­Predicting competition results from growth curves

Yoav Ram1, Eynat Deluss-Gur1, Uri Obolski1,

Maayan Bibi2, Judith Berman2 and Lilach Hadany1\*

December 30, 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, fitness, selection coefficient, experimental evolution

# Abstract

Measuring relative fitness in microbes from pairwise competition experiments is laborious and expensive. Accordingly, many microbiologists use growth curve experiments in liquid culture to estimate fitness from the maximum growth rate during the exponential growth phase. However, maximum growth rates have been shown to be an unreliable measure of fitness as indicated by discrepancies with the outcomes of pairwise competition experiments. Here we propose a new method that estimates relative fitness by predicting the results of competition experiments from growth curves data using growth and competition models.

# 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 growing in liquid culture over a range of time. 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 curve as a measure of the growth rate (Hall et al. 2014). Indeed, growth rates can be proxies of the selection coefficient, *s*, which is a standard approach for representing relative fitness in population genetics (Crow and Kimura 1970; Chevin 2011). However, the selection coefficient can be affected by other phases of growth, such as the lag phase and the stationary phase. Thus, it is not surprising that growth rates can be poor estimates of relative fitness (Concepción-Acevedo et al. 2015).

## Competition experiments

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 a mixed culture: 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 mixed culture is measured during the course of the experiment. This is done classically by plating assays that distinguish the strains using phenotypic markers (Wiser and Lenski 2015). More recently, flow cytometry has been used with fluorescently marked cells (Gallet et al. 2012) and deep sequencing read counts have been used to determine the frequencies of different alleles in the population (Bank et al. 2014; Levy et al. 2015). The selection coefficient of the strains of interest can then be estimated from changes in their frequencies during the competition experiments. These methods infer relative fitness with high precision (Gallet et al. 2012), as they directly estimate fitness from changes in frequencies over time. However, competition experiments are more laborious than growth curve 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.

Here we present a new computational method named *Curveball*: a descriptive and predictive framework for estimating growth parameters from growth curves, predicting competition results, and inferring relative fitness.

# Results

Our method includes four stages: (i) fitting growth models to monoculture growth curves data, (ii) fitting competition models to mixed culture growth curve data, (iii) using the estimated growth and competition parameters to predict the results of competition experiments, and (iv) estimating fitness from the predicted competition results.

## Growth models

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

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

(1a)

(1b)

(1c)

where is the population density, is time, is the specific growth rate in low density (see Appendix A) , is the adjustment function (see below), is the maximum density, and is the specific increase in log growth rate (see Appendix A). For a derivation of eq. 1 from first principles, see Appendix A.

The term is used to describe the deceleration in the growth of the population as it approaches the maximum density . When the specific increase in log growth rate is unity (), the deceleration is the same as in the standard logistic model and the density at the time of the maximum growth rate is half the maximum density, . When or , the deceleration is slower or faster, respectively, and the density at the time of the maximum growth rate is (Richards 1959, substituting ).

The adjustment function describes the fraction of the population (between 0 and 1) that is adjusted for growth at time . Typically, microorganisms are grown in overnight liquid culture, reaching stationary phase, and then diluted into fresh media. Therefore, populations that are adjusted to stationary phase must now adjust to growth conditions, and this might take some time. This adjustment phase is called the *lag phase*. The specific adjustment function we use here (eq. 1c) was suggested by Baranyi and Roberts (1994) due to being both computationally convenient and having a biological interpretation: 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 (eq. 1) has a closed form solution:

(2a)

, (2b)

where is the initial population density (usually determined by the dilution chosen by the experimenter).

We use six 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 , as in the standard logistic model. Another five parameter form has both rate parameters set to the same value (), which can make the fitting procedure more stable (Baranyi 1997). A four parameter form has both of the previous reductions, setting and (Baranyi 1997). Another four parameter form of the model has no lag phase, with . This is also known as the Richards model (Richards 1959), the -logistic model (Gilpin and Ayala 1973), or the generalized logistic model (see Wikipedia). 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 fifth form is the standard logistic model, in which and .

## Model fitting and selection

We fit all six model forms to the growth curve data of each strain using the *leastsq* non-linear curve fitting procedure (Jones et al. 2001; Newville et al. 2014). The standard deviation at each time point is used to weight the curve fitting 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 model parameters, is the number of data points, are the time points, is the optical 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 (Kass and Raftery 1995; Ward 2008).

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 (Fig. 1B).

## Competition model

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

(3a)

(3b)

(3c)

, (3d)

where is the density of strain and and are the values of the corresponding parameters for 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 (Fig. 1C).

This double 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.

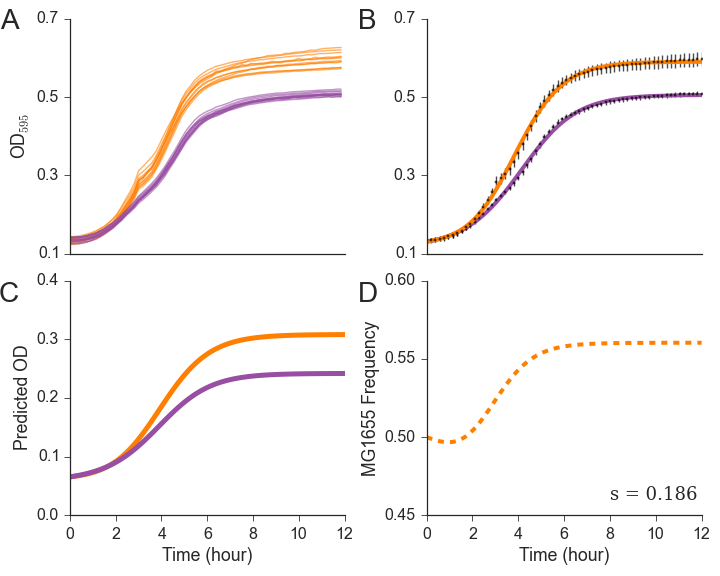


Figure 1. Example of the method applied on growth curves of two *Escherichia coli* strains (A) Growth curves data of MG1655 in orange (top lines) and DH12S in purple (bottom lines). Each line (12 per strain) represents a series of OD595 measurements from a single well in a 96-well microplate (Costar), taken every 10 minutes. Cells of either strain with Kan+Cap+ plasmids were diluted 1:20 from overnight culture and grown in 100 µl LB with 50 mg/ml Kanamycin and 34 mg/ml Chloramphenicol at 30°C in an automatic plate reader (Tecan Infinite 200Pro). The OD of cell-free wells was ~0.13. (B) Solid line: model fit; markers and error bars: mean and standard deviation of OD595 measurements from 12 wells per strain. Fitted parameters for 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; for 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, calculated by solving eq. 3. Initial OD of both strains was set to 0.067, half of the average estimated *N0* in both strains. (D) The frequency of MG1655 during the predicted competitions (dashed line). The estimated selection coefficient is *s*~0.186, calculated with eq. 4 and *t*=12. Note that the frequency of MG1655 initially declines slightly 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).

## Selection coefficient inference

One common method for estimating relative fitness or selection coefficients from pairwise 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. Eq. 4 can be applied to the predicted competition results to infer the selection coefficient of the strain of interest (Fig. 1D).

# Discussion

We present a new computational method to predict the results of competitions between two strains from the separately measured 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 dependent upon specific experimental conditions that differ for different organisms and from lab to lab; 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 (eq. 1) and a closed form analytical solution (eq. 2). Hence, it is very useful for our method: the closed form is used to fit to the 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 relative fitness. Our method improves fitness estimation from growth curves, has a clear biological interpretation, and can be used as a null model for the interpretation of competition experiments.

# Acknowledgments

We thank E. Kroll, Y. Pilpel, D. Hizi, I. Françoise, I. Frumkin, O. Dahan, A. Yona, T. Pupko, A. Eldar, I. Ben-Zion, E. Even-Tov, H. Acar, J. Barrick, and J. Masel for helpful discussions and L. Zelcbuch, N. Wertheimer, A. Rosenberg, A. Zisman, F. Yang, E. Shtifman Segal, and R. Yaari for sharing materials and experimental advice. 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).

# Appendices

## Appendix A: Deriving the growth model

Consider a population whose density or count is denoted by *N*. Let us denote the specific growth rate (or the per-capita growth rate) by *Z* (Schnute 1981):

.

Let us assume that the specific rate of *Z* is a linear function of itself:

.

Now, if we substitute we get the following differential equations system:

|  |  |  |
| --- | --- | --- |
|  | . | [A1a]  [A1b] |

Solving eqs. A1 with the boundary conditions and (where is the time after which the *N* is constant) gives:

|  |  |  |
| --- | --- | --- |
|  | . | [A2] |

However, this is also the solution to the Richards ODE (Richards 1959) which is equivalent to eq. 1 without lag :

|  |  |  |
| --- | --- | --- |
|  | . | [A3] |

Eq. A3 is an autonomous differential equation (*dN/dt* doesn't depend on *t*). To include the lag phase, Baranyi and Roberts (1994) suggested to add an adjustment function (which makes the equation non-autonomous):

|  |  |  |
| --- | --- | --- |
|  | . | [A4] |

Baranyi and Roberts suggested a Michaelis-Menten type of function (Baranyi 1997), which has two parameters: *q0* is the initial fraction of the population that is adjusted to growth and *v* is the adjustment rate, or the rate at which additional cells become adjusted to growth. Integration now leads to the solution in eq. 2.

Note that Baranyi and Roberts suggest to fix to avoid instability of the fitting procedure (Baranyi 1997).

To interpret we notice that , and together with eq. A1b, we can define as the **specific increase in log growth rate**. *r* retains its original interpretation as the **specific growth rate in very low density**:

.

## Appendix B: Deriving the competition model

We start by deriving the single-species logistic model, then derive the two-species logistic model.

### Single-species logistic model

In the context of microbial growth, we derive the logistic model from resource consumption perspective (Otto and Day 2007, 365). It is also possible to do so from a predator-prey perspective or from intraspecific interference perspective [REF].

Consider a population of consumers with density *N* and denote the density of its limiting resource by *R*. We assume that *R* starts with density *R0*and is depleted by the population growth, without being replenished (this is probably the case when the population grows in a sealed vial):

|  |  |  |
| --- | --- | --- |
|  |  | [B1a]  [B1b] |

where *h* is the resource intake rate is the conversion rate between resource mass and the consumer biomass. We now use a conservation law approach by setting which denotes the potential biomass when all resources are consumed (Dilao and Domingos 1999). We can see that *K* is constant:

,

and so we can substitute in eq. (B1b):

,

which gives the logistic differential equation, with the maximum population density *K* and the specific growth rate in very low density which is the intake rate at very low density, when all biomass is still in the resources.

## Two-species logistic equation

We now consider the case in which two species (or strains) compete for a single limiting resource:

|  |  |  |
| --- | --- | --- |
|  |  | *[B2a]*  *[B2b]*  *[B2c]* |

Similar to the one-species case, we define to find that, again, *K1* is constant:

.

We then substitute into eq. (B2b) and (B2c) and setting ("how much biomass of species 2 could be made from the made from the biomass of species 1 when it is in maximum density") and :

.

Note that this competition model is different from the one described in some textbooks, in which .

We get a similar result if each strain is limited by a different resource that both stains consume, and the intake rates depend only on the strain:

|  |  |  |
| --- | --- | --- |
|  |  | *[B3a]*  *[B3b]*  *[B3c]*  *[B3d]* |

Here, we notice first that and therefore is a constant. We then substitute in eqs. [B3a,c,d] and continue as above. This only changes the definition of .

If the intake rates depend only on the resource:

|  |  |  |
| --- | --- | --- |
|  |  | *[B4a]*  *[B4b]* |

Then we define and and again continue as above.

# References

Bank, Claudia, Ryan T. Hietpas, Alex Wong, Daniel N. A. Bolon, and Jeffrey 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): 1–35. 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.

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.

Crow, James F., and Motoo Kimura. 1970. *An Introduction to Population Genetics Theory*. Minneapolis: Burgess Pub. Co.

Dilao, Rui, and Tiago Domingos. 1999. “A General Approach to the Modelling of Trophic Chains.” *Ecological Modelling* 132 (3): 20. doi:10.1016/S0304-3800(00)00272-6.

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.

Gilpin, Michael E., and Francisco J. Ayala. 1973. “Global Models of Growth and Competition.” *Proceedings of the National Academy of Sciences of the United States of America* 70 (12 Pt 1-2): 3590–3593. doi:10.1073/pnas.70.12.3590.

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.

Jones, Eric, Travis Oliphant, Pearu Peterson, and others. 2001. “SciPy: Open Source Scientific Tools for Python.” http://www.scipy.org/.

Kass, Robert, and Adrian Raftery. 1995. “Bayes Factors.” *Journal of the American Statistical Association*: 773–795. doi:doi: 10.2307/2291091.

Levy, Sasha F., Jamie R. Blundell, Sandeep Venkataram, Dmitri a. Petrov, Daniel S. Fisher, and Gavin Sherlock. 2015. “Quantitative Evolutionary Dynamics Using High-Resolution Lineage Tracking.” *Nature* 519 (7542) (February 25): 181–186. doi:10.1038/nature14279.

Masel, Joanna. 2014. “Eco-Evolutionary ‘Fitness’ in 3 Dimensions: Absolute Growth, Absolute Efficiency, and Relative Competitiveness.” Populations and Evolution. *arXiv:1407.1024v3* (July): 1–44.

Newville, Matthew, Antonino Ingargiola, Till Stensitzki, and Daniel B. Allen. 2014. “LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python” (September 21). doi:10.5281/zenodo.11813.

Otto, Sarah P., and Troy Day. 2007. *A Biologist’s Guide to Mathematical Modeling in Ecology and Evolution*. Princeton University Press.

Richards, F. J. 1959. “A Flexible Growth Function for Empirical Use.” *Journal of Experimental Botany* 10 (2): 290–301. doi:10.1093/jxb/10.2.290.

Schnute, Jon. 1981. “A Versatile Growth Model with Statistically Stable Parameters.” *Canadian Journal of Fisheries and Aquatic Sciences* 38 (9) (September): 1128–1140. doi:10.1139/f81-153.

Ward, Eric J. 2008. “A Review and Comparison of Four Commonly Used Bayesian and Maximum Likelihood Model Selection Tools.” *Ecological Modelling* 211 (1-2): 1–10. doi:10.1016/j.ecolmodel.2007.10.030.

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.