

An evolutionary framework for systems biology

Laurence Loewe¹



Center for Systems Biology Edinburgh, University of Edinburgh



Why combine the two?

Evolutionary genetics and molecular biology have both been very successful in furthering our understanding of the natural world.

The limits of simplistic approaches.

After decades of research some familiar assumptions are reaching their limits and evolutionary biologists get increasingly interested in the molecular details of their systems. At the same time molecular biologists increasingly realise that quantitative modelling is actually worth the effort.

The flood of *omics data.

Many researchers in both fields feel that in these exciting times their field is almost being redefined. While evolutionary biology has no agreed name for 'the new' (many have been suggested), molecular and cell biology have decided to use 'systems biology' to promote the new direction to their discipline.

Evolutionary Frameworks

Evolutionary population genetics has a long tradition of mathematical modelling in biology that frequently abstracts details. The rigorous nature of the corresponding models and their extensive analysis has lead to key insights with applications well beyond evolutionary questions. The current hunt for functional sequences by scanning genomes for signatures of selection is one example. Increasingly realistic models are needed for further progress.

What is Systems Biology?

Molecular biology has a strong tradition of inferring molecular interactions from clear-cut clever experiments that frequently do not require quantitative analyses. The spectacular success of molecular and cell biology has accumulated such a wealth of knowledge that further progress in many areas increasingly depends on quantitative models. Recognising this, some molecular and cell biologists have started to collaborate with theoreticians to develop quantitative models. Assuming that the whole is more than the sum of its parts, these models aim to capture the essence of all important **intracellular interactions** of the system under investigation.

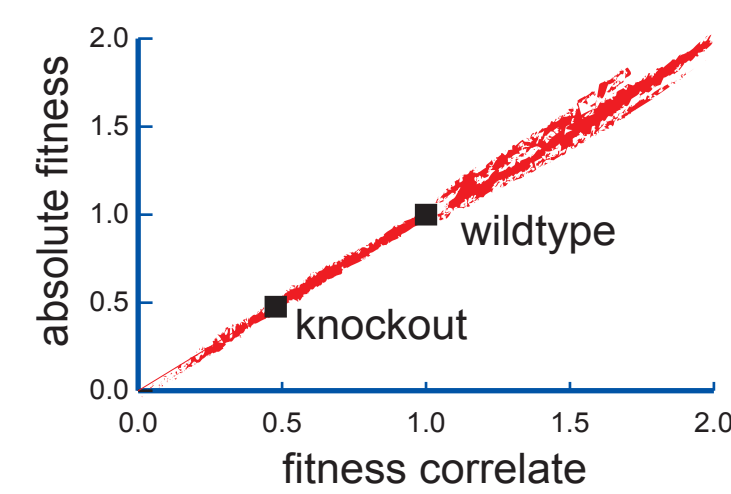
The famous systems biology cycle refines these models:

- 1 observe
- 2 improve model
- 3 predict,
- 4 to test the model goto 1

The use of fitness correlates

Since absolute total fitness is too complex to compute *ab initio*, one may use fitness correlates:

1. Pick a subsystem that *critically limits* fitness.
2. Predict the fitness limiting property from a systems biology model.
3. Calibrate the maximal deleterious effect by equating the dramatic selection coefficient measured in a knockout mutant with the fitness correlate prediction from an *in silico* knockout.
4. If scaling is linear, wildtype and knockout allow scaling of other effects.
5. If scaling is not linear, find other mutants with different effects and use
 - (a) observed selection coefficients and
 - (b) observed and/or predicted fitness correlatesto calibrate predictions of fitness from computable fitness correlates.



A powerful combination.

I propose that the enthusiasm for quantitative descriptions of intracellular processes could benefit from and contribute to the evolutionary biology objective to understand the forces that shape the existing diversity of life:

(a) Systems biology can provide much better models of how genotypes are mapped to phenotypes in nature.

(b) Evolutionary genetics can help analyse effects that are important on the long term but too small for observation in any laboratory.

Analysing Distributions of Mutational Effects (DMEs)

- 1 Use the current wildtype as neutral reference.
- 2 Link DNA changes to protein function by assuming a realistic distribution of mutational effects within proteins and within regulatory sequences.
- 3 Model natural mutagenesis by scaling frequencies of mutations with the size and composition of mutational targets.
- 4 Do many 1-step random perturbations of the wildtype and compute the fitness for each of them to >6 or >9 decimal digits.
- 5 Summarise the differences to the neutral reference as a deleterious DME (a logscale is best to visualise a majority of small effects).
- 6 The scaling of this DME depends on the quality of the fitness correlate.
- 7 Assume different intramolecular DMEs to test their influence.
- 8 Predict the **fraction of effectively neutral mutations** for the given effective population size of this species. Compare to observed neutral divergence as quality control or additional scaling mechanism.

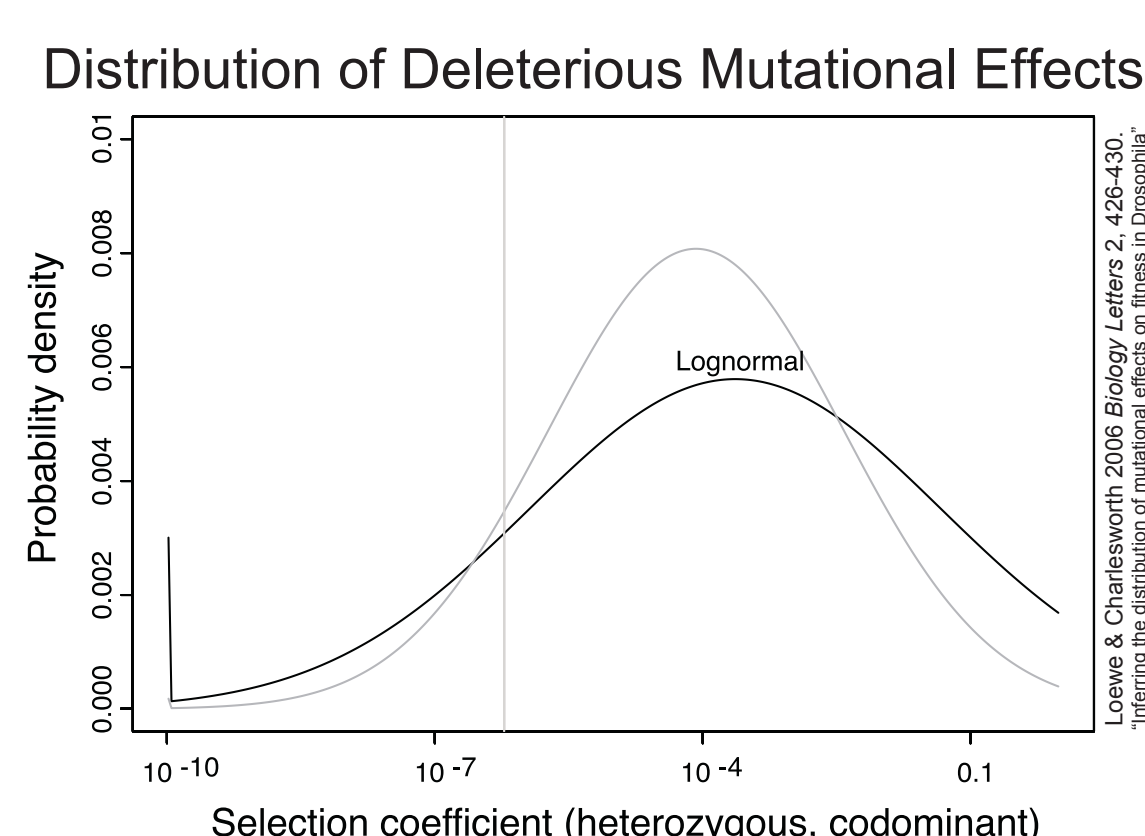
Analysing Beneficial Mutations

The methods described here for determining the DME and epistasis easily allow the investigation of two other hard problems in evolution:

A Infer the fraction of **advantageous mutations** and their distribution of effects using the same approach as analysing the deleterious DME (the quality of the scaling of the fitness correlate determines the reliability of estimates of bigger effects)

B Infer the frequency of **compensatory mutations** at the molecular level from analyses of *n*-way epistasis for large *n*.

Additional mechanistic insights from the molecular model can further help to refine these analyses.



Conclusions

An inability to meet the assumptions above severely limits the precision of results gathered by the framework above. In that case answers will only be rough and qualitative. Given the crudeness of many current models of fitness in evolutionary biology, this will nevertheless be a significant step forward, especially if many such rough models are built and common features start to emerge.

Critical Questions

- 1 Will it be possible to define good systems biology models that contain meaningful fitness correlates and that are still computationally tractable?
- 2 Will systems biologists of today consider using an evolutionary framework?
- 3 Is it necessary to predict protein functions from specific DNA sequences with high accuracy?
- 4 Do mutagenesis or evolution experiments have enough power to test the predictions from analyses of epistasis in these models?

Analysing Epistasis

Epistasis is defined as any deviation from an independent combination of mutational effects. Evolutionary genetics has explored much of the enormous consequences of particular types of epistasis, but so far models of epistasis and evolution are rather simplistic. Since all epistasis comes from the underlying networks of molecular interactions, systems biology models as defined above could help. The simplest case is **2-step epistasis**:

- 1 Start with the current wildtype as neutral reference (WT).
- 2 Like in analyses of the DME add one random mutation (A) and analyse its fitness consequences.
- 3 Then use genotype WT-A as a neutral reference for adding another random mutation (B) and analyse it.
- 4 Compare the fitness of WT with WT-AB to measure the combined effect of A and B.
- 5 Comparing the combined effect to the sum of all independent effects gives a measure of epistasis.

Such measurements of epistasis will strongly depend on the number of mutational steps analysed, so multiple numbers of steps need to be analysed to understand **n-step epistasis**. To get a thorough understanding one needs to analyse the distribution of epistatic effects in a similar way for 2, 5, 10, 100, 10³, 10⁴, 10⁵, 10⁶-step epistasis.