Measuring the burden of hundreds of BioBricks defines an evolutionary limit on constructability in synthetic biology

Noor Radde[1]*, Genevieve A. Mortensen[1]*, Diya Bhat[1], Shireen Shah[1], Joseph J. Clements[1], Sean P. Leonard[1], Matthew J. McGuffie[1], Dennis M. Mishler[1,2], Jeffrey E. Barrick[1#]

[1]Department of Molecular Biosciences, Center for Systems and Synthetic Biology, The University of Texas at Austin, Austin, TX 78712, USA

[2]The Freshman Research Initiative, College of Natural Sciences, The University of Texas at Austin, Austin, TX 78712, USA

*Equal contributions

[#]Correspondence: jbarrick@cm.utexas.edu

ABSTRACT

Engineered DNA will slow the growth of a host cell if it redirects limiting resources or otherwise interferes with homeostasis. Populations of engineered cells can rapidly become dominated by "escape mutants" that evolve to alleviate this burden by inactivating the intended function. Synthetic biologists working with bacteria rely on genetic parts and devices encoded on plasmids, but the burden of different engineered DNA sequences is rarely characterized. We measured how 301 BioBricks on high-copy plasmids affected the growth rate of *Escherichia coli*. Of these, 59 (19.6%) negatively impacted growth. The burden imposed by engineered DNA is commonly associated with diverting ribosomes or other gene expression factors away from producing endogenous genes that are essential for cellular replication. In line with this expectation, BioBricks exhibiting burden were more likely to contain highly active constitutive promoters and strong ribosome binding sites. By monitoring how much each BioBrick reduced expression of a chromosomal GFP reporter, we found that the burden of most, but not all, BioBricks could be wholly explained by diversion of gene expression resources. Overall, no BioBricks reduced the growth rate of *E. coli* by >45%, which agreed with a population genetic model that predicts such plasmids should be "unclonable" because escape mutants will take over during growth of a bacterial colony or small laboratory culture from a transformed cell. We made this model available as an interactive web tool for synthetic biology education and added our burden measurements to the iGEM Registry descriptions of each BioBrick.

Keywords: evolutionary failure, genetic stability, metabolic burden, Registry of Standard Biological Parts, International Genetically Engineered Machines (iGEM) competition

INTRODUCTION

Synthetic biologists are engineering increasingly sophisticated functions into cells and deploying these "living machines" in new and more challenging environments. For example, cells have been created with genetic circuits that perform complex sensing and logic operations,[1,2] and bacterial symbionts have been engineered to improve the productivity and health of their plant and animal hosts.[3–5] However, unlike computer code, engineered DNA sequences in cells can evolve, potentially making their functions unpredictable and unreliable.[6,7] Evolutionary failure—when less-functional or nonfunctional mutants outcompete their ancestor—can occur rapidly if an engineered function is highly burdensome to a cell or if the sequences that encode it are especially mutation-prone.[8–12] In extreme cases, a population of cells may already become dominated by "escape mutants" that have evolved inactivated variants of a designed sequence after outgrowth of a single transformed cell into a colony or small laboratory culture, making that construct essentially "unclonable". To improve the foundations of bioengineering, we need to better understand why certain DNA constructs are more burdensome to cells than others and the limits on how much burden a cell can tolerate before unwanted evolution becomes a barrier.

Because all engineered DNA constructs must use resources from the cell to replicate and express genes, these processes are the most common and predictable sources of burden.[13] Burden from replicating engineered DNA in cells is typically negligible, even for very high-copy plasmids in bacteria.[14] Instead, transcriptional resources (e.g., RNA polymerases) or translational resources (e.g., ribosomes, charged tRNAs) often become limiting when a foreign DNA construct directs a cell to synthesize RNAs and proteins. Protein overexpression studies in *E. coli* generally find that ribosomes are the most limiting factor, with a proportional decrease in the growth rates of cells as producing more heterologous protein diverts more of their ribosomes away from expressing host proteins needed for replication.[15–19] Usage of gene expression resources can be monitored using high-throughput approaches that globally profile RNA

abundance and ribosomal occupancy[20] or reporter genes with expression levels that reflect the depletion of overall cellular capacities for transcription and translation.[21]

Burden may also arise due to how specific gene products expressed from an engineered DNA construct interact with host cells. Metabolic engineering purposefully funnels precursor molecules toward a target compound by expressing foreign enzymes, altering gene regulation, and/or disrupting native pathways. These modifications will generally slow a cell's growth, and metabolic products or intermediates may also accumulate to levels that are detrimental to cellular physiology.[22–24] Expressing certain types of proteins, such as proteases and integral membrane proteins, is also known to be stressful or toxic to *E. coli* cells, due either directly to their functions or to competition with native proteins for secretion machinery.[25,26] Proteins used for orthogonal control of gene expression, like T7 RNA polymerase and dCas9, can exhibit excessive activity or off-target effects that are extremely burdensome.[27] Finally, unintentional expression of antisense and frameshifted gene products from cryptic promoters and ribosome binding sites has been shown to be an unexpected source of burden in some constructs.[9,20]

Sharing of standardized genetic parts has been a cornerstone of synthetic biology since its inception.[28,29] The Registry of Standard Biological Parts is a database of engineered DNA sequences[30] that thousands of teams have contributed to as part of their participation in the International Genetically Engineered Machines (iGEM) competition.[31,32] Most BioBrick parts are cloned into a small set of standard vector backbones, which makes these plasmids a useful "common garden" for analyzing the properties of inserts encoding different genetic parts and devices. In past studies, BioBricks have been used to compare standardized measurements of promoter strength[33] and fluorescent protein expression[34,35] across many labs. It has been proposed that genetic reliability—in the evolutionary sense of for how many cell doublings a certain level of function is maintained in a population—be listed on a data sheet describing a genetic part,[29] but this property is rarely characterized in practice. One goal of iGEM is to improve upon existing parts, and many BioBrick sequences are re-used by synthetic biology

researchers outside of iGEM. Therefore, characterizing which of these parts are evolutionarily unstable and understanding why this is the case would broadly benefit the field.

We measured the burden of 301 BioBrick plasmids from the iGEM Registry containing DNA constructs ranging from individual parts to complex devices. None of these plasmids reduced the growth rate of their *E. coli* hosts by >45%, in agreement with stochastic simulations of evolution that predict a level of burden above this threshold would make a construct "unclonable". We found that 6 BioBrick plasmids had a burden of >30%, which would be expected to be problematic on the laboratory scale, and that 19 had a burden of >20%, enough that they might fail during process scale-up or in other applications in which cells continue to divide. Several BioBrick plasmids, including two we used as controls, evolved mutations that likely reduce their burden by compromising their designed functions. Finally, we determined that depletion of gene expression resources is sufficient to explain the burden of most BioBrick plasmids, though some reduce host growth rates for other, currently unknown reasons. Our work demonstrates how standardized frameworks for measuring burden and simulating the dynamics of evolutionary failure can be used to improve the reliability of bioengineering.

RESULTS

Model of Evolutionary Failure. Growth of a cell population that has been engineered with a new DNA construct begins from a single transformed cell. As the population divides, progeny with mutations in the sequence of the designed DNA construct will arise. If these mutations alleviate a burden on the cells caused by the engineered DNA—most often by lessening or eliminating a designed function that compromises their growth—then, the mutant cells will have a competitive advantage. These higher-fitness cells will outreplicate and displace ancestral cells with the original DNA construct until they dominate within the population and function declines.

To put our experimental measurements of burden into context, we first investigated the expected timing of evolutionary failure using a differential equation model (**Fig. 1A**). This model

has two parameters. The first is the burden (b) of the engineered DNA, expressed as a percent reduction in the rate of replication of a cell containing the genetic construct. The model makes a simplifying assumption that there is one category of mutations that leads to failure of the engineered function in a way that completely alleviates its burden. The rate of these failure mutations per cell division (μ) is the second parameter. The typical dynamics for this model are that "broken" cells with a failure mutation are initially very rare but then rapidly take over a population as their fitness advantage is exponentially compounded over time (**Fig. 1B**).

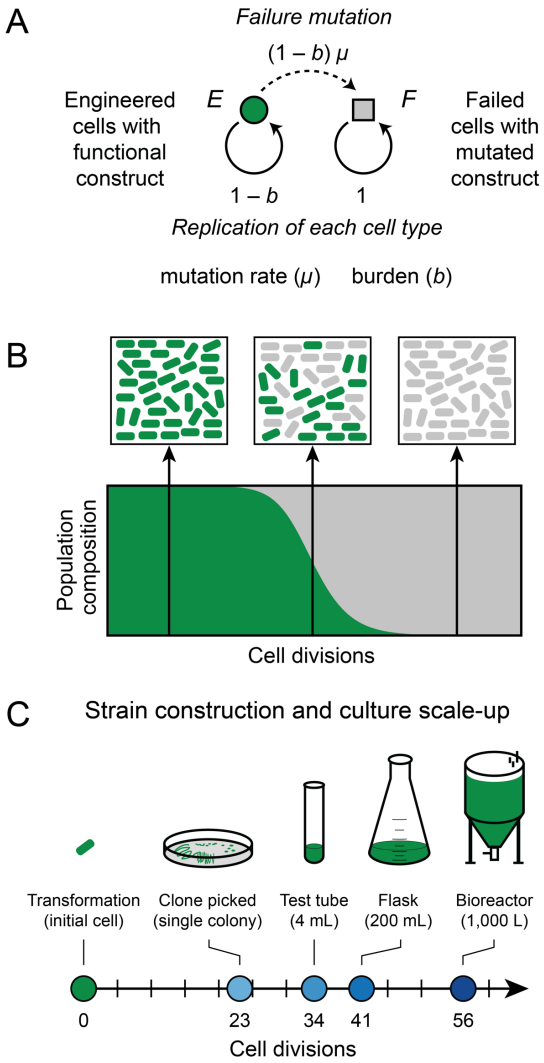


Fig. 1. Evolutionary failure of a population of engineered cells. (**A**) Graphical representation of a differential equation model with one class of failure mutations that completely alleviates the fitness burden of an engineered DNA construct on a host cell. (**B**) Population dynamics expected from this model. Subpopulations of failed cells with mutated constructs evolve and

outcompete the original engineered cells with functional constructs. Complete failure happens rapidly once the mutant cells reach a detectable frequency in the population. (**C**) Approximate numbers of cell divisions required for scale-up from a single engineered cell to laboratory and industrial processes requiring different culture sizes. (See the text and **Table S1** for details.)

We wanted to understand what magnitude of burden would be likely to lead to evolutionary failure of an engineered function during a typical scale-up process starting with a single bacterial cell picked as a colony isolate after transformation with a newly cloned plasmid or some other genome editing procedure (**Fig. 1C, Table S1**). We estimate that ~23 cell divisions occur by the time a single cell produces a normal-sized colony containing ~8 million cells on an agar plate. If this entire colony is placed into ~4 ml of LB in a test tube, it takes an additional ~11 cell divisions to reach saturation, assuming a final density of ~5 $\times 10^9$ cells/ml. Growth to a 200 mL laboratory scale at a higher cell density (e.g., in terrific broth for recombinant protein overexpression) brings the total to ~40 cell divisions. Larger-scale industrial processes can reach even higher cell densities such that ~56 cell divisions may be needed to saturate a 1,000 L bioreactor.

The rates of mutations leading to the failure of different DNA constructs can vary widely, so we tested values of this parameter spanning several orders of magnitude: from 10^{-4} to 10^{-8} per genome per cell division. One factor that plays into the mutation rate is the information content of a sequence, i.e., how many base pairs must be specified to encode its function. Longer engineered DNA sequences and those that are more densely coded are at a greater risk for inactivating mutations.[6,7] The rate of base substitutions in *E. coli* is ~5×10^{-10} per base pair per generation,[36,37] and most microbes with DNA genomes have similar mutation rates.[38,39] Thus, if a sequence contains protein-coding genes that constitute 1000 base pairs and 20% of the substitutions in these genes lead to a loss of function, the failure rate will be ~1×10^{-7} per cell division just from base substitutions. This estimate does not account for the presence of sequence repeats that can act as mutational hotspots that cause specific large deletions and small indels in certain sequence contexts at much higher rates.[40] Furthermore, selfish elements in the host genome usually contribute other types of mutations that further increase the total rate

of failure mutations. In particular, transposon insertions often inactivate genes or sequences required for gene expression in engineered DNA constructs.[11,41,42]

In the end, empirical measurements generally find a rate of ~10^{-6} per cell division for mutations that inactivate a single-gene that is located in the chromosome of *E. coli* or another bacterium.[41,43] The effective mutation rate is much higher for engineered constructs maintained on multicopy plasmids because each copy of the plasmid in a cell is at risk. If there are 100 copies of a plasmid, the chance of a plasmid with a certain mutation arising is ~100-fold higher. So, for example, the rate of reverting a stop codon in an engineered reporter construct, which is expected to only occur via one or a few single base substitutions, has been measured as ~10^{-7} rather than the value of ~10^{-9} expected if this reporter were tested in the chromosome[12]. For plasmids that lack partitioning systems like pBR322 and pUC derivatives commonly used in *E. coli*, one broken plasmid copy can rapidly lead to 100% failure of all plasmids in all cells in a population because progeny that happen to inherit more broken plasmid copies due to random segregation will outcompete those that do not. In summary, the effective rate of failure mutations in a high-copy plasmid is usually much higher than the point mutation rate; it is expected to be at least on the order of 10^{-5} and often as high as 10^{-4} per cell doubling. Though mutational hotspots and multicopy plasmid replication are not explicitly accounted for in our model, they justify exploring simulations with a wide range of mutation rates.

Previous studies of escape mutations have used the deterministic results of ordinary differential equation (ODE) models to estimate the times to failure of engineered cells.[6,11] This framework assumes that mutants appear continuously and immediately at the beginning of the simulation. However, in reality, mutations appear stochastically in single cells at very low rates, and the dynamics can vary greatly depending on whether these “jackpots” occur early or late in the growth of a population. Therefore, we compared the deterministic results for our ODE model to stochastic simulations of this model to evaluate how and when the results varied. We found that deterministic simulations consistently overestimate how unstable a construct will be for a

given combination of parameters (**Fig. 2**). The discrepancy becomes larger at lower mutation rates where it mainly reflects the waiting time needed for the rare event that generates the first mutant cell to appear in a population in the stochastic simulations, compared to the immediate appearance of these mutants in the deterministic simulations. However, there are also occasional stochastic simulation runs in which failure occurs sooner than it does in the deterministic model due to early jackpots (as seen in the panel for b = 20%, $\mu = 10^{-8}$).

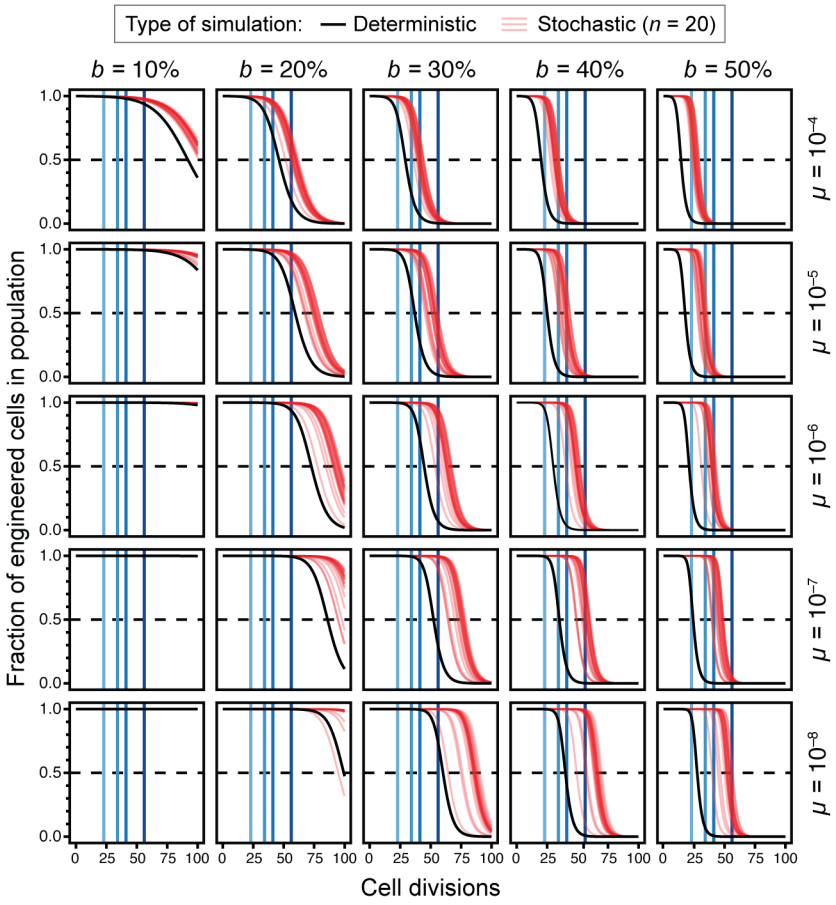


Fig. 2. Simulations of evolutionary failure times for populations of engineered cells. In each panel, the results for deterministic (black) and stochastic (red) simulations of the failure model are shown for one combination of burden (b) and failure mutation rate (μ) parameters. Vertical blue lines represent the culture scales shown in **Figure 1C**. Curves for stochastic simulations are partially transparent so that one appears pink and overlapping trajectories from multiple simulations appear red. Twenty stochastic simulations are displayed in each panel.

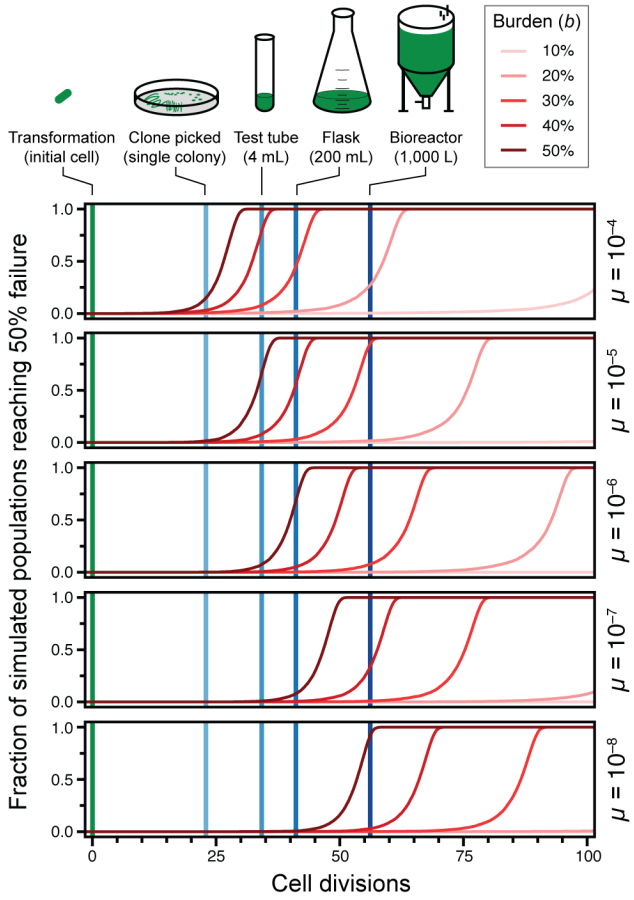


Fig. 3. Cumulative distributions of times to 50% failure in stochastic simulations. Curves represent the output from 10,000 replicate simulations for each parameter combination.

Because we expect it to better represent the true evolutionary dynamics, we further examined the results of the stochastic simulations (**Fig. 3**). They show that at a typical mutation rate of 10^{-5} per cell doubling (expected for a plasmid-borne construct) a burden of ≥50% would lead to takeover of broken mutants in a test tube culture most of the time. At a mutation rate of 10^{-4}, constructs with a burden of ≥40% would not survive on this small scale. Since one needs to grow a single transformed cell into a culture of this size to purify and sequence a plasmid to verify that it has the designed sequence, the model predicts that constructs this burdensome will be essentially "unclonable". Even for less-burdensome plasmids or for constructs experiencing lower mutation rates (for example, single-copy genes in the chromosome), the model predicts that failure may occur at larger scales if the burden reaches the 20-30% range.

We created an online version of our model that allows users to adjust the burden and failure mutation rate parameters (https://barricklab.org/burden-model). There is an option to use the stochastic or deterministic version of the model and compare the results. Additionally, users can change the effective volume and density of their culture to understand the scale at which a DNA construct with certain characteristics is likely to fail. This interactivity encourages users to explore a range of parameters and to rerun simulations multiple times to see for themselves the sizable impact of mutational stochasticity on the continuing functioning of devices constructed in living, and therefore evolving cells. We believe that this resource will be useful for educating both new and practicing synthetic biologists, as this type of random and self-reinforcing failure can be confusing and does not have a direct parallel in traditional engineering fields.

Burden of BioBrick Parts. To test whether actual engineered DNA sequences obey the evolutionary constraints predicted by our model of escape mutations, we examined a diverse collection of engineered DNA sequences created for the iGEM (International Genetically Engineered Machine) competition.[31] These BioBricks range in complexity from small DNA "parts", such as promoters and protein tags, to larger "devices" that consist of multiple genes and operons. Historically, BioBricks in the Registry of Standard Biological Parts had to be cloned into plasmids in ways that allowed them to be combined into larger constructs using a specific assembly standard.[44] As a consequence, most BioBricks in the kit distributed to iGEM teams are provided in plasmids pSB1C3, pSB1A2, or in both of these backbones (**Fig. 4A**). pSB1C3 and pSB1A2 share the same high-copy pUC origin of replication and overall organization, but they are maintained using different antibiotic resistance genes: chloramphenicol acetyltransferase (*cat*) which confers chloramphenicol resistance (Cam^R) for pSB1C3 versus β-lactamase (*bla*) which confers resistance to ampicillin and other β-lactams (Amp^R) for pSB1A2. These plasmids also differ in how expression of the cloned BioBrick part is insulated from elements in the plasmid backbone. pSB1A2 has a transcriptional terminator

upstream of the BioBrick prefix multiple cloning site. pSB1C3 has a terminator at the same site and an additional terminator downstream of the BioBrick suffix multiple cloning site.

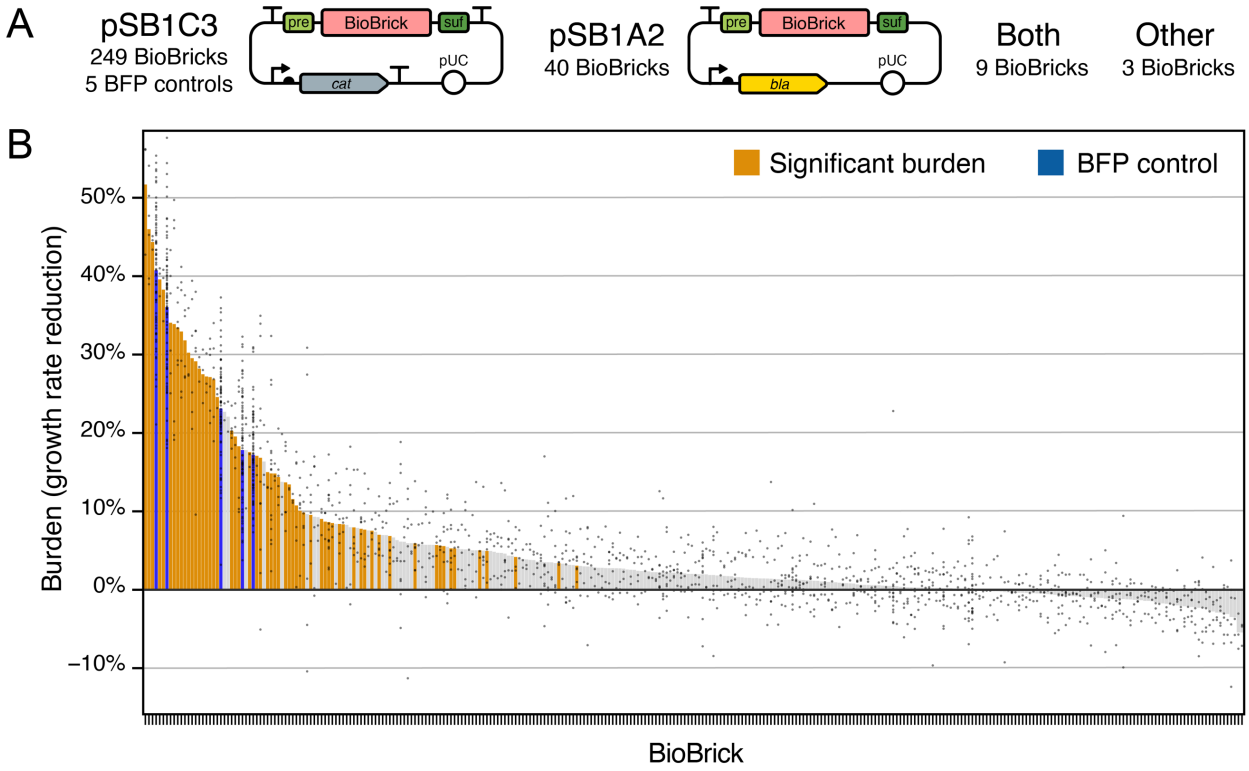


Fig. 4. Measurements of BioBrick burden. (**A**) Maps of the two plasmid backbones that housed most of the 301 BioBricks that were tested and the five BFP controls that were included in every assay. The prefix (pre) and suffix (suf) multiple cloning sites used in BioBrick assembly are shown. (**B**) Burden of each BioBrick tested. Burden is the percentage reduction in the growth rate of *E. coli* cells transformed with a BioBrick plasmid. Gray points are individual measurements. Bars are the means for all measurements of a BioBrick. For BioBricks with orange bars, the measured burden was significantly greater than zero (adjusted $p < 0.05$, one-tailed *t*-tests with Benjamini-Hochberg correction for multiple testing). Data used to create this figure are provided in **Table S2** and **Table S3**.

We measured the growth rates of *E. coli* DH10B derived cells transformed with BioBrick plasmids to determine how many of these genetic parts and devices were burdensome and to what extent. In each microplate assay, we included 5 pSB1C3-based BioBrick plasmids we constructed with different promoter and ribosome binding site combinations driving expression of blue fluorescent protein (BFP). These plasmids cause different amounts of burden and served as internal controls. We normalized growth rates between assays to account for plate-to-

plate variation based on results for the BFP controls and an additional assumption that most parts in each microplate would exhibit no burden (**Fig. S1**, **Table S2**, and **Methods**).

In total, we measured the effects of the 5 BFP control plasmids and 301 other BioBricks on *E. coli* growth (**Fig. 4B, Table S3**). Of the 301 BioBricks we characterized, we tested 249 in pSB1C3, 40 in pSB1A2, 9 in both of these plasmid backbones, and 3 housed in other similar backbones (pSB1AK3 or pSB3C5). Even though different antibiotics were added to growth media when testing BioBricks cloned into pSB1C3 and pSB1A2, there was not a significant effect of the plasmid backbone on the growth rates measured for the 9 parts tested in both plasmids ($p = 0.069$, $F_{1,56} = 3.44$, two-way ANOVA) (**Fig. S2A**). We also did not find evidence for any overall difference in the distributions of growth rates measured for parts tested in pSB1C3 versus the other three backbones ($p = 0.92$, two-sided Kolmogorov-Smirnov test) (**Fig. S2B**). Therefore, we considered all of our measurements together, irrespective of the plasmid backbone in which a BioBrick part was tested, in all further analyses.

Excluding the five BFP control plasmids, which were all burdensome, 112 of the 301 other BioBrick part plasmids (37.2% of those tested) significantly decreased *E. coli* growth rates relative to the majority of parts that had no burden before correcting for multiple testing (individual one-tailed *t*-tests, $p < 0.05$). For 31 BioBricks the growth rate burden was significantly greater than 10%, for 19 it was significantly greater than 20%, and for 6 it was significantly greater than 30% (one-tailed *t*-tests, $p < 0.05$). In agreement with our population genetic model, none of the BioBrick plasmids had a large enough burden (>45%) that they would be predicted to mutate when growing a small test-tube culture in the laboratory (one-tailed *t*-tests, $p < 0.05$). After accounting for multiple testing using the Benjamini-Hochberg procedure at a 5% false discovery rate (FDR), we can conclude that 59 of the 301 tested BioBrick parts (19.6%) exhibit some level of burden with high confidence (one-tailed *t*-tests, adjusted $p < 0.05$). **Table** 1 lists the 34 BioBricks that met this criterion and had a mean estimated burden of >10%.

Table 1. Most burdensome BioBricks

BioBrick‡	Seq&	Backbone	Burden (*b*)†	Fraction other burden (*b*o/*b*)*	Subparts¶	Function#
K523022	M	pSB1C3	51.7 ± 19.2%	*n.s.*	P_{lac} *&lacZ' crtE crtI crtB*	Carotenoid synthesis (*Pantoea ananatis*)
K733010	C	pSB1C3	46.0 ± 6.2%	0.16–0.71	P_{tms} *&endB*	Antitoxin gene (*Bacillus subtilis*)^B
J04450	NS	pSB1AK3	44.4 ± 2.2%	*n.s.*	P_{lac} *&mRFP1*	RFP reporter
K523014	C	pSB1C3	39.6 ± 4.7%	1.04–1.98	P_{lac} *&lacZ' bglX*	Cellobiose degradation
K523020	M, E	pSB1C3	38.3 ± 8.7%	*n.s.*	P_{lac} *&lacZ' INP+bglX*	Cellobiose degradation (INP, *Pseudomonas syringae*)
K608010	C	pSB1C3	34.1 ± 7.7%	NT	P_{J23110} *&GFP*	GFP reporter
K515100	C	pSB1C3	33.9 ± 15.9%	0.26–0.88	P_{veg2} *&IaaM &IaaH*	Indoleacetamide synthesis (*Pseudomonas savastanoi*)^B
J61000	m	pSB1A2	33.4 ± 4.1%	0.21–0.96	P_{cat} *&cat*	Chloramphenicol resistance
K541526	C	pSB1C3	32.9 ± 7.8%	*n.s.*	P_{veg} *&reflectin1A*	Reflectin reporter (*Euprymna scolopes*)^B
K592020	m	pSB1C3	31.8 ± 5.0%	NT	P_{fixK2} *&cI(λ)* P_{cI} *&amilCP*	Blue light sensor output (*Acropora millepora*)
J36335	m	pSB1C3	30.2 ± 12.2%	*n.s.*	P_{lac} *&kaiA* P_{lac} *&kaiC*	Circadian rhythm (*Synechococcus elongatus*)
I759017	C	pSB1C3	29.5 ± 8.3%	NT	P_{tet} [cis5] *&YFP*	YFP reporter
K346000	C	pSB1C3	29.1 ± 10.4%	*n.s.*	*&RNAP(T3)*	Phage RNA polymerase (Phage T3)
C0056	C	pSB1A2	28.2 ± 3.9%	*n.s.*	*cI434(λ)*	Mutant phage repressor (Phage λ)
K880005	C	pSB1C3	27.5 ± 8.6%	*n.s.*	P_{J23100} *&*	Gene expression
C0053	NS	pSB1C3	27.2 ± 6.5%	*n.s.*	*cII(P22)*	Phage repressor (Phage P22)
K608012	C	pSB1C3	27.1 ± 4.7%	NT	P_{J23110} *&GFP*	GFP reporter
I759014	C	pSB1C3	26.8 ± 5.8%	n.s.	P_{tet} [cis2] *&YFP*	YFP reporter
K541502	C	pSB1C3	24.6 ± 3.2%	0.42–1.91	P_{veg} *&lipA*$_{sig}$	Gene expression/secretion (*Bacillus subtilis*)^B
K395602	C	pSB1C3	20.3 ± 1.9%	0.09–0.38	P_{T7} *&MpAAT1*	Apple fragrance generator (*Malus pumila*)
K733013	C	pSB1C3	19.5 ± 3.3%	n.s.	P_{veg} &GFP	GFP reporter^B
K523013	C	pSB1C3	18.3 ± 8.8%	NT	P_{lac} *&lacZ' INP+EYFP*	EYFP reporter (INP, *Pseudomonas syringae*)
I761014	C	pSB1C3	17.5 ± 5.0%	0.21–1.33	*&cinR &cinI*	Quorum sensing (*Rhizobium leguminosarum*)
C0051	NS	pSB1C3	17.1 ± 8.2%	n.s.	*λ-cI+LVA*	Phage repressor (Phage λ)
K137018	C	pSB1C3	16.8 ± 8.2%	NT	$P_{L\text{-}lacO1}$ *&luxR* $P_{lux\text{-}R}$ *&GFP*	Quorum sensing receiver (*Aliivibrio fischeri*)
K1149051	C	pSB1C3	15.0 ± 8.4%	n.s.	P_{J23104} *&phaC1 phaA phaB1*	Polyhydroxybutyrate synthesis (*Ralstonia eutropha*)
K731721	C	pSB1C3	14.8 ± 4.4%	n.s.		Transcription terminator (Phage T7)
K639003	m	pSB1C3	14.8 ± 2.8%	n.s.	$P_{rmB\text{-}P1}$ *&lacI* $P_{L\text{-}lacO1}$ *&mCherry*	Stress sensor
K541501	C	pSB1C3	14.4 ± 3.6%	n.s.	P_{veg} &sacB$_{sig}$	Gene expression/secretion (*Bacillus subtilis*)^B
K608011	C	pSB1C3	13.7 ± 5.4%	NT	P_{J23110} *&GFP*	GFP reporter
K861172	NS	pSB1C3	13.4 ± 2.5%	n.s.	P_{cstA} *&cI(λ)*	Phage repressor (Phage λ)
K617004	C	pSB1C3	11.6 ± 1.5%	0.95–2.32	*attP(λ)* P'OP	Phage attachment site (Phage λ)
K325218	m	pSB1C3	10.8 ± 7.3%	0.76–1.55	P_{araC} *&luc(orange)*	Luciferase reporter (*Luciola cruciata*)
I712669	m	pSB1C3	10.1 ± 4.5%	NT	P_{CMV} *GFP*	GFP reporter^M

‡BioBrick accession numbers. The 34 parts shown all had an estimated burden that was significantly greater than zero after correcting for multiple testing and had a mean estimated burden value of >10%. &Results of sequencing the BioBrick plasmid: C, reported BioBrick sequence was confirmed; M, major discrepancies found in BioBrick sequence; m, minor discrepancies found in BioBrick sequence; NS, not sequenced; E, part is reported to have errors in the iGEM Registry. Full sequencing results are provided in **Table S4**. †Burden as the percentage reduction in growth rate caused by the BioBrick ± estimated 95% confidence limits. *95% confidence interval on the fraction of burden from sources other than utilization of the host cell's gene expression capacity. *n.s.*, value was not significantly greater than zero. NT, not tested because the BioBrick contains a protein that interferes with measurement of GFP fluorescence. ¶Representation of gene expression signals and genes in the BioBrick abbreviated as follows: P_x, promoter from gene or operon *x*; &, ribosome binding site; [y] other regulatory sequence. Other italicized entries are gene names. #General description of the designed function of the BioBrick. For BioBricks that contain recombinant DNA encoding genes other than fluorescent proteins, the organism of origin is shown in parentheses. Superscript *B or M*, indicates that the gene expression sequences are intended to function in *Bacillus subtilis* or mammalian cells, respectively.

BioBricks containing gene expression parts are more likely to be burdensome. Only BioBricks that express an RNA or protein product are expected to appreciably burden a host cell, as the cost of replicating plasmid DNA is generally negligible in comparison.[14] Therefore, we hypothesized that the 59 BioBricks in the high-confidence burden set would be more likely than BioBricks that had no significant burden to contain strong gene expression signals. Series of constitutive promoter parts (J23100–J23119) and ribosome binding site (RBS) parts (B0030, B0032, and B0034) with known relative strengths are commonly reused in different BioBricks.

These promoters and RBS sequences can be divided into weak, medium, and strong variants on the basis of experimental data reported in the iGEM Registry (**Fig. 5A,B**).[45]

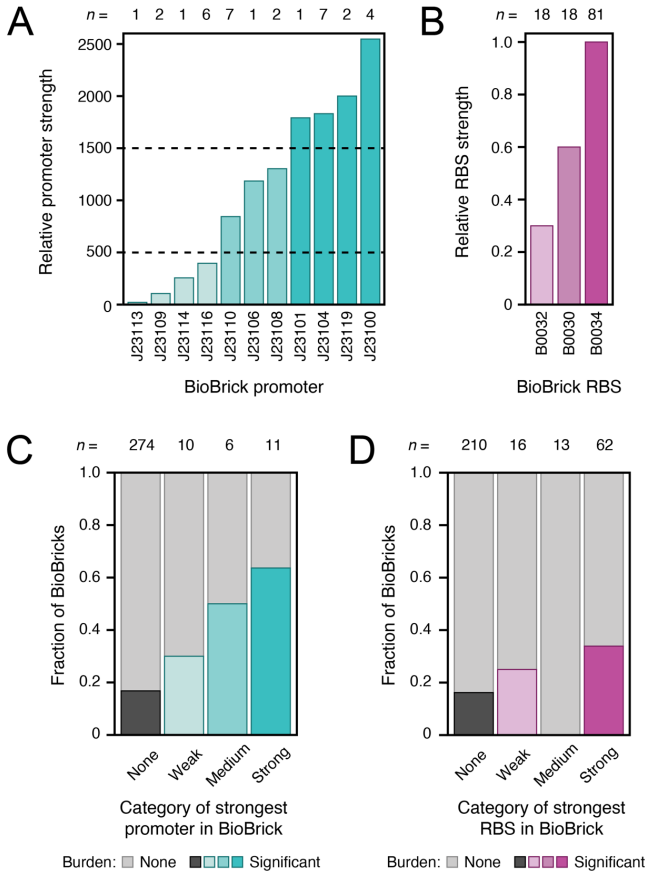


Fig. 5. Strong promoter and ribosome binding sites are more likely to be found in BioBricks exhibiting significant burden. (**A**, **B**) Relative strengths of common promoters and ribosome binding site (RBS) BioBrick parts, as reported in the iGEM Registry. The numbers of examples of each promoter or RBS in the 301 BioBricks examined in this study are indicated above the bars (*n*). Some of these BioBricks contain multiple instances of these promoter and RBS parts. Dashed lines in **A** are the thresholds used to classify promoters as weak, medium, or strong. (**C**, **D**) Fraction of BioBricks tested that exhibited significant burden when grouped by the strongest gene expression element of each type that they contain. The total numbers of parts in each category are shown above the bars (*n*).

We examined whether BioBricks that exhibited burden were more likely to include these common gene expression parts than those that were not burdensome (**Fig. 5C,D**). BioBricks that contained any of these constitutive promoters were 2.9 times as likely to be in the set of 59 BioBricks with significant burden compared to those that did not have one of these promoters (p = 0.00040, Fisher's exact test), with a trend that the stronger promoters were even more likely to

be associated with burdensome BioBricks. Similarly, BioBricks that included the strongest of the three RBS parts (B0034) were 2.1 times as likely to exhibit significant burden as BioBricks that included only the two weaker RBS variants or none of the RBS sequences in this series (p = 0.0037, Fisher's exact test). None of the BioBricks that contained the medium-strength RBS also had a constitutive promoter part, which can explain why this category noticeably deviated from the general trends. Overall, these results agree with the general expectation that strong, constitutive gene expression contributes to the burden of many BioBricks.

One case that stood out in examining these results was BioBrick K880005. It includes the strongest constitutive promoter (J23100) and RBS (B0034) from these sets, but it does not include a downstream open-reading frame. Nevertheless, K880005 is among the most burdensome BioBricks that we measured: it reduces the growth rate of *E. coli* by 27.5 ± 8.6% (95% confidence interval) (**Table 1**). The high burden of this BioBrick may put it at risk of mutating during laboratory propagation, even at the test-tube scale (**Fig. 3**). Its unexpected burden could result from transcription and/or translation of sequences downstream of the part in the BioBrick suffix sequence and plasmid backbone, even though it was tested in the pSB1C3 backbone that has transcriptional terminators designed to insulate the BioBrick.

Mutations and variability in strains with BioBrick plasmids support a burden limit on constructability. To validate the identity and integrity of the plasmids we tested, we compared whole-plasmid sequencing data for 215 BioBricks plus the 5 BFP controls to the sequences reported in the iGEM Registry (**Table S4**, and **Methods**). Excluding the controls, we sequenced 214 of the 301 BioBricks for which we had burden measurements (71.1%). Of these, 8 plasmids were initially misassigned to the wrong BioBrick and 3 others to the wrong backbone in our results before we corrected them. For 185 of the 215 sequenced plasmids (86.0%), our results perfectly matched the expected BioBrick sequences. Of the 30 others, we found relatively minor discrepancies between the sequencing data and the reported BioBrick sequences for 23, and the other 7 had major discrepancies, such as large deletions or transposon insertions.

It is not possible to determine with 100% certainty whether these discrepancies are due to errors in the designed part sequences that were submitted to the iGEM Registry or mutations that arose and took over cell populations because they reduced BioBrick burden. Most discrepancies are single-base changes or deletions that may have no effect on genetic part function. However, in the seven cases of major discrepancies we can be reasonably sure that we have observed unplanned mutations with consequences. Two BioBricks (S03749, I759016) were inactivated by insertion sequence (IS) elements that must have transposed into their sequences after construction. Two BioBricks that were closely related to the second of these (I759019, I759020) had frameshifting or large deletions. Two other parts related to one another (K523020, K523022) also contained large deletions, and the first of these was marked as "believed to contain major errors" in the iGEM Registry. Finally, most of BioBrick I732920 was deleted, and its sequence marked as "inconsistent" in iGEM Registry.

Two of the BFP control BioBrick plasmids, which our own iGEM team constructed and submitted to the iGEM Registry, demonstrate that there is a real risk of selecting cells that have mutated copies of highly burdensome plasmids soon after they are created. We noticed that there was a discrepancy in the order of the growth rates of strains carrying these plasmids in our burden assays: the two control plasmids designed to have the strongest combinations of promoters and ribosome-binding sites driving BFP expression unexpectedly exhibited the least burden. Re-testing the frozen cell stocks of the original transformants of these plasmids demonstrated that the derived stocks used in the burden assays had picked up mutations that largely alleviated the burden of these two plasmids (**Fig. S3**). The burden was reduced from 45.8% to 17.8% in one case and 41.9% to 17.2% in the other. Further supporting the instability of the two most burdensome BFP control plasmids, when we shared them with another iGEM team, they found an insertion of an IS*5* element occurred in the promoter driving BFP expression in their transformant, which reduced but did not eliminate fluorescence.

Even if the original cell giving rise to a colony that is picked after transforming a plasmid or after resteaking a stock has only intact copies of a plasmid, it may give rise to a heterogeneous population of descendant cells as it is cultured, stored as a frozen stock, and revived. As our simulations show, more burdensome plasmids will be at a greater risk of having newly evolved mutants begin to take over the population during these steps. If this type of stochastic, partial takeover of a cell population with mutants was occurring during our experiments, more burdensome BioBricks might exhibit greater variability in their measured growth rates between replicate cultures. In agreement with this hypothesis, we found a significant trend toward a higher standard error of the mean for growth rates measured for BioBrick plasmids that had higher burden ($p = 2.0\times10^{-11}$, two-tailed *t*-test for a non-zero slope) (**Fig. S4**).

In summary two lines of evidence support that "clonability" or "constructability" limits for engineered DNA are creating an upper bound on what plasmids are possible to construct and measure that might be causing us to underestimate the burden of some BioBrick designs. First, our BFP control plasmids designed to have the strongest gene expression mutated during construction and some of the BioBrick plasmids we characterized also sustained mutations that likely reduce their burden. Second, we see more variation in our measurements of growth rates for more burdensome BioBricks, which could be at least partially explained by cells with mutations that reduce plasmid burden arising and beginning to take over during our assays.

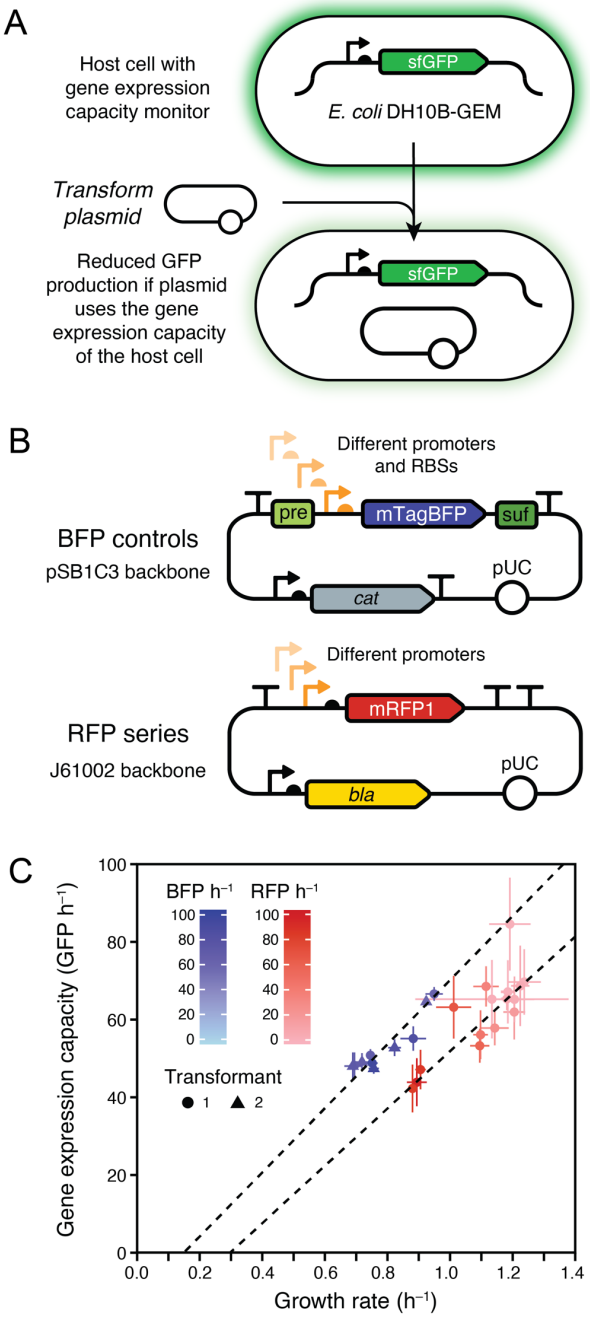
Redirecting gene expression capacity to recombinant protein production causes a proportional reduction in growth rate. The *E. coli* DH10B-GEM strain that we used as a host for testing BioBrick burden has a constitutively expressed GFP gene integrated into its chromosome (**Fig. 6A**). This GFP can be used to monitor how much the presence of a BioBrick plasmid reduces the capacity of an *E. coli* cell for expressing its native proteins.[21,46] If the main source of burden from a plasmid is due to its use of any cellular resources or machinery that are necessary to achieve translation of proteins (e.g., ribosomes), then one expects that for a given reduction in GFP expression there will be a proportional reduction in growth rate. If there is a

reduction in growth rate that is larger than expected relative to the reduction in GFP expression that is observed, then some or all of the burden comes from other sources. For example, gene products encoded on the plasmid may lead to depleting a cellular resource that is not directly related to gene expression or have a toxic effect that interferes with homeostasis.

To establish that the monitoring device worked as expected, we initially tested two series of plasmids that express other fluorescent proteins (FPs) at varying levels (**Fig. 6B**). The first was our set of 5 burdensome BFP control plasmids that have different promoter and RBS combinations. Here we used stocks of cells with the BFP plasmids that did not contain the mutations that alleviated burden noted above. The second set consisted of 14 plasmids available from the iGEM Registry that contain constitutive promoters of different strengths driving expression of RFP. These RFP constructs were not included in the prior tests of BioBrick burden because they are housed in a different plasmid backbone (J61002). In both cases, we expected that all of the burden exhibited by these plasmids would be due to recombinant FP expression depleting the translational capacity of the host cell. FP production does not use any other types of limiting cellular resources, and these FPs are not expected to be toxic to cells within the range of concentrations at which they are expressed.

In agreement with this expectation, we found that the growth rates of these strains were reduced in proportion to how much they reduced GFP expression (**Fig. 6C, Table S5, Table S6**). The Pearson correlation coefficients for this linear relationship were 0.93 and 0.81 for the BFP and RFP plasmid series, respectively. The relationship between growth rate and GFP expression differed slightly between the BFP and RFP series, but this was expected because they have different plasmid backbones and were tested under different culture conditions (see **Methods**). The growth rate reductions seen for RFP series plasmids were roughly in proportion to the amount of recombinant protein that they expressed. By contrast, strains with BFP series plasmids that experienced more gene expression burden did not necessarily produce more BFP. This discrepancy is likely related to how different combinations of promoter and RBS

420 strengths can lead to translating the same amount of protein but with more or less efficient use
421 of ribosomes.[21] As for the 301 BioBricks we tested and the unmutated BFP controls, none of the
422 RFP expression constructs had a burden of >45% in the "unclonable" range.
423



424

425 **Fig. 6. Expression of recombinant proteins from a plasmid reduces the growth rate of *E.***
426 ***coli* because it diverts some of its capacity for gene expression.** (**A**) *E. coli* DH10B-GEM
427 host strain with the gene expression capacity monitoring device that constitutively expresses

GFP integrated into its chromosome. (**B**) Maps for the BFP and RFP plasmid series. (**C**) Growth rates and fluorescent protein production rates for different BFP and RFP plasmids in *E. coli* DH10B-GEM. Dashed lines are Deming regressions showing that the reduction in growth rate is proportional to the reduction in the capacity of the host cell for protein expression within each set of strains. The rate of GFP production from the monitoring device is used as a readout of gene expression capacity. Rates of BFP and RFP production in cells with each type of plasmid are indicated by shading in the respective color. Error bars are 95% confidence limits. Two independent transformants of each BFP plasmid that were tested separately are displayed as points with different shapes. GFP and BFP production rates were measured on different relative scales and each series uses a different vector backbone and was measured under different growth conditions, so results should only be compared within each series. Data used to create this figure are provided in **Table S5** and **Table S6**.

Some BioBricks exhibit burden from sources other than gene expression. All of our measurements of BioBrick burden were conducted in the *E. coli* DH10B-GEM host strain that contained the GFP gene expression capacity monitor (**Fig. 6A**), so we next examined how GFP production correlated with the previously characterized growth rates to understand whether the burden of each BioBrick could be attributed partly or wholly to its use of the host cell's gene expression resources. If GFP production was reduced in direct proportion to the growth rate, as it was in the BFP control plasmids, this would indicate that all of the BioBrick burden was from gene expression (**Fig. 7A**). If there was significant burden with no or less-than-the-expected reduction in GFP production, then it would indicate a BioBrick was compromising *E. coli* growth for some other reason (**Fig. 7B**). Of the 301 BioBricks tested, 42 encode GFP or another protein that is expected to interfere with measuring GFP fluorescence, so they were excluded from this analysis (see **Methods**). We again used the BFP plasmids as internal controls for normalizing GFP production rates between different microplate assays (**Fig. S5** and **Methods**).

Plotting a linear relationship between the BFP plasmid controls, the no-burden BioBrick plasmids, and the origin yields the expected trade-off between growth rate and GFP production for the BFP plasmids and some of the measured BioBrick plasmids (**Fig. 7C**). However, some parts displayed a higher GFP production rate than what would be expected from the measured growth rate reduction, evidence that some or all of their burden arises for reasons other than diverting the host cell's gene expression resources. Of the 26 BioBrick parts with high-

confidence predictions of burden >10% that could be evaluated in this assay, 9 (34.6%) had a significantly greater reduction in growth rate than predicted from the change in GFP production (adjusted $p < 0.05$, one-tailed *t*-tests with Benjamini-Hochberg correction for multiple testing), indicating that a component of their burden is due to a source other than reducing the gene expression capacity of the host cell (**Table 1**).

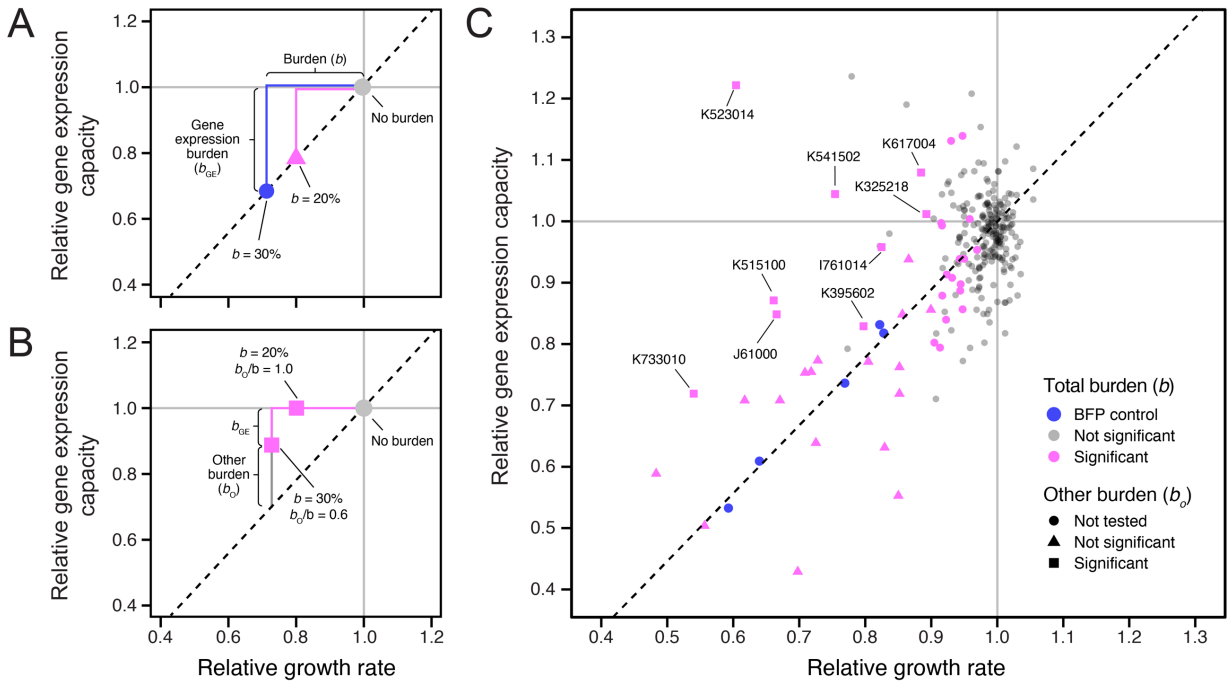


Fig. 7. Some BioBricks exhibit burden from sources other than gene expression. (**A**) Examples of expected results for two BioBricks that exhibit burden (*b*) that is wholly due to utilizing the gene expression capacity of the host cell. The reduction in growth rate is proportional to the reduction in GFP production according to a linear relationship (dashed line) that is established from measurements of control strains. (**B**) Examples of expected results for two BioBricks that exhibit burden from sources other than gene expression. (**C**) Results of measuring growth rates and GFP production rates for 259 BioBricks that do not contain fluorescent proteins that are expected to interfere with measuring GFP fluorescence in the *E. coli* host strain containing the gene expression capacity monitor. Points for each BioBrick are colored based on whether there was significant burden (reduction in growth rate). Symbols indicate whether the null hypothesis that all burden was due to utilizing the gene expression capacity of the host cell could be rejected. BioBricks with significant burden from sources other than gene expression are labeled with their accession numbers. Estimates of b_O/b for these BioBricks are shown in **Table 1**.

DISCUSSION

By measuring the burden of 301 BioBricks and performing simulations, we established an evolutionary limit on the constructability of engineered DNA sequences: none of the BioBricks we tested slowed *E. coli* growth rates by >45%. Our results are in broad agreement with other studies that have made similar measurements of growth defects and the effects of spontaneous mutations that alleviate the burden of engineered DNA on bacterial cells.[47,11] For example, researchers testing a library of plasmids expressing three fluorescent proteins found that a mutant that deleted one of these genes and took over populations after 30 generations of serial transfer had an 89% higher exponential growth rate compared to the original engineered strain,[10] which corresponds to this mutation reducing burden by 47%. Similarly, the level of burden under non-inducing conditions topped out in the 40-60% range for cells containing various constructs in the study that developed the gene expression capacity monitor we used.[21]

We found potential mutations in some BioBricks relative to their designed sequences and more variation in our measurements of more burdensome BioBricks. We also discovered that two of the BioBricks we used as internal controls for our assays unexpectedly mutated while we were using them in ways that maintained some BFP fluorescence yet reduced their burdens from near the unclonable threshold (>40%) down to levels that can be reliably maintained during growth on a laboratory scale (<20%). These results suggest that we may be underestimating the burden of some BioBrick designs, either because their plasmids were mutated before we obtained them or because new mutants arose and reached appreciable frequencies in our assays. Some discrepancies are likely due to human errors in the sequences digitally submitted to the Registry versus the original DNA samples themselves. For example, researchers might have copied over a portion of a sequence from a prior plasmid map or part entry and assumed it was correct and unchanged without ever empirically validating their construct. However, there is also both direct and anecdotal evidence that some Biobricks are prone to mutate.

One such example of evolutionary instability is for the exceptionally well-characterized BioBrick F2620 device. F2620 encodes a luciferase gene that is expressed in response to the quorum sensing molecule acyl homoserine lactone.[29] It was not one of the BioBricks we tested. F2620 was noted to reproducibly fail due to deletions between two 143-bp repeats introduced by re-use of the B0015 double terminator part. When induced, device function declined between 56 and 74 cell doublings and was entirely lost after 92. The creators originally hypothesized that failure was due to pre-existing mutant plasmid copies in their cell populations, but the instability persisted even when they re-transformed the plasmid, confirming that it was due to evolution fueled by *de novo* mutations. Our model shows how you can get deterministic-seeming failures like this if the mutation rate is sufficiently high, as it can be for repeat-mediated deletions.[40]

Few BioBricks have been characterized to the same extent as F2620. We discovered inactivating deletions or transposon insertions in seven of the BioBrick plasmids, which likely indicates that that they are also especially prone to mutational failure. As an example, the Registry page for BioBrick K523020—one of the most burdensome plasmids that we measured—contains a warning, "Part submitted to Registry is believed to contain major errors," which is probably more typical of how a user of an unstable part would understand rapid evolutionary failure due to mutations that are relieving burden. Future work could clarify whether the cases of sequence discrepancies we encountered are already mutated BioBricks, especially unstable BioBricks, or design errors by reverting the putative mutations to the designed sequences and, if successful (i.e., the change does not make them so burdensome that they are unclonable), measuring their burden. Alternatively, deep-sequencing populations of plasmids isolated from laboratory-scale cultures could be used to characterize whether they consist of mixtures of mutated and unmutated plasmids.[11,48] Surveys of plasmids in other repositories have also found that some acquire inactivating transposon insertions.[49]

The GFP gene expression monitor that we used responds to changes in a cell's global capacity for protein expression. For any one construct, this could theoretically represent

depletion of factors as diverse as the availability of RNA polymerases, ribosomes, initiation factors, charged tRNAs, amino acids, or nucleotides. However, we expect that ribosome availability is the limiting factor in all or nearly all BioBricks we tested, based on studies of recombinant gene overexpression in *E. coli*.[15–19] While we were able to establish overall trends that plasmids containing strong constitutive promoters and ribosome-binding sites had a higher chance of exhibiting burden, it was not possible to predict the gene expression component of burden *a priori* on this set of sequences. Hopefully, ongoing improvements in tools for predicting transcription and translation initiation rates trained on expanding databases of high-throughput gene expression measurements[50,51] will make this possible in *E. coli* and other organisms.

Burden can also arise for diverse reasons other than gene expression: anytime engineered DNA taxes a cellular resource to the extent that it becomes a bottleneck for cell growth. For example, genetic engineering can overwhelm protein export pathways or the capacities of different subcellular compartments.[25,26] Further case studies of the burdensome plasmids with costs not associated with gene expression could reveal the origins of these costs. It would be particularly useful to create other types of burden monitors, e.g. of protein secretion, membrane occupancy,[52] or different metabolic bottlenecks so that the relevant limiting factors could be rapidly diagnosed and systems redesigned accordingly to make them more stable. This more refined information will likely be needed to predict how the burden of a composite part or device depends on the burden of each of the genetic parts from which it is constructed. If multiple components use gene expression resources, then one might expect them to have additive effects on burden, but if they use orthogonal (i.e., distinct) limiting resources, then one may find that the combination is no more burdensome than the more burdensome of the two on its own.

We measured burden as a decrease in the exponential growth rate of *E. coli* host cells. While this was convenient for making replicated, high-throughput measurements in a microplate reader, it does not fully reflect how a DNA construct impacts the evolutionary fitness of a cell. For example, it is possible that engineering a cell changes the lag time before growth begins,[53]

559 survival during stationary phase, colony growth on agar, or survival of cryopreservation.
560 Furthermore, our approach can only be applied to understand genetic stability under laboratory
561 conditions, not in environmental contexts or host-associated microbiomes. Co-culture
562 competition assays between a strain of interest and a reference strain could be used to
563 measure fitness in a way that captures all components of fitness in any environment.[54] To make
564 these measurement high-throughput, host strains with unique sequence barcodes in their
565 chromosomes and transformed with different engineered plasmids or DNA constructs could be
566 simultaneously competed all-against-one-another in bulk competitive fitness assays.[55,56]
567 Researchers can take actions to improve the constructability and stability of especially
568 burdensome engineered DNA sequences. Most obviously, using low- or medium-copy plasmids
569 rather than high-copy ones or integrating constructs into the chromosome of a bacterium to
570 make them single-copy will often reduce burden into the cloneable and stable ranges.[10]
571 Systems have also been engineered for controlling plasmid copy number, so that DNA parts
572 can be maintained in cells at a low copy number and then amplified on demand.[14,57] Similarly,
573 reducing the burden of a construct can be achieved by altering promoter and ribosome-binding
574 site strengths or by using inducible promoters, as long as these changes are compatible with
575 device function.[10,21] Systems that regulate expression in response to the growth rate of a cell[58,59]
576 or that couple continued functioning of the engineered DNA to cell survival[60] can more directly
577 buffer against evolutionary failure. Another category of more ambitious approaches is to
578 introduce orthogonal polymerases[61] or ribosomes[62,63] into a cell to prevent synthetic constructs
579 from competing with native gene expression, though the requirement that a cell produce the
580 necessary machinery may itself be burdensome. Next, aspects of the growth environment can
581 sometimes be changed. For example, supplementing media with vitamins or altering salt
582 concentrations has been reported to stabilize certain constructs.[11,22] A final category of
583 approaches seeks to reduce the chances of mutations to improve the evolutionary stability of

584 genetic constructs.[7,64] For example, cells with lower mutation rates can be created by deleting or
585 repressing transposons [9,65] or by altering cellular processes that affect point mutation rates.[12]
586 We created an interactive model of failure mutations in a cell population that can be used to
587 explore how tuning mutation rates and construct burden affect whether a DNA construct is likely
588 to remain intact cell populations that are grown to typical different laboratory and production
589 scales. Similar deterministic[6,11] and stochastic[66] models have been developed by others. Models
590 that include individual steps in gene expression and RNA and protein degradation are also
591 beginning to be used to examine evolutionary stability.[21,67] Our model and these others still do
592 not consider or fully take into account several complications. First, rather than one category of
593 mutation leading to complete failure, there are typically multiple categories of mutations, some
594 of which only partially alleviate the burden, occurring at different rates in real systems.[10,11]
595 Equally important, plasmids are multi-copy within cells so the fitness benefit of a mutation can
596 take several generations to fully manifest and depends on how plasmids segregate between
597 daughter cells. These intricacies of plasmid evolution have been tackled by a variety of more
598 complex models that could be applied to engineered plasmids.[68] Finally, models that take into
599 account different phases of cellular growth could be used to further refine these dynamics.[69]
600 Improving our understanding of what types of synthetic DNA constructs exhibit different
601 types of burden and modeling the effects on the reliability and predictability of cellular function
602 over time is important for realizing synthetic biology applications. Researchers designing
603 engineered cells should be aware of when they are nearing a danger zone of evolutionary
604 stability where DNA designs may become unconstructable, and they should recognize that the
605 stochastic nature of evolutionary failure may lead to large variation in their experimental results,
606 failure during process scale-up, or loss of function when cells are deployed for long periods of
607 time in complex environments outside of the lab, such as in animal and plant microbiomes. Our
608 simulations and results will contribute to spreading this awareness and achieving these goals.

609 The main conclusion can be summarized as a rule of thumb: to avoid unwanted evolution of
610 engineered microbes at a laboratory scale, do not burden their growth by more than ~30%.
611
612 **METHODS**
613 **Model of evolutionary failure.** We implemented a model in R that is similar to one used by
614 Rugbjerg *et al.* to predict loss of production from an engineered cell population due to escape
615 mutations.[11] We parameterized our model such that failed (i.e., mutated) cells, *F*, have a relative
616 growth rate of one. Engineered cells, *E*, have a growth rate that is this value minus the burden,
617 *b*, of the engineered construct. The corresponding equations for how the numbers of engineered
618 cells, $E(t)$, and failed cells, $F(t)$, change over time are:

619 $$\frac{dE(t)}{dt} = (1-b)E(t) - \mu\,(1-b)E(t) \qquad (1)$$

620 $$\frac{dF(t)}{dt} = F(t) + \mu\,(1-b)E(t) \qquad (2)$$

621 Growth of cells in batch culture typically continues until a certain number of total cell doublings
622 occurs that exhausts the provided resources rather than for a certain fixed period of time.
623 Therefore, we chose to plot the dynamics of engineered and failed cell populations versus the
624 number of cell doublings, $D(t)$, that have occurred at a given time:

625 $$D(t) = \log_2\,[E(t) + F(t)] \qquad (3)$$

626 For stochastic simulations of this model, we used the adaptivetau R package.[70] We also created
627 an online version (https://barricklab.org/shiny/burden-model) that can perform deterministic and
628 stochastic simulations of this model using the Shiny R package.[71]
629 **Media and growth conditions.** *E. coli* was cultured at 37 °C in Lysogeny Broth (LB) (10 g
630 tryptone, 5 g yeast extract, 10 g NaCl per liter) with 16 g/L agar added for solid media. Unless
631 otherwise indicated, liquid cultures were grown in 18 mm × 150 mm glass test tubes with orbital
632 shaking at 200 r.p.m over a 1-inch diameter. Antibiotics were added at the following
633 concentrations: carbenicillin (100 µg/ml), chloramphenicol (20 µg/ml), kanamycin (50 µg/ml).

Gene expression monitor strain construction. *E. coli* DH10B-GEM (JEB1203), the host strain used in the burden assays, was created using plasmids and methods described in Haldimann *et al.*[72] and Ceroni *et al.*[21] Briefly, we inserted the constitutive GFP expression cassette cloned into pAH63 (Addgene #66073) into the *E. coli* chromosome at the λ integration site by electroporating this plasmid into DH10B cells containing the helper plasmid pInt-ts (Addgene #66076) and selecting for kanamycin resistant colonies. pAH63 has a *pir*-dependent R6K origin, so it does not replicate in the recipient cells. pInt-ts has a pSC101ts origin and was cured by screening colonies after further growth at the restrictive temperature of 42 °C to create DH10B-GEM. We also obtained and characterized *E. coli* DH10GFP (Addgene #109392), a strain constructed in the same way in the prior study of burden by Ceroni *et al.*[21]

We isolated genomic DNA from cultures of DH10B-GEM and DH10GFP using a PureLink Genomic DNA Mini Kit (Invitrogen). Then, we prepared Illumina libraries using 10 µg of DNA as input into a 2S Turbo DNA Library kit (Swift Biosciences) using 50% reaction volumes and a final PCR step with custom adapters that added dual 6-bp sample barcodes. Sequencing was carried out on a HiSeq X Ten by Psomagen. Reads were compared to *E. coli* DH10B genome (GenBank: NC_010473) and pAH63 plasmid sequences using *breseq*.[73,74] Split-read mappings (new junction evidence) between plasmid and chromosomal sequences verified that the GFP cassette was integrated at the expected site in both strains. There were two shared differences, a single base insertion in an intergenic region and a synonymous base substitution, between both strains and the DH10B reference genome. DH10GFP also had two additional mutations, a nonsynonymous mutation in *uspF* and an IS*4* element insertion in *mdtL*.

Transformation of BioBrick plasmids. We made DH10B-GEM competent cells as follows. A 10 ml liquid culture of cells was grown overnight in a 50 mL Erlenmeyer flask from an aliquot of the glycerol stock. The entire culture was then added to 500 ml of LB in a 2 L Erlenmeyer flask. This culture was incubated until reaching mid-exponential phase (an OD600 between 0.4 and 0.6). At this point, it was divided into 35 ml aliquots and centrifuged at room temperature for

10 minutes at 3400 × g. Then, the supernatant was removed and all cell pellets were combined by resuspended (via vortexing) in a total of 150 ml of a 10% (v/v) glycerol + 100 mM $CaCl_2$ solution chilled on ice. Next, 30 ml fractions of the cells were centrifuged again at room temperature for 10 minutes at 3400 × g. Again, the pellets were combined, resuspending in a total of 20 ml of chilled glycerol-$CaCl_2$ this time. After incubating this mixture on ice for 25 min, 200 µl aliquots were snap frozen in liquid nitrogen. Competent cells were stored at –80°C.

Heat shock was used to transform BioBrick plasmids into DH10B-GEM. This transformation method entailed transferring 2 µl of a miniprep of the plasmid of interest into 50 µl of competent cells and incubating on ice for 1 hour. After this, the mixture was placed in a 42°C heat bath for 30 seconds and then immediately placed back on ice for another 30 minutes. Next, we added 950 µl of SOC media and incubated at 37°C in a shaker incubator for at least an hour. After SOC recovery, we pelleted the cells and decanted 800 µl of the supernatant. We resuspended the pellet in the remaining 200 µl of supernatant and then plated this onto an LB agar plate with the appropriate antibiotic. After overnight incubation at 37°C, we picked a colony, grew an overnight culture in liquid LB media, added glycerol to 15% (v/v), and froze a stock at –80°C.

BFP plasmid construction. Five control plasmids expressing different levels of mTagBFP were created by assembling BioBrick parts from the iGEM registry. The mTagBFP sequence was from part plasmid K592100. It was combined with five promoter+RBS composite parts (K608002, K608003, K608004, K608006, and K608007), by using each of their pSB1C3 part plasmids as the vector backbone in a separate postfixing BioBrick assembly reaction.[44,75] For cloning, we used enzymes from New England Biolabs under standard conditions. Briefly, K592100 was double digested using XbaI and SpeI restriction enzymes in CutSmart buffer. Separately, each of the vector backbones was double digested using SpeI and PstI-HF restriction enzymes in CutSmart buffer followed by incubation with calf intestinal alkaline phosphatase for 1 h. Digested products were then gel extracted and purified using a QIAquick Gel Extraction Kit before being ligated together using T4 DNA ligase. Ligated products were

purified using butanol precipitation and then electroporated into competent TOP10 *E. coli* cells. Transformed cells were recovered in SOC for 1 hour at 37°C, followed by plating on LB agar containing chloramphenicol. After incubation at 37°C for 18 hours, we inoculated isolated colonies into fresh LB liquid media containing chloramphenicol and grew these cultures at 37°C for 18 hours. The five resulting composite BioBrick parts were deposited in the iGEM Registry as K3174002, K3174003, K3174004, K3174006, and K3174007.

Plasmid sequencing. We sequenced BioBrick plasmids isolated from the DH10B-GEM cell stocks that were used for burden assays. In addition, we sequenced plasmids isolated from the TOP10 cell stocks into which the BFP controls were first transformed. Plasmid DNA was purified using a QIAprep Spin Miniprep Kit (QIAGEN) or a PureLink Quick Plasmid Miniprep Kit (Invitrogen). We performed Sanger sequencing on multiple stocks of the BFP control plasmids, in-house Illumina sequencing on these and the other plasmid samples, and outsourced Nanopore sequencing on additional plasmid samples. For Illumina sequencing, up to 10 ng of plasmid DNA was used as input for sequencing library preparation using the 2S Turbo DNA Library kit (Swift Biosciences) with 20% reaction volumes. Custom adapters containing dual 6-bp sample barcodes were incorporated during the final PCR step. The resulting DNA libraries were pooled and sequenced on an iSeq 100 instrument. Nanopore data was obtained from Plasmidsaurus. Porechop[76] and fastp[77] were were used to trim adaptors from sequencing reads.

To analyze sequencing results, we first reconstructed the expected BioBrick plasmid sequences from information available on the iGEM Registry webpages (part sequences, vector sequences, and compatibility with different assembly standards). Then, we analyzed Illumina and Nanopore sequencing data in two ways. First, we compared reads to the expected plasmid sequences using *breseq*[73] to see if there were any discrepancies. Second, we performed *de novo* assembly of reads using either Unicycler[78] or flye,[79] annotated the resulting assemblies with pLannotate,[80] and examined them for matches to the expected parts using blastn searches[81] against a database of all BioBrick parts included in the 2018 iGEM distribution kit.

BioBrick plasmid burden assays. We performed burden assays largely as described previously.[21] Strains were revived by adding aliquots of −80 °C freezer stocks to test tubes containing LB with the antibiotic for maintaining their respective BioBrick plasmids. After overnight growth (12-18 h), we vortexed each culture for three seconds and loaded 5 µl into a Nunc MicroWell 96-well optical-bottom plate (ThermoScientific Cat. No. 265301) in triplicate. Every plate included the five control strains (JEB1204-1208), each also loaded in 5 µl in triplicate, and 12 blank wells (LB only). This arrangement allowed for a total of 23 strains to be tested per plate. To start the assay, a multichannel pipette was used to add 195 µl of LB pre-warmed to 37°C to every well with pipetting up and down several times to mix. Using a Tecan Infinite Pro M200 Plate Reader, optical density at 600 nm and GFP fluorescence (excitation: 485 nm; emission 528 nm) were recorded every 10 minutes with 7 minutes of orbital shaking during each cycle. Each plate was run for a minimum of 6 hours.

RFP and BFP plasmid burden assays. For the series of plasmids expressing RFP under control of different promoters, we performed burden assays using the normal procedure plus an additional measurement of RFP fluorescence (excitation: 585 nm; emission: 610 nm). For correlating BFP expression in the control strains to reduced GFP expression, we added a measurement of BFP fluorescence (excitation: 405 nm; emission: 453 nm). The extra fluorescence reads for the RFP and BFP experiments reduced the proportion of shaking time in each measurement cycle, resulting in slower maximum growth rates than were observed with the standard burden assay procedure. RFP samples were measured every 10 minutes with 6.5 minutes of shaking during each cycle. BFP samples were measured every 10 minutes with 7 minutes of shaking during each cycle. For the RFP series we also monitored cell density using OD660 instead of OD600 to avoid interference from RFP absorbance.[82]

Burden analysis. To analyze the burden assay data for one plate, we first subtracted the average values of all media blanks from the OD and fluorescence measurements. Next, to deal with well-to-well variation in background levels, we shifted the values to force the means of the

738 points over the first hour of measurements for each strain to match the grand mean for those
739 data points over all replicates of that strain. We then fit growth rates using nonlinear least-
740 squares regression to an exponential model: $C(t) = C_0\, e^{rt}$. We assumed that OD is directly
741 proportional to the number of cells at a given time, $C(t)$. C_0 is the initial number of cells, and r is
742 the specific growth rate. We fit C_0 and r for all sets of nine consecutive measurements (a 90-
743 minute window in the standard assay) after the OD exceeded 0.03 and recorded the largest
744 value of r as the maximum specific growth rate for that strain. To determine the fluorescent
745 protein (e.g., GFP) production rate per cell, p, we repeated this procedure while fitting
746 fluorescence values to the equation: $F(t) = F_0 + C_0\,(p/r)\,(e^{rt} - 1)$. F_0 is the initial fluorescence and
747 F(t) is the fluorescence at time t. This equation is derived by integrating the relationship $dF/dt =$
748 $p\,C(t)$. We fit F_0 and p in this model to the data while keeping C_0 and r fixed to the values
749 determined from the OD curve fit for the corresponding time window. Again, we recorded the
750 largest value of p across all time points as the maximum fluorescent protein production rate.
751 To account for plate-to-plate variation in growth and GFP production rate estimates (**Fig.**
752 **S1A, S3A**), we normalized measurements made on different plates. In our experimental design
753 a majority of the plasmids tested in each plate are expected to exhibit negligible burden. This let
754 us estimate the growth and GFP production rates corresponding to 'no-burden' for a given plate
755 by examining the distributions of values measured. Specifically, we calculated the density
756 distributions of growth and GFP production rates using a Gaussian kernel function with
757 bandwidths of 0.014 and 300, respectively, for all non-control strains. To account for multimodal
758 distributions, we took the no-burden value as the highest value among all peaks in the density
759 distribution that were at least 50% as high as the highest peak. Then, we normalized all rate
760 estimates by dividing them by the corresponding no-burden value for that plate (**Fig. S1B, S3B**).
761 The final distributions of the mean values for each BioBrick plasmid have a major peak at the
762 no-burden value with a noticeable shoulder of strains with a slightly decreased growth rate or
763 GFP production rate, in addition to some strains with much lower values (**Fig. S1C**, **S3C**).

Some BioBricks encode proteins that interfere with measuring GFP fluorescence. Therefore, for the analysis of gene expression capacity and burden, we disregarded all BioBricks described as including GFP; YFP, which has overlapping fluorescence; or the amilCP blue chromoprotein, which strongly absorbs at the wavelength monitored for GFP emission.[83] For the 26 remaining BioBricks that also had growth rate reductions that were statistically significant and mean estimated burdens ≥10%, we determined whether the observed GFP production rate was compatible with the null hypothesis that all of the burden was due to the BioBrick utilizing the gene expression capacity of the host cells. We determined the expected relationship between growth rate and GFP production rate for purely gene expression burden from measurements of the BFP control plasmids across all plates. Specifically, we used Deming regression to fit this linear relationship, which takes into account measurement errors in both dimensions, and we further required that the fit pass through the no-burden values (i.e., a normalized growth rate of 1.0 and normalized GFP production rate of 1.0). Then, we determined the chance that each BioBrick was located above the BFP regression using a two-dimensional probability distribution of each assuming maximum likelihood *t*-distributions for growth rate and GFP production rate. We took one-half of this value to estimate a one-tailed *p*-value for the hypothesis that there was significant burden for the test plasmid from a source other than utilization of the host cell's gene expression resources.

SUPPORTING INFORMATION

Fig. S1. Growth rate measurements for all microplate assays.

Fig. S2. Comparison of growth rates measured for BioBricks in different vector backbones.

Fig. S3. BFP plasmids in cell stocks used for microplate assays mutated to reduce burden.

Fig. S4. Growth rate measurements for BioBricks with higher burden exhibit more variability.

Fig. S5. GFP production rate measurements for all microplate assays.

DATA AVAILABILITY

Simulation code, unprocessed data files, analysis scripts, and plasmid assemblies have been archived in a GitHub repository (https://github.com/barricklab/iGEM2019) and on Zenodo (doi: 10.5281/zenodo.10938726). Raw plasmid and genome sequencing data are available from the NCBI Sequence Read Archive (Accession PRJNA1090925).

AUTHOR CONTRIBUTIONS

Conceptualization: Noor Radde, Genevieve A. Mortensen, Diya Bhat, Shireen Shah, Joseph J. Clements, Sean P. Leonard, Matthew J. McGuffie, Dennis M. Mishler, Jeffrey E. Barrick

Data curation: Noor Radde, Genevieve A. Mortensen, Jeffrey E. Barrick

Formal analysis: Noor Radde, Genevieve A. Mortensen, Jeffrey E. Barrick

Funding acquisition: Dennis M. Mishler, Jeffrey E. Barrick

Investigation: Noor Radde, Genevieve A. Mortensen, Diya Bhat, Shireen Shah, Joseph J. Clements, Sean P. Leonard, Matthew J. McGuffie, Jeffrey E. Barrick

Project administration: Dennis M. Mishler, Jeffrey E. Barrick

Software: Genevieve A. Mortensen, Jeffrey E. Barrick

Supervision: Sean P. Leonard, Matthew J. McGuffie, Dennis M. Mishler, Jeffrey E. Barrick

Visualization: Noor Radde, Genevieve A. Mortensen, Jeffrey E. Barrick

Writing – original draft: Noor Radde, Genevieve A. Mortensen, Jeffrey E. Barrick

Writing – review & editing: Noor Radde, Genevieve A. Mortensen, Sean P. Leonard, Dennis M. Mishler, Jeffrey E. Barrick

ACKNOWLEDGMENTS

We thank Angela Pak, Emily Garcia, Mina Kim, Alex MacAskill, Raul Lopez, and Michelle Chang for performing various experiments and participating in the UT Austin 2019 iGEM team; Daniel Deatherage and Jack Dwenger for assistance with genome and plasmid sequencing;

and Giaochau Nguyen, Vrinda Rajkumar, Marco Sanchez, Jeremey Fitzgerald, and Sidharth Kapur from the Microbe Hackers Freshman Research Initiative stream for cloning the BFP control plasmids. We thank the 2019 Michigan State University, Rice University, and Texas Tech University iGEM teams, iGEM judges, and members of the Barrick lab for useful feedback. We acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing high performance computing resources.

FUNDING

This research was supported by the National Science Foundation (CBET-1554179, IOS-2103208, and MCB-2123996), the National Science Foundation BEACON Center for the Study of Evolution (DBI-0939454), the National Institutes of Health (R01GM088344), and the U.S. Army Research Office (W911NF-20-1-0195). The University of Texas at Austin Freshman Research Initiative (FRI) acknowledges support from the Howard Hughes Medical Institute (#52008124). The University of Texas at Austin College of Natural Sciences and Department of Molecular Biosciences provided additional support for the FRI program and iGEM participation.

REFERENCES

1. Weinberg, B. H. *et al.* Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nat. Biotechnol.* **35**, 453–462 (2017).
2. Andrews, L. B., Nielsen, A. A. K. & Voigt, C. A. Cellular checkpoint control using programmable sequential logic. *Science* **361**, eaap8987 (2018).
3. Ryu, M. *et al.* Control of nitrogen fixation in bacteria that associate with cereals. *Nat. Microbiol.* 80–84 (2019) doi:10.1038/s41564-019-0631-2.
4. Isabella, V. M. *et al.* Development of a synthetic live bacterial therapeutic for the human metabolic disease phenylketonuria. *Nat. Biotechnol.* **36**, 857–864 (2018).

5. Leonard, S. P. *et al.* Engineered symbionts activate honey bee immunity and limit pathogens. *Science* **367**, 573–576 (2020).
6. Arkin, A. P. & Fletcher, D. A. Fast, cheap and somewhat in control. *Genome Biol.* **7**, 114 (2006).
7. Renda, B. A., Hammerling, M. J. & Barrick, J. E. Engineering reduced evolutionary potential for synthetic biology. *Mol. Biosyst.* **10**, 1668–1678 (2014).
8. Sleight, S. C., Bartley, B. A., Lieviant, J. A. & Sauro, H. M. Designing and engineering evolutionary robust genetic circuits. *J. Biol. Eng.* **4**, 12 (2010).
9. Umenhoffer, K. *et al.* Reduced evolvability of *Escherichia coli* MDS42, an IS-less cellular chassis for molecular and synthetic biology applications. *Microb. Cell Factories* **9**, 38 (2010).
10. Sleight, S. C. & Sauro, H. M. Visualization of evolutionary stability dynamics and competitive fitness of *Escherichia coli* engineered with randomized multigene circuits. *ACS Synth. Biol.* **2**, 519–528 (2013).
11. Rugbjerg, P., Myling-Petersen, N., Porse, A., Sarup-Lytzen, K. & Sommer, M. O. A. Diverse genetic error modes constrain large-scale bio-based production. *Nat. Commun.* **9**, 787 (2018).
12. Deatherage, D. E., Leon, D., Rodriguez, Á. E., Omar, S. K. & Barrick, J. E. Directed evolution of *Escherichia coli* with lower-than-natural plasmid mutation rates. *Nucleic Acids Res.* **46**, 9236–9250 (2018).
13. Borkowski, O., Ceroni, F., Stan, G.-B. & Ellis, T. Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Curr. Opin. Microbiol.* **33**, 123–130 (2016).
14. Rouches, M. V., Xu, Y., Cortes, L. B. G. & Lambert, G. A plasmid system with tunable copy number. *Nat. Commun.* **13**, 3908 (2022).
15. Bentley, W. E., Mirjalili, N., Andersen, D. C., Davis, R. H. & Kompala, D. S. Plasmid-encoded protein: the principal factor in the ‘metabolic burden’ associated with recombinant bacteria. *Biotechnol. Bioeng.* **35**, 668–681 (1990).

16. Vind, J., Sørensen, M. A., Rasmussen, M. D. & Pedersen, S. Synthesis of proteins in *Escherichia coli* is limited by the concentration of free ribosomes. Expression from reporter genes does not always reflect functional mRNA levels. *J. Mol. Biol.* **231**, 678–688 (1993).
17. Glick, B. R. Metabolic load and heterologous gene expression. *Biotechnol. Adv.* **13**, 247–261 (1995).
18. Stoebel, D. M., Dean, A. M. & Dykhuizen, D. E. The cost of expression of *Escherichia coli lac* operon proteins is in the process, not in the products. *Genetics* **178**, 1653–60 (2008).
19. Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z. & Hwa, T. Interdependence of cell growth and gene expression: origins and consequences. *Science* **330**, 1099–1102 (2010).
20. Espah Borujeni, A., Zhang, J., Doosthosseini, H., Nielsen, A. A. K. & Voigt, C. A. Genetic circuit characterization by inferring RNA polymerase movement and ribosome usage. *Nat. Commun.* **11**, 5001 (2020).
21. Ceroni, F., Algar, R., Stan, G.-B. & Ellis, T. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat. Methods* **12**, 415–418 (2015).
22. Sandoval, C. M. *et al.* Use of pantothenate as a metabolic switch increases the genetic stability of farnesene producing *Saccharomyces cerevisiae*. *Metab. Eng.* **25**, 1–12 (2014).
23. Burgard, A., Burk, M. J., Osterhout, R., Van Dien, S. & Yim, H. Development of a commercial scale process for production of 1,4-butanediol from sugar. *Curr. Opin. Biotechnol.* **42**, 118–125 (2016).
24. Wu, G. *et al.* Metabolic burden: cornerstones in synthetic biology and metabolic engineering applications. *Trends Biotechnol.* **34**, 652–664 (2016).
25. Gubellini, F. *et al.* Physiological response to membrane protein overexpression in *E. coli*. *Mol. Cell. Proteomics* **10**, (2011).
26. Kwon, K. *et al.* Recombinant expression and functional analysis of proteases from *Streptococcus pneumoniae*, *Bacillus anthracis*, and *Yersinia pestis*. *BMC Biochem.* **12**, 17 (2011).

27. Zhang, S. & Voigt, C. A. Engineered dCas9 with reduced toxicity in bacteria: implications for genetic circuit design. *Nucleic Acids Res.* **46**, 11115–11125 (2018).
28. Andrianantoandro, E., Basu, S., Karig, D. K. & Weiss, R. Synthetic biology: new engineering rules for an emerging discipline. *Mol. Syst. Biol.* **2**, 2006.0028 (2006).
29. Canton, B., Labno, A. & Endy, D. Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.* **26**, 787–793 (2008).
30. Registry of Standard Biological Parts. http://parts.igem.org/Main_Page.
31. Smolke, C. D. Building outside of the box: iGEM and the BioBricks Foundation. *Nat. Biotechnol.* **27**, 1099–1102 (2009).
32. Vilanova, C. & Porcar, M. iGEM 2.0—refoundations for engineering biology. *Nat. Biotechnol.* **32**, 420–424 (2014).
33. Kelly, J. R. *et al.* Measuring the activity of BioBrick promoters using an *in vivo* reference standard. *J. Biol. Eng.* **3**, 4 (2009).
34. Beal, J. *et al.* Reproducibility of fluorescent expression from engineered biological constructs in *E. coli*. *PLoS ONE* **11**, e0150182 (2016).
35. Beal, J. *et al.* Quantification of bacterial fluorescence using independent calibrants. *PLoS ONE* **13**, e0199432 (2018).
36. Wielgoss, S. *et al.* Mutation rate inferred from synonymous substitutions in a long-term evolution experiment with *Escherichia coli*. *G3 Bethesda* **1**, 183–186 (2011).
37. Lee, H., Popodi, E., Tang, H. & Foster, P. L. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E2774–E2783 (2012).
38. Drake, J. W. A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 7160–7164 (1991).
39. Lynch, M. Evolution of the mutation rate. *Trends Genet.* **26**, 345–352 (2010).

40. Jack, B. R. *et al.* Predicting the genetic stability of engineered DNA sequences with the EFM calculator. *ACS Synth. Biol.* **4**, 939–943 (2014).

41. Geng, P., Leonard, S. P., Mishler, D. M. & Barrick, J. E. Synthetic genome defenses against selfish DNA elements stabilize engineered bacteria against evolutionary failure. *ACS Synth. Biol.* **8**, 521–531 (2019).

42. Nyerges, Á. *et al.* CRISPR-interference-based modulation of mobile genetic elements in bacteria. *Synth. Biol. Oxf. Engl.* **4**, ysz008 (2019).

43. Fehér, T., Cseh, B., Umenhoffer, K., Karcagi, I. & Pósfai, G. Characterization of *cycA* mutants of *Escherichia coli*. An assay for measuring in vivo mutation rates. *Mutat. Res.* **595**, 184–190 (2006).

44. Shetty, R. P., Endy, D. & Knight, T. F. Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* **2**, 5 (2008).

45. Galdzicki, M., Rodriguez, C., Chandran, D., Sauro, H. M. & Gennari, J. H. Standard biological parts knowledgebase. *PLoS ONE* **6**, e17005–e17005 (2011).

46. Borkowski, O. *et al.* Cell-free prediction of protein expression costs for growing cells. *Nat. Commun.* **9**, 1457 (2018).

47. Sleight, S. C., Bartley, B. A., Lieviant, J. A. & Sauro, H. M. Designing and engineering evolutionary robust genetic circuits. *J. Biol. Eng.* **4**, 12–12 (2010).

48. Zhang, X., Deatherage, D. E., Zheng, H., Georgoulis, S. J. & Barrick, J. E. Evolution of satellite plasmids can prolong the maintenance of newly acquired accessory genes in bacteria. *Nat. Commun.* **10**, 5809–5809 (2019).

49. Brkljacic, J. *et al.* Frequency, composition and mobility of *Escherichia coli*-derived transposable elements in holdings of plasmid repositories. *Microb. Biotechnol.* **15**, 455–468 (2022).

50. Reis, A. C. & Salis, H. M. An automated model test system for systematic development and improvement of gene expression models. *ACS Synth. Biol.* **9**, 3145–3156 (2020).

944 51. LaFleur, T. L. Automated model-predictive design of synthetic promoters to control
945 transcriptional profiles in bacteria. *Nat. Commun.* **13**, 5159 (2022).
946 52. Zhuang, K., Vemuri, G. N. & Mahadevan, R. Economics of membrane occupancy and
947 respiro-fermentation. *Mol. Syst. Biol.* **7**, 500 (2011).
948 53. Shachrai, I., Zaslaver, A., Alon, U. & Dekel, E. Cost of unneeded proteins in *E. coli* is
949 reduced after several generations in exponential growth. *Mol. Cell* **38**, 758–67 (2010).
950 54. Barrick, J. E. *et al.* Daily transfers, archiving populations, and measuring fitness in the long-
951 term evolution experiment with *Escherichia coli*. *J. Vis. Exp.* e65342 (2023)
952 doi:10.3791/65342.
953 55. Chochinov, C. A. & Nguyen Ba, A. N. Bulk-fitness measurements using barcode sequencing
954 analysis in yeast. in *Yeast Functional Genomics* (ed. Devaux, F.) vol. 2477 399–415
955 (Springer US, New York, NY, 2022).
956 56. Li, F., Tarkington, J. & Sherlock, G. Fit-Seq2.0: An improved software for high-throughput
957 fitness measurements using pooled competition assays. *J. Mol. Evol.* **91**, 334–344 (2023).
958 57. Joshi, S. H.-N., Yong, C. & Gyorgy, A. Inducible plasmid copy number control for synthetic
959 biology in commonly used *E. coli strains*. *Nat. Commun.* **13**, 6691 (2022).
960 58. Ceroni, F. *et al.* Burden-driven feedback control of gene expression. *Nat. Methods* **15**, 387–
961 393 (2018).
962 59. Barajas, C., Huang, H.-H., Gibson, J., Sandoval, L. & Del Vecchio, D. Feedforward growth
963 rate control mitigates gene activation burden. *Nat. Commun.* **13**, 7054 (2022).
964 60. Rugbjerg, P., Sarup-Lytzen, K., Nagy, M. & Sommer, M. O. A. Synthetic addiction extends
965 the productive life time of engineered *Escherichia coli* populations. *Proc. Natl. Acad. Sci.*
966 **115**, 2347–2352 (2018).
967 61. Segall-Shapiro, T. H., Meyer, A. J., Ellington, A. D., Sontag, E. D. & Voigt, C. a. A 'resource
968 allocator' for transcription based on a highly fragmented T7 RNA polymerase. *Mol. Syst.*
969 *Biol.* **10**, 742–742 (2014).

970 62. Wang, K., Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Evolved orthogonal ribosomes
971 enhance the efficiency of synthetic genetic code expansion. *Nat Biotechnol* **25**, 770–777
972 (2007).
973 63. Orelle, C. *et al.* Protein synthesis by ribosomes with tethered subunits. *Nature* **524**, 119–124
974 (2015).
975 64. Ellis, T. Predicting how evolution will beat us. *Microb. Biotechnol.* **12**, 41–43 (2019).
976 65. Suárez, G. A., Renda, B. A., Dasgupta, A. & Barrick, J. E. Reduced mutation rate and
977 increased transformability of transposon-free *Acinetobacter baylyi* ADP1-ISx. *Appl. Environ.*
978 *Microbiol.* **83**, e01025-17 (2017).
979 66. Battaglino, B., Arduino, A. & Pagliano, C. Mathematical modeling for the design of evolution
980 experiments to study the genetic instability of metabolically engineered photosynthetic
981 microorganisms. *Algal Res.* **52**, 102093 (2020).
982 67. Nikolados, E.-M., Weiße, A. Y., Ceroni, F. & Oyarzún, D. A. Growth defects and loss-of-
983 function in synthetic gene circuits. *ACS Synth. Biol.* **8**, 1231–1240 (2019).
984 68. Hernández-Beltrán, J. C. R., San Millán, A., Fuentes-Hernández, A. & Peña-Miller, R.
985 Mathematical models of plasmid population dynamics. *Front. Microbiol.* **12**, 606396 (2021).
986 69. Nyström, A., Papachristodoulou, A. & Angel, A. A dynamic model of resource allocation in
987 response to the presence of a synthetic construct. *ACS Synth. Biol.* **7**, 1201–1210 (2018).
988 70. Johnson, P. adaptivetau: tau-leaping stochastic simulation. https://cran.r-
989 project.org/package=adaptivetau (2019).
990 71. Chang, W. *et al.* shiny: Web Application Framework for R. https://shiny.posit.co/ (2024).
991 72. Haldimann, A. & Wanner, B. L. Conditional-replication, integration, excision, and retrieval
992 plasmid-host systems for gene structure-function studies of bacteria. *J. Bacteriol.* **183**,
993 6384–6393 (2001).

994 73. Deatherage, D. E. & Barrick, J. E. Identification of mutations in laboratory-evolved microbes
995 from next-generation sequencing data using *breseq*. *Methods Mol. Biol.* **1151**, 165–188
996 (2014).
997 74. Barrick, J. E. *et al.* Identifying structural variation in haploid microbial genomes from short-
998 read resequencing data using *breseq*. *BMC Genomics* **15**, 1039–1039 (2014).
999 75. Knight, T. *Idempotent Vector Design for Standard Assembly of Biobricks*.
1000 http://dx.doi.org/http://hdl.handle.net/1721.1/21168 (2003).
1001 76. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Completing bacterial genome assemblies
1002 with multiplex MinION sequencing. *Microb. Genomics* **3**, e000132 (2017).
1003 77. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor.
1004 *Bioinformatics* **34**, i884–i890 (2018).
1005 78. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome
1006 assemblies from short and long sequencing reads. *PLOS Comput. Biol.* **13**, e1005595
1007 (2017).
1008 79. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads
1009 using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
1010 80. McGuffie, M. J. & Barrick, J. E. pLannotate: engineered plasmid annotation. *Nucleic Acids*
1011 *Res.* **49**, W516–W522 (2021).
1012 81. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421–421
1013 (2009).
1014 82. Hecht, A., Endy, D., Salit, M. & Munson, M. S. When wavelengths collide: Bias in cell
1015 abundance measurements due to expressed fluorescent proteins. *ACS Synth. Biol.* **5**,
1016 1024–7 (2016).
1017 83. Alieva, N. O. *et al.* Diversity and evolution of coral fluorescent proteins. *PLoS ONE* **3**, e2680
1018 (2008).
1019