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

Author: Daniel Boocock; Supervisor: Dr Conor Lawless

August 19, 2016

1 RESULTS

1.1 Guessing

N_0 estimated from average final cell amounts. See formula in code for two N_0 estimation.

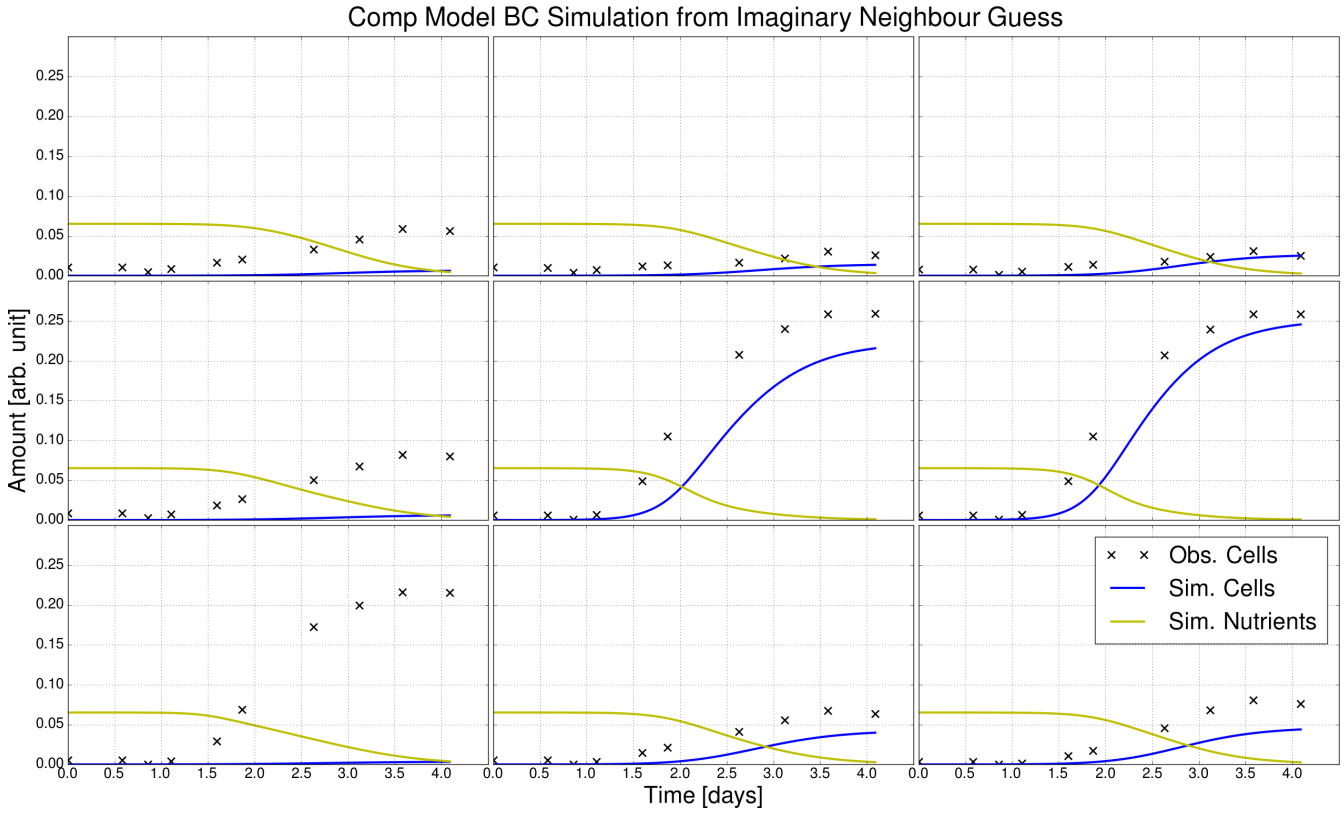

Figure 1: Competition model simulation using parameters from imaginary neighbour guessing. *Shows a 3x3 zone with top-left coordinate (5, 18) from P15 with background cdc13-1 at 27°C.*

1.2 Competition Model Fitting to P15

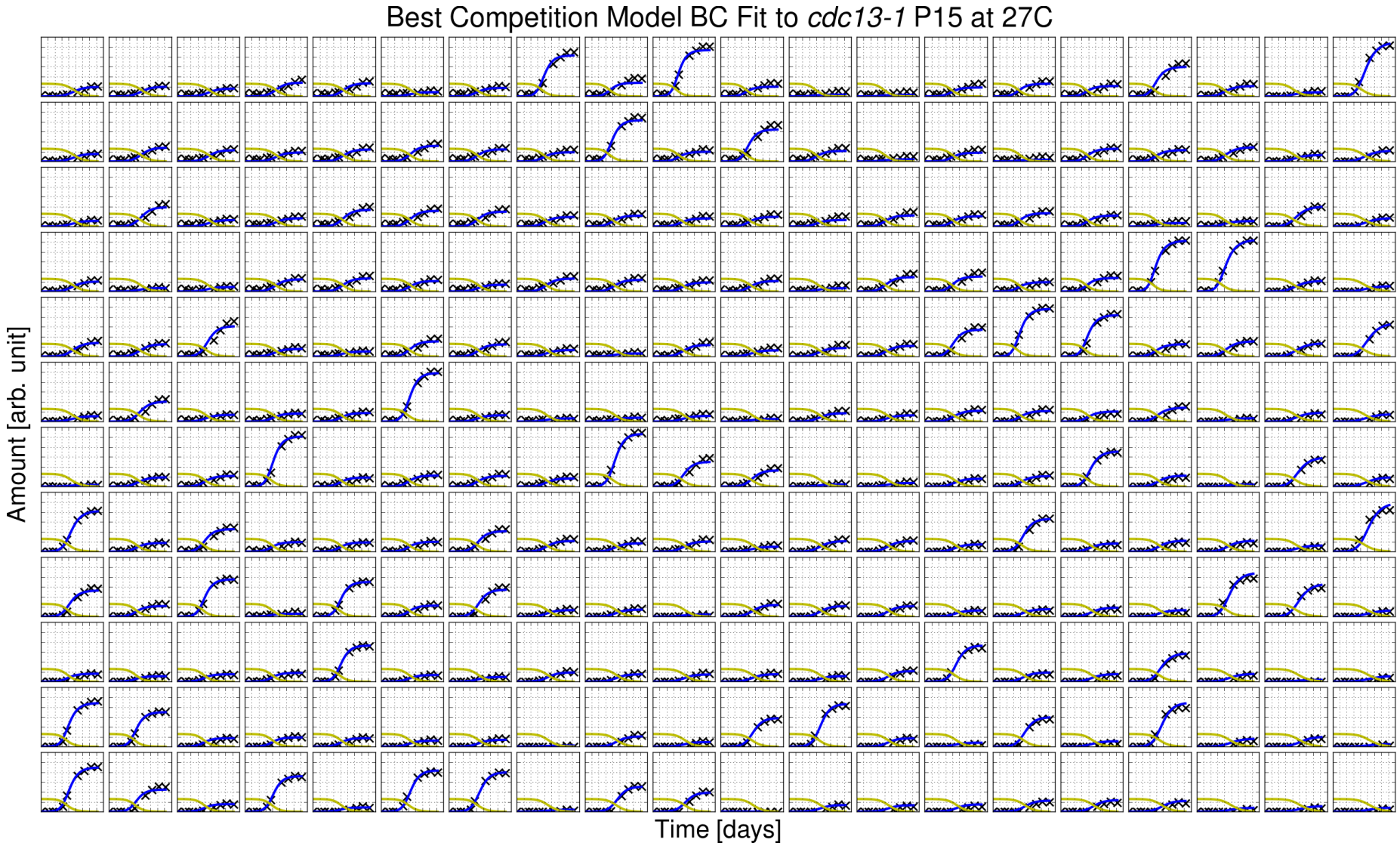

Figure 2: Fit of the competition model to a QFA plate. *Data is for a 16x24 format plate (P15) with a background mutation cdc13-1 incubated at 27°C. The plate contains 6 repeats of 50 genetic strains randomly arranged across the internal cultures. Repeats of a single strain are used for all edge cultures (removed in the plot). Model output for state variable, cell population size (blue curve), is fit to observed data (black crosses). Model predictions for unobserved variable (nutrient amount) are also plotted (yellow).*

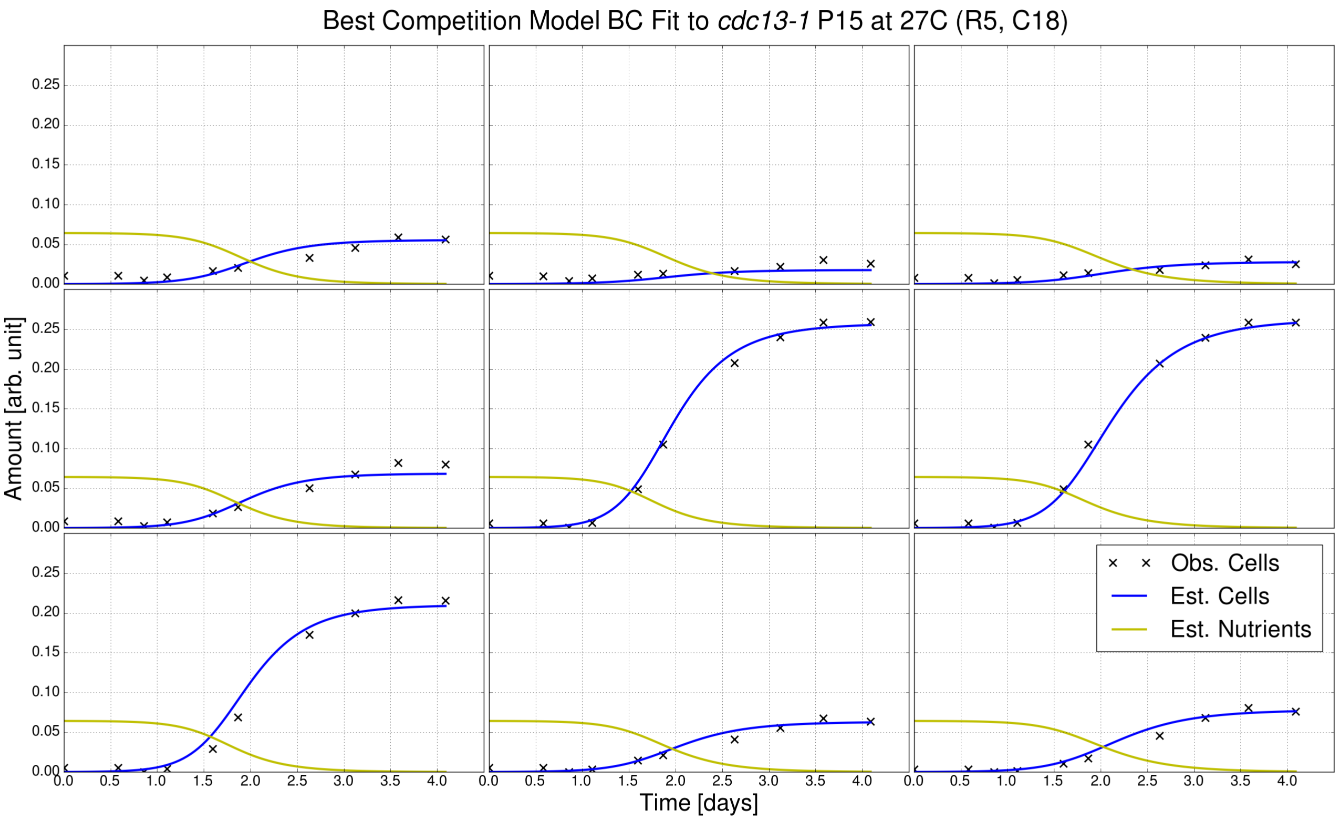

Figure 3: A 3x3 zone from Figure 2 with top-left coordinate (5, 18).

Some text Some text.

1.3 Evaluating the treatment of boundaries

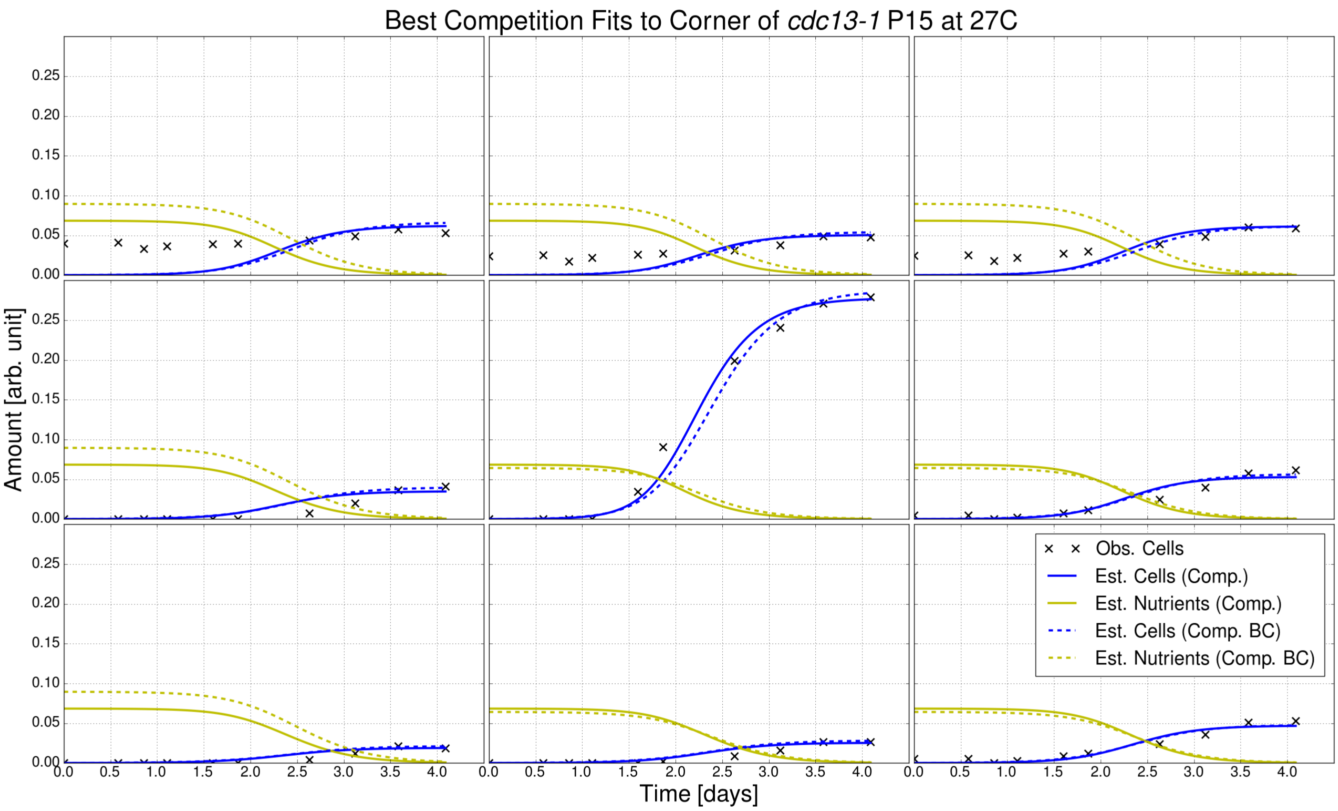

Figure 4: Treatment of boundary conditions in fits of the competition model. *The top left corner of a 16x24 QFA plate fitted with two versions of the competition model, the first has a single initial nutrient amount for all cultures, the second has a separate initial nutrient amount for edge cultures.*

Table 1: Average error in objective function for one a two N_0 parameter competition models. *Values are for the same fits as in Figure 4 and have been scaled by 10^4. Averages are for cultures belonging to the areas indicated by the column "Cultures". "Next to edge" refers to cultures one in from the edge. "Internal" refers to all cultures but the edge.*

Cultures	One N_0	Two N_0
Edge	35.9	36.5
Next to edge	9.54	7.98
Internal	6.67	6.30
All	12.4	12.2

Table 1

Some text Some text

Some text Some text.

1.4 Agreement of b rankings

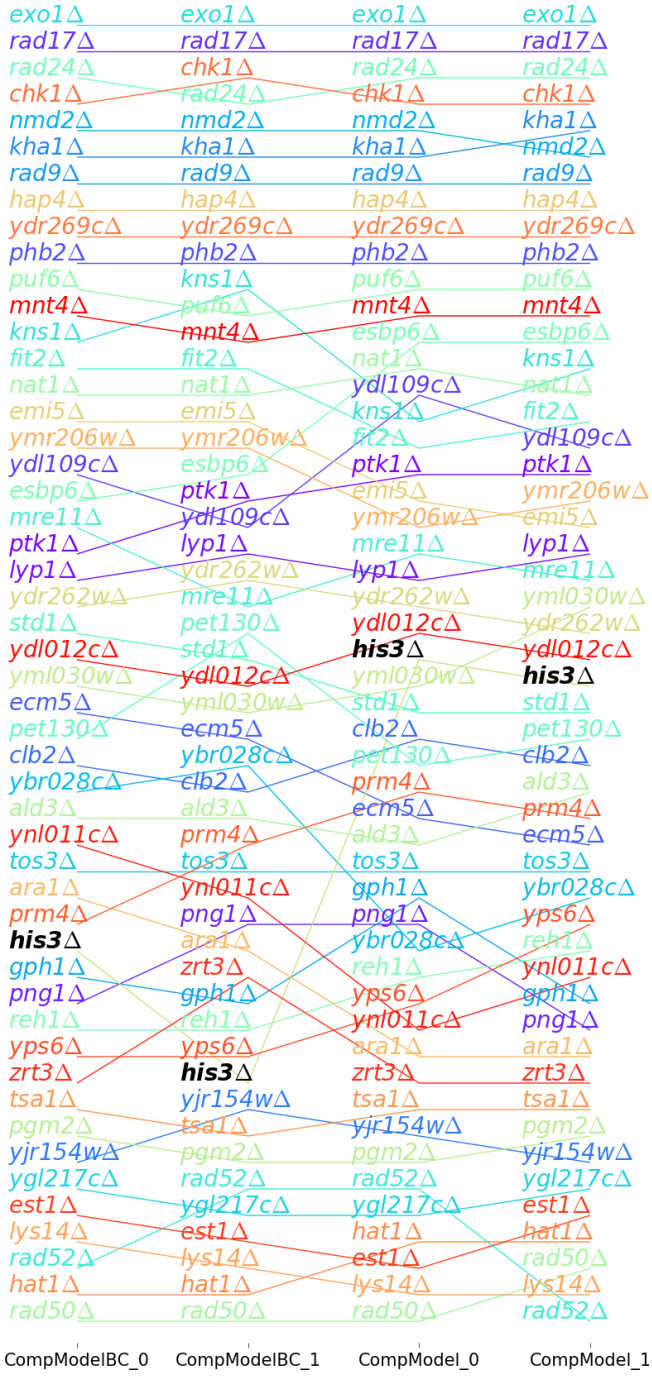

Figure 5: Comparison of b ranking for the best five competition model fits to P15. *Ranking is calculated from the mean b estimate from the six repeats or each strain.*

1.5 Comparison of fitness ranking

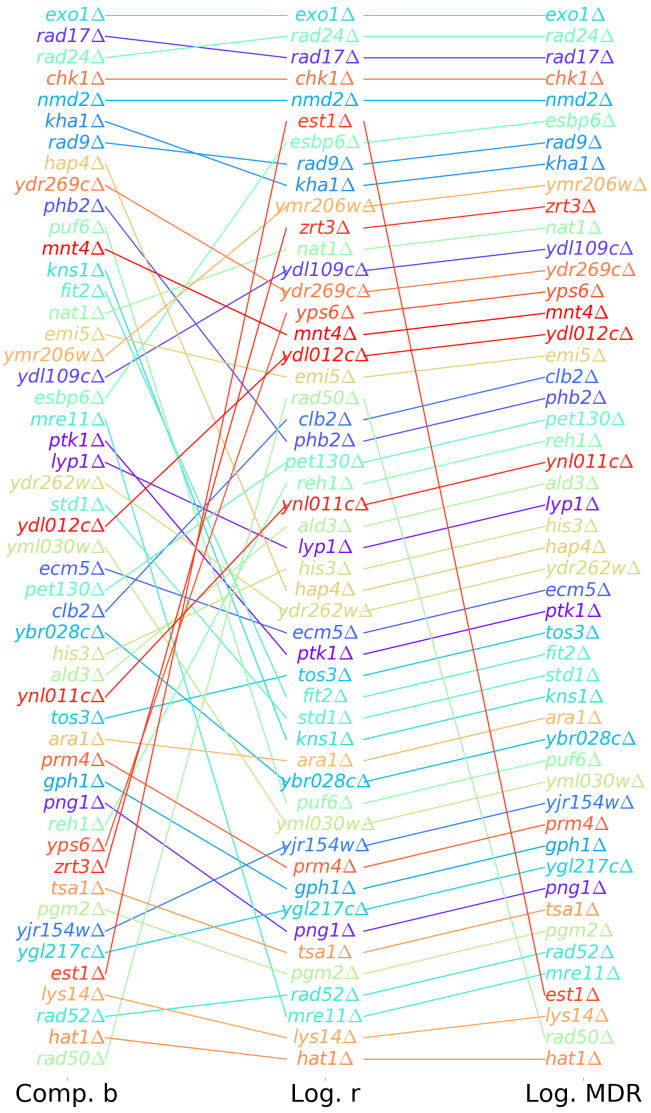

Figure 6: Comparison of r ranking for fits of the competition and logistic model to P15. *Competition model r was converted from b, N_0, and C_0 from the best competition model estimate. Logistic r was taken from fits using the QFA R package which makes heuristic checks for slow growing cultures.*

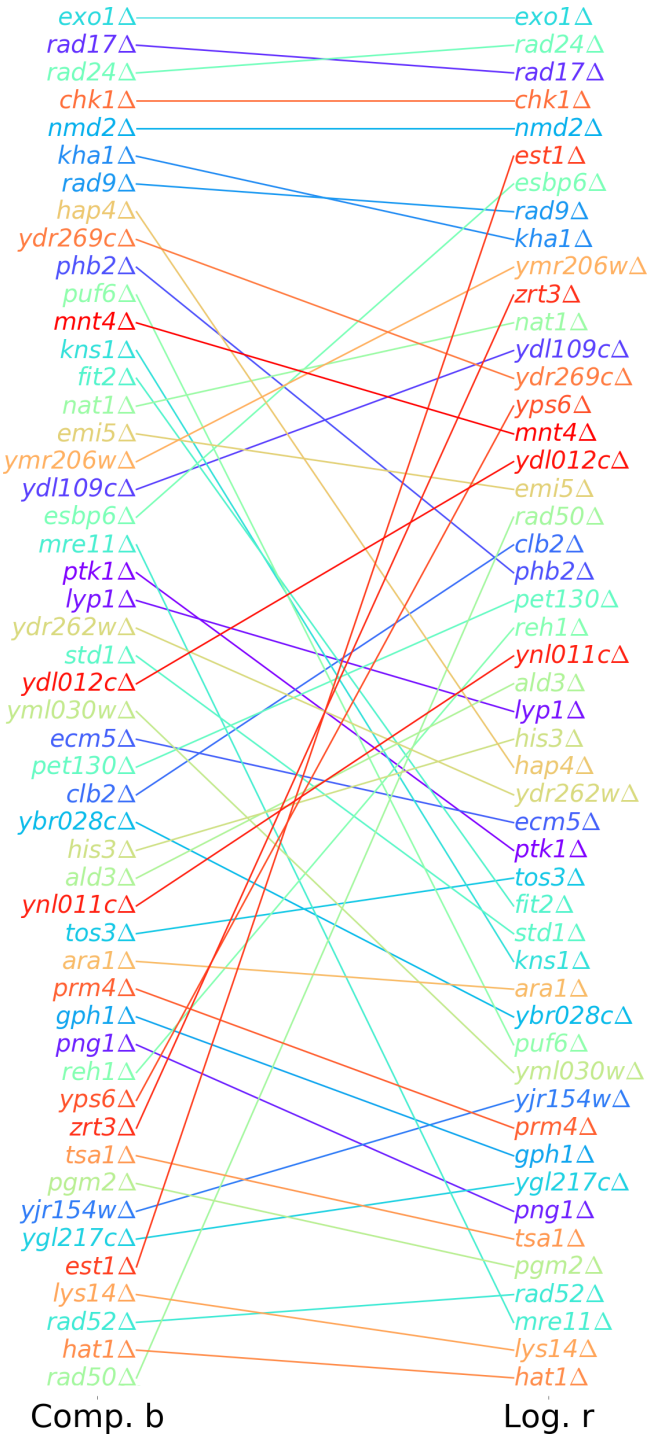

Figure 7: Comparison of r ranking for fits of the competition and logistic model to P15. *Fitnesses of genetic strains are ranked most to least fit from top to bottom. Competition model r was converted from b, N_0, and C_0 from the best competition model estimate. Logistic r and MDR were taken from logistic model fits using the QFA R package which makes heuristic checks for slow growing cultures.*

1.6 Comparison of Variation in Fitness Estimates

Use repeats on plate 15 (6 per deletion) to calculate coefficient of variation (COV) of estimated r or MDR.

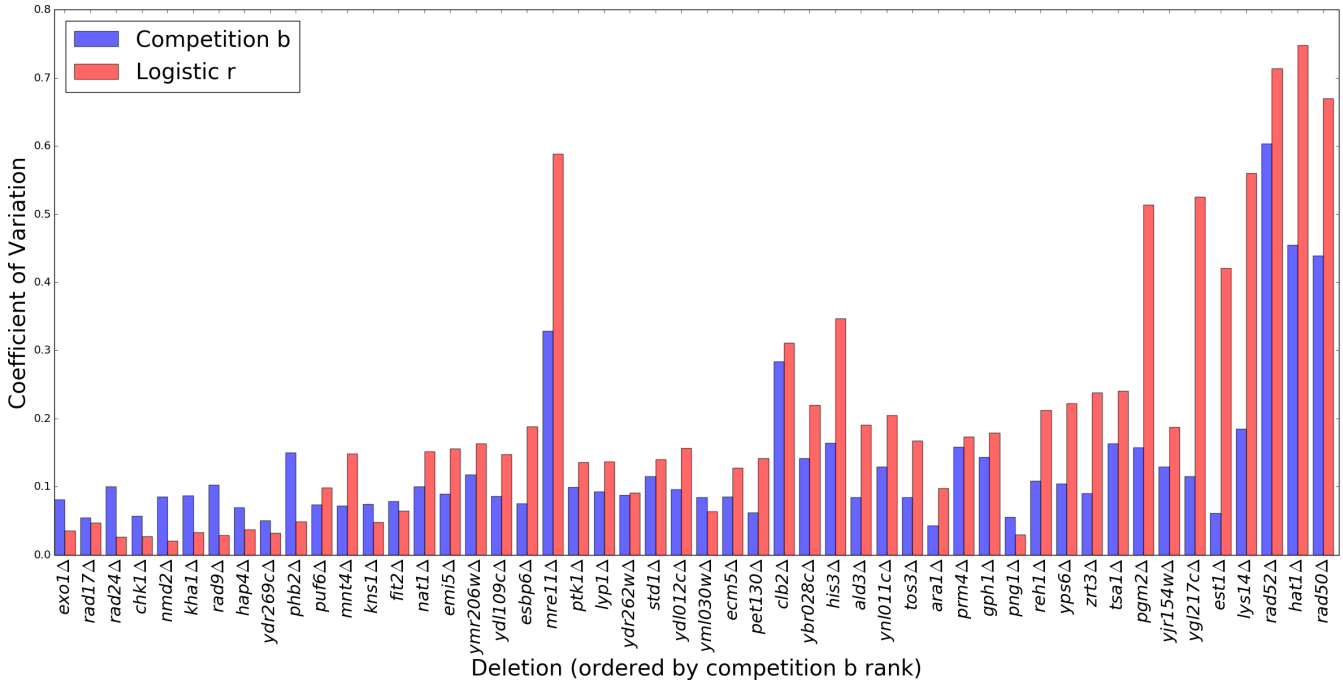

Figure 8: Coefficient of variation of r estimates. *Strains are ordered left to right along the horizontal axis by highest to lowest competition model r ranking. Fits are for the competition model, the QFA R logistic model, and the logistic equivalent model.*

1.7 Cross-plate validation

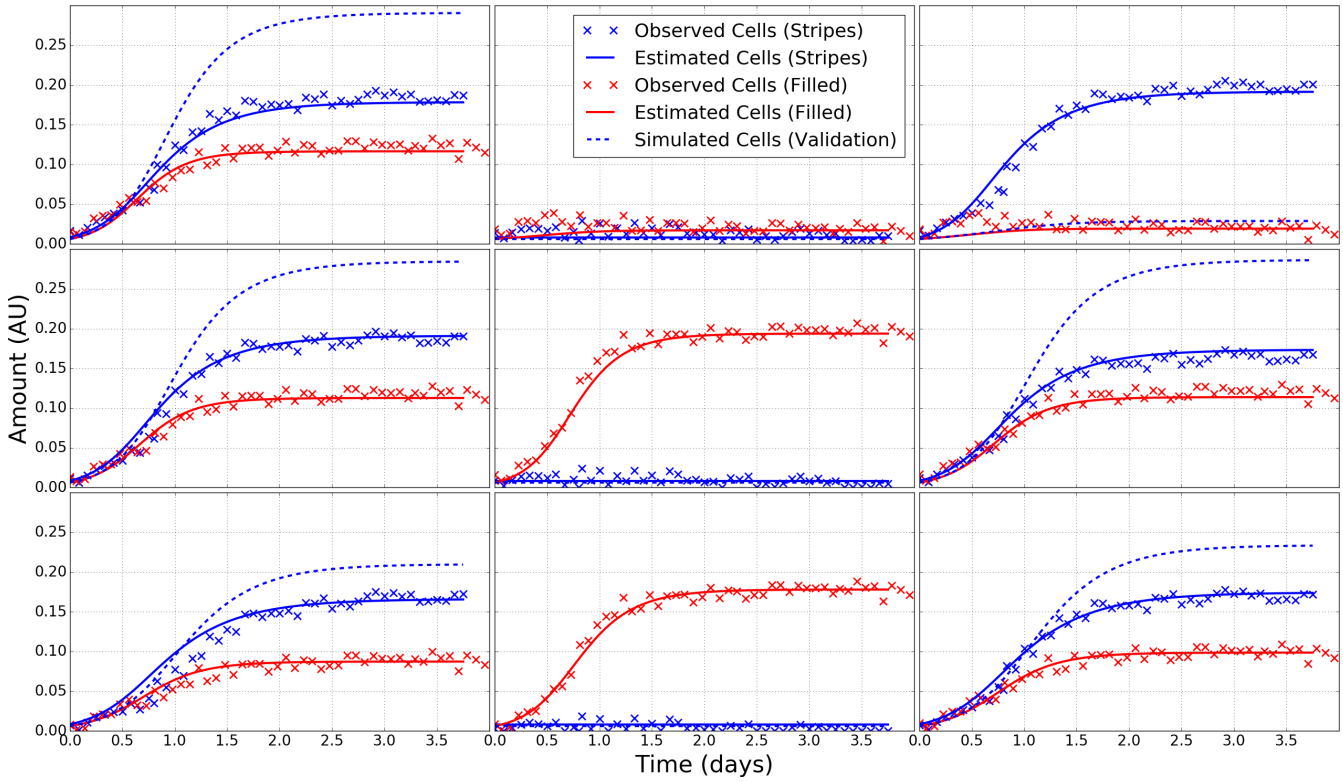

Figure 9: Calibration and validation of the competition model. *I fit the competition model to the 16x24 format "Stripes" and "Filled" plates in Figure* **??**. *The plot shows cell measurements and estimates for both plates for a 3x3 section with top left coordinates (R9, C10). I took the parameters estimates for the "Filled" plate (calibration) and set growth constant, b, to zero for cultures in the empty columns of the "Stripes" plate. I then simulated using these parameters to produce the dashed blue curve (validation). If the model is working correctly, the dashed blue curve should resemble the "Stripes" data (blue crosses).*

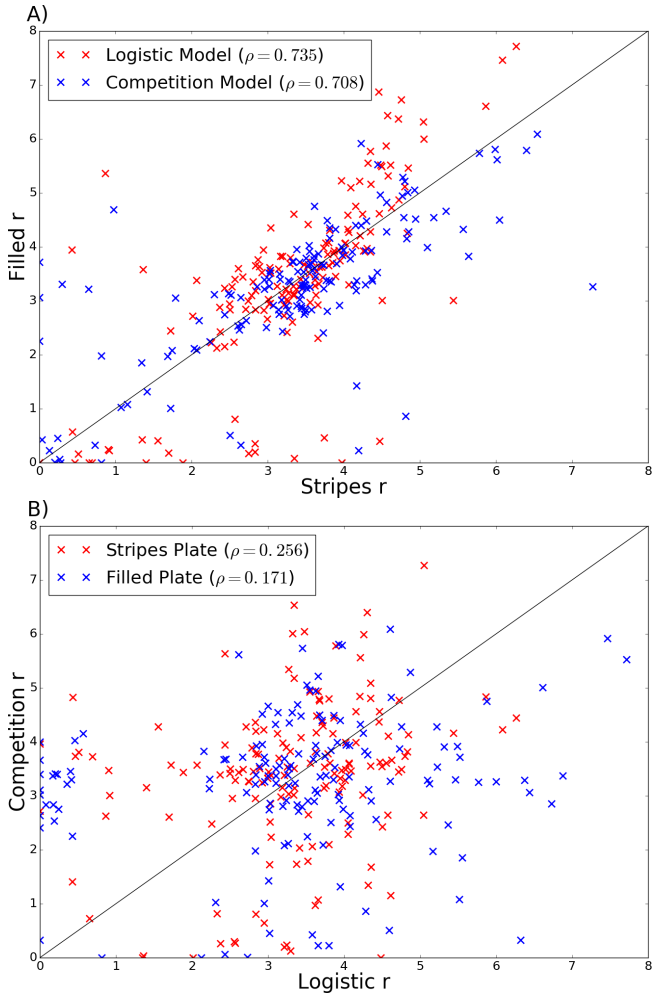

Figure 10: Correlation of r estimates for "Stripes" and "Filled" plates. *A) Correlation of r estimates between plates for logistic and competition models. B) Correlation of r estimates between logistic and competition models for both plates. I fit the competition model and independent model to the "Stripes" and "Filled" plates in Figure* **??**. *I converted competition model b to logistic model r. I only used data for cultures that were common between the two plates common and removed edge cultures. The Pearson correlation coefficient, ρ, is shown in the legends. The line $y = x$ is also plotted.*

1.8 Towards a genetic algorithm

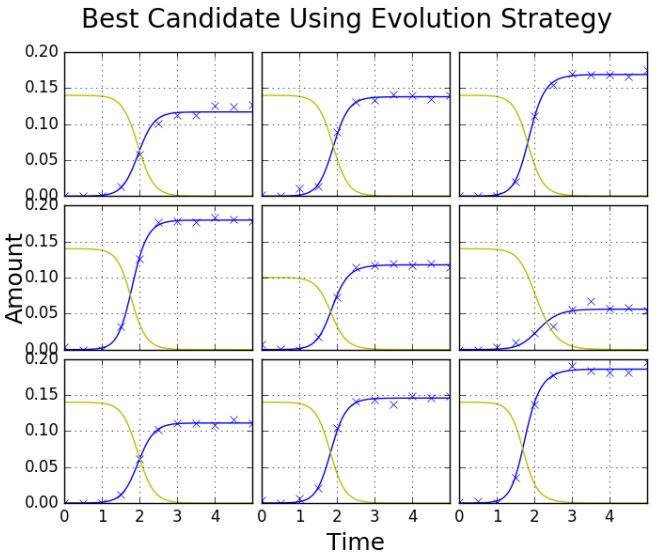

Figure 11: *Genetic algorithm fit to a 3x3 simulation. MIGHT TAKE A LITTLE BIT OF WORK TO REPRODUCE AND COULD USE PARAMETERS FROM THE BEST P15 FIT RATHER THAN JUST PICKING/RANDOMIZING. NEED TO CHECK THAT PLATE LEVEL PARAMETERS WERE ALSO EVOLVED.*

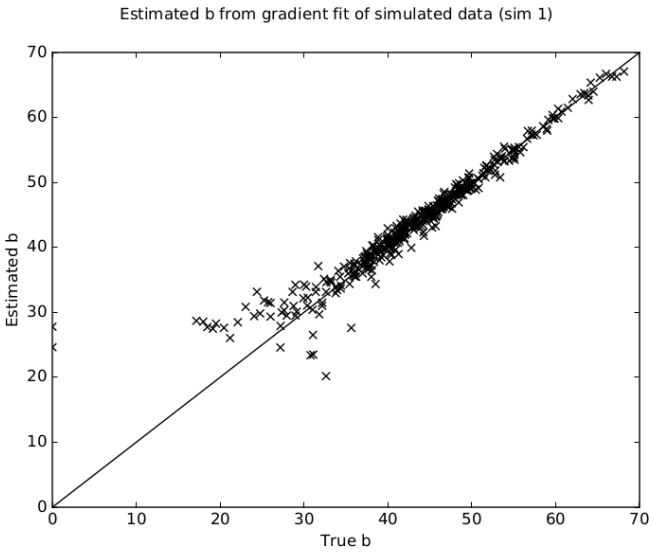

Figure 12: *Recovery of true b values from a gradient method with fixed plate level parameters. I simulated timecourses from the best five (which model? all BC?) fits to p15, fixed the true plate level parameters, and used a gradient method to recover b. This plot shows the worst case from the five sets of values.*

2 DISCUSSION

Fits of the competition model (see Figures 2 and 3) use less parameters and are qualitatively better than fits of either the logistic or generalised logistic model from the QFA R package (Addinall *et al.*, 2011; Lawless *et al.*, 2016). Competition model ranking of growth estimates for repeats on P15 (see Figure 6) agree with the logistic model rankings from Addinall *et al.* (2011) for the fastest and slowest growers (and with rankings from independent spot tests (refs) CHECK THIS).

However, there is much disagreement in the rank of other strains. (Could also do with the correlation plot for P15). The reliability of growth estimates was not improved using the competition model for the fastest growing strains on P15 (see Figure 8). This may be due to the effect of noise dominated cell observations from slow growing cultures on collectively fit parameters. This did not affect the the logistic model which fit to cultures individual. The greater reliability of estimates for slow growing cultures could be entirely due to collective fitting rather than to correcting for competition. (Could do with p-values on figure; also bold HIS3 everywhere) Unfortunately, the change in order for middle rankings is unlikely to identify new genetic interactions because significance is determined by comparison to a neutral deletion also in this range. Although improved, uncertainty in estimates for the slowest growers is still much higher than for the fastest meaning that the power to infer genetic reactions is not dramatically improved. (HIS3 has about half the variance for the competition model and this could be significant.)

Fitting the logistic model to slower growing cultures requires heuristic checks to correct for confounding between r and K. The QFA R implementation appears to have some issues. The strains *est1*Δ and *rad50*Δ have dramatic changes in ranking between logistic model r and MDR (see Figure 6). In independent spot tests (ref) these are very sick strains and I confirmed this in the raw QFA images by visual inspection. High r and low K have been erroneously fit to both strains. This is corrected for when converting to MDR which agrees with the competition model ranking and independent validation and is more similar to the fitness measure ($MDR \times MDP$) used in the original analysis by Addinall *et al.* (2011). For other cultures it appears that encroachment of fast growing cultures into neighbours is affecting cell density estimates made by Colonyzer (Lawless *et al.*, 2010). In logistic fits some growth curves are still in the exponential phase at the end of observations and this may be another fitting issue. If repeated, the plate from Addinall *et al.* (2011) should be run with a lower concentration of nutrients in the agar so that the stationary phase can be reached before cultures start to merge.

I looked at plate images from QFA and Colonyzer to investigate other discrepancies. *mre11*Δ is a weak growing strain (ref validation) which was misclassified as healthy by the competition model but not the logistic model. One repeat contained unusual heterogeneity, which may be natural or the results of contamination, and may explain the discrepancy. (Should take median value next time). *hap4*Δ appears to be much healthier than *zrt3*Δ which agrees with the competition model but not the logistic model. Although the precision of estimates is similar for both models, the competition model appears to be more accurate. Unfortunately, I lack independent data for validation of the middle strains.

Recent work Herrmann and Lawless suggests that direct measures of C_{t_0} may not be reliable due to heterogeneity between cells in the same inoculum; many cells do not grow and only the fastest growing cells contribute significantly to the final population. A plate level C_{t_0} also seems inappropriate but having extra parameters for the staring cell density of each culture is undesirable. Only a small amount of nutrients is used when cultures are small. Therefore, cultures could be grown for a short time before making direct cell density measurements that may be more accurate. QFA inocula use cells taken from the stationary phase where there might be more heterogeneity (ref). It may be possible to increase the reliability of fitness estimates by taking inocula from the exponential growth phase or using a higher starting density to average out effects.

The Stripes and Filled plates used a higher inoculum density and had very few noise-dominated cultures. Compared to P15, this would have reduced noise in collectively fit competition model estimates and would not have required heuristic checks to be employed for the logistic model. This may therefore be a fairer comparison than P15. Correlation of fitness estimates between plates in Figure **??**a was similar for both models. This is despite not finding global minima with the competition model. However, correlation between models for the same plate in Figure 10b is poor. (Could definitely do with P15 correlations to compare). There are issues with validation for both models (see Figure 9); the logistic model does not account for differences between plates at all and the competition model overcorrects. As I lack independent data for validation it is difficult to decide which to believe. (Unfortunately, these plates lack repeats so I could not study the reliability of estimates on the same plate. (I believe that we have more issues with accuracy than precision anyway). It would be informative to repeat P15 with C_{t_0} at a measurable level. In any case, it is clear that the competition model could be improved.

2.1 Future work

I was unable to find global minima using a gradient method (see Section **??**) to fit the competition model. I began work on a genetic algorithm method of solving but lacked time to complete this. I did however find that, with fixed plate level parameters, it is possible to reliably return b_i with a gradient method (see Figure 12. This offers the potential to use a hierarchical genetic algorithm where candidate plate level parameters are fixed in gradient fits of culture level parameters. Alternatively, a pure hierarchical genetic algorithm may work (i.e. where b_i are also evolved). A hierarchical Bayesian approach to fitting the competition model, similar to that of Heydari *et al.* (2016) for the logistic model, could also return global minima but might be slow. Current best fits, which are different local minima, have well correlated fitness rankings and make similar overcorrections for competition. This suggests that there is a more fundamental issue with the model.

It would be informative to validate the independent limit of the competition model to determine whether a mass action approximation is valid and whether it is correct to ignore the effect of metabolism on nutrient and final cell density. I suggest to validate first in liquid cultures, where the assumption of a well stirred mixture is more valid, and then attempt to validate for single cultures grown on agar, which more closely resembles QFA. Growth on a surface has a lower dimensionality and may be diffusion limited so a fractal kinetics model

may be required (Kopelman, 1988; Savageau, 1995). Nutrients (sugars, nitrogen, etc.) in QFA agars are of a standard composition, designed to reduce the excess of any single nutrient (check QFA paper and cite). It would be helpful to know and control the identity of the limiting nutrient using a different formula of agar. With nitrogen, rather than sugar, as the limiting nutrient, we are less likely to have to model metabolism.

Estimates of the nutrient diffusion constant k_n were fairly high such that nutrients diffused readily between neighbours and were nearly depleted when growth stopped. It may be that growth becomes limited by the diffusion of nutrients through agar before all nutrients are depleted and that nutrients are not well approximated as being evenly distributed withing the spatial scales that we model. Using a finer grid could reduce the overcorrection seen in Figure 9. Reo and Korolev (2014) use the diffusion equation (with Neumann and Dirichlet boundary conditions) to simulate nutrient dependent growth of a single bacterial culture on a pertri dish in two-dimensions. They create a sink for nutrients from culture growth and equate the flux of nutrients through culture area with the rate of increase in culture size. They model culture area as varying and keep culture density constant. This model could be adapted for QFA by keeping culture area constant and allowing culture density to vary. A mass action kinetic model of reaction (**??**) could be used for culture growth and the nutrient sink. It is computationally unfeasible to use such a detailed model to fit a whole plate but simulation could be very informative.

If we find that competition for nutrients is not responsible for the interaction between neighbours, for instance if growth becomes limited by diffusion of nutrients in the agar before nutrients from neighbours can be accessed, then we could instead model signalling by ethanol poisoning. This may be modelled similarly to how the competition model models nutrient diffusion and much of the code could be reused. (Quorum sensing via the molecule amonia could also be having an effect (ref)). If there is any combination of competition, metabolism, signalling, or arrest contributing significantly to differences in the growth of cultures and the interaction between neighbours then it will be difficult to separate them when fitting a model to data. We may have to develop ways to calibrate effects in isolation and use this information when fitting to high-throughput data. We only have observations for cells.

2.2 Improvements and recommendations

It is quicker to fit to small zones of a plate but, as these have a larger proportion of edge cultures, boundary conditions become important. Growing smaller arrays in isolation would help to speed up the development process. Edge cultures must be dealt included in fit the competition model and this may have contributed to significantly to error. When fitting the competition model noise might be better dealt with by leaving edge cultures empty. A different treatment of boundaries could also be used by modelling empty cultures outside edges rather than the approach in Section (**??**).

The design of the stripes validation experiment could be improved. Rather than filling gaps with cultures not present on the stripes plate, and for which we have no b estimates, we could fill with repeats of the cultures already present on the stripes plate. (I'm not sure it makes any difference actually whether we validate from one direction to the other). It would also have been helpful to have repeats to study differences in COV between the competition and independent models.

In order to make sure that competition effects were present in data, we made a dramatic change between the stripes and filled plates. We could have first validated the model against a smaller change, by varying between slower and faster growing cultures rather that none and very strong growing cultures. If the model works well between such plates it may work well for the majority of QFA experiments which typically have smaller differences between cultures than the data we study. If we did want to test the in an extreme case we could have inoculated fast growing cultures next certain strains and not others to try to induce a change in ranking for which the competition model might compensate better than the logistic model.

Each culture is surrounded by a different group of neighbours. The imaginary neighbour guess could be improved by using a range of b_f values to fit each culture and selecting b from the best fit. It would also be good to compare with guesses from the logistic model. I suspect that a gradient method will still fail to find a global minimum.

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.

Heydari, J. *et al.* (2016) Bayesian hierarchical modelling for inferring genetic interactions in yeast. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **65**, 3, 367–393.

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.