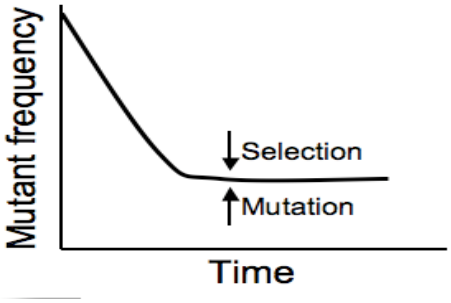
Quasispecies theory

The model

Mutation-selection balance



Quasispecies theory is a particular case of mutation-selection balance. Mutation-selection balance represents the situation when the loss of mutant due to selection is counteracted by the generation of mutant from replication errors of wild type genomes.

Quasispecies theory applies to error-prone replicons, such as RNA viruses. However, high heterogeneity is not sufficient to form a quasispecies. The theory predicts group selection based on robustness.

The basic model was developed assuming an infinite population size, but later adjusted for finite populations and other relevant factors such as complementation (see lecture #4). For those mathematically inclined who would like to see the basic details you may go to Eigen's original papers or to more recent papers:

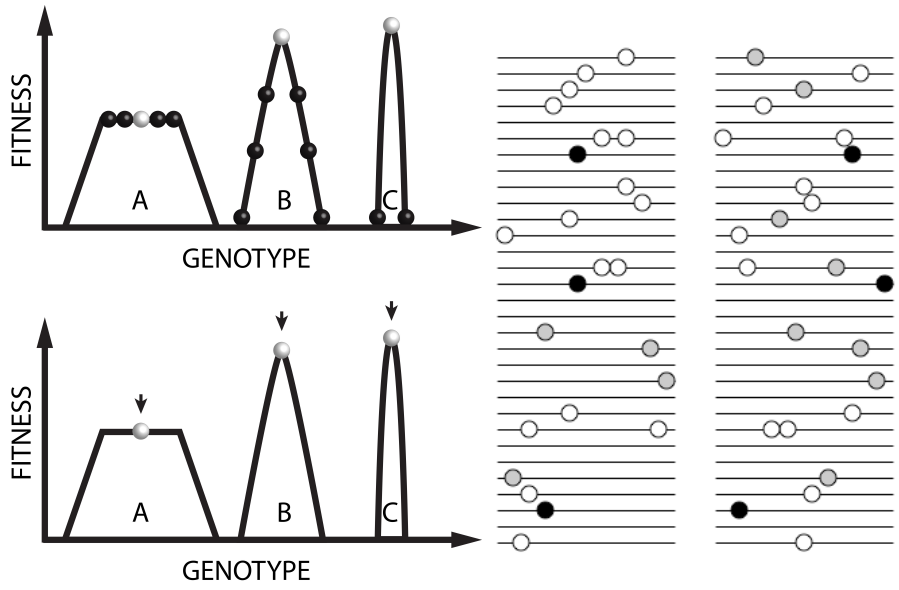
Eigen M, Schuster P (1979). The Hypercycle: A Principle of Natural Self-Organization. Berlin: Springer-Verlag. ISBN 0-387-09293-5.

Eigen, Manfred (1971). "Self organization of matter and the evolution of biological macromolecules". Die Naturwissenschaften 58 (10): 465–523. doi:10.1007/BF00623322. PMID 4942363.

Claus O Wilke. Quasispecies theory in the context of population genetics. BMC Evolutionary Biology 20055:44 DOI: 10.1186/1471-2148-5-44

Important concepts:

(Genetic) Robustness is the ability of a genome to withstand the deleterious effects of mutations.



Survival of the flattest
Survival of the fittest

Note that while we worry about deleterious mutations most, lower robustness can also be the result of a higher beneficial mutation rate.

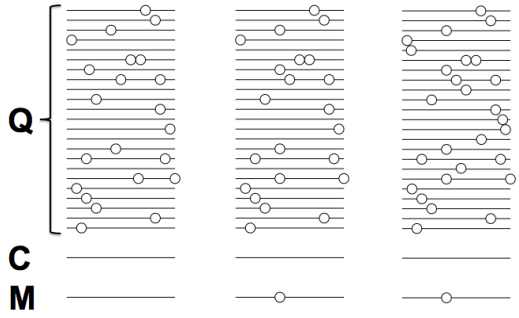
While RNA viruses as a group are under selection for high robustness not all species are equally robust and some are relatively brittle.

Consensus sequence is the sequence that contains the average nucleotide at any given position in a population.

Master sequence is the most common sequence in a population.

Error threshold is the mutation rate at which a quasispecies cannot be formed.

Mutation rate is the number or errors introduced during replication. Beware of double use of the term to describe speed of fixation!



Mutant frequency is the number of mutants relative to total virus in a replicating population. It is the result of selection after replicating with a given mutation rate.

Historical perspective

The first evidence that quasispecies theory is a good descriptor of RNA viruses came from work with phage Qβ. Domingo et al demonstrated that, contrary to the expectation at the time of constant sequences, T1 fingerprinting showed an average of 1 mutation per phage genome. The mutant variants decreased in frequency in competition with wt at a rate determined by their fitness and mutation rate.

Later on Holland's group described the loss of a high fitness mutant of VSV in competition with wt when the mutant was present in the population below a certain frequency. Note that this is not traditional frequency-dependent selection, because in that one would expect the opposite result (e.g. that the lower-fitness wt does better at lower frequencies). The described outcomes are the result of a combination of differences in robustness and bottleneck effects (see lecture 4).

Burch & Chao published analogous results using phage phi-6: viruses of similar fitness but different origins show different evolvabilty because of different beneficial/neutral/deleterious mutation rates produce different fitness distributions. Thus evolvability depends on robustness.

The cost of robustness

None to speak of under constant conditions. However, in a changing environment where there will be a need for adaptation selection cannot operate in the absence of phenotypic variation. Thus, increased robustness when it happens at the expense of beneficial variation may have negative consequences in the long-term because of insufficient adaptability.

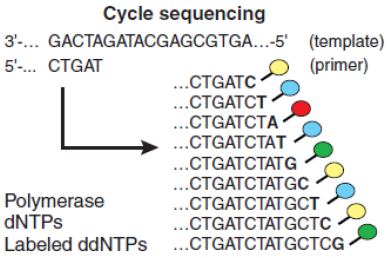
But that is the theory and in practice the two are usually correlated positively. Some theoretical studies have explained this observation based on the consideration of phenotypic robustness instead of genotypic robustness. In the case of the former the tolerance to the development of neutral networks allows the organism to access untapped combinations of mutations that otherwise would be hard to produce.

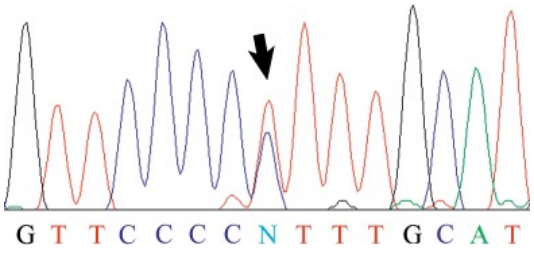
Robustness is not necessarily constant throughout the genome and some species have specific sequences that are very brittle due to selection (probably). One example is influenza, whose sequences coding for antigenic sites have an excess of non-synonymous sites that ensure changes in the external glycoproteins that will drive immune escape. While studies are lacking one would expect the same in other viral species that display evolution driven by immune selection.

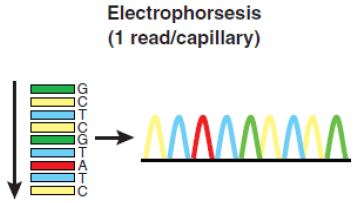
Quasispecies sequences

Because of the small size of RNA virus genomes it is easy to obtain full-length sequences and this facilitates some types of analysis. However, deciding what to sequence not trivial and different methods address different questions. The majority of sequencing data in the literature represent consensus sequence: the average nucleotide in the population. This type of information helps to learn about components of fitness. Technical note: don't forget the termini!

Sanger sequencing methods can generate more information that the consensus. Visual inspection of chromatograms can help with the identification of polymorphic sites as long as the frequency of the minority nucleotide is above 5-10%.







Notes:

1. Beware of background
2. Beware of sequencing error (need multiple readings)

The consensus doesn't tell us anything about the vast majority of the mutants in the population. Because mutant frequency at equilibrium depends on mutation rate, which is very high, we are likely missing data for the majority of the variants: we do not know how polymorphic the population is or which variants make up the quasispecies.

These questions can be addressed sequencing biological clones or molecular clones. Both methods have the limitation of numbers (how many individual can you check).

Biological clones are virus variants obtained by plaque-purification, limited dilution or a similar technique. All variant are functional virus.

Molecular clones are DNA variants obtained by RT-PCR and cloned into a plasmid. Variants may be functional of they may be dead virus. The limitations are the error rate of RT and amplification and the limitations looking at mutation linkage in distant sites.

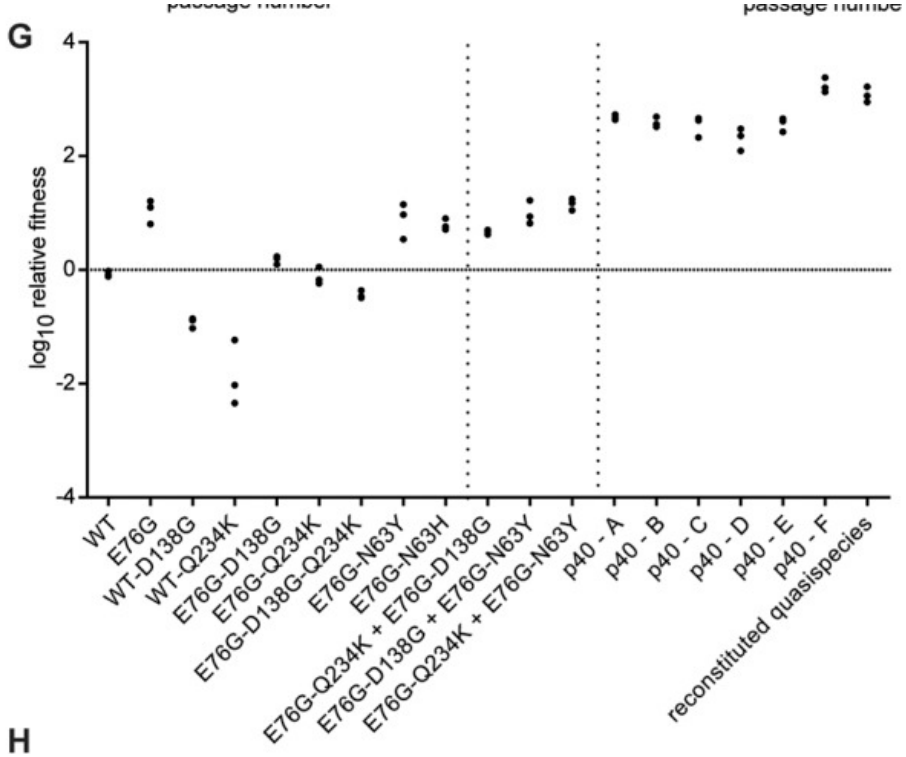
New techniques such as deep sequencing are opening new ways to address the effect of population structure in adaptation because they produce highly accurate mutant frequencies down to values around 10^{-5}.

Fitness contributions

Once the mutations in a population have been identified, the standard follow up would be the use of reverse genetics to generate individual mutants and combination of mutations to assess their fitness effects. Not much has been done because once the number of mutations go above a small handful the possibility of different combinations of mutations make the analysis unwieldy. This type of analysis, when feasible, also help to identify epistatic interactions.

Technical note: because reverse genetics involve the use of PCR amplification and recovery of virus in cell culture it is important to sequence the recovered mutant to ensure that no undesired mutations have been incorporated.

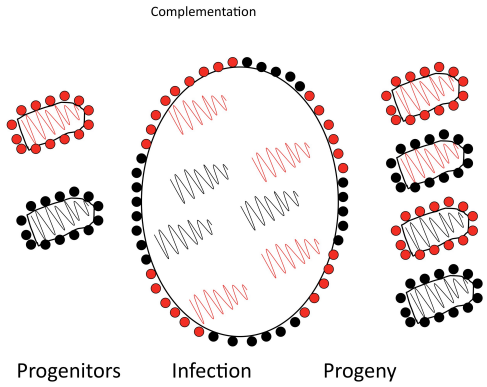
Individual mutations do not always explain a phenotype. Vignuzzi's lab demonstrated that coxsackievirus adaptation to different cell types may follow different strategies. Sometimes, mutation(s) modify specific viral functions that increase their fitness, and this is the observation noted during adaptation to HeLa cells. However, adaptation to A549 cells not only required the presence of favorable mutation, but also the reconstitution of a population with minor polymorphisms at specific sites.

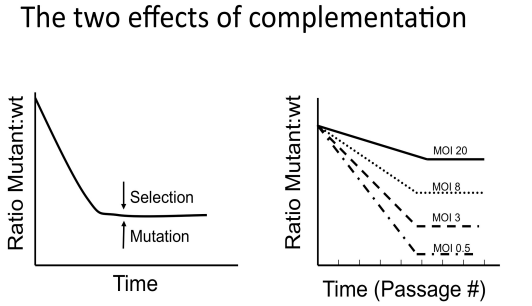


Often, but not always, uncontrolled virus replication in human viruses correlates with increased virulence and an earlier increase in the polymorphism of the population. However, remember that correlation is not the same as causation.

Mechanisms of group selection: complementation

Complementation occurs when a viral function is externally provided. When the function cannot be externally provided it is called *cis-acting* (for instance, a promoter). When it can, it is called trans-acting (for instance, a polymerase). There are a number of ways to provide a trans-acting function, but we will only consider here what happens when one virus complements another virus.



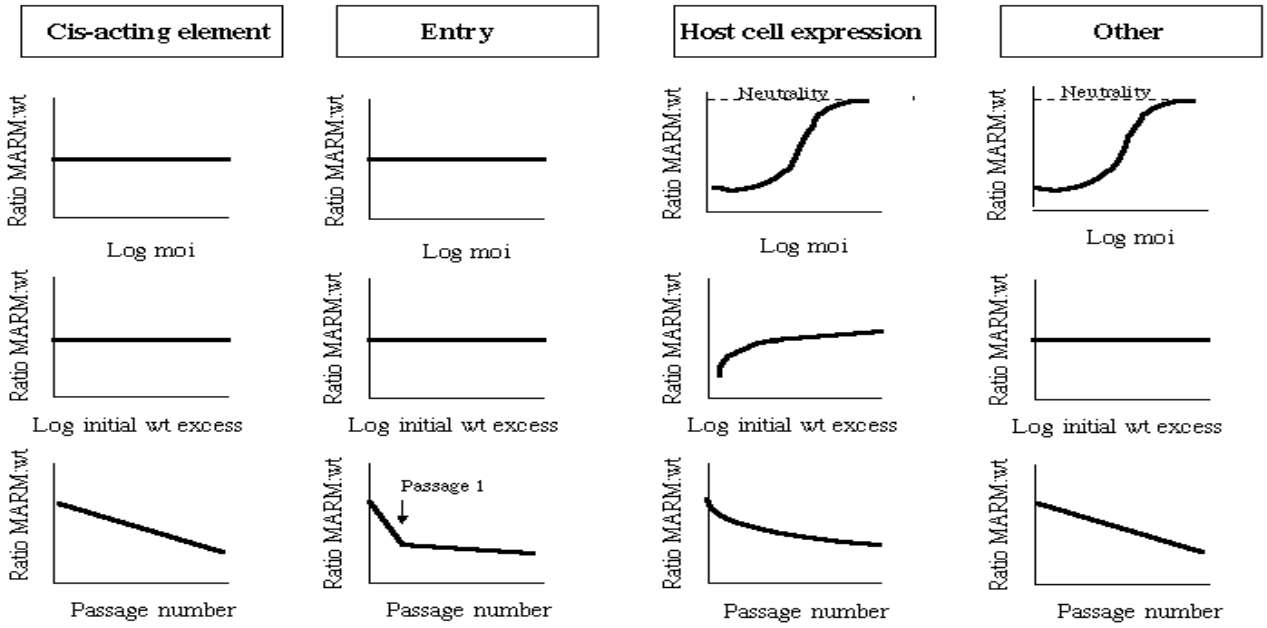


(1) Changes the speed of mutant loss.
(2) Changes the frequency of mutant at mutation-selection balance.

Evidence that complementation contributes to the fitness of the population first came from determinations of fitness distribution that showed that the average fitness of the population as a whole is larger than the average fitness of its individual component. Technical note: be careful when using plaque size as a surrogate of fitness.

A particular type of trans-complementation is phenotypic mixing and hiding. It occurs when during coinfection the external proteins of one virus enclose the genome of another virus. This may lead to the expression of new phenotypes. For instance, an antibody sensitive strain can survive in the presence of serum if the envelope proteins are from an antibody resistant strain.

Complementation can also alter the dynamics of other types of mutants, such as reshuffled genomes in negative stranders. The specific dynamics of complementation depend on the mechanism or function that is under study and the distribution of replicating genomes in the cell. In infected hosts some functions can occur in the absence of coinfection (remember poliovirus!)



Confounding factors

As we saw in lecture 1 serial passage of virus at a high MOI (>1 PFU/cell) leads to a steady decline in the titer of infectious virus. This is caused by the accumulation of defective virus particles (DIPs). Competition between DIPS and helper virus fit to Red Queen dynamics during which the helper virus generates mutations that allow escape from interference and new DIPs continue to appear that once again inhibit helper virus replication.

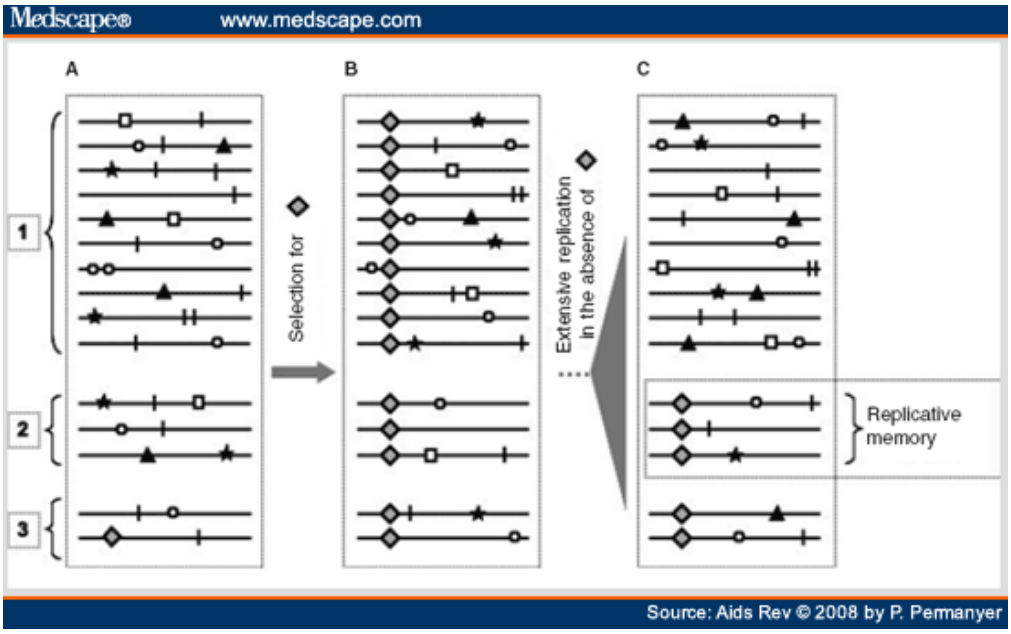
High MOI in a culture does not matter if the experiment is limited in time. In long-term experiments high MOI is a confounding factor for two reasons: first, the unexpected environmental changes under which the virus replicates and second the changes in bust size and, unless compensated for, the transmission population size (e.g. Ne changes).

Memory in viral quasispecies

Quasispecies memory represents a genetic record of the environmental pressures that a population has encountered and responded to and it has been extensively studies by Domingo's group looking at the evolution of FMDV in cell culture. Memory is observed as the presence of genomes that were dominant in past but were outcompeted after environmental changes at high frequency (e.g. higher than that expected by mutation-selection balance). Memory can be lost upon sufficiently extensive replication or in populations that go though bottlenecks and the dynamics are remarkably deterministic. Memory frequency depends on the initial fitness of the memory variant.

Memory genomes carry mutations in both cis-acting and trans-acting elements. In all cases maintenance depends on the concomitant fitness increase that the majority and minority variants undergo during evolution following red queen dynamics, but they are eventually lost (or their

frequency goes down to the expectation from mutation-selection balance) in agreement with the competitive exclusion principle. In the case of trans-acting mutations complementation also contributes to extended survival of the memory mutant.



Memory can facilitate the rapid adaptation of a population when it is reexposed to a past environment. For instance, quasispecies memory has also been found in HIV-infected patients and this is obviously relevant to the development of drug resistance. Furthermore, the HIV there are two sources of memory: genomes that have been integrated in the cell chromosome and are replicative inactive and virions that are replicating and behave as in the case of FMDV memory.

Genotypic evolution

Because of the mechanisms promoting the survival of minority mutants, either because of the biological processes that take place during replication and/or because of positive selection of complex populations, the term *fixation* does not necessarily mean that the deleterious variant has been fully outcompeted. The implicit assumption when people refer to fixed mutations is that the deleterious variant cannot be identified in the sequencing chromatograms. Nowadays with the more accurate pyrosequencing and next generation sequencing a mutation can be considered fixed when its frequency matches that predicted by mutation-selection balance (typically around 10^{-4} or 10^{-5} for mutations that do not depart too much from neutrality).

The presence of high heterogeneity in the population favors the observation of *leaping frogs*, which happen when a mutant is on its way to fixation and a more favorable mutation hits a genome that still has the wt allele in that locus but has mutations elsewhere. In this case a completely new set of mutations replaces a second set previously observed.

Nucleotide	64	1912	2109	2151	3224	3281	3846	3853	4210	7747	7773
Mutation	86 nts.	U→C	G→A	C→U	U→C	U→C	G→A	C→A	C→U	A→G	A→A+G
Region	3' Leader	P ORF	M ORF	M ORF	G ORF	G ORF	G ORF	G ORF	G ORF	L ORF	L ORF
Amino acid	NA	I→L	G→E	Y	H	I	E→A	A→D	P→L	S	L→Stop
PerA											
PerA+10											

Lethal mutagenesis

Lethal mutagenesis originated from the observation that mutation rates/site in RNA viruses can only be changed marginally. Unlike *E. coli*, for which mutator phenotypes can increase the mutation rate several orders of magnitude, ribovirus' rates cannot be increased more than a few-fold. Retroviruses, which have slightly lower mutation rates, can increase their mutation rate up to 1 order of magnitude. The reason is that the wt mutation rate is causing a mutational load at the limit of what the genomes can tolerate and keep their genetic information sufficiently intact. Thus, in the presence of RNA mutagens replication results in hypermutated genomes that are no longer viable.

Paradoxically the same high mutation rates that make lethal mutagenesis possible promote the evolution of high genetic robustness, which makes RNA viruses relatively resistant to extinction. It is not surprising then that this approach is not always successful and extinction seems to depend on the viral species. The outcome is more positive when the virus is forced through a genetic bottleneck so mutagens would be good candidates to use in cocktails with other antiviral drugs that can decrease viremia to produce the bottleneck. In the absence of bottlenecks and in species that do not reach extinction mutagenesis might have unintended consequences and generate variants with undesirable properties.

In some cases mutagenesis can result in the selection of resistant mutants. These have substitutions in the polymerases that increased their fidelity so the nucleotide is less frequently incorporated into the genome.

Because the target of lethal mutagenesis is the entire genome the consensus sequence often remains unchanged during the process, but when biological or molecular clones are characterize there is an increase in mutant frequency through time that reaches a maximum before extinction.

Increased genetic load is not the only mechanism promoting extinction and the last steps are independent of additional mutagenesis. When non-viable genomes reach a high frequency they start behaving like DIPs and interfering with the remaining viable genomes, taking up the functional proteins. Now the viable genomes, even if present, cannot replicate.

Figure Credits

Sanger sequencing: Med. Sci. 2014, 2(2), 98-126; doi:10.3390/medsci2020098. Application of Massively Parallel Sequencing in the Clinical Diagnostic Testing of Inherited Cardiac Conditions. Ivone U. S. Leong, Jonathan R. Skinner and Donald R. Love.

Multiple mutations vs. multiple mutants: Antonio V. Bordería, Ofer Isakov, Gonzalo Moratorio, Rasmus Henningsson, Sonia Agüera-González, Lindsey Organtini,6 Nina F. Gnädig, Hervé Blanc, Andrés Alcover, Susan Hafenstein, Magnus Fontes, Noam Shomron, and Marco Vignuzzi. PLoS Path. PLoS Pathog. 2015 May; 11(5): e1004838.

Quasispecies memory: Minority Report: Hidden Memory Genomes in HIV-1 Quasispecies and Possible Clinical Implications. Carlos Briones; Esteban Domingo. AIDS Rev. 2008;10(2):93-109.