

# Note on the FGM, data and inferences

## 1 Introduction

This project is aiming at predicting fitness of AMR mutants in different environments and different genotypic backgrounds and at modelling the evolution of compensatory mutations. Under this main objective, three main goals have been depicted:

1. Can we use estimates of costs of resistance obtained in one environment to reliably predict costs in a different environment?
2. Are estimates of costs of resistance measured in one genetic background sufficient to predict costs on a different genetic background?
3. Is the route to compensatory evolution similar between environments and between genetic backgrounds?

Before any discussion on these three goals, we need to define some of these notions that may have multiple definitions in the literature. First, the “cost [of resistance] is defined as the selection coefficient of resistance mutations in absence of treatments (or similarly in absence of predator, parasite or pathogens when considering resistance in the context of biotic interactions)” (Lenormand et al. 2018). This cost, as discussed in Lenormand et al. (2018), may be a blurred notion due to the infinite numbers of costs which can be defined through the infinity of different “non-treated” environments. So here, we will focus on the broader idea of the change in fitness effects of a given mutation across environments, with a particular treated environment as a reference. Second, linked to the previous one, is the notion of environment, which we choose to define here by its measurements. Thus, two different environments can be distinguished by a measurable difference in at least one abiotic and/or biotic variable independent of the genotypes we are focusing on. This means that an environment cannot be defined by the traits (including fitness) of the genotypes in the population(s) that we are studying. Third, linked to the notions of costs and environments are the estimates of fitness. Multiple estimates (or proxies) of fitness exist, one of the most commonly used is the growth rate which can be used both at the level of a genotype or at the level of a population. These estimates, must be traits highly correlated with the true fitness itself and cannot be defined by a property of the environment. Indeed, the MIC (Minimum Inhibitory Concentration) is sometimes used as a proxy of fitness for the measurements of the capacity to resist to a certain antibiotic but this estimate corresponds to a variable defining some particular environments, which thus cannot be used to compare fitnesses across environments. Finally, the notion of compensatory mutations corresponds to a particular scenario. We consider a mutant with a fitness  $X$  in environment 1. This genotype faces a new stressing environment 2, adapts to it and has a fitness  $V$  in environment 2. Then this new genotype (adapted to environment 2) faces again environment 1 and has a fitness  $Y < X$  in this environment. In this case, mutations are qualified as compensatory if they increase  $Y$  in environment 1 while keeping the fitness in environment 2 equal or higher than  $V$ .

The method chosen to achieve these objectives can be separated in two main points:

- Obtain measurements of fitness of multiple genotypes (with a set of known AMR mutations) in multiple environments and obtain replicated time series data of evolution of compensatory mutations.
- Use the fitness landscape theory and more particularly Fisher’s Geometric Model (FGM) to predict fitness changes across environments, across genetic backgrounds and through compensatory mutations.

The two methods have to be developed conjointly to be able to use as many information as possible from the data into the model. In the following are detailed the main achievable results depending of the available data and the theoretical assumptions used in the model. The first section shows the results which can be obtained using the simplest form of the FGM (with numerous theoretical assumptions). The second part discusses how alternative models with less restrictive assumptions could be fitted and used for predictions with the appropriate dataset. (The third section outlines an extension of the project considering the frequency of mutants and compensatory mutations across environments using meta-genomic data. Not done yet)

## 2 The simplest FGM:

Predicting fitnesses across environment and genotypes using the FGM implies to parametrize the model. Here is defined a constrained version of the FGM which can be easily parametrized and allows simple analytical tractability. This simple version assumes:

- Isotropy: all traits are equivalent with respect to selection or mutation.
- Constant dimensionality  $n$  across environments.
- Constant strength of selection and mutational variance (summarized by  $\lambda$ ) across environments.
- No phenotypic plasticity.
- Universal pleiotropy (no modularity or partial pleiotropy).
- Constant height of the fitness peaks across environments.

### 2.1 Parametrization:

Under these assumptions, the model can be fully parametrized for a given environment  $E_1$ , by the three parameters :  $n, \lambda, r_{max}$ .

- $n$  the number of uncorrelated phenotypic dimensions often called “complexity”.
- $\lambda$  the mutational variance (standardized by the strength of selection).
- $r_{max}$  the growth rate of the optimal phenotype in the environment  $E_1$ .

$n$  and  $\lambda$  can be measured by fitting the moments of the full distribution of mutation effects on fitness, on selection coefficient measured on random single mutants from mutation accumulation experiment or artificially engineered random single mutants or screened mutants for a given antibiotic dose (Martin & Lenormand 2006, Martin & Lenormand 2015). Alternatively, by fixing the dimensionality (for example  $n = 2$ ) we can get the mutational variance with only the mean effect of mutation on fitness. In the following we will focus on the case of  $n = 2$ .  $r_{max}$  can be obtained directly by measuring the growth rate of an optimal genotype.

### 2.2 Mapping of the genotypes and the environments in the landscape:

#### 2.2.1 The points of reference:

In order to geometrically map genotypes in the landscape we need to have 3 points of reference ( $n + 1$  in a landscape with  $n$  dimensions). These points of reference will be used for trilateration of any new point added in the landscape. The best option is to consider the genotypes adapted to 3 different environments as the reference points. The relative coordinates in the phenotypic space of these points can be obtained from the measure of the fitness of each optimal phenotype in each of the 3 environments (the fitness measured need to be converted in phenotypic distance through the inverse of the fitness function). These fitness also gives the maximal growth rate ( $r_{max}$ ) for each environment. Thus, these 3 points define a plane (for  $n = 2$ ) in which we assume that every other environments or genotypes will lie.

### 2.2.2 Trilateration of the genotypes' positions:

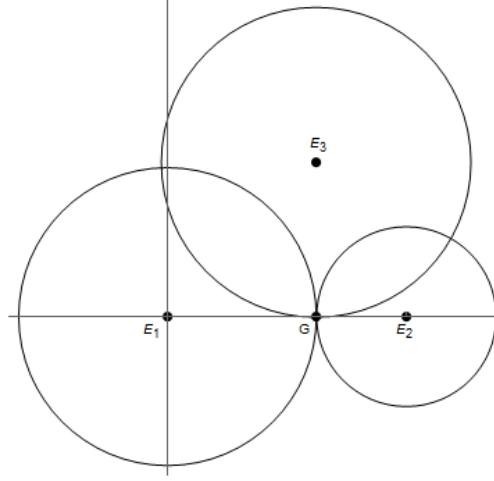


Figure 1: Example of trilateration

Using these three points of reference, the coordinates of any genotype can be obtained by measuring its fitness in the 3 environments, which gives the phenotypic distance to each optimum. These distances can then be used to derive the coordinates of the genotype using trilateration as shown in the right panel. With a collection of phenotype mapped in the landscape, it is then possible to predict the effect of mutations across environments and genotypes. To do so, mutants emerged from one given genotype among the previously mapped ones, have to be mapped in the same methods as previously. From these data, it is possible to assess the fitness of one of these mutations in any genotype and/or any environment (previously mapped). It is also possible to predict epistasis between them. Thus it is possible to predict the cost of mutations across environments and genotypes without directly measuring it.

### 2.2.3 Probability of evolving a compensatory mutation:

The FGM allows to predict the proportion of mutations which are beneficial mutations in multiple environment (see Martin & Lenormand 2015). However, the FGM does not include a genotype to phenotype map. Thus, it cannot predict the probability of emergence of a given mutation, but can predict the probability of emergence of mutations of a given effect  $s$ . Therefore the ability of the FGM to predict the emergence of compensatory mutations depends on the definition of compensation.

- If defined as “the adaptation to a second environment while keeping the same mutations of resistance adapted to the first environment”, it is not possible to give a prediction on the probability of such events.
- However, if defined as the adaptation to a second environment while keeping the same (or higher) fitness in the first environment, the FGM may give some predictions. Indeed in this definition the focus is not on a particular genotypes but on a fitnesses in both environment.

Using the geometry of the landscape, it is possible to give the probability that a mutant while have the  $n$ -uplet of growth rates  $(r_1, r_2, \dots, r_n)$  in the  $n$  environments  $E_1, E_2, \dots, E_n$  (has been done for two environments in Martin & Lenormand 2015 and can be approximated for  $K$  environments but only for  $n \gg 1$ ). Combined to an evolutionary model in the regime SSWM (or WSSM depending on the mutation rate and the mean of mutation effect on fitness), it is possible to predict the probability of fixation of such a mutation.

These results and especially those on compensation are highly dependent on the assumptions of the model. The next section is an attempt at relaxing some of these assumptions.

### 3 Alternative FGM

From the model described in the previous section, some assumptions are less “realistic” (i.e. validated by empirical results) than others and could greatly influence the “validity” of the results (i.e. ability of accurately predict the fitness of mutations across environments and genotypes). However, relaxing these assumptions imply to increase the number of parameters and thus the number/type of data needed.

#### 3.1 Constant strength of selection and mutational variance across environments:

Considering the variation of the strength of selection or the mutational variance across environments can be easily introduced in the model, as long as the model stay isotropic, but requires to fit DFEM in each environments. Moreover, using the DFEM fitting method, it is difficult to disentangle the strength of selection from the mutational variance. Note that variations in the first would impact the computation of phenotypic distances through the fitness function, whereas variation in the second would not. However, variations of the mutational variance across environments could reflect a form of plasticity, thus partially relaxing the assumptions.

#### 3.2 Constant peak height across environments:

Relaxing this hypothesis is straightforward as the fitness of the optimal genotypes in each environment should be measured in any case (however, this implies to measure absolute fitnesses, i.e. growth rates).

#### 3.3 Constant / fixed dimensionality across environments:

First, the change of dimensionality across environments could be implemented but seems a priori less pertinent biologically, as the complexity is more an attribute of the organism and not of the environment. The dimensionality measured by DFEM fitting may vary across environments, however this may reflect the variation of effective dimensions across environments rather than “true complexity” (as discussed in Lourenço et al. 2011). Thus, considering partial pleiotropy would capture these variation will being more intuitive biologically, as discussed further below. Second, considering higher dimensionality than 2 is possible and common in the FGM. However, this would increase linearly the number of reference points needed for positioning genotypes in the space and may be not be analytically tractable in high dimensions (numerical optimization of “the distance geometry problem”). One way to overcome this problem would be a dimensionality reduction by projecting the points in higher dimensions in the hyperspace of selected dimensionality. This could be achieved by multidimensional scaling methods (MDS).

#### 3.4 Isotropy and universal pleiotropy:

Considering anisotropy is the most difficult assumption to relax, as the geometry of the landscape could be greatly affected. Anisotropy, “can be introduced by mutational or selective correlations, and/or heterogeneity in mutational variance or strength of stabilizing selection across traits. These effects may in addition differ between environments, for example if different combinations of traits are favored in different environments (“syndromes”) or if mutational variances and covariances change with the environment (“phenotypic plasticity”).” (from Martin & Lenormand 2015).

First, anisotropy can reflect a certain directionality in the mutation process in the phenotypic space. This means that mutation or selection have a stronger effect in certain directions of the phenotypic space. Thus, the “orientation” and the “magnitude” of selection and mutation can change across environments which can be a way of modelling parallel evolution. This form of anisotropy can be introduced in the FGM with analytical tractability for one environment but has only been done by simulations for multiple environments. It seems that mild anisotropy could be absorbed by the isotropic model even in a case with multiple optima but for biological scenario different (genotypes maladapted to every optimum or well adapted to one) from the ones we are interested in.

Second, if we relax the assumption of universal pleiotropy, mutations may affect only a certain subset of the phenotypic traits. These subsets can be random (partial/restricted pleiotropy) or form independent clusters (modular pleiotropy). This form of anisotropy is not exclusive with the first one and can be a way to model the parallel evolution often observed in AMR evolution. Different models taking into account anisotropy and not-universal pleiotropy have derived but none for a case with multiple optima. In terms of parametrization, such models would require much more parameters, all the more so the number of dimensions is large. Time series data of replicated evolution toward the different optima would be necessary to assess the level of anisotropy or partial pleiotropy.