
Modelling Competition for Nutrients between Microbial Populations Growing on Solid Agar Surfaces

Author: Daniel Boocock; Supervisor: Dr Conor Lawless

August 14, 2016

ABSTRACT

Motivation: Growth rate is a major component of the evolutionary fitness of microbial organisms. When nutrients are plentiful, fast-growing strains come to dominate populations whereas slower-growing strains are wiped out. This makes growth rate an excellent (a useful) surrogate for the health of cells. Measuring the health of cells grown in different genetic backgrounds or environments can inform about genetic interaction and drug sensitivity. In high-throughput procedures such as QFA and SGA, arrays of microbial cultures are grown on solid agar plates and quantitative fitness estimates are determined from growth measurements. Diffusion of nutrients along gradients in nutrient density arising between fast- and slow-growing neighbours is likely to affect growth rate and fitness estimates. However, current analyses assume that cultures grow independently. We study data from QFA experiments growing *Saccharomyces cerevisiae* to test a mass action kinetic model of nutrient dependent growth and diffusion. We try to correct for competition to provide more accurate and precise fitness estimates.

Results: Don't know what to say yet.

Availability and Implementation: CANS, a Python package developed for the analysis in this paper, is freely available from [github](https://github.com/danielboocock/CANS).

Contact: daniel.boocock@protonmail.ch

1 INTRODUCTION

The bacteria *Escherichia Coli* and yeast *Saccharomyces cerevisiae* are unicellular organisms studied as a model prokaryote and eukaryote respectively. They grow in colonies, where cells may (be clones originating from a single cell or) belong to different genetic strains originating from different individual cells. In favourable conditions, growth is exponential and this makes growth rate a major component of fitness; faster growing strains quickly come to dominate the population. At a certain point growth becomes limited and a stationary phase is reached. For unicellular organisms, growth rate is equal to cell cycle progression rate and all of the genetic information must be copied before each division. As a result, evolutionary pressure has led to rapidly dividing organisms with compact genomes of essential genes. These genes have been conserved in other species over billions of years of evolution, which is, in part, what makes *E. Coli* and *S. cerevisiae* useful as model species. The eukaryote *S. cerevisiae*, is particularly useful for the study of other eukaryotes such as humans.

The growth rate of microbial organisms is measurable and is often used to determine fitness. In experiments, cell cultures are commonly grown in two types of medium: on the surface of a nutrient rich solid agar and in a liquid mixture containing nutrients. (REMOVE: In spot tests (phenotypic array), cultures are pinned or inoculated on the surface of a solid agar containing nutrients. In liquid culture assays, cultures are mixed in a liquid medium containing nutrients.) In both cases cultures are incubated and growth is observed. Identical strains can grow differently between the two mediums and disagreement in fitness estimates is currently an issue Baryshnikova *et al.* (2010a) (I couldn't find a paper specifically talking about this issue but they have a correlation plot Fig2a where correlations are worse with a liquid culture study by Jasnos and Korona; in fact the Baryshnikova paper Fig3c seems to say that they had strong correlation in their "high-resolution liquid growth profiling study"). I do not focus on this issue and exclusively study fitness screens using solid agar.

Fitness estimates can be used to infer genetic interaction or drug response and high-throughput methods allow this to be conducted on a genome-wide scale (see e.g. Costanzo *et al.* (2010); Andrew *et al.* (2013)). In a typical genetic interaction screen a strain is made with a mutation in a query gene. Double mutants are created by introducing a second deletion in this strain. By comparing the growth of double mutants with a control containing a neutral deletion, genetic interactions can be inferred. If a strain is fitter than the control then the deletion is said to suppress the defect of the query gene. If a strain is less fit than the control then the deletion is said to enhance the defect of the query gene. Either scenario suggests that the two genes interact and have

a related function. Due to redundancy, single deletions are often non-lethal. (Remove: Knock downs and conditional mutations can also be used.) This has allowed Costanzo *et al.* (2010) to explore genetic interactions for ~75% of the *S. cerevisiae* genome.

Synthetic Genetic Array (SGA) and Quantitative Fitness Analysis (QFA) are high-throughput methods for obtaining quantitative fitness estimates of microbial cultures grown on solid agar (Baryshnikova *et al.*, 2010b; Banks *et al.*, 2012). Typically one query gene and replicates of several deletions are pinned or inoculated in a rectangular array on a solid agar plate. Many plates with different query genes and deletions are grown in high-throughput to explore whole genomes. I study data from QFA which refers to quantitative estimation of fitness by measurement and fitting of growth curves. In a typical QFA procedure liquid cultures are inoculated onto solid agar (containing nutrients (already mentioned above)) in a 16x24 rectangular array of 384 spots. Inoculum density can be varied to capture more or less of the growth curve and the most dilute cultures are inoculated with ~100 starting cells (Addinall *et al.*, 2011). Plates are grown in incubation and removed to be photographed at timepoints throughout growth. Photographs are of whole plates and growth typically covers several days to capture both the exponential and stationary growth phases. Colonyzer (Lawless *et al.*, 2010) processes optical density measurements in photographs to produce a timecourse of cell density estimates for each culture. In past analysis, the logistic growth model was independently fit to the timecourse of each culture and fitness estimates were defined in terms of parameters of this model: the growth constant r and carrying capacity K . In contrast, SGA typically uses a larger array of 1536 pinned cultures and a single endpoint assay of culture area to quantify growth. The differential form and solution of the logistic model (Verhulst, 1845) (probably don't need this reference) are given in Equations 1, where C represents cell density and C_{t_0} is cell density at time zero.

$$\dot{C} = rC \left(1 - \frac{C}{K}\right) \quad (1a)$$

$$C(t) = \frac{KC_{t_0}e^{rt}}{K + C_{t_0}(e^{rt} - 1)} \quad (1b)$$

The logistic model is a simple mechanistic model describing self-limiting growth and has a sigmoidal solution. Growth begins exponentially with rate rC and curtails as the population size increases and cells begin to compete for space and nutrients (remove: or interact in some other way). Cell density reaches a final carry-

ing capacity K at the stationary phase. In QFA, nutrients must diffuse through agar to reach cells growing on the surface. It is plausible that the carrying capacity K represents the point at which nutrients either run out or growth becomes limited by the diffusion of nutrients and is approximately stationary. Fitting the logistic model to QFA data requires plate level or culture level parameters for C_{t_0} and culture level parameters for r and K making 769 or 1152 parameters per 384 culture plate.

//Could remove and just discuss MDR when I get to the results// The growth constant r could be used as a fitness measure. However, Addinall *et al.* (2011) define a more complicated fitness measure as the product of Maximum Doubling Rate (MDR) and Maximum Doubling Potential (MDP) which they calculate from logistic model parameters. MDR measures the doubling rate at the beginning of the exponential growth phase, when growth is fastest, and MDP is the number of divisions which a culture undergoes from inoculation to the stationary phase.

$$MDR = \frac{r}{\log\left(\frac{2(K-C_0)}{K-2C_0}\right)} \quad (2a)$$

$$MDP = \frac{\log\left(\frac{K}{C_0}\right)}{\log(2)} \quad (2b)$$

To improve the quality of fits, QFA now uses the generalised logistic model which requires an extra shape parameter for each culture. Standard and generalised logistic model r are not equivalent so comparison relies on MDR and MDP as fitness measures. The analysis of QFA data using both models is available through the QFA R package (Lawless *et al.*, 2016). //Could remove and just discuss MDR when I get to the results//

//Could remove//Addinall *et al.* (2011) used QFA and *S. cerevisiae* to screen for genes involved in telomere stability which is related to ageing and cancer and has implications for human health and disease. Hits from this study have been successfully followed to discover new biology (Holstein *et al.*, 2014). (To be honest I have no idea what they found in that paper. Maybe I should leave this to biologist and out of my report? We had a more general focus. Obviously I will mention the Addinall paper when I describe p15 in the methods.) //Could remove//

Since QFA aims to determine differences in the fitness of microbial strains from measurements of differences in growth, fast and slow growing cultures are often grown side-by-side. Figure 1 shows a section of a QFA plate from a study by Addinall *et al.* where this is the case. (Cultures were inoculated with approximately

equal cell density but have grown at different rates to visibly different sizes after ~ 2.5 days.) It is likely that nutrients diffuse along gradients between fast and slow growing neighbours causing growth to appear faster or slower than if it were independent. Further support for such an effect comes from the experiment shown in Figure 2 where the same cultures are grown in alternate columns on two separate plates but with cultures added or removed from the neighbouring columns in between. Cultures in Figure 2a), where neighbours were removed, grew faster and larger (how much? I can look at the data myself) than the same cultures in Figure 2b), where neighbours were added. This suggests that an interaction between neighbours is present and may be affecting fitness estimates. Current QFA analysis using the logistic model assumes that cultures grow independently and ignores possible competition effects between neighbours. The sigmoidal curve of the logistic model poorly fits QFA data in many cases and this may be due to competition effects. I aim to fit a network model of nutrient dependent growth and diffusion to QFA data to try to correct for competition and increase the accuracy and precision of fitness estimates.

Could explain the difference in dilute and more concentrated cultures. In the image captions or elsewhere?

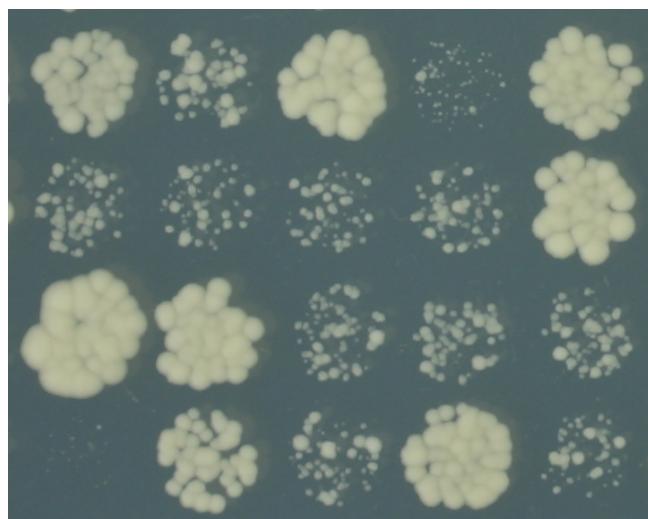


Figure 1: 4x5 section of a QFA plate. Taken from a 16x24 format solid agar plate inoculated with dilute *S. cerevisiae* cultures. Image captured at ~ 2.5 d after inoculation and incubation at 27°C .

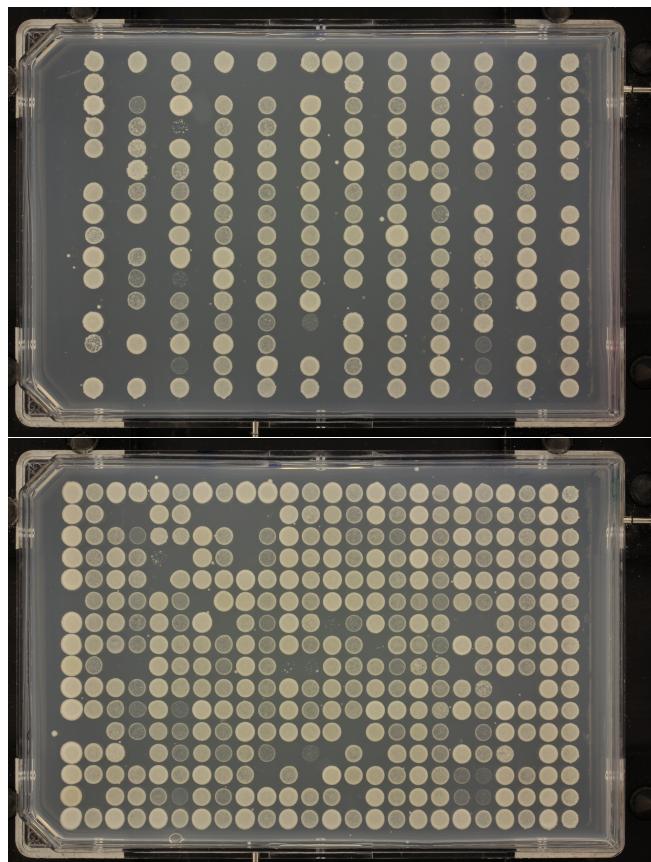


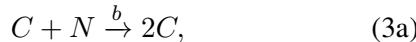
Figure 2: An experiment designed to examine competition. A) QFA plate inoculated with a more concentrated *S. Cerevisiae* inoculum (no cells inoculated on alternative columns). B) Same as in A, but with strains of similar growth rate inoculated in the positions missing in A.

Alternatively, competition effects could be dealt with experimentally by randomising the location of cultures on repeated plates. This does not require explicit knowledge or modelling of the source of competition but reduces throughput, so, if possible, a modelling approach is desirable. Poisoning of cultures by a signal molecule such as ethanol, which *S. cerevisiae* produces in the metabolism of sugars by fermentation, is another possible source of competition. QFA does not measure nutrients or signal, so if more than one source of competition exists, it will become very difficult to fit a model and randomisation may be the best approach. In an SGA study, Baryshnikova *et al.* (2010a) use statistical techniques to correct for competition between fast and slow growing neighbours in end-point assays of culture area. We hope that modelling the diffusion of nutrients explicitly and using a full network diffusion model will allow us to better correct for competition using fewer repeats. With QFA, we can also use more information by fitting whole growth curves rather than a single endpoint assay so a modelling approach offers to be more powerful. Furthermore, modelling may help us to identify and un-

derstand the source of competition. Simulation of an accurate model will allow us to compare experimental designs and explore ways to reduce competition effects.

//Diffusion Equation: I am probably going to have to reexplain this when I get to the discussion so I could just leave until then// (Reo and Korolev, 2014) use a diffusion equation model to simulate nutrient dependent growth of a single bacterial culture on a petri dish in two-dimensions. It would be too computationally intensive to fit a similar model to a full QFA plate in three-dimensions, especially if we are to use the model to process many plates from high-throughput experiments. Therefore a simpler model of nutrient diffusion is required. //Diffusion Equation//

Lawless (link blog) proposed a model of nutrient dependent growth and competition using mass action kinetics and a network diffusion model, hereinafter the competition model (3,4). A schematic of the model is drawn in Figure 3. He represents the nutrient dependent division of cells with the reaction equation,



where C is a cell, N is the amount of nutrient required for one cell division, and b is a rate constant for the reaction. The identity of the limiting nutrient N is unknown but possible candidates are sugar and nitrogen. He defines separate reactions (3) with growth constant b_i for each culture, indexed i , on a plate and uses mass action kinetics to derive rate equations for the amount of cells and nutrients associated with each culture, C_i and N_i . This gives the rate equation for C_i (4a) and the first term in the rate equation for N_i (4b).

$$\dot{C}_i = b_i N_i C_i, \quad (4a)$$

$$\dot{N}_i = -b_i N_i C_i - k_n \sum_{j \in \delta_i} (N_i - N_j). \quad (4b)$$

To arrive at the full competition model, he models the diffusion of nutrients along gradients between a culture i and its closest neighbours, δ_i , by the second term in (4b), where k_n is a nutrient diffusion constant. This can also be expressed as a series of mass action kinetic reactions,



Unlike the logistic model (1), the competition model has no analytical solution, and must instead be solved numerically. If k_n is set to zero (in the independent limit), the competition model reduces to the mass action equivalent of the logistic model, hereinafter the mass-action logistic model (blog ref)(, and has the same sigmoidal

solution). (In this limit, parameters of the competition model can be converted in terms of parameters the logistic model (see methods section)). When the competition model is fit to QFA data, C_i is observed and N_i is hidden. Inoculum density, C_{t_0} , is often below detectable levels. By assuming that inoculum density is the same for all cultures and that nutrients are distributed evenly throughout the agar at time zero, plate level initial values of cells and nutrients, C_{t_0} and N_{t_0} , can be used. k_n is assumed to be constant across the plate but must be inferred. There is a growth constant, b_i , for each of 384 cultures on a typical QFA plate making 387 parameters in total. The competition model shares more information between cultures and has fewer than half the number of the parameters of either the standard or generalised logistic model (Banks *et al.*, 2012; Lawless *et al.*, 2016).

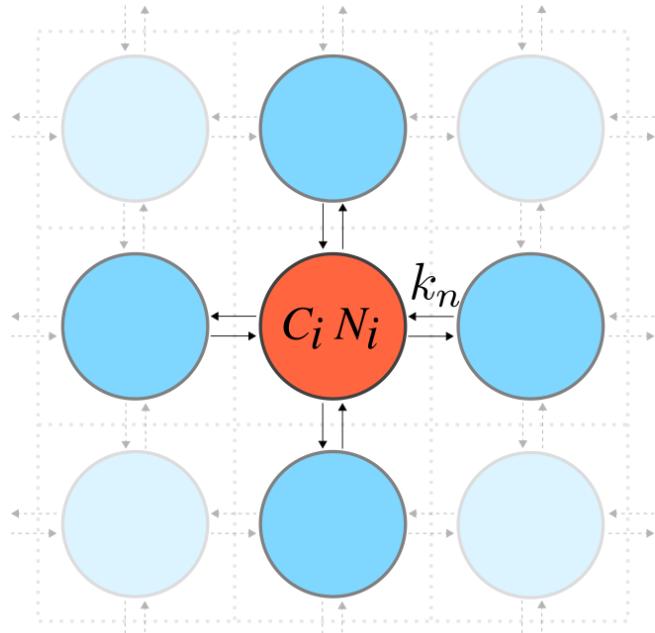


Figure 3: Schematic of the competition model. Each circle represents a culture, indexed i , growing in a rectangular array on the surface of a nutrient containing solid agar. Arrows represent a network of nutrient diffusion along gradients between cultures. C_i - amount of cells; N_i - amount of nutrients; k_n - plate level diffusion constant; darker blue circles δ_i - closest neighbours of culture i .

Mass action kinetics applies to reactions in a well stirred mixture and this is perhaps not valid for a culture growing on solid agar. The order of a reaction also affects the rate equation and the identity and quantity of the nutrient molecule in (3) is unknown. However, a mass action approximation has been successful in the Lotka-Volterra model of predator-prey dynamics (Bryman, 1992) and in signalling and reaction models inside cells (Aldridge *et al.*, 2006; Chen *et al.*, 2010) where

this assumption is also questionable. In QFA, populations begin with ~100 cells and quickly grow to reach thousands of cells so a deterministic approximation also seems valid. I am attempting to collectively fit a large model with many parameters to a set of data in a difficult optimisation problem where I only have observations for one variable (C). This involves many simulations of a model so if this is to be computationally feasible (and eventually used in the analysis of high-throughput data) then I must use a simplified model with many approximations. It is hoped that even an approximate model will measure more reliable growth parameters and better estimate fitness. This will increase our power to infer genetic interaction and drug response which could lead to further discoveries (Addinall and).

1.1 Subsection

REFERENCES

- Addinall, S.G. *et al.* (2011) Quantitative fitness analysis shows that nmd proteins and many other protein complexes suppress or enhance distinct telomere cap defects. *PLoS Genet*, **7**, 4, 1–16.
- Aldridge, B.B. *et al.* (2006) Physicochemical modelling of cell signalling pathways. *Nature cell biology*, **8**, 11, 1195–1203.
- Andrew, E.J. *et al.* (2013) Pentose phosphate pathway function affects tolerance to the g-quadruplex binder tmppyp4. *PLoS ONE*, **8**, 6, 1–10.
- Banks, A. *et al.* (2012) A quantitative fitness analysis workflow. <http://www.jove.com/video/4018/a-quantitative-fitness-analysis-workflow>
- Baryshnikova, A. *et al.* (2010a) Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nature Methods*, **7**, 12, 1017–24. Copyright - Copyright Nature Publishing Group Dec 2010; Last updated - 2014-09-19.
- Baryshnikova, A. *et al.* (2010b) Synthetic genetic array (sga) analysis in *saccharomyces cerevisiae* and *schizosaccharomyces pombe*. *Methods in enzymology*, **470**, 145–179.
- Berryman, A.A. (1992) The origins and evolution of predator-prey theory. *Ecology*, **73**, 5, 1530–1535.
- Chen, W.W. *et al.* (2010) Classic and contemporary approaches to modeling biochemical reactions. *Genes & development*, **24**, 17, 1861–1875.
- Costanzo, M. *et al.* (2010) The genetic landscape of a cell. *science*, **327**, 5964, 425–431.
- Holstein, E.M. *et al.* (2014) Interplay between nonsense-mediated mrna decay and {DNA} damage response pathways reveals that stn1 and ten1 are the key {CST} telomere-cap components. *Cell Reports*, **7**, 4, 1259 – 1269.
- Kopelman, R. (1988) Fractal reaction kinetics. *Science*, **241**, 4873, 1620–1626.
- Lawless, C. *et al.* (2010) Colonyzer: automated quantification of micro-organism growth characteristics on solid agar. *BMC Bioinformatics*, **11**, 1, 1–12.
- Lawless, C. *et al.* (2016) *qfa: Tools for Quantitative Fitness Analysis (QFA) of Arrayed Microbial Cultures Growing on Solid Agar Surfaces*. R package version 0.0-42/r678.
- Reo, Y.J. and Korolev, K. (2014) Modeling of Nutrient Diffusion and Growth Rate in Bacterial Colonies.
- Savageau, M.A. (1995) Michaelis-menten mechanism reconsidered: implications of fractal kinetics. *Journal of theoretical biology*, **176**, 1, 115–124.
- Verhulst, P. (1845) Recherches mathematiques sur la loi d'accroissement de la population. *Nouveaux memoires de l'Academie Royale des Sciences et Belles-Lettres de Bruxelles*, **18**, 14–54.