

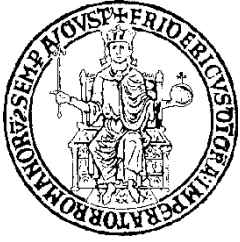
Population genetics and evolutionary dynamics I

Luca Peliti

Dipartimento di Scienze Fisiche and Unità INFN

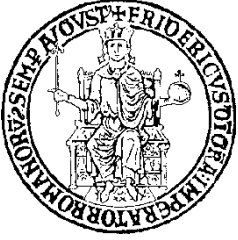
Università “Federico II” Napoli

INFN, Sezione di Napoli



Evolution and population genetics





Evolution and population genetics

The frequencies with which the genotypes occur define the gene ratios characteristic of the population, so that it is often convenient to consider a natural population not so much as an aggregate of individuals as an aggregate of gene ratios. Such a change of viewpoint is similar to that familiar in the theory of gases, where the specification of the population of velocities is often more useful than that of a population of particles.

Fisher, 1922



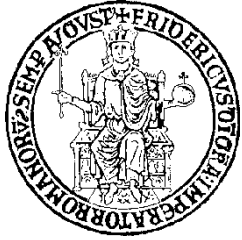
Evolution and population genetics

The frequencies with which the genotypes occur define the gene ratios characteristic of the population, so that it is often convenient to consider a natural population not so much as an aggregate of individuals as an aggregate of gene ratios. Such a change of viewpoint is similar to that familiar in the theory of gases, where the specification of the population of velocities is often more useful than that of a population of particles.

Fisher, 1922

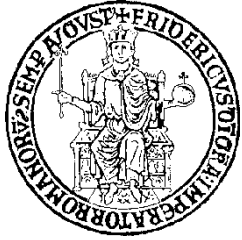
Evolution is a stochastic process of change in gene frequencies in natural populations. Since the populations making up a species consist of many individuals and since evolution extends over enormous periods of time, laws which govern the process of change are inevitably “statistical”. In this sense the genetical theory of evolution, as R. A. Fisher (1922) suggests, is comparable to the theory of gases.

Motoo Kimura, 1953



Darwinian mechanism

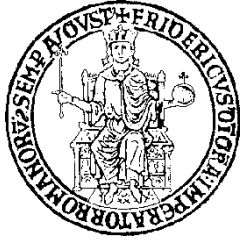
According to Darwin, evolution is the outcome of a process involving



Darwinian mechanism

According to Darwin, evolution is the outcome of a process involving

Reproduction: Like begets like

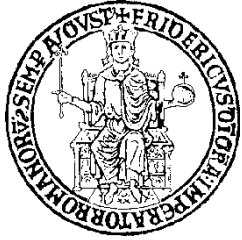


Darwinian mechanism

According to Darwin, evolution is the outcome of a process involving

Reproduction: Like begets like

Mutation: Offspring are not identical to parents



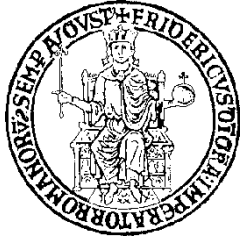
Darwinian mechanism

According to Darwin, evolution is the outcome of a process involving

Reproduction: Like begets like

Mutation: Offspring are not identical to parents

Selection: Inheritable changes may affect reproductive success



Darwinian mechanism

According to Darwin, evolution is the outcome of a process involving

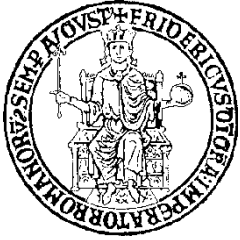
Reproduction: Like begets like

Mutation: Offspring are not identical to parents

Selection: Inheritable changes may affect reproductive
SUCCESS

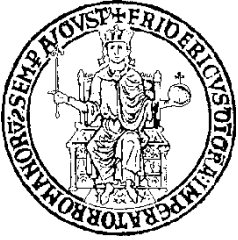
This preservation of favourable variations and the rejection of injurious variations, I call Natural Selection. Variations neither useful nor injurious would not be affected by natural selection, and would be left a fluctuating element, as perhaps we see in the species called polymorphic.

Charles Darwin, 1859



Fitness

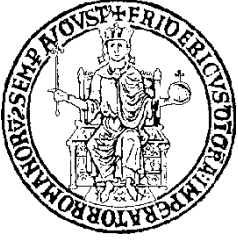
Fitness: A measure of the reproductive potential of a genotype



Fitness

Fitness: A measure of the reproductive potential of a genotype

Absolute fitness: W_G of a genotype G : *expected* number of offspring contributed by an individual of genotype G to the next generation



Fitness

Fitness: A measure of the reproductive potential of a genotype

Absolute fitness: W_G of a genotype G : *expected* number of offspring contributed by an individual of genotype G to the next generation

Relative fitness: Absolute fitnesses rescaled by an arbitrary factor



Fitness

Fitness: A measure of the reproductive potential of a genotype

Absolute fitness: W_G of a genotype G: *expected* number of offspring contributed by an individual of genotype G to the next generation

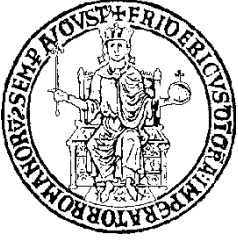
Relative fitness: Absolute fitnesses rescaled by an arbitrary factor

The concept of relative fitness allows one to separate population dynamics (increase or decrease in a population) from evolutionary dynamics, considered as the “change of gene frequencies in the population”



Observations on fitness

- ⑥ This definition holds for **nonoverlapping generations**: otherwise fitness is defined by the **Malthusian parameter**



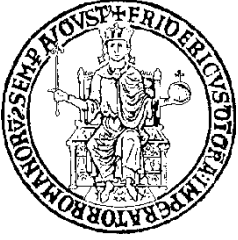
Observations on fitness

- ⑥ This definition holds for **nonoverlapping generations**: otherwise fitness is defined by the **Malthusian parameter**
- ⑥ Fitness is assigned to a class of individuals, e.g., those homozygous for allele A at a certain locus



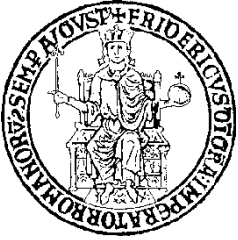
Observations on fitness

- ⑥ This definition holds for **nonoverlapping generations**: otherwise fitness is defined by the **Malthusian parameter**
- ⑥ Fitness is assigned to a class of individuals, e.g., those homozygous for allele A at a certain locus
- ⑥ The number of *actual* offspring of an individual is random: fitness only determines its *expected* number



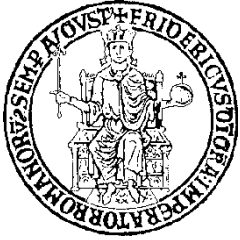
Observations on fitness

- ⑥ This definition holds for **nonoverlapping generations**: otherwise fitness is defined by the **Malthusian parameter**
- ⑥ Fitness is assigned to a class of individuals, e.g., those homozygous for allele A at a certain locus
- ⑥ The number of *actual* offspring of an individual is random: fitness only determines its *expected* number
- ⑥ Fitness depends on the environment



Observations on fitness

- ⑥ This definition holds for **nonoverlapping generations**: otherwise fitness is defined by the **Malthusian parameter**
- ⑥ Fitness is assigned to a class of individuals, e.g., those homozygous for allele A at a certain locus
- ⑥ The number of *actual* offspring of an individual is random: fitness only determines its *expected* number
- ⑥ Fitness depends on the environment
- ⑥ Populations do not have fitnesses



Sampling

Population of M^0 individuals, with N_G genotypes.

W_g : fitness of genotype g

Genotype distribution in the new population?



Sampling

Population of M^0 individuals, with N_G genotypes.

W_g : fitness of genotype g

Genotype distribution in the new population?

Assume:



Sampling

Population of M^0 individuals, with N_G genotypes.

W_g : fitness of genotype g

Genotype distribution in the new population?

Assume:

- (a) Haploid population with one-parent reproduction mechanism



Sampling

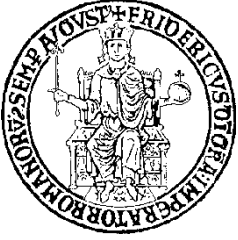
Population of M^0 individuals, with N_G genotypes.

W_g : fitness of genotype g

Genotype distribution in the new population?

Assume:

- (a) Haploid population with one-parent reproduction mechanism
- (b) Offspring and parent have the same genotype (no **mutations**)



Sampling

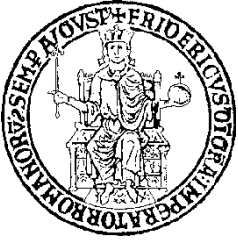
Population of M^0 individuals, with N_G genotypes.

W_g : fitness of genotype g

Genotype distribution in the new population?

Assume:

- (a) Haploid population with one-parent reproduction mechanism
- (b) Offspring and parent have the same genotype (no **mutations**)
- (c) Reproduction events are independent from one another



n_g^0 : number of individuals with genotype g in the old population

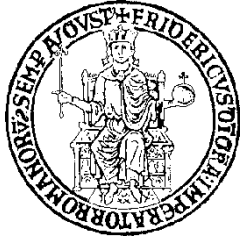
W_g : absolute fitness of genotype g

Expected number of individuals with genotype g in the next generation:

$$\nu_g = n_g^0 W_g$$

Probability that the actual number of individuals with genotype g equals n_g :

$$P_g(n_g) = \frac{e^{-\nu_g}}{n_g!} \nu_g^{n_g}$$

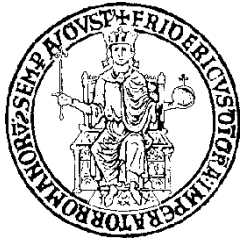


Wright-Fisher process

x_g^0 : fraction of individuals with genotype g

$$x_g^0 = \frac{n_g^0}{\sum_{g'} n_{g'}^0}$$

$$w_g = \frac{\nu_g}{\sum_{g'} \nu_{g'}}$$



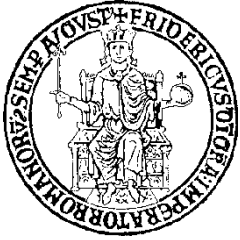
Wright-Fisher process

x_g^0 : fraction of individuals with genotype g

$$x_g^0 = \frac{n_g^0}{\sum_{g'} n_{g'}^0}$$

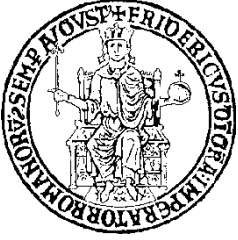
$$w_g = \frac{\nu_g}{\sum_{g'} \nu_{g'}} = \frac{W_g x_g^0}{\sum_{g'} W_{g'} x_{g'}^0}$$

$$\begin{aligned} P(n_1, \dots, n_{N_G} \mid \sum_g n_g = M) \\ = \frac{M!}{n_1! \dots n_{N_G}!} w_1^{n_1} \dots w_{N_G}^{n_{N_G}} \end{aligned}$$



Fisher's "Fundamental Theorem"

Assume $\nu_g \gg 1, \forall g$



Fisher's "Fundamental Theorem"

Assume $\nu_g \gg 1, \forall g$

$$x_g := \frac{n_g}{\sum_{g'} n_{g'}} \approx \frac{w_g}{\sum_{g'} w_{g'}}$$



Fisher's "Fundamental Theorem"

Assume $\nu_g \gg 1, \forall g$

$$x_g := \frac{n_g}{\sum_{g'} n_{g'}} \simeq \frac{w_g}{\sum_{g'} w_{g'}}$$

$$\langle W \rangle := \frac{\sum_g n_g W_g}{\sum_{g'} n_{g'}} = \sum_g x_g W_g$$

$$\langle W \rangle = \sum_g x_g W_g$$



Fisher's "Fundamental Theorem"

Assume $\nu_g \gg 1, \forall g$

$$x_g := \frac{n_g}{\sum_{g'} n_{g'}} \simeq \frac{w_g}{\sum_{g'} w_{g'}}$$

$$\langle W \rangle := \frac{\sum_g n_g W_g}{\sum_{g'} n_{g'}} = \sum_g x_g W_g$$

$$\langle W \rangle = \sum_g x_g W_g = \sum_g \frac{w_g}{\sum_{g'} w_{g'}} W_g$$



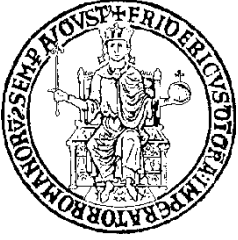
Fisher's "Fundamental Theorem"

Assume $\nu_g \gg 1, \forall g$

$$x_g := \frac{n_g}{\sum_{g'} n_{g'}} \simeq \frac{w_g}{\sum_{g'} w_{g'}}$$

$$\langle W \rangle := \frac{\sum_g n_g W_g}{\sum_{g'} n_{g'}} = \sum_g x_g W_g$$

$$\begin{aligned} \langle W \rangle &= \sum_g x_g W_g = \sum_g \frac{w_g}{\sum_{g'} w_{g'}} W_g \\ &= \frac{\sum_g x_g^0 W_g^2}{\sum_{g'} x_{g'}^0 W_{g'}} \end{aligned}$$



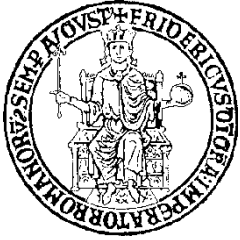
Fisher's "Fundamental Theorem"

Assume $\nu_g \gg 1, \forall g$

$$x_g := \frac{n_g}{\sum_{g'} n_{g'}} \simeq \frac{w_g}{\sum_{g'} w_{g'}}$$

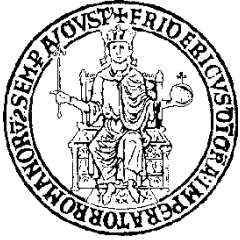
$$\langle W \rangle := \frac{\sum_g n_g W_g}{\sum_{g'} n_{g'}} = \sum_g x_g W_g$$

$$\begin{aligned} \langle W \rangle &= \sum_g x_g W_g = \sum_g \frac{w_g}{\sum_{g'} w_{g'}} W_g \\ &= \frac{\sum_g x_g^0 W_g^2}{\sum_{g'} x_{g'}^0 W_{g'}} = \frac{\langle W^2 \rangle_0}{\langle W \rangle_0} \geq \langle W \rangle_0 \end{aligned}$$



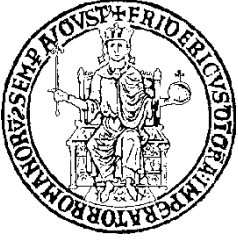
Fitness vs. entropy





Fitness vs. entropy

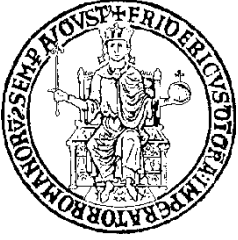
It will be noticed that the fundamental theorem proved above bears some remarkable resemblances to the second law of thermodynamics. . . Both are statistical laws; each requires the constant increase of a measurable quantity. . .



Fitness vs. entropy

It will be noticed that the fundamental theorem proved above bears some remarkable resemblances to the second law of thermodynamics. . . Both are statistical laws; each requires the constant increase of a measurable quantity. . .

. . . entropy changes lead to a progressive disorganization of the physical world, at least from the human standpoint of the utilization of energy, while evolutionary changes are generally recognized as producing progressively higher organization in the organic world.



Fitness vs. entropy

It will be noticed that the fundamental theorem proved above bears some remarkable resemblances to the second law of thermodynamics. . . Both are statistical laws; each requires the constant increase of a measurable quantity. . .

. . . entropy changes lead to a progressive disorganization of the physical world, at least from the human standpoint of the utilization of energy, while evolutionary changes are generally recognized as producing progressively higher organization in the organic world.

Fisher, 1930



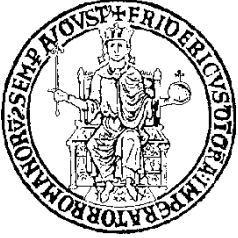
Fitness vs. entropy

It will be noticed that the fundamental theorem proved above bears some remarkable resemblances to the second law of thermodynamics. . . Both are statistical laws; each requires the constant increase of a measurable quantity. . .

. . . entropy changes lead to a progressive disorganization of the physical world, at least from the human standpoint of the utilization of energy, while evolutionary changes are generally recognized as producing progressively higher organization in the organic world.

Fisher, 1930

So we see, in physics, disorder growing inexorably in systems isolated from their surroundings; and in biology, fitness increasing steadily in populations struggling for life. Ascent here and degradation there—almost too good to be true.



Fitness vs. entropy

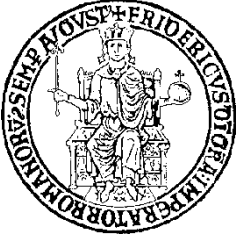
It will be noticed that the fundamental theorem proved above bears some remarkable resemblances to the second law of thermodynamics. . . Both are statistical laws; each requires the constant increase of a measurable quantity. . .

. . . entropy changes lead to a progressive disorganization of the physical world, at least from the human standpoint of the utilization of energy, while evolutionary changes are generally recognized as producing progressively higher organization in the organic world.

Fisher, 1930

So we see, in physics, disorder growing inexorably in systems isolated from their surroundings; and in biology, fitness increasing steadily in populations struggling for life. Ascent here and degradation there—almost too good to be true.

K. Sigmund, 1993

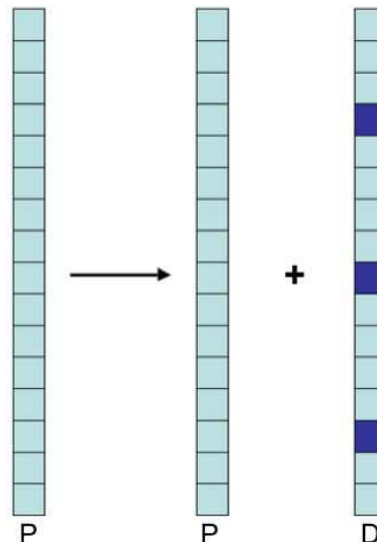


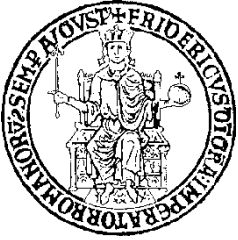
Mutations

Mutation: The offspring's genome is different from the parent's one

Substitution: The *population* genome is different from the ancestral one

Substitutions take place by *fixation* of mutations



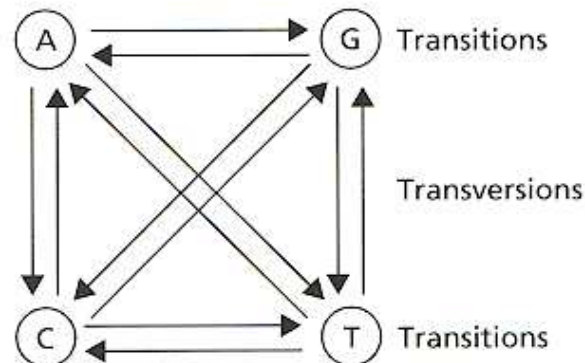


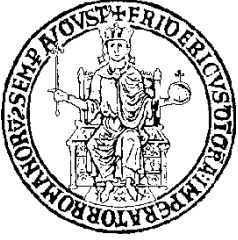
Mutation models

- ⑥ In classical population theory, one considers mutation probabilities for the alleles:

$$a_i \xleftarrow{p_{a'/a}} a_i$$

- ⑥ Modern theories consider mutations at the level of the genome:





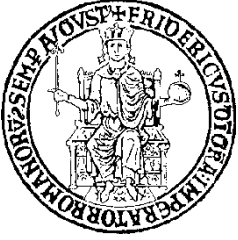
Observations

- ⑥ Possibility of insertions and deletions (*indels*)



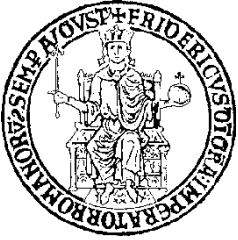
Observations

- ⑥ Possibility of insertions and deletions (*indels*)
- ⑥ Major reshufflings can take place



Observations

- ⑥ Possibility of insertions and deletions (*indels*)
- ⑥ Major reshufflings can take place
- ⑥ One assumes *uniformity* (mutation rates are the same at each site)



Observations

- ⑥ Possibility of insertions and deletions (*indels*)
- ⑥ Major reshufflings can take place
- ⑥ One assumes *uniformity* (mutation rates are the same at each site)
- ⑥ One assumes *stationarity* (*nucleotide frequencies* do not change with time)



Observations

- ⑥ Possibility of insertions and deletions (*indels*)
- ⑥ Major reshufflings can take place
- ⑥ One assumes *uniformity* (mutation rates are the same at each site)
- ⑥ One assumes *stationarity* (*nucleotide frequencies* do not change with time)
- ⑥ These assumptions are **wrong!**



Basic models

Jukes-Cantor: All mutation rates are the same

Kimura's two-parameter model (K2P): Transition rate \neq transversion rate (transitions are *more frequent* than transversions)

General reversible model: All six *pairs* of substitutions have different rates

N.B.: Data compatible with uniformity and stationarity assumption only determine *five* parameters (*J. Lobry and O. Zagordi*)



Silent and nonsilent mutations

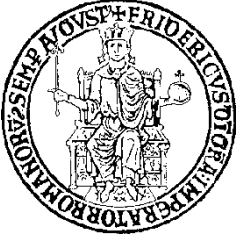
Silent mutations: Nucleotide mutations that the expressed amino acid invariant (also called *synonymous* mutations)

Neutral aminoacid substitution: A substituted amino acid can be functionally equivalent to the wild type one

Both kinds of mutations are (believed to be) **selectively neutral**, i.e., they do not change the fitness

Other mutations are **selectively significant**

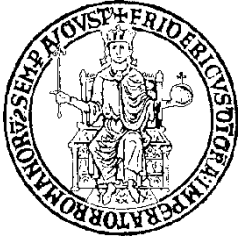
Mutation-selection (deterministic) models



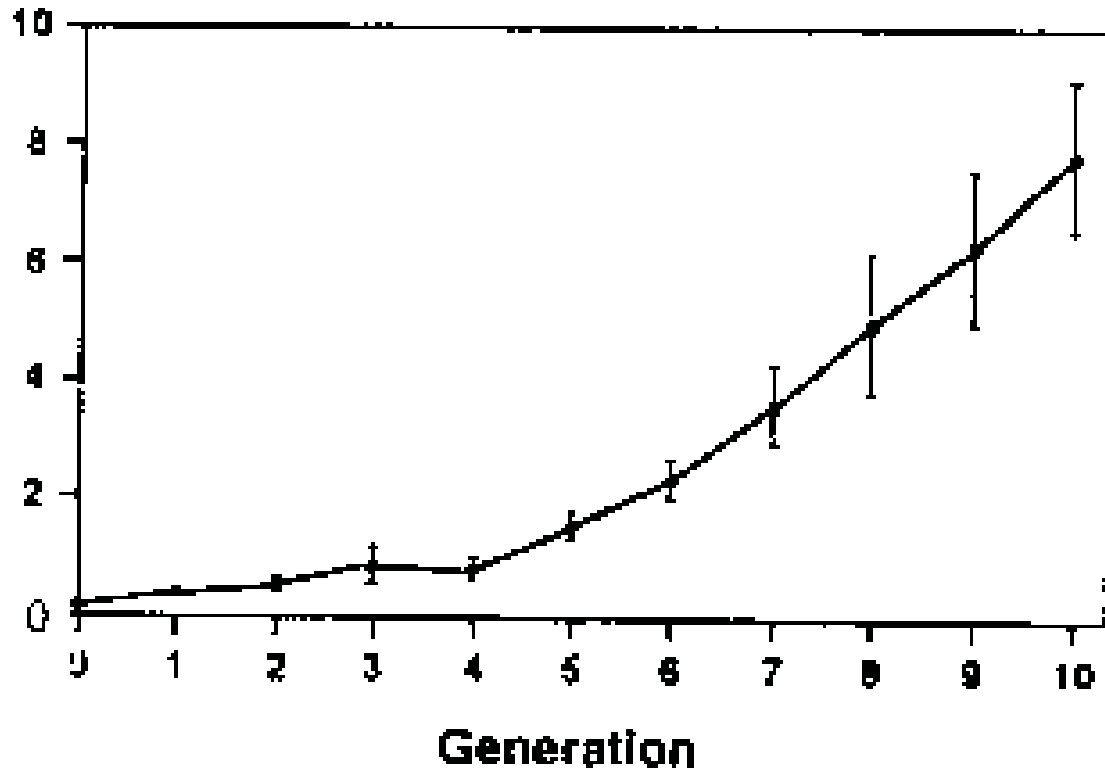
- ⑥ How do mutations qualify Fisher's theorem?
- ⑥ Mutation-selection (*quasispecies*) models are *deterministic* models that describe the interplay of selection and mutation
- ⑥ The usual setup is **sequence space**:

$$\{g\} := \{A, C, G, U\}^L$$

(sometimes reduced to a 2^L dimensional space)

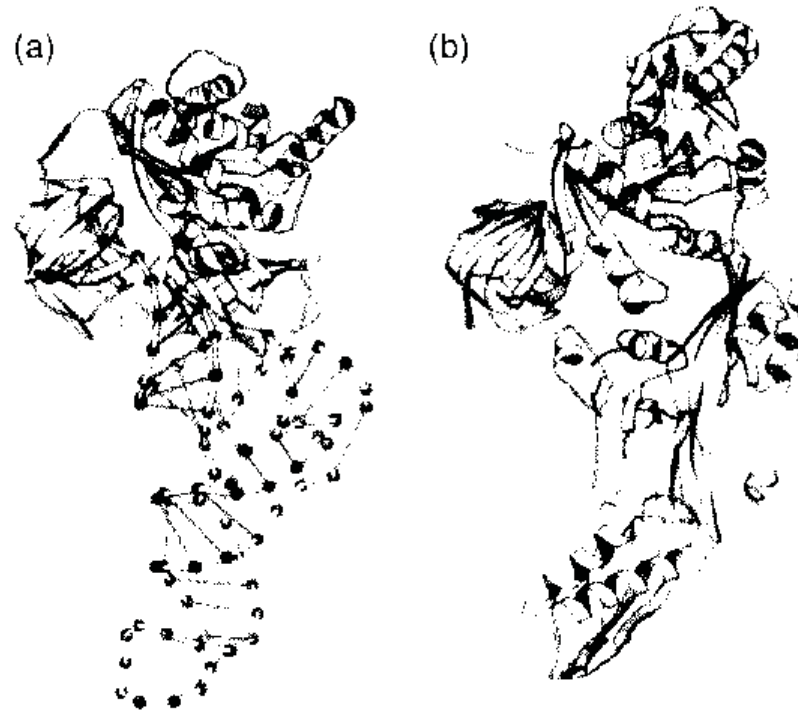


Enzymatic activity of selected RNA



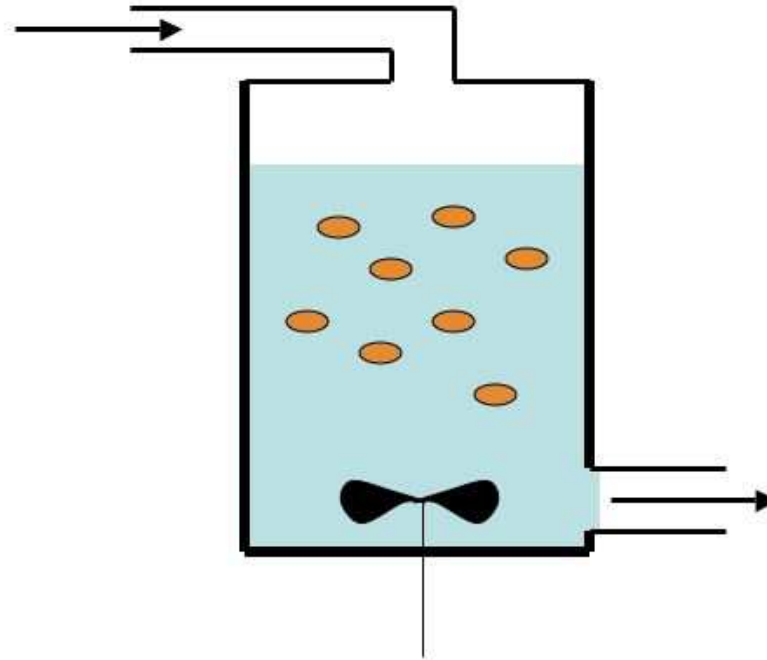
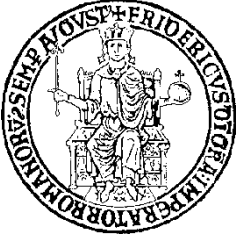
Baudry and Joyce, 1994

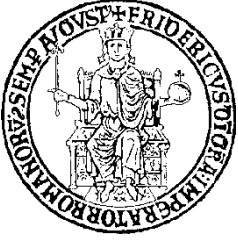
In vitro evolved RNA may mimic existing enzymes



Landweber, 1999

Chemostat: Evolution in continuous time

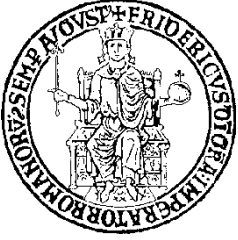




The Quasispecies (QS) model

Eigen, 1971

- ⑥ Direct RNA \longrightarrow RNA' reproduction
- ⑥ Point-mutation model by nucleotide substitution
- ⑥ Reproduction rate as fitness
- ⑥ Infinite population limit (this makes the dynamics deterministic)
- ⑥ Evolution takes place in **sequence space**



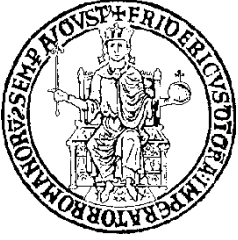
The QS model in discrete time

Assume **nonoverlapping generations**

Mutation matrix: $Q := (Q_{gg'})$: probability that a string g is generated in the attempt to reproduce a string g'

$$0 \leq Q_{gg'} \leq 1 \quad \sum_g Q_{gg'} = 1, \quad \forall g'$$

Fitness: $W_g \geq 0$ *proportional* to expected number of offspring of a string of genotype g

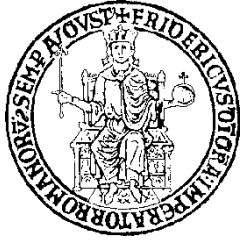


Evolution equation

Genotype fractions $x_g(t) = n_g(t) / \sum_{g'} n_{g'}$

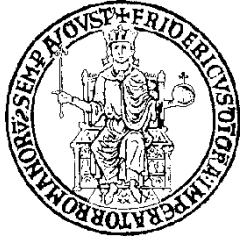
$$x_g(t+1) = \frac{1}{\langle W \rangle_t} \sum_{g'} Q_{gg'} W_{g'} x_{g'}(t)$$

$$\langle W \rangle_t = \sum_g W_g x_g(t)$$



Asymptotic behavior

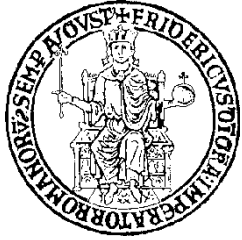




Asymptotic behavior

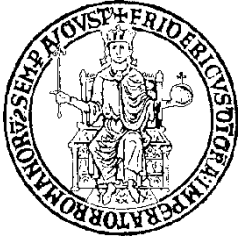


$$x_{\infty}^* = \lim_{t \rightarrow \infty} x_g(t)$$



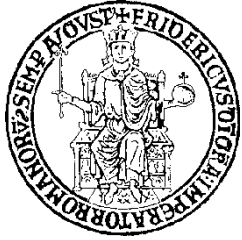
Asymptotic behavior

$$\begin{aligned}x_g^* &= \lim_{t \rightarrow \infty} x_g(t) \\y_g(0) &:= x_g(0), \quad \forall g \\y_g(t+1) &:= \sum_{g'} Q_{gg'} W_{g'} y_{g'}(t), \quad t \geq 0\end{aligned}$$



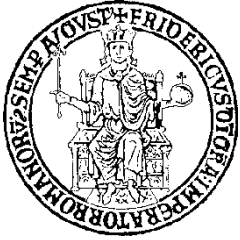
Asymptotic behavior

$$\begin{aligned}x_g^* &= \lim_{t \rightarrow \infty} x_g(t) \\y_g(0) &:= x_g(0), \quad \forall g \\y_g(t+1) &:= \sum_{g'} Q_{gg'} W_{g'} y_{g'}(t), \quad t \geq 0 \\x_g(t) &= \frac{y_g(t)}{\sum_{g'} y_{g'}(t)}, \quad \forall t\end{aligned}$$



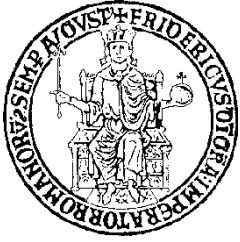
Asymptotic behavior

$$\begin{aligned}x_g^* &= \lim_{t \rightarrow \infty} x_g(t) \\y_g(0) &:= x_g(0), \quad \forall g \\y_g(t+1) &:= \sum_{g'} Q_{gg'} W_{g'} y_{g'}(t), \quad t \geq 0 \\x_g(t) &= \frac{y_g(t)}{\sum_{g'} y_{g'}(t)}, \quad \forall t \\T &= (Q_{gg'} W_{g'})\end{aligned}$$



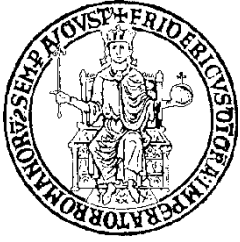
Asymptotic behavior

$$\begin{aligned}x_g^* &= \lim_{t \rightarrow \infty} x_g(t) \\y_g(0) &:= x_g(0), \quad \forall g \\y_g(t+1) &:= \sum_{g'} Q_{gg'} W_{g'} y_{g'}(t), \quad t \geq 0 \\x_g(t) &= \frac{y_g(t)}{\sum_{g'} y_{g'}(t)}, \quad \forall t \\T &= (Q_{gg'} W_{g'}) \\\vec{y}(t) &= \underbrace{T \circ T \cdots \circ T}_t \vec{x}(0) \\&\quad t \text{ times}\end{aligned}$$



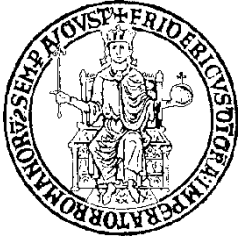
Thus





Thus

$$\vec{y}(t) \longrightarrow \lambda(t) \vec{y}^*$$



Thus

$$\vec{y}(t) \longrightarrow \lambda(t) \vec{y}^*$$

$$\vec{y}^* : \mathsf{T} \vec{y}^* = \lambda_{\max} \vec{y}^*$$



Thus

$$\vec{y}(t) \longrightarrow \lambda(t) \vec{y}^*$$

$$\vec{y}^* : T \vec{y}^* = \lambda_{\max} \vec{y}^*$$

λ_{\max} : leading eigenvalue of T (positive and non-degenerate)



Thus

$$\vec{y}(t) \longrightarrow \lambda(t) \vec{y}^*$$

$$\vec{y}^* : T \vec{y}^* = \lambda_{\max} \vec{y}^*$$

λ_{\max} : leading eigenvalue of T (positive and non-degenerate)

$$\vec{x}(t) \longrightarrow \vec{x}^* = \frac{\vec{y}^*}{\sum_g y_g^*}$$



Thus

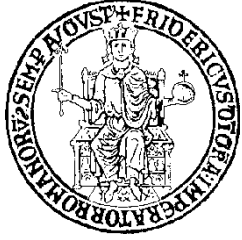
$$\vec{y}(t) \longrightarrow \lambda(t) \vec{y}^*$$

$$\vec{y}^* : T \vec{y}^* = \lambda_{\max} \vec{y}^*$$

λ_{\max} : leading eigenvalue of T (positive and non-degenerate)

$$\vec{x}(t) \longrightarrow \vec{x}^* = \frac{\vec{y}^*}{\sum_g y_g^*}$$

Note: λ_{\max} is not “experimentally” accessible



A simple mutation model



$$\mathbf{g} = (\sigma_1, \dots, \sigma_L) \quad \sigma_i = \pm 1 \quad \forall i$$

$$Q_{\mathbf{g}\mathbf{g}'} = \prod_{i=1}^L P(\sigma'_i \longrightarrow \sigma_i)$$

$$P(\sigma'_i \longrightarrow \sigma_i) = \begin{cases} \mu, & \text{if } \sigma_i = -\sigma'_i \\ (1 - \mu), & \text{if } \sigma_i = \sigma'_i \end{cases}$$

\Downarrow

$$Q_{\mathbf{g}\mathbf{g}'} = \mu^{d_H(\mathbf{g}, \mathbf{g}')} (1 - \mu)^{L - d_H(\mathbf{g}, \mathbf{g}')}$$

Hamming distance $d_H(\mathbf{g}, \mathbf{g}') := \sum_{i=1}^L (1 - \delta_{\sigma_i \sigma'_i})$

Single-peak landscapes

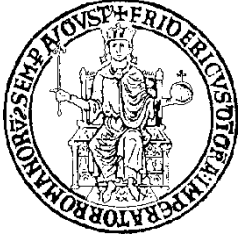


Fitness landscape: $g \longrightarrow W_g$

g_0 : **master sequence**

$$W_{g_0} \geq W_g \quad \forall g$$

Single-peak landscapes: W_g depends only on the Hamming distance $d_H(g_0, g)$



Fujiyama landscape



$$W_{\mathbf{g}} \propto w^{d_{\text{H}}(\mathbf{g}_0, \mathbf{g})}, \quad w = e^{-s} < 1$$

The stationary distribution factorizes over sites

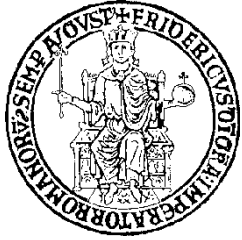
$$x_{\mathbf{g}}^* = \prod_{i=1}^L p_i(\sigma_i)$$

Define

$$p = p_i(\sigma_{0i}) \quad \mathbf{g}_0 = (\sigma_{0i})$$

Then

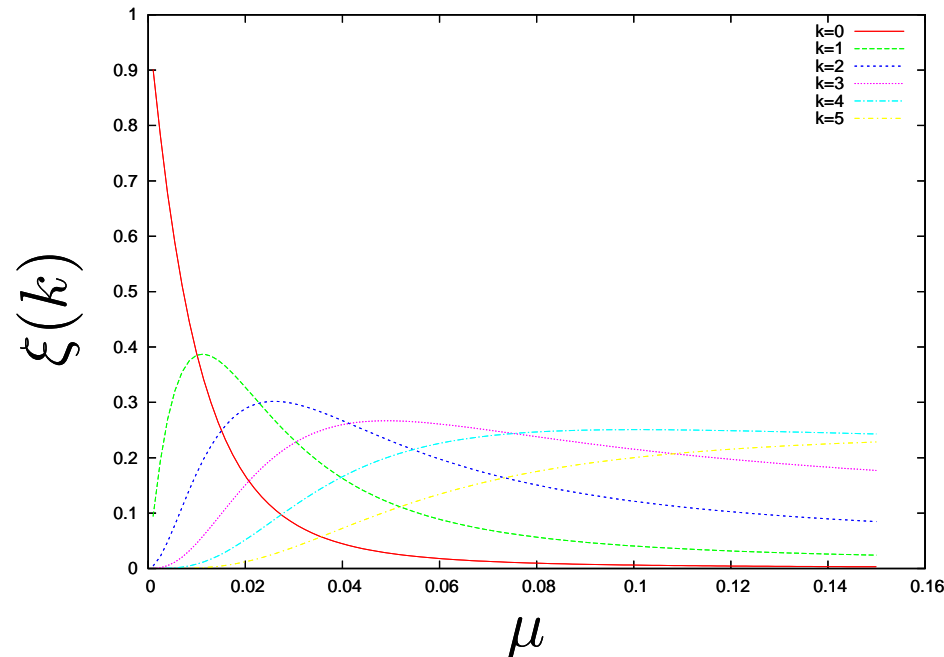
$$p = \frac{(1 - \mu)p + \mu w (1 - p)}{p + w (1 - p)}$$



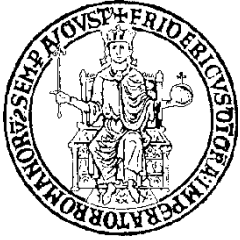
Error class distribution

Error class k : set of all individuals which differ from g_0 in k sites

$$\xi(k) := \sum_{\mathbf{g}} \delta_{d_H(\mathbf{g}, \mathbf{g}_0), k} x_{\mathbf{g}}^*$$



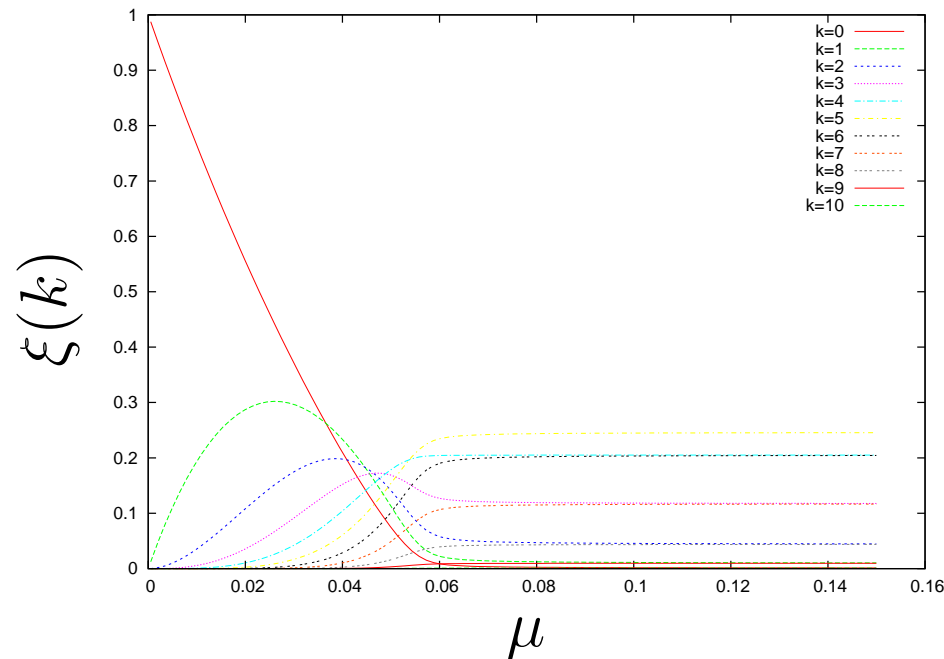
$L = 10$ and $s = 0.06$



The Error Threshold

Sharp-peak landscape

$$W_{\sigma g} = \begin{cases} 1, & \text{if } g = g_0 \\ 1 - s, & (0 < s < 1), \end{cases} \quad \text{otherwise}$$



$L = 10$ and $s = 0.4$

The infinite genome limit



Assume $L \rightarrow \infty$, $\lim_{L \rightarrow \infty} \mu L = u$

$q = e^{-u}$: probability that there are no mutations

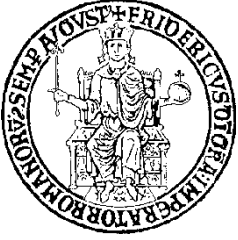
$$x_0(t+1) = \frac{q x_0(t)}{x_0(t) + (1-s)x_0(t)}$$

$$x_0^* = \begin{cases} 1 - (1-q)/s, & \text{if } (1-q) < s \\ 0, & \text{otherwise} \end{cases}$$

Maynard Smith, 1982

Threshold at

$$q_c = 1 - s$$



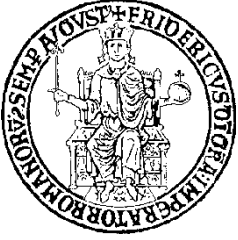
Finite genome size

For finite L , remember that q is the probability of having no mutations:

$$q_c = (1 - \mu)^{L_c} \quad \Rightarrow \quad L_c \simeq \frac{\log(1 - s)}{\log(1 - \mu)}$$

The error threshold sets an upper limit to the length L of a string which has to be reproduced

Eigen, 1971



Interpretation of the Error Threshold

Prebiotic RNA replication

μ : Probability of base replacement

s : Quality factor

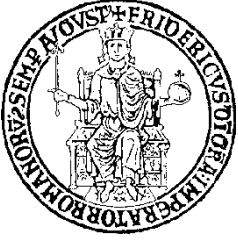
L_c : Maximum RNA length for which adaptation information can be conserved

$$\mu \sim 10^{-2} \quad \log(1 - s) \sim 1$$

$$L_c \sim 100$$

Can a RNA string of ~ 100 nucleotides exhibit enzymatic activity?

If no, information must be distributed among several strings
→ Hypercycle, replicator, etc.

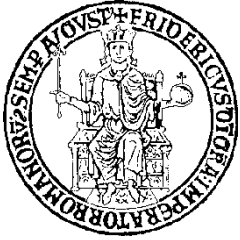


Time-dependent landscapes

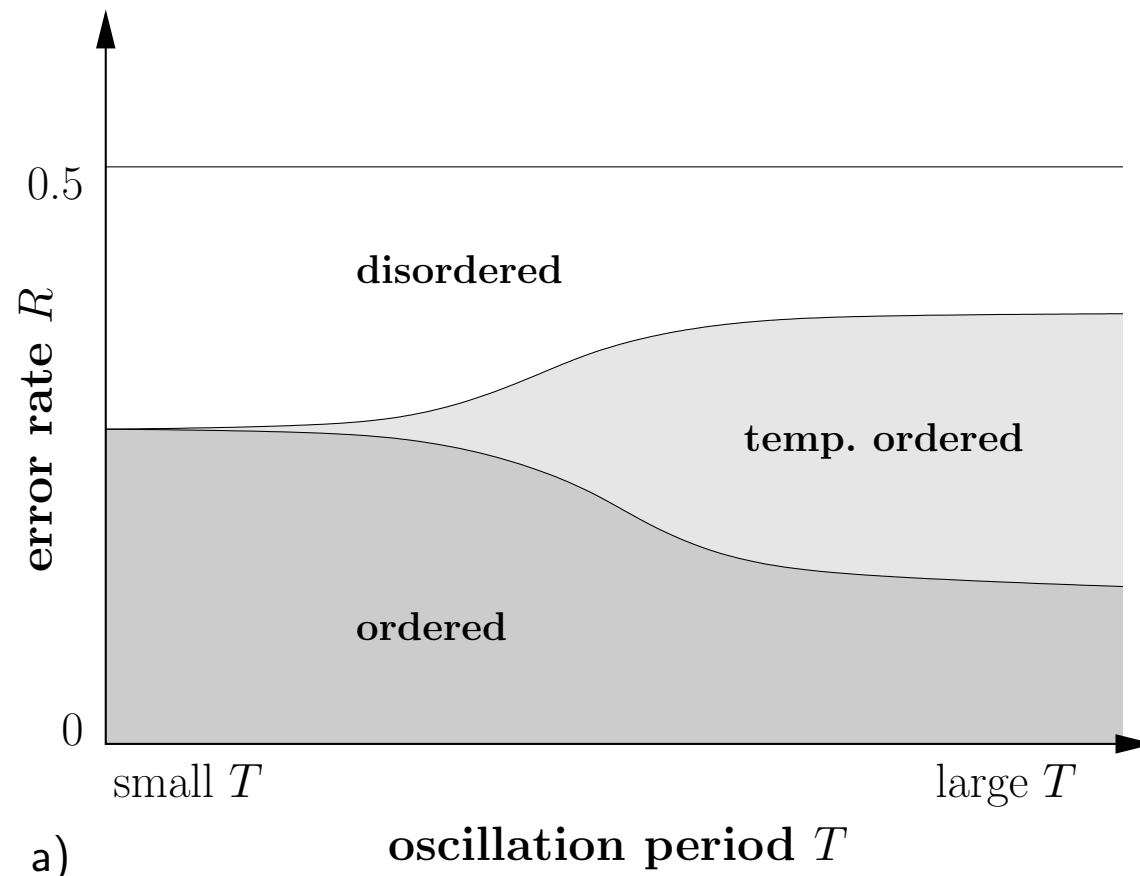
Oscillating single-peak landscape

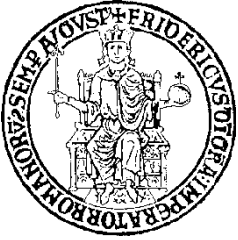
$$s(t) = s(t + \tau)$$

$$\vec{y}(t) = \underbrace{T_{t-1} \circ T_{t-2} \cdots T_{t-\tau}}_{\tau \text{ times}} \circ \underbrace{T_{t-\tau-1} \circ \cdots T_{t-2\tau}}_{\tau \text{ times}} \cdots \\ \circ \underbrace{T_{\tau} \circ T_{\tau-1} \cdots T_1}_{\tau \text{ times}} \vec{y}(0)$$

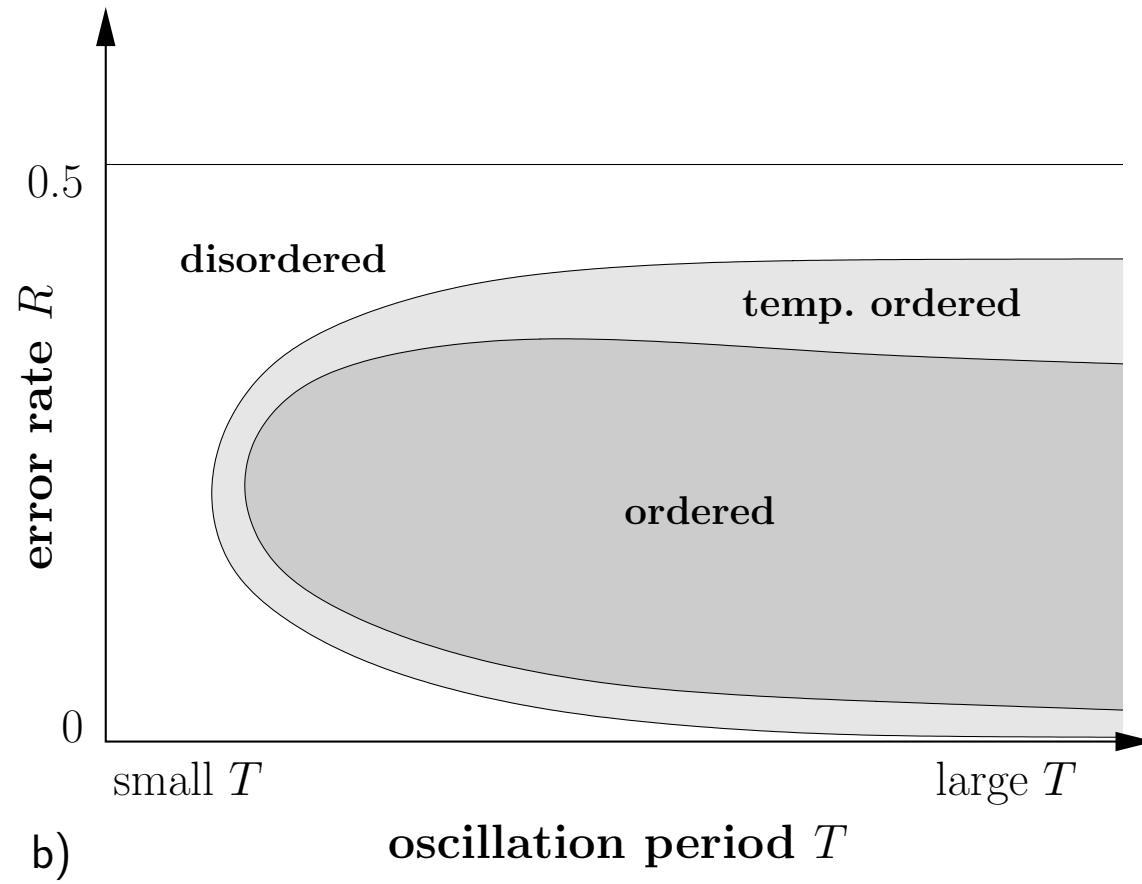


\bar{W} = non-flat landscape

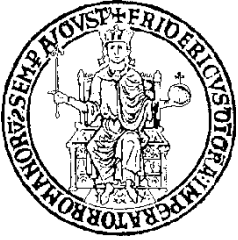




\bar{W} = flat landscape

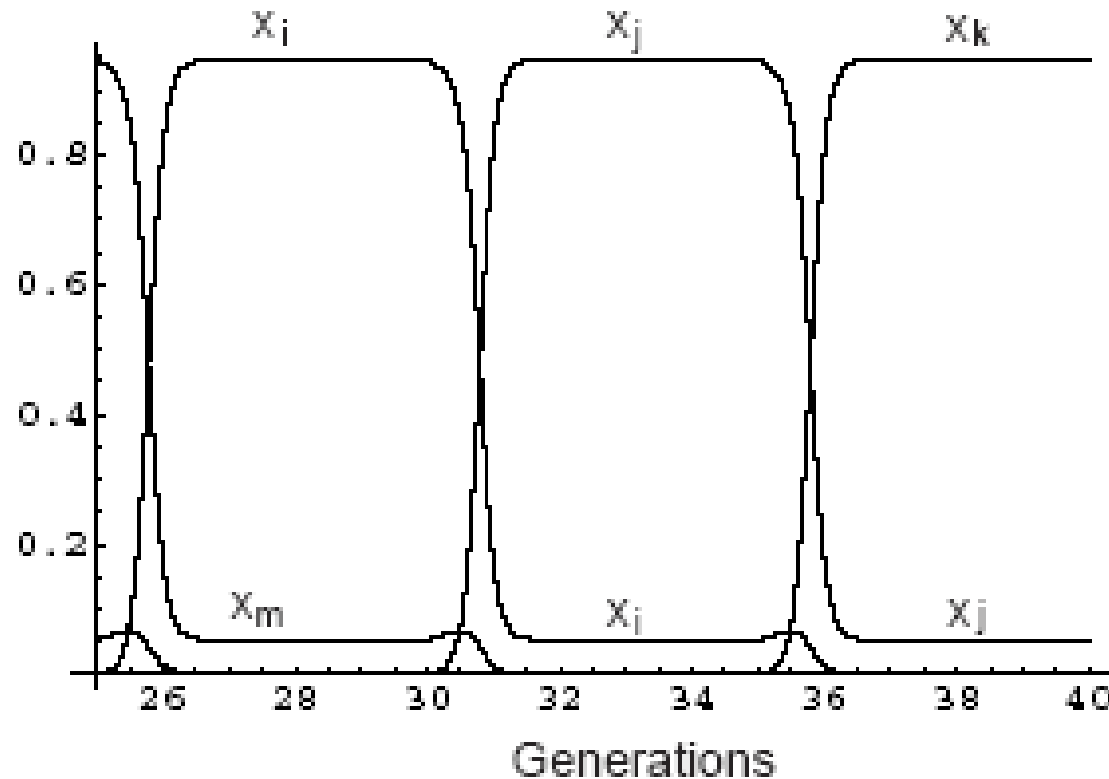


Wilke et al., 1999

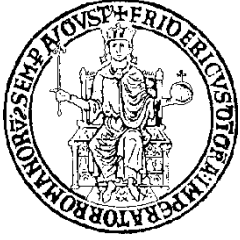


Wandering single-peak landscape

The fitness peak moves to a neighboring sequence every τ generations



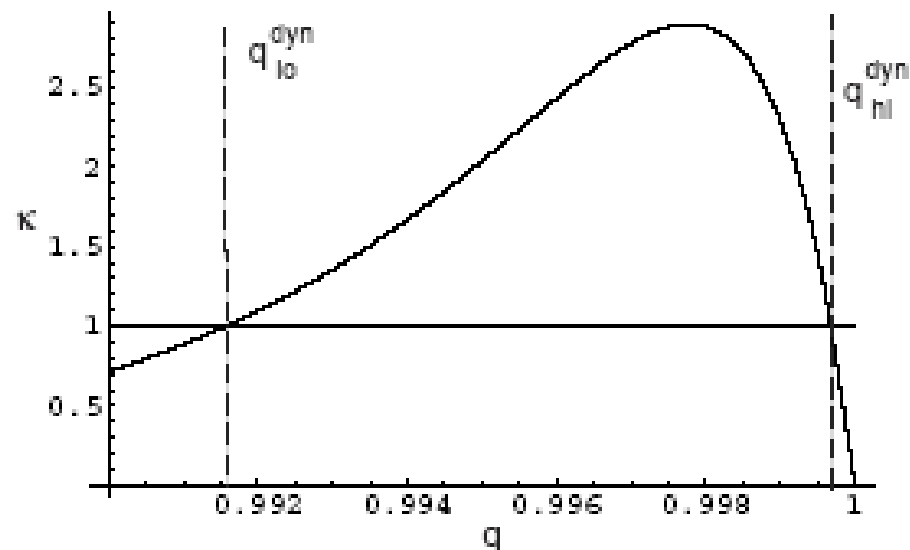
Nilsson and Snoad, 1999

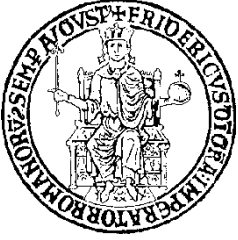


Two error thresholds

$$S := \frac{1}{1-s} \quad q := (1-\mu) \text{ (replication accuracy per base)}$$

$$\kappa := \left(e^{(q^L S - 1)\tau} - e^{(q^L - 1)\tau} \right) (1-q)S / ((S-1)q)$$

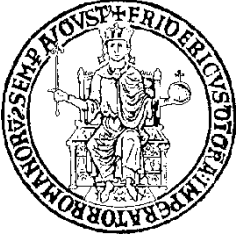




Coevolutionary virus-host dynamics

Kamp and Bornholdt, 1999

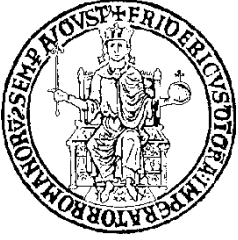
- ⑥ Two populations: virus V and immune system cells I , identified by sequence σ



Coevolutionary virus-host dynamics

Kamp and Bornholdt, 1999

- ⑥ Two populations: virus V and immune system cells I , identified by sequence σ
- ⑥ Viral population fraction z_σ , immune cells fraction y_σ
- ⑥ V growth rate affected by I concentration, and viceversa



Coevolutionary virus-host dynamics

Kamp and Bornholdt, 1999

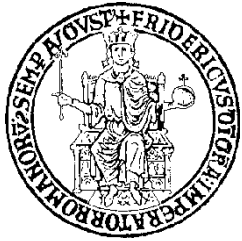
- ⑥ Two populations: virus V and immune system cells I , identified by sequence σ
- ⑥ Viral population fraction z_σ , immune cells fraction y_σ
- ⑥ V growth rate affected by I concentration, and viceversa
- ⑥ V total population not fixed



Coevolutionary virus-host dynamics

Kamp and Bornholdt, 1999

- ⑥ Two populations: virus V and immune system cells I , identified by sequence σ
- ⑥ Viral population fraction z_σ , immune cells fraction y_σ
- ⑥ V growth rate affected by I concentration, and viceversa
- ⑥ V total population not fixed
- ⑥ “Local” interaction and continuous time



Coevolution equations

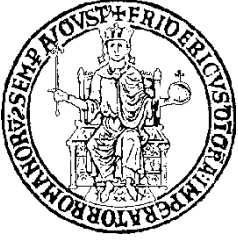
$$\dot{y}_\sigma = \sum_{\sigma'} Q_{\sigma\sigma'}^I W_{\sigma'}^I(z) y_{\sigma'} - \langle W^I \rangle_t y_\sigma$$

$$\dot{z}_\sigma = \sum_{\sigma'} Q_{\sigma\sigma'}^V W_{\sigma'}^V z_{\sigma'} - D_\sigma(y) z_\sigma$$

$$W_\sigma^I(z) = \begin{cases} S, & \text{if } \sigma \text{ is the master viral sequence} \\ \eta \ll S, & \text{otherwise} \end{cases}$$

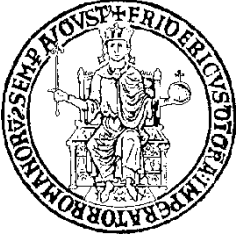
$$W_\sigma^V(z) = \begin{cases} S_V, & \text{if } \sigma \text{ is the master viral sequence} \\ \eta_V \ll S_V, & \text{otherwise} \end{cases}$$

$$D_\sigma(y) = \begin{cases} \delta > W^V, & \text{if } \sigma \text{ is the master immune sequence} \\ 0, & \text{otherwise} \end{cases}$$



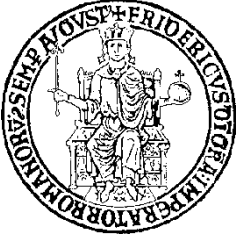
Hide and seek in sequence space

- ⑥ Once the immune system “finds out” the viral quasispecies, the viral fitness peak moves to one neighboring location



Hide and seek in sequence space

- ⑥ Once the immune system “finds out” the viral quasispecies, the viral fitness peak moves to one neighboring location
- ⑥ The viral quasispecies adapts to the new peak on a time scale τ_V



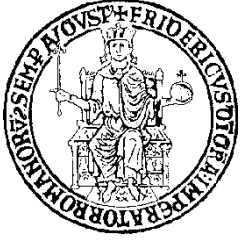
Hide and seek in sequence space

- ⑥ Once the immune system “finds out” the viral quasispecies, the viral fitness peak moves to one neighboring location
- ⑥ The viral quasispecies adapts to the new peak on a time scale τ_V
- ⑥ The immune system is adjusted to the new fitness peak on a time scale τ_I



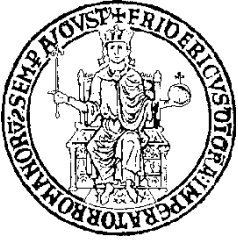
Hide and seek in sequence space

- ⑥ Once the immune system “finds out” the viral quasispecies, the viral fitness peak moves to one neighboring location
- ⑥ The viral quasispecies adapts to the new peak on a time scale τ_V
- ⑥ The immune system is adjusted to the new fitness peak on a time scale τ_I
- ⑥ The whole cycle repeats on a time scale $\tau = \tau_V + \tau_I$

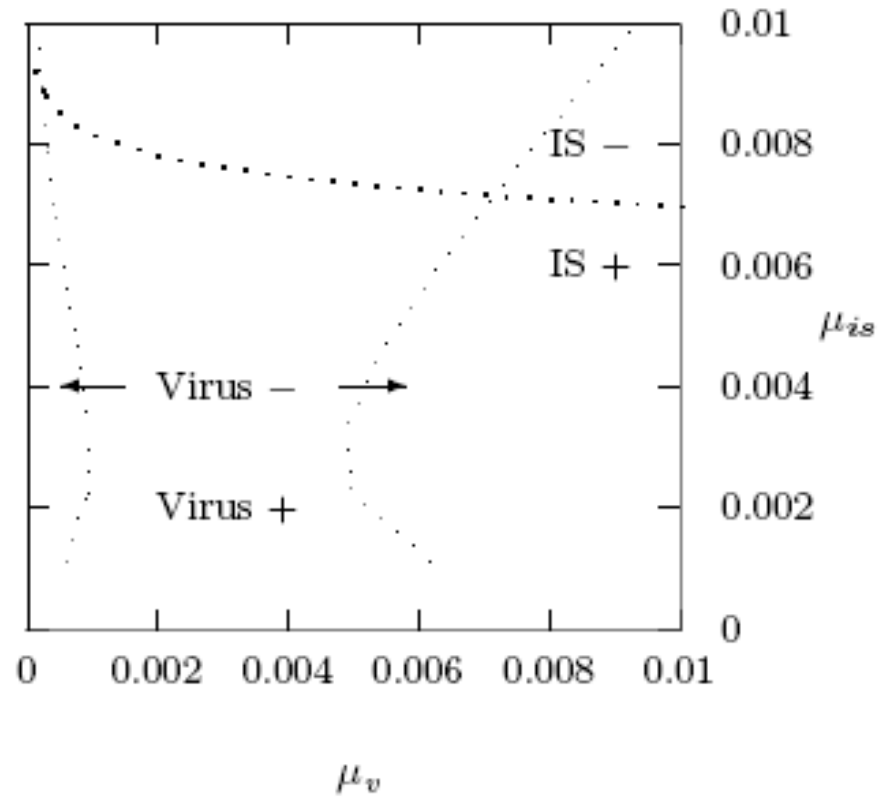


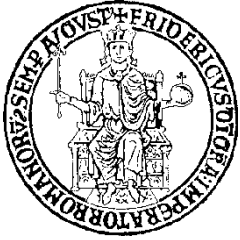
Estimating the time scales

$$\tau_V \approx -\frac{\ln(1 - q_V)}{q_V^L (S_V - \eta_V) + \delta}$$
$$\tau_I \approx -\frac{(1 - q_I)}{q_I^L (S - \eta)}$$



Phase diagram

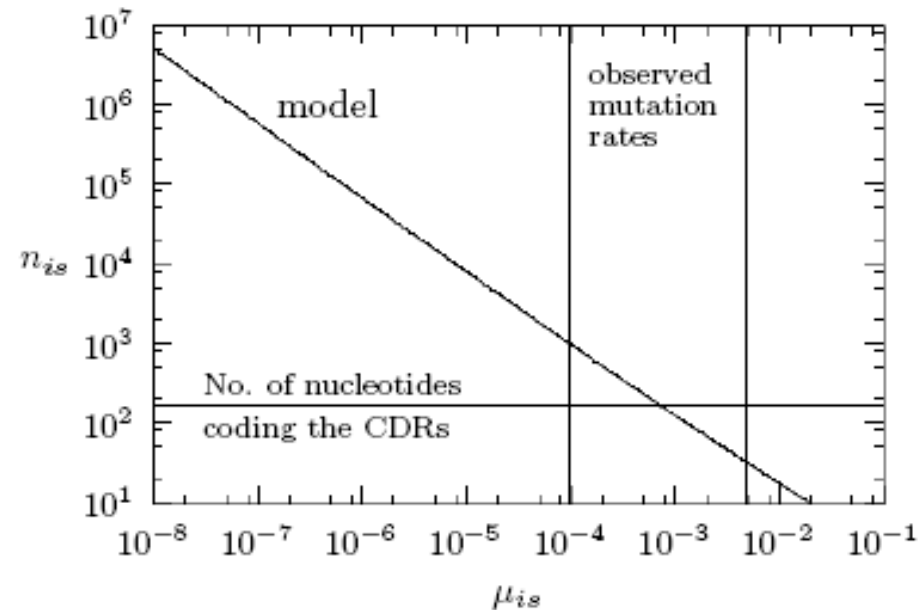


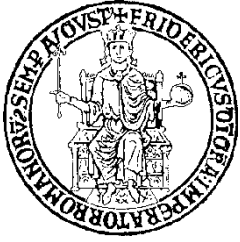


Optimality condition

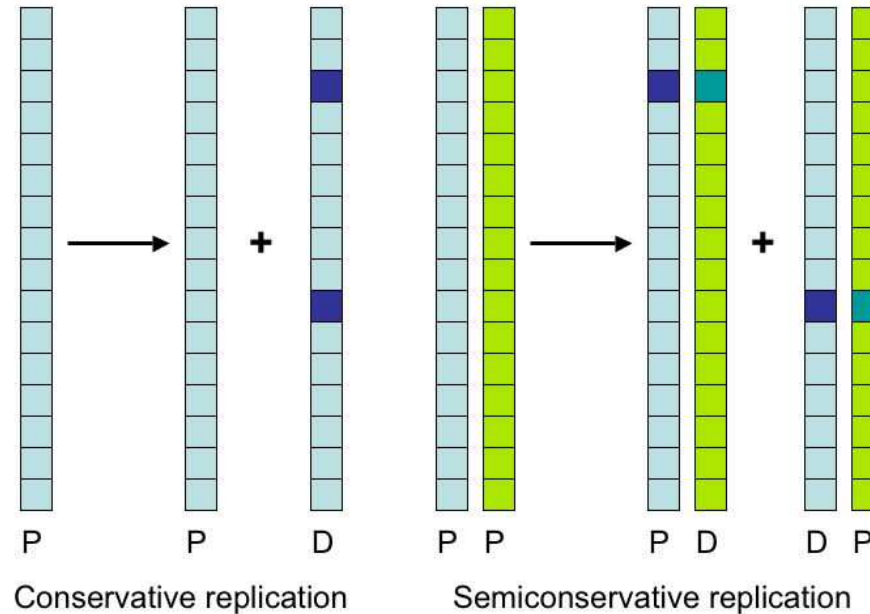
If μ_I is optimized to minimize the range of V quasispecies, we obtain the condition

$$\frac{\partial \kappa_V}{\partial \mu_I} = 0$$

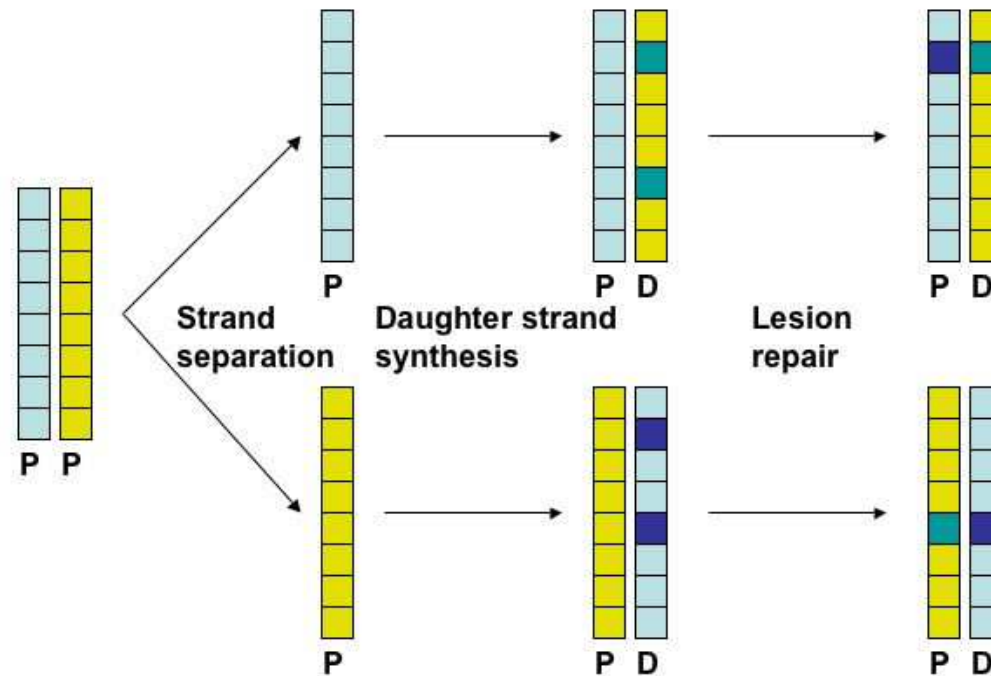
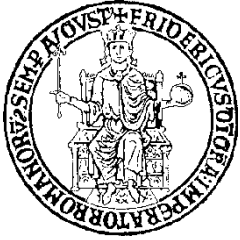


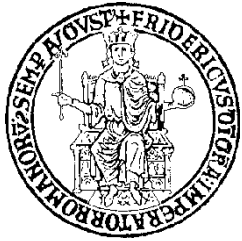


Semiconservative replication



Tannenbaum, Shakhnovich et al.





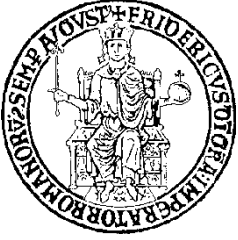
Mutation probabilities

$p(\sigma, \sigma')$: probability that σ is paired with σ' during daughter strand synthesis

$p((\sigma, \sigma'), (\sigma'', \bar{\sigma}''))$: probability that $\sigma \rightarrow \sigma''$, $\sigma' \rightarrow \bar{\sigma}''$ during lesion repair

$p(\sigma, \{\sigma'', \bar{\sigma}''\})$: probability that a “mother” strand σ generates the pair $\{\sigma'', \bar{\sigma}''\}$ in the daughter cell

$$p(\sigma, \{\sigma'', \bar{\sigma}''\}) = \sum_{\sigma'} p(\sigma, \sigma') [p((\sigma, \sigma'), (\sigma'', \bar{\sigma}'')) + p((\sigma, \sigma'), (\bar{\sigma}'', \sigma''))]$$

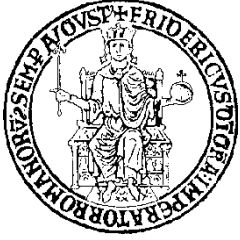


Quasispecies equations

$$n = \sum_{\{\sigma, \bar{\sigma}\}} n_{\{\sigma, \bar{\sigma}\}}$$

$$x_{\{\sigma, \bar{\sigma}\}} = n_{\{\sigma, \bar{\sigma}\}} / n$$

$$\begin{aligned} \frac{dx_{\{\sigma, \bar{\sigma}\}}}{dt} = & \sum_{\{\sigma', \bar{\sigma}'\}} \kappa_{\{\sigma', \bar{\sigma}'\}} (p(\sigma', \{\sigma, \bar{\sigma}\}) + p(\sigma', \{\sigma, \bar{\sigma}\})) \\ & - (\kappa_{\{\sigma, \bar{\sigma}\}} + \bar{\kappa}(t)) x_{\{\sigma, \bar{\sigma}\}} \end{aligned}$$



The sharp peak landscape



$$\kappa_{\{\sigma, \bar{\sigma}\}} = \begin{cases} k = 1 + s, & \text{if } \sigma = \sigma_0 \\ 1, & \text{otherwise} \end{cases}$$

Master sequence $\sigma_0 \longrightarrow$ master pair $\{\sigma_0, \bar{\sigma}_0\}$

Consider “error classes” around $\sigma_0, \bar{\sigma}_0$: w_0, \bar{w}_0

$$\frac{dw_0}{dt} = 2ke^{-\mu/2}w_0 - (k + \bar{k}(t))w_0$$

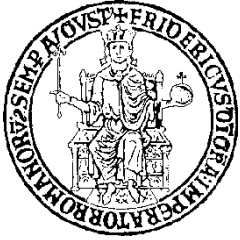


Error threshold

$$w_0 = 0, \quad \frac{k(2e^{-\mu/2} - 1) - 1}{k - 1} = w^*$$

$$w_0 \geq 0 \Rightarrow w = w^* \text{ for } \mu < \mu_c$$

$$\mu_c = 2 \ln \frac{2}{1 + 1/k}$$

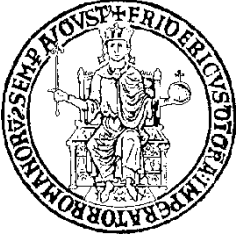


Localization length



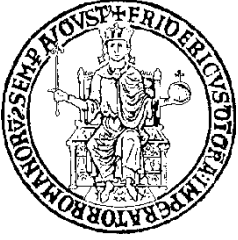
$$\langle l \rangle = \sum_{l=1}^{\infty} 2lw_l$$

$$\langle l \rangle = \mu \frac{k(2e^{-\mu/2} - 1)}{k(2e^{-\mu/2} - 1) - 1}$$

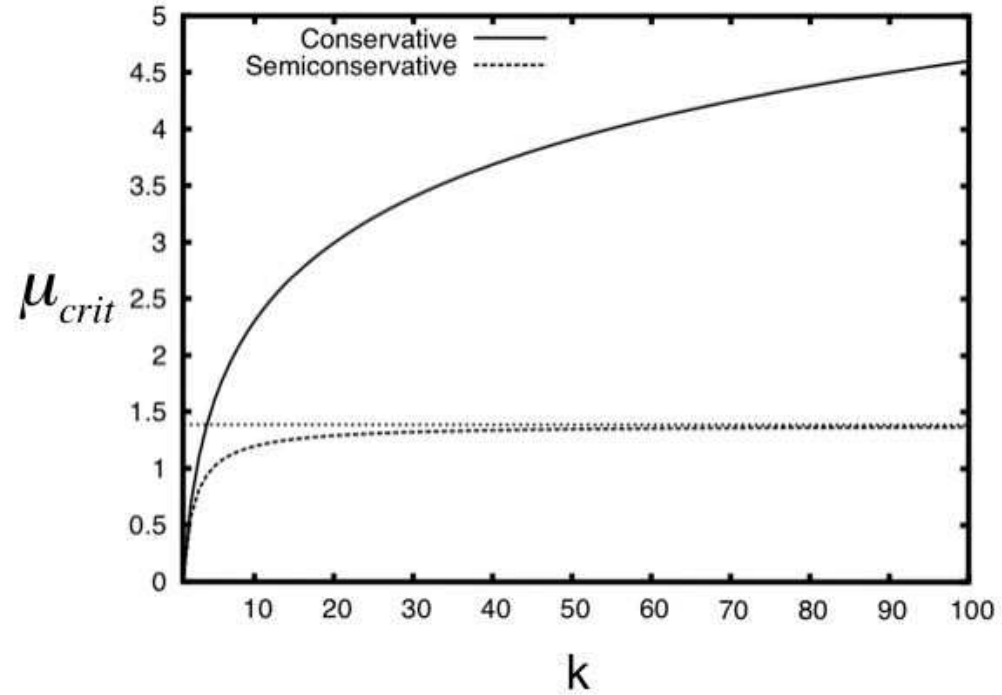


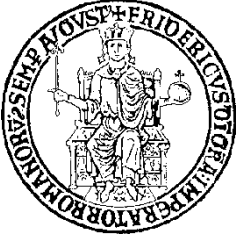
Comparison

Parameter	Conservative	Semiconservative
μ_{crit}	$\ln k$	$2 \ln \frac{2}{1+1/k}$
$x_{via} (\mu < \mu_{crit})$	$\frac{ke^{-\mu}-1}{k-1}$	$\frac{k(2e^{-\mu/2}-1)-1}{k-1}$
$\bar{\kappa}(t = \infty) (\mu < \mu_{crit})$	$ke^{-\mu}$	$k(2e^{-\mu/2} - 1)$
$\langle l \rangle (\mu < \mu_{crit})$	$\mu \frac{ke^{-\mu}}{ke^{-\mu}-1}$	$\mu \frac{k(2e^{-\mu/2}-1)}{k(2e^{-\mu/2}-1)-1}$



Critical μ

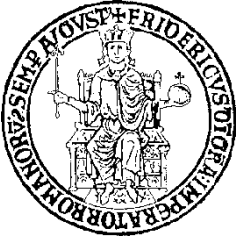




Observations

- ⑥ In conservative replication, it is possible avoid the error catastrophe by replicating sufficiently fast, because one copy of the original genome is kept standing

Tannenbaum and Shakhnovich, 2005



Observations

- ⑥ In conservative replication, it is possible avoid the error catastrophe by replicating sufficiently fast, because one copy of the original genome is kept standing
- ⑥ Semiconservative replication, by contrast, destroys the original parent genome. If the mutation rate is too high, the error catastrophe cannot be avoided by any amount of selection pressure

Tannenbaum and Shakhnovich, 2005

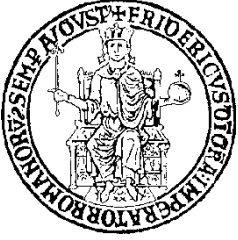


Observations

- ⑥ In conservative replication, it is possible avoid the error catastrophe by replicating sufficiently fast, because one copy of the original genome is kept standing
- ⑥ Semiconservative replication, by contrast, destroys the original parent genome. If the mutation rate is too high, the error catastrophe cannot be avoided by any amount of selection pressure
- ⑥ This feature of semiconservative replication provides a natural explanation for why mutagens can kill faster replicating cells more quickly than more slowly replicating cells

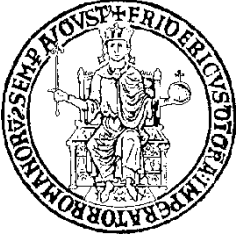
Tannenbaum and Shakhnovich, 2005

Discussion of the Quasispecies approach



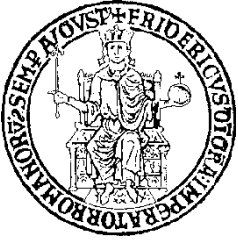
- ⑥ “Mean-field like”: Useful to identify major features

Discussion of the Quasispecies approach



- ⑥ “Mean-field like”: Useful to identify major features
- ⑥ Quasispecies: Coexistence of molecularly different sequences

Discussion of the Quasispecies approach



- ⑥ “Mean-field like”: Useful to identify major features
- ⑥ Quasispecies: Coexistence of molecularly different sequences
- ⑥ Generalizable to sexual reproduction (Higgs) and coevolution (Eigen, Schuster, Fontana)

Discussion of the Quasispecies approach



- ⑥ “Mean-field like”: Useful to identify major features
- ⑥ Quasispecies: Coexistence of molecularly different sequences
- ⑥ Generalizable to sexual reproduction (Higgs) and coevolution (Eigen, Schuster, Fontana)
- ⑥ Unreliable beyond the error threshold (neutral evolution)

Discussion of the Quasispecies approach

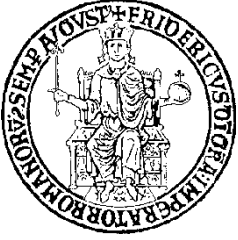


- ⑥ “Mean-field like”: Useful to identify major features
- ⑥ Quasispecies: Coexistence of molecularly different sequences
- ⑥ Generalizable to sexual reproduction (Higgs) and coevolution (Eigen, Schuster, Fontana)
- ⑥ Unreliable beyond the error threshold (neutral evolution)
- ⑥ Effects of finite population size?



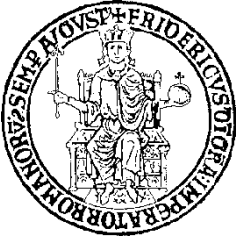
Neutral mutations?

- ⑥ Substitution rate \approx mutation rate: very *large* mutation rate ($\sim 2 \cdot 10^{-8}$ nb/generation in humans (Drake et al. 1995))



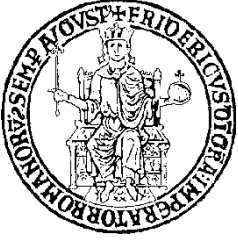
Neutral mutations?

- ⑥ Substitution rate \lesssim mutation rate: very *large* mutation rate ($\sim 2 \cdot 10^{-8}$ nb/generation in humans (Drake et al. 1995))
- ⑥ Large mutation rate \Rightarrow large substitutional load, *unless* fitness variation is small



Neutral mutations?

- ⑥ Substitution rate \lesssim mutation rate: very *large* mutation rate ($\sim 2 \cdot 10^{-8}$ nb/generation in humans (Drake et al. 1995))
- ⑥ Large mutation rate \Rightarrow large substitutional load, *unless* fitness variation is small
- ⑥ Most mutations that are substituted are neutral (or only very slightly deleterious)



Neutral mutations?

- ⑥ Substitution rate \lesssim mutation rate: very *large* mutation rate ($\sim 2 \cdot 10^{-8}$ nb/generation in humans (Drake et al. 1995))
- ⑥ Large mutation rate \Rightarrow large substitutional load, *unless* fitness variation is small
- ⑥ Most mutations that are substituted are neutral (or only very slightly deleterious)
- ⑥ Evolution on neutral mutations takes place by sampling fluctuations (*drift*)