­Predicting competitions from growth curves

Yoav Ram1\*, Eynat Deluss-Gur1, Uri Obolski1,

Maayan Bibi2, Judith Berman2, and Lilach Hadany1

July 15, 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: yoavram@post.tau.ac.il

**Keywords:** mathematical model, microbiology, evolution, ecology

# Introduction

Many experimental investigators in microbiology, genetics, and evolutionary biology use growth curves to estimate fitness. They measure the Optical Density (OD) of one or more populations of cells over several hours or even days to acquire the growth curves. The simplest way to estimate fitness from these curves is to infer the growth rate: taking the log of the curves during the exponential growth phase, using linear regression to fit a linear line to the data, and taking the slope of the line as a measure of the growth rate1. Growth rates can indeed be proxies of the *selection coefficient s*2, which is the standard way of measuring relative fitness in population genetics3. But in many cases growth curves include other growth phases in addition to the exponential growth phase: a lag phase, a deceleration phase, and a stationary phase.

Competition assays are a common fitness inference method that takes these additional growth phases into account. Competition assays include the growth of two strains in the same container – the strain of interest and a reference strain (for example, a mutant strain and a wildtype strain). From the change in frequency over the competition duration, investigators can estimate the *selection coefficient s* of the strain of interest4. Theoretically, this is a much better method to infer fitness, as it directly estimates the relative fitness rather than indirectly estimating it from proxy measures such as growth rates. However, competition assays are more laborious than growth curve assays and are typically more expansive, requiring the construction and assaying of genetic or phenotypic markers5–7.

Because competition assays require more work and/or incur high costs, many investigators do without them and use proxies of fitness such as growth rates. However, these proxies of fitness suffer from several disadvantages: (i) they fail to capture the full scope of effects contributing to differences in fitness; (ii) they are hard to compare between different studies and organisms; and (iii) they are hard to use as parameters for population genetics models that can be used to test hypotheses and predict evolutionary dynamics.

Here we propose a new method of fitness inference. Our method fits growth models to growth curves data and uses the fitted growth models to predict the results of competitions assays. The predicted competitions can then be used instead of empirical ones to infer selection coefficients.

We implemented our method using an open source Python package that can be freely used and extended (http://curveball.yoavram.com). We will also provide a friendly user interface to our method to allow other investigators easy access to analysis of their growth curves.

# 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 assays, and (iii) inferring selection from the predicted competition results.

## Growth models

Because we are interested in all growth phases – the lag phase, the exponential phase, the deceleration phase, and the stationary phase1 – we use an extension of the classic logistic model: the *Baranyi-Roberts model* 8,9.

The Baranyi-Roberts growth model is defined by the following one-species ordinary differential equation (see eqs. 1c, 3a, and 5a in 8):

(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 , the deceleration is the same as in the classic logistic model and the density at the time of the maximum growth rate (the curve's inflection point) is half the maximum density . When or this density at the time of the maximum growth rate is higher or lower, respectively (to the right or the left of the inflection point in the logistic model).

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 the growth curve experiment. Therefore, populations that are adjusted to stationary phase must now adjust to growth, 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 Roberts8 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.

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

(2a)

. (2b)

We use four versions of the Baranyi-Roberts model. The full model, BR6, is described by eq. 2 and has six parameters. BR5 is the model in which the deceleration parameter is set to one, as in the classic logistic model. This model is useful because the reduced growth during the lag phase might sometimes be inferred as . BR4 is the model without a lag phase, where . This is also known as the Richards model or the generalized logistic model. This model is useful in cases where there is no observed lag phase: either because the organism in question adjusts very rapidly or because the population is already adjusted prior to the growth experiment, usually by priming it in fresh media before the experiment. The last model is BR3, in which and . This is simply the classic logistic model, .

## Model fitting and selection

We fit all four models to the mean growth curve of each strain using a least-squares procedure10. The standard deviation at each time point is used as weights for 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 growth at time point , and is the expected growth at time point according to the model. We select the model with the lowest BIC.

As a sanity check, we also fit the data using a linear model () and check that the BIC of our selected model is larger than the BIC of the linear model by at least 6 (See 11 for significance of BIC differences).

We repeat 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* and therefore captured by the deceleration of 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

.

There are different ways to infer selection coefficients (or relative fitness) from competition dynamics. For example, Chevin2 suggests that the selection coefficient of the strain of interest is

(4)

where is the frequency of the strain of interest at time .

Lenski and colleagues4 use a different measure:

(5)

where is usually chosen to be 24 hours.

Crow and Kimura3 have suggested that the frequency of the strain of interest follows a logistic equation:

(6)

Following the suggestion of Crow & Kimura, we fit eq. 6 to the frequency time-series of the strain of interest in the predicted competitions to find an estimate for the selection coefficient *s*. Of course, the competition prediction can be readily used to infer fitness in other ways, for example by using eq. 4 or 5.

# Discussion

We present a new computational method to predict the results of competition assays based on the growth curves of each separate strain.

This method should be very useful because growth curve assays require much less effort and preparation than competition assays. For example, one common protocol for competition assays6 requires the insertion of genes coding for fluorescent proteins to the strains in question and the measurement of the two fluorescence markers using a flow cytometry reader. Another approach4 requires the deletion of the arabinose utilization gene from one of the strains. The respective assay requires plating the bacteria on agar plates specific for identification of *ara-* mutants and counting colonies after overnight growth. In contrast, growth curve assays only require the growth of the two strains and the measurement of their optical density once every 10-60 minutes. As automatic 96-well microplate readers become more and more frequent in microbiology labs, this assay can be prepared in less than 30 minutes, after which the measurements are automatically collected by the plate reader.

The growth model that we use - the Baranyi-Roberts model - has a differential equation form and a closed form solution. Hence, it is highly 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. In addition, the Baranyi-Roberts model is an extension of the logistic model which is widely used in textbooks and it has a clear biological interpretation.

Our method assumes that the two strains interact via resource competition alone; 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 assays can be compared to model predictions and a goodness of fit test (such as the Kolmogorov-Smirnov 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 interactions12.

## Conclusions

To conclude, we propose a new method to analyze growth curves and infer 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 assays.

# Acknowledgments

We thank E. Kroll, Y. Pilpel, D. Hizi, I. Frumkin, O. Dahan, A. Yona, I. Ben-Zion, and J. Masel for helpful discussions. This work was funded by the Israeli Science Foundation (XXX), the Minerva Center for Lab Evolution, Manna Center Program for Food Safety & Security, the Israeli Ministry of Science & Technology, and the Anat Krauskopf Foundation.

# References

1. Hall, B. G., Acar, H., Nandipati, A. & Barlow, M. Growth rates made easy. *Mol. Biol. Evol.* **31,** 232–238 (2014).

2. Chevin, L.-M. On measuring selection in experimental evolution. *Biol. Lett.* **7,** 210–3 (2011).

3. Crow, J. F. & Kimura, M. *An introduction to population genetics theory*. (Burgess Pub. Co., 1970). at <https://books.google.co.il/books?id=MLETAQAAIAAJ>

4. Wiser, M. J. & Lenski, R. E. A Comparison of Methods to Measure Fitness in *Escherichia coli*. *PLoS One* **10,** e0126210 (2015).

5. Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence During 2,000 Generations. *Am. Nat.* **138,** 1315 (1991).

6. Hegreness, M., Shoresh, N., Hartl, D. L. & Kishony, R. An equivalence principle for the incorporation of favorable mutations in asexual populations. *Science (80-. ).* **311,** 1615–7 (2006).

7. Levy, S. F. *et al.* Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* **advance on,** (2015).

8. Baranyi, J. & Roberts, T. a. A dynamic approach to predicting bacterial growth in food. *Int. J. Food Microbiol.* **23,** 277–294 (1994).

9. Baranyi, J. Simple is good as long as it is enough. *Commentary* 391–394 (1997). doi:10.1006/fmic.1996.0080

10. Newville, M. LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python. (2014). at <http://lmfit.github.io/lmfit-py/>

11. Kass, R. & Raftery, A. Bayes Factors. *J. Am. Stat. Assoc.* 773–795 (1995). doi:doi: 10.2307/2291091

12. Masel, J. Eco-evolutionary ‘fitness’ in 3 dimensions: absolute growth, absolute efficiency, and relative competitiveness. 1–44 (2014). at <http://arxiv.org/abs/1407.1024>