JOURNAL OF THE ROYAL SOCIETY Interface

# Patterns of antigenic diversity and the mechanisms that maintain them

Marc Lipsitch and Justin J O'Hagan

*J. R. Soc. Interface* 2007 **4**, 787-802
doi: 10.1098/rsif.2007.0229

JOURNAL OF THE ROYAL SOCIETY Interface

*J. R. Soc. Interface* (2007) **4**, 787–802
doi:10.1098/rsif.2007.0229
*Published online* 10 April 2007

# Patterns of antigenic diversity and the mechanisms that maintain them

**Marc Lipsitch[1,2,*] and Justin J. O'Hagan[1]**

[1]*Department of Epidemiology, and* [2]*Department of Immunology and Infectious Diseases, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA*

Many of the remaining challenges in infectious disease control involve pathogens that fail to elicit long-lasting immunity in their hosts. Antigenic variation is a common reason for this failure and a contributor to the complexity of vaccine design. Diversifying selection by the host immune system is commonly, and often correctly, invoked to explain antigenic variability in pathogens. However, there is a wide variety of patterns of antigenic variation across space and time, and within and between hosts, and we do not yet understand the determinants of these different patterns. This review describes five such patterns, taking as examples two bacteria (*Streptococcus pneumoniae* and *Neisseria meningitidis*), two viruses (influenza A and HIV-1), as well as the pathogens (taken as a group) for which antigenic variation is negligible. Pathogen-specific explanations for these patterns of diversity are critically evaluated, and the patterns are compared against predictions of theoretical models for antigenic diversity. Major remaining challenges are highlighted, including the identification of key protective antigens in bacteria, the design of vaccines to combat antigenic variability for viruses and the development of more systematic explanations for patterns of antigenic variation.

**Keywords: antigenic variability; mathematical model; acquired immunity; strain-specific immunity**

Edward Jenner's development of vaccination against smallpox (Mazumdar 2003) and Peter Panum's observations of immunological memory in the 1846 Faroe Islands' measles epidemic (Panum 1940) provided cornerstones for the science that would become immunology. There is no coincidence that these two early advances were accomplished with antigenically conserved viruses. Both measles and smallpox vary little in their antigenic properties either within a single infection or in the global population. Hence, infection with any strain of such a virus (e.g. measles) leads to vigorous and long-lived acquired immunity, which is effective against all known strains of the virus to which the individual may later be exposed.

Many of the early successes in vaccine development were against infections caused by such viruses (smallpox, measles, mumps, rubella and yellow fever) with little or no known variability in their antigens, for which natural infection provides lifelong immunity against all known strains of the species. For the vaccines widely available by the 1970s, with the exception of polio, antigenic variation was not an important consideration.

Many recent vaccine development projects have faced the problem of antigenic variability. Some recently developed vaccines—notably, vaccines against the encapsulated nasopharyngeal bacteria, *Haemophilus influenzae*, *Streptococcus pneumoniae*, *Neisseria meningitidis*—have approached the problem of antigenic variability by multivalent formulations that raise immunity individually targeted against one or more of the most clinically significant antigenic (capsular) types. Such approaches have been largely successful, but in the case of *S. pneumoniae*, they have led to the selection of vaccine escape variants, namely strains carrying capsular types that are not included in the vaccine (Lipsitch 1999; Kyaw *et al.* 2006). Many of the remaining challenges for vaccine design—malaria, HIV/AIDS, trypanosomiasis, gonorrhoea and others—are against pathogens for which antigenic variation is common both within infected individuals and in the pathogen population as a whole, and for which natural immunity is imperfect and/or temporary. Indeed, for many of these pathogens, antigenic variability is so pervasive that no single protective antigen has been identified (as it was for the encapsulated bacteria), widening the scope of the challenge.

It is tempting to speculate that vaccine development is most difficult for pathogens that provoke the least effective immune responses following natural infection, and that these pathogens, in turn, are disproportionately prone to significant antigenic variability. This hypothesis suggests a practical benefit in understanding the mechanisms underlying antigenic variation, in order to both design vaccines that may overcome the problem, and predict how pathogen populations will respond to the selective pressure imposed by vaccines that target

*Author for correspondence (mlipsitc@hsph.harvard.edu).

One contribution of 20 to a Theme Issue 'Cross-scale influences on epidemiological dynamics: from genes to ecosystems'.

*Received* 1 February 2007
*Accepted* 2 March 2007

particular antigenic types. Antigenic variability is also of considerable interest from the perspective of basic population biology. Understanding the forces that maintain diversity in natural populations is a central question of both ecology and population genetics, and pathogen populations form excellent model systems for investigating these principles both theoretically and empirically. As this review will argue, antigenic variability takes widely different forms in different pathogen populations. The idea that host immune pressure leads to antigenic diversity is almost a truism, but has not been refined to the point that it can explain why different infectious agents show such different patterns of antigenic variability. This article is intended to encourage further studies towards the development of a more unified theory of antigenic variation, one that can explain this variety of patterns while making testable predictions about the biological mechanisms needed to maintain these patterns.

In §1, we describe five patterns of antigenic variability observed in different pathogens. In §2, we survey each of these patterns and their possible biological underpinnings in a single pathogen species. In §3, we consider some of the general principles that have been proposed to explain antigenic diversity, including mathematical models of the effects of diversifying selection by host immunity. In §4, we briefly consider non-immune mechanisms that may lead to antigenic diversity. In §5, we discuss the strengths and weaknesses of existing explanations for antigenic variability and lay out a set of theoretical, experimental and practical (vaccine development) questions that, we believe, would merit further study towards a more general understanding of the problem.

## 1. THE PHENOMENON OF ANTIGENIC VARIABILITY

In this review, we define antigenic variability as the existence of two or more strains within a pathogen species for which the immune response provoked by one strain is more effective against the strain itself (the homologous strain) than against other (heterologous) strains. The concepts of species and strain within microbial populations are controversial because sex, in particular outcrossing sex, is possible yet optional for many important microbes. This liberated approach to sex (Levin & Bergstrom 2000) complicates the definition of species because the biological species concept, so useful for obligately sexual organisms, is not as clearly defined for those in which sex is optional and not tied to reproduction. On the other hand, the possibility of sex within nearly all microbial populations complicates the classical notion of a strain, which is based on clonal descent. We do not dispute the complexity of these arguments, but, in general, will take the notions of species and strain as if they were unproblematic.

Antigenic variability takes many forms. Here, we consider four pathogens that exemplify widely varying patterns of antigenic variability: the bacteria, *S. pneumoniae* and *N. meningitidis*; and the viruses, influenza A and HIV-1. We also consider infectious agents that rarely exhibit antigenic variability.

Pattern 1: little variation within a single host, but extensive population-wide variation that is consistent in space and time (*S. pneumoniae*).

Pattern 2: little variation within a single host, but extensive population-wide variation that can change in space and time (*N. meningitidis*).

Pattern 3: little variation within a single host, and little variation in the global population at a single time, but rapid variation over time on a scale of years (influenza A virus).

Pattern 4: extensive replacement of dominant types over time within a single host, with extensive and growing standing diversity in the global population (human immunodeficiency virus, type 1).

Pattern 5: since little or no antigenic variability is known, naturally acquired immunity and vaccine-induced immunity are universal or nearly so.

These patterns are neither fully exhaustive nor mutually exclusive; in particular, we do not consider here the pattern of within-host variation by high-frequency genetic changes (cassette expression or phase variation) used by pathogens including *Plasmodium falciparum*, trypanosomes and *Neisseria gonorrheae* (Criss *et al.* 2005; Dzikowski *et al.* 2006; Taylor & Rudenko 2006).

## 2. PATTERNS AND PATHOGEN-SPECIFIC EXPLANATIONS OF GENETIC VARIABILITY

For each of the first four patterns of antigenic diversity, we consider here some relevant details of a particular pathogen that, we believe, exemplifies the pattern. General epidemiological details of these four infectious agents are given in table 1.

### *2.1. Pattern 1: little variation within a single host, but extensive population-wide variation that is consistent in space and time (***S. pneumoniae, *figure 1*a,b*)*

Ninety serotypes of *S. pneumoniae* (pneumococcus) are recognized, and for many years, capsular serotyping was the sole method of classifying pneumococcal strains. Pneumococci are carried asymptomatically in the nasopharynx and are spread from carrier to carrier, a transmission process whereby symptomatic disease (otitis media, pneumonia, bacteremia and meningitis) is not required (Bogaert *et al.* 2004*a*). Capsular serotype—but no other known naturally variable determinant on the pneumococcus—is strongly associated with the tendency of a strain to cause disease or to be carried asymptomatically (Hanage *et al.* 2005). Type-specific anticapsular antibodies are highly effective when used for passive immunization (serum therapy; Lord & Heffron 1938) or elicited by active immunization (vaccination; Black *et al.* 2000). The primary role of anticapsular antibodies as the means by which people naturally gain immunity to *S. pneumoniae* is often assumed (Janeway *et al.* 2001) but is less clear as shown below.

In a given location, the frequency distribution of particular serotypes or serogroups (groups of closely related serotypes) in invasive disease has remained

Table 1. Epidemiologic characteristics of four 'typical' pathogens with different patterns of antigenic diversity.

| | pathogen | | | |
|---|---|---|---|---|
| | *Streptococcus pneumoniae* | *Neisseria meningitidis* | influenza | HIV-1 |
| infectious agent | gram (+)ve bacterium | gram (−)ve bacterium | RNA virus | RNA virus |
| transmission route | respiratory | respiratory | respiratory | sexual, bodily fluids |
| type of transmission | endemic, outbreaks in close quarters (Bogaert *et al.* 2004*a*; Gleich *et al.* 2000) | endemic, outbreaks in close quarters, epidemics (Baltimore 1998) | epidemic (annual) and pandemic | one pandemic is ongoing |
| seasonality | disease is seasonal, but cases occur year-round; carriage is modestly seasonal (Gray *et al.* 1982) | disease is seasonal, but cases occur year-round; carriage seasonality uncertain (Baltimore 1998) | highly seasonal (Viboud *et al.* 2004) | not importantly seasonal |
| carrier state important for transmission | yes (Bogaert *et al.* 2004*a*) | yes (Baltimore 1998) | no | transmission occurs during long asymptomatic or undiagnosed period |
| duration of carriage or infectiousness | weeks to few months (Bogaert *et al.* 2004*a*) | varies from transient to months (Baltimore 1998) | days to week | decade |
| typical basic reproductive number ($R_0$) | ? | ? | $<=2$ (Mills *et al.* 2004) | much debated; highly dependent on the population |

nearly constant over several decades (figure 1*b*; Babl *et al.* 2001; Konradsen & Kaltoft 2002 and references therein), with the exception that serotype 1 has declined in frequency since the mid-twentieth century in several places, but has then rebounded in Europe while remaining low in the US. Unfortunately, similar data are not available for nasopharyngeal carriage, which should represent the population of pneumococci under selection for transmission. The top serogroups recovered from nasopharyngeal carriage (Bogaert *et al.* 2004*a*) or invasive or mucosal disease (Hausdorff *et al.* 2000, 2005) are also geographically consistent, although not identical, across continents. Despite this consistency, the total importance of the top serogroups varies geographically, with the most limited diversity in North America; part of this difference, however, probably reflects variation in sampling (Hausdorff *et al.* 2005).

Observing that different antigenic types (serotypes or serogroups) have roughly similar frequencies across time and space suggests the possibility that some strong selective force is maintaining these frequencies. Evidence for the strength of this selective force comes from a comparative study of antibiotic resistance and serotype composition across eight regions in the United States (McCormick *et al.* 2003). In the US, as elsewhere, resistance to antibiotics has been strongly concentrated in certain serotypes; for example, 42–84% of serotype 19A isolates were penicillin resistant, compared with only 0–6.8% of serotype 4 isolates (ranges are for the eight regions). Overall, penicillin resistance varied significantly between US regions, from 14% of pneumococci in California to 36% in Georgia and Tennessee; analyses by McCormick *et al.* (2003) attributed these differences to regional variation in selection for antibiotic resistance. One might have expected that this would lead to enrichment of the resistance-prone serotypes in the high-resistance regions compared with the low-resistance regions, i.e. the serotype composition would be changed by associated linkage selection for antibiotic resistance. This was not the case; high-resistance regions showed greater proportions resistant within each serotype, but no trend towards having more of the high-resistance serotype (McCormick *et al.* 2003). An interpretation of this finding is that whatever selection was maintaining, serotype composition in the populations of these states was stronger than the selection imposed by linkage of serotype to antibiotic resistance. Conversely, when strong selection to change serotype composition has been applied, by widespread use of the pneumococcal conjugate vaccine in the US, serotype composition of the pneumococcal population has changed and with it has brought down the rates of resistance (since most of the resistant strains are of serotypes included in the vaccine; Kyaw *et al.* 2006). Thus, at least among invasive disease isolates in the US, selection on serotype can change resistance patterns more readily than selection for resistance changes serotype composition.

As with most forms of antigenic variability, the existence of diverse serotypes of pneumococcus has been attributed to diversifying selection by host immune responses. In a population of $n$ pneumococcal serotypes, with many hosts immune to many of the types, the $n+1$th serotype will have an advantage, as long as it is seen by host immune systems as being distinct from existing serotypes, because there will be more susceptible hosts in the population for this serotype. As it increases in frequency, its advantage will wane (opening the door for more diversification), while if it declines in frequency, the host population will be less exposed over time, resulting in less immunity

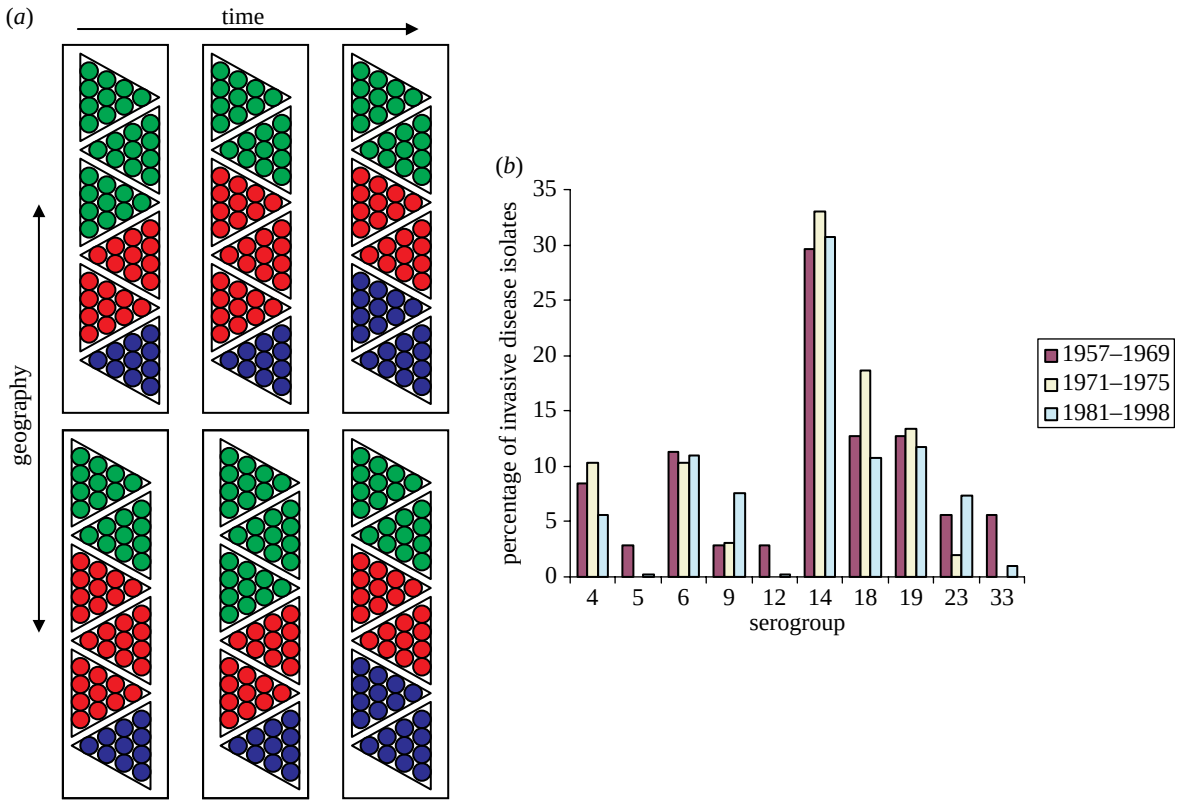


Figure 1. Antigenic diversity, pattern 1: constant high level of standing between-host diversity consistent over space and time; little within-host diversity, exemplified by *Streptococcus pneumoniae*. (*a*) Schematic of this pattern. Triangles, individual hosts within populations (rectangles); circles, pathogens within a host; colours, antigenic variants. (*b*) Serogroup representation among invasive isolates of *S. pneumoniae* at Boston City Hospital over four decades; data from Babl *et al.* (2001).

and thereby restoring its advantage. This is the classic scenario of 'negative, frequency-dependent' diversifying selection on antigens by host immunity.

This scenario, while plausible, has important limitations in the pneumococcal case. First, in its simple form, it predicts that diversity should be high everywhere and at all times, but probably growing over time and not necessarily geographically coherent. Second, we and others who have looked into this failed to find evidence that naturally acquired immunity to *S. pneumoniae* is mainly serotype specific. Invasive disease incidence from various serotypes peaks around 1 year of age and falls in parallel, suggesting a single type of immunity acquired as children age, rather than the acquisition of multiple antigenic specificities, which should appear at different ages. Moreover, these declines begin before anticapsular antibody rises measurably in children, suggesting that anticapsular antibodies are not the mechanism of reduction in disease (Lipsitch *et al.* 2005). Similar patterns have been observed in nasopharyngeal carriage (Hogberg *et al.* 2007), and associations between anticapsular antibody and protection from pneumococcal carriage have been weak in epidemiologic (Goldblatt *et al.* 2005) and experimental human studies (McCool *et al.* 2002). In mice, acquired immunity following exposure to one serotype is equally effective against that type and other types, suggesting that it is mainly directed against one or more non-capsular antigens (which are probably more conserved across pneumococci; Malley *et al.* 2005; Trzcinski *et al.* 2005), and acquired immunity to carriage in mice is dependent on CD4+T cells, but not on antibody (Trzcinski *et al.* 2005; van Rossum *et al.* 2005), thereby excluding anticapsular antibody as a mechanism.

These findings cast doubt on the classical immune diversifying selection hypothesis as an explanation for pneumococcal capsular diversity. Other possible mechanisms of diversifying selection on pneumococcal capsules are under investigation, including the possibility that different capsular types are preferentially successful in different types of hosts, perhaps determined by host genetics, age (Bogaert *et al.* 2004*a*) or other factors.

### 2.2. *Pattern 2: little variation within a single host, but extensive population-wide variation that can change in space and time* (N. meningitidis, *figure 2*a,b)

Ecologically, *N. meningitidis* (meningococcus) has much in common with *S. pneumoniae*. Meningococcus is present in the nasopharynx of approximately 10% of healthy individuals at any point in time (Cartwright *et al.* 1987; Caugant *et al.* 1988; Stephens 1999). Capsular polysaccharides are frequently found on the meningococcal surface and have been divided into 13 serogroups, six of which account for 90% of meningococcal disease: A, B, C, W-135, X and Y (Pollard 2004). The capsule is considered a major virulence factor and (with the exception of the group B capsule, which cross-reacts with human sialic acid) has traditionally been viewed as a principal target of the immune system (Goldschneider *et al.* 1969; Caugant *et al.* 2007).

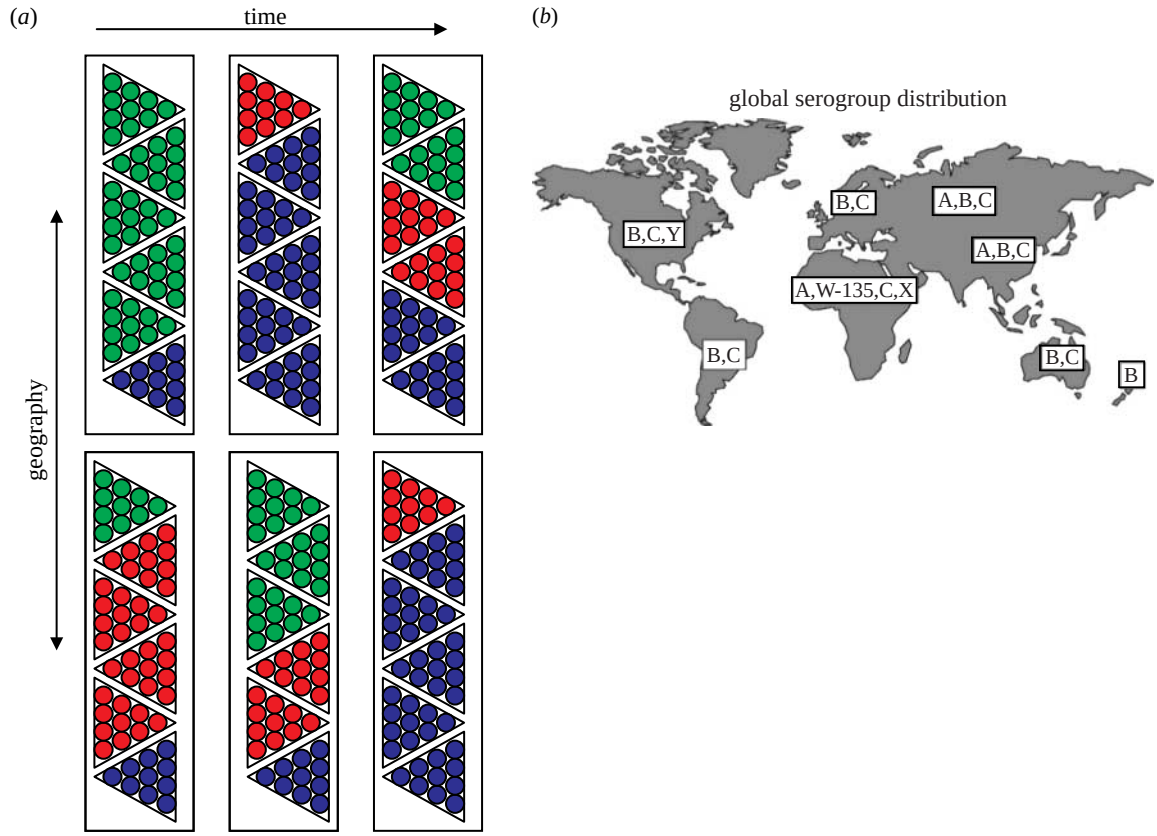


Figure 2. Antigenic diversity, pattern 2: substantial between-host diversity, changing over time and by geographic region; little within-host diversity, exemplified by *Neisseria meningitidis*. (*a*) Schematic as in figure 1*a*. (*b*) Global serogroup distribution of *N. meningitidis* strains mainly responsible for invasive disease; adapted from Stephens (2007).

Capsular serogroups also vary in their epidemiologic behaviour: traditionally, group A has been associated with annual epidemics in Africa, but recently the importance of W-135 in these epidemics has been recognized (Mueller *et al.* 2006).

Despite these similarities, the contribution of particular meningococcal strains to disease shows great variation in both space and time, in contrast to pneumococcus (figure 2*a*,*b*). Geographical variation is observed whether meningococci are classified by serogroup (figure 2*b*; Stephens 2007) or by clonal complex—a designation defined by the similarity of nucleotide sequence type (ST) of housekeeping genes, which are thought to be selectively neutral (Rosenstein *et al.* 1999; Yazdankhah *et al.* 2004). Temporal variation is also common. The incidence of disease caused by serogroup Y bacteria in the US has increased significantly from 2% of endemic disease in 1989–1991 to 39% for the period 1996–2001, whereas this serogroup is an infrequent cause of disease in Europe (Rosenstein *et al.* 1999; Raghunathan *et al.* 2004; Yazdankhah & Caugant 2004). Serogroup A has declined in incidence to very low levels in the US and Europe since World War II, but has been responsible for significant outbreaks in African countries and remains a major cause of meningitis there (Stephens 2007). It should be noted that while patterns of strains causing meningococcal disease show great geographical and temporal variation, the patterns among carriage isolates may be somewhat more homogeneous, at least within Europe (Caugant *et al.* 2007). The strong contrast drawn herein between patterns in the meningococcus and those in the pneumococcus should be seen as tentative until comparable samples of carriage isolates of the meningococcus can be assessed.

Several factors complicate the study of antigenic diversity in meningococcus. First, while the capsule (which defines meningococcal serogroup) is a vaccine antigen, its centrality as an antigen in the natural immune response is uncertain (as with the pneumococcus). Both encapsulated and acapsulate forms of the bacterium are common and each state is suggested to enable escape from immune recognition under different circumstances. While virtually all strains that cause disease are encapsulated (Dolan-Livengood *et al.* 2003; Yazdankhah & Caugant 2004; Yazdankhah *et al.* 2004), up to 30% of meningococcal isolates from carriers are non-serogroupable. It is suggested that the capsule may prevent immune recognition during systemic disease, increase survival in the environment and enhance transmission (Lipsitch & Moxon 1997; Dolan-Livengood *et al.* 2003). However, benefits of the acapsulate state have also been proposed, including evasion from a host immune response during carriage and facilitating adherence to the human nasopharyngeal epithelium (Hammerschmidt *et al.* 1996; Dolan-Livengood *et al.* 2003). Thus, any complete explanation of serogroup structure will require an explanation of the variable expression of the capsule observed in the species. Further knowledge of the regulation and role of

the capsule could shed light on the transmission and pathogenic processes of *N. meningitidis*.

The value of the capsule in defining biologically distinct meningococcal types is also open to question. Clonal types defined by housekeeping genotype (or previously by protein electrophoretic type) seem to correlate better with many biological properties than capsular serogroup does, and meningococcal strains as defined by molecular methods have been observed to change serogroup (Swartley *et al.* 1997). Assessed by either serogroup or ST, temporal and geographical variations are common (Rosenstein *et al.* 1999; Yazdankhah *et al.* 2004). Other antigens may also be protective (Pizza *et al.* 2000; Urwin *et al.* 2004), raising questions as in the pneumococcal case about the uniqueness of the capsule as an important antigen for natural immunity.

Diversity is greater among carried meningococcal strains than among those causing disease, which represent a small number of hypervirulent lineages (Caugant *et al.* 1988; Yazdankhah *et al.* 2004). As colonization with the majority of strains of *N. meningitidis* produces no symptoms, the disease state is unimportant for much of its spread (Jolley *et al.* 2000). The low frequency with which virulent strains are isolated from carriers could be due to inherent low transmissibility of these strains. Alternatively, these strains could be highly transmissible but result in colonization of short duration (Caugant *et al.* 2007). As disease is thought to occur between 1 and 14 days after acquisition (Stephens 2007), it is plausible that either clearance or disease occurs soon after acquisition of a virulent strain and that this would explain the low point prevalence of such strains in all carriage studies. If the rate of transmission was significantly higher for some of the 'hyperinvasive' strains when compared to avirulent strains, it may be that the rate of disease per acquisition is very similar for the two types (Jolley *et al.* 2000). To discern which of these possibilities is correct, it will be essential to conduct large cohort studies over many months to learn what the lengths of carriage for different strains are. It may be the case that each strain has a characteristic carriage length or that the duration of carriage falls into two categories: one for avirulent strains and another for virulent ones.

A theory of strain structure proposed by Gupta and colleagues suggests that immune selection can organize polymorphic antigens into non-overlapping strains within a pathogen population (Gupta *et al.* 1996; Gupta & Maiden 2001). The allelic diversity of two antigenic outer membrane proteins (OMP) of *N. meningitidis* is suggested to be structured in such a manner. As an adaptive immune response is specific to a particular antigenic variant, such responses will limit the transmission of pathogens that share the same antigen or that cross-react with it. If the immune response is sufficiently strong, then strains that do not share alleles at antigenic loci will be selected as they do not interfere with each others' transmission. Thus, pathogen strains with unique combinations of antigenic alleles will be produced. Gupta suggests that the *porA* gene of *N. meningitidis*, which encodes a highly diverse OMP, shows signs of being structured in such a way (Gupta *et al.* 1996); further studies on *porA* and *fetA*, another gene encoding an OMP, are described in §5.

### 2.3. *Pattern 3: little variation within a single host, and little variation in the global population at a single time, but rapid variation over time on a scale of years (influenza A virus, figure 3*a,b*)*

Antigenic variation in human influenza A (where it is called antigenic drift) is well known as the reason for vaccines being reformulated each season (Zambon 1999) and as part of the reason for individuals suffering multiple bouts of influenza A infection during their lifetime (Smith *et al.* 2004). The need to assess the antigenic composition of the virus population for purposes of vaccine formulation every year has led to extensive, albeit not always representative, collections of influenza A isolates over time.

The major immune response that is thought to be important in preventing or clearing influenza A infection is an antibody response directed against the haemagluttinin (HA) protein (Zambon 1999). Phylogenetic studies of the HA nucleotide sequence within a single HA subtype (H3) have shown a pattern of slow genetic change over time, with considerable genetic change over a time span of decades, but rather limited diversity present in the global population at a given time (figure 3*a*,*b*; Fitch *et al.* 1991).

Recent studies have complicated this picture somewhat. Smith and colleagues have shown that the variation in HA nucleotide sequence does not map continuously into variation in antigenic similarity; rather, every few years, a new 'antigenic cluster' arises that differs significantly from previous strains, in that antibodies cross-react poorly between antigenic clusters. Within an antigenic cluster, however, significant amounts of genetic variation can occur without much apparent effect on the antigenic properties of the HA molecule. Thus, the rather smooth pattern of genetic variation leads to a more punctuated pattern of antigenic variation (Smith *et al.* 2004).

In contrast to the bacteria considered above, the key role of the immune system in selecting for antigenic variation in influenza is difficult to dispute. Strains isolated years apart provoke immune responses that are poorly cross-protective (Smith *et al.* 2004). Genetic changes in the HA are concentrated in the regions of the molecule that are most often the targets of antibodies (Plotkin & Dushoff 2003). Despite this clear evidence of selection for antigenic change, recent analyses suggest that significant antibody escape mutations are relatively rare, rarely reaching measurable levels within an infected host or becoming noticeable even within a season (Holmes *et al.* 2005; Wolf *et al.* 2006). This puzzling finding is related to a more general puzzle: why does antigenic diversification in influenza proceed in such an orderly fashion, with limited diversity at one time, rather than a great explosion of diversity in the population? Two proposed answers to this issue are discussed in §3.

### *2.4. Pattern 4: extensive replacement of dominant types over time within a single host, with extensive and growing standing diversity in the global population (human immunodeficiency virus, type 1, figure 4a,b)*

The size of the HIV epidemic, its continuing spread and the difficulties associated with developing a vaccine against such a diverse pathogen have made understanding the extent and causes of its variation a priority in vaccine research. HIV-1 can be subdivided into three groups, M, N and O, each the result of a different cross-species transmission (Gao *et al.* 1999), with the M group being further subdivided into nine subtypes and multiple recombinants (Holmes 2004). While M group viruses are found throughout the globe, N and O group lineages are largely confined to West Central Africa consistent with the evidence that this is where HIV-1 first emerged (Gao *et al.* 1999). The global distribution of M group subtypes is also marked by significant spatial structure with, for example, subtype B dominating in North America and Europe, while subtype C is most frequent in sub-Saharan Africa (Hemelaar *et al.* 2006). It has been suggested that this geographical structuring is due to repeated founder events and incomplete sampling (Rambaut *et al.* 2004).

The long infectious period of HIV-1 means that evolution occurs both within and among hosts. Thus, the large discrepancy in the degree of positive selection observed between the intra-host and the population-level phylogenetic trees of HIV-1 is surprising. The within-host tree is marked by a clear temporal structure showing continual adaptation to the host immune system and high rates of lineage extinction resulting in little genetic diversity at any time point (Grenfell *et al.* 2004). This is similar to the phylogenetic structure of influenza A on a population level (Grenfell *et al.* 2004). However, spatial dynamics appears to dominate the between-host HIV structure which shows multiple coexisting strains and weaker positive selection (Grenfell *et al.* 2004).

The remarkable genetic variation of HIV-1 within a host is a significant issue in terms of drug resistance, vaccine design and host immune response. This is due to a combination of factors: high mutation rate; a short generation time of approximately 2.6 days; production of approximately $10^{10}$ new virus particles per day; and frequent recombination (Sharp *et al.* 1995; Perelson *et al.* 1996). Strong immune selection pressure from neutralizing antibodies and cytotoxic T lymphocytes (CTL) drives the fixation of escape mutants that allow it to evade the host immune system (Ogg *et al.* 1998; Richman *et al.* 2003). Why then is such intense immune-driven selection not simply reflected on a population level? Rambaut *et al.* (2004) have proposed four explanations. Firstly, transmission is associated with a significant population bottleneck; over 99% of diversity in *gag* and *env* genes is lost upon sexual or vertical transmission (Edwards *et al.* 2006). Such a reduction in diversity affects the ability of HIV-1 to adapt on a population level, as advantageous mutations which arise in a new host have a reduced chance of transmission. Secondly, host behaviour can impact the action of selection at a population level. As HIV is largely a sexually transmitted disease, variation in the rate of partner exchange will also cause genetic drift (Grassly *et al.* 1999). By chance, viruses with advantageous mutations may be present in a person with a low rate of partner exchange and may not be transmitted for this reason. Thirdly, some advantageous mutations, for example CTL escape mutants, may not arise until late in infection after which point little transmission occurs (Goulder *et al.* 1997). Finally, CTL escape mutants may be transmitted to individuals with different HLA alleles from the donor. In such instances, escape mutants that were advantageous in the previous host may incur a high fitness cost in the current host without providing any benefit in terms of immune evasion. Studies of both humans and macaques found that CTL escape mutants that incur viral fitness costs reverted quickly upon transmission to individuals lacking the donor's restricting HLA alleles (Friedrich *et al.* 2004; Leslie *et al.* 2004), and that the rate of reversion reflects the magnitude of the viral fitness cost (Fernandez *et al.* 2005; Kent *et al.* 2005; Li *et al.* 2007). Therefore, HIV-1 evolution on a population level is driven by neutral demographic and spatial dynamics and both immune-mediated positive selection (Asquith *et al.* 2006) and purifying selection for replication ability. This balance has been elegantly demonstrated by a detailed longitudinal study of seven patients, in whom almost half of all substitutions observed were associated with escape from CTL responses or reversion of presumed escape mutations from prior hosts (Li *et al.* 2007). Moreover, the sites at which changes occurred to escape CTL responses were those that are most variable (and thus presumably least functionally constrained) in the global HIV-1 population (Allen *et al.* 2005).

An effective vaccine against HIV-1 will probably depend on careful selection of which epitopes should be targeted, taking into account selection pressure by a range of HLA alleles, associated fitness costs of escape mutants and the fact that the genetic diversity of HIV will change during the long process of vaccine development. Strains belonging to different subtypes can differ by up to 35% in their envelope protein and within-subtype diversity can reach 20% (Gaschen *et al.* 2002). In comparison, roughly 10% sequence divergence in the antibody-binding HA1 region of influenza HA typically exists between antigenic clusters, and this is more than enough divergence to compromise vaccine efficacy (Smith *et al.* 2004; Koelle *et al.* 2006). The most obvious method of creating a broadly effective vaccine is to include a variety of epitopes from different strains in a single formulation. Alternatively, artificial sequences, namely consensus or ancestral sequences, could provide the highest levels of cross-protection possible by any single sequence (Gaschen *et al.* 2002; Nickle *et al.* 2003). Appropriate construction of such sequences may require unbiased sequence datasets of each prominent subtype, as well as a quantitative understanding of the above individual- and population-level processes.

### *2.5. Pattern 5: lack of diversity (figure 5)*

As noted at the outset, many important pathogens, including smallpox, measles, mumps and rubella

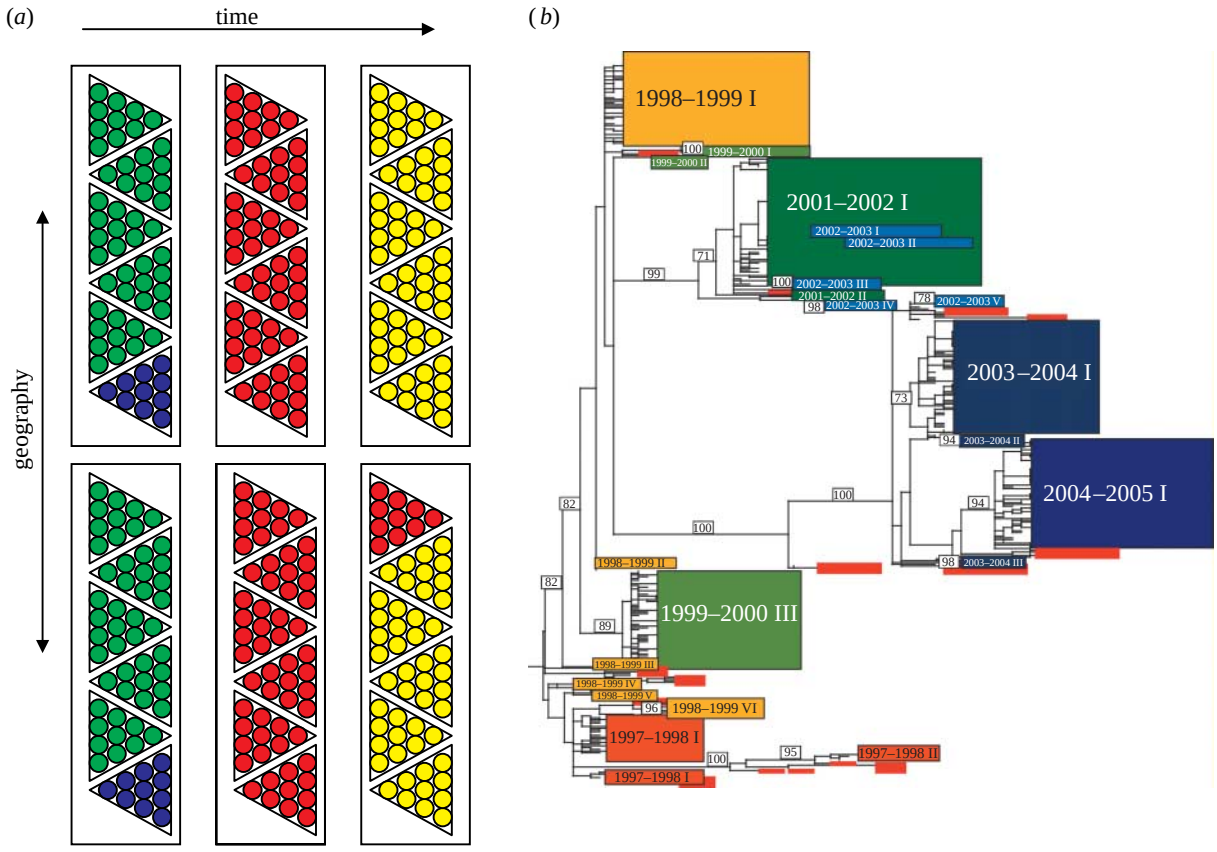


Figure 3. Antigenic diversity, pattern 3: little diversity within-host or within the host population at a given time point, but rapid change over time, exemplified by influenza A virus. (*a*) Schematic as in figure 1*a*. (*b*) Phylogeny of haemagluttinin variants present in New York state and globally 1997–2005; reprinted from figure 2 of Nelson *et al.* (2006). Colours here represent years and are not connected with (*a*).

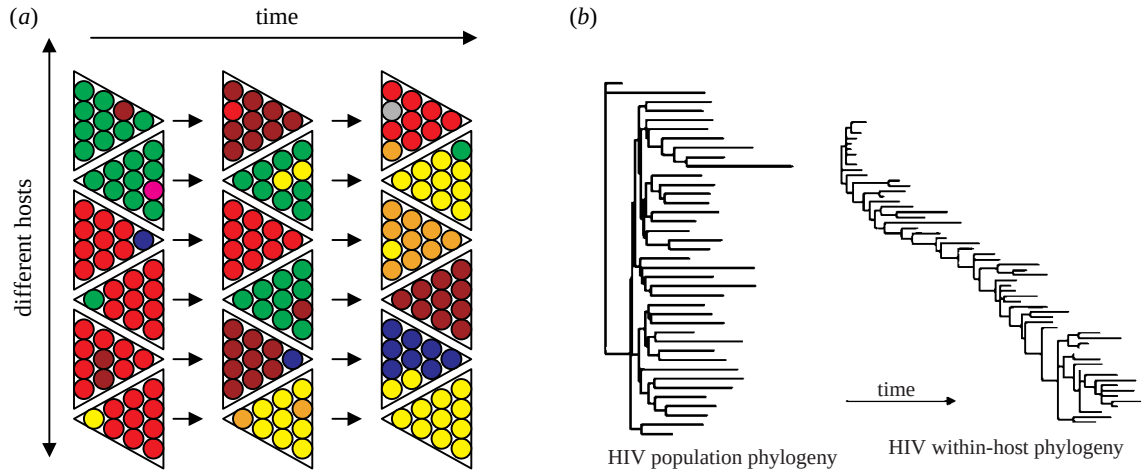


Figure 4. Antigenic diversity, pattern 4: substantial within-host diversity over time; little diversity at any single time point, and ongoing positive selection; global population does not reflect this selection but is highly diverse, exemplified by HIV-1. (*a*) Schematic as in figure 1*a*. (*b*) Within- and between-host phylogenies; reprinted from Fig. 1 of Grenfell *et al.* (2004)

viruses and several bacteria, seem to be antigenically stable throughout space and time. Put another way, for these pathogens, immune escape variants with sufficient fitness advantages to spread do not arise even on a time-scale of decades, or if they do they are rare and die out by chance. One hypothesis is that immune responses to natural infection with these agents (or to whole-pathogen vaccines) have multiple specificities, so that a mutant capable of escaping one immune response is still held in check by other immune responses. Birrer *et al.* (1981) readily selected for escape mutants of measles virus using monoclonal antibodies and suggested that the polyclonal immune response in natural infection prevents such escape. Escape

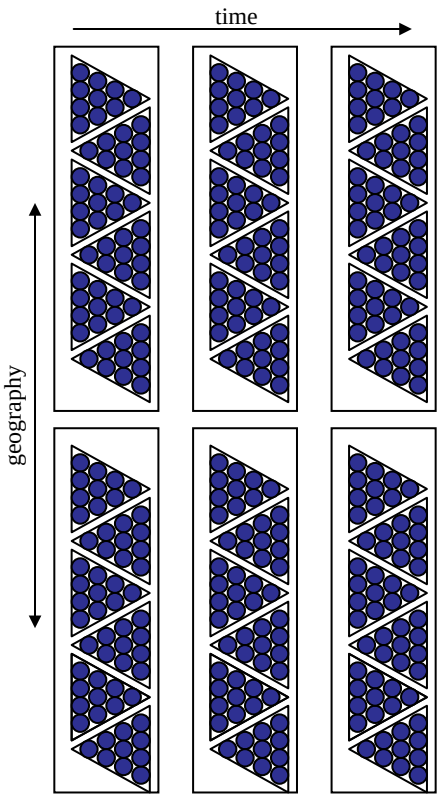


Figure 5. Antigenic diversity, pattern 4: antigenic homogeneity, within and between hosts, over time. Schematic representation as in figure1*a*.

mutations from natural immunity in hepatitis B infection are rare and appear to correlate with a narrowly focused immune response, while escape from antibody responses to the surface antigen used as a vaccine is more common (Rehermann & Nascimbeni 2005). A more ambiguous example is that of pertussis, in which two of the major antigens, pertactin and pertussis toxin, have been evolving away from the variants included in the whole-cell pertussis vaccine since that vaccine was introduced in several countries (Mooi *et al.* 2001). In the Netherlands (Mooi *et al.* 2001), these antigenic changes have been associated with a resurgence of pertussis, though the same has not been documented in other countries where the antigenic changes occur (Crowcroft & Pebody 2006). Interestingly, these escape variants have occurred while whole-cell pertussis vaccines have been in use, undermining the simple hypothesis that immunization against a range of antigens necessarily prevents escape.

## 3. IMMUNE MECHANISMS GENERATING AND MAINTAINING DIVERSITY

In the foregoing section, five patterns of antigenic diversity were described, and each was accompanied by a more or less plausible explanation relating the pattern observed to diversifying selection by host immune systems. In the ideal case, we would have more than just a good story for each pathogen and its pattern: we would have a list of characteristics which, once measured for a particular pathogen, would accurately predict the pattern of diversity observed. We have not reached this ideal—a sort of grand unified theory of antigenic diversity. However, there is a growing body of literature describing mathematical models of antigenic diversity, which taken together provides some testable hypotheses whose explanatory power for particular pathogens can be assessed. In this section, we briefly summarize some of the findings of this theoretical literature and suggest some of the relevant comparisons.

A general challenge in modelling the spread of multiple antigenic variants of a pathogen is that as the number of strains considered increases, there is an exponential increase in the number of equations required to track the population. Various elegant approaches have been employed to reduce the complexity of the problem, including assumptions of symmetry (such that all strains are identical and affect all other strains equally; Abu-Raddad & Ferguson 2004), status-based models (which track only the present ability of hosts to respond to strains, rather than their full past history of infection; Gog & Grenfell 2002; Gog & Swinton 2002), polarized immunity (in which infection confers total immunity on a proportion of hosts, rather than partial immunity on all hosts; Gog & Grenfell 2002; Abu-Raddad & Ferguson 2004) and the assumption that immunity reduces one's infectiousness to others, rather than one's risk of becoming infected (Gog & Grenfell 2002; Abu-Raddad & Ferguson 2004).

Having made varying combinations of simplifying assumptions, general models of pathogen diversity make the following predictions (summarized in table 2).

(i) If strains can be considered to exist in a 'strain space', in which distance corresponds to antigenic dissimilarity, the process of appearance and spread will tend to lead to clustering in strain space, because novel types similar to the most prevalent current types will constantly be generated, but types nearby an existing prevalent type will be at a disadvantage due to strong cross-immunity with the existing type. Therefore, highly prevalent types can exist only at some distance from one another, in strain space and/or time (Gog & Grenfell 2002; Gomes *et al.* 2002). This effect can, under certain conditions, maintain strain structure—collections of strains with non-overlapping sets of alleles at antigenic sites—despite recombination, because recombinants falling 'between' non-overlapping sets will be at a disadvantage (Gupta *et al.* 1996, 1998).

(ii) The group of strains present in a population may be stable or changing over time. Rapid strain turnover is promoted by higher transmission (Gomes *et al.* 2002), weaker cross-immunity (Gog & Grenfell 2002; Gomes *et al.* 2002) and short-lived infections (Gupta *et al.* 1998; Gog & Grenfell 2002). In the case of multilocus genotypes that determine strain structure, Gupta (Gupta *et al.* 1998) finds a more complex relationship with cross-immunity, in which turnover (either cyclic or chaotic) is greatest at intermediate levels of cross-immunity, with

Table 2. Predicted relationships between transmission characteristics of multi-strain pathogens (rows) and epidemiological features of strain diversity (columns), based on several theoretical studies.

| | | epidemiologic feature | | |
|---|---|---|---|---|
| | | total pathogen prevalence | rate of strain turnover | number of concurrent types |
| biological/transmission characteristics | $R_0$ of each strain | +(Abu-Raddad & Ferguson 2004) | + (Gomes *et al.* 2002) | +(Abu-Raddad & Ferguson 2004) |
| | extent of cross-immunity | − (Abu-Raddad & Ferguson 2004) | − (Gog & Grenfell 2002; Gomes *et al.* 2002) Complex (Gupta *et al.* 1998)[a] | − (Abu-Raddad & Ferguson 2004) |
| | duration of infectiousness relative to lifespan | +(Abu-Raddad & Ferguson 2004) | − (Gog & Grenfell 2002; Gupta *et al.* 1998) | + (Abu-Raddad & Ferguson 2004; Gog & Grenfell 2002) |

[a] Low, coexist without strain structure; medium, cycles or chaos; high, stable coexistence with strain structure.

lower levels giving unstructured stable coexistence of many types and high levels of cross-immunity promoting coexistence of non-overlapping strain types.

(iii) Total pathogen prevalence is increased for pathogens having higher transmissibility and those with longer duration or less cross-immunity (Gupta & Galvani 1999; Abu-Raddad & Ferguson 2004).

(iv) The standing diversity at any given time is greater for more transmissible pathogens and weaker cross-immunity (Abu-Raddad & Ferguson 2004) and longer infectious durations (Gog & Grenfell 2002; Abu-Raddad & Ferguson 2004).

Reassuringly, these predictions are largely qualitatively consistent across models despite differences in their simplifying assumptions, and in some cases are also consistent with previous simpler models (Dietz 1979; Gupta *et al.* 1994; Lipsitch 1997).

How well do these theoretical principles explain the observed patterns described in the previous section? Perhaps, the best match between theory and observation is for the case of influenza (the problem that motivated many of the theoretical studies): its short duration of infectiousness predicts limited diversity at one time, rapid turnover and low prevalence at any given time, which are qualitatively accurate. A problem with naive models is that diversity is generated too fast. Invoking either transient strain-transcending cross-immunity that does not depend on the distance between strains (Ferguson *et al.* 2003) or the existence of neutral variation that limits the rate of generation of antigenically novel variants (Koelle *et al.* 2006) can reduce this rate, thereby slowing the rate of strain turnover to realistic levels. In the general models, increasing the degree of antigenic distance-dependent cross-immunity (i.e. making the immunity extend to more distant strains) has a similar effect in reducing the rate at which escape variants are generated, but the details and other effects of these assumptions are different from those invoked in the influenza models (Ferguson *et al.* 2003; Koelle *et al.* 2006).

Some of the other patterns can perhaps be understood in the same framework, but the concordance is spotty. For brevity, we consider only *S. pneumoniae;* other comparisons are left as an exercise for the reader. *S. pneumoniae* differs from influenza mainly in the temporal stability of its antigenic patterns. If our hypothesis of strongly species-specific (not strain-specific) immunity is correct, then the models would predict slower turnover; the duration of pneumococcal carriage, while still short, is several-fold longer than infectiousness with influenza (Smith *et al.* 1993; Sleeman *et al.* 2006; Hogberg *et al.* 2007), also predicting slower strain turnover. Pneumococcal point prevalence is considerably higher than that of influenza, perhaps reflecting longer carriage duration and weaker overall immune responses, despite the fact that these immune responses are equally strong across different strains. These findings are concordant with the theoretical predictions just discussed (table 2). On the other hand, strong cross-immunity should lead to limited diversity at one time, yet in fact many serotypes coexist in the population.

This example suggests two considerations: first, that the qualitative predictions summarized in table 2 may be inadequate for the comparison of different pathogens, since two pathogens may differ in multiple respects, with countervailing predicted effects on some epidemiologic pattern (e.g. rate of turnover), making it impossible to predict which of the two pathogens will have a faster rate of turnover. Second, when considering individual pathogens, other factors not considered in general models may make an important contribution to the pattern of antigenic diversity. For example, immunity to pneumococcus appears to build up very gradually with exposure, and to take the form of shorter duration of carriage and reduced susceptibility to disease. While the immunity seems to be directed against conserved determinants and thus is effective against most or all serotypes, it may be more effective against some serotypes than against others (Hausdorff *et al.* 2005; Hogberg *et al.* 2007). More biological understanding may be required before general models can be adequately calibrated to particular pathogens.

## 4. NON-IMMUNE MECHANISMS THAT MAY GENERATE ANTIGENIC DIVERSITY

While much antigenic variability is undoubtedly generated by interactions between a pathogen and the immune system, it should be remembered that probably no antigen produced by a pathogen has provocation of a specific immune response (or susceptibility to that response) as its evolutionary *raison d'etre*. Antigens are components of infectious agents with diverse functions of nutrient acquisition, interaction with host cells, etc. As a result, any process that selects for genetic or phenotypic variability—in ecological language, any temporal or spatial variation in the pathogen's habitat—may directly or indirectly select for antigenic variability. Examples of such mechanisms include variation between hosts due to genetics (Moxon *et al.* 2006), behaviour, competing flora (Bogaert *et al.* 2004*b*; Regev-Yochay *et al.* 2004) or other factors that may selectively favour different variants for reasons other than immune escape. Cell-to-cell or microanatomical variation within a host (Kang & Blaser 2006) may differentially favour particular genotypes (Briles *et al.* 2005); likewise, there may be selection for variation within a host to maintain a colonizing or infecting population within the host while also producing progeny that can colonize other hosts (Boles *et al.* 2004; Briles *et al.* 2005). Such hypotheses, which, to date, have been mainly invoked to explain the evolution of highly mutable 'contingency' loci in bacteria (Moxon *et al.* 2006), are serious contenders to explain some of the antigenic variation in pathogens for which immune selection has not been demonstrated as the cause.

## 5. OPEN QUESTIONS AND FUTURE DIRECTIONS

The open questions raised by the viral examples above and those raised by the bacterial examples are rather different. We consider these in turn.

Important antigens have been identified for many important viruses, including those considered here. In such cases, the major theoretical challenge is to understand what constrains variation in the viruses—why measles and smallpox fail to escape immunity, while HIV does so within a host and influenza does so in host populations over time. It is probably not coincidental that the greatest within-host variation occurs in HIV-1 compared to these other viruses, since HIV-1 has by far the longest infection period. However, even this relationship is complex (Nowak & May 2000): short duration may be a cause of limited variability (if there is no time for variants to arise); it may be a consequence of limited variability (without immune escape, clearance occurs); and, as noted above, short duration may reduce diversity at the epidemiologic level as well (Gog & Grenfell 2002; Abu-Raddad & Ferguson 2004).

From a practical perspective, a major challenge of viral antigenic diversity is the design of vaccines. In the extreme case of HIV, in which immunodominant responses are strongly dependent on the host's HLA type (Bihl *et al.* 2006), it remains to be seen whether a vaccine can be developed that elicits strongly protective responses in a diverse population of hosts (Walker & Korber 2001). Similar considerations apply in the case of hepatitis C virus and other viruses for which vaccines remain elusive. In the case of influenza, an open challenge is to develop a vaccine that can provide 'species-wide' protection, not only against the antigenic drift variants that occur over time but also across influenza A subtypes, so that a vaccine could be usable when a new subtype appears (potentially causing a pandemic) without complete redesign (Doherty *et al.* 2006).

For both the pneumococcus and the meningococcus—and for other bacterial and eukaryotic pathogens—a central question is: what is (are) the key antigen(s) that trigger protective immunity? It will be instructive to revisit the broad distinctions made herein once the critical antigens for naturally acquired immunity are identified. From a scientific perspective, there are a number of promising approaches to identifying such antigens. These can be broadly divided into population-genetic approaches and experimental approaches. Theoretical (population-genetic) analyses predict that antigens should have several properties. First, they should be under positive selection, detectable by the pattern and rate of nucleotide substitution (Volkman *et al.* 2007) or perhaps even by the pattern of codon usage (Plotkin & Dushoff 2003; Plotkin *et al.* 2004). Second, they should be in linkage disequilibrium with one another (Gupta *et al.* 1996) despite sexual recombination in the pathogen genome. Identifying antigens by these characteristics and then assessing their pattern risks circularity (patterns of diversity assessed in antigens identified by their pattern of diversity), but certainly there is much to be learned by assessing patterns in known antigens, then using those patterns to 'fish' for candidates for additional unknown antigens. We have focused on spatio-temporal patterns in this review, but phylogenetic approaches (recently reviewed by Grenfell *et al.* 2004) use an additional dimension in the data and have been profitably applied to influenza (Wolf *et al.* 2006) among other pathogens.

In general, further comparisons between population genetic patterns of putative antigens and those of non-antigens will aid in understanding the selective forces maintaining variation. It is notable that the pattern of stable polymorphism in pneumococcal capsules described previously is observed in a species where strain types defined by 'housekeeping' genes fit well with a neutral model of evolution (Fraser *et al.* 2005). Similarly, comparison of HA evolution in influenza A against that of the whole genome has highlighted the importance of recombination and other processes that differentiate antigen evolution from 'background' evolution (Holmes *et al.* 2005). Use of these approaches will often depend on unbiased sampling of strains (Fraser *et al.* 2005; Holmes *et al.* 2005) and appropriate typing systems that gather data on targets of interest (Jolley *et al.* 2007).

For both bacteria and viruses, much work remains to understand how hosts confronted with a pathogen displaying multiple antigens 'choose' which antigens to respond to, which arms of the immune system will be

involved and which of these responses will be protective—the general phenomenon of immunodominance. From the pathogen side, it would be of interest to understand, in the presence of a host response to multiple antigens, how much, if any, advantage is garnered by a variant that escapes some but not all of these responses. If we think of a genotype-fitness mapping in antigenic space, does a broad immune response always lead to a locally flat fitness landscape (because escaping from one response has no effect on fitness in the presence of other responses)? Does a very narrow immune response create a simple adaptive landscape in which any escape from the response is favoured (unless the escape mutation destroys function)? Does a response of intermediate breadth create a landscape of the form observed for influenza A, in which many changes are neutral (perhaps because they do not sufficiently reduce immune pressure; Smith *et al.* 2004; Koelle *et al.* 2006), but the right combination of changes can produce significant escape?

Understanding the answers to such questions would be valuable for the design of vaccines, especially (as is now almost universal) the design of subunit vaccines that contain a limited number of antigens. Until such answers are available, a novel approach for the design of subunit vaccines is that of 'reverse vaccinology' (Rappuoli 2000). Reverse vaccinology uses bioinformatic techniques to find vaccine candidates that are likely to be surface expressed (hence antibody-accessible), tests these candidates for immunogenicity and protective ability in a high-throughput way and suggests a combination of antigens for a multi-subunit vaccine (Rappuoli 2000; Mora *et al.* 2006). More recently, the concept has been updated to account for within-species variation in gene content (Mora *et al.* 2006). An analysis of inter-strain diversity of group B streptococcus (GBS) found that only 80% of genes are shared by all strains compared (Tettelin *et al.* 2005), while a study of *Escherichia coli* found that just 40% of a combined set of proteins were common to all strains analysed (Welch *et al.* 2002). Such diversity in gene content within pathogen species compounds the difficulty of producing an effective vaccine in the face of frequent and extensive antigenic polymorphism. To develop a vaccine against GBS, the genome sequences of eight GBS isolates were analysed as above and tested for immunogenicity. Four antigens were found to elicit protective immunity in mice and their combination conferred protection against a large number of strains (Maione *et al.* 2005). The novel aspect of the study is that none of the proteins could be classified as universal. Three of the proteins were absent in several strains tested and the fourth showed limited surface accessibility in some of these. Use of multiple genome sequences in such a way has been termed 'pan-genomic reverse vaccinology' in reference to the concept of a 'pan-genome' that is shared by members of a microbial species by descent and horizontal gene transfer, but which is not wholly possessed by any one strain (Mora *et al.* 2006).

In a variant on this approach, Urwin *et al.* (2004) have studied variation in OMP of *N. meningitidis*. Consistent with the predictions of Gupta *et al.* (1996), these antigenic and polymorphic proteins (PorA and FetA) show a relatively strong strain structure, with associations between particular alleles that are more than expected by clonal descent alone. Urwin *et al.* (2004) suggest that such structuring can greatly simplify vaccine design. As both PorA and FetA confer protection against reinfection with a homologous strain, careful selection of variants for inclusion in a vaccine could provide broad protection. Urwin *et al.* (2004) suggest that up to 90% of meningococcal strains could be covered by a limited number of antigen combinations. In principle, such a formulation is likely to be resistant against the generation of escape mutants as such mutants can only emerge and spread if they undergo simultaneous changes at both loci. It remains to be seen how successful such strategies will be in these and other pathogens, but these latter studies raise the exciting prospect that theoretical predictions about strain structuring may be rigorously tested in natural populations, and the results put to practical application in vaccine design.

Diversifying selection on both hosts and pathogens clearly plays a large part in their ongoing coevolution. Existing theoretical models provide a framework for understanding some of the observed patterns of antigenic diversification, but many questions remain—from how to identify antigens to why the patterns of diversity vary so widely between pathogens and how our evolving understanding of these patterns can be best used to design effective vaccines against the antigenically variable pathogens that continue to elude us.

This work was supported by the US National Institutes of Health cooperative agreement 5U01GM076497 (Models of Infectious Disease Agent Study) and grant 1R01 AI48935 to M.L.

## REFERENCES

Abu-Raddad, L. J. & Ferguson, N. M. 2004 The impact of cross-immunity, mutation and stochastic extinction on pathogen diversity. *Proc. R. Soc. B* **271**, 2431–2438. (doi:10.1098/rspb.2004.2877)

Allen, T. M. *et al.* 2005 Selective escape from CD8+T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution. *J. Virol.* **79**, 13 239–13 249. (doi:10.1128/JVI.79.21.13239-13249.2005)

Asquith, B., Edwards, C. T. T., Lipsitch, M. & McLean, A. R. 2006 Inefficient cytotoxic T lymphocyte-mediated killing of HIV-1-infected cells *in vivo*. *PLoS Biol.* **4**, e90. (doi:10.1371/journal.pbio.0040090)

Babl, F. E., Pelton, S. I., Theodore, S. & Klein, J. O. 2001 Constancy of distribution of serogroups of invasive pneumococcal isolates among children: experience during 4 decades. *Clin. Infect. Dis.* **32**, 1155–1161. (doi:10.1086/319750)

Baltimore, R. S. 1998 Meningococcal infections. In *Bacterial infections of humans: epidemiology and control* (eds A. E. Evans & P. S. Brachman). New York, NY: Plenum Publishing.

Bihl, F. *et al.* 2006 Impact of HLA-B alleles, epitope binding affinity, functional avidity, and viral coinfection on the immunodominance of virus-specific CTL responses. *J. Immunol.* **176**, 4094–4101.

Birrer, M. J., Udem, S., Nathenson, S. & Bloom, B. R. 1981 Antigenic variants of measles virus. *Nature* **293**, 67–69. (doi:10.1038/293067a0)

Black, S. *et al.* 2000 Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. Northern California Kaiser Permanente Vaccine Study Center Group. *Pediatr. Infect. Dis. J.* **19**, 187–195. (doi:10.1097/00006454-200003000-00003)

Bogaert, D., De Groot, R. & Hermans, P. W. 2004*a* *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. *Lancet Infect. Dis.* **4**, 144–154. (doi:10.1016/S1473-3099(04)00938-7)

Bogaert, D., van Belkum, A., Sluijter, M., Luijendijk, A., de Groot, R., Rumke, H. C., Verbrugh, H. A. & Hermans, P. W. 2004*b* Colonisation by *Streptococcus pneumoniae* and *Staphylococcus aureus* in healthy children. *Lancet* **363**, 1871–1872. (doi:10.1016/S0140-6736(04)16357-5)

Boles, B. R., Thoendel, M. & Singh, P. K. 2004 Self-generated diversity produces "insurance effects" in biofilm communities. *Proc. Natl Acad. Sci. USA* **101**, 16 630–16 635. (doi:10.1073/pnas.0407460101)

Briles, D. E., Novak, L., Hotomi, M., van Ginkel, F. W. & King, J. 2005 Nasal colonization with *Streptococcus pneumoniae* includes subpopulations of surface and invasive pneumococci. *Infect. Immun.* **73**, 6945–6951. (doi:10.1128/IAI.73.10.6945-6951.2005)

Cartwright, K. A., Stuart, J. M., Jones, D. M. & Noah, N. D. 1987 The Stonehouse survey: nasopharyngeal carriage of meningococci and *Neisseria lactamica*. *Epidemiol. Infect.* **99**, 591–601.

Caugant, D. A., Kristiansen, B. E., Froholm, L. O., Bovre, K. & Selander, R. K. 1988 Clonal diversity of *Neisseria meningitidis* from a population of asymptomatic carriers. *Infect. Immun.* **56**, 2060–2068.

Caugant, D. A., Tzanakaki, G. & Kriz, P. 2007 Lessons from meningococcal carriage studies. *FEMS Microbiol. Rev.* **31**, 52–63. (doi:10.1111/j.1574-6976.2006.00052.x)

Criss, A. K., Kline, K. A. & Seifert, H. S. 2005 The frequency and rate of pilin antigenic variation in *Neisseria gonorrhoeae*. *Mol. Microbiol.* **58**, 510–519. (doi:10.1111/j.1365-2958.2005.04838.x)

Crowcroft, N. S. & Pebody, R. G. 2006 Recent developments in pertussis. *Lancet* **367**, 1926–1936. (doi:10.1016/S0140-6736(06)68848-X)

Dietz, K. 1979 Epidemiologic interference of virus populations. *J. Math. Biol.* **8**, 291–300.

Doherty, P. C., Turner, S. J., Webby, R. G. & Thomas, P. G. 2006 Influenza and the challenge for immunology. *Nat. Immunol.* **7**, 449–455. (doi:10.1038/ni1343)

Dolan-Livengood, J. M., Miller, Y. K., Martin, L. E., Urwin, R. & Stephens, D. S. 2003 Genetic basis for nongroupable *Neisseria meningitidis*. *J. Infect. Dis.* **187**, 1616–1628. (doi:10.1086/374740)

Dzikowski, R., Templeton, T. J. & Deitsch, K. 2006 Variant antigen gene expression in malaria. *Cell Microbiol.* **8**, 1371–1381. (doi:10.1111/j.1462-5822.2006.00760.x)

Edwards, C. T., Holmes, E. C., Pybus, O. G., Wilson, D. J., Viscidi, R. P., Abrams, E. J., Phillips, R. E. & Drummond, A. J. 2006 Evolution of the human immunodeficiency virus envelope gene is dominated by purifying selection. *Genetics* **174**, 1441–1453. (doi:10.1534/genetics.105.052019)

Ferguson, N. M., Galvani, A. P. & Bush, R. M. 2003 Ecological and immunological determinants of influenza evolution. *Nature* **422**, 428–433. (doi:10.1038/nature01509)

Fernandez, C. S. *et al.* 2005 Rapid viral escape at an immunodominant simian-human immunodeficiency virus cytotoxic T-lymphocyte epitope exacts a dramatic fitness cost. *J. Virol.* **79**, 5721–5731. (doi:10.1128/JVI.79.9.5721-5731.2005)

Fitch, W. M., Leiter, J. M., Li, X. Q. & Palese, P. 1991 Positive Darwinian evolution in human influenza A viruses. *Proc. Natl Acad. Sci. USA* **88**, 4270–4274. (doi:10.1073/pnas.88.10.4270)

Fraser, C., Hanage, W. P. & Spratt, B. G. 2005 Neutral microepidemic evolution of bacterial pathogens. *Proc. Natl Acad. Sci. USA* **102**, 1968–1973. (doi:10.1073/pnas.0406993102)

Friedrich, T. C. *et al.* 2004 Reversion of CTL escape-variant immunodeficiency viruses *in vivo*. *Nat. Med.* **10**, 275–281. (doi:10.1038/nm998)

Gao, F. *et al.* 1999 Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature* **397**, 436–441. (doi:10.1038/17130)

Gaschen, B. *et al.* 2002 Diversity considerations in HIV-1 vaccine selection. *Science* **296**, 2354–2360. (doi:10.1126/science.1070441)

Gleich, S., Morad, Y., Echague, R., Miller, J. R., Kornblum, J., Sampson, J. S. & Butler, J. C. 2000 *Streptococcus pneumoniae* serotype 4 outbreak in a home for the aged: report and review of recent outbreaks. *Infect. Control Hosp. Epidemiol.* **21**, 711–717. (doi:10.1086/501717)

Gog, J. R. & Grenfell, B. T. 2002 Dynamics and selection of many-strain pathogens. *Proc. Natl Acad. Sci. USA* **99**, 17 209–17 214. (doi:10.1073/pnas.252512799)

Gog, J. R. & Swinton, J. 2002 A status-based approach to multiple strain dynamics. *J. Math. Biol.* **44**, 169–184. (doi:10.1007/s002850100120)

Goldblatt, D. *et al.* 2005 Antibody responses to nasopharyngeal carriage of *Streptococcus pneumoniae* in adults: a longitudinal household study. *J. Infect. Dis.* **192**, 387–393. (doi:10.1086/431524)

Goldschneider, I., Gotschlich, E. C. & Artenstein, M. S. 1969 Human immunity to the meningococcus. II. Development of natural immunity. *J. Exp. Med.* **129**, 1327–1348. (doi:10.1084/jem.129.6.1327)

Gomes, M. G., Medley, G. F. & Nokes, D. J. 2002 On the determinants of population structure in antigenically diverse pathogens. *Proc. R. Soc. B* **269**, 227–233. (doi:10.1098/rspb.2001.1869)

Goulder, P. J. *et al.* 1997 Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nat. Med.* **3**, 212–217. (doi:10.1038/nm0297-212)

Grassly, N. C., Harvey, P. H. & Holmes, E. C. 1999 Population dynamics of HIV-1 inferred from gene sequences. *Genetics* **151**, 427–438.

Gray, B. M., Turner, M. E. & Dillon Jr, H. C. 1982 Epidemiologic studies of *Streptococcus pneumoniae* in infants. The effects of season and age on pneumococcal acquisition and carriage in the first 24 months of life. *Am. J. Epidemiol.* **116**, 692–703.

Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L., Daly, J. M., Mumford, J. A. & Holmes, E. C. 2004 Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**, 327–332. (doi:10.1126/science.1090727)

Gupta, S. & Galvani, A. 1999 The effects of host heterogeneity on pathogen population structure. *Phil. Trans. R. Soc. B* **354**, 711–719. (doi:10.1098/rstb.1999.0424)

Gupta, S. & Maiden, M. C. 2001 Exploring the evolution of diversity in pathogen populations. *Trends Microbiol.* **9**, 181–185. (doi:10.1016/S0966-842X(01)01986-2)

Gupta, S., Swinton, J. & Anderson, R. M. 1994 Theoretical studies of the effects of heterogeneity in the parasite population on the transmission dynamics of malaria. *Proc. R. Soc. B* **256**, 231–238. (doi:10.1098/rspb.1994.0075)

Gupta, S., Maiden, M. C., Feavers, I. M., Nee, S., May, R. M. & Anderson, R. M. 1996 The maintenance of strain structure in populations of recombining infectious agents. *Nat. Med.* **2**, 437–442. (doi:10.1038/nm0496-437)

Gupta, S., Ferguson, N. & Anderson, R. 1998 Chaos, persistence, and evolution of strain structure in antigenically diverse infectious agents. *Science* **280**, 912–915. (doi:10.1126/science.280.5365.912)

Hammerschmidt, S. *et al.* 1996 Capsule phase variation in *Neisseria meningitidis* serogroup B by slipped-strand mispairing in the polysialyltransferase gene (siaD): correlation with bacterial invasion and the outbreak of meningococcal disease. *Mol. Microbiol.* **20**, 1211–1220. (doi:10.1111/j.1365-2958.1996.tb02641.x)

Hanage, W. P., Kaijalainen, T. H., Syrjanen, R. K., Auranen, K., Leinonen, M., Makela, P. H. & Spratt, B. G. 2005 Invasiveness of serotypes and clones of *Streptococcus pneumoniae* among children in Finland. *Infect. Immun.* **73**, 431–435. (doi:10.1128/IAI.73.1.431-435.2005)

Hausdorff, W. P., Bryant, J., Kloek, C., Paradiso, P. R. & Siber, G. R. 2000 The contribution of specific pneumococcal serogroups to different disease manifestations: implications for conjugate vaccine formulation and use, part II [In Process Citation]. *Clin. Infect. Dis.* **30**, 122–140. (doi:10.1086/313609)

Hausdorff, W. P., Feikin, D. R. & Klugman, K. P. 2005 Epidemiological differences among pneumococcal serotypes. *Lancet Infect. Dis.* **5**, 83–93.

Hemelaar, J., Gouws, E., Ghys, P. D. & Osmanov, S. 2006 Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS* **20**, W13–W23. (doi:10.1097/01.aids.0000247564.73009.bc)

Hogberg, L., Geli, P., Ringberg, H., Melander, E., Lipsitch, M. & Ekdahl, K. 2007 Age and serogroup related differences in the observed duration of nasopharyngeal carriage of penicillin-resistant pneumococci. *J. Clin. Microbiol.* **45**, 948–952. (doi:10.1128/JCM.01913-06)

Holmes, E. C. 2004 The phylogeography of human viruses. *Mol. Ecol.* **13**, 745–756. (doi:10.1046/j.1365-294X.2003.02051.x)

Holmes, E. C. *et al.* 2005 Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. *PLoS Biol.* **3**, e300. (doi:10.1371/journal.pbio.0030300)

Janeway, C. A., Travers, P., Walport, M. & Shlomchik, M. 2001 *Immunobiology*, 5th edn. New York, NY: Garland Publishing.

Jolley, K. A., Kalmusova, J., Feil, E. J., Gupta, S., Musilek, M., Kriz, P. & Maiden, M. C. 2000 Carried meningococci in the Czech Republic: a diverse recombining population. *J. Clin. Microbiol.* **38**, 4492–4498.

Jolley, K. A., Brehony, C. & Maiden, M. C. 2007 Molecular typing of meningococci: recommendations for target choice and nomenclature. *FEMS Microbiol. Rev.* **31**, 89–96. (doi:10.1111/j.1574-6976.2006.00057.x)

Kang, J. & Blaser, M. J. 2006 Bacterial populations as perfect gases: genomic integrity and diversification tensions in *Helicobacter pylori*. *Nat. Rev. Microbiol.* **4**, 826–836. (doi:10.1038/nrmicro1528)

Kent, S. J., Fernandez, C. S., Dale, C. J. & Davenport, M. P. 2005 Reversion of immune escape HIV variants upon transmission: insights into effective viral immunity. *Trends Microbiol.* **13**, 243–246. (doi:10.1016/j.tim.2005.03.011)

Koelle, K., Cobey, S., Grenfell, B. & Pascual, M. 2006 Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science* **314**, 1898–1903. (doi:10.1126/science.1132745)

Konradsen, H. B. & Kaltoft, M. S. 2002 Invasive pneumococcal infections in Denmark from 1995 to 1999: epidemiology, serotypes, and resistance. *Clin. Diag. Lab. Immunol.* **9**, 358–365. (doi:10.1128/CDLI.9.2.358-365.2002)

Kyaw, M. H. *et al.* 2006 Effect of introduction of the pneumococcal conjugate vaccine on drug-resistant *Streptococcus pneumoniae*. *N. Engl. J. Med.* **354**, 1455–1463. (doi:10.1056/NEJMoa051642)

Leslie, A. J. *et al.* 2004 HIV evolution: CTL escape mutation and reversion after transmission. *Nat. Med.* **10**, 282–289. (doi:10.1038/nm992)

Levin, B. R. & Bergstrom, C. T. 2000 Bacteria are different: observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes. *Proc. Natl Acad. Sci. USA* **97**, 6981–6985. (doi:10.1073/pnas.97.13.6981)

Li, B., Gladden, A. D., Altfeld, M., Kaldor, J. M., Cooper, D. A., Kelleher, A. D. & Allen, T. M. 2007 Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution. *J. Virol.* **81**, 193–201. (doi:10.1128/JVI.01231-06)

Lipsitch, M. 1997 Vaccination against colonizing bacteria with multiple serotypes. *Proc. Natl Acad. Sci. USA* **94**, 6571–6576. (doi:10.1073/pnas.94.12.6571)

Lipsitch, M. 1999 Bacterial vaccines and serotype replacement: lessons from *Haemophilus influenzae* and prospects for *Streptococcus pneumoniae*. *Emerg. Infect. Dis.* **5**, 336–345.

Lipsitch, M. & Moxon, E. R. 1997 Virulence and transmissibility of pathogens: what is the relationship? *Trends Microbiol.* **5**, 31–37. (doi:10.1016/S0966-842X(97)81772-6)

Lipsitch, M., Whitney, C. G., Zell, E. R., Kaijalainen, M., Dagan, R. & Malley, R. 2005 Are anticapsular antibodies the primary mechanism of protection against invasive pneumococcal disease? *PLoS Med.* **2**, e15. (doi:10.1371/journal.pmed.0020015)

Lord, F. T. & Heffron, R. 1938 *Pneumonia and serum therapy*. London, UK: Commonwealth Fund.

Maione, D. *et al.* 2005 Identification of a universal Group B streptococcus vaccine by multiple genome screen. *Science* **309**, 148–150. (doi:10.1126/science.1109869)

Malley, R., Trzcinski, K., Srivastava, A., Thompson, C. M., Anderson, P. W. & Lipsitch, M. 2005 CD4+T cells mediate antibody-independent acquired immunity to pneumococcal colonization. *Proc. Natl Acad. Sci. USA* **102**, 4848–4853. (doi:10.1073/pnas.0501254102)

Mazumdar, P. M. H. 2003 In *History of immunology* (ed. W. E. Paul), 5th edn. *Fundamental immunology*, ch. 2. Philadelphia, PA: Lippincott Williams & Wilkins.

McCool, T. L., Cate, T. R., Moy, G. & Weiser, J. N. 2002 The immune response to pneumococcal proteins during experimental human carriage. *J. Exp. Med.* **195**, 359–365. (doi:10.1084/jem.20011576)

McCormick, A. W. *et al.* 2003 Geographic diversity and temporal trends of antimicrobial resistance in *Streptococcus pneumoniae* in the United States. *Nat. Med.* **9**, 424–430. (doi:10.1038/nm839)

Mills, C. E., Robins, J. M. & Lipsitch, M. 2004 Transmissibility of 1918 pandemic influenza. *Nature* **432**, 904–906. (doi:10.1038/nature03063)

Mooi, F. R., van Loo, I. H. & King, A. J. 2001 Adaptation of *Bordetella pertussis* to vaccination: a cause for its reemergence? *Emerg. Infect. Dis.* **7**, 526–528.

Mora, M., Donati, C., Medini, D., Covacci, A. & Rappuoli, R. 2006 Microbial genomes and vaccine design: refinements to the classical reverse vaccinology approach. *Curr. Opin. Microbiol.* **9**, 532–536. (doi:10.1016/j.mib.2006.07.003)

Moxon, R., Bayliss, C. & Hood, D. 2006 Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu. Rev. Genet.* **40**, 307–333. (doi:10.1146/annurev.genet.40.110405.090442)

Mueller, J. E., Borrow, R. & Gessner, B. D. 2006 Meningococcal serogroup W135 in the African meningitis belt: epidemiology, immunity and vaccines. *Expert Rev. Vaccines* **5**, 319–336. (doi:10.1586/14760584.5.3.319)

Nelson, M. I. *et al.* 2006 Stochastic processes are key determinants of short-term evolution in influenza A virus. *PLoS Pathog.* **2**, e125. (doi:10.1371/journal.Ppat.0020125)

Nickle, D. C., Jensen, M. A., Gottlieb, G. S., Shriner, D., Learn, G. H., Rodrigo, A. G. & Mullins, J. I. 2003 Consensus and ancestral state HIV vaccines. *Science* **299**, 1515–1518. (doi:10.1126/science.299.5612.1515c)

Nowak, M. A. & May, R. M. 2000 *Virus dynamics*. Oxford, UK: Oxford University Press.

Ogg, G. S. *et al.* 1998 Quantitation of HIV-1-specific cytotoxic T lymphocytes and plasma load of viral RNA. *Science* **279**, 2103–2106. (doi:10.1126/science.279.5359.2103)

Panum, P. 1940 *Observations made during the epidemic of measles on the Faroe Islands in the year 1846*. New York, NY: Delta Omega.

Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M. & Ho, D. D. 1996 HIV-1 dynamics *in vivo*: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**, 1582–1586. (doi:10.1126/science.271.5255.1582)

Pizza, M. *et al.* 2000 Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* **287**, 1816–1820. (doi:10.1126/science.287.5459.1816)

Plotkin, J. B. & Dushoff, J. 2003 Codon bias and frequency-dependent selection on the hemagglutinin epitopes of influenza A virus. *Proc. Natl Acad. Sci. USA* **100**, 7152–7157. (doi:10.1073/pnas.1132114100)

Plotkin, J. B., Dushoff, J. & Fraser, H. B. 2004 Detecting selection using a single genome sequence of *M. tuberculosis* and *P. falciparum*. *Nature* **428**, 942–945. (doi:10.1038/nature02458)

Pollard, A. J. 2004 Global epidemiology of meningococcal disease and vaccine efficacy. *Pediatr. Infect. Dis. J.* **23**, S274–S279.

Raghunathan, P. L., Bernhardt, S. A. & Rosenstein, N. E. 2004 Opportunities for control of meningococcal disease in the United States. *Annu. Rev. Med.* **55**, 333–353. (doi:10.1146/annurev.med.55.091902.103612)

Rambaut, A., Posada, D., Crandall, K. A. & Holmes, E. C. 2004 The causes and consequences of HIV evolution. *Nat. Rev. Genet.* **5**, 52–61. (doi:10.1038/nrg1246)

Rappuoli, R. 2000 Reverse vaccinology. *Curr. Opin. Microbiol.* **3**, 445–450. (doi:10.1016/S1369-5274(00)00119-3)

Regev-Yochay, G., Dagan, R., Raz, M., Carmeli, Y., Shainberg, B., Derazne, E., Rahav, G. & Rubinstein, E. 2004 Association between carriage of *Streptococcus pneumoniae* and *Staphylococcus aureus* in children. *J. Am. Med. Assoc.* **292**, 716–720. (doi:10.1001/jama.292.6.716)

Rehermann, B. & Nascimbeni, M. 2005 Immunology of hepatitis B virus and hepatitis C virus infection. *Nat. Rev. Immunol.* **5**, 215–229. (doi:10.1038/nri1573)

Richman, D. D., Wrin, T., Little, S. J. & Petropoulos, C. J. 2003 Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc. Natl Acad. Sci. USA* **100**, 4144–4149. (doi:10.1073/pnas.0630530100)

Rosenstein, N. E. *et al.* 1999 The changing epidemiology of meningococcal disease in the United States, 1992–1996. *J. Infect. Dis.* **180**, 1894–1901. (doi:10.1086/315158)

Sharp, P. M., Robertson, D. L. & Hahn, B. H. 1995 Cross-species transmission and recombination of 'AIDS' viruses. *Phil. Trans. R. Soc. B* **349**, 41–47. (doi:10.1098/rstb.1995.0089)

Sleeman, K. L., Griffiths, D., Shackley, F., Diggle, L., Gupta, S., Maiden, M. C., Moxon, E. R., Crook, D. W. & Peto, T. E. 2006 Capsular serotype-specific attack rates and duration of carriage of *Streptococcus pneumoniae* in a population of children. *J. Infect. Dis.* **194**, 682–688. (doi:10.1086/505710)

Smith, T., Lehmann, D., Montgomery, J., Gratten, M., Riley, I. D. & Alpers, M. P. 1993 Acquisition and invasiveness of different serotypes of *Streptococcus pneumoniae* in young children. *Epidemiol. Infect.* **111**, 27–39.

Smith, D. J., Lapedes, A. S., de Jong, J. C., Bestebroer, T. M., Rimmelzwaan, G. F., Osterhaus, A. D. & Fouchier, R. A. 2004 Mapping the antigenic and genetic evolution of influenza virus. *Science* **305**, 371–376. (doi:10.1126/science.1097211)

Stephens, D. S. 1999 Uncloaking the meningococcus: dynamics of carriage and disease. *Lancet* **353**, 941–942. (doi:10.1016/S0140-6736(98)00279-7)

Stephens, D. S. 2007 Conquering the meningococcus. *FEMS Microbiol. Rev.* **31**, 3–14. (doi:10.1111/j.1574-6976.2006.00051.x)

Swartley, J. S., Marfin, A. A., Edupuganti, S., Liu, L.-J., Cieslak, P., Perkins, B., Wenger, J. D. & Stephens, D. S. 1997 Capsule switching of *Neisseria meningitidis*. *Proc. Natl Acad. Sci. USA* **94**, 271–276. (doi:10.1073/pnas.94.1.271)

Taylor, J. E. & Rudenko, G. 2006 Switching trypanosome coats: what's in the wardrobe? *Trends Genet.* **22**, 614–620. (doi:10.1016/j.tig.2006.08.003)

Tettelin, H. *et al.* 2005 Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc. Natl Acad. Sci. USA* **102**, 13 950–13 955. (doi:10.1073/pnas.0506758102)

Trzcinski, K., Thompson, C., Malley, R. & Lipsitch, M. 2005 Antibodies to conserved pneumococcal antigens correlate with, but are not required for, protection against pneumococcal colonization induced by prior exposure in a mouse model. *Infect. Immun.* **73**, 7043–7046. (doi:10.1128/IAI.73.10.7043-7046.2005)

Urwin, R., Russell, J. E., Thompson, E. A., Holmes, E. C., Feavers, I. M. & Maiden, M. C. 2004 Distribution of surface protein variants among hyperinvasive meningococci: implications for vaccine design. *Infect. Immun.* **72**, 5955–5962. (doi:10.1128/IAI.72.10.5955-5962.2004)

van Rossum, A. M., Lysenko, E. S. & Weiser, J. N. 2005 Host and bacterial factors contributing to the clearance of colonization by *streptococcus pneumoniae* in a murine model. *Infect. Immun.* **73**, 7718–7726. (doi:10.1128/IAI.73.11.7718-7726.2005)

Viboud, C., Pakdaman, K., Boelle, P. Y., Wilson, M. L., Myers, M. F., Valleron, A. J. & Flahault, A. 2004 Association of influenza epidemics with global climate variability. *Eur. J. Epidemiol.* **19**, 1055–1059. (doi:10.1007/s10654-004-2450-9)

Volkman, S. K. *et al.* 2007 A genome-wide map of diversity in *Plasmodium falciparum*. *Nat. Genet.* **39**, 113–119. (doi:10.1038/ng1930)

Walker, B. D. & Korber, B. T. 2001 Immune control of HIV: the obstacles of HLA and viral diversity. *Nat. Immunol.* **2**, 473–475. (doi:10.1038/88656)

Welch, R. A. *et al.* 2002 Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **99**, 17 020–17 024. (doi:10.1073/pnas.252529799)

Wolf, Y. I., Viboud, C., Holmes, E. C., Koonin, E. V. & Lipman, D. J. 2006 Long intervals of stasis punctuated by bursts of positive selection in the seasonal evolution of influenza A virus. *Biol. Direct.* **1**, 34. (doi:10.1186/1745-6150-1-34)

Yazdankhah, S. P. & Caugant, D. A. 2004 *Neisseria meningitidis*: an overview of the carriage state. *J. Med. Microbiol.* **53**, 821–832. (doi:10.1099/jmm.0.45529-0)

Yazdankhah, S. P. *et al.* 2004 Distribution of serogroups and genotypes among disease-associated and carried isolates of *Neisseria meningitidis* from the Czech Republic, Greece, and Norway. *J. Clin. Microbiol.* **42**, 5146–5153. (doi:10.1128/JCM.42.11.5146-5153.2004)

Zambon, M. C. 1999 Epidemiology and pathogenesis of influenza. *J. Antimicrob. Chemother.* **44**, 3–9. (doi:10.1093/jac/44.suppl_2.3)