Epistasis and the structure of fitness landscapes: are experimental fitness landscapes compatible with Fisher's geometric model?

François Blanquart

Thomas Bataillon Guillaume Achaz Olivier Tenaillon

EvoEvo workshop, Conference on Complex Systems, Tuesday September 20th





Danish Research Council

Fitness landscape: definition

Adaptation by natural selection



• To predict the course of adaptation, we need to know the map between genotypes and fitness: the fitness landscape

Representation of a fitness landscape

• Wright 1932



FIGURE 1.-The combinations of from 2 to 5 paired allelomorphs.

Low-dimension representation of a fitness landscape

• Wright 1932



FIGURE 2.—Diagrammatic representation of the field of gene combinations in two dimensions instead of many thousands. Dotted lines represent contours with respect to adaptiveness.

Ruggedness requires epistasis

epistasis = interaction for fitness between two mutations
 e = log[W₁₁] + log[W₀₀] - log[W₀₁] - log[W₁₀]



Fitness landscape models with tuneable ruggedness

- Kauffman's NK model 1987
 - N loci
 - each locus acts in conjunction with K-1 loci
 - fitness is the sum of contributions from each locus
 - K = 1 is full additivity, K = N is a random landscape
- "Rough Mount Fuji" model Aita Husimi 1996
 - fitness is the sum of contributions from each locus plus random noise with variance v
 - v=0 is full additivity, large v is a random landscape

A phenotypic fitness landscape model

• Fisher's geometric model of adaptation (1930)



phenotype 1

Experimental fitness landscapes

- 5 mutations in β-lactamase gene conferring resistance to cefotaxime
- Minimum Inhibitory Concentration of $2^5 = 32$ genotypes is measured



Cefotaxime H1 – data

• Adaptation can proceed through 18 trajectories out of the 120 possible

Weinreich et al 2006

Experimental fitness landscapes

Name	Species	Environment	Mutation and genotype number	References
A1, A2	Aspergillus niger	Minimal medium	2 datasets of 5 mutations, 2 genotypes	de Visser et al. 1997 de Visser et al. 2009
B1-B10	Saccharomyces cerevisiae	Standard medium (on plates)	1711+3885 mutations, 5.4 million genotypes	Costanzo et al. 2010
C1, C2	Drosophila melanogaster	Lab environment	5 mutations, 2	Whitlock & Bourguet 2000
D	ssDNA bacteriophage ID11	E. coli (host)	9 mutations, 9 single mutants, 18 double mutants	Rokyta et al. 2011
E1, E2	Vesicular stomatitis virus	Baby hamster kidney (BHK21) cells (host)	6 mutations, 6 single mutants, 15 double mutants (E1) 28 mutations, 76 double mutants (E2)	Sanjuán et al. 2004
F	Escherichia coli	New, Low-glucose environment	5 mutations, 2	Khan et al. 2011
G	Methylobacterium extorquens	Methanol environment	4 mutations, 2	Chou et al. 2011
H1, H2	Escherichia coli	Cefotaxime (β lactam antibiotic)	5 mutations, 2	Weinreich et al. 2006 Tan et al. 2011
H3, H4	Escherichia coli	Cefotaxime (β lactam antibiotic)	4 mutations, 2	Schenk et al. 2013
11, 12, 13	Plasmodium falciparum (or vivax) DHFR gene transformed into E. coli and S. cerevisiae	Pyrimethamine (antimalarial drug)	4 mutations, 2	Lozovsky et al. 2009 Brown et al. 2010 Jiang et al. 2013

- Can we infer the underlying fitness landscapes?
- Can we compare the underlying fitness landscape across species and environments?
- Is Fisher's geometric model explaining well the structure of these genotypic landscapes?

Methods





Challenge 2: Diversity due to small number of mutations sampled



Resolution of the two challenges

- Solution to challenge 1: simulate the details of the protocol used to isolate each mutation
- Solution to challenge 2: simulate the fact that each genotypic landscapes represents a small sample of mutations among the myriad available mutations

> Approximate Bayesian Computation> Fisher's Geometric model of adaptation

Several arguments in favour of Fisher's geometric model

- Any fitness optimum can be approximated as a Gaussian so even if the landscape is multi-peaked, the population only "sees" a single peak
- Makes sense biologically to consider phenotypes instead of genotypes
- Emerges from mechanistic first principles (Martin 2014)
- Predicts a number of experimental quantities (Martin et al. 2007, Perfeito et al. 2014, Schoustra et al. 2016)
- Generates a diversity of landscapes with properties similar to experimental landscapes (Blanquart et al 2014)









Fisher's geometric model: details



- Mutation effects on phenotypes follows multivariate normal distribution with mean 0
- Additive effects of mutations in the phenotypic space
- Four parameters
 - W_{max}: maximum fitness
 - σ_{mut} : sd of mutational effect
 - n: dimensionality of phenotypic space
 - Q: shape of the peak



Overview of Approximate Bayesian Computation



Details of the evolutionary simulations



- Strong selection, weak mutation approximation
- Several types of landscapes: all combinations of 5 mutations, all pair of mutations among 8 mutations, etc.

Details of the distance measure

- Compute summary statistics for the focal dataset and for the 10⁶ simulated pseudo-datasets
- Distance between summary statistics of the simulated pseudo-data and the focal dataset, normalised

$$D(\widehat{D}, D) = \sqrt{\sum_{i=1}^{n_{\text{stat}}} \left[\frac{\widehat{S}_i - S_i}{\max(\widehat{S}_i)} \right]^2}$$

- We tested two sets of summary statistics:
 - (1) all fitness values
 - (2) mean and variance of selection coefficient, mean and variance of epistasis, maximal fitness achieved, and correlation between selection and epistasis

Cross-validation to check the accuracy of parameter inference

- We can cross-validate using simulation as focal datasets
- Prediction error

٠



- 1. "Neural network" algorithm and use of summary statistics is best
- 2. Small genotypic landscapes contain some information about $W_{_{max}},\,\sigma_{_{mut}}$ but less so about n and Q
- 3. Protocols involving selection are best, in particular the protocol with four large effect mutations

Results

"Replicate" genotypic landscapes from the same biological system have similar underlying landscapes



Costanzo et al 2010, de Visser et al 1997, Whitlock and Bourguet 2000

Two seemingly different genotypic landscapes from the same biological system have similar underlying landscapes





Sanjuan et al 2004

Similar biological systems have similar underlying landscapes



maximal fitness log(W_{max})

Chou et al 2011, Khan et al 2011, Sanjuan et al 2004, Rokyta et al 2004

Are experimental landscapes compatible with Fisher's geometric model?

- Null hypothesis: the data follows Fisher's model
- "Bayesian p-value"



median distance of pseudo-dataset to retained simulations

Statistical properties of experimental fitness landscapes are well explained by Fisher's model







Pyrimethamine 13 - dataumber of mutationshamine 13 - best simulated

Pyrimethami





p=0.83



р=0.







Conclusions

- Phenotypic fitness landscapes solve the problem of high dimensionality
 of genotypic landscapes
- Empirical landscapes contain some information about $W_{\text{max}},\,\sigma_{\text{mut}}$ but less so about Q and n
- Empirical landscapes obtained with selected mutations contain more information on the underlying landscape
- Fisher's geometric model predicts well the statistical properties of fitness landscapes
- In most systems some specific combinations of mutations have fitness not predicted by Fisher's model

Blanquart and Bataillon Genetics 2016

Thank you for your attention!