Predicting competition results from growth curves

Yoav Ram[1], Eynat Deluss-Gur[1], Uri Obolski[1],

Maayan Bibi[2], Judith Berman[2] and Lilach Hadany[1*]

July 23, 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

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 experiments. Here we propose a new method that estimates relative fitness by predicting the results of competition experiments from single strain growth curves.

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 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 a growth curve 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 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 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 the frequencies of the different strains during competition experiments. These are good methods to infer relative fitness, as they directly estimate fitness from changes 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

2

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 in an open source Python package called *Curveball* (http://curveball.yoavram.com).

Model and Results

Our method includes three stages: (i) fitting growth models to 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, and the stationary phase – we use an extension of the standard logistic model, the Baranyi-Roberts model (Baranyi and Roberts 1994; Baranyi 1997).

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

$$\frac{dN}{dt} = r\alpha(t)N\left(1 - \left(\frac{N}{K}\right)^{\nu}\right) \tag{1a}$$

$$N(0) = N_0 \tag{1b}$$

$$\alpha(t) = \frac{q_0}{q_0 + e^{-mt}} \tag{1c}$$

where N is the population density, r is the per capita growth rate, t is time, $\alpha(t)$ is the adjustment function (see below), K is the maximum density, and ν is a deceleration parameter (see below for q_0 and m).

The term $(1 - (N/K)^{\nu})$ is used to describe the deceleration in the growth of the population as it nears the maximum density K. When the deceleration parameter ν is

unity ($v = 1$), the deceleration is the same as in the standard logistic model and the density at the time of the maximum growth rate $\left(\frac{d^2N}{dt^2}(t) = 0\right)$ is half the maximum density, $\frac{K}{2}$. When $v > 1$ or $0 < v < 1$, the deceleration is slower or faster, respectively, and the density at the time of the maximum growth rate is $\dfrac{K}{(1+v)^{\frac{1}{v}}}$

(Richards 1959, substituting $W = N, A = K, v = $ m-1$, k = $ r $\cdot$ v).

The adjustment function $\alpha(t)$ is used to describe the adjustment of the population to growth conditions at the beginning of the growth curves experiment. Typically, microorganisms are grown in overnight culture to stationary phase and 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 meaning: q_0 is the initial amount of some molecule (nutrient, enzyme, etc.) that is required for growth; m is the rate in which this molecule is accumulated in the cell.

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

$$N(t) = \frac{K}{\left[1-\left(1-\left(\frac{K}{N_0}\right)^v\right)e^{-r\,vA(t)}\right]^{1/v}} \tag{2a}$$

$$A(t) := \int_0^t \alpha(s)\,ds = t + \frac{1}{m}\log\left(\frac{e^{-mt}+q_0}{1+q_0}\right), \tag{2b}$$

where $N_0 \equiv N(0)$ is the initial population density.

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 v set to unity, as in the standard logistic model. A four parameter form of the model has no lag phase, with $1/m = 0 \Rightarrow A(t) \equiv t$. This is also known as the Richards model (Richards 1959) 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 standard logistic model, in which $v = 1$ and $1/m = 0$.

4

Model fitting and selection

We fit all four model forms to the mean growth curve of each strain in the data using non-linear curve fitting (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:

$$BIC = n \cdot \log\left(\frac{\sum_{i=1}^{n}\left(N(t_i) - \widehat{N}(t_i)\right)^2}{n}\right) + k \cdot \log n,$$

where k is the number of parameters of the model, n is the number of time points t_i, $N(t_i)$ is the average density at time point t_i, and $\widehat{N}(t_i)$ is the expected density at time point t_i 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 ($N(t) = a \cdot t + b$) and check that the BIC of our selected model form is smaller than the BIC of the linear model by at least 6 [See (Kass and Raftery 1995) 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 (Fig. 1B).

Competition prediction

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

$$\frac{dN_1}{dt} = r_1 \alpha_1(t) N_1 \left(1 - \left(\frac{N_1}{K_1}\right)^{\nu_1} - \left(\frac{N_2}{K_2}\right)^{\nu_2}\right) \tag{3a}$$

$$\frac{dN_2}{dt} = r_2 \alpha_2(t) N_2 \left(1 - \left(\frac{N_1}{K_1}\right)^{\nu_1} - \left(\frac{N_2}{K_2}\right)^{\nu_2}\right) \tag{3b}$$

$$\alpha_i(t) = \frac{q_{0,i}}{q_{0,i} + e^{-m_i t}} \tag{3c}$$

$$N_i(0) = N_{0,i}, \tag{3d}$$

where N_i is the density of strain i and $r_i, K_i, \nu_i, \alpha_i, q_{0,i}$, and m_i are the values of the corresponding parameters for strain i 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

5

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 K_i for each strain.
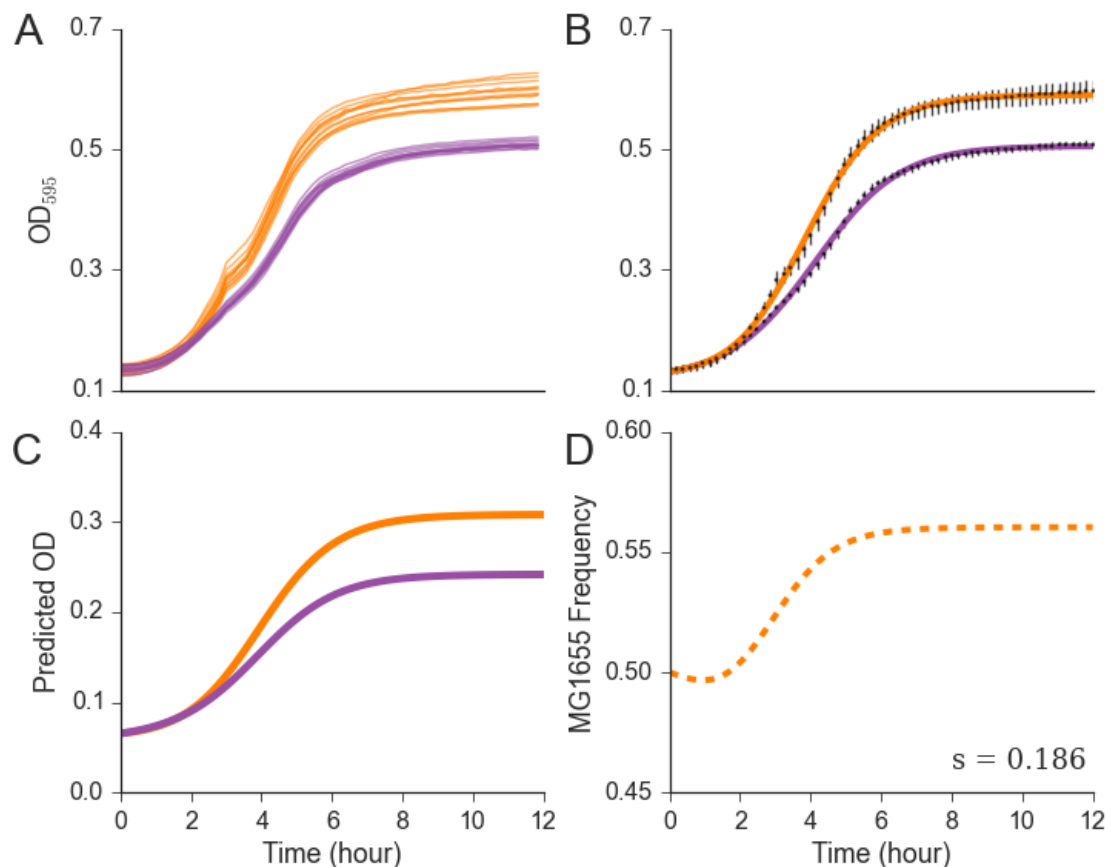

Figure 1. Example of the method applied on growth curves of two *Escherichia coli* strains (A) Growth curves data of MG1655 in orange (top lines) and DH12S in purple (bottom lines). Each line (12 per strain) represents a series of OD_{595} measurements from a single well in a 96-well microplate (Costar), 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 OD_{595} measurements from 12 wells per strain. Fitted parameters for MG1655: N_0=0.134, r=0.416, v=2.73, K=0.588, q_0=0.053, m=2.37, lag duration=1.714, maximum growth rate=0.357; for DH12S: N_0=0.13, r=0.876, v=1, K=0.505, q_0=0.15, m=0.772, lag duration=1.691, maximum growth rate=0.279. Note that the maximum growth rate $\left(\max\frac{dN}{dt}\right)$ is a function of r, v, 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 N_0 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).

Ram et al., 2015

Selection coefficient inference

One common method for estimating relative fitness or selection coefficients from pairwise competition results is (Wiser and Lenski 2015):

$$s = \log\left(\frac{N_1(t)}{N_1(0)}\right) \Big/ \log\left(\frac{N_2(t)}{N_2(0)}\right) - 1, \tag{4}$$

where N_1 and N_2 are the densities of the strains and t 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.

7

Our method assumes that the two strains interact solely via resource competition; that is, only through the factor $(1 - N_1/K_1 + N_2/K_2)$. 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

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.

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* advance on. doi:10.1038/nature14279. http://dx.doi.org/10.1038/nature14279.

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, 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. http://zenodo.org/record/11813.

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. http://jxb.oxfordjournals.org/lookup/doi/10.1093/jxb/10.2.290.

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://dx.plos.org/10.1371/journal.pone.0126210.