The unconstrained evolution of fast efficient and antibiotic resistant bacterial genomes

MS / tracking No. NATECOLEVOL–16060187A

Carlos Reding-Roman[1], Mark Hewlett[+1], Sarah Duxbury[+1], Fabio Gori[2], Ivana Gudelj[*1] & Robert Beardmore[*1]

[1]Biosciences, Geoffrey Pope, University of Exeter, Stocker Road, Exeter EX4 4QD, UK

[+]These authors contributed equally

email: r.e.beardmore@exeter.ac.uk

Evolutionary trajectories are constrained by tradeoffs when mutations that benefit one life history trait incur fitness costs in other traits. As resistance to tetracycline antibiotics by increased efflux can be associated with a 10%, or more, increase in length of the *Escherichia coli* chromosome, we sought costs of resistance associated with doxycycline. However, it was difficult to identify any because *E.coli's* growth rate (r), carrying capacity (K) and drug efflux rate *increased* during evolutionary experiments where *E.coli* was exposed to doxycycline. Moreover, these improvements remained following drug withdrawal. We sought mechanisms for this seemingly unconstrained adaptation particularly as these traits ought to tradeoff according to rK selection theory. Using prokaryote and eukaryote microbes, including clinical pathogens, we therefore show r and K can tradeoff, but need not, because of 'rK trade-ups'. r and K only tradeoff in sufficiently carbon-rich environments where growth is inefficient. We then used *E. coli* ribosomal RNA (*rrn*) knockouts to determine specific mutations, namely changes in *rrn* operon copy number, than can simultaneously maximise r and K. The optimal genome has fewer operons, and therefore fewer functional ribosomes, than the ancestral strain. It is, therefore, unsurprising for r-adaptation in the presence of a ribosome-inhibiting antibiotic, doxycycline, to also increase population size. Although *E. coli* can evolve to grow faster and to larger population sizes in the presence of antibiotics when compared to their absence, we found two costs to this improvement: an elongated lag phase and the loss of stress protection genes.

Introduction

Tradeoffs lie at the heart of a cross-kingdom research effort that seeks to explain how biodiversity is generated and maintained.[1–5] Two traits engage in an evolutionary tradeoff when beneficial mutations for one trait are deleterious for the other, and vice versa, and many theories agree[2,6–11] that genetic polymorphisms are maintained when tradeoffs have an appropriate geometry. Less clear, however, are the physical, chemical and physiological forces that create tradeoffs in the first place[12] and tradeoffs needed for the theories to work can be difficult to isolate in practise.[13–18]

It is essential for medicine that we understand tradeoffs. The term 'superbug' refers to a pathogenic microorganism that resists treatment by antibiotics with no apparent cost, or tradeoff, in terms of its pathogenicity. An evolutionary route to superbug status is thought to occur when a pathogen first adopts costly drug resistance mutations, a process that sees resistance traded against proliferation rate in antibiotic-free environments. Thereafter, other mutations compensate for those costs, yielding strains that are both drug resistant and capable of rapid proliferation.[19,20]

Tradeoffs are, however, sometimes observed in pathogens. A genomic study of a clinical pathogen using several antibiotic classes[21] showed resistance costs were dynamic during treatment for a *Salmonella* infection in which drug efflux proteins exhibited between-drug tradeoffs. During treatment, structural mutations that increased efflux rates for one antibiotic decreased them for others. Biophysical tradeoffs like this are important as they can help support genetic diversity, as was shown in a laboratory study where the rate-affinity tradeoff maintained a trimorphic transporter gene within a bacterial population.[10]

Single-protein tradeoff mechanisms, like the above, are more readily identified than organism-wide tradeoffs where costs may be hard to discern. Indeed, a classical concept, the rK tradeoff, where r is growth rate and K is organismal carrying capacity, suffers from an absence of both mechanism and data and so, along with rK selection theory, fell out of favour some time ago.[22] Others argue rK theory is relevant to understanding tumour progression and heterogeneity[23,24] but details of how rK selection theory makes predictive statements about tumour evolution are unclear. Interestingly, much the same can be said of the rate-yield tradeoff (RYTO) postulated

for microbes and cancers[25] that has also proven elusive.[14,26–28] Here 'rate' refers to growth rate and 'yield' to a metabolic conversion efficiency, c, between carbon source intake and biomass production. Some have observed a RYTO[10] while others have seen weak RYTOs, or no tradeoff at all, even positive rate-yield correlations have been observed[13,26,27] (i.e. a trade-up).

In this study of *E.coli* and fungal *Candida* we too see both rK, and rate-yield, tradeoffs and tradeups. However, we show this is only to be expected mechanistically because, for unicellular organisms, rK and rate-yield relationships derive from one underlying theory. By applying that theory, which is consistent with all our data, we resolve the geometry of the RYTO by demonstrating it has a parabolic shape that contains a tradeup. As a direct corollary we also resolve the geometry of rK tradeoffs for microorganisms, including *Candida*, by showing that it too has a parabolic shape where growth rate is maximal at intermediate population size.

The presence of rK tradeups is important for antibiotics which suppress r and K by design, which the antibiotic doxycycline achieves by slowing protein production in *Escherichia coli*. Antibiotic resistance mutations should restore r during chemotherapy but, here, the absence of rK constraints during resistance evolution allows the creation of a seeming *E. coli* mutant 'uberbug' that effluxes doxycycline more quickly, and which grows more rapidly to higher densities than its wild-type ancestor, whether or not the antibiotic is present. We were able to isolate one phenotypic and one genetic cost to this triple improvement: increased delay in the onset of exponential growth and the loss of stress protection genes contained in a prophage.

Results

We first present a mechanistic argument[10] supporting a parabolic relationship between growth rate in exponential phase of microbes and the population size those microbes achieve; recent extensions of flux balance analysis provide another.[29] As a result of the parabolic geometry of this trait pairing, rK relationships can contain both trade-ups and tradeoffs.

First consider the rK tradeup: it is straightforward to see why r and K might be positively correlated. Assume a carbohydrate, like glucose, is a limiting source of carbon, as can be the case for tumours and microbes. The size of a population of unicellular organisms can be predicted

from the carbohydrate concentration if the number of cells produced per carbon molecule, a.k.a. yield, c, is a constant:

$$K = \text{number of cells} = \overbrace{\frac{\text{number of cells}}{\text{moles of available glucose}}}^{\text{yield } (c)} \times \text{moles of available glucose} = c \times S; \qquad (1)$$

S is the concentration of that extracellular carbohydrate. This equation circumvents one problem of rK theory in which 'K *cannot be realistically expressed as a function of life history traits*'.[30] Equation (1) addresses this for microbes whereby the trait is yield, c, which is multiplied by the available nutrient to form K.[31] Equation (1) will not apply, however, to multicellular organisms where different cells in tissue develop with different nutrient to biomass conversion efficiencies.

To relate growth rate to the carbohydrate assume, for now, that r varies as a Monod function:[31]

$$r = r(S) = c \cdot \frac{V_{max} S}{k_m + S}. \qquad (2)$$

Here k_m is a half-saturation parameter, so-called because $r(k_m) = \frac{1}{2} \max\{r(S) : S \geq 0\}$ and V_{max} is the maximal uptake rate of carbohydrate into the cell. As $S = K/c$ in (1), using (2), we derive a putative rK relationship:

$$r = r(K/c) = c \cdot \frac{V_{max} \cdot K}{k_m \cdot c + K}. \qquad (3)$$

Equation (3) is consistent with microbial data[31] but, interestingly, inconsistent with classical rK theory: increasing r increases K. Assumptions typical of rK theory,[32,33] for example that a linear regression $r = a - bK$ can describe rK datasets, is not compatible with (3). To resolve this for microbes we first turn to a clinical pathogen, *Candida glabrata*.

RYTO and rK relationships in *C. glabrata*

By culturing (see Methods) *C. glabrata* in media with different concentrations of glucose as the carbon source, seeking to test equations (1-3) against data, we determined the dependence of r and K on glucose supply. Figure 1(a) summarises this dataset and the Methods section details how r and K are determined from microbial culture data. It shows a linear regression poorly captures the data (parameter-adjusted $R^2 \approx 0.13$), so does equation (3) (adjusted $R^2 \approx 0.46$) and a unimodal Poisson distribution function is a marginally better fit (adjusted $R^2 \approx 0.48$). One

might hypothesise, from Figure 1(a), that equation (3) and a linear regression apply to different subsets of the data: (3) applies at low K whereas data are approximately linear at high K. Much better would be one theory capturing all the data.

For this we summarise a calculation performed elsewhere[10,25,34] which removes the assumption in (2) that yield, c, is a constant independent of environment[10,34] by writing

$$K = \text{number of cells} = \frac{\text{cells}}{\text{ATP}} \times \overbrace{\frac{\text{ATP}}{\text{glucose}}}^{S-\text{dependent}} \times \text{glucose} = c(A) \times A(S) \times S. \tag{4}$$

We now need information about two measures of efficiency: the ATP produced per cell and the ATP per glucose in each cell. To allow progress, although not ideal, we now assume that ATP per cell does not depend on glucose availability. This allows yield, that we write $c(S)$, to depend only on glucose concentration, S. This is the simplest possible extension we could make to add realism to (2) and (3).

It has been shown,[10] to approximation, that

$$c(S) = c_{\text{hi}} \frac{1}{1+pS} + \frac{pS}{1+pS} c_{\text{lo}}, \tag{5}$$

where p is a shape parameter, c_{lo} is the lowest possible yield and c_{hi} is the highest possible yield over a non-toxic range of carbohydrate concentrations; Figure 1(c) validates (5) for *Candida* and *E.coli*. If we modify (2) and (3) to accommodate (5), we then obtain a curve in the rK-plane parameterised by S:

$$(r, K) = (r(S), K(S)) = \overbrace{\left(c_{\text{hi}} \frac{1}{1+pS} + \frac{pS}{1+pS} c_{\text{lo}} \right)}^{\text{this is } c(S) \text{ in equation (5)}} \cdot \overbrace{\left(\frac{V_{max} S}{k_m + S}, S \right)}^{\text{(uptake,sugar)}}, \tag{6}$$

where $K(S) = S \cdot c(S)$. Equation (6) can be written as a rate-yield relationship.[10,25–27,34] To see this, divide $K(S)$ by S to form $c(S)$:

$$(r, c) = (r(S), c(S)) = \left(c_{\text{hi}} \frac{1}{1+pS} + \frac{pS}{1+pS} c_{\text{lo}} \right) \cdot \left(\frac{V_{max} S}{k_m + S}, 1 \right). \tag{7}$$

Supplementary Figure S3 illustrates geometric tradeup and tradeoff forms described by (5-7). Note that we cannot fit (6) and (7) directly to rK and rate-yield data. Instead, to obtain those data, we first determine p, c_{lo} and c_{hi} from yields determined at different carbohydrate concentrations. We then determine k_m and V_{max} from r data measured at different carbohydrate concentrations.

Let us test this reasoning on *C. glabrata* rK data. Figure 1(b-d) show that (5) is an excellent descriptor of *C. glabrata* yield (adjusted $R^2 \approx \{0.995, 0.98, 0.99, 0.95, 0.98\}$ for five strains). We then eliminated parameters $V_{\max}$ and k_m in (6) using those data and the resulting datafits, and this is key, exhibit unimodal, parabola-like profiles (Figure 1, Supplementary Figures S5, S3 and S4) in which r exhibits its maximum at intermediate K. Thus, (6) reconciles the competing tradeoff and the Monod-based, tradeup, theories of rK data. Furthermore, the fit of (7) to *C. glabrata* rate-yield data is unimodal in rate and yield (Supplementary Figure S4). This is consistent with prior, untested predictions on the shape of the RYTO[25] that are corroborated by recent flux balance analyses.[29]

Figure 1 shows the dependence of r, yield and K on S for five *C. glabrata* strains. This shows K increases nonlinearly as glucose supply increases due to a decrease in efficiency at high, non-toxic glucose concentrations (relative likelihood of linear and a nonlinear regression $< 10^{-16}$), a property not observed by Monod.[31] As a result, growth rate can decrease as glucose supply increases (Figure 1(c)) and so growth rate can be maximised at intermediate glucose concentrations.

Given differences in their blood sugar levels, we sought, and found, evidence for different rK and rate-yield profiles in the fungal infections of diabetic and non-diabetic patients. Supplementary Figure S5(a) compares rK parabolae of *C. glabrata* strains 2001 and 3605, the latter isolated from a diabetic patient, which indicates a between-strain rK tradeoff: 2001 has lower r with potential for greater K than does 3605 (t-test for r: $t \approx -4.52, df = 250, p < 10^{-5}$, for K: $t \approx 7.72, df = 250, p < 10^{-12}$). Supplementary Figure S5(b) shows the mathematical rK model datafit for the *C. glabrata* strain taken from a diabetic patients has a skewed profile toward higher growth rates at lower population densities.

RYTO and rK relationships in *E.coli*

We then asked whether the parabolic rK shape was specific to eukaryotes. It is not, Figure 2 shows *Escherichia coli* has rK parabolae too. We use *E.coli* strains MG1655, wild-type, $\Delta 1, \Delta 2 - \Delta 5$ that have different numbers of ribosomal RNA operons (*rrn*) in their genomes; wild-type has 7, $\Delta 1$ has 6 operons, $\Delta 2$ has 5, and so on. Figures 1 and S5 show that claims made above for

Candida apply equally to these *E.coli* and Figure 2 shows the RYTO and rK shapes to be robust to changes in the number of *rrn* operons. These *E. coli* answer the following question: do genetic changes that increase yield, and therefore K, lead to a concomitant decrease in r? In other words, can we produce evidence of a between-strain rK tradeoff?

In response, ribosomal RNA is known to constitute a metabolic burden under carbon limitation.[35,36] By using these *E.coli* strains whereby this burden is under genetic control, we can mediate r and yield, thus K, to probe how selection for one of these traits affects the others. To this end, we establish the following property of the six *E.coli* strains that differ in the number of ribosomal RNA (*rrn*) operons. First, strains with fewer such operons have greater yields at the same carbon supply concentrations, and subsequently higher K (Figure 3(a), Supplementary Figure S8 shows p-values). According to equation (7) the number of *rrn* operons, as they mediate yield, should also mediate r, and indeed they do (Figures 3(a) and 4). This operon also mediates lag phase (Supplementary Figure S11). We thus have three inter-dependent traits, r, K and yield under genetic control, as desired, and lag phase too.

Figure 4 shows that different copy numbers of *rrn* optimise r at different glucose availabilities, although yield (therefore K) is optimised when fewest operons are present (Supplementary Figure S8). Figures 3(b) and S9 shows there is no between-genotype rate-yield tradeoff between these strains because changes in *rrn* copy number can follow a parabolic geometry exhibiting both positive and negative corrections with r and K and which occurs is contingent upon the glucose available. Supplementary Figure S7 summarises this information using three empirical and theoretical r, yield and K landscapes illustrating the dependence of these phenotypes on glucose and *rrn* operon number.

Data from the *E.coli* strain set indicate it may be possible to simultaneously improve r and K by mutational change, for example by deleting or duplicating an *rrn*, but is it? To answer this, we sought genetic and environmental conditions where population size is largest and Figure 5(a) provides a self-evident response: K is largest at greatest nutrient supply, as is obvious, and for the highest yield genomes, namely when *rrn* copy number is low. Now, Figures 5(a) and (b) show the strain with 3 and 4 *rrn* operons resides within a 2-strain cluster with maximal K but which also has the largest observed r (Supplementary Clustering Analysis). Finally, Figure 5(c)

shows that rate and yield, and therefore K, can be positively or negative correlated between genotypes, depending on the glucose background in which growth takes place. We conclude that improvements in K are not associated with a penalty in r in this strain set: deleting operons from the WT can improve both r and K.

E.coli rate-yield and rK changes following adaptation to an antibiotic

K, r and yield phenotypes are relevant to clinical infections because some antibiotics, ones said to be bacteriostatic, explicitly reduce r without lysing cells in a way that would reduce K. We therefore ask, do drug-resistance mutations that restore r during a bacteriostatic antibiotic challenge also reduce yield or K? Prior RYTO and rK theory would predict so. However, the potential for rK and rate-yield trade ups demonstrated above indicates that some resistance mutations might, instead, increase both r and K.

To study this, we propagated six replicate *E.coli* K12 (AG100) populations for 4 days of antibiotic treatment at clinical levels above the minimal inhibitory concentration (MIC) of doxycycline treating every 12h with drug (∼60 generations, MIC from 24h data in Supplementary Figure S12, see Methods). Figure 6 summarises the resulting r and K phenotypes. First, r in the presence of drug was restored to, and greater than, that of the ancestral strain in the absence of drug (Figure 6(b), $p < 0.012, df = 6, t \approx -3.5$). However, discordant with the hypothesis of a RYTO or rK tradeoff during adaptation, K also increased above that observed in both the ancestral strain and in the strain adapted to growth media containing no doxycycline (K nearly tripled: Figure 6(a), $p \approx 1.3 \times 10^{-7}, df = 6, t \approx -28.1$ for the 'Ancestral' strain, $p \approx 4.0 \times 10^{-7}, df = 6, t \approx -23.4$ for the 'Media Adapted' strain). Thus, although it was an antibiotic molecule for initial treatments, doxycycline eventually stimulated biomass production. Supplementary Figures S12 and S14 show doxycycline can also stimulate population growth within 48h at near MIC dosages.

Simultaneous rK improvements in AG100 were observed not only at super-MIC dosages, they also arose when the same adaption protocol was performed at sub-MIC dosages (see Methods). In this case, populations that had been inhibited by 70% due to the antibiotic subsequently grew to higher densities than populations cultured without antibiotic (Supplementary Figure S13).

Whole genome sequencing (WGS) of high-K populations indicated the appearance, and partial sweep, of duplications of a genomic region containing a doxycycline efflux pump operon,[37] *acr* (at frequencies 21, 25 and 42% in each of three replicates, Figures 7 and S15). It is on the basis of there being more cells with more efflux operons that we claim antibiotic resistance by efflux has increased in these populations, consistent with observations of increased r and K in the presence of doxycycline (Figures 6 and Supplementary Figure S13). WGS showed the excision of prophage *dlp12* from the AG100 genome at high frequency in high K populations (at frequencies 78, 84 and 88%). Polymorphisms common to all the drug-adapted populations observed at lower frequency were found in the insertion sequence (IS5) transposase, *insH-5*, and in the carbon starvation lipoprotein (*slp*) that mediates acid resistance (Supplementary Tables S1 and S2).

Although r, yield, K and drug resistance increased, and rK improvements remained following drug withdrawal (Figure 6), we were able to isolate a cost to this improvement: an increased duration of lag phase (Supplementary Figure S14). It is notable that although doxycycline is said to be bacteriostatic,[38] in fact here it stimulates biomass production while lengthening lag phase beyond 24h without slowing exponential growth (Supplementary Figure S14(b)). Thus, assays testing the antibiotic effect of doxycycline for 24h, or less, can conclude that it reduces growth rate when, instead, it merely increases the the waiting time to exponential growth.

Discussion

The logistic equation (8) is extremely useful when seeking growth rate and populations sizes of microbial growth data, but this does not mean rK selection theory provides any mechanistic insight into our data. Indeed, simply because the logistic model is a good fit of a growth dataset, one cannot deduce that r and K tradeoff, nor can one exploit that datafit to understand the nature of selection for r or K in that population. A range of theoretical and computational tools were required here to elucidate the selective forces in our experimental system. We needed mathematical models that causally relate growth efficiency, and therefore population size, to the rate of population growth and whole genome sequencing data highlighted the likely cause of that efficiency increase. Here this amounts to selection for the amplification of an efflux pump and the loss of a prophage from the genome.

rK theory is consistent with our datasets where r and K are negatively correlated, for example in Figure 1(a), although *any* prior claim that r and K were correlated would have been consistent with our data somewhere due to the humped, parabolic geometry we observe in rK and rate-yield data.

Ours is a laboratory study but there is a potential medical consequence of this geometry and the subsequent presence of rK tradeups: mutations that overcome chemotherapies designed to reduce growth rate can also increase population size. Here, just such adaptations created drug-resistant populations of bacteria that grow more quickly, and to higher densities, than bacteria not treated with antibiotics. These beneficial adaptations bear costs that are relevant to survival outside the laboratory. The doubling of lag time (Supplementary Figure S14, 'lag time' is parameter L in equation (9)) means the evolved strains can take twice as long to profit from increased nutrient uptake if the environment switches from a glucose-poor to a glucose-rich state. The loss of *dlp12* prophage removes protection from environmental stresses: *dlp12* is implicated in biofilm formation, an important aspect of the drug-resistance phenotype.[39] The observed *acr* amplifications can be strongly selected against in environments without antibiotic.[40] So, although a 3-dimensional phenotype of (rate, yield, resistance) shows positively correlated adaptation in lab conditions, the 5-dimensional phenotype (rate, yield, resistance, lag time, biofilm production) indicates those changes come with costs that will be exposed in other environments.

Finally, it is possible the humped rK geometry we observe applies to tumours and metazoa because equations (6) and (7) were derived from a theoretical model of a branched, glucose-processing pathway[10] common to all living cells. However, those equations are not applicable when different cells express those pathways differently or do not have pathways in common. Therefore, to generalise our findings to such heterogenous populations one could extend equations (6) and (7) to a situation where one microbial strain uses an extracellular, secondary metabolite of another microbe to grow. We hypothesise, in such a microbial community, r, yield and K will also exhibit positive correlations.

Methods

The supplement contains additional details on the methods used to analyse data. Please email the corresponding author to request any of the Matlab scripts used in this work.

Throughout, estimates of r and K are obtained from microbial growth data using the logistic equation

$$\frac{dN}{dt} = rN(1 - N/K), \qquad x(0) \geq 0, \tag{8}$$

whose solutions are modified to account for lag phase. This means r, K and L are determined from the best fit of the following 4-parameter model to microbial density timeseries:

$$N(t) = \beta + \frac{K}{1 + q \cdot e^{-r(t-L)}} \tag{9}$$

where $N(t)$ is density at time t. In this model, β is the estimated experimental blank, needed because microtitre plates used for culture absorb light, q is a composite parameter that contains the initial population size, $N(0)$, and L is lag, a proxy for the time taken for growth to enter exponential phase. Yield is calculated as the ratio between K and the glucose supplied to the growth medium (Supplementary Figures S1 and S2).

Please note that the biological parameter L is used throughout the text and supplementary and is always referred to by the term 'lag'. The parameter β is not a biological phenotype. It is an electromechanical parameter associated with the microtitre plates and plate-reading devices in which microbial population densities are measured. It refers to a constant measure of the light absorbance of the material from which those plates are constructed and the value of β derives from that material, in addition to the light absorbed by the liquid media placed in each well of the plate which allows microbial cultures to grow. This value must be properly subtracted from optical density readings to determine microbial population densities and the definition of $N(t)$ above allows us to do this in a reliable manner that accounts for potential errors in estimates of β. The parameter β might be termed the 'blank' but this term is used nowhere in the paper.

The parameters associated with equation (8) are estimated by fitting (8) to microbial growth data using `NonLinearModel.fit` in Matlab. This is a powerful nonlinear regression facility that produces a range of statistics on the robustness and suitability of that fit to the dataset in question.

Once estimates of r and K have been obtained, with associate uncertainties, because glucose has been supplied to the cells at a controlled concentration, S_0, the efficiency of biomass production is the value K/S_0.

Microbial biomass measure

Throughout our empirical work we use optical density (OD) as a proxy for *biomass*. Our reasons for doing so are 1) OD is known to correlate well with live cell counts at the densities we are using, see the Supplementary Information in these references[10,37] , 2) OD accounts for mass contained in non-viable cells which have previously sequestered carbon from the environment and therefore count as biomass, even if they no longer grow, 3) OD need not always correlate with live cell counts because cell size changes are accounted for within OD (larger cells produce a larger OD for equivalent population sizes), but this is consistent with the use of OD as a proxy for biomass, indeed, larger cells will have sequestered more carbon from the extracellular environment than smaller cells.

Other measures could be used to as a proxy for biomass, and therefore yield (a.k.a. *efficiency*: biomass produced per carbon supplied), for example i) colony forming units (CFUs, a.k.a. live cells counts), ii) total cell count performed by flow cytometry, iii) total DNA, iv) dry mass. We chose OD as it can be measured on a minute-by-minute basis for large numbers of replicates and culture conditions, allowing us to measure dynamics in the chosen biomass measure without disrupting or sampling the culture vessel, something that is not readily feasible for the other measures mentioned here.

C. glabrata rK data using 24h cultures

Overnight cultures were prepared in YPD medium (Yeast Peptone Dextrose with 20mg/ml glucose (2%w/v)) in 4ml volumes in universal tubes, via inoculation of a single colony per tube. Following 18-24 hours incubation at 30^oC and 180rpm, overnights were centrifuged and washed once in PBS solution prior to re-suspension in SC (Synthetic Complete) minimal medium at the appropriate glucose concentration. SC medium was prepared at 14 different glucose concentrations (0.25–

32 mg/ml) via autoclaving of media components (excluding glucose and 10% of the final media volume) prior to the addition of filter-sterilised glucose solution. D-Glucose was dissolved in distilled, de-ionised water to reach a final concentration of 320mg/ml (32%w/v) and sterilised through a $0.22\mu m$ filter unit. Glucose was diluted appropriately in autoclaved SC media water used to produce a final volume as required to prepare SC media with 4, 12, 20, 24 and 32mg/ml glucose concentrations. All other SC medias were prepared from a 200mg/ml (20% w/v) glucose stock solution. These experiments were repeated in triplicate.

E.coli rK data from 24h/48h cultures

$150\mu L$ of M9 minimal media was prepared using dilute K_2HPO_4 (350g), KH_2HPO_4 (100g) in 1L of deionised water and dilute trisodium citrate (29.4g), $(NH_4)_2SO_4$ (50g) and $MgSO_4$ (10.45g) in 1L of deionised water, autoclaved and diluted accordingly. Filtered, sterilised glucose and casamino acids were added from a 20% (w/v) stock. All strains were grown in M9 minimal media supplemented with 0.1% casamino acids and glucose ranging from 0 to 3mg/mL, incubated at 30^oC and shaken in a Tecan Infinite Pro for 24h with reads taken every 20min. Inoculating strains (AG100, MG1655 and *rrn* operon knockouts derived from MG1655) were prepared from overnight cultures in liquid LB medium were grown in M9 supplemented with 0.125mg/mL of glucose and centrifuged prior to inoculation. These experiments were replicated six times.

E.coli K12(AG100, MG1655 and *rrn* operon knockouts) antibiotic adaptation protocols

For extended doxycycline and no-doxycycline control treatments at 30°, a microtitre plate reader measured OD_{600nm} of cultured bacteria every 20 minutes in 96-well microtitre plates containing $150\mu L$ of liquid M9 medium supplemented with glucose (0.2%), casamino acids (0.1%) both with and without antibiotics. All cultures were shaken in a linear manner before each OD measurement. Inoculating *E.coli* were taken from a colony and cultured overnight in M9 (0.2%glucose, 0.1% casamino acids). At the end of either 12h or 24h, called the length of a 'season', a 96-pin replicator was used to transfer a 1% of volume sample (the volume transferred is approximately $1.5\mu L$) to a

new plate containing fresh growth medium and antibiotics, thus creating a new season of growth. The same environment was maintained for each subsequent transfer (8 seasons of treatment in total) and each replicate population (repeated in triplicate). The resulting OD time series were imported into Matlab R2013b to subtract the background (blank wells containing only medium) and generate all other statistics described in this document.

Based on the 24h data in Supplementary Figure S12, the minimal inhibitory concentration (MIC) of *E.coli* AG100 when exposed to doxycycline was taken as $0.6\mu g/ml$ for the purposes of this study. This is also called 1×MIC in the main text and 2×MIC is twice this concentration. When treatments of AG100 are said to be 'sub-MIC' in the main text, the doxycycline concentration was $0.2\mu g/ml$ (see Supplementary Figure S12) and this achieves a reduction in population size over a 24h period of AG100 growth that approximately equates to a 70% inhibition relative to the same *E.coli* strain cultured in the absence of doxycycline.

Sequencing and bionformatics protocols

Genetic mechanisms of adaption in antibiotic-evolved populations of *E. coli* (AG100)

To determine genetic mechanisms that might account for the K-stimulatory effect of the antibiotic doxycycline on AG100 reported in the main text, we sequenced the genomes of 3 evolved populations from the treatment which resulted in the highest cell number (measured by OD). These are indicated in Figure S13. To ensure the genetic changes were the result of antibiotic challenge, we also sequenced 3 control populations from that figure that evolved under experimental conditions but were not exposed to any antibiotics. We further sequenced 3 replicates of the ancestral population to ensure conformity between our starting strain and the published AG100 reference.

DNA Extraction

DNA was extracted from frozen samples ($150\mu L$ culture + $75\mu L$ 80% glycerol); a small quantity ($< 1\mu L$) was removed from the frozen population in a microtitre plate well using a sterile tip and inoculated into M9 media containing the same concentration of nutrients and antibiotic as

the adaptive condition as that microtitre well, therefore recapitulating the conditions of season 8 in Figure S13. DNA was extracted using an Ambion gene jet DNA extraction kit following the manufacturers instructions, with an additional ethanol wash step, and additional elution step to maximise DNA purity and yield. Samples were extracted from populations taken from season 8. DNA was also processed on a 1% agarose gel to ensure that DNA size was in excess of 10,000bp and to ensure that there was no protein contamination. DNA was accurately quantitated using the Qubit system, ensuring yields of over $25ng/\mu L$ in $50\mu L$. Paired-end DNA libraries were prepared by Exeter Sequencing Service, using the Nextera (Illumina) library preparation protocol.

DNA quality control and mapping

Reads were trimmed using FASTQC to ensure a mean phred score of at least 25 at all positions of the read. Adapter sequence trimming was performed by ESS. All sample reads were mapped to a previously published *E. coli* K12(AG100) reference genome available for download at the http site mentioned in the 'Annotation' section below.

Each sequence was indexed to the reference in fasta format using the BWA (Burrows-Wheeler alignment) index subroutine (BWA 0.7.4). Alignment was also performed using BWA and aligned files were subsequently sorted into genomic position and indexed using Samtools and filtered to remove any non-paired end matches. Coverage was analysed using genomeCoverageBed in Bedtools[41] and data were further analysed in Matlab.

Variant calling

Single Nucleotide Polymorphisms (SNPs) were called using VarScan 2.3.8. VarScan uses a heuristic method based on read depth, base quality and variant frequency in order to call SNPs. Filters were added to SNP calling to ensure that high quality calls were made. These quality filters were: minimum read depth of 60, with at least 10 reads supporting the alternative variant, a P-value of 0.05 or less and a minimum average quality of 30. Additionally VarScan calls SNPs as homologous if over 90% of the reads support one or the other variant. VarScan also detected short indels in the same manner. Pindel was used to detect structural variations, including inversions, rear-

rangements and indels. Pindel works by detecting breakpoints using a pattern-growth algorithm. This identities paired reads whereby only one of the reads is able to map to the reference. By breaking the unmapped read into shorter fragments, and re-mapping, the breakpoint of insertions or deletions can be detected. Duplication events were determined using CNVnator v0.3, with a bin size of 100, and MATLAB, using coverage data from Bedtools.

Annotation

Annotation files from the previously published AG100 genome were used in annotation of wild-type populations of AG100. These files were accessed from EBI:

`http://www.ebi.ac.uk/ena/data/view/PRJEB7832`

Annotation of the evolved genomes was also performed using RAST for initial web-based viewing, and PROKKA to annotate the consensus fasta files of all experimental replicates.

Clustering analysis

Given r and K data taken from replicated culture experiments using several genetically distinct strains of *E.coli*, we require a rationale to decide which strain optimises r and K. Rather than use multiple pairwise comparisons, instead we used an agglomerative clustering approach to categorise the whole phenotypic dataset, r and K, into some number of clusters that represent subsets of the entire dataset with similar phenotype. A natural prior expectation is that the data from a single genotype are found within one phenotypic cluster, although one such cluster may contain the data of two, or more genotypes. With this in mind, we used the following algorithm to determine the genotype that maximises the phenotypes r and K.

First choose a clustering algorithm that takes a dataset $\mathcal{D}$ and returns a categorisation of this dataset as i distinct clusters, call it $C(i, \mathcal{D})$. Now, given n genotypes and a phenotypic r dataset, $\mathcal{R}$, to categorise, for all $i = 2, ..., n$ form the output $\mathcal{C}_i := C(i, \mathcal{R})$. Let $R(\cdot)$ be the function that returns the set of phenotypic values for a given cluster.

We call a cluster *viable*, with parameter p, if a one-way anova is **unable** to detect significant differences, at level p, between all phenotypes in that cluster, this is the set $R(C(i, \mathcal{R}))$. The entire clustering is said to be *admissible* if each cluster is itself viable and, moreover, the phenotypes from any two distinct clusters **can be** detected by testing with a one-way anova at significance level p, thus $R(C(i, \mathcal{R}))$ and $R(C(j, \mathcal{R}))$ represent significantly different sets of phenotypic values for $i \neq j$.

Then, for each admissible clustering, i^*, choose the cluster label, $m(i^*) \in \{1, ..., i^*\}$, which contains the maximal phenotype:

$$\max \mathcal{R} \in R(\mathcal{C}_{m(i^*)}).$$

The optimal genotypes are said to be all those found within $\mathcal{C}_{m(i^*)}$.

The clustering method, C, used for Figure 5(b) is an agglomerative, 1-dimensional clustering analysis that was performed on the *rrn* operon mutants for both growth rate (r) and carrying capacity (K) datasets. To implement this we used the routines `clusterdata` and `anova1`, as implemented in Matlab 2013a, whereby $i^* = 3$ was determined to be the number of admissible clusters in each case when $p = 0.05$.

For the K dataset, the cluster of optimal genomes corresponding to $\{3, 4, 5\}$ (and these values are the numbers of *rrn* operons present in the genome) were deemed to be optimal according to the above algorithm, whereas $\{3, 4, 7\}$ was the analogous set determined for the r dataset. The intersection of these two sets is $\{3, 4\}$ which we therefore deem to be the optimal pair of strains (i.e. *rrn* strains $\Delta 3$ and $\Delta 4$) in the sense that these may both be said to maximise r and K across the entire dataset studied here.

Data availability

Data for figures that do not use genome sequence data is available in Excel spreadsheet format at `http://people.exeter.ac.uk/reb217/rebHomePage/data.html` (this link also contains coarse-grained DNA coverage data used in Figure 7). A complete whole-genome sequencing dataset can be accessed at `http://www.ebi.ac.uk/ena` using ENA project accession number of PRJEB15352.

Author contributions statement

Author contributions to this article are summarised as follows. Proposed research questions and hypotheses, subsequently designed experiments: RB, IG, MH, CRR; designed and wrote computer codes to analyse data: RB, CRR, FG, MH; performed experiments: CRR, MH, SD; wrote the manuscript: RB, CRR, IG.

References

[1] Herron, M. D. & Doebeli, M. Parallel evolutionary dynamics of adaptive diversification in *Escherichia coli*. *PLoS Biol* **11** (2013).

[2] Maharjan, R. *et al.* The form of a trade-off determines the response to competition. *Ecology Letters* 1267–1276 (2013). URL `http://dx.doi.org/10.1111/ele.12159`.

[3] Ben-Hur, E., Fragman-Sapir, O., Hadas, R., Singer, A. & Kadmon, R. Functional trade-offs increase species diversity in experimental plant communities. *Ecol Lett* **15**, 1276–1282 (2012).

[4] Kraaijeveld, A. R. & Godfray, H. C. Trade-off between parasitoid resistance and larval competitive ability in drosophila melanogaster. *Nature* **389**, 278–280 (1997).

[5] Fine, P. V. A. *et al.* The growth-defense trade-off and habitat specialization by plants in amazonian forests. *Ecology* **87**, S150–62 (2006).

[6] Dieckmann, U. & Doebeli, M. On the origin of species by sympatric speciation. *Nature* **400**, 354–357 (1999). URL `http://dx.doi.org/10.1038/22521`.

[7] Calsina, A. & Cuadrado, S. Small mutation rate and evolutionary stable strategies in infinite dimentional adaptive dynamics. *Mathematical Biology* **48**, 135–159 (2004).

[8] Metz, J. A. J., Geritz, S. A. H., Meszena, G., Jacobs, F. J. A. & Van Heerwaarden, J. S. Adaptive dynamics: a geometrical study of the consequences of nearly faithful reproduction. *in* Stochastic and spatial structures of dynamical systems (ed. van Strien, S. J. and Verduyn Lunel, S. M.), pp 183-231. North Holland, Amsterdam. In van Strien SJ & SM, V. L. (eds.)

Stochastic and spatial structures of dynamical systems, 183—231 (North Holland, Amsterdam, 1996).

[9] Gudelj, I., Coman, C. D. & Beardmore, R. E. Classifying the role of trade-offs in the evolutionary diversity of pathogens. *Proc. R. Soc. A* **426**, 97–116 (2006).

[10] Meyer, J. R., Gudelj, I. & Beardmore, R. Biophysical mechanisms that maintain biodiversity through trade-offs. *Nat. Commun.* **6**, 6278 (2015).

[11] Beardmore, R. E., Gudelj, I., Lipson, D. A. & Hurst, L. D. Metabolic trade-offs and the maintenance of the fittest and the flattest. *Nature* **472**, 342–346 (2011). URL `http://dx.doi.org/10.1038/nature09905`.

[12] Reznick, D., Nunney, L. & Tessier, A. Big houses, big cars, superfleas and the costs of reproduction. *Trends in Ecology and Evolution* **15**, 421 – 425 (2000). URL `http://www.sciencedirect.com/science/article/pii/S0169534700019418`.

[13] Leiby, N. & Marx, C. J. Metabolic erosion primarily through mutation accumulation, and not tradeoffs, drives limited evolution of substrate specificity in *Escherichia coli*. *PLoS Biol* **12**, e1001789 (2014). URL `http://dx.doi.org/10.1371\%2Fjournal.pbio.1001789`.

[14] Lipson, D. A. The complex relationship between microbial growth rate and yield and its implications for ecosystem processes. *Frontiers in Microbiology* **6**, 615 (2015). URL `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4468913/`.

[15] Mueller, L. D. & Ayala, F. J. Trade-off between r-selection and k-selection in drosophila populations. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 1303–1305 (1981).

[16] Sibly, R. M. & Hone, J. Population growth rate and its determinants: an overview. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **357**, 1153–1170 (2002).

[17] Luckinbill, L. S. r and K selection in experimental populations of *Escherichia coli*. *Science* **202**, 1201–1203 (1978).

[18] Luckinbill, L. S. Selection and the r/K continuum in experimental populations of protozoa. *Am. Nat.* 427–437 (1979).

[19] Nilsson, A. I. *et al.* Reducing the fitness cost of antibiotic resistance by amplification of initiator trna genes. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 6976–6981 (2006).

[20] Millan, A. S. *et al.* Positive selection and compensatory adaptation interact to stabilize non-transmissible plasmids. *Nat Commun* **5** (2014). URL `http://dx.doi.org/10.1038/ncomms6208`.

[21] Blair, J. M. A. *et al.* Acrb drug-binding pocket substitution confers clinically relevant resistance and altered substrate specificity. *Proceedings of the National Academy of Sciences* **112**, 3511–3516 (2015). `http://www.pnas.org/content/112/11/3511.full.pdf`.

[22] Stearns, S. *The Evolution of Life Histories* (Oxford University Press, 1992).

[23] Aktipis, C. A., Boddy, A. M., Gatenby, R. A., Brown, J. S. & Maley, C. C. Life history trade-offs in cancer evolution. *Nat. Rev. Cancer* **13**, 883–892 (2013).

[24] Korolev, K. S., Xavier, J. B. & Gore, J. Turning ecology and evolution against cancer. *Nat. Rev. Cancer* **14**, 371–380 (2014).

[25] Pfeiffer, T., Schuster, S. & Bonhoeffer, S. Cooperation and competition in the evolution of ATP-producing pathways. *Science* **292**, 504–507 (2001).

[26] Novak, M., Pfeiffer, T., Lenski, R. E., Sauer, U. & Bonhoeffer, S. Experimental tests for an evolutionary trade-off between growth rate and yield in E. coli. *Am Nat* **168**, 242–251 (2006).

[27] Wong, W. W., Tran, L. M. & Liao, J. C. A hidden square-root boundary between growth rate and biomass yield. *Biotechnol Bioeng* **102**, 73–80 (2009).

[28] Fitzsimmons, J., Schoustra, S., Kerr, J. & Kassen, R. Population consequences of mutational events: effects of antibiotic resistance on the r/k trade-off. *Evolutionary Ecology* **24**, 227–236 (2010). URL `http://dx.doi.org/10.1007/s10682-009-9302-8`.

[29] Mori, M., Hwa, T., Martin, O. C., De Martino, A. & Marinari, E. Constrained allocation flux balance analysis. *PLoS Comput Biol* **12**, e1004913 (2016).

[30] Stearns, S. C. The evolution of life history traits: A critique of the theory and a review of the data. *Annual Review of Ecology and Systematics* **8**, 145–171 (1977).

[31] Monod, J. The growth of bacterial cultures. *Annu. Rev. Microbiol.* **3**, 371–394 (1949).

[32] Roughgarden, J. Density-dependent natural selection. *Ecology* 453–468 (1971).

[33] Boyce, M. S. Restitution of r-and k-selection as a model of density-dependent natural selection. *Annu. Rev. Ecol. Syst.* 427–447 (1984).

[34] MacLean, R. C. & Gudelj, I. Resource competition and social conflict in experimental populations of yeast. *Nature* **441**, 498–501 (2006).

[35] Koch, A. R. & Deppe, C. S. *In vivo* assay of protein synthesizing capacity of *escherichia coli* from slowly growing chemostat cultures. *J. Mol. Biol.* **55**, 549–562 (1971).

[36] Stevenson, B. S. & Schmidt, T. M. Growth rate-dependent accumulation of rna from plasmid-borne rrna operons in escherichia coli. *J. Bacteriol.* **180**, 1970–1972 (1998).

[37] Fuentes-Hernandez, A. *et al.* Using a sequential regimen to eliminate bacteria at sublethal antibiotic dosages. *PLoS Biol* **13**, e1002104 (2015). URL `http://dx.doi.org/10.1371\%2Fjournal.pbio.1002104`.

[38] Chopra, I. & Roberts, M. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol Mol Biol Rev* **65**, 232–260 (2001).

[39] Wang, X. *et al.* Cryptic prophages help bacteria cope with adverse environments. *Nat Commun* **1**, 147 (2010). URL `http://dx.doi.org/10.1038/ncomms1146`.

[40] Laehnemann, D. *et al.* Genomics of rapid adaptation to antibiotics: convergent evolution and scalable sequence amplification. *Genome Biol Evol* **6**, 1287–1301 (2014).

Competing Interests

The authors declare that they have no competing financial interests.

Correspondence

Queries should be addressed to REB (r.e.beardmore@exeter.ac.uk)

Acknowledgements

The ribosomal *rrn* knockout strains derived from *E. coli* MG1655 were gifted by Tobias Bollenbach, strain AG100 was provided by Stuart Levy and *Candida* strains were a gift from Steve Bates who are sincerely thanked for their help.

Figures

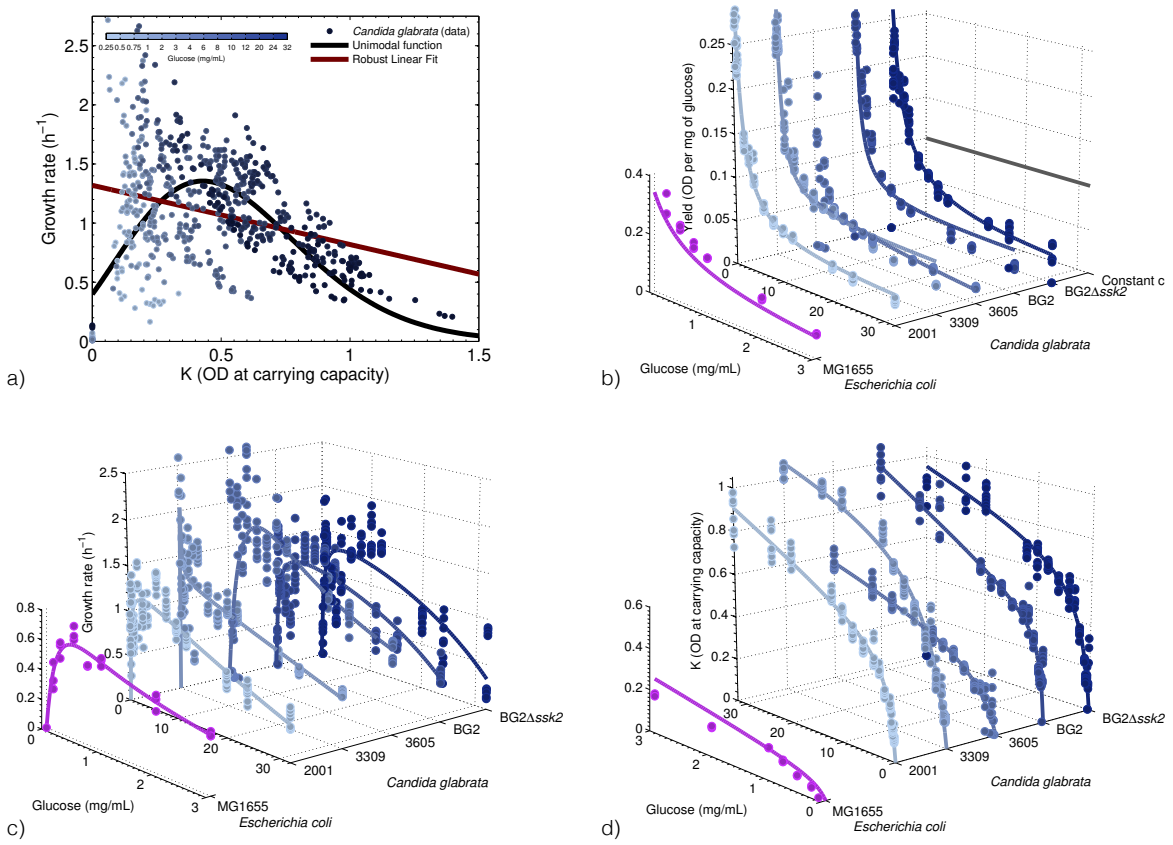


Figure 1: a) Collated rK datasets are best explained by unimodal datafits; b-d) K, r and yield datasets in fungi and bacteria are explained by equations (2-7). a) Collated growth rates (r) from different $C.\ glabrata$ strains on the y-axis and population size (K) on the x-axis; each datapoint has a corresponding glucose supply concentration. A unimodal regression based on a scaled Poisson distribution function ($r(K) = p \cdot \exp(n \cdot \ln K - K - \ln \Gamma(n+1))$ where p and n are shape parameters, Γ is the gamma function) and a robust linear regression are shown. Based on the corrected Akaike Information Coefficient, AICc, the relative likelihood of the two regressions is $< 10^{-15}$: the unimodal fit is significantly more likely. b) Biomass per mg of glucose supplied (i.e. yield) on the z-axis for $C.\ glabrata$ and $Escherichia\ coli$, with glucose supply on the x-axis and strain labels on the y-axis. Datapoints are dots whereas predictions from equation (5) are lines (adjusted $R^2 \approx \{0.995, 0.98, 0.99, 0.95, 0.98\}$ for the Candida strains). Analogously, c) and d) show r and K on the z-axis, respectively, indicating their dependence on the supply of glucose (x-axis). Model predictions (solid lines) are obtained by fitting equations (2-7) to data using `NonLinearModel.fit` in Matlab.

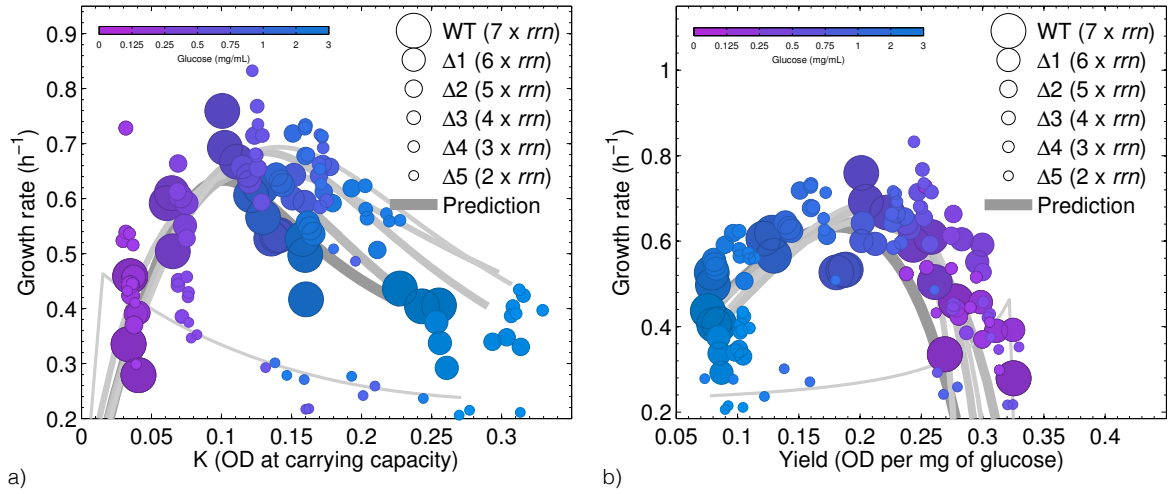


Figure 2: Empirical and theoretic rK relationships (in a) and r-yield relationships (in b); dots are data, sold lines are theoretical model fits. Consistent with theory, these are skewed parabolae for all the strains of $E.coli$ K12(MG1655) wild-type and rrn knockout strains used in this study, bar the one with fewest rrn operons. The strain with only 2 such operons (smallest dots shown) does not conform to this geometry and, unlike the other strains, has an rK tradeoff with no trade-up region.

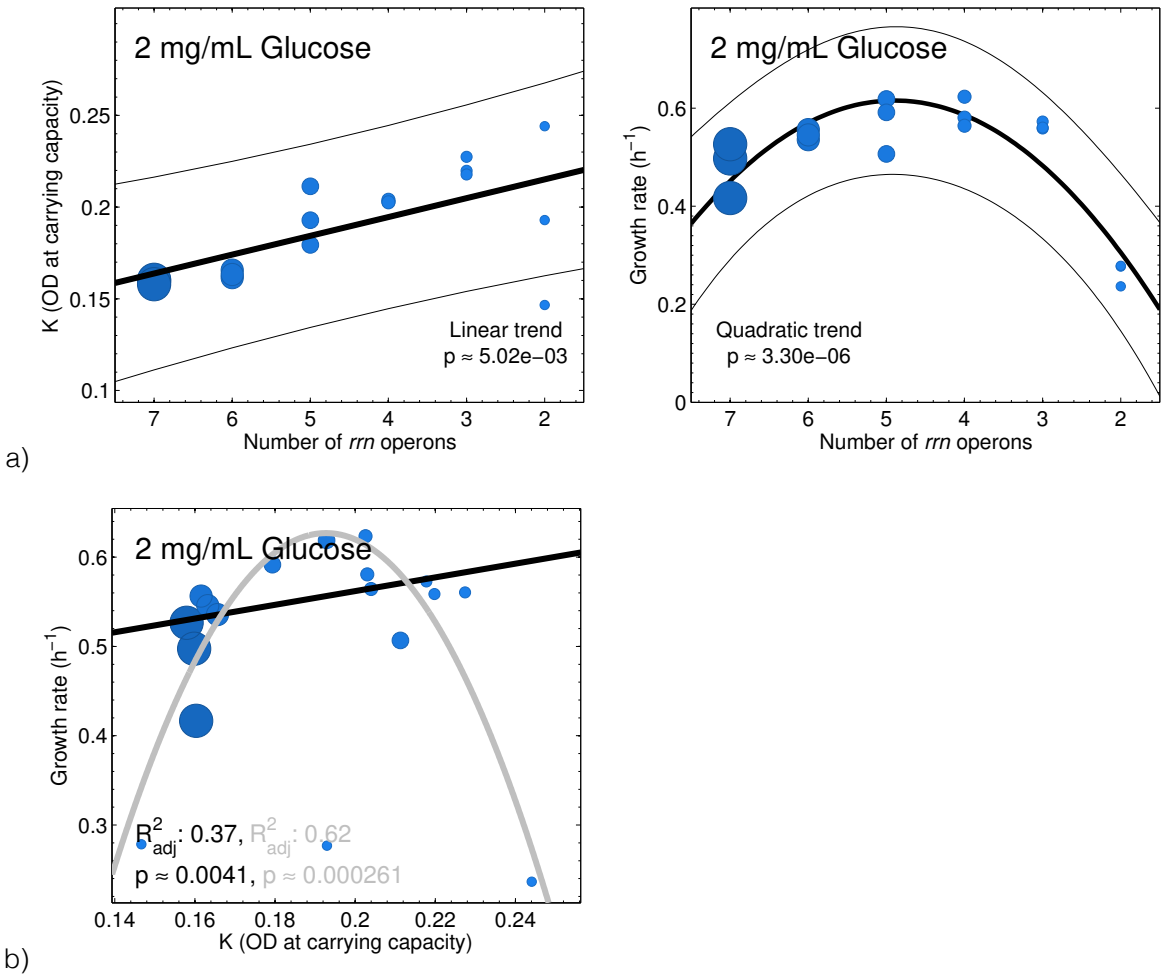


Figure 3: A demonstration that rrn operon copy number controls r and K by mediating yield: to turn K into yield, divide by 2mg/mL glucose. a) The relationship between the number of ribosomal RNA (rrn) operons, carrying capacity (K, left) and growth rate (r, right) for one glucose concentration (2mg/mL; Figures 4 and S8 contain more). Linear and quadratic regressions are shown, respectively, as lines $\pm$ estimated 95% CI (3 replicates). These data indicate greater K results from fewer operons for fixed glucose supply and that r can be maximised at intermediate operon number. b) A scatterplot of collated (r,K) values for all rrn knockout strains for one glucose concentration showing a quadratic relationship is a better descriptor of data than a linear regrewsion; note, this also serves as a between-genotype rate-yield plot. The size of the dot for each datum corresponds to rrn operon number, as per a).

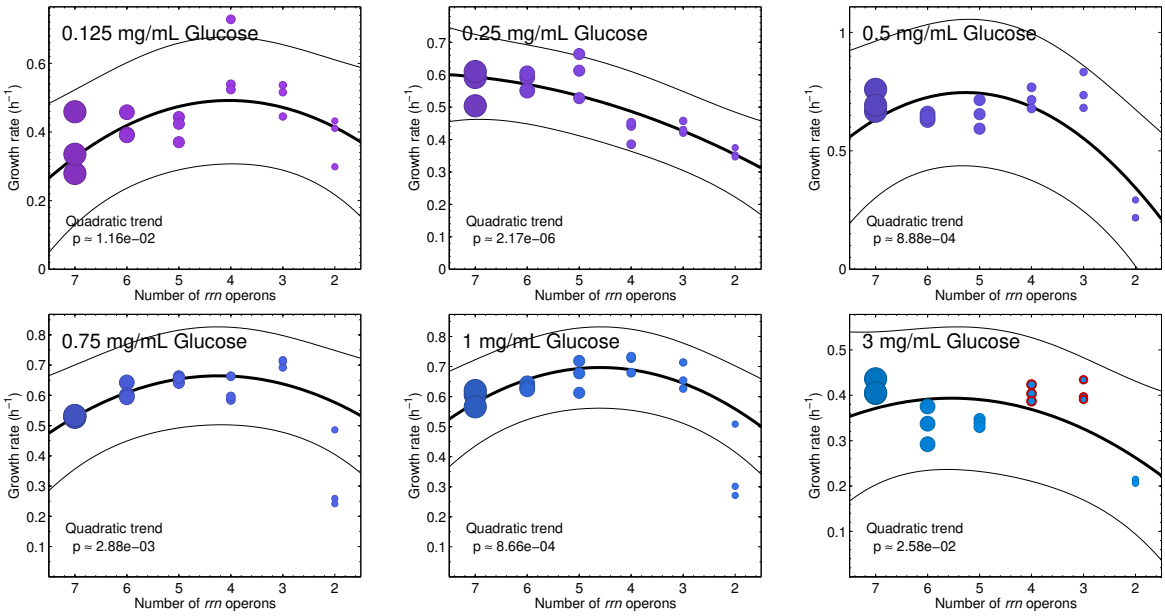


Figure 4: Determining the rrn operon copy number that maximises growth rate of the $E.\ coli$ rrn mutants. Growth rate (r) is shown on the y-axis as a function of ribosomal RNA operon copy number (x-axis) for different glucose supply concentrations. We sought a relationship between growth rate and operon copy number using linear and quadratic regressions and plotted the regression with the better fit to data according to adjusted R^2 value (solid black line, outer lines are the prediction $\pm$ 95% CI). These data indicates that intermediate rrn operon numbers can optimise r under some environmental conditions, although higher numbers of operons (the wild type $E.coli$ strain has 7) can also optimise r under other conditions. For completeness, using the quadratic regression we computed 95% confidence intervals for r-optimal rrn copy numbers for $\{0.5, 0.75, 1, 2, 3\}$mg/mL glucose, these are, respectively, $\{5.26 \pm 0.47, 4.22 \pm 0.24, 4.60 \pm 0.02, 4.89 \pm 0.11, 5.53 \pm 1.21\}$. This shows that 7 operons is not optimal for growth under all environmental conditions, although it optimal for some (0.25 and 3 mg/mL glucose). Red dots in the bottom-right panel indicate populations that optimise K from all glucose concentrations tested.

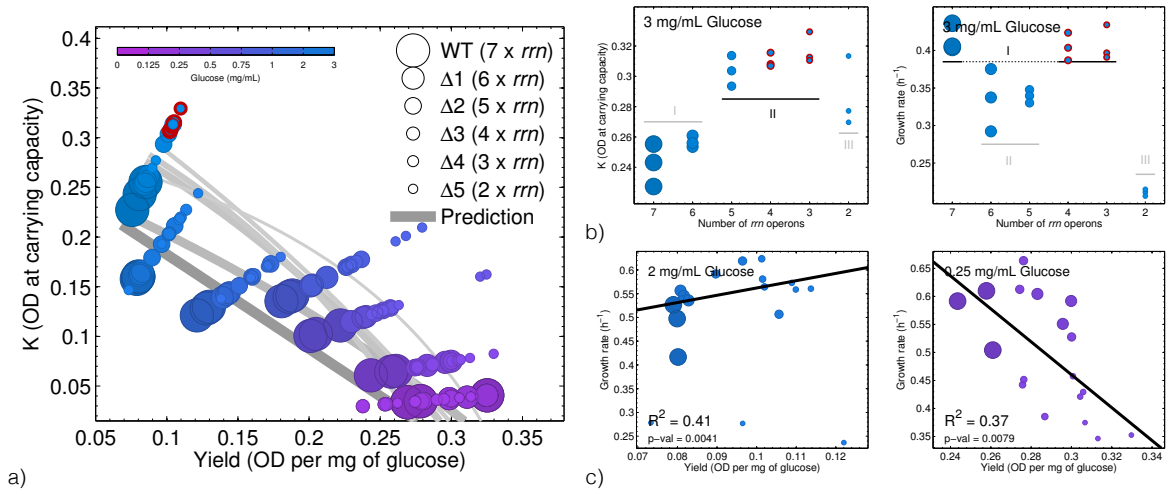


Figure 5: *E.coli* **K12(MG1655)** rrn **knockout strains** $\Delta 3$ **and** $\Delta 4$ **optimise population size, K, when glucose is high in concentration without paying a growth rate cost.** a) Empirical data (dots) and theoretical predictions (lines using models in main text) for each rrn knockout strain show the variation in K (y-axis) as a function of yield (x-axis) and ribosomal RNA (rrn) operon copy number (dot size). The cluster of combinations of rrn operon number and glucose supply concentration that maximise K are highlighted as red dots: self-evidently, K is greatest when glucose supply is greatest and growth is efficient, meaning few rrn operons (see Supplementary Figure S8). As K / yield = glucose concentration, when glucose concentration is fixed, the collated data for all strains lie on straight lines that pass through (yield,K) = (0,0). The highest yield outcome is marked as a large red circle: the WT strain cultured at the lowest glucose concentration. b) An agglomerative, 1-dimensional clustering analysis (Supplementary Methods) shows that the highest population sizes found at the highest glucose concentration tested (K-cluster II, shown left) are also achieved at the highest observed growth rates (r-cluster I, shown right). As a result, operon copy numbers that simultaneously optimise r and K are highlighted with red circles (3 and 4 copies of rrn). c) There is no between-strain tradeoff in r and K: dot sizes indicate different rrn copy number (as Figure 2) with each datum each representing observed r and K at high (left) and low (left) glucose concentration. Linear regressions (black lines) indicate that r and K can be both positively and negatively correlated, depending on the environment (Supplementary Figure S9 shows more glucose concentrations).

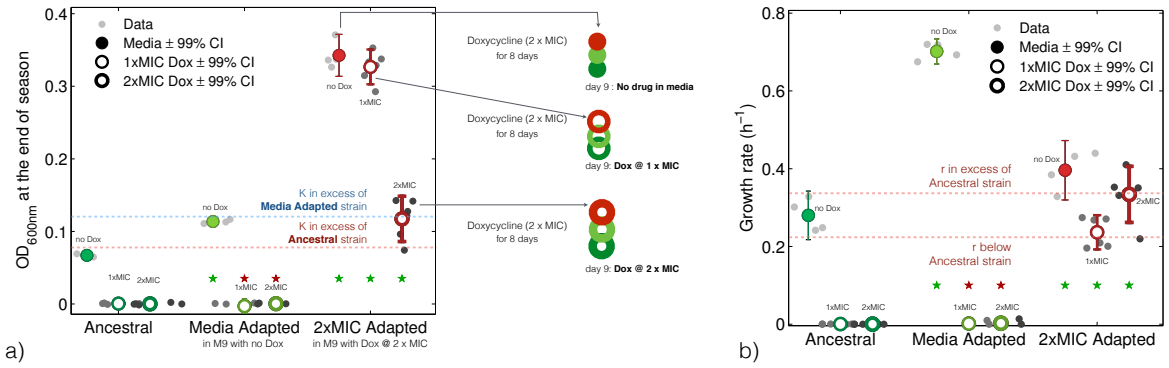


Figure 6: Simultaneous rK-adaptation from an antibiotic challenge where drug resistance also increased. K (in a, units of OD_{600nm}) and r (in b) for three strains of *E. coli* K12(AG100) labelled 'Ancestral', 'Media Adapted' and '2×MIC' Adapted (Supplementary Methods describe protocols). The latter two strains were derived from Ancestral by ∼60 generations adaption to media containing either no antibiotic (then labelled Media Adapted) or else in two times the minimum inhibitory concentration of doxycycline (then labelled 2xMIC Adapted); this is a clinical dose. All three strains were then cultured for one day (the 'day 9' data indicated) in media containing either no antibiotic (filled circle), a dose of doxycycline equivalent to the MIC (1×MIC, thin open circle) or twice the MIC (thick open circle). We tested whether differences in K (in a) and r (in b) were significant (green star) or not (red star) with respect to the Ancestral strain under identical day 9 conditions (t-test, $p < 0.05$ in each case). (a) Unexpectedly, the greatest K was observed when '2×MIC Adapted' was cultured either in the absence of doxycycline or with doxycycline present at 1×MIC. (b) Unsurprisingly, the greatest r was observed when Media Adapted was cultured in media with no doxycycline; '2×MIC Adapted' had an r phenotype in the presence of doxycycline at a 2×MIC concentration that could not be distinguished from Ancestral cultured in the absence of doxycycline; when doxycline was withdrawn, the r of strain '2×MIC Adapted' was larger than Ancestral. Thus, drug-resistance adaptation has restored r above ancestral values while also increasing K.

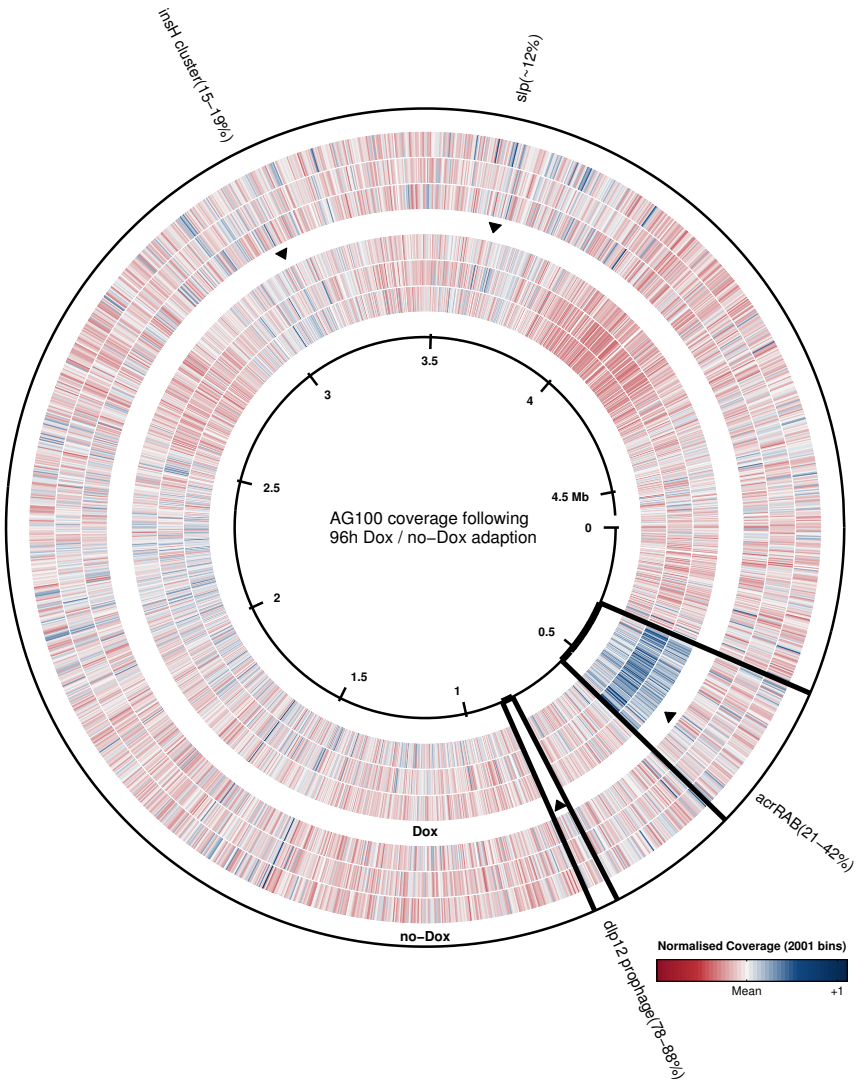


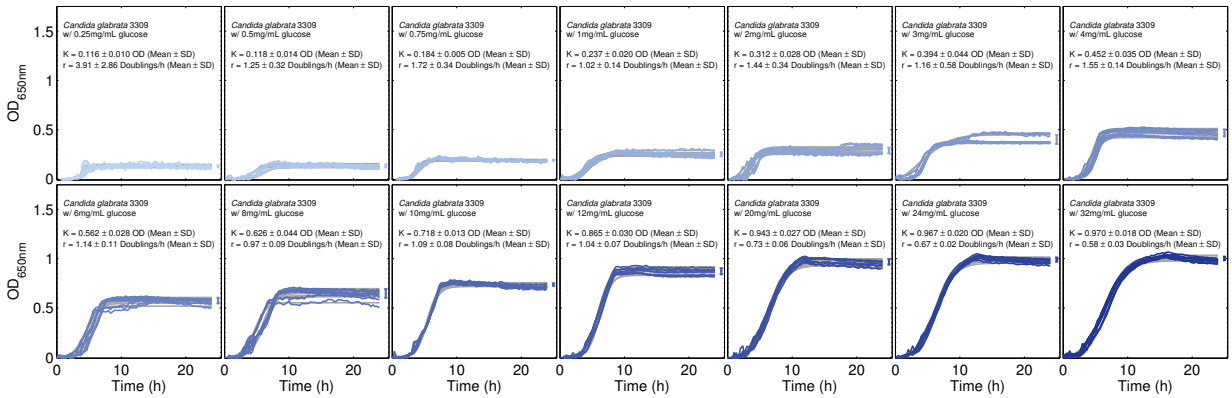
Figure 7: A DNA coverage plot for *E. coli* K12(AG100) following ∼60 generations (96h) of growth in the presence and absence of doxycycline. Data in the presence (3 inner annuli, 'Dox') and absence (3 outer annuli, 'no-Dox') indicate potential genetic mechanisms supporting positive r, K and resistance adaptation. A change in DNA detected in the sequencing protocol is shown in red (reduction) and blue (increase; and white is complete loss), respectively, with respect to no mean change (in grey). The inner ring (black) indicates genome position and three replicates for each treatment were sequenced (indicated in Figure S13). The white region marked *dlp12* shows the deletion of prophage dlp12 only during doxycycline treatment. The outer text labels summarise the estimated frequencies of mutations common to all three replicate populations, indicating positive selection for the doxycycline efflux operon, *acr*, with the prophage deletion at even higher frequency.

The unconstrained evolution of fast efficient and antibiotic resistant bacterial genomes: supplementary

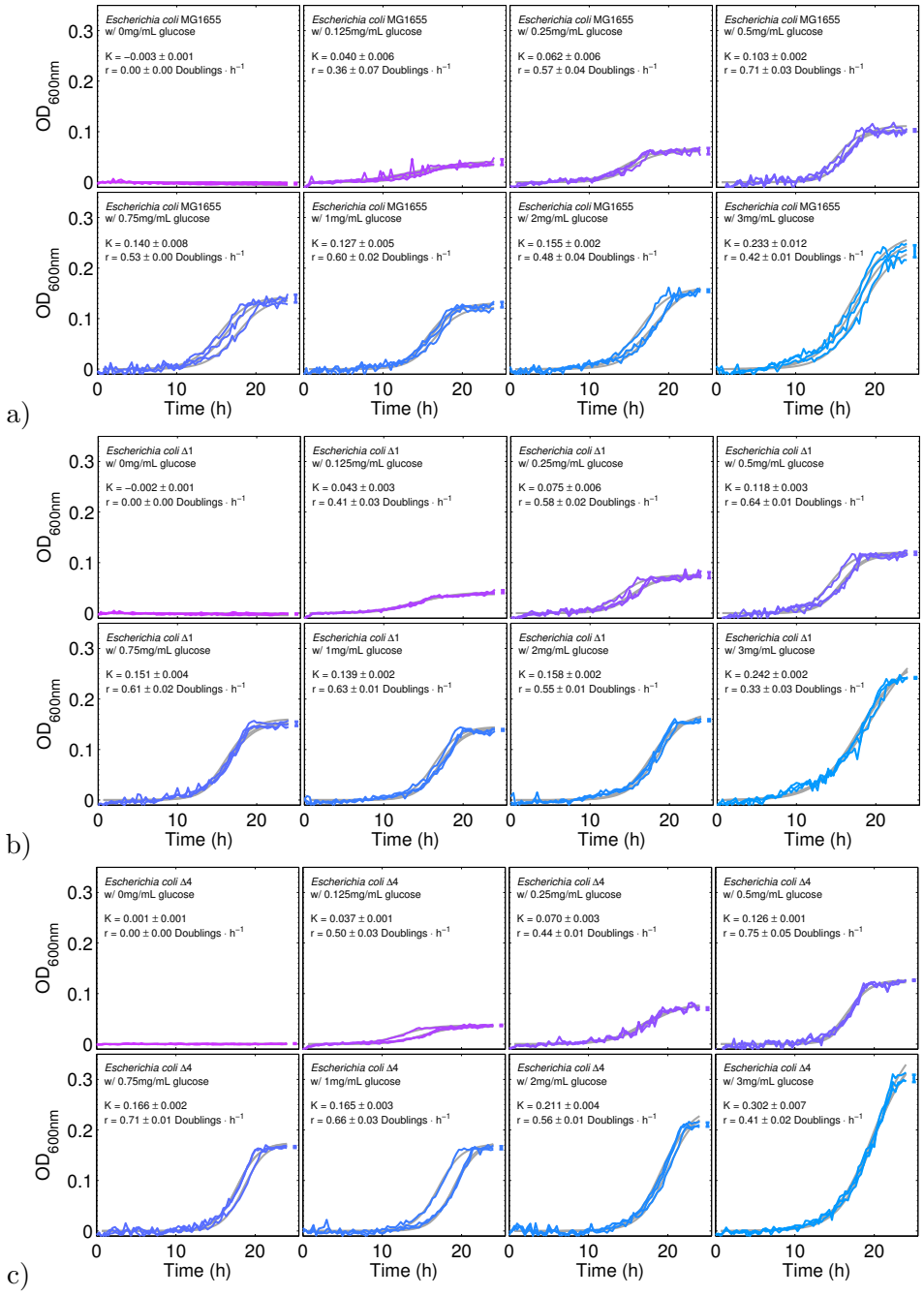
Carlos Reding-Roman, Mark Hewlett, Sarah Duxbury, Fabio Gori, Ivana Gudelj & Robert Beardmore

email: r.e.beardmore@exeter.ac.uk

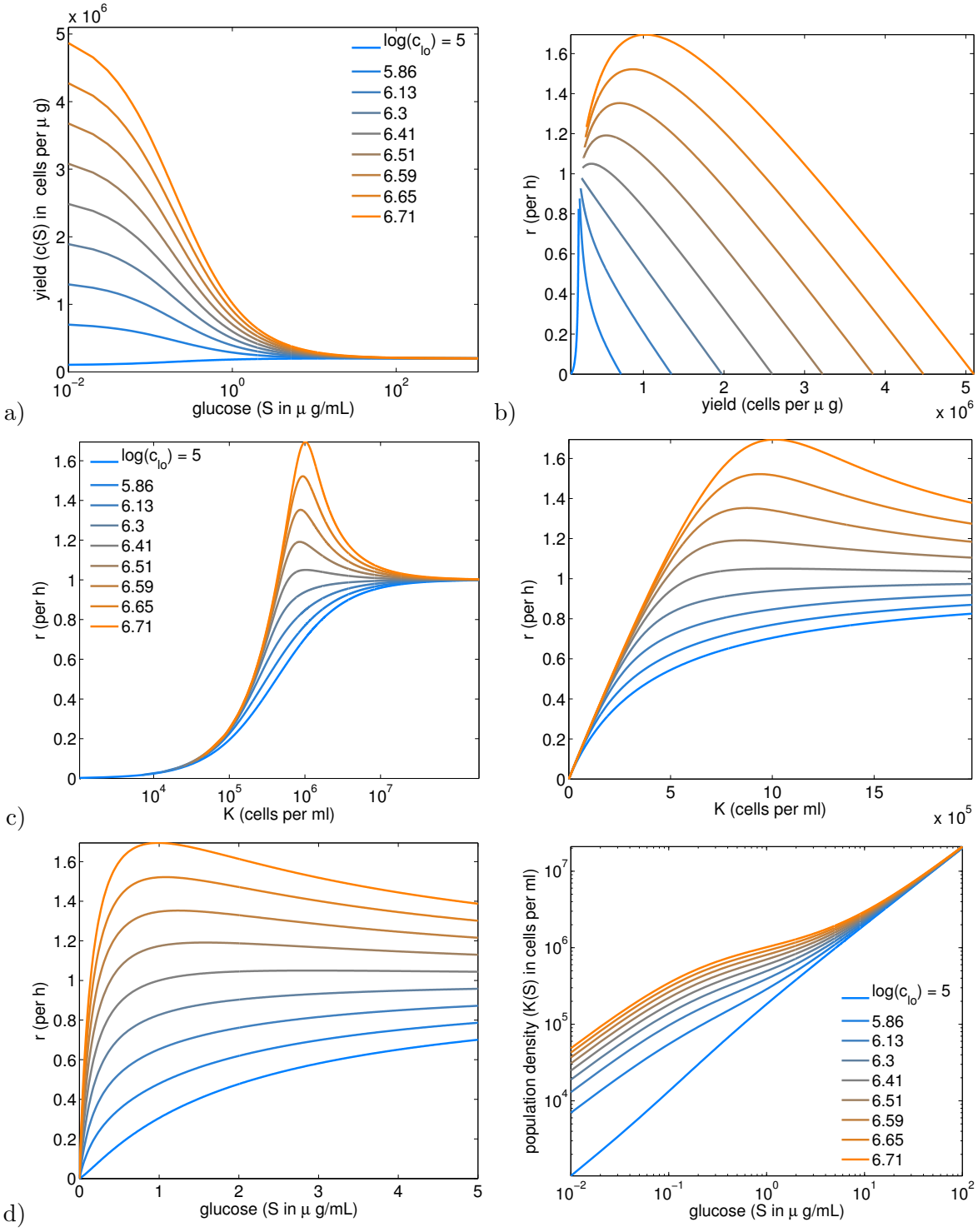
Supplementary Figures



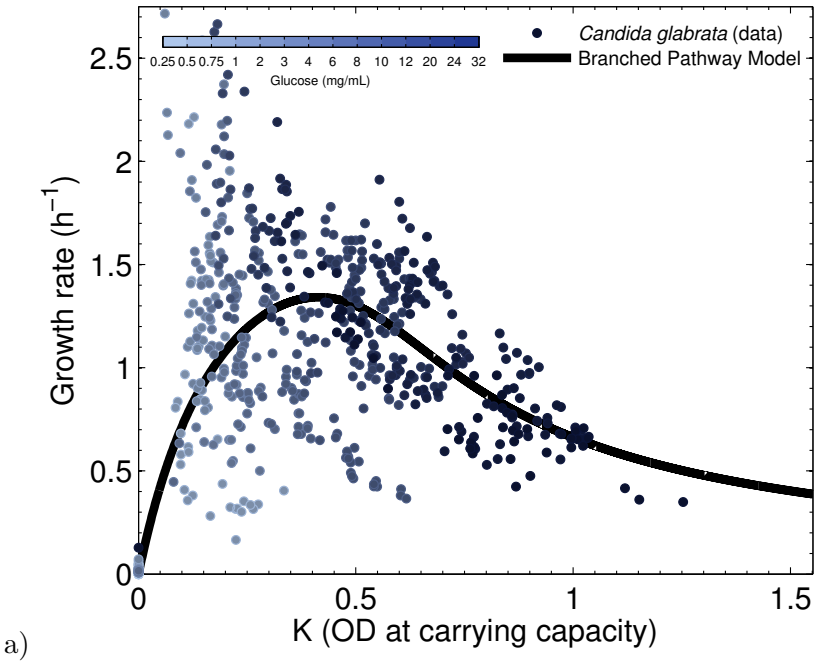
Supplementary Figure S1: Representative raw OD data (coloured lines) of the strains of *C. glabrata*. Each subplot represents the growth of the cultures over a 24h period in SC media supplemented with 0.25-32mg/mL of glucose. We used the logistic model, shown in grey but obscured by data in most cases, to calculate growth rate (r) and carrying capacity (K).



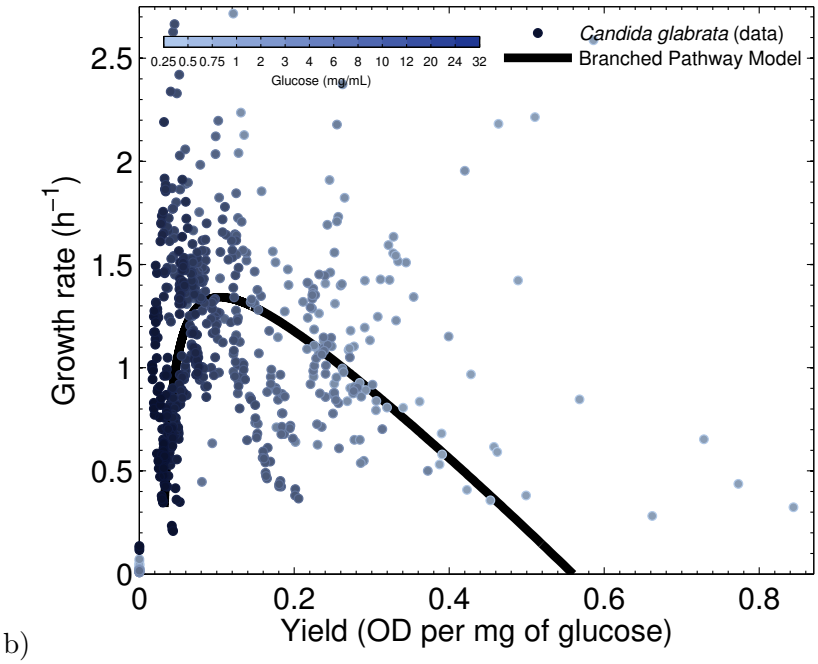
Supplementary Figure S2: Representative raw OD data (coloured lines) of *E. coli rrn* knockouts. Each subplot represents the growth dynamics of a 24h culture in minimal M9 media supplemented with 0-3mg/mL of glucose and casamino acids. The logistic model (grey) is used to calculate growth rate (r) and carrying capacity (K).


a)
x 10^6
yield (c(S) in cells per μ g)
glucose (S in μ g/mL)
log(c_lo) = 5
5.86
6.13
6.3
6.41
6.51
6.59
6.65
6.71
b)
r (per h)
yield (cells per μ g)
x 10^6
c)
r (per h)
K (cells per ml)
d)
r (per h)
glucose (S in μ g/mL)
K (cells per ml)
x 10^5
population density (K(S) in cells per ml)
glucose (S in μ g/mL)

Supplementary Figure S3: The relationship between the rate-yield and rK parabolae: one theory that morphs from a tradeup (blue) to a tradeup-tradeoff switch (orange). Starting with the expression (5) for yield: $c(S) = (c_{\rm hi} + pc_{\rm lo} \cdot S)/(1 + pS)$, note that this is a within-strain rate-yield tradeoff (RYTO) if and only if $c_{\rm lo} < c_{\rm hi}$. Panel (a) shows a plot of $c(S)$ for the values of $c_{\rm hi}$, measured in $\log_{10}$ cells per μg, as indicated in the panel legend, with $\log_{10}(c_{\rm lo})$ fixed at 5. Typical realisations of equation (7), the rate-yield relationship, then have the form shown in (b). Analogous rK relationships from (6) are then shown in (c); the two panels in (c) are identical, except one has linear and one has log-linear x-axes. As growth rate is defined to be $r(S) = c(S) \cdot V_{\rm max} S/(k_m + S)$ and recalling that $K = c(S) \cdot S$, an algebraic condition for an rK tradeoff can be derived on the parameters in this model from the requirement that there is an S^* for which growth is locally maximal: $\frac{dr}{dS}(S^*) = 0$. Models in (c) both do, and do not, meet this condition. For those that do not, r increases with K and we have an rK tradeup, where this condition is met, the rK-curve has a tradeup part that gives way to a tradeoff at large enough values of K. Panel (d) shows the two functions $r(S)$ and $K(S) = S \cdot c(S)$ in the definition of equation (6) for the same parameter values as panels (a-c). Note how in (d-right) the relationship between population size and the glucose supplied can be nonlinear.

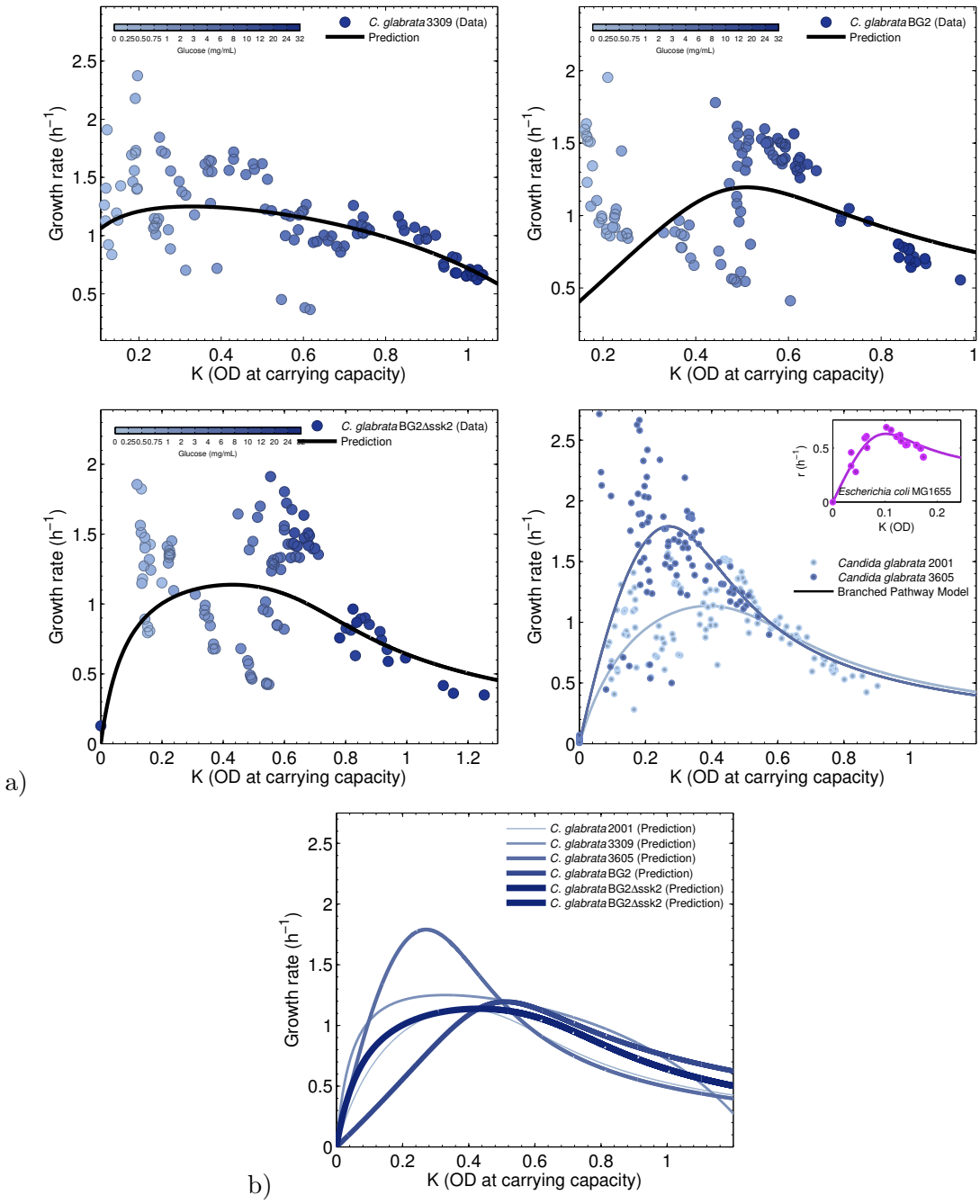


a)

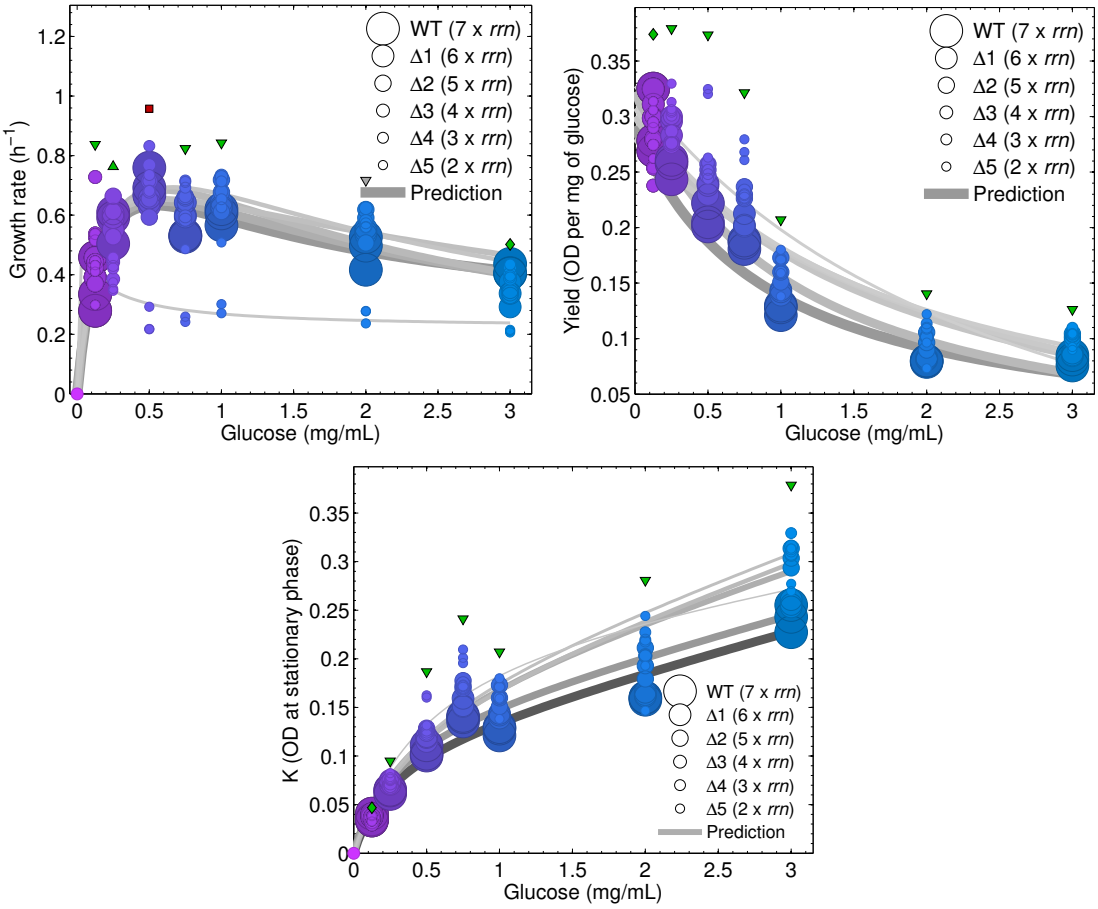


b)

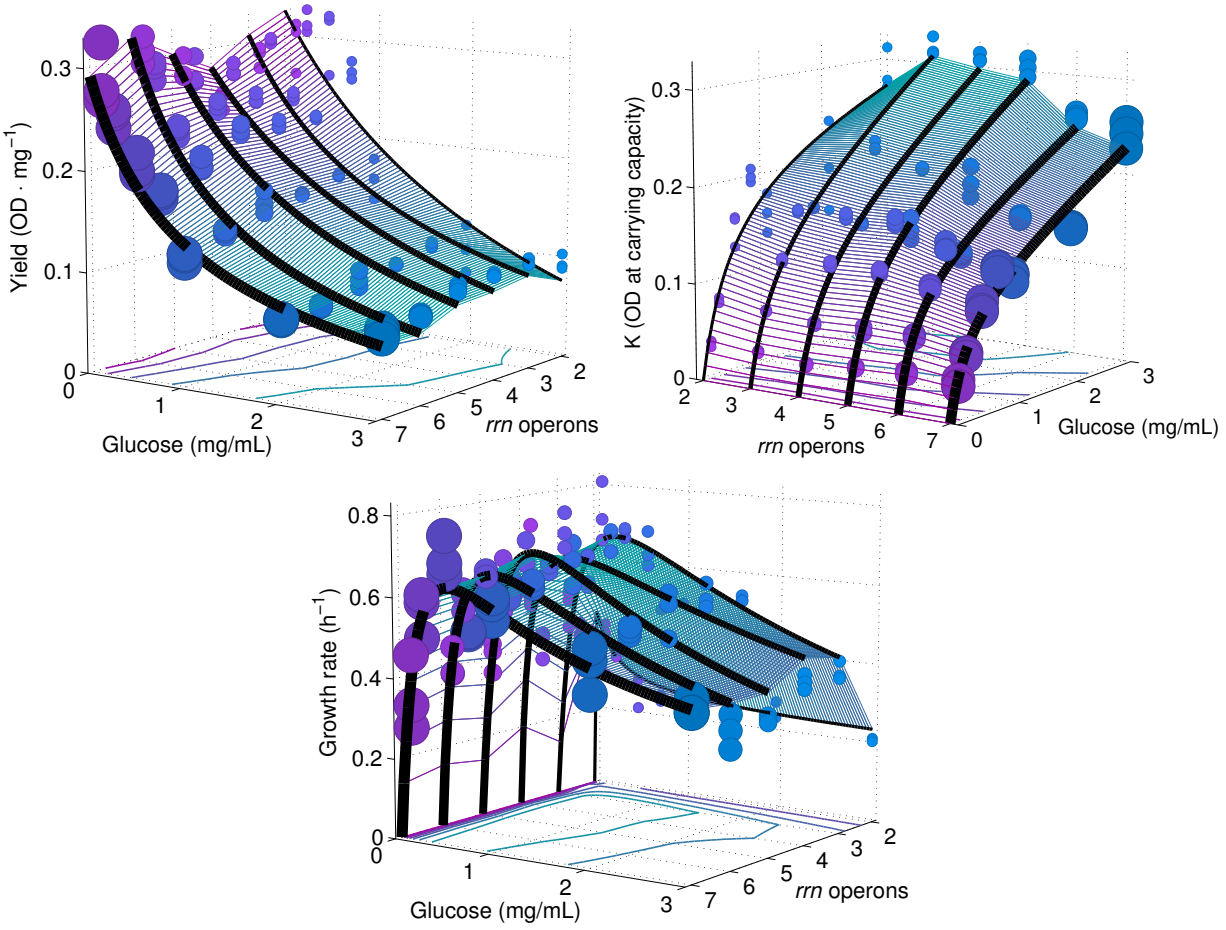
Supplementary Figure S4: Robustness of the skewed rK/rate-yield parabolae across *C. glabrata* stains. (a) Variation in growth rate (r) on the y-axis as a function of carrying capacity (K, x-axis) for data pooled from all cultured strains of *C. glabrata*. Dots are empirical data measured at different glucose concentrations (shown as different colours). The prediction from the model in equation (6) in the main text is represented in black. (b) The analogous plot to (a), but for rate-yield data, not rK data.



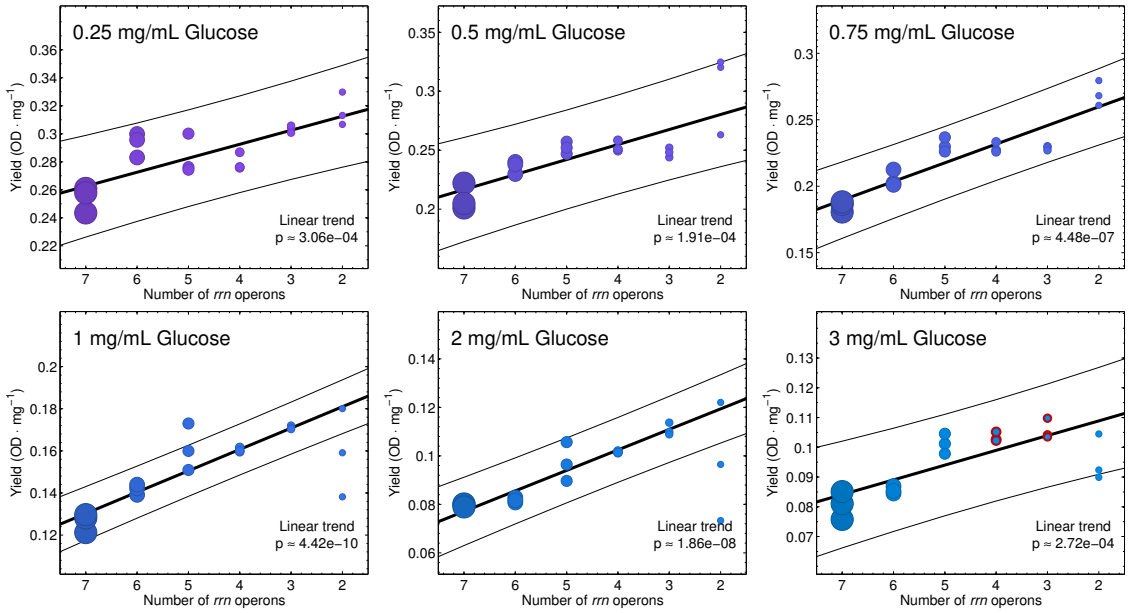
Supplementary Figure S5: rK parabolae of five *C. glabrata* strains. a) Relationships between growth rate (r, y-axis) and carrying capacity (K, x-axis) for *C. glabrata*, data are dots, robust predictions from equation (6) are lines. Bottom-right panel: rK profiles for two clinical isolates of *C. glabrata* (2001 and 3605) the latter was isolated from a diabetic patient and has higher r, but smaller K (predictions from equation (6) have adjusted $R^2 \approx \{0.982, 0.878\}$). (inset) An analogous unimodal rK geometry for *E. coli* MG1655. b) Putting the five predicted rK parabolas of *C. glabrata* in one plot highlights strain 3605 isolated from a diabetic patient with the greatest low-K, high-r skew.



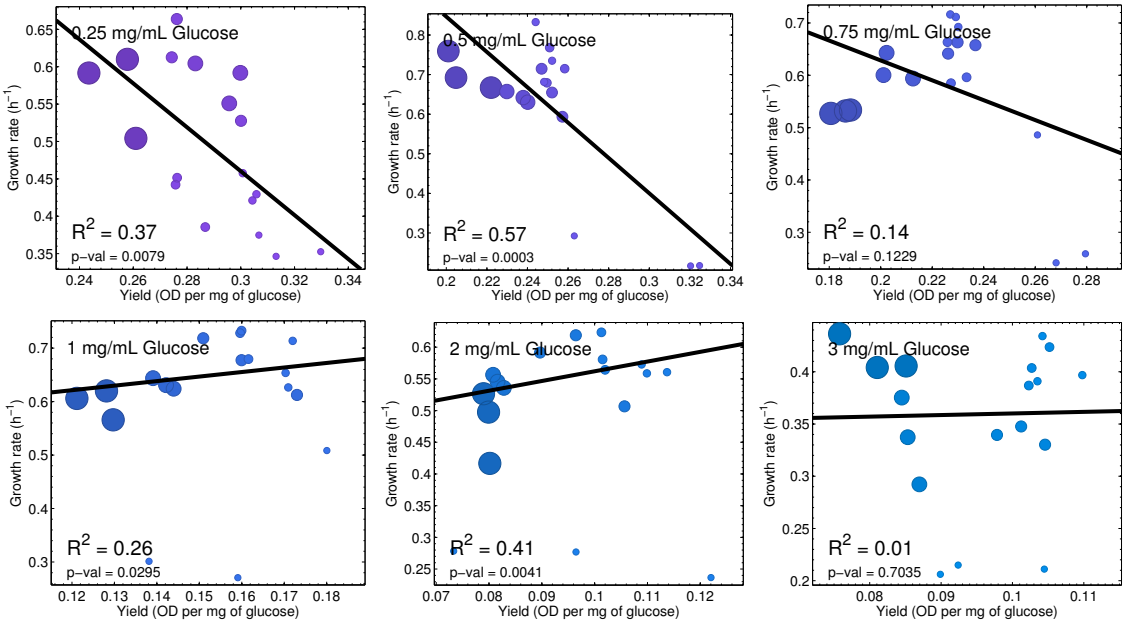
Supplementary Figure S6: r, K and yield for *E. coli rrn* knockout strains cultured at increasing glucose concentrations Growth rate (r, top left), yield (top right) and carrying capacity (K, centre) on the y-axis (dots) at different glucose concentrations (x-axis) alongside predictions from equations (5-7) (grey). For each concentration of glucose, between-strain differences in phenotype within each condition (i.e. for each fixed glucose concentration) were tested using a one-way ANOVA. Whether or not (green or red) these differences are significant ($p < 0.05$) is indicated by symbols above the data. If the symbol is a triangle, its orientation (as an up or down arrow) shows if the number of *rrn* operons is positively (up) or negatively (down) correlated with the phenotype in each plot and a square, not a triangle, is shown when no significant correlation is found either way. As a result, this analysis shows r, K and yield are negatively correlated with operon number in most conditions tested, however not all, and more details associated with these correlations can be found in Supplementary Figures 4 and S8.



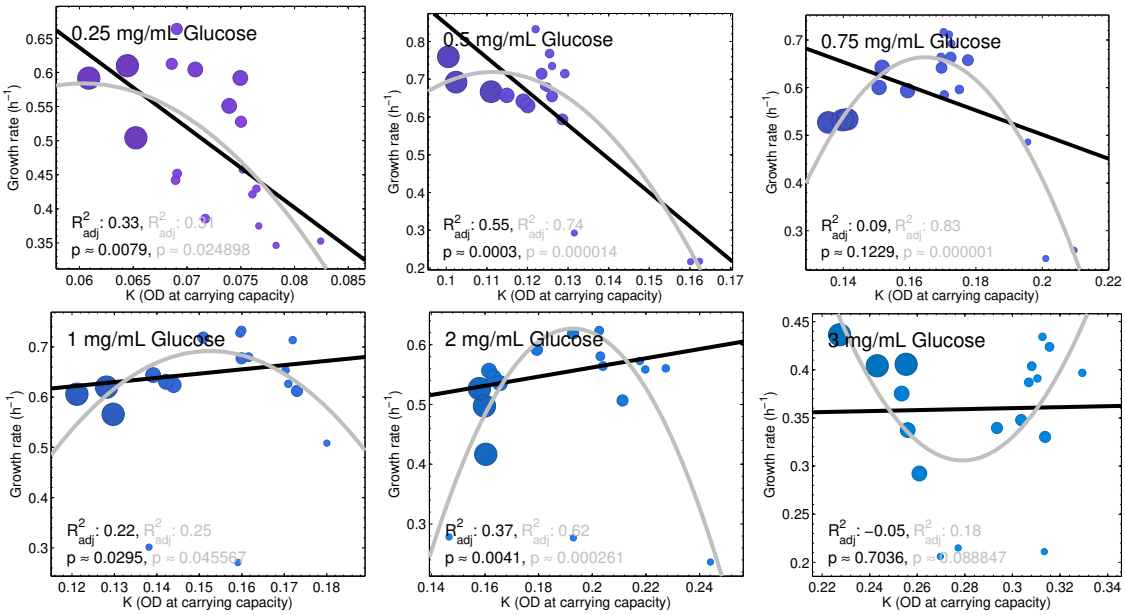
Supplementary Figure S7: Empirical yield, K and r landscapes (z-axes) showing each of these phenotypes as a function of glucose supply concentration and rrn copy number. Data (dots) are superimposed upon best-fit predictions using theory from the main text. Larger dots denote more rrn operons present in the genome, sizes as Figure 4.



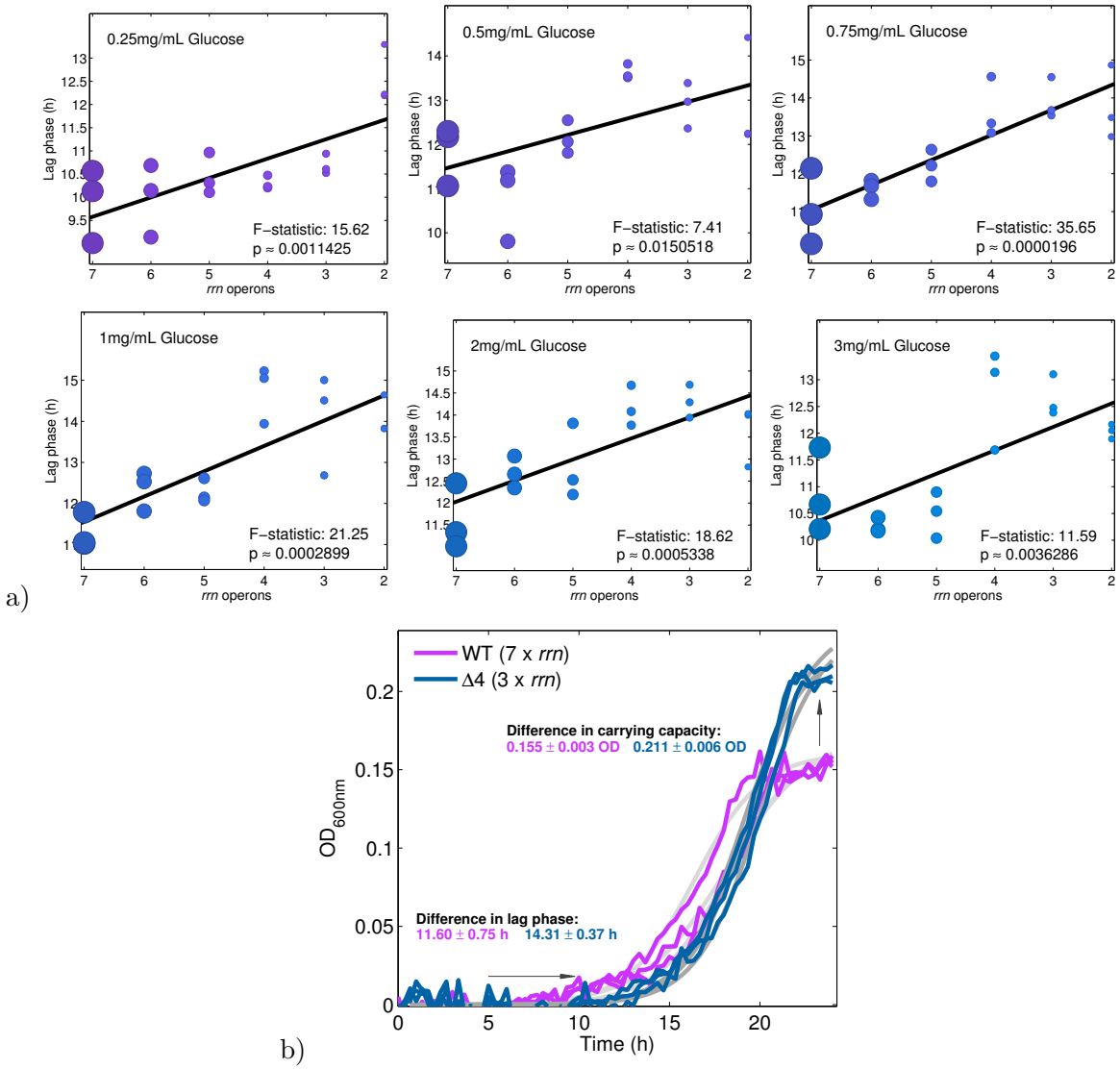
Supplementary Figure S8: Increasing copy number of the *rrn* operon reduces yield in *E. coli rrn* mutants. The dependence of biomass yield of the *rrn* knockout strains (y-axis) is shown as a function of *rrn* operon copy number (x-axis). Linear regressions (solid black line, the outer lines show the prediction $\pm$ 95% CI, n = 3) determine whether yield and operon copy number are correlated. The correlation is significantly negative in each case: note the x-axis has a negative orientation. **To obtain the analogous plot for the K phenotype, multiply y-axes by the amount of glucose in each panel, this does not change the qualitative form of each plot but it does scale the y-axis.**



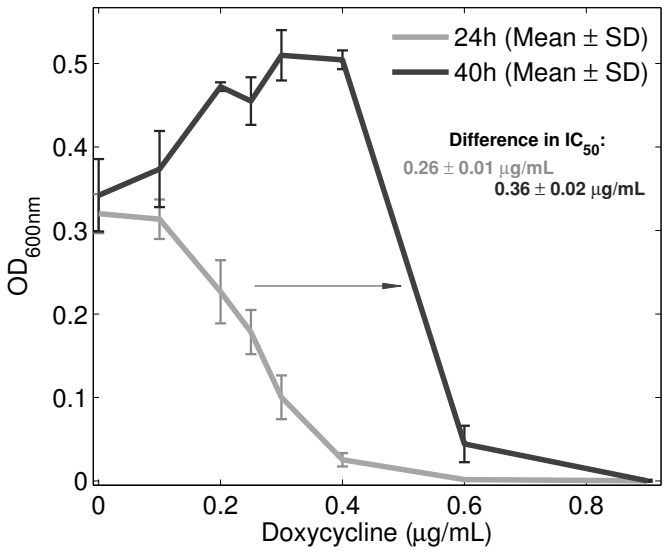
Supplementary Figure S9: Seeking between-strain tradeoffs and tradeups in *E. coli rrn* **mutants**: correlations between growth rate (r, y-axis) and carrying capacity (K, x-axis) are shown as a function of glucose supply (colour-coded) and *rrn* operon copy number (dot size). Using linear regressions (black line), we tested whether r and K are correlated. Clearly both negative and positive correlations are observed, depending on the glucose supplied and there is, therefore, no evidence of a rate-yield tradeoff between strains consistently across all environmental conditions. **To obtain the analogous plot for the K phenotype, multiply x-axes by the amount of glucose in each panel: this produces Supplementary Figure S10**.



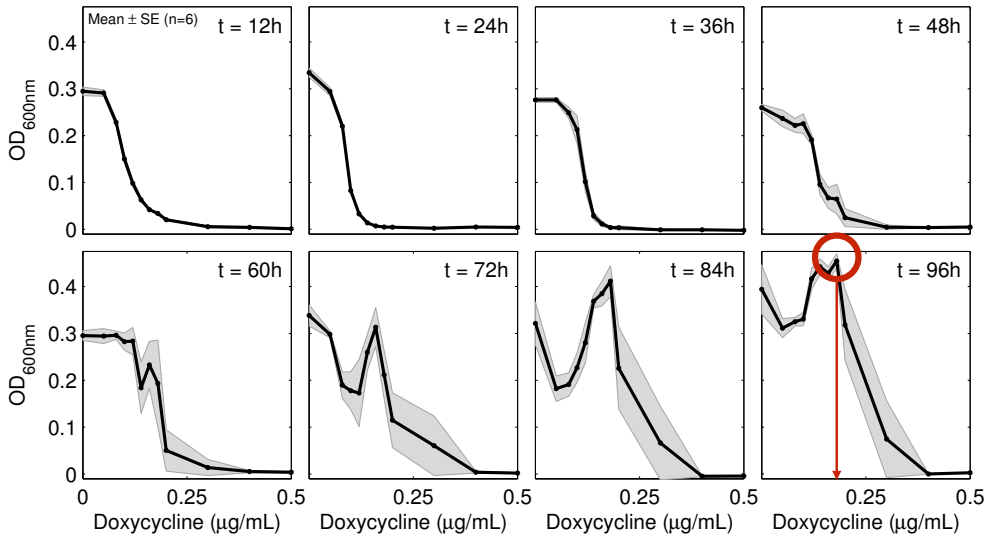
Supplementary Figure S10: Between-strain rK data in *E. coli rrn* mutants: this is a direct analogy of Supplementary Figure S9 but with K plotted on the x-axis instead of yield. A quadratic datafit is a better descriptor of the data than a linear regression in most cases (based on adjusted R^2values), but note how the form of quadratic is inverted at the highest glucose. The reasons for this are not known.



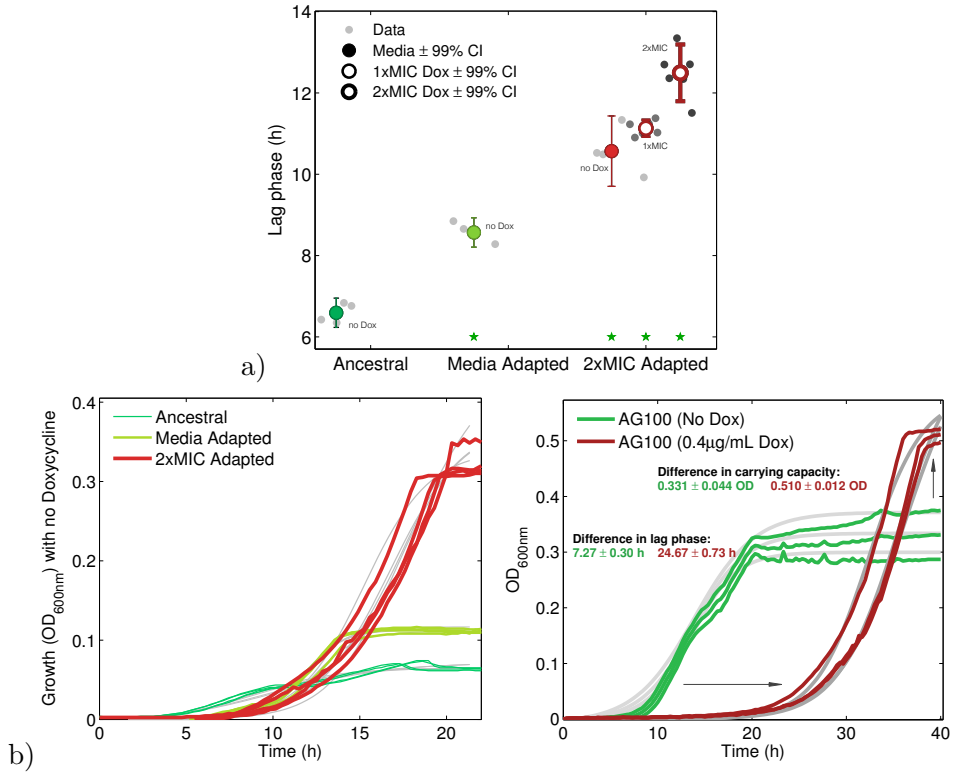
Supplementary Figure S11: Effect of copy number variation on the lag phase of $E.$ $coli$ rrn **mutants**. a) Duration of the lag phase (y-axis) as a function of rrn operon copy number (negatively oriented x-axis). Linear regressions (black line) determine whether lag and operon copy number are correlated, evidently the correlation is negative in all cases. b) A comparison of the growth kinetics of the wild-type strain of $E.$ $coli$ K12(MG1655) and the knockout strain with four rrn deletions ($\Delta 4$), where the y-axis shows the dynamics of optical density (OD_{600nm}) over a 24h period. We note that the removal of rrn operons does not alter the observed growth rate significantly in the case shown when 2mg/mL of glucose are supplied to the media (two-sided t-test, $t = -2.52, df = 4$ and $p \approx 0.065$). It does increase, however, both the length of the lag phase ($t = -5.58, df = 4$ and $p \approx 0.005$) and the carrying capacity ($t = -14.39, df = 4$ and $p \approx 0.0004$). Other glucose concentrations produce analogous observations.



Supplementary Figure S12: Doxycycline dose-response profiles for *E.coli* K12(AG100), aka 'Ancestral' in the main text. *E. coli* K12(AG100) OD_{600nm} on the y-axis during a growth cycle where measurements have been taken at two time points, one at 24h and one at 40h. The doxycycline concentration supplied is shown on the x-axis. Of note is the qualitatively different form of dose-response at the two different time points, particularly the large densities in the 40h dose-response curve. Thus, from these data, doxycycline can be a stimulant of optical density over periods longer than 24h incubation. For example, comparing OD in the zero dox condition against the OD in the 0.2ug/mL dox condition at 24h we obtain $t \approx -1.08, df = 4, p \approx 0.34$: no significant difference. Whereas comparing the OD of the zero dox condition versus that of the 0.2ug/mL dox condition at 48h we obtain a significant difference: $t \approx -9.52, df = 4, p \approx 7 \times 10^{-4}$. OD is therefore maximised in a drug treatment and not in the absence of drug. Importantly, the dose inhibiting cell growth by 50% relative to the no-drug condition (written IC_{50}) varies significantly depending on the incubation time ($t \approx -45.88, df = 2$ and $p \approx 0.0004$). **Based on the 24h data in this figure, as is standard, we take the minimal inhibitory concentration (MIC) of AG100 exposed to doxycycline to be $0.6\mu g/ml$ for the purposes of this study. This is also called $1\times$MIC in the main text; $2\times$MIC is double this value.**



Supplementary Figure S13: The course of adaptation of an *E. coli* K12(AG100) dose-response to doxycycline for ∼60 generations. Empirical dose-response data showing the growth of AG100 is represented on the y-axis, quoted as the mean optical density at 600nm ($OD_{600nm} \pm$ s.e. (the grey area), $n = 6$) with doxycycline on the x-axis. Treatment lasted four days with a new drug treatment every 12h, each period of which is termed a 'season' of growth: there are 8 seasons corresponding to each subplot in this figure. Note the 'non-monotonic overgrowth' feature of the later curves whereby some of the drug-treated populations can be as high as the density of the untreated control populations. The drug-treated sample highlighted by a red circle was subjected to whole-genome sequencing, as was the no-doxycycline treatment at the same timepoint.



Supplementary Figure S14: Comparison between the 'Ancestral', 'Media Adapted' and '2×MIC Adapted' strains of *E. coli* K12(AG100) described in the main text. a) Differences in lag phase, shown on the y-axis, for the three strains determined in three different environmental conditions: (1) in M9 media with glucose and casamino acids (called 'Media' or 'noDox'), (2) the same media with doxycycline at 1×MIC (called '1×MIC Dox') and (3) the same media supplemented with doxycycline at 2×MIC (called '2×MIC Dox'). A two-sided t-test, with $p < 0.05$, determined whether differences in lag were statistically significant (green star) or not (red star) from the Ancestral strain in the same media. Note the absence of lag phase when the Ancestral and Media Adapted strains are grown at 1x and 2xMIC: they do not growth in these conditions. b) Growth curves of the Ancestral, Media Adapted and 2xMIC Adapted strains cultured in the absence of antibiotic over 24h (left). Note how the growth of the 2xMIC Adapted strain is as fast as the other two and it reaches a greater population size in all three replicates (statistics on next page). (right) Note how the addition of 0.4μg/mL of doxycycline to media leads to a significant increase in lag phase of Ancestral's (AG100) growth ($t = -34.93, df = 2$ and $p \approx 0.0008$) and carrying capacity ($t = -8.08, df = 2$ and $p \approx 0.015$), importantly with no significant penalty in growth rate ($t = 1.19, df = 2$ and $p \approx 0.35$). Thus doxycycline here serves to increase lag phase, not decrease exponential growth rate.

Statistics required for Supplementary Figure S14(b):

T-test values that 'Media Adapted' has a **greater** r phenotype than '2xMIC Adapted':

$t \approx 10.97, df = 6, p \approx 3.4 \times 10^{-5}$.

t-test values that 'Ancestral' has a **lower** r phenotype than '2xMIC Adapted':

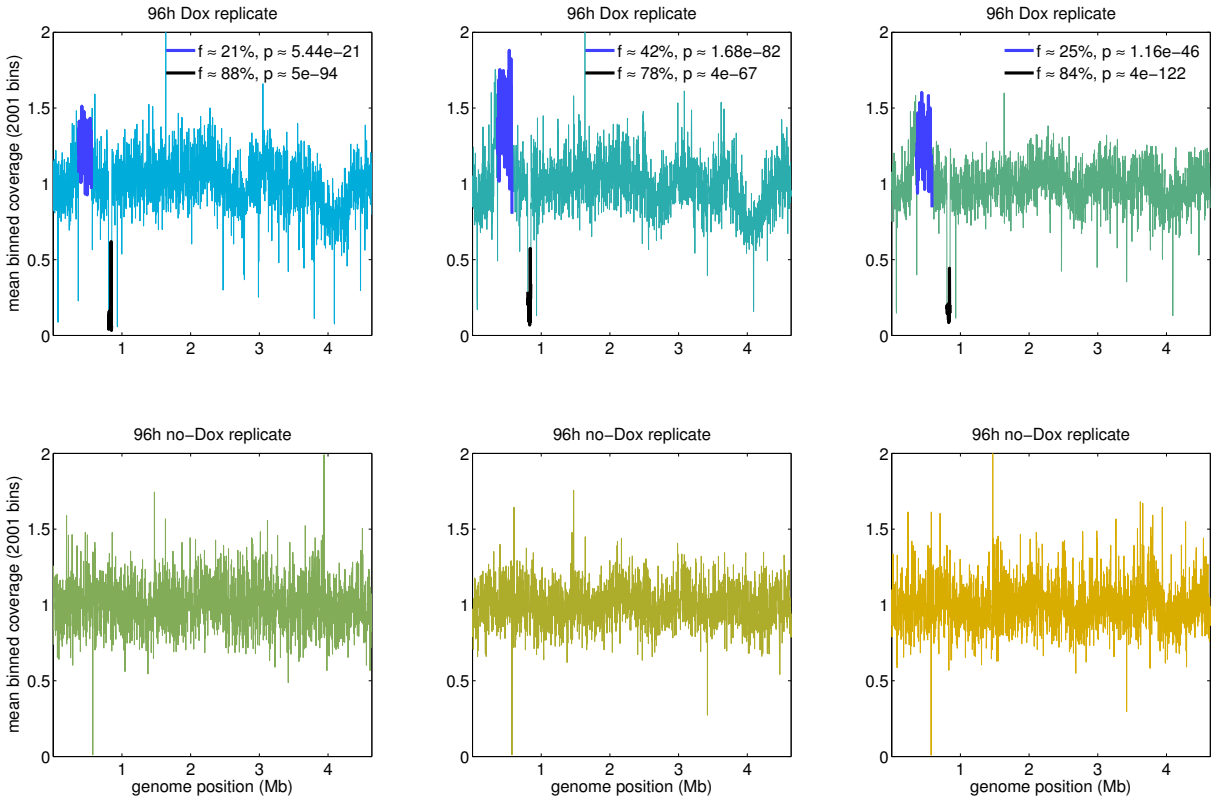
$t \approx -3.50, df = 6, p \approx 0.012$.

t-test values that 'Media Adapted' has a **smaller** K phenotype than '2xMIC Adapted':

$t \approx -23.37, df = 6, p \approx 4.02 \times 10^{-7}$.

t-test values that 'Ancestral' has a **smaller** K phenotype than '2xMIC Adapted':

$t \approx -28.12, df = 6, p \approx 1.34 \times 10^{-7}$.



Supplementary Figure S15: Illumina DNA coverage data for different treatments associated with Figures S13 and 7. (top) This shows Illumina whole-genome coverage data determined for the three doxycycline-treated replicates highlighted with a red circle in Figure S13. It shows the loss of a *dlp12* prophage region (in black) and the amplification of a large genomic region (in blue) containing the *acr* antibiotic efflux operon. The estimated frequencies of these mutations in each population are shown in each figure legend as an 'f' value. (bottom) This is analogous to the top figure but now for populations taken from Figure S13 at 96h that have not be exposed to doxycycline. Note the greater uniformity of coverage of Illumina reads across the genome: there is no detected loss of the *dlp12* prophage region nor duplication of the same *acr* efflux pump in the absence of doxycycline.

Supplementary Tables

Supplementary Table S1: SNPs common to all 96h AG100 doxycycline-adapted populations (K12-AG100$_{dox}$, three replicates)

Strain	Position	Reference	Alternate	Gene	Function	Status
K12-AG100$_{dox}$	3127332 b.p.	G	T	insH-4	transposase and	silent cgg → cgT
	3127350	A	G	insH-4	trans-activator	silent cca → ccG
	3127368	A	G	insH-4	present on insertion	silent cga → cgG
	3127402	T	C	insH-4	sequence 5 (IS5)	silent ttg → Ctg
	3127440	G	C	insH-4		silent acg → acC
	3127494	C	G	insH-4		non-silent cac → caG
	3127460	C	T	insH-4		non-silent gcg → gTg
	3648916	G	T	Slp	Outer membrane lipoprotein	non-silent atg → atT
	3648944	G	A	Slp	carbon-starvation inducible	non-silent gcg → Acg

Supplementary Table S2: Quality filters for all reported polymorphisms for 96h AG100, doxycycline-adapted populations (K12-AG100$_{dox}$).

Strain	Position	Reference	Alternate	Read Depth	High quality reads	Reference depth	Alternate depth	Frequency	P-value
K12-AG100$_{dox}$	3127332	G	T	85	77	65	12	15.58%	1.53×10^{-4}
	3127350	A	G	81	73	60	13	17.81	6.79×10^{-5}
	3127368	A	G	83	77	65	12	15.58	1.53×10^{-4}
	3127402	T	C	86	79	65	14	17.72	3.24×10^{-5}
	3127440	G	C	83	80	66	14	17.5	3.27×10^{-5}
	3127460	C	T	82	76	61	15	19.74	1.42×10^{-5}
	3127494	C	G	71	67	54	13	19.4	6.41×10^{-5}
	3421431	A	G	105	93	73	20	20.21	3.03×10^{-7}
	3648916	G	T	109	95	83	12	12.63	1.68×10^{-4}
	3648944	G	A	103	96	85	11	11.46	3.6×10^{-4}

Supplementary References

[41] Quinlan, A. & Hall, I. Bedtools: A flexible framework for comparing genomic features. *Bioinformatics* **26** (2010).