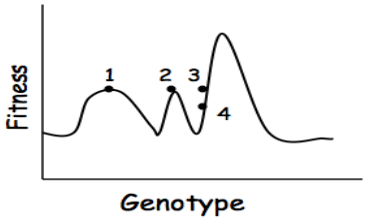
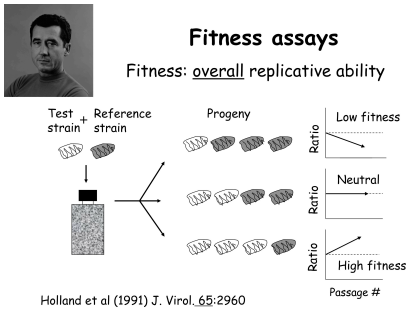
The basis of evolutionary success

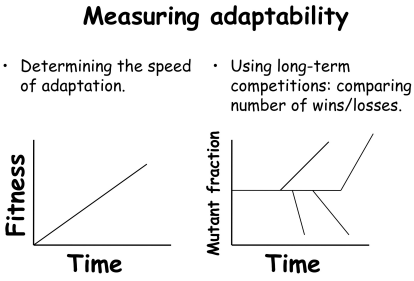


Three different factors determine the long-term evolutionary success of a population:

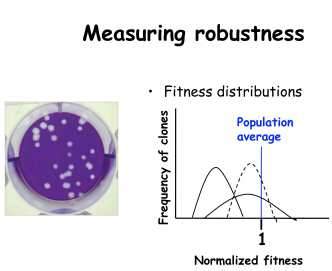
- Fitness, defined as the overall ability of a population to survive in a defined environment. Relative fitness refers to the ability of one population to survive in comparison to another population used as a standard (of neutrality). Fitness is shown as the height of peaks in fitness landscapes. Fitness has the most immediate of effects. There are many ways to measure fitness and express fitness values, the most common $W = 1+s$, where s is the selection coefficient. For the purpose of these lectures W values will represent ratios of relative frequencies determined before and after competition.

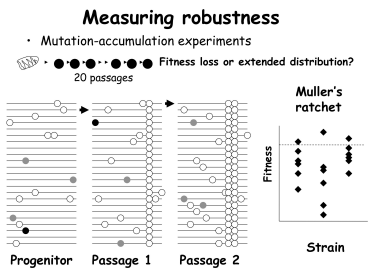


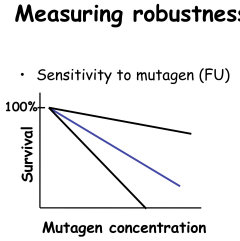
- Adaptability, defined as the ability of a population to increase fitness. Some authors use the term *evolvability*, but I prefer the former because of the possibility of fitness loss or of sequence evolution without fitness changes. Adaptability is shown in fitness landscapes as upward slopes surrounding the position of the population. There are several methods to measure adaptability, including determination of speed/level of adaptation and determination of outcomes during long-term competitions (when populations have similar fitness).



- Robustness, defined as the ability of a population to maintain a constant phenotype despite mutational pressure. In a fitness landscape high robustness corresponds to flat areas, while low robustness corresponds to sharp peaks. There is an inverse correlation between adaptability (or evolvability) and robustness. There are also several methods to measure robustness, including determination of fitness distributions, determination of divergence from neutrality during Muller's ratchet and determination of sensitivity to mutagens.







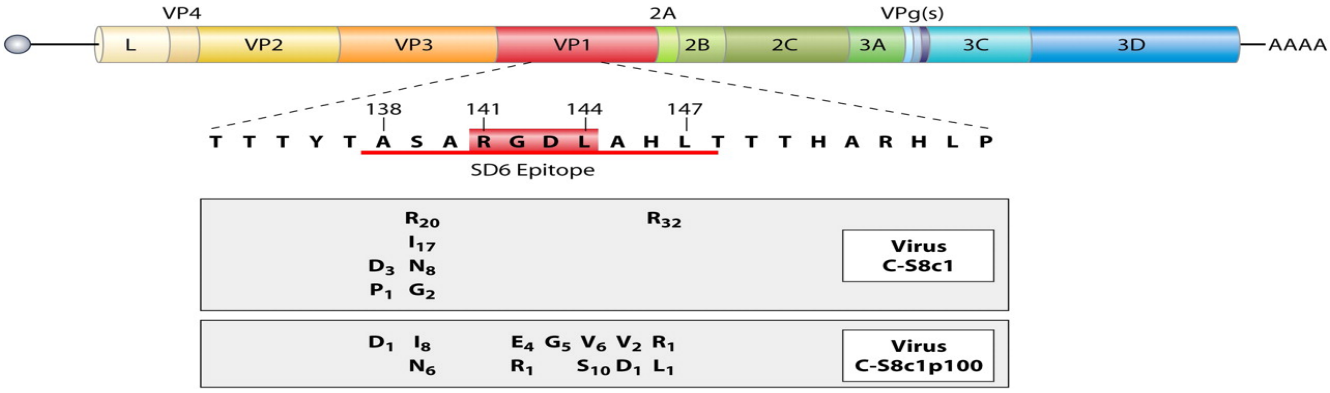
Targets of selection

In this section the goal is to highlight aspects of how selection may act upon a biological system, particularly in viruses, in ways that may be counterintuitive or are usually neglected.

Proteins

Traditionally the "importance" of a sequence has been assessed based on it evolutionary conservation. Thus, highly conserved sequences are dubbed *important* and variable sequences are considered *unimportant* or even *dispensable.*

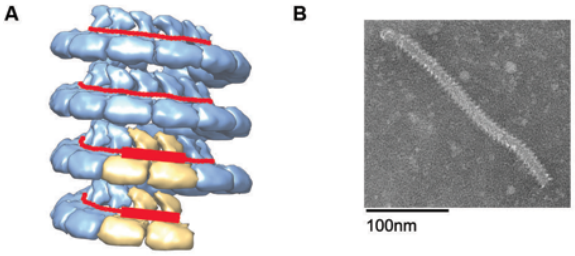
Once we take into consideration the biological function of a sequence it may be worth reconsidering. For instance, animal viruses face immune pressure and their survival depends on the ability of antigenic sites to change. **Lack** of sequence conservation may be important. The figure below shows extreme variation of the A1 antigenic site. The ability of some residues to evolve allows maintenance of the constant integrin-binding RGD motif (the antireceptor).



It is also important to avoid a second misconception: that conservation indicates an inability to change. While it may be so in some cases in (many?) others conserved sites may evolve upon specific environmental changes. The second part of the figure shows that even the highly conserved RGD motif can mutate to produce viable virus. In this case the evolution of FMDV during persistent infection resulted in the recognition of alternative receptors.

Nucleic Acids

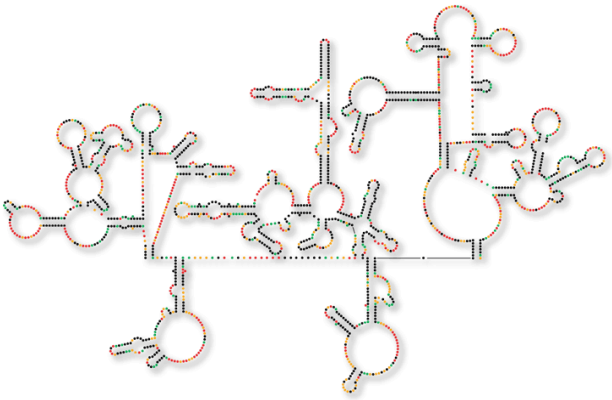
Rule of 6

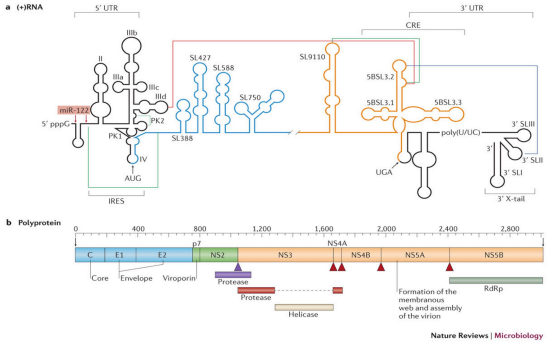


Negative-stranded RNA viruses require encapsidation of the genome (and antigenome) templates for synthesis to take place. Among the negative stranders, paramyxoviruses have the additional constraint that the structure does not leave unmatched RNA or protein. Because each nucleoprotein molecule matches perfectly with 6 nucleotides, (anti)genomic templates must be a multiple of 6 or replication cannot take place.

RNA secondary structures

Secondary RNA structures are significant for two reasons: first, because they induce mammalian TLR-3-mediated innate immunity and second, because they carry out functions during viral replication. Secondary structures are particularly important in positive-stranders because the genomes themselves are highly structured. In negative stranders the main role of secondary structures are related to mRNA.





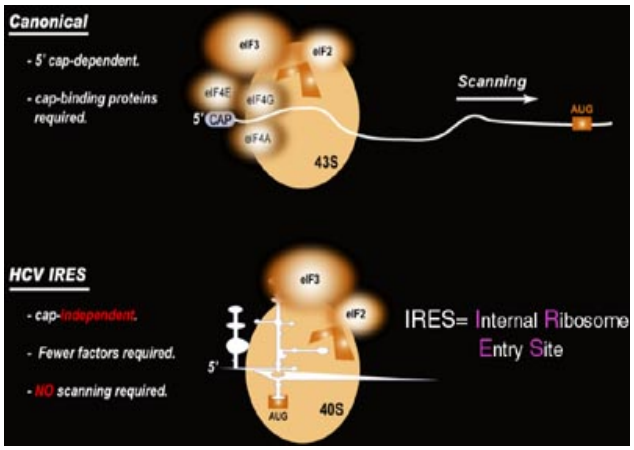
Encapsidation signals

Encapsidation signals are sequences that guide the progeny genome to its shell. These signals are frequently found in either or both genomic termini, but in other instances the situation is more complex and encapsidation signals overlap with coding sequences, as is the case in the polymerase ORF of phage MS2.

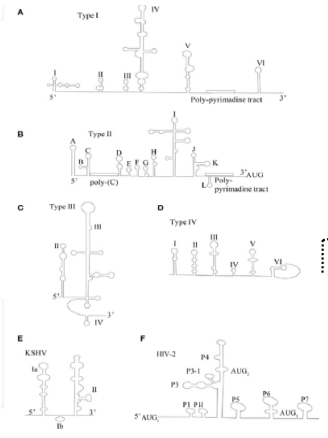
IRES element

Soon after poliovirus infection, cellular protein synthesis declines sharply, almost to zero. This is called '**shutoff**', and in this case it is the primary cause of cytophatic effect (c.p.e.). The 5' UTR contains the IRES: Internal Ribosome Entry Site or 'landing pad'. Normally, translation is initiated when ribosomes bind to the 5' methylated cap then scan along the mRNA to find the first AUG initiation codon. The IRES overcomes this & allows picornavirus RNAs to continue to be translated after degradation of cellular transcription factors. The IRES element is so critical in the life of poliovirus that mutations in its sequence drastically reduce the production of new virions in infected cells. For instance, the attenuation of poliovirus in the Sabin vaccine is largely due to mutations in the IRES (note that some of the attenuating mutations do not affect secondary structure!)

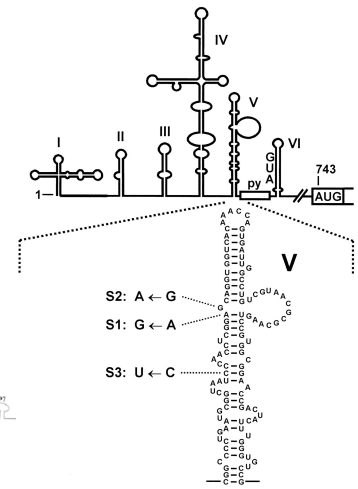
Cap-dependent and Cap-independent Translation



IRES Elements



The Sabin vaccine



Codon usage

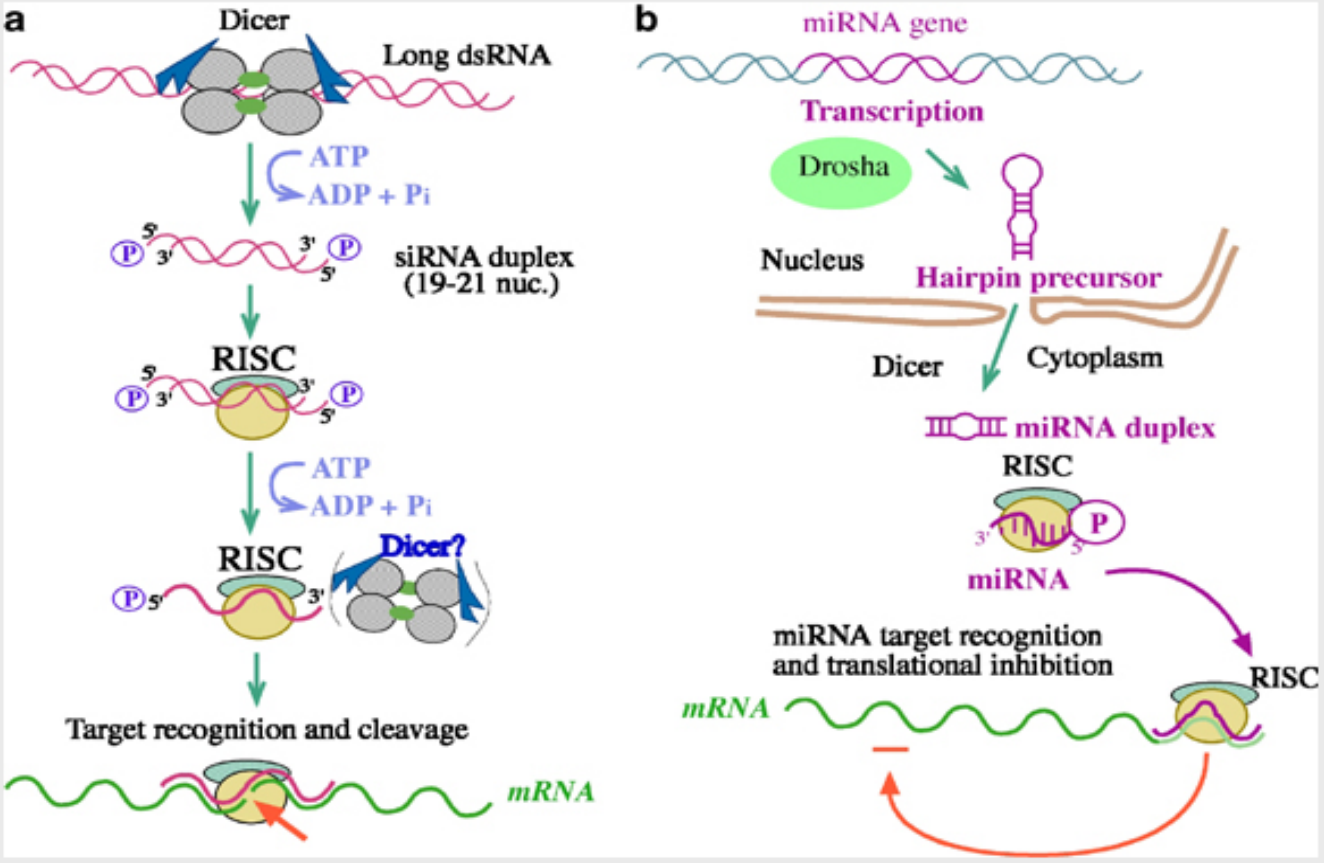
Host(s) optimization

Codon usage varies among species of eukaryotes and prokaryotes and these differences are an obvious environmental factor that play a role in virus evolution. It is easy to imagine that one way to maximize speed of biosynthesis is to use codons with abundant tRNA matches. However, it is important to remember that many viruses, particularly among the RNA viruses, are capable of infecting a wide variety of hosts, each of which may have its own codon usage bias. Replacement of wild type sequences at synonymous sites can have extremely deleterious effects, as it has been shown in poliovirus and HIV. On the other hand, HIV research has been made a bit easier by generating artificial genomes with "humanized" suboptimal sites. Nevertheless codon usage may be a useful tool to understand the ecology and evolution of some viruses.

Protein folding

Under some circumstances fast is not better. Once such example is during persistent infection, when virus is produced continually but at low level. This type of infection allows cell survival. Another case is that in which protein folding is complex, so a bit of extra time results in a well-folded, functional protein, while more speed results in a misfolded and dysfunctional molecule. This is the reason why hepatitis A virus' codon usage does not match well human codon usage, and artificial genomes that are generated with optimized codon use evolve towards suboptimization during infection.

siRNA (a) and miRNA (b)



siRNA and miRNA mediate the cleavage of target mRNA and lead to the inhibition of protein synthesis. Because the targets are only a few nucleotides long, it is easy to escape si/miRNA

targeting with single nucleotide substitutions. Both host cells and viruses encode sequences that target each other and that are susceptible to evolve.

Implications for sequence analysis

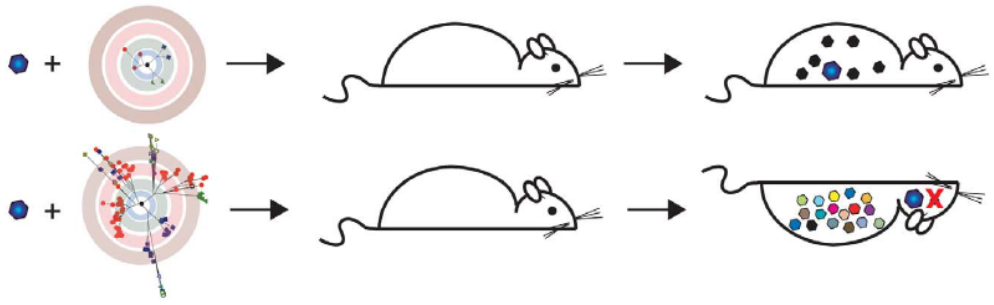
One of the most common parameters that we use to assess the forces shaping the evolution of a population is dN/dS. The theory is that because selection works (mostly) on proteins, these forces are reflected in the comparison of synonymous mutations (dS) and non-synonymous mutations (dN) accumulated through time. If random events prevail both types of mutations will accumulate equally because even if the protein doesn't work the absence of selection allows fixation. If positive selection prevails one would expect proteins to change, while the same cannot happen to synonymous mutations, resulting on an excess of nonsynonymous changes. Finally, if negative selection prevails the synonymous changes will be allowed, but the nonsynonymous changes will be eliminated, thus producing an excess of the former. While this is true for most of the biosphere, because RNA genome sequences themselves have such a variety of roles in the replication cycle of a virus, a lot of caution is needed when using dN/dS to identify selective forces (or lack thereof). HCV is one of the viruses claimed to be under negative selection because of an excess of synonymous mutations, but if was shown that may of them are compensatory mutations acquired after periods of drift to restore secondary structures.

We don't need to completely disregard dN/dS. Because processes such as receptor recognition and antibody binding depend on protein function, dN/dS may be useful to look for signals suggesting changes in tropism or immune escape. More generally transitions/transversions may be a better way to find a signal of selection.

Populations

RNA viruses are the subject of group/kin selection in two independent ways. In this section I will only address one of the mechanisms: cooperative interactions. As mentioned previously mutation rates are genetically determined and can evolve. One good way to do that is the use of nucleoside analogs such as chain terminators or mutagens. Viruses that replicate in the presence of these compounds either die or become resistant by lowering the probability of incorporating the drugs during replication, which they can do by increasing fidelity. The more accurate polymerase spends some time at each pair of complementary nucleotides and does not form covalent bonds if there is a signal of mismatch. These high-fidelity mutants have been isolated from several viral species and tend to have a lower fitness associated with the slower replication rate. In addition, poliovirus high fidelity mutants have low virulence due to their inability to reach the CNS. The critical experiment in this study was the use of mutagenized high fidelity mutant that despite its slow replication can replicate extensively, reach the brain and kill the host, while virus isolated from brain could not do the same.

Diversity determines the ability of poliovirus to extend its tropism



Evolution of virulence

Mechanisms of virulence

The pathogenesis of a virus infection is the aggregate of the effects produced in the host. These include the direct effects of virus replication and also the effects of the host response to infection. While disease is often the result of massive replication and direct cell killing by the virus, it is not always necessarily so.

Oncogenesis

It is now estimated that approximately 20% of human cancers have a viral etiology. This figure is likely to rise as epidemiologic data accrues on the etiologic association of hepatitis C virus (*Flaviviridae*) with hepatocellular carcinoma, and new viruses are recognized to be causally associated with malignancy. Normally, oncogenic viruses establish latency or become persistent infections. The major virus-malignancy systems include hepatitis B virus (HBV, *Hepadnaviridae*), hepatitis C virus (HCV), and hepatocellular carcinoma; human lymphotropic virus-type 1 (HTLV-1, *Retroviridae*) and adult T-cell leukemia/lymphoma (ATL, *Retroviridae*); Epstein-Barr virus (EBV, *Herpesviridae*) and endemic Burkitt's lymphoma, nasopharyngeal carcinoma, and Hodgkin's disease; human herpesvirus 8 (HH8, *Herpesviridae*) and Kaposi sarcoma; and human papilloma virus (HPV, *Papillomaviridae*) and cervical cancer. These malignancies tend to occur in early to mid-life and account for a substantial amount of morbidity and mortality. They are also likely to occur as "opportunistic malignancies" among individuals infected with human immunodeficiency virus type-1, particularly among those who experience prolonged survival.

Viruses cause cancer by multiple mechanisms, none of which if fully understood: (i) by continuous stimulation of cell growth due to continues cell killing by CTLs; (ii) by insertion of the genome in sites that are oncogenes, cell-cycle proteins and/or regulatory signals; and (iii) by interaction of viral proteins with regulatory signals in the cell chromosome and with proteins implicated in cell-cycle regulation and apoptosis.

Immunopathogenesis

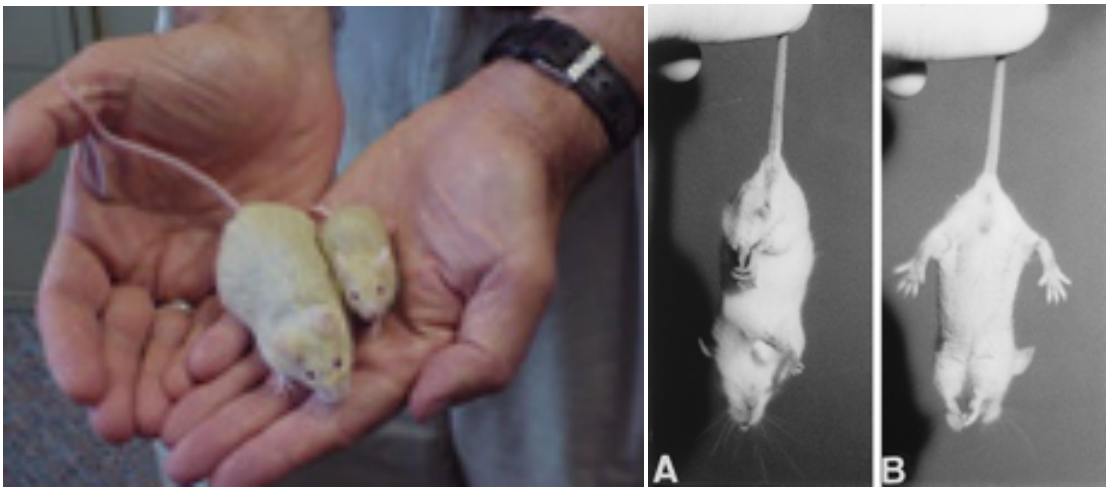
There are four different subtypes of Dengue virus (*Flaviviridae*) producing varying manifestations of the disease. Dengue fever is transmitted to humans by *Aedes* mosquitoes. A large number of infections may be sub-clinical, that is, the patients may not even be aware that they have had the disease. The infection usually manifests itself as fever with severe body pain, and may be an associated rash over parts of the body. Quite often, the disease makes no further progress and the patients recover. However, some patients may develop involvement of either of the two dreaded syndromes in Dengue - bleeding (called DHF or Dengue Hemorraghic Fever) or involvement of the brain with altered consciousness (encephalitis), both causing high mortality. This is typically observed in patients that had prior exposure to the virus (and have seroconverted). There is now a considerable body of evidence to show that antibody dependent enhancement (ADE) of infection has a role in dengue pathogenesis. ADE is thought to occur through low levels of pre-existing antibodies to the major virion envelope protein. These antibodies can bind to Fc-receptor and the antibody-Fc receptor complex can act as a surrogate receptor for the virus. As the antibody is incapable of neutralizing the virus it can enter the cell, replicate and kill it. Fc receptor-bearing cells are frequently important components of the body’s immune defenses and therefore their deaths in large numbers can serious impair the immune response and result in severe or fatal disease.

Subverting the physiology of the cell.

Lymphocytic choriomeningitis virus (LCMV, *Arenaviridae*), naturally a murine virus, can cause very different diseases depending on viral strain, mouse genotype and route of infection. One of these diseases is growth hormone deficiency syndrome (GDH). GDH is characterized by manifested as growth retardation and hypoglycemia, Infected mice show high levels of viral replication in the GH-producing cells in the anterior pituitary leading to decreased synthesis of GH mRNA and protein despite the absence of detectable virus-induced cell structural damage.

Borna Disease Virus (BDV, *Bornaviridae*) is a unique agent associated with neurologic disease in a broad array of animals, and has recently been implicated as a possible cause of human affective disorders (bipolar disorder and chronic fatigue syndrome). BDV is classified among the Mononegavirales. The virus can grow in brain cells (astrocytes and oligodendrocytes) in a variety of animal species.In the absence of an immune response, such as in cell culture, infected cells are indistinguishable from uninfected cells. However, BDV can induce alterations in cell functions that potentially affect brain homeostasis, including molecules involved in communication between neurons.

Pathogenesis in the absence of cell damage
GHD deficient mouse (right) BDV infected mice (left)



Escape from the immune system

Viruses have a variety of ways to avoid immune surveillance. Some of the mechanisms target the innate immune response, and others are directed against the adaptive immune response. There are two main ways to escape the host defenses:

Modification of the host immune response

Replication in immunoprivileged sites

Certain regions of the body are protected from the immune system. The best example is the CNS, which is protected by the blood-brain barrier. In addition, neurons do not express class I MHC. Viruses that infect neurons are safe from immune surveillance. Examples of viruses that follow this strategy are HSV (*Herpesviridae*), measles (*Paramyxoviridae*) and several alphaviruses (*Togaviridae*).

Replication in immune cells

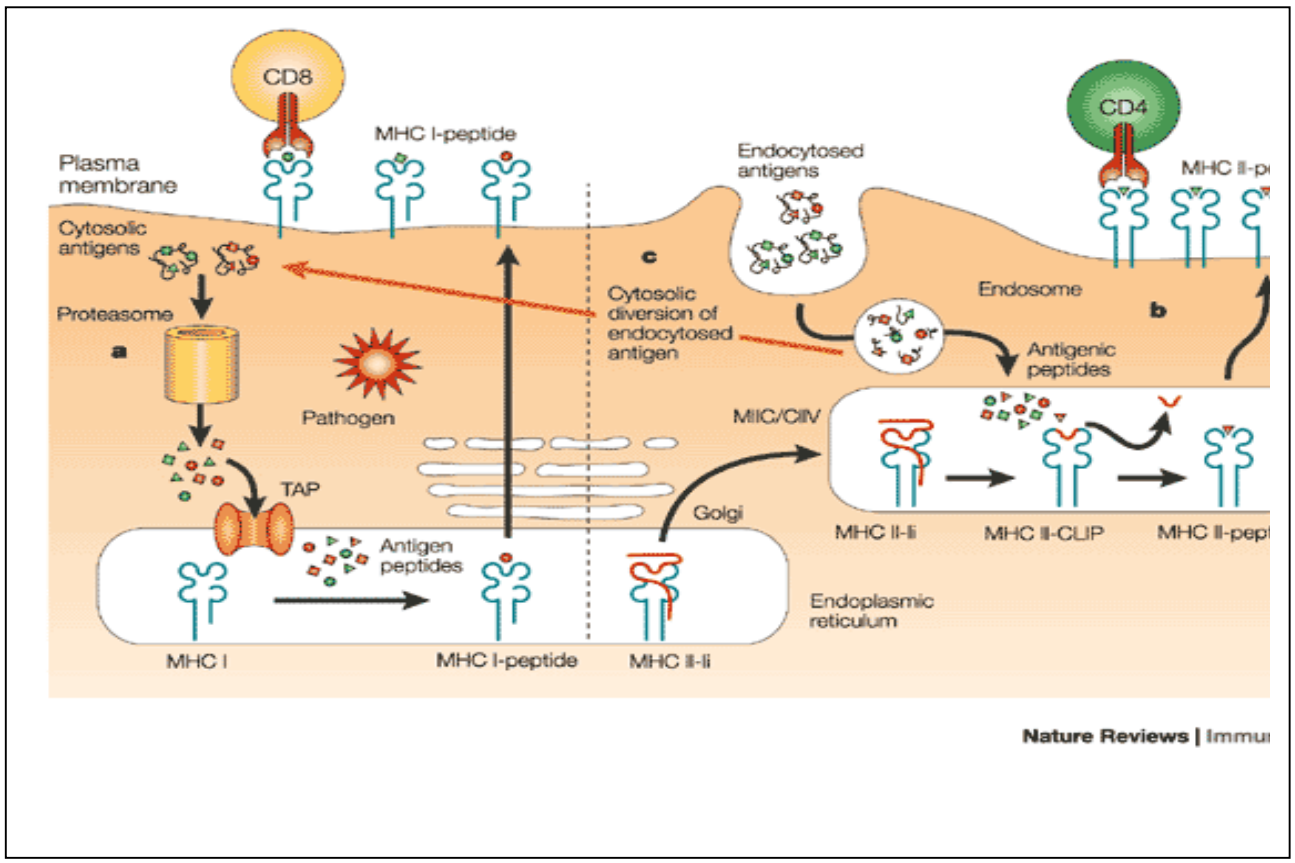
This is the simplest way to disrupt the immune response. Many acute viruses, as well as the majority of persistent viruses, target immune cells. The best know example of a virus following this strategy is HIV (*Retroviridae*). A more common and equally threatening virus in this category is measles (*Paramyxoviridae*), which infects B and T lymphocytes, and APCs. Measles mortality is in fact caused by immunosupression, which results from inhibition of lymphocyte proliferation. Indeed, earlier in the 20th century measles infection was used to treat certain immune-mediated diseases, even though they didn't know what was going on or why it worked. Ebola virus (*Filoviridae*) also replicates in immune cells, including macrophages and dendritic cells.

Changes in antiviral responses

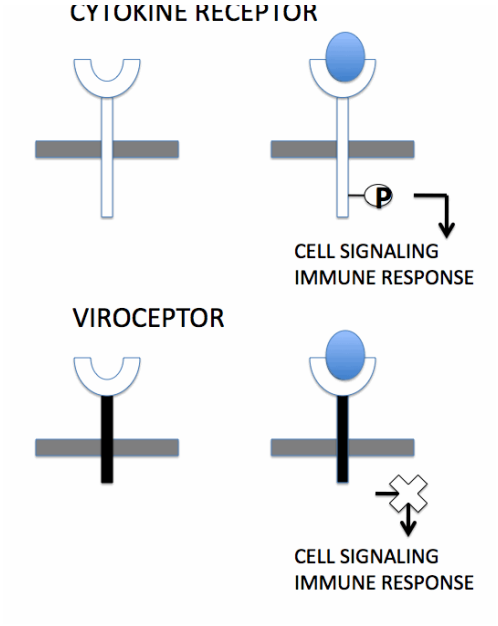
Thymic infection. The viral antigens in the infected cells will be presented by MHC and cause clonal deletion of virus-specific lymphocyte precursors, inducing tolerance.

Inhibition of antigen presentation. This is a favorite strategy for many of the big DNA viruses, like herpes virus, Epstein Barr virus or cytomegalovirus (all *Herpesviridae*). This strategy depends on the availability of virus-encoded proteins. HSV 1 has a protein, ICP47, that inhibits TAP binding and, thus, transport of peptides into the ER for MHC I presentation in he cell surface.

Antigen presentation by MHC I and MHC II



Cell-death delay. CD8+ cells clean virus-infected cells by induction of apoptosis. Several viruses produce homologues of proteins involved in cell-cycle regulations, such as bcl-2, p53 or Rb, thus extending the life of the infected cells until virions can mature and be released.



Cytokine inhibition. Also a favorite of big DNA viruses, which can produce cytokine homologues or cytokine receptor homologues (viroceptors). An example if the interferon receptor homologues, which bind interferon but do not get activated by binding.

Modification of the virus antigenicity

Down-regulation of gene expression

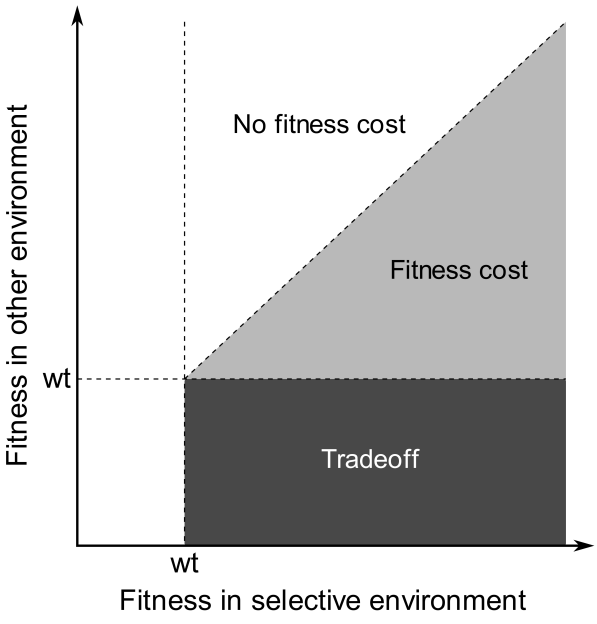
This is neither common nor very efficient, since very few MHC-antigen molecules can trigger the response. Viruses that become latent, like herpes, decrease gene expression to virtually nothing, avoiding antigen presentation.

Mutation

This is THE strategy for many, if not all, RNA viruses. Mutations can produce escape from clearance by antibodies or by CTLs. Some viruses can also accumulate mutations that affect interferon response.

Virulence and fitness

First a quick reminder that even though many use the terms *virulence* and *fitness* interchangeably (because they frequently do correlate) they are not the same. Also the number of progeny is often used as a measure of fitness, and while it is a component of fitness, the two are not the same either. In this section I will cover Ewald's concept that virulence may be selected for or against depending on environmental conditions.



Tradeoffs.

It is very hard to be able to do everything well. Developing a new skill or a new phenotypic traits are often complex processes that affect skills or phenotypes that we already have. For that reason adaptation to one environment may sometimes result in correlated phenotypic changes in a second environment. Sometimes the change in the second environment is a fitness increase similar or higher to that that would be observed if the virus us replicating in the second environment, in which case there is no fitness cost. Sometimes the fitness in the second environment will increase, but to a lesser extent than what can be observed during adaptation to the second environment. In this case we have a *cost.* There is loss of fitness in the second environment this is called *trade off*.

Virion stability and transmission.

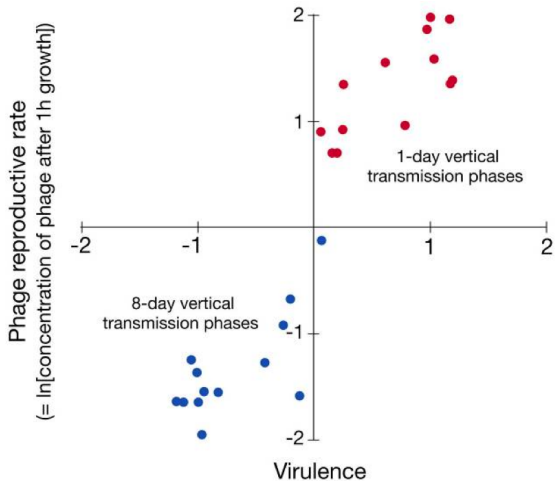
Fitness components and other phenotypic traits can be subjected to tradeoffs, and then successful transmission determines whether increase in any given trait, including virulence allows viral long-term survival. If the conditions are such that host damage improves the probability of transmission, then virulence will be selected for. Parameters that need consideration include the mode of transmission, the stability of the virus in the environment and the density of susceptible hosts.

The first goal is to avoid extinction. For that we will look at the reproductive number of R0 (R naught), which represents the average number of cases that result from a single infection. Viral strains with R0<1 go extinct, faster with smaller values of R0. If we have an infected host in an area where susceptible hosts are limiting and transmission depends on direct contact, a highly virulent strain will often kill its host before it can pass to the next host. If the viral particle is unstable in the environment, the R0 for such a strain will fall below the critical value of 1 and the strains will go extinct because it will lose infectivity before the next susceptible host comes along. In this case there will be selection for variants of reduced virulence. However, note all the "ifs" and "buts": if the virus is stable in the environment high virulence does not have a cost. If the density of susceptible hosts is high and virulence is due to high replication, virulence will be indirectly favored because higher or faster replication helps the virulent strains with chances of finding a new susceptible host. Note also that the availability of hosts can and will change through time resulting in changed in the strength of selection favoring or disfavoring virulence.

Other modes of transmission may shape differently virulence selection. For instance, vector-transmitted viruses are expected to increase virulence compared to viruses that need direct transmission because a debilitated infected host may be easier to bite in the latter case. Good experimental test of this hypothesis is currently lacking.

Vertical transmission, which occurs from a mother cell to its offspring, does not require the exit of virus from the infected cells. Vertical transmission is expected to select for lower virulence than horizontal transmission. Many insect and insect-transmitted viruses undergo vertical transmission when infected cells divide to produce eggs and horizontal transmission from insects to hosts (and sometimes among hosts). Arboviruses are typically avirulent in the insects, but more often virulent in the vertebrate host. The equivalent in the bacterial world is lysogeny, during which the phage genome is incorporated in the bacterial chromosome and transmission takes place during the replication of the bacterial chromosome. Jim Bull and coworkers showed that selection during lytic cycles promotes increased virulence, while selection during lysogeny results in decreased virulence.

Selection of virulence in phage f1



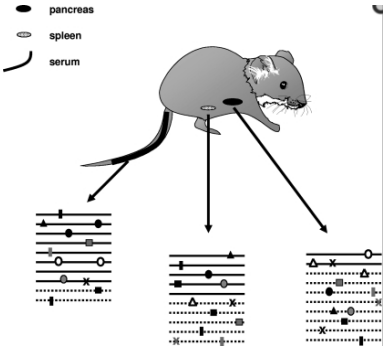
Sometimes virulence may have a direct positive effect on transmission and fitness. For instance, everything else being equal induction of diarrhea favors the release of gastrointestinal viruses to the soil and water. Another example is that of rabies virus and the behavioral changes it induces in the infected host to promote transmission by biting.

Ewald's theory is largely based on the existence of trade offs. However, as we will see in lecture 5 when we talk about environmental variation RNA viruses frequently show costs, not trade offs, during adaptation and it is important to understand the biology of a system.

Complex interactions during the expression of virulence

Another important aspect of virulence is the role of the population as an environmental factor. Sometimes we cannot predict the behavior of a population from sequence data because the observed phenotype is the result of interactions among the members of a population. Lymphocytic choriomeninitis virus (LCMV), foot-and-mouth disease virus (FMDV) and poliovirus provide some examples.

LCMV Armstrong strain does not cause GHD in mice. However, it is possible to isolate individual plaques ("clones") that, when grown *in vitro* and used to infect mice will cause GHD. The original LCMV-Armstrong harbors the mutant at a constant (~5%) rate. If mice are infected with mixtures of LCMV-Armstrong and the GHD variant the result is frequency dependent: in terms of fitness the variant is slightly defective, but in terms of virulence the variant only causes disease if present at a frequency of ~10% or higher in the mixture. This is not a problem of using infectious doses that are too low because disease is observed if the same absolute number of virions is used to inoculate mice in the absence of LCMV wild type. These results indicate that wt is suppressing the virulence of the mutant.



FMDV evolved during persistence generates variants that can kill mice. The infected animals have virus in different anatomical sites that, when recovered and used to inoculate a new animal, different degrees of virulence. The consensus sequences and levels of diversity are indistinguishable and it was proposed that the frequency of virulent genomes is the determinant of the observed phenotype.

The poliovirus Sabin vaccine is made of life-attenuated virus. The basis of attenuation is the incorporation of mutations during adaptation in cell culture that are maladaptive for human replication. The vaccine itself contains virulent variants at low-level. Furthermore, the vaccine virus often mutates and recombines in inoculated humans to produce revertants. However, while there are cases of vaccine-associated poliomyelitis, the frequency of diseases is orders of magnitude what one would expect based on the recovery of virulent wt in vaccines. Like in the case of LCMV, attenuated poliovirus masks the virulence of wt poliovirus. Protection is partially mediated by innate immunity (interferon).

Figure Credits

IRES structures: "Viral Replication", book edited by German Rosas-Acosta, ISBN 978-953-51-1055-2, Published: February 27, 2013 under CC BY 3.0 license. © The Author(s). Chapter 5 Viral Replication Strategies: Manipulation of ER Stress Response Pathways and Promotion of IRES-Dependent Translation. Paul J. Hanson, Huifang M. Zhang, Maged Gomaa Hemida, Xin Ye, Ye Qiu and Decheng Yang

IRES mutations: Impaired Binding of Standard Initiation Factors Mediates Poliovirus Translation Attenuation. Kerstin Ochs, Amandus Zeller, Lanja Saleh, Gergis Bassili, Yutong Song, Anja Sonntag and Michael Niepmann. J. Virol. January 2003 vol. 77 no. 1 115-122

Cap-dependent and cap-independent translation initiation: http://puglisi.stanford.edu/research.html.

FMDV Site A importance: Esteban Domingo Julie Sheldon and Celia Perales. Microbiol. Mol. Biol. Rev. June 2012 vol. 76 no. 2 159-216.

Genomic structure HIV: Architecture and secondary structure of an entire HIV-1 RNA genome. Nature 460, 711-716 (6 August 2009) Joseph M. Watts, Kristen K. Dang, Robert J. Gorelick, Christopher W. Leonard, Julian W. Bess Jr, Ronald Swanstrom, Christina L. Burch & Kevin M. Weeks.

Genomic structure HCV: You Li, Daisuke Yamane, Takahiro Masaki & Stanley M. LemonNat Rev Microbiol 13(9) 544–558 (2015)

Rule of six: Front. Microbiol., 12 May 2015 http://dx.doi.org/10.3389/fmicb.2015.00459. The paramyxovirus polymerase complex as a target for next-generation anti-paramyxovirus therapeutics. Robert Cox and Richard K. Plemper

Encapsidation signals HIV: Front. Microbiol., 29 December 2011 http://dx.doi.org/10.3389/fmicb.2011.00264. Structural dynamics of retroviral genome and the packaging. Yasuyuki Miyazaki, Ariko Miyake, Masako Nomaguchi and Akio Adachi

Poliovirus cycle: Host and virus determinants of picornavirus pathogenesis and tropism. J. Lindsay Whitton, Christopher T. Cornell & Ralph Feuer. Nature Reviews Microbiology 3, 765-776 (October 2005).

VSV cycle: Biochemistry, Genetics and Molecular Biology » "Methylation - From DNA, RNA and Histones to Diseases and Treatment", book edited by Anica Dricu, ISBN 978-953-51-0881-8, Published: November 28, 2012 under CC BY 3.0 license. © The Author(s). Chapter 10. Messenger RNA Cap Methylation in Vesicular Stomatitis Virus, a Prototype of Non-Segmented Negative-Sense RNA Virus. By Jianrong Li and Yu Zhang

Pathogenesis of high-fidelity poliovirus: Lauring and Andino. Quasispecies theory and the behavior of RNA viruses. PLoS Pathogens 6(7): e1001005. doi:10.1371/journal.ppat.1001005

Selection of virulence: Messenger, Molineux & Bull. Virulence evolution in a virus obeys a trade-off. Proc Biol Sci. 1999 Feb 22; 266(1417): 397–404.

Complex virulence determinants in FMDV: J Virol. 2008 Nov;82(21):10465-76. doi: 10.1128/JVI.00825-08. Epub 2008 Aug 20. Hidden virulence determinants in a viral quasispecies in vivo. Sanz-Ramos M1, Díaz-San Segundo F, Escarmís C, Domingo E, Sevilla N.