CHAPTER 15

Evolution of Antigenic Variation[1]

Steven A. Frank

Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697–2525, USA

15.1 INTRODUCTION

Infectious disease remains a major cause of morbidity and mortality. Consequently, great research effort has been devoted to parasites and to host immune responses that fight parasites.[2] This has led to rapid progress in understanding the biology of parasites, including the molecular details about how parasites invade hosts and escape host immune defenses. Vaccines have followed, sometimes with spectacular success [45].

But many parasites escape host defense by varying their *antigenic* molecules recognized by host immunity. Put another way, rapid evolution of antigenic molecules all too often prevents control of parasite populations. The challenge has been to link molecular understanding of parasite molecules to their evolutionary change and to the antigenic variation in populations of parasites.

The growth of information about antigenic variation provides a special opportunity. For example, one can find in the literature details about how single amino acid changes in parasite molecules allow escape from antibody binding, and how that escape promotes the spread of variant parasites. Evolutionary studies no longer depend on abstractions – one can pinpoint the physical basis for success or failure and the consequences for change in populations.

Molecular understanding of host–parasite recognition leads to a comparative question about the forces that shape variability. Why do some viruses escape host immunity by varying so rapidly over a few years, whereas other viruses hardly change their antigens? The answer leads to the processes that shape genetic variability and evolutionary change. The causes of variability and change provide the basis for understanding why simple vaccines work well against some viruses, whereas complex vaccine strategies achieve only limited success against other viruses.

The battle between host and parasite often comes down to the rates at which attacker and defender molecules bind or evade each other. The biochemical details of binding and recognition set the rules of engagement that shape the pace, scale, and pattern of diversity and the nature of evolutionary change [45].

The first section of this chapter lists the different ways in which parasites can gain by varying their antigenic molecules. The parasites may, for example, extend their time of infection within a particular host by changing their antigenic molecules to evade the specific recognition built up by the host immune system. Or the parasites may vary to attack hosts that had been exposed previously and had built up specific recognition against prior antigenic molecules.

The second section focuses on the nature of binding and recognition between host and parasite molecules. I summarize the different ways in which parasites generate new variants in order to escape molecular recognition.

The third section builds up the individual molecular interactions into the dynamics of a single infection within a host. The parasites spread in the host, triggering immune attack against dominant antigens. The battle within the host develops through changes in population numbers – the numbers of parasites with particular antigens and the numbers of immune cells that specifically bind to particular antigens.

The fourth and fifth sections discuss how the successes and failures of different parasite antigens within each host determine the rise and fall of parasite variants over space and time. The distribution of parasite variants sets the immune memory profiles of different hosts, which in turn

[1]This chapter is a condensed version of my book, *Immunology and Evolution of Infectious Disease*, Princeton University Press, Princeton, NJ, 2002.

[2]I use the word *parasite* for all infectious agents, including viruses, bacteria, and protozoa.

Encyclopedia of Infectious Diseases: Modern Methodologies, by M. Tibayrenc

shape the landscape in which parasite variants succeed or fail. These coevolutionary processes determine the natural selection of antigenic variants and the course of evolution in the parasite population. I discuss several methods that can be used to infer the evolutionary processes that shape antigenic variation.

15.2 WHY DO PARASITES VARY?

In this section, I describe the benefits that antigenic variation provides to parasites. These benefits help to explain why parasites vary in certain ways.

15.2.1 Extend Length of Infection

Trypanosoma brucei, the protozoan parasite responsible for human African trypanosomiasis (formerly sleeping sickness), changes its dominant antigenic surface glycoprotein at a rate of 10^{-3} to 10^{-2} per cell division [138]. The trypanosome changes to another surface coat by altering expression between different genes already present in the genome. Infections lead to successive waves of parasitemia and clearance as novel antigenic types spread and are then checked by specific immunity.

Some viruses, such as HIV, escape immune attack by mutating their dominant epitopes [86]. Mutational changes to new, successful epitopes may be rare in each replication of the virus. But the very large population size of viruses within a host means that mutations, rare in each replication, often occur at least once in the host in each parasite generation.

For parasites that produce antigenic variants within hosts, the infection continues until the host controls all variants, raises an immune response against a non-varying epitope, or clears the parasite by nonspecific defenses. Antigenic variation can extend the total time before clearance [36,47,89]. Extended infection benefits the parasite by increasing the chances for transmission to new hosts.

15.2.2 Infect Hosts with Prior Exposure

Host immune memory recognizes and mounts a rapid response against previously encountered antigens. Antigenic variants that differ from a host's previous infections escape that host's memory response. The distribution of immune memory profiles between hosts determines the success of each parasite variant.

In the simplest case, each antigenic type acts like a separate parasite that does not cross-react with other variants. As host individuals age, they become infected by and recover from different antigenic variants. Thus, the host population can be classified by resistance profiles based on the past infection and recovery of each individual [7].

Two extreme cases define the range of outcomes. On the one hand, each variant may occasionally spread epidemically through the host population. This leaves a large fraction of the hosts resistant upon recovery, driving that particular variant down in frequency because it has few hosts it can infect. The variant can spread again only after many resistant hosts die and are replaced by young hosts without prior exposure to that antigen. Variants may, on the other hand, be maintained endemically in the host population. This requires a balance between the rate at which infections lead to host death or recovery and the rate at which new susceptible hosts enter the population. The parasite population maintains as many variants as arise and do not cross-react, subject to "birth–death" processes governing the stochastic origin of new variants and the loss of existing variants.

15.2.3 Infect Hosts with Genetically Variable Resistance

Host genotype can influence susceptibility to different parasite variants. For example, MHC genotype determines the host's efficiency in presenting particular epitopes to T cells. From the parasite's point of view, a particular antigenic variant may be able to attack some host genotypes but not others.

Hill [55] pointed out that hepatitis B virus provides a good model for studying the interaction between MHC and parasite epitopes. Preliminary reports found associations between MHC genotype and whether infections were cleared or became persistent [6,57,136]. The hepatitis B virus genome is very small (about 3000 base pairs, or bp), which should allow direct study of how variation in viral epitopes interacts with the host's MHC genotype. Host genotype can also affect the structure of the cellular receptors to which parasites attach. For example, the human CCR5 gene encodes a coreceptor required for HIV-1 to enter macrophages. A 32 bp deletion of this gene occurs at a frequency of 0.1 in European populations. This deletion prevents the virus from entering macrophages [81,96,122]. It is not clear whether minor variants of cellular receptors occur sufficiently frequently to favor widespread matching variation of parasite surface antigens. Several cases of this sort may eventually be found, but in vertebrate hosts genetic variation of cellular receptors may be a relatively minor cause of parasite diversity.

15.2.4 Vary Attachment Characters

Parasite surface antigens often play a role in attachment and entry into host cells or attachment to particular types of host tissue. Varying these attachment characters allows attack of different cell types or adhesion to various tissues. Such variability can provide the parasite with additional resources or protection from host defenses. Protozoan parasites of the genus *Plasmodium* cause malaria in a variety of vertebrate hosts. Several *Plasmodium* species switch antigenic type [22]. Switching has been studied most extensively in *Plasmodium falciparum* [109]. Programmed mechanisms of gene expression choose a single gene from among many archival genetic copies for the *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) [30]. As its name implies, the parasite expresses this antigen on the surface of infected erythrocytes. PfEMP1 induces an antibody response, which likely plays a role in the host's ability to control infection [109]. PfEMP1 influences

cytoadherence of infected erythrocytes to capillary endothelia [109]. This adherence may help the parasite to avoid clearance in the spleen. Thus, antigenic variants can influence the course of infection by escaping specific recognition and by hiding from host defenses [109]. Full understanding of the forces that have shaped the archival repertoire, switching process, and course of infection requires study of both specific immune recognition and cytoadherence properties of the different antigenic variants.

HIV provides another example. This virus links its surface protein gp120 to two host-cell receptors before it enters the cell [96]. One host-cell receptor, CD4, appears to be required by most HIV variants (but see [117]). The second host-cell receptor can be CCR5 or CXCR4. Macrophages express CCR5. A host that lacks functional CCR5 proteins apparently can avoid infection by HIV, suggesting that the initial invasion requires infection of macrophages. HIV isolates with tropism for CCR5 can be found throughout the infection; this HIV variant is probably the transmissive form that infects new hosts. As an infection proceeds within a host, HIV variants with tropism for CXCR4 emerge [96]. This host-cell receptor occurs on the surface of the $CD4^+$ (helper) T lymphocytes. The emergence of viral variants with tropism for CXCR4 coincides with a drop in $CD4^+$ T cells and onset of the immunosuppression that characterizes AIDS. These examples show that variable surface antigens may sometimes occur because they provide alternative cell or tissue tropisms rather than, or in addition to, escape from immune recognition.

15.3 MECHANISMS THAT GENERATE VARIATION

In this section, I summarize the different ways in which parasites generate antigenic variants. The amount of new variation and the kinds of new variants influence antigenic polymorphism and the pace of evolutionary change [36,47,89].

15.3.1 Mutation and Hypermutation

RNA virus populations typically have high frequencies of mutants and often evolve rapidly. However, few studies have provided direct estimates of mutation rates. The limited data suggest relatively high mutation rates on the order of 10^{-4} to 10^{-5} per base per replication [31,38,39,106].

Drake et al. [38] summarized mutation rates for various microbes with DNA chromosomes. They found an amazingly consistent value of approximately 0.003 mutations per genome per generation. This value holds over genomes that vary in total size by from 6×10^3 to 4×10^7 bp; consequently, the per base mutation rates also vary over four orders of magnitude from 7×10^{-7} to 7×10^{-11}.

None of the microbes summarized in Drake et al. face intense, constant selective pressure on antigens imposed by vertebrate immunity – for example, it is unlikely that *Escherichia coli* depends on antigenic variation to avoid clearance from its hosts. It would be interesting to know if pathogens under very intense selection by host immunity have higher baseline mutation rates than related microbes under less intense immune pressure. High genome-wide mutation rates arise in bacteria by spontaneous mutator mutations, in which the mutator alleles raise the error rate during replication [38]. The mutator alleles probably are various DNA replication and repair enzymes. Ten or more genes of *E. coli* can develop mutator mutations. Assuming that each gene has about 1000 bases, the overall mutation rate of mutator loci is $10 \times 1000 \times 5 \times 10^{-10} \approx 10^{-6}$ to 10^{-5}, based on the per base mutation rate in given in Drake et al. [38]. Some mutations will be nearly neutral; others will cause extremely high mutation rates and will never increase in frequency. Typical *E. coli* cultures accumulate mutator mutants at a frequency of less than 10^{-5} [78], probably because most mutations are deleterious and therefore selection does not favor increased mutation rates. However, mutators can be strongly favored when the competitive conditions and the selective environment provide opportunities for the mutators to generate more beneficial mutations than the non-mutators [28,78]. In this case, mutators increase because they are linked with a higher frequency of beneficial mutations.

Although mutators are typically rare in freshly grown laboratory cultures, hospital isolates of *E. coli* and *Salmonella enterica* sometimes have mutator frequencies above 10^{-2} [51,65,73]. Extensive serial passage in the laboratory can also lead to high frequencies of mutators [124]. Thus, it appears that rapid change of hosts or culture conditions can increase the frequency of mutators 1000-fold relative to stable environmental conditions. As Drake et al. [38] point out, theory suggests that mutators can speed adaptation in asexual microbes [74,89,135]. It would be interesting to compare naturally occurring frequencies of mutators in stable and rapidly changing selective environments.

Targeting mutations to key loci would be more efficient than raising the genome-wide mutation rate. Various mechanisms can increase the mutation rate over short runs of nucleotides [47,111]. For example, *Streptococcus pyogenes* coats its surface with a variable M protein, of which 80 antigenically distinct variants are known [43,71]. The amino acid sequence of the M6 serotype revealed repeats in three regions of the protein [58,59].

Sequence analysis of variant M proteins suggests that mutations occur by generating both gains and losses of the duplications. These mutations probably arise by intragenic recombination between the DNA repeats, but may be created by slippage during replication. Slippage mutations over repeated DNA lead to gain or loss in the number of repeats and occur at frequencies much higher than typical replication errors [29]. The repeats of the M protein are multiples of three bases; thus changes in repeat number do not cause frameshift mutations. Some of the repeats vary slightly in base composition, so recombinations can alter sequence composition as well as total length. Fussenegger [47] reviews several other cases of bacterial cell-wall proteins that have repeated

sequences, most of which occur in multiples of 3 bp. Repeats are often associated with binding domains for other proteins or polysaccharides [145], so perhaps the ability to generate variable-length domains provides an advantage in attachment to host tissues or in escape from host immunity.

Apart from the well-known case of repeats and replication slippage, no evidence at present associates antigenic sites with higher replication errors. But this would certainly be an interesting problem to study further. One could, for example, focus on associations between mutation rate and nucleotide sequence. Comparison would be particularly interesting between epitopes that evolve rapidly and conserved regions of antigenic molecules that evolve slowly. Such comparison may help to identify aspects of nucleotide composition that promote higher error rates in replication.

In summary, microbial mutation rates per nucleotide decline with increasing genome size, causing a nearly constant mutation rate per genome per generation of about 0.003. Genome-wide hypermutation can raise the mutation rate at all sites within the genome. Such mutator phenotypes probably have altered replication enzymes. Low frequencies of mutator phenotypes have been observed in stable populations of *E. coli*, whereas fluctuating populations appear to maintain higher frequencies of mutators. In some cases, hypermutation may be targeted to certain genes by DNA repeats and other DNA sequence motifs that promote local replication errors.

15.3.2 Stochastic Switching Between Archival Copies

Many pathogens change critical surface molecules by switching expression between alternative genes. At least four types of switch mechanisms occur: replication errors that turn expression on or off, invertible promoters that change the direction of transcription, gene conversion into fixed expression sites, and transcriptional silencing of alternative genes.

15.3.2.1 Regulatory switches by replication errors of short repeats Short, repeated nucleotide sequences often lead to high error rates during replication. Repeats have recurring units typically with 1–5 bases per unit. Short, repeated DNA sequences probably lead to replication errors by slipped-strand mispairing [29,75,87]. Errors apparently arise when a DNA polymerase either skips forward a repeat unit, causing a deletion of one unit, or slips back one unit, producing a one-unit insertion.Gene expression can be turned on or off by insertions or deletions. Inserted or deleted repeats within the coding sequence cause frameshift mutations that prevent translation and production of a full protein. For example, the 11 opacity genes of *Neisseria meningitidis* influence binding to host cells and tissue tropism.These genes each have between eight and 28 CTCTT repeats, which can disrupt or restore the proper translational frame as the number of repeats changes [129,130].The limited repertoire of 11 genes and the crude on–off switching suggest that variable expression has more to do with altering cell tropism than with escape from host immunity [47].On–off switches can also be created by short repeats in transcriptional control regions. *Bordetella pertussis* controls expression of two distinct fimbriae by transcriptional switching [144]. Fimbriae are bacterial surface fibers that attach to host tissues. Particular cells produce both, only one, or neither of the fimbrial types. Sequences of about 15 C nucleotides in the transcriptional promoters of each of the two genes influence expression.The actual length of the poly-C sequence varies, probably by slipped-strand mispairing during replication. The length affects transcription of the attached gene. Thus, by the stochastic process of replication errors, the individual loci are turned on and off. Again, this sort of switching may have more to do with tissue tropism than with escape from immune recognition.

15.3.2.2 Invertible sequences *E. coli* stores two alternative fimbriae genes adjacent to each other on its chromosome [1]. A promoter region between the two genes controls transcription. The promoter triggers transcription in only one direction, thus expressing only one of the two variants. Occasionally, the promoter flips orientation, activating the alternative gene. The ends of the promoter have inverted repeats, which play a role in the recombination event that mediates the sequence inversion. *Salmonella* uses a similar mechanism to control flagellum expression [120].

Moraxella species use a different method to vary pilin expression [79,115]. The variable part of the pilin gene has alternate cassettes stored in adjacent locations. Inverted repeats flank the pair of alternate cassettes, causing the whole complex occasionally to flip orientation.The gene starts with an initial constant region and continues into one of the cassettes within the invertible complex.When the complex flips, the alternate variable cassette completes the gene. Several bacteriophage use a similar inversion system to switch genes encoding their tail fibers, which determine host range [61,66].

15.3.2.3 Gene conversion Some pathogens store many variant genes for a surface antigen, but express only one of the copies at any time. For example, there may be a single active expression site at which transcription occurs. Occasionally, one of the variant loci copies itself to the expression site by gene conversion – a type of intragenomic recombination that converts the target without altering the donor sequence. The genome preserves the archival library without change, but alters the expressed allele.

The spirochete *Borrelia hermsii* has approximately 30 alternative loci that encode an abundant surface lipoprotein [12]. There is a single active expression site when the spirochete is in mammalian hosts [13]. The expression site is changed by gene conversion to one of the variant archival copies at a rate of about 10^{-4} to 10^{-3} per cell division [14,131]. A small number of antigenic variants dominate the initial parasitemia of this blood-borne pathogen.The host then clears these initial variants with antibodies. Some of the bacteria from this

first parasitemia will have changed antigenic type. Those switches provide new variants that cause a second parasitemia, which is eventually recognized by the host and cleared. The cycle repeats several times, causing relapsing fever.

The protozoan *T. brucei* has hundreds of alternative loci that encode the dominant surface glycoprotein [16,105]. Typically, each cell expresses only one of the alternative loci. Switches in expression occur at a rate of up to 10^{-2} per cell division [138]. The switch mechanism is similar to that in *Borrelia hermsi* – gene conversion of archival copies into a transcriptionally active expression site. *T. brucei* has approximately 20 alternative transcription sites, of which only one is usually active. Thus, this parasite can also change expression by switching between transcription sites. It is not fully understood how different transcription sites are regulated.

15.3.2.4 Transcriptional silencing Changing transcriptional activation between different sites appears to be the mechanism by which *P. falciparum* regulates expression of its major surface antigen. *P. falciparum* expresses the *var* gene within erythrocytes. The gene product, PfEMP1, moves to the surface of the host cell, where it influences cellular adhesion and recognition by host immunity [37]. The *var* genes are highly diverse antigenically [133]. Each parasite exports only one PfEMP1 type to the erythrocyte surface, but a clone of parasites switches between PfEMP1 types [121]. Switching leads to a diverse population of PfEMP1 variants within a host and even wider diversity among hosts.It appears that many *var* loci are transcribed during the first few hours after erythrocyte infection, but only a single *var* gene transcript is active when PfEMP1 is translated and moved to the erythrocyte surface [30,118]. It may be that some mechanism shuts down expression of all but one locus without modifying the DNA sequence. Expression may be influenced by an interaction between an intron and the promoter, but the details need to be worked out [25,35].

15.3.3 Intragenomic Recombination

New variants of alternative genes in archival libraries may be created by recombination. For example, Rich et al. [110] found evidence for recombination between the archived loci of the variable short protein (Vsp) of *B. hermsii*.They studied the DNA sequences of 11 *vsp* loci within a single clone. These *vsp* loci are silent, archival copies that can, by gene conversion, be copied into the single expression site. The genes differ by 30–40% in amino acid sequence, providing sufficient diversity to reduce or eliminate antigenic cross-reactivity within the host.Rich et al. [110] used statistical analyses of *vsp* sequences to infer that past recombination events have occurred between archival loci. Those analyses focus on attributes such as runs of similar nucleotides between loci that occur more often than would be likely if alleles diverged only by accumulating mutations within each locus. Shared runs can be introduced into diverged loci by recombination.The archival antigenic repertoire of *T. brucei* evolves rapidly [105]. This species has a large archival library and multiple expression sites, but only one expression site is active at any time. New genes can be created within an active expression site when several donor sequences convert the site in a mosaic pattern [10,103]. When an active expression site becomes inactivated, the gene within that site probably becomes protected from further gene conversion events [102,104]. Thus, newly created genes by mosaic conversion become stored in the repertoire. Perhaps new genes in silent expression sites can move into more permanent archival locations by recombination, but this has not yet been observed. Recombination between silent, archived copies may also occur, which, although each event may be relatively rare, could strongly affect the evolutionary rate of the archived repertoire.

15.3.4 Mixing Between Genomes

New antigenic variants can be produced by mixing genes between distinct lineages. This happens in three ways.

Segregation brings together chromosomes from different lineages. Reassortment of influenza A's neuraminidase and hemagglutinin surface antigens provides the most famous example [70].The genes for these antigens occur on two separate RNA segments of the genome – the genome has a total of eight segments. When two or more viruses infect a single cell, the parental segments all replicate separately and then are packaged together into new viral particles. This process can package the segments from different parents into a new virus.New neuraminidase–hemagglutinin combinations present novel antigenic properties to the host. Rare segregation events have introduced hemagglutinin from bird influenza into the genome of human influenza [143]. The novel hemagglutinins cross-reacted very little with those circulating in humans, allowing the new combinations to sweep through human populations and cause pandemics.

Intergenomic recombination occurs when chromosomes from different lineages exchange pieces of their nucleotide sequence. In protozoan parasites such as *Plasmodium* and certain *Trypanosoma* species (e.g., *brucei*), recombination happens as part of a typical Mendelian cycle of outcrossing sex [33,64]. Recombination can occur in viruses when two or more particles infect a single cell. DNA viruses may recombine relatively frequently because they can use the host's recombination enzymes [132]. RNA viruses may recombine less often because the host lacks specific enzymes to mediate reciprocal exchange of RNA segments. However, many descriptions of RNA virus recombination have been reported [72,112]. In all cases, even rare recombination can provide an important source for new antigenic variants.*Horizontal transfer* of DNA between bacteria introduces new nucleotide sequences into a lineage [97]. Transformation occurs when a cell takes up naked DNA from the environment. Some species transform at a particularly high rate, suggesting that they have specific adaptations for uptake and incorporation of foreign DNA [48]. For example, *Neisseria* species transform frequently enough to have many apparently mosaic genes

from interspecies transfers [48,127,148], and *N. gonorrhoeae* has low linkage disequilibrium across its genome [86]. Horizontal transfer also occurs when bacteriophage viruses carry DNA from one host cell to another or when two cells conjugate to transfer DNA from a donor to a recipient [97].

15.4 INTERACTIONS WITH HOST IMMUNITY

Specific immunity favors parasites that change their antigens and escape recognition. In this section, I summarize examples of parasite escape and the consequences for antigenic diversity within hosts.

15.4.1 Natural Selection of Antigenic Variants

In several pathogens, a changing profile of antigenic variants characterizes the course of infection within a single host. Natural selection favors variants that escape immune recognition, although escape is often temporary. Selection may also favor diversification of the pathogens for the ability to attack different types of host cells. I briefly summarize a few examples.

15.4.1.1 Simian immunodeficiency virus (SIV) and HIV Soudeyns et al. [126] identified the regions of the HIV-1 envelope under strong selective pressure by analyzing the pattern of nucleotide changes in the population. They compared the rate of non-synonymous d_n nucleotide replacements that cause an amino acid change versus the rate of synonymous d_s nucleotide replacements that do not cause an amino acid change. A high d_n/d_s ratio suggests positive natural selection favoring amino acid change; a low d_n/d_s ratio suggests negative natural selection opposing change in amino acids [100]. Soudeyns et al. [126] found that regions of the envelope gene under strong positive selection corresponded to epitopes recognized by CTLs. The non-synonymous substitutions in these epitopes typically abolished recognition by a matching CTL clone. The population of viruses accumulated diversity in the dominant epitopes over the course of infection within hosts. These results suggest that CTL attack based on specific recognition drives the rapid rate of amino acid replacements in these epitopes.

Kimata et al. [67] studied properties of SIV isolated from early and late stages of infection within individual hosts. The early viruses infected macrophages, replicated slowly, and the viral particles were susceptible to antibody-mediated clearance. The late viruses infected T cells, replicated more than 1000 times faster than early viruses, and were less sensitive to antibody-mediated clearance. Kimata et al. [67] did not determine the viral amino acid changes that altered cell tropism of SIV. Connor et al. [32] found that changes in the host-cell coreceptors used by early and late HIV-1 correlated with changes in cell tropism, but it is not yet clear which changes are essential for the virus's tropic specificity. Connor et al. [32] did show that the population of early viruses used a narrow range of coreceptors, whereas the late viruses were highly polymorphic for a diverse range of host coreceptors. Clearly, the virus is evolving to use various cell types.

The relative insensitivity of late SIV to antibody apparently depended on increased glycosylation of the envelope proteins. The late viruses with increased glycosylation were not recognized by antibodies that neutralized the early viruses. Viruses that escape antibody recognition gain significant advantage during the course of infection [27,116]. Kimata et al. [67] showed that, when injected into a naive host, the late SIV did not stimulate as much neutralizing antibody as did the early SIV. Additional glycosylation apparently reduces the ability of antibodies to form against the viral surface. Presumably, the glycosylation also hinders the ability of the virus to initiate infection; otherwise both early and late viruses would have enhanced glycosylation. Both the early, macrophage-tropic SIV and the late, T cell-tropic SIV used the host coreceptor CCR5 [67]. That observation contrasts with a study of early and late HIV-1 isolated from individual hosts, in which Connor et al. [32] found that early, macrophage-tropic viruses depended primarily on the CCR5 coreceptors, whereas the population of late viruses had expanded coreceptor use to include CCR5, CCR3, CCR2b, and CXCR4.

Many other studies focus on HIV diversification within hosts (e.g. [5,49,117]).

15.4.1.2 Hepatitis C virus (HCV) Farci et al. [41] obtained HCV samples at various stages of infection within individual hosts. They sequenced the envelope genes from these samples to determine the pattern of evolution within hosts. They then compared the evolutionary pattern with the clinical outcome of infection, which follows one of three courses: clearance in about 15% of cases; chronic infection and either slowly or rapidly progressive disease in about 85% of cases; and severe, fulminant hepatitis in rare cases.

Farci et al. [41] sampled three major periods of infection: the incubation period soon after infection; during the buildup of viremia but before significant expression of specific antibodies; and after the host's buildup of specific antibodies. The sequence diversity within hosts identified two distinct regions of the envelope genes. The hypervariable region evolved quickly and appeared to be under positive selection from the host immune system, whereas other regions of the envelope genes had relatively little genetic variation and did not evolve rapidly under any circumstances. Thus, the following comparisons focus only on the hypervariable region.

Those hosts that eventually cleared the virus had similar or higher rates of viral diversification before antibodies appeared than did those patients that developed chronic infection. By contrast, after antibodies appeared, chronic infection was correlated with significantly higher viral diversity and rates of evolution than occurred when the infection was eventually cleared. It appears that hosts who cleared the infection could contain viral diversity and eventually eliminate all variants, whereas those that progressed to chronic infection could not

control viral diversification. The rare and highly virulent fulminant pattern had low viral diversity and rates of evolution. This lack of diversity suggests either that the fulminant form may be associated with a single viral lineage that has a strong virulence determinant or that some hosts failed to mount an effective immune response.

15.4.1.3 Generality of within-host evolution of antigens

HIV and HCV share several characters that make them particularly likely to evolve within hosts. They are RNA viruses, which have relatively high mutation rates, relatively simple genomes, simple life cycles, potentially high replication rates, and potentially high population sizes within hosts. HIV and HCV also typically develop persistent infections with long residence times in each host. If the mutation rate per nucleotide per replication is 10^{-5} and the population of viruses is of the order of 10^{10} within a host, then there are 10^5 point mutations at every site in every generation. For every pair of sites, there will usually be at least one virus that carries mutations at both sites. Thus, there is a tremendous influx of mutational variation. Other RNA viruses such as influenza also have high mutation rates and potentially large populations within hosts, but the hosts typically clear infections within 2 weeks. Some within-host evolution very likely occurs, but it does not play a significant role in the infection dynamics within hosts. DNA-based pathogens produce much less mutational variation per replication. But large population sizes, long infection times, and hypermutation of epitopes could still lead to significant evolution within hosts. At present, the persistent RNA infections have been studied most intensively because of their obvious potential for rapid evolutionary change. As more data accumulate, it will be interesting to compare the extent and the rate of within-host evolutionary change in various pathogens.

15.4.2 Pathogen Manipulation of Host Immune Dynamics

Pathogens use several strategies to interfere with host immunity. A parasite's exposed surface antigens or candidate CTL epitopes may lack variation because the parasite can repel immune attack. I do not know of any evidence to support this idea, but it should be considered when studying candidate epitopes and their observed level of antigenic variation.

Several reviews summarize viral methods for reducing host immunity (e.g. [4,128]). Some bacteria also interfere with immune regulation [114]. I list just a few viral examples, taken from the outline given by Tortorella et al. [137].

Some viruses interfere with MHC presentation of antigens. Cases occur in which viruses reduce MHC function at the level of transcription, protein synthesis, degradation, transport to the cell surface, and maintenance at the cell surface. The host's natural killer (NK) cells attack other host cells that fail to present MHC class I molecules on their surface. Viruses that interfere with normal class I expression use various methods to prevent NK attack, for example, viral expression of an MHC class I homolog that interferes with NK activation.

Host cells often use programed suicide (apoptosis) to control infection. Various viruses interfere with different steps in the apoptosis control pathway.

The host uses cytokines to regulate many immune functions. Some viruses alter expression of host cytokines or express their own copies of cytokines. Other viruses express receptors for cytokines or for the constant (Fc) portion of antibodies. These viral receptors reduce concentrations of freely circulating host molecules or transmit signals that alter the regulation of host defense.

15.4.3 Sequence of Variants in Active Switching from Archives

Some parasites store alternative genes for antigenic surface molecules. Each individual parasite usually expresses only one of the alternatives [36,47]. Parasite lineages change expression from one stored gene to another at a low rate. In *T. brucei*, the switch rate is about 10^{-3} or 10^{-2} per cell division [138]. Antigenic switches affect the dynamics of the parasite population within the host. For example, the blood-borne bacterial spirochete *B. hermsii* causes a sequence of relapsing fevers [11,12]. Each relapse and recovery follows from a spike in bacterial density. The bacteria rise in abundance when new antigenic variants escape immune recognition and fall in abundance when the host generates a specific antibody response to clear the dominant variants. Switches between types within a cellular lineage occur stochastically. But the sequence of variants that dominate sequential waves of parasitemia tends to follow a repeatable order in *T. brucei* [15,50] and probably in *Borrelia* [14]. Temporal separation in the rise of different antigenic variants allows trypanosomes to continue an infection for a longer period of time [141]. If all variants rose in abundance early in the infection, they would all stimulate specific immune responses and be cleared, ending the infection. If the rise in different variants can be spread over time, then the infection can be prolonged. The puzzle is how stochastic changes in the surface antigens of individual parasites can lead to an ordered temporal pattern at the level of the population of parasites within the host [3,17,44,139,140]. Five hypotheses have been developed, none of which has strong empirical support at present. I briefly describe each idea.

First, the antigenic variants may differ in growth rate. Those that divide more quickly could dominate the early phases of infection, and those that divide more slowly could increase and be cleared later in the infection [119]. Computer studies and mathematical models show that variable growth rates alone cannot easily explain wide separation in the times of appearance of different variants [3,69]. Only with a very large spread in growth rates would the slowest variant be able to avoid an immune response long enough to develop an extended duration of total infection. Aslam and Turner [8] measured the growth rates of different variants of *T. brucei* and found little difference between the variants.

Second, parasite cells may temporarily express both the old and new antigens in the transition period after a molecular switch in antigenic type [3]. The double expressers could experience varying immune pressure depending on the time for complete antigenic replacement or aspects of cross-reactivity. This would favor some transitions to occur more easily than others, leading to temporal separation in the order of appearance for different antigenic variants. This model is rather complex and has gained little empirical or popular support, as discussed in several papers [2,18,19,21,90].

Third, the switch probabilities between antigenic variants may be structured in a way to provide sequential dominance and extended infection [44]. If the transition probabilities from each variant to the other variants are chosen randomly, then an extended sequence of expression does not develop because the transition pathways are too highly connected. The first antigenic types would generate several variants that develop a second parasitemia. Those second-order variants would generate nearly all other variants in a random switch matrix. The variants may arise in an extended sequence if the parasite structures the transition probabilities into separate sets of variants, with only rare transitions between sets. The first set of variants switches to a limited second set of variants, the second set connects to a limited third set, and so on. Longer infections enhance the probability of transmission to other hosts. Thus, natural selection favors the parasites to structure their switch probabilities in a hierarchical way in order to extend the length of infection. Paget-McNicol et al. [101] also developed a model in which switch rates vary, but did not consider how natural selection might modulate switch rates.

Fourth, Recker et al. [108] noted that hosts with stronger cross-reactive immune responses against *P. falciparum* variants are more likely to sustain chronic infections. Presumably, chronic infections mean that the parasite's repertoire of antigenic surface molecules can be structured into a pattern of sequential dominance. Based on these points, Recker et al. developed a model in which host immunity develops against two distinct components of the variable surface antigens. One part of the immune response develops lasting immunity against a unique component of each antigen. Another part of the host response develops short-lived immunity against a component of the antigenic molecule that is shared by other antigenic types. With these points in mind, imagine how a malarial infection would play out. One or a few antigenic variants dominate the initial parasitemia. The host develops specific immunity against each variant. One part of the immune response is specific for each variant and long-lived, clearing each variant and preventing another dominant wave of parasitemia by that variant. Another part of the immune response against a particular variant cross-reacts with many other variants – this cross-reactive component lasts only for a short while. As the initial parasitemia develops, some cells will have switched expression to other antigenic surface variants. As the first parasitemia clears, the next wave of parasitemia will develop from those rare variants that are least affected by the cross-reactive part of the host immune response. As those favored types develop into strong parasitemia, the process repeats, favoring in the subsequent wave those variants that cross-react least with the previous wave. Molineaux et al. [88] developed a more complex model of *P. falciparum* parasitemia dynamics and host immunity. Their model includes fitted values for how the various components of immunity clear parasites and variation in growth rate of different variants. This is an interesting analysis, but with so many parameters, it is difficult to determine whether the good fit with data arises from so many degrees of freedom or from a model that properly highlights the essential features of antigenic variation.

Turner [139] proposed a fifth explanation for high switch rates and ordered expression of variants. The parasite faces a trade-off between two requirements. On the one hand, competition between parasite genotypes favors high rates of switching and stochastic expression of multiple variants early in an infection. On the other hand, lower effective rates of switching later in an infection express variants sequentially and extend the total length of infection. Many *T. brucei* infections in the field probably begin with inoculation by multiple parasite genotypes transmitted by a single tsetse fly vector [77]. This creates competition between the multiple genotypes. According to Turner [139], competition intensifies the selective pressure on parasites to express many variants – variation allows escape from specific immunity by prior infections and helps to avoid cross-reactivity between variants expressed by different genotypes. These factors favor high rates of stochastic switching. The effective rate of switching drops as the infection progresses because the host develops immunity to many variants. Effective switches occur when they produce novel variants, and the rate at which novel variants arise declines over the course of infection. Those novel variants, when they do occur, can produce new waves of parasitemia, promoting parasite transmission.

Turner's idea brings out many interesting issues, particularly the role of competition between genotypes within a host. But his verbal model is not fully specified. For example, delayed expression of some variants and extended infection depend on the connectivity of transition pathways between variants, an issue he does not discuss. The problem calls for mathematical analysis coupled with empirical study. Connectivity of transition pathways between variants plays an important role in most theories. In Agur et al.'s [3] model, host immunity acting differentially on double expressers during the switch process favors some transitions over others. In Frank's [44] model, the different rates of molecular switching between variants provides structure to transition pathways. In Recker et al.'s [108] model, short-lived and cross-reactive host immunity favors particulars sequences of antigenic dominance. Turner's [139] model is not fully specified, but to work it must also provide a tendency for some transitions to be favored over others – this may occur by chance with random and rare switching or perhaps may favor common switches early and rare switches later in the sequence, more or less as in Frank's [44] model.

Connectivity of transition pathways has not been studied empirically. Frank and Barbour [46] have recently discussed this issue based on reanalysis of earlier data from *B. hermsii*.

15.5 EXPERIMENTAL EVOLUTION

Experimental evolution manipulates the environment of a population and observes the resulting pattern of evolutionary change. This allows one to study the selective forces that shape antigenic diversity. For example, one could manipulate immune selection by exposing parasites to different regimes of monoclonal antibodies. The parasites' evolutionary response reveals the adaptive potential and the constraints that shape patterns of antigenic variation.

In this chapter, I describe experimental evolution studies of foot-and-mouth disease virus (FMDV). I also use this virus as a case study to show how different methods combine to provide a deeper understanding of antigenic variation. These approaches include structural analysis of the virion, functional analysis of epitopes with regard to binding cellular receptors, sequence analysis of natural isolates, and experimental analysis of evolving populations.

15.5.1 Antigenicity and Structure of FMDV

FMDV is an RNA virus that frequently causes disease in domesticated cattle, swine, sheep, and goats [107]. FMDV populations maintain antigenic diversity in several rapidly evolving epitopes [42,85].

The most important epitopes occur on the GH loop of the VP1 surface protein [82,84,125]. This loop has about 20 amino acids that contribute to several overlapping epitopes. These antibody-binding sites appear to be determined mostly by the amino acids in the GH peptide (a continuous epitope). Antibodies that bind to an isolated GH peptide also neutralize intact viruses.

Many antibody escape variants occur in the GH loop, leading to extensive genetic variation in this region. However, a conserved amino acid triplet, Arg-Gly-Asp (RGD), also binds to antibodies. This conserved triplet mediates binding to integrin host-cell receptors typically used in FMDV attachment and entry [20,92,125]. The GH loop of VP1 contains continuous epitopes that together define the hypervariable antigenic site A common to all serotypes. Discontinuous epitopes occur when amino acid residues from widely separated sequence locations come together conformationally to form a binding surface for antibodies. Two antigenic sites of serotypes A, O, and C have discontinuous epitopes that have received widespread attention [42,84].

15.5.2 Antibody Escape Mutants

Many antibody escape mutants have been sequenced (references in [80]). One can develop a map of natural escape variants by comparing changes in sequence with differences in binding affinity to a panel of monoclonal antibodies (Mabs).

Two problems of interpreting selective pressures arise from an escape map based on natural variants. First, field isolates do not control the multitude of evolutionary pressures on variation. Mutants may spread either in direct response to antibody pressure, in response to other selective pressures, or by stochastic fluctuations independent of selective forces. Lack of variability may result either from lack of antibody pressure or from constraining selective pressures such as binding to host receptors.

The second problem for interpreting selective pressures from natural isolates concerns lack of control over genetic background. Whether a particular amino acid site affects antibody affinity may depend on conformation-changing variants at other sites.

Site-directed mutagenesis controls amino acid replacements in a fixed genetic background. One can alter sites that do not vary naturally to test for effects on antibody binding. Site-directed mutagenesis has provided useful information for FMDV [83]. But this method can only define changes in antibody binding; it does not show how viral populations actually respond to immune pressure. Several studies have applied monoclonal or polyclonal antibodies to FMDV in laboratory culture [82,125]. This allows direct control of selective pressure by comparing lines with and without exposure to antibodies. In addition, cultures can be started with genetically monomorphic viruses to control genetic background.

Martinez et al. [80] began laboratory evolution studies from a single viral clone of serotype C. These viruses were grown on baby hamster kidney cells (BHK-21). All host cells were derived from a single precursor cell. Two separate viral lines were established. C-S8c1 developed through three successive plaque isolations. C-S8c1p100 began with C-S8c1 and developed through 100 serial passages on a monolayer of BHK-21 cells. The host cells were refreshed from independent stock in each passage and therefore did not coevolve with the virus over the passage history.

In natural isolates, extensive sequence variability in the GH loop of VP1 correlates with escape from antibody neutralization. However, the Arg-Gly-Asp (RGD) sequence near the center of this GH loop is invariant in field isolates [125]. Controlled studies of laboratory evolution provide some insight into the evolution of this region.

The monoclonal antibody SD6 binds to an epitope spanning residues 136–147 in the GH loop of VP1. Martinez et al. [80] applied selective pressure by SD6 after establishment of the separate viral lines C-S8c1 and C-S8c1p100 by growing a cloned (genetically monomorphic) isolate in the presence of the antibody and sampling escape mutants. Nucleotide sequences of escape mutants were obtained. Each mutant (except one) escaped antibody neutralization by a single amino acid change. The different locations of these mutations in the original (C-S8c1) line compared with the serially passaged (C-S8c1p100) line provide the most striking result of this study. The original line conserved the Arg-Gly-Asp (RGD) motif at positions 141–143. By contrast, the serially passaged line had

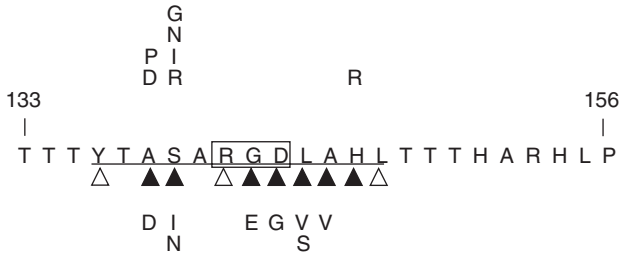

Fig. 15.1. Amino acid sequence in the central region of the VP1 GH loop of FMDV. The start and stop numbers label amino acid positions. The box shows the RGD motif at positions 141–143. The monoclonal antibody SD6 recognizes the continuous epitope defined by the underlined positions. Black triangles show positions at which most replacement amino acids greatly reduce binding by SD6; in other words, a single replacement at any of these sites creates an escape mutant. The white triangles denote positions that can tolerate certain amino acid replacements without greatly affecting antibody binding. Unmarked positions in the epitope can vary without much change in binding. The letters above the sequence summarize the escape mutants of C-S8c1 (original line); letters below the sequence summarize escape mutants of C-S8c1p100 (passaged line). Letters denote amino acids according to the standard single-letter code. Redrawn from [80].

numerous mutations within the RGD motif. Figure 15.1 contrasts the location of mutants for the two lines.

Variants in the RGD motif had not previously been observed in spite of neutralizing antibodies' affinity for this region. The RGD motif was thought to be invariant because of its essential role in binding to the host cell. Yet, the serially passaged line accumulated variants in this region. Those variants replicated with the same kinetics as the parental viruses of C-S8c1p100, with no loss in fitness. Baranowski et al. [9] showed that lineages with an altered RGD motif use an alternative pathway of attachment and entry to host cells.

Martinez et al. [80] sequenced the capsid genes from the original line, the serially passaged line, and an escape mutant of the serially passaged line. The escape mutant from the serially passaged line differed from the parental virus of this line only at a single site in the RGD region. Tolerance to replacements in the RGD region must follow from the differences accumulated by C-S8c1p100 during serial passage. Six amino acids differed between the original and serially passaged lines. Apparently, those substitutions changed cell tropism properties of the virus and allowed variation in the previously invariant RGD motif.

15.5.3 Cell Binding and Tropism

Attachment and entry to host cells impose strong natural selection on some regions of the viral surface. Experimental evolution provides one approach to analyze those selective forces, as described in the previous section. In this section, I briefly summarize further studies of amino acid variation in the FMDV capsid and the consequences for attachment and entry to host cells. Jackson et al. [62] compared the affinity of different viral genotypes for two integrin receptors, $\alpha_v\beta_3$ and $\alpha_5\beta_1$. The standard RGD motif was required for both receptors. The following amino acid at the RGD+1 position influenced relative affinity for the two integrin types. For $\alpha_v\beta_3$, several different amino acids at RGD+1 allowed binding, consistent with this receptor's multifunctional role in binding several ligands. By contrast, $\alpha_5\beta_1$ has narrower specificity, favoring a leucine at RGD+1. Jackson et al. [62] compared two viruses that differed only at RGD+1, the first with an arginine and the second with a leucine. The first virus had relatively higher affinity for $\alpha_v\beta_3$ compared with $\alpha_5\beta_1$. By contrast, the second virus had relatively higher affinity for $\alpha_5\beta_1$ compared with $\alpha_v\beta_3$. For at least some antibodies that recognize RGDL, loss of leucine at RGD+1 abolishes recognition (see Fig. 15.1) [80]. Thirty type O and eight type A field isolates had leucine at RGD+1. By contrast, five SAT-2 isolates had arginine, two Asia-1 isolates had methionine, and one Asia-1 isolate had leucine [62]. These and other data suggest that most serotypes have leucine at RGD+1 and perhaps a higher affinity for $\alpha_5\beta_1$. SAT-2 may either have greater affinity for $\alpha_v\beta_3$ or its binding may be conditioned by amino acid variants at other sites.

In another study, Jackson et al. [63] analyzed FMDV binding to a different integrin, $\alpha_v\beta_6$. This integrin binds relatively few host ligands and depends on an RGDLXXL motif with leucines at RGD+1 and RGD+4. Most FMDV isolates have leucines at those two positions. $\alpha_v\beta_3$ does not have stringent requirements at those sites, suggesting that $\alpha_v\beta_6$ may be an important natural receptor. Overall, RGDLXXL binds to the widest array of integrins, at least over those studied so far, although relative affinities for different integrins may be modulated by substitutions at RGD+1 and perhaps other sites. It would be interesting to sample isolates from various host tissues that differ in the densities of the various integrin receptors and analyze whether any substitutions appear relative to isolates in other body compartments of the same host.

Viral success in different cell types or in different hosts may depend on variations in nonstructural genes that do not mediate binding and entry to host cells. For example, Nunez et al. [95] serially passaged FMDV in guinea pigs. FMDV does not normally cause lesions in guinea pigs, but after serial passage, viral variants arose that caused disease. Among the several amino acid substitutions that arose during passage, a single change from glutamine to arginine at position 44 of gene 3A provided virulence. The function of 3A in FMDV is not known. In poliovirus, a distantly related picornavirus, 3A plays a role in virus-specific RNA synthesis. These studies show the potential power of experimental evolution in studying evolutionary forces, particularly when combined with analysis of naturally occurring variation.

15.5.4 Fitness Consequences of Substitutions

Antibody escape mutants are typically isolated in one of two ways. First, pathogens may be grown *in vitro* with antibodies. This creates selective pressure for substitutions that escape antibody recognition. Second, naturally occurring variants from field isolates may be tested against a panel of antibodies. Certain sets of antibodies may bind most isolates, allowing identification of those variants that differ at commonly recognized epitopes.

Escape variants gain a fitness advantage by avoiding antibody recognition targeted to important epitopes. However, those pathogen epitopes may also play a role in binding to host cells, in release from infected cells, or in some other aspect of the pathogen's life cycle. Functional and structural studies of amino acid substitutions provide one method of analysis. That approach has the advantage of directly assessing the mechanisms by which amino acid variants affect multiple components of parasite fitness, such as escape from antibody recognition and altered host attachment characteristics. Although functional and structural approaches can directly measure binding differences caused by amino acid substitutions in different genetic backgrounds, they cannot provide a good measure of all the fitness consequences associated with changes in genotype.

Carrillo et al. [26] used an alternative approach to analyze the consequences of amino acid substitutions. They studied the relative fitnesses *in vivo* of a parental FMDV genotype and three mutant genotypes derived from the parental type. They measured relative fitness by competing pairs of strains within live pigs. The parental type, C-S8c1, came from a C serotype isolated from a pig. The first monoclonal antibody-resistant mutant, MARM21, arose in a pig infected with C-S8c1. MARM21 differs from C-S8c1 by a single change from serine to arginine at VP1 139 (Fig. 15.1), providing escape from the monoclonal antibody SD6. The second mutant, S-3T_1, came from a blood sample of a pig 1 day after experimental inoculation with C-S8c1. That isolate had a single change from threonine to alanine at VP1 135 (Fig. 15.1). Only one of 58 monoclonal antibodies differentiated between the parental type and S-3T_1, and the difference in affinity was small. This supports the claim in Figure 15.1 that position 135 is not strongly antigenic. The third mutant, C-S15c1, derived from a field variant of type C1 isolated from a pig. This mutant type had eight amino acid differences in VP1 compared with C-S8c1. C-S15c1 did not react with monoclonal antibody SD6.

One of the three mutants was coinoculated with the parental type into each experimental pig. Two replicate pigs were used for each of the three pairs of mutant and parental types. Fever rose 1 day after infection and peaked 2 or 3 days post infection. All animals developed vesicular lesions 2–4 days post infection. For each animal, between two and seven samples were taken from lesions, and the relative proportions of the competing viruses were assayed by reactivity to monoclonal antibodies. The small sample sizes do not allow strong conclusions to be drawn. Rather, the following two results hint at what might be learned from more extensive studies of this sort. First, the parental type strongly dominated MARM21 in all seven lesions sampled from the two experimental animals, comprising between 80% and 94% of the viruses in each lesion. The MARM21 mutation appears to confer lower fitness *in vivo*, at least in the two animals tested. The lower fitness may arise because the mutant was cleared more effectively by antibodies, bound less efficiently to host cells, or had reduced performance in some other fitness component. Second, S-3T_1 abundance relative to the parental type varied strongly between lesions. In the two lesions analyzed from one animal, the parental type comprised 67 $\pm$ 3.4% and 3.2 $\pm$ 1.5% (mean $\pm$ standard deviation). In the other animal, the three lesions analyzed had parental-type percentages of 75 $\pm$ 4.1%%, 25 $\pm$ 2.8%, and 5.9 $\pm$ 1.2%. Differences in dominance between lesions also occurred between C-S15c1 and the parental type. Variations in dominance may arise from stochastic sampling of viruses that form lesions, from differences in tissue tropism, or from some other cause. Further studies of this sort may provide a more refined understanding of the multiple fitness consequences that follow from particular amino acid changes, their interactions with the genetic background of the virus, the role of different host genotypes, and the effect of prior exposure of hosts to different antigenic variants.

15.6 MEASURING SELECTION WITH POPULATION SAMPLES

Experimental evolution provides insight into kinetic and mechanistic aspects of parasite escape from host immunity. Such experimental studies clarify selective forces that influence change at certain amino acid sites. But experimental studies provide only a hint of what actually occurs in natural populations, in which selective pressures and evolutionary dynamics differ significantly from those in controlled laboratory studies. It is important to combine experimental insights with analyses of variation in natural populations. In this section, I discuss how population samples of nucleotide sequences provide information about natural selection of antigenic variation. I focus on themes directly related to the goal of this chapter – the synthesis between different kinds of biological analyses. In particular, I show how analysis of population samples complements studies of molecular structure and experimental evolution. Several books and articles review the methods to analyze population samples and the many different types of applications [23,34,56,60,68,76,91,93,94,98–100,113,147].

15.6.1 Positive and Negative Selection

The genetic code maps three sequential nucleotides (a codon) to a single amino acid or to a stop signal. The four different nucleotides combine to make $4^3 = 64$ different

codons. The 64 codons specify 20 different amino acids plus a stop signal, leading to an average of $64/21 ≈ 3 different codons for each amino acid or stop signal. This degenerate aspect of the code means that some nucleotide substitutions do not change the encoded amino acid or stop signal. Nucleotide substitutions that do not cause an amino acid change are called *synonymous*; those that do change the encoded amino acid are called *non-synonymous*. Synonymous substitutions do not affect the amino acid sequence and therefore should not be affected by natural selection of phenotype. By contrast, non-synonymous substitutions can be affected by selection because they do change the encoded protein. If there is no selection on proteins, then the same forces of mutation and random sampling influence all nucleotide changes, causing the rate of non-synonymous substitutions, d_n, to equal the rate of synonymous substitutions, d_s [76,93,100].

When natural selection favors change in amino acids, the non-synonymous substitution rate d_n rises. Thus, $d_n > d_s$ measured in a sample of sequences implies that natural selection has favored evolutionary change. This contribution of selection to the rate of amino acid change above the background measured by d_s is called *positive selection*. Parasite epitopes often show signs of positive selection as they change to escape recognition by host immunity [147].

By contrast, negative selection removes amino acid changes, preserving the amino acid sequence against the spread of mutations. Negative selection reduces the non-synonymous substitution rate, causing $d_n < d_s$. The great majority of sequences show negative selection, suggesting that most amino acid replacements are deleterious and are removed by natural selection. In cases where positive selection does occur, the non-synonymous replacements often cluster on protein surfaces involved in some sort of specific recognition. In these positively selected proteins, amino acid sites structurally hidden from external recognition often show the typical signs of negative selection.

15.6.2 Positive Selection to Avoid Host Recognition

Many examples of positive selection come from genes involved in host–parasite recognition [40,60,147]. These sequence analyses provide information about how selection has shaped the structure and function of proteins. For example, one may combine analysis of positive selection with structural data to determine which sites are exposed to antibody pressure. In the absence of structural data, sequences can be used to predict which sites are structurally exposed and can change and which sites are either not exposed or functionally constrained. I briefly summarize one example.

The tick-borne protozoan *Theileria annulata* causes disease in cattle [52]. The surface antigen Tams1 induces a strong antibody response and has been considered a candidate for developing a vaccine. However, Tams1 varies antigenically; thus studies have focused on the molecular nature of the variability to gain further insight. The structure and function of Tams1 have not been determined. Recently, Gubbels et al. [52] analyzed a population sample of nucleotide sequences to predict which domains of Tams1 change in response to host immunity and which domains do not vary because of structural or functional constraints. They found seven domains with elevated rates of non-synonymous substitutions compared with synonymous substitutions (Fig. 15.2), suggesting that these regions may be exposed to antibody pressure. Some domains had relatively little non-synonymous change, indicating that structural or functional constraints preserve amino acid sequence. These inferences provide guidance in vaccine design and point to testable hypotheses about antigenicity and structure.

15.6.3 Phylogenetic Analysis of Nucleotide Substitutions

Initial studies of selection often used small numbers of sequences, typically fewer than 100. Small sample sizes required aggregating observations across all nucleotide sites to gain sufficient statistical power. Conclusions focused on whether selection was positive, negative, or neutral when averaged over all sites. With slightly larger samples, one could do a sliding window analysis as in Figure 15.2 to infer the kind of selection averaged over sets of amino acids that occur contiguously in the two-dimensional sequence [40]. Major changes in binding and antigenicity often require only one or a few amino acid changes [45]. The analytical methods that

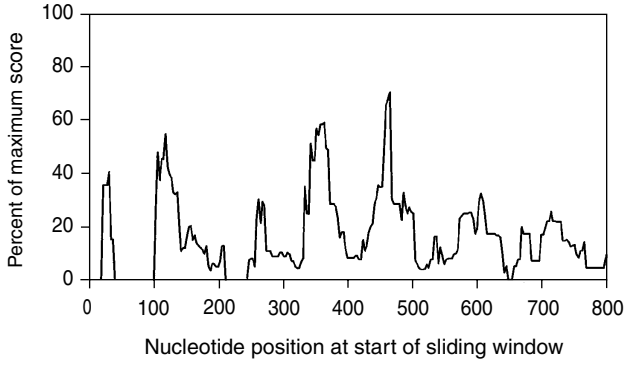

Fig. 15.2. The seven peaks identify the major regions of positive selection in the Tams1 protein. The 18 sequences analyzed in this figure have about 870 nucleotides. The analysis focused on a sliding window [40] of 60 nucleotides (20 amino acids). For each window shown on the *x*-axis, the numbers of non-synonymous and synonymous nucleotide substitutions were calculated by comparing the 18 sequences. The *y*-axis shows the strength of positive selection measured as follows. For each window of 60 nucleotides, each pair of sequences was compared. Each paired comparison was scored for the statistical significance of positive selection based on the numbers of non-synonymous and synonymous changes between the pair, with a score of zero for nonsignificant, a score of 1 for significant, and a score of 2 for highly significant. The maximum score is twice the number of comparisons; the actual score is the sum of significance values for each comparison; and the percentage of the maximum is the actual divided by the maximum multiplied by 100. From [52].

aggregate over whole sequences or sliding windows often fail to detect selection at the scale of single-site substitutions, which appears to be the proper scale for understanding antigenic evolution. Recently, larger samples of sequences have provided the opportunity to study the rates of synonymous and non-synonymous substitutions at individual nucleotide sites. Each individual substitution occurs within a lineal history of descent, that is, a change occurs between parent and offspring. To study each substitution directly, one must first arrange a sample of sequences into lineal relationships by building a phylogenetic tree. From the tree, one can infer the nucleotide sequence of ancestors, and therefore trace the history of each nucleotide change through time. Each nucleotide change can be classified as synonymous or non-synonymous. For each amino acid site, one can sum up the numbers of synonymous and non-synonymous nucleotide changes across the entire phylogeny and derive the associated rates of change. With appropriate statistics, one determines for each amino acid site whether non-synonymous changes occur significantly more or less often than synonymous changes [24,53,87,134,142,147]. The concepts of measuring positive and negative selection remain the same. However, for the first time, the statistical power has been raised to the point where analysis of population samples provides significant insight into the evolution of antigens. The power derives from studying the relative success of alternate amino acids at a single site. Important selective forces include the amino acids at other sites as well as binding properties to host immune molecules and other host receptors.

Haydon et al. [54] analyzed selection on individual amino acid sites of FMDV. Most sites showed mild to strong negative selection, as usually occurs. At 17 sites, they found evidence of significant positive selection. Twelve of these positively selected sites occurred at positions that had previously been observed to develop escape mutants in experimental evolution studies that imposed pressure by monoclonal antibodies. The other five sites indicate candidates for further experimental analysis.

Haydon et al.'s [54] study of natural isolates gives further evidence that a small number of amino acid sites determines a large fraction of antigenic evolution to escape antibody recognition. The combination of analyses on structure, experimental evolution, and natural variation provide an opportunity to study how complex evolutionary forces together determine the evolutionary dynamics of particular amino acids.

15.6.4 Predicting Evolution

Studies on positive selection in FMDV [54] and HIV [146] could not correlate amino acid substitutions at particular sites with the actual success of the viruses. In each case, selection was inferred strictly from the patterns of nucleotide substitutions in a sample of sequences.

Bush et al.'s [24] study of influenza takes the next step by associating particular amino acid substitutions with the success or failure of descendants that carry the substitutions. Influenza allows such studies because sequences have been collected each year over the past several decades, providing a history of which substitutions have led to success over time. The influenza data can be used to predict future evolution by two steps. First, previous patterns of substitutions and the successes of associated lineages suggest which amino acid sites contain variants that enhance fitness. Second, new variants arising at those key sites are predicted to be the progenitors of future lineages. Bush (this volume) discusses these methods applied to the influenza data.

15.7 SHAPE, CHARGE, BINDING KINETICS, AND EVOLUTION

The molecular shape and charge of host and parasite molecules influence binding of those molecules, which defines the nature of host–parasite recognition. Binding reactions determine the course of infection within each host, and the advantages and disadvantages of different antigenic variants of the parasite. Those advantages and disadvantages set the course of success for the different variants, changing the frequency of variants over time and space and determining the evolution of antigenic variation.

ABBREVIATIONS

AIDS:	Acquired immunodeficiency syndrome
CTL:	Cytotoxic T lymphocyte
FMDV:	Foot-and-mouth disease virus
HCV:	Hepatitis C virus
HIV:	Human immunodeficiency virus
Mabs:	Monoclonal antibodies
MHC:	Major histocompatibility complex
NK:	Natural killer
PfEMP1:	*Plasmodium falciparum* erythrocyte membrane protein
SIV:	Simian immunodeficiency virus

GLOSSARY

Antigen: A molecule that induces an immune response.

Antigenic variation: Molecular variation between individual parasites in a particular antigenic molecule, usually a specific host immune response directed against one variant does not recognize other variants as well.

Archival copies: Different genetic loci that store and do not express variant genes for an antigenic molecule.

Bacteremia: The presence or amount of bacteria in the blood.

Cross-reaction: The reaction of an antibody with an antigen other than the one that gave rise to it.

Cytoadherence: Strength of binding by a parasite to the surface of host cells.

Cytokine: Molecules secreted by certain cells of the immune system that have an effect on other cells.

Epitope, continuous, and discontinuous: The part of an antigen molecule to which an antibody attaches itself; continuous if the epitope is composed of a linear sequence of amino acids in the protein chain; discontinuous if, during protein folding, the epitope forms from disparate parts of the amino acid sequence.

Escape mutant: A genetic variant of a parasite epitope in which the original type was recognized by a particular host immune response and the mutant is not.

Glycosylation: The addition of molecular sugar components to a protein, sometimes protects an antigen from being recognized by the host immune response.

Macrophage: An immune cell that devours invading pathogens; stimulates other immune cells by presenting them with small pieces of the invader.

MHC class I: Molecules of the major histocompatibility complex that bind small peptides within cells and then present the MHC–peptide complex on the surface of cells for interaction with T cells; class I can stimulate cytotoxic T lymphocytes.

Monoclonal antibodies: An antibody produced by a single clone of cells and consisting of identical antibody molecules.

Parasitemia: The presence or amount of parasites in the blood.

Polyclonal antibodies: An antibody produced by a multiple distinct clones of cells and consisting of diverse, distinct antibody molecules.

Receptor: A particular molecule on a host cell surface to which a parasite binds.

Site-directed mutagenesis: Experimentally controlled mutation to a particular part of a gene, causing a mutational change in the targeted amino acid.

Tropism, cell: Tendency of a particular parasite variant to bind to a particular kind of cell.

Viremia: The presence or amount of viruses in the blood.

REFERENCES

1. Abraham JM, Freitag CS, Clements JR, Eisenstein BI. An invertible element of DNA controls phase variation of type 1 fimbriae of *Escherichia coli*. *Proc Natl Acad Sci USA* 1985;**82**:5724–7.
2. Agur Z. Mathematical models for African trypanosomiasis. *Parasitol Today* 1992;**8**:128–9.
3. Agur Z, Abiri D, van der Ploeg, LHT. Ordered appearance of antigenic variants of African trypanosomes explained in a mathematical model based on a stochastic switch process and immune-selection against putative switch intermediates. *Proc Natl Acad Sci USA* 1989;**86**:9626–30.
4. Alcami A, Koszinowski UH. Viral mechanisms of immune evasion. *Trends Microbiol* 2000;**8**:410–8.
5. Allen TM, O'Connor DH, Jing P, et al. Tat-specific cytotoxic T lymphocytes select for SIV escape variants during resolution of primary viraemia. *Nature* 2000;**407**:386–90.
6. Almarri A, Batchelor JR. HLA and hepatitis B infection. *Lancet* 1994;**344**:1194–5.
7. Andreasen V, Lin J, Levin SA. The dynamics of cocirculating influenza strains conferring partial cross-immunity. *J Math Biol* 1997;**35**:825–42.
8. Aslam N, Turner CMR. The relationship of variable antigen expression and population growth rates in *Trypanosoma brucei*. *Parasitol Res* 1992;**78**:661–4.
9. Baranowski E, Ruiz-Jarabo CM, Sevilla N, Andreu D, Beck E, Domingo E. Cell recognition by foot-and-mouth disease virus that lacks the RGD integrin-binding motif: flexibility in aphthovirus receptor usage. *J Virol* 2000;**74**:1641–7.
10. Barbet AF, Kamper SM. The importance of mosaic genes to trypanosome survival. *Parasitol Today* 1993;**9**:63–6.
11. Barbour AG. Immunobiology of relapsing fever. *Contrib Microbiol Immunol* 1987;**8**:125–37.
12. Barbour AG. Linear DNA of *Borrelia* species and antigenic variation. *Trends Microbiol* 1993;**1**:236–9.
13. Barbour AG, Burman N, Carter CJ, Kitten T, Bergstrom S. Variable antigen genes of the relapsing fever agent *Borrelia hermsi* are activated by promoter addition. *Mol Microbiol* 1991;**5**:489–93.
14. Barbour AG, Stoenner HG. Antigenic variation of *Borrelia hermsii*. In *Genome Rearrangement* (eds I. Herskowitz and M.I. Simon), UCLA Symposium on Molecular and Cellular Biology, New Series, Vol. 20. Alan Liss, Inc., New York, 1985.
15. Barry JD. Antigenic variation during *Trypanosoma vivax* infections of different host species. *Parasitology* 1986;**92**:51–65.
16. Barry JD. The relative significance of mechanisms of antigenic variation in African trypanosomes. *Parasitol Today* 1997;**13**:212–8.
17. Barry JD, McCulloch R. Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. *Adv Parasitol* 2001;**49**:1–70.
18. Barry JD, Turner CMR. The dynamics of antigenic variation and growth of African trypanosomes. *Parasitol Today* 1991;**7**:207–11.
19. Barry JD, Turner CMR. Mathematical models for African trypanosomiasis – reply. *Parasitol Today* 1992;**8**:129.
20. Berinstein A, Roivainen M, Hovi T, Mason PW, Baxt B. Antibodies to the vitronectin receptor (integrin $\alpha_v\beta_{3)}$ inhibit binding and infection of foot-and-mouth disease virus to cultured cells. *J Virol* 1995;**69**:2664–6.
21. Borst P, Rudenko G, Blundell PA, et al. Mechanisms of antigenic variation in African trypanosomes. *Behring Institute Mitteilungen* 1997;**99**:1–15.
22. Brannan LR, Turner CMR, Phillips RS. Malaria parasites undergo antigenic variation at high rates in vivo. *Proc R Soc Lond Ser B Biol Sci* 1994;**256**:71–5.
23. Bush RM. Predicting adaptive evolution. *Nat Rev Genet* 2001;**2**:387–92.
24. Bush RM, Bender CA, Subbarao K, Cox NJ, Fitch WM. Predicting the evolution of influenza A. *Science* 1999;**286**:1921–5.

25. Calderwood MS, Gannoun-Zaki L, Wellems TE, Deitsch KW. *Plasmodium falciparum var* genes are regulated by two regions with separate promoters, one upstream of the coding region and a second within the intron. *J Biol Chem* 2003;**278**: 34125–32.
26. Carrillo C, Borca M, Moore DM, Morgan DO, Sobrino F. In vivo analysis of the stability and fitness of variants recovered from foot-and-mouth disease virus quasispecies. *J Gen Virol* 1998;**79**:1699–706.
27. Chackerian B, Rudensey LM, Overbaugh J. Specific N-linked and O-linked glycosylation modifications in the envelope V1 domain of simian immunodeficiency virus variants that evolve in the host after recognition by neutralizing antibodies. *J Virol* 1997;**71**:7719–27.
28. Chao L, Cox EC. Competition between high and low mutating strains of *Escherichia coli*. *Evolution* 1983;**37**:125–34.
29. Charlesworth B, Sniegowski P, Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 1994;**371**:215–20.
30. Chen Q, Fernandez V, Sundstrom A, et al. Developmental selection of *var* gene expression in *Plasmodium falciparum*. *Nature* 1998;**394**:392–5.
31. Coffin JM. Retroviridae: the viruses and their replication. In *Fields Virology* (eds B.N. Fields, D.M. Knipe, and P.M. Howley), 3rd edn. Lippincott-Raven, Philadelphia, 1996, pp. 763–843.
32. Connor RI, Sheridan KE, Ceradini D, Choe S, Landau NR. Change in coreceptor use correlates with disease progression in HIV-1-infected individuals. *J Exp Med* 1997;**185**:621–8.
33. Conway DJ, Roper C, Oduola AMJ, et al. High recombination rate in natural populations of *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 1999;**96**:4506–11.
34. Crandall KA (ed.). *The Evolution of HIV*. Johns Hopkins University Press, Baltimore, MD, 1999.
35. Deitsch KW, Calderwood MS, Wellems TE. Cooperative silencing of elements in *var* genes. *Nature* 2001;**412**:875–6.
36. Deitsch KW, Moxon ER, Wellems TE. Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections. *Microbiol Mol Biol Rev* 1997;**61**:281–93.
37. Deitsch KW, Wellems TE. Membrane modifications in erythrocytes parasitized by *Plasmodium falciparum*. *Mol Biochem Parasitol* 1996;**76**:1–10.
38. Drake JW, Charlesworth B, Charlesworth D, Crow JF. Rates of spontaneous mutation. *Genetics* 1998;**148**:1667–86.
39. Drake JW, Holland JJ. Mutation rates among RNA viruses. *Proc Natl Acad Sci USA* 1999;**96**:13910–3.
40. Endo T, Ikeo K, Gojobori T. Large-scale search for genes on which positive selection may operate. *Mol Biol Evol* 1996;**13**: 685–90.
41. Farci P, Shimoda A, Coiana A, et al. The outcome of acute hepatitis C predicted by the evolution of viral quasispecies. *Science* 2000;**288**:339–44.
42. Feigelstock DA, Mateu MG, Valero ML, Andreu D, Domingo E, Palma EL. Emerging foot-and-mouth disease virus variants with antigenically critical amino acid substitutions predicted by model studies using reference viruses. *Vaccine* 1996; **14**:97–102.
43. Fischetti VA. Streptococcal M protein. *Sci Am* 1991;**264**(6): 32–9.
44. Frank SA. A model for the sequential dominance of antigenic variants in African trypanosome infections. *Proc R Soc Lond Ser B Biol Sci* 1999;**266**:1397–401.
45. Frank SA. *Immunology and Evolution of Infectious Disease*. Princeton University Press, Princeton, NJ, 2002.
46. Frank SA, Barbour AG. Within-host dynamics of antigenic variation. *Infect Genet Evol* 2006;6(12):141–6.
47. Fussenegger M. Different lifestyles of human pathogenic procaryotes and their strategies for phase and antigenic variation. *Symbiosis* 1997;**22**:85–153.
48. Fussenegger M, Rudel T, Barten R, Ryll R, Meyer TF. Transformation competence and type-1 pilus biogenesis in *Neisseria gonorrhoeae*: a review. *Gene* 1997;**192**:125–34.
49. Goulder PJR, Brander C, Tang Y, et al. Evolution and transmission of stable CTL escape mutations in HIV infection. *Nature* 2001;**412**:334–8.
50. Gray AR. Antigenic variation in a strain of *Trypanosoma brucei* transmitted by *Glossina morsitans* and *G. palpalis*. *J Gen Microbiol* 1965;**41**:195–214.
51. Gross MD, Siegel EC. Incidence of mutator strains in *Escherichia coli* and coliforms in nature. *Mutat Res* 1981;**91**:107–10.
52. Gubbels M-J, Katzer F, Hide G, Jongejan F, Shiels BR. Generation of a mosaic pattern of diversity in the major merozoite-piroplasm surface antigen of *Theileria annulata*. *Mol Biochem Parasitol* 2000;**110**:23–32.
53. Hasegawa M, Rienzo AD, Kocher TD, Wilson AC. Toward a more accurate time scale for the human mitochondrial DNA tree. *J Mol Evol* 1993;**37**:347–54.
54. Haydon DT, Bastos AD, Knowles NJ, Samuel AR. Evidence for positive selection in foot-and-mouth disease virus capsid genes from field isolates. *Genetics* 2001;**157**:7–15.
55. Hill AVS. The immunogenetics of human infectious diseases. *Annu Rev Immunol* 1998;**16**:593–617.
56. Hillis DM, Moritz C, Mable BK (eds). *Molecular Systematics*. Sinauer Associates, Sunderland, MA, 1996.
57. Hohler T, Gerken G, Notghi A, et al. HLA-DRB1*1301 and *1302 protect against chronic hepatitis B. *J Hepatol* 1997;**26**:503–7.
58. Hollingshead SK, Fischetti VA, Scott JR. Complete nucleotide sequence of type 6 M protein of the group A streptococcus: repetitive structure and membrane anchor. *J Biol Chem* 1986;**261**:1677–86.
59. Hollingshead SK, Fischetti VA, Scott JR. Size variation in group A streptococcal protein is generated by homologous recombination between intragenic repeats. *Mol Gen Genet* 1987;**207**:196–203.
60. Hughes AL. *Adaptive Evolution of Genes and Genomes*. Oxford University Press, Oxford, 1999.
61. Iida S, Meyer J, Kennedy KE, Arber W. A site-specific, conservative recombination system carried by bacteriophage P1: mapping the recombinase gene *cin* and the crossover sites *cix* for the inversion of the C-segment. *EMBO J* 1982;**1**:1445–53.
62. Jackson T, Blakemore W, Newman JWI, et al. Foot-and-mouth disease virus is a ligand for the high-affinity binding conformation of integrin $\alpha_5\beta_1$: influence of the leucine residue within the RGDL motif on selectivity of integrin binding. *J Gen Virol* 2000;**81**:1383–91.

63. Jackson T, Sheppard D, Denyer M, Blakemore WE, King AMQ. The epithelial integrin $\alpha_v\beta_6$ is a receptor for foot-and-mouth disease virus. *J Virol* 2000;**74**:4949–56.
64. Jenni L, Marti S, Schweizer J, et al. (1986). Hybrid formation between trypanosomes during cyclical transmission. *Nature* 2000;**322**:173–5.
65. Jyssum K. Observations on two types of genetic instability in *Escherichia coli*. *Acta Pathol Microbiol Scand* 1960;**48**:113–20.
66. Kamp D, Kahmann R, Zipser D, Broker TR, Chow LT. Inversion of the G DNA segment of phage Mu controls phage infectivity. *Nature* 1978;**271**:577–80.
67. Kimata JT, Kuller L, Anderson DB, Dailey P, Overbaugh J. Emerging cytopathic and antigenic simian immunodeficiency virus variants influence AIDS progression. *Nat Med* 1999;**5**:535–41.
68. Kimura M. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, 1983.
69. Kosinski RJ. Antigenic variation in trypanosomes: a computer analysis of variant order. *Parasitology* 1980;**80**:343–57.
70. Lamb RA, Krug RM. *Orthomyxoviridae*: the viruses and their replication. In *Fields Virology* (eds D.M. Knipe and P.M. Howley), 4th edn. Lippincott-Raven, Philadelphia, 2001, pp. 1487–531.
71. Lancefield RC. Current knowledge of type-specific M antigens of group A streptococci. *J Immunol* 1962;**89**:307–13.
72. Laukkanen T, Carr JK, Janssens W, et al. Virtually full-length subtype F and F/D recombinant HIV-1 from Africa and South America. *Virology* 2000;**269**:95–104.
73. LeClerc JE, Li B, Payne WL, Cebula TA. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* 1996;**274**:1208–11.
74. Leigh EG. Natural selection and mutability. *Am Nat* 1970;**104**:301–5.
75. Levinson G, Gutman GA. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 1987;**4**:203–21.
76. Li W-H. *Molecular Evolution*. Sinauer Associates, Sunderland, MA, 1997.
77. MacLeod A, Turner CMR, Tait A. A high level of mixed *T. brucei* infections in tsetse flies detected by three hypervariable minisatellites. *Mol Biochem Parasitol* 1999;**102**:237–48.
78. Mao EF, Lane L, Lee J, Miller JH. Proliferation of mutators in a cell population. *J Bacteriol* 1997;**179**:417–22.
79. Marrs CF, Reuhl WW, Schoolnik GK, Falkow S. Pilin gene phase variation of *Moraxella bovis* is caused by an inversion of the pilin gene. *J Bacteriol* 1988;**170**:3032–9.
80. Martinez MA, Verdaguer N, Mateu MG, Domingo E. Evolution subverting essentiality: dispensability of the cell attachment Arg-Gly-Asp motif in multiply passaged foot-and-mouth disease virus. *Proc Natl Acad Sci USA* 1997;**94**:6798–802.
81. Martinson JJ, Chapman NH, Rees DC, Liu Y-T, Clegg JB. Global distribution of the CCR5 gene 32-basepair deletion. *Nat Genet* 1997;**16**:100–3.
82. Mateu MG. Antibody recognition of picornaviruses and escape from neutralization: a structural view. *Virus Res* 1995;**38**:1–24.
83. Mateu MG, Escarmis C, Domingo E. Mutational analysis of discontinuous epitopes of foot-and-mouth disease virus using an unprocessed capsid promoter precursor. *Virus Res* 1998;**53**:27–37.
84. Mateu MG, Hernandez J, Martinez MA, et al. Antigenic heterogeneity of a foot-and-mouth disease virus serotype in the field is mediated by very limited sequence variation at several antigenic sites. *J Virol* 1994;**68**:1407–17.
85. Mateu MG, Silva DA, Rocha E, et al. Extensive antigenic heterogeneity of foot-and-mouth disease virus of serotype C. *Virology* 1988;**167**:113–24.
86. McMichael AJ, Phillips RE. Escape of human immunodeficiency virus from immune control. *Annu Rev Immunol* 1997;**15**:271–96.
87. Meyer TF. Molecular basis of surface antigen variation in *Neisseria*. *Trends Genet* 1987;**3**:319–24.
88. Molineaux L, Diebner HH, Eichner M, Collins WE, Jeffery GM, Dietz K. *Plasmodium falciparum* parasitaemia described by a new mathematical model. *Parasitology* 2001;**122**:379–91.
89. Moxon ER, Rainey PB, Nowak MA, Lenski RE. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr Biol* 1994;**4**:24–33.
90. Munoz-Jordan JL, Davies KP, Cross GA. Stable expression of mosaic coats of variant surface glycoproteins in *Trypanosoma brucei*. *Science* 1996;**272**:1795–7.
91. Nee S, Holmes EC, Rambaut A, Harvey PH. Inferring population history from molecular phylogenies. *Philos Trans R Soc Lond Ser B Biol Sci* 1995;**349**:25–31.
92. Neff S, Sa-Carvalho D, Rieder E, et al. Foot-and-mouth disease virus virulent for cattle utilizes the integrin $\alpha_v\beta_3$ as its receptor. *J Virol* 1998;**72**:3587–94.
93. Nei M. *Molecular Evolutionary Genetics*. Columbia University Press, New York, 1987.
94. Nei M, Kumar S. *Molecular Evolution and Phylogenetics*. Oxford University Press, Oxford, 2000.
95. Nunez JI, Baranowski E, Molina N, et al. A single amino acid substitution in nonstructural protein 3A can mediate adaptation of foot-and-mouth disease virus to the guinea pig. *J Virol* 2001;**75**:3977–83.
96. O'Brien SJ, Dean M. In search of AIDS-resistance genes. *Sci Am* 1997;**277**:44–51.
97. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. *Nature* 2000;**405**:299–304.
98. Otto SP. Detecting the form of selection from DNA sequence data. *Trends Genet* 2000;**16**:526–9.
99. Overbaugh J, Bangham CRM. Selection forces and constraints on retroviral sequence variation. *Science* 2001;**292**:1106–9.
100. Page RDM, Holmes EC. *Molecular Evolution: A Phylogenetic Approach*. Blackwell Scientific, Oxford, 1998.
101. Paget-McNicol S, Gatton M, Hastings I, Saul A. The *Plasmodium falciparum var* gene switching rate, switching mechanism and patterns of parasite recrudescence described by mathematical modelling. *Parasitology* 2002;**124**:225–35.
102. Pays E. Gene conversion in trypanosome antigenic variation. *Prog Nucleic Acid Res Mol Biol* 1985;**32**:1–26.
103. Pays E. Pseudogenes, chimeric genes and the timing of antigen variation in African trypanosomes. *Trends Genet* 1989;**5**:389–91.
104. Pays E, Lheureux M, Steinert M. The expression-linked copy of the surface antigen gene in *Trypanosoma* is probably the one transcribed. *Nature* 1981;**292**:265–7.

105. Pays E, Nolan D. Expression and function of surface proteins in *Trypanosoma brucei*. *Mol Biochem Parasitol* 1998;**91**:3–36.
106. Preston BD, Dougherty JP. Mechanisms of retroviral mutation. *Trends Microbiol* 1996;**4**:16–21.
107. Racaniello VR. *Picornaviridae*: the viruses and their replication. In *Fields Virology* (eds D.M. Knipe and P.M. Howley), 4th edn. Lippincott-Raven, Philadelphia, 2001, pp. 685–722.
108. Recker M, Nee S, Bull PC, Kinyanjui S, Marsh K, Newbold C, Gupta S. Transient cross-reactive immune responses can orchestrate antigenic variation in malaria. *Nature* 2004;**429**: 555–8.
109. Reeder JC, Brown GV. Antigenic variation and immune evasion in *Plasmodium falciparum* malaria. *Immunol Cell Biol* 1996;**74**:546–54.
110. Rich SM, Sawyer SA, Barbour AG. Antien polymorphism in *Borrelia hermsii*, a clonal pathogenic bacterium. *Proc Natl Acad Sci USA* 2001;**98**:15038–43.
111. Ripley LS. Predictability of mutant sequences: relationships between mutational mechanisms and mutant specificity. *Ann New York Acad Sci* 1999;**870**:159–72.
112. Robertson DL, Hahn BH, Sharp PM. Recombination in AIDS viruses. *J Mol Evol* 1995;**40**:249–59.
113. Rodrigo AG, Learn GH, Jr (eds). *Computational and Evolutionary Analysis of HIV Molecular Sequences*. Kluwer Academic Publishers, Boston, 2000.
114. Rottem S, Naot Y. Subversion and exploitation of host cells by mycoplasmas. *Trends Microbiol* 1998;**6**:436–40.
115. Rozsa WF, Marrs CF. Interesting sequence differences between the pilin gene inversion regions of *Moraxella lacunata* ATCC 17956 and *Moraxella bovis* Epp63. *J Bacteriol* 1991;**173**:4000–6.
116. Rudensey LM, Kimata JT, Long EM, Chackerian B, Overbaugh J. Changes in the extracellular envelope glycoprotein of variants that evolve during the course of simian immunodeficiency virus SIVMne infection affect neutralizing antibody recognition, syncytium formation, and macrophage tropism but not replication, cytopathicity, or CCR-5 coreceptor recognition. *J Virol* 1998;**72**:209–17.
117. Saha K, Zhang J, Gupta A, Dave R, Yimen M, Zerhouni B. Isolation of primary HIV-1 that target CD8+ T lymphocytes using CD8 as a receptor. *Nat Med* 2001;**7**:65–72.
118. Scherf A, Hernandez-Rivas R, Buffet P, et al. Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of *var* genes during intra-erythrocytic development in *Plasmodium falciparum*. *EMBO J* 1998;**17**:5418–26.
119. Seed JR. Competition among serologically different clones of *Trypanosoma brucei gambiense* in vivo. *J Protozool* 1978;**25**:526–9.
120. Silverman M, Zieg J, Hilmen M, Simon M. Phase variation in *Salmonella*: genetic analysis of a recombinational switch. *Proc Natl Acad Sci USA* 1979;**76**:391–5.
121. Smith JD, Chitnis CE, Craig AG, et al. Switches in expression of *Plasmodium falciparum var* genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* 1995;**82**:101–10.
122. Smith MW, Dean M, Carrington M, et al. Contrasting genetic influence of CCR2 and CCR5 variants on HIV-1 infection and disease progression. *Science* 1997;**277**:959–65.
123. Smith MJ, Smith NH, O'Rourke M, Spratt BG. How clonal are bacteria? *Proc Natl Acad Sci USA* 1993;**90**:4384–8.
124. Sniegowski PD, Gerrish PJ, Lenski RE. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* 1997;**387**:703–5.
125. Sobrino F, Saiz M, Jimenez-Clavero MA, et al. Foot-and-mouth disease virus: a long known virus, but a current threat. *Vet Res* 2001;**32**:1–30.
126. Soudeyns H, Paolucci S, Chappey C, et al. Selective pressure exerted by immunodominant HIV-1-specific cytotoxic T lymophocyte responses during primary infection drives genetic variation restricted to the cognate epitope. *Eur J Immunol* 1999;**29**:3629–35.
127. Spratt BG, Bowler LD, Zhang Q-Y, Zhou J, Smith JM. Role of interspecies transfer of chromosomal genes in the evolution of penicillin resistance in pathogenic and commensal *Neisseria* species. *J Mol Evol* 1992;**34**:115–25.
128. Spriggs MK. One step ahead of the game: viral immunomodulatory molecules. *Annu Rev Immunol* 1996;**14**:101–30.
129. Stern A, Brown M, Nickel P, Meyer TF. Opacity genes in *Neisseria gonorrhoeae*: control of phase and antigenic variation. *Cell* 1986;**47**:61–71.
130. Stern A, Meyer TF. Common mechanism controlling phase and antigenic variation in pathogenic neisseriae. *Mol Microbiol* 1987;**1**:5–12.
131. Stoenner HG, Dodd T, Larsen C. Antigenic variation of *Borrelia hermsii*. *J Exp Med* 1982;**156**:1297–311.
132. Strauss EG, Straus JH, Levine AJ. Virus evolution. In *Fundamental Virology* (eds B.N. Fields, D.M. Knipe, and P.M. Howley), 3rd edn. Lippincott-Raven, Philadelphia,1996, pp. 141–59.
133. Su X, Heatwole VM, Wertheimer SP, et al. A large and diverse gene family (var) encodes 200–350 kD proteins implicated in the antigenic variation and cytoadherence of *Plasmodium falciparum*–infected erythrocytes. *Cell* 1995;**82**:89–100.
134. Suzuki Y, Gojobori T. A method for detecting positive selection at single amino acid sites. *Mol Biol Evol* 1999;**16**: 1315–28.
135. Taddei F, Radman M, Smith JM, Toupance B, Gouyon PH, Godelle B. Role of mutator alleles in adaptive evolution. *Nature* 1997;**387**:700–2.
136. Thursz M, Kwiatkowski D, Allsopp CEM, Greenwood BM, Thomas HC, Hill AVS. Association of an HLA class II allele with clearance of hepatitis B virus infection in The Gambia. *N Engl J Med* 1995;**332**:1065–9.
137. Tortorella D, Gewurz BE, Furman MH, Schust DJ, Ploegh HL. Viral subversion of the immune system. *Annu Rev Immunol* 2000;**18**:861–926.
138. Turner CMR. The rate of antigenic variation in fly-transmitted and syringe-passaged infections of *Trypanosoma brucei*. *FEMS Microbiol Lett* 1997;**153**:227–31.
139. Turner CMR. Antigenic variation in *Trypanosoma brucei* infections: an holistic view. *J Cell Sci* 1999;**112**:3187–92.
140. Turner CMR. A perspective on clonal phenotypic (antigenic) variation in protozoan parasites. *Parasitology* 2002;**125**: S17–23.
141. Vickerman K. Trypanosome sociology and antigenic variation. *Parasitology* 1989;**99**:S37–47.
142. Wakely J. Substitution rate variation among sites in hypervariable region I of human mitochondrial DNA. *J Mol Evol* 1993;**37**:613–23.

143. Webster RG, Shortridge KF, Kawaoka Y. Influenza: interspecies transmission and emergence of new pandemics. *FEMS Immunol Med Microbiol* 1997;**18**:275–9.
144. Willems R, Paul A, van der Heide HGJ, ter Avest AR, Mooi FR. Fimbrial phase variation in *Bordetella pertussis*: a novel mechanism for transcriptional regulation. *EMBO J* 1990;**9**:2803–9.
145. Wren BW. A family of clostridial and streptococcal ligand-binding proteins with conserved C-terminal repeat sequences. *Mol Microbiol* 1991;**5**:797–803.
146. Yamaguchi-Kabata Y, Gojobori T. Reevaluation of amino acid variability of the human immunodeficiency virus type 1 gp120 envelope glycoprotein and prediction of new discontinuous epitopes. *J Virol* 2000;**74**:4335–50.
147. Yang ZH, Bielawski JP. Statistical methods for detecting molecular adaptation. *Trends Ecol Evol* 2000;**15**:496–503.
148. Zhou J, Spratt BG. Sequence diversity within the *argF*, *fbp* and *recA* genes of natural isolates of *Neisseria meningitidis*—interspecies recombination within the *argF* gene. *Mol Microbiol* 1992;**6**:2135–46.