High(er) order epistasis and strength of drift

Florian Labourel

Why on earth should we worry about effects of epistasis: an idea coming from cellular models of fitness?

Enzyme-substrate reactions (see top panel of the figure below) rely on the binding (at a rate k_f) between these two molecules, that can then either dissociate (at a rate k_r, these pointless meetings decrease the rate of the reaction) or undergo the process of catalysis (at a rate k_{cat}) that releases the enzyme and a newly formed product. Therefore, Evolution should in first approach maximise both k_f and k_{cat}. The process can be described through Michaelis-Menten equation, where the flux at steady-state is given by the following equation:

$$v = k_{cat}.[E_{\mathrm{tot}}].\frac{[S]}{K_{\mathrm{M}} + [S]}, \tag{1}$$

under the assumption that the concentration $[S]$ is approximately constant and that of $[ES]$ is at equilibrium $K_M = \dfrac{k_{\mathrm{r}} + k_{\mathrm{cat}}}{k_{\mathrm{f}}}$ represents an affinity term: the lower the better.

In this framework, it is possible to study the kinetic efficiency of enzymes. When we do so, the results is quite unexpected: enzymes seem to belong to a zoo where variation is the rule (see three panels below). One may then wonder why most enzymes are so far from their kinetic optimum (and so different one from another).

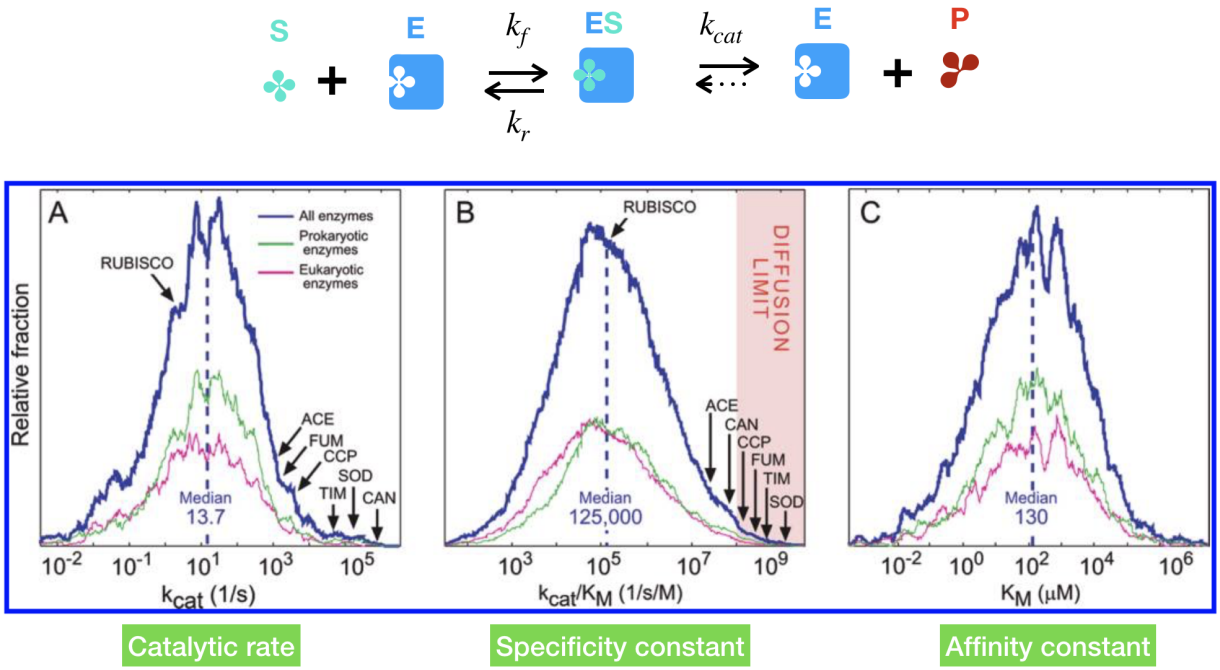

Figure 1: Enzyme zoo: widespread (and yet to be explained) variation in enzyme kinetics (source: Bar-Even et al., 2011)

When introducing more parameters, enzymes generally evolve on plateau-like fitness landscapes (results from many diferent theories: control of flux (Kacser and Burns, 73; Rapoport and Heinrich, 74; Dykhuizen, 88), protein stability, metabolix flux analysis; see right panel on the following figure); nonetheless, we have shown that several parameters are involved in the fitness landscapes on which enzymes evolve, one of which is the level of the flux an enzyme has to sustain:

But still, enzymes should be far more efficient and N_e should be a (if not the) major driver of their evolution as population genetics assuming realistic mutational landscapes shown:

One explanation of why Evolution underperforms (if it is the case - some biophysicists think that cells may be tuned to sustain specific flux levels rather than ; others think that some physics constraints may be impossible to escape such that they disable Evolution to act as it should be) is the existence of epistasis between enzymes.

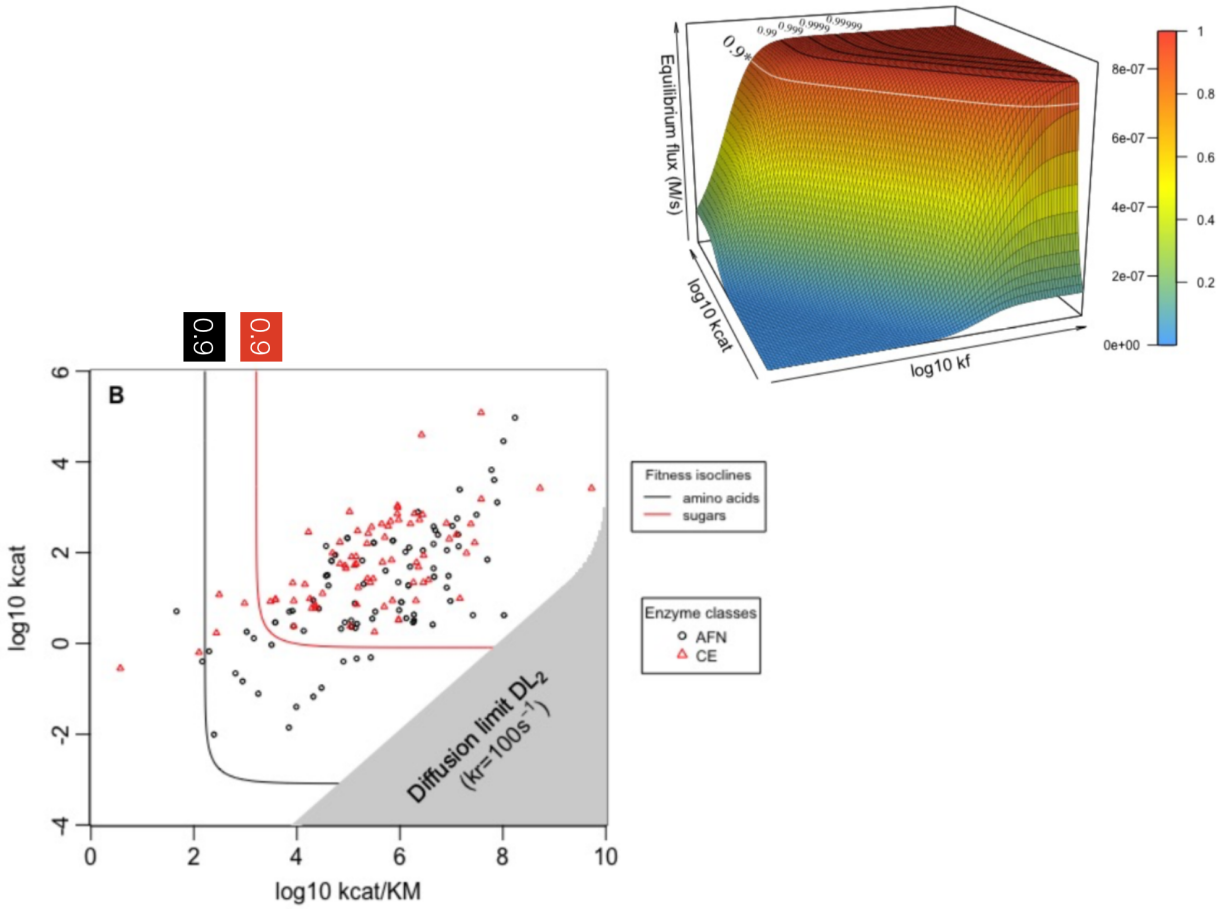

Figure 2: Enzyme zoo, some pieces of explanation: the levels of flux in a given pathway

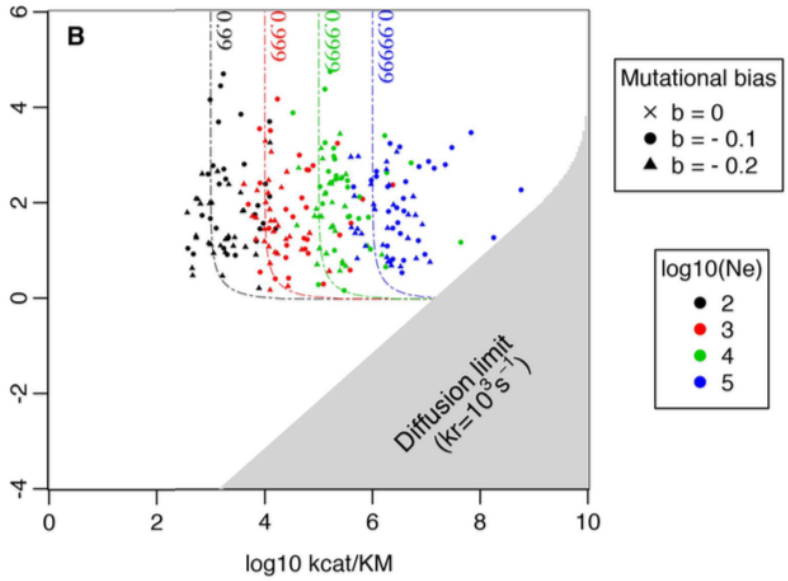

Figure 3: Enzyme zoo: the paradox of

Brief recall about epistasis

Epistasis: process by which the effect of the mutation is dependent on the genetic background in which it appears. In other words, it is the part of variance that cannot be explained by effects of additivity or dominance. When two mutations are present, there is a departure from the prediction of additivity (additivity: first case in the figure below; different kind of epistasis are shown next to it).

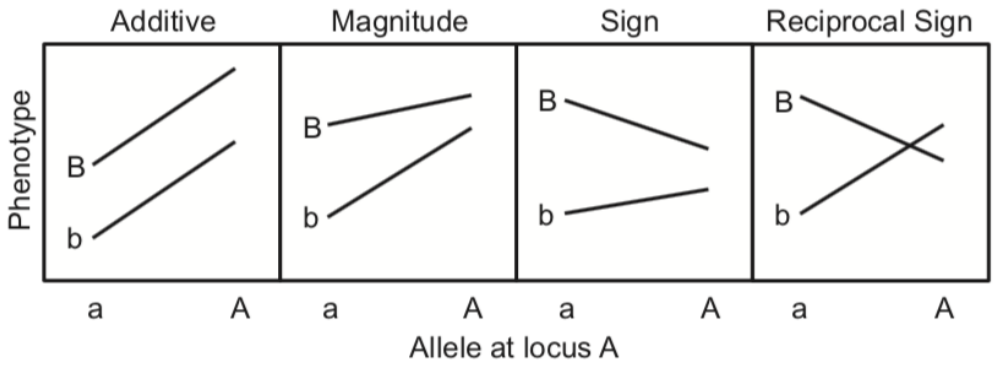

Figure 4: Weinreich definition: epistasis (three panels on the right) is a measure of our surprise

Epistasis has long been discussed in the case of the fitness landscape analogy that describes Evolution as a climbing process that may or may not be deterministic. More particularly, it has been demonstrated that in a constant environment, where organisms can be represented on one and only one fitness landscape, rugged landscape can arise if and only if there is some reciprocal sign epistasis, which produces valleys (and potentially ruggedness) separating different peaks that opens the door to evolutionary contingency.

The idea here is to study the case of complementary epistasis when it involves a large number of loci/genes/modules all needed in order to perform a given task at a given level (eg. enzymes involved in

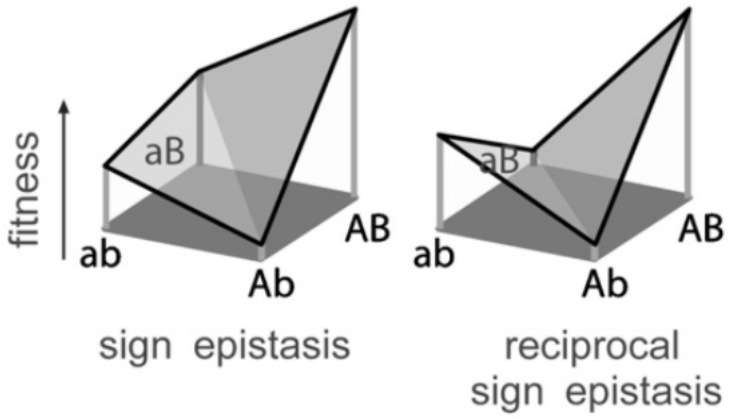

Figure 5: Epistasis and fitness

a metabolic pathway: if one is worse than the others, the flux is decreased; different cellular pathways: producing a lot of a given amino-acid may be useless if ribosomes are too slow to use them, or if nucleotids are produced at a slower pace (relatively to that of the aforementionned amino-acid; a very efficient stomach might be useless if the gut or the liver underperforms).

The particular case of complementary epistasis

Complementary epistasis : both genes are needed to give rise to a new phenotype as shown below (A and B can only produce their effect if they are expressed together; otherwise, none of them has any effect at all (to study it in a generic case) - neutral mutation). This is well-known for pigment coloration in corn, for instance, where changing the color relies on the presence of two mutations (in some cases).

Under such assumptions, increasing fitness relies on a two-step process: first, a dormant advantageous mutation has to evolve through drift as its effect on fitness is perfectly neutral. Only then fitness can be improved thanks to a second advantageous mutation occuring on the complementary gene.

As it is cumbersome to draw hypervolume with many dimensions, we got focused on the simplistic case of two complementary genes. One way then wonder what would happen if higher order epistasis comes into play.

High(er) order epistasis : epistatic interactions depend on a high number of interactions (>>2 genes)

First studies on the influence of the combination of drift and high(er) order epistasis

Some authors have recently raised the need to address this process (in the case of enzyme turn-over numbers k_{cat}s for instance). In their model, the fitness of cells result from the complex combination of thousands of enzymes whose k_{cat}s can undergo mutations. Mutation fixation of one variant is computed by a random draw from a binomial distribution with (Kimura:61)'s formula for fixation probability.

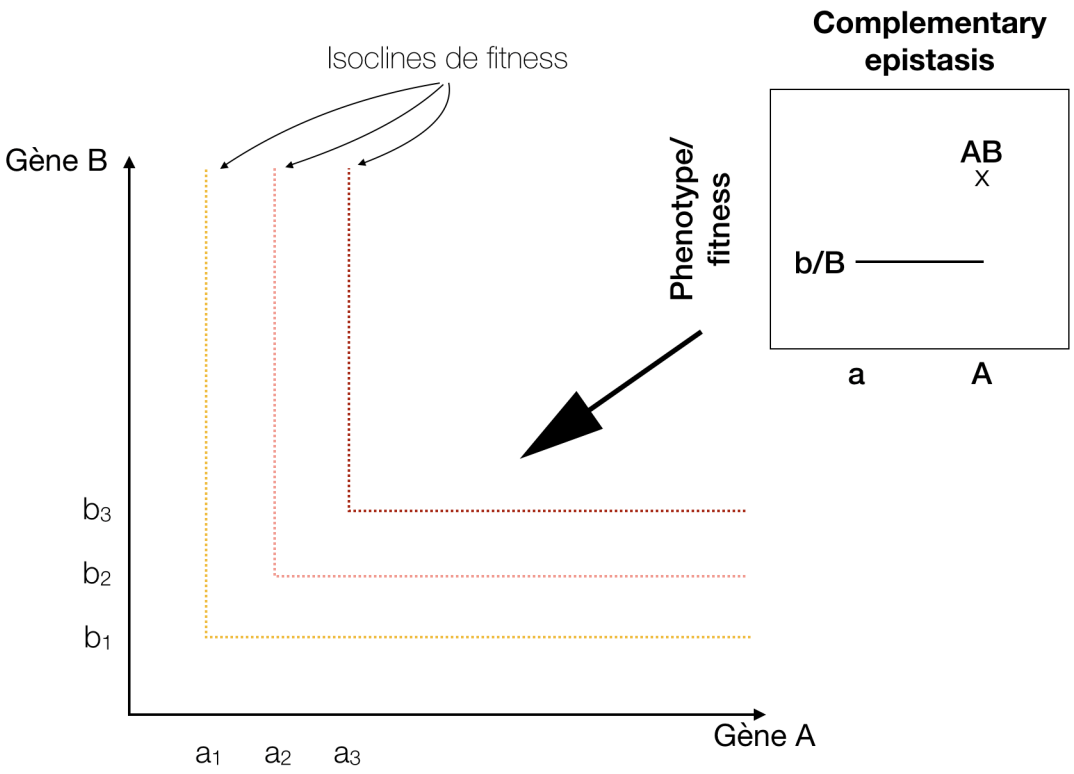

Figure 6: Complementary epistasis and fitness isoclines

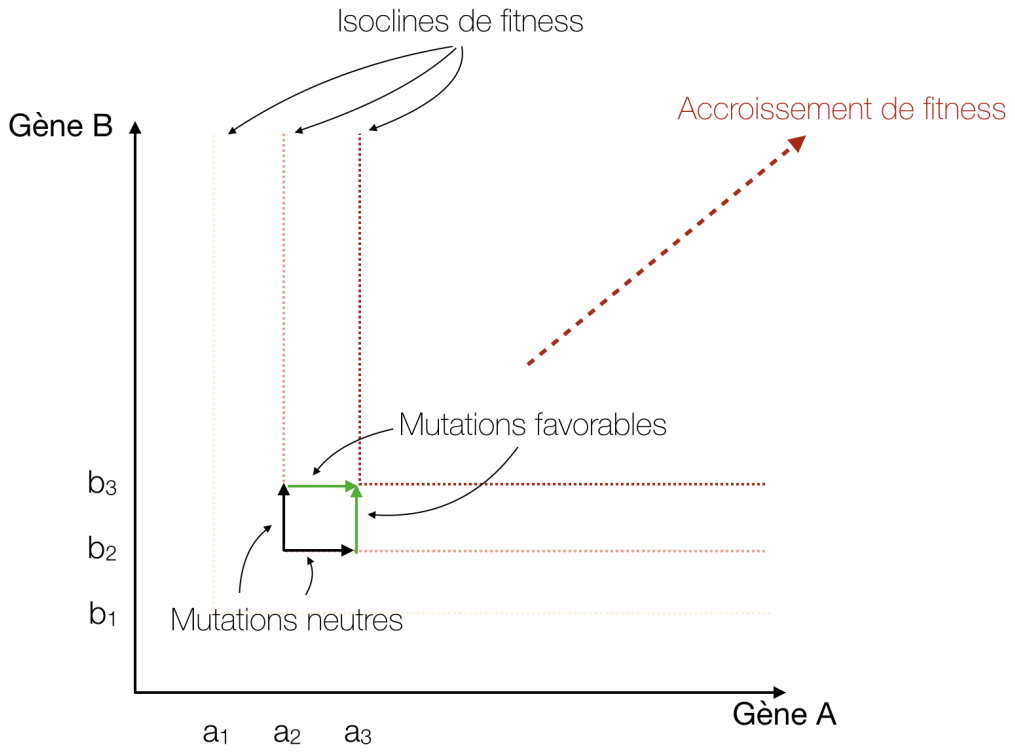

Figure 7: Complementary epistasis and evolutionary trajectories: only the last beneficial mutation gives an extra fitness to its carrier

ARTICLE

DOI: 10.1038/s41467-018-07649-1 OPEN

Modeling genome-wide enzyme evolution predicts strong epistasis underlying catalytic turnover rates

David Heckmann[1], Daniel C. Zielinski[1] & Bernhard O. Palsson[1,2]

And they have shown that it may account for the wide variability among enzymes efficiency within an organism, as their outcomes recnstruct most of this variability:

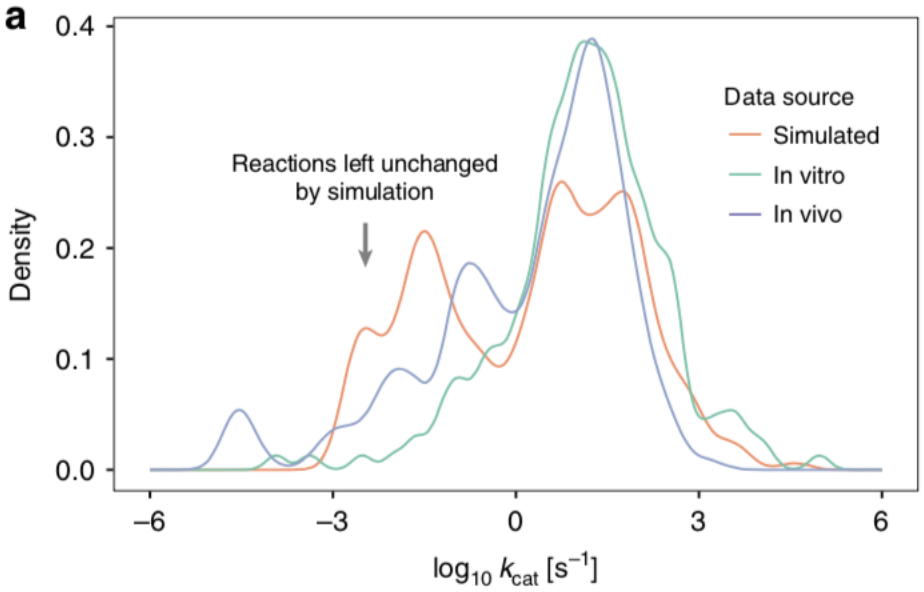

Figure 8: Turn-over numbers hit a ceiling in the different pathways

Here, the authors fixed some enzymes to moderate values for there seems to exist an upper-limit to some of them (dubious to me in first approach, see their explanations in the quotation reported below). This strong assumption limits the interest of their findings to me and might explain alone that enzymes are stuck to low values in their model.

The aim of the present document is to test what would happen when an enzyme's fitness depends on its genetic background in a complementary fashion (and in other cases for which this framework could make some sense) and, more specifically, how biased mutations - towards deleterious effects - would affect the evolutionary process.

A toy model to test the effect within a general framework

1) Fitness is determined by the selective value of the worst enzyme of an organism. Eg. set of 3 enzymes with selective value (0.9,0.95,1) yields a fitness of 0.9

As biological catalysts are limited to natural amino acids to stabilize transition states, it is expected that many reactions will have a distinct biophysical upper limit to the turnover rate that is lower than the theoretical limit resulting from diffusion rate of collisions. As certain reaction mechanisms were shown to consistently exhibit high k_{cat}s[16], we use the enzyme commission (EC) number to decide on a candidate set of 569 biophysically unconstrained reactions (see Methods). The remaining 1087 enzymes were considered biophysically constrained and were fixed to the median of in vitro k_{cat} measurements (13.7 s^{-1}). In the context of evolutionary predictions, the number of enzymes in the constrained and unconstrained set are more meaningfully compared in terms of reactions that are contributing to growth. Flux variability analysis[38] for aerobic growth on glucose reveals that 278 growth-relevant reactions (see Methods) are unconstrained, while 183 carry a biophysical constraint; the majority of in silico growth-relevant reactions is thus evolving without upper limit.

Figure 9: Explanations about constrained enzymes from Heckamnn et al.(2019)

2) A time-step corresponds to a mutational event conerning one gene/module of the set. Each mutation can then invade the population or be removed during the timestep (no polymorphism occurs here, to stay as simple as possible). Its probability of fixation depends on the gain of fitness (to the haploid organism) it provides, according to Kimura equation (Kimura,1961):

```
#Fixation probability
u_exact<-function(s,Ne,p){
  return((1-exp(-2*Ne*s*p))/(1-exp(-2*Ne*s)))
}
```

Note: If a mutation improves an enzyme that is not the worst one, its probability of fixation is that of a neutral mutation.

3) Mutations are drawn randomly from beta distributions with different properties (4 cases), as shown below:

Figure 10: Distribution of fitness effects of mutations in the different cases

Note: the distribution of fitness effects is dependent on the present value of the gene (latent) fitness. In other words, a better enzyme has a higher mutational average fitness (eg. an enzyme with fitness 1 (maximum) have more mutations that yield a fitness f=1 than en enzyme with fitness 0.5).

4) The initial fitness value of each enzyme/gene/locus is set to the rather high value of 0.9, sufficiently far away from high values to avoid boundary conditions problems (unclear why it is a problem at this point, may be creating some nearly absorbing states). #, the (Kimura) expected outcome of Evolution when there is a high mutational bias towards deleterious effects.

Results of the toy model

Simulations are ran over an average of $100N_e$ mutational events occuring on each module, with 30 replicates for each set of parameters. The first case considers here considers that $N_e = 10$, for which we report fitness outcomes when the evolutionary steady-state has been reached:

Figure 11: Very preliminary simulation outcomes ($N_e = 10$) : effect of complementary epistasis on fitness at evolutionary steady-state

The following reults display what happens for fitness when $N_e = 100$:

Figure 12: Very preliminary simulation outcomes ($N_e = 100$) : effect of complementary epistasis on fitness at evolutionary steady-state

The following reults display what happens for fitness when $N_e = 1000$:

Figure 13: Very preliminary simulation outcomes ($N_e = 1000$) : effect of complementary epistasis on fitness at evolutionary steady-state

One limit of these toy models lies in the choice of the distribution, which may trap the fitness in absorbing states (at very low values or very high values), especially when N_e increases, as very few mutations of the neutral area are sampled. Note that for the last results, the scale was adjusted in the bottom panels to facilitate their visualization.

What one can see is that outcomes are largely dependent on both the fitness effect distribution and the amount of complementary modules/genes (locus?) such that this process may help explain why most mutations are evolving under neutral evolution even if they might be necessary pieces to increase fitness. More importantly, it might help reconcile the paradox between the fact that many genes seem to evolve under neutrality while these genes still lie far from their theoretical optimum (eg. of enzymes above).