# Simple Models of Evolutive Population Dynamics

Franco Bagnoli

October 20, 2010

# Replicators

- Living cells (mono and multicellular organisms, sexual and asexual, etc.): DNA, mithocondries,
- Viruses.
- Transposable elements, plasmids.
- Computer viruses.
- Memes (cultural replicators).
- Other chemical replicators (prions, crystals...) ?.

What is interesting, is that these entities may interact: viruses and cells, memes and hoaxes....

# Structure

- An out-equilibrium "chemical" background, that sets the "rules" and furnishes the energy.
- Information contents (genotype).
- Appearence (phenotype).
- Persistence and diffusion (reproduction).
- Variability (mutations, sexual reproduction).
- Selection.

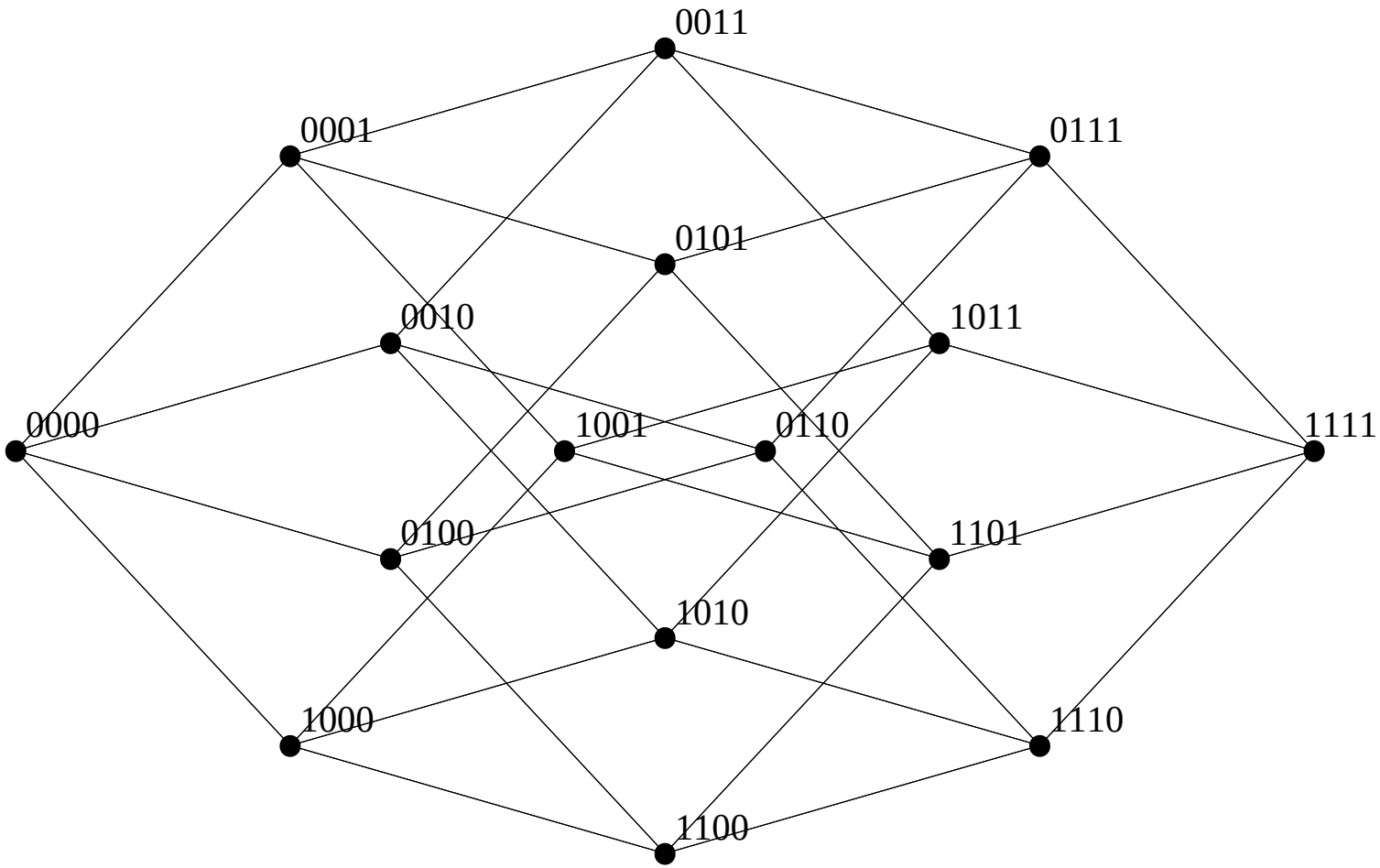
# Rules of the game

- The individual (genotype and phenotype).
- The population (an array of individuals).
- The pool (a network of interactions): may be all-to-all or spatially located on a lattice, or in plain space.
- Pick up an individual and compute its probability of survival (fitness). If unfortunate, free the space...
- Otherwise, if there is available free space (and, for sexual organisms, if there is an available partner), replicate it, with mutations.
- Let it run for a long time and ... voilá: Adam and Eve...

# The Genotype

- DNA: sequence of symbols, with informative (genes) and structural contents: the latter is (also) part of the phenotype.
- Gene: a piece of genetic information that has some function and persists long enough (protein-coding, structure, parasitic, ...).
- Simplification: instead of sequences, let us consider the genotype as a bag of genes (synonymous of loci). But there are interesting questions about packing information (viruses), structural constraints, etc.
- Discrete number of alleles (say 2): the genome is an array of zeros and ones.
- $g = (g_1, g_2, \ldots, g_L)$, $g_i \in \{0, 1\}$.

# Sequence space



# Variability

- Mutations in the germ line (during life or replication of genome) are transmitted to offsprings.
- Sexual reproduction (recombination) shuffles genes.
- Point mutations are like local steps in the genotypic space.
- Genes may be deeply changed by duplications, transpositions, inversions, deletions, insertion of viral genes, plasmids (bacteria), chromosome duplication and fusion, etc (long-range jumps).

# The Phenotype

- Phenotype: array of quantitative traits $f = (f_1, \ldots, f_M)$. It may include "internal" traites or extended ones (like the beaver ponds).
- The phenotype is determined by the genotype and the history of individual (age, experiences). Let us suppose for the moment that $f = p(g)$.
- In generale $f_i$ is a function of many genes, and any gene affect many traits (this is why genetic engineering is hard). Some trait may depend on one or very few genes (for instance "structural" genes.
- $f_i(g_j, g_k, \ldots)$ may be a linear (non-epistatic) or nonlinear (epistatic) function of $g_j, g_k, \ldots$.
- The phenotipic space may be one or high-dimensional (depends on what one is intested in).

# Selection (fitness)

- Let us assume for the moment that the the probability of survival $A$ depends on the phenotype $A(f)$.
- In general, it depends also on the distribution of other phenotypes in the network of contacts (hares do not like to stay near to lynxs, the latters have opposite opinions).
- The probability of passing (and spreading) the genotype to the following generation depends also on available space, which may be considered a global coupling.
- We can have evolution without selection (neutral evolution).

# A microscopic model

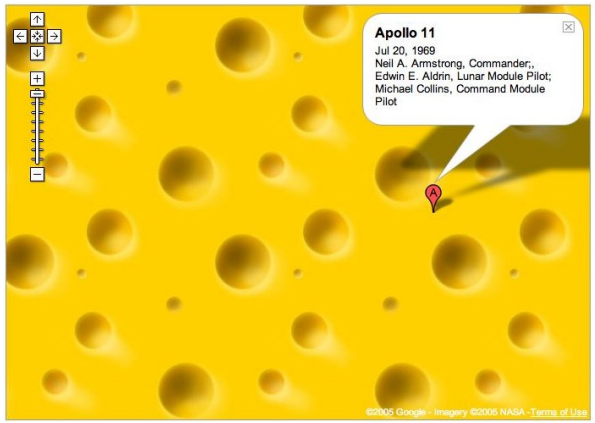
- The population is an array of genotypes (numbers in base 2 or strings). Time may be continuous or discrete. We consider fixed-size populations.
- For continous time, one computes the fitness of the individual, compared to the average one the population, and from this the survival probability is computed. If the individual do not survives, another, randomly chosen one, is replicated (with errors) and occupy its position.
- For discrete times, one has to replace the whole population, picking individuals at random and offering the possibility of passin (with errors) to the next generation, with a probability that is the fitness divided by the average fitness of the population.

# Fitness (or adaptive) Landscape

- Sewall Wright introduced the concept of adaptive landscape, similar to hill climbing.
- In our formalism, $A = A(g)$ and does not depend on the distribution of phenotypes.
- However, there is still global competition: Red Queen effect.
- In practice, the fitness is given by $A(g)/\langle A \rangle$.

# Evolution on a flat landscape

- The genotypic space is highly dimensional.
- A tiny fraction of all combinations are viable.
- There are many holes in this space. The remaining path is presumable locally flat (neutral evolution).
- Point mutations allow to define an hypermetric distance (molecular clock, philogenetic trees).
- Long-range jumps in genotipic space may cause a "small world" effect (not a tree).

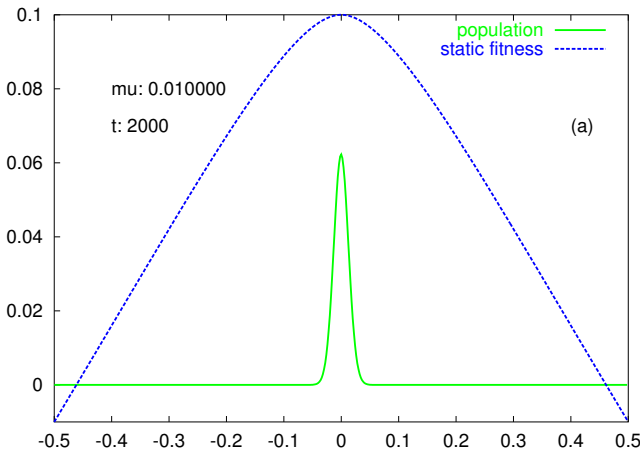


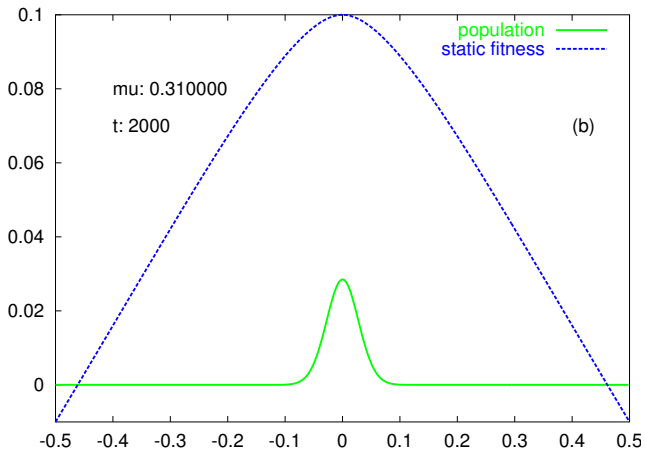
# Evolution on a fitness landscape

- For vanishing mutations the average fitness $\langle A \rangle$ is a nondecreasing function of time (Fisher theorem),
- And the asymptotic population distribution is a delta peak at the global maximum of the fitness (master sequence).
- For finite mutations, the master sequence is surrounded by a cloud of mutants with lower fitness (quasispecies).
- Broader peaks may "win" over sharper and slightly higher ones.
- Coexistence is fragile. Anyhow, coexisting strains have the same average fitness (Gause principle).
- The portions of the genome subjected to higher selective pressure are less mutable than "neutral" ones.

# Quasispecies.

- In “equilibrium”, infinite population, static fitness landscape, no mutations: just one strain (Fisher theorem).
- Mutations widen the distribution (quasispecies) and lower its average fitness



# Fitness depending on age: senescence (Penna model)

- Assume that gene $i$ is activate at age $i$.
- Genes $g_i = 0$ are good, $g_i = 1$ are bad.
- If three bad genes are activated the individual dies.
- At age 8 or greater, an individual can reproduce (variable population).
- Bad genes concentrate at age greater than 8, causing semelparousy.
- Introducing parental cares and death risk by childbirth, one may for instance introduce menopause.

# Infinite population

In the case of infinite population, one can use the probability distribution $p(g) = \lim_{N\to\infty} \frac{\sum_i [g_i = g]}{N}$.
The evolution equation is, for discrete time intervals

$$p(g, t+1) = \frac{A(g)}{\langle A \rangle} p(g, t) + \text{mutations}$$

or, for continuous time

$$\frac{\partial p(g,t)}{\partial t} = (A(g) - \langle A \rangle) p(g, t) + \text{mutations}$$

Evolution is a kind of pattern formation in genotypic space.

# Mutations in sharp fitness landscape

In fitness landscapes, mutations play a fundamental role:

- Individual histories are like random walks on a potential (fitness=energy, mutation=temperature).
- Random walks interact only on a global way (competition for space).
- Transitions (speciations) are dominated by search time (finding a niche).
- Searching efficiency depends on the smoothness of fitness landscape.
- Mutations may cause error threshold (disappearance of master sequence), and mutational meltdown (extinction in variable population, keeping the quasispecies structure).

# Smothness of fitness landscape

S. Kaufmann (the origins of order) introduced a "general" framework for fitness landscapes:

- There are L phenotypic traits, corresponding to L genes (L is N for Kauffman)
- Each trait $i$ depends on $K$ genes, by a function $f_i(g_i, \ldots, g_{i+k})$ (genes may be pick at random).
- The total fitness is $f = \sum_i f_i$.
- If $K = 1$, gene effects are additive (non-epistatic interactions): smooth landscape, one maximum.
- If $K = L$ and functions are random, the fitness landscape is random, and there are $2^L/(L-1)$ maxima (no search possible).
- For small values of $K$ maxima are correlated (they share some part of genome): recombination (sexual reproduction) is advantageous.

# Quasispecies, error threshold, mutational meltdown

Evolution on a sharp phenotypic ($u$) landscape $A(u)$: $A(0) = A_0$, others $A(u \neq 0) = A_*$.
Population: $n(0) = n_0$, $n(1) = n_1$, $n(u > 1) = n_*$. No back mutations.

$$n_0' = \left(1 - \frac{N}{K}\right)(1 - \mu_s) A_0 n_0, \tag{1}$$

$$n_1' = \left(1 - \frac{N}{K}\right)\left((1 - \mu_s) A_* n_1 + \mu_s A_0 n_0\right), \tag{2}$$

$$n_*' = \left(1 - \frac{N}{K}\right) A_*(n_* + \mu_s n_1). \tag{3}$$

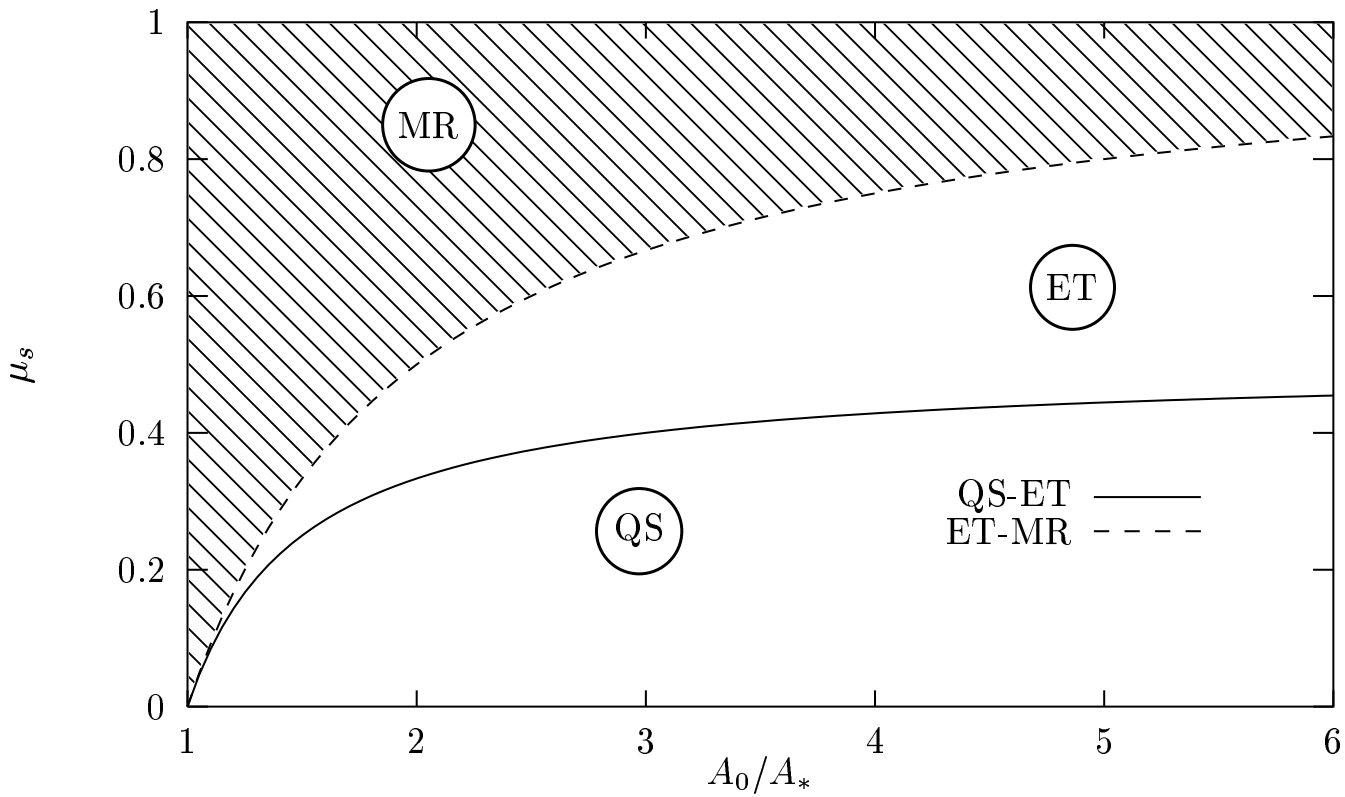
and

$$\langle A \rangle = \frac{A_0 n_0 + A_*(n_1 + n_*)}{N}$$

The steady state is given by $\mathbf{n}' = \mathbf{n}$. There are three possible fixed points $\mathbf{n}^{(i)} = \left(n_0^{(i)}, n_1^{(i)}, n_*^{(i)}\right)$: $\mathbf{n}^{(1)} = (0,0,0)$ $(N^{(1)} = 0)$, $\mathbf{n}_2 = (0,0,K(1-1/\langle A_*\rangle))$ $(N^{(2)} = n_*^{(2)})$ and

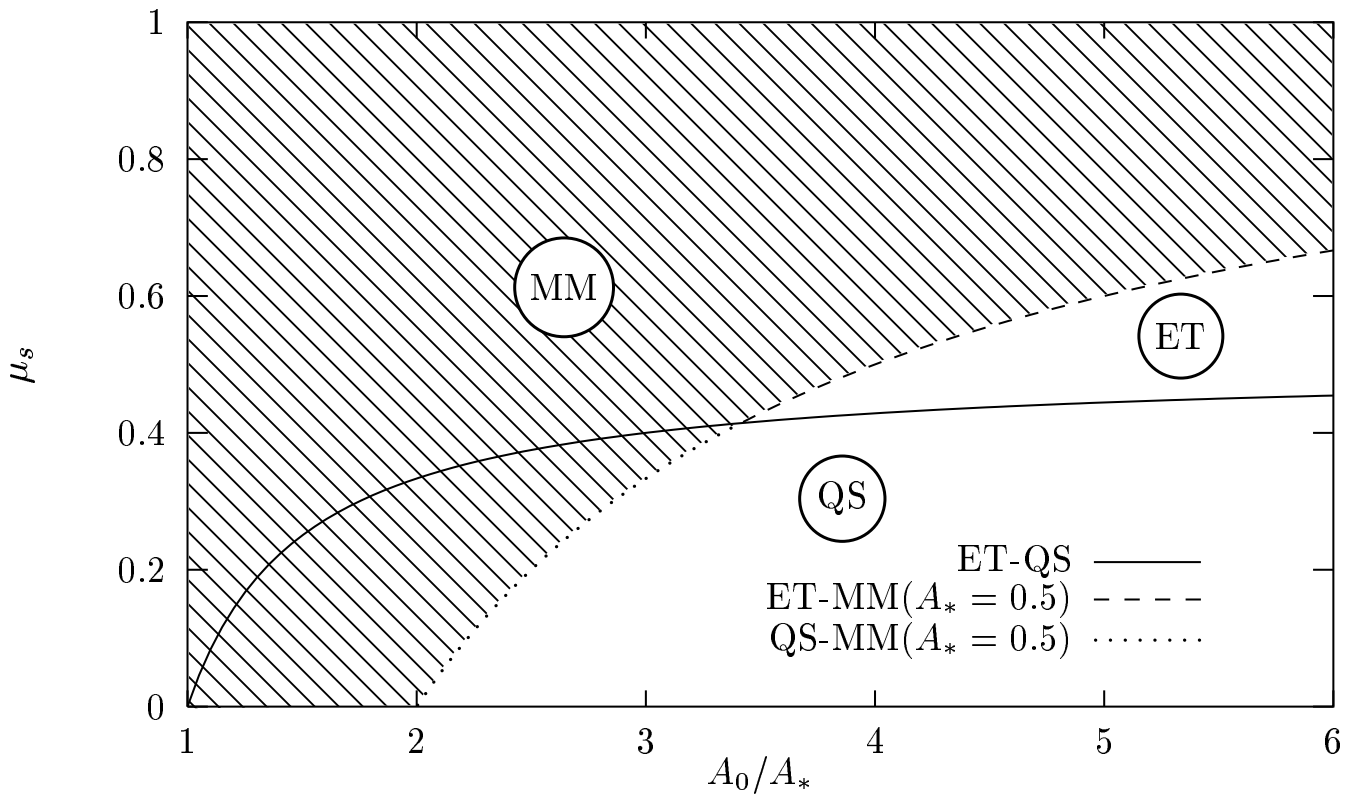
$$\mathbf{n}^{(3)} = \begin{cases} n_0^{(3)} & = N^{(3)} \dfrac{(1-\mu_s)A_0 - A_*}{A_0 - A_*}, \\ n_1^{(3)} & = N^{(3)} \dfrac{\mu_s}{1-\mu_s} \dfrac{A_0((1-\mu_s)A_0 - A_*)}{(A_0 - A_*)^2}, \\ n_*^{(3)} & = N^{(3)} \dfrac{\mu_s^2}{1-\mu_s} \dfrac{A_0 A_*}{(A_0 - A_*)^2}, \\ N^{(3)} & = 1 - \dfrac{1}{A_0(1-\mu_s)}. \end{cases}$$

# Quasispecies, error threshold, mutational meltdown



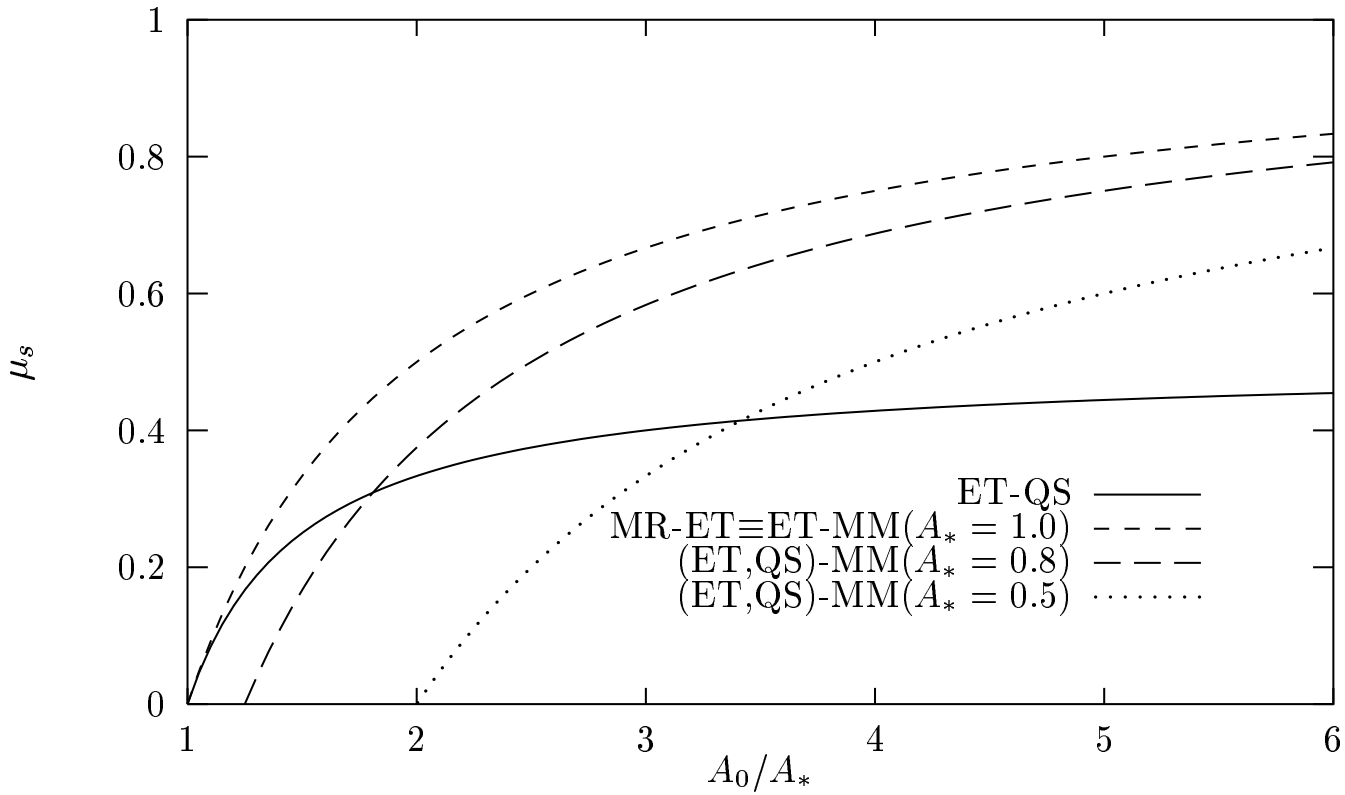

$A_* > 1$

# Quasispecies, error threshold, mutational meltdown



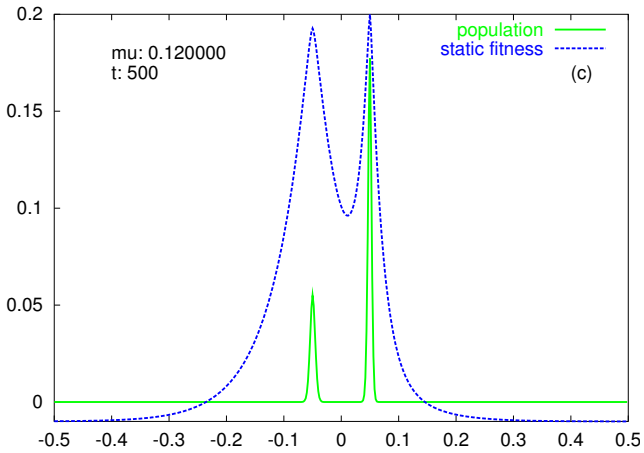

$A_* < 1$

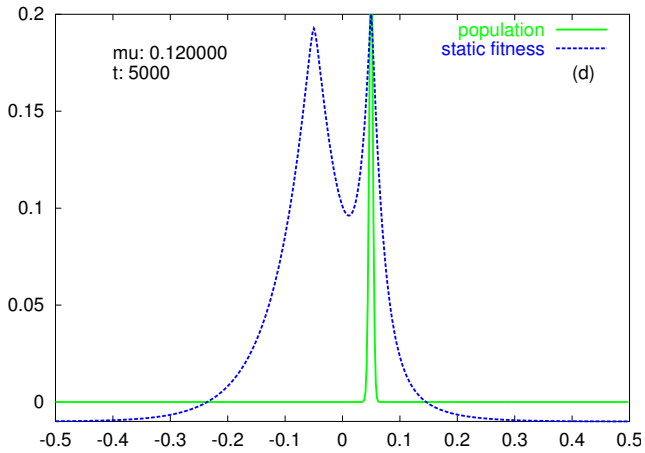
# Quasispecies, error threshold, mutational meltdown

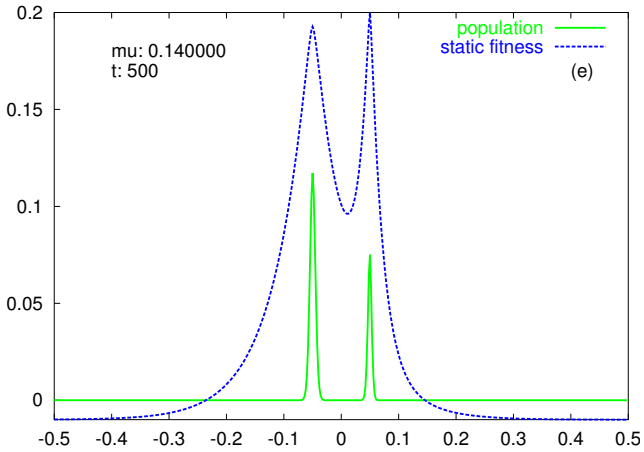


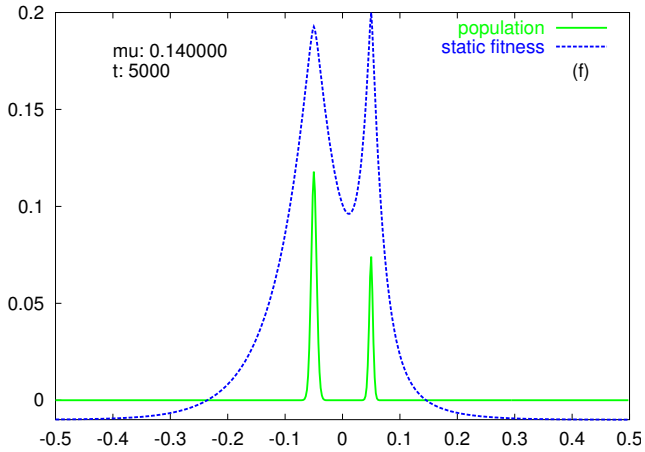
# Coexistence.

- Possible coexistence, but very sensitive to mutation rate



# Competition

The fitness actually depends also on the rest of population.

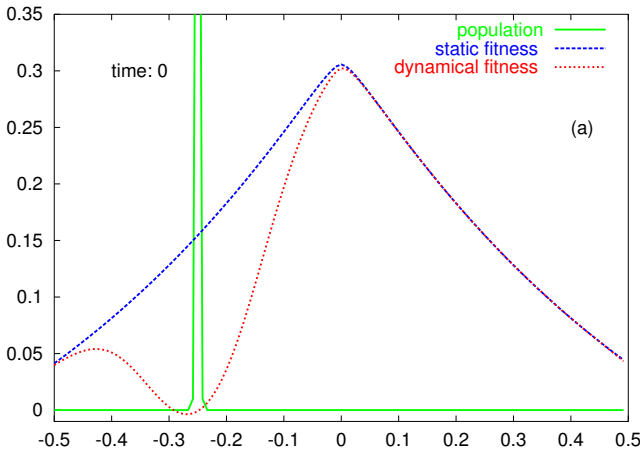
$$\frac{\partial p(g)}{\partial t} = \frac{A(g,\mathbf{p})}{\langle A \rangle} p(g) + \text{mutations},$$

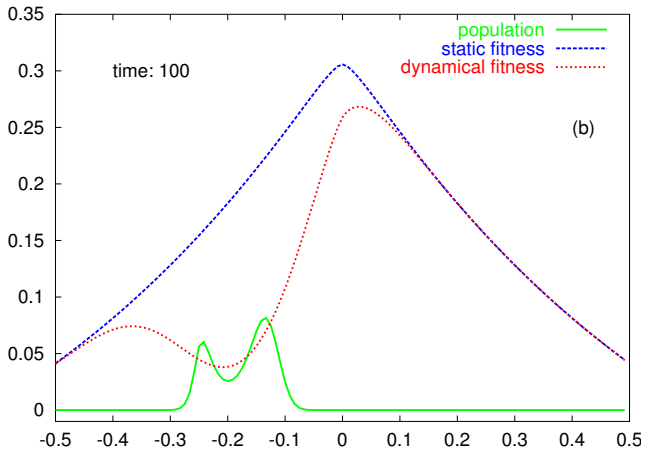
$$A(g,\mathbf{p}) = \exp\left( H_0(g) + \sum_{g' \simeq g} H_1^{(i)}(g,g')p(g') + \sum_{g' \neq g} H_1^{(e)}(g,g')p(g') + \dots \right).$$

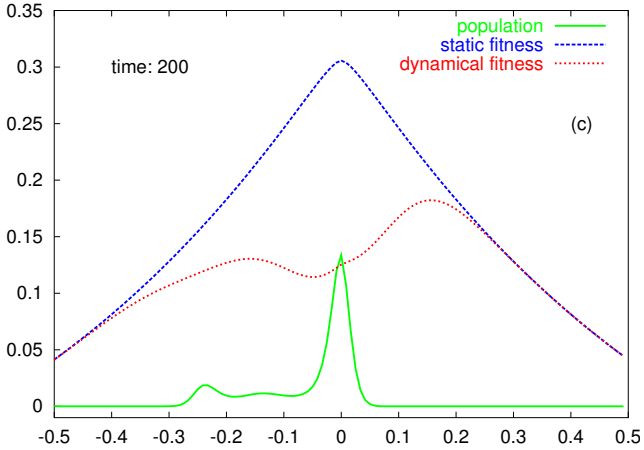
- One can include simple competition terms: intraspecies ($H_1^{(i)}$) and interspecies ($H_1^{(e)}$) contributions.
- Intraspecie competition broadens the distribution, favouring dispersion.
- Interspecies competition stabilizes coexistence.
- Predation induces competition between predators and between preys.

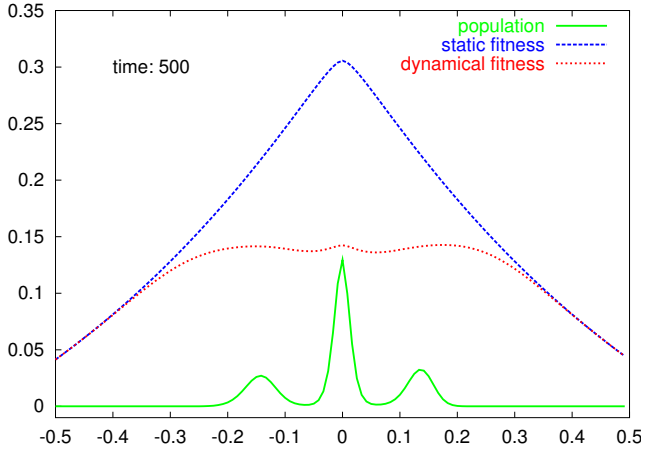
# Competition.

- Short-range competition stabilizes coexistence



# The structure of evolutive patterns

Is evolution dominated by mutations (punctuated equilibrium) or by selection?

- Does evolution walks on a fitness landscape (evolutive times determined by mutations) or is it dominated by coevolution?
- Once that one removes the parts of genotypic space that are unviable (proteins that do not fold, not working metabolic pathways...), how much space is left? Does phylogenetic trees represent a vanishing portion of the available space or not?
- Are transitions (exctinctions, speciations) result of random walks (hopeful monsters), slow effects of coevolution or externally driven?

# Sex, finally!

- Sex (actually, recombination) increases variability without increasing mortality by mutations (that also prevent increasing the genome length due to error threshold).
- Sex/recombination is advantageous for escaping predation or parassitism.
- With diploidity (or horizontal trasmission of genes) genes has more freedom to rearrange (but they have to cooperate, if interested in the survival or the organism...).
- Although sex may be expensive, it may be advantageous in variable environments.
- Sexual selection can be much more effective than fitness for speciation.

# Doebeli and Dieckman model (Nature **400**, 354)

- Diploid or haploid individuals
- Three sets of genes (loci): ecological character $x$, mating preferences $m$ and marker trait $y$.
- The static fitness landscape id unimodal in $x$ (favors intermediate $x$).
- Competition however, favors differentiation on $x$.
- Mating alleles are $+$ and $-$, $m$ is the sum. $m > 0$ is assortative, $m = 0$ is random and $m < 0$ is disassortative.
- Two cases: mating depends on $x$ or $y$.
- When mating depends on the ecological trait $x$, sympatric speciation occours quickly.
- Also when mating depends on the neutral mark $y$, sympatric speciation occours, but more slowly.

# Genes and brains

- Brains "are used" by genes as a fast-adaptation device.
- Genes are shaped by brains, expeciallythrough sexual selection.
- Genes are switched on and off by brains, but brains are determined by genes...

# Games

- Game theory has changed the way we look at evolution.
- We can think to a population of agents (automata), that interact (play the game).
- A strategy can be seen as a "rule" for an automata, either deterministic (pure rules) or stochastic (mixed rules).
- The ingredients to the strategy are present and past states.
- The selection of the best strategy depends on the payoff (evolutionary stable strategies).
- Individuals whose strategies give higher payoff reproduce (and mutate).

# Example: evolution of cooperation

- Interspecies interaction can be divided into predation (or parassitism), cooperation and competition.
- Predation may lead to interspecies competition (your enemy is my friend).
- Cooperation is strongly related to competiton: there is no"natural" cooperation (genes are selfish).
- Nevertheless, cooperation is common: genes, social insect, muticellular organisms, slime molds, even humans...
- So, what are the evolutionary basis of cooperation?

# Payoff

Suppose that there are only two states (0 and 1 or cooperate (C) and defect (D)), and that the strategy depends only on recent states.
Payoff table (for one of participant):

| payoff | C | D |
|---|---|---|
| C | $\alpha$ | $\beta$ |
| D | $\delta$ | $\gamma$ |

Let us denote with $p$ the probability of "playing" $C$ (and $1-p$ the probability of playing $D$). The payoff is related to fitness:

$$p' = \frac{\alpha p^2 + \beta p(1-p)}{\alpha p^2 + \beta p(1-p) + \delta p(1-p) + \gamma(1-p)^2}$$

neglecting fluctuation, the asymptotic state is given by the fixed points of the previous equations.

# Scenarios (Nowak, Science **314**, 1560 (2006)

A society of cooperators has an higher fitness that a society of defectors, but is in general susceptible to invasions or mutation by defectors. Possible scenarios are:

- (D) If the only stable point is $p^* = 0$, defectors always dominates.
- (ESS) If there are two stable points, $p^* = 0$ and $p^* = 1$, but the latter has only a tiny basin, then cooperators are evolutive stable strategie (ESS): a single mutant cannot invade, but multiple ones can, helped by fluctuations. $\alpha > \beta$.
- (RD) If the basin of $p^* = 1$ extends up to $p_0 = 1/2$, it is favoured by fluctuations. $\alpha + \gamma > \beta + \delta$
- (AD) The basin extends up to $p_0 = 2/3$ can be shown to be related to stability respect to fixation of a single mutant (Kimura theory). $\alpha + 2\gamma > \beta + 2\delta$
- (C) Finally, if the basin extends up to $p_0 \simeq 1$, cooperators always dominate (except for $p_0 = 1$).

# Standard payoff

| payoff | C | D |
|---|---|---|
| C | $b - c$ | $-c$ |
| D | $b$ | 0 |

Benefits (may be long term like parental cares) have to be greater than costs. However, defectors always wins.

# Kin selection

Haldane one said: "I will jump inot the river to save two brothers or eight cousins". If you cooperate with a relative that shares a fraction $r$ of your genes, then your payoff (well, that of your genes) is augmented by a fraction $r$ of the payoff of your opponent.

| payoff | C | D |
|---|---|---|
| C | $(b-c)(1+r)$ | $br-c$ |
| D | $b-rc$ | 0 |

For $\frac{c}{b} < r$, cooperation is ESS,RD,AD.
Kin selection says that cooperation among relatives (cells in multicellular organisms, social insects) derives by selfishness of genes.

# Direct reciprocity

For a "one shot" game the best strategy is to defect. But Axelrod discovered by computer experiments that for repeated games between two opponents it is best to cooperate and to forgive: tit for tat or similar strategies.
In this case the parameter is the probability $w$ of another encounter (or the expected number $1/w$ of rounds).

| payoff | C | D |
|---|---|---|
| C | $(b-c)/(1-w)$ | $-c$ |
| D | $b$ | 0 |

- ESS for $c/b < w$
- RD for $c/b < w/(2-w)$
- AD for $c/b < w/(3-2w)$

# Reputation (indirect reciprocity)

For humans, reputation is a valuable quantity. It is defined as the average cooperation-to defection record. It may be known with a probability $q$. If you know that the opponent is a defector, defect, otherwise cooperate.

| payoff | C | D |
|---|---|---|
| C | $(b-c)$ | $-c(1-q)$ |
| D | $b(1-q)$ | 0 |

- ESS for $c/b < q$
- RD for $c/b < q/(2-q)$
- AD for $c/b < q/(3-2q)$

# Network reciprocity

Human societies are structurate networks. For a given connectivity $k$ of a node, a cooperator pays a cost $c$ and each of the neighbors receive a benefit $b$.

| payoff | C | D |
|---|---|---|
| C | $(b-c)$ | $H-c$ |
| D | $b(1-H)$ | 0 |

with $H = \frac{(b-c)k-2c}{(k+1)(k-2)}$.
ESS, RD, AD for $c/b < 1/k$

# Group selection

Group selection is based on the higher payoff of cooperation, but one has to find a mechanism for stabilizing it against defectors. The idea is that the society automatically splits into $m$ groups of size $n$. Cooperators help only inside groups, and successful groups split more often.

| payoff | C | D |
|---|---|---|
| C | $(b-c)(n+m)$ | $bm-c(m+n)$ |
| Theoretical and D | $bn$ | $0$ |

ESS, RD, AD for $c/b < m/(m+n)$.

# Conclusions

- Theoretical and computer studies cannot say "this model explains this observed effect"
- But the can check the consistency of hypothesis with an effect.
- They may also be useful in testing the robustness with respect to parameters.