

Review

Safety by design: Biosafety and biosecurity in the age of synthetic genomics

Stefan A. Hoffmann,¹ James Diggans,² Douglas Densmore,³ Junbiao Dai,⁴ Tom Knight,⁵ Emily Leproust,² Jef D. Boeke,^{6,7} Nicole Wheeler,^{8,*} and Yizhi Cai^{1,*}

SUMMARY

Technologies to profoundly engineer biology are becoming increasingly affordable, powerful, and accessible to a widening group of actors. While offering tremendous potential to fuel biological research and the bioeconomy, this development also increases the risk of inadvertent or deliberate creation and dissemination of pathogens. Effective regulatory and technological frameworks need to be developed and deployed to manage these emerging biosafety and biosecurity risks. Here, we review digital and biological approaches of a range of technology readiness levels suited to address these challenges. Digital sequence screening technologies already are used to control access to synthetic DNA of concern. We examine the current state of the art of sequence screening, challenges and future directions, and environmental surveillance for the presence of engineered organisms. As biosafety layer on the organism level, we discuss genetic biocontainment systems that can be used to create host organisms with an intrinsic barrier against unchecked environmental proliferation.

INTRODUCTION

The past decades have seen a tremendous decrease in the cost of DNA synthesis whereas methods for nucleic acid manipulation, assembly and delivery have become more sophisticated and versatile. This has enabled the emergence of the field of synthetic genomics, achieving the total synthesis of viral^{1–3} and bacterial genomes.^{4,5} Soon, with the Synthetic Yeast Sc2.0 project⁶ nearing completion, the first synthetic eukaryotic genome will be available. The possibility to create designer genomes provides fundamentally new opportunities to probe and alter biological functions. This will have tremendous economic and societal impacts in biomedicine, biomanufacturing, bioenergy, agriculture, bioconservation and other areas. Around the globe, biofoundries have been established,⁷ enabling unprecedented throughput of assembly and testing of genetic constructs. At the same time DNA 'printers', desktop devices producing DNA by enzymatic instead of chemical synthesis, are coming to the market, decentralizing access to synthetic DNA. These devices currently focus on synthesis of shorter sequences (30–80 nucleotides) but the technology has the potential to synthesize sequences of more than 300 nucleotides. Furthermore, 'do-it-yourself biology' (DIY BIO) communities have enjoyed increasing popularity, democratizing genetic engineering technology and infrastructure, with limited institutional oversight.

Although these technologies and developments offer vast potential to accelerate biological research and to improve human life through a variety of applications, they also come with inherent risks of accidental or intentional creation and dissemination of potentially harmful biological entities.⁸ Biosafety risks may be posed by unintentional spread of biological agents with the potential to negatively affect human-, agricultural- and ecological health. Special consideration must be given to use cases of engineered organisms in non-contained environments, for example, in medical, agricultural or bioremediation applications. In these cases, it is necessary to ensure that genetically modified microorganisms or engineered genes do not proliferate unchecked. In addition to biosafety concerns, synthetic genomics technologies can pose biosecurity risks because of their potential for dual use. For instance, it is possible to obtain virions solely from sequence information, without requiring actual viral isolates, as demonstrated early on during the ongoing pandemic for SARS-CoV-2,⁹ an RNA virus with a relatively large viral genome. Clearly, this capability can be used to accelerate biological research and to quickly combat global health challenges posed

¹Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester M1 7DN, UK

²Twist Bioscience, 681 Gateway Boulevard, South San Francisco, CA 9408, USA

³Department of Electrical and Computer Engineering, Boston University, 610 Commonwealth Avenue, Boston, MA 02215, USA

⁴CAS Key Laboratory of Quantitative Engineering Biology, Guangdong Provincial Key Laboratory of Synthetic Genomics and Shenzhen Key Laboratory of Synthetic Genomics, Shenzhen Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

⁵Ginkgo Bioworks, 27 Drydock Avenue, Boston, MA 02210, USA

⁶Institute for Systems Genetics, and Department of Biochemistry & Molecular Pharmacology, NYU Langone Health, 435 East 30th Street, New York, NY 10016, USA

⁷Department of Biomedical Engineering, NYU Tandon School of Engineering, Brooklyn, NY 11201, USA

⁸Institute of Microbiology and Infection, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

*Correspondence:
n.wheeler@bham.ac.uk
(N.W.),
yizhi.cai@manchester.ac.uk
(Y.C.)

<https://doi.org/10.1016/j.isci.2023.106165>



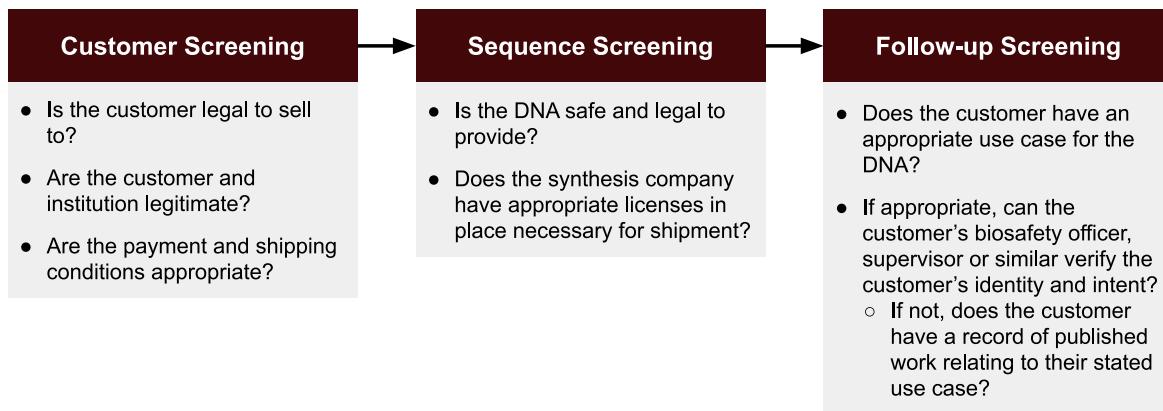


Figure 1. Overview of DNA synthesis screening components

by novel pathogens. However, the same technologies and their relative ease of access can enable malicious state or non-state actors to recreate and engineer dangerous pathogens, with potentially catastrophic consequences. Thus, it is imperative and strategic that safeguarding frameworks and technologies are developed and implemented to safely and securely deliver the revolutionary societal and economic benefits promised by synthetic genomics.

In this article, we review measures and technologies to comprehensively address biosafety and biosecurity risks arising from synthetic genomics capabilities, with a focus on genetically engineered microorganisms (GEMs). We discuss risk mitigation technologies on the digital and the biological layer, DNA sequence screening and genetic biocontainment. Sequence screening aims to control access to genetic material of concern and is already routinely performed by many DNA synthesis providers for orders of gene-length sequences. Genetic biocontainment systems are intended to curb proliferation of GEMs or their DNA outside of specified conditions.

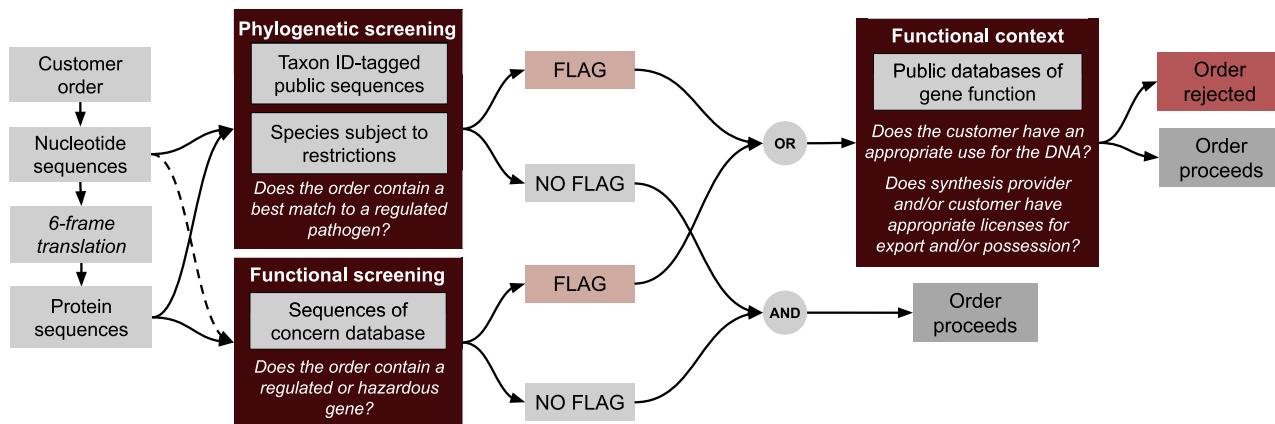
DNA SEQUENCE SCREENING

A clear and immediately feasible intervention point for controlling the emergence of harmful synthetic genes or organisms is the production of DNA used in engineering biology. The 2007 report "Synthetic Genomics: Options for Governance"¹⁰ identified a suite of possible actions by companies, customers, and regulators for preventing biosecurity incidents and concluded that gene and oligonucleotide synthesis companies screening DNA orders was the most effective and tractable action for reducing this risk.

Today, a small number of DNA synthesis firms specialize in synthesizing gene- and pathway-length pieces of double-stranded DNA, which are sometimes incorporated into living cells before shipment to a customer. This industry has been supplying gene-length synthetic DNA to customers for more than 15 years.¹¹ The need to screen orders placed with these companies was recognized early in the genesis of synthetic biology,¹¹ sparking the formation of the International Gene Synthesis Consortium (IGSC), a consortium of gene synthesis companies who have committed to screening DNA synthesis orders and sharing best practices.

Regulation of access to synthetic DNA

No governments currently require screening or mandate how it is performed, so any companies performing screening of DNA synthesis orders today do so voluntarily. In 2010, the US Department of Health and Human Services (HHS) issued the "Screening Framework Guidance for Providers of Synthetic Double-stranded DNA" (<https://www.phe.gov/Preparedness/legal/guidance/syndna/Documents/syndna-guidance.pdf>)" to provide a high-level framework to DNA providers on how to screen orders. The Guidance states that dsDNA providers should implement customer and sequence screening and perform follow-up screening if the customer or sequence screening raises a concern (Figure 1). In issuing its guidance, HHS defined several key characteristics of each step but allowed flexibility in how organizations implemented this screening. In 2022, HHS released a proposed revision to the guidance (<https://www.federalregister.gov/d/2022-09210>) that expands the reach of

**Figure 2. Summary of technical approaches to DNA screening**

Orders undergo 6-frame translation to retrieve protein sequences. Protein and nucleotide sequences are compared to publicly available protein and nucleotide sequences labeled with the taxonomic ID the sequence originates from, and best matches within the order are identified. Any best matches that correspond to regulated pathogens are flagged if they meet a minimum length requirement for screening (200 bases under 2010 US Government guidance but this threshold can differ from company to company). Sequences are also screened against a database of individual sequences of concern (e.g., controlled toxins). Hits to these sequences are also flagged. Flagged sequences undergo additional follow-up screening, where information available on the flagged sequence's function is compared with the customer's stated use case, publication history and institutional affiliation.

the 2010 guidance to oligonucleotides and a wider range of potentially hazardous sequences. The International Gene Synthesis Consortium (IGSC), a group of gene synthesis providers, published the IGSC "Harmonized Screening Protocol (International Gene Synthesis Consortium. Harmonized Screening Protocol v2.0. 2017. Available: <https://genesynthesisconsortium.org/wp-content/uploads/IGSCHarmonizedProtocol11-21-17.pdf>" in 2010. This protocol aligned with the 2010 guidance but is more detailed.

Other countries do not currently have their own guidelines on how to conduct DNA screening, with international companies often referring to the IGSC guidelines to inform screening practices. However, a number of countries are reviewing their biosecurity strategies in the light of the COVID pandemic, for example the United Kingdom, (<https://www.gov.uk/government/consultations/biological-security-strategy-call-for-evidence/public-feedback/biological-security-strategy-summary-of-public-response>) which has named democratized access to DNA synthesis as key risk area for consideration identified through public consultation. The World Health Organization's 2022 Global guidance framework for the responsible use of the life sciences (<https://www.who.int/publications/item/9789240056107>) also highlights DNA synthesis as one of its seven illustrative scenarios to illustrate challenges and gaps in the governance of biorisks. The guidance framework recommends DNA synthesis companies screening according to the IGSC's protocol as a risk mitigation measure to prevent the misuse of synthetic DNA.

An overview of a typical screening workflow is given in Figure 2. The details of screening implementations differ by provider and customer location given the lack of specifics in terms of regulatory requirements. Any orders that are flagged during sequence screening must undergo additional screening to assess the customer's intended use, the ability to perform the described work and the biological function of the ordered DNA. This follow-up investigation is the most time-consuming step in DNA synthesis screening. It typically requires a staff member with a PhD in bioinformatics or similar training to perform the work. Follow-up screening may also involve contacting the relevant regulatory authorities, such as the Bureau of Industry and Security in the US Commerce Department, to confirm whether a specific sequence is subject to an export control license requirement or not. As the cost of DNA synthesis drops and demand increases, this manual evaluation of orders is becoming less practical, creating a need for automatable approaches.

Technical approaches to screening

IGSC members screen orders against a Regulated Pathogen Database, assembled and curated by the IGSC to include data from all organisms on the US Select Agent and Toxin list (<https://www.selectagents.gov/sat/list.htm>), the Australia Group Common Control List (<https://www.dfat.gov.au/publications/minisite/theaustraliagroupnet/site/en/controllists.html>), and other national lists of regulated

pathogens and toxins. These lists cover internationally recognized threat organisms and viruses (such as *Variola major*, the virus that causes smallpox, or *Bacillus anthracis*, the causative agent of anthrax) and toxins like tetrodotoxin. Members screen a six-frame translation of each order to evaluate encoded biological functions rather than just DNA sequence identity.

Although the core goal of sequence screening is to determine whether an ordered sequence could be used to do harm,¹² current regulatory controls focus on orders that contain sequence that is unique to regulated viruses or toxins or to genes that 'endow or enhance' pathogenicity from regulated bacteria, fungi and parasites. These restrictions often include exemptions for specified vaccine strains or inactivated/inviolate forms of these pathogens (<https://www.ecfr.gov/current/title-42/chapter-I/subchapter-F/part-73>).

A major challenge posed by current screening tools is a high false-positive rate - benign sequences are flagged for manual review too frequently.¹² A common source of false positives are so-called 'housekeeping genes' from regulated pathogens (genes not directly involved in pathogenesis, although some believe these should still be considered sequences of concern¹³). In addition to clear false positives, there are also ambiguities about which functions should be viewed as hazardous and thus have their access restricted. The genomic era has revealed that many virulence factors can be found in non-pathogenic bacteria, particularly those involved in interaction with a host.¹⁴ For example, in *Escherichia coli*, pathogenicity islands coding for fimbrial adhesins can be detected in commensal strains of the normal intestinal microflora. The "Best Match" criteria recommended in the 2010 HHS Guidance is intended to address this problem at least partially by ensuring that only sequences unique to regulated pathogens fall under regulatory controls.

In addition to those sequences subject to formal regulatory control, there is a broader set of sequences often referred to as 'sequences of concern' – sequences that might be subject to some form of misuse but that do not pose sufficient danger to merit regulatory control. Such sequences can sometimes cause similar although less-severe illnesses.¹⁵ DNA synthesis companies screen to ensure customers will use DNA products responsibly and detecting such sequences of concern, while not legally mandated by any government, falls squarely within this goal. There are also limitations to these systems – DNA synthesis companies do not generally have access to the broader context for use of a sequence or how multiple sequences may act in concert in a specific chassis organism.¹⁶

Current HHS guidelines specify that regions of 200 bases or more that are a best match to a regulated gene or pathogen should be flagged for further attention, so many methods restrict screening and analysis of hits to regions of this length or longer. Local sequence alignment algorithms (e.g., BLAST) are still the most popular for screening orders, partly because flagged hits are easily interpretable. However, these approaches are slower than newer, k-mer-based methods that count exact matches between short fragments of defined length k (called k-mers) without performing an alignment step, to gain an approximation of sequence identity within a much shorter time frame. Matches between all k-mers can be calculated, or a 'sketch' can be computed for each sequence, which retains a defined number of the lexicographically smallest kmers from the full set of observed k-mers in a sequence (<https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0997-x>). K-mer similarities between sequences, however, degrade with increasing sequence divergence much faster than alignment-based sequence similarity (e.g., for $k = 31$, a popular choice, sequences below 80% nucleotide identity cannot be detected),¹⁷ creating potential sensitivity problems.

The path forward for DNA synthesis screening

As DNA synthesis advances, the cost of producing DNA has gone down (<http://www.synthesis.cc/synthesis/category/Carlson+Curves>) but the cost of performing screening has not. Without a legal requirement for screening (which no country has yet imposed), firms place themselves at an economic disadvantage if they choose to invest in- and perform routine screening. This disincentive must be addressed if screening is to remain an economically viable practice for competitive businesses.¹⁸ Options include improving screening to reduce false positives, outsourcing follow-up investigation to specialized vendors, financially compensating companies for screening,¹⁹ or making screening a legal requirement.²⁰ This last option, however, would require broad international adoption to ensure a level economic playing field for the deeply international DNA synthesis market.

In terms of knowing whether DNA synthesis screening is accurately detecting sequences of concern, there are no broadly available standardized performance benchmarks that screening software must meet to

Existing screening solutions

The Black List Sequence Screening Pipeline (BLiSS) (<https://www.aiche.org/sbe/conferences/synthetic-biology-engineering-evolution-design-seed/2016/proceeding/paper/bliss-black-list-sequence-screening-pipeline>) was developed by the U.S. Department of Energy Joint Genome Institute with guidance from the US Department of Health and Human Services. BLiSS detects “sequences of concern” of at least 200 nucleotides in length. A “Best Match” approach is used to determine whether any of the sequences of concern are unique to Select Agents or Toxins, or CCL-listed agents, toxins or genetic elements. All hits are broken up using a 200 bp sliding window to determine the Best Match for each fragment. The pipeline was adapted using over 10,000 sequences submitted to the JGI for synthesis to reduce false positives.

SeqScreen¹⁸ was designed to sensitively assign taxonomic classifications, functional annotations, and biological processes of interest to short nucleotide sequences of unknown origin (40 bp-1,000 bp). An open-source solution, it is not explicitly designed for identifying biosecurity threats, but can be used for this purpose. Highly trained individuals read a vast amount of literature to provide accurate annotations for each sequence in their database of concern. The teams are exploring machine learning approaches as a way of automating a portion of this work by predicting a suite of biological processes of interest on large protein databases before incorporation into the SeqScreen framework.

ThreatSeq (<https://www.battelle.org/markets/health/chemical-and-biological-threats/biosecurity-pandemic-preparedness/threatseq>) is a commercial solution from Battelle for screening synthetic DNA orders that identifies similarity to specific sequences of concern and Select Agent genomes. They use a tiered reporting style to reflect different levels of risk posed by different orders and the scientific uncertainty around the relevance of different sequences to pathogenicity. ThreatSeq’s database currently contains more than 10,000 Sequences of Concern, which are constantly under review and encompass all regulated organisms, viruses and toxins. The ThreatSeq metadata is the product of several years of government-funded curation efforts, an indicator of the high cost required to produce a set of annotated sequence records.

SecureDNA (<https://www.securedna.org/>) differs from previous approaches by using a one-way hash algorithm to create digests of sequence regions to obscure customer orders from the screening provider and Sequences of Concern from both customers and DNA providers. The database contains short strings of numbers and letters (a digest) produced by a hash function from k-mers of both protein and nucleotide sequences. These k-mers are drawn from conserved regions of Sequences of Concern and mutated variants predicted to preserve the biological function of the original sequence. This digest-based approach allows the inclusion of sequences of concern without revealing their identity.

The Nuclear Threat Initiative is developing an International Common Mechanism (<https://www.nti.org/analysis/articles/common-mechanism-prevent-illicit-gene-synthesis/>) for DNA synthesis screening in consultation with an international technical consortium of industry, policy and academic stakeholders. The Common Mechanism is intended to act as a common standard for screening worldwide and a baseline capability that DNA providers can adopt. This platform identifies genes of concern and best matches to regulated pathogens. It has been designed to accommodate different lists of controlled organisms and provides customizable reporting according to individual regions’ legal requirements.

demonstrate that the software is performing well. Having regular benchmarks of available tools and well-designed test sets would help evaluate whether new offerings are fit for purpose and to monitor whether existing tools remain robust to growing reference sequence databases and changing regulatory controls.¹² Test datasets would need to be updated regularly to prevent over-training of screening algorithms on a specific test set.

The bulk (by number of unique ordered sequences) of the synthetic DNA (and RNA) market is for shorter-length oligonucleotide constructs (less than 200 bp in length). Previously, industry-wide screening of oligonucleotide orders was considered intractable because of the far greater volume of companies and orders in this market, the faster turnaround time expected by customers, and the higher false-positive rate of screening shorter sequences.¹⁹ However, the newly proposed HHS guidance has expanded recommendations for screening to low-volume orders of sequences 50 bases or longer and high-mass orders (>1 mM) of 20 bases or longer. The screening of these shorter sequence lengths poses a new technical challenge, as the probability of spurious similarity between orders and Sequences of Concern increases as sequence length decreases.

In addition to commercial providers of synthesized DNA, some firms sell benchtop DNA synthesizers for individual laboratory use. The development of these technologies is now reaching some crucial milestones,

which increases the urgency of universalizing screening. First, the accuracy of DNA synthesis by benchtop devices is anticipated to improve substantially (<https://www.nature.com/articles/s41587-020-0695-9>); second, companies are now producing DNA of length ~300 bases.²¹ As a result, there is growing interest in requirements to screen customers buying these devices, screen the sequences produced by these devices for potential hazards, and log the synthesized sequences (<https://www.jcvi.org/sites/default/files/assets/projects/dna-synthesis-and-biosecurity/report-complete.pdf>).

Current sequence screening is predominantly based on DNA derived from control lists of organisms of concern rather than genes or biological functions of interest/concern.¹¹ There is interest in moving away from this agent-based screening toward function-oriented screening for Sequences of Concern that could pose a serious risk if misused.¹⁶ Virulence factor databases could provide important candidates for Sequences of Concern.²² However, the use of these databases has limitations. First, these databases often focus on common human pathogens, which are not the only high-priority pathogens from a biosecurity perspective. Second, most virulence factors can also be found in non-pathogens,¹⁴ raising uncertainty about whether they should be flagged as hazards and if it is practical to conduct follow-up screening on these orders. Third, many virulence factors presumably remain undiscovered. A recent effort to define Functions of Sequences of Concern (FunSOCs)²³ may address some of the uncertainty around which virulence factors constitute intrinsic hazards and which do not. However the best way to identify and catalog these is still a matter of debate, as this potentially poses an information hazard.²⁴

Another potential future direction of DNA screening is the adoption of screening earlier in the process of ordering DNA, such as in synthetic biology design software, or as part of institutional requisition systems, allowing institutional biosafety officers to submit pre-approval of sequences with DNA synthesis orders to expedite screening. Technical challenges arise in ensuring that screening is used in practice and updated regularly, but implementing screening at these earlier stages could result in a lower burden of follow-up screening for DNA providers and ensure that customers receive their orders without unnecessary delay.

Environmental surveillance

Sequence screening plays an important role in restricting access to DNA of concern, but it also offers the potential to detect biological threats through metagenomic analysis of environmental samples. Wastewater metagenomics has been successfully employed in the surveillance of a number of infectious diseases and public health threats,²⁵ including COVID-19.²⁶ Samples are relatively accessible compared to clinical samples, and are culture-independent, meaning pathogens with strict growth requirements (e.g., viruses that only replicate in human cells) can still be detected. Metagenomics can be expensive and technically challenging, but there is an opportunity to leverage this surveillance to serve a host of purposes, including monitoring for the release of engineered organisms.

The depth of sequencing required to correctly distinguish engineered and non-engineered organisms is not currently known, as genetic engineering detection and attribution is typically performed on known sequences or amplified pre-selected markers of genetic engineering. Given sufficient data coverage, sequence data could be fed into algorithms trained to recognize engineered DNA and even to attribute these engineered organisms to a lab of origin.²⁷ It is unknown how these methods would perform if a laboratory changed the organism they were working on to engineer a hazard. It is also unknown how these methods would perform on incomplete sequences or even mixed metagenomic samples.

For environmental metagenomics to be most effective, long-term funding and implementation will be needed to establish baseline rates and diversity of genetic markers of engineering in the environment, track trends and identify disruptions in these patterns. This is particularly important for machine learning-based methods that are typically only trained on engineered sequences and do not disclose what they are using to flag engineering. Established surveillance for GEMs in the environment would be able to act as a warning system for potential biosafety and biosecurity threats emanating from non-sanctioned releases of engineered organisms.

GENETIC SAFEGUARDS

Clearly, it is vital to curb known and unknown risks to public health, agriculture and ecology, arising from the inadvertent or deliberate release of GEMs into natural environments. The measures outlined above aim to control access to DNA sequences of concern and detect engineered organisms in environmental samples. Effective and reliable ways to actually prevent unwanted proliferation of GEMs or their recombinant

Table 1. Major groups of genetic biocontainment systems

Containment system	Advantages	Disadvantages
Gene knockouts creating auxotrophies	<ul style="list-style-type: none"> Easy to implement Resistant against mutational escape 	<ul style="list-style-type: none"> Required natural metabolite can be available in natural environments e.g., by cross-feeding
De-repression of suicide genes	<ul style="list-style-type: none"> Large available repertoire of suitable effectors Coupling to environmental sensing possible 	<ul style="list-style-type: none"> Prone to mutational inactivation of effector gene, exacerbated by incomplete effector repression
Control of essential genes	<ul style="list-style-type: none"> Control approaches along the flow of genetic information can be combined Coupling to environmental sensing possible Largely resistant against mutational escape 	<ul style="list-style-type: none"> Typically, small number of essential genes suitable, large-scale screenings necessary for their identification
Orthogonal central dogma	<ul style="list-style-type: none"> Both semantic and trophic barrier to natural organisms possible Highly resistant against mutational escape 	<ul style="list-style-type: none"> Difficult to develop and implement

DNA in natural environments are another important pillar of biosafety. Solving this question is imperative to gain regulatory and public acceptance of open-environment applications of engineered microorganisms, e.g., for bioremediation, agricultural or medical purposes. Although methods of physical containment for such scenarios, for instance through microencapsulation,²⁸ have been proposed, we will focus on in-built genetic features that prevent survival of engineered organisms outside of defined environments and/or sexual or horizontal transfer of genetic material to natural organisms.

Genetic safeguard systems acting on the organism level are typically designed to limit survival to permissive environments characterized by the availability or absence of certain compound(s). Under these permissive conditions a wild-type like fitness is highly desirable to not create evolutionary pressure favoring loss of the genetic biocontainment system. Owing to the potential for quick replication of microorganisms, even slight fitness advantages gained by escape mutations will eventually take over in populations of GEMs under permissive conditions, voiding the genetic control system. If the survival conditions are not met, constituting a restrictive environment, the deployed system is supposed to robustly prevent proliferation and optimally induce death of the engineered organisms.

According to the NIH Guidelines for Research Involving Recombinant or Synthetic Nucleic Acid Molecules (https://osp.od.nih.gov/wp-content/uploads/NIH_Guidelines.pdf), survival of the organism or escape of the recombinant DNA should be less than 10^{-8} for contained host-vector systems (EK2). Considering that a single test tube of densely grown *E. coli* easily contains 10^9 cells, this degree of stringency is insufficient for open-environment uses of biological entities with self-replicative potential. The combination of several orthogonal genetic containment systems should allow containment stringencies several orders of magnitude higher. Also, effective containment is expected to be maintained far longer, as evolutionary enrichment of escape mutations against one of the systems can be prevented through redundancy of containment with additional orthogonal systems. Here, genetic biocontainment systems of a wide range of technology readiness levels are discussed (Table 1).

Gene knockouts

The necessity of working with safe organisms was already recognized in the early days of genetic engineering. Just a few years after the first animal gene had been cloned in 1973, the *E. coli* strain χ 1776 was introduced.²⁹ It was highly sensitive against bile and dependent on availability of thymine/thymidine and diaminopimelic acid, rendering it incapable of colonizing the guts of mammals even in the absence of competing microbes.³⁰ Despite even being employed in industrial settings, its use was eventually ceased in favor of strains that were easier to work with.

However, the containment approach of using gene knockouts for reduction of competitive fitness or creation of metabolic deficiencies has been widely used since. Many currently used laboratory strains have such genetic features. For instance, *E. coli* K-12 derived laboratory strains typically feature a *relA* mutation, reducing starved survival,³¹ and often also possess one or multiple auxotrophies, the inability to produce essential metabolites. Likewise, the commonly used *Saccharomyces cerevisiae* lab strains BY4743 and W303 and their haploids feature a number of auxotrophies. These metabolic deficiencies are routinely leveraged in the lab by using complementing genes as selective markers for plasmid maintenance or genomic integration. They also decrease competitive fitness under dynamically changing conditions and nutrient limitations³² which are typical for natural environments. Auxotrophic containment has been employed in applied medical research.³³ *Lactococcus lactis* has been engineered for intestinal delivery of human interleukin 10 to combat inflammatory bowel disease. By replacing its thymidylate synthase gene with the human *IL10* gene, insertion of the transgene rendered it auxotroph for thymidine. Knockout-based auxotrophies are likely the most commonly used genetic biosafety feature currently, albeit not a universally secure one. The nucleic acids, vitamins or amino acids that organisms have been made dependent on may be readily available in some natural environments, e.g., by cross-feeding from other organisms.

A combination of knockouts of endogenous phosphate transporters and introduction of two heterologous genes for transport and metabolism of phosphite has been used to make cyanobacteria dependent on phosphite as a phosphorus source.³⁴ This metabolic bypass both restricts the engineered organisms' survival to controlled environments offering phosphite, and offers a means to preclude contamination with natural, phosphate-dependent organisms if phosphite is the only source of phosphorus.

As was already shown with χ 1776 and bile, knockouts can also be used to make microorganisms susceptible to certain environmental conditions instead of creating dependence on exogenous supply of a metabolite. A recent example is the knockout of fluoride exporters in yeast, resulting in sensitivity to fluoride at concentrations acceptable in drinking water.³⁵ Because they rely on gene knockouts, these containment strategies are thought to be evolutionarily robust, requiring a mutational gain of function for escape. However, the specific environmental requirements of restrictive conditions mean these strategies cannot be employed as a generic biocontainment method against survival in natural environments.

Clearly, for safe open-environment uses of engineered microorganisms there is a need for specifically designed containment systems to make survival dependent on the availability of a select compound, e.g., one generally not found in natural environments. This need has been recognized and in the past decades notable research into dedicated genetic systems has been conducted. These active containment systems typically confine survival by using lethal factors and/or by regulating the activity of essential genes.

Suicide genes

The use of lethal genetic factors, or "suicide genes", as effectors in active biocontainment systems has been a prominent strategy, particularly in prokaryotes. This is largely thanks to the number of discovered prokaryotic toxin-antitoxin systems, giving researchers a repertoire of effective toxic genes for construction of suicide systems in bacteria. In proof-of-concept studies, toxin genes were driven by inducible promoters.^{36,37} Conditions leading to de-repression of the promoter constitute restrictive environments, causing toxin expression and cell death. Closer to the potential use case of bioremediation, similar suicide systems were extended with a sensing system for a chemical contaminant to be removed, triggered when the chemical is depleted. Notable examples^{38–40} made use of a promoter inducible by benzoic acid derivatives. Using prokaryotic toxin-antitoxin systems has the advantage that the cognate toxin-antitoxin interaction can be leveraged for toxin control.⁴¹ This has also been used for a plasmid biocontainment system engineering specific host-plasmid dependence.⁴² A list of toxin-antitoxin systems suitable for biocontainment has been compiled and reviewed elsewhere.⁴³

Cas proteins have been used as an efficient killing effector in both prokaryotes and eukaryotes.^{44–46} Used with guide RNAs targeting repetitive regions, a high toxicity results from the numerous induced double-strand breaks. The bipartite nature of the system, requiring both the Cas9 protein and the guide RNA, allows control over expression of both and evidently affords very low basal toxicity.

With suicide-based systems, killing can be impaired by loss of function of the suicide gene or acquired resistance. Basal expression of the suicide gene evolutionarily promotes these escapes. Consequently,

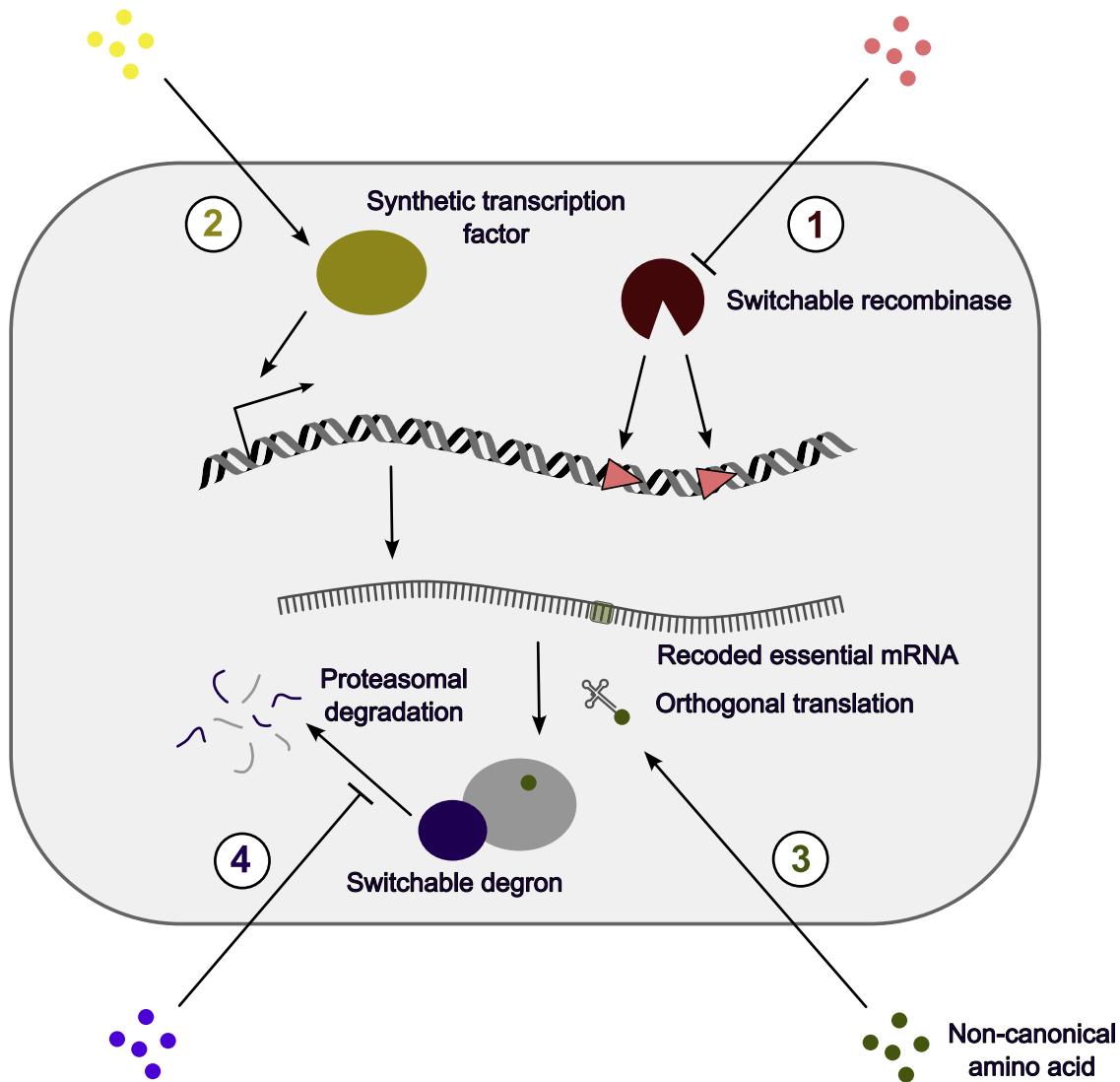


Figure 3. Biocontainment by control over essential genes

Control can be attained by independent systems at different points along the flow of genetic information, allowing the free combination of different systems. (1) Recombinational switches lead to the excision of essential genes in restrictive conditions. (2) Transcriptional switches drive expression of essential proteins only if specific chemical inducers are available. (3) Post-translational switches direct essential proteins for degradation, leading to cell death unless a ligand is supplied. (4) Genetic code expansion can be leveraged to make cell survival dependent on the exogenous supply of a non-canonical amino acid.

stringent control of the suicide gene expression is paramount and, for instance, has been implemented by bifurcated control of suicide gene expression.⁴⁰ However, even sophisticated multi-layered suicide circuits with stringent control of suicide gene expression can suffer from increasing escape frequency by mutational inactivation, as evidenced for the so-called Deadman kill switch within a few days of passaging.⁴⁷

Control of essential genes

Controlling essential genes is a potentially more mutation-resistant containment approach; unlike suicide genes, biocontainment escape is usually not possible by easily obtainable loss-of-function mutations. This control can be achieved in a number of ways—notably by making transcription or translation of essential genes, or the stability or function of essential proteins conditional (Figure 3). Another possibility is the conditional removal of essential genes, for instance by an estradiol-activated Cre recombinase fusion protein.⁴⁸ In this approach, the recombinase effectively acts as a ‘suicide gene’.

Conditional transcription of essential genes has been demonstrated as a viable approach for biocontainment in yeast.^{48,49} In this particular system suitable essential genes are under the control of *GAL* promoters. The chimeric transcriptional activator GEV⁵⁰ drives expression from these promoters only in the presence of estradiol. The cells are thus strictly dependent on the presence of either galactose or estradiol. Controlling the expression of an essential histone gene in this way yielded an escape rate of 10^{-6} . Exerting the control on two histone genes reduced this rate to as low as 10^{-9} , but resulted in a significant fitness defect under permissive conditions.⁴⁸ A systematic search for more suitable control targets from 250 essential genes yielded three essential genes giving escape rates of less than 10^{-7} individually with minimal changes to the transcriptome.⁴⁹ There is a comparable yeast chimeric transcription factor using zinc finger DNA-binding domains, which activates transcription from a synthetic promoter in the presence of estradiol (ZEV).⁵¹ In contrast to GEV, ZEV has no apparent off-target effects and therefore was also investigated by Agmon et al. for the three top hits from their screening. Recently, the ZEV system was applied to 1,022 (of ~1,100) essential yeast genes, and the growth profile in a range of estradiol concentrations was assessed for each.⁵² This screen yielded hundreds of essential genes resulting in estradiol dependent growth under GEV control, providing a resource to identify essential genes suitable for biocontainment leveraging this system. Another attractive feature of this system is that nanomolar concentrations of estradiol required add minimal cost to fermentations, and this low abundance also means they will be difficult to detect by potential bad actors seeking to bypass these containment systems.

Control over translation of essential genes can be obtained in different ways. One developed approach makes use of well-established allosteric transcription factors LacI and TetR for transcriptional control over riboregulators, which in turn affect translation of the target essential gene.⁵³ Another method leverages orthogonal translation systems (OTSs) selectively incorporating non-canonical amino acids at repurposed stop codons. An OTS consists of two components, an aminoacyl-tRNA synthetase and a cognate tRNA, which are orthogonal to the endogenous tRNAs and synthetases. The tRNA, loaded with the non-canonical amino acid by the orthogonal aminoacyl-tRNA, has an anticodon recognizing the repurposed stop codon, leading to ribosomal incorporation of the non-canonical amino acid instead of a translational stop. A stop codon reassigned in this way is introduced into an essential gene, such that translation of the functional full-length protein requires presence of the non-canonical amino acid. In the absence of the non-canonical amino acid a truncated, non-functional gene product is being made, resulting in cell death. This approach, creating "synthetic auxotrophy" for the non-canonical amino acid, has been demonstrated in a highly engineered *E. coli* strain devoid of TAG stop codons and release factor 1 (RF1). The introduction of two to three TAG stops at carefully chosen essential gene positions and optimization of the OTS⁵⁴ or the genomic background⁵⁵ resulted in escape frequencies $<10^{-11}$ in non-permissive medium, constituting an extremely tight control of cell survival. The impending completion of the first synthetic eukaryotic genome in the Sc2.0 project will yield an *S. cerevisiae* strain without TAG stop codons. This provides an opportunity to develop highly stringent OTS-based biocontainment in yeast. Furthermore, the same approach can be used to prevent horizontal gene transfer of genes of interest: by introduction of reassigned stop codons into these genes, they will not yield functional gene products when introduced into natural organisms.

Besides controlling transcription and translation of essential genes, the stability of the essential proteins themselves can be made conditional. This can be achieved by fusing them to regulatable degrons. Conditional survival by fusing suitable degrons to essential proteins in yeast has been demonstrated with the auxin inducible degron (AID),^{56,57} a protein degradation system taken from plants, and with the SMASH tag, a self-cleaving degron whose removal can be inhibited by a small molecule.⁵⁸ In the case of the AID, the tag that is fused to the protein of interest is fairly small, but the system requires heterologous expression of the TIR1 protein, which is not found in non-plant eukaryotic cells. Addition of auxin leads to polyubiquitination of the tag, marking it for proteasomal degradation. With SMASH, the tag consists of the degron itself at the distal end to the fusion, a protease from hepatitis C virus, and a cognate protease cleavage site. Self-cleavage of the tag leads to its removal, leaving behind only part of the cleavage site. Addition of a specific protease inhibitor prevents the self-removal of the tag, leading to protein degradation because of the attached degron. Although these studies show that the survival of microorganisms can be controlled by conditional stability of essential proteins, the operational logic of both systems may not be suitable for typical biocontainment applications with this approach, though: the addition of specific small molecules, rather than their absence, leads to depletion of the tagged essential protein and thus cell death. Another class of switchable degrons called destabilizing domains (DD)⁵⁹ might make it possible to leverage conditional essential protein stability for a practical biocontainment system. These are protein domains engineered to be unstable in the absence of their

ligand and thus degraded with a protein they are fused to. Addition of the ligand leads to domain stabilization, preventing protein depletion. With DDs being comparably large tags, effects on protein expression and function are to be expected, likely making it necessary to carefully choose targeted essential genes to get stringent containment without reducing fitness under permissive conditions.

Finally, the function of essential proteins can be a potential target for exerting control over organism survival. A previously described containment system inspired by a protein engineering approach creating *de novo* allosteric enzymes, making enzyme activity dependent on the availability of a ligand.⁶⁰ Using directed evolution, five essential enzymes in *E. coli* have been engineered to require a ligand for their activity.⁶¹ Combining three engineered, benzothiazole dependent enzymes in a single strain gave an escape frequency of $<3 \times 10^{-11}$ over two days, clearly making this a promising approach.

Orthogonal central dogma

Instead of engineering individual safeguarding systems, globally altering the genetic code or the information carrier altogether can create a powerful barrier to genetic interaction with natural organisms. This protection against horizontal gene transfer owing to genetic code incompatibility is referred to as semantic biocontainment. The alteration can also make the organism dependent on synthetic compounds, creating a synthetic auxotrophy, i.e., trophic biocontainment.

Engineering a genetic code that is incompatible to natural organisms is a conceptually attractive way to curb risks arising from genetic exchange of GMOs with the natural environment. Substantially altering the way mRNA messages are decoded into proteins means that any genetic material exchanged between natural and synthetic organisms will not yield functional proteins. A cell-free translation study swapping serine, leucine and alanine codons found functional protein exclusively expressed if the genetic code of mRNA and tRNAs was matched, demonstrating the potential to create a powerful genetic separation to natural organisms.⁶² However, creating an organism with sense codon reassignment is a daunting task considering the vast biochemical, thermodynamic, and semantic complexity and interconnectedness of translation. For instance, the same *in vitro* study reported lowered protein expression from re-coded mRNA by matched translation systems, probably because of altered mRNA folding.⁶² Accordingly, a similarly profound change of an organism's genetic code would likely require optimization of codon usage in hundreds of essential genes.⁶³ Still, there are abundant examples of non-standard genetic codes in nature.⁶⁴ In fact, eukaryotes across a wide range of phylogenetic groups possess mitochondrial genetic codes deviating from the standard code in different ways. The comparably small number of proteins encoded by mitochondrial genomes likely constitutes a much smaller evolutionary barrier to changes of the genetic code. Non-standard nuclear codes do not appear to be as diverse, and most of these aberrant codes have a reassignment of a stop codon to a sense codon. However, some *Candida* yeast species decode a canonical leucine codon as a serine, a change that evolutionarily likely was preceded by code ambiguity.⁶⁵ This mechanism could possibly be used to evolve sense codon reassessments by a stepwise tRNA dosage change from one isoacceptor class to another. Another path to sense codon reassignment is through codon compression followed by codon reassignment. Using a synthetic genomics bottom-up approach, a viable *E. coli* strain free of two serine codons and the amber stop codon, Syn61, has recently been created.⁵ This enabled the deletion of the respective decoding tRNAs and release factor 1, conferring immunity to bacteriophages.⁶⁶ The freed-up codons were subsequently reassigned with orthogonal translation systems to encode non-canonical amino acids. Likewise, reassignment to endogenous amino acids can be used to create two-way barrier to genetic interaction with natural organisms.⁶⁷ However, there are constraints surrounding tRNA crosstalk and the dependence of many aminoacyl-tRNA synthetases on the anticodon for tRNA recognition. This type of recoding may find itself in tension with the goals of bio-security screening of synthetic DNA. Synthetic DNA designed for use in organisms using alternative genetic codes, when translated using standard genetic codes, would not be identifiable as Sequences of Concern even if they encoded such a sequence in the target non-natural genetic code.

To achieve genetic code incompatibility, instead of reassessing existing codons, entirely new codons can be created. This possibility has been demonstrated by systems specifically decoding quadruplet instead of natural triplet codons.^{68,69} Another avenue for codon creation is the expansion of DNA polymer chemistry using non-natural base pairs. A few groups have managed to create unnatural DNA base pairs in DNA polymers, which are faithfully replicated in PCR reactions and can be transcribed into RNA *in vitro*.^{70,71} The Romesberg group has succeeded in maintaining an unnatural base pair (dNaM-dTPT3) in *E. coli*⁷² and reported translation of mRNA and tRNA containing unnatural base pairs, enabling ribosomal decoding of

codons with unnatural bases.⁷³ Still, achieving DNA replication, transcription and translation of unnatural base pairs as efficiently and faithfully as natural base pairs currently appears to be a distant feat. Once achieved, it will open up a powerful way to create semantic biocontainment and synthetic auxotrophy.

An even more ambitious undertaking is altering the central dogma altogether by using an alternative information-carrying polymer. Several non-natural nucleic acid backbone chemistries have been demonstrated to be suited for information storage, and polymerases have been engineered both to specifically synthesize these XNAs (xeno nucleic acids) from DNA templates and to transcribe XNAs back to DNA.⁷⁴ DNA could be produced in *in vivo* by the cellular machinery of *E. coli* from different non-canonical backbones.^{75,76} Whether XNAs can take on functions of and potentially replace natural nucleic acids in living cells remains yet to be seen. The thought of synthetic organisms with a different central dogma is certainly enticing, not least from a biocontainment point of view.⁷⁷

CONCLUSION

With increasing affordability of DNA synthesis and rapid technological advances in synthetic genomics, an ever-growing number of actors are capable of engineering or building genomes. The nascent decentralized capabilities of creating designer genomes challenge existing biosecurity and biosafety frameworks, which need to be adapted and updated to address these challenges.

Because synthetic genomics technologies can remove the need for biological isolates to create certain pathogens, restricting access to synthetic DNA of concern is a major safeguard against its harmful use. Already widely implemented measures of customer and sequence screening discussed above are largely a result of industry self-regulation. Notably, however, sequence screening has an economic cost, e.g., compiling and maintaining an extensive database of sequences of concern and suitable screening algorithms with the computational costs necessary to perform these screens. Furthermore, sensitive detection comes with trade-offs relating to computational expense and false positive hit rate with associated costs of manual follow-up procedures. These costs provide an economic disincentive for synthesis companies to implement rigorous screening procedures, particularly for smaller companies under legislations with little regulation of the synthetic DNA market. The Nuclear Threat Initiative has proposed to develop a common screening mechanism, providing both a comprehensive database and screening algorithms, to partially externalize the costs of sequence screening. Reducing these economic disincentives should be complemented by coordinated regulation of the global synthetic DNA market, perhaps as part of existing multilateral arrangements like the Australia Group. Regulating synthetic DNA, however, comes with challenges not encountered with 'conventional' biological agents. Sequence screening algorithms can potentially be gamed and do not generally have sufficient awareness of the broader customer goals across sequence providers and orders (for instance, in the way synthetic DNA is (re-)assembled). The emergence of desktop devices enabling decentralized DNA synthesis stresses the question of whether sequence screening should be done locally or centrally, and how to prevent attempts at circumventing oversight. Clearly, developing scalable solutions that incentivize effective control of synthetic sequences of concern while balancing the needs of the various actors and stakeholders is not trivial.

A related issue is that of ensuring biosafety of physically created (partially) synthetic organisms, particularly organisms that contain sequences of concern or are intended for applications in open environments. For both cases, it appears to be prudent to deploy genetic systems for intrinsic containment, confining the organisms' survival to specific, defined environments. This can bridge the conflicting needs for risk reduction on the one hand and for convenience and cost-effectiveness of cultivation on the other. In contrast to restricting access to DNA of potential concern, genetic biocontainment systems largely require compliant actors, as the systems embedded in the host organism's genome can be genetically inactivated. The notable exception is creating an orthogonal central dogma which would allow for semantically confining synthetic genes to synthetic host organisms with a genetic 'firewall' to natural organisms.

For research settings, current NIH guidelines list approved host-vector systems offering biocontainment with escape rates of less than 1 in 10^8 , lowering biosafety level (BL) requirements of work conducted with them. Synthetic genomics with increasingly versatile genome editing tools have brought about a shift from vector-based to genome-based work. To accommodate this, well-characterized genetic biocontainment systems controlling organism survival could similarly be approved for reduction of BL classification. Undoubtedly, widespread implementation of suitable genetic biocontainment systems can make work with biological agents much safer. As described above, there is already a variety of different systems available,

many of which have orthogonal modes of action and can be combined to achieve containment much more stringent than NIH guidelines mandate. Achieving extremely stringent genetic containment will also play a pivotal enabling role for approval of field applications of GEMs, which will require control over cell counts that are several orders of magnitude higher than 10^8 . In addition to requirements regarding stringency and evolutionary stability, industrial-scale use cases will necessitate genetic biocontrol that is also cost-effective and environmentally unproblematic.

The rapidly changing landscape of synthetic genomics, enabling new capabilities and applications of engineering biological systems, requires advanced biosafety and biosecurity frameworks to ensure responsible research and commercial use. Thus, the continued development, review and adaptation of digital and biological safety mechanisms as major components in these frameworks is indispensable.

ACKNOWLEDGMENTS

This work is supported by grants from EPSRCEP/V05967X/1, BBSRCB/W014483/1, Volkswagen Foundation “Life? Initiative” (ref. 94 771) to YC, and from NSF (2027045 and 2211040) to DD and (1026068, 1443299, 1616111 and 1921641) to JDB. JD is supported by Royal Society Newton Advanced Fellowship (NAF\R2\180590), National Natural Science Foundation of China (31725002), Bureau of International Cooperation, Chinese Academy of Sciences (172644KYSB20180022), Guangdong Provincial Key Laboratory of Synthetic Genomics (2019B030301006) and Shenzhen Outstanding Talents Training Fund. The authors thank James J. Collins (MIT) for discussions on genetic biocontainment.

DECLARATION OF INTERESTS

D.D. is a co-founder of Lattice Automation, Inc. and Asimov Inc. Both companies create engineered biological systems using software and automation. J.D.B. is a Founder and Director of CDI Labs, Inc., a Founder of and consultant to Neochromosome, Inc, a Founder, SAB member of and consultant to Re-Open Diagnostics, LLC, and serves or served on the Scientific Advisory Board of the following: Sangamo Therapeutics, Inc., Modern Meadow, Inc., Rome Therapeutics, Inc., Sample6, Inc., Tessera Therapeutics, Inc. and the Wyss Institute. T.K. is co-founder of Ginkgo Bioworks, Inc. EL is CEO and co-founder of Twist Bioscience, Inc., and J.D. is an employee of Twist Bioscience, Inc.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in their field of research or within their geographical location. One or more of the authors of this paper self-identifies as a gender minority in their field of research. One or more of the authors of this paper self-identifies as a member of the LGBTQIA+ community. One or more of the authors of this paper received support from a program designed to increase minority representation in their field of research.

REFERENCES

1. Cello, J., Paul, A.V., and Wimmer, E. (2002). Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science* 297, 1016–1018. <https://doi.org/10.1126/science.1072266>.
2. Becker, M.M., Graham, R.L., Donaldson, E.F., Rockx, B., Sims, A.C., Sheahan, T., Pickles, R.J., Corti, D., Johnston, R.E., Baric, R.S., and Denison, M.R. (2008). Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. *Proc. Natl. Acad. Sci. USA* 105, 19944–19949. <https://doi.org/10.1073/pnas.0808116105>.
3. Noyce, R.S., Lederman, S., and Evans, D.H. (2018). Construction of an infectious horsepox virus vaccine from chemically synthesized DNA fragments. *PLoS One* 13, e0188453. <https://doi.org/10.1371/journal.pone.0188453>.
4. Gibson, D.G., Benders, G.A., Andrews-Pfannkoch, C., Denisova, E.A., Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., Algire, M.A., et al. (2008). Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* 319, 1215–1220. <https://doi.org/10.1126/science.1151721>.
5. Fredens, J., Wang, K., de la Torre, D., Funke, L.F.H., Robertson, W.E., Christova, Y., Chia, T., Schmied, W.H., Dunkelmann, D.L., Beránek, V., et al. (2019). Total synthesis of *Escherichia coli* with a recoded genome. *Nature* 569, 514–518. <https://doi.org/10.1038/s41586-019-1192-5>.
6. Richardson, S.M., Mitchell, L.A., Stracquadanio, G., Yang, K., Dymond, J.S., DiCarlo, J.E., Lee, D., Huang, C.L.V., Chandrasegaran, S., Cai, Y., et al. (2017). Design of a synthetic yeast genome. *Science* 355, 1040–1044. <https://doi.org/10.1126/science.aaf4557>.
7. Hillson, N., Caddick, M., Cai, Y., Carrasco, J.A., Chang, M.W., Curach, N.C., Bell, D.J., Le Feuvre, R., Friedman, D.C., Fu, X., et al. (2019). Building a global alliance of biofoundries. *Nat. Commun.* 10, 2040. <https://doi.org/10.1038/s41467-019-10079-2>.
8. Biodefense in the Age of Synthetic Biology (2018). <https://doi.org/10.17226/24890>.
9. Thi Nhu Thao, T., Labroussaa, F., Ebert, N., V'Kovski, P., Stalder, H., Portmann, J., Kelly, J., Steiner, S., Holwerda, M., Kratzel, A., et al. (2020). Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature* 582, 561–565. <https://doi.org/10.1038/s41586-020-2294-9>.
10. Garfinkel, M.S., Endy, D., Epstein, G.L., and Friedman, R.M. (2007). Synthetic genomics | options for governance. *Biosecur. Bioterror.* 5, 359–362. <https://doi.org/10.1089/bsp.2007.0923>.
11. Bügl, H., Danner, J.P., Molinari, R.J., Mulligan, J.T., Park, H.O., Reichert, B., Roth, D.A.,

Wagner, R., Budowle, B., Scripp, R.M., et al. (2007). DNA synthesis and biological security. *Nat. Biotechnol.* 25, 627–629. <https://doi.org/10.1038/nbt0607-627>.

12. Diggans, J., and Leproust, E. (2019). Next steps for access to safe, secure DNA synthesis. *Front. Bioeng. Biotechnol.* 7, 86. <https://doi.org/10.3389/fbioe.2019.00086>.

13. Casagrande, R., and Cerles, A.A. (2020). NTI/WEF white paper on DNA synthesis advances the discussion but misses the mark. *Health Secur.* 18, 67–68. <https://doi.org/10.1089/hs.2020.0004>.

14. Niu, C., Yu, D., Wang, Y., Ren, H., Jin, Y., Zhou, W., Li, B., Cheng, Y., Yue, J., Gao, Z., and Liang, L. (2013). Common and pathogen-specific virulence factors are different in function and structure. *Virulence* 4, 473–482. <https://doi.org/10.4161/viru.25730>.

15. Nawrocki, E.M., Bradshaw, M., and Johnson, E.A. (2018). Botulinum neurotoxin-encoding plasmids can be conjugatively transferred to diverse clostridial strains. *Sci. Rep.* 8, 3100. <https://doi.org/10.1038/s41598-018-21342-9>.

16. Elsworth, R.A.L., Diaz, C., Yang, J., de Figueiredo, P., Ternus, K., and Treangen, T. (2020). Synthetic DNA and biosecurity: nuances of predicting pathogenicity and the impetus for novel computational approaches for screening oligonucleotides. *PLoS Pathog.* 16, e1008649. <https://doi.org/10.1371/journal.ppat.1008649>.

17. Bradley, P., den Bakker, H.C., Rocha, E.P.C., McVean, G., and Iqbal, Z. (2019). Ultrafast search of all deposited bacterial and viral genomic data. *Nat. Biotechnol.* 37, 152–159. <https://doi.org/10.1038/s41587-018-0010-1>.

18. Balaji, A., Kille, B., Kappell, A.D., Godbold, G.D., Diep, M., Elsworth, R.A.L., Qian, Z., Albin, D., Nasko, D.J., Shah, N., et al. (2022). SeqScreen: accurate and sensitive functional screening of pathogenic sequences via ensemble learning. *Genome Biol.* 23, 133. <https://doi.org/10.1168/s13059-022-02695-x>.

19. DiEuliis, D., Carter, S.R., and Gronvall, G.K. (2017). Options for synthetic DNA order screening, revisited. *mSphere* 2, e00319-17. <https://doi.org/10.1128/mSphere.00319-17>.

20. West, R., and Gronvall, G.K. (2020). California shows the way for biosecurity in commercial gene synthesis. *Nat. Biotechnol.* 38, 1021. <https://doi.org/10.1038/s41587-020-0667-0>.

21. Eisenstein, M. (2020). Enzymatic DNA synthesis enters new phase. *Nat. Biotechnol.* 38, 1113–1115. <https://doi.org/10.1038/s41587-020-0695-9>.

22. Sayers, S., Li, L., Ong, E., Deng, S., Fu, G., Lin, Y., Yang, B., Zhang, S., Fa, Z., Zhao, B., et al. (2019). Victors: a web-based knowledge base of virulence factors in human and animal pathogens. *Nucleic Acids Res.* 47, D693–D700. <https://doi.org/10.1093/nar/gky999>.

23. Godbold, G.D., Kappell, A.D., LeSassier, D.S., Treangen, T.J., and Ternus, K.L. (2022). Categorizing sequences of concern by function to better assess mechanisms of microbial pathogenesis. *Infect. Immun.* 90, e0033421. <https://doi.org/10.1128/IAI.00334-21>.

24. Esveld, K.M. (2018). Inoculating science against potential pandemics and information hazards. *PLoS Pathog.* 14, e1007286. <https://doi.org/10.1371/journal.ppat.1007286>.

25. Falman, J.C., Fagnant-Sperati, C.S., Kossik, A.L., Boyle, D.S., and Meschke, J.S. (2019). Evaluation of secondary concentration methods for poliovirus detection in wastewater. *Food Environ. Virol.* 11, 20–31. <https://doi.org/10.1007/s12560-018-09364-y>.

26. Zhu, Y., Oishi, W., Maruo, C., Saito, M., Chen, R., Kitajima, M., and Sano, D. (2021). Early warning of COVID-19 via wastewater-based epidemiology: potential and bottlenecks. *Sci. Total Environ.* 767, 145124. <https://doi.org/10.1016/j.scitotenv.2021.145124>.

27. Lewis, G., Jordan, J.L., Relman, D.A., Koblenz, G.D., Leung, J., Dafoe, A., Nelson, C., Epstein, G.L., Katz, R., Montague, M., et al. (2020). The biosecurity benefits of genetic engineering attribution. *Nat. Commun.* 11, 6294. <https://doi.org/10.1038/s41467-020-19149-2>.

28. Tang, T.C., Tham, E., Liu, X., Yehl, K., Rovner, A.J., Yuk, H., de la Fuente-Nunez, C., Isaacs, F.J., Zhao, X., and Lu, T.K. (2021). Hydrogel-based biocontainment of bacteria for continuous sensing and computation. *Nat. Chem. Biol.* 17, 724–731. <https://doi.org/10.1038/s41589-021-00779-6>.

29. Curtiss, R., 3rd (1978). Biological containment and cloning vector transmissibility. *J. Infect. Dis.* 137, 668–675. <https://doi.org/10.1093/infdis/137.5.668>.

30. Wells, C.L., Johnson, W.J., Kan, C.M., and Balish, E. (1978). Inability of debilitated *Escherichia coli* chi 1776 to colonise germ-free rodents. *Nature* 274, 397–398. <https://doi.org/10.1038/274397a0>.

31. Hecker, M., Schroeter, A., Träder, K., and Mach, F. (1986). Role of *relA* mutation in the survival of amino acid-starved *Escherichia coli*. *Arch. Microbiol.* 143, 400–402. <https://doi.org/10.1007/BF00412809>.

32. Kokina, A., Kibilds, J., and Liepins, J. (2014). Adenine auxotrophy—be aware: some effects of adenine auxotrophy in *Saccharomyces cerevisiae* strain W303-1A. *FEMS Yeast Res.* 14, 697–707. <https://doi.org/10.1111/1567-1364.12154>.

33. Steidler, L., Neirynck, S., Huyghebaert, N., Snoeck, V., Vermeire, A., Goddeeris, B., Cox, E., Remon, J.P., and Remaut, E. (2003). Biological containment of genetically modified *Lactococcus lactis* for intestinal delivery of human interleukin 10. *Nat. Biotechnol.* 21, 785–789. <https://doi.org/10.1038/nbt840>.

34. Motomura, K., Sano, K., Watanabe, S., Kanbara, A., Gamal Nasser, A.H., Ikeda, T., Ishida, T., Funabashi, H., Kuroda, A., and Hirota, R. (2018). Synthetic phosphorus metabolic pathway for biosafety and contamination management of cyanobacterial cultivation. *ACS Synth. Biol.* 7, 2189–2198. <https://doi.org/10.1021/acssynbio.8b00199>.

35. Yoo, J.I., Seppälä, S., and O’Malley, M.A. (2020). Engineered fluoride sensitivity enables biocontainment and selection of genetically-modified yeasts. *Nat. Commun.* 11, 5459. <https://doi.org/10.1038/s41467-020-19271-1>.

36. Bej, A.K., Perlin, M.H., and Atlas, R.M. (1988). Model suicide vector for containment of genetically engineered microorganisms. *Appl. Environ. Microbiol.* 54, 2472–2477. <https://doi.org/10.1128/aem.54.10.2472-2477.1988>.

37. Knudsen, S.M., and Karlström, O.H. (1991). Development of efficient suicide mechanisms for biological containment of bacteria. *Appl. Environ. Microbiol.* 57, 85–92. <https://doi.org/10.1128/aem.57.1.85-92.1991>.

38. Contreras, A., Molin, S., and Ramos, J.L. (1991). Conditional-suicide containment system for bacteria which mineralize aromatics. *Appl. Environ. Microbiol.* 57, 1504–1508. <https://doi.org/10.1128/aem.57.5.1504-1508.1991>.

39. Jensen, L.B., Ramos, J.L., Kaneva, Z., and Molin, S. (1993). A substrate-dependent biological containment system for *Pseudomonas putida* based on the *Escherichia coli* *gef* gene. *Appl. Environ. Microbiol.* 59, 3713–3717. <https://doi.org/10.1128/aem.59.11.3713-3717.1993>.

40. Szafranski, P., Mello, C.M., Sano, T., Smith, C.L., Kaplan, D.L., and Cantor, C.R. (1997). A new approach for containment of microorganisms: dual control of streptavidin expression by antisense RNA and the T7 transcription system. *Proc. Natl. Acad. Sci. USA* 94, 1059–1063. <https://doi.org/10.1073/pnas.94.4.1059>.

41. Stirling, F., Bitzan, L., O’Keefe, S., Redfield, E., Oliver, J.W.K., Way, J., and Silver, P.A. (2017). Rational design of evolutionarily stable microbial kill switches. *Mol. Cell* 68, 686–697.e3. <https://doi.org/10.1016/j.molcel.2017.10.033>.

42. Wright, O., Delmans, M., Stan, G.B., and Ellis, T. (2015). GeneGuard: a modular plasmid system designed for biosafety. *ACS Synth. Biol.* 4, 307–316. <https://doi.org/10.1021/sb500234s>.

43. Wright, O., Stan, G.B., and Ellis, T. (2013). Building-in biosafety for synthetic biology. *Microbiology (Read.)* 159, 1221–1235. <https://doi.org/10.1099/mic.0.066308-0>.

44. Broto, A., Gaspari, E., Miravet-Verde, S., Dos Santos, V.A.P.M., and Isalan, M. (2022). A genetic toolkit and gene switches to limit *Mycoplasma* growth for biosafety applications. *Nat. Commun.* 13, 1910. <https://doi.org/10.1038/s41467-022-29574-0>.

45. Caliando, B.J., and Voigt, C.A. (2015). Targeted DNA degradation using a CRISPR device stably carried in the host genome. *Nat. Commun.* 6, 6989. <https://doi.org/10.1038/ncomms7989>.

46. Rottinghaus, A.G., Ferreiro, A., Fishbein, S.R.S., Dantas, G., and Moon, T.S. (2022). Genetically stable CRISPR-based kill switches for engineered microbes. *Nat. Commun.* 13, 672. <https://doi.org/10.1038/s41467-022-28163-5>.

47. Chan, C.T.Y., Lee, J.W., Cameron, D.E., Bashor, C.J., and Collins, J.J. (2016). 'Deadman' and 'Passcode' microbial kill switches for bacterial containment. *Nat. Chem. Biol.* 12, 82–86. <https://doi.org/10.1038/nchembio.1979>.

48. Cai, Y., Agmon, N., Choi, W.J., Ubide, A., Stracquadanio, G., Caravelli, K., Hao, H., Bader, J.S., and Boeke, J.D. (2015). Intrinsic biocontainment: multiplex genome safeguards combine transcriptional and recombinational control of essential yeast genes. *Proc. Natl. Acad. Sci. USA* 112, 1803–1808. <https://doi.org/10.1073/pnas.1424704112>.

49. Agmon, N., Tang, Z., Yang, K., Sutter, B., Ikushima, S., Cai, Y., Caravelli, K., Martin, J.A., Sun, X., Choi, W.J., et al. (2017). Low escape-rate genome safeguards with minimal molecular perturbation of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 114, E1470–E1479. <https://doi.org/10.1073/pnas.1621250114>.

50. McIsaac, R.S., Silverman, S.J., McClean, M.N., Gibney, P.A., Macinskas, J., Hickman, M.J., Petti, A.A., and Botstein, D. (2011). Fast-acting and nearly gratuitous induction of gene expression and protein depletion in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 22, 4447–4459. <https://doi.org/10.1091/mbc.E11-05-0466>.

51. McIsaac, R.S., Oakes, B.L., Wang, X., Dummit, K.A., Botstein, D., and Noyes, M.B. (2013). Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Res.* 41, e57. <https://doi.org/10.1093/nar/gks1313>.

52. Arita, Y., Kim, G., Li, Z., Friesen, H., Turco, G., Wang, R.Y., Clime, D., Usaj, M., Hotz, M., Stoops, E.H., et al. (2021). A genome-scale yeast library with inducible expression of individual genes. *Mol. Syst. Biol.* 17, e10207. <https://doi.org/10.1525/msb.202110207>.

53. Gallagher, R.R., Patel, J.R., Interiano, A.L., Rovner, A.J., and Isaacs, F.J. (2015). Multilayered genetic safeguards limit growth of microorganisms to defined environments. *Nucleic Acids Res.* 43, 1945–1954. <https://doi.org/10.1093/nar/gku1378>.

54. Mandell, D.J., Lajoie, M.J., Mee, M.T., Takeuchi, R., Kuznetsov, G., Norville, J.E., Gregg, C.J., Stoddard, B.L., and Church, G.M. (2015). Biocontainment of genetically modified organisms by synthetic protein design. *Nature* 518, 55–60. <https://doi.org/10.1038/nature14121>.

55. Rovner, A.J., Haimovich, A.D., Katz, S.R., Li, Z., Grome, M.W., Gassaway, B.M., Amiram, M., Patel, J.R., Gallagher, R.R., Rinehart, J., and Isaacs, F.J. (2015). Recoded organisms engineered to depend on synthetic amino acids. *Nature* 518, 89–93. <https://doi.org/10.1038/nature14095>.

56. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., and Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat. Methods* 6, 917–922. <https://doi.org/10.1038/nmeth.1401>.

57. Yesbolatova, A., Saito, Y., Kitamoto, N., Makino-Ito, H., Ajima, R., Nakano, R., Nakao, H., Fukui, K., Gamo, K., Tominari, Y., et al. (2020). The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. *Nat. Commun.* 11, 5701. <https://doi.org/10.1038/s41467-020-19532-z>.

58. Chung, H.K., Jacobs, C.L., Huo, Y., Yang, J., Krumm, S.A., Plemper, R.K., Tsien, R.Y., and Lin, M.Z. (2015). Tunable and reversible drug control of protein production via a self-excising degron. *Nat. Chem. Biol.* 11, 713–720. <https://doi.org/10.1038/nchembio.1869>.

59. Banaszynski, L.A., Chen, L.C., Maynard-Smith, L.A., Ooi, A.G.L., and Wandless, T.J. (2006). A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* 126, 995–1004. <https://doi.org/10.1016/j.cell.2006.07.025>.

60. Deckert, K., Budiardjo, S.J., Brunner, L.C., Lovell, S., and Karanicolas, J. (2012). Designing allosteric control into enzymes by chemical rescue of structure. *J. Am. Chem. Soc.* 134, 10055–10060. <https://doi.org/10.1021/ja301409g>.

61. Lopez, G., and Anderson, J.C. (2015). Synthetic auxotrophs with ligand-dependent essential genes for a BL21(DE3) biosafety strain. *ACS Synth. Biol.* 4, 1279–1286. <https://doi.org/10.1021/acssynbio.5b00085>.

62. Fujino, T., Tozaki, M., and Murakami, H. (2020). An amino acid-swapped genetic code. *ACS Synth. Biol.* 9, 2703–2713. <https://doi.org/10.1021/acssynbio.0c00196>.

63. Napolitano, M.G., Landon, M., Gregg, C.J., Lajoie, M.J., Govindarajan, L., Mosberg, J.A., Kuznetsov, G., Goodman, D.B., Vargas-Rodriguez, O., Isaacs, F.J., et al. (2016). Emergent rules for codon choice elucidated by editing rare arginine codons in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 113, E5588–E5597. <https://doi.org/10.1073/pnas.1605861113>.

64. Knight, R.D., Freeland, S.J., and Landweber, L.F. (2001). Rewiring the keyboard: evolvability of the genetic code. *Nat. Rev. Genet.* 2, 49–58. <https://doi.org/10.1038/35047500>.

65. Santos, M.A., Cheesman, C., Costa, V., Moradas-Ferreira, P., and Tuite, M.F. (1999). Selective advantages created by codon ambiguity allowed for the evolution of an alternative genetic code in *Candida* spp. *Mol. Microbiol.* 31, 937–947. <https://doi.org/10.1046/j.1365-2958.1999.01233.x>.

66. Robertson, W.E., Funke, L.F.H., de la Torre, D., Fredens, J., Elliott, T.S., Spinck, M., Christova, Y., Cervettini, D., Böge, F.L., Liu, K.C., et al. (2021). Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science* 372, 1057–1062. <https://doi.org/10.1126/science.abg3029>.

67. Zürcher, J.F., Robertson, W.E., Kappes, T., Petris, G., Elliott, T.S., Salmond, G.P.C., and Chin, J.W. (2022). Refactored genetic codes enable bidirectional genetic isolation. *Science* 378, 516–523. <https://doi.org/10.1126/science.add8943>.

68. Anderson, J.C., Wu, N., Santoro, S.W., Lakshman, V., King, D.S., and Schultz, P.G. (2004). An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci. USA* 101, 7566–7571. <https://doi.org/10.1073/pnas.0401517101>.

69. Neumann, H., Wang, K., Davis, L., Garcia-Alai, M., and Chin, J.W. (2010). Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* 464, 441–444. <https://doi.org/10.1038/nature08817>.

70. Hirao, I., and Kimoto, M. (2012). Unnatural base pair systems toward the expansion of the genetic alphabet in the central dogma. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* 88, 345–367. <https://doi.org/10.2183/pjab.88.345>.

71. Malyshev, D.A., and Romesberg, F.E. (2015). The expanded genetic alphabet. *Angew. Chem., Int. Ed. Engl.* 54, 11930–11944. <https://doi.org/10.1002/anie.201502890>.

72. Malyshev, D.A., Dhami, K., Lavergne, T., Chen, T., Dai, N., Foster, J.M., Corrêa, I.R., Jr., and Romesberg, F.E. (2014). A semi-synthetic organism with an expanded genetic alphabet. *Nature* 509, 385–388. <https://doi.org/10.1038/nature13314>.

73. Zhang, Y., Ptacin, J.L., Fischer, E.C., Aerni, H.R., Caffaro, C.E., San Jose, K., Feldman, A.W., Turner, C.R., and Romesberg, F.E. (2017). A semi-synthetic organism that stores and retrieves increased genetic information. *Nature* 551, 644–647. <https://doi.org/10.1038/nature24659>.

74. Pinheiro, V.B., Taylor, A.I., Cozens, C., Abramov, M., Renders, M., Zhang, S., Chaput, J.C., Wengel, J., Peak-Chew, S.Y., McLaughlin, S.H., et al. (2012). Synthetic genetic polymers capable of heredity and evolution. *Science* 336, 341–344. <https://doi.org/10.1126/science.1217622>.

75. Pezo, V., Liu, F.W., Abramov, M., Froeyen, M., Herdewijn, P., and Marlière, P. (2013). Binary genetic cassettes for selecting XNA-templated DNA synthesis in vivo. *Angew. Chem., Int. Ed. Engl.* 52, 8139–8143. <https://doi.org/10.1002/anie.201303288>.

76. Luo, M., Groaz, E., Froeyen, M., Pezo, V., Jaziri, F., Leonczak, P., Schepers, G., Rozenski, J., Marlière, P., and Herdewijn, P. (2019). Invading *Escherichia coli* genetics with a xenobiotic nucleic acid carrying an acyclic phosphonate backbone (ZNA). *J. Am. Chem. Soc.* 141, 10844–10851. <https://doi.org/10.1021/jacs.9b04714>.

77. Marlière, P. (2009). The farther, the safer: a manifesto for securely navigating synthetic species away from the old living world. *Syst. Synth. Biol.* 3, 77–84. <https://doi.org/10.1007/s11693-009-9040-9>.