Why is there “sex”?
And
Is “sex” with others better than “sex” with yourself?

Chris Stephens
C3 - Centro de Ciencias de la Complejidad y Instituto de Ciencias Nucleares
Universidad Nacional Autónoma de México
Seminar FQ, UNAM
21/04/2017

What is sex?

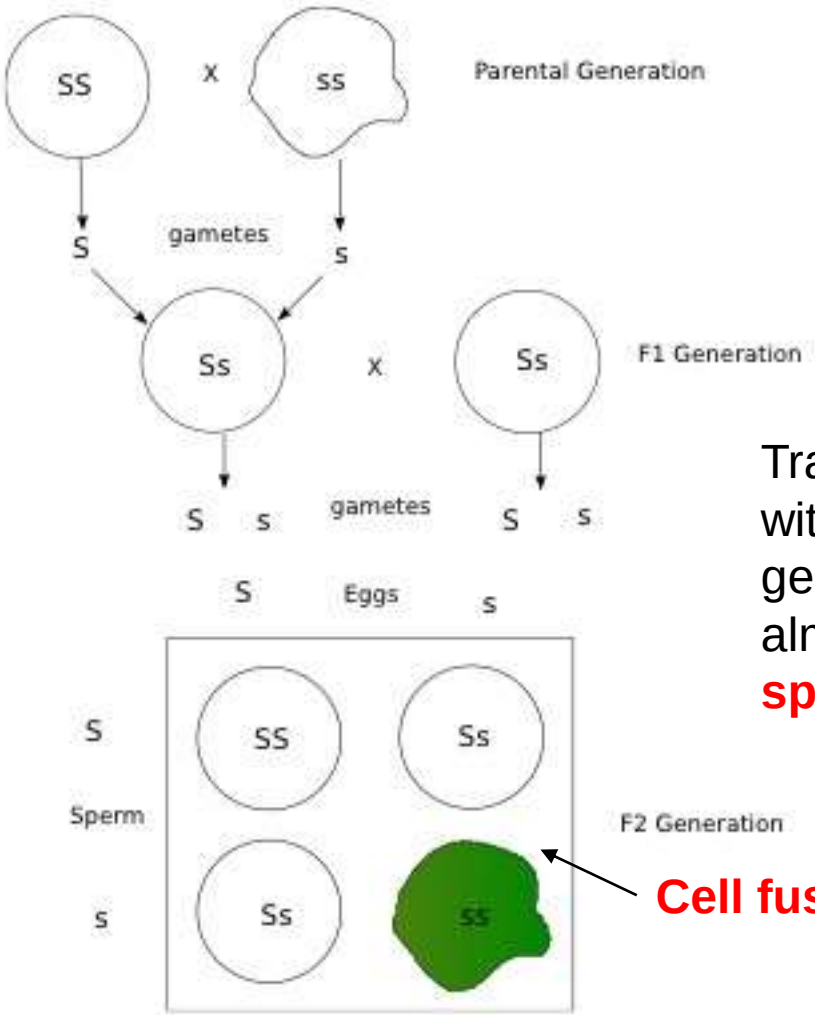


Fig. 2.1 Representation of Mendel's experiment: hybrid (Ss) smooth peas are bred together in generation F_1, resulting in a three-to-one ratio of smooth to wrinkled peas in generation F_2. Wrinkled peas are produced because the hybrid smooth peas have both a dominant gene (S) and a recessive one (s). When the recessive genes combine (ss), the peas appear wrinkled.

The production of new living organisms by combining genetic information from **two** individuals of **different types** (sexes).

Traditional sex is associated with the **vertical** transfer of genetic material and occurs almost exclusively **within a species**

There are two distinct sources of variation

Cell fusion

Homologous recombination

Dominance means that only part of the genome is expressed

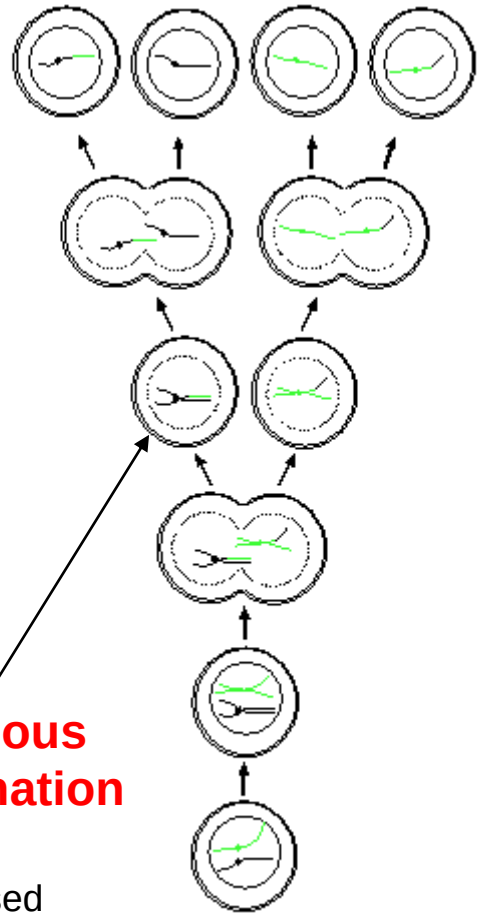
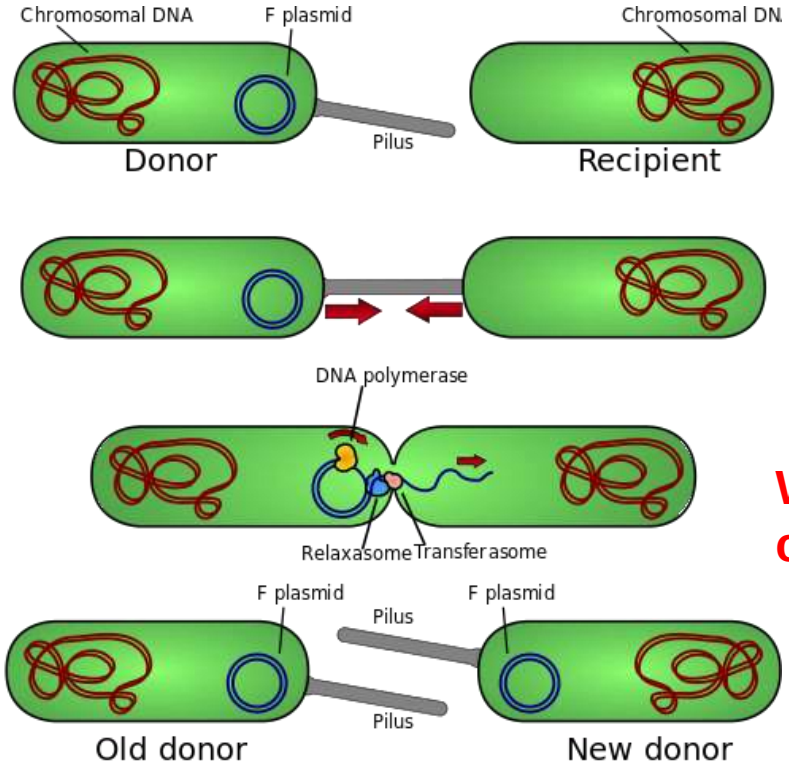


Fig. 18.1 The [illegible]

What is sex?

"Sex" can also occur via
Lateral Gene Transfer

For example, in bacterial conjugation

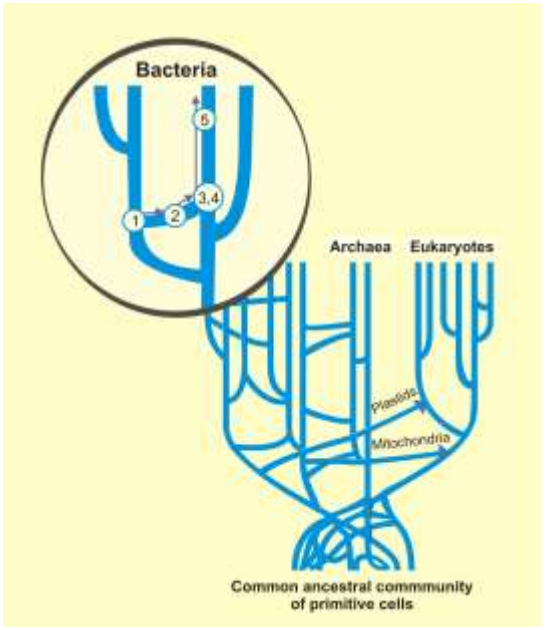


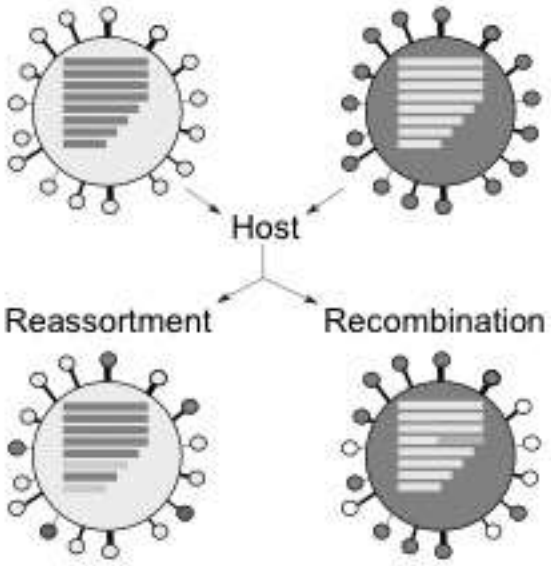
In a world with LGT your family tree is a lot more complicated!

You can even have more than two "parents"!

There are varying degrees of homology

What genetic material can be transferred?





What is sex?

I will consider “sex” in the most general sense to be the **recombination** of genetic material from one or more genomes to create a new “offspring” genome.

It is distinct from mutation in that it does not produce de novo genetic sequences. It uses already created genetic sequences and mixes and redistributes them.

The Genotype-Phenotype Map

Indirect genotype-phenotype map and gene expression

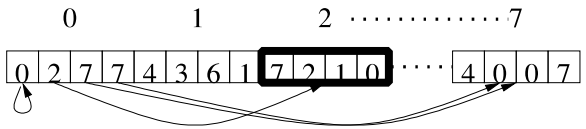
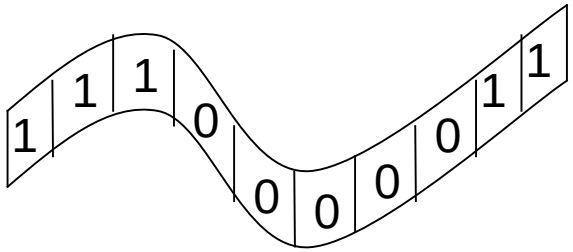


Fig. 1. An example of the switchboard gene in operation. Each codon in the switchboard gene (in the current version, always located in the first position) acts as an index into the *entire* genome. Notice how this particular switchboard gene indexes itself.

Direct genotype-phenotype map and gene expression



One gene – one protein

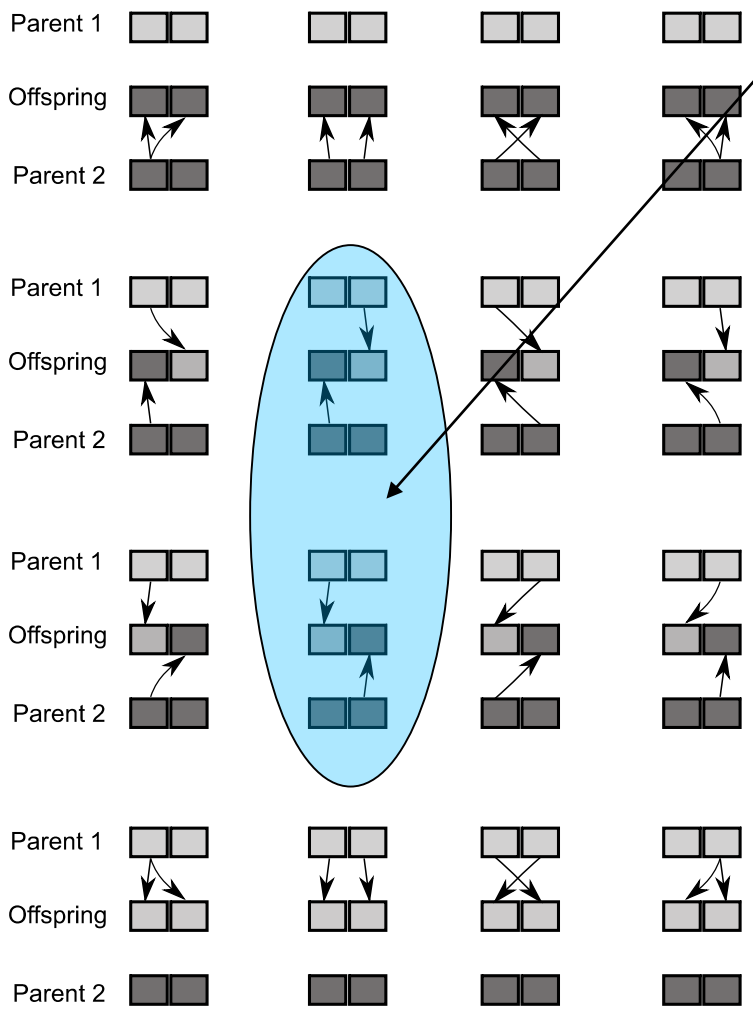
Not all genetic material is expressed
Coding can be redundant

```
+++X XXX*
+X+X +XX*
```

Both genomes code for 4X

Genomes are potentially enormously fragile

Representing “sex” in all its varieties



Homologous versus non-homologous recombination

And this is only for fixed length genomes!

Generalised Crossover Mask (GCM) $r = (m, v)$

Example 11.1 *Standard one-point crossover for $\ell = 3$*
As a first example of how the representation of a GCM works, let us consider standard one-point crossover for $\ell = 3$. The associated traditional crossover masks are 100 and 110, each invoked with probability $\frac{1}{2}$. These are equivalent to the GCMs $r_1 = (100, (1, 2, 3))$ and $r_2 = (110, (1, 2, 3))$.

Example 11.4 *Class II transposition as generalised recombination for $\ell = 4$*
An example of transposition considered as a GRD in the setting of a four-locus system is the following:

$$p_c(1111, (4, 1, 2, 3)) = p_c(1111, (1, 3, 2, 4)) = p_c(1111, (4, 3, 1, 2)) = \frac{1}{3}$$

This GRD represents a four-locus system where there is equal probability for the occurrence of three different transposition events. In the first, the fourth locus — last gene on the chromosome — is cut from its position and transposed to the first locus. In the second, the gene at the third locus is cut from its position and transposed to the second locus. Finally, the last example is a double transposition, where the fourth gene is cut and then pasted to the first locus while the third gene is cut and then pasted into the second position.

Fig. 11.1 Action of the 16 generalised recombination masks in Equation 11.3 uals of length $\ell = 2$.

Why does “sex” exist?

Why bother with two sexes if one is enough?

The two-fold cost of sex in sexual versus asexual reproduction.
Moreover, you need to find a mate, attract it, avoid being eaten, forego resources,…
Also, from a selfish gene standpoint you’re only transmitting 50% of your genes.
Finally, recombination can break up useful gene combinations

- Sexual reproduction (homologous recombination and sex) provides variation on which natural selection can act.
    - Sex doesn’t have to increase variation relative to selection and even if it does it doesn’t mean it increases fitness

- Sexual reproduction can better remove deleterious mutations
    - Finite population intuition

- Sexual reproduction can better produce advantageous
    - Finite population intuition

- Recombination was designed to maintain reproductive fidelity (reduce variation!) in DNA. “Sex” is then just an “accidental” by-product.

- Other more complicated, not universal sounding, reasons

How do we test hypotheses?

1. Interpreting and explaining observational evidence
2. Developing mathematical models
    a) What model framework?

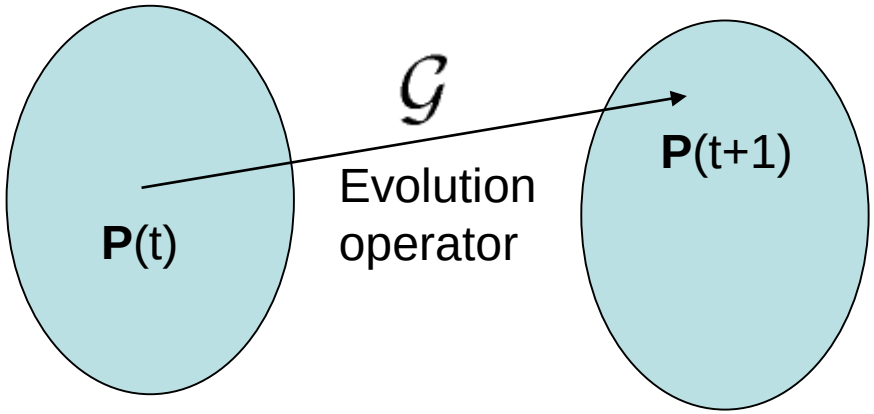
Of course, if we are to understand the benefits of recombination in the context of a mathematical model, a requirement is that the model itself captures the very mechanisms by which it is useful in the first place. This then leads us to ask if the apparent inability to find an agreed universal advantage for recombination is due to the fact that the considered models are incapable of modeling the benefits — a defect of the model — or, rather, that the benefits are not transparent in the analyses of the models that have been studied. If the models themselves are inadequate then new models with new features must be developed. On the contrary, if the analyses themselves are at fault, one must understand why.

Population genetics

What is Genetic Dynamics?

Population of “objects” – “genotypes”

$$\mathbf{P}(t) = (P_1(t), P_2(t), \cdots, P_\Omega(t))$$



determines the state of the population at time t; Ω is the dimension of the space of states of an “object”; for linear chromosomes with binary alleles $\Omega = 2^N$

Space of populations $\mathcal{P}$

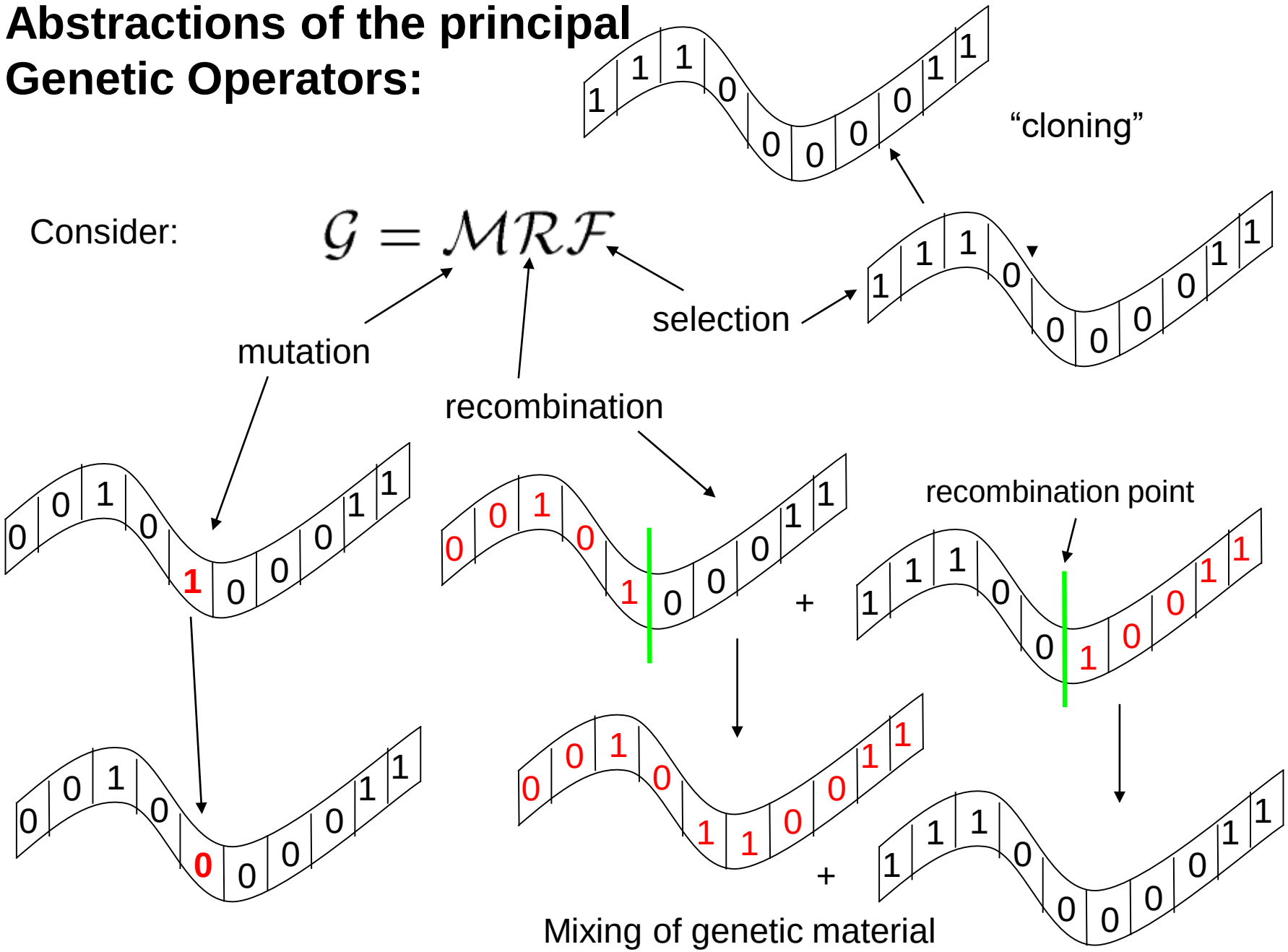
General evolution equation

$$\mathbf{P}(t+1) = \mathcal{G}(\mathbf{P}(t), \mathbf{p})$$

p represents a set of parameters associated with the evolution operator

Expected next population for finite Populations.
Describes evolution?
Fixed length strings…

Abstractions of the principal Genetic Operators:



In mathematics…

Finite population model determined by Markov chain. In the infinite population limit for haploids:

$$P_I(t+1) = M_I{}^J((1-p_c)P'_J(t) + p_c \sum_m p_c(m)\lambda_J{}^{KL}(m)P'_K(t)P'_L(t))$$

That's most of standard population genetics and evolutionary computation!

Implicit summation over repeated indices

$M_I{}^J$ Probability to mutate genotype J to genotype I

p_c Probability to implement recombination

$p_c(m)$ Probability that given recombination takes place it is implemented with mode m

$P'_I(t)$ Probability to select genotype I $P'_I(t) = \frac{f(I)}{\bar{f}(t)} P_I(t)$

$\lambda_J{}^{KL}(m)$ Conditional probability for "child" J given "parents" K and L and a mode m

Don't recombine it with another

Select an object J

Select two "parents" K and L ("phase space")

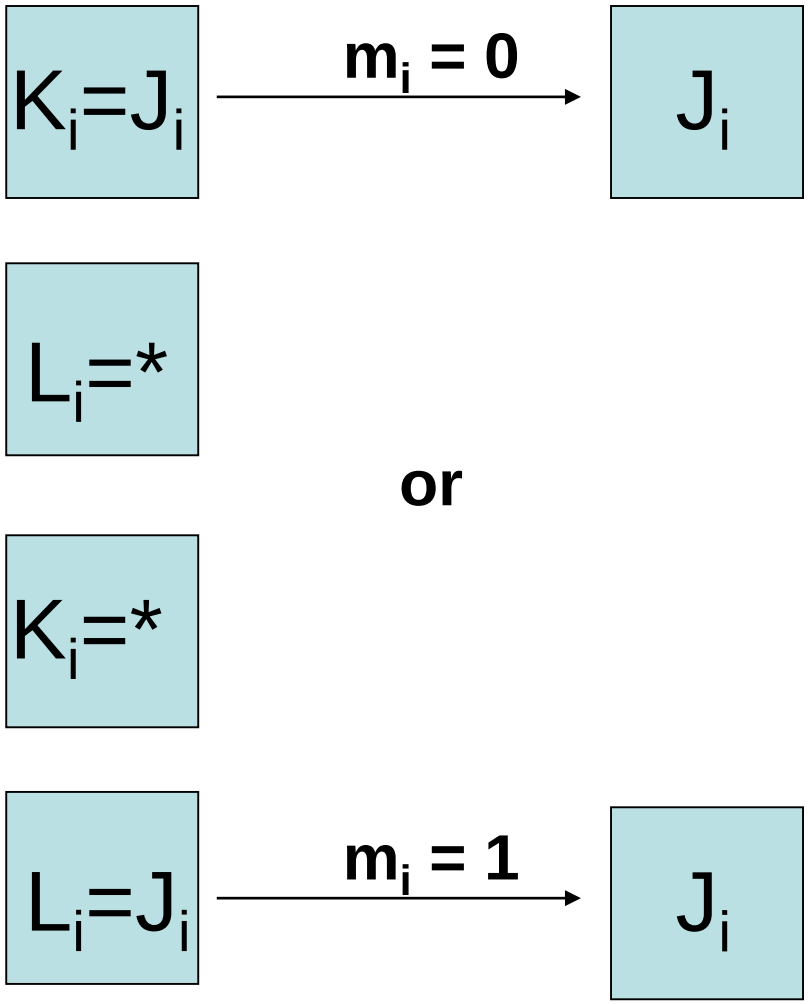
$$P_I(t+1) = M_I{}^J((1-p_c)P'_J(t) + p_c \sum_m p_c(m)\lambda_J{}^{KL}(m)P'_K(t)P'_L(t))$$

Mutate it to object I

Recombine them with respect to a recombination mode m applied with probability $p_c p_c(m)$ to obtain a "child" J

- **Ω coupled non-linear difference equations**
- **Population genetics has spent the last 70 years trying to deal with them**
  - Go to reduced number of loci
- **In object basis there are Ω^3 different λ_J^{KL} - that's a lot!**
  - Most of them are 0!

Recombinative Dynamics



In recombination, at every locus, one of the parental alleles is ***always*** coarse-grained

Every m defines a particular coarse-graining

Here its “homologous” recombination which means that the ith locus of the child string comes from the ith locus of a parent string. This formalism generalises to the case where the ith locus of the child comes from ANY locus of the parent

Recombinative Dynamics

$$\lambda_J{}^{KL}(m) = \prod_{i=1}^{N} \lambda_{J_i}{}^{K_i L_i}(m_i) = \prod_{i=1}^{N}((1-m_i)\delta_{J_i}{}^{K_i} + m_i \delta_{J_i}{}^{L_i})$$

Product of locus-wise projection operators

$$\sum_{K_i} \sum_{L_i} ((1-m_i)\delta_{J_i}{}^{K_i} + m_i \delta_{J_i}{}^{L_i})$$

If $m_i = 0$ (take allele for first locus of "child" from first locus of first parent) then

$$\sum_{K_i} \sum_{L_i} ((1-m_i)\delta_{J_i}{}^{K_i} + m_i \delta_{J_i}{}^{L_i}) P_{K_1 \cdots K_i \cdots K_N}(t) P_{L_1 \cdots L_i \cdots L_N}(t)$$

$$= P_{K_1 \cdots J_i \cdots K_N}(t) P_{L_1 \cdots *_i \cdots L_N}(t)$$

where $*_i$ means we have marginalized the probability at the ith locus

Similarly, for $m_i = 1$

$$= P_{K_1 \cdots *_i \cdots K_N}(t) P_{L_1 \cdots J_i \cdots L_N}(t)$$

Recombinative Dynamics

$$P_I(t+1) = M_I{}^J((1-p_c)P'_J(t) + p_c \sum_m p_c(m) P'_{J_m}(t) P'_{J_{\bar{m}}}(t))$$

So?!

Where's λ_J^{KL} gone?

Every m, i.e., coarse-graining mode, for given target object J defines a "Building Block" J_m. At the same time this uniquely defines a conjugate Building Block $J_{\bar{m}}$ that is the set complement of J in J_m.

This coarse-graining can also be implemented as a coordinate transformation using a transformation matrix

$$\Lambda = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^{\otimes N}$$

In this basis $\lambda_J^{KL}(m)$ for a given m has only one non-zero entry and it's on the skew diagonal

Recombinative Dynamics

- Thus we see how recombination “works” by taking BBs and recombining them into strings
- If $\Delta_I(m) = P'_I - P'_{I_m} P'_{I_{\bar{m}}}$ (Selection Weighted Linkage Disequilibrium Coefficient) > 0 then recombination is bad for the formation of that string and good if < 0 (more construction then destruction.
- But if we want to “solve” the dynamics have to know what happens to the BBs! E.g. what´s the equation for I_m? Need to coarse grain the string equation

Recombinative Dynamics

$$P_{I_m}(t) = \sum_{\{i:m_i=1\}} P_{I_1 I_2 \cdots I_i \cdots I_N}$$

Projection operator $\mathcal{R}(\eta, \eta')$

$$\mathcal{R}(\eta, \eta'') = \mathcal{R}(\eta, \eta')\mathcal{R}(\eta', \eta'')$$

Renormalization (semi)-group

$$P_{I_m}(t+1) = M_{I_m}^{\ J_m}((1-p_c)P'_{J_m}(t) + p_c \sum_{m'} p_c(m') P'_{J_{m m'}}(t) P'_{J_{m \bar{m}'}}(t))$$

BBs of the BB J_m

Note the form invariance under the coarse graining
Strings are built up from BBs which in turn have their BBs
which … the hierarchy ends at BBs with only one locus, e.g. ***1*****

How to measure the benefits of “sex”?

$$D_{P_I}(t) = (P_I(t+1) - P_I(t))$$

Does it give you more of a particular fit string I?

$$D_{P_I}(t) = \left(\frac{f_I}{\bar{f}(t)} - 1\right) P_I(t) - p_c \sum_m p_c(m) \Delta_I(m,t)$$

$$D_{\bar{f}_{r+s}}(t) = \bar{f}_{r+s}(t+1) - \bar{f}_s(t+1)$$

Does it improve overall population fitness relative to selection only dynamics?

$$= \frac{\bar{f}^2_{s+r}(t)}{\bar{f}_{s+r}(t)} - \frac{\bar{f}^2_s(t)}{\bar{f}_s(t)} - p_c \sum_{m,I} p_c(m) f_I \Delta_I(m,t)$$

Indirect effect from recombination

Direct effect from recombination

Note how $\Delta_I(m)$ plays a crucial role in determining the efficacy of recombination. If $\Delta_I(m)$ >/< 0 you get less/more of the fit string due to recombination

Note how the benefits depend on the fitness landscape and on the actual population

Fitness Landscapes

$$f_x = f_{x_1 x_2 \ldots x_{i_\ell}}$$

$$= F^{(0)} + \sum_{i_1=1}^{\ell} F^{(1)}_{i_1} x_{i_1} + \sum_{i_1=1}^{\ell-1} \sum_{i_2=i_1+1}^{\ell} F^{(2)}_{i_1 i_2} x_{i_1} x_{i_2}$$

$F^{(n)}_{i_1 i_2 \ldots i_n}$ represents an epistatic interaction between n alleles located at loci $i_1, i_2, \ldots, i_n$ and $x_{i_n} = 0, 1$

$$+ \sum_{i_1=1}^{\ell-2} \sum_{i_2=i_1+1}^{\ell-1} \sum_{i_3=i_2+1}^{\ell} F^{(3)}_{i_1 i_2 i_3} x_{i_1} x_{i_2} x_{i_3} + \cdots + F^{(\ell)}_{i_1 i_2 \ldots i_\ell} x_{i_1} x_{i_2} \ldots x_{i_\ell}$$

For two loci $$f_{x_1 x_2} = F^{(0)} + \sum_{i_1=1}^{2} F^{(1)}_{i_1} x_{i_1} + F^{(2)}_{12} x_1 x_2$$

$$f_{00} = F^{(0)} = a,$$

$$f_{01} = F^{(0)} + F^{(1)}_2 = a + b_2,$$

$$f_{10} = F^{(0)} + F^{(1)}_1 = a + b_1,$$

$$f_{11} = F^{(0)} + F^{(1)}_1 + F^{(1)}_2 + F^{(2)}_{12} = a + b_1 + b_2 + c$$

For an additive (modular) landscape $F^{(2)}_{12} = 0$. For a multiplicative landscape $F^{(0)} F^{(2)}_{12} = F^{(1)}_1 F^{(1)}_2$. For a redundant (modular) landscape $F^{(2)}_{ij} = -F^{(1)}_i = -F^{(1)}_j$ which, as mentioned, can be understood in terms of a Boolean "OR", fitness being the same if either one or both alleles are optimal. For a NIAH landscape $F^{(1)}_1 = F^{(1)}_2 = 0$ which, in contrast to the redundant landscape, corresponds to a Boolean "AND" as fitness is only different if both alleles are optimal.

For two loci with binary alleles…

Take 11 as optimal genotype $\Delta_{11}(t) = \frac{1}{f^2(t)}(a(a + b_1 + b_2 + c)P_{11}(t)P_{00}(t)$

6 parameter model, 5 if we set b1 = b2 $- (a + b_1)(a + b_2)P_{01}(t)P_{10}(t)).$

First set population bias to zero $P_{ij}(0) = 1/4$

Competition between "constructive" and "destructive" effects of recombination

$$\Delta_{11}(0) = \frac{(c - b^2)}{(1 + b + c/4)^2}.$$

For a multiplicative landscape $c = b^2$ and $\Delta_{11} = 0$ i.e. recombination has no effect

For an additive landscape $c = 0$ $\Delta_{11}(t) = -b^2/(1 + b) < 0$

For a deceptive landscape $b < 0$, but $c > -2b$ and so $\Delta_{11}(t) > 0$

For positive additive epistasis $c > b^2$ $\Delta_{11}(t) > 0$

For negative additive epistasis $c < 0$ $\Delta_{11}(t) = -b(1 + b)/(1 + 3b/4)^2$

For two loci with binary alleles…

$$\sum_{ij} f_{ij}\,\Delta_{ij}(t) = c\Delta_{11}(t) = \frac{c(c - b^2)}{(1 + b + c/4)^2}.$$

For this term to give a positive contribution to the change in average population fitness we require $c(c - b^2) < 0$ In the additive limit c = 0 there is no contribution from this term. For c > 0 we require $c < -b^2$, i.e., weak positive epistasis; while for c < 0 the contribution is negative. Thus, naively what we find for our two metrics is contradictory – at least over one generation starting with an homogeneous population. So,…

- Consider full space of fitness landscapes
- Consider full space of initial populations
- Iterate the dynamical equations over multiple generations

for infinite population and without mutation

Recombination finding optimal genotypes

Initial population $P_{00} \approx 1$

$P_{00}(0) = 0.8999,\ P_{01}(0) = 0.05,\ P_{10}(0) = 0.05,\ P_{11}(0) = 0.0001$

Blue/red = recombination
Advantageous/disadvantageous

Blue/red = recombination
Disadvantageous/advantageous

Multiplicative landscapes

Additive landscapes

Maximal Negative epistasis

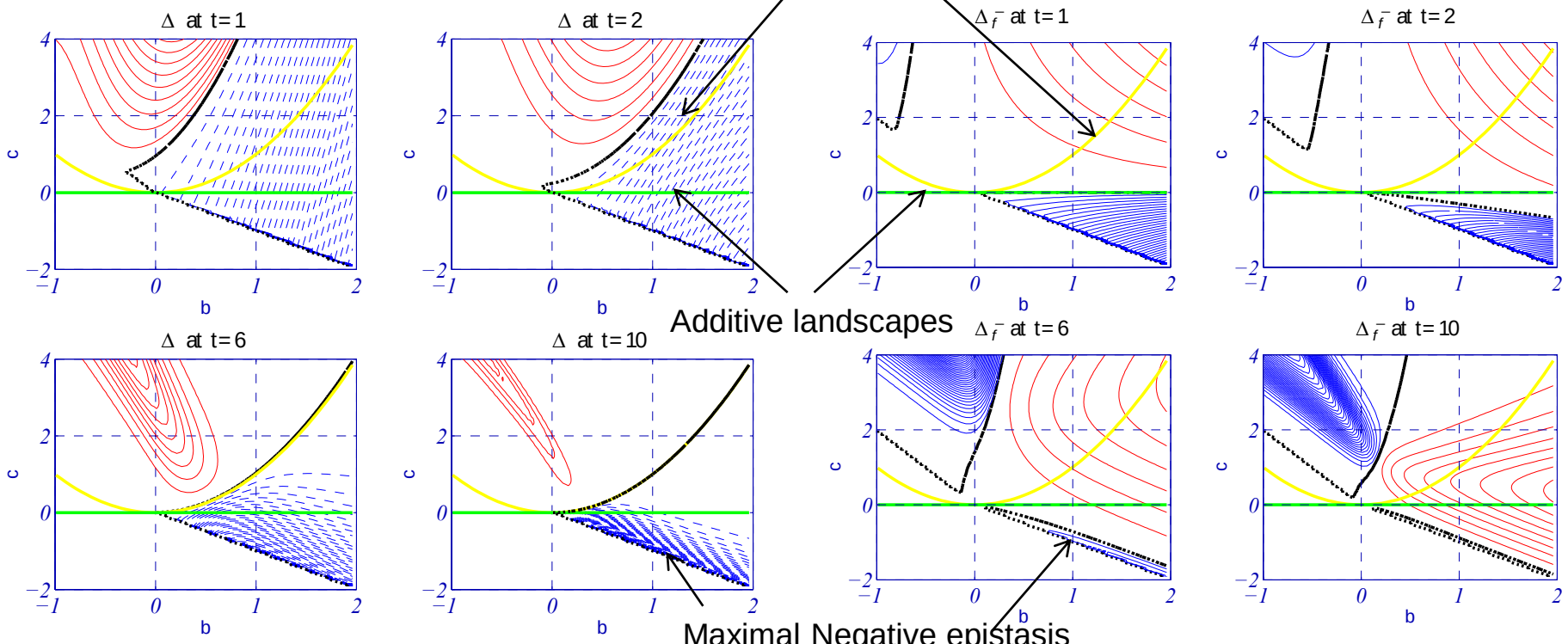


Fig. 1. Value of Δ at different generations for two-locus two-allele system as a function of fitness landscape, characterized by *b* and *c*. The initial population is $P_{00}(0) = 0.8999$, $P_{01}(0) = 0.05$, $P_{10}(0) = 0.05$, $P_{11}(0) = 0.0001$. The $\Delta = 0$ plane has been marked to distinguish between conditions in which recombination is favorable ($\Delta < 0$) or not. The curve on the plane is $c = b^2$, the condition for a multiplicative landscape.

Fig. 2. Value of $D_{f_{r+s}}^-$ at different generations for the two-locus two-allele system as a function of fitness landscape, characterized by *b* and *c*. The initial population is $P_{00}(0) = 0.8999$, $P_{01}(0) = 0.05$, $P_{10}(0) = 0.05$, $P_{11}(0) = 0.0001$. The $D_{f_{r+s}}^- = 0$ plane has been marked to distinguish between conditions in which recombination is favorable ($D_{f_{r+s}}^- > 0$) or not.

$b > -1,\ c > -2b$ and $c > -b$

Recombination getting rid of deleterious mutants

Initial population $P_{11} \approx 1$

$P_{11}(0) = 0.8999$, $P_{10}(0) = 0.05$, $P_{01}(0) = 0.05$ and $P_{00}(0) = 0.0001$

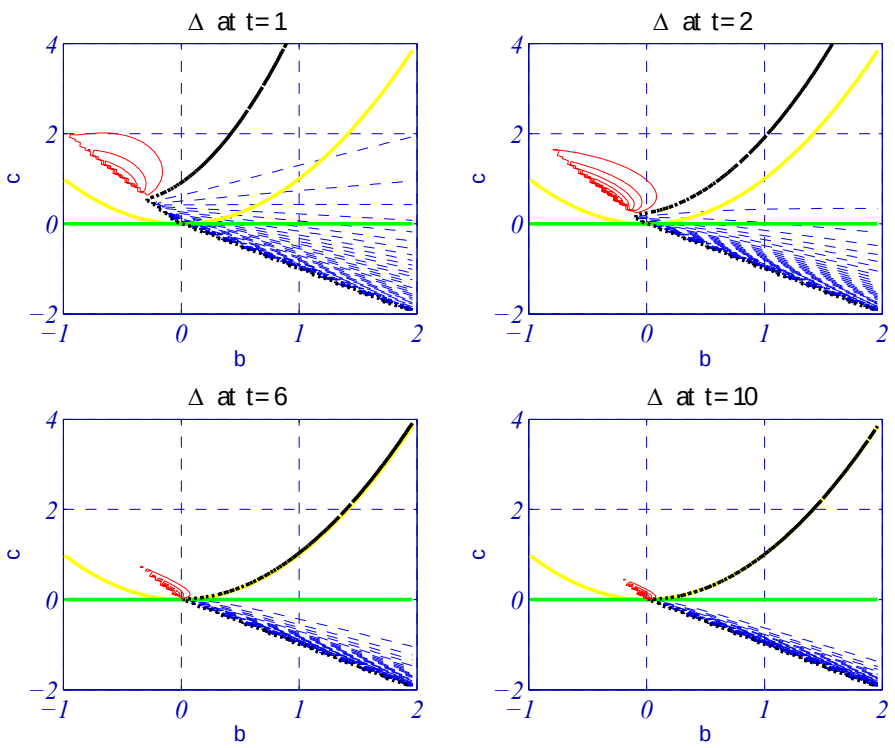


Fig. 3. Value of Δ at different generations for two-locus two-allele system as a function of fitness landscape, characterized by *b* and *c*. The initial population is $P_{11}(0) = 0.8999$, $P_{10}(0) = 0.05$, $P_{01}(0) = 0.05$ and $P_{00}(0) = 0.0001$. The Δ = 0 plane has been marked to distinguish between conditions in which recombination is favorable (Δ < 0) or not. The curve on the plane is $c = b^2$, the condition for a multiplicative landscape.

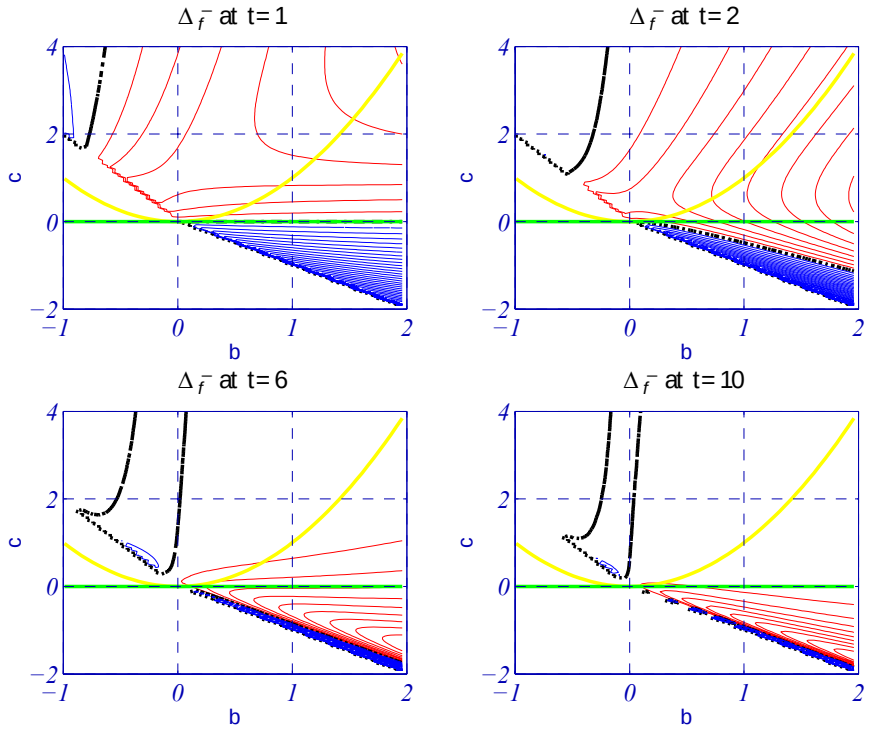


Fig. 4. Value of $D^{-}_{f_{r+s}}$ at different generations for the two-locus, two-allele system as a function of fitness landscape, characterized by *a* and *c*. The initial population is $P_{11}(0) = 0.8999$, $P_{10}(0) = 0.05$, $P_{01}(0) = 0.05$ and $P_{00}(0) = 0.0001$. The $D^{-}_{f_{r+s}} = 0$ plane has been marked to distinguish between conditions in which recombination is favorable ($D^{-}_{f_{r+s}} > 0$) or not.

Recombination and Building Blocks

Initial population $P_{11} \approx 0$, $P_{00} \approx \frac{1}{2}$, $P_{01} \approx P_{10} \approx \frac{1}{4}$

$P_{11}(0) = 0.0001$, $P_{10}(0) = 0.25$ $P_{01}(0) = 0.25$ and $P_{00}(0) = 0.4999$

Large supply of BBs

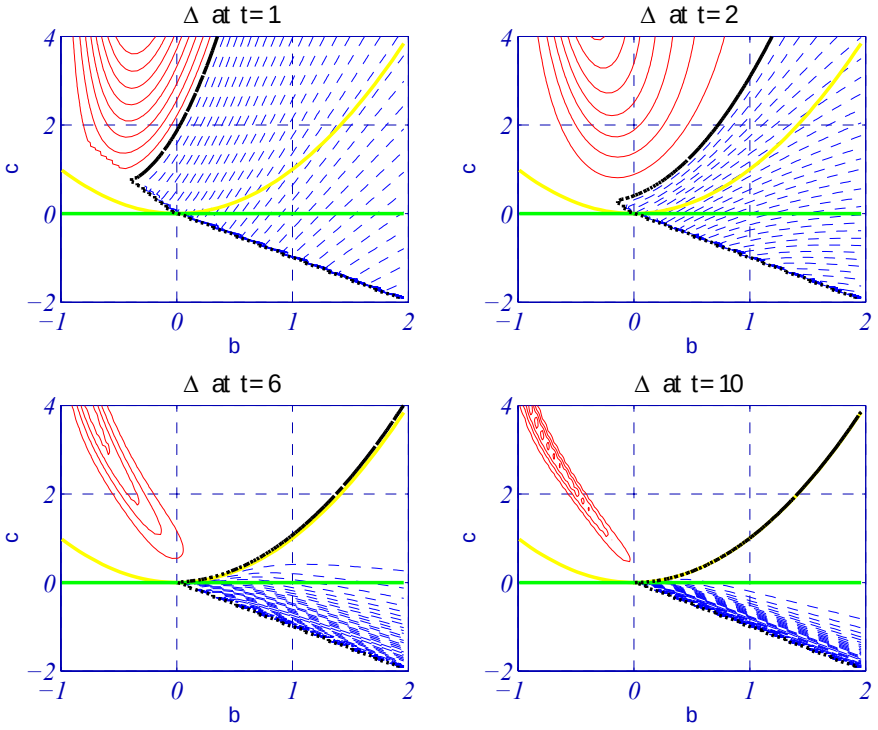


Fig. 5. Value of Δ at different generations for two-locus two-allele system as a function of fitness landscape, characterized by b and c. The initial population is $P_{00}(0) = 0.4999$, $P_{01}(0) = 0.25$, $P_{10}(0) = 0.25$, $P_{11}(0) = 0.0001$. The $\Delta = 0$ plane has been marked to distinguish between conditions in which recombination is favorable ($\Delta < 0$) or not. The curve on the plane is $c = b^2$, the condition for a multiplicative landscape.

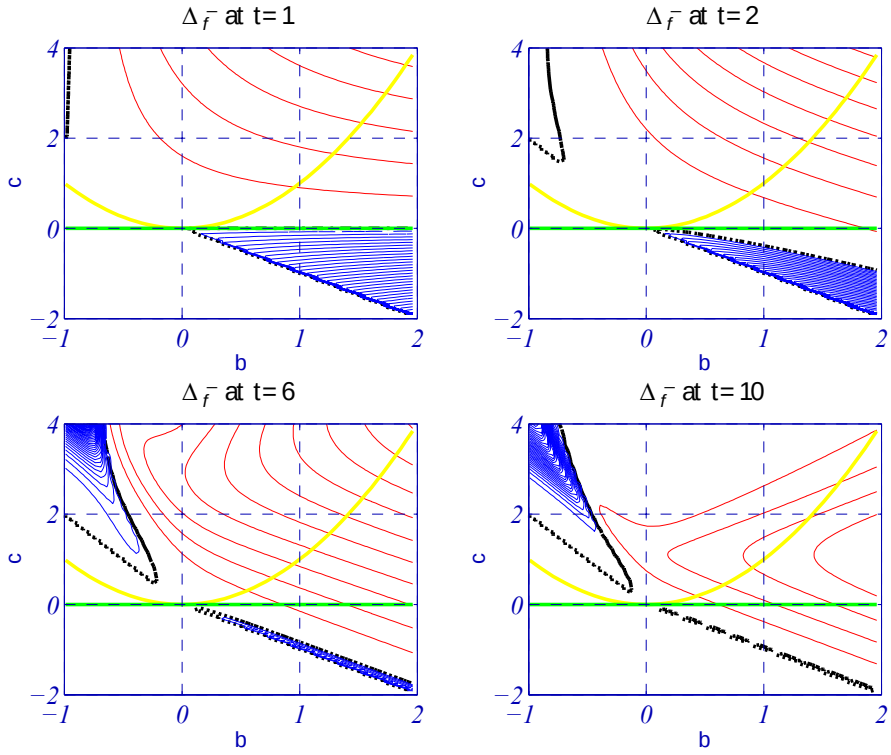


Fig. 6. Value of $D_{\bar{f}_{r+s}}$ at different generations for the two-locus two-allele system as a function of fitness landscape, characterized by b and c. The initial population is $P_{11}(0) = 0.0001$, $P_{10}(0) = 0.25$, $P_{01}(0) = 0.25$ and $P_{00}(0) = 0.4999$. The $D_{\bar{f}_{r+s}} = 0$ plane has been marked to distinguish between conditions in which recombination is favorable ($D_{\bar{f}_{r+s}} > 0$) or not.

Recombination with no population bias

Initial homogeneous population $P_{ij} = 0.25$

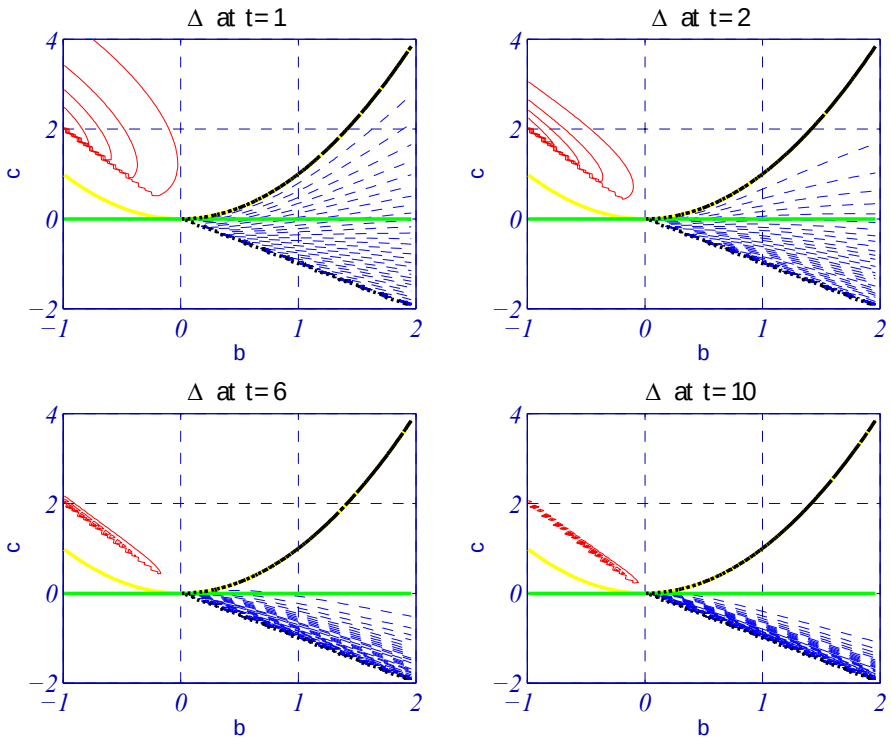


Fig. 8. Value of Δ at different generations for two-locus two-allele system as a function of fitness landscape, characterized by *b* and *c*. The initial population is $P_{ij}(0) = 0.25$. The $\Delta = 0$ plane has been marked to distinguish between conditions in which recombination is favorable ($\Delta < 0$) or not. The curve on the plane is $c = b^2$, the condition for a multiplicative landscape.

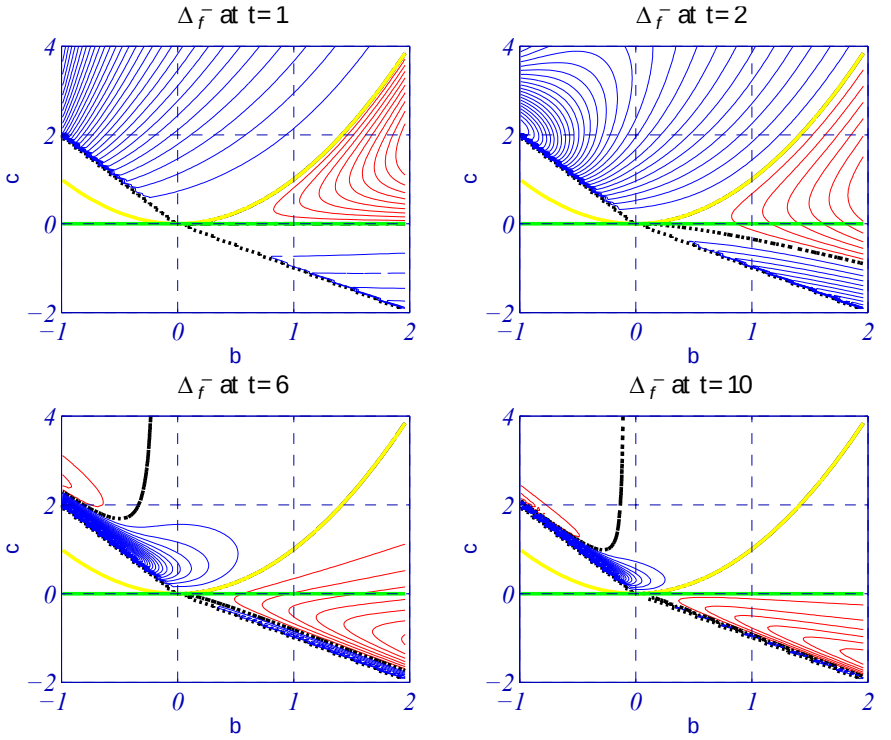


Fig. 9. Value of $D_{f^-_{r+s}}$ at different generations for the two-locus two-allele system as a function fitness landscape, characterized by *b* and *c*. The initial population is $P_{ij}(0) = 0.25$. The $D_{f^-_{r+s}} =$ plane has been marked to distinguish between conditions in which recombination is favorab ($D_{f^-_{r+s}} > 0$) or not.

Recombination as a function of population bias

Additive landscape $a = 1, b_1 = b_2 = 1, c = 0$

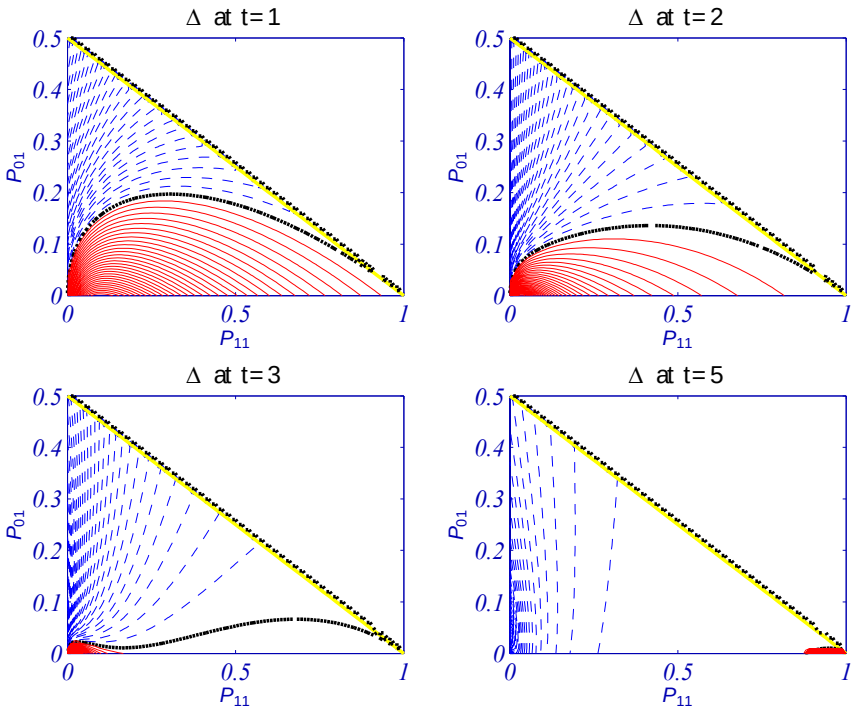


Fig. 10. Value of Δ at different time steps for a two-locus two-allele system with an additive fitness landscape ($a = 1, b_1 = b_2 = 1, c = 0$) for different values of the initial population given by P_{11} and $P_{10}(= P_{01})$. The Δ = 0 plane has been marked to distinguish between conditions in which recombination is favorable (Δ < 0) or not.

The more BBs initially the more "search" is aided

Neutral landscape: $b_1 = b_2 = c = 0, a \neq 0$ ($A = 1$)

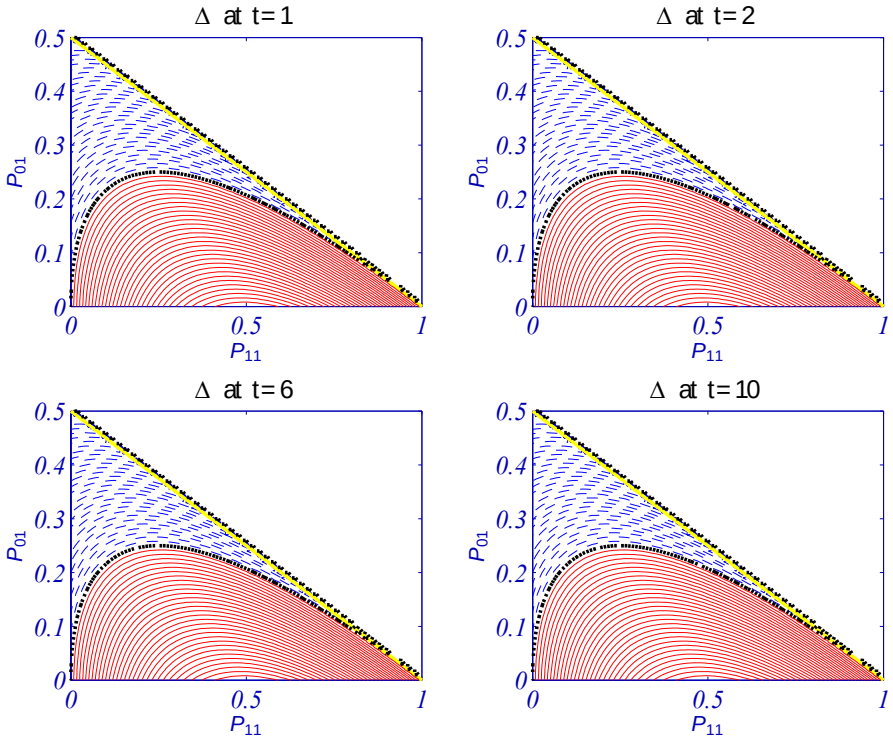


Fig. 11. Value of Δ at different time steps for a two-locus two-allele system with a neutral ($b_1 = b_2 = c = 0, a \neq 0$) fitness landscape for different values of the initial population given by $P_{11}(0)$ and $P_{10}(0) = P_{01}(0)$. The Δ = 0 plane has been marked to distinguish between conditions in which recombination is favorable (Δ < 0) or not.

Any population is a fixed point for a flat landscape

Recombination as a function of population bias

Multiplicative landscape $a = 1,\ b_1 = b_2 = 2,\ c = 4$

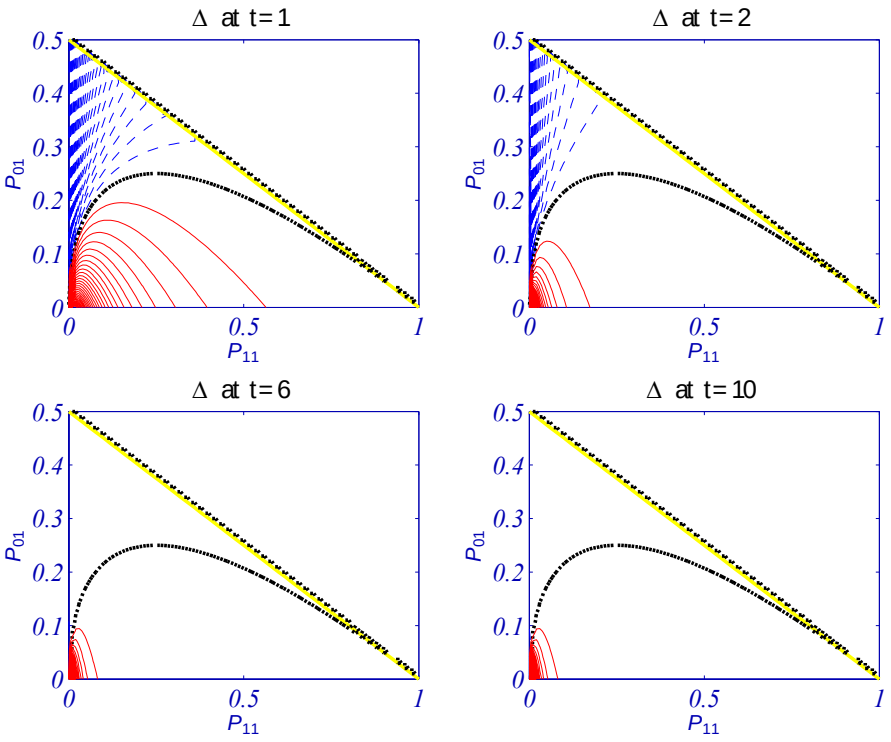


Fig. 12. Value of Δ at different time steps for a two-locus, two-allele system with a multiplicative fitness landscape ($a = 1$, $b_1 = b_2 = 2$, $c = 4$) for different values of the initial population given by P_{11} and $P_{10}(= P_{01})$. The Δ = 0 plane has been marked to distinguish between conditions in which recombination is favorable (Δ < 0) or not.

Needle-In-A-Haystack, $b_1 = b_2 = 0,\ c \neq 0,\ a \neq 0$ ($A = \frac{a}{a+c}$)

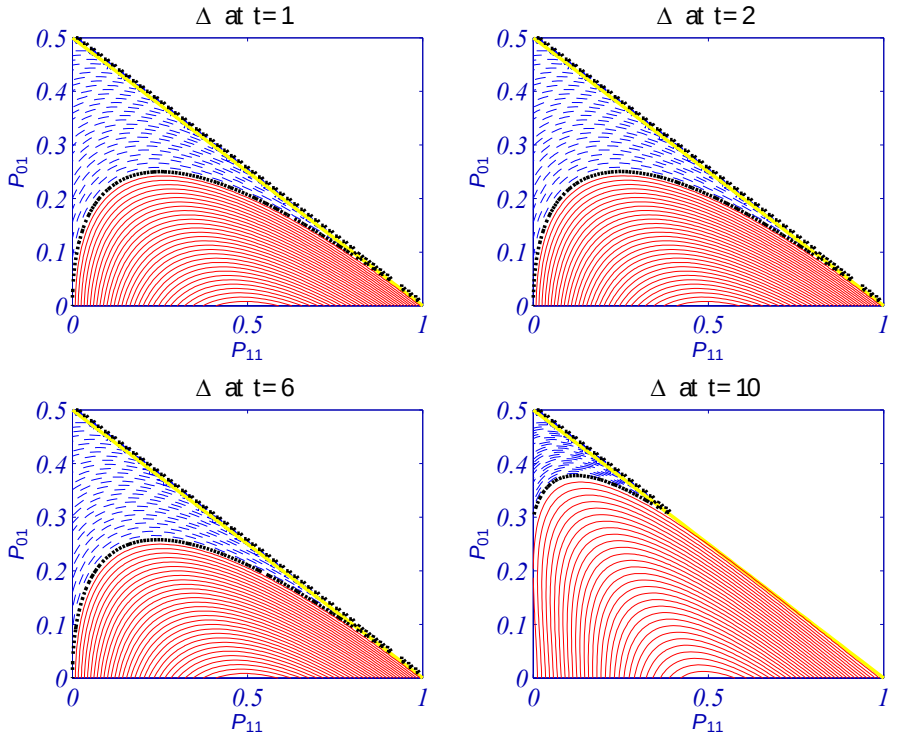


Fig. 13. Value of Δ at different generations for a two-locus two-allele system with a "Needle in a haystack" fitness landscape ($b_1 = b_2 = 0$, $c = 0.001$, $a = 1$) for different values of the initial population given by P_{11} and $P_{10}(= P_{01})$. The Δ = 0 plane has been marked to distinguish between conditions in which recombination is favorable (Δ < 0) or not.

For a highly positively epistatic landscape, recombination is only useful in the transient search regime

Recombination as a function of population bias

Deceptive landscape, $a = 1$, $b = -0.5$, $c = 2$

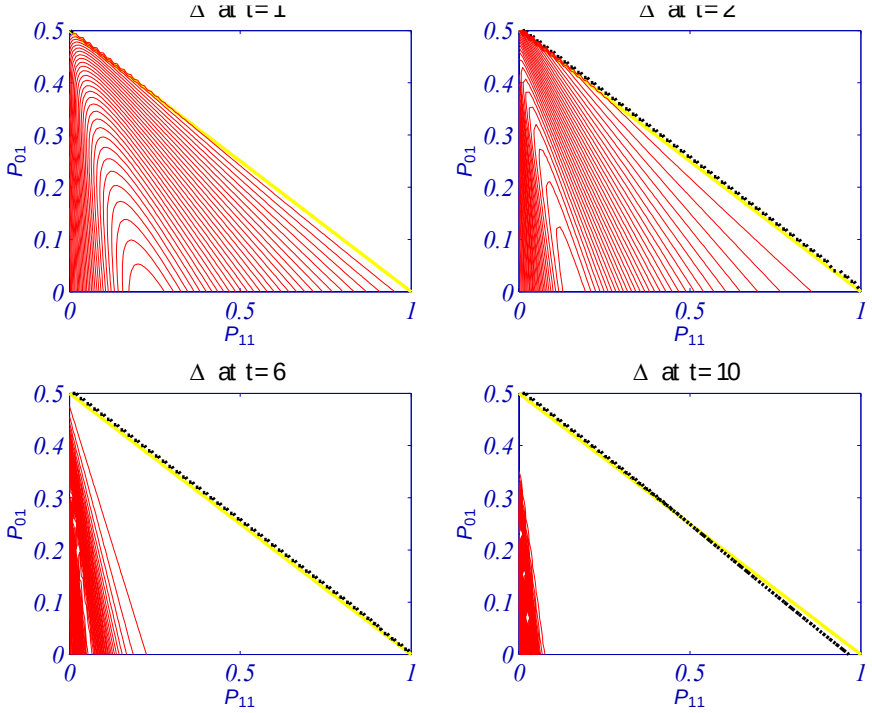


Fig. 16. Value of Δ at different generations for a two-locus two-allele system with a deceptive fitness landscape ($b_1 = b_2 = -0.5$, $c = 2$, $a = 1$) for different values of the initial population given by P_{11} and $P_{10}(= P_{01})$. The $\Delta = 0$ plane has been marked to distinguish between conditions in which recombination is favorable ($\Delta < 0$) or not.

Landscape with genetic redundancy, $a = 1$, $b = 1$, $c = -1$

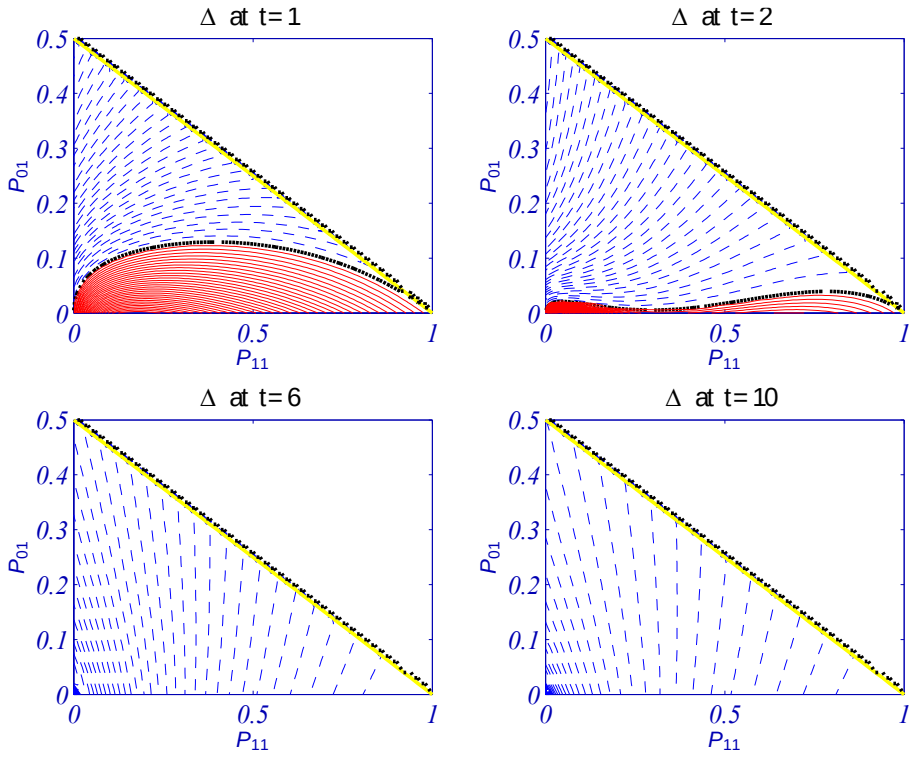


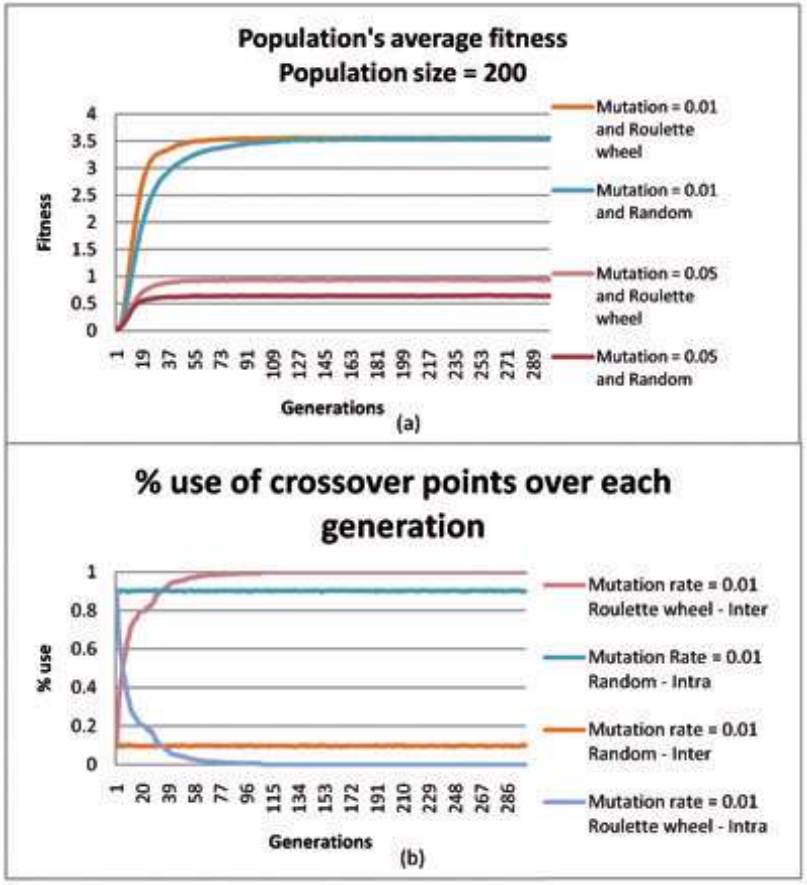
Fig. 15. Value of Δ at different generations for a two-locus two-allele system with a fitness landscape with genetic redundancy ($b_1 = b_2 = 1$, $c = -1$, $a = 1$) for different values of the initial population given by P_{11} and $P_{10}(= P_{01})$. The $\Delta = 0$ plane has been marked to distinguish between conditions in which recombination is favorable ($\Delta < 0$) or not.

Sex and the fitness landscape

- There are two distinct regimes in which recombination (sex) is beneficial
  - the “search” regime, when recombination acts principally as a search operator and is quasi-independent of the fitness landscape
  - the “modular” regime, which is emergent and “universal” (independent of initial conditions and true for both performance metrics) and which is valid only for a relatively small part of landscape space
- The modular regime is characterized by “modular” fitness landscapes
  - Quasi-additive
  - Redundant (negative epistasis/Boolean OR)

- The benefits of “sex” are a trade off between the creative and destructive effects of recombination as a function of landscape
- Although it seems that “sex” is only favoured asymptotically in a small part of landscape space – “modular” landscapes – it is precisely those landscapes that dominate biology
- Modularity (quasi-additivity) and redundancy pervade biological fitness landscapes and their underlying structural hierarchies (nucleotides, exons/introns, genes, gene complexes, chromosomes,…)

- However, just as in physics, modularity at a higher structural level is based on a higher degree of epistasis at a lower structural level (intra- versus inter-genic epistasis)
  - e.g., exons within a gene are more epistatically linked than exons in distinct genes just as atoms in a molecule are more tightly bound than atoms in different molecules.
- How does “sex”/recombination respect such hierarchies?
- By being subject to evolution...



Modular landscape of four blocks of 8 genes with a NIAH landscape in each block

Two types of recombination
- Random
- Adaptive

Adaptive recombination allowed for a feedback mechanism from the offspring fitness. Fitter/less fit offspring → Increase/decrease recombination rate at that locus → Recombination hotspots

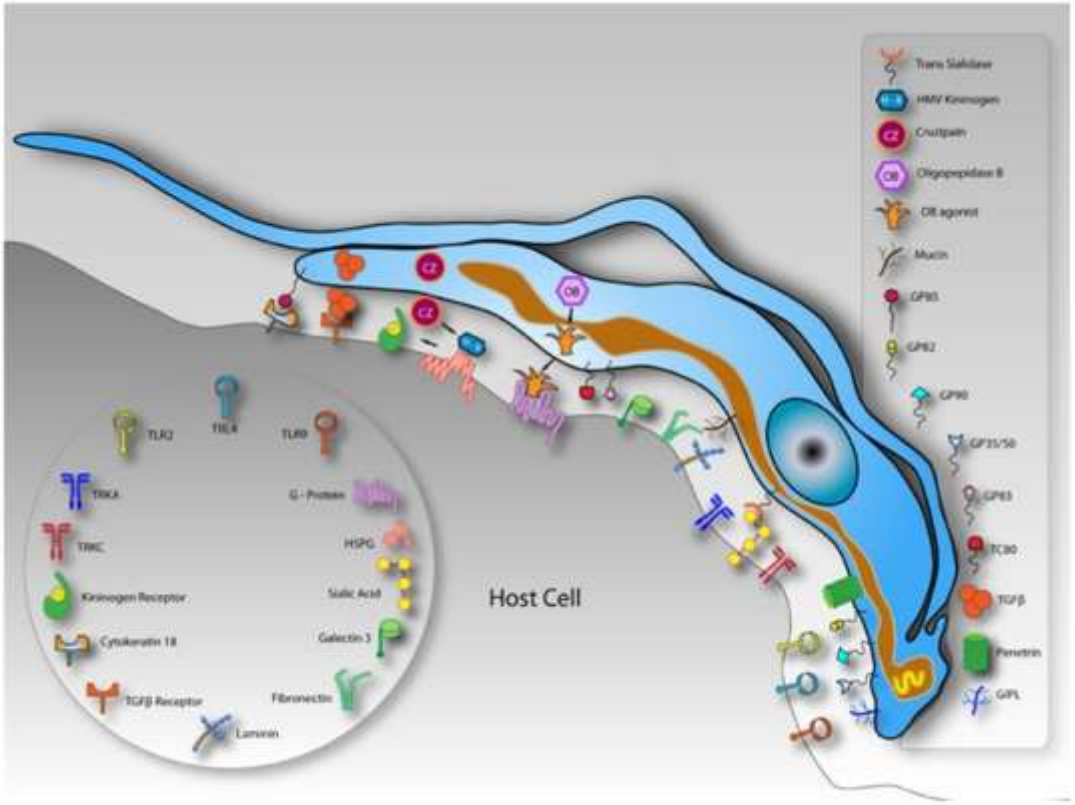
Figure 1: Results using block size 8 and a NIAH landscape. In (a) we can observe that the roulette wheel recombination always has better performance. In (b) we show the usage of crossover points (intra-block vs. inter-block) each generation.

Ortegon, Hartasanchez y Stephens
GECCO'10, July 7–11, 2010, Portland, Oregon, USA.
ACM 978-1-4503-0072-8/10/07.

So, we've considered genetic variation through "sex" meaning recombination of genetic material thought of as coming from more than one "type" leading to a different phenotype but in the context of a trivial genotype-phenotype map

"Sex with yourself" will consider genetic variation arising from recombination of genetic material coming from only one "type" leading to a different phenotype but in the context of a non-trivial genotype-phenotype map

T. Cruzi



Etiological agent of Chagas disease; a zoonosis endemic in Mexico affecting more than 8 million people in the Americas.

About 12,000 genes, thousands of which are members of particular gene families and about 25% of which are associated with surface proteins

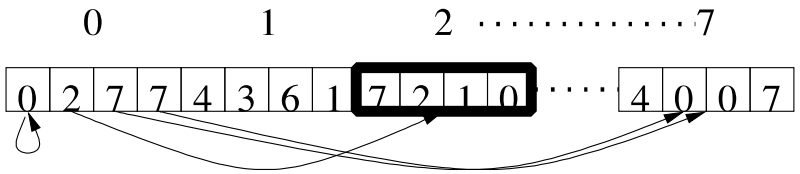
Infects:
Multiple vectors (Triatomines), multiple hosts (mammals), Multiple human cell types/tissues

Every one of these environments is a different challenge to the pathogen

It needs to generate immense phenotypic diversity and that requires enormous genetic plasticity. But its mainly clonal!

Sex with yourself, because finding the right partner can be difficult!

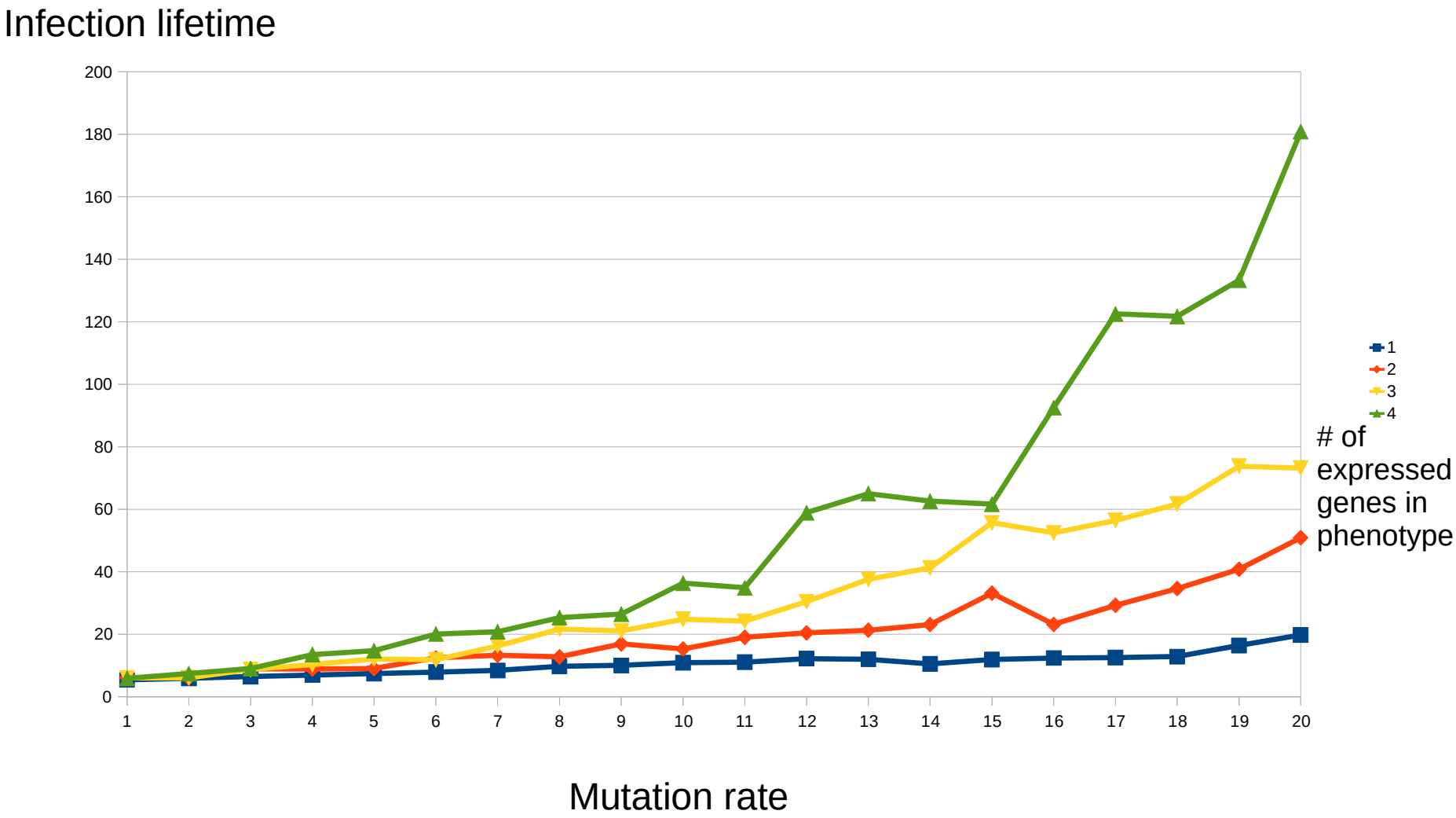
Indirect genotype-phenotype map and
gene expression



If of all the genes (multiple copy number) of a gene family that code for a given surface protein (mucin, transialidase,…) only a subset are expressed in the phenotype then expression of different subsets will lead to different phenotypes. There are then ${}^{N}C_{m}$ possible phenotypes, where N is the number of genes in the family and m is the number that are expressed.

Each subset expressed is generated by recombining already existing genetic material.

Elementary model of T. cruzi vs. the immune system



Conclusions

- Despite being one of the most important properties of biological systems there is still no generally agreed explanation of why “sex” exists
- There are different characterizations of “sex” – all are associated with some form or other of recombination of genetic material
- Considered recombination in the context of a two locus- two allele model in the space of “all” possible landscapes and “all” possible initial populations – the full parameter space
- There are two distinct regimes where sex/recombination provides an advantage in terms of our performance metrics
  - Search regime – quasi-independent of landscape
  - Modular regime – emergent and universal
- Recombination is asymptotically favoured only for a small subset of landscapes
  - Modular – quasi additive
  - Redundant – negative epistasis
- These landscapes types are the basis for all of biology

Conclusions

- A hierarchy of epistatic interactions link different levels of biological structure
- Recombination masks/rates have coevolved with biological fitness landscapes to reduce destruction of highly epistatically linked genetic combinations and increase creation of new, evolutionarily innovative combinations of modules (meta-evolution) leading to new phenotypes
- Normal “sex” requires at least two participants – types – so is “suppressed”
- Genetic variation can also be generated internally by horizontal transfer
- Phenotypic variation can also be potentially generated without direct genetic variation by variable expression using an indirect genotype-phenotype map
- T, cruzi potentially uses such an apparatus to survive in multiple environments in its lifecycle