# Survival Guide for Viruses

The dynamics of different species can be described by the following equation:

$$\frac{dN_i}{dt} = f_i N_i$$

$N_i$ – population of the $i$-th species

$f_i$ – fitness of the $i$-th species

In general, $f_i$ can depend on $N_i$ and $t$,

$$f_i = f_i(N_1, N_2, \cdots; t)$$

When analyzing evolutionary dynamics, it is convenient to introduce the relative population:

$$x_i \equiv N_i / N$$

$N$ – total population

From the definition, its easy to see that

$$\sum_i x_i = \frac{1}{N} \sum_i N_i = 1$$

The dynamics of $x_i$ after the "gauge" transformation is captured by the well-known replicator equation :)

# Replicator Equation

Let's derive the dynamics for $x_i(t)$ :

$$\frac{dx_i}{dt} = \frac{d}{dt}\left(\frac{N_i}{N}\right) = \frac{1}{N}\frac{dN_i}{dt} - \frac{N_i}{N^2}\frac{dN}{dt} \quad \left(-\sum_j \frac{dN_j}{dt}\right)$$

$$= \frac{1}{N}\cdot f_i N_i - \frac{N_i}{N^2}\sum_j f_j N_j$$

$$= f_i x_i - x_i \sum_j f_j x_j \quad \leftarrow \text{fitness of the ecosystem } \phi$$

$$\rightarrow \boxed{\frac{dx_i}{dt} = (f_i - \phi)\, x_i}$$

Replicator equation ☺

Note that the presence of $\phi$ makes the replicator eq. non-linear even if all fitness $f_i$ = const. Why?

## Zero-sum GAME 零和遊戲 ($\phi = 0$)

$$\frac{dN}{dt} = \sum_i \frac{dN_i}{dt} = \sum_i f_i N_i = \left(\sum_i f_i x_i\right) N$$

$$\rightarrow \frac{dN}{dt} = \phi N$$

If $\phi = 0$, the total population is constant.

# Fitness Landscape – important concept

Suppose a virus with binary genome of length $L$.

| | | | |
|---|---|---|---|
| wild type | 000000 | $f_0$ | |
| mutated | 010010 | $f_i$ | $i = 1, 2, \cdots, 2^L - 1$ |

Introduce the **Hamming distance $d_{ij}$** between two sequences $d_{ij} = d_{ji}$ by point mutations.

$$m_{ij} = u^{d_{ij}} (1-u)^{L-d_{ij}}$$

mutation matrix

One can plot the fitness $f_i$ versus the genome sequence → **fitness landscape**

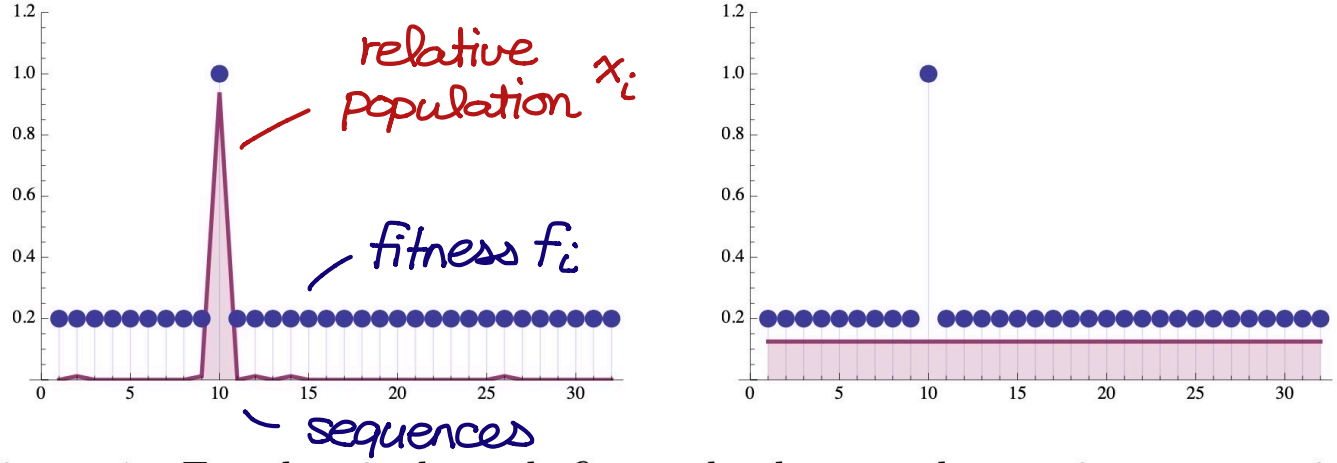



Figure 1: For the single-peak fitness landscape, there exists a mutation threshold $u_c$. For $u < u_c$, frequency profile is localized near the fitness peak. On the other hand, for $u > u_c$, an extended state (not necessarily uniform) emerges and the notion of quasi-species no longer exists.

# Quasi-species Equation

In the presence of mutation matrix $m_{ij}$, the replicator equation no longer holds. The viral population is described by the quasi-species eq.

$$\frac{dx_i}{dt} = \sum_j m_{ij} f_j x_j - \phi x_i$$

$m_{ij}$: mutations from sequence $j$ to sequence $i$

It is rather interesting that the QS eq. can be cast into quantum mechanical form (in the imaginary time) by the gauge transformation.

$$\Psi_i(t) = \sqrt{f_i}\, x_i(t)\, e^{W(t)}, \qquad \dot{W}(t) = \phi(t)$$

$$\frac{d\Psi_i}{dt} = \sqrt{f_i}\frac{dx_i}{dt} e^W + \sqrt{f_i} x_i \frac{dW}{dt} e^W$$

$$= e^W \left[ \sum_j \sqrt{f_i} m_{ij} f_j x_j - \cancel{\sqrt{f_i}\phi x_i} + \cancel{\sqrt{f_i} x_i \phi} \right]$$

$$\rightarrow \quad \frac{d\Psi_i}{dt} = \sum_j \sqrt{f_i f_j}\, m_{ij}\, \Psi_j$$

Introduce the Hamilton matrix $H_{ij}$

$$H_{ij} = -\sqrt{f_i f_j}\, m_{ij} \qquad H_{ij} = H_{ji}$$

So, the imaginary time Schrödinger eq. is

$$\boxed{\frac{d\Psi_i}{dt} = -\sum_j H_{ij}\Psi_j}$$

The dynamics can be constructed from the basis of the eigenvectors $|E_\alpha\rangle$ $\alpha = 0, 1, 2, \cdots, 2^L - 1$

If $|\Psi(t)\rangle$ is an eigenstate of $H$,

$$\frac{d|\Psi\rangle}{dt} = -H|\Psi\rangle = -E_\alpha|\Psi\rangle$$

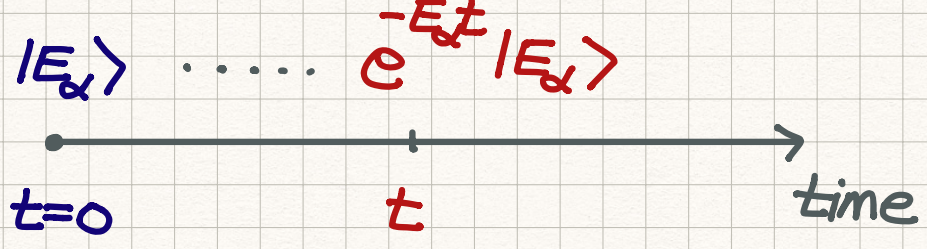
$$|\Psi(t)\rangle = |\Psi(0)\rangle e^{-E_\alpha t} = e^{-E_\alpha t}|E_\alpha\rangle$$

In general, one can decompose the initial state into eigenvectors,

$$|\Psi(0)\rangle = \sum_\alpha |E_\alpha\rangle\langle E_\alpha|\Psi(0)\rangle$$

$$= \sum_\alpha C_\alpha |E_\alpha\rangle$$

# Quantum dynamics of the eigenstates is simple




$$\rightarrow \quad |\Psi(t)\rangle = \sum_\alpha c_\alpha e^{-E_\alpha t} |E_\alpha\rangle$$

In the long-time limit,

$$|\Psi(t)\rangle \sim c_0 e^{-E_0 t} |E_0\rangle$$

determine the population $x_i$

only the ground state survives ☺

Inverse gauge transformation:

$$x_i = \frac{1}{\sqrt{F_i}} \Psi_i \, e^{-W(t)}$$

Because $\sum_i x_i = 1$

$$\left(\sum_i \frac{\Psi_i}{\sqrt{F_i}}\right) e^{-W} = 1 \quad \rightarrow \quad e^{W} = \sum_i \frac{\Psi_i}{\sqrt{F_i}}$$

We can now compute $x_i$ from $\Psi_i$:

$$x_i = \frac{\frac{1}{\sqrt{F_i}} \Psi_i}{\sum_j \frac{1}{\sqrt{F_j}} \Psi_j}$$

# How can a virus survive 🦠?

## Error threshold

$$u_c L = \ln\left(\frac{f_p}{f}\right)$$

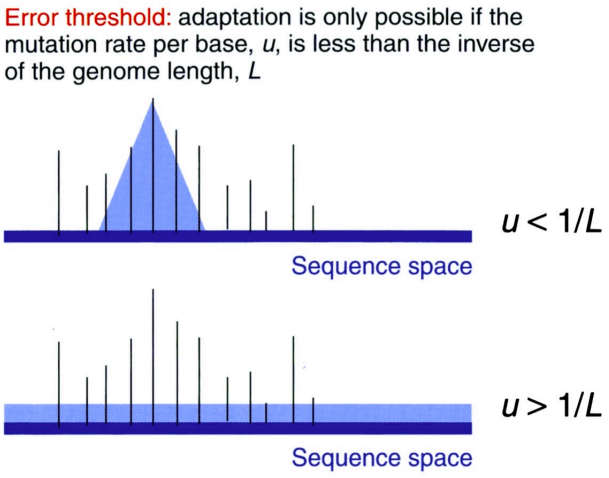


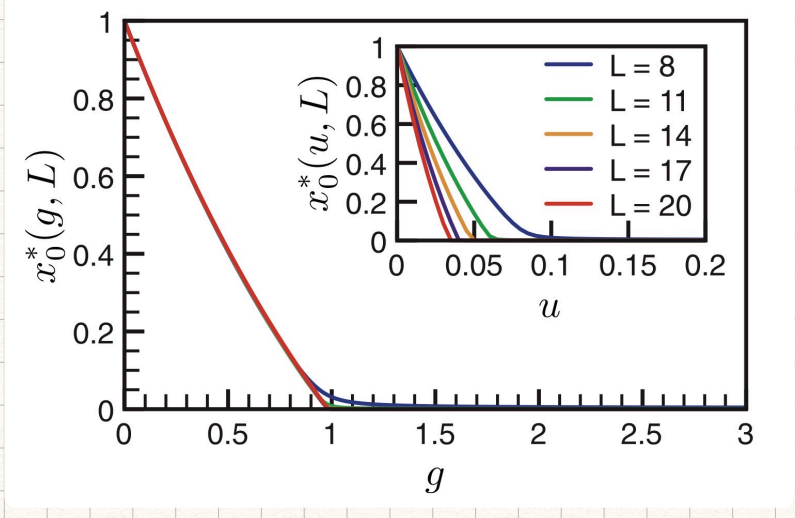

**Figure 3.6** Error threshold: a quasispecies can only maintain a peak in a fitness landscape if the mutation rate is less than the inverse of the genome length. This is a very general and beautiful result that must hold for any living organism. The beauty is not spoilt by two qualifying remarks that are necessary: (i) the genome length, $L$, has to be defined properly to include only those positions that affect fitness and (ii) there are some pathological landscapes where a peak can be maintained beyond the error threshold, for example if the peak is "infinitely" high or so wide that its presence can be felt by the majority of all possible sequences.

**Table 3.1** Genome length (in bases), mutation rate per base, and mutation rate per genome for organisms ranging from DNA viruses to humans

| Organism | Genome length in bases | Mutation rate per base | Mutation rate per genome |
|---|---|---|---|
| RNA viruses | | | |
| *Lytic viruses* | | | |
| Qβ | $4.2 \times 10^3$ | $1.5 \times 10^{-3}$ | 6.5 |
| Polio | $7.4 \times 10^3$ | $1.1 \times 10^{-4}$ | 0.84 |
| VSV | $1.1 \times 10^4$ | $3.2 \times 10^{-4}$ | 3.5 |
| Flu A | $1.4 \times 10^4$ | $7.3 \times 10^{-6}$ | 0.99 |
| *Retroviruses* | | | |
| SNV | $7.8 \times 10^3$ | $2.0 \times 10^{-5}$ | 0.16 |
| MuLV | $8.3 \times 10^3$ | $3.5 \times 10^{-6}$ | 0.029 |
| RSV | $9.3 \times 10^3$ | $4.6 \times 10^{-5}$ | 0.43 |
| Bacteriophages | | | |
| M13 | $6.4 \times 10^3$ | $7.2 \times 10^{-7}$ | 0.0046 |
| λ | $4.9 \times 10^4$ | $7.7 \times 10^{-8}$ | 0.0038 |
| T2 and T4 | $1.7 \times 10^5$ | $2.4 \times 10^{-8}$ | 0.0040 |
| *E. coli* | $4.6 \times 10^6$ | $5.4 \times 10^{-10}$ | 0.0025 |
| Yeast (*S. cerevisiae*) | $1.2 \times 10^7$ | $2.2 \times 10^{-10}$ | 0.0027 |
| *Drosophila* | $1.7 \times 10^8$ | $3.4 \times 10^{-10}$ | 0.058 |
| Mouse | $2.7 \times 10^9$ | $1.8 \times 10^{-10}$ | 0.49 |
| Human (*H. sapiens*) | $3.5 \times 10^9$ | $5.0 \times 10^{-11}$ | 0.16 |

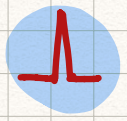
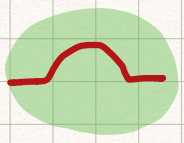
*Sources:* Drake (1991, 1993) and Drake et al. (1998).
*Note:* Most organisms have a mutation rate per genome which is less than one, as predicted by the error threshold theory. Why Qβ and VSV have such a high mutation rate is at present unexplained.

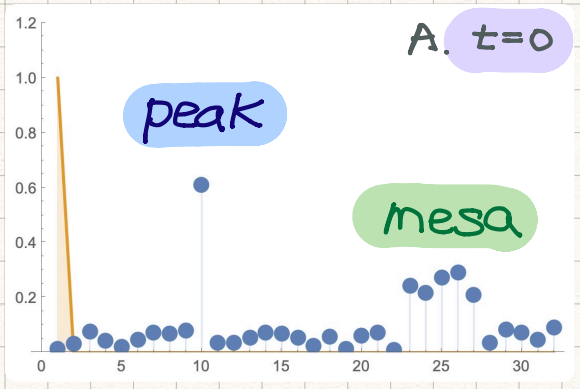
## Evolutionary Dynamics by M.A. Nowak

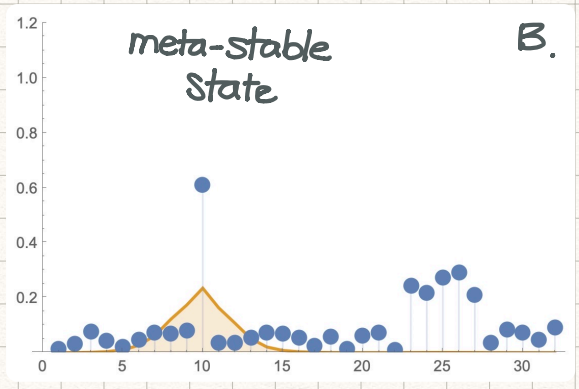


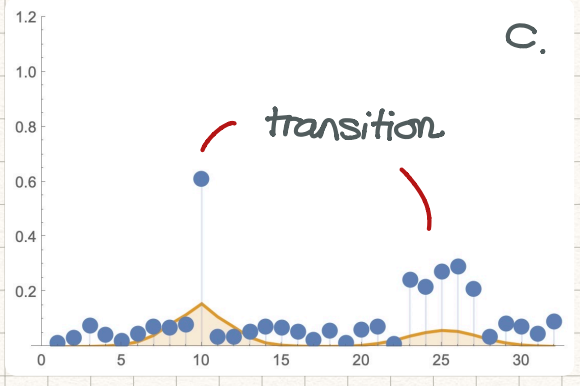

We can use the wild type population $x_0^*$ in the long-time limit as the "order parameter".

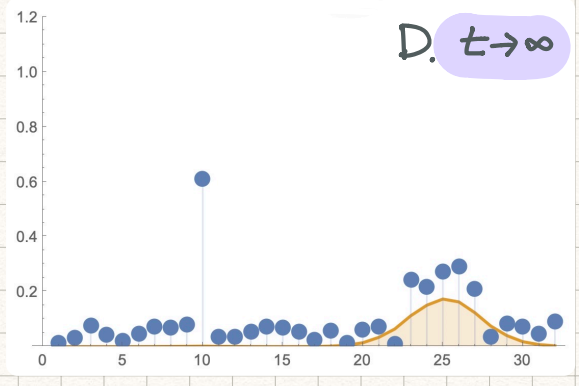
$$x_0^* = \lim_{t \to \infty} x_0(t)$$

# Competitions between  and  quasi-species




When the mutation rate is not-too-small yet not-too-large, the winner is NOT the QS with highest fitness. It is the QS with reasonable tolerance will dominate in the end.