­Predicting competition results from growth curves

Yoav Ram1, Eynat Deluss-Gur1, Uri Obolski1,

Maayan Bibi2, Judith Berman2, and Lilach Hadany1\*

July 21, 2015

1 Dept. Molecular Biology and Ecology of Plants, Tel Aviv University, Tel Aviv 69978, Israel

2 Dept. of Molecular Microbiology and Biotechnology,Tel Aviv University, Tel Aviv 69978, Israel

\*Corresponding author: lilach.hadany@gmail.com

**Keywords:** mathematical model, microbiology, evolution, ecology, fitness, selection, experimental evolution

# Abstract

Measuring relative fitness by pairwise competition experiments is laborious and expensive. Accordingly, many investigators estimate fitness from the maximum growth rate during exponential growth. However, maximum growth rates have been shown to be an unreliable measure of fitness as indicated by discrepancies between these parameters and the outcomes of pairwise competition assays. Here we propose a new method that predicts the results of competition assays using single strain growth curves and estimates of competitive fitness. This method can also be used to infer interactions between competing strains and to provide biological interpretation to selective differences between different strains.

# Introduction

## Growth curves

Growth curves are commonly used to estimate fitness in microbiology, genetics, and evolutionary biology. Growth curves are acquired by measuring the optical density (OD) of one or more populations of cells over a range of time periods. The simplest way to infer fitness from growth curves is to estimate the growth rate during the exponential growth phase. This is done by taking the log of the mean of the growth curves during the exponential growth phase and using linear regression to estimate the slope of the curves as a measure of the growth rate (Hall et al. 2014). Indeed, growth rates can be proxies of the *selection coefficient s* (Chevin 2011), which is a standard way of measuring relative fitness in population genetics (Crow and Kimura 1970). However, the selection coefficient can be affected by other phases of a growth curve: the length of lag phase, the degree to which the ‘diauxic shift’ phase deaccelerates division cycles and stationary phase. Thus, it is not surprising that growth rates can be poor estimates of relative fitness (Concepción-Acevedo et al. 2015).

## Pairwise competition experiments

Fitness competition experiments infer relative fitness in a manner that accounts for all growth phases. In competition experiments, two or more strains are grown together in the same vessel – a reference strain and one or more strains of interest (for example, a wild-type reference strain and a mutant strain of interest). The frequency of each strain in the population is measured during the course of the experiment. This is done classically by plating assays that distinguish the strains using genetic markers. More recently, flow cytometry (Gallet et al. 2012) has been used with fluorescently marked cells and deep sequencing read counts have been used to determine the frequencies of different alleles in the population (Bank et al. 2014)]. The selection coefficient of the strains of interest can rthen be estimated from changes in the frequencies of the different strains during competition experiments, This is a good method to estimate relative fitness, as it directly estimates fitness from change in frequencies over time. However, competition experiments are more laborious than growth curves experiments and are typically more expensive, requiring the construction and assaying of genetic or phenotypic markers (Concepción-Acevedo et al. 2015, and references therein). Therefore, many investigators prefer to use proxies of fitness such as growth rates.

## Predicting competition results from growth curves

Here we propose a new framework for fitness inference. We fit growth models to growth curves data and use the fitted growth models to predict the results of competition experiments. The predicted competitions can then be used instead of empirical ones to estimate selection coefficients.

We implemented our method using *Curveball*, an open source Python package that can be freely used and extended; in the future we intend to develop a user friendly web site that will allow other investigators to analyze growth curves (http://curveball.yoavram.com).

# Methods

Our method includes three stages: (i) fitting growth models to the growth curves data, (ii) using the fitted models to predict the results of competition experiments, and (iii) estimating selection from the predicted competition results.

## Growth model

Because we are interested in several growth phases – the lag phase, the exponential phase, the deceleration phase, and the stationary phase – we use an extension of the classic logistic model, the Baranyi-Roberts model (Baranyi and Roberts 1994; Baranyi 1997).

The Baranyi-Roberts model is defined by the following one-species ordinary differential equation [see eqs. 1c, 3a, and 5a in (Baranyi and Roberts 1994)]:

(1a)

(1b)

(1c)

where is the population density, is the per capita growth rate, is time, is the adjustment function (see below), is the maximum density, and is a deceleration parameter.

The term is used to describe the deceleration in the growth of the population as it nears the maximum density. When the deceleration parameter is one (), the deceleration is the same as in the classic logistic model and the density at the time of the maximum growth rate is half the maximum density, . When or the density at the time of the maximum growth rate is higher or lower than , respectively.

The adjustment function is used to describe the adjustment of the population to a new environment. Typically, microorganisms are grown in overnight culture and diluted into fresh media for growth curve experiments. Therefore, populations that are adjusted to stationary phase must now adjust to growth conditions, and this takes some time. This adjustment phase is usually called the *lag phase*. The specific adjustment function we use here was suggested by Baranyi and Roberts (1994) due to being both computationally convenient and having a biological meaning: is the initial amount of some molecule (nutrient, enzyme, etc.) that is required for growth; is the rate in which this molecule is accumulated in the cell.

The Baranyi-Roberts differential equation has a closed form solution:

(2a)

. (2b)

We use four forms of the Baranyi-Roberts model. The full model is described by eq. 2 and has six parameters. A five parameter form of the model has the deceleration parameter set to one, as in the classic logistic model. This form is useful because the reduced growth during the lag phase might sometimes be inferred as . A four parameter form of the model has no lag phase, with . This is also known as the Richards model or the generalized logistic model. This form of the model is useful in cases where there is no observed lag phase: either because the population adjusts very rapidly or because it is already adjusted prior to the growth experiment, usually by priming it in fresh media before the experiment. The fourth form is the classic logistic model, in which and .

## D:\workspace\curveball_project\ms\combined.png

Figure 1. Example of the method applied on growth curves of two *Escherichia coli* strains (A) Growth curves of MG1655 in orange (top lines) and DH12S in purple (bottom lines). Each line represents a series of OD595 measurements from a single well in a 96-well microplate (Costar), taken every 10 minutes. Cells with Kan+Cap+ plasmids were diluted 1:20 from overnight culture and grown in 100µl LB (with Kanamycin and Cholarmphenicol) at 30°C in an automatic plate reader (Tecan Infinite 200Pro). The OD of cell-free wells was ~0.13. (B) Model fit (solid line) and OD595 data (markers: mean, errorbars: standard deviation) of the two strains. Fitted parameters: MG1655, *N0*=0.134, *r*=0.416, *ν*=2.73, *K*=0.588, *q0*=0.053, *m*=2.37, lag duration=1.714, maximum growth rate=0.357; DH12S, *N0*=0.13, *r*=0.876, *ν*=1, *K*=0.505, *q0*=0.15, *m*=0.772, lag duration=1.691, maximum growth rate=0.279. Note that the maximum growth rate is a function of *r, ν*, and *K*. (C) Predicted OD in competitions between the two strains. Initial OD of both strains was set to 0.067. (D) The frequency of MG1655 during the predicted competitions (dashed line). The estimated selection coefficient is *s*=0.2, calculated with Eq. 4 and *t*=12. Note that initially the frequency of MG1655 declines due to a longer lag phase, but then increases due to faster growth and a higher maximum density. Calculating the selection coefficient from the maximum growth rates would have yielded *s*=0.192 (Chevin 2011 eq. 2.3).

## Model fitting and selection

We fit all four model forms to the mean growth curve of each strain using a least-squares procedure on data from ## of parallel growth curves for two *E. coli* strains with different growth curve parameters (Newville 2014). The standard deviation at each time point is used to weight the least-squares procedure so that time points with lower variance are more heavily weighted and therefore better fitted.

We then calculate the Bayesian Information Criteria (BIC) of each model fit,

,

where is the number of parameters of the model, is the number of time points , is the average density at time point , and is the expected density at time point according to the model. We select the model form with the lowest BIC.

As a sanity check, we also fit the data using a linear model () and checked that the BIC of our selected model form was larger than the BIC of the linear model by at least 6 [See (Kass and Raftery 1995) for significance of BIC differences].

We repeated the model fitting procedure for the growth curves data of each strain to produce estimates for all six parameters as well as confidence intervals on these estimates.

## Competition prediction

We introduce the two-strain Baranyi-Roberts model, which, to the best of our knowledge, has not been used before:

(3a)

(3b)

(3c)

, (3d)

where is the value of parameter of strain which we get from the model fitting procedure. This equation system is then solved by numerical integration, resulting in a prediction of the competition dynamics.

This two-strain competition model explicitly assumes that all the interactions between the two strains can be attributed to *resource competition*. Therefore, all interactions are described by the deceleration of the growth rate of each strain in response to growth of the other strain. We do not however assume the same limiting resource or resource efficiency for both strains, as we use different maximum densities for each strain.

## Selection coefficient inference

Perhaps the most common method for estimating relative fitness or selection coefficients from pair-wise competition results is (Wiser and Lenski 2015):

(4)

where and are the densities of the strains and is time, usually chosen to be 24 hours.

# Discussion

We present a new computational method to predict the results of competitions between two strains from the separate growth curves of each strain.

This method should be useful, because growth curve experiments require much less effort and resources than pairwise competition experiments (Concepción-Acevedo et al. 2015; Wiser and Lenski 2015; Hegreness et al. 2006; Gallet et al. 2012). As automatic 96-well microplate readers become more and more common in microbiology labs, growth curve experiments can be set up in less than 30 minutes, after which the measurements are automatically collected by the plate reader (Hall et al. 2014; Concepción-Acevedo et al. 2015).

Current methods for estimation of fitness from growth curves use the growth rate as a proxy of fitness. The growth rate and other proxies of fitness have several disadvantages: (i) they can't capture the full scope of effects contributing to differences in fitness; (ii) they are difficult to compare between different studies and organisms; and (iii) they can't be used as parameters in standard population genetics models that test hypotheses and predict evolutionary dynamics. In contrast, our method integrates several growth phases into the fitness estimation, and our growth model can be extended to include other phases and factors of growth, such as biphasic growth and cell death.

The growth model that we use - the Baranyi-Roberts model - has a differential equation form and a closed form solution. Hence, it is very useful for our method: the closed form is used to fit to growth curve data while the differential equation is used to predict the competition dynamics.

Our method assumes that the two strains interact solely via resource competition; that is, only through the factor . If the investigators know or suspect that additional interactions exist (*i.e.*, density-dependent interactions such as social or sexual selection, mutualism, and interference), our model can serve as a null hypothesis: the results of competition experiments can be compared to model predictions and a goodness of fit test can be used to decide if additional interactions are significant. Moreover, these additional interactions can be measured, either in terms of the difference in selection coefficients (between the coefficient calculated from the empirical results and coefficient calculated from the model prediction) or by fitting the empirical results to an extended model that includes density-dependent interactions (Masel 2014).

## Conclusions

We propose a new method to analyze growth curves, predict competition results, and estimate fitness. Our method is easy to use, has a clear biological interpretation, and can also be used as a null model for the interpretation of competition experiments.

# Acknowledgments

We thank E. Kroll, Y. Pilpel, D. Hizi, I. Frumkin, O. Dahan, A. Yona, I. Ben-Zion, J. Barrick, H. Acar, and J. Masel for helpful discussions. This research has been supported in part by the Israel Science Foundation 1568/13 (LH) and 340/13 (JB), the Minerva Center for Lab Evolution (LH), Manna Center Program for Food Safety & Security (YR), the Israeli Ministry of Science & Technology (YR), the Anat Krauskopf Foundation (YR), TAU Global Research and Training Fellowship in Medical and Life Science and The Naomi Foundation (MB), the European Research Council (FP7/2007-2013)/ERC grant 340087 (JB).

# References

Bank, C., R. T. Hietpas, A. Wong, Daniel N. A. Bolon, and J. D. Jensen. 2014. “A Bayesian MCMC Approach to Assess the Complete Distribution of Fitness Effects of New Mutations: Uncovering the Potential for Adaptive Walks in Challenging Environments.” *Genetics* 196 (3) (January 7): 841–852. doi:10.1534/genetics.113.156190. http://www.genetics.org/cgi/doi/10.1534/genetics.113.156190.

Baranyi, József. 1997. “Simple Is Good as Long as It Is Enough.” *Commentary* (1996): 391–394. doi:10.1006/fmic.1996.0080.

Baranyi, József, and Terry a. Roberts. 1994. “A Dynamic Approach to Predicting Bacterial Growth in Food.” *International Journal of Food Microbiology* 23: 277–294. doi:10.1016/0168-1605(94)90157-0.

Chevin, Luis-Miguel. 2011. “On Measuring Selection in Experimental Evolution.” *Biology Letters* 7 (2) (April 23): 210–3. doi:10.1098/rsbl.2010.0580. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3061150&tool=pmcentrez&rendertype=abstract.

Concepción-Acevedo, Jeniffer, Howard N. Weiss, Waqas Nasir Chaudhry, and Bruce R. Levin. 2015. “Malthusian Parameters as Estimators of the Fitness of Microbes: A Cautionary Tale about the Low Side of High Throughput.” *Plos One* 10 (6): e0126915. doi:10.1371/journal.pone.0126915. http://dx.plos.org/10.1371/journal.pone.0126915.

Crow, James F., and Motoo Kimura. 1970. *An Introduction to Population Genetics Theory*. Minneapolis: Burgess Pub. Co. https://books.google.co.il/books?id=MLETAQAAIAAJ.

Gallet, Romain, Tim F. Cooper, Santiago F. Elena, and Thomas Lenormand. 2012. “Measuring Selection Coefficients below 10-3: Method, Questions, and Prospects.” *Genetics* 190 (1) (January): 175–86. doi:10.1534/genetics.111.133454. http://www.genetics.org/cgi/content/abstract/190/1/175.

Hall, Barry G., Hande Acar, Anna Nandipati, and Miriam Barlow. 2014. “Growth Rates Made Easy.” *Molecular Biology and Evolution* 31 (1): 232–238. doi:10.1093/molbev/mst187.

Hegreness, Matthew, Noam Shoresh, Daniel L. Hartl, and Roy Kishony. 2006. “An Equivalence Principle for the Incorporation of Favorable Mutations in Asexual Populations.” *Science* 311 (5767) (March): 1615–7. doi:10.1126/science.1122469. http://www.ncbi.nlm.nih.gov/pubmed/16543462.

Kass, Robert, and Adrian Raftery. 1995. “Bayes Factors.” *Journal of the American Statistical Association*: 773–795. doi:doi: 10.2307/2291091. http://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476572.

Masel, Joanna. 2014. “Eco-Evolutionary ‘Fitness’ in 3 Dimensions: Absolute Growth, Absolute Efficiency, and Relative Competitiveness.” Populations and Evolution (July): 1–44. http://arxiv.org/abs/1407.1024.

Newville, Matthew. 2014. “LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python.” http://lmfit.github.io/lmfit-py/.

Wiser, Michael J, and Richard E. Lenski. 2015. “A Comparison of Methods to Measure Fitness in *Escherichia Coli*.” *PLOS ONE* 10 (5): e0126210. doi:10.1371/journal.pone.0126210. http://biorxiv.org/content/early/2015/03/06/016121.