bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Modeling the growth of organisms validates a general relation between metabolic costs and natural selection

Efe Ilker[1,2] and Michael Hinczewski[2]

[1] *Physico-Chimie Curie UMR 168, Institut Curie, PSL Research University, 26 rue d'Ulm, 75248 Paris Cedex 05, France*

[2] *Department of Physics, Case Western Reserve University, Cleveland OH 44106*

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Metabolism and evolution are closely connected: if a mutation incurs extra energetic costs for an organism, there is a baseline selective disadvantage that may or may not be compensated for by other adaptive effects. A long-standing, but to date unproven, hypothesis is that this disadvantage is equal to the fractional cost relative to the total resting metabolic expenditure. This hypothesis has found a recent resurgence as a powerful tool for quantitatively understanding the strength of selection among different classes of organisms. Our work explores the validity of the hypothesis from first principles through a generalized metabolic growth model, versions of which have been successful in describing organismal growth from single cells to higher animals. We build a mathematical framework to calculate how perturbations in maintenance and synthesis costs translate into contributions to the selection coefficient, a measure of relative fitness. This allows us to show that the hypothesis is an approximation to the actual baseline selection coefficient. Moreover we can directly derive the correct prefactor in its functional form, as well as analytical bounds on the accuracy of the hypothesis for any given realization of the model. We illustrate our general framework using a special case of the growth model, which we show provides a quantitative description of overall metabolic synthesis and maintenance expenditures in data collected from a wide array of unicellular organisms (both prokaryotes and eukaryotes). In all these cases we demonstrate that the hypothesis is an excellent approximation, allowing estimates of baseline selection coefficients to within 15% of their actual values. Even in a broader biological parameter range, covering growth data from multicellular organisms, the hypothesis continues to work well, always within an order of magnitude of the correct result. Our work thus justifies its use as a versatile tool, setting the stage for its wider deployment.

Introduction

Discovering optimality principles in biological function has been a major goal of biophysics [1–6], but the competition between genetic drift and natural selection means that evolution is not purely an optimization process [7–9]. A necessary complement to elucidating optimality is clarifying under what circumstances selection is actually strong enough relative to drift in order to drive systems toward local optima in the fitness landscape. In this work we focus on one key component of this problem: quantifying the selective pressure on the extra metabolic costs associated with a genetic variant. We validate a long hypothesized relation [10–12] between this pressure and the fractional change in the total resting metabolic expenditure of the organism.

The effectiveness of selection versus drift hinges on two non-dimensional parameters [13]: i) the *selection coefficient* s, a measure of the fitness of the mutant versus the wild-type. Mutants will have on average $1+s$ offspring relative to the wild-type per wild-type generation

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

time; ii) the *effective population* N_e of the organism, the size of an idealized, randomly mating population that exhibits the same decrease in genetic diversity per generation due to drift as the actual population (with size N). For a deleterious mutant ($s < 0$) where $|s| \gg N_e^{-1}$, natural selection is dominant, with the probability of the mutant fixing in the population exponentially suppressed. In contrast if $|s| \ll N_e^{-1}$, drift is dominant, with the fixation probability being approximately the same as for a neutral mutation [7]. Thus the magnitude of N_e^{-1} determines the "drift barrier" [14], the critical minimum scale of the selection coefficient for natural selection to play a non-negligible role.

The long-term effective population size N_e of an organism is typically smaller than the instantaneous actual N, and can be estimated empirically across a broad spectrum of life: it varies from as high as $10^9 - 10^{10}$ in many bacteria, to $10^6 - 10^8$ in unicellular eukaryotes, down to $\sim 10^6$ in invertebrates and $\sim 10^4$ in vertebrates [12, 15]. The corresponding six orders of magnitude variation in the drift barrier N_e^{-1} has immense ramifications for how we understand selection in prokaryotes versus eukaryotic organisms, particularly in the context of genome complexity [16–18]. For example, consider a mutant with an extra genetic sequence relative to the wild-type. We can separate s into two contributions, $s = s_c + s_a$ [12]: s_c is the baseline selection coefficient associated with the metabolic costs of having this sequence, i.e. the costs of replicating it during cell division, synthesizing any associated mRNA / proteins, as well as the maintenance costs associated with turnover of those components; s_a is the correction due to any adaptive consequences of the sequence beyond its baseline metabolic costs. For a prokaryote with a low drift barrier N_e^{-1}, even the relatively low costs associated with replication and transcription are often under selective pressure [11, 12], unless $s_c < 0$ is compensated for an $s_a > 0$ of comparable or larger magnitude [19]. For the much greater costs of translation, the impact on growth rates of unnecessary protein production is large enough to be directly seen in experiments on bacteria [1, 20]. In contrast, for a eukaryote with sufficiently high N_e^{-1}, the same s_c might be effectively invisible to selection, even if $s_a = 0$. Thus even genetic material that initially provides no adaptive advantage can be readily fixed in a population, making eukaryotes susceptible to non-coding "bloat" in the genome. But this also provides a rich palette of genetic materials from which the complex variety of eukaryotic regulatory mechanisms can subsequently evolve [12, 21].

Part of the explanatory power of this idea is the fact that the s_c of a particular genetic variant should in principle be predictable from underlying physical principles. In fact, a very plausible hypothesis is that $s_c \approx -\delta C_T / C_T$, where C_T is the total resting metabolic expenditure of an organism per generation time, and δC_T is the extra expenditure of the mutant versus the wild-type. This relation can be traced at least as far back as the famous "selfish DNA" paper of Orgel and Crick [10], where it was mentioned in passing. But its true usefulness was only shown more recently, in the notable works of Wagner [11] on yeast and Lynch & Marinov [12] on a variety of prokaryotes and unicellular eukaryotes. By doing a detailed biochemical accounting of energy expenditures, they used the relation to derive

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

values of s_c that provided intuitive explanations of the different selective pressures faced by different classes of organisms. The relation provides a Rosetta stone, translating metabolic costs into evolutionary terms. And its full potential is still being explored, most recently in describing the energetics of viral infection [22].

Despite its plausibility and long pedigree, to our knowledge this relation has never been justified in complete generality from first principles. We do so through a general bioenergetic growth model, versions of which have been applied across the spectrum of life [23–25], from unicellular organisms to complex vertebrates. We show that the relation is universal to an excellent approximation across the entire biological parameter range.

Growth model

Let $\Pi(m(t))$ [unit: W] be the average power input into the resting metabolism of an organism (the metabolic expenditure after locomotion and other activities are accounted for [24]). $\Pi(m(t))$ can be an arbitrary function of the organism's current mass $m(t)$ [unit: g] at time t. This power is partitioned into maintenance of existing biological mass (i.e. the turnover energy costs associated with the constant replacement of cellular components lost to degradation), and growth of new mass (i.e. synthesis of additional components during cellular replication) [26]. Energy conservation implies

$$\Pi(m(t)) = B(m(t))m(t) + E(m(t))\frac{dm}{dt}, \tag{1}$$

Here $B(m(t))$ [unit: W/g] is the maintenance cost per unit mass, and $E(m(t))$ [unit: J/g] is the synthesis cost per unit mass. We allow both these quantities to be arbitrary functions of $m(t)$.

Though we will derive our main result for the fully general model of Eq. (1), we will also explore a special case: $\Pi(m(t)) = \Pi_0 m^\alpha(t)$, $B(m(t)) = B_m$, $E(m(t)) = E_m$, with scaling exponent α and constants Π_0, B_m, and E_m [25]. Allometric scaling of $\Pi(m(t))$ with $\alpha = 3/4$ across many different species was first noted in the work of Max Kleiber in the 1930s [27], and with the assumption of time-independent $B(m(t))$ and $E(m(t))$ leads to a successful description of the growth curves of many higher animals [23, 24]. However, recently there has been evidence that $\alpha = 3/4$ may not be universal [28, 29]. Higher animals still exhibit $\alpha < 1$ (with debate over $\alpha = 2/3$ versus $3/4$ [30]), but unicellular organisms have a broader range $\alpha \lesssim 2$. Thus we will use the model of Ref. [25] with an arbitrary species-dependent exponent α. While the resulting description is reasonable as a first approximation, particularly for unicellular organisms, one can easily imagine scenarios where the exponent and maintenance costs might vary between different developmental stages [31]. For the case of maintenance in endothermic animals, which in our approach includes all non-growth-related expenditures, more energy per unit mass is allocated to heat production as the organism matures [32], effectively increasing the cost of maintenance. In

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the Supplementary Information (SI) Sec. V [33] we show how the generalized model works in this scenario, using experimental growth data from two endothermic bird species [34]. Thus it is useful to initially consider the model in complete generality.

Baseline selection coefficient for metabolic costs

To derive an expression for s_c for the growth model of Eq. (1), we first focus on the generation time t_r, since this will be affected by alterations in metabolic costs. t_r is the typical age of reproduction, defined explicitly for any population model in SI Sec. I, where we relate it to the population birth rate r through $r = \ln(R_b)/t_r$ [35, 36]. Here R_b is the mean number of offspring per individual. Let $\epsilon = m_r/m_0$ be the ratio of the mass $m_r = m(t_r)$ at reproductive maturity to the birth mass $m_0 = m(0)$. For example in the case of symmetric binary fission of a unicellular organism, $R_b \approx \epsilon \approx 2$ (see SI Sec. III for a discussion of ϵ in more general models of cell size homeostasis). Since $m(t)$ is a monotonically increasing function of t for any physically realistic growth model, we can invert Eq. (1) to write the infinitesimal time interval dt associated with an infinitesimal increase of mass dm as $dt = dm\, E(m)/G(m)$ where $G(m) \equiv \Pi(m) - B(m)m$ is the amount of power channeled to growth, and we have switched variables from t to m. Note that $G(m)$ must be positive over the m range to ensure that $dm/dt > 0$. Integrating dt gives us an expression for t_r,

$$t_r = \int_{m_0}^{\epsilon m_0} dm \frac{E(m)}{G(m)}. \tag{2}$$

If we are interested in finding s_c for a genetic variation, we can focus on the additional metabolic costs due to that variation. For the purposes of calculation, this means treating the mutation as if it does not alter biological function in any other respect, including the ability of the organism to assimilate energy for its resting metabolism through uptake of nutrients or foraging. If the mutation actually had only metabolic cost effects, the full selection coefficient $s = s_c$. However generically mutations can affect both metabolic costs and power input (and/or other adaptive aspects), so $s = s_c + s_a$, with a correction term s_a due to the adaptive effects [12]. In the latter case s_c can still be calculated as shown below (ignoring adaptive effects) and interpreted as the baseline contribution to selection due to metabolic costs. While we do not focus on s_a here, our theory can be readily extended to consider adaptive contributions as well, as illustrated in SI Sec. VII, including aspects like spare respiratory capacity. This broader formalism is summarized in Fig. S3.

Proceeding with the s_c derivation, the products of the genetic variation (i.e. extra mRNA transcripts or translated proteins) may alter the mass of the mutant, which we denote by $\tilde{m}(t)$. The left-hand side of Eq. (1) remains $\Pi(m(t))$, where $m(t)$ is now the *unperturbed* mass of the organism (the mass of all the pre-variation biological materials). The power input $\Pi(m(t))$ depends on $m(t)$ rather than $\tilde{m}(t)$ since only $m(t)$ contributes to the processes that allow the organism to process nutrients, in accordance with the assumption that power

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

input is unaltered in order to calculate s_c. It is also convenient to express our dynamics in terms of $m(t)$ rather than $\tilde{m}(t)$, since the condition defining reproductive time t_r remains unchanged, $m(t_r) = \epsilon m_0$, or in other words when the unperturbed mass reaches ϵ times the initial unperturbed mass m_0. Thus Eq. (1) for the mutant takes the form $\Pi(m(t)) = \tilde{B}(m(t)) + \tilde{E}(m(t))dm(t)/dt$, where $\tilde{B}(m(t)) = B(m(t)) + \delta B$ and $\tilde{E}(m(t)) = E(m(t)) + \delta E$ are the mutant maintenance and synthesis costs. For simplicity, we assume the perturbations δB and δE are independent of $m(t)$, though this assumption can be relaxed. In SI Sec. IV, we show a sample calculation of δB and δE for mutations in *E.coli* and fission yeast involving short extra genetic sequences transcribed into non-coding RNA. This provides a concrete illustration of the framework we now develop.

Changes in the metabolic terms will perturb the generation time, $\tilde{t}_r = t_r + \delta t_r$, and consequently the birth rate $\tilde{r} = r + \delta r$. The corresponding baseline selection coefficient s_c can be exactly related to $\tilde{s}_c \equiv -\delta t_r/t_r$, the fractional change in t_r, through $s_c = R_b^{\tilde{s}_c/(1-\tilde{s}_c)} - 1$ (see SI Sec. I). This relation can be approximated as $s_c \approx \ln(R_b)\tilde{s}_c$ when $|\tilde{s}_c| \ll 1$, the regime of interest when making comparisons to drift barriers $N_e^{-1} \ll 1$. In this regime $\tilde{s}_c \approx \delta r/r$, the fractional change in birth rate. While we focus here on the the simplest case of exponential population growth, where $\tilde{s}_c$ is time-independent, we generalize our approach to density-dependent growth models, where $\tilde{s}_c$ varies between generations, in SI Sec. VI. $\tilde{s}_c$ can be written in a way that directly highlights the contributions of δE and δB to $\tilde{s}_c$. To facilitate this, let us define the average of any function $F(m(t))$ over a single generation time t_r as $\langle F \rangle \equiv t_r^{-1} \int_0^{t_r} dt\, F(m(t))$. Changing variables from t to m, like we did above in deriving Eq. (2), we can write this equivalently as $\langle F \rangle = \int_{m_0}^{\epsilon m_0} dm\, F(m)p(m)$, where $p(m) \equiv t_r^{-1} dt/dm = t_r^{-1} E(m)/G(m)$. The value $p(m)dm$ is just the fraction of the generation time that the organism spends growing from mass m to mass $m + dm$. Expanding Eq. (2) for t_r to first order in the perturbations δE and δB, the coefficient $\tilde{s}_c = -\delta t_r/t_r = -\sigma_E \delta E/\langle E \rangle - \sigma_B \delta B/\langle B \rangle$, with positive dimensionless prefactors

$$\sigma_E \equiv \langle E \rangle \langle E^{-1} \rangle, \quad \sigma_B \equiv \langle B \rangle \langle \Theta^{-1} \rangle. \tag{3}$$

Here $\Theta(m) \equiv G(m)/m$, and $F^{-1}(m) \equiv 1/F(m)$ for any F. The magnitude of σ_B versus σ_E describes how much fractional increases in maintenance costs matter for selection relative to fractional increases in synthesis costs. We see that both prefactors are products of time averages of functions related to metabolism. See SI Sec. II for a detailed derivation of Eq. (3), and also Eq. (4) below.

Relating the baseline selection coefficient to the fractional change in total resting metabolic costs

The final step in our theoretical framework is to connect the above considerations to the total resting metabolic expenditure C_T of the organism per generation time t_r, given

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

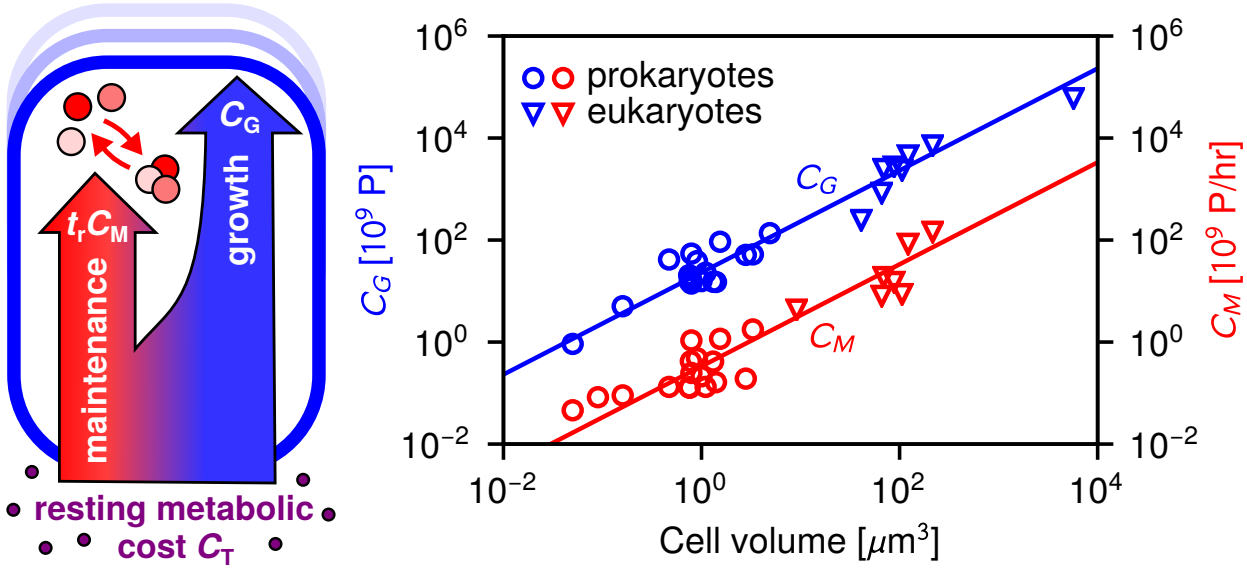

FIG. 1. The growth C_G (blue) and maintenance C_M (red) contributions to an organism's total resting metabolic cost $C_T = C_G + t_r C_M$ per generation time t_r. The symbols (circles = prokaryotes, triangles = unicellular eukaryotes) represent data tabulated in Ref. [12]. C_G and C_M have units of 10^9 P (phosphate bonds hydrolyzed), and 10^9 P/hr respectively. The lines represent best fits to the theoretical expressions for C_G and C_M from the allometric growth model.

by $C_T = \zeta \int_0^{t_r} dt\, \Pi(m(t)) = \zeta t_r \langle \Pi \rangle$. To compare with the experimental data of Ref. [12], compiled in terms of phosphate bonds hydrolyzed [P], we add the prefactor ζ which converts from units of J to P. Assuming an ATP hydrolysis energy of 50 kJ/mol under typical cellular conditions, we set $\zeta = 1.2 \times 10^{19}$ P/J. The genetic variation discussed above perturbs the total cost, $\tilde{C}_T = C_T + \delta C_T$, and the fractional change $\delta C_T / C_T$ can be expressed in a form analogous to $\tilde{s}_c$, namely $\delta C_T / C_T = \sigma'_E \delta E / \langle E \rangle + \sigma'_B \delta B / \langle B \rangle$, with

$$\sigma'_E \equiv \langle E \rangle \langle \Pi \rangle^{-1} \langle \Pi E^{-1} \rangle, \quad \sigma'_B \equiv \langle B \rangle \langle \Pi \rangle^{-1} \langle \Pi \Theta^{-1} \rangle, \tag{4}$$

where again the prefactors are expressed in terms of time averages over metabolic functions. The connection between s_c and $\delta C_T / C_T$ can be constructed by comparing Eq. (3) with Eq. (4). We see that $\tilde{s}_c = -\delta C_T / C_T$ for all possible perturbations δE and δB only when $\sigma_E = \sigma'_E$ and $\sigma_B = \sigma'_B$. We derive strict bounds on the differences between the prefactors (SI Sec. II), which show that the relation is exact when: i) $\Pi(m)$ is a constant independent of m; and/or ii) $E(m)$ and $\Theta(m)$ are independent of m. Outside these cases, the relation $\tilde{s}_c \approx -\delta C_T / C_T$ is an approximation. To see how well it holds, it is instructive to investigate the allometric growth model described earlier, where $\Pi(m(t)) = \Pi_0 m^\alpha(t)$, $E(m(t)) = E_m$, $B(m(t)) = B_m$.

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

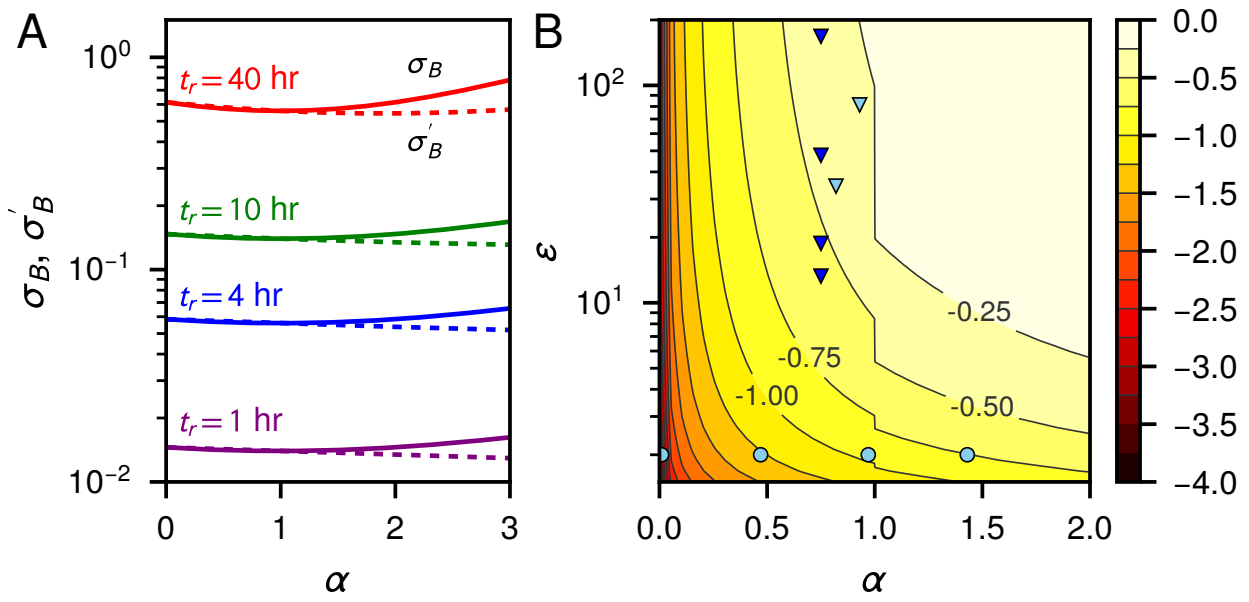

FIG. 2. A: σ_B (solid curves) from Eq. (3) and σ'_B (dashed curves) from Eq. (4) versus α, for the allometric growth model with $E_m = 2,600$ J/g, $B_m = 7 \times 10^{-3}$ W/g, and $\epsilon = 2$. At any given α, the parameter Π_0 for each pair of curves (different colors) is chosen to correspond to particular reproductive times t_r, indicated in the labels. B: Contour diagram showing the logarithm of the maximum possible discrepancy $\log_{10} |1 - \sigma'_B/\sigma_B|$ for any allometric growth model parameters, as a function of α and ϵ. To illustrate biological ranges α and ϵ, the symbols correspond to data for various species (circles = unicellular, triangles = multicellular) drawn from the growth trajectories analyzed in Ref. [25] (light blue) and Ref. [23] (dark blue). SI Sec. III shows a detailed species list.

Testing the relation in an allometric growth model

We use model parameters based on the metabolic data of Ref. [12], covering a variety of prokaryotes and unicellular eukaryotes. This data consisted of two quantities, C_G and C_M, which reflect the growth and maintenance contributions to C_T. Using Eq. (1) to decompose $\Pi(m(t))$, we can write $C_T = C_G + t_r C_M$, where $C_G = \zeta \int_{m_0}^{\epsilon m_0} dm\, E(m) = \zeta(\epsilon - 1) m_0 E_m$ is the expenditure for growing the organism, and $C_M = \zeta \langle Bm \rangle = \zeta B_m \langle m \rangle$ is the mean metabolic expenditure for maintenance per unit time. C_G and C_M scale linearly with cell volume (SI Sec. III), and best fits to the data, shown in Fig. 1, yield global interspecies averages: $E_m = 2,600$ J/g and $B_m = 7 \times 10^{-3}$ W/g. As discussed in the SI, these values are remarkably consistent with earlier, independent estimates, for unicellular and higher organisms [24, 25, 37, 38].

Since $E(m(t)) = E_m$ is a constant in the allometric growth model, $\sigma_E = 1$ from Eq. (3), and $\sigma_E = \sigma'_E$ holds exactly from Eq. (4). So the only aspect of the approximation that needs to be tested is the similarity between σ_B and σ'_B. Fig. 2A shows σ_B versus σ'_B for the range $\alpha = 0 - 3$, which includes the whole spectrum of biological scaling [28] up to $\alpha = 2$,

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

plus some larger α for illustration. For a given α, the coefficient Π_0 has been set to yield a certain division time $t_r = 1-40$ hr, encompassing both the fast and slow extremes of typical unicellular reproductive times. In all cases σ'_B is in excellent agreement with σ_B. For the range $\alpha \leq 2$ the discrepancy is less than 15%, and it is in fact zero at the special points $\alpha = 0, 1$. Clearly the approximation begins to break down at $\alpha \gg 1$, but it remains sound in the biologically relevant regimes. Note that σ_B values for $t_r = 1$ hr are ~ 0.01, reflecting the minimal contribution of maintenance relative to synthesis costs in determining the selection coefficient for fast-dividing organisms. This limit is consistent with microbial metabolic flux theory [39], where maintenance is typically neglected, so $\tilde{s}_c = -\delta C_T/C_T$ exactly (since only $\sigma_E = \sigma_{E'} = 1$ matters). As t_r increases, so does σ_B and hence the influence of maintenance costs, so by $t_r = 40$ hr, σ_B is comparable to σ_E.

To make a more comprehensive analysis of the validity of the $\tilde{s}_c \approx -\delta C_T/C_T$ relation, we do a computational search for the worst case scenarios: for each value of α and ϵ, we can numerically determine the set of other growth model parameters that gives the largest discrepancy $|1 - \sigma'_B/\sigma_B|$. Fig. 2B shows a contour diagram of the results on a logarithmic scale, $\log_{10} |1 - \sigma'_B/\sigma_B|$, as a function of α and ϵ. Estimated values for α and ϵ from the growth trajectories of various species are plotted as symbols to show the typical biological regimes. While the maximum discrepancies are smaller for the parameter ranges of unicellular organisms (circles) compared to multicellular ones (triangles), in all cases the discrepancy is less than 50%. To observe a serious error (σ'_B a different order of magnitude than σ_B), one must go to the large α, large ϵ limit (top right of the diagram) which no longer corresponds to biologically relevant growth trajectories.

Validity of the relation in more complex growth scenarios

Going beyond the simple allometric model, SI Sec. V analyzes avian growth data, where the metabolic scaling exponent varies between developmental stages. We find $\sigma_E = \sigma'_E = 1$ and the discrepancy $|1 - \sigma'_B/\sigma_B| \leq 30\%$. SI Sec. VI considers density-dependent growth, illustrated by examples of bacteria competing for a limited resource in a chemostat and predators competing for prey. Remarkably, when these systems approach a stationary state in total population and resource/prey quantity, we find $\sigma_E = \sigma'_E = 1$, $\sigma_B = \sigma'_B = (B_m \ln R_b)/(E_m d \ln \epsilon)$, where d is the dilution rate in the chemostat, or the predator death rate. The simple expression for σ_B allows straightforward estimation of the maintenance contribution to selection. For the chemostat that contribution can be tuned experimentally through the dilution rate d.

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Conclusion

We thus reach the conclusion that the baseline selection coefficient for metabolic costs can be reliably approximated as $s_c \approx -\ln(R_b)\delta C_T/C_T$. As in the original hypothesis [10–12], $-\delta C_T/C_T$ is the dominant contribution to the scale of s_c, with corrections provided by the logarithmic factor $\ln(R_b)$. Our derivation puts the relation for s_c on a solid footing, setting the stage for its wider deployment. It deserves a far greater scope of applications beyond the pioneering studies of Refs. [11, 12, 22]. Knowledge of s_c can also be used to deduce the adaptive contribution $s_a = s - s_c$ of a mutation, which has its own complex connection to metabolism [40] (see also SI Sec. VII). The latter requires measurement of the overall selection coefficient s, for example from competition/growth assays, and the calculation of s_c from the relation, assuming the underlying energy expenditures are well characterized. The s_c relation underscores the key role of thermodynamic costs in shaping the interplay between natural selection and genetic drift. Indeed, it gives impetus to a major goal for future research: a comprehensive account of those costs for every aspect of biological function, and how they vary between species, what one might call the "thermodynome". Relative to its more mature omics brethren—the genome, proteome, transcriptome, and so on—the thermodynome is still in its infancy, but fully understanding the course of evolutionary history will be impossible without it.

The authors thank useful correspondence with M. Lynch, and feedback from B. Kuznets-Speck, C. Weisenberger, and R. Snyder. E.I. acknowledges support from Institut Curie.

[1] E. Dekel and U. Alon, Nature **436**, 588 (2005).

[2] K. A. Dill, K. Ghosh, and J. D. Schmit, Proc. Natl. Acad. Sci. **108**, 17876 (2011).

[3] P. R. ten Wolde, N. B. Becker, T. E. Ouldridge, and A. Mugler, J. Stat. Phys. **162**, 1395 (2016).

[4] D. Hathcock, J. Sheehy, C. Weisenberger, E. Ilker, and M. Hinczewski, IEEE Trans. Mol. Biol. Multi-Scale Commun. **2**, 16 (2016).

[5] C. Zechner, G. Seelig, M. Rullan, and M. Khammash, Proc. Natl. Acad. Sci. **113**, 4729 (2016).

[6] S. Fancher and A. Mugler, Phys. Rev. Lett. **118**, 078101 (2017).

[7] M. Kimura, Genetics **47**, 713 (1962).

[8] T. Ohta, Nature **246**, 96 (1973).

[9] T. Ohta and J. H. Gillespie, Theor. Popul. Biol. **49**, 128 (1996).

[10] L. E. Orgel and F. H. Crick, Nature **284**, 604 (1980).

[11] A. Wagner, Mol. Biol. Evol. **22**, 1365 (2005).

[12] M. Lynch and G. K. Marinov, Proc. Natl. Acad. Sci. **112**, 15690 (2015).

[13] J. H. Gillespie, *Population genetics: a concise guide* (JHU Press, 2010).

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[14] W. Sung, M. S. Ackerman, S. F. Miller, T. G. Doak, and M. Lynch, Proc. Natl. Acad. Sci. **109**, 18488 (2012).

[15] B. Charlesworth, Nature Rev. Genet. **10**, 195 (2009).

[16] M. Lynch and J. S. Conery, Science **302**, 1401 (2003).

[17] M. Lynch, Mol. Biol. Evol. **23**, 450 (2005).

[18] E. V. Koonin, BMC Biol. **14**, 114 (2016).

[19] I. Sela, Y. I. Wolf, and E. V. Koonin, Proc. Natl. Acad. Sci. **113**, 11399 (2016).

[20] M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa, Science **330**, 1099 (2010).

[21] R. J. Taft, M. Pheasant, and J. S. Mattick, Bioessays **29**, 288 (2007).

[22] G. Mahmoudabadi, R. Milo, and R. Phillips, Proc. Natl. Acad. Sci. **114**, E4324 (2017).

[23] G. B. West, J. H. Brown, and B. J. Enquist, Nature **413**, 628 (2001).

[24] C. Hou, W. Zuo, M. E. Moses, W. H. Woodruff, J. H. Brown, and G. B. West, Science **322**, 736 (2008).

[25] C. P. Kempes, S. Dutkiewicz, and M. J. Follows, Proc. Natl. Acad. Sci. **109**, 495 (2012).

[26] S. Pirt, Proc. R. Soc. Lond. B **163**, 224 (1965).

[27] M. Kleiber, Hilgardia **6**, 315 (1932).

[28] J. P. DeLong, J. G. Okie, M. E. Moses, R. M. Sibly, and J. H. Brown, Proc. Natl. Acad. Sci. **107**, 12941 (2010).

[29] F. J. Ballesteros, V. J. Martinez, B. Luque, L. Lacasa, E. Valor, and A. Moya, Sci. Rep. **8**, 1448 (2018).

[30] C. R. White and R. S. Seymour, Proc. Natl. Acad. Sci. **100**, 4046 (2003).

[31] D. S. Glazier, Biol. Rev. **80**, 611 (2005).

[32] J. Werner, N. Sfakianakis, A. D. Rendall, and E. M. Griebeler, J. Theor. Biol. **444**, 83 (2018).

[33] See Supplementary Information, which includes Refs. [41-82].

[34] M. W. Dietz and R. H. Drent, Physiol. Zool. **70**, 493 (1997).

[35] R. M. May, Am. Nat. **110**, 496 (1976).

[36] V. M. Savage, J. F. Gillooly, J. H. Brown, G. B. West, and E. L. Charnov, Am. Nat. **163**, 429 (2004).

[37] M. E. Moses, C. Hou, W. H. Woodruff, G. B. West, J. C. Nekola, W. Zuo, and J. H. Brown, Am. Nat. **171**, 632 (2008).

[38] A. Maitra and K. A. Dill, Proc. Natl. Acad. Sci. **112**, 406 (2015).

[39] J. Berkhout, E. Bosdriesz, E. Nikerel, D. Molenaar, D. de Ridder, B. Teusink, and F. J. Bruggeman, Genetics **194**, 505 (2013).

[40] M. N. Price and A. P. Arkin, Genome Biol. Evol. **8**, 1917 (2016).

[41] J. Grilli, M. Osella, A. S. Kennard, and M. C. Lagomarsino, Physical Review E **95**, 032411 (2017).

[42] J. Pfleger, M. He, and M. Abdellatif, Cell Death Dis. **6**, e1835 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[43] S. W. Choi, A. A. Gerencser, and D. G. Nicholls, J. Neurochem. **109**, 1179 (2009).

[44] H. Yaginuma, S. Kawai, K. V. Tabata, K. Tomiyama, A. Kakizuka, T. Komatsuzaki, H. Noji, and H. Imamura, Sci. Rep. **4**, 6522 (2014).

[45] K. P. Nickens, J. D. Wikstrom, O. S. Shirihai, S. R. Patierno, and S. Ceryak, Mitochondrion **13**, 662 (2013).

[46] M. D. Brand and D. G. Nicholls, Biochem. J. **435**, 297 (2011).

[47] J. T. Sauls, D. Li, and S. Jun, Curr. Opin. Cell Biol. **38**, 38 (2016).

[48] G. Facchetti, F. Chang, and M. Howard, Curr. Opin. Sys. Biol. **5**, 86 (2017).

[49] J. L. Martinez, A. Fajardo, L. Garmendia, A. Hernandez, J. F. Linares, L. Martínez-Solano, and M. B. Sánchez, FEMS Microbiol. Rev. **33**, 44 (2008).

[50] P. Blanco, S. Hernando-Amado, J. Reales-Calderon, F. Corona, F. Lira, M. Alcalde-Rico, A. Bernardini, M. Sanchez, and J. Martinez, Microorganisms **4**, 14 (2016).

[51] E. Ilker and M. Hinczewski, in preparation (2019).

[52] M. Zampieri, T. Enke, V. Chubukov, V. Ricci, L. Piddock, and U. Sauer, Mol. Sys. Biol. **13**, 917 (2017).

[53] C. Dahlberg and L. Chao, Genetics **165**, 1641 (2003).

[54] A. H. Melnyk, A. Wong, and R. Kassen, Evol. Appl. **8**, 273 (2015).

[55] S. Pawar, A. I. Dell, and V. M. Savage, Nature **486**, 485 (2012).

[56] J. R. Karr, J. C. Sanghvi, D. N. Macklin, M. V. Gutschow, J. M. Jacobs, B. Bolival Jr, N. Assad-Garcia, J. I. Glass, and M. W. Covert, Cell **150**, 389 (2012).

[57] A. Marantan and A. Amir, Phys. Rev. E **94**, 012405 (2016).

[58] R. M. Sibly and J. H. Brown, Am. Nat. **173**, E185 (2009).

[59] F. Courchamp, L. Berec, and J. Gascoigne, *Allee effects in ecology and conservation* (Oxford Univ. Press, 2008).

[60] H. L. Smith, "The Rosenzweig-Macarthur predator-prey model," `https://math.la.asu.edu/~halsmith/Rosenzweig.pdf`.

[61] C. S. Holling, Can. Entomol. **91**, 385 (1959).

[62] P. Turchin, *Complex population dynamics: a theoretical/empirical synthesis* (Princeton Univ. Press, 2003).

[63] J. Monod, Annu. Rev. in Microbiol. **3**, 371 (1949).

[64] M. L. Rosenzweig and R. H. MacArthur, Am. Nat. **97**, 209 (1963).

[65] A. Novick and L. Szilard, Science **112**, 715 (1950).

[66] A. Novick and L. Szilard, Proc. Natl. Acad. Sci. **36**, 708 (1950).

[67] D. Gresham and J. Hong, FEMS Microbiol. Rev. **39**, 2 (2014).

[68] J. P. De Magalhães, J. Costa, and O. Toussaint, Nucleic Acids Res. **33**, D537 (2005).

[69] K. E. Sears, A. J. Kerkhoff, A. Messerman, and H. Itagaki, Physiol. Biochem. Zool. **85**, 159 (2012).

[70] V. Callier and H. F. Nijhout, PLoS ONE **7**, e45455 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[71] D. S. Glazier, A. G. Hirst, and D. Atkinson, Proc. R. Soc. B **282**, 20142302 (2015).

[72] A. Farlow, H. Long, S. Arnoux, W. Sung, T. G. Doak, M. Nordborg, and M. Lynch, Genetics **201**, 737 (2015).

[73] J. Charlesworth and A. Eyre-Walker, Mol. Biol. Evol. **23**, 1348 (2006).

[74] F. Chang, Mol. Biol. Cell **28**, 1819 (2017).

[75] R. Raghavan, E. A. Groisman, and H. Ochman, Genome Res. **21**, 1487 (2011).

[76] H. S. Leong, K. Dawson, C. Wirth, Y. Li, Y. Connolly, D. L. Smith, C. R. Wilkinson, and C. J. Miller, Nat. Commun. **5**, 3947 (2014).

[77] L.-M. Chevin, Biol. Lett. **7**, 210 (2011).

[78] R. E. Ricklefs, Proc. Natl. Acad. Sci. **107**, 10314 (2010).

[79] A. Wides and R. Milo, "Understanding the dynamics and optimizing the performance of chemostat selection experiments," arXiv preprint arXiv:1806.00272 (2018).

[80] R. Milo and R. Phillips, "Cell biology by the numbers," `http://book.bionumbers.org/` (2017).

[81] M. Nei, Molecular biology and evolution **22**, 2318 (2005).

[82] M. Tokuyama and I. Oppenheim, Physica A **216**, 85 (1995).

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Information

CONTENTS

I. Derivation of the relation between s_c and $\tilde{s}_c$ 14

II. Derivation of main text Eqs. (3)-(4) and related bounds 16

III. Fitting of allometric growth model to experimental data 17

IV. Sample calculation of the baseline selection coefficient: short, non-coding RNA in *E. coli* and fission yeast 19

V. Generalized growth model with varying developmental scaling regimes 21

VI. Generalizing the s_c and $\tilde{s}_c$ relation for density-dependent population models 25

VII. Understanding the roles of baseline metabolic costs versus adaptive effects in selection 31

I. Derivation of the relation between s_c and $\tilde{s}_c$

In the main text we posited an approximate relation $s_c \approx \ln(R_b)\tilde{s}_c$ between the baseline selection coefficient s_c and the fractional change in growth rate $\tilde{s}_c$ due to a genetic variation. Here we derive an exact relation between the two quantities, generalizing the approach used in Ref. [77] for the specific case of binary fission. We then show how the approximation used in the main text arises in the limit $|\tilde{s}_c| \ll 1$.

Consider a group of wild-type organisms with population $N_w(t)$ as a function of time. Over each generation time t_r the population grows by a factor of R_0, the net reproductive rate, so that after n generations,

$$N_w(nt_r) = (R_0)^n N_w(0). \tag{S1}$$

Here we focus on the simple case of exponential growth where both R_0 and t_r are time-independent. Below in Sec. VI we consider more complex scenarios in systems with density-dependent population growth. For a growing population $R_0 > 1$. Most generally R_0 and t_r are defined through $R_0 = \int_0^\infty dt\, p(t)f(t)$ and $t_r = R_0^{-1}\int_0^\infty dt\, tp(t)f(t)$, where $p(t)$ is the probability of the organism to survive from birth to age t, and $f(t)$ is the average fecundity at age t [35, 36]. A common simplification is to assume $f(t)$ is sharply peaked around age t_r, $f(t) \approx R_b\delta(t - t_r)$ with mean number of offspring R_b, so that $R_0 \approx R_b p(t_r)$ [25, 36]. One example is binary fission in bacteria, where $R_b = 2$, and $p(t_r) \approx 1$ if one neglects cell deaths

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(and there is no cell removal), so $R_0 \approx R_b = 2$. In the continuum time approximation to population growth, $n = t/t_r$ and Eq. (S1) becomes

$$\begin{aligned} N_w(t) = R_0^{t/t_r} N_w(0) &= e^{(t/t_r)\ln R_0} N_w(0) \\ &\approx e^{(t/t_r)\ln(R_b p(t_r))} N_w(0) \\ &\equiv e^{(r-d)t} N_w(0) \end{aligned} \tag{S2}$$

describing Malthusian growth with net rate $r - d$, the difference between the birth rate $r = \ln(R_b)/t_r$ and death rate $d = -\ln(p(t_r))/t_r$.

Now consider the growth of a mutant population $N_m(t)$. Under the baseline assumption on which we focus in the main text (neglecting any adaptive effects), the mutant has the same mean number of offspring R_b and death rate d, but metabolic cost differences affecting growth perturb the mean generation time to $\tilde{t}_r = t_r + \delta t_r$. This leads to a modified birth rate $\tilde{r} = r + \delta r = \ln(R_b)/(t_r + \delta t_r)$, and a growth equation

$$N_m(t) = e^{(\tilde{r}-d)t} N_m(0). \tag{S3}$$

From Eqs. (S2)-(S3) and the definitions of r and $\tilde{r}$, the ratio of the mutant to wild type populations is given by

$$\frac{N_m(t)}{N_w(t)} = \frac{N_m(0)}{N_w(0)} e^{t\delta r} = \frac{N_m(0)}{N_w(0)} R_b^{t\left(\frac{1}{t_r+\delta t_r} - \frac{1}{t_r}\right)}. \tag{S4}$$

On the other hand, after n wild-type generations ($t = nt_r$) the ratio of the two populations is related to the selection coefficient (as conventionally defined in population genetics) through

$$\frac{N_m(nt_r)}{N_w(nt_r)} = \frac{N_m(0)}{N_w(0)} (1 + s_c)^n. \tag{S5}$$

Plugging $t = nt_r$ into Eq. (S4) and comparing to Eq. (S5), we see that

$$s_c = R_b^{-\frac{\delta t_r}{t_r + \delta t_r}} - 1. \tag{S6}$$

If we define $\tilde{s}_c \equiv -\delta t_r/t_r$, then Eq. (S6) can be rewritten as

$$s_c = R_b^{\frac{\tilde{s}_c}{1-\tilde{s}_c}} - 1. \tag{S7}$$

Note that in the case where $|\tilde{s}_c| \ll 1$, or $|\delta t_r| \ll t_r$, we can also write $\tilde{s}_c \approx \delta r/r$, and so interpret $\tilde{s}_c$ as the fractional change in birth rate. In this same limit we can expand Eq. (S7) for small $\tilde{s}_c$,

$$s_c = \ln(R_b)\tilde{s}_c + \frac{1}{2}\ln(R_b)\left(2 + \ln(R_b)\right)\tilde{s}_c^2 + \cdots. \tag{S8}$$

Keeping only the leading order term, linear in $\tilde{s}_c$, yields the approximation $s_c \approx \ln(R_b)\tilde{s}_c$.

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

II. Derivation of main text Eqs. (3)-(4) and related bounds

To derive Eq. (3) in the main text, we start with the equation for t_r [main text Eq. (2)]:

$$t_r = \int_{m_0}^{\epsilon m_0} dm \frac{E(m)}{G(m)} = \int_{m_0}^{\epsilon m_0} dm \frac{E(m)}{\Pi(m) - B(m)m}, \tag{S9}$$

with $G(m) \equiv \Pi(m) - B(m)m$. Under the perturbations $E(m) \to E(m) + \delta E$ and $B(m) \to B(m) + \delta B$, the generation time is altered to

$$\begin{aligned} t_r + \delta t_r &= \int_{m_0}^{\epsilon m_0} dm \frac{E(m) + \delta E}{\Pi(m) - (B(m) + \delta B)m} \\ &\approx t_r + \delta E \int_{m_0}^{\epsilon m_0} dm \frac{1}{\Pi(m) - B(m)m} + \delta B \int_{m_0}^{\epsilon m_0} dm \frac{mE(m)}{(\Pi(m) - B(m)m)^2}, \end{aligned} \tag{S10}$$

where in the second line we have Taylor expanded the integrand to first order in δE and δB. We can thus write $\tilde{s}_c \equiv -\delta t_r / t_r$ as

$$\tilde{s}_c = -\frac{\delta t_r}{t_r} = -\frac{\delta E}{t_r} \int_{m_0}^{\epsilon m_0} dm \frac{1}{G(m)} - \frac{\delta B}{t_r} \int_{m_0}^{\epsilon m_0} dm \frac{mE(m)}{G^2(m)}. \tag{S11}$$

Using the definitions $p(m) \equiv t_r^{-1} E(m)/G(m)$, $\langle F \rangle \equiv \int_{m_0}^{\epsilon m_0} dm\, F(m) p(m)$, and $\Theta(m) \equiv G(m)/m$, we can rewrite Eq. (S11) as

$$\tilde{s}_c = -\langle E^{-1} \rangle \delta E - \langle \Theta^{-1} \rangle \delta B, \tag{S12}$$

where $F^{-1}(m) \equiv 1/F(m)$ for any function F. Finally we define the prefactors σ_E and σ_B through the relation

$$\tilde{s}_c = -\sigma_E \frac{\delta E}{\langle E \rangle} - \sigma_B \frac{\delta B}{\langle B \rangle}. \tag{S13}$$

Comparing Eq. (S12) to (S13) we find the result of main text Eq. (3):

$$\sigma_E = \langle E \rangle \langle E^{-1} \rangle, \quad \sigma_B = \langle B \rangle \langle \Theta^{-1} \rangle. \tag{S14}$$

The derivation of main text Eq. (4) proceeds analogously, starting with the total resting metabolic expenditure per generation,

$$C_T = \zeta \int_0^{t_r} dt\, \Pi(m(t)) = \zeta \int_{m_0}^{\epsilon m_0} dm \frac{dt}{dm} \Pi(m) = \zeta \int_{m_0}^{\epsilon m_0} dm \frac{E(m)\Pi(m)}{\Pi(m) - B(m)m}, \tag{S15}$$

where we have changed variables in the integral from t to m and used the fact that $dt/dm = E(m)/G(m)$. Under the perturbations $E(m) \to E(m) + \delta E$ and $B(m) \to B(m) + \delta B$, the expenditure is altered to

$$\begin{aligned} C_T + \delta C_T &= \zeta \int_{m_0}^{\epsilon m_0} dm \frac{(E(m) + \delta E)\Pi(m)}{\Pi(m) - (B(m) + \delta B)m} \\ &= C_T + \zeta \delta E \int_{m_0}^{\epsilon m_0} dm \frac{\Pi(m)}{\Pi(m) - B(m)m} + \zeta \delta B \int_{m_0}^{\epsilon m_0} dm \frac{E(m)\Pi(m)m}{(\Pi(m) - B(m)m)^2} \\ &= C_T + \zeta t_r \langle \Pi E^{-1} \rangle \delta E + \zeta t_r \langle \Pi \Theta^{-1} \rangle \delta B. \end{aligned} \tag{S16}$$

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Using the fact that Eq. (S15) can also be written as $C_T = \zeta t_r \langle \Pi \rangle$, we can use Eq. (S16) to express the ratio $\delta C_T / C_T$ as

$$\frac{\delta C_T}{C_T} = \frac{\langle \Pi E^{-1} \rangle}{\langle \Pi \rangle} \delta E + \frac{\langle \Pi \Theta^{-1} \rangle}{\langle \Pi \rangle} \delta B. \tag{S17}$$

Comparing this to the expression defining σ'_E and σ'_B,

$$\frac{\delta C_T}{C_T} = \sigma'_E \frac{\delta E}{\langle E \rangle} + \sigma'_B \frac{\delta B}{\langle B \rangle}, \tag{S18}$$

we find the result of main text Eq. (4):

$$\sigma'_E \equiv \langle E \rangle \langle \Pi \rangle^{-1} \langle \Pi E^{-1} \rangle, \quad \sigma'_B \equiv \langle B \rangle \langle \Pi \rangle^{-1} \langle \Pi \Theta^{-1} \rangle. \tag{S19}$$

The degree to which $\tilde{s}_c$ can be approximated as $\delta C_T / C_T$ depends on the similarity of the prefactor σ_E to σ'_E, and σ_B to σ'_B. Their relative differences can be written as:

$$\begin{aligned} \left| 1 - \frac{\sigma'_E}{\sigma_E} \right| &= \left| 1 - \frac{\langle \Pi E^{-1} \rangle}{\langle \Pi \rangle \langle E^{-1} \rangle} \right| \leq \kappa(\Pi) \kappa(E^{-1}), \\ \left| 1 - \frac{\sigma'_B}{\sigma_B} \right| &= \left| 1 - \frac{\langle \Pi \Theta^{-1} \rangle}{\langle \Pi \rangle \langle \Theta^{-1} \rangle} \right| \leq \kappa(\Pi) \kappa(\Theta^{-1}). \end{aligned} \tag{S20}$$

where $\kappa(F) \equiv \sqrt{\langle F^2 \rangle - \langle F \rangle^2} / \langle F \rangle$ and we have used the Cauchy-Schwarz inequality. These bounds imply two cases when $\tilde{s}_c$ is exactly equal to $\delta C_T / C_T$: i) $\kappa(\Pi) = 0$, which means $\Pi(m)$ is a constant independent of m; ii) $\kappa(\Pi) > 0$ and $\kappa(E^{-1}) = \kappa(\Theta^{-1}) = 0$, which means $E(m)$ and $\Theta(m)$ are independent of m.

III. Fitting of allometric growth model to experimental data

As discussed in the main text, we can decompose C_T into two components, $C_T = C_G + t_r C_M$, where $C_G = \zeta \int_{m_0}^{\epsilon m_0} dm\, E(m)$ is the expenditure for growing the organism, and $C_M = \zeta \langle B m \rangle$ is the mean metabolic expenditure for maintenance per unit time. For the allometric growth model, these contributions to C_T simplify to $C_G = \zeta(\epsilon - 1) m_0 E_m$ and $C_M = \zeta B_m \langle m \rangle$. Ref. [12] noted that C_G and C_M collected from experimental data scaled nearly linearly with cell volume, with allometric exponents of 0.97 ± 0.04 and 0.88 ± 0.07 respectively. In fact, the simplest version of the allometric model predicts exactly linear scaling, using the following assumptions. Since the data tabulated in Ref. [12] covers prokaryotes and unicellular eukaryotes, we take $\epsilon = 2$. Since the mass of the organism varies between m_0 and $2m_0$ over time t_r, we approximate $\langle m \rangle \approx (3/2) m_0$. Note that setting $\epsilon = 2$ assumes symmetric binary fission, though unicellular organisms can also exhibit asymmetric fission where $\epsilon \neq 2$ [57]. However since the variation in ϵ is typically within a factor of two of the symmetric case, any errors introduced by this assumption, and the approximation for $\langle m \rangle$, will not change the order of magnitude of the estimated model parameters. For simplicity we

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

are also ignoring variation in birth sizes m_0 over the population, which in the unicellular case is closely related to the possible mechanisms of cell size homeostasis, so-called sizer, adder, or timer behaviors [47, 48]. In the current context, once a steady state size distribution is reached, we can define ϵ as the ratio of mean size at division to the mean birth size m_0, which will have different interpretations depending on the mechanism. In the sizer picture $\epsilon = m_{\rm r}/m_0$, where $m_{\rm r}$ is the target cell mass at which division occurs. In the adder picture, where cells divide after adding a mass Δm, we have $\epsilon = 1+\Delta m/m_0$. In the timer mechanism, where division occurs after a specified time, additional conditions are needed to achieve a steady state size distribution, for example linear growth (in which case timer and adder are equivalent) or a mixed model coupling sizer and timer behavior [41]. In all cases, capturing the full effects of these mechanisms would require modeling the variation in birth sizes and their evolution over time, which we leave for a future work.

We relate the experimentally observed cell volume V to the mean cell mass $\langle m \rangle$ by assuming a typical cell is 2/3 water (density $\rho_{\rm wat} = 10^{-12}$ g/μm^3) and 1/3 dry biomass (density $\rho_{\rm dry} \approx 1.3 \times 10^{-12}$ g/μm^3) [80]. Hence $\langle m \rangle = (2\rho_{\rm wat} + \rho_{\rm dry})V/3 \equiv \rho_{\rm cell} V$. We thus find:

$$C_G = (2/3)\zeta E_m \rho_{\rm cell} V, \qquad C_M = \zeta B_m \rho_{\rm cell} V. \tag{S21}$$

For each expression we have only one unknown parameter, E_m and B_m respectively. Best fits to the Ref. [12] data, shown in main text Fig. 1B, yield global interspecies averages of the parameters, $E_m = 2,600$ J/g and $B_m = 7 \times 10^{-3}$ W/g.

The fitted values are consistent with earlier approaches, once water content is accounted for (i.e. to get E_m per dry biomass, multiply the value by ≈ 3, so $E_m^{\rm dry} = 7,800$ J/g). The synthesis cost E_m has a very narrow range across many species, with $E_m = 1,100 - 1,800$ J/g in bird and fish embryos, and $4,000 - 7,500$ J/g for mammal embryos and juvenile fish, birds, and mammals [37]. This energy scale seems to persist down to the prokaryotic level, with $E_m^{dry} = 3,345$ J/g estimated for *E. coli* [25]. $E_m^{\rm dry}$ also appears in a different guise as the inverse of the "energy efficiency" ε of *E. coli* growth in the model of Ref. [38]; converting the optimal observed $\varepsilon \approx 15$ dry g/(mol ATP) yields $E_m^{dry} = \zeta/\varepsilon = 3,333$ J/g, consistent with the other estimates cited above, as well as our fitted value. The ratio B_m/E_m was estimated for various species in Ref. [25], and found to vary in the range $10^{-6} - 10^{-5}$ s^{-1} from prokaryotes to unicellular eukaryotes, entirely consistent with our fitted value of $B_m/E_m = 3 \times 10^{-6}$ s^{-1}. The scale shifts for larger, multicellular species, but not dramatically. For example for a subset of mammals with scaling $\alpha = 3/4$, adult mass sizes $m_a = 10 - 6.5 \times 10^5$ g, and typical values of $\Pi_0 \approx 0.022$ W/g$^{3/4}$, $E_m \approx 7000$ J/g [24], we get a range of $B_m/E_m = 10^{-7} - 10^{-6}$ s^{-1}. We thus have confidence that the growth model provides a description of the metabolic expenditures (in terms of growth and maintenance contributions) that is consistent both with the empirical data of Ref. [12] and parameter expectations based on a variety of earlier approaches.

For the symbols in the contour diagram of main text figure Fig. 2B, we used parameters

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

extracted from growth trajectories analyzed in Ref. [25] (light blue) and Ref. [23] (dark blue). Circles (left to right) are unicellular organisms ($\epsilon = 2$): *T. weissflogii*, *L. borealis*, *B. subtilis*, *E. coli*. Triangles (top to bottom) are multicellular organisms: guinea pig, *C. pacificus*, hen, *Pseudocalanus sp.*, guppy, cow. For the multicellular case the plotted values of ϵ correspond to asymptotic adult mass in units of m_0. This is an upper bound on ϵ, though the actual ϵ should typically be comparable [23, 78].

IV. Sample calculation of the baseline selection coefficient: short, non-coding RNA in *E. coli* and fission yeast

To illustrate a calculation of baseline selection coefficients in the framework developed in the main text, let us consider a specific biological example: a mutant with a short (< 200 bp) sequence in the genome that is transcribed into non-coding RNA, and which is not present in the wild-type. We will focus on two organisms, the prokaryote *E. coli* and the unicellular eukaryote *S. pombe* (fission yeast). To date we know that at least some subset of non-coding RNA transcripts have functional roles in these organisms [75, 76]. The evolution of such regulatory sequences will be shaped both by the selective advantage s_a of having the sequence in the genome, and the baseline disadvantage s_c from the extra energetic costs of copying and transcription.

Before calculating s_c, we first establish the validity of the growth model for these organisms. The model parameters fitted for the data from prokaryotes and unicellular eukaryotes in Fig. 1B of the main text are $E_m = 2,600$ J/g and $B_m = 7 \times 10^{-3}$ W/g. The corresponding growth and maintenance contributions to the total resting metabolic cost per generation, C_G and C_M, are given by Eq. (S21). Using $\zeta = 1.2 \times 10^{19}$ P/J (recall that P corresponds to ATP or ATP equivalents hydrolyzed), $\rho_{\text{cell}} = 1.1 \times 10^{-12}$ g/μm^3, and typical cell volumes $V_{E.coli} = 1$ μm^3 [80], $V_{S.pombe} = 106$ μm^3 [74], we find: $C_G^{E.coli} = 2.30 \times 10^{10}$ P, $C_M^{E.coli} = 3.34 \times 10^8$ P/hr, $C_G^{S.pombe} = 2.43 \times 10^{12}$ P, $C_M^{S.pombe} = 3.54 \times 10^{10}$ P/hr. These agree well in magnitude with the literature estimates compiled in the SI of Ref. [12] (all normalized to 20°C): $C_G^{E.coli} = 1.57 \times 10^{10}$ P, $C_M^{E.coli} = 2.13 \times 10^8$ P/hr, $C_G^{S.pombe} = 2.35 \times 10^{12}$ P, $C_M^{S.pombe} = 8.7 \times 10^9$ P/hr. Thus the globally fitted E_m and B_m values are physically reasonable for both organisms.

The extra sequence in the mutant leads to perturbations in both synthesis cost per unit mass, δE, and maintenance cost per unit mass, δB. To calculate the first, we use the following estimates based on the analysis in Ref. [12]: for a sequence of length L, the total DNA-related synthesis cost is $d_\xi L$, where the label $\xi =$ *E. coli* or *S. pombe*. Here the prefactor $d_{E.coli} \approx 101$ P and $d_{S.pombe} \approx 263$ P. If the steady-state average number of corresponding mRNA trascripts in the cell is N_r, the additional ribonucleotide synthesis costs are $\approx 46 N_r L$

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

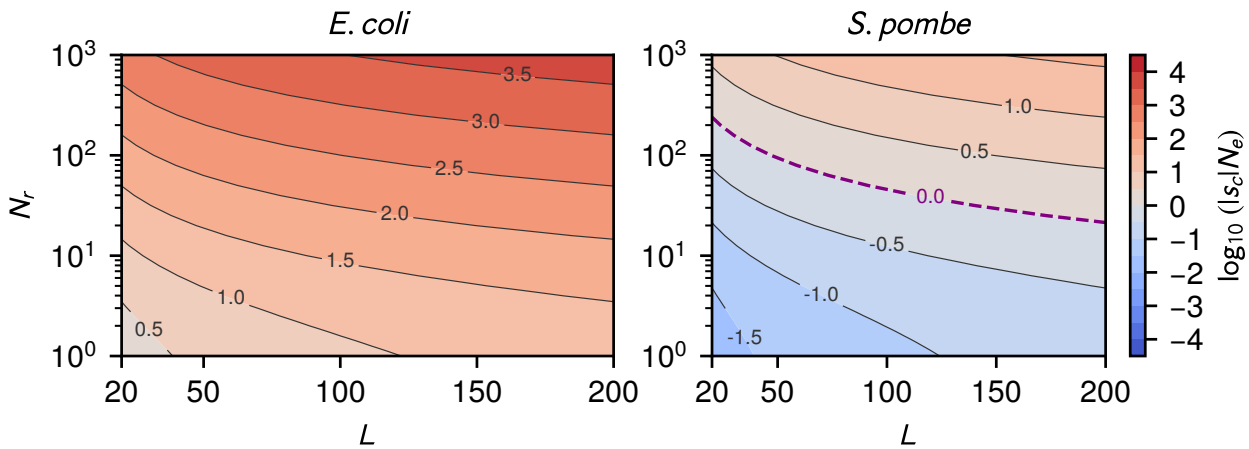

FIG. S1. Contour diagrams of $\log_{10}(|s_c|N_e)$ as a function of sequence length L and mean RNA transcript number N_r per cell for *E. coli* (left) and *S. pombe* (right). The dashed line in the diagram on the right corresponds to $|s_c| = N_e^{-1}$.

in units of P. Hence we have, per unit mass,

$$\delta E \approx \frac{\zeta^{-1} L}{\rho_{\text{cell}} V_\xi} \left(d_\xi + 46 N_r\right), \tag{S22}$$

with the ζ^{-1} prefactor converting from P to J, so that δE has units of J/g. The same analysis [12] yields the maintenance cost per unit time for replacing transcripts after degradation, $\approx 2 N_r L \gamma_\xi$ in units of P/s, where $\gamma_{E.coli} = 0.003$ s^{-1} and $\gamma_{S.pombe} = 0.001$ s^{-1} are the RNA degradation rates for the two organisms. Per unit mass, the maintenance perturbation δB is given by

$$\delta B \approx \frac{2\zeta^{-1} L N_r \gamma_\xi}{\rho_{\text{cell}} V_\xi}, \tag{S23}$$

in units of W/g.

The final step is to calculate the prefactors σ_E and σ_B from Eq. (3) in the main text. For this we need to choose a particular growth model exponent α, and we set $\alpha = 1$, corresponding to the assumption of exponential cell mass growth. In this case $\sigma_E = 1$ for both organisms, while $\sigma_B^{E.coli} = 0.0070$, $\sigma_B^{S.pombe} = 0.060$. The choice of α has a minimal influence on the prefactors: $\sigma_E = 1$ exactly for any model with a constant function $E(m) = E_m$. Moreover, any α value in the biologically relevant range of $0 \leq \alpha \leq 2$ yields a σ_B value within 5% of the $\alpha = 1$ result for each organism.

Putting everything together, we now can calulate all the components of main text Eq. (3) for $\tilde{s}_c$, namely σ_E, σ_B, δE, δB, $\langle E \rangle = E_m$, and $\langle B \rangle = B_m$. Had we chosen instead to use the $\delta C_T / C_T$ approximation of main text Eq. (4), the only discrepancy would have been in the fact that $\sigma'_B \neq \sigma_B$, since $\sigma'_E = \sigma_E = 1$. However the discrepancy is small, with $|1 - \sigma'_B/\sigma_B| < 0.09$ for both organisms in the range $0 \leq \alpha \leq 2$.

Fig. S1 shows contour diagrams of $\log_{10}(|s_c|N_e)$ as a function of L and N_r for *E. coli* and *S. pombe*. Here $s_c = \ln(2)\tilde{s}_c$, assuming $R_b = 2$, and the effective population sizes are

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$N_e^{E.coli} = 2.5 \times 10^7$ [73], $N_e^{S.pombe} = 1.2 \times 10^7$ [72]. For *E.coli*, with its smaller metabolic expenditures per generation relative to fission yeast, the cost of the extra sequence is always significant: $|s_c| > N_e^{-1}$ for the entire range of L and N_r considered, even for the smallest length ($L = 20$ bp) and a single transcript per cell on average, $N_r = 1$. Thus there will always be strong selective pressure to remove the extra sequence, unless s_c is compensated for by a comparable or greater adaptive advantage s_a. In contrast, for *S. pombe* there is a regime of L and N_r where $|s_c| < N_e^{-1}$ (the region below the dashed line). Here the selective disadvantage of the extra sequence is weaker than genetic drift, and such a genetic variant could fix in the population at roughly the same rate as a neutral mutation even if it conferred no selective advantage, $s_a = 0$. While this makes fission yeast more tolerant of genomic "bloat" relative to *E. coli*, initially non-functional extra genetic material could subsequently facilitate the development of novel regulatory mechanisms.

V. Generalized growth model with varying developmental scaling regimes

In the last section of the main text, we explored the validity of the baseline selection coefficient relation in the simplest version of the growth model, with allometric power input $\Pi(m(t)) = \Pi_0 m^\alpha(t)$ and time-independent synthesis and maintenance factors $E(m(t)) = E_m$, $B(m(t)) = B_m$. While this may be a reasonable approximation for various kinds of organisms [23–25] (particularly those without major physiological / morphological changes during their development [31]), there is evidence that in certain cases a single power law scaling with mass cannot accurately capture the resting metabolic rate (a detailed review is provided in Ref. [31]). This includes insects [69, 70] and marine invertebrates [71] that progress through several distinct developmental stages or instars, as well as endothermic birds when comparing juvenile and adult metabolism [34]. In these cases the resting metabolic input $\Pi(m(t))$ may scale approximately like a power-law in $m(t)$ during individual stages of development, but the exponent may vary from stage to stage. Additionally, the rate of energy expenditure for maintenance, $B(m(t))m(t)$ in our model, may also not have a simple time-independent prefactor $B(m(t)) = B_m$, since the cost of maintaining a unit of mass can also vary throughout development: in juvenile endothermic birds, before internal heat production has reached its mature level and while there is no energy allocation toward reproductive functions, the maintenance costs are lower than in adult birds [32]. There are also factors that influence metabolic scaling, for example the dimensionality of the environment in which the consumer organism encounters the resources on which it subsists [55]. Another example is a recent model where the partitioning of energy between metabolic processes and heat loss leads to a $\Pi(m)$ that is a linear combination of power law terms with different exponents (2/3 and 1) [29]. Thus it is important to have a general theory, one where $\Pi(m(t))$, $B(m(t))$ and even $E(m(t))$ could be arbitrary functions reflecting changes in the organism throughout development. We note that the mathematical formalism developed in the main

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

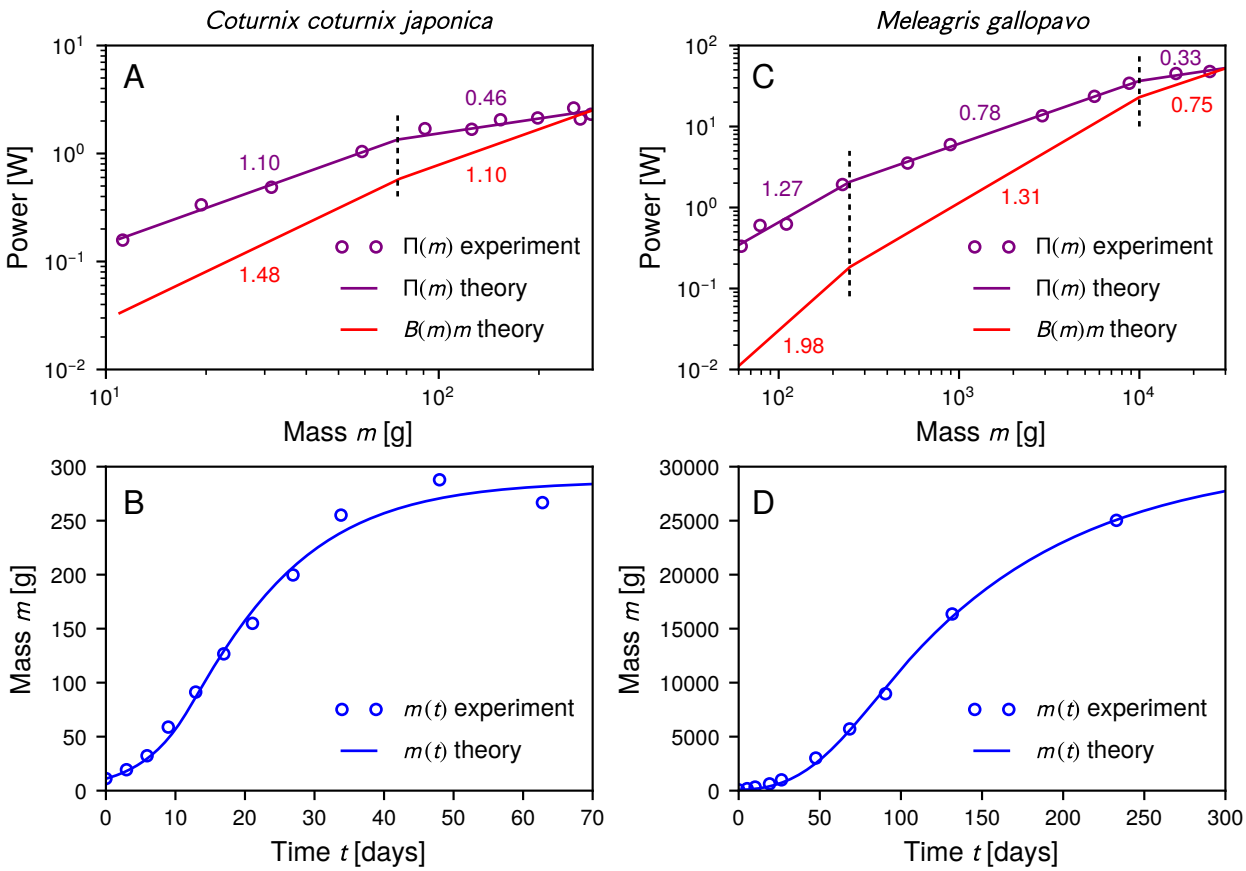

FIG. S2. Fits of the generalized growth model to experimental results (taken from Ref. [34]) for male members of two galliform species: Japanese quail (*Coturnix coturnix japonica*, left column); turkey (*Meleagris gallopavo*, right column). A, C: Power input $\Pi(m)$ as a function of mass m (purple), with symbols indicating experimental data and lines a piecewise continuous power-law fit to the data. The power $B(m)m$ expended in maintenance is shown in red, and is calculated by fitting the growth model to the mass vs. time data in panel B. Numbers next to the lines indicate the power-law exponents in each regime, while vertical dashed lines show the regime boundaries. B, D: Mean mass m versus time t, with symbols denoting experimental data and the curve the theoretical best-fit to the generalized growth model.

text, including Eqs. (1-5), is indeed fully general, independent of any specific choice for these functions.

To illustrate how this formalism can be applied for more complex growth models with varying developmental regimes, we analyze empirical data [34] from two endothermic, galliform bird species: Japanese quail (*Coturnix coturnix japonica*) and turkey (*Meleagris gallopavo*). Fig. S2 A,C shows the experimental resting metabolic rates $\Pi(m)$ (symbols) as a function of m. Before applying our selection coefficient theory, we first establish the details of the growth model for each organism. Following the analysis of Ref. [34], we fit $\Pi(m)$ using distinct power-law scaling exponents across different mass regimes (two regimes for quail, three for turkey). The theory fits thus appear as piecewise continuous linear functions

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(purple lines) on the log-log graphs of panels A and C, with the exponents indicated above each regime. The specific forms of the fitting functions are:

$$\text{quail:} \qquad \Pi(m) = \begin{cases} 0.0116\, m^{1.10} & m \leq 75 \\ 0.185\, m^{0.46} & m > 75 \end{cases}$$
$$\text{turkey:} \qquad \Pi(m) = \begin{cases} 0.00191\, m^{1.27} & m \leq 247 \\ 0.0288\, m^{0.78} & 247 < m \leq 10^4 \\ 1.70\, m^{0.33} & m > 10^4 \end{cases} \tag{S24}$$

where the units of m are g and of $\Pi(m)$ are W. These forms for $\Pi(m)$ fit the experimental data closely, and exhibit a trend seen commonly in organisms with distinct metabolic scaling at different life stages: the power-law exponent progressively decreases as the organism matures (see for example Type III and Type IV scaling behavior as classified by Glazier in Ref. [31]). And though the exponents vary between regimes, they all still fall in the typical biological range $\lesssim 2$.

The resting metabolic power input determines the mass trajectory $m(t)$ of the organism through the energy conservation equation [Eq. (1) in the main text]:

$$\Pi(m(t)) = B(m(t))m(t) + E(m(t))\frac{dm}{dt}. \tag{S25}$$

Since we know the empirical $m(t)$ curves (symbols in Figs. S2 B,D), we can use these to find best-fit forms for $B(m(t))$ and $E(m(t))$. For the latter we will assume the cost of synthesizing a new unit mass are constant throughout development, $E(m(t)) = E_m$. This assumption is based on the fact that estimated E_m values are broadly consistent among many different organisms at different life stages, as described in Sec. III above, generally of the order of magnitude $E_m \sim \mathcal{O}(10^3)$ W/g. For $B(m)$ we allow a more general fitting form, with the product $B(m)m$ (the maintenance power consumption) having piecewise continuous power-law scaling, with the same mass regimes as $\Pi(m)$. Note that in our model $B(m)m$ subsumes all the parts of the resting metabolic power that are consumed in processes other than growth. This is a broader definition of "maintenance" than in models which distinguish the uses of that power, for example separately keeping track of energy used for heat production and reproduction [32]. However for our purposes it is sufficient to collectively track the total non-growth expenditures through $B(m)m$. The time $t(m)$ to reach a mass m, starting from initial mass $m_0 = 11$ g (quail), 70 g (turkey), is analogous to Eq. (2) of the main text:

$$t(m) = \int_{m_0}^{m} dm \frac{E_m}{\Pi(m) - B(m)m}. \tag{S26}$$

By inverting this to get $m(t)$ and comparing to the empirical data for the mass trajectories, we can find theoretical best-fits for E_m and the functional forms of $B(m)m$. The results are

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$E_m = 6005$ J/g (quail), 6413 J/g (turkey) and

$$\begin{aligned} \text{quail:} \qquad B(m)m &= \begin{cases} 9.53 \times 10^{-4}\, m^{1.48} & m \leq 75 \\ 0.00506\, m^{1.10} & m > 75 \end{cases} \\ \text{turkey:} \qquad B(m)m &= \begin{cases} 3.29 \times 10^{-6}\, m^{1.98} & m \leq 247 \\ 1.37 \times 10^{-4}\, m^{1.31} & 247 < m \leq 10^4 \\ 0.0241\, m^{0.75} & m > 10^4 \end{cases} \end{aligned} \tag{S27}$$

where $B(m)m$ has units of W and m has units of g. The fitted theory results (solid curves in Fig. S2 B,D) agree very well with the $m(t)$ experimental data. The fitted values of E_m are also consistent with the range $(1400 - 7500$ J/g) seen for various juvenile bird species in an earlier analysis [24]. As expected, the fraction of power input available for growth, $1 - B(m)m/\Pi(m)$, decreases as the organism develops, eventually reaching zero at the asymptotic adult mass where $B(m)m$ intersects $\Pi(m)$.

With the details of the growth model established, evaluating our expressions for σ_E and σ_B [Eq. (3) of the main text] and comparing them to σ'_E and σ'_B [Eq. (4) of the main text] is relatively straightforward. Since $E(m(t)) = E_m$ is time-independent, $\sigma_E = \sigma'_E = 1$. This leaves only σ_B and σ'_B, given by:

$$\sigma_B = \langle B \rangle \langle \Theta^{-1} \rangle, \qquad \sigma'_B \equiv \langle B \rangle \langle \Pi \rangle^{-1} \langle \Pi \Theta^{-1} \rangle, \tag{S28}$$

where the averages are given by the following integrals:

$$\begin{aligned} \langle B \rangle &= \frac{1}{t_r} \int_{m_0}^{\epsilon m_0} \frac{B(m) E_m}{\Pi(m) - B(m)m}, \\ \langle \Theta^{-1} \rangle &= \frac{1}{t_r} \int_{m_0}^{\epsilon m_0} \frac{m E_m}{(\Pi(m) - B(m)m)^2}, \\ \langle \Pi \rangle &= \frac{1}{t_r} \int_{m_0}^{\epsilon m_0} \frac{\Pi(m) E_m}{\Pi(m) - B(m)m}, \\ \langle \Pi \Theta^{-1} \rangle &= \frac{1}{t_r} \int_{m_0}^{\epsilon m_0} \frac{m \Pi(m) E_m}{(\Pi(m) - B(m)m)^2}. \end{aligned} \tag{S29}$$

The time for reproductive maturity is $t_r = 52$ days (quail), 365 days (turkey), based on mean values from the AnAge online database [68]. From the theoretically fitted $m(t)$ curves this translates to $m(t_r) = \epsilon m_0$ where $\epsilon = 25$ (quail), 416 (turkey). The integrals in Eq. (S29) can be evaluated numerically, since all the expressions in the integrand are known based on the growth model fitting. The results for the prefactors in Eq. (S28) are:

$$\begin{aligned} \text{quail:} \qquad \sigma_B &= 7.43, \qquad \sigma'_B = 9.60 \\ \text{turkey:} \qquad \sigma_B &= 11.04, \qquad \sigma'_B = 14.05 \end{aligned} \tag{S30}$$

The discrepancy $|1 - \sigma'_B/\sigma_B| = 0.27 - 0.29$ in both cases. This is comparable to the simple allometric growth model (single power-law) cases investigated in the main text, where the

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

discrepancy was always less than 50%. Thus the relation $s_c \approx -\ln(R_b)\delta C_T/C_T$ continues to hold even for more complex growth models in organisms where metabolic scaling varies with developmental stage.

VI. Generalizing the s_c and $\tilde{s}_c$ relation for density-dependent population models

The derivation in Sec. I above, relating the baseline selection coefficient s_c to the fractional change in growth rate, $\tilde{s}_c = -\delta t_r/t_r \approx \delta r/r$, can be generalized to cases where the population growth of the wild-type and mutant organisms is density-dependent. This can occur for example when there is competition for a limited resource shared between the wild-type and mutant (i.e. a nutrient in the case of bacterial growth, or a prey population for a predatory organism), or when there are other external constraints on growth as the overall population increases. As a result, the reproductive time t_r may change from generation to generation, for example lengthening as the resource is depleted and organismal growth is slowed. We will denote $t_r^{(k)}$ to be the mean reproductive time (duration) of the kth generation, and the cumulative time span of n generations as $\tau_n = \sum_{k=1}^{n} t_r^{(k)}$, with $\tau_0 \equiv 0$. Eq. (S5), defining the per-generation selection coefficient, can be adapted to this scenario as:

$$\frac{N_m(\tau_n)}{N_w(\tau_n)} = \frac{N_m(0)}{N_w(0)} \prod_{k=1}^{n} (1 + s_c^{(k)}) \tag{S31}$$

where we have introduced a baseline selection coefficient $s_c^{(k)}$ for the kth generation that can in general vary from each generation to the next. Eq. (S31) implies

$$\begin{aligned} s_c^{(k)} = \frac{N_m(\tau_k)N_w(\tau_{k-1})}{N_w(\tau_k)N_m(\tau_{k-1})} - 1 &\approx \ln\left(\frac{N_m(\tau_k)N_w(\tau_{k-1})}{N_w(\tau_k)N_m(\tau_{k-1})}\right) \\ &= \int_{\tau_{k-1}}^{\tau_k} dt \left(\frac{d}{dt}\ln N_m(t) - \frac{d}{dt}\ln N_w(t)\right) \end{aligned} \tag{S32}$$

where the approximation assumes selection coefficients $|s_c^{(k)}| \ll 1$, the typical case we consider. The integral expression for $s_c^{(k)}$ will be useful for evaluating the baseline selection coefficient later. To make further progress, we will define a general density-dependent growth model, and subsequently illustrate it with several examples.

Let us consider wild type and mutant populations of organisms that depend on a shared resource whose quantity $S(t)$ varies in time. Two examples where this occurs, discussed below, are bacterial populations competing in a chemostat [65–67], and predators competing for the same prey species [60, 62, 64]. The resting metabolic power input has a general functional form $\Pi(m(t); S(t), N_{\text{tot}}(t))$, which depends on the current mass $m(t)$ of the organism, the amount of available resource $S(t)$, and potentially also on the total population $N_{\text{tot}}(t) = N_w(t) + N_m(t)$, i.e if there is a cooperative feeding interaction or other ecological mechanism leading to an Allee effect [59]. Our focus will be on time scales covering many

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

generations, and we will assume changes in $\Pi(m(t); S(t), N_{\text{tot}}(t))$ over a single generation are small enough that we can approximate $\Pi(m(t); S(t), N_{\text{tot}}(t)) \approx \Pi(m(t); S(\tau_{k-1}), N_{\text{tot}}(\tau_{k-1}))$ for $\tau_{k-1} \leq t \leq \tau_k$. Then from Eq. (2) of the main text we can write the reproductive time of the kth generation as

$$t_r^{(k)} \approx \int_{m_0}^{\epsilon m_0} dm \, \frac{E(m)}{\Pi(m; S(\tau_{k-1}), N_{\text{tot}}(\tau_{k-1})) - B(m)m} \tag{S33}$$

for a wild-type organism with synthesis costs $E(m)$ and per-mass maintenance costs $B(m)$. In analogy to the discussion in Sec. I, the corresponding birth rate for the kth generation is $r(S(\tau_{k-1}), N_{\text{tot}}(\tau_{k-1})) = \ln(R_b)/t_r^{(k)}$, which now depends on $S(\tau_{k-1})$ and $N_{\text{tot}}(\tau_{k-1})$. To model the population dynamics over time scales much longer than a generation, we make the usual continuum time approximation, $r(S(\tau_{k-1}), N_{\text{tot}}(\tau_{k-1})) \to r(S(t), N_{\text{tot}}(t))$, and posit a population model for $N_w(t)$ of the form

$$\frac{dN_w}{dt} = r(S(t), N_{\text{tot}}(t))N_w(t) - dN_w(t). \tag{S34}$$

Here d is the rate at which the wild-type population is removed from the system i.e. dilution by outflow of solution in a chemostat, or death. If r was time-independent, the dynamics of $N_w(t)$ described by Eq. (S34) would reduce to Eq. (S2) in Sec. I.

We can derive an analogous equation for the mutant population $N_m(t)$. Under our baseline assumption about the mutant organism, it has modified synthesis costs $\tilde{E}(m) = E(m) + \delta E$ and maintenance costs $\tilde{B}(m) = B(m) + \delta B$, and hence through the analogue of Eq. (S33) it will have mean generation times $\tilde{t}_r^{(k)} = t_r^{(k)} + \delta t_r^{(k)}$ perturbed by $\delta t_r^{(k)}$ relative to the wild type. In the continuum time population model, this translates to a mutant birth rate $\tilde{r}(S(t), N_{\text{tot}}(t)) = r(S(t), N_{\text{tot}}(t)) + \delta r(S(t), N_{\text{tot}}(t))$, modified by some term $\delta r(S(t), N_{\text{tot}}(t))$, and population dynamics governed by

$$\frac{dN_m}{dt} = \tilde{r}(S(t), N_{\text{tot}}(t))N_m(t) - dN_m(t) \tag{S35}$$

Here d is the same as in Eq. (S34), since we assume the mutant is removed at the same rate through dilution or other processes as the wild type. As in the wild type case, for time-independent r the solution of Eq. (S35) for $N_m(t)$ yields the corresponding Sec. I result, Eq. (S3).

The final equation necessary to complete the description of the system is for the resource $S(t)$. To derive it, first consider the relationship between resource consumption and metabolic expenditure. Over the course of generation k, the mean resting metabolic expenditure of a wild type cell per unit time (the average input power) is given by

$$\begin{aligned} P_T^{(k)} &\approx \frac{1}{t_r^{(k)}} \int_{\tau_{k-1}}^{\tau_k} dt \, \Pi(m(t); S(\tau_{k-1}), N_{\text{tot}}(\tau_{k-1})) \\ &= \frac{1}{t_r^{(k)}} \int_{m_0}^{\epsilon m_0} dm \, \frac{E(m)\Pi(m; S(\tau_{k-1}), N_{\text{tot}}(\tau_{k-1}))}{\Pi(m; S(\tau_{k-1}), N_{\text{tot}}(\tau_{k-1})) - B(m)m}. \end{aligned} \tag{S36}$$

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Since this is the generalization of $\langle\Pi\rangle$ from the main text, the corresponding total resting metabolic expenditure $C_T^{(k)}$ in the kth generation is $C_T^{(k)} = \zeta t_r^{(k)} P_T^{(k)}$. The mean amount of resource consumed per unit time scales with the mean input power as $\gamma^{-1} P_T^{(k)}$, where the yield parameter γ sets the conversion rate (flux of resource needed to sustain a certain power input). Since $\gamma^{-1} P_T^{(k)}$ depends on the substrate quantity and total population, let us denote it as a function $\gamma^{-1} P_T^{(k)} \equiv \Gamma(S(\tau_{k-1}), N_{\text{tot}}(\tau_{k-1}))$, the rate of resource consumption per wild type cell. If the analogous rate for the mutant is $\tilde{\Gamma}$, then the equation for $S(t)$ takes the following form in the continuum approximation,

$$\frac{dS}{dt} = F(S(t)) - \Gamma(S(t), N_{\text{tot}}(t)) N_w(t) - \tilde{\Gamma}(S(t), N_{\text{tot}}(t)) N_m(t). \tag{S37}$$

The function $F(S)$ describes the net production rate of the resource.

From Eqs. (S32), (S34)-(S35) we find:

$$\begin{aligned} s_c^{(k)} = \int_{\tau_{k-1}}^{\tau_k} dt\, \delta r(S(t), N_{\text{tot}}(t)) &\approx t_r^{(k)} \delta r(S(\tau_{k-1}), N_{\text{tot}}(\tau_{k-1})) \\ &= \ln(R_b) \frac{\delta r(S(\tau_{k-1}), N_{\text{tot}}(\tau_{k-1}))}{r(S(\tau_{k-1}), N_{\text{tot}}(\tau_{k-1}))} \\ &\equiv \ln(R_b) \tilde{s}_c^{(k)} \end{aligned} \tag{S38}$$

where we have used the fact that $\tau_k - \tau_{k-1} = t_r^{(k)} = \ln(R_b)/r(S(\tau_{k-1}), N_{\text{tot}}(\tau_{k-1}))$, and approximated the integral by assuming $\delta r(S(t), N_{\text{tot}}(t))$ varies slowly within a generation. We thus have a result directly analogous to the $s_c \approx \ln(R_b)\tilde{s}_c$ relation derived in Sec. I, but now due to density dependence the fractional change in growth rate $\tilde{s}_c^{(k)}$ can in general be different for each generation k.

It is interesting to note that there are scenarios where $\tilde{s}_c^{(k)}$ becomes independent of k. One case is when the synthesis cost is assumed independent of the organism's mass, $E(m) = E_m$, and the maintenance term is negligible $B(m) \approx 0$. This would be a rough approximation for fast-dividing organisms like bacteria. Using the fact that $\tilde{s}_c^{(k)}$ can also be written as $\tilde{s}_c^{(k)} = -\delta t_r^{(k)}/t_r^{(k)}$, we can see from Eq. (S33) that a mutant with synthesis cost $E_m + \delta E$ would correspond to $\tilde{s}_c^{(k)} = -\delta E/E_m$. In the notation of the main text, this would mean $\sigma_E = 1$, and σ_B is not defined in this case because maintenance is ignored. Note that the σ_E result here is independent of generation number k, or any specific details of the density dependence.

Another scenario where $\tilde{s}_c^{(k)}$ becomes independent of k is in the long time limit for cases where the dynamical system of Eqs. (S34)-(S35),(S37) converges to a stationary solution with non-zero total population. In other words $N_{\text{tot}}(t)$ and $S(t)$ go to nonzero asymptotic values as $t \to \infty$, which we denote at N_{tot}^s and S^s. This requires that $r_{\max}$, the maximum possible value of $r(S, N_{\text{tot}})$, satisfies $r_{\max} > d$, and in addition $F(S^s) > 0$. Combinining Eqs. (S34)-(S35) we know that

$$\frac{d}{dt} \ln \frac{N_m(t)}{N_w(t)} = \frac{1}{N_m(t)} \frac{dN_m(t)}{dt} - \frac{1}{N_w(t)} \frac{dN_w(t)}{dt} = \delta r(S(t), N_{\text{tot}}(t)). \tag{S39}$$

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

After an initial transient, the right-hand side relaxes to a constant value $\delta r(S^s, N^s_{\rm tot})$. For small δr, the regime of interest in our work, we see that the ratio $N_m(t)/N_w(t)$ changes very slowly at long times. The system has two possible outcomes: for $\delta r > 0$ the ratio $N_m(t)/N_w(t)$ goes to infinity, corresponding to the mutant eventually taking over the whole population, $N_m(t) \to N^s_{\rm tot}$. The stationary state values N^s_w, N^s_m, S^s, satisfy relations that make the left-hand sides of Eqs. (S34)-(S35),(S37) zero:

$$\tilde{r}(S^s, N^s_m) = d, \qquad N^s_m = \frac{F(S^s)}{\tilde{\Gamma}(S^s, N^s_m)}, \qquad N^s_w = 0. \tag{S40}$$

Alternatively when $\delta r < 0$, $N_m(t)/N_w(t)$ goes to zero, with the wild type taking over, $N_w(t) \to N^s_{\rm tot}$. Here the stationary state values satisfy:

$$r(S^s, N^s_w) = d, \qquad N^s_w = \frac{F(S^s)}{\Gamma(S^s, N^s_w)}, \qquad N^s_m = 0. \tag{S41}$$

From Eq. (S38) the asymptotic relaxation dynamics at long times between mutant and wild type populations have a selection coefficient proportional to

$$\tilde{s}^{(k)}_c \approx \frac{\delta r(S^s, N^s_{\rm tot})}{r(S^s, N^s_{\rm tot})} \tag{S42}$$

for large k. Note that Eq. (S42) also applies if we start in the stationary state with a fully wild type population, and subsequently a mutant with small δr appears in the population.

To illustrate the utility of Eq. (S42), it is instructive to consider two model systems:

1) *Bacteria in a chemostat:* The canonical model of a bacterial populations growing in a chemostat [67] assumes resting metabolic power input $\Pi(m; S; N_{\rm tot})$ takes the form $\Pi(m; S; N_{\rm tot}) = \Pi_{\max}(m)S/(K_s + S)$, which depends on a resource S but not explicitly on $N_{\rm tot}$. Here $\Pi_{\max}(m)$ is the power input under unlimited resource, and the second term is a phenomenological hyperbolic function, first posited by Monod [63], that expresses approximately linear dependence of the power on resource for small S, and saturation for S much greater than some threshold value K_s. We assume a simple allometric growth model with exponent $\alpha = 1$ typical of bacteria: $\Pi_{\max}(m) = \Pi_0 m$ for some constant Π_0, $E(m) = E_m$, $B(m) = B_m$. The integrals in Eq. (S33) for $t^{(k)}_r$ and Eq. (S36) for $P^{(k)}_T$ can be evaluated directly, and the resulting birth rate $r(S(t))$ and resource consumption rate $\Gamma(S(t))$ take the form

$$r(S(t)) = \frac{1}{E_m}\left(\Pi_0 \frac{S(t)}{K_s + S(t)} - B_m\right), \qquad \Gamma(S(t)) = \frac{\Pi_0 m_0 S(t)}{\gamma(K_s + S(t))\ln 2}, \tag{S43}$$

where we have taken $R_b = \epsilon = 2$ for simplicity. (Other values of ϵ would just add an overall prefactor.) If B_m is neglected, a common simplification for bacteria, $r(S)$ takes the form known as Monod's law, $r(S) = r_{\max}S/(K_s + S)$ with a maximum rate $r_{\max} = \Pi_0/E_m$. However for completeness we will not ignore B_m, and use Eq. (S43) instead. For typical

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

chemostat experiments, cell death is negligible relative to D, the dilution rate. This is the rate at which the solution (including the resource and bacterial cells) is removed from the system. Hence we set $d = D$. The final piece of the chemostat model is the net resource production, $F(S) = F_0 - DS$, where F_0 represents the incoming resource from a reservoir, and $-DS$ is the output flow of resource due to dilution. The dilution rate in chemostat experiments is typically of the order of $D \sim \mathcal{O}(0.1)$ hr^{-1} [79].

Given these model details, Eqs. (S34)-(S35),(S37) have a a stable stationary solution with $N^s_{\text{tot}} > 0$ in the long time limit if $D < (\Pi_0 - B_m)/E_m$. For concreteness let us take $\delta r < 0$, so the stationary state satisfies Eq. (S41). Using Eq. (S43) to calculate δr for perturbations to E_m and B_m, we plug into Eq. (S42) in the stationary limit to get

$$\tilde{s}_c^{(k)} \approx -\frac{1}{E_m}\delta E_m - \frac{1}{E_m r(S^s)}\delta B_m. \tag{S44}$$

Hence

$$\sigma_E = 1, \qquad \sigma_B = \frac{B_m}{E_m r(S^s)} = \frac{B_m}{E_m D}. \tag{S45}$$

We thus see that the relative maintenance contribution σ_B to the baseline selection coefficient can be tuned by changing D, making the chemostat a versatile experimental tool to explore maintenance versus synthesis cost selection pressures in bacterial populations. As discussed in Sec. III above, empirical data for a broad swath of unicellular organisms [12] yields a global estimate $B_m/E_m = 3 \times 10^{-6}$ s^{-1}. For a typical experimental range $D = 0.05 - 0.5$ hr^{-1}, we then know that σ_B would vary between 0.22 and 0.022.

As noted above, the total resting metabolic expenditure in the kth generation $C_T^{(k)} = \zeta t_r^{(k)} P_T^{(k)}$, which can be written as $C_T^{(k)} \approx \zeta\gamma \ln(R_b)\Gamma(S^s)/r(S^s)$ for large k when the system approaches the stationary state. Using Eq. (S43) to calculate $\delta C_T^{(k)}$ for perturbations to E_m and B_m, we find

$$\frac{\delta C_T^{(k)}}{C_T^{(k)}} \approx -\frac{1}{E_m}\delta E_m - \frac{1}{E_m D}\delta B_m, \tag{S46}$$

and so

$$\sigma'_E = 1, \qquad \sigma'_B = \frac{B_m}{E_m D}. \tag{S47}$$

Comparing Eqs. (S45) and (S47), we see that $\sigma_E = \sigma'_E$, $\sigma_B = \sigma'_B$. Thus the approximation $\tilde{s}_c^{(k)} \approx \delta C_T^{(k)}/C_T^{(k)}$ works well in this stationary limit.

2) *Predator-prey dynamics:* The second example is predator-prey system where the predator wild type organism of population N_w has a resting metabolic rate described by some allometric exponent α, so that $\Pi(m; S) = \Pi_0(S)(m/m_0)^\alpha$, with a function $\Pi_0(S)$ that depends on the population of a prey organism S. For convenience we have explicitly factored out $m_0^{-\alpha}$, in order that $\Pi_0(S)$ have units of W. As in the allometric model of the main text, $E(m) = E_m$, $B(m) = B_m$. From Eq. (S33) the predator reproductive time for $\alpha \neq 1$ is given

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

by:

$$t_r^{(k)} = \frac{E_m}{B_m(1-\alpha)} \ln\left(\frac{\Pi_0(S(t)) - m_0 B_m}{\Pi_0(S(t)) - m_0 \epsilon^{1-\alpha} B_m}\right). \tag{S48}$$

The population growth rate $r(S(\tau_{k-1})) = \ln(R_b)/t_r^{(k)}$. Since the exponent α for metazoans is typically close to 1 (as in the common choice of $\alpha = 3/4$ [23, 24]), we will simplify the expression for $r(S(\tau_{k-1}))$ by expanding it to first order in α around $\alpha = 1$. In the continuum version, where $S(\tau_{k-1}) \to S(t)$, the growth rate then takes the form

$$r(S(t)) \approx \mu \Pi_0(S(t)) - \nu, \tag{S49}$$

where

$$\mu \equiv \frac{[2-(1-\alpha)\epsilon]\ln R_b}{2E_m m_0 \ln \epsilon}, \qquad \nu \equiv \frac{B_m \ln R_b}{E_m \ln \epsilon}. \tag{S50}$$

This approximation works best when $|1-\alpha|\ln\epsilon \ll 2$. The first term in the $r(S(t))$ expression is the contribution to population growth of metabolic power input through prey consumption, while the second term describes the degree to which the growth rate is reduced by having some of that power channeled into maintenance. We approximate Eq. (S36) for $P_T^{(k)}$ to lowest order in $1-\alpha$, giving the following expression for the resource consumption rate,

$$\Gamma(S(t)) \approx \frac{(\epsilon - 1)\Pi_0(S(t))}{\gamma \ln \epsilon}. \tag{S51}$$

Eqs. (S49)-(S51) can be plugged into the dynamical system of Eqs. (S34)-(S35),(S37). To complete the description of the predator-prey system, we make several additional assumptions: we interpret d as the death rate of the predator, and choose a prey reproduction function $F(S) = bS(1-S/K)$, where b is the maximum net rate at which the prey species reproduces itself, K is the prey carrying capacity. Finally we choose $\Pi_0(S) = \Pi_0^{\max} S/(S_c + S)$ for some constants $\Pi_0^{\max}$ and S_c, where $\Pi_0^{\max}$ is the maximum metabolic power input in the limit of unlimited prey abundance, and S_c is the critical prey population above which saturation of $\Pi_0(S)$ occurs. This has the form of a Holling type II predator-prey functional response [61, 62], which assumes that when prey becomes overabundant, the consumption rate is limited by the finite time required for the predator to process each kill, hence leading to saturation of $\Pi_0(S)$. Given the above assumptions, the dynamical system of Eqs. (S34), (S37) reduces to the Rosenzweig-Macarthur (RM) predator-prey model [60, 64] in the absence of a mutant predator population, and with maintenance B_m neglected. Here we consider the generalized RM model with $B_m > 0$ and with populations of both wild type and mutant predators competing for the same prey.

To apply Eq. (S42), we need to establish the properties of the stationary solution of Eqs. (S34)-(S35),(S37), along the lines of the stability analysis for the original RM model in Ref. [60]. For concreteness let us assume $\delta r < 0$, so the stationary solution should satisfy

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Eq. (S41). Then a stationary solution exists with $N^s_{\rm tot} > 0$ if $d < \mu\Pi_0^{\rm max} K(K+S_c)^{-1} - \nu$, and takes the form

$$N^s_w = \frac{b\gamma\mu S^s(K-S^s)\ln\epsilon}{(\epsilon-1)K(d+\nu)}, \qquad N^s_m = 0, \qquad S^s = \frac{S_c(d+\nu)}{\mu\Pi_0^{\rm max} - \nu - d}. \tag{S52}$$

This solution is stable when $S^s > (K-S_c)/2$, which will be the parameter regime on which we focus. When the stationary solution is unstable, the system exhibits limit cycles, which are beyond the scope of this analysis (though in such cases Eq. (S38) would still hold, with an $s_c^{(k)}$ that becomes periodic in k in the long time limit).

Putting everything together, we can use Eqs. (S49)-(S50) to calculate δr for perturbations to E_m and B_m, and we then evaluate Eq. (S42) at the stationary state,

$$\tilde{s}_c^{(k)} \approx -\frac{1}{E_m}\delta E_m - \frac{1}{E_m d}\left(\frac{\ln R_b}{\ln\epsilon}\right)\delta B_m. \tag{S53}$$

Thus in this case,

$$\sigma_E = 1, \qquad \sigma_B = \frac{B_m}{E_m d}\left(\frac{\ln R_b}{\ln\epsilon}\right). \tag{S54}$$

The final result for σ_B has a very similar form to the bacterial chemostat example above, Eq. (S45), except for an additional factor of $\ln R_b/\ln\epsilon$, and d being the predator death rate, rather than a dilution rate. As discussed in Sec. III above, the ratio B_m/E_m is expected to be somewhat different for multicellular versus unicellular species, with for example $B_m/E_m = 10^{-7} - 10^{-6}$ s^{-1} for a range of mammals [24]. The scale of the dimensionless factor $\ln R_b/\ln\epsilon$ can also be estimated. For example, data from a variety of fissiped carnivore and insectivore mammals yields a range $\ln R_b/\ln\epsilon \approx 0.2 - 0.4$ [58]. Given these estimates, we will get $\sigma_B > \sigma_E = 1$ when the mean predator lifetime d^{-1} is on the order of a year or higher. If we do the analogous $C_T^{(k)}$ perturbation analysis as in the previous example, we again find that $\sigma_E = \sigma'_E$, $\sigma_B = \sigma'_B$, validating the approximation $\tilde{s}_c^{(k)} \approx \delta C_T^{(k)}/C_T^{(k)}$ for large k.

VII. Understanding the roles of baseline metabolic costs versus adaptive effects in selection

In the main text our focus has been on understanding the baseline contribution to selection, s_c, arising from changes in metabolic costs (perturbations to synthesis and maintenance). However in the general case a genetic variant will have additional perturbations to its phenotype contributing to the selection coefficient, yielding an overall coefficient $s = s_c + s_a$ with an adaptive correction s_a [12]. In this section we illustrate how our bioenergetic growth formalism can be extended to consider perturbations beyond metabolic costs, yielding expressions for s_a. We show that even with these additional perturbations, s_c retains the form of our original formalism, and can be generally related to associated fractional changes in the total resting metabolic expenditure C_T. However the relationship between s_a and C_T is

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

more complicated, dependent on system-specific details of the organism and its environment. Thus investigating the relative magnitude of s_c and s_a in individual biological cases, and the ways in which s_a might or might not compensate for the metabolic costs encapsulated in s_c, opens a rich set of questions for future study.

We proceed by generalizing the derivations in SI Secs. and to include two examples of adaptive effects for the mutant organisms: changes in the death rate d, and changes in resting metabolic power input $\Pi(m(t))$. These are not the only quantities in the theory that could potentially be subject to adaptation: other possibilities include the the mean number of offspring R_b or the reproductive maturity parameter ϵ. However these two examples illustrate the general scheme for how to include adaptive effects in the formalism. A more comprehensive treatment of the interplay between baseline metabolic costs and adaptive effects in selection is the subject of an upcoming work [51].

In the original derivation of SI Sec. I, where we considered only the baseline selective contribution due to metabolic costs, we assumed the mutant has the same mean number of offspring R_b and death rate $d = -\ln(p(t_r))/t_r$ as the wild type. We now partially relax that assumption: we still keep R_b the same, but allow the mutant death rate $\tilde{d} = -\ln(p(t_r + \delta t_r))/(t_r + \delta t_r)$ to be different than d by some amount $\delta d \equiv \tilde{d} - d$. For example if the survival probability $p(t)$ was exponential for the wild-type, $p(t) = \exp(-\eta t)$, and the rate η was modified in the mutant to $\tilde{\eta}$, then $\delta d = \tilde{\eta} - \eta$. One case where such considerations would come into play would be bacterial strains growing in an environment with antibiotics: a variant that acquired antibiotic resistance would have a lowered death rate, $\delta d < 0$, but also possibly non-trivial metabolic costs associated with maintaining the resistance [52–54], as discussed in more detail below. Since the mutant now has both altered birth rate $\tilde{r} = r + \delta r = \ln(R_b)/(t_r + \delta t_r)$ and death rate $\tilde{d}$, the population growth equation [Eq. (S3)] becomes

$$N_m(t) = e^{(\tilde{r} - \tilde{d})t} N_m(0), \tag{S55}$$

and the ratio of mutant to wild-type populations [Eq. (S4)] is changed to

$$\frac{N_m(t)}{N_w(t)} = \frac{N_m(0)}{N_w(0)} e^{t(\delta r - \delta d)} = \frac{N_m(0)}{N_w(0)} R_b^{t\left(\frac{1}{t_r + \delta t_r} - \frac{1}{t_r}\right)} e^{-t\delta d}. \tag{S56}$$

Since we now have both metabolic costs and adaptive effects, Eq. (S5) now defines the full selection coefficient s instead of just s_c, evaluated after n wild-type generations, $t = nt_r$,

$$\frac{N_m(nt_r)}{N_w(nt_r)} = \frac{N_m(0)}{N_w(0)} (1 + s)^n. \tag{S57}$$

Comparing Eq. (S57) to Eq. (S56), we can write

$$s = R_b^{-\frac{\delta t_r}{t_r + \delta t_r}} e^{-t_r \delta d} - 1 = R_b^{\frac{\tilde{s}}{1 - \tilde{s}}} e^{-t_r \delta d} - 1, \tag{S58}$$

where $\tilde{s} \equiv -\delta t_r / t_r \approx \delta r / r$, with the latter approximation valid when $|\tilde{s}| \ll 1$. Note that in this general case δr (or equivalently δt_r) describes the perturbation to the growth rate

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

both from the altered metabolic costs as well as adaptive effects (as we will see below when we consider changes in $\Pi(m(t))$). We will distinguish the metabolic versus adaptive contributions shortly. If the magnitudes of δd and δt_r are small relative to t_r, Eq. (S58) to first order in the perturbations δd and δt_r is given by

$$s \approx -t_r \delta d + \ln(R_b)\tilde{s}. \tag{S59}$$

To complete the description of s, we now would like to expand out the $\tilde{s}$ contribution on the right-hand side of Eq. (S59) into contributions from metabolic and adaptive perturbations, generalizing the derivation of SI Sec. II for $\tilde{s}_c$. Starting with Eq. (S9) for t_r, we consider how t_r will be altered under the simultaneous perturbations: $E(m) \to E(m) + \delta E$, $B(m) \to B(m) + \delta B$, $\Pi(m) \to \Pi(m) + \delta\Pi(m)$. The first two are the baseline metabolic changes in synthesis and maintenance costs we considered before. The last one represents the possibility that the genetic variant could also have modifications in the resting metabolic power input function $\Pi(m(t))$. As before we assume δE and δB are independent of $m(t)$ for simplicity, but we allow the power input perturbation $\delta\Pi(m(t))$ to have some general $m(t)$-dependent functional form. Eq. (S10) then becomes

$$\begin{aligned} t_r + \delta t_r &= \int_{m_0}^{\epsilon m_0} dm \frac{E(m) + \delta E}{(\Pi(m) + \delta\Pi(m)) - (B(m) + \delta B)m} \\ &\approx t_r + \delta E \int_{m_0}^{\epsilon m_0} dm \frac{1}{G(m)} + \delta B \int_{m_0}^{\epsilon m_0} dm \frac{E(m)}{\Theta(m)G(m)} - \int_{m_0}^{\epsilon m_0} dm \, \frac{E(m)\delta\Pi(m)}{G^2(m)}, \end{aligned} \tag{S60}$$

where we have expanded to first order in the perturbations, and $G(m) = \Pi(m) - B(m)m$, $\Theta(m) = G(m)/m$. As in SI Sec. II, we can rewrite the perturbation terms on the right-hand side of Eq. (S60) in terms of averages over $p(m) = t_r^{-1}E(m)/G(m)$, and get an expression for $\tilde{s}$,

$$\begin{aligned} \tilde{s} &= -\langle E^{-1}\rangle \delta E - \langle \Theta^{-1}\rangle \delta B + \langle G^{-1}\delta\Pi\rangle \\ &= \underbrace{-\sigma_E \frac{\delta E}{\langle E\rangle} - \sigma_B \frac{\delta B}{\langle B\rangle}}_{\tilde{s}_c} + \underbrace{\langle G^{-1}\delta\Pi\rangle}_{\tilde{s}_a}, \end{aligned} \tag{S61}$$

where the synthesis/maintenance prefactors have the same definition as before: $\sigma_E = \langle E\rangle\langle E^{-1}\rangle$, $\sigma_B = \langle B\rangle\langle \Theta^{-1}\rangle$. The first two terms of Eq. (S61) are thus $\tilde{s}_c$ [Eq. (S13)], the baseline contribution to $\tilde{s}$ from perturbations to metabolic synthesis/maintenance costs. The new adaptive contribution $\tilde{s}_a = \langle G^{-1}\delta\Pi\rangle$ comes from the perturbation to the power input, and has a straightforward physical interpretation: since $G(m)$ is the power available for growth (once maintenance costs are subtracted away from the power input), the average $\langle G^{-1}\delta\Pi\rangle$ is just the fractional size of the power input perturbation $\delta\Pi(m)$ relative to $G(m)$, averaged over the course of a generation. Plugging Eq. (S61) into Eq. (S59), we get our final

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

expression for s,

$$s \approx \underbrace{\ln(R_b)\tilde{s}_c}_{s_c} + \underbrace{\ln(R_b)\tilde{s}_a - t_r\delta d}_{s_a}. \tag{S62}$$

The first term is s_c, what we had calculated originally under the baseline assumption, and the remaining terms we can identify as s_a, the correction due to adaptive effects beyond the baseline assumption.

As we did for the baseline case, we can compare the contributions to $\tilde{s}$ in Eq. (S61) from the different perturbations to the associated changes in the total resting metabolic expenditure per generation, $C_T = \zeta t_r \langle \Pi \rangle$. Again proceeding analogously to SI Sec. II, Eqs. (S15)-(S19), we find the fractional change

$$\frac{\delta C_T}{C_T} = \underbrace{\sigma'_E \frac{\delta E}{\langle E \rangle} + \sigma'_B \frac{\delta B}{\langle B \rangle}}_{\equiv \left(\frac{\delta C_T}{C_T}\right)_c} + \underbrace{\langle \Pi \rangle^{-1} \langle \delta \Pi \rangle - \langle \Pi \rangle^{-1} \langle \Pi G^{-1} \delta \Pi \rangle}_{\equiv \left(\frac{\delta C_T}{C_T}\right)_a}, \tag{S63}$$

where we have split $\delta C_T / C_T$ into the baseline cost contribution $(\delta C_T/C_T)_c$ and the adaptive correction $(\delta C_T/C_T)_a$. Here σ'_E, σ'_B have the same definitions as before, given in Eq. (S19). As we argued in the main text, the prefactors σ'_E, σ'_B are comparable to σ_E, σ_B respectively across a wide range of biologically relevant growth scenarios. Thus the maintenance/synthesis contributions to $\tilde{s}$ in Eq. (S61) have a direct reflection in the baseline contribution to $\delta C_T/C_T$ in Eq. (S63). In other words $\tilde{s}_c \approx -(\delta C_T/C_T)_c$, the baseline relation that is the central focus of the main text.

For the adaptive components, the relation between $\tilde{s}_a$ from Eq. (S61) and $(\delta C_T/C_T)_a$ from Eq. (S63) is not so straightforward. Moreover the full adaptive contribution to the selection coefficient, s_a from Eq. (S62), includes a term reflecting the change in death rate, $-t_r\delta d$, which has no analogue in $(\delta C_T/C_T)_a$. As a concrete example, consider the allometric growth model described in the main text, where $\Pi(m(t)) = \Pi_0 m^\alpha(t)$, $E(m(t)) = E_m$, $B(m(t)) = B_m$. Let us focus on the simplest case, $\alpha = 1$, corresponding to exponential cell mass growth, where the baseline relation is exact, $\tilde{s}_c = -(\delta C_T/C_T)_c$ since $\sigma_E = \sigma'_E = 1$ and $\sigma_B = \sigma'_B$. If the power input perturbation occurs through the prefactor, $\Pi_0 \to \Pi_0 + \delta\Pi_0$, then $\delta\Pi(m) = \delta\Pi_0 m$. We can then evaluate both $\tilde{s}_a$ and $(\delta C_T/C_T)_a$:

$$\begin{aligned} \tilde{s}_a &= \langle G^{-1}\delta\Pi \rangle = \frac{\delta\Pi_0}{\Pi_0 - B_m}, \\ \left(\frac{\delta C_T}{C_T}\right)_a &= \langle \Pi \rangle^{-1}\langle \delta\Pi \rangle - \langle \Pi \rangle^{-1}\langle \Pi G^{-1}\delta\Pi \rangle = \frac{\delta\Pi_0}{\Pi_0} - \frac{\delta\Pi_0}{\Pi_0 - B_m}. \end{aligned} \tag{S64}$$

So we see that even in this simple case $\tilde{s}_a$ and $(\delta C_T/C_T)_a$ differ by an extra term in the latter. The significance of this difference depends on the details of the specific system under consideration. A schematic summary of the above results is shown in Fig. S3.

The relative contributions of s_c and s_a to selection in particular biological systems may also depend on aspects of the environment, with s_c or s_a playing a leading role in different

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

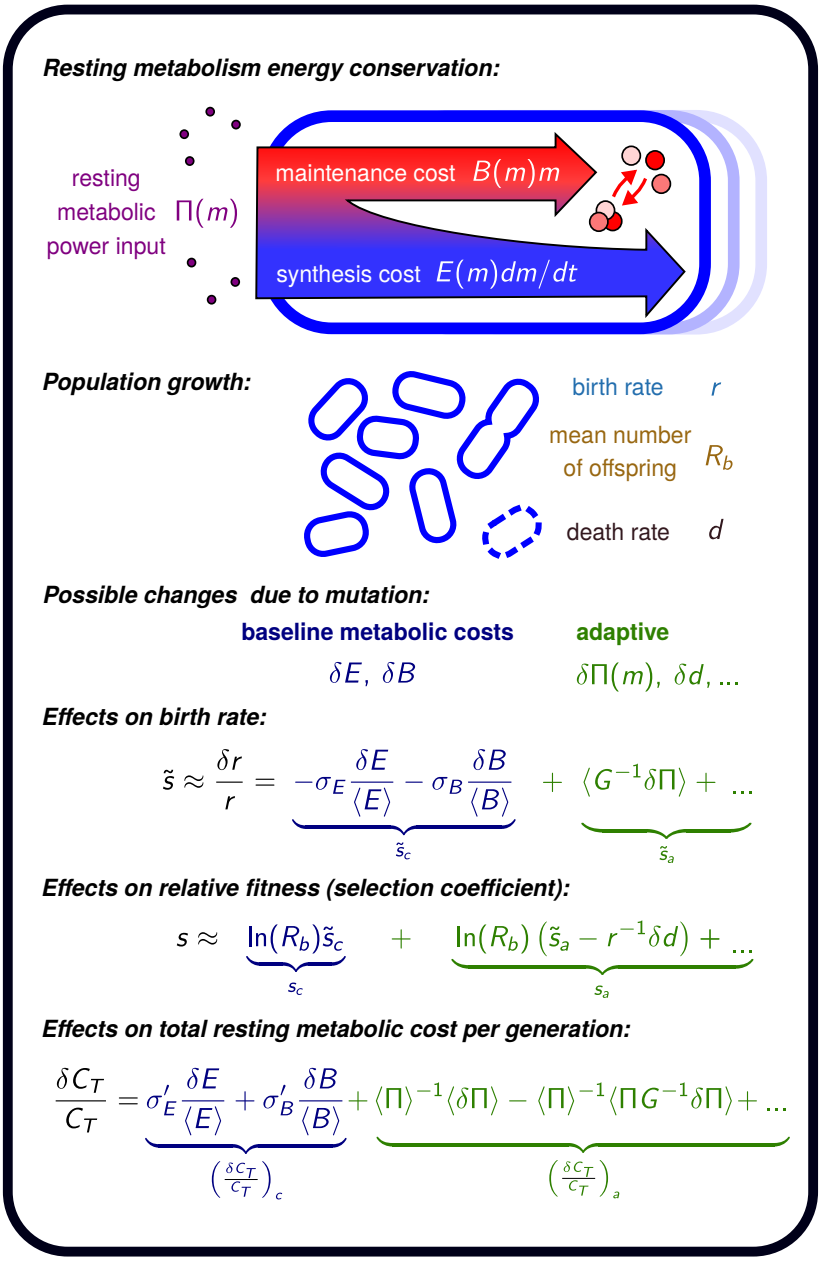

FIG. S3. Schematic summary of the basic theoretical formalism. Energy conservation for individual organisms implies a growth relation for mass $m(t)$: resting metabolic power input $\Pi(m)$ is expended in maintenance $B(m)m$ and synthesis $E(m)dm/dt$ [main text Eq. (1)]. The overall population has a birth rate $r = \ln(R_b)/t_r$ and death rate d, where R_b is the mean number of offspring and t_r is the mean generation time [related to metabolism through main text Eq. (2)]. Mutations can introduce perturbations in the baseline metabolic costs δE, δB as well as adaptive effects like $\delta\Pi(m)$, δd. These have several consequences: i) modifying the generation time by δt_r [SI Eq. 60], which in turn changes the birth rate by δr and leads to a relative fitness difference quantified by the selection coefficient s; ii) modifying the total resting metabolic expenditure C_T over a generation by an amount δC_T. All these changes can be expressed in terms of time averages of metabolic quantities integrated over the wild type generation time t_r, i.e. $\langle F\rangle \equiv t_r^{-1}\int_0^{t_r} dt F(m(t))$. The definitions of $\sigma_E, \sigma_B, \sigma'_E, \sigma'_B$ are given in main text Eqs. (3-4), and $G(m) \equiv \Pi(m) - B(m)m \geq 0$ is the power channeled to growth. The baseline metabolic contributions due to δE and δB we denote with subscript c, and the remaining (adaptive) contributions with subscript a.

bioRxiv preprint doi: https://doi.org/10.1101/358440; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

contexts. For bacterial strains evolving in the presence of an antibiotic, the adaptive benefit of a mutation that confers antibiotic resistance (i.e. lowering the death rate d in the example above) will be reflected in a significant positive fitness contribution $s_a > 0$. But the molecular mechanisms that lead to resistance typically come with non-negligible metabolic costs ($s_c < 0$) that slow bacterial growth in the absence of drug, for example over-expression of energy-consuming multi-drug efflux pumps [50], or switching to less energy-efficient biochemical pathways that are not targeted by the drug [52]. While s_a may more than compensate for s_c when the antibiotic is present, the resistant strains are at least initially at a disadvantage in environments without the antibiotic, where s_c is likely the dominant contribution to s. Ref. [54] conducted meta-analysis of competitive fitness assays between wild type and resistant bacteria in drug-free conditions, focusing on cases where the resistance was conferred by a single mutation. They found mean values of s for mutants resistant to different drugs to range between ≈ -0.03 to -0.28 for eight different bacterial species. Given the large effective populations of bacteria, these values are large enough to put the mutant strains under strong selective pressure to reduce the metabolic price of maintaining resistance. Intriguingly the net result of this pressure under subsequent drug-free evolution is typically not the loss of the resistance, but rather additional mutations that compensate for the metabolic costs [52–54].

Another scenario where environmental conditions can determine the significance of metabolic costs to selection is for systems that exhibit spare respiratory capacity: an energetic reserve that allows them to maintain ATP levels to cope with stress that increases energy demand [42, 43, 45, 46]. While this is particularly relevant for multicellular eukaryotes, it may also be facilitated in prokaryotic populations by large cell-to-cell variations in ATP concentrations [44]. If such spare capacity exists, then under unstressed conditions a mutation with extra baseline metabolic costs may get absorbed by the reserve, with these costs not translating into a penalty in terms of cell growth rate (up to a certain cost threshold). In other words, after factoring in the spare capacity the effective δE and δB of this mutation would have a smaller magnitude (or be zero) compared to the case where no spare capacity existed, and hence the magnitude of s_c would also be smaller. However these costs would not necessarily always remain hidden from selection: if the environment switched to stress conditions with higher energy demands over the long term, these mutants would have a lower reserve capacity to cope. In this case there would be less buffering of the costs, and a larger magnitude of s_c for the mutants.

In summary, quantitative understanding of how s splits into s_c and s_a components, and their relative importance under different conditions, will be very useful in the probing the metabolic influences on evolution in biological systems, and an important target for future work [51].