

DISSERTATION

**Xenobiology in synthetic biology: From
non-standard amino acid incorporation to
biosecurity implications**

Submitted by

Svenja Vinke

For the Degree Doctor of Science

Bielefeld University

Faculty of Biology

Bielefeld, March 2023

This thesis is dedicated to my parents Eva and Torsten.

I love you and thank you for everything.

The practical work of this thesis has been performed at Bielefeld University in the Institute of Microbial Genomics and Biotechnology, from February 2020 until March 2023 as well as Harvard University from January 2022 until February 2023 under the supervision of Prof. Dr. Jörn Kalinowski.

Examiner

Prof. Dr. Jörn Kalinowski
Chair of Microbial Genomics and Biotechnology
Faculty of Biology & CeBiTec, Bielefeld University

Parts of this thesis have been published in:

1. Nanopore sequencing-guided directed evolution of aminoacyl-tRNA synthetases for the incorporation of non-standard amino acids into proteins. **Svenja Vinke** and Levin Joe Klages, Christian Rückert, Laura Schlüter, Louisa Sophie Lehmann, Sophia Nölting and Jörn Kalinowski. Submitted to MDPI Biology. 2023
2. Swapped genetic code blocks viral infections and gene transfer. Akos Nyerges, **Svenja Vinke**, Regan Flynn, Siân V. Owen, Eleanor A. Rand, Bogdan Budnik, Eric Keen, Kamesh Narasimhan, Jorge A. Marchand, Maximilien Baas-Thomas, Min Liu, Kangming Chen, Anush Chiappino-Pepe, Fangxiang Hu, Michael Baym, George M. Church. *Nature*. 2023. 615: 720–727
3. The Dual-Use Education Gap: Awareness and Education of Life Science Researchers on Nonpathogen-Related Dual-Use Research. **Svenja Vinke**, Irina Rais, and Piers Millett. *Health Security*. 2022. 2: 35-42.
4. How to teach life sciences students about dual-use research—a view from the field. **Svenja Vinke** and Michèle Gemünden. *Canadian Journal of Microbiology*. 2022. 69(1): 62-71

The supplementary material of the respective publication was added when applicable.

1 Abstract

In the era of sustainability, synthetic biology has emerged as a game changing field, holding immense potential to create innovative solutions to restore the ecological balance. The ability to produce peptides and proteins that can aid in achieving sustainability goals greatly depends on the development of effective protein design methods. However, conventional protein design approaches are limited by the 20 standard amino acids. This limitation calls for the exploration of alternative strategies for protein design that can expand the range of amino acids and allow the creation of novel proteins with enhanced functionalities. This challenge can be addressed by incorporating xenobiology in protein design strategies.

Xenobiology seeks to expand the genetic code beyond the four naturally occurring nucleotides (A, C, T and G) and the 20 standard amino acids to create new synthetic DNA and protein blocks.

To incorporate non-standard amino acids (nsAA) during translation, an orthogonal aminoacyl-tRNA synthetase (aaRS) needs to be introduced to endow the cell with the ability to charge a tRNA that is able to suppress a repurposed codon with the nsAA of choice. To make this system work a perfect aaRS and a codon that can be repurposed without influencing endogenous translation, apart from the incorporation of the nsAA, are needed. Current evolution methods for aaRS are limited in library design and alternative continuous evolution systems are not straightforward and can be hard to evaluate.

To address this first bottleneck, we established and validated a Nanopore sequencing guided phage-assisted evolution strategy. The evolution of a L-2-nitrophenylalanyl aaRS was used to demonstrate the power of this new evolution method. By designing a new software solution to filter sequencing errors, it was possible to deep sequence the complete CDS throughout the whole evolution. This resulted in the evaluation of the entire evolution by constructing evolutionary trees with data indicating the strength of every mutation enriched throughout the evolution. By performing the evolution in triplicates, it was shown that independent experiments produced proteins that show the same amino acid exchanges under the same selection pressure. This uncovered which mutation results in local and global maxima of the aaRS activity and showed that mutations outside the ligand binding pocket have a strong influence on enzyme activity and specificity.

Next to the aaRS, the choice of the repurposed codon has a strong influence on protein production with the nsAA. While in the evolution the most common used codon, the *amber* stop codon, was used to incorporate the non-standard amino acid, genetically recoded organisms are an alternative to create codons that do not interfere with the translational machinery. Genetically recoded organism are strains with one or several codons exchanged to synonymous

codons throughout all protein CDSs. In Syn61 Δ3, the *amber* stop codons and two of the serine codons are exchanged. In addition to production of non-standard amino acid containing proteins, strains also offer the ability to resist viral replication. When a phage enters Syn61 Δ3 cells, the 3 recoded codons cannot be read by the cell anymore, and the phage proteins are not produced, impairing viral replication. However, several phages bring their own tRNAs being able to replicate in the recoded strains.

To block this replication, leucine-serine swapped tRNAs were developed in this work. Through library and selection, we identified viral leucine tRNAs that tolerate an anticodon swap to the forbidden serine codons. These tRNAs can out-compete the viral tRNAs and lead to leucine incorporation at the serine codons, which leads to misfolding of the phage proteins and blocking phage replication. This swap can also be used to contain genetic information, making this system combined with genetically recoded organism a powerful tool for xenobiology.

Both publications demonstrate the high pace at which this research discipline develops. However, to make this field grow in a responsible way, biosafety and biosecurity regulations need to grow at the same pace. Since biosecurity governance is mainly limited to pathogen-related research, the assessment and mitigation on the xenobiology field needs to be done by the researchers. This work shows that there is a dramatic lack of knowledge regarding dual-use risks and little effective efforts to change this situation.

To make xenobiology research live up to its potential, responsibility, safety and security need to become a crucial aspect in project planning and execution, starting by educating young scientists about these topics over establishing suitable frameworks in governance.

2 List of abbreviations

2NPA	2-nitro-phenylalanine
aaRS	tRNA-aminoacyl-synacyl synthetase
amp	ampicillin
bp	Base pairs
BWC	Biological weapons convention
CDS	Coding DNA sequence
cm	chloramphenicol
DNA	Deoxyribonucleic acid
EF-Tu	Elongation Factor Tu
GMO	Genetically modified organism
GRO	Genetically recoded organism
GFP	Green fluorescent protein
h	Hours
LC	liquid chromatography
MALDI	Matrix associated laser desorption ionization
MOI	Multiplicity of infection
MS	mass spectrometry
nsAA	Non-standard amino acid
nsbp	Non-standard base pair
ONT	Oxford Nanopore Technologies
ori	Origin of replication
PACE	Phage-assisted continuous evolution
PANCE	Phage-assisted non-continuous evolution
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PFU	Plaque forming units
PylRS	Pyrrolysyl tRNA-aminoacyl-synacyl synthetase
RBS	Ribosome binding site
t	Time
TAE	Tris-acetate-EDTA
tet	tetracycline
Tris	Tris(hydroxymethyl)aminomethane
tRNA	transfer ribonucleic acid
TyrRS	Tyrosyl tRNA-aminoacyl-synacyl synthetase
WT	wildtype

Contents

1 Abstract	9
2 List of abbreviations	12
3 Introduction	16
3.1 Xenobiology	16
3.2 Non-standard amino acids	18
3.2.1 Incorporation of non-standard amino acids in <i>E. coli</i>	19
3.2.2 <i>Methanocaldococcus jannaschii</i> tyrosyl tRNA-aminoacyl synthetase . .	21
3.3 Directed evolution	22
3.3.1 Directed evolution of tRNA-aminoacyl synthetases	24
3.3.2 Phage-assisted evolution of tRNA-aminoacyl-synthetases	26
3.4 Genetically recoded organism	28
3.5 Biosecurity	29
3.5.1 Regulation of Xenobiology research	29
4 Results	32
4.1 Nanopore sequencing-guided directed evolution of aminoacyl-tRNA synthetases for the incorporation of non-standard amino acids into proteins	32
4.1.1 Abstract	32
4.1.2 Introduction	32
4.1.3 Materials and Methods	34
4.1.4 Results	38
4.1.5 Discussion	47
4.2 Swapped genetic code blocks viral infections and gene transfer	50
4.2.1 Abstract	50
4.2.2 Introduction	50
4.2.3 Materials and Methods	51
4.2.4 Results	67
4.2.5 Discussion	81
4.3 The Dual-Use Education Gap: Awareness and Education of Life Science Re- searchers on Nonpathogen-Related Dual-Use Research	84
4.3.1 Abstract	84
4.3.2 Introduction	84
4.3.3 Materials and Methods	86

4.3.4	Results and Discussion	88
4.3.5	Conclusion	95
4.4	How to teach life sciences students about dual-use research—a view from the field	96
4.4.1	Abstract	96
4.4.2	Introduction	96
4.4.3	Materials and Methods	98
4.4.4	Results and Discussion	98
4.4.5	Conclusion	112
5	Discussion	115
5.1	tRNA-aminacyl-synthetase evolution	115
5.1.1	Directed evolution of aaRS pairs	115
5.1.2	Deep-Sequencing guided evaluation of directed evolution experiments .	117
5.1.3	Challenges using PACE or PANCE	118
5.1.4	Alternative aaRS/tRNA pairs	120
5.2	Alternatives to the <i>amber</i> codon	122
5.3	Biosecurity risks	124
5.3.1	General problems with risk assessment in Synthetic biology	124
5.3.2	Responsibility	125
5.3.3	Biosecurity risks of Xenobiology	126
5.3.4	Necessary mitigation steps for biosecurity risks	127
6	Conclusion	129
7	Acknowledgments	130
8	References	130
9	Appendices	154

3 Introduction

3.1 Xenobiology

Xenobiology, derived from the Greek word *xenos* meaning “stranger/alien”, originally was synonymous to the description of astrobiology (Wooster, 1961). With the rise of synthetic biology, the definition expanded. In synthetic biology, xenobiology describes the field of research that investigates or designs biological systems that differ from the conventional 4 base pair-translation into mRNA- 20 amino acid dogma, which is fairly universal in life on earth (Kubyshkin *et al.*, 2018; Budisa *et al.*, 2020). With this new, and still evolving definition, this field of research contains diverse research directions, from origin of life studies, that investigate why life on earth evolved into this 4 bp-20 amino acid system (Preiner *et al.*, 2020), to astrobiology trying to find other life forms (Chyba & Hand, 2005; McKay, 2004), over to synthetic biology trying to create those (Rennekamp, 2019). In general xenobiology on synthetic biology level can be split into 3 research disciplines: The work on xeno-DNA and RNA (Pinheiro, 2012), recoding organism to carry an altered genetic code (Lajoie *et al.*, 2013), and the incorporation of non-standard amino acids (nsAA) (Noren *et al.*, 1989).

The diverse phenotypes of life on this planet are based on the encoding capabilities of DNA, RNA and polypeptides. The chemical properties of their building blocks, nucleic acids and amino acids are limited. The expansion of nucleic acid and amino acid choices offers new possibilities that overcome the limitations by introducing new chemical groups that enable to catalyze new reactions and fold in different conformations.

Xeno-DNA and RNA are biopolymers that are build entirely or partly of modified nucleic acids. The modification can be at the nucleobase, sugar or the triphosphate, depending on the desired function. Chemical modification can be added to the N7 in purines as well as to the C5 in pyrimidines to create DNAzymes, improved aptamer selection, or to mimic natural C5 pyrimidine modification (Hollenstein *et al.*, 2009; Vaught *et al.*, 2010; Gommers-Ampt & Borst, 1995). These modifications can usually be introduced without altering DNA properties if the modification is introduced in only a few sites (Hasegawa *et al.*, 2006; Kuwahara *et al.*, 2006; Weisbrod & Marx, 2008). However, big modifications, such as large hydrophobic fluorophores, result in altered physio-chemical properties (Jäger & Famulok, 2004; Jäger *et al.*, 2005).

Using nucleobases with altered binding properties to the counterstrand can be used to expand the genetic alphabet. Significant achievements in this area include the creation of functional alternative nucleobase pairs with altered donor-acceptor pattern or hydrophobic interactions (Hoshika *et al.*, 2010; Yang *et al.*, 2006; Kimoto *et al.*, 2009; Seo *et al.*, 2011).

Alternatively to modifying the nucleobase, changing the sugar or the phosphodiester bond can increase storage capabilities of DNA and cause the loss of the ability to cross-talk with extant biology (Schmidt, 2010). The main challenge working with Xeno-DNA and RNA to expand the genetic alphabet *in vivo* is to reprogram the cellular machinery to accept these bases (Pinheiro & Holliger, 2012). The cellular machinery is evolved to get rid of bases in the DNA that show a wrong acceptor-donor pattern, since these pose a high risk of DNA mutation. One mechanism where this is the case is isocytosine which is a tautomerization product that gets deaminated to uracil by cytosine deaminase CodA in *E. coli* cells (Hall *et al.*, 2011). Another problem is to change the translational machinery to accept new RNA bases. This has not been achieved *in vivo* for nucleobases with altered acceptor-donor pattern yet, and for hydrophobic bases, incorporation of the new nucleobase was only accepted at base position two in the anticodon (Fischer *et al.*, 2020).

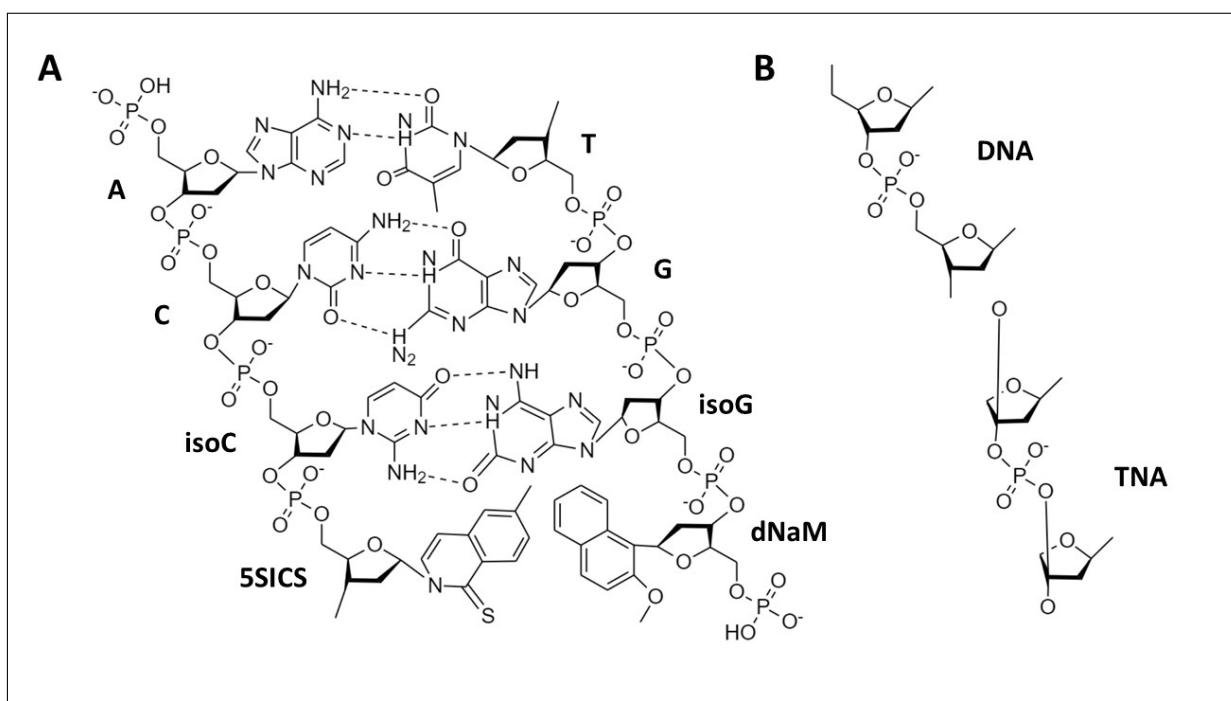


Figure 1: A Donor-acceptor pattern of the canonical nucleobases A, T, C and G as well as the two most used non-standard base pairs isocytosine and isoguanine first described by Switzer *et al.* in 1989 and 5SICS and dNaM first described by Seo *et al.* in 2009 (Switzer *et al.*, 1989; Seo *et al.*, 2009). B Threose nucleic acid developed by Eschenmoser in 1999 as alternative to the standard DNA backbone (Eschenmoser, 1999).

The incorporation of one unnatural base pair poses a great opportunity for synthetic biology. One new base pair alone offers the opportunity to use 152 new DNA codons that can be used

for incorporation of multiple nsAAs. Even the creation of completely synthetic DNA with only non-standard base pairs was proposed by (Schmidt, 2010).

All in all, unnatural bases pose an immense potential in various fields of synthetic biology, but there is still a long way to synthetic organism with completely altered DNA, or even the stable maintenance of just one unnatural base pair.

3.2 Non-standard amino acids

Non-standard amino acids (nsAAs) are amino acids that do not belong to the pool of the proteinogenic 22 amino acids. Technically, every chemical molecule with a carboxylic acid group (-COOH) and an amine group (-NH₂) is considered an amino acid. nsAA occur naturally in cells where they function for example as precursor or degradation product in amino acid metabolism, metabolite in e.g., non-ribosomal peptide synthesis or even as signal molecules such as the neurotransmitter GABA (Koyack & Cheng, 2006; Coxon *et al.*, 2005; Kuriyama & Sze, 1971). The chemical properties of the standard 22 amino acids are limited and a lot of them posses side chains with very similar structure and chemical properties (e.g. leucine and isoleucine) (Braun & Venkatarajan, 2001). This leaves a lot of unexplored space in protein design, because a lot of chemical groups are not available in the pool of amino acids that can be naturally incorporated into proteins by translation.

The side chains of novel nsAAs can be very divers, thus proteins with before unimaginable functions can be derived from this new form of protein design. So far more than 150 nsAAs have been incorporated with a wide range of function (Liu & Schultz, 2010). As diverse as the functions of this amino acids are, are their applications. One example is the enhancement or introduction of completely new catalytic functions as demonstrated e.g. by Zhang 2002, by the incorporation of alkenes. Other examples include the incorporation of an alkyne by Deiters in 2005 and the incorporation of m-acetyl-l-phenylalanine and p-acetyl-l-phenylalanine by Zhang in 2003(Zhang *et al.*, 2002; Deiters & Schultz, 2005; Zhang *et al.*, 2003).

A very commonly used application for nsAA is the introduction of side groups for easy chemical modification of nsAAs, such as click-chemistry. Additionally other groups used for common chemical reactions to incorporate probes that can then be used to investigate protein structure or function through, e. g., IR, NMR or fluorescence have been incorporated (Ye *et al.*, 2009; Taki *et al.*, 2002; Cellitti *et al.*, 2008; Fekner *et al.*, 2009; Hagiwara *et al.*, 2008; Bae *et al.*, 2003). This has been demonstrated by Wang et al. in 2006 by the incorporation of the fluorescent amino acid l-(7-hydroxycoumarin-4-yl) ethylglycine to label the myoglobin (Wang *et al.*, 2006).

In general photoreactive amino acid show great promise for the regulation and study of a lot of cellular processes as demonstrated by various groups (Chen *et al.*, 2009; Deiters *et al.*, 2006;

Lemke *et al.*, 2007; Wu *et al.*, 2004; Chen *et al.*, 2007; Chin *et al.*, 2002). Prominent examples have been photocaging, which is the protection of amino acid side chains by chemical groups that are split off upon light irradiation, as demonstrated by Deiters in 2006, by photocaging tyrosine and thus regulating β -galactosidase (Deiters *et al.*, 2006). Another example is the incorporation of L-2-nitro-phenylalanine, which can cleave the peptide backbone at its side of incorporation upon light irradiation with light of a certain wavelength, showing potential to be used as tool for activation and deactivation of a wide range of proteins and peptides (Peters *et al.*, 2009).

Another area of high interest is the mimicking of naturally occurring post-translational modifications (PTMs). This area is from high interest, due to the fact that production of proteins with PTMs from organism that are hard or expensive to cultivate often is no option for large scale production of the desired protein. The introduction of the PTM by the introduction of an nsAA which already carries the desired chemical group enables production of an identical protein in *E. coli*, lowering the production cost significantly (Lemke *et al.*, 2007; Rogerson *et al.*, 2015; Vázquez *et al.*, 2003; Zhang *et al.*, 2008; Liu *et al.*, 2009; Neumann *et al.*, 2008). This has been shown for a wide range of examples, e.g. the incorporation of phosphoserine in ubiquitin and a kinase (Rogerson *et al.*, 2015). This will be one of the most relevant research areas in protein design of proteins containing nsAA, due to the high interest of the pharmaceutical industry, based on the huge potential of producing biopharmaceuticals cheaper and with novel functions.

3.2.1 Incorporation of non-standard amino acids in *E. coli*

Chemical peptide synthesis can be used to incorporate any amino acid of choice into a peptide chain, but is very expensive and thus unsuitable for large scale production of proteins (Isidro-Llobet *et al.*, 2019).

Since most applications require incorporation of the nsAA at one or more defined site in a target protein, another approach was needed. The most prominent is the earlier mentioned approach of re-purposing a codon, mostly the *amber* stop codon, first shown by Furter *et al.* in 1998 (Furter, 1998). This approach has proven as a powerful tool to incorporate a wide variety of nsAAs (Liu & Schultz, 2010).

To understand how the suppression of the *amber* stop codon works, understanding of the natural end of translation process is crucial. In most unedited biological systems, at least some protein CDS end with an *amber* stop codon (Lajoie *et al.*, 2013). During translation, when the ribosome encounters this end of the coding sequence, there is no corresponding tRNA, the ribosome stalls, causing the release factor one binding, enabling the ribosome to detach from the RNA and release the peptide chain (Capecchi, 1967).

In some organisms such as the archaea *Methanosaeca mazei*, this system differs drastically. In this organism an additional amino acid, pyrrolysine, is encoded and incorporated in response to the *amber* stop codon (Rother & Krzycki, 2010). *M. mazei* does this by possessing an orthogonal tRNA with the anticodon to the *amber* stop codon and an orthogonal tRNA-aminoacyl synthetase, which can charge pyrrolysine to this tRNA.

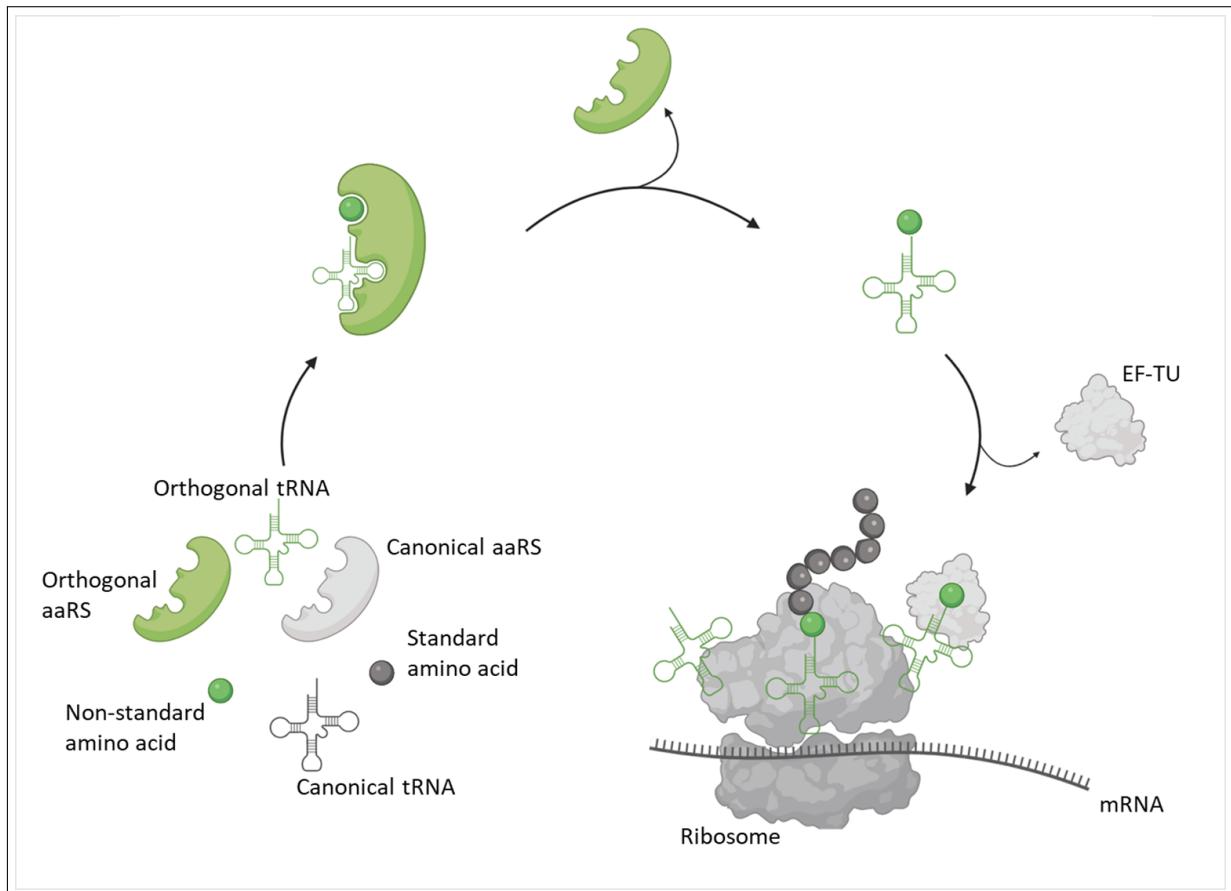


Figure 2: Incorporation of non standard amino acids. In addition to the canonical amino acids, tRNAs and tRNA-aminoacyl synthetases, an orthogonal tRNA-aminoacyl synthetase and tRNA with an anticodon suppressing the *amber* stop codon are produced within the cell. The orthogonal tRNA-aminoacyl synthetase charges a nsAA to the *amber* suppressor tRNA. This charged tRNA gets transported to the ribosome by the elongation factor Tu. The tRNA suppresses the *amber* codon and the nsAA gets incorporated at the sites in the polypeptide chain, every time an *amber* codon is in frame.

By taking this biological system as an inspiration of how to incorporate aaRS during translation using an orthogonal tRNA-aminoacyl synthetase pair, research groups have been working on translating this system for the use in other organisms, mainly *E. coli* (Namy *et al.*, 2007). The *M. mazei* PylRS can be used to incorporate pyrrolysine in *E. coli* and has been used to incorporate

a wide range of amino acid, most of them very similar to pyrrolysine (Liu & Schultz, 2010). Recently, Koch *et al.* have succeeded to use rational design of the PylRS to incorporate small aromatic amino acids but before, the substrate acceptance of the enzyme and also stability problems. Those have only partially been solved by adding solubility tags, prevented PylRS to be the ideal candidate for nsAA incorporation (Koch *et al.*, 2021a; Koch *et al.*, 2021b). These limitations in PylRS use have caused the rise of another aaRS that have been more broadly used since it overcomes most limitations in nsAA incorporation: The *Methanocaldococcus jannaschii* tyrosyl aaRS.

3.2.2 *Methanocaldococcus jannaschii* tyrosyl tRNA-aminoacyl synthetase

The *Methanocaldococcus jannaschii* tyrosyl tRNA-aminoacyl synthetase (TyrRS), first described as tool for genetic code expansion by Wang *et al.* in 2001, is the native aaRS charging tyrosine to tRNAs in the hyperthermophilic methanogen *M. jannaschii*, first described as *Methanococcus jannaschii* (Wang & Schultz, 2001; Jones *et al.*, 1983).

The *M. jannaschii* TyrRS is a homodimeric enzyme that is significantly shorter than its bacterial counterparts with a small anticodon loop (Steer & Schimmel, 1999). The crystal structure of the *M. jannaschii* TyrRS in complex with its tRNA, resolved in 2003, reveals the structural basis for the orthogonality to bacterial aaRS (Kobayashi *et al.*, 2003). The strict recognition of the C1-G72 base pair, and the discriminator base A73 of the tRNA differ significantly from the recognition by different residues in bacteria and thus prevents cross-talking with the *E. coli* TyrRS and tRNA-Tyr (Kobayashi *et al.*, 2003).

The binding pocket is located deep within the enzyme, making it difficult to adapt the binding pocket for larger amino acids, but also enables a high specificity towards the ligand. Only small changes in the amino acid sequence are necessary to adapt the enzyme to recognize alternative ligands like shown for a vast range of nsAAs (Liu & Schultz, 2010).

When the native ligand tyrosine is bound, hydrogen bonds to Y32, D158, E36, Q173, Y151 and Q155 are formed, resulting in a small conformational change within the binding pocket (Zhang *et al.*, 2005). To generate TyrRS variant that recognize other or nsAAs, library design focuses on these amino acids (Liu & Schultz, 2010). Beside the binding pocket, other domains are also involved in ligand specificity and binding. For example, the loop 73-83, positioned at the entrance of the binding pocket, provides a hydrophobic lid over the binding pocket in reaction to the conformational changes upon tyrosine binding and to prevent the activated tyrosine to react with water during the catalytic reaction catalyzed by the aaRS (Zhang *et al.*, 2005).

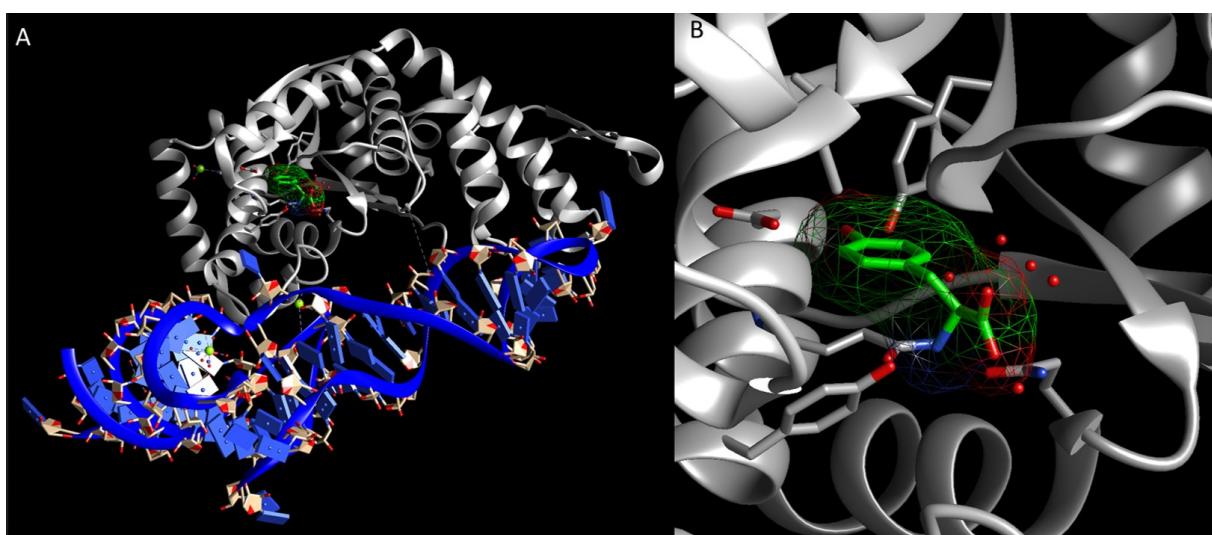


Figure 3: *Methanococcus jannaschii* in complex with its tRNA (RCSB identifier: 1j1u (Kobayashi *et al.*, 2003)). A Structure of the aaRS monomer (gray) in complex with the tRNA (blue) that gets charged by the other enzyme in the TyrRS dimer with tyrosine (green). B binding pocket of the *M. jannaschii* TyrRS. The amino acids that form hydrogen bonds once the ligand tyrosine (green) is bound are shown in stick presentation, while the backbone is shown in ribbon presentation.

To this date, it has been the most used orthogonal aaRS/tRNA pair, even though the nsAAs incorporated, are mainly tyrosine analogous (Liu & Schultz, 2010).

3.3 Directed evolution

Directed evolution is the artificial design and performance of darwinian evolution *in vivo* or *in vitro* to evolve parts of biological systems. It usually involves the design of a library of a diversified biological component. This component can be a whole cells, protein, peptide, mRNA or DNA molecule that is then selected towards a desired property, often increased functionality or specificity.

Every design of a directed evolution experiment contains two essential components: The library design, and the selection design including readout.

When focusing on evolving a target protein, which most direct evolution processes focus on our days, the DNA sequence of the protein must be diversified. For this process several methods have been developed over the last centuries. They can be split in two categories: Rational and random diversification.

Random diversification is usually done by random mutagenesis or including random bases throughout the CDS or parts of the CDS of a target protein. The most prominent example

for this is error-prone PCR. Error-prone PCR was developed in 1995 by Fromant *et al.* and makes use of changing PCR conditions to increase the error rate in replication during amplification (Fromant *et al.*, 1995). First changing the concentration of MgCl₂ was used to cause this effect. later, more refined protocols including the incorporation of non-standard bases that increase error rate have been used. Error-prone PCR can be used to diversify the whole length CDS or parts of it if more detailed information about the protein suggest this would be enough to achieve the desired property (Gruet *et al.*, 2012; Cirino *et al.*, 2003).

Error-prone PCR has the disadvantage that it also includes the mutation to stop codons and thus results in a lot on non-functional library clones. Since the selection usually has a limit of library clones that can be screened (often determined by transformation efficiency and high-throughput suitability of the selection), more refined design strategies than error prone are often necessary. If some information about the target protein is available, e.g. the amino acid residues involved in the desired reaction, site directed mutagenesis can be a useful alternative. In site directed mutagenesis, oligonucleotides are used to randomize only certain codons to change the desired amino acid codon to NNN or NNK (NNK avoids stop codon formation at the chosen position). This way, smaller libraries can be designed and conserved regions stay untouched enabling a more elegant library design. There are methods allowing the incorporation of several NNN/K oligonucleotides over the CDS length e.g. SMOOT or Darwin assembly (Cozens & Pinheiro, 2018; Cerchione *et al.*, 2020).

Some directed evolution techniques, e.g. directed evolution, require a diversification method that works *in vivo*. For this approach error-prone polymerases and expression plasmids that express proteins and subunits that increase the mutation rate within the cells have been developed (Alexander *et al.*, 2014; Camps *et al.*, 2003; Badran & Liu, 2015). Badran *et al.* developed a series of mutation plasmids that increase the mutation rate during replication to be used in directed evolution experiments (Badran & Liu, 2015).

Error-prone polymerases contain a mutant subunit that is usually involved in proof-reading during replication. This subunit is now unable to fulfill this function (Camps *et al.*, 2003). It has been shown that the error-prone variant of PolIII containing the epsilon subunit encoded by mutD(dnaQ) enables efficient increase of the mutation rate (Cox & Horner, 1986). This system has been refined by fusing the error-prone polymerase to an nCas enabling targeted mutation of desired DNA regions (Halperin *et al.*, 2018).

In general the choice of the diversification methods relies heavily on the target protein and available selection. If a lot about the target protein is known, e.g. the crystal structure, targeting specific residues is more likely to yield a good library. If nothing is known, error-prone is a

good approach, and if the selection requires continuous mutation, the *in vitro* methods are the only option.

Next to choosing a suitable diversification, the design of the selection is the second crucial design choice. The desired property change of the target protein needs to be linked to a selection pressure or readout. This can be fairly easy, e.g. the selection of an improved antibiotic resistance where the library is selected by simply increasing the concentration of the respective antibiotic, to very complex systems involving enzyme cascades.

In the field of directed evolution it is debated if directed evolution necessarily needs to involve a selection pressure or if screening the clones (e.g. for a fluorescence reporter) is also considered directed evolution. For the scope of this work, directed evolution will only be considered for a selection pressure based selection, and not screening of the library.

To design a selection pressure for directed evolution, a selection pressure needs to be found that can be overcome by the production of the target which shall be evolved. This can be directly enabled by the choice of target protein, like the example of the antibiotic resistance mentioned above, or linked through artificial pathways. The choice of selection is heavily dependent of the target protein and there is no general applicable strategy.

3.3.1 Directed evolution of tRNA-aminoacyl synthetases

In regards to directed evolution, tRNA-aminoacyl synthetaseacyl-synthetases that are supposed to charge *amber* suppressor tRNA, have the advantage that they can be easily linked towards the translation of a target protein. Therefore, the target protein needs to contain *amber* codon throughout their CDS at sites where the incorporation of the non-standard or standard amino acid has no influence on the protein function. Based on this, a library of aaRS can easily be selected for activity. The main problem with aaRS evolution is that next to their activity it is also crucial that these are specific, meaning they only incorporate the amino acid of choice and no canonical amino acid. Due to this, evolution of aaRS is usually split in two selection rounds: A positive selection for activity and a negative selection for specificity.

In the positive selection the aaRS library gets co-transformed with the positive selection plasmid which contains a marker, usually an antibiotic resistance like chloramphenicol or kanamycin with *amber* codons at permissive sites (Wang *et al.*, 2006; Melançon & Schultz, 2009; Liu & Schultz, 2010). The co-transformants are then plated on agar plates containing increasing concentrations of the antibiotic and the nsAA. Clones that survive on the antibiotic are then isolated and co-transformed with the negative selection plasmid (Wang *et al.*, 2006).

The negative selection plasmid contains the CDS of a toxin that contains *amber* codons at permissive sites, e.g. the barnase, an enzyme that causes lethal RNA degradation if expressed

in *E. coli* (Wang *et al.*, 2006; Prior *et al.*, 1991). The co-transformants are then plated on plates not containing the nsAA. Clones that incorporate wrong amino acids will still be able to suppress the *amber* codon while clones with a specific aaRS variant will not be able to express the full-length protein. Thus only clones with a specific aaRS can survive this selection.

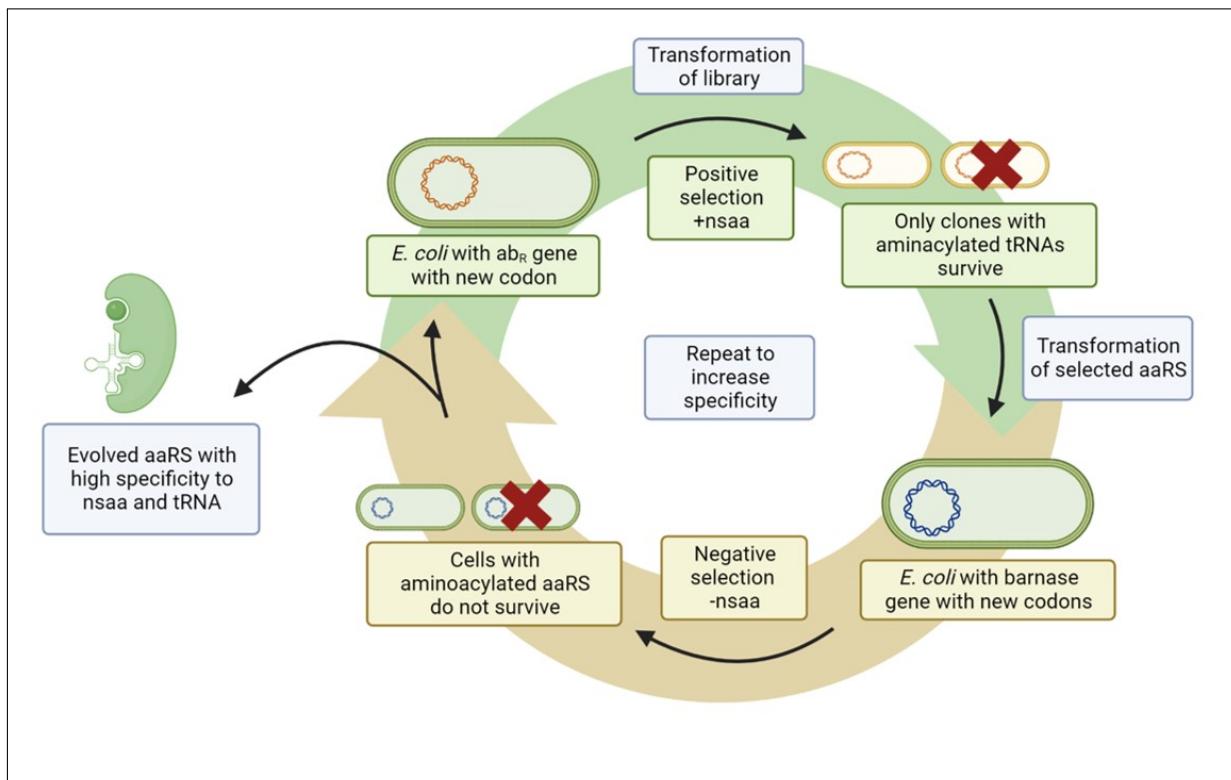


Figure 4: Schematic overview over two round aaRS selection. An aaRS library is created and selected in iterative rounds of positive and negative selection, leading to the identification of the aaRS clone(s) with improved activity and specificity. In the positive selection, the aaRS is co-expressed with an antibiotic resistance containing *amber* codons at permissive sites. Only cells with an active aaRS are able to incorporate the nsAA of choice for these codons and generate full-length antibiotic resistance referring proteins. In the negative selection, the nsAA is not present in the media. The aaRS is co-expressed with a toxic barnase with *amber* codons at permissive sites. Only clones with an unspecific aaRS charge the tRNAs and generate full length toxin.

Usually positive and negative selection gets performed for several rounds ending with a positive selection to ensure the best chance to prevent inactive aaRS clones.

There have been several variations of this basic method of aaRS selection including different resistances in the positive selection and different toxins in the negative selection (Chatterjee *et al.*, 2012; Ruan *et al.*, 2008). Only few selections have explored a one selection cycle method,

e.g. the selection based on N-terminal protein degradation by Kunjapur *et al.* in 2018 (Kunjapur *et al.*, 2018).

All of these selection methods have the disadvantages that they do not allow continuous mutation of the aaRS library and often focus only on the aaRS residues in the binding pocket of the amino acid, resulting in libraries that leave out exploring the activity increase of aaRS by mutations that are involved for example in stability or tRNA biding.

3.3.2 Phage-assisted evolution of tRNA-aminoacyl-synthetases

To solve the problem of insufficient library generation and come closer to darwinian evolution, Esvelt *et al.* developed Phage-assisted continuous evolution (PACE) in 2011.

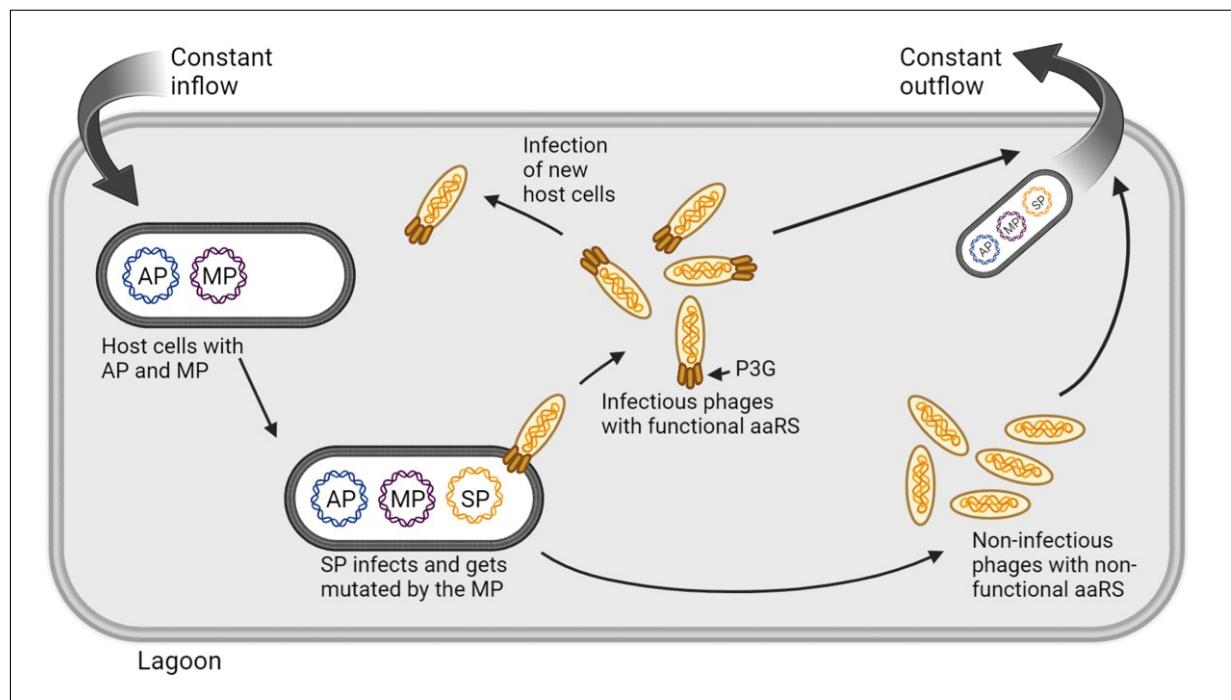


Figure 5: Schematic overview over phage-assisted continuous evolution(PACE). There is a constant in- and outflow of host cells into the reaction ch *amber*, the lagoon. The host cells contain an accessory plasmid encoding for pIII production, dependent on target protein activity, and a mutation plasmid, that when induced increases the error-rate during DNA replication. Selection phages are added to the lagoon to start the selection. They infect the host cells, get replicated and MP4 introduces mutations during replication. If the phage variant is able to produce an active variant, they are able to induce pIII production from the accessory plasmid and assemble functional phages. Those get released from the cell and can infect new host cells. The phage production is proportional to target protein activity, this way the most active phage gets replicated faster than less active phages.

The PACE system contains 3 fundamental components:

1. A mutation plasmid that expresses proteins that increase the error-rate during DNA replication, and thus enables continuous mutation of our target CDS during evolution.
2. A selection phage that is an M13 phage where the CDS for pIII, which is an essential protein for replication of this phage is deleted and instead the CDS of the protein that is supposed to be evolved, the aaRS is cloned.
3. The accessory or selection plasmid which only enables pIII expression if the aaRS is active.

PACE relies on the use of an M13 phage. M13 phages are *E. coli* bacteriophages widely used in biochemistry research e.g. for the production of single stranded DNA or phage display (Rakonjac *et al.*, 2016). The phage has a single strand 6407 kb genome that carries the CDS for 11 genes (Sattar *et al.*, 2015). That are either involved in replication, DNA packaging, transportation of the phage outside the cell or building the phage coat (Rakonjac *et al.*, 2016). The phage coat is built out of five proteins, PIII and PVI are building the head of the phage, pIII is involved in f-pili binding in the infection process and also release of the newly assembled phage from the infected cell. pVIII is the major coat protein, and with approx. 2700 copies the most prevalent in the phage coat, it builds a cylindrical body of approx. 6 nm with and approx. 900 nm length that ends in pVII and pIX building the phage tail (Sattar *et al.*, 2015). During PACE, this phage is modified to work as a unit that carries the genetic information about a protein that is supposed to get evolved instead of the pIII CDS. This way, the selection phage is not able to replicate on its own, since it is lacking one of the essential proteins for release from its host cell and the infection of new cells.

The mutation plasmids used in PACE have been developed by Badran.2015 *et al.* in 2015 (Badran & Liu, 2015). They are a range of different plasmids with different mutation strength. The mutation ability of these plasmids relies on the expression of a range of proteins that are involved in DNA replication and proof-reading, but the variants expressed increase the error-frequency in DNA replication, e. g., *dnaQ926*, *dam*, *seqA*, *emrR*, *ugi* and *cda* encoded by the MP4 plasmid (Badran & Liu, 2015). Depending on the desired mutation rate, one of the mutation plasmids is chosen and transformed into the selection cells. After induction of protein production, approx. 30 min before phage infection of the host cells, the phage DNA gets mutated when replicated in the cell. Due to the higher replication rate, the phage accumulates mutations a lot faster compared to the host cell genome and plasmids (Esvelt *et al.*, 2011).

The last component of PACE is the accessory or selection plasmid. This plasmid needs to enable pIII production dependent on the target protein activity, in case of aaRS evolution, it does so by having two amino acid codons in the CDS replaced by *amber* stop codons (Bryson *et al.*,

2017). If there is no active aaRS present translation stops at this codon. If there is active aaRS present, the nsAA is incorporated at that position, full length pIII is produced and an infectious phage can be assembled. The design of the accessory plasmid is highly dependent on the target protein.

3.4 Genetically recoded organism

Alternatively to incorporating non-standard DNA bases to generate new codon, organisms can also be recoded to change synonymous DNA codons to blank codons. *E. coli*, and all other life forms have 64 DNA codon, encoding for 20 amino acids (with the exceptions of some species incorporating also selenocysteine and pyrrolysine) and three stop codons (Krzycki, 2005). In theory encoding of twenty amino acids and one stop codon should be possible by 21 codons, but for several amino acids, multiple amino acid codon encode for the same amino acids, and three codons encode a translational stop. Serine for example is encoded by 6 DNA codons. To have more codon available for amino acid encoding, several research groups have explored the possibility to use codons that are very little used to encode nsAAs. In 1998 Furter *et al.* used the *amber* stop codon, the least used stop codon in *E. coli* to incorporate the nsAA *p*-fluorophenylalanine with the native phenylalanyl aaRS from *Saccharomyces cerevisiae* (Furter, 1998). Later, other groups used the least used leucine codon (Kwon & Choi, 2016). But all suppression of little used codons in the cells has an influence on the cellular machinery. Proteins where translation usually stops at these codons get elongated, destroying or influencing the function depending on the protein.

To avoid this problem, but also harnessing the opportunity that comes with synonymous codon, the Church lab in 2013 published the first Genetically Recoded Organism (C321) (Lajoie *et al.*, 2013). C321 is an *E. coli* strain where all *amber* stop codons have been changed to the synonymous codons TGA and TAA. This way, the cellular machinery is not influenced by an introduced *amber* suppression machinery. They later also optimized this strain by deleting the release factor I, which is responsible for termination of translation at *amber* and TAA stop codons and is in competition with the suppressor machinery, and thus leads to impure protein yield.

The approach of redesigning the genome of organism to delete synonymous codons has since then be expanded to another strain which has synonymous arginine codons deleted to offer two blank codons, to the to this date most recoded organism Syn61 Δ3 (Napolitano *et al.*, 2016; Fredens *et al.*, 2019). Syn61 Δ3 published by the Chin Lab in 2019 is an *E. coli* strain with the *amber* stop codon, and two serine codons, TCA and TCG, are recoded to synonymous codons in all protein CDSs. The two corresponding serine tRNAs encoded by *serT* and *serU* have been

deleted, leaving Syn61 as the first semi-synthetic organism with a 61 codon genetic code. This way, 3 different nsAAs can be incorporated as shown by Robertson *et al.* 2021 (Robertson *et al.*, 2021).

Genetically recoded organism are also shown to be resistant to bacteriophages as shown by Robertson *et al.* in 2021 for Syn61 Δ3 (Robertson *et al.*, 2021). when a bacteriophage inserts its genome into recoded cells and the CDS of the phage proteins necessary for phage propagation contain the forbidden codon, translation stops at these codons and no functional phage can be assembled.

3.5 Biosecurity

Biosecurity considerations are a crucial part of responsible research, but most life-sciences scientists learn only about biosafety during their education. Language barriers and insufficient education result in the misconception that biosafety and biosecurity describe the same concept, but despite certain overlaps in risk mitigation, biosafety and biosecurity are not the same. Biosafety describes “The containment principles, technologies and practices that are implemented to prevent the unintentional exposure to pathogens and toxins, or their accidental release”, while biosecurity describes “institutional and personal security measures designed to prevent the loss, theft, misuse, diversion or intentional release of pathogens and toxins”(World Health Organization, 2006).

3.5.1 Regulation of Xenobiology research

Synthetic biology research risk is mainly assessed based on the organism and DNA involved in the research study. Organism are sorted into biosafety level groups, starting with Biosafety Level 1, posing no risk to humans and/or the environment, over two with moderate risk but well treatability, over three with high risk and moderate treatability over to the highest risk level biosafety level 4 which includes organism and viruses with a high risk of death for humans (European Union, 2000; Chosewood & Wilson, 2010). Research Laboratories are also classified as BSL1-4, and work with a certain organism requires the lab to have at least the corresponding biosafety level. Organism that pose significantly high risks are also very strictly related by export control to prevent their spread to individuals that might misuse these (Roberts, 1998).

When synthetic DNA is ordered to be introduced into other organism, it is screened in most cases for known toxins encoding CDS and origin in BSL3 or 4 organism (Elworth *et al.*, 2020). In case DNA gets flagged as toxic, it can only be ordered if the research institution shows the

qualification to work with this kind of DNA (International Gene Synthesis Consortium — The Promotion of Biosecurity, 2017).

Most countries and research institutes require scientists to be specially trained for their laboratory safety category to ensure unintentional accidents. Work in BSL4 level laboratories most of the time also requires background checks for employees (Skvorc & Wilson, 2011).

On international level, bioweapon use is supposed to be prevented by the Biological Weapons Convention (BWC), an international treaty opened for signature on 10 April 1972 and entered into force on 26 March 1975 (UNODA, 2023). To date 183 states have signed the treaty and thus declared to not use or develop biological weapons (Alexander, 2023). The BWC is more aimed at state party actors and does specifically state that all regulations do not include research. Thus individual scientists and research institutes are not regulated through this treaty(UNODA, 2023).

In contrast to the BWC, United Nations Security Resolutions 1540, shall “refrain from providing any form of support to non-State actors that attempt to develop, acquire, manufacture, possess, transport, transfer or use nuclear, chemical or biological weapons and their means of delivery, in particular for terrorist purposes” (United Nations Office for Disarmament Affairs, 2004). This way all members of the United Nations are obligated to ensure to prevent individuals from acquiring or developing biological weapons. How to achieve this aim is determined by the individual countries.

Countries take their responsibility to prevent the misuse of biomedical research differently serious. Some countries have very strong biosafety and biosecurity regulations, at least when it comes to biosafety level 3 or 4, e.g. the Netherlands who even have their own Biosecurity initiative from the Government or the US who just last year started with a widely funded biosecurity initiative (Rijksinstituut voor Volksgezondheid en Milieu, 2023; The White House, 2022).

All of these regulations, if in place, focus on regulating pathogen access and pathogen related dual-use research. Which is research focusing on BSL3 and 4 organism.

In most, if not all cases in the past, xenobiology research focused only on working with BSL1 organism and DNA. Because of this, regulation that is specific for BSL 3 and 4 labs, as well as pathogen related regulations like export control, or additional training of employees did so far not apply to this research.

This means that the researchers alone were responsible to access and mitigate biosafety risks in their own research. The only additional layer of safety is that some bigger national and international research funding agencies require a dual-use risk assessment in the grant application process (National Science Foundation, 2023; Deutsche Forschungsgemeinschaft, 2023). If the

research is not funded by those, the research team has to show self-initiative in biosecurity risk prevention. This risk mitigation approach is called the self-regulatory approach.

To make the self-regulatory approach in biosecurity work, widespread awareness and education about biosecurity is necessary. Scientists need to be aware of their responsibility to assess biosecurity risk and also educated to assess and mitigate those risks.

An international survey from 2018 showed that scientists in the synthetic biology community are not aware what biosecurity risks like dual use are (Vinke *et al.*, 2022). They also state that the education at their universities was poor regarding those issues and that they would like to learn more about this topic. The vast majority stated that they see a risk for misuse in their research discipline.

4 Results

4.1 Nanopore sequencing-guided directed evolution of aminoacyl-tRNA synthetases for the incorporation of non-standard amino acids into proteins

4.1.1 Abstract

Expanding protein design possibilities with non-standard amino acids (nsAAs) is of great interest to the pharmaceutical and synthetic biology industries. Peptides and proteins with nsAAs can be used to produce complex proteins with post-translational modifications in *E. coli* and introduce new-to-nature functions into proteins. To harness the full power of this new tool, precise incorporation of a certain nsAA needs to be achieved *in vivo* and the production must result in a homogeneous protein product. This can only be achieved by evolving highly active and specific aminoacyl/tRNA-synthetases (aaRS). Current directed evolution methods show severe limitations in library design and continuous evolution methods lack suitable sequencing evaluation to clearly interpret which mutations enrich. This work introduces how to use long-read Nanopore sequencing during a non-continuous phage-assisted evolution experiment to guide the choice of mutants, analyze which mutations never enrich and analyze the protein landscape with the final goal to identify the best candidate aaRS. With the help of this new technique, combined with new sequencing scripts to distinguish sequencing errors from mutations, it is shown here that full evolutionary trees can be created based on the sequencing data. The analysis showed that mutations occurring in the directed evolution of a L-2-nitro-phenylalanine aaRS are reproducible over independent evolutions and that mutations of amino acid residues not directly involved in ligand binding have a strong influence on aaRS activity and specificity. With this newly developed long-read deep sequencing approach, the understanding of the entire protein landscape is possible, and the method can be easily adapted to a wide range of other proteins.

4.1.2 Introduction

Protein design is naturally limited to the 20 canonical amino acids. These only offer a small variety of chemical properties and sizes. To overcome this limitations, non-standard amino acids (nsAAs) can be incorporated to expand the possibilities in protein design. This has been shown for over 150 amino acids, incorporating functions like photoreactivity, new catalytic abilities, fluorescence, click-chemistry and mimic post-translational modifications, e.g. the incorporation of the photo-cleavable nsAA L-2-nitro-phenylalanine (Liu & Schultz, 2010; Murakami *et al.*,

2000; Fekner *et al.*, 2009; Rogerson *et al.*, 2015; Peters *et al.*, 2009; Ye *et al.*, 2009; Taki *et al.*, 2002; Cellitti *et al.*, 2008; Fekner *et al.*, 2009; Hagiwara *et al.*, 2008; Bae *et al.*, 2003).

The incorporation of nsAAs at a certain position in a target protein can be achieved by repurposing the *amber* stop codon, the least used stop codon in *E. coli* (Bossi, 1983). To incorporate a nsAA in response to the *amber* stop codon an orthogonal *amber* suppressor tRNA and an orthogonal aminoacyl-tRNA synthetase (aaRS), which is capable to charge the nsAA to the *amber* suppressor tRNA need to be introduced. Orthogonal aaRS are often derived from archaea, like the most commonly used scaffold enzyme, the tyrosyl aaRS from *Methanocaldococcus jannaschii* (Liu & Schultz, 2010). This aaRS and its matching tRNA are already orthogonal in *E. coli* and accept an anticodon change to the *amber* codon anticodon, but need to be evolved to accept nsAAs as ligand instead of tyrosine (Anderson & Schultz, 2003; Kobayashi *et al.*, 2003). Several directed evolution methods for aaRS have been published most relying on a two round selection system where a library of the aaRS is designed and then selected for activity in a positive selection and for specificity in a negative selection (Liu *et al.*, 1997; Santoro *et al.*, 2002). The limitation of this selection is mainly the restriction in library design, since the maximum number of clones is limited by the transformation efficiency and screening capacity of the lab. Due to this limitation, rarely more than 10 amino acid residues are mutated, mostly focusing on amino acid residues directly involved in ligand binding.

An alternative to this selection is the use of phage-assisted continuous evolution (PACE) for aaRS evolution as published by Bryson *et al.* in 2017 (Bryson *et al.*, 2017). This approach still relies on a similar two round selection pressure but enables ongoing mutation and selection of the whole protein CDS, overcoming the limitations in library design posed by the classical directed evolution method (Esveld *et al.*, 2011; Badran & Liu, 2015; Bryson *et al.*, 2017).

However, PACE systems are complex to set up and so far evaluation of the evolution was performed using Sanger or Illumina sequencing of single clones, missing the chance to gain insights about the whole protein landscaped by analyzing which mutations lead to an enrichment of the phage and which mutations destroy the protein function (Esveld *et al.*, 2011; Bryson *et al.*, 2017).

To add to the advantages of the PACE system, the goal of this work was to design an easier non-continuous phage-assisted directed evolution system for aaRS evolution and develop a deep-sequencing guided evaluation of continuous directed evolution experiments.

The evaluation of PACE was so far performed using Sanger sequencing of individual clones that survived the selection. This approach is highly limited by the high throughput processing capabilities of the lab leading to a limitation in the number of clones that can be evaluated. This could be overcome using Illumina sequencing, which due to the maximum read length of 2x300

nt, does not allow to detect the co-occurrence of mutations further apart (Profaizer *et al.*, 2015; Tan *et al.*, 2019). The alternative Nanopore sequencing enables long-read sequencing but has a high error-rate compared to Sanger and Illumina sequencing, so that in sequencing on single read basis, mutations and sequencing errors are difficult to distinguished (Wang *et al.*, 2021).

4.1.3 Materials and Methods

Plasmid cloning

The test plasmid for aaRS activity was cloned on the basis of the aaRS expression plasmid published by iGEM Bielefeld-CeBiTec 2017, BBa_K2201202. The template for the 2NPA WT enzyme was amplified with the oligonucleotides 2NPARS_scaffold_rev (cgttgaaactgaagtacaggcgtt-gcggattgg) and 2NPARS_glnS_F (gtttacgctttaggaaatccatatggacgag) from the plasmid BBa_K220-1200 published by iGEM Bielefeld CeBiTec 2017, and cloned into the expression plasmid which was amplified with Scaffold_NPARS_fwd (ccgcaaacgcctgtacttcagttcaaacgctaaattgc-ctg) and glnS_NPARS_R (ctcgccatatggattcccaaagcg). All PCRs were performed using NEB Q5 HotStart Mastermix according to the manufacturer's protocol with an annealing temperature of 60 °C. The fragments were PCR purified using the NEB Monarch PCR purification kit and assembled using the NEB 2x Gibson Mastermix according to the manufacturer's protocol. A reporter for the aaRS strength was added to the same plasmid by introducing mRFP with an *amber* codon at position 132. The backbone with the oligonucleotides glnS_mRFP_fwd (CAT-CATCATCATCATtaagattatcaattaaaaactaacatggcgttcgcgttc) and Prefix_rev (ctctagaagcg-gccgcgaattc). The mRFP was amplified from the standard iGEM part E1010 with the oligonucleotides mRFP_glnS_rev (ATGATGATGATGATGAGcaccgggtgg) and Prefix_fwd (gaattcgcg-gccgctttagag). Both fragments were PCR purified using the NEB Monarch PCR purification kit and assembled using the NEB 2x Gibson MM according to the manufacturers protocol. The *amber* codon was introduced by site directed mutagenesis using the oligos mRFP_Asp132TAG_F (ggtccggatgcagaaaaaccatg) and mRFP_Asp132TAG_R (accatggttttctgcata-acccgaccgtcgacggaaagttgtaccac).

Phage cloning

The coding DNA sequence (CDS) of the wildtype 2NPARS was amplified using the oligonucleotides 2NPARS_GOI_F (aggctccttggagcccttttatggacgagttgagatgattaagcg) and 2NPARS_GOI_R (aaaaggcgacattcaaccgattggagggagttacgttggaaactgaagtacaggc) and the BioBrick BBa_K2201- 200 as template. The phage backbone was amplified using the oligonucleotides M13_GOI_F (ctccctcaatcggtgaatgtcgccctttgtctt) and M13_GOI_R (aaaaaaaaaggctccaaaggagcccttaattg-

tatcggttatac) using the wildtype M13 as template. The *amber* codon ending the translation of proteinIV was mutated to TAA via site directed mutagenesis using the oligos pIV_TAA_F (cgtaaaagcaaccatagtacgcgcctgtaacggcgcattaaggcgccgg) and pIV_TAA_R (tacaggcgcgactatggttgcggacg).

All PCR fragments were amplified using the NEB Q5 Hotstart Mastermix according to the manufacturer's protocol with an annealing temperature of 60 °C. All fragments were DpnI digested (NEB) overnight and purified using the NEB Monarch PCR purification kit according to the manufacturer's instructions. The fragments were assembled using the NEB Gibson MM according to the Manufacturer's protocol and transformed into electrocompetent S2060 cells with the pJC175e plasmid. The cells were regenerated for 45 min and plated undiluted and in 1:100 dilutions in standard phage plaque assays as described by (Miller *et al.*, 2020). Plaques were picked into PBS the following day and the colony PCR was performed with the oligonucleotides nano_fwd (gcctcagegaccgaatatatc) and nano_rev (cctgaacaaagttagaggtaattg) using NEB Q5 Hotstart Mastermix according to the manufacturer's protocol. Positive clones were enriched in S2060 pJC175e cultures, precipitated using PEG8000/NaCl, resuspended in PBS and stored at 4°C until the start of the evolution. The whole phage was sequenced using Nanopore sequencing.

Phage-assisted non-continuous evolution

For Phage-assisted non-continuous evolution (PANCE), the host cells have been prepared as follows: S2060 chemocompetent cells have been transformed with either pDB026e for the positive selection host cells or pDB036d for the negative selection host cells and plated on 2xYT plates containing ampicillin (100 µg/mL) and tetracycline (15 µg/mL). Individual clones from the plates were picked into a 3 mL 2xYT culture with ampicillin(100 µg/mL) and tetracycline (15 µg/mL), which was then used to prepare chemocompetent cells in the presence of 1 % glucose. The competent cells were transformed with MP4 and plated on 2xYT plates with ampicillin (100 µg/mL), tetracycline (15 µg/mL) and chloramphenicol (20 µg/mL). Single colonies were picked and used to prepare glycerol stocks that were then used as starter for each selection round.

Each selection round was performed as follows: A 50 mL 2xYT culture with ampicillin (100 µg/mL), tetracycline (15 µg/mL) and chloramphenicol (20 µg/mL) was grown at 37 °C until the OD₆₀₀ reached approx. 0.3. Then the culture was induced with 0.5 % arabinose and grown for another 30 min. Afterwards the culture was infected with the selection phage. In the beginning of each evolution, the phage was added to the culture at a MOI of 1. For the following selection rounds, 1 mL of the previous culture's supernatant was used for infection. The cultures were

cultivated for 12 h, then spun down at 5000 g for 30 min and the supernatant was filtered through 0.22 µm filters and stored at 4 °C until further use.

Nanopore sample preparation and sequencing

For the amplicon-based sequencing approach, the phages were amplified from 1 µL supernatant sample with 35 rounds of PCR using the Q5 High-Fidelity 2X Master Mix (NEB) to linearize the region of interest according to the manufacturers protocol. The primers for amplifications were nano_fwd (gcctcagegaccgaatatac) and nano_rev (cctgaacaaagttagggtaattg), the annealing temperature was set to 60 °C, and the elongation time to 1 min. The sequencing preparation was performed using the Ligation Sequencing with native barcodes (SQK-LSK109 with SQK-NBD112.24) from Oxford Nanopore Technology according to the manufacturer's protocol. Afterwards, the sample pool was sequenced on a GridION with a R9.4 flowcell (Oxford Nanopore Technology).

Nanopore evaluation and tree generation

For sequence trimming (amplicon_trimming_phages.py) and indel error correction (Indel_correction_phage.py), two python programs were written which are attached in the appendix. For the comparison of the sequences (script_phages.pl) a PERL script was written which is uploaded at github/attached in the appendix. For the evaluation of the amino acid changes (calc_pos.pl) a further PERL script was written which is attached in the appendix.

For the construction of the trees, a neighborhood joining tree was generated with the tool Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). This phylogenetic tree was then visualized using the interactive Tree Of Life tool (iTOL, <https://itol.embl.de/>).

aaRS activity test

The 2NPARS mutants identified in the bioinformatic analysis that were supposed to be tested for their activity were generated using site directed mutagenesis of the aaRS activity test plasmid. Therefore, forward oligonucleotides were designed approx. 30 bp downstream of the codon to be mutated, and reverse oligonucleotides upstream of the codon, with the new codon and the reverse sequence of the forward oligonucleotide as tail. All PCR fragments were amplified using the NEB Q5 HotStart Mastermix according to the manufacturer's protocol with an annealing temperature of 60 °C. All fragments were DpnI digested overnight and purified using the NEB Monarch PCR purification kit according to the manufacturer's instructions. The fragments were assembled using the NEB Gibson MM according to the Manufacturer's protocol

and transformed into BL21 chemocompetent cells. After regeneration, the cells were plated on 2xYT plates with kanamycin (25 µg/mL). Single colonies were picked, plasmids were isolated and sequenced using Nanopore Sequencing.

The clones with the correct sequence were used to start a 2xYT starter with kanamycin (25 µg/mL). After overnight incubation the starter was used to inoculate six 2xYT 3 mL cultures with kanamycin (25 µg/mL) and an OD₆₀₀ of 0.1. The cultures were cultivated for 3 h and then induced with 0.5 % arabinose. The cultures were cultivated for 48 h, then spun down at 10 000g for 2 min, washed with PBS, resuspended in 1 mL PBS and 150 µL were transferred to a 96 well plate with clear bottom for fluorescence measurement. All samples were measured using the infinite M200 plate reader by Tecan using an excitation wavelength of 580 nm and detect the emission at 610 nm.

MS evaluation

The incorporation of 2NPA for the codon TAG in mRFP was analyzed by Orbitrap LC-MS/MS from a whole proteome sample. Cell pellets (1 mL) from Gly38Ser mutant cultivation experiments were resuspended in 100 µL of 100 mM ammonium bicarbonate (Honeywell). Afterwards, 100 µL trifluoroethanol (TFE, Sigma Aldrich) and 5 µL 200 mM dithiothreitol (DTT, Thermo Fisher Scientific) were added to the cells. The samples were mixed and incubated for 60 min at 60 °C. 20 µL of 200 mM iodoacetamide (IAA, Sigma Aldrich) was added to the samples which were then mixed and incubated at room temperature for 90 min in the dark. After incubation, 5 µL of 200 mM DTT was added, and the samples were mixed and left for 60 min at room temperature. After protein isolation, the proteins were digested using the enzyme trypsin. 435 µL of MS-grade water (TH. Geyer) and 435 µL of 100 mM ammonium bicarbonate were added to the suspensions. Subsequently, 10 µL of MS-grade Trypsin Gold (1 µg µL⁻¹, Promega) was added and the samples were incubated overnight at 37 °C. The next day, the digested protein solutions were purified using Sep-Pak C18 columns according to the manufacturer's protocol (Sep-Pak Vac 1cc Cartridge, Waters). For LC-MS/MS measurements, the purified peptide mixtures were diluted to a final concentration of 1 µg µL⁻¹. Peptide identification was carried out with a QExactive Plus Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific) coupled to an Ultimate 3000 RSLC nano system (Thermo Fisher Scientific). 1 µL of the samples were injected to the system and separated by an Acclaim PepMap 100 C18 column (3 µm particle size, 75 µm diameter, Thermo Fisher Scientific). The flow rate was set at 0.3 µL min⁻¹ and a 60 min gradient from 5% to 40% acetonitrile was applied. Afterwards, the MS/MS data were evaluated to identify the peptide with the incorporated 2NPA using Proteome Discoverer 4.2 (Thermo Fisher Scientific). The data were searched against an

E. coli K12 protein database including the newly formed protein sequence using the Sequest HT algorithm. The incorporation of 2NPA (+77.02655 Da) was added as a new modification to the analysis software.

4.1.4 Results

To facilitate the detection of mutations that are not in the ligand binding pocket a L-2-nitrophenyl-alanine aaRS developed by Peters *et al.* 2009 was used as starting point for the evolution (Peters *et al.*, 2009). This enzyme has 10 mutation when compared to the wildtype *M. jannaschii* tyrosyl aaRS. All of these mutated amino acid residues were originally in contact with the ligand or in contact the the amino acid in direct contact with the ligand (Fig. 6). The CDS was cloned into the selection phage and evolution was performed in three independent replicates in 50 mL shake flask batch cultivations. All evolutions were performed in 7 rounds of positive selection with 6 rounds of negative selection in between, each for 12 h. During evolution the phage titer was monitored and found to decrease during the course of evolution. After the evolution, each culture was spun down and the evolved CDS was amplified using the supernatant directly as template. The PCR amplicons were sequenced using ligation-based Nanopore sequencing (LSK109).

For the analysis of libraries based on Nanopore sequencing data the usage of a consensus sequences is not possible because this would filter out the mutations that occur on the single read basis. To detect the mutations of single clones, single reads need to be evaluated separately and be at a quality that enables to distinguish between mutations and sequencing errors.

For the analysis strategy, the region of interest (ROI) was localized, and the ends of the sequences were trimmed so that the sequences within the library had the same start and ending points (amplicon_trimming_phages.py). The reads were than mapped using minimap2 against the origin sequence for the library, to create CIGAR strings for the indel correction. Indel correction was performed to minimize differences due to a sequencing error based frame shifts and to obtain reads with the same length to analyze the number of identical reads with mutations. Therefore, reads were compared with the cigar string and for each deletion the nucleotide of the original sequence was inserted and for each insertion the inserted nucleotide was deleted. With these modifications, only sequences with the same length, same start and end region and sequences without indel errors remain (indel_correction_phage.py). For this purpose, the DNA sequence is translated into the protein sequence. These protein sequences were then compared to the protein sequence of the original sequence to sort out sequences with a high number of mutations that may result from low quality reads. Sequences were then compared with each other,

Nanopore sequencing-guided directed evolution of aminoacyl-tRNA synthetases for the incorporation of non-standard amino acids into proteins

and the number of identical sequences were added into the sequence header so that sequences can be sorted by occurrence (script_phages.pl).

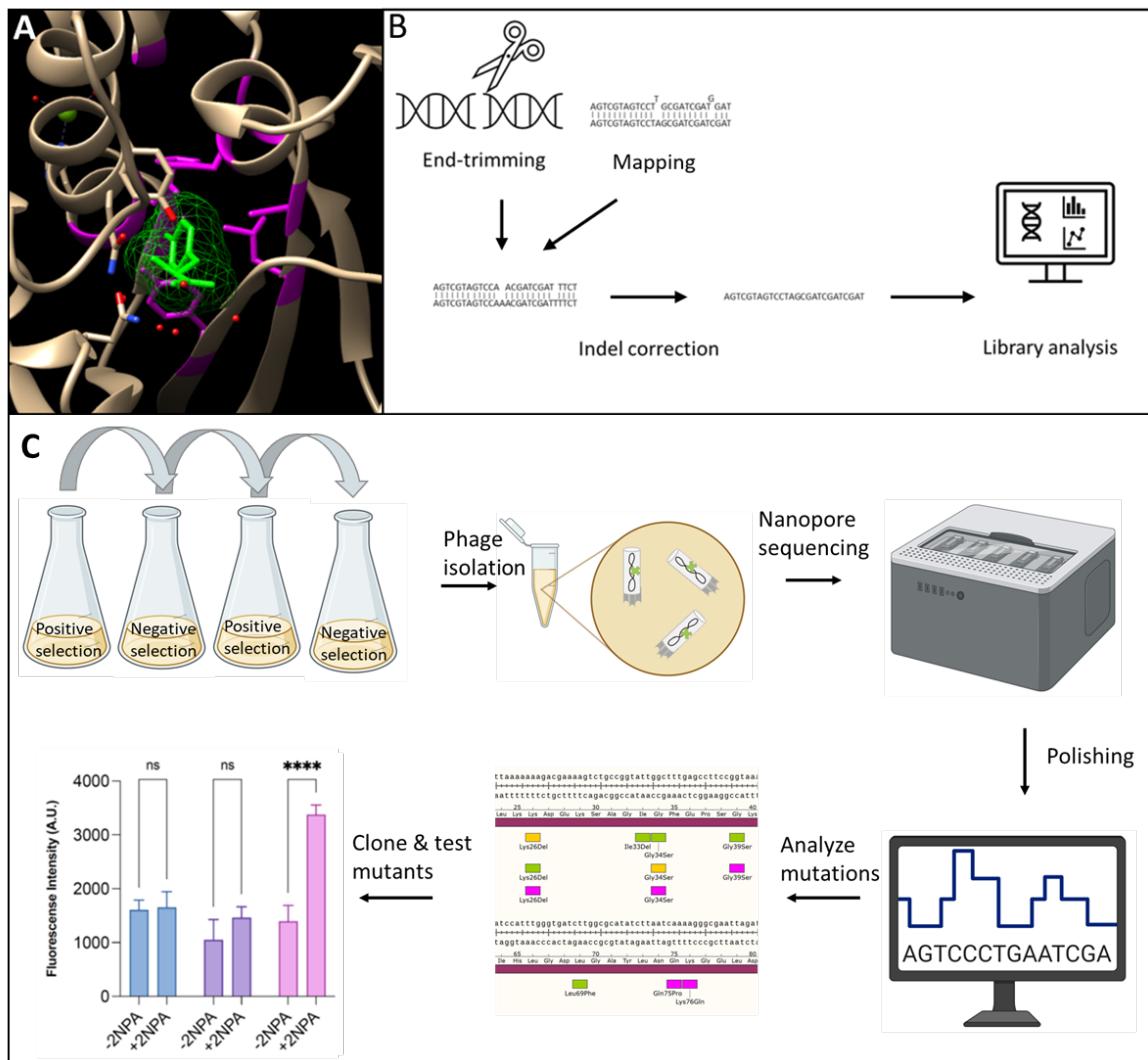


Figure 6: A Crystal structure of the wildtype *M. jannaschii* TyrRS. Amino acid residues mutated by Peters *et al.* 2009 have been colored magenta. B Workflow of the bioinformatic pipeline for the sequencing raw data processing. The sequences were first trimmed to the region of interest and mapped to the origin sequence. Afterwards the cigar string from the mapping was used for the indel correction of the trimmed sequences before the library was analyzed. C Schematic workflow of the PANCE experiment. Iterative rounds of batch cultivations of positive and negative selection were performed each infected with 1 mL supernatant from the previous culture. The aaRS CDS has been PCR amplified directly from the supernatant and the amplicons sequenced with Nanopore sequencing. After Sequencing raw data processing and bioinformatic analysis as described in B, variant have been identified that enrich throughout the evolution. These variants have been tested in fluorescence based activity measurements.

After raw data analysis, all sequences were sorted into five categories (supplementary data Tab. 18): wildtype, internal stop codons, sequences not passing the sequencing error filter, mutants passing all filters, and enriching mutants passing all filters.

A lot of sequences contained in frame stop codons. Of the total sequencing reads those have been between 11.29 % to 93 %, while showing no tendency of occurring only in certain evolutions or evolution rounds. Most likely these numbers are caused by very successful negative selections, where inactive variants with stop codons are likely to enrich. To ensure those stop codons are not resulting in truncated but more active variants, the variants were manually controlled for the localization of the internal stop. This revealed no tendency towards stop codons only occurring downstream of the active site. This indicates that those mutations are most likely no activity enhancing mutations, but lead to the inactivation of the enzyme. Of the total reads, there was a relatively low number of wildtype (between 0.72 % and 6.4 %), again showing no trend to occur in only certain evolutions or rounds. This number shows that the wildtype has likely some activity, and that the coexistence of different mutations with different activities in the 50 mL culture volume is possible. The sequencing error filter out 1.67-24.1 %, given the high amount of raw data reads, this still enables to deep sequence more than 20,000 reads in well concentrated samples. The number of total reads was in correlation with the phage titer and can be increased by preparing more amplicon of low titer samples.

Sequences that not fall under these three categories are mutants that can be evaluated. A particularly interesting part of this category are mutants with more than 3 occurrences, since those are highly likely enriched mutants. This enriched mutants category makes up between 0.79% and 9.21 % of the total reads, with no trend in occurring in specific evolutions or rounds. This category makes up between 6.6 % to 25.0 % of filter passing reads.

Enriching reads over 2 % of total reads in the enriched reads category have been further analyzed since those mutations seem to increase the replication rate of the phages most, indicating higher aaRS activity. This data was used to design evolutionary trees to compare the enriching mutations of all three evolutions (Fig. 7). Each data point represents a over 2 % enriched mutation of one of the evolution rounds of the three independent evolutions. Identical mutation have been colored in the same color. Mutation occurring only in one evolution round once have been left white. The tree shows that multiple rounds from the three independent different evolution enrich the same mutations, showing that a lot of enriching mutations are reproducible over independent directed evolution experiments. This indicates these mutations are not coincidental appearances that enrich because of another mutation in the same phage and most likely have an influence on activity of the aaRS.

Nanopore sequencing-guided directed evolution of aminoacyl-tRNA synthetases for the incorporation of non-standard amino acids into proteins

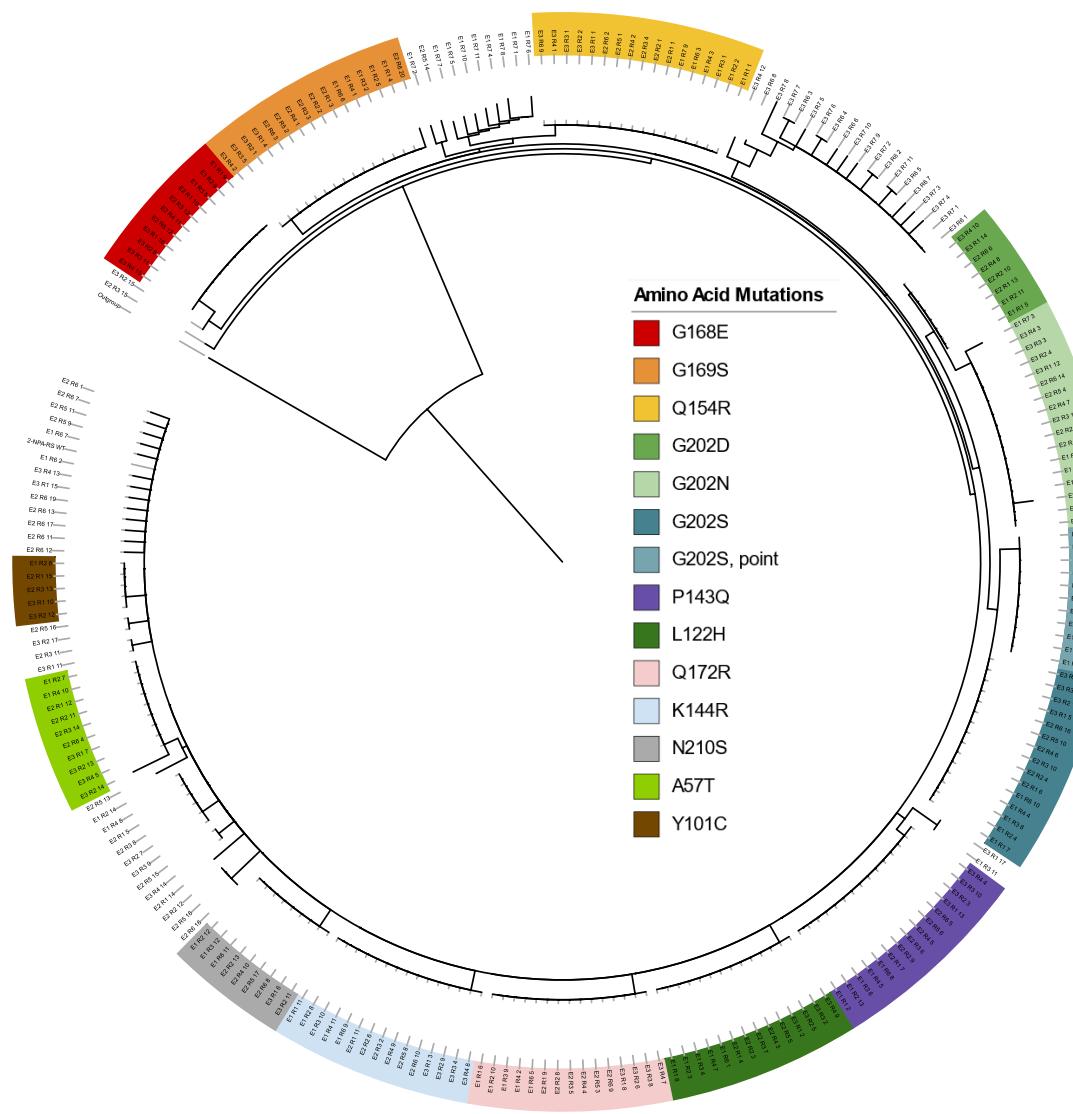


Figure 7: Evolutionary tree for all clones enriched above 2 % of the total sequencing reads in all positive selection rounds of all three of evolutions. Each data point represents a point mutation causing an amino acid exchange that enriched above 2 % of the total enriched reads. Identical mutations are highlighted in the same color.

Mutation occurring in a lot of evolution rounds are the residues G168E, often occurring in combination with G169S, Q154R and G202S (in two codon variations). This evaluation gives a good idea which residues do have an influence in the desired enzyme properties, while showing also which residues are highly likely conserved by not appearing in this tree.

The data also enables to analyze single evolution rounds more in depth. When looking at evolution 1 round 1, the evolutionary tree (Fig. 8) shows that some of the enriched candidates un-

derwent several rounds of CDS mutation and selection. In classical directed evolution systems this would have required a new library cloning step after each evolution round. This evaluation also enables the discovery of potential co-dependent mutations.

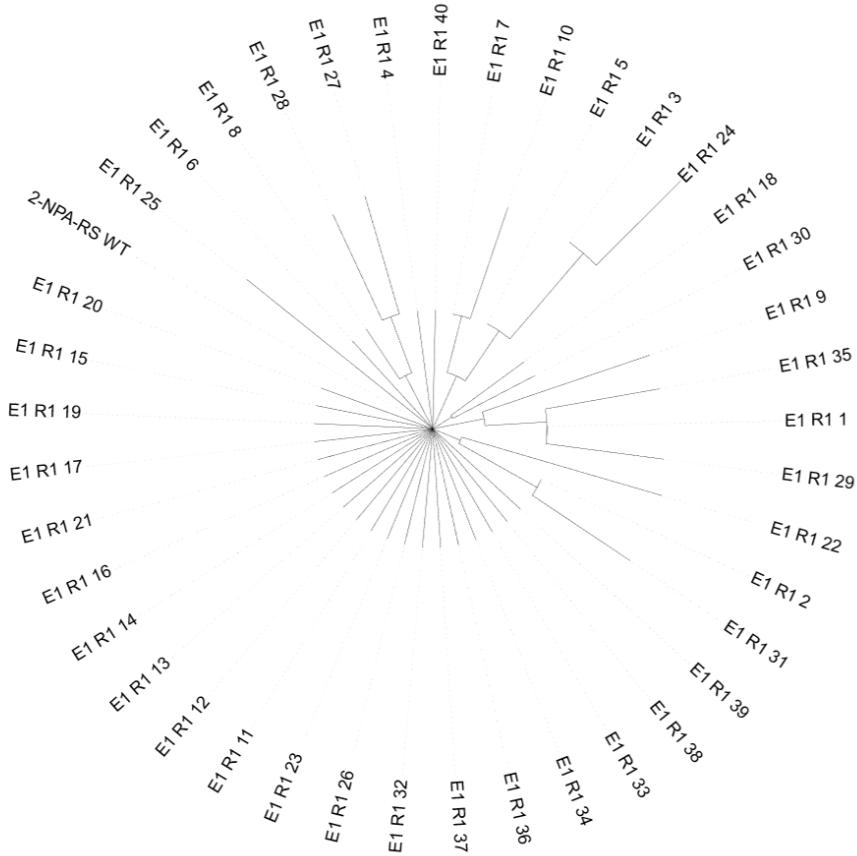


Figure 8: Evolutionary tree for all clones enriching above 2 percent of the total enriched sequencing reads in positive selection round one of evolution one. Each data point represents one mutant.

Another important information next to which mutants enrich is the ratio of the different enriched mutations, since higher enriched mutants are more likely to show higher activity. Mutations that enrich over 2 percent total reads in the enriched read category were analyzed regarding their ratio (Table 7), showing three different kinds of mutations. The first category are the mutants that are the same as the mutations that have already been displayed in the evolutionary tree of all three evolutions as reoccurring mutations (Fig. 9, highlighted in different colors). White spots in these mutations can be explained by not making the cutoff of 2 % of total enriched

clones, but they were still present in all evolutions. The second category are some mutations that occur in only a few evolutions but do not enrich further than the first category. These mutations are most likely silent mutations that enrich by chance. The third category are mutations that occur rarely but take over almost the whole evolution like S113P occurring only in evolution 3 or His44Tyr of evolution 1. These mutations might represent mutations towards local minima. One particularly interesting mutation in this category is Glu141TAG, that took over evolution 2. In this mutant 2NPA is incorporated into the aaRS. In theory not able to survive negative selection, since no incorporation of 2NPA is possible in this selection round, but this mutant seem to enrich so rapidly, that it can survive negative selection by being diluted. In general lower phage titer in the last rounds of evolution seems to increase the likelihood for evolutions take over the whole selection.

Next to amino acid exchanges also two codon usage modifications that were found in almost all evolutions were observed. GTG167GTA and GGT199GGA, both neighboring residues or even being residues that seem to enrich rapidly throughout all evolutions. It was also observed that almost all amino acid exchanges resulted from A to G mutations, showing a clear prevalence for this mutation caused by the choice of mutation system.

Nanopore sequencing-guided directed evolution of aminoacyl-tRNA synthetases for the incorporation of non-standard amino acids into proteins

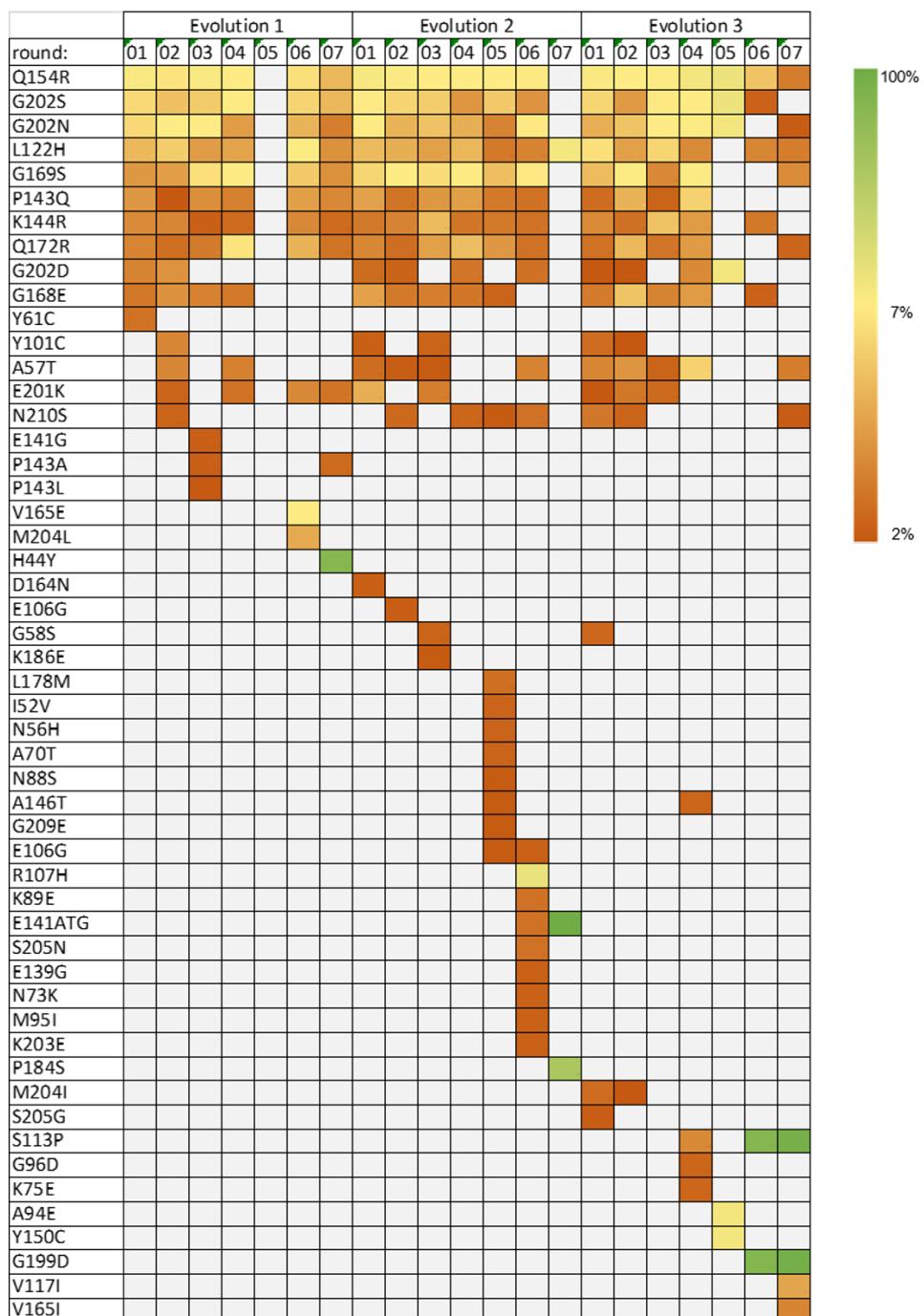


Figure 9: Percent shares of all amino acid mutations that enrich over two percent of the enriched reads category of the total reads in each positive selection of the three independent evolution experiments.

To get an idea what mutations actually refer the desired properties, different mutants were cloned and tested regarding their activity and specificity. The assay chosen was an mRFP fluorescence measurements using whole cells. This measurement, which already poses a relatively high background allows the activity determination of highly active aaRS. The wildtype showed no signal above only background, proving its low activity. Three experiments measuring for the influence on aaRS activity and specificity were performed. The first analysis had the aim to determine the influence of single mutations that were identified as interesting candidates from the evolutionary tree and ratio of the evolution (Fig. 10A). For the second analysis, random combinations of mutations that occur together and/or showed higher activity during the single mutations activity assay were cloned and tested (Fig. 10B) and in the third analysis the codon usage altering mutations were analyzed (Fig. 10D).

The first measurement showed that the influence of most mutations on the activity is too low to actually lift them above our assays minimum required activity to distinguish activity from background noise, but we did identify three single point mutations that improve activity significantly, two of them being the mutations that only occurred in one evolution, but fast taking over almost the whole evolution: His44Tyr from evolution 1 and Ser113Pro from evolution 3. The *amber* stop codon mutation that took over almost all of evolution 2 did not show any activity boost as the other two, but this one also occurred almost exclusively in combination with other mutations, indicating a dependency on other mutations to increase activity, or a coincidental mutation that occurred with the other mutations and just prevailed because of the activity increase of those other mutations.

To further increase activity, mutations have also been tested in random combinations leading to the identification of a 4 amino acid exchange activity mutant (Gly34Ser, Glu141TAG, Gly168Glu and Gly169Ser) showing even higher activity as the single mutations. Further introducing mutations into this mutant led to an activity decrease.

To prove the incorporated amino acid is indeed 2NPA, MS-Orbitrap analysis has been performed. To make the most out of the high sensitivity of this machine, whole proteome analysis was performed instead of analysis of isolated protein. A sample from the cultivation of the Gly38Ser mutant was used for whole proteome isolation and trypsin digest, followed by LC-Orbitrap MS-measurement. By introducing 2NPA as post-translational modification in the Proteome Discoverer workflow, the incorporation of 2NPA can be analyzed in an automated manner. This workflow can also be used to analyze 2NPA mis-incorporation in other proteins of the proteome, which was not observed for this sample. 2NPA incorporation was shown in several fragments, proving the incorporation at the respective position in mRFP (Fig. 9C).

Nanopore sequencing-guided directed evolution of aminoacyl-tRNA synthetases for the incorporation of non-standard amino acids into proteins

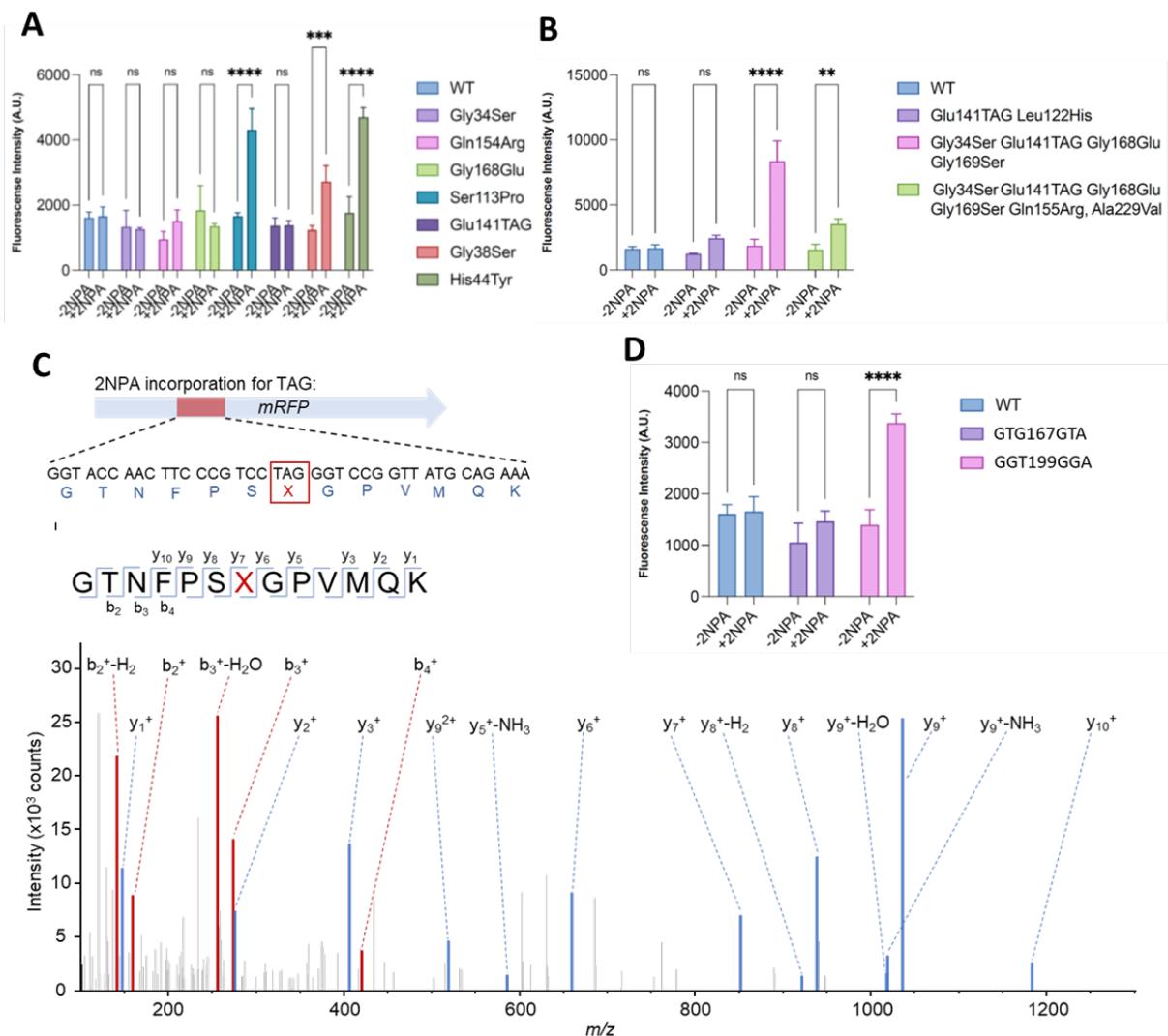


Figure 10: A Activity measure nt of single amino acid exchange mutants that were identified from the Nanopore sequencing data as enriched candidates. -nsAA samples contained no amino acid in the media, +nsAA samples were cultivate din the presence 1 mM 2NPA in the media. Bar represents the mean. Error bar shows standard deviation based on n = 3 independent experiments. B Activity measurements of randomly combined amino exchanged candidates with improved activity compared to the wildtype. -nsAA samples contained no amino acid in the media, +nsAA samples were cultivate din the presence 1 mM 2NPA in the media. Bar represents the mean. Error bar shows standard deviation based on n = 3 independent experiments C Orbitrap MS-spectra of one of the mRFP peptides containing 2NPA. D Activity measurement of mutants with altered codon usage. -nsAA samples contained no amino acid in the media, +nsAA samples were cultivate din the presence 1 mM 2NPA in the media. Bar represents the mean. Error bar shows standard deviation based on n = 3 independent experiments.

4.1.5 Discussion

PANCE combined with deep-sequencing has been shown to be a powerful tool for the evolution of proteins that can be evolved using the PACE system. However, aaRS selection in the PACE system is imperfect. In the positive selection variants that are unspecific and incorporate endogenous amino acids can enrich. In the negative selection non-functional variants can enrich, wasting a lot of host cells for variants with undesired properties. This system can only be overcome by the design of a one-round selection system for aaRS, which have not been designed for PACE of aaRS yet. This system could also solve the problem of the low phage titer observed during this evolution experiment.

The ongoing mutation during the PANCE run showed that previous library design methods have missed out on a lot of mutations that are relevant for activity and specificity. This has also been shown by Bryson *et al.*, who generated a truncated aaRS variant of the *Methanosarcia mazei* pyrrolysine aaRS with drastically improved activity (Bryson *et al.*, 2017). Variants with stop codons throughout the CDS resulting in truncated variants as well as randomized amino acid residues outside the binding pocket are very rare in aaRS library design. Thus, PACE offers a substantial improvement, not only in regards to finding more active variants but also broaden the understanding which amino acid residues are directly involved in activity and specificity.

To get this big picture of all amino acid residues influence in the protein sequence, deep sequencing of the whole directed evolution experiment is crucial. So far evaluation of PACE experiments mainly focused on sequencing individual clones of phages after selection rounds, mainly by Sanger sequencing. This method often limits to sequence only hundreds of phages, if even that much. This in combination with the only short read length and thus inability to analyze co-dependent mutations, and not repeating evolutions and comparing which mutations occur in all evolutions, results in only analyzing a fraction of the whole evolution landscape.

These limitations can be overcome by using a long-read sequencing method that sequences single reads. Nanopore sequencing offers the possibility to overcome these limitations, but has a highly increased error-rate compared to Sanger and Illumina. To overcome these limitations, either technical improvements in flowcell design are necessary, which Oxford Nanopore Technology is working on, showing an increase in accuracy of their R7 flowcell from 87% of the R9 flowcell to 97 % of the 9.4.1 flowcell (Minei *et al.* 2018; Sanderson *et al.* 2023). This trend in better accuracy most likely will continue to a level where single read sequencing is accurate enough to be confident to call only mutations and not mistake sequencing errors for mutations. Until then, improved evaluation software is the only solution to be able to distinguish errors from mutations. The advantage of the sample being a CDS where we only expect amino acid exchanges and no indels, since those would most likely destroy the function of the whole protein,

enables to use algorithms that analyze only point mutations that cause amino acid exchanges. This method works well for enriched variants with point mutations, but would not be able to detect true indels or frameshifts that lead to truncated or elongated variants. For the *M. jannaschii* aaRS truncated variants have not been reported to show improved activity. However, this has been reported for pyrrolysine aaRS (Bryson *et al.*, 2017).

To improve the evaluation, sequencing of the whole phage could be an options, if the phage is in a suitable concentration, which in this experiment would have only been possible by cultivating large amounts. However a phage with a higher titer is a good candidate for this approach, since it completely gets rid of the PCR bias. Another alternative sample preparation is isolation of the phage DNA followed by rolling circle amplification with one specific oligonucleotide, to generate repeats of the CDS that than could be used to assemble a consensus sequence from each individual read and thus correct the sequencing errors through sequence coverage.

In Nanopore sequencing, indel errors occur mainly in homopolymeric segments and lead to frame shifts, so indel errors must be corrected for single mutations analysis. Frame shifts would alter the entire amino acid sequence, resulting in too many changes for the analysis of a library (Goodwin *et al.*, 2016; Claverie, 1993). To avoid this, an DNA sequence can be analyzed using comparison with the original sequence and subsequent correction of indel errors using cigar strings (Li *et al.*, 2009). With our analysis strategy, indel error correction that can overcome the frame shift problem was achieved. Further improvement in sequencing accuracy may also lead to longer analyzable regions for DNA libraries.

However, the current evaluation method enable to evaluate the whole sequence space of the aaRS and is thus a dataset that can be used to study darwinian evolution on a one CDS level. The data can be mined to check which mutations occur only once, thus most likely destroy the protein function; occur rarely, likely showing no improvement in the evolution; and which mutations improve replication rate of the phage greatly by improving the aaRS activity or specificity greatly. Mutations can be analyzed that only enrich in the negative selection, these most likely destroy the protein function or only enrich in the positive selection, indicting the incorporation of endogenous amino acids.

The analyzed mutation showed a big influence of amino acid residues on activity and specificity which lay outside the amino acid binding pocket. Some of those could be explained by the wildtype crystal structure to be involved in tRNA binding. However, some of the mutations show no obvious structure function relationship when compared to the wildtype crystal structure and would be an interesting candidate for protein crystallography to analyze their influence on the protein structure.

Nanopore sequencing-guided directed evolution of aminoacyl-tRNA synthetases for the incorporation of non-standard amino acids into proteins

The indirect analysis of aaRS activity based on fluorescence followed by MS analysis is a fairly standard method in aaRS activity and specificity determination and showed to be sufficient for this kind of experiment. However, the MS analysis via Orbitrap that was also able to detect nsAA incorporation on proteome level is a very valuable tool for nsAA incorporation analysis, since it enables the analysis of cross-talk between the cell machinery and the aaRS and also incorporation sites at undesired sites in the proteome.

Lastly, the improved incorporation of 2-NPA was achieved by an improved directed evolution system that can be applied to all proteins that are evolvable in continuous directed evolution settings. The potential of 2-NPA in protein design has so far not been harnessed. It could be a very valuable tool in designing light induced elution methods in protein purification and various activation and de-activation systems for a wide range of proteins including antibodies.

4.2 Swapped genetic code blocks viral infections and gene transfer

4.2.1 Abstract

Engineering the genetic code of an organism has been proposed to provide a firewall from natural ecosystems by preventing viral infections and gene transfer (Church, G. M. & Regis, E., 2014; Lajoie *et al.*, 2013; Ma & Isaacs, 2016; Robertson *et al.*, 2021; Ostrov *et al.*, 2016; Fujino *et al.*, 2020). However, numerous viruses and mobile genetic elements encode parts of the translational apparatus (Abrahão *et al.*, 2018; Morgado & Vicente, 2019; Al-Shayeb *et al.*, 2020), potentially rendering a genetic-code-based firewall ineffective. Here we show that such mobile transfer RNAs (tRNAs) enable gene transfer and allow viral replication in *Escherichia coli* despite the genome-wide removal of 3 of the 64 codons and the previously essential cognate tRNA and release factor genes. We then establish a genetic firewall by discovering viral tRNAs that provide exceptionally efficient codon reassignment allowing us to develop cells bearing an amino acid-swapped genetic code that reassigns two of the six serine codons to leucine during translation. This amino acid-swapped genetic code renders cells resistant to viral infections by mistranslating viral proteomes and prevents the escape of synthetic genetic information by engineered reliance on serine codons to produce leucine-requiring proteins. As these cells may have a selective advantage over wild organisms due to virus resistance, we also repurpose a third codon to biocontain this virus-resistant host through dependence on an amino acid not found in nature(Mandell *et al.*, 2015). Our results may provide the basis for a general strategy to make any organism safely resistant to all natural viruses and prevent genetic information flow into and out of genetically modified organisms.

4.2.2 Introduction

The genetic code allows organisms to exchange functions through horizontal gene transfer (HGT) and enables recombinant gene expression in heterologous hosts. However, the shared language of the same code permits the undesired spread of antibiotic-, herbicide- and pesticide-resistance genes and allows viruses to cause diseases. By exploiting the shared nature of the genetic code, recombinant DNA technologies revolutionized our ability to produce small molecules, peptides, biologics and enzymes in vast quantities; however, the production cell cultures remained susceptible to viral contamination. Viral contamination in cell cultures remains a real risk with severe consequences: over the past four decades, dozens of viral contamination cases were documented in industry (Zou *et al.*, 2022; Barone *et al.*, 2020; Baltz, 2018). HGT also threatens the safe use of genetically modified organisms (GMOs) by enabling the spread of their engineered genetic information into natural ecosystems. Despite the impact of viral infec-

tions and HGT and the growing economic and societal role of GMOs and recombinant DNA, so far, no technology exists that could prevent viral infections and the escape of engineered genetic information from genetically modified biological systems.

It is widely believed that genomically recoded organisms, whose genomes have been systematically redesigned to confer an alternative genetic code, would offer genetic isolation from natural ecosystems by obstructing the translation of horizontally transferred genetic material (Church, G. M. & Regis, E., 2014; Lajoie *et al.*, 2013; Ma & Isaacs, 2016; Robertson *et al.*, 2021; Ostrov *et al.*, 2020), including resistance to both viral infections and HGT. Indeed, the genome-wide removal of TAG stop codons and release factor 1 (RF1) from *E. coli*, which abolishes the ability of cells to terminate translation at TAG stop codons, provides substantial but not complete resistance to bacteriophages (Lajoie *et al.*, 2013; Ma & Isaacs, 2016). Most recently, a strain of *E. coli*, Syn61Δ3, was created with a synthetic recoded genome in which all annotated instances of two serine codons, TCG and TCA (together TCR), and the TAG stop codon were replaced with synonymous alternatives, and the corresponding serine tRNA genes (serU and serT) and RF1 (prfA) have been deleted (Robertson *et al.*, 2021; Fredens *et al.*, 2019). Although the compressed genetic code of Syn61Δ3 provided resistance to five viruses⁴, it could not prevent the escape of its engineered genetic information.

However, despite the potentially broad virus and gene transfer resistance of compressed genetic codes and widespread industrial applicability of these organisms, how natural genetic material could breach genetic-code-based resistance remained unanswered. Numerous viruses and mobile genetic elements encode parts of the translational apparatus, ranging from single tRNA genes and release factors up to lacking only ribosomal genes^{7,8,9}. These genes allow mobile genetic elements to reduce their dependency on host translational processes (Yang *et al.*, 2021; Peters *et al.*, 2022; Borges *et al.*, 2022; Abe *et al.*, 2011; Alamos *et al.*, 2018; Santamaría-Gómez *et al.*, 2021). Therefore, the selection pressure posed by the compressed genetic code of genomically recoded organisms might facilitate the rapid evolution of viruses and mobile genetic elements capable of crossing a genetic-code-based barrier. Here we show that horizontally transferred tRNA genes can readily substitute cellular tRNAs and thus abolish genetic-code-based resistance to viral infections and HGT. We provide an example of a virus-resistant, biocontained bacterial host that prevents both incoming and outgoing HGT, which paves the way towards engineering multi-virus-resistant cell lines from any organisms and the safe use of GMOs in natural environments by eliminating HGT.

4.2.3 Materials and Methods

Bacterial media and reagents

Lysogeny broth Lennox (LBL) was prepared by dissolving 10 g l⁻¹ tryptone, 5 g l⁻¹ yeast extract and 5 g l⁻¹ sodium chloride in deionized H₂O and sterilized by autoclaving. Super optimal broth (SOB) was prepared by dissolving 20 g l⁻¹ tryptone, 5 g l⁻¹ yeast extract, 0.5 g l⁻¹ sodium chloride, 2.4 g l⁻¹ magnesium sulfate and 0.186 g l⁻¹ potassium chloride in deionized H₂O and sterilized by autoclaving. 2×YT medium consisted of 16 g l⁻¹ casein digest peptone, 10 g l⁻¹ yeast extract, 5 g l⁻¹ sodium chloride. LBL and 2×YT agar plates were prepared by supplementing LBL medium or 2×YT with agar at 1.6% w/v before autoclaving. Top agar for agar overlay assays was prepared by supplementing LBL medium with agarose at 0.7% w/v before autoclaving. SM buffer (50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 8 mM MgSO₄, 0.01% gelatin) was used for storing and diluting bacteriophage stocks (Geno Technology). bipA was obtained from PepTech Corporation (USA).

Bacteriophage isolation

Bacteriophages were isolated from environmental samples from Massachusetts, USA, collected in the third quarter of 2021 (samples 1, 2 and 4–13) and first quarter of 2022 (samples 1–3; Extended Data Table 1a), and by using *E. coli* Syn61Δ3(ev5) (from the laboratory of Jason W. Chin (Addgene strain no. 174514)) as the host. For aqueous samples, including sewage, we directly used 50 ml filter-sterilized filtrates, whereas samples with mainly solid components, such as soil and animal faeces, were first resuspended to release phage particles and then sterilized by centrifugation and subsequent filtration. Environmental samples and the sterilized filtrates were stored at 4 °C in the dark. This protocol avoided the inactivation of chloroform-sensitive viruses. Sterilized samples were then mixed with exponentially growing cultures of Syn61Δ3(ev5) in SOB supplemented with 10 mM CaCl₂ and MgCl₂. Infected cultures were grown overnight at 37 °C aerobically and then filter sterilized by centrifugation at 4,000g for 15 min and filtered through a 0.45-µm PVDF Steriflip disposable vacuum filter unit (Millipore-Sigma). Next, 1 ml from each sterilized enriched culture was mixed with 10 ml exponentially growing Syn61Δ3 (optical density at 600 nm (OD_{600nm}) = 0.2), supplemented with 10 mM CaCl₂ and MgCl₂, and mixed with 10 ml 0.7% LBL top agar. Top agar suspensions were then poured on top of LBL agar plates in 145 × 20-mm Petri dishes (Greiner Bio-One). Petri dishes were incubated overnight at 37 °C and inspected for phage plaques the next day. Areas with visible lysis or plaques were excised, resuspended in SM buffer, and diluted to single plaques on top agar lawns containing 99% Syn61Δ3 and 1% MDS42 cells. We note that adding trace amounts of MDS42 cells increased the visibility of plaques, and clear plaques, indicating phage replication on the recoded host, could be easily picked. Dilutions and single-plaque isolations were repeated four times for each plaque to purify isogenic phages. Finally, high-titre stocks were prepared by mixing sterilized suspensions from single plaques with exponentially growing

MDS42 cells ($OD_{600\text{nm}} = 0.3$) in SOB supplemented with 10 mM CaCl₂ and MgCl₂. Phage-infected samples were grown at 37 °C until complete lysis (about 4 h) and then sterilized by filtration.

Bacteriophage culturing

Bacteriophage stocks were prepared by a modified liquid lysate Phage on Tap protocol in LBL medium (Bonilla *et al.*, 2016). High-titre lysates were prepared from single plaques by picking well-isolated phage plaques into SM buffer and then seeding 3–50 ml early-exponential-phase cultures of *E. coli* MDS42 cells with the resulting phage suspension in SOB supplemented with 10 mM CaCl₂ and MgCl₂. Phage-infected samples were grown at 37 °C until complete lysis and then sterilized by filtration. High-titre phage lysates were stored at 4 °C in the dark. Phages were archived as virocells and stored at -80 °C in the presence of 25% glycerol for long-term storage.

Phage replication assay

Phages containing genomic TCR-suppressor tRNASer(UGA) genes (based on Supplementary Data 1), corresponding to National Center for Biotechnology Information (NCBI) GenBank numbers MZ501046, MZ501058, MZ501065, MZ501066, MZ501067, MZ501074, MZ501075, MZ501089, MZ501096, MZ501098, MZ501105 and MZ501106 (ref. 24), were obtained from DSMZ (Germany). Exponential-phase cultures ($OD_{600\text{nm}} = 0.3$) of MDS42 and Syn61Δ3(ev5) were grown in SOB supplemented with 10 mM CaCl₂ and MgCl₂ at 37 °C. Cultures were infected with phage at a multiplicity of infection (MOI) of approximately 0.001. Simultaneously, the same amount of each phage was added to sterile SOB supplemented with 10 mM CaCl₂ and MgCl₂ to act as a cell-free control for input phage calculation. Replication assays with the T6 bacteriophage in Syn61Δ3 in Fig. 11d were carried out by infecting exponential-phase cultures at a MOI of 0.01 with T6 phage. The figure shows total T6 titre after 24 h of incubation. Infected cultures were grown at 37 °C with shaking at 250 r.p.m. After 24 h, cultures were transferred to 1-ml tubes and centrifuged at 19,000g to remove cells and cellular debris, and the clarified supernatant was serially diluted in SM buffer to enumerate output phage concentration. A 1.5 µl volume of the diluted supernatants was applied to LBL 0.7% top agar seeded with MDS42 cells and 10 mM CaCl₂ and MgCl₂ using a 96-fixed-pin multi-blot replicator (VP407, V&P Scientific). Following 18 h of incubation at 37 °C, plaques were counted, and the number of plaques was multiplied by the dilution to calculate the phage titre of the original sample.

Single-step phage growth curve

An exponential-phase culture ($OD_{600\text{nm}} = 0.3$) of Syn61Δ3 was grown in 50 ml SOB supplemented with 10 mM CaCl₂ and MgCl₂ at 37 °C with shaking at 250 r.p.m. Cultures were then

spun down and resuspended in 3 ml SOB supplemented with 10 mM CaCl₂ and MgCl₂, and 1-ml samples were infected with REP12 phage at a MOI of 0.01. Infected cultures were incubated at 37 °C for 10 min without shaking for phage attachment and then washed twice with 1 ml SOB by pelleting cells at 4,000g for 3 min. Infected cells were then diluted into 50 ml SOB supplemented with 10 mM CaCl₂ and MgCl₂ and incubated at 37 °C with shaking at 250 r.p.m. Every 20 min, a 1-ml sample was measured out into a sterile Eppendorf tube containing 100 µl chloroform, immediately vortexed, and then placed on ice. Phage titres were determined by centrifuging chloroform-containing cultures at 6,000g for 3 min and then serially diluting supernatants in SM buffer and spotting 1 µl dilutions to LBL 0.7% top agar plates seeded with MDS42 cells and 10 mM CaCl₂ and MgCl₂. Following 18 h of incubation at 37 °C, plaques were counted, and the number of plaques was multiplied by the dilution to calculate the phage titre of the original sample.

Bacteriophage genome sequencing, assembly and annotation

Genomic DNA of bacteriophages was prepared from high-titre (that is, $\geq 10^{10}$ plaque-forming units (PFUs) ml⁻¹) stocks after DNase treatment using the Norgen Biotek Phage DNA Isolation Kit (catalogue no. 46800) according to the manufacturer's guidelines and sequenced at the Microbial Genome Sequencing Center (MiGS; Pittsburgh, PA, USA). Sequencing libraries were prepared using the Illumina DNA Prep kit and IDT 10-base pair (bp) UDI indices and sequenced on an Illumina NextSeq 2000, producing 150-bp paired-end reads. Demultiplexing, quality control and adapter trimming were carried out with bcl-convert (v3.9.3). Reads were trimmed to Q28 using BBduk from BBTools. Phage genomes were then assembled de novo using SPAdes 3.15.2 in –careful mode with an average read coverage of 10–50×. Assembled genomes were then annotated using Prokka version 1.14.6 (Seemann, 2014) with default parameters, except that the PHROGs HMM database⁵¹ was used as input to improve phage functional gene annotations.

Bacterial genome sequencing and annotation

Genomic DNA from overnight saturated cultures of isogenic bacterial clones was prepared using the MasterPure Complete DNA and RNA Purification Kit (Lucigen) according to the manufacturer's guidelines and sequenced at the MiGS (Pittsburgh, PA, USA). Sequencing libraries were prepared using the Illumina DNA Prep kit and IDT 10-bp UDI indices and sequenced on an Illumina NextSeq 2000, producing 150-bp paired-end reads. Demultiplexing, quality control and adapter trimming were carried out with bcl-convert (v3.9.3). Reads were then trimmed to Q28 using BBduk from BBTools and aligned to their corresponding reference by using Bowtie2 2.3.0 (Langmead & Salzberg, 2012) in –sensitive-local mode. Single-nucleotide polymorphisms (SNPs) and indels were called using breseq (version 0.36.1) (Deatherage & Barrick,

2014). Only variants with a prevalence higher than 75% were voted as mutations. Following variant calling, mutations were also manually inspected within the aligned sequencing reads in all cases.

The de novo sequencing and genome assembly of Syn61Δ3(ev5) (from a single-colony isolate of Addgene strain no. 174514) was carried out by generating 84,136 Oxford Nanopore (ONT) long reads by PCR-free library generation (Oxford Nanopore) on a MinION Flow Cell (R9.4.1) and 4.5×10^6 150-bp paired-end reads on an Illumina NextSeq 2000. Quality control and adapter trimming were carried out with bcl2fastq 2.20.0.445 and porechop 0.2.3_seqan2.1.1 for Illumina and ONT sequencing, respectively. Next, we carried out hybrid assembly with Illumina and ONT reads by using Unicycler 0.4.8 by using the default parameters. Finally, the resulting single, circular contig representing the entire genome was manually inspected for errors in Geneious Prime 2022.1.1 and annotated on the basis of sequence homology by using the BLAST function (version 2.11.0) implemented in Geneious Prime 2022.1.1 based on *E. coli* K12 MG1655 (NCBI ID: U00096.3) as a reference. Gene essentiality was determined on the basis of (Goodall *et al.*, 2018).

Transcriptome analysis of phage-infected cells

We explored transcriptomic changes and mRNA production in phage-infected Syn61Δ3 cells by carrying out a modified single-step growth experiment and collected samples at 20-min intervals. A 50 ml volume of early-exponential-phase ($OD_{600nm} = 0.15$) Syn61Δ3 cells (corresponding to 2×10^{10} CFUs) growing at 37 °C, 250 r.p.m. in SOB containing 10 mM CaCl₂ and MgCl₂ was spun down at room temperature and resuspended in 1 ml SOB. A 50 µl volume of this uninfected sample was immediately frozen in liquid N₂ and stored at -80 °C until RNA extraction. Next, 900 µl of this cell suspension was mixed with 10 ml prewarmed REP12 phage stock (that is, about 7×10^{10} PFUs to achieve a MOI of about 4) in SOB containing 10 mM CaCl₂ and MgCl₂, and then incubated at 37 °C for 10 min without shaking for phage absorption. Following phage attachment, samples were spun down, washed with 1 ml SOB twice to remove unabsorbed phages, and then resuspended in 10 ml SOB containing 10 mM CaCl₂ and MgCl₂. Samples were then incubated at 37 °C, 250 r.p.m. After 20- and 40-min post-infection, we spun down a 1-ml cell suspension from each sample, and the cell pellets were frozen in liquid N₂ and stored at -80 °C until RNA extraction. As expected, after 60 min post-infection, no cell pellet was visible. Phage infections were carried out in three independent replicates. Total RNA from frozen samples was extracted by using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions and the extracted RNA was DNase-treated with Invitrogen RNase-free DNase (Thermo Fisher Scientific). Sequencing library preparation was then carried out using Stranded Total RNA Prep Ligation kit with Ribo-Zero Plus for rRNA depletion

and by using 10-bp IDT for Illumina indices (all from Illumina). Sequencing was carried out on a NextSeq 2000 instrument in 2×50 -bp paired-end mode. Demultiplexing, quality control and adapter trimming were carried out with bcl-convert (v3.9.3). cDNA reads were aligned to their corresponding reference by using Bowtie2 2.3.0 (Langmead & Salzberg, 2012) in –sensitive-local mode, and read count and expression metrics were determined by using Geneious Prime 2022.1.1 (Biomatters). Finally, differential expression analysis was carried out using DESeq2 ((Love *et al.*, 2014); version 1.38.3) with default settings.

tRNA-seq sample preparation

We explored tRNA expression levels and changes in phage-infected Syn61Δ3 cells by carrying out a modified single-step growth experiment with high MOI and cell mass. An early-exponential-phase culture ($OD_{600nm} = 0.2$) of Syn61Δ3 cells (corresponding to approximately 5×10^{10} CFUs) growing at 37°C , 250 r.p.m. in SOB containing 10 mM CaCl₂ and MgCl₂ was spun down at room temperature and resuspended in 1.1 ml SOB. A 100 μl volume of this uninfected sample was immediately frozen in liquid N₂ and stored at -80 °C until tRNA extraction. Next, 1,000 μl of this cell suspension was mixed with 20 ml prewarmed REP12 phage stock (that is, about 10¹² PFUs to achieve a MOI of about 20) in SOB containing 10 mM CaCl₂ and MgCl₂, and then incubated at 37°C for 10 min without shaking for phage absorption. Following phage attachment, samples were spun down, the supernatant containing unabsorbed phages was removed, and the cell pellet was then resuspended in 7 ml SOB containing 10 mM CaCl₂ and MgCl₂. Samples were then incubated at 37°C , 250 r.p.m. Immediately after phage attachment and after 20- and 40-min post-infection, 1-ml cell suspensions from each sample were spun down, and cell pellets were frozen in liquid N₂ and stored at -80 °C until total RNA extraction. Phage infections were carried out in two independent replicates.

We analysed the total tRNA content of Ec_Syn61Δ3-SL cells expressing KP869110.1 viral tRNA24Leu(UGA) and tRNA24Leu(CGA) by pelleting cells from 5 ml mid-exponential-phase ($OD_{600nm} = 0.3$) culture at 4,000g and flash-freezing the cell pellet in liquid N₂.

We extracted tRNAs by lysing samples at room temperature for 30 min in 150 μl lysis buffer containing 8 mg ml⁻¹ lysozyme (from chicken egg white, no. 76346-678, VWR), 10 mM Tris-HCl pH 7.5 and 1 μl murine RNase inhibitor (New England Biolabs). Samples were then mixed with 700 μl Qiazol reagent (no. 79306, Qiagen) and incubated for 5 min at room temperature. Next, 150 μl chloroform was added, vortexed and incubated until phase separation. Samples were then spun at 15,000g for 15 min in a cooled centrifuge. The supernatant was transferred to a new Eppendorf tube and mixed with 350 μl 70% ethanol. Larger RNA molecules were then bound to an RNeasy MinElute spin column (no. 74204, Qiagen), and the flow-through was mixed with 450 μl of 100% ethanol, and tRNAs were bound to a new RNeasy MinElute spin

column. The tRNA fraction was then washed first with 500 µl wash buffer (no. 74204, Qiagen), next with 80% ethanol, and then eluted in RNase-free water. The eluted tRNAs were deacylated in 60 mM pH 9.5 borate buffer (J62154-AK, Alfa Aesar, Thermo Fisher Scientific) for 30 min and then purified using a Micro Bio-Spin P-30 gel column (7326251, Bio-Rad).

tRNA-seq library preparation, sequencing and data analysis

We prepared tRNA cDNA libraries by reverse-transcribing tRNAs using the TGIRT-III template-switching reverse transcriptase (TGIRT50, InGex) according to the manufacturer's instructions. In brief, we prepared reaction mixtures containing 1 µl (about 100 ng) of the deacylated tRNAs, 2 µl of 1 µM TGIRT DNA–RNA heteroduplex (prepared by hybridizing equimolar amounts of rCrUrUrUrGrArGrCrCrUrArArUrGrCrCrUrGrArArArGrArUrCrGrGrArArGrArGrCrArCrAr-CrGrUrCrUrArGrUrUrCrUrArCrArGrUrCr-CrGrArCrGrArU/3SpC3/ and ATCGTCGGACT-GTAGAACTAGACGTGTGCTCTTCCGATCTTCAGGCATTAGGCTCAAAGN oligonucleotides), 4 µl 5× TGIRT reaction buffer (2.25 M NaCl, 25 mM MgCl₂, 100 mM Tris-HCl, pH 7.5), 2 µl of 100 mM DTT, 9 µl RNase-free water and 1 µl TGIRT-III, and incubated at room temperature for 30 min to initiate template-switching. Next, 1 µl of 25 mM dNTPs (Thermo Fisher Scientific) was added to the reaction mixture, and samples were incubated at 60 °C for 30 min to carry out reverse transcription. RNA was then hydrolysed by NaOH, and then neutralized by HCl, and the cDNA library was purified using the MinElute PCR purification kit. cDNAs were then ligated to a preadenylated DNA adapter, /5Phos/GATCNNNAGATCGGAAGAGCGTCGT-GT/3SpC3/ (in which NNN denotes an N, NN or NNN spacer), to increase library diversity during sequencing (preadenylated oligonucleotides were prepared by a 5' DNA adenylation kit (E2610L) using thermostable 5' App DNA/RNA ligase (M0319L, both from New England Biolabs) following the manufacturer's protocol). The cDNA library was purified using the MinElute PCR purification kit (Qiagen) and amplified using Q5 Host-Start High-Fidelity 2× Master Mix (New England Biolabs). PCR products were then size selected to remove adapter dimers below 200 bp using three subsequent size-selection rounds with a Select-a-Size DNA Clean & Concentrator Kit (D4080, Zymo Research). Finally, amplicon libraries were barcoded using the IDT 10-bp UDI indices (Illumina) and sequenced on an Illumina MiSeq to produce 250-bp paired-end reads. Read-demultiplexing was carried out with bcl-convert (v3.9.3). Paired-end reads were then aligned to their reference sequences by using Geneious assembler, implemented in Geneious Prime 2022.1.1, allowing a maximum of ten SNPs within tRNA reads compared to their reference. These settings allowed us to map lower-fidelity TGIRT-III-transcribed cDNA reads to their corresponding reference sequence without cross-mapping to tRNAs sharing sequence homology. tRNA reads from Ec_Syn61Δ3-SL cells expressing KP869110.1 viral tRNA24Leu(UGA) and tRNA24Leu(CGA) were mapped without allowing the presence of

SNPs in sequencing reads to distinguish tRNA₂₄Leu(UGA) and tRNA₂₄Leu(CGA) that differs by only a single SNP within the anticodon region.

Genome editing and biocontainment of Syn61Δ3

We first generated a deficient recombination variant of Syn61Δ3(ev5) by eliminating the expression of the genomic recA gene using Cas9-assisted recombineering. recA-deletion experiments were carried out by first transforming Syn61Δ3(ev5) cells with a plasmid carrying a pSC101 origin of replication, a constitutively expressed chloramphenicol resistance marker, SpCas9 and tracrRNA (from pCas9 (refs. 56,57), Addgene no. 42876), and the λRed operon, consisting of gam, exo and bet (from pORTMAGE311B (Szili *et al.*, 2019), Addgene no. 120418). Next, cells were made electrocompetent using a standard protocol^{56,57} for Cas9-assisted recombineering and transformed with 2 µl of 100 µM 90-nucleotide-long single-stranded DNA oligonucleotide inserting a stop codon and a frameshift mutation into recA (Supplementary Data 3). Successful edits were selected by cotransforming 1 µg from a variant of the pCRISPR plasmid (Jiang *et al.*, 2013; Umenhoffer *et al.*, 2017) carrying a 5'-AGTTGATACCTCGCCGTAG-3' guide sequence to cleave the genomic recA sequence in unedited cells. All plasmids were recoded to lack TCR and TAG codons in protein-coding genes, and synthesized by GenScript USA Inc. The resulting Syn61Δ3(ev5) ΔrecA strain was validated by whole-genome sequencing and then evolved for increased fitness (see the below section entitled Adaptive laboratory evolution of Syn61Δ3). Finally, the replacement of the genomic adk gene of Syn61Δ3(ev5) ΔrecA (ev1) with the bipA-dependent adk.d6 variant¹⁰ was carried out by first transforming cells with a plasmid carrying a constitutively expressed MjTyrRS-derived bipA aaRS (variant 10, based on (Kunjapur *et al.*, 2018)) together with its associated tRNA under the control of a proK tRNA promoter and an aminoglycoside-(3)-N-acetyltransferase gene, conferring gentamycin resistance, all on a plasmid containing a p15A origin of replication (Supplementary Data 3). Next, we integrated the adk.d6 variant by Cas9-assisted recombineering as described above, but instead of oligonucleotide-mediated recombineering, we transformed 4 µg of a dsDNA cassette carrying the full-length adk.d6 variant with 400-bp flanking genomic homology (constructed by GenScript USA Inc.; Supplementary Data 3). Cells were grown in the presence of 200 µM bipA in 2×YT medium throughout the entire procedure. Successful edits were selected using a dual-targeting crRNA expression construct, carrying 5'-GCAATGCGTATCATTCTGCT-3' and 5'-GCCGTCAACTTCGCGTATT-3' guide sequences (from GenScript). Positive colonies were selected by screening colonies with allele-specific PCR (Supplementary Data 3) and validated by whole-genome sequencing. Finally, the escape rate of the resulting Syn61Δ3(ev5) ΔrecA (ev1) adk.d6 strain was determined as described earlier¹⁰, but instead of chloramphenicol, cells were grown in the presence of 10 µg ml⁻¹ gentamycin in 2×YT. Plates were incubated for 7

days at 37 °C. Escape rate measurements were carried out in triplicate; ± indicates standard deviation.

Adaptive laboratory evolution of Syn61Δ3

We carried out standard adaptive laboratory evolution in rich bacterial medium for 30 days (about 270 cell generations) on Syn61Δ3(ev5) ΔrecA cells to increase fitness. At each transfer step, 10⁹–10¹⁰ bacterial cells were transferred into 500 ml LBL medium containing 1.5 g l⁻¹ Tris/Tris-HCl and incubated aerobically for 24 h at 37 °C, 250 r.p.m. in a 2,000-ml Erlenmeyer flask with a vented cap. Following 30 transfers, bacterial cells were spread onto LBL agar plates, and an individual colony was isolated and subjected to whole-genome sequencing. The identified mutations in the resulting evolved variant, Syn61Δ3(ev5) ΔrecA (ev1), are listed in Supplementary Data 3.

Doubling-time measurements

To determine growth parameters under standard laboratory conditions, saturated overnight cultures of *E. coli* Syn61Δ3(ev5), Syn61Δ3(ev5) ΔrecA and its evolved variant Syn61Δ3(ev5) ΔrecA (ev1) were diluted 1:200 into 50 ml of 2×YT and LBL in a 300-ml Erlenmeyer flask with vented cap and incubated aerobically at 37 °C, 250 r.p.m. Ec_Syn61Δ3-SL cells were characterized similarly, but by using 2×YT containing 50 µg ml⁻¹ kanamycin. All growth measurements were carried out in triplicate. OD_{600nm} measurements were taken every 20 min for 8 h or until stationary phase was reached on a CO8000 Cell Density Meter (WPA). The doubling time was calculated for each independent replicate by log2-transforming OD_{600nm} values and calculating the doubling time based on every six consecutive data points during the exponential growth phase. We calculated the doubling time (1/slope) from a linear fit to log2 derivatives of the six data points within this window and reported the shortest doubling time for each independent culture. Curve fitting, linear regression and doubling-time calculations were carried out with Prism9 (GraphPad). Error bars show ±s.d.

tRNA annotation

We detected tRNA genes in the Viral genomic NCBI Reference Sequence Database (accessed 2 January 2022) and in the genomes of individual phage isolates by using tRNAscan-SE 2.0.9 in bacterial (-B), archaeal (-A) or eukaryotic (-E) maximum sensitivity mode (-I -max) (Chan *et al.*, 2021). tRNAscan-SE detection parameters were chosen according to the predicted host of the corresponding viral strain.

Mobile tRNAome tRNA library generation and selection

We generated our mobile tRNAome expression library by synthesizing tRNAscan-SE-predicted tRNAs from diverse sources (Supplementary Data 1), driven by a strong bacterial proK tRNA

promoter and followed by two transcriptional terminators as 10-pmol single-stranded DNA oligonucleotide libraries (10 pmol oPool, from Integrated DNA Technologies). Oligonucleotides were resuspended in 1× TE buffer and then amplified using 5' phosphorylated primers. Ampli-cons were then blunt-end ligated into pCR4Blunt-TOPO (Invitrogen, Zero Blunt TOPO PCR Cloning Kit) for 18 h at 16 °C and then purified by using the Thermo Scientific GeneJET PCR Purification Kit. We then electroporated 50 ng purified plasmid in five parallel electroporations into 5 × 40 µl freshly made electrocompetent cells of MDS42 and Syn61Δ3(ev5). Before electrotransformation, bacterial cells were made electrocompetent by growing cells after a 1:100 dilution in SOB until mid-log phase ($OD_{600\text{nm}} = 0.3$) at 32 °C and then washing cells three times using ice-cold water. Electroporated cultures were allowed to recover overnight at 37 °C and then plated to LBL agar plates containing 50 µg ml⁻¹ kanamycin in 145 × 20 mm Petri dishes (Greiner Bio-One). Plates were incubated at 37 °C until colony formation. Approximately 1,000–5,000 colonies were then washed off from selection plates, and plasmids were extracted by using the Monarch Plasmid Miniprep Kit (New England Biolabs). The tRNA insert from isolated plasmids was then amplified with primers bearing the standard Nextera Illumina Read 1 and Read 2 primer binding sites, barcoded using the IDT 10-bp UDI indices, and sequenced on an Illumina NextSeq 2000, producing 150-bp paired-end reads. Demultiplexing was carried out with bcl-convert (v3.9.3). Paired-end reads were then trimmed using BBduk from BBTools (in Geneious Prime 2022.1.1, Biomatters), merged and aligned to their reference sequences by using Geneious assembler, implemented in Geneious Prime 2022.1.1, allowing maximum a single SNV within the tRNA read.

tRNALeu(YGA) library generation and selection

We identified leucine tRNAs that can translate TCR codons as leucine by carrying out two consecutive screens with plasmid libraries expressing an anticodon-loop-mutagenized 65,536-member library of leuU tRNA variants and a smaller, 13-member tRNALeu(YGA) expression library consisting of bacteriophage-derived leucine tRNA variants, both bearing two tRNAs under the control of a proK promoter and with anticodons swapped to UGA and CGA. To construct a 65,536-member leuU tRNA library, we synthesized an expression construct consisting of a proK promoter–leuUUGA–spacer–leuUCGA–proK terminator sequence, in which the anticodon loop of both leuU tRNAs has been fully randomized, as an oPool library (Supplementary Data 3; Integrated DNA Technologies). Next, we amplified these leuU variants by using Q5 Hot Start High-Fidelity Master mix using 5'-phosphorylated primers and then ligated the library into a plasmid backbone containing a high-copy-number pUC origin of replication and an APH(3')-I aminoglycoside O-phosphotransferase (aph3Ia29×Leu-to-TCR) gene in which all 29 instances of leucine-coding codons were replaced with TCR serine codons (synthesized as

a gBlock dsDNA fragment by Integrated DNA Technologies). The ligation was carried out at a 3:1 insert-to-vector ratio and by using T4 DNA ligase (New England Biolabs) for 16 h at 16 °C according to the manufacturer's instructions. Finally, the ligation product was purified using the GeneJet PCR purification kit (Thermo Fischer Scientific). We constructed the second, 13-member tRNA_{Leu}(YGA) expression library (Supplementary Data 3) consisting of bacteriophage-derived leucine tRNA variants bearing a UGA and CGA anticodon by using the same method as for our leuUYGA library. Following library generation, 100 ng from each library was electroporated into freshly made electrocompetent cells of Syn61Δ3 (ev5) ΔrecA (ev1) and recovered in SOB at 37 °C for 16 h, 250 r.p.m. After recovery, the cells were plated to 2×YT agar plates containing kanamycin at 200 µg ml⁻¹ concentration, and selection plates were incubated at 37 °C until colony formation. Finally, plasmids from clones were purified using a Monarch plasmid miniprep kit (New England Biolabs) and subjected to whole-plasmid sequencing (SNPsaurus, Eugene, OR, USA).

Virus resistance analysis of Ec_Syn61Δ3-SL cells

An exponential-phase culture ($OD_{600\text{nm}} = 0.3$) of the corresponding strain was grown in 3 ml SOB supplemented with 10 mM CaCl₂ and MgCl₂ and 75 µg ml⁻¹ kanamycin at 37 °C with shaking. Cultures were then spun down and resuspended in 1 ml SOB supplemented with 10 mM CaCl₂ and MgCl₂ and infected with a 1:1 mixture of all 12 Syn61Δ3-lytic phage isolates from this study (Extended Data Table 1b) at a MOI of 0.1. Infected cultures were incubated at 37 °C without shaking for 10 min for phage attachment and then washed three times with 1 ml SOB supplemented with 10 mM CaCl₂ and MgCl₂ by pelleting cells at 4,000g for 3 min. Infected cells were then diluted into 4 ml SOB supplemented with 10 mM CaCl₂ and MgCl₂ and 75 µg ml⁻¹ kanamycin and incubated at 37 °C with shaking at 250 r.p.m. After 24 h of incubation, 500-µl samples were measured out into a sterile Eppendorf tube containing 50 µl chloroform, immediately vortexed and then placed on ice. Phage infection experiments were carried out in three independent replicates. Phage titres were determined by centrifuging chloroformed cultures at 6,000g for 3 min and then plating 5 µl of the supernatant directly or its appropriate dilutions mixed with 300 µl MDS42 cells in LBL 0.7% top agar with 10 mM CaCl₂ and MgCl₂. Following 18 h of incubation at 37 °C, plaques were counted, and the number of plaques was multiplied by the dilution to calculate the phage titre of the original sample.

Phage enrichment experiments were carried out by mixing 50 ml early-exponential-phase cultures ($OD_{600\text{nm}} = 0.2$) of bipA-biocontained Ec_Syn61Δ3-SL carrying pLS1 and pLS2 plasmids with 10 ml environmental sample mix, containing the mixture of samples 2–13 from our study (Extended Data Table 1b). We maximized phage diversity in the environmental sample mix by using freshly collected and filter-sterilized sewage. Infected Ec_Syn61Δ3-SL cells with

the corresponding plasmid were grown overnight in SOB supplemented with 200 mM bipA, 10 mM CaCl₂ and MgCl₂, and 75 µg ml⁻¹ kanamycin at 37 °C with shaking at 250 r.p.m. On the next day, cells were removed by centrifugation at 4,000g for 20 min, and the supernatant was filter-sterilized using a 0.45-µm filter. Next, 5 ml of the sterilized sample was mixed again with 50 ml early-exponential-phase cultures (OD_{600nm} = 0.2) of the corresponding strain, incubated for 20 min at 37 °C for phage absorption, pelleted by centrifugation at 4,000g for 15 min, and then resuspended in 50 ml SOB supplemented with 200 mM bipA, 10 mM CaCl₂ and MgCl₂, and 75 µg ml⁻¹ kanamycin. Infected cultures were then incubated at 37 °C with shaking at 250 r.p.m. Cultures were grown overnight and then sterilized by centrifugation at 4,000g for 15 min and filtered through a 0.45-µm PVDF Steriflip disposable vacuum filter unit (MilliporeSigma). Finally, phage titres were determined by using MDS42 cells as above. Phage enrichment experiments were carried out in two independent replicates. The lytic phage titre of the unenriched sample mix was determined by diluting 100 µl of the input environmental sample mix, containing a mixture of samples 2–13 (Extended Data Table 1), in SM buffer and 10 µl samples from each dilution steps were mixed with late-exponential-phase MDS42 cells and 4 ml 0.7% top agar, and then poured on top of LBL agar plates. Plates were incubated until plaque formation at 37 °C.

Construction of pLS plasmids

All pLS plasmids listed in Supplementary Data 3 were synthesized as gBlocks by IDT and circularized either by ligation with T4 DNA ligase (New England Biolabs), or, in the case of pSC101 and RK2 plasmid-derived variants, by isothermal assembly using the HiFi DNA Assembly Master Mix (New England Biolabs). Following assembly, purified plasmid assemblies were electroporated into Ec_Syn61Δ3-SL cells carrying pLS1. pLS1 and pLS2 were designed to express two distinct combinations of the previously identified phage tRNA^{Leu(YGA)} tRNAs in antiparallel orientation to avoid repeat-mediated instability (Hossain *et al.*, 2020), driven by the strong bacterial proK and SLP2018-2-101 promoters, together with a high-copy-number pUC origin of replication, the aph3Ia29xLeu-to-TCR and aminoglycoside-(3)-N-acetyltransferase18xLeu-to-TCR marker genes. Transformants carrying either pLS1 or pLS2 were identified by transforming assemblies into Syn61Δ3(ev5) ΔrecA (ev1) and selecting for kanamycin resistance. Finally, plasmids from antibiotic-resistant clones were purified using a Monarch plasmid miniprep kit (New England Biolabs) and subjected to whole-plasmid sequencing (SNPsaurus, Eugene, OR, USA).

Escape rate analysis of viral Leu-tRNAYGA and pLS plasmids

We analysed the ability of pLS plasmids to function outside Ec_Syn61Δ3-SL cells by transforming extracted plasmids into *E. coli* K12 MG1655. Plasmids were purified from biocontained

*Ec*_Syn61Δ3-SL cells, carrying either pLS1 or pLS2 to express tRNA_{Leu}(YGA), or pLS1 together with pLS3-5, by using the PureLink Fast Low-Endotoxin Midi Plasmid Purification Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. Next, we electroporated 1 µg from each plasmid preparation into freshly made electrocompetent cells of *E. coli* K12 MG1655. Cells were made electrocompetent by diluting an overnight SOB culture of MG1655 1:100 into 500 ml SOB in a 2,000-ml flask and growing cells aerobically at 32 °C with shaking at 250 r.p.m. At OD_{600nm} = 0.3, cells were cooled on ice and then pelleted by centrifugation and resuspended in 10% glycerol-in-water. Cells were washed four times with 10% glycerol-in-water and then resuspended in 400 µl 20% glycerol-in-water. A 1,000 ng quantity from each plasmid sample was then mixed with 80 µl electrocompetent cells and electroporated by using standard settings in two 1-mm electroporation cuvettes by using standard electroporator settings (1.8 kV, 200 Ohm, 25 µF). Electroporations were carried out in three independent replicates. Electroporated cells were then resuspended in 1 ml SOB, and the culture was allowed to recover overnight at 37 °C with shaking at 250 r.p.m. Finally, 500 µl from each recovery culture was plated to LBL agar plates containing antibiotics corresponding to the given pLS plasmid's resistance marker (15 µg ml⁻¹ gentamycin plus 50 µg ml⁻¹ kanamycin in the case of pLS1 and 2; 100 µg ml⁻¹ carbenicillin for pLS4, 30 µg ml⁻¹ chloramphenicol for pLS3 and pLS6, and 20 µg ml⁻¹ gentamycin for pLS5) in 145 × 20-mm Petri dishes (Greiner Bio-One). Plates were incubated at 37 °C for 7 days and inspected for growth. Electroporation efficiency measurements were carried out by electroporating a plasmid carrying a pUC origin of replication and kanamycin resistance (pUC-KanR) into MG1655 electrocompetent cells under identical conditions.

Cloning of REP12 tRNA operon

We analysed the incorporated amino acid in Syn61Δ3 cells bearing the tRNA operon and its native promoter from the REP12 phage by subcloning the genomic tRNA operon into a low-copy-number plasmid containing an RK2 origin of replication and a chloramphenicol-acetyltransferase marker, both recoded to contain no TCR or TAG codons. The genomic tRNA operon of REP12 was PCR amplified using the Q5 Hot Start Master Mix (New England Biolabs) from extracted phage gDNA and purified using the GeneJet PCR purification kit (Thermo Fisher Scientific). Next, 100 ng of the amplified tRNA operon was assembled into the linearized pRK2-cat backbone using the HiFi DNA Assembly Master Mix (New England Biolabs). After incubation for 60 min at 50 °C, the assembly was purified using the DNA Concentrator & Clean kit (Zymo Research) and transformed into Syn61Δ3(ev5) ΔrecA (ev1) cells expressing an MSKGPGKVPAGVPGXGVPGVGKGGT-elastin peptide fused to sfGFP with a terminal 6×His tag (in which X denotes the analysed codon, TCA or TCG) on a plasmid containing a

kanamycin resistance gene and a pUC origin of replication (Mohler *et al.*, 2017). Following an overnight recovery at 32 °C, cells were plated to 2×YT agar plates containing kanamycin and chloramphenicol. Finally, plasmid sequences in outgrowing colonies were validated by whole-plasmid sequencing. Elastin(16TCR)–sfGFP–His6 expression measurements have been carried out as described below.

Elastin(16TCR)–sfGFP–His6 expression measurements

We assayed the amino acid identity of the serine tRNAUGA tRNAs of MZ501075 and REP12 (see Supplementary Data 3 for sequence information) by coexpressing selected tRNAs and a constitutively expressed MSKGPGKVPAGVPGXGVPGVGKGGT-elastin peptide fused to sfGFP with a terminal 6×His tag (in which X denotes the analysed codon, TCA or TCG)62 on a plasmid containing a kanamycin resistance gene and a pUC origin of replication in Syn61Δ3(ev5). Similarly, the pRK2-REP12 plasmid carrying the REP12 phage tRNA operon under the control of its native promoter was coexpressed with the same pUC plasmid carrying no tRNA genes. As a control, we used the same elastin–sfGFP–His6 expression construct in which position X has been replaced with an alanine GCA codon. For fluorescence and MS/MS measurements, we diluted cultures 1:100 from overnight starters into 50 ml 2×YT in 300-ml shake flasks containing the corresponding antibiotics and cultivated for 48 h at 37 °C, 200 r.p.m. aerobically. We then determined sfGFP expression levels in samples by pelleting and washing 1 ml of the culture with phosphate-buffered saline and resuspending cell pellets in 110 µl BugBuster Protein Extraction Reagent (MilliporeSigma). Reactions were incubated for 5 min and then spun down at 13,000 r.p.m. for 10 min. The fluorescence of the BugBuster-treated supernatants and the OD_{600nm} of the original culture were measured using the Synergy H1 Hybrid Reader (BioTek) plate reader using bottom-mode analysis with excitation at 480 nm and emission measurement at 515 nm, with the gain set to 50. Fluorescence values were normalized on the basis of OD_{600nm} data. The remaining 49 ml culture was spun down, and the cell pellet was resuspended in 2 ml BugBuster Protein Extraction Reagent (MilliporeSigma) and incubated at room temperature for 5 min. The lysed cell mixture was spun down at 13,000 r.p.m. for 10 min, and the supernatant was mixed in a 1:1 ratio with His-Binding/Wash Buffer (G-Biosciences) and 50 µl HisTag Dynabeads (Thermo Fischer Scientific). Following an incubation period of 5 min, the beads were separated on a magnetic rack and washed with 300 µl His-Binding/Wash Buffer and phosphate-buffered saline three times. After the last wash step, the bead pellets containing the bound elastin–sfGFP–His6 protein samples were frozen at -80 °C until MS/MS sample preparation. Protein production experiments were carried out in three independent replicates.

LC/MS–MS analysis of tryptic elastin–sfGFP–His6

Samples from elastin–sfGFP–His6 expression experiments were digested directly on HisTag Dynabeads according to the FASP digest procedure⁶³. In brief, samples were washed with 50 mM triethylammonium bicarbonate (TEAB) buffer and then rehydrated with 50 mM TEAB–trypsin solution, followed by a 3-h digest at 50 °C. Digested peptides were then separated from HisTag Dynabeads and concentrated by spinning and drying samples at 3,000 r.p.m. using a SpeedVac concentrator. Samples were then solubilized in 0.1% formic acid-in-water for subsequent analysis by MS/MS. LC–MS/MS analysis of digested samples was carried out on a Lumos Tribrid Orbitrap mass spectrometer equipped with an Ultimate 3000 nano-HPLC (both from Thermo Fisher Scientific). Peptides were separated on a 150-μm-inner-diameter microcapillary trapping column packed first with 2 cm of C18 ReproSil resin (5 μm, 100 Å, from Dr. Maisch) followed by a 50-cm analytical column (PharmaFluidics). Separation was achieved by applying a gradient from 4% to 30% acetonitrile in 0.1% formic acid over 60 min at 200 nl min⁻¹. Electrospray ionization was carried out by applying a voltage of 2 kV using a custom electrode junction at the end of the microcapillary column and sprayed from metal tips (PepSep). The MS survey scan was carried out in the Orbitrap in the range of 400–1,800 m/z at a resolution of 6 × 10⁴, followed by the selection of the 20 most intense ions for fragmentation using collision-induced dissociation in the second MS step (CID-MS2 fragmentation) in the ion trap using a precursor isolation width window of 2 m/z, automatic gain control setting of 10,000 and a maximum ion accumulation of 100 ms. Singly charged ion species were excluded from CID fragmentation. The normalized collision energy was set to 35 V and an activation time of 10 ms. Ions in a 10-ppm m/z window around ions selected for MS/MS were excluded from further selection for fragmentation for 60 s.

The raw data were analysed using Proteome Discoverer 2.4 (Thermo Fisher Scientific). Assignment of MS/MS spectra was carried out using the Sequest HT algorithm by searching the data against a protein sequence database, including all protein entries from *E. coli* K12 MG1655, all proteins sequences of interest (including the elastin–sfGFP fusion protein), and other known contaminants such as human keratins and common laboratory contaminants. Quantitative analysis between samples was carried out by label-free quantification between different samples. Sequest HT searches were carried out using a 10-ppm precursor ion tolerance and requiring the N and C termini of each peptide to conform with trypsin protease specificity while allowing up to two missed cleavages. Methionine oxidation (+15.99492 Da), deamidation (+0.98402 Da) of asparagine and glutamine amino acids, phosphorylation at serine, threonine and tyrosine amino acids (+79.96633 Da) and N-terminal acetylation (+42.01057 Da) were set as variable modifications. We then determined the amino acid incorporated at position X in our elastin–sfGFP–6×His construct by analysing changes compared to phenylalanine. To cover all

20 possible amino acid exchange cases at the X position, we carried out five separate searches with four different amino acids as possible variable modifications in each search. All cysteines were analysed without modification as no alkylation procedure was included in our workflow. An overall false discovery rate of 1% on both protein and peptide level was achieved by carrying out target-decoy database search using Percolator (Käll *et al.*, 2008).

Total proteome analysis and the detection of serine-to-leucine mistranslation events We analysed the translation of viral proteins in Ec_Syn61Δ3-SL cells (Syn61Δ3(ev5) ΔrecA (ev1) expressing a proK promoter-driven Leu9-tRNAYGA construct from the *Escherichia* phage OSYSP (GenBank ID MF402939.1) and APH(3')-I aminoglycoside O-phosphotransferase (aph3Ia29 × Leu-to-TCR), on a high-copy-number pUC plasmid), by carrying out a modified single-step growth experiment and subsequent time-course MS/MS-based proteome analysis. An early-exponential-phase culture ($OD_{600\text{nm}} = 0.2$) of Ec_Syn61Δ3-SL cells (corresponding to approximately 4×10^{10} CFUs) growing at 37 °C, 250 r.p.m. in SOB containing 10 mM CaCl₂, MgCl₂ and 75 µg ml⁻¹ kanamycin was spun down at room temperature and resuspended in 1.1 ml SOB containing 10 mM CaCl₂, MgCl₂ and 75 µg ml⁻¹ kanamycin. A 100 µl volume of this uninfected sample was immediately frozen in liquid N₂ and stored at -80 °C until proteome analysis. Next, 1,000 µl of this cell suspension was mixed with 10 ml prewarmed REP12 phage stock (that is, about 5×10^{11} PFUs to achieve a MOI of about 12) in SOB containing 10 mM CaCl₂, MgCl₂ and 75 µg ml⁻¹ kanamycin, and then incubated at 37 °C for 10 min without shaking for phage absorption. Following phage attachment, samples were spun down, the supernatant containing unabsorbed phages was removed, and the cell pellet was resuspended in 5 ml SOB containing 10 mM CaCl₂ and MgCl₂. Samples were then incubated at 37 °C, 250 r.p.m. After 20- and 40-min post-infection, 1-ml cell suspensions were spun down, and cell pellets were frozen in liquid N₂ and stored at -80 °C until total protein extraction. Samples from control and phage-infected Syn61Δ3 cells were then digested by using the FASP digest procedure (Wiśniewski *et al.*, 2009). In brief, samples were washed with 50 mM TEAB buffer on a 10-kDa-cutoff filter (Pall Corp) and then rehydrated with 50 mM TEAB-trypsin solution, followed by a 3-h digestion at 37 °C. Digested peptides were then extracted and separated into ten fractions by using the Pierce High pH Reversed-Phase Peptide Fractionation Kit according to the manufacturer's protocol (Thermo Fisher Scientific). Following fractionation, peptides were concentrated and dried by spinning samples at 3,000 r.p.m. using a SpeedVac concentrator. Samples were then solubilized in 0.1% formic acid-in-water for subsequent analysis by MS/MS. LC-MS/MS analysis of digested samples was carried out on a Lumos Tribrid Orbitrap Mass Spectrometer equipped with an Ultimate 3000 nano-HPLC (both from Thermo Fisher Scientific). Peptides were separated on a 150-µm-inner-diameter microcapillary trapping col-

umn packed first with 2 cm of C18 Reprosil resin (5 µm, 100 Å, from Dr. Maisch) followed by a 50-cm analytical column (PharmaFluidics). Separation was achieved by applying a gradient from 5% to 27% acetonitrile in 0.1% formic acid over 90 min at 200 nL min⁻¹. Electrospray ionization was carried out by applying a voltage of 2 kV using a custom electrode junction at the end of the microcapillary column and sprayed from metal tips (PepSep). The MS survey scan was carried out in the Orbitrap in the range of 400–1,800 m/z at a resolution of 6 × 104, followed by the selection of the 20 most intense ions for fragmentation using collision-induced dissociation in the second MS step (CID-MS2 fragmentation) in the ion trap using a precursor isolation width window of 2 m/z, automatic gain control setting of 10,000 and a maximum ion accumulation of 100 ms. Singly charged ion species were excluded from CID fragmentation. The normalized collision energy was set to 35 V and an activation time of 10 ms. Ions in a 10-ppm m/z window around ions selected for MS/MS were excluded from further selection for fragmentation for 60 s.

The raw data were analysed using Proteome Discoverer 2.4 (Thermo Fisher Scientific). Assignment of MS/MS spectra was carried out using the Sequest HT algorithm by searching the data against a protein sequence database, including all protein entries from *E. coli* K12 MG1655, all protein sequences of the corresponding REP12 bacteriophage and the Aph3Ia APH(3')-I amino-glycoside O-phosphotransferase, as well as other known contaminants such as human keratins and common laboratory contaminants. Quantitative analysis between samples was carried out by label-free quantification between different samples. Sequest HT searches were carried out using a 10-ppm precursor ion tolerance and requiring the N and C termini of each peptide to conform with trypsin protease specificity while allowing up to two missed cleavages. Methionine oxidation (+15.99492 Da), deamidation (+0.98402 Da) of asparagine and glutamine amino acids, phosphorylation at serine, threonine and tyrosine amino acids (+79.96633 Da) and N-terminal acetylation (+42.01057 Da) were set as variable modifications. Special modification of serine-to-leucine amino acid exchange (+26.052036 Da) on all serine amino acid positions was used as a variable modification. All cysteines were analysed without modification as no alkylation procedure was included in our workflow. An overall false discovery rate of 1% on both protein and peptide levels was achieved by carrying out target–decoy database search using Percolator (Käll *et al.*, 2008).

4.2.4 Results

Mobile tRNAs abolish virus resistance

We first investigated whether tRNAs of mobile genetic elements can substitute cellular tRNAs and support viral infection in cells with a compressed genetic code. We sampled the

mobile tRNAome, tRNA genes encoded by horizontally transferred genetic elements, by computationally screening thousands of viral genomes for the presence of tRNA genes, and then synthesizing 1,192 tRNA genes from phylogenetically diverse plasmids, transposable elements and bacteriophages infecting members of the Enterobacteriaceae family (Supplementary Data 1). Next we assayed these tRNAs for their ability to produce functional tRNAs in an *E. coli* host and substitute genomic tRNA genes to translate TCR codons. As depicted in Fig. 11a, this high-throughput assay is based on an *E. coli* strain with a synthetic recoded genome in which all annotated instances of two sense serine codons (TCG, TCA) and a stop codon (TAG) were replaced with synonymous alternatives, and the corresponding serU, serT tRNA genes and RF1 (prfA) have been deleted. This strain, *E. coli* Syn61Δ3, thereby relies on a 61-codon genetic code and prevents the expression of protein-coding genes containing TCR codons. Candidate tRNAs have been synthesized and cloned into a plasmid carrying each tRNA under a strong constitutive promoter together with an nptII40TCA,68TCG,104TCA,251TCG aminoglycoside O-phosphotransferase antibiotic-resistance gene containing TCA codons at positions 40 and 104 and TCG codons at positions 68 and 251. In wild-type *E. coli* cells bearing the canonical genetic code, nptII40TCA,68TCG,104TCA,251TCG confers resistance to kanamycin through serine incorporation at positions 40, 68, 104 and 251, and the production of full-length APH(3')-II aminoglycoside O-phosphotransferase. In Syn61Δ3, however, the production of this resistance-conferring gene product is inhibited owing to the lack of serU- and serT-encoded tRNASer(UGA) and tRNASer(CGA) needed for TCR-codon decoding. Therefore, in our screen, only plasmid variants that are expressing tRNAs capable of decoding TCR codons will survive kanamycin selection. The transformation of this plasmid library into Syn61Δ3 and subsequent selection in the presence of kanamycin yielded thousands of colonies, indicating the presence of TCR-translating tRNAs in our library. Pooled extraction of plasmid variants from kanamycin-resistant colonies followed by amplicon sequencing of their tRNA insert identified 62 tRNA sequences capable of promoting nptII40TCA,68TCG,104TCA,251TCG expression (Fig. 11b and Supplementary Data 1). These tRNAs represent 89% of all predicted TCR-codon-recognizing tRNAs in our library and share (Zürcher *et al.*, 2022) 7–61.1% (median = 46.2%) similarity to the endogenous serU tRNA of *E. coli*. In agreement with the anticodon composition of mobile serine tRNAs, most tRNA hits contained a UGA anticodon and carried the identity elements necessary for recognition by the host's SerS serine-tRNA ligase (Fig. 11b).

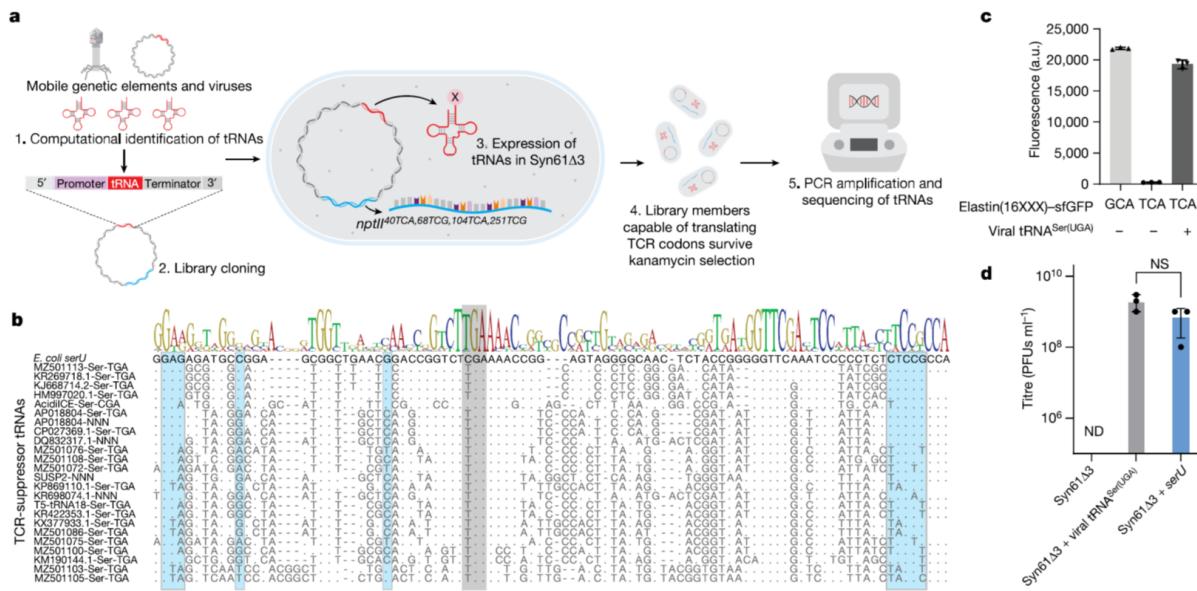


Figure 11: a, We screened the mobile tRNAome for tRNAs that can simultaneously translate TCA and TCG (together TCR) codons by computationally identifying tRNA genes in mobile genetic elements (1) and then synthesizing select candidates as an oligonucleotide library and cloning these variants into a plasmid vector carrying a *nptII*40TCA,68TCG,104TCA,251TCG marker (conferring kanamycin resistance) (2). Following the transformation of this library into *Syn61Δ3* (3), in which the deletion of *serU* (encoding tRNASer(CGA)) and *serT* (encoding tRNASer(UGA)) makes TCG and TCA codons unreadable, only variants carrying functional TCR-suppressor tRNAs survive kanamycin selection (4). Finally, high-throughput sequencing of the tRNA inserts from kanamycin-resistant clones identified suppressor tRNAs (5). b, Multiple sequence alignment of mobile TCR-codon-translating tRNAs. Grey shading indicates the anticodon region; the host's serine–tRNA ligase identity elements are shown in blue. c, Viral serine tRNAUGA translates the TCA codon. *Syn61Δ3* expressing elastin(16GCA(alanine))-sfGFP–His6 served as a wild-type expression control, and the elastin(16TCR)-sfGFP–His6 expression was compared with and without the coexpression of the tRNASer(UGA) of the *Escherichia* phage IrisVonRoten (Maffei *et al.*, 2021). XXX represents the analysed codon, TCA or GCA. Bar graph represents the mean; error bars represent s.d. based on $n = 3$ independent experiments; a.u., arbitrary fluorescence units. d, The expression of viral TCR-suppressor tRNAs and *serU* (tRNASer(CGA)) restores the replication of the T6 bacteriophage in *Syn61Δ3*. Circles represent data from $n = 3$ independent experiments, error bars represent s.d., and the bar graph represents the mean. ND, below the detection limit (that is, $\geq 10^3$ plaque-forming units (PFUs) ml⁻¹); NS, not significant ($P = 0.116$) based on unpaired two-sided Student's t-test.

Notable examples include the UAG-anticodon-containing serine tRNA of the laboratory model coliphage T5, tRNAs from plasmids of multidrug-resistant *E. coli* isolates (GenBank IDs AP018804 and CP023851) and the Ser-tRNACGA of the integrative conjugative element of Acidithiobacillus ferrooxidans. The presence of mobile tRNAs in integrative conjugative elements is espe-

cially concerning as these mobile genetic elements can carry up to 38 tRNAs corresponding to all 20 amino acids in a single operon and are capable of excision and transfer into neighbouring bacterial cells (Alamos *et al.*, 2018; Bustamante, 2012). In agreement with the findings of previous studies (Bowden *et al.*, 1997; Fujino *et al.*, 2020; Abrahão *et al.*, 2018), our computational screen also showed that mobile tRNA genes are not limited to mobile genetic elements of bacteria. Computational analysis of viruses infecting vertebrates and archaea revealed the presence of sense- and stop-codon-suppressor tRNAs in both groups, suggesting that mobile tRNAs are prevalent across viruses infecting prokaryotic, archaeal and eukaryotic hosts (Supplementary Data 2).

We confirmed the predicted serine amino acid identity of the TCR-codon-recognizing tRNAs by coexpressing a selected tRNA hit with an elastin(16TCA)-sfGFP-His6 construct harbouring a single TCA codon at position 16. The coexpression of the tRNASer(UGA) of the *Escherichia* phage IrisVonRoten (Maffei *et al.*, 2021) together with the elastin(16TCA)-sfGFP-His6 construct conferred near wild-type level expression (Fig. 111c), and tryptic digestion followed by reverse-phase liquid chromatography and tandem mass spectrometry (LC–MS/MS) confirmed serine incorporation at the TCA position (Extended Data Fig. 11a).

Next, we investigated whether mobile tRNAome-derived tRNAs could promote viral replication. A previous study demonstrated that Syn61Δ3 resists infection by multiple bacteriophages, including the Enterobacteria phage T6 (ref. 4). Infecting Syn61Δ3 with the T6 phage recapitulated these results. By contrast, the infection of Syn61Δ3 harbouring a bacteriophage-derived Ser-tRNAUGA gene with T6 resulted in rapid lysis, indicating that tRNA genes that reside in viral genomes can substitute cellular tRNAs and promote phage infection (Fig. 11d).

The discovery of diverse TCR-codon-translating tRNAs on horizontally transferred genetic elements indicates that mobile tRNA genes are widespread and can readily complement the lack of cellular tRNAs to promote viral replication and HGT.

Viruses infecting a recoded organism

We next investigated whether lytic viruses of Syn61Δ3 exist. We infected Syn61Δ3 cells with 11 coliphages whose genomes harbour TCR-translating tRNA genes on the basis of our plasmid-based screen (Fig. 11b). Notably, none of these eleven phages could overcome the recoded host's genetic isolation, indicating that the presence of tRNA genes on viral genomes does not directly rescue viral replication in recoded organisms (Extended Data Fig. 11b).

We next attempted to isolate lytic viruses from diverse environmental samples by carrying out a standard two-step enrichment-based phage isolation protocol and using Syn61Δ3 as the host. First, bacteria-free filtrates of environmental and wastewater samples ($n = 13$, from Massachusetts (USA), Extended Data Table 1a) were mixed with Syn61Δ3 and grown until station-

ary phase. Next, bacterial cells were removed, and we analysed the presence of lytic phages by mixing sample supernatants with Syn61Δ3 in soft-agar overlays. Five samples produced visible lysis. Viral plaque isolation from these samples followed by DNA sequencing and de novo genome assembly identified 12 new phage strains. All identified phages belong to the Caudovirales order and the Myoviridae family, taxa rich in tRNA-encoding bacteriophages⁸ (Extended Data Table 1b). Computational identification of tRNA genes revealed the presence of tRNA operons in all phage isolates, with 10 to 27 tRNA genes in each genome (Supplementary Data 1). Notably, all isolates harboured TCR-codon-translating serine tRNAs with a UGA anticodon that we identified in our earlier nptII40TCA,68TCG,104TCA,251TCG suppressor screen (Fig. 11b). One isolate, REP1, also harboured a predicted homing endonuclease (Fig. 12c). Homing endonucleases encoded in tRNA operons have been shown to be responsible for the horizontal transfer of tRNA gene clusters (Brok-Volchanskaya *et al.*, 2008). Phage isolates showed more than two orders of magnitude difference in viral titres after replication on recoded cells (Fig. 12a). One of the most virulent isolates, REP12, required 60 min to complete a replication cycle at 37 °C in Syn61Δ3 (Fig. 12b).

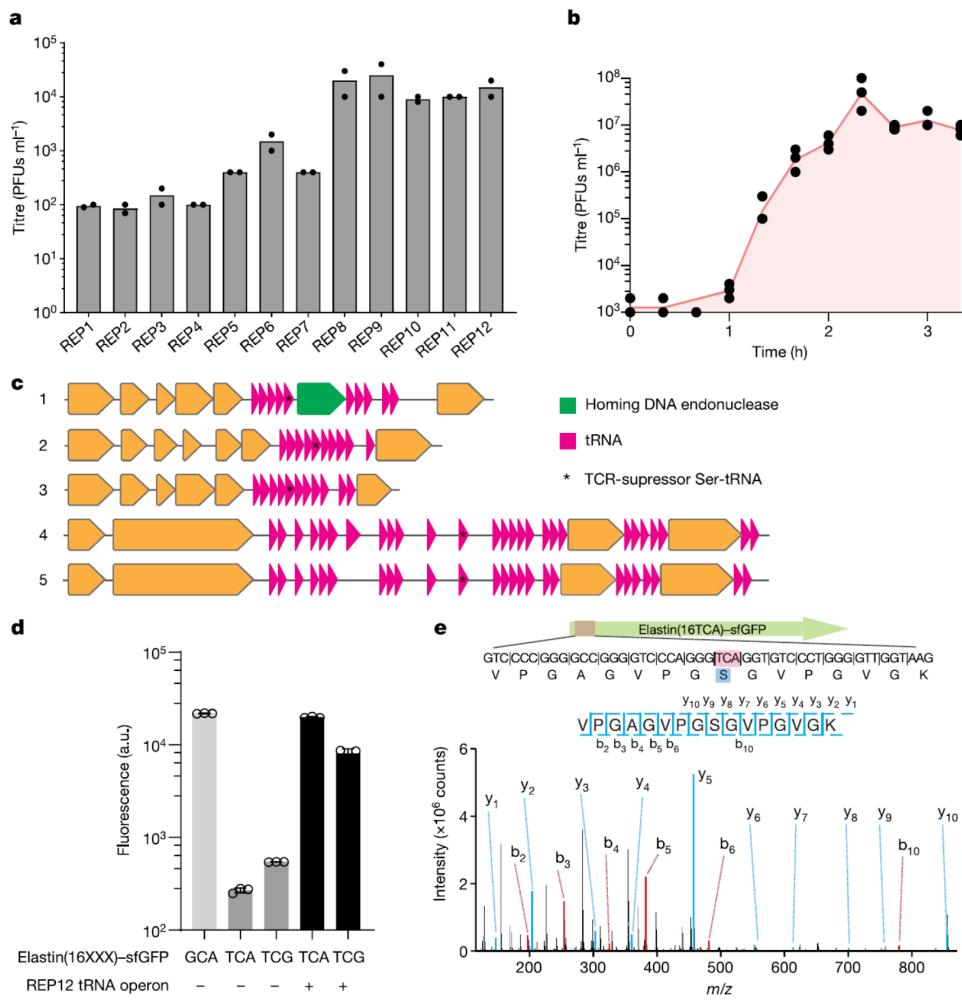


Figure 12: a, Titre of Syn61Δ3 phage isolates after replication on Syn61Δ3. Circles represent data from $n = 2$ independent experiments; bar graphs represent the mean. b, Single-step growth curve of the REP12 lytic Syn61Δ3 phage. Single-step growth was carried out in $n = 3$ independent experiments; red line represents the mean, and circles represent the total viral titre. c, Genomic maps of tRNA operons in lytic Syn61Δ3 phages. Magenta arrows represent predicted tRNA genes; asterisks denote tRNA genes identified in our earlier TCR codon suppressor screen (Fig. 11b); the green arrow represents a homing endonuclease gene, and orange arrows represent protein-coding genes. Phage operon numbers correspond to the following REP phages—1, REP1; 2, REP2; 3, REP4; 4, REP6; 5, REP12. d, tRNAs expressed by the viral tRNA operon translate TCR codons. Syn61Δ3 expressing elastin(16GCA(alanine))-sfGFP-His6 served as a wild-type expression control, and the elastin(16TCR)-sfGFP-His6 expression was compared with and without the coexpression of the REP12 viral tRNA operon. XXX represents the analysed codon, TCA, TCG or GCA. a.u., arbitrary fluorescence units; bar graphs represent the mean. Error bars represent s.d. based on $n = 3$ independent experiments. e, tRNAs expressed by the viral tRNA operon decode TCR codons as serine. The amino acid identity of the translated TCA codon within elastin(16TCA)-sfGFP-His6 was confirmed by MS/MS from Syn61Δ3 cells containing the REP12 tRNA operon and its cognate promoter. The figure shows the amino acid sequence and MS/MS spectrum of the analysed elastin(16TCR) peptide. MS/MS data were collected once.

The isolated viral strains infecting Syn61Δ3 show that bacteriophages that can overcome viral resistance based on sense codon recoding exist and are widespread in environmental samples. However, the presence of TCR-codon-translating serine tRNAs on viral genomes does not directly rescue viral replication in the recoded strain.

Viral tRNAs substitute cellular tRNAs We next investigated how tRNA-encoding viruses evade genetic-code-based resistance. Time-course transcriptome analysis of REP12-phage-infected Syn61Δ3 cells during the viral replication cycle revealed early and high-level expression of the viral tRNA operon (Supplementary Fig. 19). In agreement with this observation, the computational prediction of bacterial promoters driving the tRNA array indicated the presence of multiple strong constitutive promoters upstream of the tRNA operon region. We then investigated the time-course kinetics of tRNA expression in Syn61Δ3 cells that were infected with our REP12 phage by carrying out tRNA sequencing (tRNA-seq). Time-course tRNA-seq experiments revealed remarkably high-level expression of the viral tRNASer(UGA) immediately after phage attachment (that is, a relative viral tRNASer(UGA) abundance of 56.1% ($\pm 5\%$) compared to the host serV tRNA; Extended Data Fig. 20). Throughout the entire phage replication cycle, the phage tRNASer(UGA) remained one of the most abundant viral tRNA species inside infected Syn61Δ3 cells (Extended Data Fig. 20). We next investigated whether phage tRNASer(UGA) participates in translation by analysing the presence of their mature form. The gene encoding the tRNASer(UGA) in the genome of REP12 does not encode the universal 5'-CCA tRNA tail, which allows for amino acid attachment as well as for interaction with the ribosome. Therefore, CCA tail addition must happen before these tRNAs can participate in translational processes. The sequencing-based analysis of phage tRNASer(UGA) ends detected CCA tail addition in 62.9% ($\pm 1.9\%$) of all tRNA-seq reads immediately after phage attachment, indicating that mature tRNASer(UGA) is instantly produced after host infection (Extended Data Fig. 22).

We also investigated transcriptomic changes in Syn61Δ3 during phage replication. Analysis of the host transcriptome after phage infection revealed upregulation in genes responsible for tRNA maturation and modification. Upregulated genes include queG, encoding epoxyqueuosine reductase (which catalyses the final step in the de novo synthesis of queuosine in tRNAs) (Miles *et al.*, 2011), and trmJ, encoding tRNA Cm32/Um32 methyltransferase (Liu *et al.*, 2015) (which introduces methyl groups at the 2'-O position of U32 of several tRNAs, including tRNASer(UGA)), suggesting the potential post-transcriptional modification of phage-derived tRNAs (Extended Data Fig. 23).

Finally, we also validated the role of phage tRNASer(UGA) tRNAs in decoding TCR codons. We first cloned the REP12 viral tRNA operon containing the hypothetical tRNASer(UGA) and its predicted promoter into a plasmid vector. Coexpression of this tRNA operon with an

elastin(16TCA)-sfGFP-His6 and elastin(16TCG)-sfGFP-His6 construct, harbouring either a single TCA or TCG codon at position 16, respectively, resulted in high-level elastin-sfGFP-His6 expression (Fig. 12d). Next, tryptic digestion followed by LC/MS–MS analysis confirmed serine incorporation in response to both the TCA and TCG codon in these elastin(16TCR)-sfGFP-His6 samples (Fig. 12e and Supplementary Fig. 20). As expected, the coexpression of the same elastin(16TCA)-sfGFP-His6 construct with the only tRNASer(UGA) of the viral tRNA operon conferred a similar effect, and LC/MS–MS analysis confirmed the role of this tRNA in decoding viral TCR codons as serine (Supplementary Fig. 20).

Together these results show that lytic phages of Syn61Δ3 overcome genetic-code-based viral resistance by rapidly complementing the cellular tRNA pool with virus-encoded tRNAs.

Swapped genetic code resists viruses

We predicted that establishing an artificial genetic code, in which TCR codons encode an amino acid different from their natural serine identity, would create a genetic firewall that safeguards cells from HGT and infection by tRNA-encoding viruses. In an amino acid-swapped genetic code, viral tRNAs would compete with host-expressed tRNAs that decode TCR codons as a non-serine amino acid resulting in the mistranslation of viral proteins. As the genetic isolation provided by amino acid-swapped genetic codes is expected to correlate with the hydrophobicity and polarity distance between exchanged amino acids (Schmidt & Budisa, 2020), we sought to reassign TCR serine codons to leucine, the amino acid most distant from the natural serine meaning of the TCR codon. The more distant tryptophan, phenylalanine, tyrosine and isoleucine exchanges were not investigated as their cognate aaRSs are nonpermissive to anti-codon mutations and expected to require extensive enzyme and/or tRNA engineering to function (Yang *et al.*, 2006; Kobayashi *et al.*, 2003; Giegé *et al.*, 1998). Although alternative forms of swapped codes emerged in the past, these codes either targeted the exchange of chemically more similar amino acids (that is, serine to alanine, histidine and proline) or remained tested only *in vitro* (Fujino *et al.*, 2020; Schmidt & Kubyshkin, 2021; Church *et al.*, 2007; Zürcher *et al.*, 2022).

To establish a serine(TCR)-to-leucine swapped genetic code (Fig. ??a), we utilized Syn61Δ3, which genome-wide lacks annotated instances of TCR codons and their corresponding tRNA genes, and sought to identify tRNAs capable of efficiently translating TCR codons as leucine. We modified our high-throughput tRNA library screen (Fig. 11a) to evolve TCR suppressors from the endogenous *E. coli* leuU tRNA carrying a TCA and TCG decoding anticodon. We coexpressed a 65,536-member mutagenized library of the anticodon-swapped leuU tRNA gene in which the anticodon loop of both tRNAs has been fully randomized, together with aph3Ia29×Leu-to-TCR, a kanamycin-resistance-conferring gene in which all 29 instances of

leucine codons were replaced with TCR serine codons. In this system, only anticodon-swapped leuU variants capable of simultaneously translating all 29 TCR codons as leucine would confer resistance to kanamycin. High kanamycin selection pressure in combination with 29 instances of TCR codons in aph3Ia was expected to select tRNA variants that provide wild-type-level translation efficiency for TCR codons. We identified two distinct leuU variants by applying ‘high’ kanamycin concentration (that is, 200 µg ml⁻¹) as selection pressure to Syn61Δ3 cells carrying the anticodon-swapped tRNA library. These variants, carrying tRNAs containing distinct anticodon loop mutations (Extended Data Fig. 23a), were then infected with a cocktail of all 12 phage isolates (Extended Data Table 1) that are capable of lysing Syn61Δ3 at a 10:1 cell-to-phage ratio (that is, a multiplicity of infection of 0.1). Instead of plaque and efficiency-of-plating assays (Loś *et al.*, 2008; Abedon & Yin, 2009) with selected phage isolates and assaying viral titre on the resistant strain, we quantified virus resistance using a mixture of all 12 Syn61Δ3-infecting viral isolates, and measured virion release from infected cells after 24 h by plating culture supernatants on a virus-susceptible wild-type *E. coli* host. This can reveal even slow or non-plaque-forming viral replication in the target cells^{35,36} (Supplementary Note).

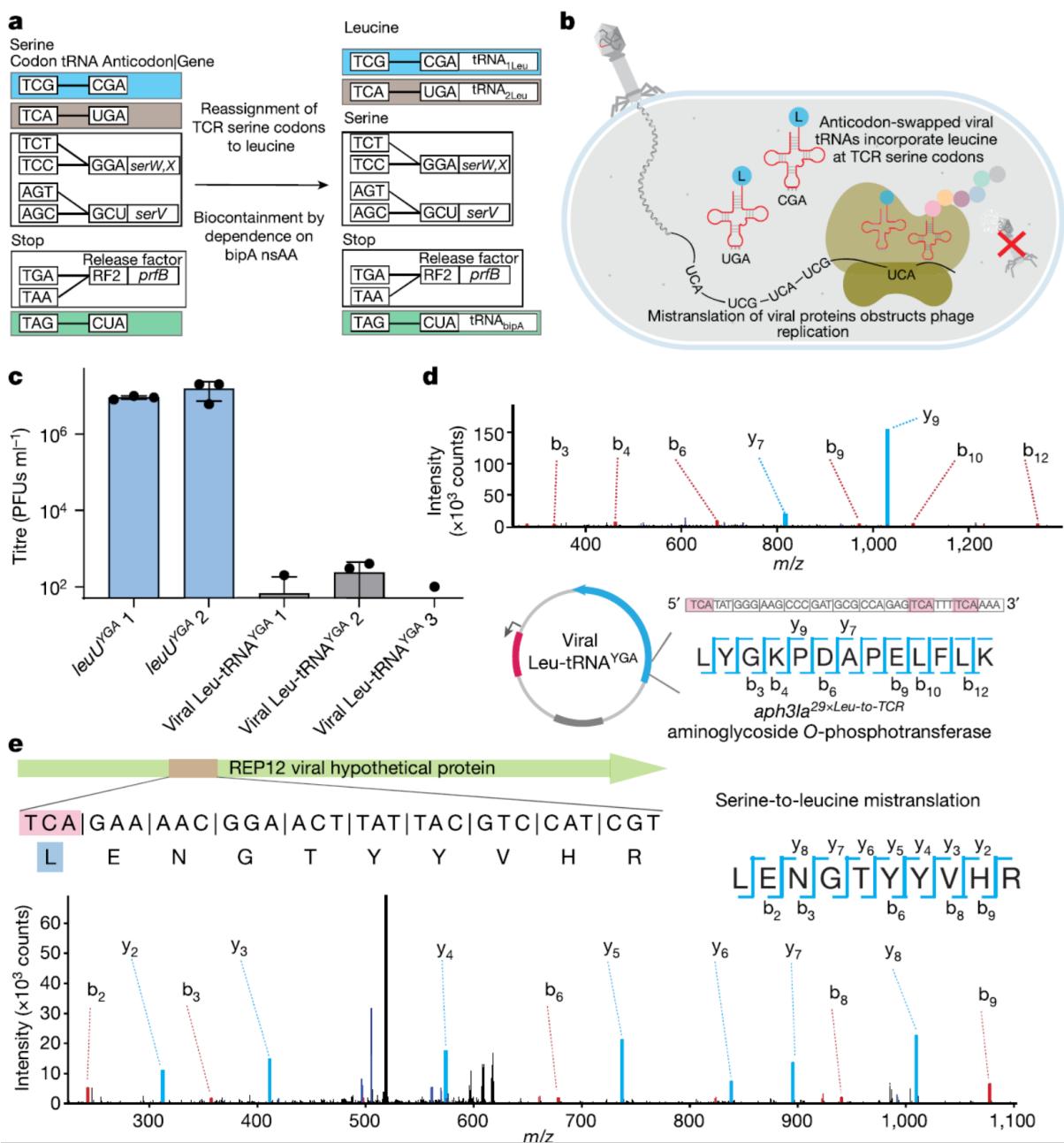


Figure ??:a, The creation of a genomically recoded organism, Ec_Syn61Δ3-SL, from *E. coli*, in which both TCA and TCG—naturally serine-meaning—codons are translated as leucine. The introduction of bacteriophage-derived Leu-tRNA_{UGA} and Leu-tRNA_{CGA} to Syn61Δ3 reassigns TCA and TCG codons to leucine, and the reassignment of the TAG stop codon to encode bipA in an essential gene of the host ensures the biocontainment of Ec_Syn61Δ3-SL. b, Schematic of viral infection in Ec_Syn61Δ3-SL. The reassignment of the sense codons TCA and TCG to leucine in Ec_Syn61Δ3-SL provides multi-virus resistance by mistranslating the viral proteome. c, Bacteriophage-

derived Leu-tRNAUGA and Leu-tRNACGA expression in Syn61Δ3 provides multi-virus resistance. The figure shows the titre of lytic Syn61Δ3 phages following the infection of the corresponding Leu-tRNAYGA-expressing Syn61Δ3 strain with a mixture of 12 distinct REP Syn61Δ3 phages. All experiments were carried out in three independent replicates; circles represent data from $n = 3$ independent experiments; bar graphs represent the mean; the error bars represent s.d. d, The reassignment of TCR codons to leucine within the coding sequence of aph3Ia29×Leu-to-TCR in Syn61Δ3-LS was confirmed by MS/MS. The figure shows the amino acid sequence and MS/MS spectrum of the detected aph3Ia29×Leu-to-TCR peptide and its coding sequence. MS/MS data were collected once. e, Mistranslated viral protein synthesis in Ec_Syn61Δ3-SL. The figure shows the amino acid sequence and MS/MS spectrum of a bacteriophage-expressed protein, together with its viral genomic sequence, in which the naturally serine-coding TCA codon is mistranslated as leucine. The experiment was carried out by infecting Ec_-Syn61Δ3-SL cells, expressing Leu9-tRNAYGA from the Escherichia phage OSYSP, with the REP12 phage, and the proteome of infected cells was analysed by MS/MS. MS/MS data were collected once. Notably, all selected

leuU library members allowed robust phage replication, with phage titres reaching about 10⁷ plaque-forming units per millilitre after 24 h (Fig. ??c). We reasoned that viral replication in the presence of TCR-suppressing leuU variants is due to the lower suppression efficiency of their tRNAs compared to that of competing phage-carried serine tRNAs, which leads to rapid viral takeover. Viral tRNASer(YGA) species, that are tRNASer(UGA) and tRNASer(CGA), might: have higher aminoacylation efficiency by their corresponding *E. coli* aminoacyl-tRNA ligase than our selected leuU variants; have higher affinity towards the bacterial ribosome; and/or better evade phage- and host-carried tRNA-degrading effector proteins (Yang *et al.*, 2021; Wang *et al.*, 2022; Tomita *et al.*, 2000).

On the basis of this observation, we reasoned that bacteriophage-encoded tRNAs might provide much higher suppression efficiencies for their cognate codons than their native *E. coli* counterpart. Therefore, and to exploit this superior translation efficiency of bacteriophage-encoded tRNAs and their resistance to phage-carried tRNA-degrading enzymes, we next constructed a small, focused library that coexpressed YGA-anticodon-swapped mutants of 13 phage-encoded leucine tRNAs, together with the aph3Ia29×Leu-to-TCR aminoglycoside O-phosphotransferase gene. The transformation of this library into Syn61Δ3 cells and subsequent ‘high’-concentration (that is, 200 µg ml⁻¹) kanamycin selection identified three distinct tRNAs enabling robust growth. Identified tRNAs showed only 48.3–37.9% similarity to *E. coli* leuU but carried most of the canonical *E. coli* leucine-tRNA ligase identity elements (Extended Data Fig. 23b). Furthermore, the analysis of the total tRNA content of these cells by tRNA-seq confirmed the presence of synthetic phage Leu-tRNAYGA tRNAs with similar abundances to those of the cellular endogenous serine tRNAs (that is, a relative expression level of 172% and 140% for Leu-tRNAUGA and Leu-tRNACGA, respectively, compared to serV (Extended Data Fig. 24)).

Next, as in our previous infection assay, phage-tRNAYGA-expressing cells were infected with a mixture of 12 distinct, lytic phages of Syn61Δ3 at a multiplicity of infection of 0.1. The analysis of phage titre in culture supernatants after 24 h showed a marked drop compared to the input phage inoculum, suggesting that anticodon-swapped viral leucine tRNAs block phage replication (Fig. ??c).

We then investigated the mechanism of phage resistance in *E. coli* cells carrying virus-derived tRNA_{Leu}(YGA) tRNAs (Ec_Syn61Δ3 serine-to-leucine swap, or Ec_Syn61Δ3-SL in short; Fig. ??b) by carrying out total proteome analysis. Untargeted deep proteome analysis of uninfected cells by MS/MS validated the translation of TCR codons as leucine in Ec_Syn61Δ3-SL (Fig. ??d). Time-course untargeted proteome analysis after bacteriophage infection revealed extensive mistranslation at TCR codons in newly synthesized phage proteins (Fig. ??e and Supplementary Fig. 21), indicating that an amino acid-swapped genetic code broadly obstructs viral protein synthesis. In agreement with earlier reports that showed the partial recognition of TCT codons by tRNA_{UAGA} (Takai *et al.*, 1999; Kunjapur *et al.*, 2021) and owing to the extreme sensitivity of our untargeted proteomics assay, serine-to-leucine mistranslation was also detected at TCT codon positions in Ec_Syn61Δ3-SL cells (Extended Data Fig. 25). The recognition of TCT codons by phage tRNA_{Leu}(YGA) tRNAs might also be responsible for the fitness decrease of Ec_Syn61Δ3-SL cells compared to those of its ancestor strain (that is, a doubling time of 69.3 min, compared to 44.29 min for the parental Syn61Δ3 strain in rich 2×YT medium (Extended Data Fig. 26)). Alternatively, the fitness decrease of Ec_Syn61Δ3-SL might also be attributable to the presence of TCR codons in essential genes of Syn61Δ3. According to our genome analysis, at least four essential genes of Syn61Δ3 (*mukE*, *ykfM*, *yjbS* and *safA*) contain TCR codons and become mistranslated in Ec_Syn61Δ3-SL (Supplementary Data 4).

Finally, we also sought to develop a tightly biocontained version of Ec_Syn61Δ3-SL because a virus-resistant strain might have a competitive advantage in natural ecosystems due to the lack of predating bacteriophages. Synthetic auxotrophy based on the engineered reliance of essential proteins on human-provided nonstandard amino acids (nsAAs; for example, L-4,4-biphenylalanine (bipA)) offers tight biocontainment that remains stable under long-term evolution (Mandell *et al.*, 2015; Kunjapur *et al.*, 2021; Rovner *et al.*, 2015). Therefore, we generated a recombination-deficient (that is, ΔrecA), biocontained version of Ec_Syn61Δ3-SL bearing a bipA-dependent essential adk gene and the bipA aminoacyl-tRNA synthetase–tRNAbipA(CUA) system by first carrying out adaptive laboratory evolution on a recA-knockout Syn61Δ3, and then replacing the genomic adk copy with its bipA-dependent variant (Nyerges *et al.*, 2016) (Methods). This strain maintained the low escape frequency of previously reported single-gene synthetic auxotrophs10 (that is, 2.9×10^{-6} ($\pm 5.9 \times 10^{-7}$) escapees per colony-forming units

(CFUs) escape frequency) and provided robust growth. We also tested the viral resistance of Ec_Syn61Δ3-SL under mock environmental conditions by repeating our phage enrichment and isolation process with a mixture of 12 environmental samples, including fresh sewage (Extended Data Table 1 and Supplementary Note), but could not detect plaque-forming phages in culture supernatants (Extended Data Fig. 28).

Together, these results demonstrate that reassigning the sense codons TCA and TCG to leucine *in vivo* provides broad protection against viruses, including new mixtures of viruses directly from environmental samples, and the TAG stop codon can be simultaneously utilized to bio-contain this virus-resistant strain through dependence on an amino acid not found in nature.

Genetic-code-based bidirectional firewall Finally, we developed a set of plasmid vectors that we systematically addicted to an amino acid-swapped genetic code in which leucine is encoded as TCR codons. GMOs are increasingly deployed for large-scale use in agriculture, therapeutics, bioenergy and bioremediation. Consequently, it is critical to implement robust biocontainment strategies that prevent the unintended proliferation of GMOs and protect natural ecosystems from engineered genetic information. Although efficient biocontainment strategies exist (for example, bipA nsAA-based synthetic auxotrophy, as in Ec_Syn61Δ3-SL), current methods fail to prevent the HGT-based escape of engineered genetic information. Synthetic addiction to a swapped genetic code offers a solution to this problem. Using our Ec_Syn61Δ3-SL cells expressing virus-derived tRNA_{Leu}(YGA), we therefore developed a set of plasmid vectors that depend on TCR codons to express leucine-containing proteins and thus can function only in cells that efficiently translate TCR codons as leucine (Fig. 13a). These plasmids, called the pLS plasmids, offer four widely used, orthogonal antibiotic-resistance markers in combination with four mutually orthogonal low- to high-copy-number origins of replication for stable maintenance in Ec_Syn61Δ3-SL cells (Fig. 13b and Supplementary Data 3). Antibiotic-resistance genes and proteins necessary for pLS plasmid replication encode leucine as TCR—naturally serine-meaning—codons and, therefore, fail to function in cells bearing the canonical genetic code. The addiction of resistance markers and replication proteins to an artificial genetic code ensures that pLS plasmids can stably and safely maintain synthetic genetic functions but restrict the functionality of these genes to Ec_Syn61Δ3-SL cells. Using our pLS plasmids, we achieved serine-to-leucine codon reassignment simultaneously at about 14,000 codon positions in a single cell, thus demonstrating the exceptional codon reassignment efficiency of our viral tRNA-based system.

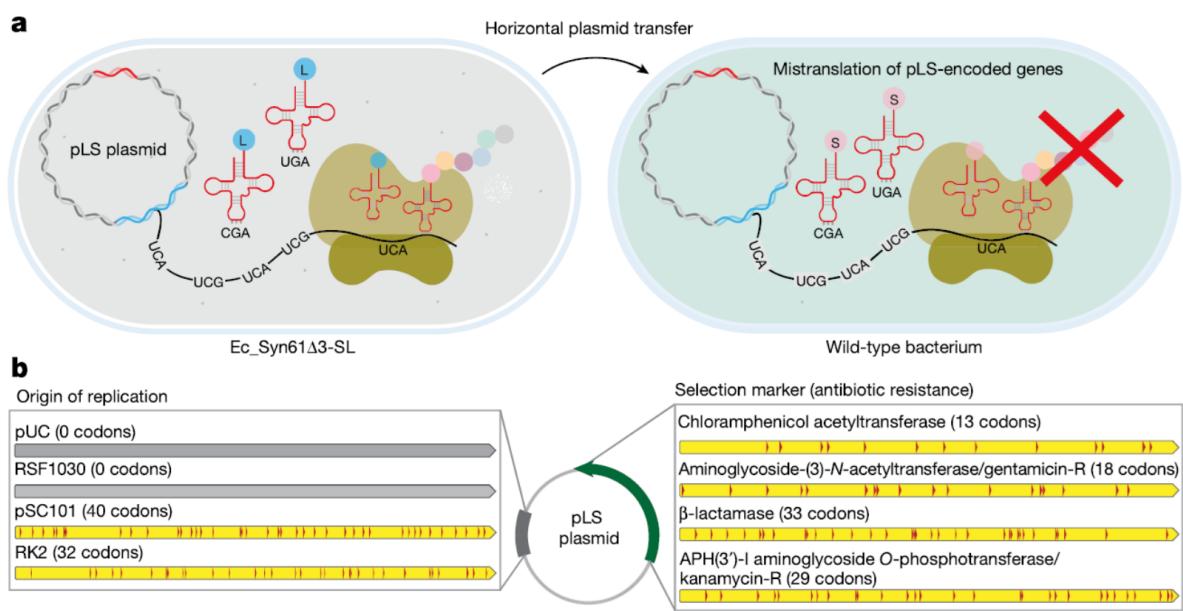


Figure 13: a, We developed a set of plasmid vectors, termed the pLS plasmids, that rely on TCR codons to express leucine-containing proteins. pLS plasmids function only in Ec_Syn61 Δ 3-SL expressing bacteriophage-derived synthetic tRNA $\text{Leu}(\text{YGA})$ tRNAs, and the encoded proteins of pLS plasmids become mistranslated in cells bearing the canonical genetic code. b, The pLS plasmids offer multiple mutually orthogonal antibiotic-resistance markers together with low- to high-copy-number origins of replication that are addicted to an artificial genetic code in which leucine is encoded as TCR codons. The number in parenthesis marks the number of Leu(TCR) codons in each gene. Detailed sequence information and a description of the pLS plasmids are available in Supplementary Data 3.

We then tested the ability of our pLS vectors to function in cells bearing the standard genetic code by electroporating plasmid DNA from six variants (that is, pLS1–6) into wild-type *E. coli* K12 MG1655 cells but could not detect escapees carrying pLS plasmids (that is, an electroporation efficiency of ≈ 1 CFU per microgram, whereas the electroporation efficiency of the recipient cells was $3 \times 10^9 \pm 1.1 \times 10^9$ CFUs per microgram (\pm s.d., $n = 3$ independent experiments) based on a pUC-KanR control plasmid). The escape of pLS plasmids was similarly prevented when the expression cassette for phage tRNA $\text{Leu}(\text{YGA})$ was encoded within the plasmid backbone (that is, pLS1 and pLS2), indicating that anticodon-swapped viral tRNAs are severely toxic to wild-type cells. On the basis of these results, we also expect that, similarly to genes of pLS, any leucine-requiring protein, up to entire chromosomes, can be addicted to Ec_Syn61 Δ 3-SL by recoding target genes to encode one or more leucine positions as TCR codons.

In sum, the addiction of pLS plasmids to an artificial genetic code in which leucine is encoded as TCR codons, in combination with nsAA-based synthetic auxotrophy, promises escape-free biocontainment for engineered genetic information.

4.2.5 Discussion

Previous works provided support for using rational genetic code engineering to isolate GMOs from natural ecosystems by preventing viral infections and HGT^{2,3,4}; however, how natural viruses and transferred genes could breach genetic-code-based resistance remained unanswered. By systematically investigating HGT into *E. coli* Syn61Δ3, a synthetic organism with a compressed genetic code⁴, we demonstrated that tRNAs expressed by bacteriophages, plasmids and integrative conjugative elements readily substitute cellular tRNAs and abolish genetic-code-based isolation. We discovered 12 viruses in environmental samples that can infect *E. coli* Syn61Δ3, despite its compressed genetic code (Fig. 12a). These bacteriophages express up to 27 tRNAs, including a functional tRNASer(UGA) needed to replace the host's deleted tRNAs. These findings suggest that the selection pressure posed by compressed genetic codes can facilitate the rapid evolution of viruses and mobile genetic elements capable of crossing a genetic-code-based barrier. This hypothesis is further supported by the co-localization of tRNAs and homing endonucleases (Fig. 12c) that catalyse copy-and-paste tRNA operon transfer²⁵. We have also shown that mobile tRNAs are not limited to bacteria, as multiple archaeal and eukaryotic viruses also carry predicted tRNA genes.

Next, as a combined solution against HGT and viral infections, including tRNA-expressing viruses, we have created a new type of biocontained *E. coli* strain, Ec_Syn61Δ3-SL, carrying an amino acid-swapped genetic code. Ec_Syn61Δ3-SL achieves unprecedented gene-transfer resistance by the reassignment of TCR codons to leucine—an amino acid distant in physicochemical properties compared to their natural serine identity—using reprogrammed, virus-derived tRNAs. In contrast to previous studies^{2,3,4,33}, our study confirmed the virus resistance of this strain using a broad range of viruses. We found that viral tRNAs provide exceptional codon reassignment efficiency and are superior for establishing altered genetic codes compared to cellular tRNAs. We also showed that the swapped genetic code of Ec_Syn61Δ3-SL simultaneously prevents viral replication and the escape of synthetic genetic information into wild organisms. As a direct application, we addicted the most widely used plasmid vectors to express leucine-containing proteins with TCR codons and developed plasmids that cannot function in natural organisms. By demonstrating efficient serine-to-leucine reassignment, we made biological containment possible at the level of genes, to operons, to entire chromosomes. Adaptive laboratory evolution, which has substantially improved doubling times of organisms with altered genetic

codes (Robertson *et al.*, 2021; Wannier *et al.*, 2018), will be applied to Ec_Syn61Δ3-SL to help move towards industrial use.

Potential limitations of our work include the sampled viral diversity and the inability of our viral resistance assays to sample multi-step, long-term evolutionary processes of viruses. Although our tests of the viral resistance of Ec_Syn61Δ3-SL included 12 lytic Syn61Δ3 phages and a complex mixture of environmental samples, including fresh sewage, we cannot rule out the existence of Ec_Syn61Δ3-SL-infecting phages in Earth's biome. Our assays provided an example of Ec_Syn61Δ3-SL culture contamination with soil, sewage and faecal material (Extended Data Table 1a); however, lysogenic viruses or jumbo phages and megaphages, with genomes of more than 200 and 500 kilobase pairs, respectively, could have remained undetected in our assays owing to their frequent inability to form visible plaques (Al-Shayeb *et al.*, 2020; Serwer *et al.*, 2007). We reason that bacteriophages with few TCR codons in essential genes (for example, *Escherichia* phage EC6098 with only 33 TCR positions in its 6 protein-coding genes (Kirchberger & Ochman, 2020)) have the highest potential to overcome an amino acid-swapped code. Alternatively, virus-expressed tRNA-degrading proteins (Wang *et al.*, 2022; Tomita *et al.*, 2000) could evolve to selectively destroy mistranslating tRNAs and thus promote viral escape; however, both mechanisms would require multiple simultaneous mutations and thus are unlikely to occur. Future work will explore whether such viruses can evade the amino acid-swapped genetic code of Ec_Syn61Δ3-SL.

After our manuscript was publicly submitted, another study³³ published findings consistent with some of ours, identifying two tRNAs that could overcome the virus resistance of Syn61Δ3 and noting the utility of swapped genetic codes to obstruct HGT. The use of weak *E. coli* tRNAs was insufficient to produce true resistance to a broad set of viruses in our experiments. The apparent success with such weak tRNAs in (Zürcher *et al.*, 2022) may be due to testing only two (related) viral strains and using efficiency-of-plating (Loś *et al.*, 2008; Abedon & Yin, 2009) and conjugation assays (Supplementary Note). Furthermore, the lower expression level of mistranslating tRNAs used in (Zürcher *et al.*, 2022), in combination with the use of *E. coli* tRNAs still susceptible to virus-expressed tRNA-degrading enzymes (Wang *et al.*, 2022; Tomita *et al.*, 2000), might result in compromised resistance in the presence of complex environmental viral communities. The authors of (Zürcher *et al.*, 2022) did not implement biocontainment strategies to limit the unwanted proliferation of their potentially virus-resistant host.

These findings fundamentally affect ongoing prokaryotic and eukaryotic genome recoding projects (Ostrov *et al.*, 2016; Chen *et al.*, 2022; Boeke *et al.*, 2016; Dai *et al.*, 2020), including our aim to engineer a virus-resistant stop-codon-recoded human cell line and a 57-codon strain of *E. coli*, as some of the identified tRNAs are expected to enable viral replication in these recoded organ-

isms. Therefore, we are now implementing the swapped genetic code of this work to ensure the virus and HGT resistance of these engineered hosts. We expect that our results will have implications for the safe use of GMOs in open environments by establishing a generalizable method for genetic code alteration that simultaneously prevents viral predation in natural ecosystems and blocks incoming and outgoing HGT with natural organisms. Follow-up works could utilize biocontained swapped genetic codes to prevent transgene release during bioremediation, offer containment for open cultures of engineered photoautotrophs for carbon sequestration, and enable safe microbiome engineering and vaccine production by addicting living treatments to use a swapped code. The combination of genome recoding and codon swap may provide a universal strategy to make any species resistant to many or all natural viruses.

4.3 The Dual-Use Education Gap: Awareness and Education of Life Science Researchers on Nonpathogen-Related Dual-Use Research

4.3.1 Abstract

With the rise of synthetic biology, dual-use research risks are not confined to pathogen-related research. However, existing measures to mitigate the risks of dual-use research, such as export control, are still designed to hinder access to pathogens and do not address the risks of nonpathogen-related dual-use research. The current self-regulatory approach requires scientists to be aware of their responsibility and know how to assess risks and establish countermeasures. The purpose of this study was to examine the state of knowledge about dual-use research among life science students and to test an alternative teaching approach on the importance of considering biosecurity risks for teams participating in the International Genetically Engineered Machine (iGEM) competition. We conducted an international survey from July 18 to September 13, 2018, which was completed by 192 respondents from 29 countries and 74 universities. Based on the results of the survey, we designed and tested a learning workshop on dual-use research within the iGEM community. Results from the workshop and the survey show that educational machinery so far have failed to integrate teaching about dual-use research issues.

4.3.2 Introduction

The World Health Organization (WHO) defines dual-use life science research as “knowledge and technologies generated by legitimate life science research that may be appropriated for illegitimate intentions and applications” (World Health Organization, 2012).

One of the most obvious examples of dual-use research, and the one oversight focuses on, is research on pathogens. This research is needed to cope with outbreaks and pandemics, but pathogens can also cause severe harm when used by people with malicious intent. To prevent the misuse of pathogen-related research, in most parts of the world countermeasures such as export control and strict laboratory biorisk management are in place. In addition to those measures, sufficient education for researchers on how best to mitigate pathogen-related research risks is also important (National Academies of Sciences, Engineering, and Medicine, 2018).

Dual-use research is defined not only by the particular organisms used in research; certain activities, such as 7 types of risky studies or “experiments of concern”, are also commonly considered to be dual-use research (National Research Council, 2004). This includes experiments where pathogens are modified to become more transmissible or more virulent (Selgelid, 2016), such as the experiment performed by Herfst *et al.* in 2012, in which his team modified the influenza A (H5N1) virus to gain airborne transmissibility between ferrets (Herfst *et al.*, 2012).

Frameworks like the 7 experiments of concern primarily focus on animal or agricultural pathogens, but research does not have to involve pathogens to have potential for misuse. Examples of such research include the gene editing-based Insect Allies project (Bextine, 2018), whose potential to be used as a bioweapon has been extensively discussed in the past (Reeves *et al.*, 2018), and the potential malicious use of gene drives (Gurwitz, 2014).

We became aware of this problem while working on our own project as part of the International Genetically Engineered Machine (iGEM) competition in 2018. Our team, from the Center for Biotechnology at Bielefeld University (Bielefeld-CeBiTec), aimed to produce metal nanoparticles from waste using *Escherichia coli* (iGEM Bielefeld CeBiTec, 2018). During the course of the project, a team member asked, “What would happen if these ‘metal-eating bacteria’ are used to degrade functional electronics?” At that time, all team members and most of our supervisors had never heard of the term “dual-use research” or knew who to talk to after identifying a potential for misuse of their research project.

If the question of misuse had not been raised by our team, we might never have realized that our project could potentially be used to cause harm. All of the team’s organisms, parts, and research activities were suited for a biosafety laboratory level 1, the lowest safety category in Germany. This suggests there was minimal safety risk from the project. The company synthesizing the DNA would not have flagged the orders as potentially hazardous, because none of the sequences were associated with pathogenicity. Additionally, none of the organisms or parts used fall under export control lists, such as the Australia Group Common Control List (Australia Group, 2007). If the work of researchers is not captured by any of these regulatory regimes, it can be difficult for the researchers to identify and manage dual-use risks. Researchers working with risk group 1 organisms, like *E. coli* laboratory strains, are less likely to have come across or considered dual-use research issues (National Academies of Sciences, Engineering, and Medicine, 2017). This was the case for our iGEM team, in which none of the students had ever worked with pathogens nor had any training on dual-use research issues. In general, the university relies on the self-initiative of the primary investigator to offer dual-use research training. If these educational gaps had been filled, the researchers could have identified dual-use research risks earlier, during the project planning phase, enabling them to redesign their project to be safer. They would still get the desired research outcome, but in a safer research environment with lower risks.

This approach of dual-use research risk mitigation by the researchers themselves has several advantages; however, many researchers are hesitant to learn more about dual-use research because they fear that regulations are too critical and will result in them having to give up research projects. This impression of biosecurity considerations as a hindrance to free research needs

to be met with good practice examples that show, to the contrary, that biosecurity considerations can ease the realization of research projects. For our 2018 iGEM project, a simple change of the substrate from solid metals to soluble metal ions, like those found in mining wastewater, greatly decreased the potential for misuse. By identifying the dual-use research risk in the planning phase and adapting the substrate early in the project, we reduced the dual-use research potential and avoided additional work by not wasting resources on adapting the project at a later stage. Our team also later realized that changes made to manage the dual-use research risks had other benefits, even increasing the feasibility of the project, since the new uptake strategy was easier.

To further understand dual-use research awareness among life science students and researchers worldwide, we decided to use the iGEM community as a case study. The iGEM competition is the largest synthetic biology competition in the world. The competition had 353 teams in 2019, reaching more than 40 countries and over 6,500 participants (iGEM Foundation, no date[a]). iGEM has a dedicated biosafety and biosecurity program, and the website of this program offers information on dual-use research (iGEM Foundation, no date[b]). In addition, nearly all results are open source. Consequently, anyone with access to the internet has access to all information about the projects. Due to the enormous impact of this access and the fact that teams are forbidden to work with organisms categorized as risk group 3 or 4 (which include the most dangerous pathogens), iGEM offers the ideal environment to study how researchers handle nonpathogen-related dual-use research.

4.3.3 Materials and Methods

We conducted an international survey from July 18 to September 13, 2018, which was completed by 192 respondents from 29 countries and 74 universities. The survey was posted online and advertised through the social media channels of the iGEM Bielefeld-CeBiTec team.

Based on the results of the survey, we designed and tested a learning workshop on dual-use research within the iGEM community. Past education projects emphasize the importance of teaching biosecurity issues using active learning approaches like team-based learning, especially given that biosecurity is an unfamiliar topic to life science students (Novossiolova T, 2016). To test a continuous approach to establish better awareness for dual-use research issues, we tested the workshop 6 times within and outside of the iGEM community in 2020 and 2021. The interactive workshop focuses on 2 case studies. In contrast to team-based learning strategies, the workshop was designed to fit within a 2-hour timeframe while still facilitating an active learning approach. The workshop aimed to introduce students to the basic concepts of

evaluating and classifying biosafety and biosecurity risks, with a focus on nonpathogen-related dual-use research.

Due to the ongoing COVID-19 pandemic, workshops were held via Zoom. Students were separated into 4 groups with up to 5 students in each group, and 2 groups were assigned to each case study. For each case study, 1 group watched a 3-minute video introduction before beginning, whereas the other group did not. The video contained WHO definitions of dual-use research, biosafety, and biosecurity (World Health Organization, 2012; World Health Organization, 2018). We used this video introduction to determine whether teaching definitions would help students to correctly assess and classify risks.

The 4 groups then learned about their case study in the form of a short presentation. We decided to use ambiguous cases that, depending on the definition used, may or may not be classified as dual-use research. Both case studies were nonpathogen related. The 2 case studies were:

- The Insect Allies project by the Defense Advanced Research Projects Agency, which aimed to use insects to deliver gene therapy that would engineer agriculturally relevant plants (Bextine, 2018).
- An example gene drive solution to combat Amaranthus palmeri in the southern states of the United States, which was taken from the 2016 National Academies of Science, Engineering, and Medicine report, Gene Drives on the Horizon.¹⁶ The aim of the project was to create gene drives that would reduce or eliminate glyphosate-resistant A palmeri on agricultural fields. content...

The teams had 30 minutes to respond to the following questions:

- What are the biosafety risks of the project?
- What are the biosecurity risks of the project?
- Would you classify the project as a dual-use research project? Why or why not?

Each team then had 3 to 5 minutes to present their results to the group. After the team presentations, workshop facilitators gave a more detailed presentation entitled: “Dual-Use and Dual-Use Research of Concern – Between Science Misuse and Science Communication”. The facilitators then discussed the concept and intended design of the workshop with participants.

We held the dual-use research workshop 5 times as an online event for iGEM participants and once for the life science community as a whole, in cooperation with the Alberta RNA Research and Training Institute and SynBio Canada. Each group had a facilitator who helped keep track of the time and ensured protocols were followed during discussions.

4.3.4 Results and Discussion

The aim of the initial survey was to determine whether it was an isolated incident that the Bielefeld iGEM team had never heard of dual-use research risks or if the lack of awareness was more widespread. Previous investigations on dual-use research awareness have shown an overall lack of awareness among life scientists; for example, the 2016 Gene Drives on the Horizon report (National Academies of Sciences, Engineering, and Medicine, 2016) states that education on biosecurity issues in the United States is rarely introduced in a systematic way at undergraduate, graduate, and postgraduate levels, unless the researcher works with pathogens and even then, the focus is on biosafety. In addition, a survey conducted among postgraduate researchers in Pakistan showed that 58.2% of survey respondents had never heard of the term “dual-use research of concern”(Sarwar *et al.*, 2019).

In the survey conducted for our study, the first question was whether respondents knew the meaning of the term “dual use” in the context of science. Only 41% ($n = 79$) knew the meaning of the term in the context of science. Most respondents did not know, and a few chose the wrong definition (Figure 14).

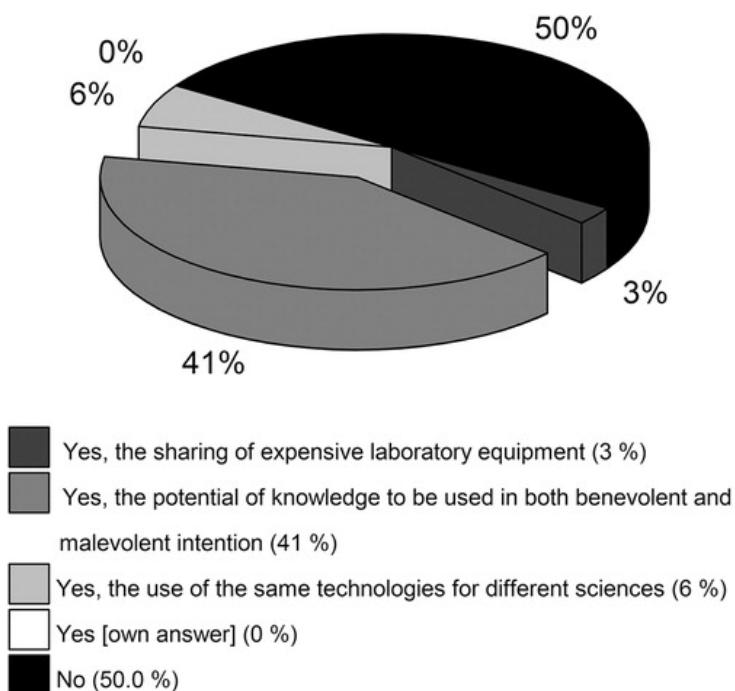


Figure 14: Responses to the survey question: “Do you know what ‘dual use’ in the context of science is?”

To better understand how students think about biosecurity-related risks, we asked if students were concerned about the potential misuse of research in their discipline. Over half ($n = 104$,

54%) of respondents agreed they were concerned, while 30% ($n = 57$) said they were not concerned and 16% ($n = 31$) had no opinion. This shows that even if students are not aware of the definition of “dual use”, they are concerned about biosecurity risks.

When asked if they thought results from dual-use research should be published and freely accessible, almost two-thirds ($n = 117$, 61%) of respondents stated that results should be published, with censorship if necessary (Figure 15). This response shows a relatively high degree of comfort with adapting publications because of sensitive research findings, which surprised us since an argument we often heard against increasing awareness of dual-use research was that it might lead to censorship and hinder freedom in research.

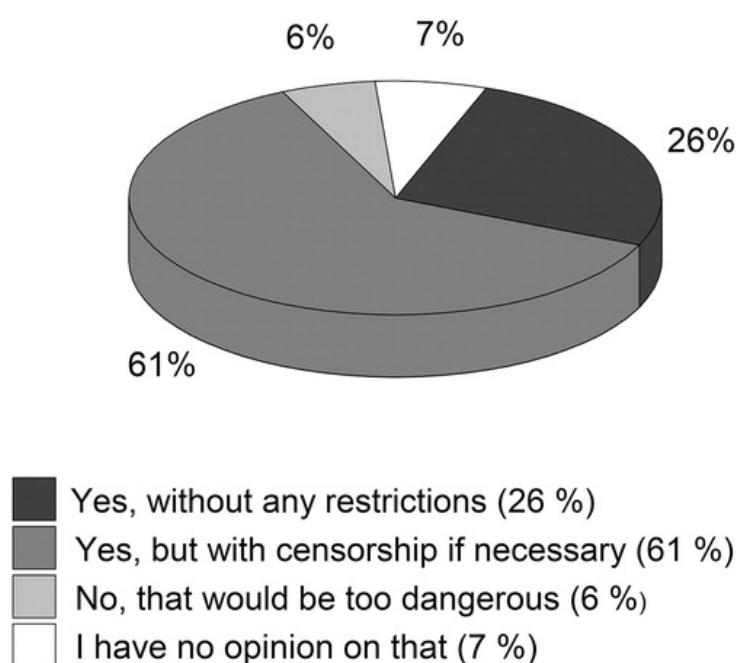


Figure 15: Responses to the survey question: “Should results from dual use research of concern be published freely accessible in your opinion?”

We then asked the respondents about their experience with the concept of dual use in their education and training, including how well their university or department teaches or informs students about the topic in their curricula (Figure 16), and in what form of teaching course dual-use issues are taught—choices included lecture, seminar, research project, other, or not at all. Only 16% ($n = 31$) of the students stated that their department does a good job in including dual-use issues in the curricula, whereas 4% ($n = 8$) answered “poor”. These results do not seem surprising given the overall lack of knowledge as mentioned earlier. They also align with the responses to the question: “How does your university inform about dual-use

and dual-use research of concern?" More than half of the respondents stated that their university does not inform them at all (data shown in the Supplemental Material, available at www.liebertpub.com/doi/suppl/10.1089/hs.2021.0177 In Citavi anzeigen).

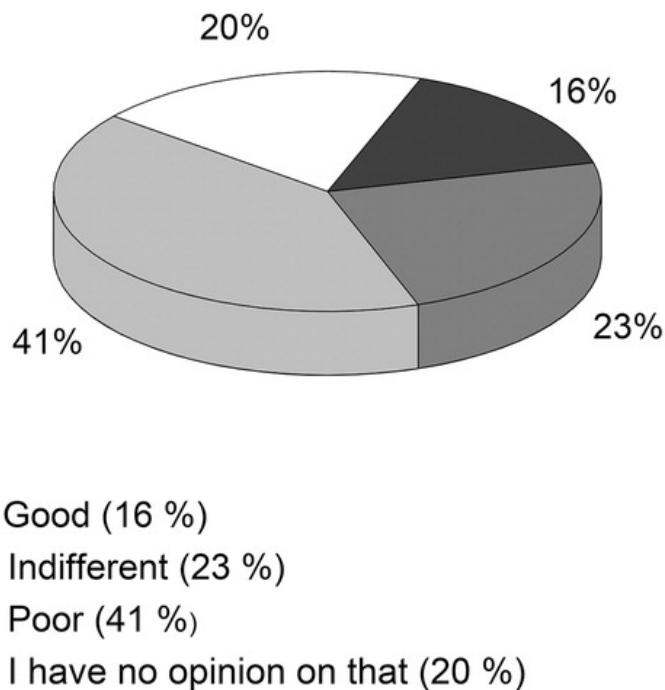


Figure 16: Responses to the survey question: "How well does your university/department enlighten about the topics 'dual use' and 'dual-use research of concern'?"

We also wanted to know if the survey respondents were interested in learning more about these topics as part of their university education. The majority ($n = 146$, 76%) of respondents stated that they want more education about dual-use issues, while the rest had either no opinion ($n = 17$, 14%) or no interest $n = 19$, 10

In the last question, students were asked if they knew who to contact if they had concerns about the effects of their research and publications; 58% ($n = 111$) responded "no" and 42% ($n = 81$) responded "yes."

The lack of awareness related to dual-use research has long been a problem. In 2011, Minehata *et al.* summarized their findings from multiple surveys and concluded that although international attention on dual-use research issues has increased, the lack of awareness among life science students was a severe problem (Minehata *et al.*, 2013). According to our survey findings, this is still the case. To improve awareness, coordinated international approaches are needed to integrate teaching into life science curricula. We need to design and test new learning

opportunities and biosecurity policy concepts to ensure that safety measures grow at the same pace as life science research (Novossiolova T, 2016).

Based on the results of the workshop, we were able to draw some general conclusions that can help to design and improve dual-use research education programs:

Students should consider biosecurity risks that are closely related to their laboratory work and will require support when considering those risks. Nearly all teams participating in the learning workshop focused initially on defining or remembering the difference between biosafety and biosecurity. All teams started with a discussion on biosafety risks closely related to their work in the laboratory, such as unintentional release or mutations arising from their experimental design that could lead to unwanted effects. Their 2 main challenges were (1) classifying risks as biosafety or biosecurity and (2) identifying biosecurity risks. As a result, the students later named the unfamiliarity of biosecurity risks. After the discussions and presentations, the students said that the process of discussing the mistakes in their evaluation as a group increased their confidence in being able to better identify dual-use research.

If questions are well structured, providing definitions may be counterproductive. All teams working on the Insect Allies case study classified the project as dual-use research and most team presentations mentioned its potential to be used as a bioweapon. For the *A. palmeri* case study, the teams had different opinions about whether to classify it as dual-use research. The team that had not been taught the definitions did not classify the project as dual-use research, arguing that the technology of using gene drives harbors the potential to cause harm, but that the specific project of gene drives in *A. palmeri* did not fulfill the criteria of potential for direct misuse. The teams that received the introduction with definitions did not identify more risks or better classify the risks as biosafety or biosecurity related. According to the facilitators, the teams that did not receive an introduction with definitions had gained a better understanding of the terms “biosafety”, “biosecurity”, and “dual-use research”. During the discussions, teams that received an introduction to the definitions said the video was useful, especially in realizing the difference between biosafety and biosecurity. However, the teams without the definitions knew there was a difference because of the way the questions were formulated, which led quickly to a discussion within the group about what those differences might be. This approach may have led to a better understanding of the concepts, because those students took a deeper look at the definitions. Ultimately, we want students to assess whether their project could cause harm. To achieve this goal, teaching definitions could be counterproductive since risks might be overlooked with the justification that the risks do not closely match one of the existing definitions.

Active learning approaches work better than presenting definitions. If students must come up with risks and definitions themselves, they take a deeper look into these risks, which leads to better understanding of the general matter even if the definitions might not be entirely correct. Allowing our workshop participants to work on case studies first and then evaluate their work as a group ensured they had the best understanding of the topic. Although we believe that courses based on active learning work better in person, the online format enabled students from around the world to work together on projects, which generated discussions of regional attitudes and approaches to dual-use research risk reduction and education.

Students recognize gaps in their education related to biosecurity and would like to learn more. In the discussion after the workshop presentations, the students stated that none of them had learned about the difference between biosafety and biosecurity during their university education; in fact, some languages translate both terms into the same word, which had led to misunderstandings. All of the students in the workshops said they wished to learn about biosecurity risks in their university education, which mirrors the results from our survey. None of the students had ever attended a lecture or seminar covering these issues. They also told us that “beyond the lab” they learn only about ethical issues and biosafety as far as it concerns laboratory safety. All of the students also said they had underestimated the importance of considering biosecurity issues and were unfamiliar with the case studies, even though the Insect Allies project had been widely discussed in the biosecurity community (Reeves *et al.*, 2018; Baumgartner, 2018; Durkin, 2018).

Considering dual-use risks early in the project facilitates the research process and prevents wasting resources. For the *A palmeri* case study, we suggested 2 possible targets for a gene drive: a glyphosate-resistant enzyme or sex-specific genes that would cause the entire population to collapse. As the students discussed risks related to these 2 research approaches, they quickly realized that even if both approaches followed the same aim, the second approach had a higher potential to be misused. Considering dual-use research risks early in the project planning phase enables the researchers to implement mitigation measures, preventing wasting time and materials on research that is not sufficient for implementation.

Only a fraction of the students showed enough interest in biosecurity to attend a voluntary course. Workshop participation was voluntary, but contrary to the findings of the survey, in which 76% ($n = 146$) of the respondents stated they want to learn more about dual-use research, about 45 iGEM 2021 participants attended one of the workshops, which is less than 1% of the 7,314 iGEM 2021 participants (iGEM Foundation, 2021). The absence of mandatory dual-use research education conveys the impression that biosecurity considerations are not a

fundamental part of life science education. Several organizations have made biosecurity-related teaching material available online, but they still rely on the self-initiative of teachers to incorporate these materials into their courses (Novossiolova T, 2016; University of Bath, 2022). The state of biosecurity knowledge suggests that this approach is not sufficient, perhaps because teachers, like their students, are not aware of biosecurity risks or because they do not consider biosecurity risks important enough to integrate into curricula. As a result, students have the impression that biosecurity is not as important as other topics taught in their courses and, therefore, not an integral part of their research. In addition, teaching programs mainly include examples of pathogen-related dual-use research, conveying the impression that nonpathogen-related research has no relevance to biosecurity.

The data from the survey and the workshop show that most life science students have never heard of dual-use research and that the educational machinery so far has failed to integrate teaching about dual-use research issues into life science curricula. We should note that the survey data are from 2018, and that the COVID-19 pandemic and teaching approaches conducted since 2018 may have increased awareness on dual-use research. Nonetheless, this study identifies key problems and presents education approaches regarding nonpathogen dual-use research within the iGEM community that can be applied to the entire life science community.

The key findings from this study are:

Dual-use research issues exist in nonpathogen-related research, but existing oversight frameworks primarily focus on pathogen-related dual-use research.

Enabling researchers to identify dual-use research risks early in the planning stage allows them to revise the experimental design, both reducing risk of misuse and preventing wasted resources.

Too few life science researchers understand the meaning of dual-use research.

The majority ($n = 146$, 76%) of survey respondents stated they wanted more education on dual-use research issues.

Education on dual-use research issues needs to be widespread to convey the importance of considering dual-use research risks as a fundamental part of research.

Nonpathogen-related dual-use research examples need to be included in existing and future learning opportunities.

The main implication of these findings for iGEM is that it needs to decide whether to have a stronger focus on teaching biosecurity issues. iGEM is able to quickly test and implement new educational projects and has the big advantage of a designated safety and security program (Millett *et al.*, 2019). Participants already have an obligation to assess safety and security risks related to their projects, but some projects still present challenges to the existing safety and security program, including nonpathogen-related dual-use research (Millett *et al.*, 2020). The

workshop might improve understanding and consideration of dual-use research issues among teams, but it is not realistic to think that workshops at the regional meetup would be able to reach every iGEM team. To gain wider reach, educational programs should be implemented on a larger scale and earlier in the iGEM season, when teams are beginning to plan their projects. Beyond iGEM, implementing suitable educational programs about dual-use research in the larger life science community will be more difficult, because more actors need to be involved and more time and money will be needed to change a whole educational machinery. The first step is to integrate not only examples of pathogen-related research but also nonpathogen-related dual-use research into existing learning opportunities such as FutureLearn's "Next Generation Biosecurity: Responding to 21st Century Biorisks" online course by the University of Bath and Biosecure (University of Bath, 2022) or in teaching materials such as the Biological Security Education Handbook: The Power of Team-Based Learning (Novossiолова Т, 2016). The second step is to ensure that these teaching opportunities reach the students.

The results of this study show that it is not enough to publish teaching materials and rely on the self-initiative of universities to implement them. A biosecurity outreach campaign is needed to convince all actors to implement teaching about risks. In addition, more actors need to be involved, including future employers and the life science research industry. If future employers do not value students with additional qualifications in biosecurity, students have less incentive to participate in workshops and other learning opportunities.

An alternative to voluntary learning based on incentives would be a compulsory dual-use research education. The advantage of compulsory learning is that students would understand that biosecurity considerations are a fundamental part of life science research. For this approach to be successful, the basics of biosecurity must be discussed in at least 1 lecture of every life science study curriculum. If lecturers do not see the necessity of integrating biosecurity issues in the curriculum, it might be worth inviting guest lecturers who have experience with biosecurity policy. It is important to note that universities cannot be forced to implement mandatory biosecurity education. To facilitate this approach, an open dialogue is needed between universities and biosecurity policymakers to ensure that universities understand why such an approach is beneficial and which frameworks are needed to support implementation.

A mandatory teaching approach that covers only the basic concepts of dual-use research has the advantage of requiring minimal time invested. For students who want to learn more about dual-use research issues, additional voluntary learning opportunities could be offered. It should be noted that several aspects of responsible research, such as broader ethical guidelines, are underrepresented in teaching curricula (Mackelprang *et al.*, 2021). Combining various aspects of responsible life science research into a single study course through, for example, guest lectures

on responsible science, ethics, and biosecurity like those offered by the Centre for Biosecurity and Biopreparedness, could therefore be useful (Centre for Biosecurity and Biopreparedness, 2021).

International organizations involved in life science education, networking, and biological threat reduction can play an important role in supporting and building up educational networks. As previously mentioned, some organizations have already started to offer teaching materials or workshops and/or are aware of the importance of raising awareness; for example, the Organisation for the Prohibition of Chemical Weapons (OPCW) future priorities from 2018 stated that “public engagement, education, and awareness-raising must become an integral part of OPCW activities” (Organisation for the Prohibition of Chemical Weapons, 2018).

4.3.5 Conclusion

Overall, our findings from the survey and workshops conducted within the iGEM community show that more opportunities for learning are needed to bridge the dual-use education gap among life science researchers, but that increasing dual-use research awareness on a broader scale is still within reach. In the past, implementing new biosecurity learning opportunities had not been a priority. To correct this oversight, we need to incorporate teaching about nonpathogen-related dual-use research. Current regulation strategies for dual-use research rely almost entirely on self-regulation, and thus their awareness. All actors including biosecurity policymakers, teachers, and industry need to work together and invest time and resources to create educational programs that ensure life science research is conducted in a safe and responsible way. Doing so will prevent future researchers from being unaware of dual-use research risks that have the potential to cause widespread harm.

4.4 How to teach life sciences students about dual-use research—a view from the field

4.4.1 Abstract

To reduce biological risks, raising awareness for Dual Use issues already at the level of university education is essential. Currently, most life sciences education programs do not incorporate biosecurity and Dual Use in their regular curricula. Consequently, the responsibility rests with individual lecturers and depends on their initiative to incorporate Dual Use topics into teaching activities. Students interested in biosecurity and Dual Use topics often only have the option to educate themselves in external or online courses. Here, we provide practical guidance on how to initiate and integrate a Dual Use education program within the curriculum and provide a selection of existing teaching materials. In addition, we suggest key learning objectives to guide the planning of Dual Use courses. Different course formats like lectures, seminars or stand-alone events are discussed regarding their advantages, disadvantages and suitability for conveying the learning objectives to different educational stages and audiences. As a minimum, we recommend the incorporation of Dual Use issues into at least one mandatory course. Ideally, students should additionally participate in in-depth seminars which can be voluntary and offered in cooperation with external organisations.

4.4.2 Introduction

Teaching how to reduce biorisks is a crucial part in the education of life sciences researchers. Biorisks are risks posed by unsuitable biosafety, biosecurity and Dual Use research precautions that lead to unintentional (biosafety) or intentional (biosecurity) harm of animals, humans, the environment, or national security. Dual Use research, a biosecurity risk, is life sciences research with the potential to be misused by third parties to cause harm.

While the teaching strategies in this study are beneficial for all areas of biorisk teaching, in this paper the focus will be on teaching Dual Use issues, issues which are often neglected in life sciences curricula. The mitigation of these risks rely on the individual researcher, while for other biosecurity and biosafety risks, institutional, national, and even international frameworks are in place. The self-regulatory approach regarding Dual Use research requires life sciences researchers to be able to assess and mitigate these risks.

For this publication, life sciences are considered as all sciences that deal with living organisms, including humans, nonhuman animals and the environment; or products of living organisms or incorporating components derived directly or synthetically from living organisms; and including but not limited to biology, biotechnology, genomics, proteomics, bioinformatics, pharma-

ceutical and bio-medical research and technologies. Therefore, people with diverse academic backgrounds contribute to the advances in this broad field. Additionally, educational programs maintain different focusses even within the same discipline. Consequently, there is no prototypical life sciences education program or curriculum. However, all life sciences studies share that epistemology is the core of these research disciplines. So even if there may be a variation in focus, all disciplines share a core foundation which should be used as basis to incorporate teaching Dual Use risks into curricula.

Life sciences research aims to better understand, manipulate or even to engineer living systems to improve health, quality of life, or the environment. The field evolves rapidly, promising benefits for society by introducing new products or technologies and expanding basic knowledge. This rapid progress is further compounded by the fact that necessary equipment becomes cheaper and easier to use. While the increasing innovation speed and accessibility enables fast scientific progress, it also increases the potential of misuse(Evans & Selgelid, 2015).

The mitigation of Dual Use risks aims to prevent misuses of legitimate life sciences research out-puts, while preserving the freedom of research. According to the US National Academy of Sciences, Engineering and Medicine, researchers who are not working on pathogens are mostly unaware of Dual Use issues and education is rarely introduced at the under-, post-graduate or graduate student level (National Academies of Sciences, Engineering, and Medicine, 2017). A survey amongst post-graduate students from the life science community of Pakistan found that 58 % of the participants have never heard of the term “Dual Use Research of Concern” (Sarwar *et al.*, 2019). Mirroring these results, only 41 % of the participants in a study within the synthetic biology research community know what Dual Use means in the context of science (Vinke *et al.*, 2022).

The survey results are not too surprising, since besides special courses on bioethics, most life sciences study programs have little to no humanity courses (National Academies of Sciences, Engineering, and Medicine, 2018). The lack of awareness of Dual Use is a long-known problem (National Research Council, 2011; Minehata *et al.*, 2013). Although there is a broad agreement that awareness-raising is a crucial element in mitigating Dual Use risks, the surveys demonstrate wide-spread education is still missing.

Different elements contribute to the reduction of biosecurity risks ranging from controlling access to certain materials, agents, genetic sequences and equipment with export control and industrial self-regulation down to the level of individual scientists through education and awareness raising. Education becomes even more important as laboratory equipment becomes more accessible and as the access to genetic material gets more complicated to control; e.g. laboratory equipment is widely available, companies start to offer benchtop DNA synthesisers,

and genomes or even organisms can be designed *de novo* (Rennekamp, 2019; Venetz *et al.*, 2019). Therefore, governance of Dual Use needs to evolve to mitigate risks of non-pathogen related research and synthetic organisms by de-signing risk assessment frameworks that are highly adaptable to include risks posed by novel scientific findings and methods (Inglesby *et al.*, 2019; Millett *et al.*, 2020)

All these challenges highlight the need for a continuous assessment and mitigation of potential Dual-Use risks throughout the entire cycle of a research project by all actors involved from research project planning to the publication of potentially sensitive results. Implementation of risk assessment from the initial planning steps by the researchers themselves is especially important, preventing wasteful use of resources and allowing the incorporation of biosafety (prevent unintended release) and biosecurity (prevent unauthorized access) considerations in the project design. Next to widespread education, other actors like funding agencies and publishers carry responsibility for what kind of research is carried out and what is published.

4.4.3 Materials and Methods

General awareness of Dual Use issues is essential to make considerations of Dual Use risks a common practice within the life sciences research community. To promote a sense of responsibility for planning and conducting research in a way that mitigates the risk of misuse, instruction about these issues needs to be implemented into university curricula worldwide. Here we propose a list of key learning objectives (Table 1) to be covered in Dual Use courses for life science students and review different course formats as well as materials available online.

This article provides a starting point for the implementation of Dual Use education in different contexts to motivate and inspire educators who are often new to the field of biosecurity and Dual Use. Therefore, we also discuss and refer to our personal experiences from implementing such courses.

4.4.4 Results and Discussion

University classes in the main curricula heavily focus on providing subject specific factual knowledge, complemented with training in transferable skills like scientific writing, scientific integrity or ethics. To prepare the students for a lifelong learning and implement a culture of responsible science, Dual Use education aims to equip students with critical thinking skills and a mindset to integrate into their future professional life. It should trigger and enable their critical reflection on life science research in the context of misuse potential and make them aware of their role in promoting benefits for society while reducing both predictable, as well as unpredictable risks. We see Dual Use awareness as the first step and a key requirement for the

behavioural change needed to reduce potential misuse risks, such as adjusting the research design or the publication of research findings. Awareness alone might not prevent the intent of misuse. However, the combination of awareness with preventive and risk mitigating measures is essential to make misuse more difficult. Therefore, Dual Use awareness is important for every career path, whether in academia, industry, or science publishing. Measuring the success of Dual Use education beyond knowledge is challenging (Revill J., 2010). Workshops held in Pakistan to address the lack of awareness identified in the survey mentioned above increased knowledge in key questions. Furthermore, 96.9% of the participants shared the gained knowledge either informally (80.6%), in lectures (16.2%) or workshops (3.2%) (Sarwar *et al.*, 2019; Arfin Qasmi *et al.*, 2021). This demonstrates the potential positive effect of educational efforts beyond the classroom.

Learning means successively adding and linking new pieces of knowledge, a process that is facilitated by taking prior knowledge into account and primarily driven by intrinsic motivation (Schneider & Stern, 2010; Linn, 2012; Cook & Artino, 2016). Factors like emotions, values as well as the larger societal context interact with motivation and learning (Schneider and Stern 2010; Cook and Artino 2016). Education about Dual Use should therefore emphasize the relevance of the topic outside the classroom by linking it to the students' practical research experience as well as their values or goals to contribute to benefits for society. To build-up the conceptual change and to encourage a reflection about the potential of offensive misuse of life science research in contrast to the students' peaceful intentions, students should be provided with meaningful learning activities aligned with the learning objectives to obtain the necessary skillset (for background information about constructive alignment see: (Biggs, 1996; Anderson & Krathwohl, 2001; Krathwohl, 2002; Biggs J. & Tang, 2011)).

The summary of key objectives in Table 1 integrates existing Dual Use course frameworks like the course handbook “Preventing Biological Threats, What you can do” (Whitby *et al.*, 2015; Novossiolova T, 2016) as well as personal recommendations from the authors.

Table 1: Overview of recommended learning objectives. The essential learning outcomes are in bold, the cognitive level of the underlying learning objectives (1-13) roughly increases from top to bottom.e.

	By the end of a Dual Use course, students should be able to ...	Cognitive process
	... explain the goals and anticipated benefits of their research.	Understand (explain)

	... check their research to identify and evaluate and judge potential harms and risks to biosafety and biosecurity..	Remember (recognize/identify), evaluate (check, judge)
	... design and implement appropriate precautionary and mitigating risk reduction strategies in their research design..	Apply (implement), create (design)
1.	... recall occasions when life science research has been misused for warfare purposes and explain the historical connection between scientific advances and their military application.	Remember (recall), understand (explain)
2.	... recall events to illustrate the potential interest of non-state actors like terrorist groups in biological weapons and to explain the insider threat.	Remember (recall), Understand (explain, exemplify/illustrate)
3.	... summarize and interpret the developments of the Chemical Weapons (CWC) as well as the Biological and Toxin Weapons Convention (BTWC).	Understand (summarize, interpret)
4.	... discuss the role of scientific advances in the BTWC and CWC and explain the challenges they pose to the Conventions.	Understand (explain)
5.	... recognize the obligation of states to prevent non-state actors from access to weapons of mass destruction (UNSCR 1540).	Remember (recognize)
6.	... categorize and distinguish different types of biological risks and illustrate how they are influenced by scientific advances.	Understand (classifying/categorizing), analyze (differentiate/distinguish)
7.	... explain and differentiate the concepts of biosafety, biosecurity and Dual Use.	Understand (explain), analyze (differentiate)
8.	... explain the concept and attribute their role as life scientists within the Web-of-Prevention.	Understand (explain), analyze (attribute)
9.	... name and describe key actors and regulations that are concerned with Dual Use and identify a point of contact in case of concerns.	Remember (identify)

10.	... illustrate and explain why certain technological developments are under debate in the context of their Dual Use potential.	<i>Understand</i> (exemplify/illustrate, explain)
11.	... outline the different dimensions of responsibility and possible roles scientists can have at different stages of research projects and how to take action to take this responsibility.	<i>Analyze</i> (organize/outline)
12.	... check research projects by differentiating and hypothesizing the potential to apply scientific results or methods both with benevolent or malevolent intentions to judge their misuse potential.	<i>Analyze</i> (differentiate), <i>evaluate</i> (check), <i>create</i> (generate/hypothesize)
13.	... conclude and design measures to implement best biosafety and biosecurity practices to reduce the identified risks.	<i>Understand</i> (infer/conclude), <i>create</i> (plan, design)

The essential learning outcome should be the ability to reflect on benefits and risks and the implementation of measures to reduce these risks (Table 1, bold). The list of 13 learning objectives can be seen as a summary containing discrete building blocks serving as a basis on which educators can develop relevant course content to reach this goal. To promote the transfer of knowledge, higher cognitive processes from understand and apply, through analyze up to evaluate and create have to be activated beyond the underlying low cognitive level remember (for further details see (Anderson & Krathwohl, 2001)). The formulations of the learning objectives (Table 1) intend to reflect the different cognitive levels of the learning activities required for the envisioned learning outcome (Anderson & Krathwohl, 2001). The learning objectives are structured in terms of content and roughly increase in cognitive complexity but do not necessarily reflect a chronological order. Different learning activities and assessments should be designed according to, and aligned with, the respective learning objectives to provide students with opportunities to acquire the skills necessary to achieve their particular learning outcomes (Biggs, 1996; Anderson & Krathwohl, 2001).

Lectures or background reading material can be useful elements for teaching hard facts in this context, or the circumstances that led to the installation of the Biological and Toxin Weapons Convention, as well as the Chemical Weapons Convention (learning objective 3). However, interactive elements should be integrated into such a course to enable the students to assume responsibility and take on an active role throughout their professional life by engaging in debates at the intersection of science, policy, and society (learning objective 11, Table 1)). Applying

the concepts of Active Learning (students engage actively with the content through e.g., case studies or debates) and the Flipped Classroom concept (learning about the topic at home and applying the knowledge during the course) can improve critical thinking skills (Styers *et al.*, 2018).

It is often challenging for scientists to recognize the risks and misuse potential in their own research projects, especially when it comes to non-pathogen related research, as previously published examples highlight (Vinke *et al.*, 2022). While students seem familiar with biosafety measures, they struggle with biosecurity considerations like anticipating the malevolent use of their research (learning objective 7, Table 1). Many students in our courses reported that they never considered misuse risks before. To prepare the students to reflect with respect to unintentional uses of their own research in a stepwise manner, a course may roughly move through the cognitive levels of the learning objectives from low to high. Providing a historical background of the role of research in chemical and biological warfare, terrorism, and insider incidents (learning objectives 1 and 2, Table 1) can be a first step to provide a different view on scientific and technological advances (learning objective 6, Table 1). Analyzing example projects with high misuse potential (learning objective 12, Table 1) helps the students to recognize different ways to apply methods depending on contexts and purposes before applying it to their own research (essential learning outcomes, Table 1, bold). They will also realize that no clear answers can be given in this topic area. The question of whether certain research should be performed or not usually cannot be answered just with a simple yes or no. To assess respective risks and benefits, factors like the regulatory environment (learning objective 9, Table 1), the ease and accessibility to (mis)apply the methods along with the possibility to design alternative approaches need to be considered (learning objectives 12 and 13, Table 1).

In addition, it is important for students to be aware that international treaties and resolutions against the development and dissemination of biological and chemical weapons exist that cover states as well as non-state actors (learning objectives 3 and 5, Table 1). Examining the development of the Biological and Toxin Weapons Convention as well as the Chemical Weapons Convention, including their strengths and weaknesses (learning objective 4, Table 1) and how they are enforced, is essential to provide a broader context in which Dual Use issues are discussed. The topics of international treaties offer plenty of opportunities for reflection with respect to the role of scientific advances and how they might pose challenges to the conventions (learning objectives 4, 5 and 6).

To become aware of their role and responsibilities as scientists, including ways and opportunities to become active (learning objective 11, Table 1), students profit from a comprehensive approach to Dual Use issues that combines detailed background knowledge in conjunction with

opportunities to practice. Such an integration will ultimately enable the students to connect the concepts and experiences to their reality as life scientists. Over the course of their careers, they might experience different levels of awareness of and willingness to consider the misuse potential of research outcome within their respective research communities. An interactive teaching approach that promotes discussions and asks students to work on collaborative case studies can serve as a training opportunity to discuss Dual Use issues with peers, to test arguments and to be prepared for a possible confrontation with direct misuse potential of their own future work. It has been argued that discussions are valuable both to learn content as well as the skills needed for such assessments (Larson, 2000). Practicing in a “safe space” without fearing repercussions is especially important, as students might fear negative consequences when raising concerns in front of their research groups or supervisors. This is especially important for teaching and raising Dual Use issues since the risk assessment is less standardized like for biosafety risks. Thus, the perceived risk can vary, and raising of Dual Use issues by students is often seen as hypercritical. Supervisors might even fear that raising these issues might evoke their research to be associated with Dual Use risks which leads to further regulations and administrative work. Students should become aware of how to get support in these cases and how to interact with both university administration and the public.

In summary, Dual Use education aims to enable students to critically reflect on research approaches to reduce potential risks. It therefore needs to provide the necessary background knowledge while also providing time and opportunities for the students to directly interact with the content and their peers. Discussions, case studies, and scenario-based exercises are examples how this can be implemented. Cooperative learning elements are valuable in allowing students to leave a course with a mind- and a skill set that they can incorporate into their future professional life.

Teaching modes and online available courses/resources

Currently, only few universities offer dedicated Dual Use courses (Vinke *et al.*, 2022). Integration of Dual Use courses into existing curricula faces several challenges, ranging from finding suitable lecturers to restructuring densely packed curricula. The following section provides examples how this integration can be achieved and reviews different instructional modes. Hoping to motivate educators to initiate such activities, a collection of online available teaching material and courses can be found in Table 1. The collection includes resources for material that can directly be used in the classroom as well as opportunities for self-paced training of the educators themselves.

Courses within university curricula

Dual Use issues can either be introduced as a stand-alone lecture, a seminar within an existing life science lecture, as a seminar covering different aspects of responsible research, or be included in the safety introductions for a laboratory course.

Safety introductions can address the biosafety risks of research or a project. They could also offer the opportunity to educate students about the fact that biosecurity addresses a different set of risks, although measures for biosafety and biosecurity often overlap. As Dual Use risks vary a lot between projects, safety introductions are not ideal for a general overview and risk to distract the students from the essential focus on measures to ensure their safety in the laboratory. Additionally, the Dual Use concept is complex and the incorporation in usually short safety introductions will already require a basic understanding. However, potential project specific Dual Use risks should be included to convey the importance and necessity of Dual Use risk considerations at any stage of life science research. Lectures and seminars are the most feasible formats for Dual Use education. Universities and their departments must decide whether they are mandatory or not, as well as the amount of time allocated within the curricula. Mandatory Dual Use education would demonstrate the importance of assessing potential risks as fundamental part of life science research, if not of research in general.

Regarding the format, Dual Use could either be the main topic for a semester-long course, or a dedicated single lecture embedded within a life sciences course. A short introduction within a mandatory lecture reaches all students and can be a first step to convey that Dual Use education is an integral part of life sciences education, not just an add-on qualification. A follow-up course must balance the need to go into more detail within the existing time constraints of the curricula; hence, a dedicated Dual Use course has the advantage to allow a broad coverage of aspects reflected in the learning objectives, while a short teaser sequence can only have limited impact with a subset of the objectives.

Seminars provide room for discussions and facilitate Active Learning approaches, which have been demonstrated to convey the Dual Use problem better than lectures (Novossiolova T, 2016). Students can evaluate case studies, discuss current and future governance approaches and emerging technologies. In contrast to a lecture, interactive seminars challenge the students to apply their knowledge about Dual Use issues. Taken together, seminars are more suitable to cover the learning objectives that allow practicing and applying risk assessment skills. Such seminars are most suitable in the late undergraduate curriculum or at the graduate level when participants understand the science behind and can focus primarily on the risk assessment. However, the number of participants in seminars is usually limited. Expanding the format to a larger number

of students by splitting into breakout groups could be an option but would require additional instructors.

Example 1: ETH Zurich seminar

At the ETH Zurich, students are required to earn credit points in the “Science in Perspective” program (<https://gess.ethz.ch/en/studies/science-in-perspective.html>). This program allows students to reflect on natural and engineering sciences within a broader societal context. One of the program courses aims to raise awareness for Dual Use issues in the life sciences. It is open to students from various life and engineering science fields as well as different education levels. It has been designed based on the course concept “Preventing Biological Threats, What You can Do”, developed at the University of Bradford (Whitby *et al.*, 2015; Novossiolova T, 2016) (see Table 2). The semester-long seminar combines introductory lectures and guest presentations with a strong focus on active-learning elements and group discussions adopted from the Novossiolova handbook (Whitby *et al.*, 2015; Novossiolova T, 2016). This seminar is designed as a comprehensive course covering a broad range of topics ranging from biological warfare history to biosecurity implications of the latest scientific advances. Extensive time is allocated to group work and class discussions based on material provided to the students in advance. Students have the chance to practice and apply the learned concepts during individual analysis of their own research projects and develop a code of conduct flyer.

The ETH Zurich example demonstrates how voluntary courses on Dual Use can be integrated into programs that offer a broader view on societal implications of science in general. To motivate students to participate, it is important to recognize their participation through credit points earned counting against their respective degree requirements.

Courses led by third parties

Missing expertise in teaching Dual Use can be overcome by cooperating with third parties. These external experts can be Non-Governmental Organizations (NGOs), Non-Profit Organizations (NPOs), student associations or government institutions working in biosecurity governance, non-proliferation, or risk assessment. A cooperation between these groups and universities can be structured in several ways, such as inviting guest lecturers from external organisations or by designing special stand-alone events.

Inviting guest lecturers offers the advantage that they can more easily be integrated into existing courses than organizing students to have a course at an external organization and takes pressure from lecturers to design new courses on a topic that is not their main expertise. The perspective of people working in the biosecurity field makes the topic a more concrete, providing incentives to engage with the topic in-depth and potentially even inspiring students to see biosecurity

policy as a possible future career path. A key challenge of this approach is to establish the contact between external experts and the university.

Stand-alone events are the least time-consuming option as they do not take up time from the regular curriculum. Their disadvantage remains that they are typically not recognized with credit points towards university degrees which reduces their attractivity for the students. Other incentives like certificates or micro credentials should therefore be considered to demonstrate their value. External online courses can be an alternative for students and researchers where universities do not integrate Dual Use education into their curricula. Due to the virtual format, they have the advantage of inter-national availability and can thus facilitate the multilateral dialogue on Dual Use risk mitigation.

Example 2: Malice Analysis Workshop (Engineering Biology Research Consortium)

The Engineering Biology Research Consortium (EBRC) is a NPO aiming at building a community to accelerate advancing engineering biology research (Table 2). The EBRC's designated security working group attempts to raise awareness for security issues in engineering biology research (Ta-ble 2). One of its activities is the Malice Analysis Workshop (Table 2), which aims to enable re-searchers to assess security concerns within their research. According to participants, students re-ceive a short introduction including historic examples of science misuse, and life sciences research projects are assessed for their misuse potential based on a risk assessment workflow provided by the EBRC. Subsequently, the participants are divided into small groups to assess their own research projects and discuss the results as a group. The workshop concludes with a short summary by each group and a discussion.

The Malice analysis Workshop is an example for an Active Learning approach that facilitates the application of knowledge on Dual Use issues. Due to the short timeframe (less than 4 hours), not all learning objectives mentioned in Table 2 are introduced, but the participants learn to assess their own project for possible concerns. The workshop requires prior knowledge in life sciences research to participate in the discussion of the different projects. Thus, it is a example for a course addressing students close to graduation, or post-graduates who want to learn skill sets that enable them to assess their projects for misuse potential.

E-learning modules

For students and researchers wanting to educate themselves, self-directed e-learning courses are a good alternative to in-person classes. Although they lack the room for personal exchange and discussion, they provide access to people from all over the world. Online courses are available from different organisations and offer a broad range, e.g., focussing more on disarmament and non-proliferation or laboratory biosafety and biosecurity, respectively. A selected list of courses and materials is summarized in Table 1.

Example 3: Next generation Biosecurity Course: Responding to 21st Century Biorisks
(University of Bath, 2022)

The Next Generation Biosecurity Course is a free, six-week online course by Biosecu.re and the University of Bath, covering laboratory biosafety, biosecurity, and bioethics in general (University of Bath, 2022) . Participants learn the relevant theoretical background of all learning objectives (except responsibilities on Dual Use that are specific to individual regions and institutions). The course combines videos, texts, quizzes and case-studies and participants have the opportunity to comment on every section. Participation in the course is free of charge, however receiving the certificate of successful completion will cost a fee. The material can also be used for in-person training courses. The Next Generation Biosecurity Course offers a great alternative for all students and researchers that do not have the opportunity to receive in-person training at their universities.

The large number of examples and case-studies included enable participants to apply the concepts to their own research. One major drawback of this course is the format as an online course. The participants are not able to discuss what they learned directly facing each other and therefore cannot practice within the course. However, the course materials can be used to develop with only limited effort a new seminar course that offers both: the theoretical background taught by the course materials and interactive elements including follow-up discussions and assessments as a group. This approach was demonstrated by the German Association for Synthetic Biology (GASB) by offering students to join an online learning group that meets once a week after working on the course modules to discuss the course content with biosecurity experts (GASB 2022).

Table 1: Overview of selected material from the “Guides to Training and Information Resources on the Culture of Biosafety, Biosecurity, and Responsible Conduct in the Life Sciences 2019 and 2021” (International Working Group on Strengthening the Culture of Biosafety, Biosecurity, and Responsible Conduct in the Life Sciences (IWG) 2021) . (individuals: material useful for self-guided education, e.g., students, professionals, researchers, lecturers; lecturers: resources to design a course).

Developer/Source	Name	Target group	Material, comments	Link (accessed on Oct 13, 2021)
University of Bradford, UK	Preventing Biological Threats: What You Can Do (Whitby, 2015)	lecturers	modular course handbook and guide for team-based learning: reading, quizzes, tasks	https://bradscholars.brad.ac.uk/handle/10454/7-821
Bradford Disarmament Research Centre, UK, National Defence Medical College, JP, Landau Network Centro Volta, IT	Educational Module Resource (EMR), Bioethics	lecturers	modular resources, background information material, example cases	https://www.bradford.ac.uk/bioethics/educational-module-resources-emr/english-language-version-of-educational-module-resource-emr/
Infectious Disease Institute, UG; Funded by: Nuclear Threat Initiative (NTI) 2018 Next Generation for Biosecurity Competition, Next Generation Global Health Security Network	Act Like a Pro	individuals	e-Learning modules: scenarios, information material, quizzes	https://elearning.idi.co.ug/course/index.php?categoryid=19

	Nuclear Threat Initiative	Tutorials on Biological, Chemical and Nuclear Threats	individuals	basic online tutorial and quiz, non-proliferation focus	https://tutorials.nti.org/
	Biosecure Ltd and the University of Bath, on Future Learn	Biosecurity for the Next Generation: Responding to Biological Risks in the 21st Century	individuals	online course: videos, quizzes, material may be used by lecturers	https://www.futurelearn.com/courses/biosecurity
	Johns Hopkins University, on Coursera	Engineering Life: Synbio, Bioethics & Public Policy	individuals	online course	https://www.coursera.org/learn/synbioethicsab-out
	ASBA International		individuals	online-courses, training tools, resources, biosafety focus	https://absa.org/online-education
	European Biosafety Association, Federal Experts Security Advisory Panel (FESAP)		lecturers	teaching material, case studies, quizzes, biosafety and biosecurity culture	https://www.ebsaweb.eu/biosafety/biosecurity-resources/awareness-of-biosafety-and-biosecurity
	S3: Science Safety Security, US		individuals, lecturers	information material on Oversight of Dual Use Research of Concern	www.phe.gov/s3/dualuse/Pages/default.aspx
	Engineering Biology Research Consortium (EBRC)	Malice Analysis	individuals	online or in-person seminars	https://ebrc.org/ebrc-malice-analysis-workshop/

Public Health Agency of Canada	Introduction to Dual-Use in Life Sciences Research	individuals, lecturers	e-learning module	https://training-formation.phac-aspc.gc.ca/index.php?
Contributors: Canadian Food Inspection Agency, Health Canada, Public Health Agency of Canada	Biosafety and Biosecurity Portal	individuals	e-learning modules, resources	https://www.canada.ca/en/services/health/biosafety-biosecurity.html
Federation of American Scientists	Case studies in Dual Use – biological research	lecturers, individuals	e-learning modules with case studies	https://biosecurity.fas.org/education/dualuse/index.html
Centers for Disease Control and Prevention (CDC), US		individuals, lecturers	online courses, resources and tools, biosafety focus	https://www.cdc.gov/safelabs/index.html
Yale Center for Public Health Preparedness, Yale School of Public Health, US	Biosafety, Biosecurity and the Evaluation of Biohazards Course	lecturers	teaching material, presentations, handouts, videos	https://my.absa.org/biosafety-/intro.html
CITI Program	Dual Use Research of Concern and other Biosafety and Biosecurity Courses	individuals	online courses	https://about.citiprogram.org/courses/?fwp_series=242
World Organization for Animal Health (OIE)	Guidelines for Responsible Conduct in Veterinary Research – Identifying, Assessing and Managing Dual Use	individuals	information brochure	https://www.oie.int/app/uploads/2021/03/a-guidelines-veterinary-research.pdf

EU Non-Proliferation and Disarmament Consortium	EU Non-Proliferation and Disarmament eLearning Courses	individuals	e-learning modules	https://nonproliferation-elearning.eu
Sciensano, Service Biosafety and Biotechnology (SBB), BE	Belgian Biosafety Server	individuals	informational resources	<p>https://www.biosafety.be</p> <p>Biosecurity Office, NL Biosecurity Self-scan Toolkit, Vulnerability Scan, Dual Use Quickscan</p> <p>individuals, lecturers information resources, self-evaluation tools</p> <p>https://www.bureaubiosecurity.nl/en/about-biosecurity-office</p> <p>www.biosecurityselfscan.nl/mainMenu.html</p> <p>https://biosecurityvulnerabilityscan.nl</p> <p>https://dualusequickscan.com/</p>

4.4.5 Conclusion

The life sciences community needs to recognize that Dual Use considerations must be implemented into research project planning, which requires Dual Use education as a general element within the life sciences. A dialogue between all actors within the research ecosystem, including funding agencies, publishers, biorisk management policy makers from governmental or research institutes focusing on biorisk management, as well as the educational and research communities is needed to convince universities and research institutions of the necessity and benefits of such educational programs.

Independent of the mode of instruction, any Dual Use course for life scientists must communicate the importance of assessing Dual Use risks of research projects and emphasise the nature of this as a fundamental part of all research project planning. Researchers need to be aware of their responsibility to evaluate and mitigate potential misuse risks. Scientists should lead by example and pass on these values by teaching their students about Dual Use and the importance of the reflective process as an integral part of research project design.

To reach as many students as possible, education projects like the handbook “Preventing Biological Threats, What You Can Do” mentioned above are especially useful. Such ‘Teach the Teacher’ programs facilitate the implementation of courses on biosecurity and Dual Use issues at universities that currently do not have this expertise and trained personnel.

Based on the comparison of the different teaching formats, we recommend a two-stage Dual Use education approach. The first stage should be mandatory awareness raising at the undergraduate level. This can be a single lecture within a mandatory life sciences lecture series/course with the topic “Dual Use potential of life sciences research” that aims at drawing attention to the existence of Dual Use issues, as well as their consideration as a fundamental part in project planning. This first stage must reach every student and can only provide a short overview covering only select learning objectives (e.g., Learning objectives 1-8, Table 1).

For students further advanced in their degrees who started to design and work on their own research projects, or who want to engage with Dual Use issues more in-depth, we recommend a second stage education program such as a voluntary seminar. A specialised seminar is valuable for students who work on projects with a known or rather direct Dual Use potential. A full seminar at this second stage can also cover the learning objectives more broadly and focus on those with higher complexity (e.g., Learning objectives 9-13).

However, an inherent problem of Dual Use is that it is often not clearly identifiable. Even if a good initial Dual Use risk assessment took place, potential for misuse of results, methods or technologies might only become apparent over time, and thus projects require continuous evaluation of Dual Use risks. Researchers need to practice risk assessment on a regular basis. A

single lecture is probably neither sufficient to reach this level of awareness nor able to consider Dual Use in its whole complexity. Thus, a recommendation for mandatory in-depth courses seems appropriate especially for study programmes related to fields of research with high Dual Use potential. However, smaller voluntary seminars would be a good starting point. Ideally, students will act as multipliers by applying their knowledge and implementing risk assessment and mitigation measures into their professional life. Course sizes should be small to offer a learning environment that emphasises active participation in discussions. Therefore, seminars offer the ideal platform to incorporate Active Learning elements and discussions to establish the necessary foundational knowledge and to build a mindset of responsible research. These seminars can be offered by the universities directly or through other organisations with expertise in biorisk management. Semester-long seminars also offer the possibility to explore various aspects of responsible research, including societal, legal, biosafety and bioethics aspects. Students are usually also unfamiliar with these equally important aspects of life sciences and education on these topics will benefit from the same learning environment that facilitates learning about Dual Use issues.

Because the life sciences are inherently multidisciplinary, Dual Use risk courses can be designed to be relevant for other disciplines as well. Disciplines that influence the life sciences and enable progress in a particular area, like engineering, maths, computer science and chemistry, can be grouped together in a common course on Dual Use risks. Such a course needs to include the relevant control regimes and suitable case studies, which can lead to a mutual benefit and broader understanding of the relevance of Dual Use. Here, lessons learned from other disciplines might be adapted to solve existing challenges within the life sciences.

In general, new Dual Use teaching concepts need to be tested, which raises the question of how to evaluate Dual Use education. Past studies have used surveys similar to traditional course evaluations to assess if the students improved at answering key questions (Sarwar *et al.*, 2019; Arfin Qasmi *et al.*, 2021; Vinke *et al.*, 2022). Evaluations focused on knowledge increase do not necessarily investigate the students' ability to assess and mitigate Dual Use risks which are, as mentioned above, the most important learning objectives. More follow-up questionnaires focusing on the intended behavioural change and asking if and how participants integrate the knowledge in their work, how and where they might share it, would be very valuable (Arfin Qasmi *et al.*, 2021). Future evaluations should incorporate case studies on Dual Use projects with the task to assess risks and give recommendations how to mitigate the risks. Based on ongoing evaluations like this, teaching forms and learning objectives need to be adapted and the general recommendations from this study might need to be tailored to the individual course audience.

In summary, here we aim to encourage life science educators to actively engage in implementing teaching approaches within universities worldwide and to raise awareness about Dual Use issues. The rapid progress in life sciences research increases the Dual Use risks, and the widespread teaching about Dual Use issues will not be established by relying on the self-motivated lecturers alone. All stakeholders including scientists, the scientific community, ethics committee members, institutional/repository managers, biosafety officers, security officials, regulators, institutional and other authorities, civil society networks, and publics, politicians, industry, and academia need to work together on concepts and strategies how to fund and implement suitable training concepts to make life sciences researchers aware of Dual Use issues, as well as to make advances in this research area responsibly, safely, and securely.

5 Discussion

5.1 tRNA-aminacyl-synthetase evolution

5.1.1 Directed evolution of aaRS pairs

Directed evolution of tRNA-aminoacyl-synthetases, even if established decades ago, is still an imperfect system. Several methods have been developed, from the classical two-round selection system, focusing on the incorporation of the nsAA in a resistance in the positive selection and incorporation of the nsAA in a toxin in the negative selection like described in chapter 3.2.1, over PACE of the aaRS, described in 3.2.2,

The classical two-round selection system suffers mainly from the limitation in library design. The maximum of library size depends on the maximal achievable transformation efficiency, in commercial competent cells this can be up to $3 \cdot 10^9$ per vial (New England Biolabs, 2023). This number needs to be divided by the desired library coverage, for this work we assumed that 10-fold coverage ensures that most clones are represented. Furthermore, the maximal transformation efficiency is lowered by using larger plasmid. The measurement of the maximal transformation efficiency is usually done with pUC-19, a relatively small plasmid. The last influence on the maximal transformation efficiency is determined how many vial of competent cells can be transformed and plated. All in all, it is unrealistic to assume that libraries, encoded by a small plasmid, can be achieved in a standard lab that have more than $3 \cdot 10^{10}$ clones, smaller if a bigger plasmid is used. This would mean that only 7-8 amino acids could be randomized throughout the whole CDS. Randomizing several amino acid encoding codons quickly leads to a very big library, resulting in a tedious process when more than 8 codons needs to be randomized. The choice of amino acids to mutate also is a difficult choice. Classical libraries mainly focus on the amino acid residues directly involved in ligand binding. In this thesis it was shown that many other residues also have a high impact on activity and specificity. Amino acids in the binding pocket that are not in direct contact with the ligand, but with the amino acids that bind the ligand, have been shown to change specificity and activity. Also amino acids involved in tRNA binding have been shown to increase activity (Vinke *et al.*, 2023). Some aaRS selection protocols try to address this problem by first selecting a library with randomized amino acids that are directly involved in ligand binding, and then use the resulting clones as template for error-prone PCR. This has been shown to be a useful approach by Cervettini in 2020 (Cervettini *et al.*, 2020). This approach explores the unused space of activity and specificity increase by identifying other relevant residues, but most likely will not be able to identify a lot of co-dependent mutations, since the chance of those occurring is rather low.

All in all, Library cloning and iteration of consecutive positive and negative selection are a very laborious and time consuming process that most likely will not lead to the ideal clone. This limitation in library design can only be overcome by using a continuous mutation and selection process like PACE.

Another problem of the positive and negative selection is the choice of the ideal selection marker, while used commonly in positive selection, chloramphenicol has been shown to show a rather low dynamic range. This makes it difficult to separate based on activity if the activity of clone is not drastically higher than the background activity (Grasso *et al.*, 2022). Grasso et al. have shown that ampicillin shows an improved dynamic range and should thus be used as marker. The choice of the marker is also relevant based on what resistance the cell most likely will not achieve based on natural resistance development. By using an antibiotic resistance that is very rarely overcome by compensatory mutations of the host cell, the occurrence of false positive colonies can be reduced drastically.

The toxin in the negative selection of choice has been the barnase, but since this protein is very toxic to *E. coli*, it is rather difficult to clone. This resulted in some groups shifting to the use of inducible toxins like the *E. coli* uracil phosphoribosyltransferase UPRT encoded by the *upp* gene. UPRT is the natural *E. coli* uracil phosphoribosyltransferase and catalyses the conversion of 5-fluorouracil to 5-fluoro-dUMP, which inhibits thymidylate synthase, causing cell death (Melançon & Schultz, 2009). The use of this selection system requires the deletion of the *upp* gene from the *E. coli* genome.

All two round selection systems have the problem that they select for undesirable traits due to their nature of being split into positive and negative selection. While in the classic evolution this is less of a problem, since only aaRS clones get selected from an existent and non-changing pool of variants, this is a bigger problem during PACE. In the positive selection aaRSs that incorporate any amino acid are being enriched. Since most nsAA introduce bigger side chains that involve enlargement of the binding pocket, it is probably more likely for the aaRS to adapt to be able to bind the smaller canonical amino acids, that are more similar to their native ligand, or become fairly promiscuous and accept several aaRS in the binding pocket.

In the negative selection aaRSs get enriched which are specific or non-functional. Again the undesired trait, non-functionality, is a trait that is more likely to be achieved than perfect specificity for the nsAA.

This flaw in the selection system results in enrichment of clones that predominantly are false positives that need to be selected out by the other selection round. This most likely leads to never having the perfect combinations of both traits, running into local minima of activity and

specificity and a lot of host cells being “used up” by false positive variants. All of these flaws could be prevented by designing a one-round selection system that works in PACE

5.1.2 Deep-Sequencing guided evaluation of directed evolution experiments

Nanopore sequencing as evaluation of directed evolution has been proven to be an effective tool for the identification of variants with desired properties, but it is also an imperfect evaluation tool. In the experiments presented in this work, it has been shown that the high error-rate makes it impossible to distinguish between mutations and sequencing errors on individual variant level. However, in PACE the enrichment of the phages allows to filter for variant that show the same mutation. Thus, the mutations are less likely to be a sequencing error and instead are phage variants that enriched because it encodes for an aaRS with desired properties. This paired with the evaluation developed in this work, that filters out indel sequencing errors, enables the reliable evaluation of a high number of individual clones.

The described evaluation only has the disadvantage that true indel mutations, even if unlikely in the case of aaRS, would be filtered out since they are interpreted to be a sequencing error. It has been shown for PyIRS that a truncated enzyme is actually more active (Bryson *et al.*, 2017). This kind of mutation would not have been able to be identified with the current evaluation, which would have characterized it as an error. The evaluation has also been shown to not deal well with long homopolymers of the same base. For example the predicted Lys-deletion at amino acid position 26 is only an error in sequencing, because several codon encoded by AAA. This resulted in misinterpretation of a three-base pair deletion, which is accepted by the evaluation, since it does not cause a frameshift in the CDS (Vinke *et al.*, 2023). The evaluation also includes several PCR amplification steps which could falsify the ratio of aaRS mutants toward each other by introducing a PCR bias by enriching certain variants by chance or due to easier PCR amplification.

The challenges listed above can be addressed by using different approaches. The PCR bias could be avoided by directly sequencing isolated M13 ssDNA. The amplification of the aaRS was only favored because of the higher read rate, which can be addressed by optimizing phage DNA isolation and sequencing library preparation for low-input template concentration and ssDNA sequencing.

The increased error-rate can be addressed by using other sequencing methods like Pacbio sequencing since it also allows the sequencing of long reads (PacBio, 2023). It is more expensive and does not enable unnatural base sequencing which might be of interest in aaRS and tRNA design in the future, thus waiting for technological improvements in Flowcell design in Nanopore sequencing might be the better approach.

Another alternative to decrease error-rate is to establish rolling circle amplification for sequencing preparation. The circular single-stranded M13 phage DNA should pose a great template and rolling circle amplification has been shown to increase the accuracy in Nanopore sequencing drastically when used for smaller plasmids as template (Maguire & Guan, 2022). This method combined with better ssDNA isolation would offer a great opportunity to overcome the challenge of the proposed evaluation.

5.1.3 Challenges using PACE or PANCE

PACE or PANCE have proven to be very efficient in evolving enzymes if a suitable selection pressure can be established. The automated system is fast, and the continuous mutation and selection gets rid of all limitation in library design that other systems face, resulting in a very powerful tool in directed evolution. PACE however, is not perfect, some processes are still very laborious, especially the phage titer determination in form of plaque assays. To enable perfect automated regulation the system would need an online detection of the phages, that is then used as benchmark to regulate the flow rate, inducer concentration, and selection pressure.

It also would improve the system to use a targeted mutation mechanism, like EvolvR, a system that uses a nCas guided error-prone polymerase instead of a whole DNA replication targeting system and thus enables the mutation of a tunable DNA-window (Halperin *et al.*, 2018). Base editing systems like described by Tong *et al.* in 2023 could also help to target known residues involved in binding to increase the mutation rate at specific residues (Tong *et al.*, 2023).

Without online phage titer determination, it can come to phage washout, like it would have been the case of the evolution described in this work, but also described for e.g., the evolution of adenine base editors by Neugebauer *et al.* in 2022 (Neugebauer *et al.*, 2022). The washout is mainly based on low expression rates of pIII, which is caused by a very high selection pressure, either through the reaction taking too long or very low activities of the target enzyme. This can be partly solved by changing the design of the selection, e.g. the stringency can be reduced by editing of a T7 polymerase that then amplifies pIII, or introduction of drift plasmids, that enable a low constitutive pIII expression independent of the selection pressure (Bryson *et al.*, 2017; Carlson *et al.*, 2014). In general, the aaRS evolution seemed to have a too high selection pressure showing that changing to the T7-polymerase approach described by (Bryson *et al.*, 2017) is most likely a better approach than editing pIII directly. Changing to the pIII approach is an option once the aaRS is already shows moderate to high activity.

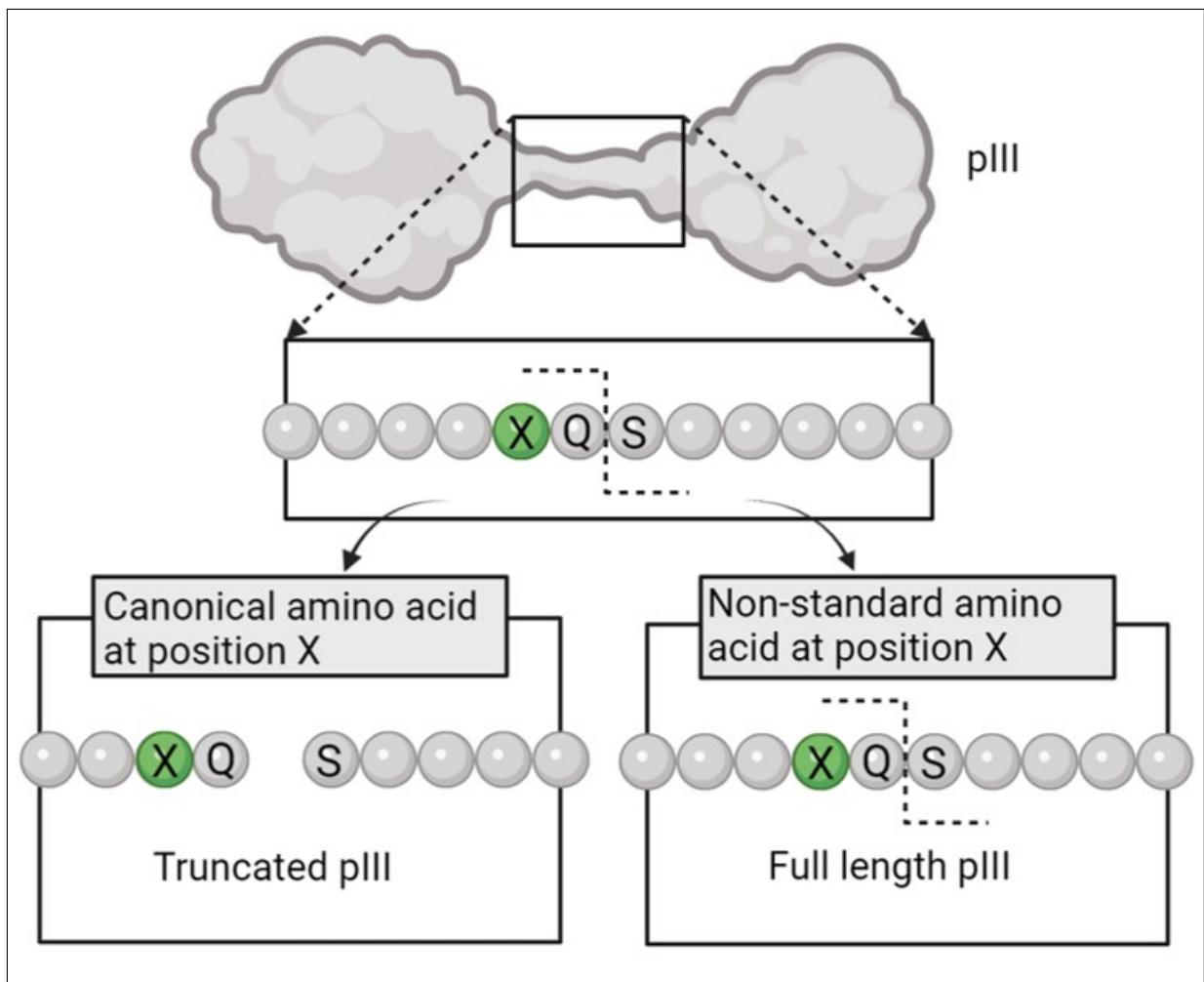


Figure 17: Schematic overview over the proposed one round selection system based on proteases. The protease has an amino acid residue in the cutting sequence where it accepts all canonical amino acid. If one of the canonical amino acid gets charged by the aaRS, the protease cuts pIII. Non-standard amino acids are to be big to allow the protease to still recognize the recognition sequence leading to no degradation of the pIII.

The main problem in the aaRS selection with PACE is the two round selection system, a way more effective evolution would be a one round selection. To make a one round selection system work, a mechanism capable of differentiating between standard and nsAAs is necessary. There are very few described like the N-terminal protein degradation based selection described by Kunjapur *et al.* in 2018, which is imperfect regarding the differentiating ability for some amino acids, or the biocontainment mechanism in essential genes developed by Mandell *et al.* in 2015 (Kunjapur *et al.*, 2018; Mandell *et al.*, 2015). This mechanism will likely be very hard to be coupled to pIII expression. However, there are processes that are very good in differentiating

between certain amino acids, one of them being protease recognition of cutting sites. This system could be used to degrade pIII when a standard amino acid is incorporated into the peptide change, and a bigger nsAA would block degradation. However, this system would only work for amino acids that show big structural differences towards the standard amino acids, while structurally similar amino acids still need to be evolved in a classic two-round selection system.

5.1.4 Alternative aaRS/tRNA pairs

The *M. jannaschii* TyrRS has been shown to be a powerful tool in genetic code expansion. It has been used to incorporate more than 100 different nsAAs (Liu & Schultz, 2010). However, there are still a lot of nsAA that so far have not been able to be incorporated, like glycosylated or amino acids with big fluorescent side chains. Also, nsbps in the anticodon seem to pose a big challenge to the commonly used aaRS pairs. It is not clear which part of the translational machinery, aaRS, EF-Tu or ribosome causes this problem, but for some nsbps, e.g., isoC and isoG in the anticodon, incorporation *in vitro* has been possible, showing that the problem lies in generating a functional aaRS (Bain *et al.*, 1992).

To generate a functional aaRS, the choice of evolution system is crucial as discussed and shown in this work, but a big part is also the choice of a suitable scaffold aaRS. The *M. jannaschii* TyrRS was used commonly due to a lot of beneficial features. It shows very high thermal stability, making it unlikely to destroy the enzyme even if a lot of amino acid residues are exchanged (Fechter *et al.*, 2001). It shows very little toxicity and it is small in size compared to other aaRS (Steer & Schimmel, 1999). However, it does posses a relatively small binding pocket located deep within the enzyme, a feature that increases specificity, but also might prevent the incorporation of big amino acids (Kobayashi *et al.*, 2003). These problems, plus the need for further orthogonal aaRS/tRNA pairs for the incorporation of multiple amino acids, led to the discovery and usage of several other aaRS/tRNA pairs for genetic code expansion, but the *M. jannaschii* TyrRS is still the most used one (Liu & Schultz, 2010).

The second most used aaRS/tRNA pairs are pyrrolysine aaRS (PylRS). PylRS are a class of enzymes derived from archaea, naturally capable of incorporating pyrrolysine encoded by an *amber* stop codon in these organism (Srinivasan *et al.*, 2002). The first, and most used pair of those is the *Methanosarcia mazei* PylRS. In the last years, different classes of PylRS have been identified at PylRS pairs that are distant enough in phylogeny have been identified that they are orthogonal towards each other (Willis & Chin, 2018). With this approach Willis *et al.* were able to generate several sets of three distinct PylRS pairs, capable of incorporating three different nsAAs in the same system. PylRS do have two major issues. They demonstrate bad thermostability, leading to approaches exploring to add solubility tags to increase their activity

at 37 °C (Koch *et al.*, 2021b). The bad thermostability leads to insolubility issues, wasting cell resources in recombinant protein production, but also makes it more likely that mutations lead to misfolding. Furthermore, PylRS have been limited to the use of pyrrolysine analogues, until in 2021 Koch *et al.* introduced a rational design strategy to enable to incorporate small aromatic amino acids as well (Koch *et al.*, 2021a).

Several research groups have tested a lot of other aaRS pairs, but few of them have shown better properties. A lot of aaRS have shown toxicity like the *Pyrococcus horikoshii* or phosphoserine aaRS, making it very hard to produce suitable amounts of nsAA containing protein, and also to select in positive and negative selection (Anderson *et al.*, 2004; Rogerson *et al.*, 2015). A non-toxic alternative has been the use of *Saccharomyces cerevisiae* pairs like the phenylalanyl and tryptophanyl aaRS/tRNA pairs (Ellefson *et al.*, 2014; Kwon *et al.*, 2006). Those showed activity, but have not been broadly used, partly because they are not orthogonal for the use in mammalian cells, but remain a potentially useful pair for another orthogonal pair if multiple nsAAs need to be incorporated.

The most promising candidate for future work, the *Archaeoglobus fulgidus* TyrRS, has been identified by Cervettini *et al.* by using the t-REX assay to screen and test aaRS for suitability and orthogonality in genetic code expansion (Cervettini *et al.*, 2020). The *A. fulgidus* TyrRS has been shown to be even more active than the *M. jannaschii* TyrRS in nsAA incorporation with very low toxicity (Cervettini *et al.*, 2020). It is also very stable and has been shown to be evolvable to incorporate nsAA, such as demonstrated by Cervettini *et al.* for the incorporation of iodo-phenylalanine (Kuratani *et al.*, 2006; Cervettini *et al.*, 2020).

The choice of a suitable aaRS always depends on the application, but considerations of additional aaRS pairs in aaRS design increase the chances of finding a suitable pair. The establishment of selection systems for several of them it will be necessary, given that in the future most likely multiple nsAAs shall be incorporated within the same cell. To achieve this aim, multiple orthogonal aaRS/tRNA pairs are needed.

For amino acids, that so far have not been able to be incorporated by aaRS, chemically charging has been established for *in vitro* assays (Robertson *et al.*, 1991). Recently, an alternative approach has been developed using RNA enzymes instead of chemically charging. The so called flexizymes are single stranded RNAs, capable of charging a vast range of nsAAs to tRNAs with a lot higher yields compared to the chemical charging assay (Murakami *et al.*, 2006; Tharp *et al.*, 2021). So far, flexizymes remain unspecific in regards to tRNA and nsAA, but it might be possible to use them directly in *in vitro* experiments without purification, or maybe even *in vivo*, when evolved to become specific regarding their ligand.

5.2 Alternatives to the *amber* codon

This work mainly made use of the *amber* stop codon to incorporate nsAAs. A process that has been proven to work well to incorporate nsAAs, but also has its limitations. *amber* stop codons, in presence of the release factor 1 still lead to terminations of translation, resulting in the competition between the release factor 1 (RF-1) and the *amber* suppressor tRNA in systems where RF-1 is not deleted (Schwark *et al.*, 2018). This leads to inhomogeneous protein yield. In case of RF-1 deletion, it has been shown that endogenous tRNAs can suppress the *amber* codon due to the wobble effect. This leads to incorporation of mainly glutamine and another competitor in nsAA incorporation at the *amber* stop codon position (Gan & Fan, 2017). The incorporation rate depends on the affinity of the tRNA to the anticodon and if there is a sufficient amount of charged *amber* suppressor tRNA with high suppression efficiency, these out-compete the wobbling tRNAs. Thus, the effect of misincorporation in nsAA are usually limited to a very small portion of the protein yield, making this system suitable for most basic research process, but if the production of the target protein or peptide requires highly homogeneous protein, these processes can cause problems. Even if the target protein can be produced with this system, *amber* codon suppression will always lead to influencing the endogenous translational machinery, leading to growth inhibition caused by elongation of peptides during translation that would usually be stopped by the *amber* codon.

An alternative or addition to the *amber* suppression is using other codons, like little used codons in the genome, e.g. the least used leucine codon, as demonstrated by Kwon *et al.* in 2016 for the incorporation of 2-naphthylalanine (Kwon & Choi, 2016). This approach faces the same problem as *amber* stop codons that it interferes a lot with the endogenous translation. An alternative is the use of quadruplet codons (Anderson *et al.*, 2004; Wang & Schultz, 2004). Quadruplet codons enable the incorporation of multiple amino acids, but usually are low yield production systems, that need to be combined with orthogonal ribosomes to ensure homogeneous yield (Gan & Fan, 2017; Dunkelmann *et al.*, 2020). Quadruplets in a non-orthogonal system leads to a lot of cross-talk with the endogenous translational machinery in both ways: Incorporation of nsAAs in endogenous proteins, and incorporation of canonical amino acids causing frameshifts in the target protein. Combined with the low yield, this system is only suitable in combination with orthogonal ribosomes optimized for quadruplet incorporation, but before the introduction of GROs, it was the only system allowing the incorporation of multiple nsAAs, resulting in vast popularity of this approach.

The concept of GROs enables the incorporation of multiple nsAAs in one cell without crosstalk with the endogenous machinery. The currently published GROs, C321, the AGR-deletion strain, and Syn61 Δ3 have demonstrated that the currently constructed GROs can not hold up to this

promise. For the recoding, *E. coli* MG1655 has been used, an *E. coli* strain with an already reduced genome, but still good growth, that can easily be manipulated for genome engineering (Blattner *et al.*, 1997). For recoding all CDSs have been annotated, and computationally all codons that are supposed to be recoded have been identified and exchanged to synonymous codons. These mutations have then been introduced at the Chin lab (Fredens *et al.*, 2019). The problem is that genome annotations are still not perfect. This is less a problem for C321, since *amber* codon suppression is tolerated from the cell even in non-recoded cells without causing a big growth deficit. It is a problem in Syn61 Δ3 and will most likely be a problem in rEcoli-57, since both rely on the same genome annotation. When the genome of Syn61 Δ3 is sequenced and compared with the most recent *E. coli* genome annotation, a lot of non-recoded genes occur, including several essential genes. This updates in genome annotation mainly are changes in the start of the CDS that have been previously mis-annotated. This leads to deletion or alternative start sites in these proteins, since the forbidden codons are not recoded. To make Syn61 Δ3 still able to grow, despite of stop codons in essential genes, either means that these genes are wrongly described as essential genes, unlikely for e.g. *mukE*, involved in cell separation, or Syn61 Δ3 being able to suppress these codons by other tRNAs through wobbling. Increased wobbling is a trait that likely evolved in Syn61 Δ3 since it increases fitness throughout the whole creation process of this strain, but this adaption will cause inhomogenous yield once the new blank codons are used for nsAA incorporation. The presence of the forbidden codons in essential genes also may prevent the use of certain nsAAs if these destroy the function of the essential genes once incorporated into that protein. In general the existent GROs with multiple codons need to be engineered to grow better, which includes the recoding of the remaining genes, but also a lot of regulatory elements, like promoter motifs are likely to be destroyed that need to be identified and optimized.

All these limitations make it very difficult to work with GROs at the moment. They grow significantly slower, genetic engineering is a lot harder, e.g. regeneration after transformation usually takes 8 h, and there are very few plasmids described to work in these strains, since they require to be recoded as well. The benefits of this system including the incorporation of only 3 amino acids, of whom one is recommended to be used for biocontainment, are not big enough to choose this system over already existing non-GRO systems for nsAA incorporation. It will likely take years to optimize the GRO strains to become sufficient for this. However, as demonstrated in this work, these strains are an outstanding tool for biocontainment, containment of genetic information and to block viral transfer. These abilities are valuable enough to justify optimization of these strains even if the workload is immense.

When it comes to incorporation of nsAAs it might be useful to consider other approaches such as the use of non-standard base pairs for encoding the nsAA. This approach, even if still at the very beginning, is unlikely to influence cell growth and engineering abilities like observed in GROs. It also allows the incorporation of way more nsAAs than GROs. To make this approach feasible, the whole DNA replication, transcription, and translational machinery has to be evolved to work with this new base pair. Proof-reading and degradation pathways might also need to be engineered, depending on the desired nsbp. This has been shown to be possible for isoC and isoG *in vitro* (Cherbuliez & Bernhard, 1932). The aaRS seems to be the problem when it comes to using isoC and isoG *in vivo* to encode for an amino acid (Cherbuliez & Bernhard, 1932). For base pairs with hydrophobic interaction it has been demonstrated *in vivo* that incorporation of a nsAA for a nsbp is possible (Bain *et al.*, 1992). For dNaM and 5SICS, the ribosome seems to have trouble accepting the nsAA in the first position, reducing the number of possible codons drastically. This is a big challenge since the third base is not an ideal position for the nsbp as well, because of its low specificity compared to the first and second base in the anticodon. This means without optimization of the ribosome only the incorporation of 16 nsAA would be possible without expected wobble effects. Both systems, nsbps with modified acceptor-donor pattern and the hydrophobic bases developed by the Romesberg lab pose a lot of potential once the cellular machinery is better optimized to incorporate those. However, isoC and isoG have the advantage of being more similar to endogenous bases, making it more likely to be able to establish a biosynthesis pathway. Biosynthesis of the bases lowers the cost for using this approach immensely, but also takes away the possibility of biocontainment on DNA level.

5.3 Biosecurity risks

5.3.1 General problems with risk assessment in Synthetic biology

Synthetic biology risk assessment is difficult, since risks can not be evaluated in standardized numbers, like e.g. in food security. So far, synthetic biology was not known to be used by malicious actors to cause harm, which gives us no data about what the aftermath of misuse is. Thus, biosecurity risks of synthetic biology can only be predicted, but not really objectively assessed. One expert might see the risk of a technology to be misused, while others see no risk at all. This problem is also increased by no uniform definitions of e.g. dual-use and gain-of-function research. Various national and international organizations use their own definition, often broad to be applicable to a lot of projects. This is useful in some situations, e.g. just

because a project does not fit a definition 100 % does not mean it does not cause harm, but makes it very difficult to discuss biosecurity issues on an international level.

Current regulations in general do not hold up to the challenges we will face in the future. Synthetic biology enables scientists to design potentially dangerous system without the use of gene synthesis of known pathogenic variants and benchtop synthesizers will soon enable gene synthesis without oversight. Next to that, synthetic biology becomes easier to use and thus more people are able to perform experiments, that decades ago only a handful of people could perform. This increases the number of potential malicious actors in this field. Also the BWC lacks a unit to investigate potential breaches of the treaty like the Chemical Weapons convention, making it hard to proof bioweapons use and thus prevents prosecution.

A few national and international organizations have published frameworks that are supposed to assist scientists in assessing risks, but most scientists are not aware of biosecurity risks in general, and even less of the existence of those frameworks. In general it is debatable if a framework that is designed in a way to be applicable for a large number of research projects is the best way to identify risks. This approach could maybe even poses a loophole for continuing with a specific project that poses a biosecurity risk with the justification that the risks assessed in the general frameworks do not apply for this project.

In general the best system for risk reduction will always be a multiple layer approach, starting with the scientist in the project planning phase, internal biosecurity boards, over funding, publishing and responsible use of the technology.

5.3.2 Responsibility

The current structure for risk mitigation is not sufficient for effective risk mitigation, but the establishment of new considerations results in the question who is responsible to ensure this process takes place. One could argue it is the responsibility of each UN-country under UNSCR 1450 to ensure that individuals are not able to misuse technologies. Since most countries rely on the self-regulatory approach for biosecurity-risk mitigation combined with existing biosafety regulations, each of those countries would need to ensure that scientists are aware and educated to ensure risk mitigation. This is a problem in countries where educational institutions are not obligated to integrate curricula activities that are suggested by the politic. A law that ensures the freedom of education, but in this case makes it the universities responsibility to integrate biosecurity education, the government can only suggest to do so. Most universities struggle with integration biosecurity courses, for several reasons, including unawareness of the importance, already packed curricula and the lack of lecturers with experience in that area. This can only be solved by reaching out to universities, offering to help out with external experts and funding

for individual researchers and groups working on biosecurity issues. A process that can be initiated by the government by offering funds for biosecurity research for universities and non-proliferation think tanks.

Another problem is the lack of contact persons in case an individual researcher or student identifies a biosecurity risk. Often universities offer no information on who to contact in case of concerns and especially early career researcher will not report issues in case they fear repercussions.

All in all, it is the responsibility of each researcher to ensure their research is beneficial for society, but most are not aware of biosecurity risks. This can only be changed by education. Putting the universities in the place of providing this education if they are interested to educate researchers to be responsible individuals. The government is responsible to support this process, which is easiest by providing funds. In general all actors involved in the process of a scientific project are responsible for its effect on society. A sufficient biosecurity risk mitigation process will only be possible if all actors, from researchers, over universities, funding agencies, and publishers will ask, expect, and support risk assessment. Under UNSCR 1450 it is in the responsibility of the government to ensure this process takes place.

5.3.3 Biosecurity risks of Xenobiology

Xenobiology poses an increased biosecurity risk because it changes the main biological compartments of life, which makes it more difficult to detect with the methods we have established to detect biopolymers.

When it comes to Xeno-DNA, the detection poses a great problem. DNA with non-standard bases or artificial backbones can not be sequenced with Sanger, Illumina or PacBio-Sequencing (Yan *et al.*, 2019). The only available sequencing method that can detect them at the moment would be Nanopore-Sequencing, but even Nanopore-Sequencing might not be possible if the base modification is so big that it clogs the pores. Even if a non-standard base pair is sequenced with Nanopore, the signal could be close to the standard base and it is most likely to be miscalled as one of the standard bases and thus not detected as being part of another DNA encoding scheme. The encoding abilities also enable that organism might be used to smuggle information, even though this is a risk posed by natural DNA as well.

The modification of the backbone, as well as the incorporation of certain nsAAs, can also increase the stability or reduce the stability of biopolymers, enabling the sabotage of production or introducing stability into toxins. Proteins that include D-amino acids for example are not able to be degraded by proteases anymore (Feng & Xu, 2016).

In general nsAAs pose a high risk to be used as toxins since a lot of fungi toxins are small polypeptides with non-standard amino acids produced by non-ribosomal processes (Martínez-Núñez & López, 2016). This process could be mimicked with nsAAs incorporated by *E. coli* to produce these toxins in a higher scale. nsAAs can also be used to hide proteins from the immune system or influence it, since they can be used to mimic PTMs and in the future might also be used to enable the presentations of customized glycans on the protein surface.

Recoded organism can be used as host for the production of nsAA containing proteins, but their bigger biosecurity risk is posed by their virus resistance. This is a problem since they are not sensitive to phage-therapy anymore.

5.3.4 Necessary mitigation steps for biosecurity risks

The mitigation of the listed biosecurity risks is probably a lot easier than of other projects, e.g. gain-of-function experiments in vaccine research. The main aspect that needs to be built in as a safety mechanism is the awareness of researchers that it is their responsibility to consider risks in their own research, since at the moment, they are the only effective stakeholder in risk mitigation. They should also be encouraged to take part in meetings where the risk of emerging technologies is discussed, so they can report the state of research and at the same time give recommendations how biorisk assessment and mitigation frameworks can be changed to best react to this new research discipline. One example for this is the emerging technology session of the biological weapons convention. This way they would be in direct contact with stakeholders that can directly implement precautions, e.g. law enforcement, which would be the first responder in case a bioweapons use is suspected.

To best mitigate the risks listed above, a few actions can be directly implemented to best prevent these risks. The first one is the update of technologies for detection of biopolymers to also be suitable to detect xeno-biopolymers. On protein level this can be done by adapting MS-software. Post-translational modification detection applications of programs like Excalibur can be used to identify nsAAs as demonstrated in this work. Therefore, nsAAs are added as new PTMs to the program which is then able to search for them in specific peptides. If searched in the whole proteome a lot of computational power is required. This way of detection can easily be integrated in routine MS-study of proteome, peptide and protein samples.

On DNA-level screening can be easily implemented by using Nanopore sequencing. Recently, companies and governments try to establish approaches for better and more broad biosurveillance (The White House, 2022; Concentric by Ginkgo, 2023). By identifying over-proportional enrichment of DNA sequences, potential threats can be better identified. Especially samples like wastewater and hospital wastewater are of interest. It is relatively easy to isolate DNA

from those samples and we would expect to find DNA of (human) pathogens enriching there in case of an epi- or pandemic (The Nucleic Acid Observatory Consortium, 2021). Working in cooperation with the already existing initiatives for biosurveillance and recommend those to add Nanopore-sequencing in case they use another sequencing method is an easy first step. This should enable cost-effective and reliable surveillance for non-standard bases and backbone modifications, just by keeping an eye on the quality score of the base editor.

Xenobiology research using *E. coli* falls under the category of non-pathogen related research. Thus, DNA screening methods would not flag any of ordered synthetic DNA as potentially dangerous. Which given the current risk analysis is a reasonable assessment, but when xenobiology becomes easier to use and thus more individuals would be enabled to access and use xenobiology, the risk of misuse increases. In case we see this increased risk, screening mechanisms need to be updated from a “list” to a “risk” based approach. This would include changing risk assessment from synthetic DNA screening for genes of known pathogens to an reliable assessment of what the synthetic DNA is most probably be used for and assessing the risk of the use scenario. This is a general approach that needs to be implemented in DNA screening, also for other disciplines, but has so far been challenging, since no one seems to know how. There have been approaches by several DNA synthesis companies to improve risk assessment done by these companies which led to the founding of the International Gene Synthesis Consortium in 2009 to work on these challenges (International Gene Synthesis Consortium — The Promotion of Biosecurity, 2017).

Another aspect that needs to be taken into consideration in DNA screening is to train software to recognize DNA-codon shuffling in GROs. In GROs, empty codons can be used to encode for other amino acids, as demonstrated by the leucine-serine swap (Nyerges *et al.*, 2023). In theory even more codons could be “shuffled” to encode for other amino acids, and thus reducing the similarity score for the ordered DNA sequence to the initial pathogenic variants, even if the protein is still identical. The software needs to be able to detect the shuffling of at least 7 DNA codons since rEcoli-57 will have 7 codons exchanged. This requires more computational power, but should be a relatively easy to implement in current DNA screening mechanisms.

Next to these technical changes, the researchers who are involved in advancing this field need to be better connected to biosecurity policy makers to get a voice in risk prediction and mitigation suggestions. Therefore, scholarships for young professionals to work on the interface of science and policy would be a beneficial, but also inviting lead scientists of the field to biosecurity meetings would substantially improve the current situation.

In general the mindset of the scientific community has to change from focusing just on the “hard”-science in the lab to see their research as strongly connected with society to be able

to better analyze societal implications of their work in the laboratory. This shift of mindset would necessarily include the perception of biosafety and biosecurity risks as fundamental part of research and not an administrative burden.

6 Conclusion

nsAA incorporation is currently the most promising xenobiology discipline for a high commercial value application, since nsAA incorporation would enable to design and produce a lot of novel biopharmaceuticals or lower the existing production cost. To enable nsAA incorporation on industrial scale, *in vivo* incorporation needs to be optimized. Protein production of homogeneous recombinant proteins with nsAAs requires an optimal aaRS and expression host with suitable growth properties for large scale production. To achieve this, GROs would need to be drastically optimized, including recoding of the non-recoded genes, optimization of regulatory elements destroyed during recoding and strain optimization for improved growth and engineering properties. GROs with multiple exchanged codons show impaired growth, but offer unique opportunities for biocontainment and the blocking of viral infection. This makes these organisms very interesting for a lot of applications, once they are optimized.

An alternative would be to design *E. coli* strains capable of incorporating the nsAA encoded by a nsbp. These strains would need to be optimized regarding their abilities to retain the nsbp in their DNA and adapt the translational machinery to accept the nsbp. This system would have the advantage of not changing the genome as much as GROs, resulting in very little growth reduction compared to GROs. It also enables the incorporation of multiple amino acids.

Both systems require also an optimal aaRS, which is highly active and specific. The work presented in this thesis has demonstrated that continuous mutation and selection paired with deep-sequencing offers a huge potential for aaRS design. It has also been shown that the used PACE system is imperfect because of the split in two selection rounds. This limitation can only be overcome by designing a one-round selection system, like proposed in chapter 5.1.3.

Xenobiology is still in the beginning of this research field, with a lot of potential still unlocked. This will change in the future. To make this change in a responsible way, biosecurity precautions need to be updated at the same pace that the field grows. To ensure this, scientists need to be educated to be able to assess and mitigate biosecurity risks. They also need to be included in biosecurity policy to give recommendation on how to mitigate risks, predict future directions of this research discipline and advise on which regulations are useful and can be realistically implemented.

7 Acknowledgments

I love my PhD topic, but doing a PhD is hard, and to get in the fortunate position to become a PhD student in the first place is also not a given. I would not have started this journey, nor submitted this thesis, without an incredible support system.

The most important part of this system are my parents, Eva and Torsten. All of this would have been impossible without you. You always supported me and my decisions, encouraged me to think big, to be creative, and to value education. I feel so loved from the two of you and I am grateful for all the chances you two enabled me in life. I love you very much!

My parents have been my main support system in all this, but the rest of my family sure was a big factor as well, from my aunt Dodo, great-grandma Uroma Frieda, my grandparents Oma Liesschen, Opa Dieter, Oma Erna and Opa Gerd, my cousin Tascha, to my little brother Erik, all of you have always encouraged me from being very little to go for my dreams and support me wherever you could. I am so grateful for all your support!

Next to family there is also chosen family. Thank you Rami for supporting me so much. I could not wish for a kinder, smarter and more supportive partner. I am the luckiest person in the world to have you. I can't wait for our joint projects in the future and I love you so much!

My family has been joined by a lot of people that I got to know when I started my PhD. The first person I would like to thank here is my supervisor Jörn. You have been the best supervisor I could have wished for! I never met any other supervisor that saw his main task in helping their supervisees grow the way you do. I would never have become the scientist I am today without your support throughout so many years. You encouraged me to go for my own ideas, to go abroad, to think out of the box. I hope I can give what you gave to me to people I supervise one day. Thank you!

I would also like to thank George Church, and Ehmke and Steffi Pohl for being the best mentors and supervisors I could wish to work with on my stays abroad.

I also had friends on this journey that became so close that I consider them family. Laura, Levin and Carina, thank you for helping me through my PhD (and Bachelor's and Master's). You are the best listeners, and advice giver's, the best emotional support, and the most beautiful souls any person could wish to be surrounded with in their lives. Thank you for being there for me.

I also would like to thank all my Boston friends, especially Caro, Piers and Tessa, my Volti girls and Martina, the rest of the Microbial genomics and biotechnology group, especially Christian, the Bielefelder Mbtler, and my iGEM team(s), especially Yannic, Chris and Markus, for being part of my life and my PhD journey. All interactions and all your kindness made my life better, helped me, and taught me on my way. Thank you for this!

References

1. Abe, T., Ikemura, T., Sugahara, J., Kanai, A., Ohara, Y., Uehara, H., Kinouchi, M., Kanaya, S., Yamada, Y., Muto, A. & Inokuchi, H. tRNADB-CE 2011: tRNA gene database curated manually by experts. *Nucleic Acids Research* **39**, D210–3 (2011).
2. Abedon, S. T. & Yin, J. Bacteriophage plaques: theory and analysis. *Methods in molecular biology (Clifton, N.J.)* **501**, 161–174 (2009).
3. Abrahão, J., Silva, L., Silva, L. S., Khalil, J. Y. B., Rodrigues, R., Arantes, T., Assis, F., Boratto, P., Andrade, M., Kroon, E. G., Ribeiro, B., Bergier, I., Seligmann, H., Ghigo, E., Colson, P., Levasseur, A., Kroemer, G., Raoult, D. & La Scola, B. Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nature communications* **9**, 749 (2018).
4. Al-Shayeb, B., Sachdeva, R., Chen, L.-X., Ward, F., Munk, P., Devoto, A., Castelle, C. J., Olm, M. R., Bouma-Gregson, K., Amano, Y., He, C., Méheust, R., Brooks, B., Thomas, A., Lavy, A., Matheus-Carnevali, P., Sun, C., Goltsman, D. S. A., Borton, M. A., Sharrar, A., Jaffe, A. L., Nelson, T. C., Kantor, R., Keren, R., Lane, K. R., Farag, I. F., Lei, S., Finstad, K., Amundson, R., Anantharaman, K., Zhou, J., Probst, A. J., Power, M. E., Tringe, S. G., Li, W.-J., Wrighton, K., Harrison, S., Morowitz, M., Relman, D. A., Doudna, J. A., Lehours, A.-C., Warren, L., Cate, J. H. D., Santini, J. M. & Banfield, J. F. Clades of huge phages from across Earth’s ecosystems. *Nature* **578**, 425–431 (2020).
5. Alamos, P., Tello, M., Bustamante, P., Gutiérrez, F., Shmaryahu, A., Maldonado, J., Levicán, G. & Orellana, O. Functionality of tRNAs encoded in a mobile genetic element from an acidophilic bacterium. *RNA Biology* **15**, 518–527 (2018).
6. Alexander, D. L., Lilly, J., Hernandez, J., Romsdahl, J., Troll, C. J. & Camps, M. Random mutagenesis by error-prone pol plasmid replication in Escherichia coli. *Methods in molecular biology (Clifton, N.J.)* **1179**, 31–44 (2014).
7. Alexander, K. UNODA Treaties 2023. <https://treaties.unoda.org/t/bwc>.
8. Anderson, J. C. & Schultz, P. G. Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression. *Biochemistry* **42**, 9598–9608 (2003).
9. Anderson, J. C., Wu, N., Santoro, S. W., Lakshman, V., King, D. S. & Schultz, P. G. An expanded genetic code with a functional quadruplet codon. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 7566–7571 (2004).

10. Anderson, L. & Krathwohl, D. R. *A taxonomy for learning, teaching and assessing. A revision of Bloom's taxonomy of educational objectives* (Longman, New York, 2001).
11. Arfin Qasmi, S., Sarwar, S. & Azheruddin, M. Capacity building for the identification, mitigation, and communication of DURC in Pakistan: A cross-sectional study. *Journal of Biosafety and Biosecurity* **3**, 141–146 (2021).
12. Ashton, P. M., Nair, S., Dallman, T., Rubino, S., Rabsch, W., Mwaigwisya, S., Wain, J. & O'Grady, J. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nature Biotechnology* **33**, 296–300. https://idp.nature.com/authorize/casa?redirect_uri=https://www.nature.com/articles/nbt.3103&casa_token=jerugej_wuiaaaaaa:qb5-qidalaoi6ckyuq6ty_7ffymmaxb7kean4zw3ny_fakahnmo7dy3vle29hjeepxtfbnukzzqxt3yb (2015).
13. Australia Group. *Australia Group common control lists* (ed Australia Group) 2007. <https://www.dfat.gov.au/publications/minisite/theaustraliagroupnet/site/en/controllists.html>.
14. Badran, A. H. & Liu, D. R. Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nature communications* **6**, 8425 (2015).
15. Bae, J. H., Rubini, M., Jung, G., Wiegand, G., Seifert, M. H. J., Azim, M. K., Kim, J. S., Zumbusch, A., Holak, T. A., Moroder, L., Huber, R. & Budisa, N. Expansion of the genetic code enables design of a novel gold class of green fluorescent proteins. *Journal of Molecular Biology* **328**, 1071–1081 (2003).
16. Bain, J. D., Switzer, C., Chamberlin, A. R. & Benner, S. A. Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code. *Nature* **356**, 537–539 (1992).
17. Baltz, R. H. Bacteriophage-resistant industrial fermentation strains: from the cradle to CRISPR/Cas9. *Journal of industrial microbiology & biotechnology* **45**, 1003–1006 (2018).
18. Barone, P. W., Wiebe, M. E., Leung, J. C., Hussein, I. T. M., Keumurian, F. J., Bouressa, J., Brussel, A., Chen, D., Chong, M., Dehghani, H., Gerentes, L., Gilbert, J., Gold, D., Kiss, R., Kreil, T. R., Labatut, R., Li, Y., Müllberg, J., Mallet, L., Menzel, C., Moody, M., Monpoeho, S., Murphy, M., Plavsic, M., Roth, N. J., Roush, D., Ruffing, M., Schicho, R., Snyder, R., Stark, D., Zhang, C., Wolfrum, J., Sinskey, A. J. & Springs, S. L. Viral contamination in biologic manufacture and implications for emerging therapies. *Nature Biotechnology* **38**, 563–572 (2020).

19. Baumgartner, E. Viruses spread by insects to crops sound scary - The military calls it food security. *New York Times*. <https://www.nytimes.com/2018/10/04/science/darpa-gene-editing.html> (2018).
20. Bextine, B. *Insect allies* (ed Defense Advanced Research Projects Agency) 2018. <http://www.darpa.mil/program/insect-allies>.
21. Biggs, J. Enhancing teaching through constructive alignment. *Higher Education* **32**, 347–364 (1996).
22. Biggs J. & Tang, K. *Teaching for Quality Learning at University* (McGraw-Hill Education, UK, 2011).
23. Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. & Shao, Y. The complete genome sequence of Escherichia coli K-12. *Science (New York, N.Y.)* **277**, 1453–1462 (1997).
24. Boeke, J. D., Church, G., Hessel, A., Kelley, N. J., Arkin, A., Cai, Y., Carlson, R., Chakravarti, A., Cornish, V. W., Holt, L., Isaacs, F. J., Kuiken, T., Lajoie, M., Lessor, T., Lunshof, J., Maurano, M. T., Mitchell, L. A., Rine, J., Rosser, S., Sanjana, N. E., Silver, P. A., Valle, D., Wang, H., Way, J. C. & Yang, L. GENOME ENGINEERING. The Genome Project-Write. *Science* **353**, 126–127 (2016).
25. Bonilla, N., Rojas, M. I., Netto Flores Cruz, G., Hung, S.-H., Rohwer, F. & Barr, J. J. Phage on tap-a quick and efficient protocol for the preparation of bacteriophage laboratory stocks. *PeerJ* **4**, e2261 (2016).
26. Borges, A. L., Lou, Y. C., Sachdeva, R., Al-Shayeb, B., Penev, P. I., Jaffe, A. L., Lei, S., Santini, J. M. & Banfield, J. F. Widespread stop-codon recoding in bacteriophages may regulate translation of lytic genes. *Nature Microbiology* **7**, 918–927 (2022).
27. Bossi, L. Context effects: translation of UAG codon by suppressor tRNA is affected by the sequence following UAG in the message. *Journal of Molecular Biology* **164**, 73–87 (1983).
28. Bowden, R. J., Simas, J. P., Davis, A. J. & Efstathiou, S. Murine gammaherpesvirus 68 encodes tRNA-like sequences which are expressed during latency. *The Journal of general virology* **78 (Pt 7)**, 1675–1687 (1997).
29. Braun, W. & Venkatarajan, M. S. New quantitative descriptors of amino acids based on multidimensional scaling of a large number of physical?chemical properties. *Molecular modeling annual* **7**, 445–453 (2001).

30. Brok-Volchanskaya, V. S., Kadyrov, F. A., Sivogrivov, D. E., Kolosov, P. M., Sokolov, A. S., Shlyapnikov, M. G., Kryukov, V. M. & Granovsky, I. E. Phage T4 SegB protein is a homing endonuclease required for the preferred inheritance of T4 tRNA gene region occurring in co-infection with a related phage. *Nucleic Acids Research* **36**, 2094–2105 (2008).
31. Bryson, D. I., Fan, C., Guo, L.-T., Miller, C., Söll, D. & Liu, D. R. Continuous directed evolution of aminoacyl-tRNA synthetases. *Nature Chemical Biology* **13**, 1253–1260 (2017).
32. Budisa, N., Kubyshkin, V. & Schmidt, M. Xenobiology: A Journey towards Parallel Life Forms. *Chembiochem : a European journal of chemical biology* **21**, 2228–2231 (2020).
33. Bustamante, P. ICEAfe1, an actively excising genetic element from the biomining bacterium Acidithiobacillus ferrooxidans. *J. Mol. Microbiol. Biotechnol.* **22**, 399–407 (2012). *J. Mol. Microbiol. Biotechnol.* **22**, 399 (2012).
34. Camps, M., Naukkarinen, J., Johnson, B. P. & Loeb, L. A. Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 9727–9732 (2003).
35. Capecchi, M. R. Polypeptide chain termination in vitro: isolation of a release factor. *Proceedings of the National Academy of Sciences of the United States of America* **58**, 1144–1151 (1967).
36. Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A. & Liu, D. R. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nature Chemical Biology* **10**, 216–222 (2014).
37. Cellitti, S. E., Jones, D. H., Lagpacan, L., Hao, X., Zhang, Q., Hu, H., Brittain, S. M., Brinker, A., Caldwell, J., Bursulaya, B., Spraggon, G., Brock, A., Ryu, Y., Uno, T., Schultz, P. G. & Geierstanger, B. H. In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *Journal of the American Chemical Society* **130**, 9268–9281 (2008).
38. Centre for Biosecurity and Biopreparedness. *Responsible science, ethics and biosecurity* 2021. <https://www.biosecurity.dk/english/educational-activities/responsible-science-ethics-and-biosecurity>.

39. Cerchione, D., Loveluck, K., Tillotson, E. L., Harbinski, F., DaSilva, J., Kelley, C. P., Keston-Smith, E., Fernandez, C. A., Myer, V. E., Jayaram, H. & Steinberg, B. E. SMOOT libraries and phage-induced directed evolution of Cas9 to engineer reduced off-target activity. *PLoS ONE* **15**, e0231716 (2020).
40. Cervettini, D., Tang, S., Fried, S. D., Willis, J. C. W., Funke, L. F. H., Colwell, L. J. & Chin, J. W. Rapid discovery and evolution of orthogonal aminoacyl-tRNA synthetase-tRNA pairs. *Nature Biotechnology* **38**, 989–999 (2020).
41. Chan, P. P., Lin, B. Y., Mak, A. J. & Lowe, T. M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Research* **49**, 9077–9096 (2021).
42. Chatterjee, A., Xiao, H. & Schultz, P. G. Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in *Escherichia coli*. *Proceedings of the National Academy of Sciences* **109**, 14841–14846 (2012).
43. Chen, J. J., Tsai, C.-H., Cai, X., Horhota, A. T., McLaughlin, L. W. & Szostak, J. W. Enzymatic primer-extension with glycerol-nucleoside triphosphates on DNA templates. *PLoS ONE* **4**, e4949 (2009).
44. Chen, S., Schultz, P. G. & Brock, A. An improved system for the generation and analysis of mutant proteins containing unnatural amino acids in *Saccharomyces cerevisiae*. *Journal of Molecular Biology* **371**, 112–122 (2007).
45. Chen, Y., Hysolli, E., Chen, A., Casper, S., Liu, S., Yang, K., Liu, C. & Church, G. Multiplex base editing to convert TAG into TAA codons in the human genome. *Nature communications* **13**, 4482 (2022).
46. Cherbuliez, E. & Bernhard, K. Croton seed (1) crotonoside. *Helv. Chim. Act*, 978–982 (1932).
47. Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of *Escherichiacoli*. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 11020–11024 (2002).
48. Chosewood, L. C. & Wilson, D. E. *Biosafety in microbiological and biomedical laboratories* 5th ed. (U.S. G.P.O. and For sale by the Supt. of Docs., U.S. G.P.O, Washington, 2010).
49. Church, G., Baynes, B. & Pitcher, E. *patent* (2007).
50. Church, G. M. & Regis, E. *Regenesis: How Synthetic Biology Will Reinvent Nature and Ourselves* (Basic Books, 2014).

51. Chyba, C. F. & Hand, K. P. Astrobiology: The Study of the Living Universe. *Annual Review of Astronomy and Astrophysics* **43**, 31–74 (2005).
52. Cirino, P. C., Tang, Y., Takahashi, K., Tirrell, D. A. & Arnold, F. H. Global incorporation of norleucine in place of methionine in cytochrome P450 BM-3 heme domain increases peroxygenase activity. *Biotechnology and Bioengineering* **83**, 729–734 (2003).
53. Claverie, J. M. Detecting frame shifts by amino acid sequence comparison. *Journal of Molecular Biology* **234**, 1140–1157 (1993).
54. Concentric by Ginkgo. *Home - Concentric by Ginkgo* 2023. <https://www.concentricbyginkgo.com/>.
55. Cook, D. A. & Artino, A. R. Motivation to learn: an overview of contemporary theories. *Medical Education* **50**, 997–1014 (2016).
56. Cox, E. C. & Horner, D. L. DNA sequence and coding properties of mutD(dnaQ) a dominant Escherichia coli mutator gene. *Journal of Molecular Biology* **190**, 113–117 (1986).
57. Coxon, K. M., Chakauya, E., Ottenhof, H. H., Whitney, H. M., Blundell, T. L., Abell, C. & Smith, A. G. Pantothenate biosynthesis in higher plants. *Biochemical Society transactions* **33**, 743–746 (2005).
58. Cozens, C. & Pinheiro, V. B. Darwin Assembly: fast, efficient, multi-site bespoke mutagenesis. *Nucleic Acids Research* **46**, e51 (2018).
59. Dai, J., Boeke, J. D., Luo, Z., Jiang, S. & Cai, Y. Sc3.0: revamping and minimizing the yeast genome. *Genome Biology* **21**, 205 (2020).
60. Deatherage, D. E. & Barrick, J. E. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods in molecular biology (Clifton, N.J.)* **1151**, 165–188 (2014).
61. Deiters, A., Groff, D., Ryu, Y., Xie, J. & Schultz, P. G. A genetically encoded photocaged tyrosine. *Angewandte Chemie International Edition* **45**, 2728–2731 (2006).
62. Deiters, A. & Schultz, P. G. In vivo incorporation of an alkyne into proteins in Escherichia coli. *Bioorganic & Medicinal Chemistry Letters* **15**, 1521–1524 (2005).
63. Deutsche Forschungsgemeinschaft. *Zum Umgang mit Sicherheitsrelevanter Forschung* 2023. https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/sicherheitsrelevante_forschung/index.html.

64. Dunkelmann, D. L., Willis, J. C. W., Beattie, A. T. & Chin, J. W. Engineered triply orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids. *Nature Chemistry* **12**, 535–544 (2020).
65. Durkin, E. *US plan to genetically alter crops via insects feared to be biological war plan* (ed The Guardian) 2018. <https://www.theguardian.com/environment/2018/oct/04/us-plan-to-genetically-alter-crops-via-insects-feared-to-be-biological-war-plan>.
66. Ellefson, J. W., Meyer, A. J., Hughes, R. A., Cannon, J. R., Brodbelt, J. S. & Ellington, A. D. Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nature Biotechnology* **32**, 97–101 (2014).
67. Elworth, R. A. L., Diaz, C., Yang, J., de Figueiredo, P., Ternus, K. & Treangen, T. Synthetic DNA and biosecurity: Nuances of predicting pathogenicity and the impetus for novel computational approaches for screening oligonucleotides. *PLOS Pathogens* **16**, e1008649 (2020).
68. Eschenmoser, A. Chemical etiology of nucleic acid structure. *Science* **284**, 2118–2124 (1999).
69. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011).
70. European Union. *Directive 2000/54/EC - biological agents at work — Safety and health at work EU-OSHA* 2000. <https://osha.europa.eu/en/legislation/directives/exposure-to-biological-agents/77>.
71. Evans, N. G. & Selgelid, M. J. Biosecurity and Open-Source Biology: The Promise and Peril of Distributed Synthetic Biological Technologies. *Science and Engineering Ethics* **21**, 1065–1083 (2015).
72. Fechter, P., Rudinger-Thirion, J., Tukalo, M. & Giegé, R. Major tyrosine identity determinants in *Methanococcus jannaschii* and *Saccharomyces cerevisiae* tRNA(Tyr) are conserved but expressed differently. *European journal of biochemistry* **268**, 761–767 (2001).
73. Fekner, T., Li, X., Lee, M. M. & Chan, M. K. A pyrrolysine analogue for protein click chemistry. *Angewandte Chemie International Edition in English* **48**, 1633–1635 (2009).
74. Feng, Z. & Xu, B. Inspiration from the mirror: D-amino acid containing peptides in biomedical approaches. *Biomolecular concepts* **7**, 179–187 (2016).

75. Fischer, E. C., Hashimoto, K., Zhang, Y., Feldman, A. W., Dien, V. T., Karadeema, R. J., Adhikary, R., Ledbetter, M. P., Krishnamurthy, R. & Romesberg, F. E. New codons for efficient production of unnatural proteins in a semisynthetic organism. *Nature Chemical Biology* **16**, 570–576 (2020).
76. 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., Uttamapinant, C., Llamazares, A. G., Elliott, T. S. & Chin, J. W. Total synthesis of Escherichia coli with a recoded genome. *Nature* **569**, 514–518 (2019).
77. Fromant, M., Blanquet, S. & Plateau, P. Direct random mutagenesis of gene-sized DNA fragments using polymerase chain reaction. *Analytical biochemistry* **224**, 347–353 (1995).
78. Fujino, T., Tozaki, M. & Murakami, H. An Amino Acid-Swapped Genetic Code. *ACS Synthetic Biology* **9**, 2703–2713 (2020).
79. Furter, R. Expansion of the genetic code: site-directed p-fluoro-phenylalanine incorporation in Escherichia coli. *Protein Science* **7**, 419–426 (1998).
80. Gan, Q. & Fan, C. Increasing the fidelity of noncanonical amino acid incorporation in cell-free protein synthesis. *Biochimica et biophysica acta. General subjects* **1861**, 3047–3052 (2017).
81. Giegé, R., Sissler, M. & Florentz, C. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Research* **26**, 5017–5035 (1998).
82. Gommers-Ampt, J. H. & Borst, P. Hypermodified bases in DNA. *FASEB journal* **9**, 1034–1042 (1995).
83. Goodall, E. C. A., Robinson, A., Johnston, I. G., Jabbari, S., Turner, K. A., Cunningham, A. F., Lund, P. A., Cole, J. A. & Henderson, I. R. The Essential Genome of Escherichia coli K-12. *mBio* **9** (2018).
84. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics* **17**, 333–351 (2016).
85. Grasso, K. T., Singha Roy, S. J., Osgood, A. O., Yeo, M. J. R., Soni, C., Hillenbrand, C. M., Ficareta, E. D. & Chatterjee, A. A Facile Platform to Engineer Escherichia coli Tyrosyl-tRNA Synthetase Adds New Chemistries to the Eukaryotic Genetic Code, Including a Phosphotyrosine Mimic. *ACS central science* **8**, 483–492 (2022).
86. Gruet, A., Longhi, S. & Bignon, C. One-step generation of error-prone PCR libraries using Gateway technology. *Microbial Cell Factories* **11**, 14 (2012).

87. Gurwitz, D. Gene drives raise dual-use concerns. *Science* **345**, 1010 (2014).
88. Hagiwara, Y., Hasegawa, T., Shoji, A., Kuwahara, M., Ozaki, H. & Sawai, H. Acridone-tagged DNA as a new probe for DNA detection by fluorescence resonance energy transfer and for mismatch DNA recognition. *Bioorganic & medicinal chemistry* **16**, 7013–7020 (2008).
89. Hall, R. S., Fedorov, A. A., Xu, C., Fedorov, E. V., Almo, S. C. & Raushel, F. M. Three-dimensional structure and catalytic mechanism of cytosine deaminase. *Biochemistry* **50**, 5077–5085 (2011).
90. Halperin, S. O., Tou, C. J., Wong, E. B., Modavi, C., Schaffer, D. V. & Dueber, J. E. CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature* **560**, 248–252 (2018).
91. Hasegawa, T., Shoji, A., Kuwahara, M., Ozaki, H. & Sawai, H. Synthesis and property of DNA labeled with fluorescent acridone. *Nucleic Acids Symposium Series*, 145–146 (2006).
92. Herfst, S., Schrauwen, E. J. A., Linster, M., Chutinimitkul, S., de Wit, E., Munster, V. J., Sorrell, E. M., Bestebroer, T. M., Burke, D. F., Smith, D. J., Rimmelzwaan, G. F., Osterhaus, Albert D. M. E. & Fouchier, R. A. M. Airborne transmission of influenza A/H5N1 virus between ferrets. *Science* **336**, 1534–1541 (2012).
93. Hollenstein, M., Hipolito, C. J., Lam, C. H. & Perrin, D. M. A self-cleaving DNA enzyme modified with amines, guanidines and imidazoles operates independently of divalent metal cations (M2+). *Nucleic Acids Research* **37**, 1638–1649 (2009).
94. Hoshika, S., Chen, F., Leal, N. A. & Benner, S. A. Artificial genetic systems: self-avoiding DNA in PCR and multiplexed PCR. *Angewandte Chemie (International ed. in English)* **49**, 5554–5557 (2010).
95. Hossain, A., Lopez, E., Halper, S. M., Cetnar, D. P., Reis, A. C., Strickland, D., Klavins, E. & Salis, H. M. Automated design of thousands of nonrepetitive parts for engineering stable genetic systems. *Nature Biotechnology* **38**, 1466–1475 (2020).
96. iGEM Bielefeld CeBiTec. *Nanofactory* 2018. <http://2018.igem.org/Team:Bielefeld-CeBiTec>.
97. iGEM Foundation. *About iGEM* (ed iGEM Foundation) <https://2020.igem.org/About>.
98. iGEM Foundation. *Safety and security hub* <http://2018.igem.org/Safety>.

99. iGEM Foundation. *iGEM 2021 Thank you!* 2021. https://2021.igem.org/Main_Page.
100. Inglesby, T., Cicero, A., Rivers, C. & Zhang, W. Biosafety and biosecurity in the era of synthetic biology: Meeting the challenges in China and the U.S. *Journal of Biosafety and Biosecurity* **1**, 73–74 (2019).
101. International Gene Synthesis Consortium — The Promotion of Biosecurity. *Home — International Gene Synthesis Consortium* 2017. <https://genesynthesisconsortium.org/>.
102. Isidro-Llobet, A., Kenworthy, M. N., Mukherjee, S., Kopach, M. E., Wegner, K., Gallou, F., Smith, A. G. & Roschangar, F. Sustainability Challenges in Peptide Synthesis and Purification: From R&D to Production. *The Journal of Organic Chemistry* **84**, 4615–4628 (2019).
103. Jäger, S. & Famulok, M. Generation and enzymatic amplification of high-density functionalized DNA double strands. *Angewandte Chemie International Edition* **43**, 3337–3340 (2004).
104. Jäger, S., Rasched, G., Kornreich-Leshem, H., Engeser, M., Thum, O. & Famulok, M. A versatile toolbox for variable DNA functionalization at high density. *Journal of the American Chemical Society* **127**, 15071–15082 (2005).
105. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology* **31**, 233–239 (2013).
106. Jones, W. J., Leigh, J. A., Mayer, F., Woese, C. R. & Wolfe, R. S. *Methanococcus jannaschii* sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent. *Archives of Microbiology* **136**, 254–261 (1983).
107. Käll, L., Storey, J. D. & Noble, W. S. Non-parametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry. *Bioinformatics (Oxford, England)* **24**, i42–8 (2008).
108. Kimoto, M., Kawai, R., Mitsui, T., Yokoyama, S. & Hirao, I. An unnatural base pair system for efficient PCR amplification and functionalization of DNA molecules. *Nucleic Acids Research* **37**, e14 (2009).
109. Kirchberger, P. C. & Ochman, H. Resurrection of a global, metagenomically defined gokushovirus. *eLife* **9** (2020).

110. Kobayashi, T., Nureki, O., Ishitani, R., Yaremcuk, A., Tukalo, M., Cusack, S., Sakamoto, K. & Yokoyama, S. Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nature structural biology* **10**, 425–432 (2003).
111. Koch, N. G., Baumann, T. & Budisa, N. Efficient Unnatural Protein Production by Pyrrolysyl-tRNA Synthetase With Genetically Fused Solubility Tags. *Frontiers in Bioengineering and Biotechnology* **9**, 807438 (2021).
112. Koch, N. G., Goettig, P., Rappaport, J. & Budisa, N. Engineering Pyrrolysyl-tRNA Synthetase for the Incorporation of Non-Canonical Amino Acids with Smaller Side Chains. *International journal of molecular sciences* **22** (2021).
113. Koyack, M. J. & Cheng, R. P. Design and synthesis of beta-peptides with biological activity. *Methods in molecular biology (Clifton, N.J.)* **340**, 95–109 (2006).
114. Krathwohl, D. R. A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice* **41**, 212–218 (2002).
115. Krzycki, J. A. The direct genetic encoding of pyrrolysine. *Current Opinion in Microbiology* **8**, 706–712 (2005).
116. Kubyshkin, V., Acevedo-Rocha, C. G. & Budisa, N. On universal coding events in protein biogenesis. *Biosystems* **164**, 16–25 (2018).
117. Kunjapur, A. M., Napolitano, M. G., Hysolli, E., Noguera, K., Appleton, E. M., Schubert, M. G., Jones, M. A., Iyer, S., Mandell, D. J. & Church, G. M. Synthetic auxotrophy remains stable after continuous evolution and in coculture with mammalian cells. *Science Advances* **7** (2021).
118. Kunjapur, A. M., Stork, D. A., Kuru, E., Vargas-Rodriguez, O., Landon, M., Söll, D. & Church, G. M. Engineering posttranslational proofreading to discriminate nonstandard amino acids. *Proceedings of the National Academy of Sciences* **115**, 619–624 (2018).
119. Kuratani, M., Sakai, H., Takahashi, M., Yanagisawa, T., Kobayashi, T., Murayama, K., Chen, L., Liu, Z.-J., Wang, B.-C., Kuroishi, C., Kuramitsu, S., Terada, T., Bessho, Y., Shirouzu, M., Sekine, S.-i. & Yokoyama, S. Crystal structures of tyrosyl-tRNA synthetases from Archaea. *Journal of Molecular Biology* **355**, 395–408 (2006).
120. Kuriyama, K. & Sze, P. Y. Blood-brain barrier to H3-gamma-aminobutyric acid in normal and amino oxyacetic acid-treated animals. *Neuropharmacology* **10**, 103–108 (1971).

121. Kuwahara, M., Suto, Y., Minezaki, S., Kitagata, R., Nagashima, J.-i. & Sawai, H. Substrate property and incorporation accuracy of various dATP analogs during enzymatic polymerization using thermostable DNA polymerases. *Nucleic acids symposium series* (2004), 31–32 (2006).
122. Kwon, I. & Choi, E. S. Forced Ambiguity of the Leucine Codons for Multiple-Site-Specific Incorporation of a Noncanonical Amino Acid. *PLoS ONE* **11**, e0152826 (2016).
123. Kwon, I., Wang, P. & Tirrell, D. A. Design of a bacterial host for site-specific incorporation of p-bromophenylalanine into recombinant proteins. *Journal of the American Chemical Society* **128**, 11778–11783 (2006).
124. Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H.-R., Haimovich, A. D., Kuznetsov, G., Mercer, J. A., Wang, H. H., Carr, P. A., Mosberg, J. A., Rohland, N., Schultz, P. G., Jacobson, J. M., Rinehart, J., Church, G. M. & Isaacs, F. J. Genomically recoded organisms expand biological functions. *Science* **342**, 357–360 (2013).
125. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).
126. Larson, B. E. Classroom discussion: a method of instruction and a curriculum outcome. *Teaching and Teacher Education* **16**, 661–677 (2000).
127. Lemke, E. A., Summerer, D., Geierstanger, B. H., Brittain, S. M. & Schultz, P. G. Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nature Chemical Biology* **3**, 769–772 (2007).
128. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* **25**, 2078–2079 (2009).
129. Linn, M. C. The Knowledge Integration Perspective on Learning and Instruction. *Cambridge University Press*, 243–264 (2012).
130. Liu, C. C., Cellitti, S. E., Geierstanger, B. H. & Schultz, P. G. Efficient expression of tyrosine-sulfated proteins in *E. coli* using an expanded genetic code. *Nature Protocols* **4**, 1784–1789 (2009).
131. Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annual Review of Biochemistry* **79**, 413–444 (2010).

132. Liu, D. R., Magliery, T. J., Pastrnak, M. & Schultz, P. G. Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins *in vivo*. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 10092–10097 (1997).
133. Liu, R.-J., Long, T., Zhou, M., Zhou, X.-L. & Wang, E.-D. tRNA recognition by a bacterial tRNA Xm32 modification enzyme from the SPOUT methyltransferase superfamily. *Nucleic Acids Research* **43**, 7489–7503 (2015).
134. Loś, J. M., Golec, P., Wegrzyn, G., Wegrzyn, A. & Loś, M. Simple method for plating *Escherichia coli* bacteriophages forming very small plaques or no plaques under standard conditions. *Applied and Environmental Microbiology* **74**, 5113–5120 (2008).
135. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).
136. Ma, N. J. & Isaacs, F. J. Genomic Recoding Broadly Obstructs the Propagation of Horizontally Transferred Genetic Elements. *Cell Systems* **3**, 199–207 (2016).
137. Mackelprang, R., Aurand, E. R., Bovenberg, R. A. L., Brink, K. R., Charo, R. A., Delborne, J. A., Diggans, J., Ellington, A. D., Fortman, J. L. C., Isaacs, F. J., Medford, J. I., Murray, R. M., Noireaux, V., Palmer, M. J., Zoloth, L. & Friedman, D. C. Guiding Ethical Principles in Engineering Biology Research. *ACS Synthetic Biology* **10**, 907–910 (2021).
138. Maffei, E., Shaidullina, A., Burkolter, M., Heyer, Y., Estermann, F., Druelle, V., Sauer, P., Willi, L., Michaelis, S., Hilbi, H., Thaler, D. S. & Harms, A. Systematic exploration of *Escherichia coli* phage-host interactions with the BASEL phage collection. *PLOS Biology* **19**, e3001424 (2021).
139. Maguire, S. & Guan, S. Rolling circle reverse transcription enables high fidelity nanopore sequencing of small RNA. *PLoS ONE* **17**, e0275471 (2022).
140. Mandell, D. J., Lajoie, M. J., Mee, M. T., Takeuchi, R., Kuznetsov, G., Norville, J. E., Gregg, C. J., Stoddard, B. L. & Church, G. M. Biocontainment of genetically modified organisms by synthetic protein design. *Nature* **518**, 55–60 (2015).
141. Martínez-Núñez, M. A. & López, V. E. L. y. Nonribosomal peptides synthetases and their applications in industry. *Sustainable Chemical Processes* **4**, 1–8 (2016).
142. McKay, C. P. What is life—and how do we search for it in other worlds? *PLOS Biology* **2**, E302 (2004).

143. Melançon, C. E. & Schultz, P. G. One plasmid selection system for the rapid evolution of aminoacyl-tRNA synthetases. *Bioorganic & Medicinal Chemistry Letters* **19**, 3845–3847 (2009).
144. Miles, Z. D., McCarty, R. M., Molnar, G. & Bandarian, V. Discovery of epoxyqueuosine (oQ) reductase reveals parallels between halorespiration and tRNA modification. *Proceedings of the National Academy of Sciences* **108**, 7368–7372 (2011).
145. Miller, T. E., Beneyton, T., Schwander, T., Diehl, C., Girault, M., McLean, R., Chotel, T., Claus, P., Cortina, N. S., Baret, J.-C. & Erb, T. J. Light-powered CO₂ fixation in a chloroplast mimic with natural and synthetic parts. *Science* **368**, 649–654 (2020).
146. Millett, P., Binz, T., Evans, S. W., Kuiken, T., Oye, K., Palmer, M. J., van der Vlugt, C., Yambao, K. & Yu, S. Developing a Comprehensive, Adaptive, and International Biosafety and Biosecurity Program for Advanced Biotechnology: The iGEM Experience. *Applied Biosafety* **24**, 64–71 (2019).
147. Millett, P., Isaac, C. R., Rais, I. & Rutten, P. The synthetic-biology challenges for biosecurity: examples from iGEM. *The Nonproliferation Review* **27**, 443–458 (2020).
148. Minehata, M., Sture, J., Shinomiya, N. & Whitby, S. Implementing biosecurity education: approaches, resources and programmes. *Science and Engineering Ethics* **19**, 1473–1486 (2013).
149. Minei, R., Hoshina, R. & Ogura, A. De novo assembly of middle-sized genome using MinION and Illumina sequencers. *BMC Genomics* **19**, 700. <https://link.springer.com/article/10.1186/s12864-018-5067-1> (2018).
150. Mohler, K., Aerni, H.-R., Gassaway, B., Ling, J., Ibba, M. & Rinehart, J. MS-READ: Quantitative measurement of amino acid incorporation. *Biochimica et biophysica acta. General subjects* **1861**, 3081–3088 (2017).
151. Morgado, S. & Vicente, A. C. Global In-Silico Scenario of tRNA Genes and Their Organization in Virus Genomes. *Viruses* **11**, 180 (2019).
152. Murakami, H., Hohsaka, T., Ashizuka, Y., Hashimoto, K. & Sisido, M. Site-directed incorporation of fluorescent nonnatural amino acids into streptavidin for highly sensitive detection of biotin. *Biomacromolecules* **1**, 118–125 (2000).
153. Murakami, H., Ohta, A., Ashigai, H. & Suga, H. A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nature Methods* **3**, 357–359 (2006).

154. Namy, O., Zhou, Y., Gundlapalli, S., Polycarpo, C. R., Denise, A., Rousset, J.-P., Söll, D. & Ambrogelly, A. Adding pyrrolysine to the *Escherichia coli* genetic code. *FEBS Letters* **581**, 5282–5288 (2007).
155. 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., Söll, D. & Church, G. M. Emergent rules for codon choice elucidated by editing rare arginine codons in *Escherichia coli*. *Proceedings of the National Academy of Sciences* **113**, E5588–97 (2016).
156. National Academies of Sciences, Engineering, and Medicine. *Gene Drives on the Horizon: Advancing Science, Navigating Uncertainty, and Aligning Research with Public Values* (2016).
157. National Academies of Sciences, Engineering, and Medicine. Dual use research of concern in the life sciences: Current issues and controversies (2017).
158. National Academies of Sciences, Engineering, and Medicine. *Governance of Dual Use Research in the Life Sciences: Advancing Global Consensus on Research Oversight: Proceedings of a Workshop* (ed The National Academies Press) Washington, DC, 2018. <https://doi.org/10.17226/25154>.
159. National Research Council. *Biotechnology Research in an Age of Terrorism* (ed The National Academies Press) Washington, DC, 2004. <https://doi.org/10.17226/10827>.
160. National Research Council. Challenges and opportunities for education about dual use issues in the life sciences (2011).
161. National Science Foundation. *Chapter II: Proposal Preparation Instructions* 2023. <https://beta.nsf.gov/policies/pappg/23-1/ch-2-proposal-preparation#2E6>.
162. Neugebauer, M. E., Hsu, A., Arbab, M., Krasnow, N. A., McElroy, A. N., Pandey, S., Doman, J. L., Huang, T. P., Raguram, A., Banskota, S., Newby, G. A., Tolar, J., Osborn, M. J. & Liu, D. R. Evolution of an adenine base editor into a small, efficient cytosine base editor with low off-target activity. *Nature Biotechnology*, 1–13 (2022).
163. Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nature Chemical Biology* **4**, 232–234 (2008).

164. New England Biolabs. *NEB 5-alpha Competent E. coli* 2023. <https://www.neb.com/products/c2987-neb-5-alpha-competent-e-coli-high-efficiency#Product%20Information>.
165. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. & Schultz, P. G. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244**, 182–188 (1989).
166. Novossiolova T. *Biological Security Education Handbook: The Power of Team-Based Learning* (University of Bradford, Bradford Disarmament Research, 2016).
167. Nyerges, Á., Csörgő, B., Nagy, I., Bálint, B., Bihari, P., Lázár, V., Apjok, G., Umenhoffer, K., Bogos, B., Pósfai, G. & Pál, C. A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *Proceedings of the National Academy of Sciences* **113**, 2502–2507 (2016).
168. Nyerges, a., Vinke, S., Flynn, R., Owen, S. V., Rand, E. A., Budnik, B., Keen, E., Narasimhan, K., Marchand, J. A., Baas-Thomas, M., Liu, M., Chen, K., Chiappino-Pepe, A., Hu, F., Baym, M. & Church, G. M. A swapped genetic code prevents viral infections and gene transfer. *Nature*, 1–8 (2023).
169. Organisation for the Prohibition of Chemical Weapons. *Open-Ended Working Group on Future Priorities of the OPCW recommendations to the Fourth Special Session of the Conference of the State Parties to Review the Operation of the Chemical Weapons Conventions* (ed Organisation for the Prohibition of Chemical Weapons) 2018. <https://www.opcw.org/sites/default/files/documents/2018/07/rc4wp01%28e%29.pdf>.
170. Ostrov, N., Landon, M., Guell, M., Kuznetsov, G., Teramoto, J., Cervantes, N., Zhou, M., Singh, K., Napolitano, M. G., Moosburner, M., Shrock, E., Pruitt, B. W., Conway, N., Goodman, D. B., Gardner, C. L., Tyree, G., Gonzales, A., Wanner, B. L., Norville, J. E., Lajoie, M. J. & Church, G. M. Design, synthesis, and testing toward a 57-codon genome. *Science* **353**, 819–822 (2016).
171. Ostrov, N., Nyerges, a., Chiappino-Pepe, A., Rudolph, A., Baas-Thomas, M. & Church, G. M. Synthetic genomes with altered genetic codes. *Current Opinion in Systems Biology* **24**, 32–40 (2020).
172. PacBio. *How HiFi sequencing works - PacBio* 2023. <https://www.pacb.com/technology/hifi-sequencing/how-it-works/>.

173. Peters, F. B., Brock, A., Wang, J. & Schultz, P. G. Photocleavage of the polypeptide backbone by 2-nitrophenylalanine. *Chemistry & Biology* **16**, 148–152 (2009).
174. Peters, S. L., Borges, A. L., Giannone, R. J., Morowitz, M. J., Banfield, J. F. & Hettich, R. L. Experimental validation that human microbiome phages use alternative genetic coding. *Nature communications* **13**, 5710 (2022).
175. Pinheiro. Polymerase engineering: from PCR and sequencing to synthetic biology. **vol III** (2012).
176. Pinheiro, V. B. & Holliger, P. The XNA world: progress towards replication and evolution of synthetic genetic polymers. *Current opinion in chemical biology* **16**, 245–252 (2012).
177. Preiner, M., Asche, S., Becker, S., Betts, H. C., Boniface, A., Camprubi, E., Chandru, K., Erastova, V., Garg, S. G., Khawaja, N., Kostyrka, G., Machné, R., Moggioli, G., Muchowska, K. B., Neukirchen, S., Peter, B., Pichlhöfer, E., Radványi, Á., Rossetto, D., Salditt, A., Schmelling, N. M., Sousa, F. L., Tria, F. D. K., Vörös, D. & Xavier, J. C. The Future of Origin of Life Research: Bridging Decades-Old Divisions. *Life* **10** (2020).
178. Prior, T. I., FitzGerald, D. J. & Pastan, I. Barnase toxin: a new chimeric toxin composed of pseudomonas exotoxin A and barnase. *Cell* **64**, 1017–1023 (1991).
179. Profaizer, T., Coonrod, E. M., Delgado, J. C. & Kumánovics, A. Report on the effects of fragment size, indexing, and read length on HLA sequencing on the Illumina MiSeq. *Human Immunology* **76**, 897–902 (2015).
180. Rakonjac, J., Das, B. & Derda, R. Editorial: Filamentous Bacteriophage in Bio/Nano/Technology, Bacterial Pathogenesis and Ecology. *Frontiers in microbiology* **7**, 2109 (2016).
181. Reeves, R. G., Voeneky, S., Caetano-Anollés, D., Beck, F. & Boëte, C. Agricultural research, or a new bioweapon system? *Science* **362**, 35–37 (2018).
182. Rennekamp, A. J. Synthetic Organisms Simplify Biology. *Cell* **178**, 1–3 (2019).
183. Revill J. *Education and Ethics in the life sciences: strengthening the prohibition of biological weapons* (ANU E Press, 2010).
184. Rijksinstituut voor Volksgezondheid en Milieu. *Biosecurity Office — Bureau Biosecurity* 2023. <https://www.bureaubiosecurity.nl/en>.
185. Roberts, B. Export controls and biological weapons: new roles, new challenges. *Critical Reviews in Microbiology* **24**, 235–254 (1998).
186. Robertson, S. A., Ellman, J. A. & Schultz, P. G. A general and efficient route for chemical aminoacylation of transfer RNAs. *Journal of the American Chemical Society* **113**, 2722–2729 (1991).

187. 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., Buse, S., Maslen, S., Salmond, G. P. C. & Chin, J. W. Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science* **372**, 1057–1062 (2021).
188. Rogerson, D. T., Sachdeva, A., Wang, K., Haq, T., Kazlauskaitė, A., Hancock, S. M., Huguenin-Dezot, N., Muqit, M. M. K., Fry, A. M., Bayliss, R. & Chin, J. W. Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. *Nature Chemical Biology* **11**, 496–503 (2015).
189. Rother, M. & Krzycki, J. A. Selenocysteine, pyrrolysine, and the unique energy metabolism of methanogenic archaea. *Archaea* **2010** (2010).
190. 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. & Isaacs, F. J. Recoded organisms engineered to depend on synthetic amino acids. *Nature* **518**, 89–93 (2015).
191. Ruan, B., Paloura, S., Sabina, J., Marvin-Guy, L., Kochhar, S., Larossa, R. A. & Söll, D. Quality control despite mistranslation caused by an ambiguous genetic code. *Proceedings of the National Academy of Sciences* **105**, 16502–16507 (2008).
192. Santamaría-Gómez, J., Rubio, M. Á., López-Igual, R., Romero-Losada, A. B., Delgado-Chaves, F. M., Bru-Martínez, R., Romero-Campero, F. J., Herrero, A., Ibba, M., Ochoa de Alda, J. A. G. & Luque, I. Role of a cryptic tRNA gene operon in survival under translational stress. *Nucleic Acids Research* **49**, 8757–8776 (2021).
193. Santoro, S. W., Wang, L., Herberich, B., King, D. S. & Schultz, P. G. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nature Biotechnology* **20**, 1044–1048 (2002).
194. Sarwar, S., Ilyas, S., Khan, B. A., Lohman, D. C., Hafeez, S., Rafique, M., Rashid, F., Akhtar, J., Khan, S. & Amoah, A. O. Awareness and Attitudes of Research Students Toward Dual-use Research of Concern in Pakistan: A Cross-sectional Questionnaire. *Health security* **17**, 229–239 (2019).
195. Sattar, S., Bennett, N. J., Wen, W. X., Guthrie, J. M., Blackwell, L. F., Conway, J. F. & Rakonjac, J. Ff-nano, short functionalized nanorods derived from Ff (f1, fd, or M13) filamentous bacteriophage. *Frontiers in microbiology* **6**, 316 (2015).
196. Schmidt, M. Xenobiology: a new form of life as the ultimate biosafety tool. *BioEssays* **32**, 322–331 (2010).

197. Schmidt, M. & Budisa, N. *Art as We Don't Know It - Alternative Biofacts – Life as we don't (yet) know it* (Aalto University, 2020).
198. Schmidt, M. & Kubyshkin, V. How To Quantify a Genetic Firewall? A Polarity-Based Metric for Genetic Code Engineering. *ChemBioChem* **22**, 1268–1284 (2021).
199. Schneider, M. & Stern, E. The cognitive perspective on learning, 69–90 (2010).
200. Schwark, D. G., Schmitt, M. A. & Fisk, J. D. Dissecting the Contribution of Release Factor Interactions to Amber Stop Codon Reassignment Efficiencies of the Methanocaldococcus jannaschii Orthogonal Pair. *Genes* **9** (2018).
201. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics (Oxford, England)* **30**, 2068–2069 (2014).
202. Selgelid, M. J. Gain-of-Function Research: Ethical Analysis. *Science and Engineering Ethics* **22**, 923–964 (2016).
203. Seo, Y. J., Malyshev, D. A., Lavergne, T., Ordoukhalian, P. & Romesberg, F. E. Site-specific labeling of DNA and RNA using an efficiently replicated and transcribed class of unnatural base pairs. *Journal of the American Chemical Society* **133**, 19878–19888 (2011).
204. Seo, Y. J., Matsuda, S. & Romesberg, F. E. Transcription of an expanded genetic alphabet. *Journal of the American Chemical Society* **131**, 5046–5047 (2009).
205. Serwer, P., Hayes, S. J., Thomas, J. A. & Hardies, S. C. Propagating the missing bacteriophages: a large bacteriophage in a new class. *Virology Journal* **4**, 21 (2007).
206. Skvorc, C. & Wilson, D. E. Developing a behavioral health screening program for BSL-4 laboratory workers at the National Institutes of Health. *Biosecurity and bioterrorism : biodefense strategy, practice, and science* **9**, 23–29 (2011).
207. Srinivasan, G., James, C. M. & Krzycki, J. A. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* **296**, 1459–1462 (2002).
208. Steer, B. A. & Schimmel, P. Domain-domain communication in a miniature archaeabacterial tRNA synthetase. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 13644–13649 (1999).
209. Styers, M. L., van Zandt, P. A. & Hayden, K. L. Active Learning in Flipped Life Science Courses Promotes Development of Critical Thinking Skills. *CBE—Life Sciences Education* **17**, ar39 (2018).
210. Switzer, C., Moroney, S. E. & Benner, S. A. Enzymatic incorporation of a new base pair into DNA and RNA. *Journal of the American Chemical Society* **111**, 8322–8323 (1989).

211. Szili, P., Draskovits, G., Révész, T., Bogár, F., Balogh, D., Martinek, T., Daruka, L., Spohn, R., Vásárhelyi, B. M., Czikkely, M., Kintses, B., Grézal, G., Ferenc, G., Pál, C. & Nyerges, Á. Rapid Evolution of Reduced Susceptibility against a Balanced Dual-Targeting Antibiotic through Stepping-Stone Mutations. *Antimicrobial Agents and Chemotherapy* **63** (2019).
212. Takai, K., Okumura, S., Hosono, K., Yokoyama, S. & Takaku, H. A single uridine modification at the wobble position of an artificial tRNA enhances wobbling in an Escherichia coli cell-free translation system. *FEBS Letters* **447**, 1–4 (1999).
213. Taki, M., Hohsaka, T., Murakami, H., Taira, K. & Sisido, M. Position-specific incorporation of a fluorophore-quencher pair into a single streptavidin through orthogonal four-base codon/anticodon pairs. *Journal of the American Chemical Society* **124**, 14586–14590 (2002).
214. Tan, G., Opitz, L., Schlapbach, R. & Rehrauer, H. Long fragments achieve lower base quality in Illumina paired-end sequencing. *Scientific Reports* **9**, 2856 (2019).
215. Tharp, J. M., Walker, J. A., Söll, D. & Schepartz, A. Initiating protein synthesis with non-canonical monomers in vitro and in vivo. *Methods in enzymology* **656**, 495–519 (2021).
216. The Nucleic Acid Observatory Consortium. *A Global Nucleic Acid Observatory for Biodefense and Planetary Health* 2021. <http://arxiv.org/pdf/2108.02678v1.pdf>.
217. The White House. FACT SHEET: President Biden to Launch a National Biotechnology and Biomanufacturing Initiative. *The White House*. <https://www.whitehouse.gov/briefing-room/statements-releases/2022/09/12/fact-sheet-president-biden-to-launch-a-national-biotechnology-and-biomanufacturing-initiative/> (2022).
218. Tomita, K., Ogawa, T., Uozumi, T., Watanabe, K. & Masaki, H. A cytotoxic ribonuclease which specifically cleaves four isoaccepting arginine tRNAs at their anticodon loops. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 8278–8283 (2000).
219. Tong, H., Wang, X., Liu, Y., Liu, N., Li, Y., Luo, J., Ma, Q., Wu, D., Li, J., Xu, C. & Yang, H. Programmable A-to-Y base editing by fusing an adenine base editor with an N-methylpurine DNA glycosylase. *Nature Biotechnology*, 1–5 (2023).

220. Umenhoffer, K., Draskovits, G., Nyerges, Á., Karcagi, I., Bogos, B., Tímár, E., Csörgő, B., Herczeg, R., Nagy, I., Fehér, T., Pál, C. & Pósfai, G. Genome-Wide Abolishment of Mobile Genetic Elements Using Genome Shuffling and CRISPR/Cas-Assisted MAGE Allows the Efficient Stabilization of a Bacterial Chassis. *ACS Synthetic Biology* **6**, 1471–1483 (2017).
221. United Nations Office for Disarmament Affairs. *UN Security Council Resolution 1540 (2004)* (ed United Nations Office for Disarmament Affairs) 2004. <https://www.un.org/disarmament/wmd/sc1540/>.
222. University of Bath, B. *Next generation biosecurity: responding to 21st century biorisks* (ed FutureLearn) 2022. <https://www.futurelearn.com/courses/biosecurity>.
223. UNODA. *Biological Weapons Convention* 2023. <https://www.un.org/disarmament/biological-weapons/>.
224. Vaught, J. D., Bock, C., Carter, J., Fitzwater, T., Otis, M., Schneider, D., Rolando, J., Waugh, S., Wilcox, S. K. & Eaton, B. E. Expanding the chemistry of DNA for in vitro selection. *Journal of the American Chemical Society* **132**, 4141–4151 (2010).
225. Vázquez, M. E., Nitz, M., Stehn, J., Yaffe, M. B. & Imperiali, B. Fluorescent caged phosphoserine peptides as probes to investigate phosphorylation-dependent protein associations. *Journal of the American Chemical Society* **125**, 10150–10151 (2003).
226. Venetz, J. E., Del Medico, L., Wölflé, A., Schächle, P., Bucher, Y., Appert, D., Tschan, F., Flores-Tinoco, C. E., van Kooten, M., Guennoun, R., Deutsch, S., Christen, M. & Christen, B. Chemical synthesis rewriting of a bacterial genome to achieve design flexibility and biological functionality. *Proceedings of the National Academy of Sciences* **116**, 8070–8079 (2019).
227. Vinke, S., Klages, L. J., Rückert, C., Schlüter, L., Lehmann, L. S., Nölting, S. & Kalinowski, J. Nanopore sequencing guided directed evolution of aminoacyl/tRNA-synthetases. *Unpublished* (2023).
228. Vinke, S., Rais, I. & Millett, P. The Dual-Use Education Gap: Awareness and Education of Life Science Researchers on Nonpathogen-Related Dual-Use Research. *Health security* **20**, 35–42 (2022).
229. Wang, J., Xie, J. & Schultz, P. G. A genetically encoded fluorescent amino acid. *Journal of the American Chemical Society* **128**, 8738–8739 (2006).

230. Wang, J., Yashiro, Y., Sakaguchi, Y., Suzuki, T. & Tomita, K. Mechanistic insights into tRNA cleavage by a contact-dependent growth inhibitor protein and translation factors. *Nucleic Acids Research* **50**, 4713–4731 (2022).
231. Wang, L. & Schultz, P. G. A general approach for the generation of orthogonal tRNAs. *Chemistry & Biology* **8**, 883–890 (2001).
232. Wang, L. & Schultz, P. G. Expanding the genetic code. *Angewandte Chemie International Edition* **44**, 34–66 (2004).
233. Wang, Y., Zhao, Y., Bollas, A., Wang, Y. & Au, K. F. Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology* **39**, 1348–1365 (2021).
234. Wannier, T. M., Kunjapur, A. M., Rice, D. P., McDonald, M. J., Desai, M. M. & Church, G. M. Adaptive evolution of genomically recoded *Escherichia coli*. *Proceedings of the National Academy of Sciences* **115**, 3090–3095 (2018).
235. Weisbrod, S. H. & Marx, A. Novel strategies for the site-specific covalent labelling of nucleic acids. *Chemical Communications*, 5675–5685 (2008).
236. Whitby, S., Novossiolova, T., Walther, G. & Dando, M. *Preventing biological threats: what you can do. a guide to biological security issues and how to address them* (University of Bradford, Bradford Disarmament Research Centre, 2015).
237. Willis, J. C. W. & Chin, J. W. Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. *Nature Chemistry* **10**, 831–837 (2018).
238. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nature Methods* **6**, 359–362 (2009).
239. Wooster, H. Xenobiology. *Science (New York, N.Y.)* **134**, 223–225 (1961).
240. World Health Organization. *Biorisk management : laboratory biosecurity guidance 2006*. <https://apps.who.int/iris/handle/10665/69390>.
241. World Health Organization. *Responsible Life Sciences Research for Global Health Security: A Guidance Document* Geneva, 2012. <https://apps.who.int/iris/handle/10665/70507>.
242. World Health Organization. *Fact sheet: biosafety and biosecurity 2018*. https://www.who.int/influenza/pip/BiosecurityandBiosafety_EN_20Mar2018.pdf?ua=1%5b06/11/20.
243. Wu, N., Deiters, A., Cropp, T. A., King, D. & Schultz, P. G. A genetically encoded photocaged amino acid. *Journal of the American Chemical Society* **126**, 14306–14307 (2004).

244. Yan, S., Li, X., Zhang, P., Wang, Y., Chen, H.-Y., Huang, S. & Yu, H. Direct sequencing of 2'-deoxy-2'-fluoroarabinonucleic acid (FANA) using nanopore-induced phase-shift sequencing (NIPSS). *Chemical Science* **10**, 3110–3117 (2019).
245. Yang, J. Y., Fang, W., Miranda-Sanchez, F., Brown, J. M., Kauffman, K. M., Acevero, C. M., Bartel, D. P., Polz, M. F. & Kelly, L. Degradation of host translational machinery drives tRNA acquisition in viruses. *Cell Systems* **12**, 771–779.e5 (2021).
246. Yang, Z., Hutter, D., Sheng, P., Sismour, A. M. & Benner, S. A. Artificially expanded genetic information system: a new base pair with an alternative hydrogen bonding pattern. *Nucleic Acids Research* **34**, 6095–6101 (2006).
247. Ye, S., Huber, T., Vogel, R. & Sakmar, T. P. FTIR analysis of GPCR activation using azido probes. *Nature Chemical Biology* **5**, 397–399 (2009).
248. Zhang, C.-M., Liu, C., Slater, S. & Hou, Y.-M. Aminoacylation of tRNA with phosphoserine for synthesis of cysteinyl-tRNA(Cys). *Nature Structural & Molecular Biology* **15**, 507–514 (2008).
249. Zhang, Y., Wang, L., Schultz, P. G. & Wilson, I. A. Crystal structures of apo wild-type *M. jannaschii* tyrosyl-tRNA synthetase (TyrRS) and an engineered TyrRS specific for O-methyl-L-tyrosine. *Protein science : a publication of the Protein Society* **14**, 1340–1349 (2005).
250. Zhang, Z., Smith, B. A. C., Wang, L., Brock, A., Cho, C. & Schultz, P. G. A new strategy for the site-specific modification of proteins in vivo. *Biochemistry* **42**, 6735–6746 (2003).
251. Zhang, Z., Wang, L., Brock, A. & Schultz, P. G. The selective incorporation of alkenes into proteins in *Escherichia coli*. *Angewandte Chemie International Edition* **41**, 2840–2842 (2002).
252. Zou, X., Xiao, X., Mo, Z., Ge, Y., Jiang, X., Huang, R., Li, M., Deng, Z., Chen, S., Wang, L. & Lee, S. Y. Systematic strategies for developing phage resistant *Escherichia coli* strains. *Nature communications* **13**, 4491 (2022).
253. Zürcher, J. F., Robertson, W. E., Kappes, T., Petris, G., Elliott, T. S., Salmond, G. P. C. & Chin, J. W. Refactored genetic codes enable bidirectional genetic isolation. *Science* **378**, 516–523 (2022).

Appendices

Nanopore sequencing-guided directed evolution of aminoacyl-tRNA synthetases for the incorporation of non-standard amino acids into proteins

	unchanged	with internal stop	too different (>7 changes)	positive	above threshold (>=3 hits) of the positives
Evo1Round1	3.37%	56.50%	7.40%	32.74%	6.75%
Evo1Round2	6.04%	11.29%	15.93%	66.74%	5.67%
Evo1Round3	5.69%	13.08%	15.15%	66.08%	5.11%
Evo1Round4	5.07%	13.04%	15.98%	65.91%	4.84%
Evo1Round5	0.47%	93.06%	1.67%	4.80%	0.79%
Evo1Round6	5.54%	12.25%	15.60%	66.61%	4.50%
Evo1Round7	2.95%	12.69%	15.94%	68.42%	7.97%
Evo2Round1	2.89%	61.11%	6.92%	29.09%	5.10%
Evo2Round2	5.87%	12.32%	15.15%	66.66%	5.37%
Evo2Round3	5.08%	15.74%	15.51%	63.66%	4.20%
Evo2Round4	4.93%	12.50%	16.75%	65.83%	4.48%
Evo2Round5	3.73%	37.48%	11.23%	47.56%	4.65%
Evo2Round6	2.78%	53.80%	7.56%	35.86%	4.30%
Evo2Round7	0.86%	71.66%	5.60%	21.89%	5.48%
Evo3Round1	6.40%	15.29%	14.24%	64.08%	6.06%
Evo3Round2	5.73%	12.63%	16.04%	65.61%	5.34%
Evo3Round3	5.70%	12.60%	15.92%	65.77%	4.76%
Evo3Round4	4.78%	12.70%	15.82%	66.70%	3.54%
Evo3Round5	0.72%	87.76%	2.75%	8.77%	1.41%
Evo3Round6	1.05%	36.73%	15.74%	46.48%	8.03%
Evo3Round7	0.93%	12.54%	24.10%	62.42%	9.21%

Figure 18: Percentage shares of different classes of library members of all positive selection rounds of the three evolutions. Unchanged variants are wildtype sequence. With internal stop are variants with an inframe stop codon. Too different are variants with more than 7 amino acid exchanges. Positive variants are variants with less than 7 amino acid exchanges. Above threshold are the variants that show at least 3 reads of a positive variant and thus can be assumed to be enriched.

amplicon_trimming_phages.py:

```
# importing the package sys
import sys
# starting the main method
def main():
    sequence=[]
    cut_fwd=[]
    fastqseq=sys.argv[1]
```

```

outp=sys.argv[2]

# defining a start and end sequence for the region of interest
fwd="ATTGGCTTT"
rev="GAAGAGATC"

seq_rev={}
number=0

# opening the fastq data to get all sequences which contain the start
and end sequence
with open (fastqseq, "r") as seq:
for line in seq:
if fwd in line and rev in line:
sequence.append(line)

# the sequences which contain the start and end sequence, were cutted
on that positions so that the sequence starts with fwd and ends
with rev.

for line in sequence:
cut = line.split(fwd, 1)[1]
roi = cut.split(rev, 1)[0]
#print(roi)
if 500<len(roi)< 700:
cut_fwd.append(roi)

# the cutted sequences were safed in a fasta file. The sequence
header is described by a the word sequence and a number

with open(outp, 'w') as data:
g=0
for i in cut_fwd:
data.write(">Sequence_"+str(g)+"\n")
data.write("%s\n" % i)
g=g+1

if __name__ == '__main__':
main()

```

Indel_correction_phage.py:

```

import sys
def main():
    # defining the reference DNA sequence
    ref="gagcttccgtaaaatccattggacattaccttcagatcaaaaaatgatcgatcttcagaa
        tgccggcttcgacattatcatccattgggtgatcttgcgcatacttaatcagaaggcgaattagat
        gaaatccgtaaagtggagattataacaaaaggatttcgaggcaatgggtctgaaagccaaatacg
        ttatggatcagagttaagtctggacaaggattactcttaacgttatcgtagccctaaaaccacttt
        aaaacgtgctcgccgtccatggagttaaattgcgcgcaggacgagaacccgaaagttgcagaagtgatc
        tatccgatcatgcaggtgaacacctatcactatgacggagttagacgttagctgtgggggcatggaacagc
        gcaagattcacatgctggctcgtaactgctgcctaaaaagtagttagtattcataatcctgtactgac
        gggatttagtgagaaaggcaagatgagtcctcgaaaggaaactttatcgccgtagacgactcacct"#
        atggacgagtttagatgattaagcgcaacacaagtgaaattattccgaagaagagttacg
        cgaggtatccatcaaaaggcgaattagatgaaatccgtaaagtggagattataacaaaaggattcg
        aggcaatgggtctgaaagccaaatacgttatggatcagagttaagtctggacaaggattactcttaa
        cgttatcgtagccctaaaaccactttaaaacgtgctcgccgtccatggagttaaattgcgcgcag
        gacgagaacccgaaagttgcagaagtgatctatccgatcatgcaggtgaacacctatcactatgacggag
        tagacgttagctgtgggggcatggaacagcgcaagattcacatgctggctcgtaactgctgcctaaaa
        agtagtatgtattcataatcctgtactgacgggattagatggagaaggcaagatgagtcctcgaaagg
        aactttatcgccgtagacgactcacctgaagagatccgcgcgaagatcaaaaggcatactgcccggcgg
        gagttgtgaaggcaaccctattatggagattgccaagtatfffftggaaatccccctgaccatcaagcg
        cccggaaaaattcggcggagacttgaccgtgaactcgtagggagttagaatcctgttcaaaaataaa
        gaactgcattatggatctgaaaatgcagttgctgaagagttgattaaaatttggaaaccaatccgca
        aacgcctg"
    ref=ref.upper()
    fasta=[]
    seq=[]
    cigar=[]
    sequ=sys.argv[1]
    cig=sys.argv[2]
    outpu=sys.argv[3]

    # opening the fasta file with the the cutted sequences and safed the
    # sequences into
    # a list
    with open(sequ,"r") as sequence:
        for line in sequence:
            if line.startswith(">"):
                line= line.strip("\n")
                fasta.append(line)
            else:
                line=line.strip("\n")

```

```

seq.append(line)

# opening a text document which contains the cigar sting from the
sequence mapping and splitted the cigare string on the different
values

with open(cig,"r") as code:
for line in code:
if line.startswith("@"):
continue
else:
splitted=[]
i=line.split("\t")
x= i[5]
for i in x:
if i=="M" or i=="I" or i=="D" or i=="S" or i=="H":
d=x.split(i,1)
splitted.append(int(d[0]))
splitted.append(i)
x=x.lstrip(d[0])
x=x.lstrip(i)
list1=[]
for d in splitted:
list1.append(d)

cigar.append(list1)

fin=[]
a=0
z=0
bxyz=len(seq)
#print(b)

# Splitted the sequence on the different values of the cigar
string so that insertions and deletions on the DNA sequence
can be corrected.

for line in cigar:
finseq=[]
counter=0
if not z<bxyz:
break

```

```
r=ref
p=seq[z]
for i in line:
    if i=="M":
        a=counter-1
        b=line[a]
        c=p[:b]
        f=r[:b]
        finseq.append(c)
        p=p.replace(c,"",1)
        r=r.replace(f,"",1)

    elif i=="I":
        a=counter-1
        b=line[a]
        c=p[:b]
        p=p.replace(c,"",1)

    elif i=="D":
        a=counter-1
        b=line[a]
        f=r[:b]
        finseq.append(f)
        r=r.replace(f,"",1)

    elif i=="S":
        a=counter-1
        b=line[a]
        c=p[:b]
        f=r[:b]
        finseq.append(f)
        p=p.replace(c,"",1)
        r=r.replace(f,"",1)

    elif i=="H":
        a=counter-1
        b=line[a]
        c=p[:b]
        f=r[:b]
        finseq.append(f)
        p=p.replace(c,"",1)
        r=r.replace(f,"",1)
```

```

counter=counter+1
a="" . join(finseq)
fin.append(a)
z=z+1

# writes the results into an new fasta file.

with open(outpu,'w') as data:
g=0
for i in fin:
data.write(">Sequence_"+str(g)+"\n")
data.write("%s\n" % i)
g=g+1

if __name__ == '__main__':
main()

```

script_phages.pl:

```

use strict;

use String::Approx 'amatch';

use Getopt::Std;

getopts('v');

our ($opt_v);

# search parameters
my $threshold = 3;
my $nc = 7;

# search sequences
my $fwd ="ATTGGCTTT";
my $rev_fwd="GAAGAGATC";

my $rev="GCAGTTCACG";
my $fwd_rev="AAGATCGATC";

my %seqs;

```

```
# translation table
my %cu = ( GCA => 'A',
            GCC => 'A',
            GCG => 'A',
            GCT => 'A',
            TGC => 'C',
            TGT => 'C',
            GAC => 'D',
            GAT => 'D',
            GAA => 'E',
            GAG => 'E',
            TTC => 'F',
            TTT => 'F',
            GGA => 'G',
            GGC => 'G',
            GGG => 'G',
            GGT => 'G',
            CAC => 'H',
            CAT => 'H',
            ATA => 'I',
            ATC => 'I',
            ATT => 'I',
            AAA => 'K',
            AAG => 'K',
            CTA => 'L',
            CTC => 'L',
            CTG => 'L',
            CTT => 'L',
            TTA => 'L',
            TTG => 'L',
            ATG => 'M',
            AAC => 'N',
            AAT => 'N',
            CCA => 'P',
            CCC => 'P',
            CCG => 'P',
            CCT => 'P',
            CAA => 'Q',
            CAG => 'Q',
            AGA => 'R',
            AGG => 'R',
            CGA => 'R',
            CGC => 'R',
```

```

CGG => 'R',
CGT => 'R',
AGC => 'S',
AGT => 'S',
TCA => 'S',
TCC => 'S',
TCG => 'S',
TCT => 'S',
ACA => 'T',
ACC => 'T',
ACG => 'T',
ACT => 'T',
GTA => 'V',
GTC => 'V',
GTG => 'V',
GTT => 'V',
TGG => 'W',
TAC => 'Y',
TAT => 'Y',
TAA => 'Z',
TGA => 'Z',
TAG => '*',
);

# define the reference sequence
my $rfn ="ATTGGCTTGAGCCTCCGGTAAAATCCATTGGGACATTACCTCAGATCAAAAAAAT
GATCGATCTCAGAATGCCGGCTTCGACATTATCATCCATTGGGTGATCTGGCGCATATCTTAATCA
GAAGGGCGAATTAGATGAAATCCGTAAGATTGGAGATTATAACAAAAGGTATCGAGGCAATGGGTCT
GAAAGCCAATACGTTATGGATCAGAGTTAAGTCTGGACAAGGATTCTACTCTAACGTTATCGTTT
AGCCCTAAACCACTTAAACGTGCTCGCGCTCCATGGAGTTAATTGCGCGAGGACGAGAACCC
GAAAGTGCAGAAGTGATCTATCCGATCATGCAGGTGAACACCTATCACTATGACGGAGTAGACGTAGC
TGTGGGGGCATGGAACAGCGCAAGATTACATGCTGGCTCGTGAAGTGCCTAAAAAAGTAGTATG
TATTCTATAATCCTGTACTGACGGATTAGGTGGAGAAGGCAAGATGAGCTCCTCGAAAGGAAACTTAT
CGCGTAGACGACTCACCTGAAGAGATC";

my (@rfncodons) = ($rfn =~ m/(\w{3})/ig);

# translation of the reference
my $raa;
foreach my $rfncodon (@rfncodons) {
    $raa .= $cu{$rfncodon};
}

```

```

open FILE, $ARGV[0];

my $wt;
my $stop;
my $diff;
my $pos;
my $total;

while (<FILE>) {
    chomp;
    unless (m/^>(\S+)/) {
        $total++;

        my $seq = $_;
        if ($seq =~ m/^$fwd(.+)\$rev_fwd$/) {
            $seq = $1;
        }

        # added the search sequences to the sequence
        $seq = "ATTGGCTTT" . $seq . "GAAGAGATC";
        my $len = length($seq);

        my (@codons) = ($seq =~ m/(\w{3})/ig);

        # translation of the sequences
        my $aaseq;
        foreach my $codon (@codons) {
            $aaseq .= $cu{$codon};
        }

        # comparison of the reference with the sequences,
        # where the sequences without AA changes
        # will be sorted out
        if ($aaseq eq $raa) {
            print STDERR "Input matches reference\n"
            if ($opt_v);
            $wt++;
            next;
        }
    }
}

```

```

# this will sort sequences out which have internal
stops
if ($aaseq =~ m/Z/) {
    print STDERR "Input contains stop\n" if
    ($opt_v);
    $stop++;
    next;
}

my @matches = amatch($raa, [ $nc ], $aaseq);

# this will sort sequences out which are too different
from the reference
unless ($#matches == 0) {
    print STDERR "Input too different
from reference\n"
    if ($opt_v);
    $diff++;
    next;
}

#this will sort sequences out which
have not a similar
length as the reference
$seqs{$seq}++ if ($len > 500 && $len < 700);
$pos++;

}

}

# here we count the identical sequences and sort them for the
output

my %bycounts;
my @seqs;

my $aboveth;

foreach my $seq (sort (keys(%seqs))) {
    push @{$bycounts{$seqs{$seq}}}, $seq;
}

```

```

my @list = (sort by_number(keys(%bycounts))) ;

foreach my $num (@list) {
    foreach my $seq (@{$bycounts{$num}}) {
        if ($seqs{$seq} >= $threshold) {
            push(@seqs, $seq);
            $aboveth += $num;
        }
    }
}

my %counts;

my %filtered;

foreach my $num (@list) {

    $counts{$num} = $#{$bycounts{$num}} + 1;

    next if ($num < $threshold);

    my $subnum = 1;

    foreach my $seq (@{$bycounts{$num}}) {
        print ">bc$ARGV[1] \_$num\_$subnum\n$seq\n";

        $subnum++;
    }
}

# here we generate an output for statistical analysis

my $frac = sprintf("%2.2f", ($wt / $total * 100));
print STDERR "Unchanged: $frac % ($wt of $total)\n";

my $frac = sprintf("%2.2f", ($stop / $total * 100));
print STDERR "With internal stop: $frac % ($stop of $total)\n";

```

```

my $frac = sprintf("%2.2f", ($diff / $total * 100));
print STDERR "Too different (> $nc changes): $frac %
($diff of $total)\n";

my $frac = sprintf("%2.2f", ($pos / $total * 100));
print STDERR "Positive: $frac % ($pos of $total)\n";

my $frac = sprintf("%2.2f", ($aboveth / $total * 100));
my $frac2 = sprintf("%2.2f", ($aboveth / $pos * 100));
print STDERR "Above threshold (>= $threshold): $frac % / $frac2 %
($aboveth
of $total / $pos)\n";

sub by_number {
    $b <=> $a;
}

```

calc_pos.pl:

```

use strict;

use Getopt::Std;

getopts('s');

our($opt_s);

# translation table
my %cu = ( GCA => 'A',
GCC => 'A',
GCG => 'A',
GCT => 'A',
TGC => 'C',
TGT => 'C',
GAC => 'D',
GAT => 'D',
GAA => 'E',
GAG => 'E',
TTC => 'F',
TTT => 'F',
GGA => 'G',

```

GGC => 'G',
GGG => 'G',
GGT => 'G',
CAC => 'H',
CAT => 'H',
ATA => 'I',
ATC => 'I',
ATT => 'I',
AAA => 'K',
AAG => 'K',
CTA => 'L',
CTC => 'L',
CTG => 'L',
CTT => 'L',
TTA => 'L',
TTG => 'L',
ATG => 'M',
AAC => 'N',
AAT => 'N',
CCA => 'P',
CCC => 'P',
CCG => 'P',
CCT => 'P',
CAA => 'Q',
CAG => 'Q',
AGA => 'R',
AGG => 'R',
CGA => 'R',
CGC => 'R',
CGG => 'R',
CGT => 'R',
AGC => 'S',
AGT => 'S',
TCA => 'S',
TCC => 'S',
TCG => 'S',
TCT => 'S',
ACA => 'T',
ACC => 'T',
ACG => 'T',
ACT => 'T',
GTA => 'V',
GTC => 'V',

```

GTG => 'V',
GTT => 'V',
TGG => 'W',
TAC => 'Y',
TAT => 'Y',
TAA => 'Z',
TGA => 'Z',
TAG => '*',
'---' => '-',
);

# reference sequence
my $rfn ="ATTGGCTTGAGCCTCCGGTAAAATCCATTGGACATTACCTCAGATCAAAAAAA
TGATCGATCTTCAGAATGCCGGCTTCGACATTATCATCCATTGGGTGATCTGGCGCATATCTTAAT
CAGAAGGGCGAATTAGATGAAATCGTAAGATTGGAGATTATAACAAAAAGGTATTCGAGGCAATGGG
TCTGAAAGCCAATACGTTATGGATCAGAGTTAAGTCTGGACAAGGATTCTACTCTTAACGTTATC
GTTTAGCCCTTAAAACCACTTAAAACGTGCTCGCCGCTCCATGGAGTTAATTGCGCGCGAGGACGAG
AACCCGAAAGTTGCAGAAGTGATCTATCGATCATGCAGGTGAACACCTATCACTATGACGGAGTAGA
CGTAGCTGTGGGGGCATGGAACAGCGCAAGATTCACATGCTGGCTCGTAECTGCTGCCTAAAAAG
TAGTATGTATTCTATAATCCTGTACTGACGGGATTAGGTGGAGAAGGCAAGATGAGCTCCTCGAAAGGA
AACTTATGCCGTAGACGACTCACCTGAAGAGATC";

```

```

open FILE, $ARGV[0];

my %vars;
my $id;
my $count;

my $total;

my %subs;
my $subtotal;

my $totalseqs;

while (<FILE>) {
    chomp;
    if (m/^>(\S+_(\S+)_.+)/) {
        $id      = $1;
        $count   = $2;
        $total += $count;

```

```

        $totalseqs++;

    } else {
        my $seq = $_;

        my $name = sprintf("%21.21d", (int(rand(1000000000000000000000000000)) + 1));

        open OUT, ">/tmp/toalign$name.fasta";

        print OUT ">Ref\n$rfn\n>Seq\n$seq\n";
        close OUT;

        # does a alignment of the sequence against the
        # reference system("/usr/bin/edialign -sequences /tmp/
        # toalign$name.
        # fasta -outfile
        # /tmp/edialign$name -outseq /tmp/aligned$name.fasta 2>
        # /dev/null >/dev
        # /null");
    }

    my %seqs;

    open IN, "/tmp/aligned$name.fasta";

    my $idn;
    while (<IN>) {
        chomp;
        if (m/^>(\S+)/) {
            $idn = $1;
        } else {
            $seqs{$idn} .= uc($_);
        }
    }
    unlink "/tmp/toalign$name.fasta";
    unlink "/tmp/aligned$name.fasta";
    unlink "/tmp/edialign$name";

    $seq = $seqs{'Seq'};
    my $rfnnew = $seqs{'Ref'};

```

```

next if ((length($seq) % 3) != 0);
next if ((length($rfnnew) % 3) != 0);

# define triplets of th sequences of the aligned
sequences the reference and the aligned
reference
my (@codons) = ($seq =~ m/(\S{3})/ig);
my (@rfncodons) = ($rfn =~ m/(\S{3})/ig);
my (@rfnnewcodons) = ($rfnnew =~ m/(\S{3})/ig);

# analyzing the Indels after the alignment to correct
the aligned      triplets if there are three
"-"splitted into two triplets
for (my $i = 0, my $k = 0; $i <= $#codons; $i++, $k++) {
    if ($codons[$i] =~ m/\w{2}-/) {
        $codons[$i] = substr($codons[$i], 0, 2)
        . substr
        ($codons
         [$i+1], 2, 1);
        $codons[$i+1] = '---';
    } elsif ($codons[$i] =~ m/\w-{2}/) {
        $codons[$i+1] = substr($codons[$i], 0, 1)
        . substr
        ($codons
         [$i+1], 1, 2);
        $codons[$i] = '---';
    }
    if ($rfnnewcodons[$i] =~ m/\w{2}-/) {
        $rfnnewcodons[$i] = substr
        ($rfnnewcodons[$i], 0, 2)
        . substr($rfnnewcodons[$i+1], 2, 1);
        $rfnnewcodons[$i+1] = '---';
    } elsif
    ($rfnnewcodons[$i] =~ m/\w-{2}/) {
        $rfnnewcodons[$i+1] = substr
        ($rfnnewcodons[$i], 0, 1)
        . substr($rfnnewcodons[$i+1], 1, 2);
        $rfnnewcodons[$i] = '---';
    }
}

```

```

next if ($codons[$i] eq $rfnnewcodons[$i]);

my $aa = $cu{$codons[$i]};
my $raa = $cu{$rfnnewcodons[$i]};

# with option -s sequences with the same AA
# as the reference can be included
# into the
unless ($opt_s) {
    next if ($aa eq $raa);
}

my (@nucs) = ($codons[$i] =~ m/(\S)/ig);
my (@rfnnucs) = ($rfnnewcodons[$i] =~
m/(\S)/ig);

for (my $j = 0; $j < 3; $j++) {
    next if ($nucs[$j] eq $rfnnucs[$j]);
    next if ($nucs[$j] eq '-' || $rfnnucs[$j] eq '-');
    $subtotal++;
    $subs{$rfnnucs[$j]}{$nucs[$j]}++;
}
$vars{$raa . ($k + 1) . $aa . "\t" .
$rfnnewcodons[$i] . "\t" . $codons[$i]} += $count;

$k-- if ($raa eq '-');

# notification if there are insertions or
# deletions in the amino acids sequence
print STDERR "Insertion: $id\n" if ($raa eq '-');
print STDERR "Deletion: $id\n" if ($aa eq '-');
}

my @counts;

foreach my $var (keys(%vars)) {
    push (@{$counts{$vars{$var}}}, $var);
}

```

```
}

# analizing single nucleotids which does not contain to any tripllett
for (my $count = $#counts; $count > 0; $count--) {
    if (defined($counts[$count])) {
        my $frac = sprintf("%2.2f", ($count / $total * 100));
        foreach my $var (@{$counts[$count]}) {
            print "$var\t$count\t$frac\n";
        }
    }
}

# analyzing which nucleotid changes occure and their frequency
print STDERR "\n\n";
foreach my $ref (sort(keys(%subs))) {
    foreach my $rep (sort(keys(%{$subs{$ref}}))) {
        my $frac = sprintf("%2.2f", ($subs{$ref}{$rep} /
$subtotal
* 100));
        print STDERR "$ref\t$rep\t$subs{$ref}{$rep}\t$frac
\n";
    }
}
# notification of the number of total sequences
print STDERR "\nTotal sequences: $total\n";
```

Swapped genetic code blocks viral infections and gene transfer

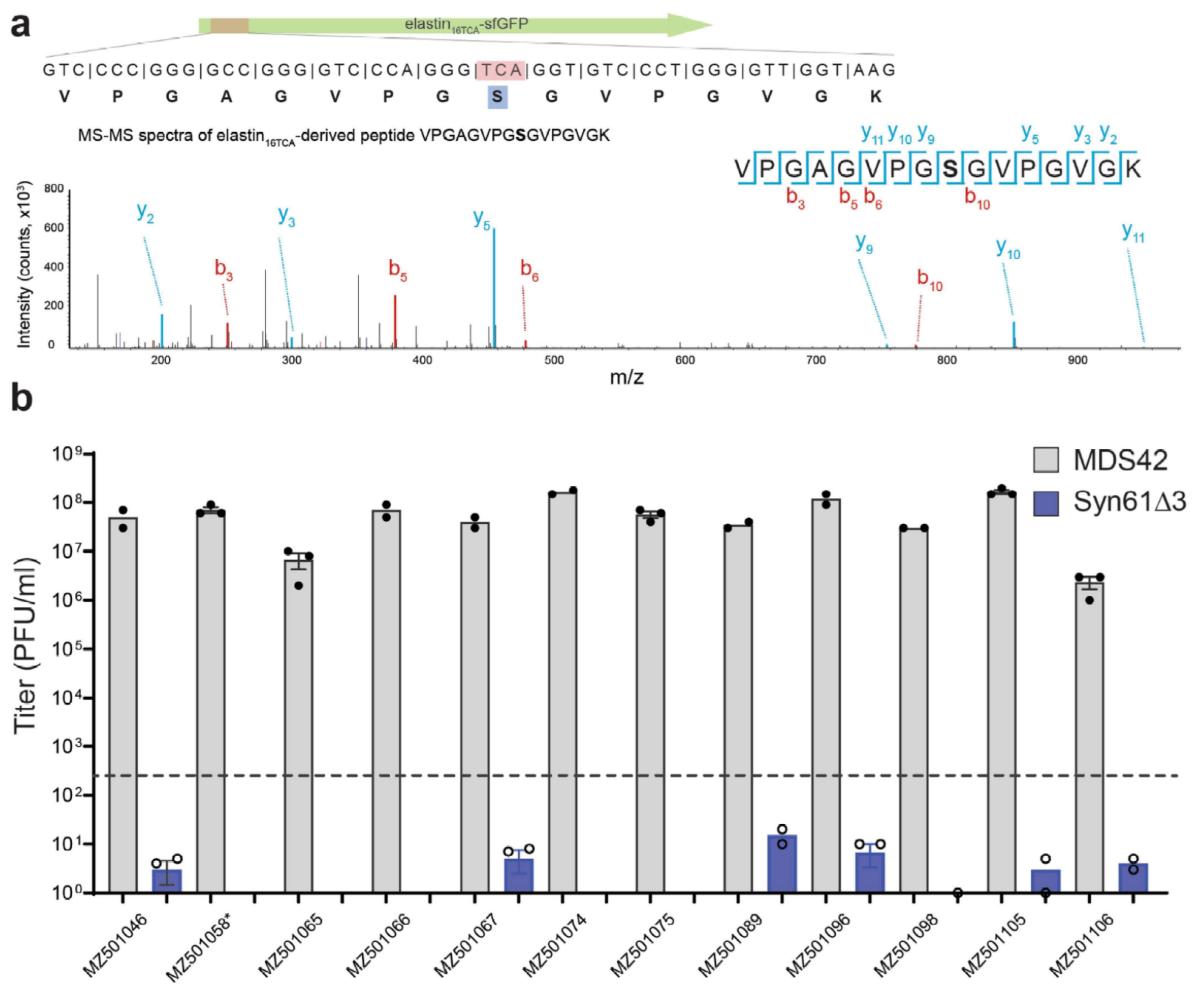


Figure 19: a) Viral TCR suppressor tRNAs decode TCA codons as serine. The amino acid identity of the translated TCR codon within elastin16 TCA-sfGFP-His6 was confirmed by tandem mass spectrometry from Syn61Δ3 expressing the tRNA-SerUGA of *Escherichia* phage IrisVonRoten (GenBank ID MZ501075)24. The figure shows the amino acid sequence and MS/MS spectrum of the analyzed elastin16 TCA peptide. MS/MS data was collected once. b) Syn61Δ3 obstructs the replication of viruses containing genomic tRNA-SerUGA. Figure shows the titer of twelve tRNA gene-containing coliphages24, after 24 h of growth on MDS42 and Syn61Δ3. All analyzed bacteriophages, except MZ501058, contain a genomic tRNA-SerUGA tRNA that provides TCR suppressor activity based on our screen (Fig. 11b, Supplementary Data 11). Early exponential phase cultures of MDS42 and Syn61Δ3 were infected at an MOI of 0.001 with the corresponding phages, and free phage titers were determined after 24 h of incubation. Measurements were performed in $n = 3$ independent experiments (i.e., MDS42 + MZ501058, MZ501065, MZ501075, MZ501105, MZ501106; Syn61Δ3 + MZ501046, MZ501067, MZ501066, MZ501096, MZ501074, MZ501098) or in $n = 2$ independent experiments (i.e., MDS42 + MZ501046, MZ501066, MZ501067, MZ501074, MZ501089, MZ501096, MZ501098; Syn61Δ3 + MZ501058, MZ501065, MZ501075, MZ501089, MZ501105, MZ501106); dashed line represents input phage titer without bacterial cells (i.e., a titer of 420 PFU/ml); dots represent data from independent experiments; bar graphs represent the mean; error bars represent the SEM based on $n = 3$ independent experiments.

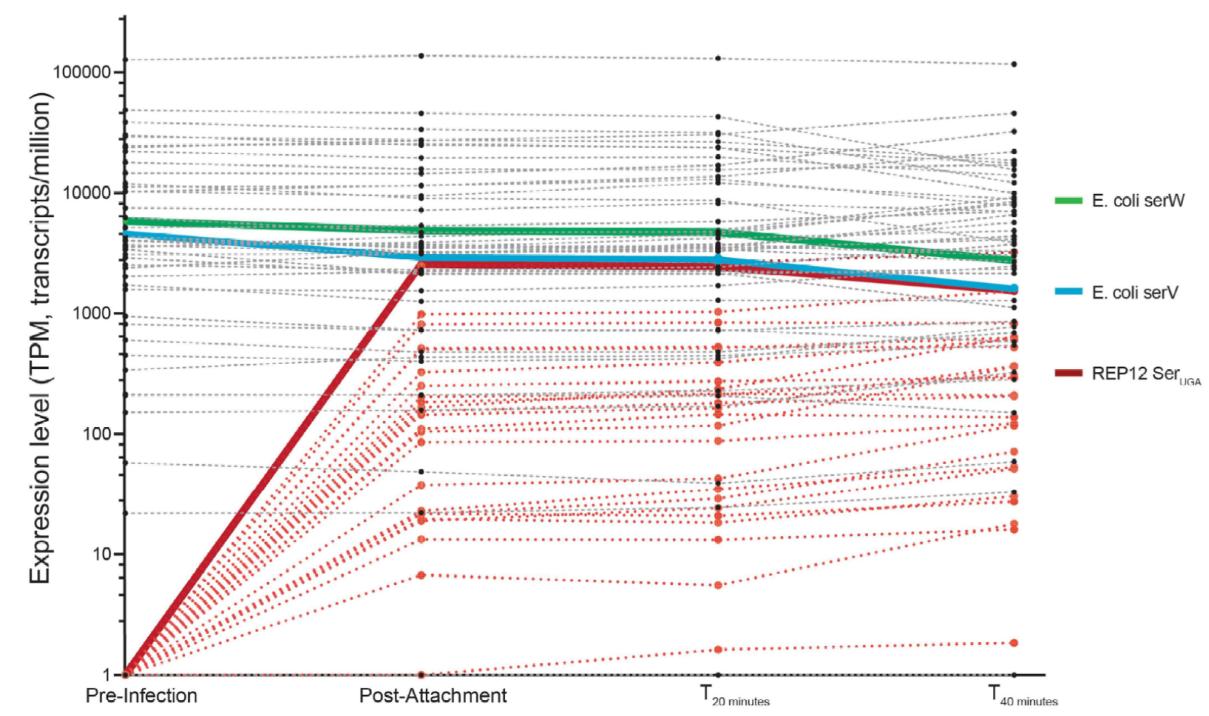


Figure 20: The time-course kinetics of host and viral tRNA expression in *Syn61Δ3* cells following REP12 phage infection was quantified using tRNAseq (Methods). The endogenous serV and serW tRNAs of the host *Syn61Δ3* are highlighted in green and blue, respectively, while the tRNA-SerUGA of the REP12 virus is highlighted in red. REP12 viral tRNAs are shown in light red; endogenous tRNAs of *Syn61Δ3* are shown in gray. Data represent mean TPM (transcript/million). Source data is available within this paper.

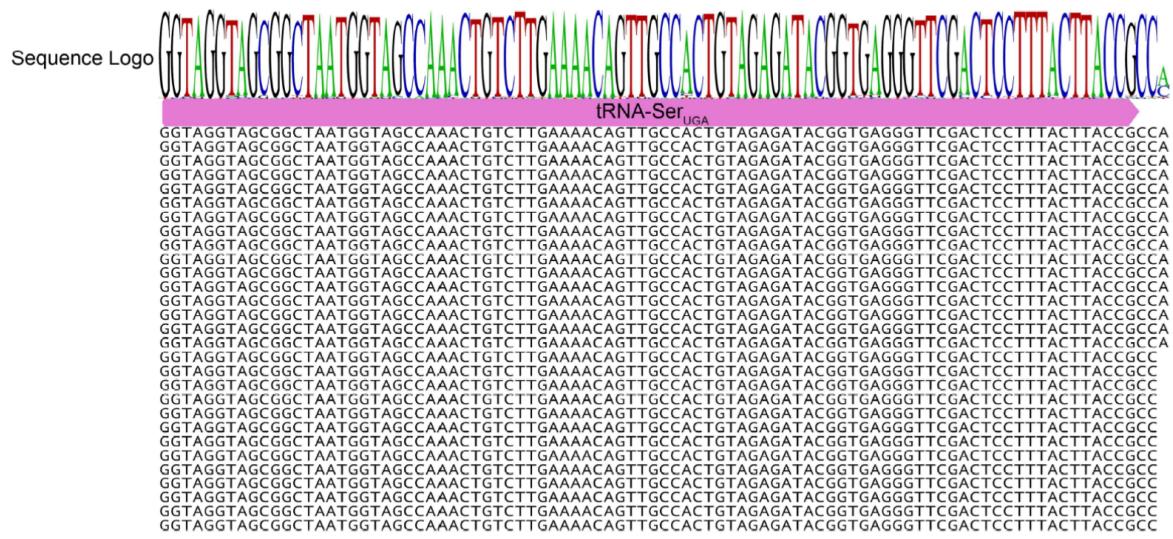


Figure 21: The tRNA tail is modified into CCA even if the phage-encoded tRNA-SerUGA does not encode a CCA end. The sequence of genomic phage-encoded tRNA-SerUGA is highlighted in magenta. Black letters indicate example tRNAseq sequencing reads from REP12 infected Syn61Δ3 cells directly after phage attachment based on a single experiment.

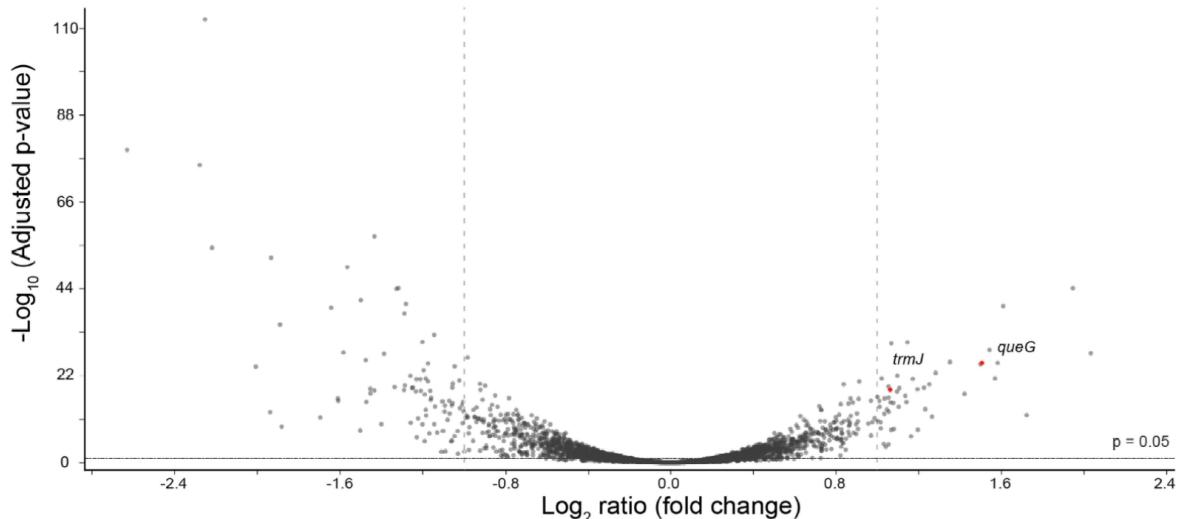


Figure 22: Volcano plot shows the differential expression between uninfected and REP12-infected Syn61Δ3 cells, 40 min post-infection, based on n = 3 independent experiments (Methods). QueG encodes epoxyqueuosine reductase that catalyzes the final step in the de novo synthesis of queuosine in tRNAs. TrmJ encodes tRNA Cm32/Um32 methyltransferase that introduces methyl groups at the 2'-O position of U32 of several tRNAs, including tRNA-SerUGA. Differential expression and -Log₁₀ adjusted p-values were calculated using the DESeq2 algorithm⁵⁵. Source data is available within this paper.

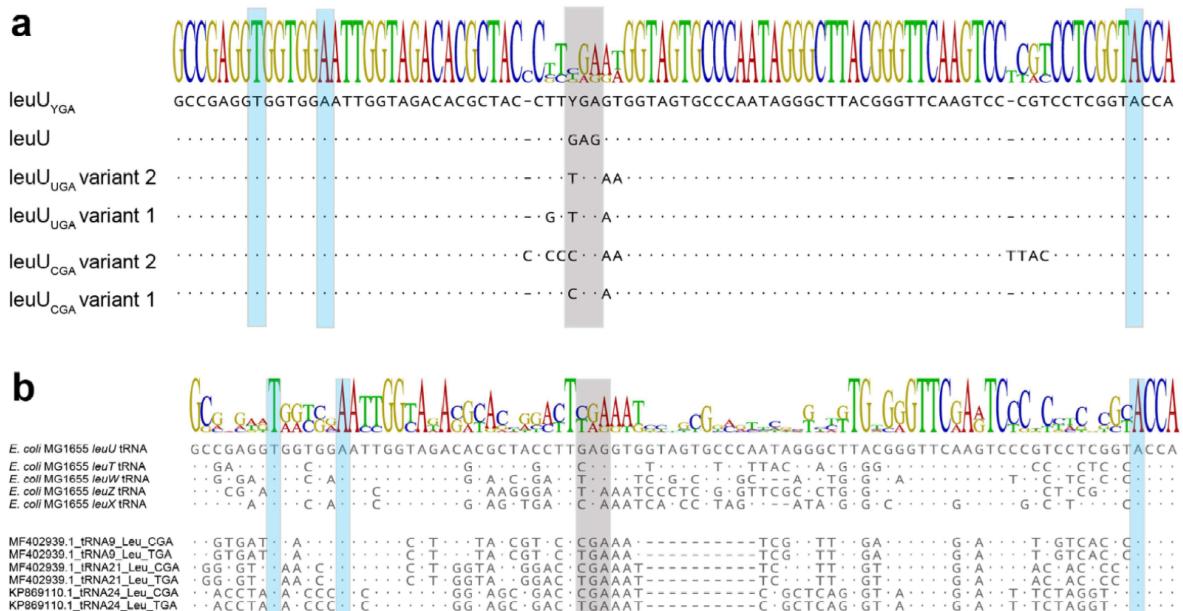


Figure 23: a) Multiple sequence alignment of leuUYGA variants selected in aminoglycoside O-phosphotransferase expression screen, compared to *E. coli* leuU and the YGA anticodon-swapped *E. coli* leuU tRNA variant. Grey shading indicates the anticodon region, and the host's LeuS leucine-tRNA-ligase identity elements³¹ are shown in blue. Sequence information of the leuUYGA variants is available in Supplementary Data 3. b) Multiple sequence alignment of phage-derived tRNA-LeuYGA variants selected in the aph3Ia29xLeu→TCR aminoglycoside O-phosphotransferase expression screen, compared to endogenous *E. coli* leucine tRNAs. Grey shading indicates the anticodon region, while the host's LeuS leucine-tRNA-ligase identity elements³¹ are shown in blue. Sequence information of the phage-derived tRNA-LeuYGA variants is available in Supplementary Data 3.

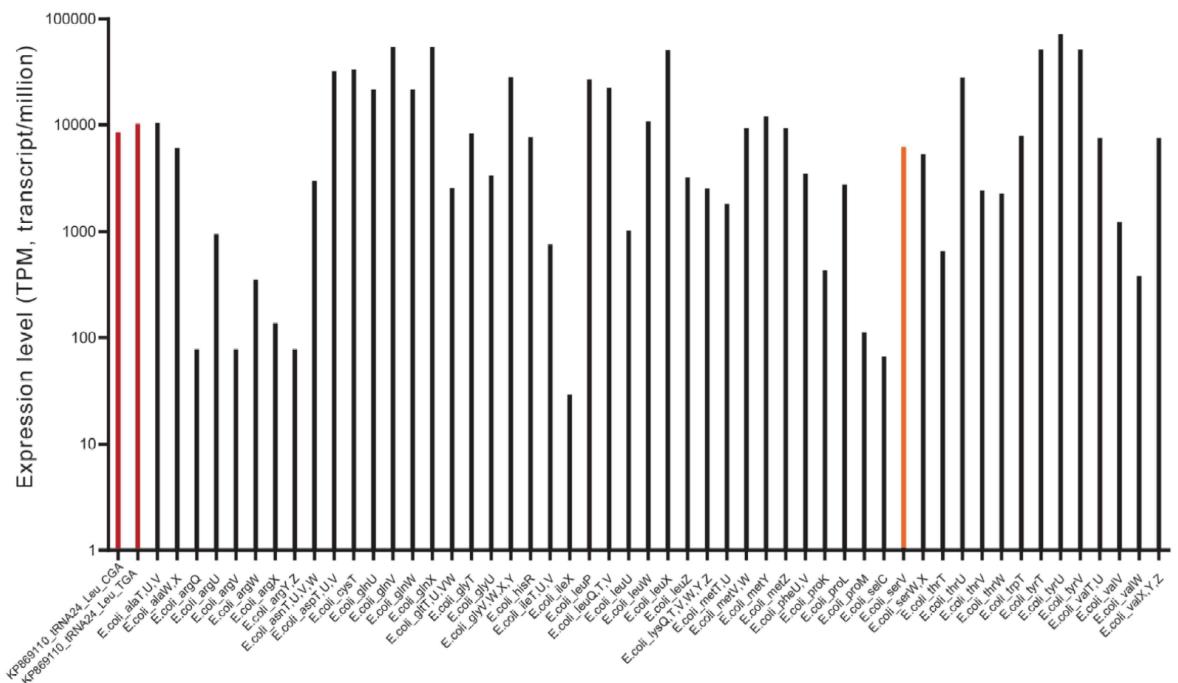
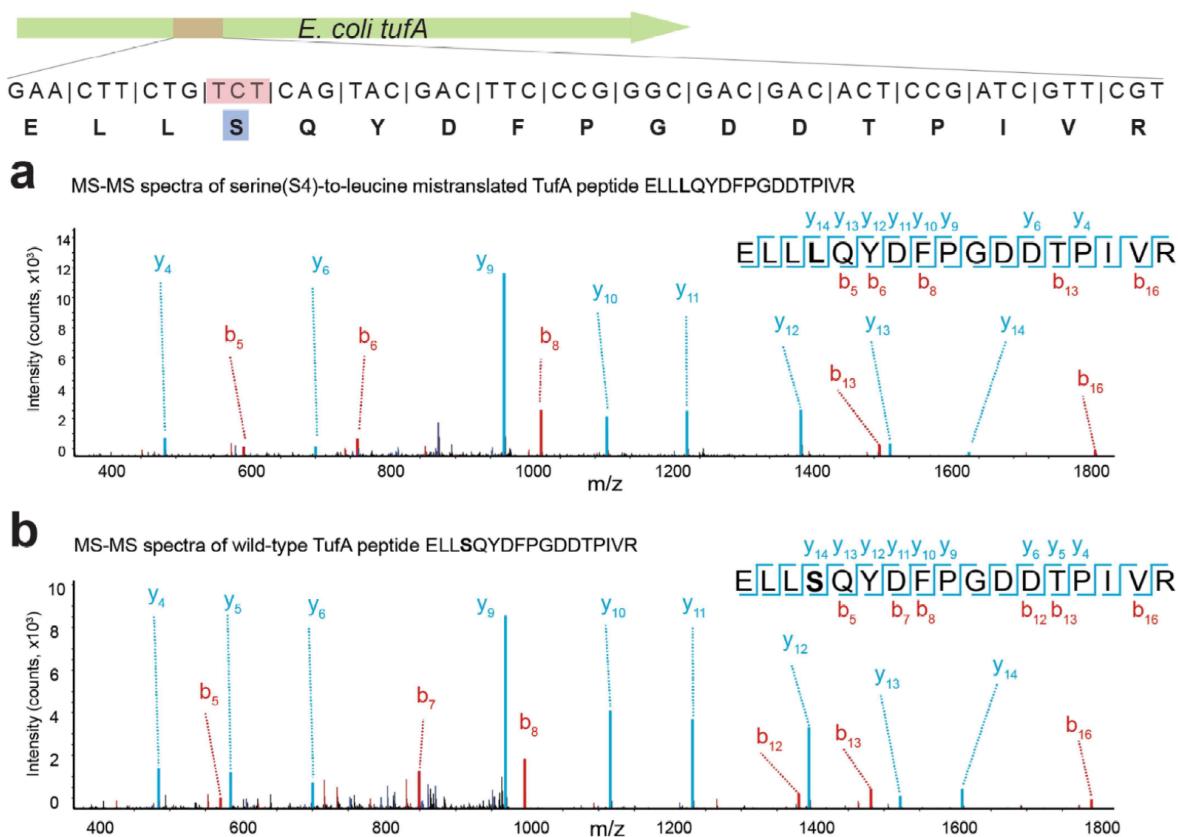


Figure 24: tRNA levels of Ec_Syn61Δ3-SL were quantified using tRNAsq (Methods). Viral Leu-tRNAUGA and Leu-tRNACGA are highlighted in red, and the host's endogenous *E. coli* serV tRNA is highlighted in orange. tRNAsq data was collected once. Data represent TPM (transcript/million). Source data is available within this paper.



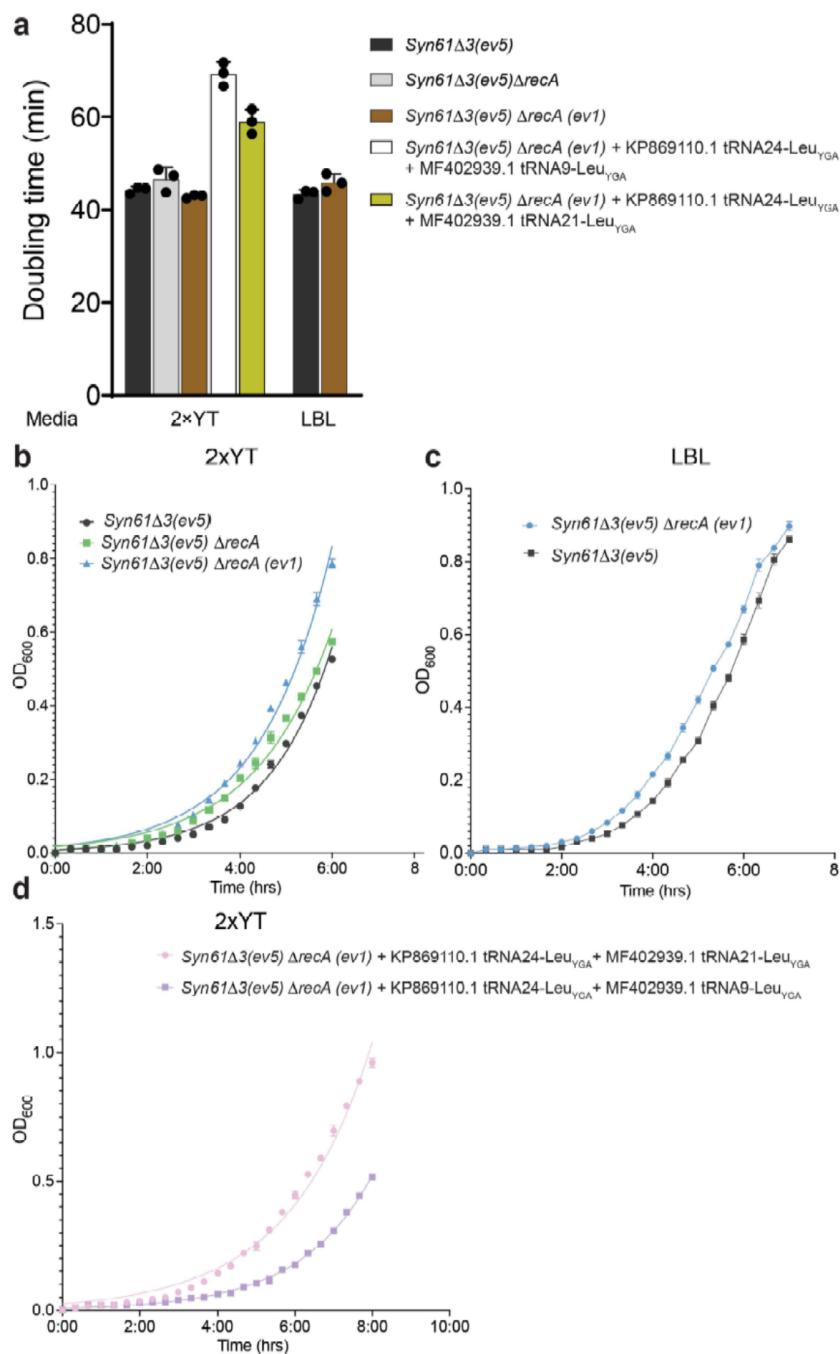


Figure 26: a) Doubling times of *Syn61 Δ 3(ev5)*, *Syn61 Δ 3(ev5) Δ recA*, *Syn61 Δ 3(ev5) Δ recA(ev1)*, and *Ec-Syn61 Δ 3-SL*, calculated based on growth curves (shown in panels b, c, d) in rich bacterial media under standard laboratory conditions. b) Growth curves of *Syn61 Δ 3(ev5)*, *Syn61 Δ 3(ev5) Δ recA*, and *Syn61 Δ 3(ev5) Δ recA(ev1)* in LBL broth. c) Growth curves of *Syn61 Δ 3(ev5)* and *Syn61 Δ 3(ev5) Δ recA(ev1)* in 2xYT broth. d) Growth curves of *Ec-Syn61 Δ 3-SL* in 2xYT broth containing 50 μ g/ml kanamycin. Three independent cultures were grown aerobically in vented shake flasks at 37 °C, and OD₆₀₀ measurements were taken during exponential growth (Methods). Data curves and bars represent the mean. Error bars show standard deviation based on n = 3 independent experiments.

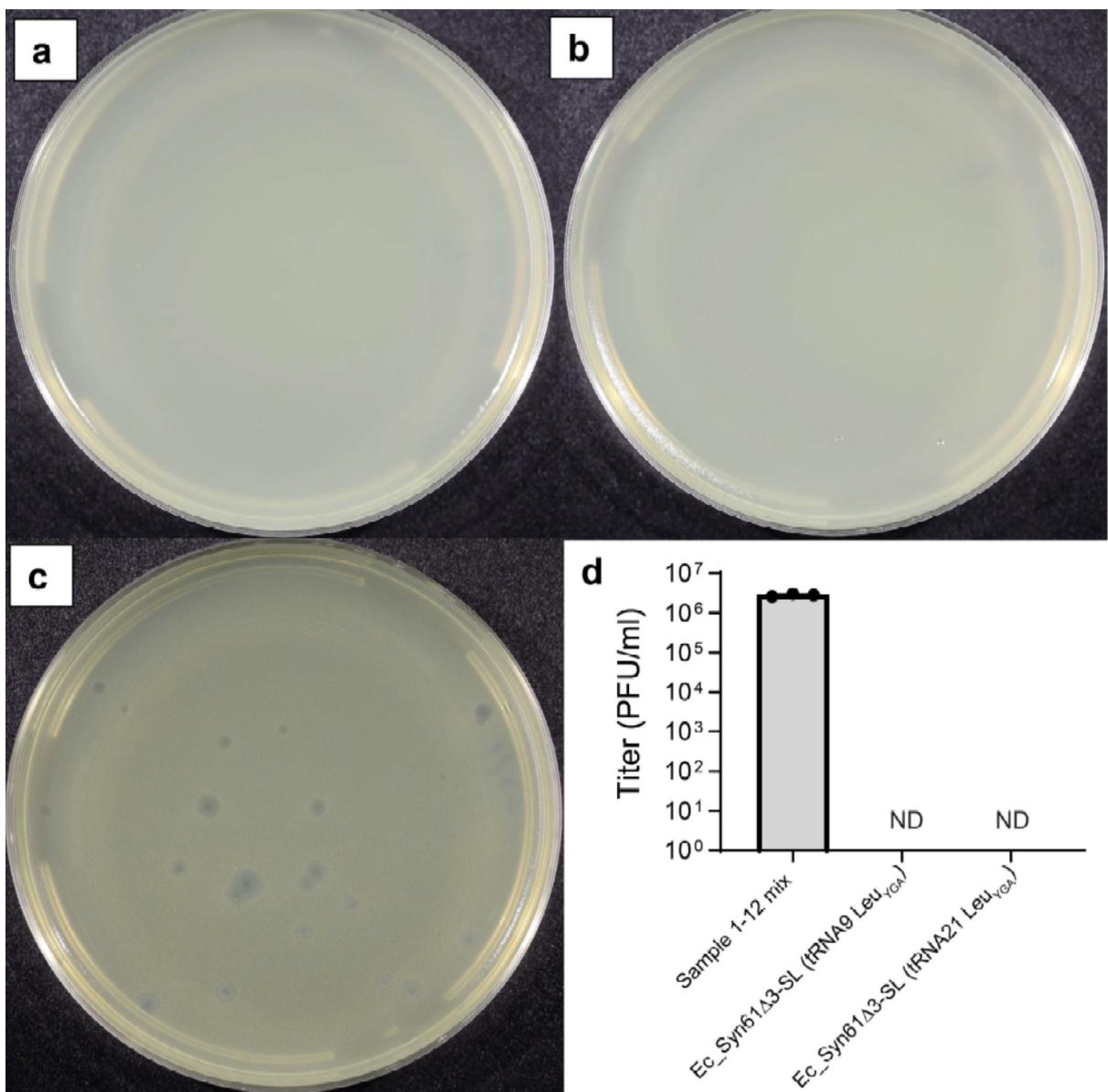


Figure 27: (a) Phage enrichment experiment using Ec_Syn61Δ3-SL, expressing KP869110.1 tRNA24 LeuYGA and MF402939.1 tRNA9 LeuYGA, as host. (b) Phage enrichment experiment using Ec_Syn61Δ3-SL, expressing KP869110.1 tRNA24 LeuYGA and MF402939.1 tRNA21 LeuYGA, as host. Phage enrichment experiments were performed by mixing early exponential cultures of Ec_Syn61Δ3-SL with 10 ml environmental sample mix containing the mixture of Sample 2–13 from our study (Extended Data Table 1a). After two enrichment cycles (Methods, Supplementary Note), filter-sterilized culture supernatants were mixed with phage-susceptible *E. coli* MDS42 cells in top agar and plated on LBL agar plates to determine viral titer. Enrichment experiments were performed in $n = 2$ independent replicates with the same result. (c) Lytic *E. coli* MDS42 phage plaques after 103-fold dilution of the environmental sample mix. (d) Lytic phage titer of the environmental sample mix, before and after enrichment on Ec_Syn61Δ3-SL. Dots represent the viral titer of the unenriched sample based on three independent experiments, measured on *E. coli* MDS42 cells. ND represents no plaques detected. Bar represents the mean. Error bar shows standard deviation based on $n = 3$ independent experiments.

a

Sample ID	Sample description	Sample treatment	Sample source location
1	Wastewater (primary effluent)	Chloroform treated	Massachusetts, USA
2	Wastewater (primary effluent)	0.22 µm filtered	Massachusetts, USA
3	River water	0.22 µm filtered	Massachusetts, USA
4	Soil from horse and alpaca enclosure	0.22 µm filtered	Massachusetts, USA
5	Water from porta-potty overflow	0.22 µm filtered	Massachusetts, USA
6	Irrigation water near wild rat nest	0.22 µm filtered	Massachusetts, USA
7	Soil from pasture-raised chicken enclosure	0.22 µm filtered	Massachusetts, USA
8	Soil from egg-laying chicken shed	0.22 µm filtered	Massachusetts, USA
9	Fresh pig feces	0.22 µm filtered	Massachusetts, USA
10	Mixed fresh horse and cow feces	0.22 µm filtered	Massachusetts, USA
11	Farm compost pile	0.22 µm filtered	Massachusetts, USA
12	Farm compost pile	0.22 µm filtered	Massachusetts, USA
13	Water from pig and goose water trough	0.22 µm filtered	Massachusetts, USA

b

Isolate	Genome size (kb)	Taxonomic classification
REP1	167.4	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Tevenvirinae; Tequattrovirus
REP2	167.9	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Tevenvirinae; Tequattrovirus
REP3	167.9	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Tevenvirinae; Tequattrovirus
REP4	166.8	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Tevenvirinae; Tequattrovirus
REP5	88.9	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Ounavirinae; Felixounavirus
REP6	87.2	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Ounavirinae; Felixounavirus
REP7	88.9	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Ounavirinae; Felixounavirus
REP8	85.8	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Ounavirinae; Felixounavirus
REP9	85.8	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Ounavirinae; Felixounavirus
REP10	85.7	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Ounavirinae; Felixounavirus
REP11	85.7	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Ounavirinae; Felixounavirus
REP12	85.8	Viruses; Duplodnaviria; Heunggongvirae; Uroviricota; Caudoviricetes; Caudovirales; Myoviridae; Ounavirinae; Felixounavirus

Figure 28: a) Environmental samples analyzed in this study for the presence of lytic viruses of Syn61Δ3. Samples containing lytic phages of Syn61Δ3 are highlighted in gray. b) Genome size and taxonomy of lytic viruses infecting Syn61Δ3. Annotated genome sequences of all REP phage isolates are available in the Supplementary Data of this paper. REP is an abbreviation of Recoded *E. coli* Phage. The annotated genomes of REP phages have been deposited to NCBI GenBank under Accession numbers OQ174500, OQ174501, OQ174502, OQ174503, OQ174504, OQ174505, OQ174506, OQ174507, OQ174508, OQ174509, OQ174510, and OQ174511.

The Dual-Use Education Gap: Awareness and Education of Life Science Researchers on Nonpathogen-Related Dual-Use Research.

Supplementary material 1 - Detailed description of the DUR workshop and results

The DUR workshop was held five times as online event for iGEMers (July 04, 2020; August 20, 2020; September 15, 2021; September 16, 2021; September 17, 2021) and one time for the whole life-sciences community in cooperation with the Alberta RNA Research Training Institute (ARRTI) and SynBio Canada (June 9, 2021). During the current pandemic, the workshop was held via Zoom. The students were separated into four groups, two groups each worked on the same case study. One of the two groups working on the same case study, got a 3 minute introduction in the form of a video. The video contains the definitions for DURC, biosafety and biosecurity from the WHO (World Health Organization, 2006). The four groups then got a DURC example project. The teams worked on two case studies.

1.) The Insect Allies Project

The Insect Allies project by the Defense Advanced Research Projects Agency (DARPA), aims to deliver gene therapy using insects as carriers to engineer traits of agricultural relevant plants (Bextine, 2018). In this project we were mainly interested in finding out whether the students are able to classify risks as biosafety or biosecurity risks.

2.) The Amaranthus Palmeri Project

An example of a possible gene drive solution to combat Amaranthus palmeri in the southern states of the USA from the book “Gene drives on the horizon” (National Academies of Sciences, Engineering, and Medicine, 2016). Aim of the project is to create gene drives in the glyphosate resistant Amaranthus palmeri to reduce or eliminate the weed on agricultural fields in the Southern United States. In this example, we wanted to find out if and why/why not students classify this project as DURC.

The case studies were sent to the students as powerpoint presentations. At the end of the presentations were slides containing 4 questions that the students had to answer:

4. What are the biosafety risks of the project?
5. What are the biosecurity risks of the project?
6. Would you classify the project as DURC project? Why/ Why not?
7. What consequences would you expect if this project would be realized?

After 30 min working on the case studies, the teams presented their results followed by the more detailed presentation with the topic Dual-use and Dual-use Research of Concern - Between Science Misuse and Science Communication. The presentation contained the definitions for Dual-

use, Dual-use Research, biosafety and biosecurity. It highlights the risks of unawareness of Dual-use Research of Concern and the importance of science communication. Furthermore, the example projects are analyzed and an example of failed science communication was presented. Finally, the results of the survey of 2018 and possible countermeasures against biosecurity issues were presented. After the presentation, the concept of the workshop and the intention why we designed the workshop were discussed with the participants. Nearly all groups focused in the beginning on defining or remembering the difference between biosafety and biosecurity. All groups started with discussing biosafety risks closely related to their work in the lab, like unintentional release or mutations arising from their experimental design which could lead to unwanted effects when the example projects would be implemented in the real world. The definitions of biosafety and biosecurity proposed by the teams without prior introduction of the definitions were very close than the definitions proposed by the WHO. Only one of the groups defined biosecurity risks as the risks the project poses to the environment and biosafety risks as the risks the project poses to humans. The problem for the teams was more to classify different risks as biosafety or biosecurity, even if they had or discussed the definitions earlier.

When it came to deciding if the group wants to classify their case study as DURC the groups working on the Insect Allies Project all classified the project as DURC and most presentations mentioned the potential of this project to be used as a bioweapon.

For the Amaranthus palmeri Project the groups differed in their opinion if they would classify this project as DURC. One group without the definitions classified this project not as DURC with the argument that that the technology of using “gene drive” harbours the potential to cause harm, but the specific project of gene drives in Amaranthus palmeri did not fulfil the criteria of potential of directly being misused, when using the first gene drive target. A group with the definition classified this project as DURC even if they just heard the definition that included the “potential to be directly used to cause harm”. We would have expected to see this effect the other way round.

The three minute introduction did not enable the teams to find noticeable better answers to the questions. In the following discussion, the teams who had the three minute introduction told us that the video was useful, especially to realize the difference between biosafety and biosecurity. However, the teams without the definitions knew that there had to be a difference because of the way the questions were formulated which led quickly to the discussion within the group what the difference might be and the outcome of the discussion was close to the actual ones. This might even lead to a better understanding of the definitions, because the students of these groups took a deeper look at the definitions. This seems to underline the same effect seen earlier in the classification as a DURC project or non DURC project: If students have to

come up with definitions themselves, they take a deeper look in these definitions causing a better understanding of the general matter even if the definition might be slightly wrong. This indicates that for the best understanding, letting students come up with their own definitions first and then giving them established definitions probably is the approach that ensures the best understanding of the thematic. The last questions asked what the students would imagine as consequences if their example project would be implemented in the real word. Most teams mentioned mainly the socio-ecological risks and two of the groups mentioned a risk of inequality since both projects harbour the potential of financial benefits for the people who use them, but the technology might not be able to be used by everybody. Only one of the four teams mentioned the possible positive impact of using the technology while all other groups focused on the negative consequences. The groups who have been working with the Insect Allies Project stated again that the project might be used as a bioweapon. In the discussion after the talk, the students stated that none of them was aware of the definitions or difference of biosafety and biosecurity. All students would wish to learn about biosecurity risks in their university education, mirroring the results from the survey. None of the students had a lecture or seminar including these issues. All students told us that “beyond the lab” they only learn about ethical issues and biosafety as far as it concerns lab safety. All students also said that they understated the importance of considering biosecurity issues and were not aware that there are projects as extensively discussed as the Insect Allies Project.

Supplementary material 2 - Survey results and raw data

Question	Possible answers	Responses (Responses in %)
Do you know what Dual-use Research of Concern is?	a) Yes, research with materials that are dangerous if they are released into the environment. b) Yes, research that needs a safety laboratory. c) Yes, research creating results that pose a considerable risk to people and the environment. d) Yes [own answer]. e) No.	18 (9.4) 4 (2.1) 46 (24) 12 (6.3) 112 (58.3)
Do you know what Dual-use in the context of science is?	a) Yes, the sharing of expensive laboratory equipment. b) Yes, the potential of knowledge to be used in both benevolent and malevolent intention. c) Yes, the use of the same technologies for different sciences. d) Yes [own answer]. e) No.	5 (2.6) 79 (41.1) 12 (6.3) 0 (0) 96 (50)
Are you concerned that research results from your discipline could be misused?	a) Yes. b) No. c) I have no opinion on that.	104 (54.2) 57 (29.7) 31 (16.1)
Should results from Dual-use Research of Concern be published freely accessible in your opinion?	a) Yes, without any restrictions. b) Yes, but with censorship if necessary. c) No, that would be too dangerous. d) I have no opinion on that.	50 (26) 118 (61.5) 11 (5.7) 13 (6.8)
How well does your university/ department enlighten about the topics Dual-use and Dual-use Research of Concern?	a) Good. b) Indifferent. c) Poor. d) I have no opinion on that.	30 (15.6) 45 (23.4) 79 (41.1) 38 (19.8)
How does your university inform about Dual-use and Dual-use Research of Concern? [Participants were able to select multiple answers.]	a) Lecture. b) Seminar. c) Research project. d) Other. e) Not at all.	44 31 17 16 105

Would you like to have a more pronounced enlightenment about Dual-use and Dual-use Research of Concern at your university?	a) Yes. b) No. c) I have no opinion on that.	146 (76) 20 (10.4) 26 (13.5)
Do you have a contact person to ask if you have concerns about the effects of your research and publications?	a) Yes. b) No.	80 (41.7) 110 (57.3)

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation eigenständig und ohne unerlaubte Hilfsmittel angefertigt habe. Ich versichere, dass ich keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, sowie Zitate kenntlich gemacht habe. Diese Dissertation wurde in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Bielefeld, den 22.03.2023

Svenja Vinke