationship between coevolution and pathogen diversification. We summarize theoretical and experimental works showing that coevolution can foster pathogen specialization, but that more frequently these events are associated with host shifts and subsequent speciation, than with cospeciation. As a conclusion, in contrast to previous assumptions, we now have substantial evidences that coevolutionary dynamics of hosts and pathogens do not favor long-term cospeciation patterns, and that the idea that cospeciation prevails among host-parasite associations is invalid.

424 words

## *Introduction*

There is a renewed interest in the reciprocal influence between hosts and pathogenic parasites (see box 1 for a discussion on the words “parasites” versus “pathogens”). This is prompted by the need of controlling devastating diseases, of identifying or developing biopests against invasive species, and of deciphering the processes of life diversification, as parasitism is a widely spread life style (Poulin and Morand 2004).

Host-parasite interactions can occur at short time-scales, from a single parasite cycle in the case of the opportunistic infection of a host species on which the disease is not self-sustaining, to very long time-scales covering several speciation events. In this chapter, we deal with the consequences of host-parasite interactions on the evolution of each partner, and we place these consequences in the context of relevant time-scales. Traditionally, two time-scales are distinguished. The first regards reciprocal selection pressure between the host and its pathogenic parasite, leading to changes in allele frequencies across successive generations. This is referred to as “coevolution” in the narrow sense (Clayton and Moore 1997). The second time-scale encompasses several speciation events. When speciation occurs concomitantly for the hosts and their parasites, it is referred to as “cospeciation” (Page 2003). Alternatively, speciation in the parasite may occur without speciation of the host, as often results from host-shifts. “Coevolution” is used by some authors to describe long-term dynamics of cospeciation, but this can be misleading as we will see, and we will rather use the term in its narrow sense, i.e. reciprocal selection pressure and micro-evolutionary changes.

Identification of the molecular basis of host-pathogen interactions is not the focus of this chapter. It is however a prerequisite on which most studies on coevolution rely: molecular interactions do not necessarily lead to reciprocal selection, *i.e.* to coevolution, but provide a tractable model for what is expected if selection does occur. The first major achievement on this topic came from the work of Flor on flax and its associated rust disease (see Loegering (1987) for a review of Flor's work). By crossing resistant and susceptible flax cultivars, Flor determined that the host allele conferring resistance to one race of the pathogen was different from the allele responsible for resistance to another pathogen race. Similarly, he made a genetic analysis of factors in the rust pathogen *Melampsora lini* that determined the nature of the interaction with the flax cultivars (Flor 1956). He concluded: "These facts suggest that the infectious range of each physiological race of the pathogen is conditioned by *pairs of factors* that are specific for each different resistant or immune factor possessed by the host variety." This concept was thereafter referred to as the "Gene-For-Gene" (GFG) interaction (Loegering and Ellingboe 1987): the protein encoded by the so-called “avirulence” locus (*AVR*) in the pathogen can activate the product of the resistance locus (*RES*) in the host, which prevents the establishment of a harmful infection; if the interaction between the pathogen's “avirulence” locus and

the host's resistance locus does not occur because of a mutation in one partner, the infection proceeds (see box 2). For interactions between animals and their pathogens, a similar model called "Matching Allele" (MA) (Frank 1992) considers a specific molecular interaction between host and pathogen gene products, but the recognition is not by the host to resist infection but rather recognition by the pathogen is necessary for infection of the host (see Table 1). This last model is thought to better fit interactions found between animal hosts and their pathogens (Little et al. 2006).

Haldane (Haldane 1954) noted that in such a framework of molecular recognition, rare alleles should be selected for in the organism that benefits from recognition avoidance. For instance under a GFG interaction, the selection pressure acting on hosts favors resistance against the most common pathogen allele (referred to as *avr* because in that condition, the pathogen is avirulent). This leads to an advantage for the pathogen carrying a mutation that confers infectivity (the *avr*<sup>-</sup> sometimes also written *vir* allele) so long as it remains rare. Similarly under the MA model, the host benefits from a rare allele that prevents recognition by the pathogen. As a consequence, selection acting on a specific allele should vary according to its frequency, which was termed "Frequency Dependent Selection" (FDS). Haldane (Haldane 1954) developed the first population genetic models for testing how allelic frequencies should evolve across generations under FDS, in particular inferring how allele frequencies should evolve among host-pathogen associations. Van Valen (Van Valen 1973) pointed that these systems required a continued evolution for each partner to survive. He referred to such dynamics as "Red Queen" dynamics in reference to Lewis Carroll's tale "Beyond the mirror": the Red Queen character explains to Alice that in her world, 'it takes all the running you can do, to keep in the same place'. This comparison highlights how important coevolution is for the long-term survival of each species involved. Note that this paradigm has had far-reaching implications in other areas of biology, such as the advantages of sexual versus clonal reproduction in presence of parasitism (Hamilton et al. 1990), referred to as "the Red Queen Hypothesis for the maintenance of sex."

Further development of Haldane's models led to two expected scenarios for allele frequency evolution among hosts and pathogens, depending on whether or not the alleles under selection should reach fixation at the population level. In the "arms race" model, new variants are advantageous and have high selective coefficients so that directional selection drives them to fixation one after the other. In the "trench warfare" model, alleles are only advantageous as long as they are rare, so that their frequency varies periodically; the corresponding selection has been depicted as "balancing selection" or "fluctuating selection", and the corresponding variation as a "balanced polymorphism". The arms race and trench warfare models are now confronted by real data covering time ranges from one hundred to several thousand generations (Wichman et al. 2005; Gandon et al. 2008) and theoretically enriched with considerations on metapopulation structure as explained in section 1. Note that all these models focus on specialized pathogens, i.e. pathogens having the ability to infect a single host species. This assumption makes sense as specialization is far more common than

generalism in instances as diverse as phytophagous insects (Dres and Mallet 2002), fungal pathogens (Giraud et al. 2008), avian parasites (Proctor and Owens 2000). Generalism is however more common among plant viruses (Garcia-Arenal et al. 2003). The factors favoring specialization are discussed in section 3.2.

Following the consideration of coevolutionary dynamics, the second time-scale we will address for host-pathogen interactions covers the long-term processes of speciation. Note that in that case, no distinction seems necessary between pathogenic and non-pathogenic parasites (see Box 1) so that we will more broadly speak of host-parasite interactions. The often obligate and specialized interactions of hosts and parasites may suggest that, when the host lineage experiences a bifurcation, its associated parasites will simultaneously become isolated. Speciation in one lineage then tracks speciation in the other, which is called cospeciation. On the contrary, new host-parasite combinations may arise, for instance by parasite specialization onto a novel host species and subsequent speciation, which is often called host shift. The idea of cospeciation has been promoted by pioneering work on avian parasites such as those of Kellogg (Kellogg 1913) and Fahrenholz (Fahrenholz 1913) at the beginning of the twentieth century. Those authors noted that closely related avian parasites, as evidenced from similar phenotypic features, were harbored by closely related host species. The narrow host distribution of parasites led researchers such as Eichler, Rothschild, Clay, etc. (Hoberg et al. 1997) to hypothesize that parasites could be used as a character to infer phylogenetic relationships among hosts taxa. Similar hypotheses were proposed for plant pathogens (Savile 1979). Obtaining extra phylogenetic information was important at a time where sequences were not available. Reciprocally, host taxa were often used as taxonomic criteria for parasites classification (see for instance Downey (1962)). In both cases, if one has been used to infer the phylogeny of the other, the two phylogenies will be congruent, but this is a circular argument. Consequently, some authors concluded without strong evidence that a process of parallel divergence has occurred, i.e. cospeciation between hosts and parasites (Hoberg et al. 1997). This process was made popular through the Fahrenholz rule “parasites phylogeny mirrors that of its host” (1913). Other evolutionary process such as host shifts, parasite duplication, extinctions (see Part 2 for details on these events) were considered less seriously. Evidences for cospeciation in host-parasite associations were therefore long inappropriate. It is not until the late 1980's that robust phylogenies built independently for both hosts and parasites were used to specifically test for cospeciation patterns.

A third goal of the current chapter is to examine the relationship between coevolution and pathogen diversification. It is often assumed that short-term coevolution should lead to cospeciation on the long-term, although the rationale underlying this idea is never fully articulated. In fact, many recent studies that compare host and pathogen phylogenies, as well as theoretical developments on the parameters controlling specialization and speciation, seem to invalidate this idea.

This chapter is therefore divided in three parts. We first review knowledge on coevolution linking both theory and data (Part 1) and the state of the art on cospeciation research describing methods and case studies (Part 2). We then review the models and evidence suggesting that coevolution can foster pathogen specialization and speciation, but that such events are more often associated with host shifts and subsequent speciation, than with cospeciation (Part 3).

We will recurrently use plant-pathogen systems as case examples both when dealing with coevolution and cospeciation, for multiple reasons. First, population genetics studies at the molecular level can build upon the well-known genetics and functional data available in plant-pathogen interactions (Dangl and Jones 2001; Jones and Dangl 2006). A second advantage of plant systems is the possibility of linking molecular sequence data with phenotypic data. This occurs for example when testing the outcome of infection for different alleles of resistance genes (Rose et al. 2005). Thirdly, plants are good models for studying coevolution in natural ecosystems with various types of pathogens (bacteria, virus, fungi, nematodes). Finally, lessons can be learned from crop systems where humans impose strong evolutionary constraints on the hosts as well as on the pathogens. Despite evident limitations, agricultural systems provide useful insights on coevolution because extensive datasets reporting genetic and phenotypic diversities are collected at various spatial and temporal scales, and the molecular function of some resistance genes is well known.

**BOX 1 - Definitions: Parasite vs pathogen, virulence vs infectivity**

"**Parasite**" comes from the Greek *para sitos*, which originally was applied to a person who ate freely at the table of someone else. The term still most often is defined by a nutritional relationship, such as in the Oxford English Dictionary: "**2. a. Biol.** An organism that lives on, in, or with an organism of another species, obtaining food, shelter, or other benefit; (now) *spec.* one that obtains nutrients at the expense of the host organism, which it may directly or indirectly harm. The term *parasite* originally included (and is still sometimes used for) organisms that are now considered to be commensals, mutualists, epiphytes, or saprophytes, as well as birds or other animals that habitually steal food from, or use the nests of, other species."

In the case of mutualism, the parasites lives in or on the host, takes nutrients from it, but also provides advantages to its host, such as the synthesis of specific metabolic compounds, help in pollination for plants, etc. The main use of *parasite* still deals with organisms harming their host. However, there is a continuum between parasitism and symbiosis such that disentangling between the two is not always trivial and in many cases can be context dependent, for instance in grass endophytes (Müller and Krauss 2005).

"**Pathogen**" comes from the Greek *pathos* meaning suffering or disease. It is this concept of causing harm that always characterizes a "pathogen" but is not an essential trait of a "parasite" in the broad

sense. Pathogens do not include predators, herbivores, or allergens such as pollens; and an essential characteristic to the “pathogen” is an infection process of growth and development in or on the host. Thus, pathogens are all parasites where growth is made possible by nutrients taken from the host. Interestingly however, pathogens do not always cause harm because of the nutrient exploitation (i.e. their parasitic status), but rather harm can result from the host's response to infection, which then makes the relationship pathogenic. This is, for example, the case for plant diseases with the Hypersensitive Response (e. g. rapid cell death induced by the host in response to infection) as well as for some animal diseases (e.g. excessive fever during malaria infection can cause neural dysfunctions).

In this study we will use the word “pathogen” when the negative selection pressure is essential to the process studied, and the word “parasite” when the result of the interaction, either harmful or beneficial for the host, is not determinant for the process we discuss.

Following conventions in the plant pathology vocabulary we use the term “**virulence**” as the qualitative ability of the parasite to infect a host. However, we will favor the word “**infectivity**” as being synonymous for this infection ability according to its use in the zoology literature. The issue arises because “virulence” in the zoology literature is used for the amount of damage, i.e. host fitness reduction caused by a parasite, a concept referred to as “aggressiveness” in plant pathology (Sacristan and Garcia-Arenal 2008)

## 1. Recent advances on coevolution: models and experimental data

### 1.1 Coevolutionary cycles

As mentioned above, reciprocal selection is expected between hosts and pathogens, and two models describing the molecular targets of this selection have been used, the Gene-For-Gene (GFG) model and the Matching allele (MA) model (Box 2 and Table 1). The dynamics of allele frequencies rely on the action of frequency-dependent selection (FDS). Under both models, a host allele for resistance will increase when it confers resistance to the most prevalent pathogen allele whereas all hosts without this resistance allele suffer disease. This selection holds until pathogens evolve infectivity on this host genotype, which in turn brings the system back to the initial state. The system is cyclic, and these cycles are called coevolutionary cycles. A common assumption of models is that the cyclic nature of coevolution occurs if resistance and infectivity cannot be additive so that universally infectious pathogen or universally resistant plants are excluded as possibilities. This occurs for instance if a gain in resistance or infectivity has a fitness cost in the absence of non-infectious pathogens or of resistant hosts respectively, an assumption that we will discuss later (Frank 1992).



Such cycles under a GFG model is presented in Figure 1: starting from a stage where a specific non-infectious allele (so-called “avirulence” or *avr* allele) is common, the corresponding host resistance allele (*RES*) has an advantage and its frequency increases (A); in turn, this increased frequency of resistance among hosts increases selection for pathogen infectivity (increase in *avr*<sup>-</sup> also called *vir* allele at the same locus) (B); the system thus reaches a state with high frequencies of both resistance and infectivity, so that the allele in host population does in fact not confer resistance anymore; this host genotype may then decrease in frequency if it carries a fitness cost (C); finally as hosts do not carry the resistance allele anymore, corresponding infectious (*avr*<sup>-</sup>/*vir*) alleles among pathogens may be selected against if they confer a lower fitness than *avr* alleles (D). Such dynamics have also been called the “boom and bust” cycle in the plant pathology literature (Leonard 1977; Leonard and Czocho 1980; Barrett 1988).

In host-pathogen evolution models, each host resistance allele is thus first selected for, and then selected against, depending on the frequency of corresponding pathogen alleles and associated costs. Such selection is more specifically described as negative frequency dependent selection (NFDS) as it selects against alleles found in high frequency (Frank 1992). More recently it has been defined as “indirect negative frequency dependent selection” referred to as iNFDS or iFDS (Tellier and Brown 2007) because the selection coefficient at the host genes (or respectively the pathogen genes) depends on the frequency of the corresponding pathogen alleles (or host alleles) (Frank 1992).

**Box 2- Two main models of molecular recognition mechanism underlying infectivity: GFG and MA models**

**The Gene-for-Gene (GFG) model** is based on plant-pathogen systems. It assumes that, for pathogens, there are two classes of alleles at the locus involved in the ability to cause an infection: one class to which some host genotypes are resistant (so called “avirulent” or *avr* alleles because in plant pathology, virulence refers to the qualitative ability to infect a host genotype, see box 1), and the other, called “virulent” allele *avr*<sup>-</sup> or *vir* allowing infectivity on the resistant host genotypes (Agrawal and Lively 2002). Note in addition that *avr*<sup>-</sup> (*vir*) alleles can correspond to loci for which no corresponding host resistant allele exists, specifically if they do not target proteins of the host.

The hosts exhibit a corresponding locus for which there are two classes of alleles respectively called “resistant” alleles *RES* and “susceptible” alleles *res*<sup>-</sup>. Hosts carrying the susceptible alleles can be infected by pathogens carrying either allele at the “avirulence” locus; hosts carrying the resistant *RES* allele are only susceptible to pathogens carrying the “virulent” (infectious) allele *avr*<sup>-</sup> (once again, also termed sometimes *vir*).

The underlying mechanism is a protein produced by the resistant plants that recognize the pathogen’s product of the “avirulent” allele, which induces a defense reaction preventing infection. The product of the infectious *avr*<sup>-</sup> allele is not recognized by the *RES* allele such that infection can occur. The non-infectious *avr* allele is dominant, and so is the resistant *RES* allele.

When multiple loci are involved in the interaction between hosts and pathogens, a host can resist a pathogen if it has at least one resistant allele corresponding to a pathogen locus for which the pathogen carries an *avr* (non infectious) allele. The pathogen can infect the host if it carries infectious allele *avr* at each locus corresponding to resistant host alleles. The multi-locus GFG system is thus characterized by the existence of a “super” infectious pathogen which can penetrate and grow in all hosts whatever their genotypes (Agrawal and Lively 2002).

As a summary, there is an advantage to pathogens carrying “virulent” allele, as these pathogens are able to infect a broader spectrum of hosts than pathogens carrying “avirulent” alleles. Similarly, there is an advantage to hosts carrying resistant alleles, as these hosts are able to resist infection from a broader array of pathogens than hosts carrying susceptible alleles. Fitness costs associated with maintaining a superfluous *avr* allele or *RES* alleles have been invoked to explain why super-infectivity or super-resistance do not reach fixation and thus to explain the maintenance of variation in populations. This assumption is discussed in part 1.4.

**Under the matching-alleles (MA) models**, a first possibility is that a host must match a pathogen genotype to resist to this pathogen. A second possibility (also referred to as inverse Matching-Allele model, a pathogen genotype must match a host genotype to perform successful infection, so that reciprocally, a host resists any pathogens that carry no allele able of recognizing this host. In this case, it is the mutation of host genotype that is first selected by the most common pathogen genotype, which in turn favors the increase in frequency of a pathogen genotype able to recognize this mutated host. This system is relevant for many animal viruses that attach to host cells via protein-protein interactions. When multiple loci are involved in the interaction, a pathogen can infect a host if it carries an allele compatible with that of the host at any of the infectivity loci.

The multi-locus MA model is characterized by the absence of super-infectious pathogens (and of super-resistant hosts) because performing infection on a host (or resisting a specific pathogen) means carrying a combination of alleles that impedes being infectious to hosts (or resisting to pathogens) carrying at least one other allele.

The underlying molecular mechanism in the inverse MA model is that the pathogen needs to block all the host molecules capably of triggering an immune response. In a classic MA model, the mechanism is that one host allele would recognize one given antigenic allele in the parasite (as for the MHC for example).

Note that in both GFG and MA models, one organism is benefiting from avoiding recognition (the pathogen in GFG and in the classical MA; the host in inverse MA) which determines who is "chasing" whom in the cycle.

Table 1- Summary of the interaction results for Gene-For-Gene (GFG) and the Matching-Alleles (MA) models of molecular recognition mechanism. A- GFG model. B- inverse MA model. C- classical MA model.

Note that for plants, the wild-type alleles are written in capital letters, but in pathogens like bacteria, all alleles are written in lower-case letters.

<b>A</b>		Allele at the pathogen locus	
		<i>avr</i> allele ("avirulent")	<i>avr</i> <sup>-</sup> (or <i>vir</i> for "virulent")
Allele at the host locus	<i>RES</i> resistant allele	Lack of infection	Successful infection
	<i>res</i> susceptible allele	Successful infection	Successful infection

<b>B</b>		Allele at the pathogen locus		
		<i>P1</i>	<i>P2</i>	<i>P3</i>
Allele at the host locus	<i>H1</i>	Successful infection	Lack of infection	Lack of infection
	<i>H2</i>	Lack of infection	Successful infection	Lack of infection
	<i>H3</i>	Lack of infection	Lack of infection	Successful infection

<b>C</b>		Allele at the pathogen locus		
		<i>P1</i>	<i>P2</i>	<i>P3</i>
Allele at the host locus	<i>H1</i>	Lack of infection	Successful infection	Successful infection
	<i>H2</i>	Successful infection	Lack of infection	Successful infection
	<i>H3</i>	Successful infection	Successful infection	Lack of infection

## 1.2 Arms race versus trench warfare models: expectations on genetic diversity

Coevolution can trigger two different outcomes on allelic frequencies dynamics under the GFG assumptions. In the arms race model (Holub 2001; Woolhouse et al. 2002), recurrent fixation of new alleles occurs in the pathogen population due to strong indirect NFDS (Tellier and Brown 2007). New mutants at pathogen infectivity loci, i.e. virulent, are expected to increase in frequency in the population because they have a selective advantage as rare variants. Specificity of this model relies in the assumption that the selection in favor of an infectious *avr*<sup>-</sup> allele does not lessen as the allele becomes prominent so that it comes to fixation (1; Figure 2a). In turn, due to a fitness cost, the corresponding resistance allele drops in frequency in host population (2; Figure 2a). Subsequently, any new non-infectious *avr* (avirulent) mutant among pathogens will then be selected for because the corresponding infectious allele (*avr*<sup>-</sup>) carries a fitness cost (C; Figure 2a) (and see 1.4 for discussion on these costs). This will select for new resistance among hosts. Allele frequencies are thus expected to

vary continuously as indicated in Figure 3a, each increase corresponding to a new allele. The dynamics is then said unstable because polymorphism in both host and pathogen populations is not maintained over the long term, but rather occurs with periodic variations dependent upon the occurrence of novel mutations.

The second scenario is known as “trench warfare” (Stahl et al. 1999; Stahl and Bishop 2000; Holub 2001; Woolhouse et al. 2002). Alleles selected for do not reach fixation, because the strength of directional selection is too weak or even negative when they are prominent. The selection process, alternating between positive selection when the allele is rare and negative selection when it is frequent, is referred to as “balancing selection”. Stabilization is expected in the end (Figure 2b) with polymorphism being maintained as a steady state. It has recently been shown that such stable dynamics are obtained if there is direct NFDS in addition to indirect NFDS (see 1.4 for details).

Expectations on genetic diversity under both of the arms race and trench warfare models have been explored for understanding potential dynamics in comparison with data from natural populations. Long term balanced polymorphism such as under a trench warfare model (Figure 3b) is expected to lead to high genetic divergence between alleles around the selected sites in the host and pathogen and to low levels of Linkage Disequilibrium (LD) with any other locus when recombination is effective. As the same alleles are expected to be maintained over the long term, a higher degree of differentiation can accumulate than at other genes.

In contrast polymorphism is transient under the arms race model (Figure 3a), with novel alleles being recurrently fixed. The loci under selection and their flanking regions should therefore exhibit the molecular signature of hitchhiking, with very low level of diversity (Maynard Smith and Haig 1974). Typical expectations for sequence data under the arms race model are thus a valley of reduced genetic diversity centered on the site of selection, and decreasing levels of Linkage Disequilibrium (LD), when going further away from the site under selection (Maynard Smith and Haig 1974; Kim and Stephan 2002; Li and Stephan 2005).

These predictions can be tested using DNA sequence data to determine which of the two models prevails in nature, or what are the genes subjected to coevolution dynamics. The principle of most population genetic studies aiming at detecting selection is to study the pattern of diversity ( $\theta$ ), linkage disequilibrium, and summary statistics describing the frequency spectrum (Tajima's  $D$  (Tajima 1989)). Various statistical tests can detect deviation from neutral evolution using such information (Fay and Wu 2001; Aguileta et al. 2009). Numerous other genes should first be analyzed to control for demographic events affecting the genome as a whole such as bottlenecks, expansion and metapopulation structure. Demographic events can indeed create deviation from standard neutral evolution, mimicking effects of selection. Controlling for these parameters before any inference on selection is therefore critical. A valuable way to detect selection is to test if summary statistics at a candidate locus are outliers compared to an empirical distribution for the whole genome. Such distributions can be obtained from sequences of numerous other genes or simulated using coalescence (Li and Stephan 2005; Barreiro et al. 2008; Aguileta et al. 2009).

With the development of whole genome sequencing projects, molecular patterns of selection can be investigated ever more easily. However, it must be highlighted that detecting footprints of selection in a plant or a pathogen does not necessarily implies coevolution, i.e. reciprocal selection pressure between a host and its pathogen. Selection can be simply due to adaptation, and in fact, disentangling between coevolution and adaptation requires analyzing both partners which is rarely done.

### 1.3 Trench warfare model among plant hosts, arms race model among microbial hosts and pathogens

As explained above, plant-pathogen systems are highly convenient for investigating dynamics of host-pathogen interactions. They were among the first systems in which footprints of selection were investigated using DNA sequences.

*Arabidopsis thaliana* in particular has served as a model because of the availability of large amounts of sequence data. A number of *A. thaliana* genes shown to be involved in resistance against pathogens exhibit selection footprints (Stahl et al. 1999; Palomino et al. 2002; Tian et al. 2002; Mauricio et al. 2003; Meyers et al. 2003; Tian et al. 2003; Allen et al. 2004; Bakker et al. 2006; Shen et al. 2006; Ding et al. 2007; Orgil et al. 2007). The resistance genes *RPP13*, *Rps2*, *Rpm1*, for which several alleles were known, harbor a balancing selection pattern, and corresponding avirulence genes have been identified (Mauricio et al. 2003; Shen et al. 2006; Desveaux et al. 2007). These resistance genes in *A. thaliana* thus seem to match the trench warfare model. Other putative resistance genes were identified based on the presence of a Leucine Rich Repeat (LRR), a domain prone to nucleotide binding carried by all plant resistance genes. Whole genome survey of genes with LRR domains revealed few genes with signature of selection (Bakker et al. 2006), suggesting that resistance genes exhibiting balancing selection could be the exception rather than the rule. In the meantime, other functions than disease resistance have been described for some of the LRR containing genes, for example protein-protein interactions (Tameling and Joosten 2007). Also some of the LRR carrying loci could be pseudogenes. The LRR- genes exhibiting selection footprints could thus be those that are in fact involved in pathogen resistance, whereas loci with no selection footprints would be involved in other processes.

Evidence for balancing selection has also been found in candidate resistance genes in wild tomato species, *Solanum pimpinellifolium* (Caicedo and Schaal 2004) and *S. peruvianum* (Rose et al. 2005; Rose et al. 2007), in common bean *Phaseolus vulgaris* (De Meaux et al. 2003; de Meaux and Neema 2003), in lettuce (Meyers et al. 1998; Kuang et al. 2004), and in maize (Tiffin et al. 2004; Moeller and Tiffin 2005). No evidence for arms race model in plant hosts has been identified so far to

our knowledge. The rare cases where such selection has been suggested could not clearly exclude neutrality (Stranger and Mitchell-Olds 2005).

With regard to evidence for selection upon pathogen loci, most studies have focused on “avirulence” genes, that code for elicitors and effectors of host resistance (see the reviews by (Aguileta et al. 2009; Stukenbrock and McDonald 2009)). Necrosis inducing proteins, also called phytotoxins, harbor signatures of positive selection in many systems (see reviews by Misas-Villamil (2008), Stahl (2000), and Tiffin (2006)), for instance in *Phytophthora infestans*, the potato pathogen responsible for the Irish famine in the XIXth (Liu et al. 2005), and in *Botrytis*, the agent of the grey mould affecting most fruits during storage and still causing important economical losses (Staats et al. 2007). Some “avirulence” genes have been shown to exhibit signs of positive selection (Birch et al. 2006; Van der Merwe et al. 2009). Interestingly, another frequent outcome of selection imposed by the host seems to be the complete deletion of “avirulence” genes (Stukenbrock and McDonald 2007).

Host evolution thus seems to conform to the trench warfare model and pathogen evolution the arms race model in plant-pathogen systems although both partners have been studied in parallel in a small number of systems only (Mauricio et al. 2003; Shen et al. 2006). That the two partners evolve following different dynamics can seem paradoxical, but it may be due to differences in host and pathogen metapopulation structure, life history traits such as generation time (see section 1.4) or in fitness costs of selected mutations. These differences may however not hold for all host-pathogen systems. In animal-pathogen systems, balancing selection reflecting the trench warfare model has also been detected in the pathogens, for instance in *Plasmodium falciparum* (Conway et al. 2000; Verra et al. 2006), in addition to their animal hosts at immune defense loci *Mhc* (Vogel et al. 1999). Microorganisms such as bacteria and their parasitic phages are a class of systems where infectivity and resistance evolution is particularly easy to investigate under controlled conditions, triggering enlightening studies on experimental evolution (Lopez-Pascua and Buckling 2008; Pepin et al. 2008; Poullain et al. 2008; Cairns et al. 2009). The bacteria and phage evolution can be directly followed due to the easy preservation of hosts and pathogens at each stage of the experiments. Phenotypic evolution fitted the arms race model for several phage-bacteria systems, including *Escherichia coli* or *Pseudomonas fluorescens* as a host (Buckling and Rainey 2002; Wichman et al. 2005). Phenotypic evolution of infectivity by *Pasteuria* bacteria onto their *Daphnia* hosts could also be studied over many generations thanks to their preservation on natural pond sediments, reflecting pathogen evolutionary dynamics that matched the trench warfare model (Decaestecker et al. 2007; Gandon et al. 2008). Whether the associations behave like a GFG or MA system i.e. who is being detected and who benefits from avoiding recognition, or a mixture of both (Agrawal and Lively 2002), is another question, more difficult to assess, which has not been thoroughly investigated to our knowledge so far.

As the number of studies increases, the picture of host-pathogen evolution becomes more diverse, calling for new models to integrate parameters that can affect the patterns of coevolution, or to circumvent hypotheses of the early models that are irrelevant in some systems. We present in the next section new theoretical developments that have stepped in this direction.

#### 1.4 Mathematical framework for improved understanding of polymorphism maintenance

A recurrent assumption in coevolutionary studies is that pathogen alleles conferring infectivity to a specific host genotype are costly, i.e. reduce fitness on susceptible host genotypes. Experimental evidence has been provided in some disease systems. For instance, in the bacterial blight pathogen of rice, *Xanthomonas oryzae* pv. *oryzae*, spore transmission on a specific host genotype was negatively correlated with spore transmission on the other host genotype i.e. each strain succeeded on a specific host lineage (Vera Cruz et al. 2000). Some other examples of trade-offs have been identified (Jensen et al. 2006; Bahri et al. 2009) ( see also the review by Alizon (2009)). However, they are unexpectedly rare, so that universality of such costs remain debated. We redirect interested readers towards specific articles on this topic (Bergelson and Purrington 1996; Brown 2003; Sacristan and Garcia-Arenal 2008; Salvaudon et al. 2008).

To circumvent the initial assumption of conditional fitness costs of resistance and infectivity, recent models have investigated whether realistic assumptions could lead to a reduction in the frequency of infectious (*avr*) alleles without such costs. In fact, it was theoretically shown that fitness costs of resistance and infectivity are necessary in infinite (or very large) population models for generating coevolutionary cycles (Tellier and Brown 2007; Tellier and Brown 2007). In contrast, coevolutionary cycles could occur in finite (small) population models with strong stochastic processes without a cost for infectivity (Damgaard 1999; Thrall and Burdon 2002; Salathe et al. 2005). These processes are detailed below.

Metapopulation structure, the most obvious ecological complexity important for host-pathogen coevolution, was the first important model component to be introduced. Spatial structure and limited gene flow among plant and pathogen demes were shown to affect the rate of coevolution under the GFG model: in highly viscous metapopulation, i.e. when gene flow is very low, asynchrony among demes can appear, leading to balancing selection (Sasaki et al. 2002). In a metapopulation with demes starting at different host and pathogen allele frequencies, the rate of cycling is slowed down by the viscous structure, and therefore fixation of resistance or infectivity (“*Avir*” alleles”) takes much longer than in an unstable host-pathogen system with a single population (Sasaki et al. 2002; Gavrillets and Michalakis 2008). In this case, the metapopulation per se creates “statistical polymorphism”, i.e. transient polymorphism which is maintained for a very long period of time before fixation of alleles ultimately occurs (Allen 1975). Metapopulation structure can also create local adaptation patterns in

host and pathogen populations depending on their relative rates of among-deme migration (Kaltz and Shykoff 1998; Gandon 2002). For instance, if the pathogen exhibits higher migration rates than the host, it will be more efficient to track the host resistance alleles in the different populations. The pathogen is then ahead of the host in an “arms race” scenario (Gandon, 2002). Expectations for the trench warfare signature of balancing selection under those scenarios however remain to be derived theoretically. Similarly, the methods for testing scenarios of coevolution (arms race or trench warfare) based on sequence data need to be improved by expanding  $F_{ST}$  based methods (Beaumont and Balding 2004; Beaumont 2005). Such methods distinguish local selective sweeps (arms race) from global balancing selection (trench warfare at the metapopulation level) in simple island models. It is unlikely that these simple scenarios are applicable to complex host-pathogen systems with metapopulation structure in stepping stone and/or expansion, as well as for balancing selection with different alleles in the demes.

Other recent studies have investigated the mechanisms by which the outcome of coevolutionary cycles is stable or unstable. A necessary, but not sufficient, condition is that, in addition to iNFDS i.e. selection due to allelic frequency of its partner, there is direct negative FDS, i.e. selection relying on the allelic frequency of the organism itself (Tellier, 2007 #895). Conditions promoting direct NFDS and therefore polymorphism maintenance in host and pathogen populations are 1) polycyclic life cycles of pathogens i.e. pathogens going through several generations per host generation (Tellier and Brown 2007; Tellier and Brown 2007), 2) high auto-infection rate of pathogens, i.e. high proportion of spores re-infecting the same host plant across successive pathogen generations (Barrett 1980). In contrast, high levels of induced resistance (a quantitative resistance mechanism triggered by infection and mediated by systemic signals, in contrast to constitutive defense mechanisms as described previously) diminished direct NFDS and therefore polymorphism (Tellier and Brown 2008). Regarding the host side, direct NFDS is expected to act on resistance frequency when the host has long lived seed banks (Tellier and Brown, unpublished data). This occurs because seed bank store past selective events in the form of previous host frequencies. The difference between host allele frequencies in the seed bank and in the above ground population dampens co-evolutionary cycles. Finally, general features regarding the host-pathogen populations' contacts promote the impact of direct FDS: in spatially structured populations where the environment varies among demes with different selective factors for host and pathogen alleles, direct NFDS is acting, promoting stable polymorphism. Preliminary work also shows that higher mutation rates and linkage disequilibrium between loci would increase direct NFDS {Leonard, 1997 #296\; Tellier and Brown, unpublished data}. Note that direct NFDS, and thus stable polymorphism, is expected to occur in a vast range of plant-parasite systems. For example polycyclic disease is a common feature of many bacterial and fungal parasites (Pei et al. 2005; Williamson et al. 2007), and many annual plants species exhibit seed banks.



Besides population structure and key determinants on selection action, epistasis is a newly considered parameter that might be critical in understanding coevolution dynamics. Recent advances in genomic studies reveal that plant resistance genes are often clustered in genomes and submitted to deletion/duplication events. It is argued that resistance genes in these clusters evolve by a birth and death process (Michelmore and Meyers 1998; Bergelson et al. 2001). Adding linkage disequilibrium in a simple two-locus GFG model does not create direct FDS, and thus does not affect stability of polymorphism (Sasaki 2000; Tellier and Brown 2007). However, if epistasis occurs among the different genes, modifying for instance the costs of resistance alleles, polymorphism at host and pathogen genes can be enhanced. This occurs, for example, when assuming that the cost of resistance genes is high when few are present in the genome and that the individual cost of adding new genes diminishes with increasing the number of *RES* genes. Similarly, stable polymorphism occurs when assuming that the individual cost of adding new virulence genes in the pathogen genome would increase (Tellier and Brown 2007). We are not aware of any study investigating the interaction between several host resistance loci. Among pathogens, a study on *Xanthomonas axonopodis* questioned additivity of fitness costs: when knocking down an increasing number of “avirulence” loci, the fitness cost measured in the bacilli was increasing in a non linear manner (Wichmann and Bergelson 2004). This is in accordance with the negative non linear correlation between spore production and the number of host alleles overcome by the pathogen, identified in natural populations of *Melampsora lini* (Thrall et al. 2002). However, further experimental studies and theoretical models tackling the evolution of duplication and resistance gene families are needed to better describe the dynamics of multi-locus GFG system.

Apart from implementing more complex data, new mathematical models are needed to predict features other than the maintenance of genetic diversity. For instance, Gandon et al (2008) modeled the evolution of the mean fitness of pathogens under an arms race model and under a trench warfare model. The outcome can allow discriminating between the models in microorganisms for which measurements of fitness are easy (Buckling and Rainey 2002). In fact, it has been shown using such approaches that predator-prey systems (Hanifin et al. 2008), as well as bacteria-bacteriophages (Gandon et al. 2008), evolve according to the arms race model.

As a conclusion, recent mathematical approaches incorporate more realistic parameters to understand more deeply coevolutionary dynamics. Some of these works have already begun to shed light on possible causes for stable polymorphism observed in plant-pathogen systems (see section 1.3): by promoting direct NFDS, numerous biological and ecological factors, such as high auto-infection rate and polycyclic disease, promote stable allele polymorphism and thus trench warfare dynamics (Tellier and Brown 2007). Metapopulation structure, high mutation rates, migration among demes and high rates of extinction-recolonization, can even generate stable polymorphism over long periods of time without fitness costs of resistance and infectivity (Thrall and Burdon 2002).

Apart from these life history and infectivity traits, other features may be important for host-pathogen interactions, such as quantitative impact of pathogen on hosts (aggressivity/virulence), and is also subjected to genotype x genotype interactions as evidenced by recent studies (Salvaudon et al. 2005; Salvaudon et al. 2007). These features could impact coevolution dynamics, although this is still poorly explored. We believe that the understanding of coevolution will benefit from testing the effect of the parameters described by the “Geographic Mosaic of coevolution” (Thompson 1994) such as host genotype  $\times$  parasite genotype  $\times$  Environment (G $\times$ G $\times$ E) interactions (Laine and Tellier 2008), although increasing complexity can obscure what processes are really essential to coevolutionary dynamics.

## 2. Cospeciation

### 2.1 Theoretical framework and methods to test for cospeciation

This part deals with consequences of host-pathogen interaction over longer time scales. In particular, it reviews theory and data about the prevalence of cospeciation versus speciation following host shifts, and the methods used to infer these past events. The theory and data in this section do not assume that there should be a negative impact on host fitness so that we will more broadly speak of “parasites” (see Box1).

Macroevolutionary evolution of host-parasite associations being usually not observable in a researcher's lifespan, methods for inferring the effects of their interaction have been developed based on the comparison of their phylogenetic trees. These methods, referred to as “co-phylogenetic methods”, rely on the idea that two interacting lineages, if having diversified only by cospeciation, will show completely congruent phylogenies (Fig. 4A), whereas evolutionary events other than cospeciation (Fig. 4B-F) will decrease the congruence (Brooks and McLennan 1991). Events that reduce congruence include host shifts (Fig. 4B), where part of the parasite species adapts to a new host and speciation occurs, duplication of the parasite species without duplication in the host, also called intra-host speciation (Fig. 4C), partial extinction or “missing the boat” (Fig. 4D), when the speciation in the host lineage is not followed by a speciation in the parasites, host jump (Fig. 4E), when the parasite leaves its host of origin and colonizes a new one, and extinction of the parasite lineage (Fig. 4F). Partial extinction and host jumps can be seen as a combination of two other events (respectively cospeciation and extinction, and host shift and extinction) so that the likelihood of these specific events will not be further discussed.

Co-phylogenetic methods can be classified into two main classes. The first class aims at reconstructing the evolutionary history of the lineages, i.e. inferring the nature and frequency of

different evolutionary scenarios (event-based methods) from the comparison of the phylogenetic trees. The second class tests the overall congruence between the trees, i.e. topological similarity and/or symmetry in time of divergence between hosts and parasites, and considers high levels of congruence for evidence of frequent cospeciations (tests of congruence). We will explain in more details hereafter these two approaches and give a brief overview of the existing methods. In a last part, we will discuss some limitations of these methods in the light of recent theoretical results on the possibility to obtain congruence among host and parasite trees without cospeciation.

## 2.2 Event based methods

The first method developed was the *Brooks Parsimony Analysis (BPA)* (Brooks 1981). It opened the way for event-based methods but considered parasites as character states of the hosts. It was modified by the same author ten years later (Brooks and McLennan 1991). The parasitic character states can be assigned to each branch in the phylogeny of the hosts and the most parsimonious reconstruction will be the one considering the smallest number of states in the host's phylogeny. As an illustration, if host and parasite phylogenies are topologically identical, at each branch in the host phylogeny is assigned one “parasite” trait, and cospeciations are then considered as the only mode of speciation. If the two phylogenies include numerous taxa and have very different topologies and if some hosts are associated to more than one parasite, the problem can become intractable, as multiple and very different reconstructions can be equally parsimonious. In practice, parasite information is transformed into additive binary code and mapped onto the host tree. BPA has been widely used in the 80's and the early 90's but has also been heavily criticized, in particular because the results given by BPA require a lot of *a posteriori* interpretations (Page 1994).

*Component analysis, reconciliation analysis and TreeMap 1.* Another method proposed by Page in 1990 (Page 1990), called “reconciliation analysis”, did not consider the parasites as character states anymore, but as proper evolutionary lineages. This method, implemented in the program COMPONENT (Page 1993), estimates the minimum number of extinctions and duplications that are needed to reconcile the host and the parasite phylogenies. It does not allow host shifts and therefore has only been used sparsely. Page proposed a modification of this method a few years later (Page 1994) allowing host-switching. This new method, implemented in the program TreeMap 1 (Page 1994), tries to reconcile host and parasite phylogenies by maximizing the number of cospeciations and minimizing the number of host-shifts. No constraints on the number of duplications, extinctions and number of parasites present on the ancestral strains are specified. The major advantage of this method is to give a graphical representation of the history of the association. This representation however can become unreadable for associations in which the topologies of the host and parasite trees are very different and for multiple associations between hosts and parasites. Furthermore, the number of parasites infecting ancestral host species can be assumed to be unreasonably high (Refrégier et al.

2008). Treemap 1 can, in addition, calculate the number of cospeciation events for hypothetical trees obtained by random permutations between hosts and/or parasites tips. This provides an indirect test to assess whether the number of cospeciation events in the observed host-parasite trees is higher than for random phylogenies. Note that reconciliation analysis infers the most likely scenario by maximizing the number of cospeciations and minimizing the number of host-switches, i.e. it assumes *a priori* that cospeciation is more likely than host-switch and other events, which has been largely debated (Ronquist 1995).

*Cost-based methods.* Ronquist (1995) developed a series of methods allowing the user to attribute a cost to each evolutionary event (cospeciation, host shift, duplication and extinction). These methods find the most parsimonious scenario by minimizing its total cost. Some of these methods considered only two or three types of events, others could take into account the same four events as Treemap 1 (for a review of these methods, see Ronquist 2003). The more popular cost-based method is implemented in the software Treefitter (Ronquist 1995). It differs from Treemap 1 by two main aspects. First, Treefitter does not try to map one tree onto another as does Treemap 1. It only estimates the number of events of each type that can explain the two phylogenies and associates to each event a probability that it arose by chance. This probability is calculated by permutations of the host and/or the parasite leaves. Second, in order to find the optimal number of each event, Treefitter minimizes the total cost of the reconstruction, obtained by multiplying the individual costs of each event by their number and summing. The advantages of Treefitter compared to Treemap 1 are (i) associating a probability to each event and (ii) letting the costs of each event to be set by the user. However, some constraints on the event costs are still present, like cospeciations that cannot be more costly than host-switches.

One of the weaknesses of Treemap 1 is that it can give a very large number of reconstructions (Dowling, 2002). There are two reasons for this; first, Treemap 1 scores each reconstruction solely by the number of cospeciation events and ignores other events when scoring. As many reconstructions can have the same number of cospeciation events, Treemap 1 can yield many solutions (Page and Charleston 1998). Second, as noticed by Ronquist (1995), Treemap 1 does not guarantee that reconstructions involving more than one host-switch are feasible. The last version of Treemap, Treemap 2, allowed getting rid of these problems, first by associating a cost to each event, and second by the implementation of the method “jungles” (Charleston 1998), an algorithm allowing the fast identification of the most optimal reconstructions taking costs into account and ensuring the feasibility of each reconstruction (host switches only performed between hosts present at the same time) (for details on the method and its implementation in Treemap 2, see Charleston 1998; Charleston and Perkins 2003).

*Bayesian methods.* The methods presented above suffer from two main problems: they consider that the phylogenies of the host and the parasites are known, and, apart from BPA, they consider the cospeciations as being more likely than host switches. The bayesian method developed by Huelsenbeck et al. (Huelsenbeck et al. 2000; Huelsenbeck et al. 2003) is free from these problems. It is however still largely incomplete, only considering host-switches and cospeciations and being only applicable for a 1:1 correspondence between hosts and parasites. The goal of this method is to determine the most likely evolutionary scenario that can explain the host and parasite sequences and not their phylogenies. This approach is based on two simple stochastic models, one regarding host-switches and one regarding DNA substitution. The two models are mixed and treated by Bayesian analysis.

The hypotheses of the model concerning host shifts (Huelsenbeck et al. 2000) are: (1) In the absence of host-switches, both phylogenies are identical, (2) differences between host and parasite phylogenies are only explained by host-switches, (3) one host is associated to a single parasite and (4) when a parasite switches, it excludes the parasite that was initially present on the host it switches to. This model allows obtaining the probability of an evolutionary scenario (number of switches, source and target hosts, and timing of the switches) given a host tree, the divergence times in this tree and a rate of host switches. As this method does not consider the host tree as known, another model, concerning the maximum likelihood of DNA substitution model, is added to the one described above to reconstruct the phylogenies. It allows finding the probability of observing a dataset (a sequence alignment) given some substitution parameters (transition/transversion rates, etc.). The peculiarity of this second model is that it considers that the probability of observing a given alignment for the parasites is determined by the host tree topology, the speciation dates in the host tree, the different parameters in the substitution model, and an evolutionary scenario (number of switches, source and target hosts, and timing of the switches).

These two mixed models include a high number of parameters. The Bayesian inference method allows dealing with all these parameters and estimating the rate and the number of host-switches. It also allows calculating the probability of a given scenario.

### **2.3 Topology- and distance-based methods**

All the methods presented above are based on the idea that host and parasite phylogenies should be identical (congruent) in the absence of host switches, extinction, and duplication. This idea results from principles enounced by Fahrenholz (1913): “Parasite phylogeny mirrors host phylogeny” and Szidat (1940): “primitive host harbour primitive parasites”. These “laws” (Fahrenholz 1913; Szidat 1940) long prevailed and led to the development of numerous statistical methods aiming at testing the

congruence between host and parasite phylogenies. These methods can be divided into different classes depending on the null hypothesis that is tested (similarity or independence, Huelsenbeck et al. 2003) and on the data that are used for the test (trees, distance matrixes, or raw sequence alignments, Light and Hafner 2008)

*Tests of independence.* The principle of this kind of tests relies on comparing the topological or genetic distance of the focal host-parasite association to a distribution of distances obtained by generating a large number of random trees. If the distance of interest is significantly smaller than expected by chance, the association is considered as significantly congruent. This is similar to the type of tests implemented in Treemap (see section 2.2).

Applied on tree topologies, independence tests differ by the way trees are constructed, by the method used to generate random trees, and by the distance method used to calculate pairwise distance between trees. One of the weaknesses of these methods is that a large number of random trees have to be generated *de novo* for each new comparison of trees by de Vienne et al. (2007; 2009) proposed a new test of tree independence that uses previously simulated associations. The distance method is based on Maximum Agreement Subtree (MAST), i.e. the largest identical tree for host and parasite phylogenies obtained by removing branches. The distance is proportional to the number of branches that has to be pruned from the two trees. A large number of pairs of random binary trees with varying numbers of tips have been generated and the size of the MAST of all possible associations was calculated. The distribution of the size of the MAST, given the number of leaves (terminal branches of the tree), has then been fitted by an exponential function. The test compares the size of the MAST obtained with the trees from the association of interest to the distribution of the size of the MASTs for random trees. If this size of the MAST is in the 5% left of the distribution, the trees are considered as more congruent than expected by chance. The test returns an index giving the degree of congruence (Icong index) as well as the associated P-value (de Vienne et al. 2007; Kupczok and von Haeseler 2008).

Tests of independence have also been used to test for temporal congruence. The occurrence of repeated cospeciation events indeed means simultaneous occurrence of speciation events (i.e. temporal congruence) between hosts and parasites, and thus proportional branch length and identical dates for the nodes in the compared phylogenies. A first method (Hafner et al. 1994) tests if the two species have accumulated the same amount of genetic differences. Input data are the host-parasite species associations and alignment of one or several specific loci independently for hosts and for parasites. These alignments are used to calculate distance matrices. The significance of the correlation between the two matrices is then assessed using a Mantel test (Hafner et al. 1994). A similar method compares matrices of branch lengths from host and parasite trees in the same way (Hafner et al. 1994; Page 1996). If molecular clocks are available for both host and parasites, one can use the estimated absolute

ages of the nodes in the two trees. Identifying identical ages for each node is in fact the only way for inferring cospeciation. Indeed, identical relative divergence times, as deduced from proportional branch lengths, can exist in some host-parasite associations without speciation times being identical (Charleston 2002). Note however that Mantel tests, while accounting for statistical non independence in matrices, do not account for phylogenetic non independence (Felsenstein 1985). The data on divergence at ancient nodes include the same information as divergence points at more recent nodes along the same branches (Felsenstein 1985; Schardl et al. 2008). All the points used in the distance matrices are thus phylogenetically non independent, which should prevent the use of a Mantel test.

Parafit (Legendre et al. 2002) is a software program implementing a test of independence between host and parasite genetic or patristic distances (patristic distances are genetic distances reevaluated in the light of the preferred phylogenetic reconstruction). The major advantages of this method are (i) to deal with the cases where multiple parasites are associated to one host or where multiple hosts associated to one parasite and (ii) to allow testing the contribution of each individual host-parasite link to the total statistics of congruence. The principle of the method is as follows: the host sequences and/or tree and the parasite sequences and/or tree are transformed into distance matrices that are themselves transformed into principal coordinates matrices (matrices B and C). A third matrix containing binary data encodes the links between host and parasites (matrix A). A fourth matrix (matrix D) is obtained as the product of the matrices B and C weighted by the values in matrix A. The sum of the squared distances in matrix D gives a value of the overall similarity between trees (ParaFitGlobal). This value is compared to a distribution of ParaFitGlobal values obtained by permutations of the lines in matrix A. If the observed similarity value is found in less than 5% of the values obtained after permutations, the association is considered as more congruent than expected by chance. Parafit also allows testing the importance of each individual link in the overall congruence between trees by removing the links one-by-one in matrix A and looking at the effect on the ParaFitGlobal value. Note finally that giving a value of 1 to each branch length in the host and the parasite phylogenies renders the test similar to a simple test of independence between tree topologies (see above). In any case, the same problem of non-independence regarding phylogeny (Felsenstein 1985) also applies for this method.

Recently, Schardl et al. (2008) proposed a modification that can apply to programs such as Parafit, that takes into account the non-independence between pairs of species belonging to the same branch, and that uses a method similar as that proposed earlier by (Felsenstein 1985). The algorithm called MRCAlink (MRCA for Most Recent Common Ancestors) identifies phylogenetically independent pairs between host and parasite trees. The reduced host and parasites matrices can then be compared as previously described (see Schardl et al. 2008 for more details).

The main problem with all the methods testing for independence is that they consider the phylogenies and their branch lengths as known when performing the test for congruence although they have been statistically inferred (Huelsenbeck et al. 1997; Huelsenbeck et al. 2003).

*Tests of similarity or identity.* This kind of test computes a probability that the hosts and parasites present congruent phylogenies, using a Bayesian approach. The probability of observing the actual molecular diversity if phylogenies are congruent is computed. Distances or topologies are not calculated separately for each partner before comparison, but rather it is tested whether the data can have resulted from a common topology.

Huelsenbeck (1997) first proposed such a test where only topologies of the phylogenies were considered. He implemented the approach via two methods, one based on Maximum Likelihood and the other on Bayesian inference. Both tests have the clear advantage of taking sequence alignments as inputs and not trees directly.

For the first test, two maximum likelihood values are computed:  $[\max(l_0)]$ , computed under the hypothesis that the host and parasite trees are identical, and  $[\max(l_1)]$ , under the hypothesis that the two trees are possibly different. The statistics ( $\Lambda_{obs}$ ) is then the ratio of these two likelihoods.

$$\Lambda_{obs} = \frac{\max(l_0)}{\max(l_1)}$$

The significance of  $\Lambda_{obs}$  is determined using parametric bootstrap: many data sets are simulated under the assumption that the null hypothesis is correct (the topologies are identical), and  $\Lambda$  is computed for each data set. If  $\Lambda_{obs}$  is greater than 95% of the simulated  $\Lambda$ s, the null hypothesis is rejected.

The second test is based on Bayesian inference and gives the probability for the host and parasite phylogenies to be identical. The basic idea of the test is to calculate the posterior probability of observing a phylogeny ( $\tau$ ) given sequences ( $\Pr(\tau | \text{sequences})$ ) for host and parasite sequences, and sum these probabilities over all possible rooted trees with a given number of leaves. The probability for host and parasite phylogenies to be identical is thus given by:

$$\sum_{i=1}^{B(s)} \Pr[\tau_i | \text{HostSequences}] \times \Pr[\tau_i | \text{ParasiteSequences}]$$

where  $B(s)$  is the number of rooted trees with  $s$  leaves.

Huelsenbeck et al. (1997; 2003) proposed a method testing for temporal congruence using a method very similar to the maximum likelihood method presented above, except that the null



hypothesis was that the speciations have occurred at the same time, the alternative hypothesis being that the speciations have occurred at different times.

## **2.4 Studies of natural associations reveal the prevalence of host shifts**

All methods presented above present a number of problems that have been highlighted along with their description. The key issue that has not been discussed until now is the fact that these methods have all been developed with the idea that congruence between host and parasite phylogenies was a result of frequent cospeciations between host and parasite phylogenies, whereas incongruence was a result of host switches, extinctions, duplications, and other evolutionary scenarios.

However, some experimental and theoretical studies recently showed that congruence between host and parasite phylogenies could be obtained without cospeciation (Charleston and Robertson 2002; de Vienne et al. 2007) and hypothesizing that host switches preferentially occur towards closely related hosts. These results place a renewed emphasis to the temporal, and not solely topological, congruence tests between host and parasite phylogenies as they become the only tests that can truly assess the occurrence of cospeciations.

After more than fifty years of studies on congruence in host-parasite associations, convincing examples of cospeciations between hosts and parasites seem to represent exceptions rather than the rule. In many systems, first studies seem to provide evidence of cospeciations, but they later turned out to be non significant after more exhaustive sampling or more rigorous analyses. This was for instance the case of the association between fungus-growing (attine) ants and their microbial associates. This system first revealed clade-to-clade correspondences between the farming ants, their cultivars and the garden parasites (Currie et al. 2003). Recently however, larger sampling and confrontation of microbial symbiont diversity with microbial saprophytic diversity showed that lineages frequently re-associated over both short time frames and periods encompassing several speciation events (Gerardo et al. 2006; Mueller et al. 2008). Another famous association defeated regarding cospeciation pattern, including the yucca-yucca moth (Smith et al. 2008), were revealed to exhibit a significant lack of synchrony in between host and parasite speciation times.

Overall, complete congruence between host and parasite trees is almost never found (chapter 4, Thompson 1994) and host switches have been proposed to be the main mode of speciation in parasites in many systems involving plant viruses, plant fungi, plant parasitoids, animal viruses (Lopez-Vaamonde et al. 2001; Roy 2001; Charleston and Robertson 2002; Lopez-Vaamonde et al. 2003; Jackson 2004; Staats et al. 2005; Braby and Trueman 2006; Spatafora et al. 2007; Refrégier et al. 2008)

Besides, a study focusing on the association between anther smut and their caryophyllaceous hosts showed that failure to properly delimit species can lead several methods to wrongly conclude in favour of cospeciation, because generalist species were found on closely related hosts. This suggests that other studies may have overestimated the rate of cospeciation as species delimitation in parasites is often difficult and as generalist parasites usually infect closely related hosts (Refrégier et al. 2008).

One exception, and maybe the only convincing example of parallel diversification of hosts and parasites, is the well known association between pocket gophers and their chewing lice (Hafner et al. 1994; Hafner et al. 2003). This “textbook example” of cophylogeny played a central role in the development of the methods presented above, but was finally revealed to be an exception rather than the rule. Interestingly, this has been linked to the life history and ecology of these symbionts and their hosts: pocket gophers (Rodentia: Geomyidae) are herbivorous rodents that spend most of their life in tunnels that they do not share with other individuals. Species of Pocket gophers are mainly allopatric, decreasing the probability for their parasite to switch to other hosts. Moreover, the Chewing lice (family Trichodectidae) are obligate parasites whose entire life cycle takes place on the host. The combination of a solitary and allopatric life style of the host and a limited dispersion ability of the parasite can be seen as a reason for absence of host shifts. In contrast, heteromyid gophers that have a more social behaviour, and their sucking lice exhibit lower levels of congruence (Light and Hafner 2008). Perfect congruence between two species phylogenies, as in the case of pocket gophers and the chewing lice, can thus be due to parallel speciation that only results from allopatry of hosts (and therefore also allopatry of parasites), a process sometimes coined as cocladogenesis (Thompson 1994). In that case, congruence between the species trees is only passive and has little to do with coevolution.

Finally, even if cospeciation leading to congruence seems to be mostly anecdotic, the topology of the host phylogeny has an effect in shaping the topology of the parasite phylogeny. Host switches, which seem to be an important mode of speciation in parasites, have been shown to occur preferentially between closely related host species, in associations as diverse as plants and their fungal parasites (Jackson 2004; Refrégier et al. 2008) to animal viruses (Charleston and Robertson 2002). This is in accordance with cross inoculation studies showing that parasites have a higher fitness on host related to their host of origin as evidenced among plant fungal parasites (de Vienne, unpublished results, Gilbert and Webb 2007), drosophila nematodes (Perlman and Jaenike 2003) or acanthocephalan hosted by cockroaches (Moore and Gotelli 1996). As a conclusion, cospeciation seems to be rare in natural host-parasite association, but cophylogenetic studies still remain interesting, showing that host phylogeny impacts parasite diversification (Bonfante and Genre 2008; Refrégier et al. 2008; Hibbett and Matheny 2009).

### **3. From short-term to long-term host-pathogen interactions: trans-specific genetic diversity and relationship between coevolution and ecological speciation (but not cospeciation)**

We have outlined above recent advances of the theory of coevolution and recent results found on polymorphism at genes involved in host and pathogen interactions in natural populations. We have then outlined the approaches used for testing for cospeciation, and the inferences regarding diversification of natural host-pathogen associations. This has already hinted at how coevolution and cospeciation occur at two very different time-scales and proceed by completely different processes. Cospeciation is, however, often confusingly called “coevolution” in the literature, the idea being that cospeciation should be the ultimate result of coevolution. We will attempt in this section to link the two time scales.

First, we will briefly review data regarding the footprints of selection acting across speciation events in host-parasite systems. Although species of pathogens can be found interspersed with non pathogens in phylogenies, the pathogenic lifestyle is often retained across several speciation events (Berbee 2001; James and et al. 2006), Coevolution may therefore act over the long term, although not necessarily on the same genes or under the same selective pressure in close species. Second, we will investigate the expected relationships between coevolution and pathogen speciation: we will review theoretical works on how coevolution can promote pathogen specialization and ecological speciation but we will see that such divergence can occur without generating a cospeciation pattern, as evidenced by both experimental and theoretical studies.

#### **3.1 Trans-specific polymorphism and divergence at the loci involved in coevolution**

Expected interspecific diversity at the loci involved in host-pathogen interaction depends upon coevolutionary processes in which these loci are involved. Under a long-term arms race, different alleles should be fixed in different species whereas under a trench warfare model, trans-specific polymorphism, i.e. shared polymorphism maintained in close species, can be produced if drift-type effects are not too strong (Hamilton 1993).

An increasing number of studies focuses on detecting genes under positive selection by comparing the number of synonymous versus non synonymous substitutions ( $d_n/d_s$ ) between closely related species (Nielsen 2005; Aguileta et al. 2009). Genes exhibiting such positive selection footprints have been found in hosts and parasites, with divergent alleles fixed in different species (i.e. evolving under diversifying selection). This pattern is expected under host-pathogen coevolution following the arms race model. Interestingly, one of the gene categories with the highest number of genes under positive selection in Primates is involved in immune system functions (Nielsen, 2005). In pathogens, where several examples pointed to an arms race occurring in the short term (section 1.3), the presence

of highly different alleles in related species was expected. Diversifying selection was in fact identified at the avirulence genes of the wheat fungal pathogens *Pyrenophora tritici*, *P. repentis* and *Phaeosphaeria nodorum* (Stukenbrock et al. 2007), *Melampsora spp.* (Van der Merwe et al. 2009) and in the oomycetes of the *Phytophthora* genus (Win et al. 2007). In animal pathogens, diversifying selection has been documented in *Plasmodium* (Putapornpit et al. 2008) and in canine viruses (Shackelton et al. 2005; McCarthy et al. 2007).

Trans-specific polymorphism (that may also be referred to as balanced polymorphism applying on large time scales) represents in contrast cases where multiple allelic classes are maintained in sister species by virtue of being more ancient than the species (Richman 2000). This is expected under the trench warfare model. In plants, trench warfare seems to be prevalent in the short term at loci involved in interaction with their pathogens (see section 1.3), so that trans-specific polymorphism could be expected. We found a single report of trans-specific polymorphism at resistance genes in plants, in the *Lactuca* genus (Kuang et al. 2004). However, introgression can also be responsible for observations of shared polymorphism (Armour et al. 2008), which was not formally excluded on *Lactuca*. Trans-specific polymorphism has been documented in some animal hosts at the Major Histocompatibility Complex (MHC) loci: in penguins (Kikkawa et al. 2009), primates (Bonhomme et al. 2008), rabbits (Su and Nei 1999) and amphibians (Nonaka et al. 2000). Trans-specific polymorphism has also been documented in some pathogens. One example is the genes controlling toxin production loci in *Fusarium graminearum*, a fungal species complex of pathogens on wheat and barley (Ward 2002).

Coevolution therefore has consequences on trans-specific polymorphism and divergence between pathogen species on the one hand and host species on the other end. This suggests that coevolution on a specific pair of loci is a long lasting process that can act across speciation events.

### 3.2 From coevolution to specialization, models and observations

The question remains whether coevolution could not only keep occurring during speciation events, but also drive pathogen divergence. We will first review experimental data and theoretical studies investigating whether coevolution can promote adaptation to different niches. More specifically, we will see how different populations specialize onto different hosts aided by coevolution, leading to host races, and then achieve speciation, i.e. become reproductively isolated.

*A priori*, one may think that any species should be selected for exploiting broad ecological niches, i.e. becoming a generalist, as this should diminish efforts in food collection as well as reduce intraspecific competition. Among pathogens, one could expect invasion of genotypes able to infect many different species, especially when host populations are patchy and temporally unstable.

However, broad host ranges are quite rare in nature (for an exception, see *Botrytis cinerea* cf Williamson et al. 2007). Relative paucity of generalist pathogens can be due to trade-offs between abilities to infect different host species. Such trade-off exist as shown by serial passage experiments: passaging a microbe onto a host species different from its original one recurrently leads to a decrease in fitness on the original host (Ebert 1998), and this property has been widely used to develop attenuated microbes that can serve as vaccines, for instance for hepatitis vaccine and for BCG vaccine protecting from some forms of tuberculosis (Karron et al. 1988; Mahairas et al. 1996). This can occur if alternative hosts differ at the exact site recognized by the pathogen, so that recognizing one of them impedes recognizing the other. However, in several studies of experimental evolution, both specialists and generalists seem to emerge (Little et al. 2006; Poullain et al. 2008). The instability in host abundance has been alternatively invoked as a factor explaining the evolution of generalists in natural systems (Jaenike 1990; Norton and Carpenter 1998) and has received recent experimental support (Soler et al. 2009).

How specialization can emerge in the absence of trade-off and in the presence of large host populations has been investigated theoretically. Specialization could evolve due to higher adaptation swiftness to each host species (Whitlock 1996; Kawecki 1998), a process again coined as “red queen dynamics” {Whitlock, 1996 #1950\; see also introduction for wider presentation of this concept}. The model by Kawecki (1998) considers a biallelic locus coding for specialization, one allele enabling specialization, the other generalization, and two loci controlling for infectivity, one for each host species. Simulations show that if recurrent selection for new alleles at the infectivity loci occurs due to coevolution, then specialization will be selected for due to the faster adaptation of specialized pathogens as compared to generalists. Indeed, selection acts every generation in specialized pathogens whereas it only proceeds part of the time on generalists distributed among several host species. This gives higher chance for specialized pathogens to invade. Additional simulations show that the model is robust to different assumptions on the genetic bases of infectivity (Kawecki 1998). In addition, once one species is specialized on a narrow niche, the other species suffer less competition in the complementary niches, so that preference for these other niches can indirectly be promoted (Whitlock 1996). As a summary, specialization, i.e. the formation of host races in pathogens, can be directly promoted by coevolution due to an impossibility to succeed onto several different hosts or due to higher adaptation swiftness of specialists, and indirectly due to competition with specialized pathogens.

### 3.3 Specialization and pathogen speciation, theoretical considerations

Divergence among pathogens via the evolution of host races leads to the emergence of specialist species only if reproductive isolation also evolves (Giraud et al. 2008). This corresponds to

ecological speciation where the species occupying different niches (i.e. hosts) become reproductively isolated one from another. The possibility of ecological speciation has been supported by many different studies on systems as different as herbivorous insects, vertebrates and plants (for a review, see Hendry 2007). Factors favoring the evolution of reproductive isolation among populations adapted to different ecological niches include low dispersal (Hendry et al. 2007), mating specifically on the niche on which the population is specialized (Rice 1984), either due to adapted behavior (Funk 1998), or to specific life history trait as for microbial pathogens mating within hosts after infection (Giraud et al. 2006), or to physical linkage between the loci controlling niche choice and mate choice (Slatkin 1996). For instance, pea aphids harbor tightly linked loci controlling respectively host preference and mating preference (Hawthorne and Via 2001); phytophagous insects are selected against mating with congeners feeding on a different plant species (Johnson et al. 1996; Nosil et al. 2002; Egan et al. 2008), and fungal ascomycete plant pathogens necessarily mate within their host plants, allowing the genes responsible for adaptation to pleiotropically cause reproductive isolation (Le Gac and Giraud 2008). As a result, pathogen specialization is expected to contribute to their diversification by speciation. The speed at which this speciation occurs (Huysse et al. 2005) of course relies on many parameters including pathogen generation time, host generation time, dispersal, and effective population size.

Pathogens hence tend to specialize, at least when host availability is not critical. As a consequence, they are expected to form two different species as the host lineage splits. This should lead to widespread cospeciation patterns, but solely provided pathogens long remain associated with one specific host lineage. This last hypothesis is rarely made explicit nor tested, and cospeciation is considered as the most likely scenario. Reasons for disruption of a host-pathogen pair are however numerous: pathogens may go extinct or become less numerous so that they can miss a host speciation event, they can duplicate within their hosts, or speciate by host shift. Extinctions should be quite frequent in parasites, due for instance to the evolution of resistance among hosts (Thrall et al. 1993). Extinctions (sorting events) are allowed and are often inferred to be numerous when using reconciliation methods (see for instance Refrégier et al, 2008). Their frequency may even be overestimated using these methods as well as that of parasite duplication. Indeed, if extinction took place on a host lineage and then this lineage was recolonized by a host shift, reconstructions favor a duplication event having occurred long before, even if this requires many more extinctions to reconcile the two phylogenies than the reconstruction involving a host shift and a single extinction. Duplication of parasites, i.e. intra-host speciation, is in contrast expected to be highly unlikely because it is the specialization onto different ecological niches that drives speciation (see above). Cases of intrahost speciation are nevertheless known, which can be due to specialization onto different niches within the host body (Simková et al. 2004), or to divergence speed being much higher in parasites than

in hosts so that transient allopatry insufficient to ensure host speciation may allow parasite speciation (Hendry et al. 2007).

Eventually, the only way to assess whether host shifts are less frequent in nature than cospeciation is to infer the frequency of past host shifts in natural host-parasite associations, to detect incipient host shifts and to assess the possibility of host shifts via cross-inoculations. We have seen in part 2.4 that host shifts seem prominent in many cases even when parasites are specialists (Lopez-Vaamonde et al. 2001; Roy 2001; Charleston and Robertson 2002; Braby and Trueman 2006; Refrégier et al. 2008). The possibility of host shifts is also suggested by the success of cross species infection was found in some cases to decrease with the genetic distance with the original host (Moore and Gotelli 1996; Perlman and Jaenike 2003; Gilbert and Webb 2007) but was clearly different from zero for a wide range of related host species in many systems. In other cases, the ability of one pathogen to infect novel hosts was found to be linked to phenotypic traits independent from host phylogeny (Clayton and Moore 1997; Bush et al. 2006), so that host shifts were again detected as a likely event. Finally, clear examples of incipient host shifts have been detected among plant fungal pathogens such as the anther smut (Antonovics et al. 2002; Hood et al. 2003; Lopez-Villavicencio et al. 2005), and among animal pathogens such as HIV virus infecting primates (Charleston and Robertson 2002) and the so-called swine fever H1N1 now infecting humans (McConnell 2009).

We thus reviewed evidence that host shifts may be prevalent among host-pathogen associations, as shown by inferences made from past histories of some associations, cross-species infection studies and the detection of incipient host shifts. One may ask in return how such host shifts can allow retaining the footprints of long-term coevolution described above (section 3.1). In fact, as already mentioned, host shifts mainly occur towards species related to the host-of-origin. The same loci can thus remain involved in host-pathogen coevolution in the novel association. As a consequence, it is likely that host shifts can occur with continued coevolution at the host and pathogen loci involved in the interaction.

General considerations of the long-term consequences of coevolution thus suggest that it favors pathogen diversification via specialization onto different host species, and that this trend towards specialization is more likely to be accompanied by host shifts and subsequent speciation rather than by cospeciation.

## Conclusion

Interaction between hosts and pathogens is an important topic of evolutionary biology, with implications in medicine and agriculture. Several lines of evidence have confirmed that both hosts and pathogens act as a selective pressure on their partner, which has been termed “coevolution”. Selection can in fact be detected at the loci involved in host-pathogen interactions. The way genetic diversity is

shaped at these loci follows either the arms race model or the trench warfare model. The reasons why a given host pathogen system would match more one model or the other still need to be better characterized by examination of a larger number of natural systems.

On a larger time scale, theoretical studies showed that coevolution can trigger specialization, and even ecological speciation. This process can occur relatively rapidly, in less than one hundred generations (Hendry et al. 2007). Such parasite speciation was long expected to follow the Fahrenholz rule of cospeciation (“parasite phylogeny mirrors that of the host”), but we have seen that theoretical considerations indicate that speciation following host shifts are as likely as cospeciation. In parallel, the previous studies suggesting the predominance of hosts and parasites cospeciation have lost robustness with the advent of larger sampling and more powerful tools to compare phylogenies. In many instances, parasites were found to have diverged more recently than their hosts, by host shifts. The only cases where cospeciations seem to have occurred are those for which hosts and parasites disperse at low rates. In that case, divergence of hosts and parasites may be due primarily to allopatry, and not due to the tightness of their interaction. It remains however true that hosts and parasites, especially when these are pathogenic, exert a strong selection pressure on their respective partners, which does have a great influence on their diversification as we outlined. Intriguingly, active speciation has been detected specifically in Primate lineages that host more parasites (Nunn et al. 2004), so that that this influence may well be reciprocal.

In conclusion, the concept of cospeciation has yielded the development of very interesting tools to compare phylogenies and to identify whether host shifts occur more frequently between more closely related hosts or between hosts with more similar ecological traits. It also has focused on systems with very interesting ecological features, but the basis of this concept, namely that tight physiological interaction leads to parallel speciation, has been largely invalidated. We suggest that the term “coevolution” should be used only to mean reciprocal selection pressure in host and pathogen systems. Links between coevolution and the pattern of pathogen and host speciation remains to be more fully explored.

### **Acknowledgements**

We acknowledge the grants ANR 06-BLAN-0201 and ANR 07-BDIV-003. AT acknowledges the Volkswagen Stiftung (I/82752) for financial support. We apologize to all those colleagues whose work we might have missed to cite in this article.





## Figure legends

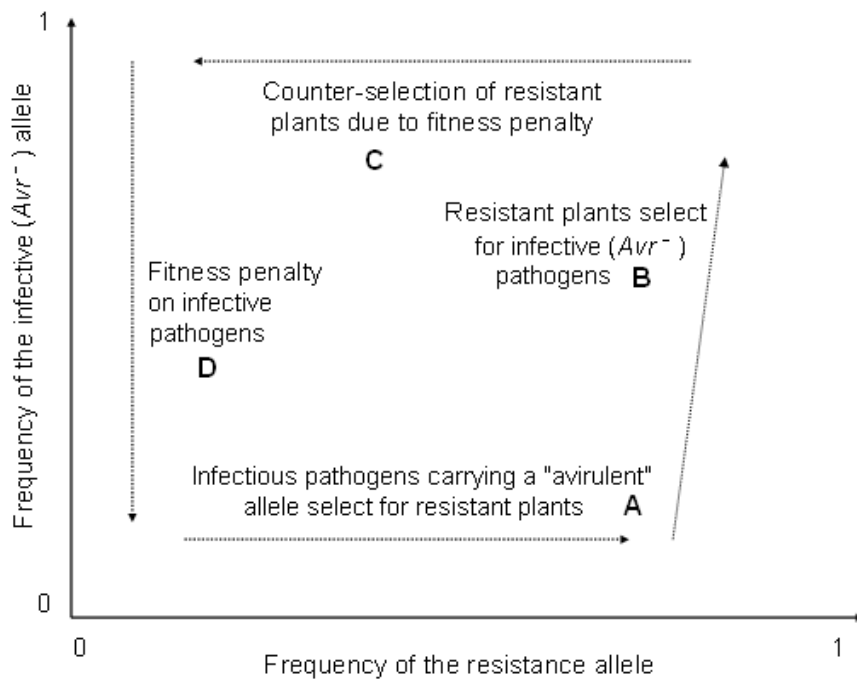
**Figure 1:** Natural selection in coevolving GFG plant-parasite interactions.

**Figure 2:** Unstable (a) and stable (b) cycles of coevolution in a GFG relationships.

**Figure 3:** Outcome of coevolution for unstable (a) and stable (b) cycles shown in Figure 2. The arms race model emerges from an unstable cycle of coevolution where recurrent selective sweeps occur in host and parasite populations (a). In a trench warfare model, long term polymorphism is kept following a balancing selection model emerging from stable cycles (b).

**Figure 4:** Six evolutionary events than can arise during the coevolution of host and parasites. Grey lines represent the host lineages; black lines represent the parasite lineages. A: cospeciation, B: host-shift, C: duplication, D: partial extinction or “missing the boat”, E: host-jump, F: total extinction (from Ronquist, 2003).

**Figure 1:** Natural selection in coevolving GFG plant-parasite interactions



**Figure 2:** Unstable (a) and stable (b) cycles of coevolution in a GFG relationships

Figure 2a

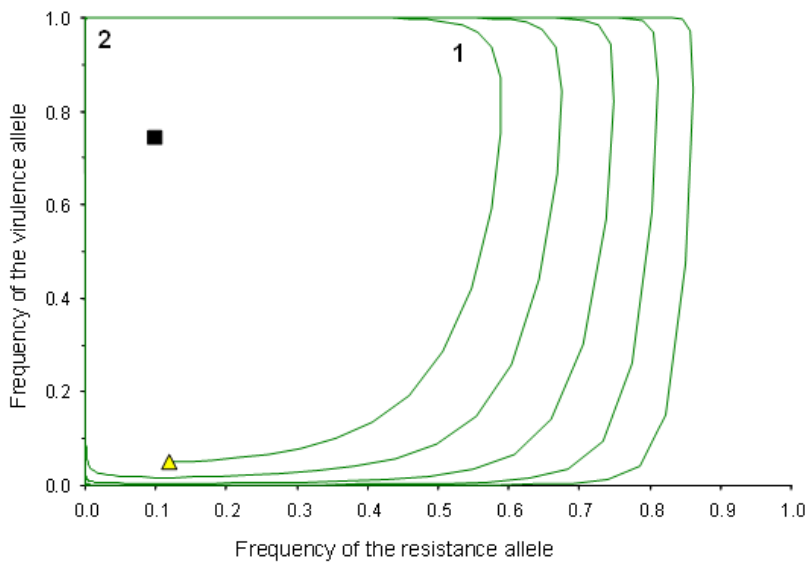
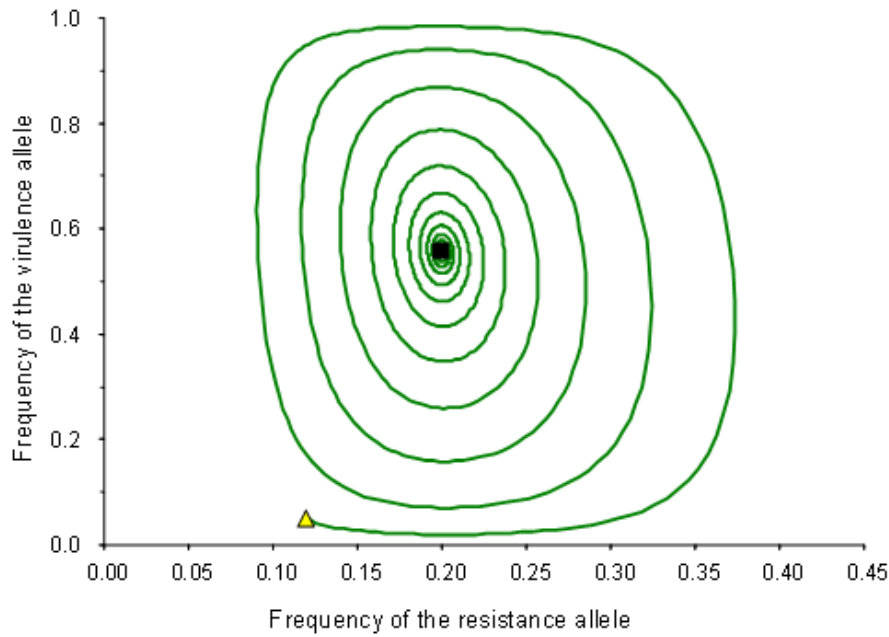


Figure 2b



**Figure 3:** Outcome of coevolution for unstable (a) and stable (b) cycles shown in Figure 2. The arms race model emerges from an unstable cycle of coevolution (a) where recurrent selective sweeps occur in host and parasite populations. In a trench warfare model, long term polymorphism is kept following a balancing selection model emerging from stable cycles (b).

Figure 3a

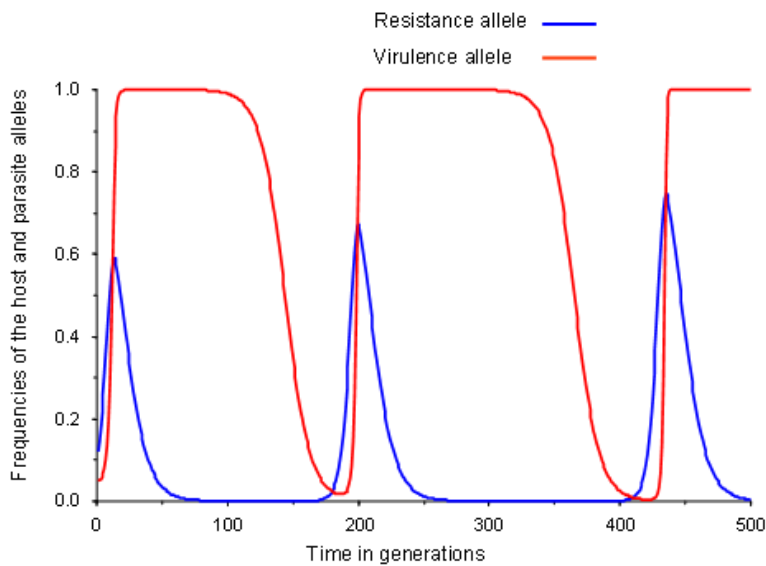
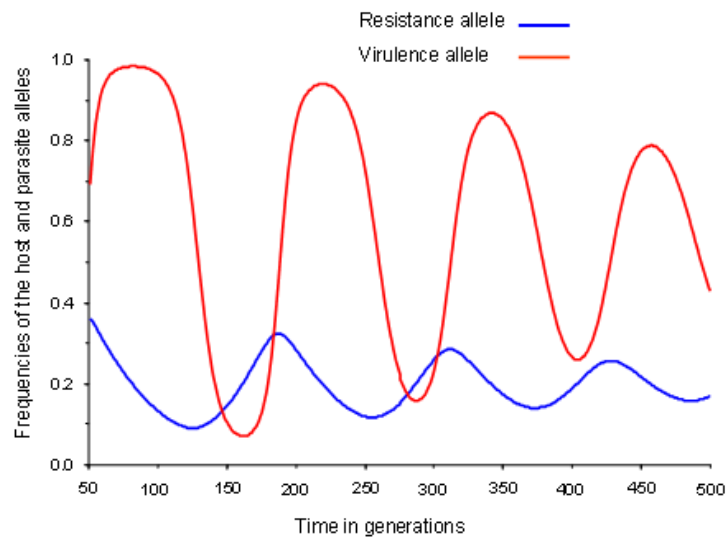


Figure 3b



**Figure 4**



- Agrawal, A. and C. M. Lively (2002). "Infection genetics: gene-for-gene versus matching-alleles models and all points in between." Evolutionary Ecology Research **4**(1): 79-90.
- Aguileta, G., G. Refrégier, R. Yockteng, E. Fournier and T. Giraud (2009). "Rapidly evolving genes in pathogens: Methods for detecting positive selection and examples among fungi, bacteria, viruses and protists." Infection, Genetics and Evolution **9**: 656-670.
- Alizon, S. and A. H. N. M. M. V. Baalen (2009). "Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future." JEB **9999**(9999).
- Allen, J. C. (1975). "Mathematical models of species interactions in time and space." The American Naturalist **109**: 319-342.
- Allen, R. L., P. D. Bittner-Eddy, L. J. Grenville-Briggs, J. C. Meitz, A. P. Rehmany, L. E. Rose and J. L. Beynon (2004). "Host-parasite coevolutionary conflict between Arabidopsis and downy mildew." Science **306**(5703): 1957-1960.
- Antonovics, J., M. Hood and J. Partain (2002). "The ecology and genetics of a host shift: Microbotryum as a model system." American Naturalist **160**: S40-S53.
- Armour, D. J., J. M. Mackie, J. M. Musial and J. A. G. Irwin (2008). "Transfer of anthracnose resistance and pod coiling traits from Medicago arborea to M-sativa by sexual reproduction." Theoretical and Applied Genetics **117**(2): 149-156.
- Bahri, B., O. Kaltz, M. Leconte, C. de Vallavieille-Pope and J. Enjalbert (2009). "Tracking costs of virulence in natural populations of the wheat pathogen, *Puccinia striiformis* f. sp. *tritici*." BMC Evolutionary Biology **9**(26).
- Bakker, E. G., C. Toomajian, M. Kreitman and J. Bergelson (2006). "A genome-wide survey of R gene polymorphisms in Arabidopsis." Plant Cell **18**(8): 1803-1818.
- Barreiro, L. B., G. Laval, H. Quach, E. Patin and L. Quintana-Murci (2008). "Natural selection has driven population differentiation in modern humans." Nature Genetics **40**(3): 340-345.
- Barrett, J. A. (1980). "Pathogen Evolution in Multilines and Variety Mixtures." Zeitschrift Fur Pflanzenkrankheiten Und Pflanzenschutz-Journal of Plant Diseases and Protection **87**(7): 383-396.
- Barrett, J. A. (1988). "Frequency-Dependent Selection in Plant Fungal Interactions." Philosophical Transactions of the Royal Society of London Series B-Biological Sciences **319**(1196): 473-483.
- Beaumont, M. A. (2005). "Adaptation and speciation: what can F-st tell us?" Trends in Ecology & Evolution **20**(8): 435-440.

- Beaumont, M. A. and D. J. Balding (2004). "Identifying adaptive genetic divergence among populations from genome scans." Molecular Ecology **13**(4): 969-980.
- Berbee, M. L. (2001). "The phylogeny of plant and animal pathogens in the Ascomycota." Physiological and Molecular Plant Pathology **59**: 165-187.
- Bergelson, J., M. Kreitman, E. A. Stahl and D. C. Tian (2001). "Evolutionary dynamics of plant R-genes." Science **292**(5525): 2281-2285.
- Bergelson, J. and C. B. Purrington (1996). "Surveying patterns in the cost of resistance in plants." American Naturalist **148**(3): 536-558.
- Birch, P. R. J., A. P. Rehmany, L. Pritchard, S. Kamoun and J. L. Beynon (2006). "Trafficking arms: oomycete effectors enter host plant cells." Trends in Microbiology **14**(1): 8-11.
- Bonfante, P. and A. Genre (2008). "Plants and arbuscular mycorrhizal fungi: an evolutionary-developmental perspective." Trends in Plant Science **13**(9): 492-498.
- Bonhomme, M., G. G. M. Doxiadis, C. M. C. Heijmans, V. Vervoort, N. Otting, R. E. Bontrop and B. Crouau-Roy (2008). "Genomic plasticity of the immune-related Mhc class I B region in macaque species." Bmc Genomics **9**.
- Braby, M. F. and J. W. H. Trueman (2006). "Evolution of larval host plant associations and adaptive radiation in pierid butterflies." Journal of Evolutionary Biology **19**(5): 1677-1690.
- Brooks, D. R. (1981). "Hennig's parasitological method: A proposed solution." Systematic Zoology **30**: 229-249.
- Brooks, D. R. and D. H. McLennan (1991). Phylogeny, Ecology, and Behavior: A Research Program in Comparative Biology. Chicago, University of Chicago Press.
- Brown, J. K. M. (2003). "A cost of disease resistance: paradigm or peculiarity?" Trends in Genetics **19**(12): 667-671.
- Buckling, A. and P. B. Rainey (2002). "Antagonistic coevolution between a bacterium and a bacteriophage." Proc. R. Soc. Lon. B **269**: 931-936.
- Bush, S. E., E. Sohn and D. H. Clayton (2006). "Ecomorphology of parasite attachment: Experiments with feather lice." Journal of Parasitology **92**(1): 25-31.
- Caicedo, A. L. and B. A. Schaal (2004). "Heterogeneous evolutionary processes affect R gene diversity in natural populations of *Solanum pimpinellifolium*." Proceedings of the National Academy of Sciences of the United States of America **101**(50): 17444-17449.
- Cairns, B. J., A. R. Timms, V. A. A. Jansen, I. F. Connerton and R. J. H. Payne (2009). "Quantitative Models of In Vitro Bacteriophage-Host Dynamics and Their Application to Phage Therapy." Plos Pathogens **5**(1).
- Charleston, M. A. (1998). "Jungles: a new solution to the host/parasite phylogeny reconciliation problem." Mathematical Biosciences **149**: 191-223.

- Charleston, M. A. and S. L. Perkins (2003). Lizards, malaria, and jungles in the Caribbean. Tangled Trees: Phylogeny, Cospeciation and Coevolution. R. D. M. Page. Chicago, University of Chicago Press: 65-92.
- Charleston, M. A. and D. L. Robertson (2002). "Preferential host switching by primate lentiviruses can account for phylogenetic similarity with the primate phylogeny." Systematic Biology **51**: 528-535.
- Clayton, D. H. and J. Moore (1997). Host-Parasite Evolution, General principles and avian models. New-York, USA, Oxford University Press Inc.
- Clayton, D. H. and J. Moore (1997). Introduction. Host-Parasite Evolution, General principles and avian models. New-York, USA, Oxford University Press Inc.: 1-7.
- Conway, D. J., D. R. Cavanagh, K. Tanabe, C. Roper, Z. S. Mikes, N. Sakihama, K. A. Bojang, A. M. J. Oduola, P. G. Kremsner, D. E. Arnot, B. M. Greenwood and J. S. McBride (2000). "A principal target of human immunity to malaria identified by molecular population genetic and immunological analyses." Nature Medicine **6**(6): 689-692.
- Currie, C. R., B. Wong, A. E. Stuart, T. R. Schultz, S. A. Rehner, U. G. Mueller, G. H. Sung, J. W. Spatafora and N. A. Straus (2003). "Ancient tripartite coevolution in the attine ant-microbe symbiosis." Science **299**(5605): 386-388.
- Damgaard, C. (1999). "Coevolution of a plant host-pathogen gene-for-gene system in a metapopulation model without cost of resistance or cost of virulence." Journal of Theoretical Biology **201**(1): 1-12.
- Dangl, J. L. and J. D. G. Jones (2001). "Plant pathogens and integrated defence responses to infection." Nature **411**(6839): 826-833.
- De Meaux, J., I. Cattani-Toupance, C. Lavigne, T. Langin and C. Neema (2003). "Polymorphism of a complex resistance gene candidate family in wild populations of common bean (*Phaseolus vulgaris*) in Argentina: comparison with phenotypic resistance polymorphism." Molecular Ecology **12**(1): 263-273.
- de Meaux, J. and C. Neema (2003). "Spatial patterns of diversity at the putative recognition domain of resistance gene candidates in wild bean populations." Journal of Molecular Evolution **57**: S90-S102.
- de Vienne, D. M., T. Giraud and O. C. Martin (2007). "A congruence index for testing topological similarity between trees." Bioinformatics **23**(23): 3119-3124.
- de Vienne, D. M., T. Giraud and O. C. Martin (2009). "In response to comment on 'A congruence index for testing topological similarity between trees'." Bioinformatics **25**(1): 150-151.
- Decaestecker, E., S. Gaba, J. A. M. Raeymaekers, R. Stoks, L. Van Kerckhoven, D. Ebert and L. De Meester (2007). "Host-parasite 'Red Queen' dynamics archived in pond sediment." Nature **450**: 870-873.



- Desveaux, D., A. U. Singer, A. J. Wu, B. C. McNulty, L. Musselwhite, Z. Nimchuk, J. Sondek and J. L. Dangl (2007). "Type III effector activation via nucleotide binding, phosphorylation, and host target interaction." *Plos Pathogens* **3**(3).
- Ding, J., W. L. Zhang, Z. Q. Jing, J. Q. Chen and D. C. Tian (2007). "Unique pattern of R-gene variation within populations in Arabidopsis." *Molecular Genetics and Genomics* **277**(6): 619-629.
- Downey, J. C. (1962). "Host-plant relations as data for butterfly classification." *Systematic Zoology* **11**(4): 150-159.
- Dres, M. and J. Mallet (2002). "Host races in plant-feeding insects and their importance in sympatric speciation." *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **357**(1420): 471-492.
- Ebert, D. (1998). "Evolution - Experimental evolution of parasites." *Science* **282**(5393): 1432-1435.
- Egan, S. P., P. Nosil and D. J. Funk (2008). "SELECTION AND GENOMIC DIFFERENTIATION DURING ECOLOGICAL SPECIATION: ISOLATING THE CONTRIBUTIONS OF HOST ASSOCIATION VIA A COMPARATIVE GENOME SCAN OF NEOCHLAMISUS BEBBIANAE LEAF BEETLES." *Evolution* **62**(5): 1162-1181.
- Fahrenheit, H. (1913). "Ectoparasiten und Abstammungslehre." *Zoologischer Anzeiger* **41**: 371-374.
- Fay, J. C. and C.-I. Wu (2001). "The neutral theory in the genomic era." *Current Opinion in Genetics and Development* **11**: 642-646.
- Felsenstein, J. (1985). "Confidence-limits on phylogenies - An approach using the bootstrap." *Evolution* **39**(4): 783-791.
- Felsenstein, J. (1985). "Phylogenies and the Comparative Method." *The American Naturalist* **125**(1): 1.
- Flor, H. H. (1956). "The Complementary Genic Systems in Flax and Flax Rust." *Advances in Genetics Incorporating Molecular Genetic Medicine* **8**: 29-54.
- Frank, S. A. (1992). "Models of plant pathogen coevolution." *Trends in Genetics* **8**(6): 213-219.
- Funk, D. J. (1998). "Isolating a role for natural selection in speciation: host adaptation and sexual isolation in *Neochlamisus bebbianae* leaf beetles." *Evolution* **52**: 1744-1759.
- Gandon, S. (2002). "Local adaptation and the geometry of host-parasite coevolution." *Ecology Letters* **5**(2): 246-256.
- Gandon, S., A. Buckling, E. Decaestecker and T. Day (2008). "Host-parasite coevolution and patterns of adaptation across time and space." *Journal of Evolutionary Biology* **21**(6): 1861-1866.
- Garcia-Arenal, F., A. Fraile and J. M. Malpica (2003). "Variation and evolution of plant virus populations." *International Microbiology* **6**(4): 225-232.
- Gavrilets, S. and Y. Michalakis (2008). "Effects of environmental heterogeneity on victim-exploiter coevolution." *Evolution* **62**: 3100-3116.

- Gerardo, N., U. Mueller and C. Currie (2006). "Complex host-pathogen coevolution in the Apterostigma fungus-growing ant-microbe symbiosis." BMC Evolutionary Biology **6**(1): 88.
- Gilbert, G. S. and C. O. Webb (2007). "Phylogenetic signal in plant pathogen-host range." Proceedings of the National Academy of Sciences **104**(12): 4979-4983.
- Giraud, T., G. Refregier, M. Le Gac, D. M. de Vienne and M. E. Hood (2008). "Speciation in fungi." Fungal Genetics and Biology **45**(6): 791-802.
- Giraud, T., L. M. M. A. Villareal, F. Austerlitz, M. Le Gac and C. Lavigne (2006). "Importance of the Life Cycle in Sympatric Host Race Formation and Speciation of Pathogens." Phytopathology **96**(3): 280-287.
- Hafner, M. S., J. W. Demastes, T. A. Spradling and D. L. Reed (2003). Cophylogeny Between Pocket Gophers and Chewing Lice. Tangled trees. R. D. M. Page. Chicago, The University of Chicago Press: 195-220.
- Hafner, M. S., P. D. Sudman, F. X. Villablanca, T. A. Spradling, J. W. Demastes and S. A. Nadler (1994). "Disparate rates of molecular evolution in cospeciating hosts and parasites." Science **265**: 1087-1090.
- Haldane, J. B. S. (1954). The statistics of evolution. Evolution as a process. J. Huxley, A. C. Hardy and E. B. Ford. London, Allen and Unwin: 190-121.
- Hamilton, W. D. (1993). "Haploid Dynamic Polymorphism in a Host with Matching Parasites - Effects of Mutation Subdivision, Linkage, and Patterns of Selection." Journal of Heredity **84**(5): 328-338.
- Hamilton, W. D., R. Axelrod and R. Tanese (1990). "Sexual Reproduction as an Adaptation to Resist Parasites (a Review)." Proceedings of the National Academy of Sciences of the United States of America **87**(9): 3566-3573.
- Hanifin, C. T., E. D. J. Brodie and E. D. I. Brodie (2008). "Phenotypic Mismatches Reveal Escape from Arms-Race Coevolution." PLoS Biol **6**(3): e60.
- Hawthorne, D. J. and S. Via (2001). "Genetic linkage of ecological specialization and reproductive isolation in pea aphids." Nature **412**: 904-907.
- Hendry, A. P., P. Nosil and L. H. Rieseberg (2007). "The speed of ecological speciation." Functional Ecology **21**(3): 455-464.
- Hibbett, D. and P. B. Matheny (2009). The relative ages of ectomycorrhizal mushrooms and their plant hosts estimated using Bayesian relaxed molecular clock analyses. **7**: 13.
- Hoberg, E. P., D. R. Brooks and D. Siegel-Causey (1997). Host-parasite co-speciation: history, principles, and prospects. Host-Parasite Evolution, General principles and avian models. New-York, USA, Oxford University Press Inc.: 212-235.
- Holub, E. B. (2001). "The arms race is ancient history in *Arabidopsis*, the wildflower." Nature Reviews Genetics **2**(7): 516-527.

- Hood, M. E., J. Antonovics and H. Heishman (2003). "Karyotypic similarity identifies multiple host-shifts of a pathogenic fungus in natural populations." Infection, Genetics and Evolution **2**(3): 167-172.
- Huelsenbeck, J. P., B. Rannala and B. Larget (2000). "A bayesian framework for the analysis of cospeciation." Evolution **54**(2): 352-364.
- Huelsenbeck, J. P., B. Rannala and B. Larget (2003). A statistical perspective for reconstructing the history of host-parasite associations. Tangled trees. R. D. M. Page. Chicago, The University of Chicago Press: 93-119.
- Huelsenbeck, J. P., B. Rannala and Z. Yang (1997). "Statistical tests of host-parasite cospeciation." Evolution **51**(2): 410-419.
- Huyse, T., R. Poulin and A. Théron (2005). "Speciation in parasites: a population genetics approach." Trends in parasitology **21**: 469-475.
- Jackson, A. P. (2004). "A reconciliation analysis of host switching in plant-fungal symbioses." Evolution **58**(9): 1909-1923.
- Jaenike, J. (1990). "HOST SPECIALIZATION IN PHYTOPHAGOUS INSECTS." Annual Review of Ecology and Systematics **21**: 243-273.
- James, T. and et al. (2006). "Reconstructing the early evolution of Fungi using a six-gene phylogeny." Nature **443**: 818-822.
- Jensen, K. H., T. Little, A. Skorping and D. Ebert (2006). "Empirical support for optimal virulence in a castrating parasite." Plos Biology **4**(7): 1265-1269.
- Johnson, P. A., F. C. Hoppensteadt, J. J. Smith and G. L. Bush (1996). "Conditions for sympatric speciation: a diploid model incorporating habitat fidelity and non-habitat assortative mating." Evolutionary Ecology **10**: 187-205.
- Jones, J. D. G. and J. L. Dangl (2006). "The plant immune system." Nature **444**(7117): 323-329.
- Kaltz, O. and J. A. Shykoff (1998). "Local adaptation in host-parasite systems." Heredity **81**: 361-370.
- Karron, R. A., R. Daemer, J. Ticehurst, E. Dhondt, H. Popper, K. Mihalik, J. Phillips, S. Feinstone and R. H. Purcell (1988). "STUDIES OF PROTOTYPE LIVE HEPATITIS-A VIRUS-VACCINES IN PRIMATE MODELS." Journal of Infectious Diseases **157**(2): 338-345.
- Kawecki, T. J. (1998). "Red Queen Meets Santa Rosalia: Arms Races and the Evolution of Host Specialization in Organisms with Parasitic Lifestyles." Am. Nat. **152**: 635-651.
- Kellogg, V. L. (1913). "Distribution and species-forming of ecto-parasites." American Naturalist **47**: 129-158.
- Kikkawa, E. F., T. T. Tsuda, D. Sumiyama, T. K. Naruse, M. Fukuda, M. Kurita, R. P. Wilson, Y. LeMaho, G. D. Miller, M. Tsuda, K. Murata, J. K. Kulski and H. Inoko (2009). "Trans-species polymorphism of the Mhc class II DRB-like gene in banded penguins (genus *Spheniscus*)." Immunogenetics **61**(5): 341-352.

- Kim, Y. and W. Stephan (2002). "Detecting a local signature of genetic hitchhiking along a recombining chromosome." Genetics **160**(2): 765-777.
- Kuang, H., S. S. Woo, B. C. Meyers, E. Nevo and R. W. Michelmore (2004). "Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce." Plant Cell **16**(11): 2870-2894.
- Kupczok, A. and A. von Haeseler (2008). "Comment on "A congruence index for testing topological similarity between trees"." Bioinformatics **in press**.
- Laine, A. L. and A. Tellier (2008). "Heterogeneous selection promotes maintenance of polymorphism in host-parasite interactions." Oikos **117**(9): 1281-1288.
- Le Gac, M. and T. Giraud (2008). "Existence of a pattern of reproductive character displacement in Basidiomycota but not in Ascomycota." J Evol Biol **21**: 761-772.
- Legendre, P., Y. Desdevises and E. Bazin (2002). "A Statistical Test for Host-Parasite Coevolution." Systematic Biology **51**(2): 217 - 234.
- Leonard, K. J. (1977). "Selection Pressures and Plant Pathogens." Annals of the New York Academy of Sciences **287**(FEB25): 207-222.
- Leonard, K. J. and R. J. Czochoz (1980). "Theory of Genetic Interactions among Populations of Plants and Their Pathogens." Annual Review of Phytopathology **18**: 237-258.
- Li, H. and W. Stephan (2005). "**Maximum-Likelihood Methods for Detecting Recent Positive Selection and Localizing the Selected Site in the Genome.**" Genetics **171**: 377-384.
- Li, H. P. and W. Stephan (2005). "Maximum-likelihood methods for detecting recent positive selection and localizing the selected site in the genome." Genetics **171**(1): 377-384.
- Light, J. E. and M. S. Hafner (2008). "Codivergence in heteromyid rodents (Rodentia : Heteromyidae) and their sucking lice of the genus Fahrenholzia (Phthiraptera : Anoplura)." Systematic Biology **57**(3): 449-465.
- Little, T. J., K. Watt and D. Ebert (2006). "Parasite-host specificity: Experimental studies on the basis of parasite adaptation." Evolution **60**(1): 31-38.
- Liu, Z., J. I. B. Bos, M. Armstrong, S. C. Whisson, L. d. Cunha, T. Torto-Alalibo, J. Win, A. O. Avrova, F. Wright, P. R. J. Birch and S. Kamoun (2005). "Patterns of Diversifying Selection in the Phytotoxin-like scr74 Gene Family of *Phytophthora infestans*." Molecular Biology and Evolution **22**(3): 659-672.
- Loegering, W. Q. and A. H. Ellingboe (1987). "FLOR, H.H. - PIONEER IN PHYTOPATHOLOGY." Annual Review of Phytopathology **25**: 59-66.
- Lopez-Pascua, L. D. C. and A. Buckling (2008). "Increasing productivity accelerates host-parasite coevolution." Journal of Evolutionary Biology **21**(3): 853-860.
- Lopez-Vaamonde, C., H. Charles, J. Godfray and J. M. Cook (2003). "Evolutionary dynamics of host-plant use in a genus of leaf-mining moths." Evolution **57**(8): 1804-1821.

- Lopez-Vaamonde, C., J.-Y. Rasplus, G. D. Weiblen and J. M. Cook (2001). "Molecular phylogenies of fig-wasps: partial cocladogenesis of pollinators and parasites." *Molecular Phylogenetics and Evolution* **21**(1): 55-71.
- Lopez-Villavicencio, M., J. Enjalbert, M. E. Hood, J. A. Shykoff, C. Raquin and T. Giraud (2005). "The anther smut disease on *Gypsophila repens*: a case of parasite sub-optimal performance following a recent host shift?" *Journal of Evolutionary Biology* **18**(5): 1293-1303.
- Mahairas, G. G., P. J. Sabo, M. J. Hickey, D. C. Singh and C. K. Stover (1996). "Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent M-bovis." *Journal of Bacteriology* **178**(5): 1274-1282.
- Mauricio, R., E. A. Stahl, T. Korves, D. C. Tian, M. Kreitman and J. Bergelson (2003). "Natural selection for polymorphism in the disease resistance gene *Rps2* of *Arabidopsis thaliana*." *Genetics* **163**(2): 735-746.
- Maynard Smith, J. and J. Haig (1974). "The hitch-hiking effect of a favourable gene." *Genet. Res. (Cambridge)* **23**: 23-35. .
- McCarthy, A. J., M. A. Shaw and S. J. Goodman (2007). "Pathogen evolution and disease emergence in carnivores." *Proceedings of the Royal Society B-Biological Sciences* **274**(1629): 3165-3174.
- McConnell, J. (2009). "Influenza begs many questions." *Lancet* **373**(9675): 1590-1590.
- Meyers, B. C., D. B. Chin, K. A. Shen, S. Sivaramakrishnan, D. O. Lavelle, Z. Zhang and R. W. Michelmore (1998). "The major resistance gene cluster in lettuce is highly duplicated and spans several megabases." *Plant Cell* **10**(11): 1817-1832.
- Meyers, B. C., A. Kozik, A. Griego, H. H. Kuang and R. W. Michelmore (2003). "Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*." *Plant Cell* **15**(4): 809-834.
- Michelmore, R. W. and B. C. Meyers (1998). "Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process." *Genome Research* **8**(11): 1113-1130.
- Misas-Villamil, J. C. and R. A. L. v. d. Hoorn (2008). "Enzyme-inhibitor interactions at the plant-pathogen interface." *Current Opinion in Plant Biology* **11**: 380-388.
- Moeller, D. A. and P. Tiffin (2005). "Genetic diversity and the evolutionary history of plant immunity genes in two species of *Zea*." *Molecular Biology and Evolution* **22**(12): 2480-2490.
- Moore, J. and N. J. Gotelli (1996). "Evolutionary Patterns of Altered Behavior and Susceptibility in Parasitized Hosts." *Evolution* **50**(2): 807-819.
- Mueller, U. G., D. Dash, C. Rabeling and A. Rodrigues (2008). "COEVOLUTION BETWEEN ATTINE ANTS AND ACTINOMYCETE BACTERIA: A REEVALUATION." *Evolution* **62**(11): 2894-2912.
- Müller, C. B. and J. Krauss (2005). "Symbiosis between grasses and asexual fungal endophytes." *Current Opinion in Plant Biology* **8**: 450-456.

- Nielsen, R. (2005). "Molecular signatures of natural selection." Annual Review of Genetics **39**: 197-218.
- Nonaka, M., C. Yamada-Namikawa, M. F. Flajnik and L. Du Pasquier (2000). "Trans-species polymorphism of the major histocompatibility complex-encoded proteasome subunit LMP7 in an amphibian genus, *Xenopus*." Immunogenetics **51**(3): 186-192.
- Norton, D. A. and M. A. Carpenter (1998). "Mistletoes as parasites: Host specificity and speciation." Trends in Ecology & Evolution **13**(3): 101-105.
- Nosil, P., B. J. Crespi and C. P. Sandoval (2002). "Host-plant adaptation drives the parallel evolution of reproductive isolation." Nature **417**: 440-443.
- Nunn, Charles L., S. Altizer, W. Sechrest, Kate E. Jones, Robert A. Barton and John L. Gittleman (2004). Parasites and the Evolutionary Diversification of Primate Clades. **164**: S90-S103.
- Orgil, U., H. Arakit, S. Tangchaiburana, R. Berkey and S. Xiao (2007). "Intraspecific genetic variations, fitness cost and benefit of RPW8, a disease resistance locus in *Arabidopsis thaliana*." Genetics **176**(4): 2317-2333.
- Page, R. D. M. (1990). "Component analysis: A valiant failure? ." Cladistics **6**: 119-136.
- Page, R. D. M. (1993). User's manual for COMPONENT, Version 2.0., The Natural History Museum, London.
- Page, R. D. M. (1994). "Parallel phylogenies: reconstructing the history of host-parasite assemblages." Cladistics **10**: 155-173.
- Page, R. D. M. (1996). "Temporal congruence revisited: comparison of mitochondrial DNA sequence divergence in cospeciating pocket gophers and their chewing lice." Systematic Biology **45**(2): 151-167.
- Page, R. D. M. (2003). Tangled trees. Phylogeny, cospeciation and coevolution. Chicago, The University of Chicago Press.
- Page, R. D. M. and M. A. Charleston (1998). "Trees within trees: phylogeny and historical associations." Trends in Ecology and Evolution **13**(9): 356-359.
- Palomino, M. M., B. C. Meyers, R. W. Michelmore and B. S. Gaut (2002). "Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*." Genome Research **12**(9): 1305-1315.
- Pei, M. H., C. Bayon and C. Ruiz (2005). "Phylogenetic relationships in some *Melampsora* rusts on Salicaceae assessed using rDNA sequence information." Mycological Research **109**(4): 401-409.
- Pepin, K. M., J. Domsic and R. McKenna (2008). "Genomic evolution in a virus under specific selection for host recognition." Infection Genetics and Evolution **8**(6): 825-834.
- Perlman, S. J. and J. Jaenike (2003). "Infection success in novel hosts: an experimental and phylogenetic study of *Drosophila*-parasitic nematodes." Evolution Int J Org Evolution **57**(3): 544-57.

- Poulin, R. and S. Morand (2004). Parasite Biodiversity. Washington.
- Poullain, V., S. Gandon, M. A. Brockhurst, A. Buckling and M. E. Hochberg (2008). "The evolution of specificity in evolving and coevolving antagonistic interactions between a bacteria and its phage." Evolution **62**(1): 1-11.
- Proctor, H. and I. Owens (2000). "Mites and birds: diversity, parasitism and coevolution." Trends in Ecology & Evolution **15**(9): 358-364.
- Putaporntip, C., S. Seethamchai, V. Suvannadhat, T. Hongsrimuanga, J. Sattabongkot and S. Jongwutiwes (2008). "Selective pressure on the merozoite surface protein-1 genes of *Plasmodium vivax*, *P. knowlesi* and *P. cynomolgi*." Asian Biomed. **2**(2): 123-134.
- Refrégier, G., M. Le Gac, F. Jabbour, A. Widmer, J. A. Shykoff, R. Yockteng, M. E. Hood and T. Giraud (2008). "Cophylogeny of the Anther Smut Fungi and their Caryophyllaceous Hosts: Prevalence of Host Shifts and Importance of Delimiting Parasite Species for Inferring Cospeciation " BMC Evolutionary Biology **In press**.
- Rice, W. R. (1984). "Disruptive selection on habitat preference and the evolution of reproductive isolation: a simulation study." Evolution **38**: 1251-1260.
- Richman, A. (2000). "Evolution of balanced genetic polymorphism." Molecular Ecology **9**(12): 1953-1963.
- Ronquist, F. (1995). "Reconstructing the history of host-parasite associations using generalised parsimony." Cladistics **11**: 73-89.
- Ronquist, F. (2003). Parsimony analysis of coevolving species associations. Tangled trees. R. D. M. Page. Chicago, The University of Chicago Press: 22-64.
- Rose, L. E., C. H. Langley, A. J. Bernal and R. W. Michelmore (2005). "Natural variation in the Pto pathogen resistance gene within species of wild tomato (*Lycopersicon*). I. Functional analysis of Pto alleles." Genetics **171**(1): 345-357.
- Rose, L. E., R. W. Michelmore and C. H. Langley (2007). "Natural variation in the Pto pathogen resistance gene within species of wild tomato (*Lycopersicon*). II. Population genetics of Pto." Genetics **175**: 1307-1319.
- Roy, B. A. (2001). "Patterns of association between crucifers and their flower-mimic pathogens: host-jumps are more common than coevolution or cospeciation." Evolution **55**(1): 41-53.
- Sacristan, S. and F. Garcia-Arenal (2008). "The evolution of virulence and pathogenicity in plant pathogen populations." Molecular Plant Pathology **9**(3): 369-384.
- Salathe, M., A. Scherer and S. Bonhoeffer (2005). "Neutral drift and polymorphism in gene-for-gene systems." Ecology Letters **8**(9): 925-932.
- Salvaudon, L., T. Giraud and J. A. Shykoff (2008). "Genetic diversity in natural populations: a fundamental component of plant-microbe interactions." Current Opinion in Plant Biology **11**: 1-9.

- Salvaudon, L., V. Heraudet and J. A. Shykoff (2005). "Parasite-host fitness trade-offs change with parasite identity: Genotype-specific interactions in a plant-pathogen system." Evolution **59**(12): 2518-2524.
- Salvaudon, L., V. Heraudet and J. A. Shykoff (2007). "Genotype-specific interactions and the trade-off between host and parasite fitness." Bmc Evolutionary Biology **7**.
- Sasaki, A. (2000). "Host-parasite coevolution in a multilocus gene-for-gene system." Proceedings of the Royal Society of London Series B-Biological Sciences **267**(1458): 2183-2188.
- Sasaki, A., W. D. Hamilton and F. Ubeda (2002). "Clone mixtures and a pacemaker: new facets of Red-Queen theory and ecology." Proceedings of the Royal Society of London Series B-Biological Sciences **269**(1493): 761-772.
- Savile, D. B. O. (1979). "Fungi as aids in higher plant classification." Botanical Review **45**(4): 377-503.
- Schardl, C. L., K. D. Craven, S. Speakman, A. Stromberg, A. Lindstrom and R. Yoshida (2008). "A novel test for host-symbiont codivergence indicates ancient origin of fungal endophytes in grasses." Systematic Biology **57**(3): 483-498.
- Shackelton, L. A., C. R. Parrish, U. Truyen and E. C. Holmes (2005). High rate of viral evolution associated with the emergence of carnivore parvovirus. **102**: 379-384.
- Shen, J. D., H. Araki, L. L. Chen, J. Q. Chen and D. C. Tian (2006). "Unique evolutionary mechanism in R-genes under the presence/absence polymorphism in *Arabidopsis thaliana*." Genetics **172**(2): 1243-1250.
- Simková, A., S. Morand, E. Jobet, M. Gelnar and O. Verneau (2004). "**MOLECULAR PHYLOGENY OF CONGENERIC MONOGENEAN PARASITES (DACTYLOGYRUS): A CASE OF INTRAHOST SPECIATION.**" Evolution **58**(5): 1001-1018.
- Slatkin, M. (1996). "In defense of founder-flush theories of speciation." The American Naturalist **147**(4): 493-505.
- Smith, C. I., W. K. W. Godsoe, S. Tank, J. B. Yoder and O. Pellmyr (2008). "DISTINGUISHING COEVOLUTION FROM COVICARIANCE IN AN OBLIGATE POLLINATION MUTUALISM: ASYNCHRONOUS DIVERGENCE IN JOSHUA TREE AND ITS POLLINATORS." Evolution **62**(10): 2676-2687.
- Soler, M., M. M. Vivaldi and A. P. Moller (2009). "Geographic distribution of suitable hosts explains the evolution of specialized gentes in the European cuckoo *Cuculus canorus*." BMC Evolutionary Biology **9**: 88.
- Spatafora, J. W., G. H. Sung, J. M. Sung, N. L. Hywel-Jones and J. F. White (2007). "Phylogenetic evidence for an animal pathogen origin of ergot and the grass endophytes." Molecular Ecology **16**(8): 1701-1711.



- Staats, M., P. v. Baarlen, A. Schouten, J. A. L. v. Kan and F. T. Bakker (2007). "Positive selection in phytotoxic protein-encoding genes of *Botrytis* species." Fungal Genetics and Biology **44**: 52-63.
- Staats, M., P. van Baarlen and J. A. L. van Kan (2005). "Molecular phylogeny of the plant pathogenic genus *Botrytis* and the evolution of host specificity." Molecular Biology and Evolution **22**(2): 333-346.
- Stahl, E. A. and J. G. Bishop (2000). "Plant-pathogen arms races at the molecular level." Current Opinion in Plant Biology **3**(4): 299-304.
- Stahl, E. A., G. Dwyer, R. Mauricio, M. Kreitman and J. Bergelson (1999). "Dynamics of disease resistance polymorphism at the Rpm1 locus of *Arabidopsis*." Nature **400**(6745): 667-671.
- Stranger, B. E. and T. Mitchell-Olds (2005). "Blackwell Publishing, Ltd. Nucleotide variation at the myrosinase-encoding locus, TGG1, and quantitative myrosinase enzyme activity variation in *Arabidopsis thaliana*." Molecular Ecology **14**: 295-309.
- Stukenbrock, E. H., S. Banke, M. Javan-Nikkhah and B. A. McDonald (2007). "Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation." Molecular Biology and Evolution **24**(2): 398-411.
- Stukenbrock, E. H. and B. A. McDonald (2007). "Geographical variation and positive diversifying selection in the host-specific toxin SnToxA." Molecular Plant Pathology **8**(3): 321-332.
- Stukenbrock, E. H. and B. A. McDonald (2009). "Population Genetics of Fungal and Oomycete Effectors Involved in Gene-for-Gene Interactions." Molecular Plant-Microbe Interactions **22**(4): 371-380.
- Su, C. and M. Nei (1999). "Fifty-million-year-old polymorphism at an immunoglobulin variable region gene locus in the rabbit evolutionary lineage." Proceedings of the National Academy of Sciences of the United States of America **96**(17): 9710-9715.
- Szidat, L. (1940). "Beiträge zum Aufbau eines natürlichen Systems der Trematoden. I. Die Entwicklung von *Echinocercaria choanophila* U. Szidat zu *Cathaemasia hians* und die Ableitung der Fasciolidae von den Echinostomidae." Z. Parasitenkd. **11**: 239-283.
- Tajima, F. (1989). "Statistical-Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism." Genetics **123**(3): 585-595.
- Tameling, W. I. L. and M. H. A. J. Joosten (2007). "The diverse roles of NB-LRR proteins in plants." Physiological and Molecular Plant Pathology **71**: 126-134.
- Tellier, A. and J. K. M. Brown (2007). "Polymorphism in multilocus host-parasite coevolutionary interactions." Genetics **177**(3): 1777-1790.

- Tellier, A. and J. K. M. Brown (2007). "Stability of genetic polymorphism in host-parasite interactions." Proceedings of the Royal Society B-Biological Sciences **274**: 809-817.
- Tellier, A. and J. K. M. Brown (2008). "The relationship of host-mediated induced resistance to polymorphism in gene-for-gene relationships." Phytopathology **98**: 128-136.
- Thompson, J. N. (1994). The coevolutionary process. Chicago and London.
- Thrall, P. H., J. Antonovics and D. H. Hall (1993). "Host and pathogen coexistence in sexually transmitted and vector-borne diseases characterized by frequency-dependent disease transmission." The American Naturalist **142**(3): 543-552.
- Thrall, P. H. and J. J. Burdon (2002). "Evolution of gene-for-gene systems in metapopulations: the effect of spatial scale of host and pathogen dispersal." Plant Pathology **51**(2): 169-184.
- Thrall, P. H., J. J. Burdon and J. D. Bever (2002). "Local adaptation in the *Linum marginale*-*Melampsora lini* host- pathogen interaction." Evolution **56**(7): 1340-1351.
- Tian, D., M. B. Traw, J. Q. Chen, M. Kreitman and J. Bergelson (2003). "Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*." Nature **423**(6935): 74-77.
- Tian, D. C., H. Araki, E. Stahl, J. Bergelson and M. Kreitman (2002). "Signature of balancing selection in *Arabidopsis*." Proceedings of the National Academy of Sciences of the United States of America **99**(17): 11525-11530.
- Tiffin, P., R. Hacker and B. S. Gaut (2004). "Population genetic evidence for rapid changes in intraspecific diversity and allelic cycling of a specialist defense gene in *Zea*." Genetics **168**(1): 425-434.
- Tiffin, P. and D. A. Moeller (2006). "Molecular evolution of plant immune system genes." Trends in Genetics **22**(12): 662-670.
- Van der Merwe, M. M., M. W. Kinnear, L. G. Barrett, P. N. Dodds, L. Ericson, P. H. Thrall and J. J. Burdon (2009). Positive selection in AvrP4 avirulence gene homologues across the genus *Melampsora*: -.
- Van Valen, L. (1973). "A New Evolutionary Law." Evolutionary Theory **1** 1: 1-30.
- Vera Cruz, C. M., J. F. Bai, I. Ona, H. Leung, R. J. Nelson, T. W. Mew and J. E. Leach (2000). "Predicting durability of a disease resistance gene based on an assessment of the fitness loss and epidemiological consequences of avirulence gene mutation." Proceedings of the National Academy of Sciences of the United States of America **97**(25): 13500-13505.
- Verra, F., W. Chokejindachai, G. D. Weedall, S. D. Polley, T. W. Mwangi, K. Marsh and D. J. Conway (2006). "Contrasting signatures of selection on the *Plasmodium falciparum* erythrocyte binding antigen gene family." Molecular and Biochemical Parasitology **149**(2): 182-190.
- Vogel, T. U., D. T. Evans, J. A. Urvater, D. H. O'Connor, A. L. Hughes and D. I. Watkins (1999). "Major histocompatibility complex class I genes in primates: co-evolution with pathogens." Immunological Reviews **167**: 327-337.

- Ward, T., Bielawski JP., Kistler HC., Sullivan E. O'Donnell K., (2002). "Ancestral polymorphism and adaptative evolution in the trichothecene mycotoxin gene cluster of phytopathogenic *Fusarium*." Proc. Natl. Acad. Sci. USA **99**(14): 9278-9283.
- Whitlock, M. C. (1996). "The red queen beats the Jack-of-all-trades: the limitations of phenotypic plasticity and niche breadth." Am. Nat. **148**: S65-S77.
- Whitlock, M. C. (1996). "The Red Queen Beats the Jack-Of-All-Trades: The Limitations on the Evolution of Phenotypic Plasticity and Niche Breadth." The American Naturalist **148**(s1): S65.
- Wichman, H. A., J. Wichman and J. J. Bull (2005). "Adaptive molecular evolution for 13,000 phage generations: A possible arms race." Genetics **170**(1): 19-31.
- Wichmann, G. and J. Bergelson (2004). "Effector genes of *Xanthomonas axonopodis* pv. *vesicatoria* promote transmission and enhance other fitness traits in the field." Genetics **166**(2): 693-706.
- Williamson, B., B. Tudzynsk, P. Tudzynski and J. A. L. van Kan (2007). "Botrytis cinerea: the cause of grey mould disease." Molecular Plant Pathology **8**(5): 561-580.
- Williamson, B., B. Tudzynski, P. Tudzynski and J. A. L. van Kan (2007). "Botrytis cinerea: the cause of grey mould disease." MOLECULAR PLANT PATHOLOGY **8**(5): 561-580.
- Win, J., W. Morgan, J. Bos, K. V. Krasileva, L. M. Cano, A. Chaparro-Garcia, R. Ammar, B. J. Staskawicz and S. Kamoun (2007). "Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes." Plant Cell **19**(8): 2349-2369.
- Woolhouse, M. E. J., J. P. Webster, E. Domingo, B. Charlesworth and B. R. Levin (2002). "Biological and biomedical implications of the co-evolution of pathogens and their hosts." Nature Genetics **32**(4): 569-577.