nature biotechnology

PERSPECTIVE

https://doi.org/10.1038/s41587-019-0171-6

Overcoming genetic heterogeneity in industrial fermentations

Peter Rugbjerg and Morten O. A. Sommer*

Engineering the synthesis of massive amounts of therapeutics, enzymes or commodity chemicals can select for subpopulations of nonproducer cells, owing to metabolic burden and product toxicity. Deep DNA sequencing can be used to detect undesirable genetic heterogeneity in producer populations and diagnose associated genetic error modes. Hotspots of genetic heterogeneity can pinpoint mechanisms that underlie load problems and product toxicity. Understanding genetic heterogeneity will inform metabolic engineering and synthetic biology strategies to minimize the emergence of nonproducer mutants in scaled-up fermentations and maximize product quality and yield.

Industrial-scale biomanufacturing of therapeutics, enzymes and chemicals relies on cultivating large volumes of engineered cells in fed-batch or continuous bioprocesses. Clonal expansion of high-performing producer cells can favor the emergence of low-producing escape variants that have shed the load from metabolic burden or toxicities imposed by product synthesis or secretion[1]. Eventually, these low-producing or nonproducing variants can reduce product yield and quality, making genetic homogeneity a crucial, yet often overlooked parameter of process scale-up. Reliable scale-up remains an enormous technical and economic barrier for commercialization of bioprocesses with genetically engineered organisms[2–4] and requires substantial research and development, as exemplified by the complex production scale-up for 1,4-butanediol by Genomatica[5] and β-farnesene by Amyris[6] (Box 1).

Genetic heterogeneity has been reported in industrial processes producing penicillin with *Penicillium chrysogenum*[7], L-arginine with *Corynebacterium*[8] and human antibodies with Chinese hamster ovary (CHO) cell lines[9]. At Genentech, for example, early mutation during CHO transfection resulted in up to 27% of the cell population producing an antibody with a single amino acid substitution[9]. Genetic heterogeneity can be problematic at industrial scales owing to the production 'load' on producer lines, which arises from metabolic burden and toxicities[10,11]. The reduced growth rate associated with load means that it is a crucial cell factory parameter. Load affects yield, rate and titer at scale, and reductions in load can bring huge benefits. For example, reducing load from 30% to 25% would prevent a 17% decline in yield and titer over the 60 generations that are required for large-scale batch fermentation in a typical industrial biotech process (Fig. 1)[11]. Controlling the underlying spontaneous mutation rate (escape rate) is also important. Escape rate depends on the susceptibility of the expression constructs to host mutation modes. Together, load and escape rate depend on the nature of the product, genetic design, host organism and growth conditions, although general principles apply across products.

The decreasing costs and improved precision of DNA sequencing enable its use in cultured production populations to pinpoint and quantify genetic escape rates of subpercentage heterogeneities in even small-scale cultures[11]. Such subpopulation-level mutant detection enables strain engineers to identify problematic genetic instability early in the development process (Fig. 1) and visualize the dynamics of mutated subpopulations to determine how and why they outcompete the initial producer clone[11]. Together with recently developed synthetic biology approaches to metabolism-responsive cell growth[12,13] and pathway regulation[14,15], developments in deep DNA sequencing present opportunities for detecting and minimizing genetic heterogeneity in bioprocesses and in turn improving fermentation performance at production scale.

Here we discuss the diverse mechanisms underlying the problem of genetic heterogeneity in industrial biotech scale-up. We provide basic and applied examples of production load and escape, and review potential strategies to detect and mitigate problems with genetic heterogeneity.

Metabolic burdens, toxicities and production load

Engineered biotechnological production usually reduces cell fitness (growth rate, viability and maximum cell density) relative to parental nonproducing or low-producing cells. Overall, production load (Fig. 2a) often exceeds 15–30% in high-yielding processes (Table 1). Replication burden from introducing additional DNA can result from depleted initiation factors, nucleotides and DNA polymerase[16]. The bioenergetic cost of replicating chromosomal or plasmid gene copies per se is generally negligible in *Escherichia coli* and haploid *Saccharomyces cerevisiae*[17–19]. However, replicative or plasmid burden can be confounded by downstream 'costs' of encoded proteins.

Production load can be sizable and can limit production rates. Loads usually scale with the amount of protein produced[18,19]. Overall costs can be divided into transcriptional (from competition for free RNA polymerase and ribonucleotides) and translational (from competition for free ribosomes and charged tRNAs[19]). Post-translation, protein-related loads result from protein folding and energy required for secretion[20] or glycosylation. Nonfunctional, misfolded proteins waste synthesis capacity and substrate, and can induce cytotoxicity and stress or unfolded protein responses[21,22]. Intracellular protein toxicity is often observed when heterologous protein makes up 30% of the internal proteome[19]. This toxicity may be exacerbated by enzymatically active products, host interactions (for example, membrane integration[23] or DNA binding[24]) and aggregate formation. Protein burden can be quantified by fitness measurements, such as introducing point mutations or stop codons to estimate effects from transcription cost or catalytic function[19,25]. Underlying mechanisms of load can be investigated by transcriptomics[26,27].

Metabolite-producing pathways also compete with endogenous pathways to impose a load on productive cells (Fig. 2a). Many

The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, Denmark. *e-mail: msom@bio.dtu.dk

Box 1 | Maintaining genetic integrity in industrial isoprenoid strains

Hallmark metabolic engineering projects for the commercial production of the antimalarial drug precursor amorphadiene at Sanofi, and β-farnesene at Amyris, suffered from initial low cell viability owing to several toxic pathway intermediates[6,89]. Pathways were initially designed in *E. coli*, but toxicity of intermediates led to problems with low cell densities in fermentation. Enzyme expression had to be rebalanced, and membrane-stabilizing chemicals were added to growth media to overcome a stress[90], which appeared to select for genetic heterogeneities in the production pathway genes[11].

S. cerevisiae was a better production host for these pathways, but producer populations were compromised by low cell viability and emergence of 'broken' nonproducer mutants. Heterogeneity in yeast producer populations was mitigated at Amyris by developing a switch to control precursor availability, thereby effectively postponing burden until an optimal time for maximal yield[91].

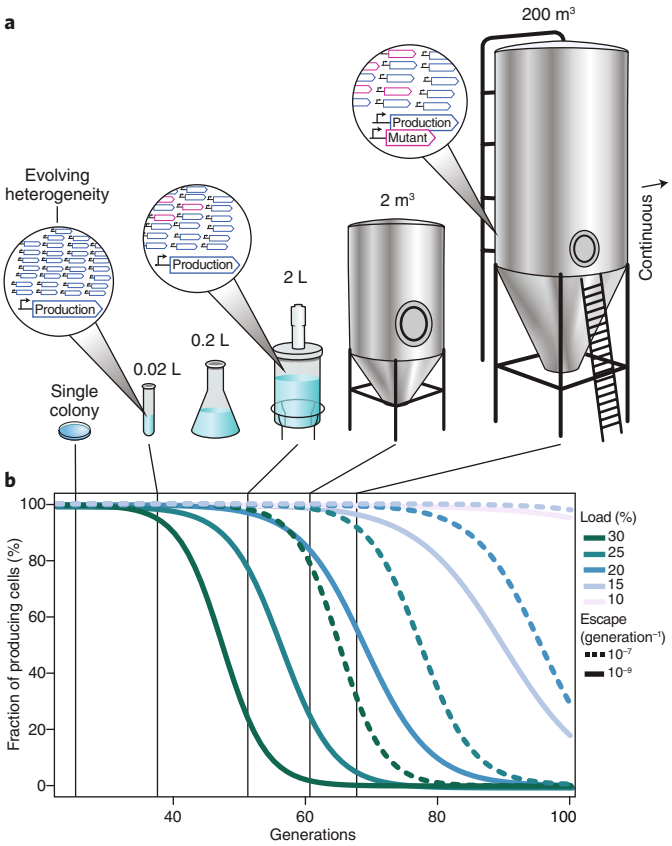

Fig. 1 | Evolution of populations during industrial fermentation. a, The fitness cost of high-level production creates a selective advantage for nonproducer cells that lose the ability to produce the desired product during industrial fermentation. Deep DNA sequencing can be applied to track evolving genetic heterogeneity early in a population of growing production cells. **b**, Mathematical models can predict the evolution of heterogeneity in the fermentation population over time[11], leading to reduction in the fraction of producing population, assuming a spontaneous escape rate for complete production decline and alleviated load (fitness cost relative to nonproducing cell).

engineered pathways drain cellular energy and central substrates such as acetyl-CoA, ATP, cofactors and coenzymes such as *S*-adenosyl methionine, NADPH and iron-sulfur clusters (Table 1). Such depletion may also force metabolic flux toward unintended, toxic metabolic nodes, such as acetate[28]. Metabolite depletions and build-ups can be investigated by extracellular supplementation[27] and time-course metabolic flux analysis[29]. In metabolite-producing fermentations, the toxicity of pathway intermediates and byproducts often results in a substantial production load because these compounds are intracellular (Table 1)[30,31]. Pathway toxicity can be investigated by measuring the fitness of strains after sequential enzyme introduction[25] or extracellular supplementation, such as with homocysteine in the L-methionine pathway[32]. The mechanism behind toxicity is not always obvious, but functional genomics and transcriptomics can aid in its understanding[33].

End product titers in successfully commercialized fermentation processes reach >50–100 g/l, but at this concentration most molecules are toxic (Table 1); of note, intracellular concentrations may be even higher.

Genetic error modes in bioprocesses

All biological populations, whether engineered or not, undergo continuous evolution by various mechanisms at rates ranging from 10^{-2} to 10^{-10} per base pair per generation (Table 2). In fermentations to produce metabolites and enzymes, base substitutions mainly cause loss of enzyme activity. CHO cell lines used for mammalian therapeutic protein production exhibit loss-of-function mutation rates of roughly 10^{-8} per base pair per generation[34]. Such rates can be sufficient to compromise product quality, even at small cultivation scales, as demonstrated by cell lines with point mutations in genes encoding antibody fragments[35–37]. Bacterial and fungal hosts have basal base-substitution rates of 10^{-10} per base pair per generation (Table 2)[38]. Even so, genomic regions carrying a production load may appear as mutation hotspots because of subsequent selection of a single mutation event; in protein-producing *E. coli*, base substitutions that affect the function of the LacI repressor-coding gene lead to lower production[23]. At the sequence level, hairpin structures and mononucleotide repeats as short as three bases can promote escape via DNA polymerase slippage, increasing the probability of insertion/deletions (indels) (Fig. 2b, Table 2)[39,40].

Homologous recombination frequently causes production instability owing to reuse of regulatory promoter/terminator sequences[41] and multicopy chromosomal integration of production cassettes. In some producer cells, chromosomal gene copy numbers are amplified during construction but cannot be maintained under production conditions. In these cases, gene loop-outs result in production decline over time, as observed in penicillin-producing *Penicillium chrysogenum* at GlaxoSmithKline[7] or protein-producing CHO cells[42]. Catalyzed by endogenous DNA repair pathways, recombination between direct repeats can lead to deletion or tandem duplication of the flanked region, while inverted repeats recombine and form inversions (Fig. 2b). Gene conversions between highly similar regions also occur (for example, weakening an *E. coli* product promoter by recombination with its wild-type promoter variant[43]). Recombination is induced by local transcription and increases with repeat lengths accelerating from around 20 base pairs[44,45] (Table 2). Nonhomologous recombination occurs at lower rates and also operates via strand slippage, but is generally constrained to short, intramolecular repeats (Table 2)[44,46].

Transposable elements present in the chromosomes of microbial and mammalian cells can move spontaneously, causing gene disruption or modulating transcription of flanking genes[47] (Fig. 2b). Bacterial insertion sequence (IS) elements present in most industrial workhorse bacteria can disrupt production genes or operons. Different IS subfamilies transpose at different rates, using various mechanisms and target site specificities[47] (Table 2), and can fully disrupt a heterologous mevalonate pathway in 70 generations[11].

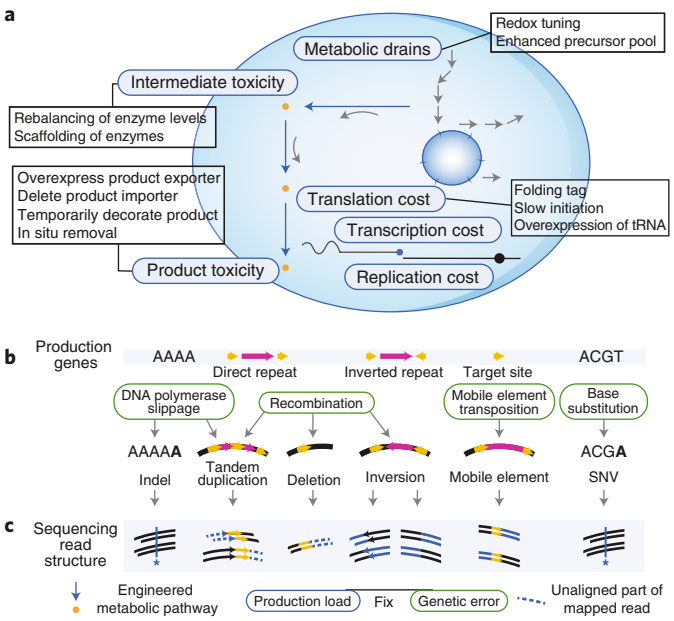

Fig. 2 | Causal factors and counterstrategies for genetic heterogeneity in large-scale biomanufacturing. a, Spontaneous escaper cells are enriched due to production loads resulting from metabolic burdens (for example, resource pool depletion) and toxicities associated with engineered production. **b**, Spontaneous genetic escape modes. **c**, Subpopulations of escaper cells can be identified following deep sequencing by analysis of broken unaligned read ends (dashed blue) that otherwise map to the reference (black) and often involves a repeated region (orange). SNV, single-nucleotide variant; *, base substitution.

Similarly, *S. cerevisiae* hosts about 50 transposable Ty elements[48] (Table 2) and 38% of the CHO K1 genome is transposable elements[49]. The potential long-term biotechnological impact of eukaryotic transposable elements is unknown. Instead, epigenetic regulation of transgenes is thought to trigger instability in CHO cultures[50]. Plasmid-borne expression systems in antibiotic-free biotechnological production previously led to considerable loss rates (Table 2). Segregational plasmid instability can be solved by chromosomal integration or plasmid addiction systems[51].

In industrial biomanufacturing, escape rates may be higher than in research laboratory conditions because of extracellular stresses from shearing, osmotic pressure, pH and starvation[10,52]. Intracellular production and associated reactive oxygen species can activate bacterial DNA-damage, chaperone (heat shock) and SOS responses[53,54] that induce mobile element transposition and more error-prone DNA polymerases with 10^2–10^3-fold higher base-substitution rates[55,56] (Table 2). In eukaryotic hosts, protein production can induce stress functions related to the unfolded protein response[57]. Overall, industrial bioprocesses are challenged by a wide range of mechanisms that generate genetic heterogeneity in the production population.

Detection of genetic heterogeneity

Diagnosing the genetic liabilities of a cell factory is crucial to commercializing and optimizing existing and new processes. Deviation in large-scale fermentation behavior is monitored by extensive quality control programs, often following biopharmaceutical guidelines[58]. These internal programs track process stability by online analyses of multivariate parameters, such as growth rates and broth metabolites[59] to detect unstable seed trains or guide ideal batch runtimes. To determine fermentation stability, end-of-production cells are banked and compared with the working clone bank for titer and maintenance of production genes. This phenotyping can be complemented with second- and third-generation DNA sequencing methods to investigate mutation accumulation during culture of production and model organisms. This work promises to identify and quantify load-carrying genes and genetic error modes specific to certain clones, designs or processes[11,36,60].

Identification of mutations in single colonies from end-of-production samples often requires only limited sequencing coverage of 20–30-fold depth because clones are assumed to be homogenous[23,43]. However, multiple competing mutations may co-exist in a cultivated population, requiring sampling of many clones. Deep DNA sequencing of entire production populations over time enables in situ quantification of competing subpopulations, but resolution of true genetic variation in populations can be challenging. Currently, the lowest detected genetic production errors are at frequencies of 10^{-4} (refs. [11,36]), but limits depend on variant type, sequencing technique, coverage, read length and error rate. For routine, short-read Illumina sequencing, true base substitutions below 10^{-2} frequency are difficult to distinguish from artificial errors that arise during sample library preparation and sequencing[61,62]. For this reason, base substitutions and indels are often only called above 10^{-3}–10^{-2} frequencies[36,62]. Comparison with a homogenous sample, or samples collected over time, aids in discriminating between true and artificial variants that lack statistically significant population dynamics[11,62]. Sequencing errors can be better distinguished from true mutant subpopulations by attaching short, unique molecular barcode sequences to sample fragments before sequencing, improving the detection limit to around 10^{-7} (ref. [62]). In contrast to base substitutions, the >10 base pair split-ended signatures of structural variants allow more unambiguous detection (for example, down to 10^{-4} frequency for transposable element insertions[11]; Fig. 2c). Combined with a match from the opposing junction, such split-end reads confidently discriminate among several structural variant forms (Fig. 2). Split ends can be ambiguous, for instance, at tandem duplications. Also, structural variants in populations are difficult to accurately quantify by short reads because of problems with establishing the true sequence coverage for structural variants with different structural compositions[63]. However, high sensitivity may enable detection of detrimental structural variants as early as in the master clone bank. Identification of such detrimental mutations in master or working clone banks could allow selection of mutation-reduced inocula. The high sequence coverage (>10^4-fold) requirements of these detection methods may necessitate targeted sequencing of production pathway at the risk of overlooking unexpected escape mechanisms. Third-generation long-read technologies, such as single-molecule real-time and nanopore sequencing, are likely to improve structural variant resolution and have been used to detect low-frequency base substitution in CHO production genes[36].

Mass spectrometry is used to assess therapeutic protein heterogeneity and can also identify non-genetic heterogeneities that occur through mistranslation at high sensitivity, such as down to 0.1% at Roche[37]. De novo screening by mass spectrometry is difficult[9], but untargeted analyses at, for instance, Gilead revealed genetic heterogeneities of unexpected short peptide extensions to antibody fragments[64], suggesting a scope of gene variability beyond routinely assessed amino acid misincorporation[65]. Flow cytometry is also used for real-time quantification of population heterogeneity: staining of product with fluorescent dyes uncovered low-producing CHO subpopulations at Janssen[66].

Reduction of genetic escape rates

Once genetic heterogeneity is diagnosed, several bottom-up strategies to overcome it can be used. IS elements are one common concern in bacterial production strains and have been deleted (for example, in *E. coli* and *Corynebacterium glutamicum*[67,68]). Resulting

Table 1 | Loads that affect producer fitness

Load	Method to diagnose load	Metabolite or protein affecting fitness	Reference
Intermediate or byproduct toxicity (end product: toxic metabolite)	Step-by-step pathway introduction	1,3-propanediol: methylglyoxal	31
		β-farnesene: farnesyl pyrophosphate	91
	Pathway deep-seq	Mevalonic acid: HMG-CoA	25
	Functional genomics	L-methionine: Homocysteine	32
	Supplementation	Resveratrol: *p*-coumaric acid	92
End-product toxicity	Supplementation	Butanol	93
	Pathway deep-seq	Artemisinic acid	94
		Vanillin	82
		L-serine	95
Protein folding stress and toxicity	Transcriptomics	Orotidine-5′-phosphate decarboxylase	22
	Proteomics	Immunoglobulin G	57
Endogenous substrate, energy, cofactor depletion	Metabolic flux analysis	Cytochrome P450: Heme pool	27
	Supplementation	Spider silk protein: $tRNA^{Gly}$ pool	78
	Transcriptomics	L-lysine: triacetic cycle intermediate pools	96
		Succinic acid: glutamate pool	97
		Fatty acids: unbalanced NADH pool	79
Protein burden	Metabolic flux analysis	Green fluorescent protein	19
	Transcriptomics	Insulin analog precursor	26
		Maltase	20
Replication burden		Selection gene product	17
		Depleted replication factors	16

IS-free hosts propagate biosynthetic pathways and protein-coding genes more stably. Genomatica eliminated chromosomal IS elements in their 1,4-butanediol-producing *E. coli* to increase genome stability[5]. And the synthetic yeast genome project Sc2.0 is synthesizing *S. cerevisiae* chromosomes de novo without retrotransposon Ty elements, which may curb genetic heterogeneity in future yeast bioprocesses[48]. CRISPR-based engineering has considerably shortened the time to remove *E. coli* IS activity[69]. It should be noted that not all common IS subfamilies are problematic[11]. Although structural variants, including transposable elements, appear to be a common type of escape mutations that is important to prevent, the base-substitution rate can be improved by deletion of stress-responsive, more error-prone DNA polymerases[67]. Another promising strategy uses directed evolution to lower plasmid mutation rates[70].

For synthetic genes, avoiding hypermutable sequence motifs can be achieved by codon optimization. Short repeats and mononucleotide stretches (>3 base pairs) should be eliminated to minimize slipping-strand mutagenesis[39] (Table 2), although their impact in bioprocesses remains unknown. Homologous recombination has been reduced in commercialized *E. coli* strains by deleting *recA*. This strategy is particularly relevant for homologous regions >20 base pairs, which are commonly repeated in plasmids and chromosomes[23], but non-*recA* pathways also mediate recombination[44]. Separation of repeated, chromosomally integrated genes by essential genetic regions may deselect direct loop-outs[71]. Recombination between several sequence copies may also be reduced by limiting sequence identity using heterologous promoter variants[72] or synonymously coded protein variants[73]. Multiple gene copies may offer more efficient redundancy against escape, when they are chromosomally integrated, compared to multicopy plasmids where uneven segregation might amplify intracellular escape variants. However, chromosomal integrations should target stable regions, as identified in CHO cells at Novo Nordisk[74]. Bacterial genes should be encoded on the leading strand to avoid mutagenic head-on encounters between DNA and RNA polymerases[45]. Finally, inoculating fermentations from clone banks based on single cells is essential to limit early mutation accumulation by creating an evolutionary bottleneck.

Reduction of production load

Several options are available to limit excessive production load, depending on its biological cause. Same-species co-cultures that divide pathway labor at an easily diffusible intermediate product may reduce the metabolic burden and improve production[75], potentially by restricting promiscuous reactions. However, these strategies in turn require careful population management.

For cost-effective protein expression, slower transcription and slower initial translation generally reduce costs per synthesized protein in *E. coli*[76] by increasing transcript reusage and reducing the overall translational rate. This strategy can be combined with codon optimization to avoid tRNA depletion. Protein costs can also be reduced by a combination of decreased growth temperature, folding-enhancing tags, and overexpression of chaperones and rare tRNAs[77], such as in production of glycine-rich spider silk proteins in *E. coli*[78].

Many nutrient pools are intricately connected to central carbon metabolism, which complicates their replenishment. Nonetheless, pool refilling, for example, of ATP or NADH, is possible by rebalancing central metabolic flux[29]. This strategy may be aided by metabolic modeling[28] or adaptive evolution when production is growth-coupled[10,79].

A potent strategy to limiting intermediate toxicities in metabolic pathways is balancing enzyme concentrations or efficiencies to minimize pools of toxic intermediates[25]. Ideally, toxic byproduct formation should be eliminated by using less promiscuous pathway enzymes[80] or through conversion into nontoxic substances, as was engineered for 1,3-propanediol[31]. Efficient extracellular product export and in situ removal may be important to reduce the load from end-product toxicity and can be achieved, for example, by functional exporter selection from metagenomic libraries[81]. Alternatively, intermittent product decoration can limit end-product toxicity, such as by glycosylation of vanillin in a biomimicking strategy at Evolva[82].

The effect of production loads can be reduced by delaying production through two-stage processes that separate biomass growth and production[83]. First-generation systems used external inducers such as isopropyl β-D-1-thiogalactopyranoside to

Table 2 | Genetic error modes that constrain production stability in bacteria and yeast

Type	Cause	Rate (per generation)	Reference
Disruption by mobile element	Site-selective transposition by *E. coli* IS element or *S. cerevisiae* Ty element	10^{-5}–10^{-8} gene^{-1}	11,52,98
Base substitution	Proof-reading DNA polymerase error Stress-induced DNA polymerase error	10^{-10} bp^{-1} 10^{-7}–10^{-8} bp^{-1}	55,99
Large gene deletion, conversion, tandem duplication or inversion	Recombination (wild type)	$<10^{-8}$ for 25 bp repeat^{-1} 10^{-5} for 200 bp repeat^{-1}	44
	Recombination (non-RecA/Rad51)	$<10^{-8}$ for 25 bp repeat^{-1} 10^{-7} for 200 bp repeat^{-1}	44
Replication slippage	>3-bp repeat	$>10^{-4}$ repeat^{-1}	40
Plasmid loss	Mononucleotide repeat Nonselective medium or no plasmid addiction	10^{-2}–$<10^{-6}$	100

Bp, base pair.

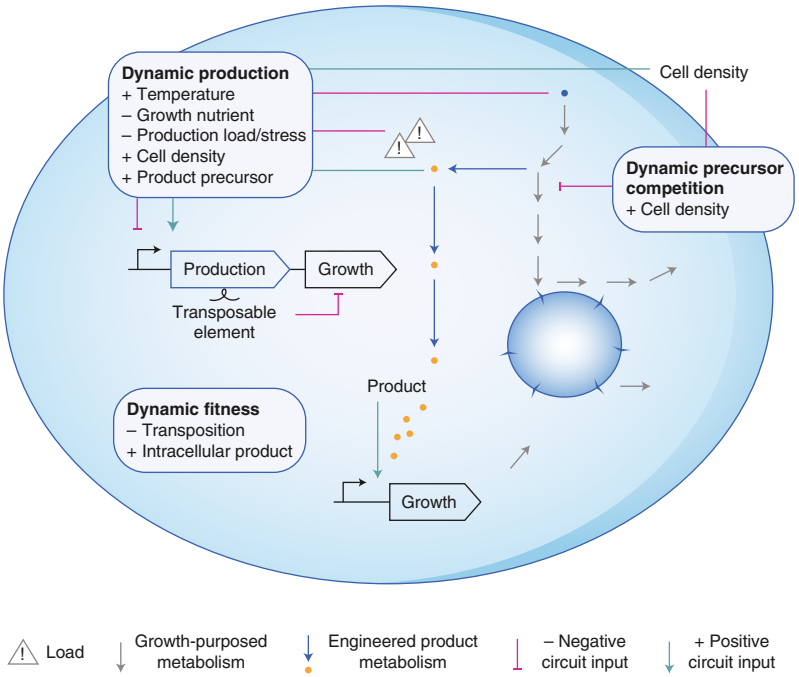

Fig. 3 | Synthetic control solutions. Circuits can reprogram cellular fitness, substrate flux or production to respond dynamically to process-specific stimuli functioning as negative or positive regulators.

initiate population-wide production at late fermentation phases; however, many inducers are cost-prohibitive in large-scale production[3]. Instead, production promoters can be repressed by limited nutrients such that they are only available during biomass formation[84]. Although generally very effective, induction systems are often leaky. The effect of cellular turnover in stationary phase is still not well understood but may permit late-stage population heterogeneity, which will require other solutions.

Synthetic circuits to minimize population heterogeneity

Cell factory synthetic circuits that regulate cell metabolism (or fitness) in response to extracellular, population-wide signals or stochastically changing, intracellular milieu of single cells create new opportunities to limit unfavorable heterogeneity by reducing production load or punishing subperforming populations. At the population level, changing temperature to degrade a production repressor can generate a two-stage process as was engineered for L-methionine at Evonik[85] (Fig. 3). Similarly, auto-induction at high cell densities can be programmed using quorum-sensing systems based on secretion and sensing of quorum molecules to replace physical or chemical induction[86].

At the single-cell level, new synthetic circuits have emerged that sense intracellular signals to dynamically control metabolic product pathways. Intracellular concentration and precursor or product flux[86], along with pathway stress and burden[15,87], have all been integrated to exert single-cell dynamic metabolic control (Fig. 3). These circuits feedback-regulate the engineered metabolism to improve production performance and reduce production load. For example, precursor stress can dynamically balance production pathways and surpass the production levels possible with constitutive ensemble expression[87] (Fig. 3). Growth-directed fluxes that compete with the product pathway may be dynamically regulated as metabolic valves: in response to adequate biomass buildup, flux competing with the pathway can be knocked down to substantially improve

yields[14] (Fig. 3). Future pathways may integrate numerous signals; double-layered circuits responsive to a pathway intermediate and quorum-sensed cell density demonstrate the potential of dynamic regulation to increase product titers[14] and separate growth from production[83] (Fig. 3).

In addition to feedback regulation, synthetic circuits can also address specific, recurring escape modes. Introducing an essential growth gene into a metabolite production gene operon (Fig. 3) can postpone production decline and perturb the composition of disruptive IS subfamilies[11]; however, rapid domination by alternative escape modes challenges stability if production load remains high.

Top-down strategies that are agnostic to specific escape modes are useful for synthetically managing long-term population dynamics based on intracellular performance using fitness-regulating genes. Circuits sensing a heterologous product can ensure that growth promotion occurs only in cells that accumulate a certain amount of product (Fig. 3)[12,13]. By engineering a circuit for population quality control, a phenotypically high-performing production population was sensed using a product-specific biosensor and enriched over time using antibiotic and auxotrophic selection genes to determine fitness[12]. The resulting population produced more fatty acid by exploiting nongenetic phenotypic variation. However, antibiotic resistance genes are precluded from most fermentation processes. Instead, non-conditionally essential genes can be regulated to achieve a self-sustained state of synthetic product addiction[13]. Product addiction alleviated diverse genetic heterogeneities during mevalonic acid production for at least 40 additional generations[13]. In product-addiction designs, the sensitivity of the product biosensor must be matched to the intracellular product concentration of producer populations. Product-addition designs are limited by the scant availability of natural biosensors, albeit new biosensors can be engineered[88]. Furthermore, synthetic fitness-regulating circuits may also escape owing to sensor load[24]. Therefore, tuning circuits to minimize growth abrogation under permissive conditions is key to ensuring long lifetimes. However, unlike related uses in synthetic library selection[81] and biocontainment, much milder selection pressures are necessary to maintain an initially pure production population, because contamination is fought only from within. Analogous to synthetic selections of enzyme libraries[81], redundant fitness regulators may better enforce an imposed production load and maintain homogeneity in large-scale or even continuous fermentation. With increasingly complex synthetic circuitry possible, several biological signals could be integrated into future cell factories to manage production loads and prevent genetic escape variants from enriching in a fermentation population.

Outlook

To realize stable, high-yielding and predictable cell factories at scale, engineered producer lines must avoid load and genetic escape. Early diagnosis and prioritization of evolving failure modes through deep DNA sequencing promises to bridge the gap between lab-scale strain development and scaled-up production. It is now feasible to identify mutation-prone, load-carrying genes, and platform- and product-specific genetic edits and conditions that enable stable populations both in early designs and commercialized long-term processes. To diagnose heterogeneity early, increasing sequencing precision and depth will be important. Combining short- and long-read sequencing technologies will likely improve resolution of structural variants. Unique molecular barcoding will increase confidence at currently uncharacterized depths of $<10^{-4}$ frequency where subtle sequencing biases may produce artefacts.

An emerging question will be the dynamics of intracellular heterogeneity when multiple gene copies are added for redundancy and increased production. Single-cell deep sequencing will improve our understanding of systems with a graded versus bimodal failure. Wider applications include prescreening new genetic designs, selecting mutationless clone banks and real-time screening of long-running or continuous bioprocesses. This might potentially be aided by machine learning trained with multivariate data from failed or prolonged fermentation batches to recognize population mutation patterns.

Avoiding load altogether in an engineered high-performing cell factory will be difficult and possibly unnecessary. Precise, late-stage switches could effectively postpone load if they strongly and swiftly induce the full population. Synthetic gene circuits may allow production organisms to withstand evolutionary pressures by regulating the fitness of diverging subpopulations in response to metabolic performance. Particularly high-performing populations could be synthetically promoted or maintained over time; however, complex synthetic circuits can also pose a load on the cell[24] and dynamic rewiring of metabolic flux and fitness with biosensors will require tuning to avoid eventual circuit collapse. Furthermore, more predictable natural and engineered biosensors for precursors and products must be developed, and their product sensitivity must be able to differentiate relevant intracellular concentrations in large-scale fermentors.

If signals are integrated correctly, fitness-regulating circuits promise to optimize any given single cell to perform to its maximum in the fermentation tank. These developments will contribute to closing the gap between early R&D and commercial production, enabling and accelerating processes toward the commercial scale.

Received: 7 October 2018; Accepted: 28 May 2019;
Published online: 08 July 2019

References

1. Glick, B. R. Metabolic load and heterologous gene expression. *Biotechnol. Adv.* **13**, 247–261 (1995).
2. Nielsen, J. & Keasling, J. Engineering cellular metabolism. *Cell* **164**, 1185–1197 (2016).
3. Lee, S. Y. & Kim, H. U. Systems strategies for developing industrial microbial strains. *Nat. Biotechnol.* **33**, 1061–1072 (2015).
4. Van Dien, S. From the first drop to the first truckload: Commercialization of microbial processes for renewable chemicals. *Curr. Opin. Biotechnol.* **24**, 1061–1068 (2013).
5. 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).
6. Leavell, M. D., McPhee, D. J. & Paddon, C. J. Developing fermentative terpenoid production for commercial usage. *Curr. Opin. Biotechnol.* **37**, 114–119 (2016).
7. Newbert, R. W., Barton, B., Greaves, P., Harper, J. & Turner, G. Analysis of a commercially improved *Penicillium chrysogenum* strain series: involvement of recombinogenic regions in amplification and deletion of the penicillin biosynthesis gene cluster. *J. Ind. Microbiol. Biotechnol.* **19**, 18–27 (1997).
8. Azuma, T., Nakanishi, T. & Sugimoto, M. Isolation and characterization of a stable L-arginine producer from continuous culture broth of *Corynebacterium acetoacidophilum*. *J. Ferment. Technol.* **66**, 279–284 (1988).
9. Harris, R. J. et al. Assessing genetic-heterogeneity in production cell lines: detection by peptide mapping of a low level Tyr to Gln sequence variant in a recombinant antibody. *Nat. Biotechnol.* **11**, 1293–1297 (1993).
10. Zelder, O. & Hauer, B. Environmentally directed mutations and their impact on industrial biotransformation and fermentation processes. *Curr. Opin. Microbiol.* **3**, 248–251 (2000).
11. Rugbjerg, P., Myling-Petersen, N., Porse, A., Sarup-Lytzen, K. & Sommer, M. O. A. Diverse genetic error modes constrain large-scale bio-based production. *Nat. Commun.* **9**, 787 (2018).
12. Xiao, Y., Bowen, C. H., Liu, D. & Zhang, F. Exploiting nongenetic cell-to-cell variation for enhanced biosynthesis. *Nat. Chem. Biol.* **12**, 339–344 (2016).
13. 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.* **115**, 2347–2352 (2018).
14. Gupta, A., Reizman, I. M. B., Reisch, C. R. & Prather, K. L. J. Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. *Nat. Biotechnol.* **35**, 273–279 (2017).
15. Ceroni, F. et al. Burden-driven feedback control of gene expression. *Nat. Methods* **15**, 387–393 (2018).

16. Yano, H. et al. Evolved plasmid-host interactions reduce plasmid interference cost. *Mol. Microbiol.* **101**, 743–756 (2016).
17. Karim, A. S., Curran, K. A. & Alper, H. S. Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications. *FEMS Yeast Res.* **13**, 107–116 (2012).
18. 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).
19. Kafri, M., Metzl-Raz, E., Jona, G. & Barkai, N. The cost of protein production. *Cell Rep.* **14**, 22–31 (2016).
20. Klein, T. et al. Overcoming the metabolic burden of protein secretion in *Schizosaccharomyces pombe* - A quantitative approach using 13C-based metabolic flux analysis. *Metab. Eng.* **21**, 34–45 (2014).
21. Villaverde, A. & Carrió, M. M. Protein aggregation in recombinant bacteria: biological role of inclusion bodies. *Biotechnol. Lett.* **25**, 1385–95 (2003).
22. Geiler-Samerotte, K. A. et al. Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. *Proc. Natl Acad. Sci.* **108**, 680–685 (2011).
23. Kwon, S. K., Kim, S. K., Lee, D. H. & Kim, J. F. Comparative genomics and experimental evolution of *Escherichia coli* BL21(DE3) strains reveal the landscape of toxicity escape from membrane protein overproduction. *Sci. Rep.* **5**, 16076 (2015).
24. Nielsen, A. A. K. et al. Genetic circuit design automation. *Science* **352**, aac7341–aac7341 (2016).
25. Pitera, D. J., Paddon, C. J., Newman, J. D. & Keasling, J. D. Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*. *Metab. Eng.* **9**, 193–207 (2007).
26. Kazemi Seresht, A. et al. Long-term adaptation of *Saccharomyces cerevisiae* to the burden of recombinant insulin production. *Biotechnol. Bioeng.* **110**, 2749–2763 (2013).
27. Michener, J. K., Nielsen, J. & Smolke, C. D. Identification and treatment of heme depletion attributed to overexpression of a lineage of evolved P450 monooxygenases. *Proc. Natl Acad. Sci. USA* **109**, 19504–19509 (2012).
28. Carneiro, S., Ferreira, E. C. & Rocha, I. Metabolic responses to recombinant bioprocesses in *Escherichia coli*. *J. Biotechnol.* **164**, 396–408 (2013).
29. Wu, G. et al. Metabolic burden: cornerstones in synthetic biology and metabolic engineering applications. *Trends Biotechnol.* **34**, 652–664 (2016).
30. Linster, C. L., Van Schaftingen, E. & Hanson, A. D. Metabolite damage and its repair or pre-emption. *Nat. Chem. Biol.* **9**, 72–80 (2013).
31. Zhu, M. M., Skraly, Fa & Cameron, D. C. Accumulation of methylglyoxal in anaerobically grown *Escherichia coli* and its detoxification by expression of the *Pseudomonas putida* glyoxalase I gene. *Metab. Eng.* **3**, 218–225 (2001).
32. Tuite, N. L., Fraser, K. R. & Byrne, C. P. O. Homocysteine toxicity in *Escherichia coli* is caused by a perturbation of branched-chain amino acid biosynthesis. *J. Bacteriol.* **187**, 4362–4371 (2005).
33. Kizer, L., Pitera, D. J., Pfleger, B. F. & Keasling, J. D. Application of functional genomics to pathway optimization for increased isoprenoid production. *Appl. Environ. Microbiol.* **74**, 3229–41 (2008).
34. Tindall, K. R. & Stankowski, L. F. Molecular analysis of spontaneous mutations at the gpt locus in Chinese hamster ovary (AS52) cells. *Mutat. Res. Genet. Toxicol.* **220**, 241–253 (1989).
35. Zhang, Z., Shah, B. & Bondarenko, P. V. G/U and certain wobble position mismatches as possible main causes of amino acid misincorporations. *Biochemistry* **52**, 8165–8176 (2013).
36. Cartwright, J. F., Anderson, K., Longworth, J., Lobb, P. & James, D. C. Highly sensitive detection of mutations in CHO cell recombinant DNA using multi-parallel single molecule real-time DNA sequencing. *Biotechnol. Bioeng.* **115**, 1485–1498 (2018).
37. Zeck, A. et al. Low level sequence variant analysis of recombinant proteins: An optimized approach. *PLoS One* **7**, e40328 (2012).
38. Lee, H., Popodi, E., Tang, H. & Foster, P. L. PNAS Plus: Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc. Natl Acad. Sci.* **109**, E2774–E2783 (2012).
39. Renda, B. A., Hammerling, M. J. & Barrick, J. E. Engineering reduced evolutionary potential for synthetic biology. *Mol. BioSyst.* **10**, 1668–1678 (2014).
40. Moxon, R., Bayliss, C. & Hood, D. Bacterial contingency loci: the role of simple sequence dna repeats in bacterial adaptation. *Annu. Rev. Genet.* **40**, 307–333 (2006).
41. Beekwilder, J. et al. Polycistronic expression of a β-carotene biosynthetic pathway in *Saccharomyces cerevisiae* coupled to β-ionone production. *J. Biotechnol.* **192**, 383–392 (2014).
42. Yoshikawa, T. et al. Amplified gene location in chromosomal dna affected recombinant protein production and stability of amplified genes. *Biotechnol. Prog.* **16**, 710–715 (2000).
43. Schlegel, S., Genevaux, P. & de Gier, J. W. De-convoluting the genetic adaptations of *E. coli* C41(DE3) in real time reveals how alleviating protein production stress improves yields. *Cell Rep.* **10**, 1758–1766 (2015).
44. Lovett, S. T., Hurley, R. L., Sutera, V. A., Aubuchon, R. H. & Lebedeva, M. A. Crossing over between regions of limited homology in *Escherichia coli*. RecA-dependent and RecA-independent pathways. *Genetics* **160**, 851–859 (2002).
45. Jinks-Robertson, S. & Bhagwat, A. S. Transcription-associated mutagenesis. *Annu. Rev. Genet.* **48**, 341–359 (2014).
46. Bzymek, M. & Lovett, S. T. Instability of repetitive DNA sequences: The role of replication in multiple mechanisms. *Proc. Natl Acad. Sci. USA* **98**, 8319–8325 (2001).
47. Mahillon, J. & Chandler, M. Insertion sequences. *Microbiol. Mol. Biol. Rev.* **62**, 725–774 (1998).
48. Richardson, S. M. et al. Design of a synthetic yeast genome. *Science* **355**, 1040–1044 (2017).
49. Xu, X. et al. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat. Biotechnol.* **29**, 735–41 (2011).
50. Kim, J. Y., Kim, Y. & Lee, G. M. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Appl. Microbiol. Biotechnol.* **93**, 917–930 (2012).
51. Kroll, J., Klinter, S., Schneider, C., Voß, I. & Steinbüchel, A. Plasmid addiction systems: Perspectives and applications in biotechnology. *Microb. Biotechnol.* **3**, 634–657 (2010).
52. Scholes, D. T., Banerjee, M., Bowen, B. & Curcio, M. J. Multiple regulators of Ty1 transposition in *Saccharomyces cerevisiae* have conserved roles in genome maintenance. *Genetics* **159**, 1449–1465 (2001).
53. Fahnert, B., Lilie, H. & Neubauer, P. Inclusion bodies: formation and utilisation. in *Physiological Stress Responses in Bioprocesses. Advances in Biochemical Engineering* **89**, 93–142 (Springer, 2004).
54. Hoffmann, F. & Rinas, U. Stress induced by recombinant protein production in *Escherichia coli*. in *Physiological Stress Responses in Bioprocesses* **89**, 73–92 (Springer, 2004).
55. Tippin, B., Pham, P. & Goodman, M. F. Error-prone replication for better or worse. *Trends Microbiol.* **12**, 288–295 (2004).
56. Foster, P.L. Stress-induced mutagenesis in bacteria. in *Critical Reviews in Biochemistry and Molecular Biology* **42**, 373–397 (Springer, 2007).
57. Bailey, L. A., Hatton, D., Field, R. & Dickson, A. J. Determination of Chinese hamster ovary cell line stability and recombinant antibody expression during long-term culture. *Biotechnol. Bioeng.* **109**, 2093–2103 (2012).
58. International Conference on Harmonisation & FDA. Guidance on quality of biotechnological/biological products: Derivation and characterization of cell substrates used for production of biotechnological/biological products. *US Fed. Regist.* **63**, 50244–50249 (1998).
59. Mercier, S. M., Diepenbroek, B., Wijffels, R. H. & Streefland, M. Multivariate PAT solutions for biopharmaceutical cultivation : current progress and limitations. *Trends Biotechnol.* **32**, 329–336 (2014).
60. Yusufi, F. N. K. et al. Mammalian systems biotechnology reveals global cellular adaptations in a recombinant CHO cell line. *Cell Syst.* **4**, 530–542.e6 (2017).
61. Nakamura, K. et al. Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res.* **39**, e90 (2011).
62. Salk, J. J., Schmitt, M. W. & Loeb, L. A. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nat. Rev. Genet.* **19**, 269–285 (2018).
63. Deatherage, D. E., Traverse, C. C., Wolf, L. N. & Barrick, J. E. Detecting rare structural variation in evolving microbial populations from new sequence junctions using breseq. *Front. Genet.* **5**, 1–16 (2015).
64. Rehder, D. S. et al. Expression vector-derived heterogeneity in a therapeutic IgG4 monoclonal antibody. *MAbs* **11**, 145–152 (2019).
65. Harris, R. P. & Kilby, P. M. Amino acid misincorporation in recombinant biopharmaceutical products. *Curr. Opin. Biotechnol.* **30**, 45–50 (2014).
66. Dorai, H. et al. Early prediction of instability of chinese hamster ovary cell lines expressing recombinant antibodies and antibody-fusion proteins. *Biotechnol. Bioeng.* **109**, 1016–1030 (2012).
67. Csorgo, B., Feher, T., Timar, E., Blattner, F. R. & Posfai, G. Low-mutation-rate, reduced-genome *Escherichia coli*: An improved host for faithful maintenance of engineered genetic constructs. *Microb. Cell Fact.* **11**, 11 (2012).
68. Choi, J. W., Yim, S. S., Kim, M. J. & Jeong, K. J. Enhanced production of recombinant proteins with *Corynebacterium glutamicum* by deletion of insertion sequences (IS elements). *Microb. Cell Fact.* **14**, 207 (2015).
69. Nyerges, Á. et al. CRISPR-interference based modulation of mobile genetic elements in bacteria. *Synth. Biol.* **4**, ysz008 (2019).
70. 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).

71. Mikkelsen, M. D. et al. Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metab. Eng.* **14**, 104–111 (2012).
72. Peng, B., Wood, R. J., Nielsen, L. K. & Vickers, C. E. An expanded heterologous GAL promoter collection for diauxie-inducible expression in *Saccharomyces cerevisiae*. *ACS Synth. Biol.* **7**, 748–751 (2018).
73. Rugbjerg, P., Knuf, C., Förster, J. & Sommer, M. O. A. Recombination-stable multimeric green fluorescent protein for characterization of weak promoter outputs in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **15**, fov085 (2015).
74. Kaas, C. S., Kristensen, C., Betenbaugh, M. J. & Andersen, M. R. Sequencing the CHO DXB11 genome reveals regional variations in genomic stability and haploidy. *BMC Genomics* **16**, 160 (2015).
75. Zhang, H. & Wang, X. Modular co-culture engineering, a new approach for metabolic engineering. *Metab. Eng.* **37**, 114–121 (2016).
76. Frumkin, I. et al. Gene architectures that minimize cost of gene expression. *Mol. Cell* **65**, 142–153 (2017).
77. Sørensen, H. P. & Mortensen, K. K. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell Fact.* **4**, 1 (2005).
78. Xia, X.-X. et al. Native-sized recombinant spider silk protein produced in metabolically engineered *Escherichia coli* results in a strong fiber. *Proc. Natl Acad. Sci. USA* **107**, 14059–14063 (2010).
79. Yu, T. et al. Reprogramming yeast metabolism from alcoholic fermentation to lipogenesis. *Cell* **174**, 1549–1558.e14 (2018).
80. Erb, T. J., Jones, P. R. & Bar-Even, A. Synthetic metabolism: metabolic engineering meets enzyme design. *Curr. Opin. Chem. Biol.* **37**, 56–62 (2017).
81. Genee, H. J. et al. Functional mining of transporters using synthetic selections. *Nat. Chem. Biol.* **12**, 1015–1022 (2016).
82. Hansen, E. H. et al. De novo biosynthesis of vanillin in fission yeast (*Schizosaccharomyces pombe*) and baker's yeast (*Saccharomyces cerevisiae*). *Appl. Environ. Microbiol.* **75**, 2765–2774 (2009).
83. Venayak, N., Anesiadis, N., Cluett, W. R. & Mahadevan, R. Engineering metabolism through dynamic control. *Curr. Opin. Biotechnol.* **34**, 142–152 (2015).
84. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
85. Figge, R. & Vasseur, P. Use of inducible promoters in the production of methionine. EP2513322A1 (2018).
86. Xu, P. Production of chemicals using dynamic control of metabolic fluxes. *Curr. Opin. Biotechnol.* **53**, 12–19 (2018).
87. Dahl, R. H. et al. Engineering dynamic pathway regulation using stress-response promoters. *Nat. Biotechnol.* **31**, 1039–46 (2013).
88. Taylor, N. D. et al. Engineering an allosteric transcription factor to respond to new ligands. *Nat. Methods* **13**, 177–183 (2016).
89. Paddon, C. J. & Keasling, J. D. Semi-synthetic artemisinin: A model for the use of synthetic biology in pharmaceutical development. *Nat. Rev. Microbiol.* **12**, 355–367 (2014).
90. Pitera, D.J., Newman, J.D., Kizer, J.L., Keasling, J.D. & Pfleger, B.F. Methods for increasing isoprenoid and isoprenoid precursor production by modulating fatty acid levels. US 8114645 B2 (2012).
91. Sandoval, C. M. et al. Use of pantothenate as a metabolic switch increases the genetic stability of farnesene producing *Saccharomyces cerevisiae*. *Metab. Eng.* **25**, 215–226 (2014).
92. Halls, C. & Yu, O. Potential for metabolic engineering of resveratrol biosynthesis. *Trends Biotechnol.* **26**, 77–81 (2008).
93. Reyes, L. H., Almario, M. P., Winkler, J., Orozco, M. M. & Kao, K. C. Visualizing evolution in real time to determine the molecular mechanisms of n-butanol tolerance in Escherichia coli. *Metab. Eng.* **14**, 579–590 (2012).
94. Ro, D. K. et al. Induction of multiple pleiotropic drug resistance genes in yeast engineered to produce an increased level of anti-malarial drug precursor, artemisinic acid. *BMC Biotechnol.* **8**, 1–14 (2008).
95. Mundhada, H. et al. Increased production of L-serine in *Escherichia coli* through adaptive laboratory evolution. *Metab. Eng.* **39**, 141–150 (2017).
96. Koffas, M. A. G., Jung, G. Y. & Stephanopoulos, G. Engineering metabolism and product formation in *Corynebacterium glutamicum* by coordinated gene overexpression. *Metab. Eng.* **5**, 32–41 (2003).
97. Raab, A. M., Gebhardt, G., Bolotina, N., Weuster-Botz, D. & Lang, C. Metabolic engineering of *Saccharomyces cerevisiae* for the biotechnological production of succinic acid. *Metab. Eng.* **12**, 518–525 (2010).
98. Durfee, T. et al. The Complete genome sequence of *Escherichia coli* DH10B: Insights into the biology of a laboratory workhorse. *J. Bacteriol.* **190**, 2597–2606 (2008).
99. Serero, A., Jubin, C., Loeillet, S., Legoix-Né, P. & Nicolas, A. G. Mutational landscape of yeast mutator strains. *Proc. Natl Acad. Sci. USA* **111**, 1897–1902 (2014).
100. Summers, D. K. The kinetics of plasmid loss. *Trends Biotechnol.* **9**, 273–8 (1991).

Acknowledgements

We thank A. Porse, F. Lino and C. Hjort for helpful comments. The research leading to these results has received funding from the Novo Nordisk Foundation, Denmark, grant number NNF10CC1016517, and from the European Union Seventh Framework Programme (FP7-KBBE-2013-7-single-stage) under grant agreement 613745, Promys.

Author contributions

P.R. and M.O.A.S. outlined and wrote the manuscript.

Competing interests

P.R. and M.O.A.S. are inventors of a pending patent application (WO2017055360) within product addiction filed by the Technical University of Denmark.

Additional information

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence should be addressed to M.O.A.S.

Peer review information: Andy Marshall was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

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