

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

Received: 8 April 2024

Accepted: 18 July 2024

Published online: 24 July 2024

 Check for updates

Noor Radde^{1,3}, Genevieve A. Mortensen ^{1,3}, Diya Bhat ¹, Shireen Shah¹, Joseph J. Clements¹, Sean P. Leonard ¹, Matthew J. McGuffie¹, Dennis M. Mishler^{1,2} & Jeffrey E. Barrick ¹ 

Engineered DNA will slow the growth of a host cell if it redirects limiting resources or otherwise interferes with homeostasis. Escape mutants that alleviate this burden can rapidly evolve and take over cell populations, making genetic engineering less reliable and predictable. Synthetic biologists often use genetic parts encoded on plasmids, but their burden is rarely characterized. We measured how 301 BioBrick plasmids affected *Escherichia coli* growth and found that 59 (19.6%) were burdensome, primarily because they depleted the limited gene expression resources of host cells. 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. We made this model available online for education (<https://barricklab.org/burden-model>) and added our burden measurements to the iGEM Registry. Our results establish a fundamental limit on what DNA constructs and genetic modifications can be successfully engineered into cells.

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 the 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 engineered DNA constructs use resources from the cell to replicate and express genes, these processes are the most common and predictable sources of burden¹³. Transcriptional resources (e.g., RNA polymerases) or translational resources (e.g., ribosomes, charged tRNAs) often become limiting when a foreign DNA construct directs the synthesis of RNAs and proteins that are not native to the cell. Protein overexpression studies in *E. coli* generally find that ribosomes are the most limiting factor. Translating these foreign proteins decreases the growth rates of cells in proportion to how many ribosomes are redirected away from producing host proteins^{14–18}. Usage of gene expression resources can be monitored using high-throughput approaches that globally profile RNA abundance and ribosomal occupancy^{19,20} or reporter genes with expression levels that reflect the

¹Department of Molecular Biosciences, Center for Systems and Synthetic Biology, The University of Texas at Austin, Austin, TX, USA. ²The Freshman Research Initiative, College of Natural Sciences, The University of Texas at Austin, Austin, TX, USA. ³These authors contributed equally: Noor Radde, Genevieve A. Mortensen. ✉ e-mail: jbarrick@cm.utexas.edu

depletion of overall cellular capacities for transcription and translation²¹.

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 burdensome²⁷. 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 sequences³⁰ 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 strength³³ and fluorescent protein expression^{34,35} across many labs. It has been proposed that genetic reliability—in the evolutionary sense of how many cell doublings it takes for an engineered function to decay within a population—be listed on a data sheet describing a genetic part²⁹, but this property is rarely characterized in practice. One goal of iGEM is to improve upon existing parts, and many BioBrick sequences are reused 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 standardized frameworks for measuring burden and simulating the dynamics of evolutionary failure that 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

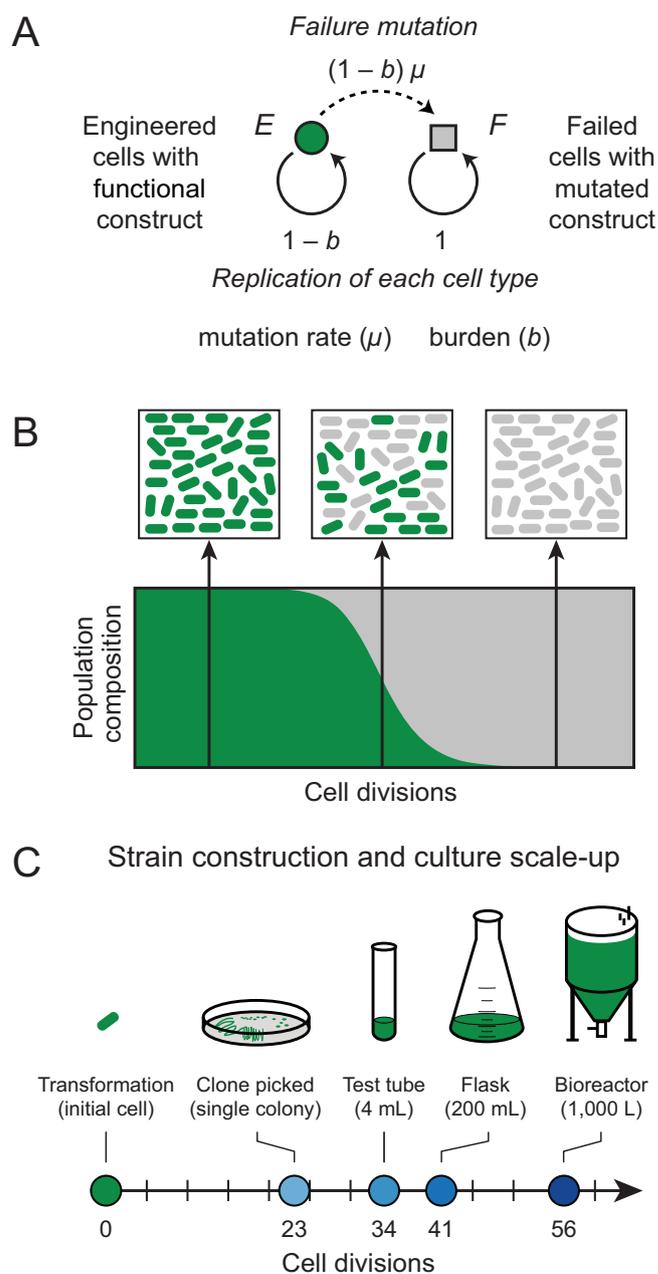


Fig. 1 | Evolutionary failure of a population of engineered cells. A Graphical representation of a differential equation model with one class of failure mutations that completely alleviates the fitness burden of an engineered DNA construct on a host cell. **B** Population dynamics expected from this 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 that occur as a single engineered cell is grown into cultures on different laboratory and industrial scales after its creation (see Supplementary Data 1). Source data are provided with this paper.

there is one category of mutations that leads to the 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).

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, Supplementary Data 1). 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 $\sim 5 \times 10^9$ 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 1000 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 affects mutation rates 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 less robust to base changes are at a greater risk for inactivating mutations^{6,7}. The rate of base substitutions in *E. coli* is $\sim 5 \times 10^{-10}$ 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 $\sim 1 \times 10^{-7}$ per cell division just from base substitutions. This estimate does not account for the possibility that there are mutational hotspots in a sequence, such as mononucleotide repeats, that can cause certain mutations to occur at much higher rates^{40,41}. 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^{11,42,43}.

In the end, empirical measurements generally find a rate of $\sim 10^{-6}$ per cell division for mutations that inactivate a single gene that is located in the chromosome of *E. coli* or another bacterium^{42,44}. 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 mutations reverting a stop codon in a reporter construct, which can only occur via one or a few single base substitutions, has been measured as $\sim 10^{-7}$ per cell division rather than the value of $\sim 10^{-9}$ expected if this reporter were tested in the chromosome¹². 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 division. 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 events 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 compared to stochastic models (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 jackpot events when a mutation occurs early in the growth of a population (as seen in the panel for $b = 20\%$, $\mu = 10^{-8}$).

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 $\geq 50\%$ 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 $\geq 40\%$ 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 allows users to explore a range of parameters and rerun simulations multiple times to see for themselves how mutations can affect the functional lifetimes of devices constructed in living, and therefore evolving cells.

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) competition³¹. 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 standard⁴⁵. As a consequence, most BioBricks in the kit distributed to iGEM teams are provided in plasmids pSBIC3, pSBIA2, or in both of these backbones (Fig. 4A). pSBIC3 and pSBIA2 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 pSBIC3 versus β -lactamase (*bla*) which confers resistance to ampicillin and other β -lactams (Amp^R) for pSBIA2. These plasmids also differ in how the expression of the cloned

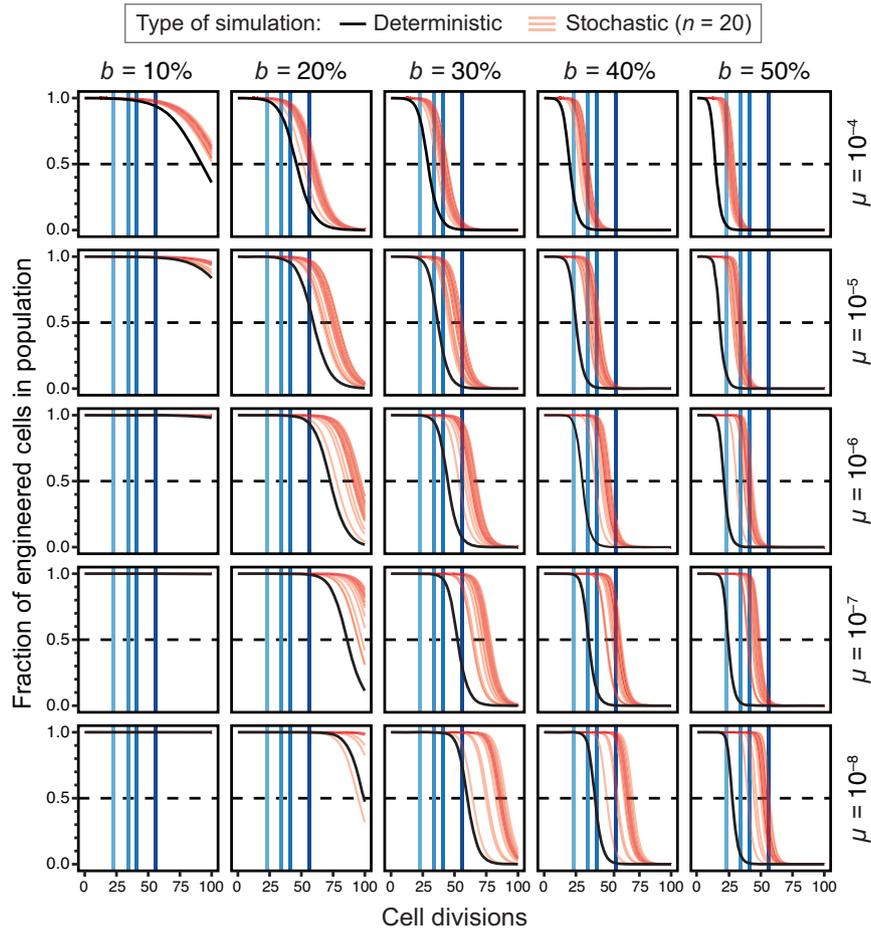


Fig. 2 | Simulations of evolutionary failure times for populations of engineered cells. In each panel, the results for deterministic (black) and stochastic (red) simulations of the failure model are shown for one combination of burden (b) and failure mutation rate (μ) parameters. Vertical blue lines represent the culture

scales shown in Fig. 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.

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.

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 (Supplementary Fig. 1). 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 (Supplementary Fig. 2, Supplementary Data 2, and “Methods” section).

In total, we measured the effects of the 5 BFP control plasmids and 301 other BioBricks on *E. coli* growth (Fig. 4B, Supplementary Data 3). 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 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) (Supplementary Fig. 3A). 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) (Supplementary Fig. 3B). 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 procedure⁴⁶ 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%.

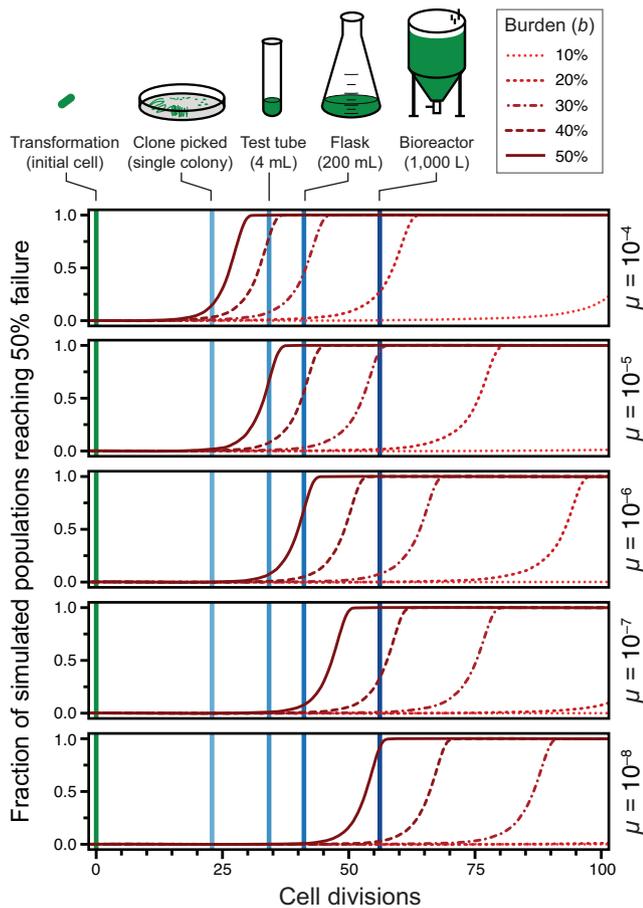


Fig. 3 | Cumulative distributions of times to 50% failure in stochastic simulations. Each curve shows the values of this summary statistic of failure for 10,000 simulations with a given parameter combination. More variability in the time to 50% failure leads to a flatter curve.

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 comparison⁴⁷. Therefore, we hypothesized that the 59 BioBricks in the high-confidence burden set would be more likely than BioBricks that were not burdensome 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)⁴⁸.

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 burdensome BioBricks 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 slow *E. coli* growth 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 $27.5 \pm 8.6\%$ (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 flanking rho-independent terminators designed to insulate the BioBrick, there is commonly some level of transcriptional read-through of these elements^{49,50}.

BioBrick burden is not correlated with organism of origin

Next, we tested whether BioBricks that incorporated genetic part sequences from certain types of organisms were more likely to exhibit burden. We classified BioBricks according to the most divergent organism from which their sequences originated and examined twelve classification schemes in which we categorized similar organisms at different levels of taxonomic resolution, only considered protein-coding features, and/or omitted fluorescent proteins when assigning a BioBrick's organism of origin ("Methods" section, Supplementary Data 3, Supplementary Fig. 4). We found no variation in the distributions of normalized growth rates of *E. coli* strains carrying BioBrick plasmids with sequences derived from different types of organisms for any of the classification schemes (Kruskal–Wallis tests, $p > 0.5$) (Supplementary Table 1). The chances that a BioBrick was classified in the high-confidence burden set of 59 BioBricks also did not vary significantly with source organism category for eleven of the twelve classification schemes (likelihood-ratio tests comparing binomial models, $p > 0.05$) (Supplementary Table 1). The one exception, which only barely reached statistical significance ($p = 0.047$), found that BioBricks with sequences from bacteria other than γ -proteobacteria were more likely to be burdensome compared to sequences from *E. coli*, γ -proteobacteria other than *E. coli*, or eukaryotes (Supplementary Fig. 4B). Overall, we conclude that organism of origin did not consistently correlate with whether a BioBrick would exhibit burden or the amount of that burden.

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 (Supplementary Data 4 and "Methods" section). 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 the 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

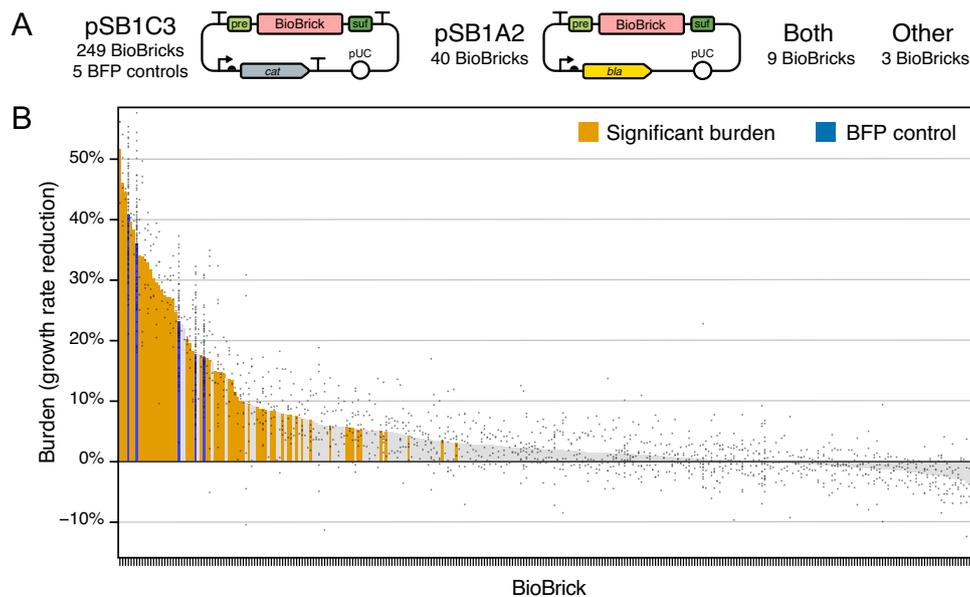


Fig. 4 | Measurements of BioBrick burden. **A** Maps of the two plasmid backbones that housed most of the 301 BioBricks that were tested and the five BFP controls that were included in every assay. The prefix (pre) and suffix (suf) multiple cloning sites used in BioBrick assembly are shown. **B** Burden 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). Supplementary Data 2 contains the results of all microplate assays. Supplementary Data 3 has calculated statistics. Source data are provided with this paper.

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 the 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 (Supplementary Fig. 5). The burden was reduced from 45.8% to 17.8% in one case and from 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 ISS 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 restreaking from 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 ($p = 2.0 \times 10^{-11}$, two-tailed t -test for a non-zero slope) (Supplementary Fig. 6).

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^{21,51}. 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.

Table 1 | Most burdensome BioBricks

BioBrick ^a	Seq ^b	Burden (b) ^c	Fraction other burden (b _o /b) ^d	Subparts ^e	Function ^f
K523022	M	51.7 ± 19.2%	n.s.	<i>P_{lac} & lacZ' crtE crtI crtB</i>	Carotenoid synthesis (<i>Pantoea ananatis</i>)
K733010	C	46.0 ± 6.2%	0.16–0.71	<i>P_{trms} & endB</i>	Antitoxin gene (<i>Bacillus subtilis</i>) ^g
J04450	NS	44.4 ± 2.2%	n.s.	<i>P_{lac} & mRFP1</i>	RFP reporter
K523014	C	39.6 ± 4.7%	1.04–1.98	<i>P_{lac} & lacZ' bglX</i>	Cellobiose degradation
K523020	M, E	38.3 ± 8.7%	n.s.	<i>P_{lac} & lacZ' INP+bglX</i>	Cellobiose degradation (INP, <i>Pseudomonas syringae</i>)
K608010	C	34.1 ± 7.7%	NT	<i>P_{J23110} & GFP</i>	GFP reporter
K515100	C	33.9 ± 15.9%	0.26–0.88	<i>P_{veg2} & laaM & laaH</i>	Indoleacetamide synthesis (<i>Pseudomonas savastanoi</i>) ^g
J61000	m	33.4 ± 4.1%	0.21–0.96	<i>P_{cat} & cat</i>	Chloramphenicol resistance
K541526	C	32.9 ± 7.8%	n.s.	<i>P_{veg} & reflectin1A</i>	Reflectin reporter (<i>Euprymna scolopes</i>) ^g
K592020	m	31.8 ± 5.0%	NT	<i>P_{fixK2} & cl(λ) P_{ci} & amilCP</i>	Blue light sensor output (<i>Acropora millepora</i>)
J36335	m	30.2 ± 12.2%	n.s.	<i>P_{lac} & kaiA P_{lac} & kaiC</i>	Circadian rhythm (<i>Synechococcus elongatus</i>)
I759017	C	29.5 ± 8.3%	NT	<i>P_{tet} [cis5] & YFP</i>	YFP reporter
K346000	C	29.1 ± 10.4%	n.s.	<i>& RNAP(T3)</i>	Phage RNA polymerase (Phage T3)
C0056	C	28.2 ± 3.9%	n.s.	<i>ci434(λ)</i>	Mutant phage repressor (Phage λ)
K880005	C	27.5 ± 8.6%	n.s.	<i>P_{J23100} &</i>	Gene expression
C0053	NS	27.2 ± 6.5%	n.s.	<i>ciI(P22)</i>	Phage repressor (Phage P22)
K608012	C	27.1 ± 4.7%	NT	<i>P_{J23110} & GFP</i>	GFP reporter
I759014	C	26.8 ± 5.8%	n.s.	<i>P_{tet} [cis2] & YFP</i>	YFP reporter
K541502	C	24.6 ± 3.2%	0.42–1.91	<i>P_{veg} & lipA_{sig}</i>	Gene expression/secretion (<i>Bacillus subtilis</i>) ^g
K395602	C	20.3 ± 1.9%	0.09–0.38	<i>P_{T7} & MpAAT1</i>	Apple fragrance generator (<i>Malus pumila</i>)
K733013	C	19.5 ± 3.3%	n.s.	<i>P_{veg} & GFP</i>	GFP reporter ^g
K523013	C	18.3 ± 8.8%	NT	<i>P_{lac} & lacZ' INP + EYFP</i>	EYFP reporter (INP, <i>Pseudomonas syringae</i>)
I761014	C	17.5 ± 5.0%	0.21–1.33	<i>& cinR & cinI</i>	Quorum sensing (<i>Rhizobium leguminosarum</i>)
C0051	NS	17.1 ± 8.2%	n.s.	<i>λ-cl + LVA</i>	Phage repressor (Phage λ)
K137018	C	16.8 ± 8.2%	NT	<i>P_{L-lacO1} & luxR P_{lux-R} & GFP</i>	Quorum sensing receiver (<i>Aliivibrio fischeri</i>)
K1149051	C	15.0 ± 8.4%	n.s.	<i>P_{J23104} & phaC1 phaA phaB1</i>	Polyhydroxybutyrate synthesis (<i>Ralstonia eutropha</i>)
K731721	C	14.8 ± 4.4%	n.s.		Transcription terminator (Phage T7)
K639003	m	14.8 ± 2.8%	n.s.	<i>P_{rrmB-P1} & lacI P_{L-lacO1} & mCherry</i>	Stress sensor
K541501	C	14.4 ± 3.6%	n.s.	<i>P_{veg} & sacB_{sig}</i>	Gene expression/secretion (<i>Bacillus subtilis</i>) ^g
K608011	C	13.7 ± 5.4%	NT	<i>P_{J23110} & GFP</i>	GFP reporter
K861172	NS	13.4 ± 2.5%	n.s.	<i>P_{cstA} & cl(λ)</i>	Phage repressor (Phage λ)
K617004	C	11.6 ± 1.5%	0.95–2.32	<i>attP(λ) P'OP</i>	Phage attachment site (Phage λ)
K325218	m	10.8 ± 7.3%	0.76–1.55	<i>P_{araC} & luc(orange)</i>	Luciferase reporter (<i>Luciola cruciata</i>)
I712669	m	10.1 ± 4.5%	NT	<i>P_{CMV} GFP</i>	GFP reporter ^m

^aBioBrick accession numbers. The 34 parts shown all had an estimated burden that was significantly greater than zero after correcting for multiple testing (one-tailed t-tests, Benjamini–Hochberg adjusted $p < 0.05$) and had a mean estimated burden value of $>10\%$.

^bResults 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 Supplementary Data 4.

^cBurden as the percentage reduction in growth rate caused by the BioBrick ± estimated 95% confidence limits.

^d95% 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 (one-tailed t-tests, Benjamini–Hochberg adjusted $p < 0.05$). NT not tested because the BioBrick contains a protein that interferes with the measurement of GFP fluorescence.

^eRepresentation of gene expression signals and genes in the BioBrick abbreviated as follows: P_x , promoter from gene or operon x ; λ , ribosome-binding site; $[y]$ other regulatory sequence. Other italicized entries are gene names.

^fGeneral 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.

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 the burden noted above. The second set consisted of 14 plasmids available from the iGEM Registry that contain constitutive promoters of different strengths driving the 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, Supplementary Fig. 7, Supplementary Data 5, Supplementary Data 6). 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

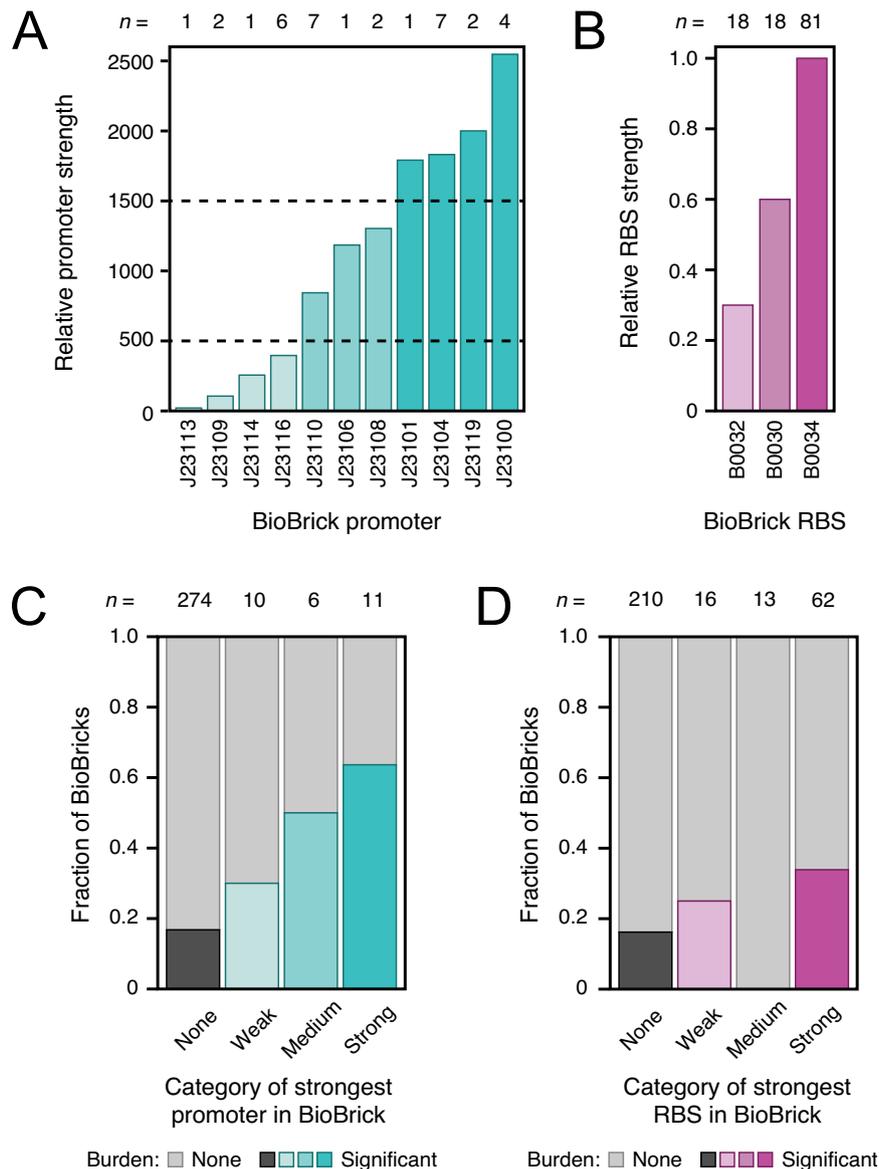


Fig. 5 | Strong promoter and ribosome-binding sites are more likely to be found in burdensome BioBricks. A, B Relative strengths of common promoters and ribosome-binding site (RBS) BioBrick parts, as reported in the iGEM Registry. The number of examples of each promoter or RBS in the 301 BioBricks examined 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 that exhibited significant burden (one-tailed t -tests, Benjamini-Hochberg adjusted $p < 0.05$) 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). Source data are provided with this paper.

tested under different culture conditions (see “Methods” section). 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 strengths can lead to translating the same amount of protein but with more or less efficient use of ribosomes²¹. As for the 301 BioBricks we tested and the unmutated BFP controls, none of the RFP expression constructs had a burden of >45% in the unclonable range.

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 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” section). We again used the BFP plasmids as internal controls for normalizing GFP production rates between different microplate assays (Supplementary Fig. 8 and “Methods” section).

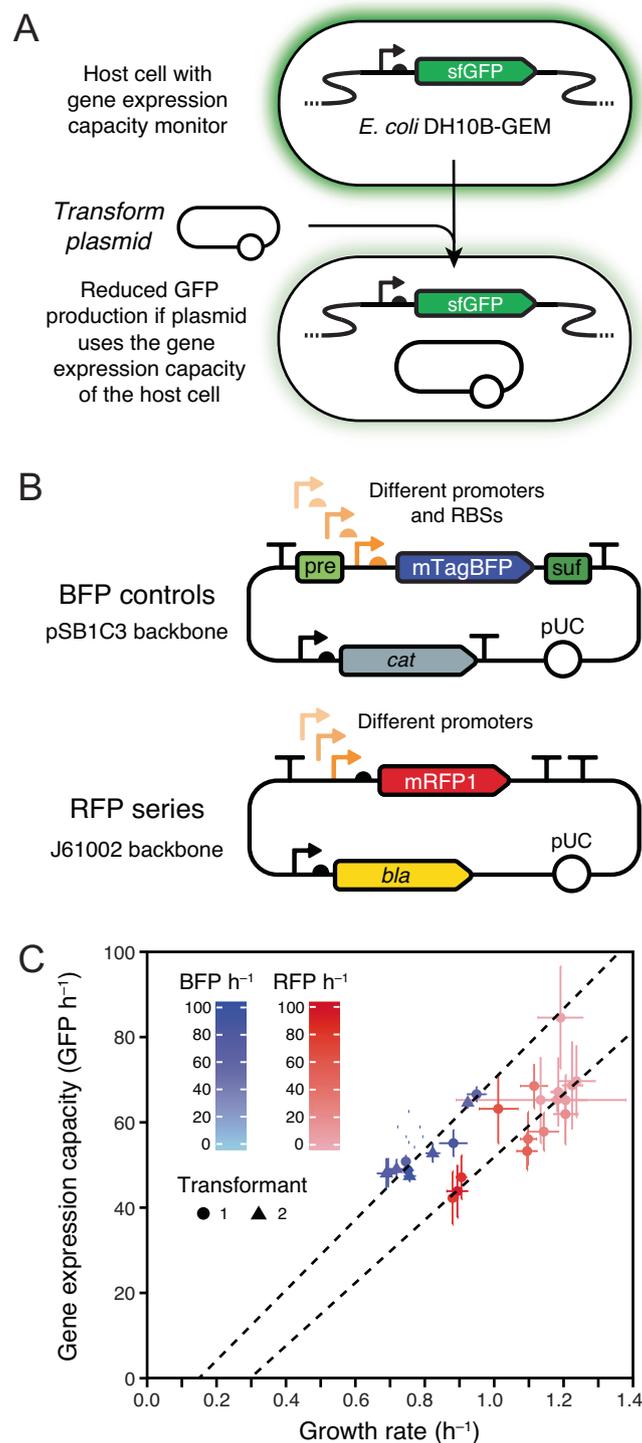


Fig. 6 | Expression of recombinant proteins from a plasmid reduces the growth rate of *E. coli* because it diverts some of its capacity for gene expression. **A** *E. coli* DH10B-GEM host strain with the gene expression capacity monitoring device that constitutively expresses GFP integrated into its 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. Means are

plotted with error bars that are 95% confidence limits. Two independent transformants of each BFP plasmid that were tested separately are displayed as points with different shapes. There were eight technical replicates (microplate wells) for each BFP plasmid transformant and four for each RFP plasmid tested. 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. Growth rates, GFP production rates, and RFP/BFP production rates for individual plasmids are plotted in Supplementary Fig. 7. Supplementary Data 5 contains the results of all microplate assays, and Supplementary Data 6 has calculated statistics. Source data are provided with this paper.

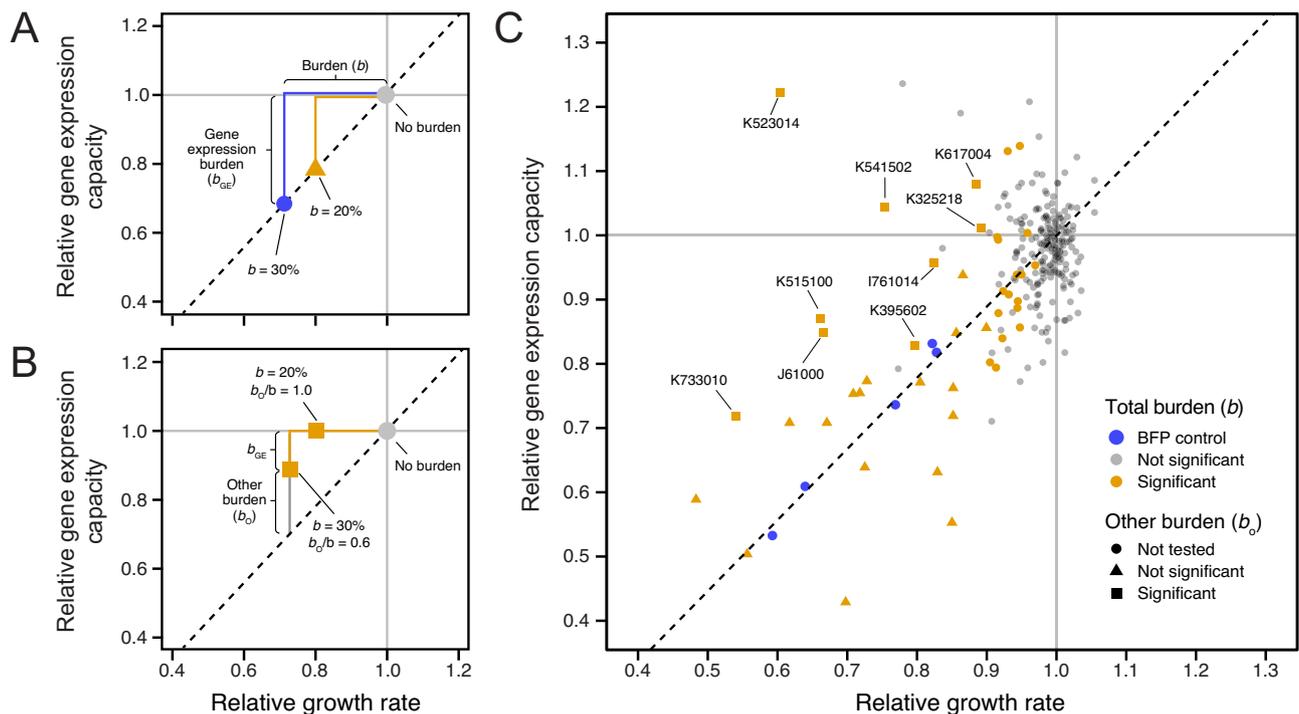


Fig. 7 | Some BioBricks exhibit burden from sources other than gene expression. **A** Examples of expected results for two BioBricks that exhibit burden (b) that is wholly due to utilizing the gene expression capacity of the host cell. The reduction in growth rate is proportional to the reduction in 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** Mean growth rates and GFP production rates measured 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 the burden (reduction in growth rate) was significantly greater than zero (one-tailed t -tests, Benjamini–Hochberg adjusted $p < 0.05$). Symbols indicate whether the null hypothesis that all burden was due to utilizing the gene expression capacity of the host cell could be rejected (one-tailed t -tests, Benjamini–Hochberg adjusted $p < 0.05$). Only the 26 BioBricks with burden significantly greater than zero, a mean estimated burden $>10\%$, and no interference from other fluorescent proteins were tested. 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. Source data are provided with this paper.

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 a high-confidence prediction of burden and a mean estimated 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).

Discussion

By measuring the burden of 301 BioBricks and performing computational 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^{8,11}. 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 strain¹⁰, 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 used²¹.

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 burden 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, which encodes an inducible luciferase gene²⁹. While not one of the BioBricks we tested, F2620 was previously noted to reproducibly fail in <100 generations

due to deletions between two 143-bp repeats introduced by re-use of the B0015 double terminator part. The creators confirmed that this failure consistently occurred due to mutations that happened after plasmid transformation. Our model shows how you can get deterministic failures like this if the mutation rate is sufficiently high, as it can be for repeat-mediated deletions⁴⁰. We discovered inactivating deletions or transposon insertions in seven BioBrick plasmids, which likely indicates 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 the burden.

Future work could clarify whether the cases of sequence discrepancies we encountered are already mutated 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^{11,52}. Surveys of plasmids in other repositories have also found that some acquire inactivating transposon insertions⁵³. Our findings support calls for researchers to report the full sequences of plasmids they create and submit to repositories such as the iGEM Registry^{54,55} and caution that one should also verify the sequences of plasmid stocks obtained from repositories. This information will make it possible to recognize when evolution is undermining DNA constructs and experimental results.

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 protein overexpression in *E. coli*^{14–18}. 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. With the limited set of BioBricks we tested, we were also unable to examine whether the burden of individual genetic parts can be used to predict the burden of complex devices constructed from combinations of these parts. Ongoing improvements in tools for predicting transcription and translation initiation rates and expanding databases of high-throughput measurements^{56,57} may eventually make these types of predictions possible. Our finding that the type of organism from which a BioBrick’s sequence originated was not consistently correlated with its burden agrees with a high-throughput study of horizontal gene transfer from 79 prokaryotic genomes to *E. coli*⁵⁸. It concluded that differences in gene expression were more important than the source organism’s relatedness to *E. coli* in explaining why plasmids containing certain genes could not be cloned in this host.

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 occupancy⁵⁹, 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 the 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 begins⁶⁰, survival during stationary phase, colony growth on agar, or survival of cryopreservation. Furthermore, our approach can only be applied to understand genetic stability under laboratory conditions, not in environmental contexts or host-associated microbiomes. Co-culture competition assays between a strain of interest and a reference strain could be used to measure fitness in a way that captures all components of fitness in any environment⁶¹. To make these measurements high-throughput, host strains with unique sequence barcodes in their chromosomes and transformed with different engineered plasmids or DNA constructs could be simultaneously competed all-against-one-another in bulk competitive fitness assays^{62,63}. Our experiments were all in a cloning strain of *E. coli*. It would be interesting to examine how burden varies in strains optimized for other applications, such as recombinant protein production.

Researchers can take actions to improve the constructability and stability of especially burdensome engineered DNA sequences. Most obviously, using low- or medium-copy plasmids rather than high-copy ones or integrating single-copy constructs into the chromosome of a bacterium will often reduce the burden into the cloneable and stable ranges¹⁰. Systems have also been engineered for controlling plasmid copy number, so that DNA parts can be maintained in cells at a low copy number and then amplified on demand^{47,64}. Similarly, reducing the burden of a construct can be achieved by altering promoter and ribosome-binding site strengths or by using inducible promoters, as long as these changes are compatible with device function^{10,21}. Systems that regulate expression in response to the growth rate of a cell^{65,66} or that couple continued functioning of the engineered DNA to cell survival⁶⁷ can more directly buffer against evolutionary failure. Another category of more ambitious approaches is to introduce orthogonal polymerases⁶⁸ or ribosomes^{69,70} into a cell to prevent synthetic constructs from competing with native gene expression, though the requirement that a cell produce the necessary machinery may itself be burdensome. Next, aspects of the growth environment can sometimes be changed. For example, supplementing media with vitamins or altering salt concentrations has been reported to stabilize certain constructs^{11,22}. A final category of approaches seeks to reduce the chances of mutations to improve the evolutionary stability of genetic constructs^{7,71}. For example, cells with lower mutation rates can be created by deleting or repressing transposons^{9,72} or by altering cellular processes that affect point mutation rates^{12,73}.

We created an interactive model of failure mutations in a cell population that can be used to explore how tuning mutation rates and construct burden affect whether a DNA construct is likely to remain intact within cell populations that are grown to typical laboratory and production scales. Similar deterministic^{6,11} and stochastic^{74,75} models have been developed by others. Models that include individual steps in gene expression and RNA and protein degradation are also beginning to be used to examine evolutionary stability^{21,76}. Our model and these others still do not consider or fully take into account several complications⁷¹. First, rather than one category of mutation leading to complete failure, there are typically multiple categories of mutations,

some of which only partially alleviate the burden, occurring at different rates in real systems^{10,11}. Second, burden, plasmid copy number, and mutation rates are not necessarily independent parameters and may exhibit cell-to-cell variation. For example, overexpression of recombinant proteins can activate stress responses and increase mutation rates⁷³. Third, plasmids are multicopy within cells so the fitness benefit of a mutation can take several generations to fully manifest and depends on how plasmids segregate between daughter cells. These intricacies of plasmid evolution have been tackled by a variety of more complex models that could be applied to engineered plasmids⁷⁷. Finally, models that take into account different phases of cellular growth could be used to further refine these dynamics⁷⁸.

Modeling cellular function over time in a way that takes evolution into account is important for making synthetic biology more reliable and predictable. This random, self-reinforcing failure mode does not have a direct parallel in traditional engineering fields. Researchers designing engineered cells should be aware of when they are nearing a danger zone of evolutionary stability where DNA designs may become unconstructable. They should also recognize that the stochastic nature of evolutionary failure may lead to large variation in their experimental results, failure during process scale-up, or loss of function when cells are deployed for long periods of time in complex environments outside of the lab, such as in animal and plant microbiomes. Our interactive model (<https://barricklab.org/burden-model>) can be used for educating both new and practicing synthetic biologists about evolutionary constraints. In the long term, it is critical that we improve our understanding of what synthetic DNA constructs exhibit burden and why. Our results from measuring the burden of many BioBricks contribute to this goal. The overarching conclusion of our study can be summarized as a rule of thumb: to avoid the specter of unwanted evolution, don't attempt to engineer a microbial cell in a way that slows its growth rate by >30%.

Methods

Model of evolutionary failure

We implemented a model in R that is similar to one used by Rugbjerg et al. to predict loss of production from an engineered cell population due to escape mutations¹¹. We parameterized our model such that failed (i.e., mutated) cells, F , have a relative growth rate of one. Engineered cells, E , have a growth rate that is this value minus the burden, b , of the engineered construct. The corresponding equations for how the numbers of engineered cells, $E(t)$, and failed cells, $F(t)$, change over time are:

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

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

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

$$D(t) = \log_2[E(t) + F(t)] \quad (3)$$

For stochastic simulations of this model, we used the `adaptivetau` R package⁷⁹. We also created an online version (<https://barricklab.org/shiny/burden-model>) that can perform deterministic and stochastic simulations of this model using the Shiny R package⁸⁰.

Media and growth conditions

E. coli was cultured at 37 °C in Lysogeny Broth (LB) (10 g tryptone, 5 g yeast extract, 10 g NaCl per liter) with 16 g/L agar added for solid media. Unless otherwise indicated, liquid cultures were grown in 18 mm × 150 mm glass test tubes with orbital shaking at 200 r.p.m over a 1-inch diameter. Antibiotics were added at the following 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.⁸¹ 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 *plnt-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. *plnt-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 2 S 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*^{82,83}. 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 IS4 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 the 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 min 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 CaCl₂ solution chilled on ice. Next, 30 ml fractions of the cells were centrifuged again at room temperature for 10 min at 3400 × *g*. Again, the pellets were combined, resuspending in a total of 20 ml of chilled glycerol-CaCl₂ 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 from the iGEM 2018 DNA Distribution Kit 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 h. 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 min. 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 insert 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^{45,84}. 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 h at 37 °C, followed by plating on LB agar containing chloramphenicol. After incubation at 37 °C for 18 h, we inoculated isolated colonies into fresh LB liquid media containing chloramphenicol and grew these cultures at 37 °C for 18 h. 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. Porechop⁸⁵ and fastp⁸⁶ 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*⁸² to see if there were any discrepancies. Second, we performed de novo assembly of reads using either Unicycler⁸⁷ or flye⁸⁸, annotated the resulting assemblies with pLannotate⁸⁹, and examined them for matches to the expected parts using blastn searches⁹⁰ 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 previously²¹. 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 min with 7 min of orbital shaking during each cycle. Each plate was run for a minimum of 6 h.

RFP and BFP plasmid burden assays

For the series of plasmids expressing RFP under the 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 min with 6.5 min of shaking during each cycle. BFP samples were measured every 10 min with 7 min of shaking during each cycle. For the RFP series, we also monitored cell density using OD660 instead of OD600 to avoid interference from RFP absorbance⁹¹.

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 points over the first hour of measurements for each strain to match the grand mean for those data points over all replicates of that strain. We then fit growth rates using nonlinear least-squares regression to an exponential model: $C(t) = C_0 e^{rt}$. We assumed that OD is directly proportional to the number of cells at a given time, $C(t)$. C_0 is the initial number of cells, and r is the specific growth rate. We fit C_0 and r for all sets of nine consecutive measurements (a 90-min window in the standard assay) after the OD exceeded 0.03 and recorded the largest value of r as the maximum specific growth rate for that strain. To determine the fluorescent protein (e.g., GFP) production rate per cell, p , we repeated this procedure while fitting fluorescence values to the equation: $F(t) = F_0 + C_0 (p/r) (e^{rt} - 1)$. F_0 is the initial fluorescence and $F(t)$ is the fluorescence at time t . This equation is derived by integrating the relationship $dF/dt = p C(t)$. We fit F_0 and p in this model to the data while keeping C_0 and r fixed to the values determined from the OD curve fit for the corresponding time window. Again, we recorded the largest value of p across all time points as the maximum fluorescent protein production rate.

To account for plate-to-plate variation in growth and GFP production rate estimates (Supplementary Figs. 2A, and 8A), we normalized measurements made on different plates. In our experimental design, a majority of the plasmids tested in each plate are expected to exhibit negligible burden. This let us estimate the growth and GFP production rates corresponding to ‘no-burden’ for a given plate by examining the distributions of values measured. Specifically, we calculated the density distributions of growth and GFP production rates using a Gaussian kernel function with bandwidths of 0.014 and 300, respectively, for all non-control strains. To account for multimodal distributions, we took the no-burden value as the highest value among all peaks in the density distribution that were at least 50% as high as the highest peak. Then, we normalized all rate estimates by dividing them by the corresponding no-burden value for that plate (Supplementary Figs. 2B, and 8B). The final distributions of the mean values for each BioBrick plasmid have a major peak at the no-burden value with a noticeable shoulder of strains with a slightly decreased growth rate or

GFP production rate, in addition to some strains with much lower values (Supplementary Figs. 2C, and 8C).

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 emission⁹². 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 rejecting the null hypothesis that all of a plasmid's burden was from utilizing the host cell's gene expression resources.

BioBrick source organism analysis

We classified BioBricks into eight categories depending on the organisms in which their sequences originated (*E. coli*, other Enterobacterales, other γ -proteobacterium, other proteobacterium, other bacterium, fungus, plant, animal, or synthetic) (Supplementary Data 3). BioBricks with multiple parts were classified according to the most divergent category. Virus- and phage-derived sequences were classified with their host species, and codon-optimized constructs were counted with the species in which the amino acid sequence originated. We considered two additional categorization schemes that reduced the number of organism categories to four or two, plus four variants of each of the three categorization schemes in which we did not include features that were not protein-coding genes or did not include fluorescent proteins in assigning the organism of origin to a BioBrick (Supplementary Fig. 4). For all twelve overall classification schemes, we examined whether there were any trends in burden with respect to source organism in two ways. First, we performed a Kruskal–Wallis test to determine whether BioBricks with different categories of source organisms had systematic differences in their normalized growth rates. Second, we determined whether the chances that BioBricks were in the set of 59 burdensome BioBricks differed across organism categories by comparing binomial regression models with and without this factor using likelihood-ratio tests. *P*-values for all twenty-four statistical tests are provided in Supplementary Table 1.

Statistics & reproducibility

BioBrick plasmids from the iGEM DNA Distribution Kit were transformed based on their locations in these microplates without first examining their sequences or functions. Outlier measurements in the burden assays were identified by manually examining growth curve fits and omitted from the analysis, as documented in the raw data files⁹³. No statistical method was used to predetermine the sample size. The experiments were not randomized. The Investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw and processed data files and plasmid assemblies are available in a GitHub repository that has been archived in Zenodo (<https://doi.org/10.5281/zenodo.11528027>)⁹³. Raw plasmid and genome sequencing data are available from the NCBI Sequence Read Archive (Accession: PRJNA1090925). Source data are provided with this paper.

Code availability

Simulation code and analysis scripts are available in the associated GitHub repository⁹³.

References

- 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).
- Andrews, L. B., Nielsen, A. A. K. & Voigt, C. A. Cellular checkpoint control using programmable sequential logic. *Science* **361**, eaap8987 (2018).
- Ryu, M. et al. Control of nitrogen fixation in bacteria that associate with cereals. *Nat. Microbiol.* <https://doi.org/10.1038/s41564-019-0631-2> (2019).
- Isabella, V. M. et al. Development of a synthetic live bacterial therapeutic for the human metabolic disease phenylketonuria. *Nat. Biotechnol.* **36**, 857–864 (2018).
- Leonard, S. P. et al. Engineered symbionts activate honey bee immunity and limit pathogens. *Science* **367**, 573–576 (2020).
- Arkin, A. P. & Fletcher, D. A. Fast, cheap and somewhat in control. *Genome Biol.* **7**, 114 (2006).
- Renda, B. A., Hammerling, M. J. & Barrick, J. E. Engineering reduced evolutionary potential for synthetic biology. *Mol. Biosyst.* **10**, 1668–1678 (2014).
- 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).
- Umenhoffer, K. et al. Reduced evolvability of *Escherichia coli* MDS42, an IS-less cellular chassis for molecular and synthetic biology applications. *Microb. Cell Fact.* **9**, 38 (2010).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- Glick, B. R. Metabolic load and heterologous gene expression. *Biotechnol. Adv.* **13**, 247–261 (1995).
- 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**, 1653–1660 (2008).
- 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).

19. Gorochowski, T. E. et al. Absolute quantification of translational regulation and burden using combined sequencing approaches. *Mol. Syst. Biol.* **15**, 1–15 (2019).
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. Proteom.* **10**, M111.007930 (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* **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. USA* **109**, E2774–E2783 (2012).
38. Drake, J. W. A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl Acad. Sci. USA* **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. Horton, J. S., Flanagan, L. M., Jackson, R. W., Priest, N. K. & Taylor, T. B. A mutational hotspot that determines highly repeatable evolution can be built and broken by silent genetic changes. *Nat. Commun.* **12**, 6092 (2021).
42. 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).
43. Nyerges, Á. et al. CRISPR-interference-based modulation of mobile genetic elements in bacteria. *Synth. Biol. Oxf. Engl.* **4**, ysz008 (2019).
44. 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).
45. Shetty, R. P., Endy, D. & Knight, T. F. Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* **2**, 5 (2008).
46. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
47. Rouches, M. V., Xu, Y., Cortes, L. B. G. & Lambert, G. A plasmid system with tunable copy number. *Nat. Commun.* **13**, 3908 (2022).
48. Galdzicki, M., Rodriguez, C., Chandran, D., Sauro, H. M. & Gennari, J. H. Standard biological parts knowledgebase. *PLoS ONE* **6**, e17005–e17005 (2011).
49. Chen, Y.-J. et al. Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods* **10**, 659–664 (2013).
50. Tarnowski, M. J. & Gorochowski, T. E. Massively parallel characterization of engineered transcript isoforms using direct RNA sequencing. *Nat. Commun.* **13**, 434 (2022).
51. Borkowski, O. et al. Cell-free prediction of protein expression costs for growing cells. *Nat. Commun.* **9**, 1457 (2018).
52. 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).
53. 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).
54. Peccoud, J. et al. Essential information for synthetic DNA sequences. *Nat. Biotechnol.* **29**, 22–22 (2011).
55. Thuronyi, B. W., DeBenedictis, E. A. & Barrick, J. E. No assembly required: time for stronger, simpler publishing standards for DNA sequences. *PLoS Biol.* **21**, e3002376 (2023).
56. 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).
57. LaFleur, T. L. Automated model-predictive design of synthetic promoters to control transcriptional profiles in bacteria. *Nat. Commun.* **13**, 5159 (2022).
58. Sorek, R. et al. Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* **318**, 1449–1452 (2007).
59. Zhuang, K., Vemuri, G. N. & Mahadevan, R. Economics of membrane occupancy and respiro-fermentation. *Mol. Syst. Biol.* **7**, 500 (2011).
60. Shachrai, I., Zaslaver, A., Alon, U. & Dekel, E. Cost of unneeded proteins in *E. coli* is reduced after several generations in exponential growth. *Mol. Cell* **38**, 758–767 (2010).
61. Barrick, J. E. et al. Daily transfers, archiving populations, and measuring fitness in the long-term evolution experiment with *Escherichia coli*. *J. Vis. Exp.* <https://doi.org/10.3791/65342> (2023).
62. Chochinov, C. A. & Nguyen Ba, A. N. Bulk-fitness measurements using barcode sequencing analysis in yeast. In *Yeast Functional Genomics* (ed. Devaux, F.) 2477–399–415 (Springer US, New York, NY, 2022).
63. Li, F., Tarkington, J. & Sherlock, G. Fit-Seq2.0: an improved software for high-throughput fitness measurements using pooled competition assays. *J. Mol. Evol.* **91**, 334–344 (2023).

64. Joshi, S. H.-N., Yong, C. & Gyorgy, A. Inducible plasmid copy number control for synthetic biology in commonly used *E. coli* strains. *Nat. Commun.* **13**, 6691 (2022).
65. Ceroni, F. et al. Burden-driven feedback control of gene expression. *Nat. Methods* **15**, 387–393 (2018).
66. Barajas, C., Huang, H.-H., Gibson, J., Sandoval, L. & Del Vecchio, D. Feedforward growth rate control mitigates gene activation burden. *Nat. Commun.* **13**, 7054 (2022).
67. Rugbjerg, P., Sarup-Lytzen, K., Nagy, M. & Sommer, M. O. A. Synthetic addiction extends the productive life time of engineered *Escherichia coli* populations. *Proc. Natl Acad. Sci. USA* **115**, 2347–2352 (2018).
68. Segall-Shapiro, T. H., Meyer, A. J., Ellington, A. D., Sontag, E. D. & Voigt, C. A. A ‘resource allocator’ for transcription based on a highly fragmented T7 RNA polymerase. *Mol. Syst. Biol.* **10**, 742–742 (2014).
69. Wang, K., Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat. Biotechnol.* **25**, 770–777 (2007).
70. Orelle, C. et al. Protein synthesis by ribosomes with tethered subunits. *Nature* **524**, 119–124 (2015).
71. Ellis, T. Predicting how evolution will beat us. *Microb. Biotechnol.* **12**, 41–43 (2019).
72. Suárez, G. A., Renda, B. A., Dasgupta, A. & Barrick, J. E. Reduced mutation rate and increased transformability of transposon-free *Acinetobacter baylyi* ADP1-ISx. *Appl. Environ. Microbiol.* **83**, e01025–17 (2017).
73. Csörgo, B., Fehér, T., Tímár, E., Blattner, F. R. & Pósfai, G. Low-mutation-rate, reduced-genome *Escherichia coli*: an improved host for faithful maintenance of engineered genetic constructs. *Microb. Cell Fact.* **11**, 11–11 (2012).
74. Battaglini, B., Arduino, A. & Pagliano, C. Mathematical modeling for the design of evolution experiments to study the genetic instability of metabolically engineered photosynthetic microorganisms. *Algal Res.* **52**, 102093 (2020).
75. Nuismer, S. L., C Layman, N., Redwood, A. J., Chan, B. & Bull, J. J. Methods for measuring the evolutionary stability of engineered genomes to improve their longevity. *Synth. Biol.* **6**, ysab018 (2021).
76. Nikolados, E.-M., Weiße, A. Y., Ceroni, F. & Oyarzún, D. A. Growth defects and loss-of-function in synthetic gene circuits. *ACS Synth. Biol.* **8**, 1231–1240 (2019).
77. Hernández-Beltrán, J. C. R., San Millán, A., Fuentes-Hernández, A. & Peña-Miller, R. Mathematical models of plasmid population dynamics. *Front. Microbiol.* **12**, 606396 (2021).
78. Nyström, A., Papachristodoulou, A. & Angel, A. A dynamic model of resource allocation in response to the presence of a synthetic construct. *ACS Synth. Biol.* **7**, 1201–1210 (2018).
79. Johnson, P. *adaptivetau: Tau-leaping Stochastic Simulation*. <https://cran.r-project.org/package=adaptivetau> (2019).
80. Chang, W. et al. *shiny: Web Application Framework for R*. <https://shiny.posit.co/> (2024).
81. Haldimann, A. & Wanner, B. L. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J. Bacteriol.* **183**, 6384–6393 (2001).
82. Deatherage, D. E. & Barrick, J. E. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using *breseq*. *Methods Mol. Biol.* **1151**, 165–188 (2014).
83. Barrick, J. E. et al. Identifying structural variation in haploid microbial genomes from short-read resequencing data using *breseq*. *BMC Genom.* **15**, 1039 (2014).
84. Knight, T. *Idempotent Vector Design for Standard Assembly of BioBricks*. <https://dspace.mit.edu/handle/1721.1/21168> (2003).
85. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb. Genom.* **3**, e000132 (2017).
86. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
87. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Comput. Biol.* **13**, e1005595 (2017).
88. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
89. McGuffie, M. J. & Barrick, J. E. pLannotate: engineered plasmid annotation. *Nucleic Acids Res.* **49**, W516–W522 (2021).
90. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
91. Hecht, A., Endy, D., Salit, M. & Munson, M. S. When wavelengths collide: bias in cell abundance measurements due to expressed fluorescent proteins. *ACS Synth. Biol.* **5**, 1024–1027 (2016).
92. Alieva, N. O. et al. Diversity and evolution of coral fluorescent proteins. *PLoS ONE* **3**, e2680 (2008).
93. Noor Radde et al. Measuring the burden of hundreds of BioBricks defines an evolutionary limit on constructability in synthetic biology. *GitHub Repository: barricklab/igem2019*. <https://doi.org/10.5281/zenodo.11528027> (2024).

Acknowledgements

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 Giao Chau Nguyen, Vrinda Rajkumar, Marco Sanchez, Jeremy 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. This research was supported by the National Science Foundation (CBET-1554179 to J.E.B., IOS-2103208 to J.E.B., and MCB-2123996 to J.E.B.), the National Science Foundation BEACON Center for the Study of Evolution (DBI-0939454 subcontract to J.E.B.), the National Institutes of Health (R01GM088344 to J.E.B.), and the U.S. Army Research Office (W911NF-20-1-0195 to J.E.B.). 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.

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.

Competing interests

M.J.M. is employed by Plasmidsaurus, which provides plasmid sequencing services for synthetic biology. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-50639-9>.

Correspondence and requests for materials should be addressed to Jeffrey E. Barrick.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024