Random adversarial threshold search enables specific, secure, and automated DNA synthesis screening

The SecureDNA Team

Summary

Searching for exact matches to unique signatures of hazardous genes enables secure and automated DNA synthesis screening.

Abstract

In 1992, members of the Aum Shinrikyo cult tried but failed to obtain Ebolavirus for use as a biological weapon[1]. Today, many individuals can assemble viruses from synthetic DNA that is not screened for hazards[2,3]. A major barrier to universal screening is the high rate of false alarms requiring expensive human curation. Here we develop, test, and implement “random adversarial threshold” (RAT) search, a highly specific approach that looks for exact matches to short peptide windows and predicted functional equivalents found in hazards but not in any unrelated genes. To determine whether bad actors could obtain replication-competent viruses by incorporating mutations to evade screening, we built databases to protect nine windows found in M13 bacteriophage virus and launched ~21,000 attacks at each window by experimentally building and measuring the fitness of variants with up to six amino acid changes. Finding that defensible windows capable of reliably blocking attacks shared certain predictable features, we identified similar windows from the Australia Group list of pathogens, constructed databases of variants, and wrote software enabling cryptographically secure screening of synthesis orders. RAT search offers a way to safeguard biotechnology by automating DNA synthesis screening.

Introduction

The ongoing COVID-19 pandemic has underscored the danger posed by exponentially spreading biological agents. Virus assembly protocols and inexpensive commercial *de novo* DNA synthesis services have made many hazardous agents accessible to a large and growing number of individuals with relevant technical skills[4]. Future advances may one day lead to publicly available genetic blueprints for pandemic agents considerably more destructive than SARS-CoV-2[5].

Fortunately, most individuals with the skills required to create potential pandemic agents cannot synthesize DNA on their own. Members of the International Gene Synthesis Consortium (IGSC), a trade industry group committed to biosecurity, screen commercial DNA synthesis orders above a certain length for sequences that match the Regulated Pathogen Database[6].

The IGSC firms deserve praise for voluntarily prioritizing safety because doing so is costly: current screening methods based on BLAST generate many false alarms from unrelated sequences that require evaluation by human experts[3,7,8]. As the price of synthetic DNA continues to fall, the effective cost of screening grows.[3] Unfortunately, ~20% of commercially synthesized DNA is generated by non-members who do not screen. Since the list of IGSC members is public, the extent to which current screening meaningfully restricts access to bioweapons is questionable. Even if all current providers did screen, the anticipated arrival of benchtop machines enabling immediate on-site gene synthesis[5] may open another window of vulnerability[3].

Creating a system to screen all commercial and academic DNA synthesis for current and emerging hazards[5,7] demands a method triggering negligibly few false alarms.

We hypothesized that automated DNA synthesis screening could be achieved by “random adversarial threshold” (RAT) search, a strategy that relies on exact-match screening against a database comprising unique signatures of hazards (Fig. 1a). Here we sought to determine whether our approach is specific and secure enough to fully automate DNA synthesis screening.

Results

Random adversarial threshold search (254)

To screen DNA via random adversarial threshold search, a database of hazard signatures is created for comparison to orders (Fig. 1b). We randomly choose 19-amino-acid peptide windows from the protein-coding gene(s) of a hazard for inclusion in the database using a distribution function biased towards defensible windows (see "security analysis" below). 57 nucleotides is short enough that reliably assembling hazards from smaller DNA pieces would be challenging[9] since most methods use much larger oligonucleotides[10]; current approaches only screen 200 base pair fragments and above. Next, a list of peptide variants predicted to be functional is computed for each window using one or more variant effect predictors[11–17]. Variants are curated to remove peptides matching to unrelated sequences in repositories, with the remaining variants included in the database (Supp. Fig. 1). Curation of database entries ensures the system will never wrongly flag unrelated sequences known to science, reducing false alarms. Finally, DNA synthesis orders are searched for exact matches to database entries, and any matched orders are rejected. Given sufficiently high specificity (see below), RAT search can in principle be fully automated, avoiding the requirement for human curation characteristic of current screening procedures.

The identities of all database entries can be guarded using multi-party oblivious encryption, which protects the privacy of DNA synthesis orders sent to be screened and can prevent attacks aiming to determine all of the windows protected for a given hazard. The cryptographic approach is provably secure and may enable screening for emerging hazards without disclosing their identities, a possibility discussed separately[18].

Theoretical specificity analysis

Database curation ensures that RAT search will never flag an unrelated sequence from any known repository. Other sources of false alarms include purely random matches, legitimate research on related pathogens, and oligo libraries encoding variants of known sequences.

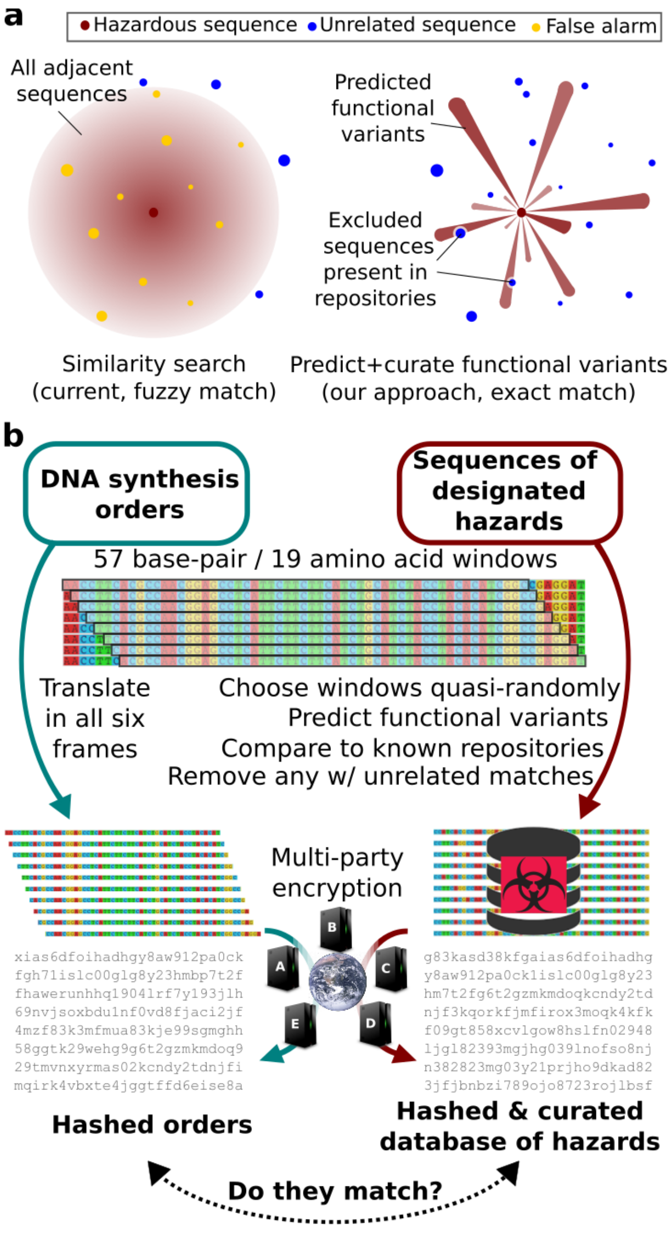

a
● Hazardous sequence ● Unrelated sequence ● False alarm
All adjacent sequences
Predicted functional variants
Excluded sequences present in repositories
Similarity search (current, fuzzy match)
Predict+curate functional variants (our approach, exact match)
b
DNA synthesis orders
Sequences of designated hazards
57 base-pair / 19 amino acid windows
Translate in all six frames
Choose windows quasi-randomly
Predict functional variants
Compare to known repositories
Remove any w/ unrelated matches
Multi-party encryption
A
B
C
D
E
xias6dfoihadhgy8aw912pa0ck
fgh71islc00glg8y23hmbp7t2f
fhawerunhhq1904lrf7y193jlh
69nvjsoxbdu1nf0vd8fjaci2jf
4mzf83k3mfmua83kje99sgmghh
58ggtk29wehg9g6t2gzmkmdoq9
29tmvnxyrmas02kcndy2tdnjfi
mqirk4vbxte4jggtffd6eise8a
g83kasd38kfgaias6dfoihadhg
y8aw912pa0ck1islc00glg8y23
hm7t2fg6t2gzmkmdoqkcndy2td
njf3kqorkfjmfirox3moqk4kfk
f09gt858xcvlgow8hslfn02948
ljgl82393mgjhg039lnofso8nj
n382823mg03y21prjho9dkad82
3jfjbnbzi789ojo8723rojlbsf
Hashed orders
Hashed & curated database of hazards
Do they match?

Figure 1 | Improving DNA synthesis screening
a) Current screening methods yield false alarms from similar but unrelated sequences, requiring human curation (left). Predicting functional variants of k-mers, curating them to avoid false alarms, and searching orders for exact matches could automate screening (right). b) In random adversarial threshold (RAT) search, predicted functional variants of randomly chosen k-mer windows are computed, curated to remove unrelated sequences in repositories, and randomly added to a database. Since adversaries don't know which windows or how many variants are included, evasion requires them to include many mutations throughout the gene or genome and gamble that the result will still be functional. Window sizes of 19 amino acids are large enough to avoid random false alarms. Efficient exact match search allows cryptographic methods to protect the privacy of DNA synthesis orders and the entries in the hazards database[18].

The frequency of random matches depends on the quantity of DNA synthesized and the number of sequences in the database. The probability that the forward or reverse translation of a random 57-mer will match any given database entry is $3.8*10^{-25}$. If we assume a hazards database of 10^{10} entries[19] and that 10^{15} unique (not total) oligonucleotides will be synthesized in 2030, we expect fewer than one random false positive for the entire world's DNA synthesis in that year[20].

To ensure that RAT search does not delay or interfere with legitimate research on related genes and genomes, database hits are checked to determine whether the exact sequence is also present in a related gene that the customer has permission to work with. GenBank/EMBL/ENA accession numbers are extracted from biosafety registration documents listing genes and organisms to create an *exemption list*. Each entry in the hazards database is associated with all accession numbers encoding that peptide.

When an order matches a database entry, the associated accession numbers are compared to the customer's exemption list and the order approved if one is present (Supp. Fig. 2).

Database hits may also result from orders seeking to generate variants of known sequences for experiments such as deep mutational scanning. These cannot be prevented through database curation because the variant sequences are not present in databases. However, orders for oligo pools supporting these types of experiments are typically ordered separately and are algorithmically recognizable, meaning that the wild-type sequence used to generate a suitable oligo pool can be extracted and checked against the hazard database and exemption list as normal. As a last resort, an entire oligo order can be added to the laboratory's exemption list.

Together, these methods suggest that searching for exact matches to functionally essential sequences from hazards will be specific enough to achieve automated screening. Later, we will measure the specificity of RAT search by screening actual customer orders.

Theoretical security analysis

RAT search is only useful insofar as it can reliably detect any given hazard. Including many wild-type sequence windows across a gene or genome will unfailingly detect any order seeking to synthesize that gene or genome[21]. However, a human adversary might incorporate at least one nonsynonymous change every 19 amino acids and still generate a functional sequence. Detecting orders generated by intelligent adversaries requires a careful consideration of how best to include functional variants in the hazard database.

Suppose an adversary seeks hazard W. For each window w_i, there are three possible outcomes:

1. w_i is present in the database, causing the synthesis order to be rejected
2. w_i escapes detection, but imposes a fitness cost c_i that reduces the functionality of W
3. w_i escapes detection without cost

Success requires the adversary to achieve the third outcome for most w_i to preserve function. We define the *random adversarial threshold R* as the probability that an adversary with perfect knowledge of the fitness of each variant – but ignorant of which windows and variants are defended – will be detected on attempting to synthesize functional W.

In theory, defending all variants that do not completely abolish the function of W at a single essential window w_i

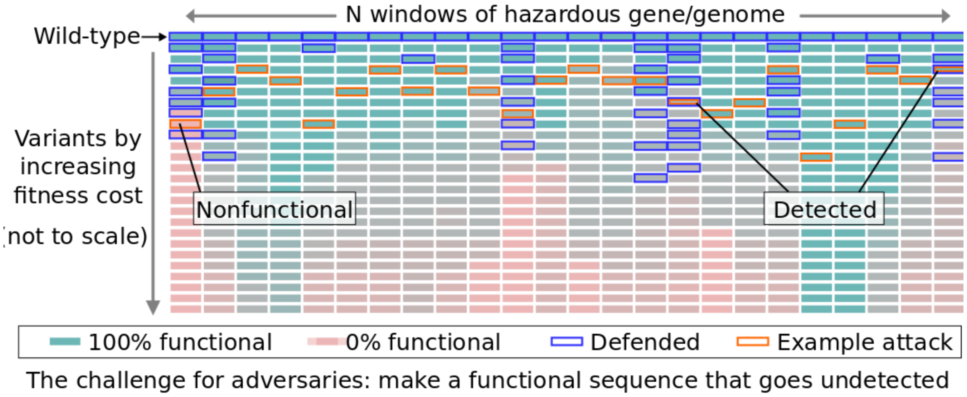

N windows of hazardous gene/genome
Wild-type
Variants by increasing fitness cost (not to scale)
Nonfunctional
Detected
100% functional
0% functional
Defended
Example attack
The challenge for adversaries: make a functional sequence that goes undetected

Figure 2 | Predicted functional variants are included in the database to maximize the adversarial threshold: the probability that an adversary with perfect knowledge of fitness will be detected on choosing a variant for each window. The likelihood of detection rises with the defender's predictive capacity.

can perfectly thwart the adversary, achieving R= 1. In practice, we can only imperfectly predict the fitness of variants.

We can maximize R while minimizing false alarms for a given set of windows by selecting the window predicted to be least tolerant of variation and adding variants with the highest predicted function scores (Fig. 2). At some point, our predictive accuracy will decline until it is better to start defending the window with the second-fewest predicted functional variants. As accuracy declines at the second window, we will either return to the first window or move on to the third.

Since real adversaries do not have perfect knowledge of fitness, we will deviate from the deterministically optimal strategy by stochastically screening for randomly chosen variants from quasi-randomly chosen windows. This forces the adversary to modify all windows throughout the hazard to have a chance at avoiding detection, reducing the odds of functional W.

Experimentally testing security

To experimentally measure the likelihood that an adversary could evade screening for diverse library sizes and window characteristics, we chose to treat the harmless M13 virus that infects *E. coli* as a "hazard". We began by analyzing M13 peptide windows using funtrp, a computational method that categorizes residues within proteins as "neutral" in tolerating most any mutation, "rheostat" in suffering reduced fitness from many but not all mutations, or "toggle" in losing all function when mutated (Supp. Fig. 3). From four different M13 proteins, we selected nine total windows with fairly low to very low neutral values and a range of rheostat and toggle scores.

Next, two "blue team" members constructed databases of 10^6 predicted functional variants for each window using a Metropolis-Hastings algorithm that combined the funtrp scores of each residue with the BLOSUM62 matrix of observed substitutions across proteins (Extended Data Fig. 4). While many variant effector predictors are markedly superior to BLOSUM62[17], our method serves as a baseline for the efficacy of defense that can certainly be improved upon.

"Red team" members experimentally tested the security of RAT search by launching up to 21,000 attacks at each of the nine windows (Fig. 3a). We ordered oligonucleotide pools with all possible combinations of the four most common substitutions at the six positions with the highest neutral scores, pairwise substitutions of all amino acids at those six positions, and all possible single substitutions, plus some predicted deleterious mutations as controls, generated libraries of variants, and measured their individual effects on phagemid replicative fitness (Supp. Fig. 4). Together, our libraries were equivalent to ~10^{36} combinatorial attacks on the databases.

We defined "functional" variants as those with a measured fitness of at least 0.05 relative to wild-type, which is the level at which the most infectious virus known can just barely spread in an unprotected population[22]. Of the functional attacks on the most defensible window, fully 92% were blocked (Fig. 3b). That is, even an adversary with perfect knowledge of fitness who already possesses the other 99% of the wild-type M13 genome sequence was likely to be detected and thwarted just at this one window. While the other windows were less easily defended (Fig. 3c), most windows could still block ~40-50% of attacks (Fig. 3d-e).

Importantly, we observed that the average "toggle" score generated by funtrp for each of the nine windows was predictive of R at that window (Fig. 3e). Analyzing false positive and false negative rates for prediction at each window allowed us to estimate the optimal number of database entries to devote to each window, underscoring differences in ease of defense (Supp. Fig. 5).

Together, these tools allow us to adopt deterministically optimal strategies for defense for adversaries lacking knowledge of window selection: we can pick a desired number of windows to defend, allocate a total number of database entries, set a target value of R, then distribute entries so as to meet R and leave the remainder to be allocated per game theoretical considerations.

Collectively, our 10^6 database entries per window blocked 99.96% of functional attacks on this virus by an adversary with perfect knowledge of fitness. Since real adversaries lacking such knowledge are forced to include at least one nonsynonymous mutation for every nineteen amino acids throughout the genome, and we can afford to allocate 10^9 total entries per hazard, these results underscore the security of RAT search.

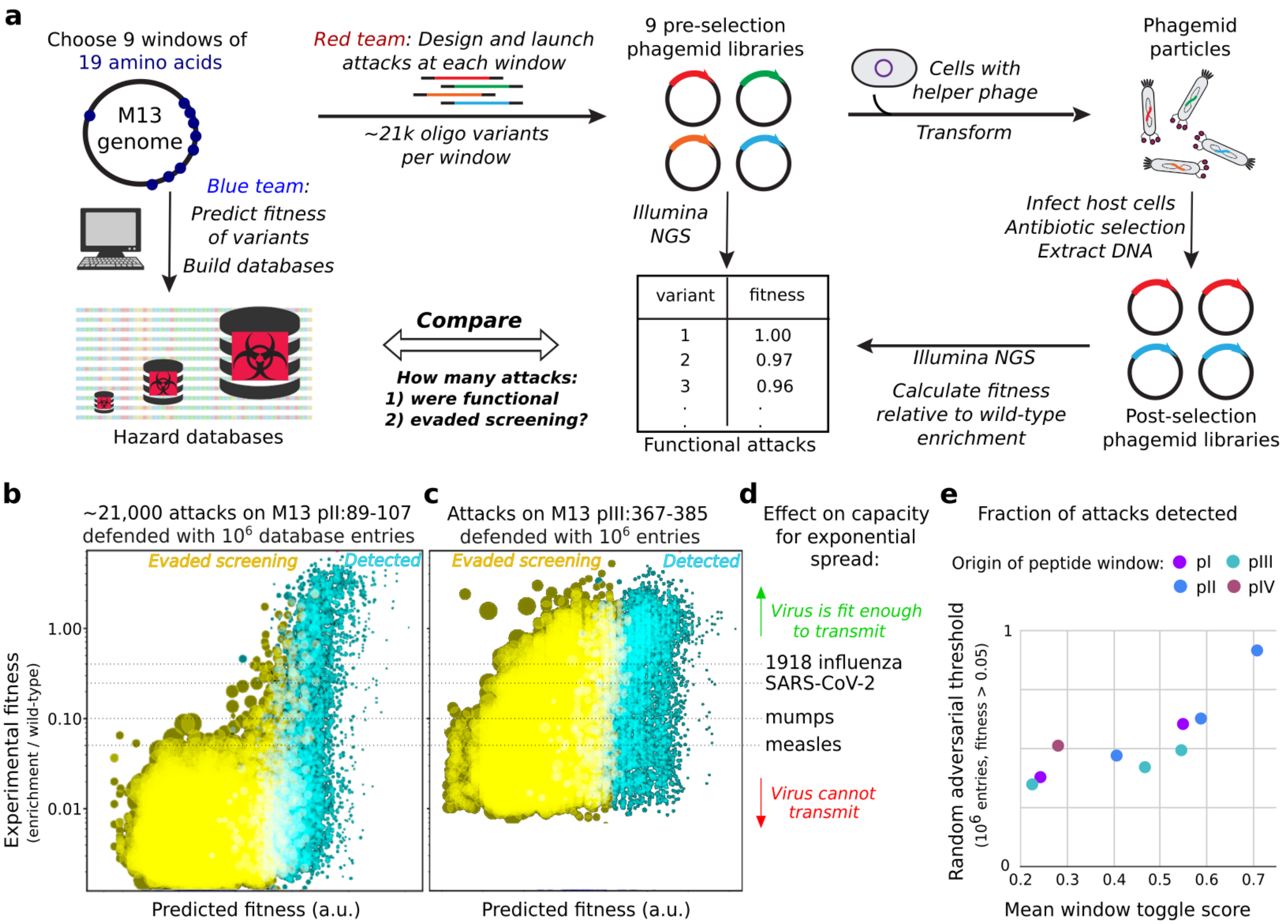


Figure 3 | Incorporating mutations into the genomic blueprint of a virus cannot escape screening. a) Two team members built defensive databases by predicting functional variants for nine different windows in the genome of M13 bacteriophage. Two others launched ~21,000 attacks at each window by synthesizing variants and using a phagemid assay to measure the fitness of each variant, which we defined as enrichment relative to the wild-type sequence. b) At the most defensible window, located within the M13 pII endonuclease, 92% of attacks yielding variants with fitness above 0.05 were thwarted by screening. c) At a moderately defensible window located within the M13 pIII receptor-binding protein, 49% of such attacks were thwarted, underscoring the importance of window choice. d) Potential pandemic pathogens can tolerate only so many mutations impairing fitness before they are no longer capable of sustained transmission. The corresponding fitness lines depict these fitness values for 1918 influenza (R_0~2.5), SARS-CoV-2 (R_0~4), mumps (R_0~10), and measles (R_0~18), which is the most infectious virus known. e) The fraction of attacks detected, which corresponds to the random adversarial threshold, as a function of the average funtrp toggle score for each of the nine windows given a database with 10^6 entries and a fitness cutoff of 0.05 (sufficient to prevent the sustained spread of measles).

Defending against known hazards

Having established a reasonably efficient strategy for randomized defense, we next applied it to the set of well-known hazards considered most important for screening. We analyzed the genomes of all viruses, microbial pathogens, and known genes encoding toxins or pathogenicity islands from the Australia group list for toggle scores using funtrp. Notably, all of the viruses save for swine vesicular disease virus harbored multiple windows with scores above 0.5, for which allocating 10^6 entries blocked over half of attacks (Fig. 3e). Since there are fewer than 100 Australia Group pathogens with publicly available genes or genomes, we can afford to compute and include variants for over a hundred windows per hazard. This will also effectively block the synthesis of related agents.

After optimizing our Metropolis-Hastings algorithm, we chose to create an initial public hazards database comprising 10^7 variants for each of 10 different windows, 10^6 variants for another 40 windows, and $4*10^5$ for a further 150 windows. We subsequently added smaller DNA sequences for many wild-type sequence windows and synonymous codon substitutions, as well as to a fraction of predicted functional variants, to preclude the assembly of hazards from smaller sequences. We can screen against the entire database at once, or optionally use only a fraction of entries until a likely attack is detected, at which point additional entries can be added (Supp. Fig. 6).

Experimentally testing specificity

Now that we have a functional RAT search database capable of screening against all Australia Group pathogens, we are currently working with DNA synthesis providers to empirically measure the false positive rate using real customer order data.

Discussion

Protocols enabling the generation of functional viruses from synthetic DNA have given an increasing number of actors the capacity to build potential pandemic pathogens. As our knowledge of how to build such agents improves, so will the risk of misuse.

Well-resourced state actors cannot be prevented from making DNA, but they are vastly outnumbered by individuals and small groups who possess the necessary technical skills in biology. Screening all commercial DNA synthesis would substantially increase the difficulty of obtaining exponential agents for these non-state actors. A future in which fewer than a hundred groups can build pandemic-class agents is considerably safer than one in which such constructs are accessible to tens of thousands.

Our results suggest that predicting functional variants of randomly chosen windows from hazards, curating them to remove any that match unrelated sequences from repositories, and searching DNA synthesis orders for exact matches would be superior to current methods and could automate screening for public and emerging hazards (Table 1). Implementation could eventually become universal given appropriate incentives favoring incorporation into benchtop DNA synthesizers and assemblers, and eventually into next-generation enzymatic DNA synthesis machines intended for large-scale providers. Once the implementation has proven secure, cryptographic methods similar to those we employ to protect customer data might be used to screen for emerging hazards without inadvertently increasing the likelihood of misuse[18].

By offering a way to fully automate and secure DNA synthesis screening, random adversarial threshold search could substantially mitigate the global catastrophic risks posed by increasingly widespread access to pandemic-class biological agents.

Table 1 | Characteristics of RAT search versus similarity search for DNA synthesis screening

	RAT search	Current similarity search screening
Speed	O(1); very fast	O(database size); slow
Minimum window size	19 amino acids	>= 200 base pairs
False alarm rate	Checking database vs repositories -> no nonrandom false alarms	Many matches to unrelated genes; needs human curation
Fully automatable	Yes	No, requires human curation
Compatible with benchtop synthesizers/assemblers	Yes, given Internet connection	No, requires human curation
Can screen for emerging hazards without disclosure	Yes[18]	No, requires disclosure; too inefficient to encrypt at scale[23]

Methods

Window selection: peptides

The proteins of bacteriophage M13 (Accession NC_003287.2) were analyzed using funtrp[16] to identify peptide windows with few predicted neutral positions and varying numbers of toggle and rheostat positions across and within proteins (Supplementary Table 1, Extended Data Figure 4).

From PI:
- High toggle: MAVYFVTGKLGSGKTLVSV (PI:1)
- High rheostat: YSYLTPYLSHGRYFKPLNL (PI:219)

From PII:
- High toggle: VEIKASPAKVLQGHNVFGT (PII:89)
- Higher neutral, mostly toggle: NFYPCVEIKASPAKVLQGH (PII:84)
- Low neutral overall with several mid-neutral: LLDVNATTISRIDAAFSAR (PII:131)

From PIII:
- High toggle: PQSVECRPFVFGAGKPYEF (PIII:367)
- High toggle: FRGVFAFLLYVATFMYVFS (PIII: 395)
- High rheostat: YANYEGCLWNATGVVVCTG (PIII:48)

From PIV:
- High rheostat: IATTVNLRDGQTLLLGGLT (PVI:360)

Specifically, windows were chosen to enable the following comparisons:

Given few neutrals, how important is the number of toggle vs rheostat positions between proteins?

Sequence (**protein:**amino acid start)	Neutral	Rheostat	Toggle
VEIKASPAKVLQGHNVFGT (**PII**:89)	1.12	3.48	14.4
MAVYFVTGKLGSGKTLVSV (**PI**:1)	1.46	7.11	10.43
FRGVFAFLLYVATFMYVFS (**PIII**: 395)	1.6	8.37	9.03

Given few neutrals, how important is the number of toggle vs rheostat positions within proteins?

MAVYFVTGKLGSGKTLVSV (**PI**:1)	1.46	7.11	10.43
YSYLTPYLSHGRYFKPLNL (**PI**:219)	1.23	13.62	4.15
PQSVECRPFVFGAGKPYEF (**PIII**:367)	2.6	5.76	10.64
YANYEGCLWNATGVVVCTG (**PIII**:48)	3.03	12.17	3.8

Given more neutral positions, is it better to choose windows with more toggle or rheostat positions?

NFYPCVEIKASPAKVLQGH (**PII**:84)	3.47	4.39	11.14
YANYEGCLWNATGVVVCTG (**PIII**:48)	3.03	12.17	3.8
IATTVNLRDGQTLLLGGLT (**PIV**:360)	3.08	10.42	5.5

For windows with comparable mean neutral scores, is it better if they are concentrated or spread out?

NFYPCVEIKASPAKVLQGH (**PII**:84)	3.47	4.39	11.14
LLDVNATTISRIDAAFSAR (**PII**:131)	3.48	7.43	8.09

PII:84 neutrals: 0 0 0 0 0.01 0.02 0.03 0.03 0.03 0.04 0.07 0.10 0.11 0.13 0.13 0.44 0.58 0.83 0.92
PII:131 neutrals: 0 0 0 0 0.01 0.02 0.02 0.03 0.03 0.06 0.08 0.11 0.16 0.31 0.38 0.38 0.48 0.54 0.87

Procedurally generating variants for each 19aa peptide window

1. We included the wild-type sequence (1)
2. We included all one-mutants at each position (19*19 = 361)
3. At the six positions predicted to be most neutral, we added all combinations of the four predicted least pathological substitutions according to BLOSUM62 (5^6 = 15625) (overlaps with one-muts and WT at 4*6+1=25)
4. As negative controls, we included up-to-six mutants of neutral positions using the two most pathological substitutions according to BLOSUM62 (3^6 = 729) (overlaps with one-muts and WT at 2*6+1=13)
5. We added all pairwise combinations of all possible substitutions at the six most neutral positions (19^2 * 15 pairwise combinations = 5415) (overlaps with 4 most tolerated at 4^2 * 15=240) (overlaps with 2 most pathological at 2^2*15=60)

   Total: 1 + 361 - 13 + 15625 - 25 + 5415 - 240 - 60= 21,793 peptide variants at each window

Procedurally generating nucleic acid variants for each 42-mer window

1. We analyzed the region of DNA along with 60bp of flanking sequence in NUPACK in order to assess the likely secondary structure in a knowledge-agnostic manner.
2. For each probably-unpaired, we included all 1-mutants.
3. For each probably-paired, we included all 2-mutants.
4. We included all 2-combinations of the top 10 highest probability single and paired bases.
5. We included all up-to-three mutants of the three highest probability stem pairs and the three highest probability unpaired bases.

However, when we ordered oligos and constructed libraries of variants at these positions, we found that they could not be reliably sequenced, presumably due to secondary structures. We consequently chose to concentrate RAT screening evaluation on peptide windows, but will include 42-mers comprising many synonymous mutants of the wild-type sequence and many predicted functional variants to block the assembly of hazards from short fragments by the small number of non-state actors who are capable of doing so.

Construction of phagemid libraries

Oligo libraries comprising variants for each 19aa peptide window were synthesized as a pool by Twist Bioscience. Individual libraries were amplified by PCR and ligated into a phagemid backbone—encoding an ampicillin resistance gene, containing an M13 phage origin of replication, and designed for library variant expression upon induction by IPTG—using NEBuilder Hifi DNA Assembly Master Mix (NEB, E2621L). Nucleic acid libraries within the origin of replication were ligated into phagemid backbones expressing wild-type pIII. All libraries were then precipitated with isopropanol, transformed into electrocompetent DH5α cells (NEB, C2989K), and plated on 2XYT-carbenicillin-1% glucose; after overnight growth at 37 °C, colonies were counted to ensure >50-fold library coverage. Colonies were scraped with 2XYT and plasmid DNA extracted with the ZymoPURE II Plasmid Maxiprep Kit (Zymo Research, D4203); the extracted plasmid DNA was then precipitated with isopropanol. These plasmid libraries constitute the "pre-selection libraries."

Construction of helper cells

M13cp[24], a plasmid containing all M13 phage genes but with a p15a origin and a chloramphenicol resistance gene replacing the phage origin of replication, was used to construct helper plasmids. Primer pairs were designed for the precise deletion of genes I, II, III, and IV from M13cp following PCR amplification and ligation using the In-Fusion Snap Assembly Master Mix (Takara Bio, 638944). The resulting helper plasmids were transformed into DH5α competent cells (NEB, C2987H), yielding four individual helper cell lines (M13cp-dg1, M13cp-dg2, M13cp-dg3, and M13cp-dg4). The helper cells were made electrocompetent for subsequent same-day transformations. Helper cells are capable of extruding phagemid particles when transformed with a phagemid library variant with a functional gene (complementing the missing phage gene in the helper plasmid) and origin of replication (Supplementary Figure 1).

Phagemid growth

Phagemid libraries were transformed into their corresponding helper cells (nucleic acid variant libraries were transformed into M13cp-dg3) by electroporation and plated on 2XYT-carbenicillin-chloramphenicol-1% glucose. After overnight growth at 37 °C, colonies were counted to ensure >15-fold library coverage. Colonies were scraped with 50 mL 2XYT, the bacterial pellet washed sequentially 3x with 50 mL 2XYT, then a 1:1000 dilution used to inoculate a 50 mL phagemid growth culture in 2XYT with maintenance antibiotics and 1% glucose. The culture was grown to OD_{600} = 0.5 with shaking at 37 °C and 250 rpm, at which point the culture was centrifuged and the media replaced with 2XYT containing maintenance antibiotics and 1 mM IPTG. The culture was grown for 16 h at 37 °C and 250 rpm, after which phagemid-containing supernatants were collected by culture centrifugation and filtration through a 0.22 µm filter.

Phagemid infection

Phagemid-containing supernatants were added to 2.5 mL S2060 cells (streptomycin-resistant, Addgene #105064) grown to OD_{600} = 0.5 and allowed to infect at 37 °C and 250 rpm for 1 h. The resulting infected cultures were plated on 2XYT-carbenicillin-streptomycin-1% glucose to select for phagemid-containing cells. After overnight growth at 37 °C, colonies were scraped with 50 mL 2XYT and plasmid DNA extracted with the ZymoPURE II Plasmid Maxiprep Kit (Zymo Research, D4203). These plasmid libraries constitute the "post-selection libraries."

Illumina NGS sequencing

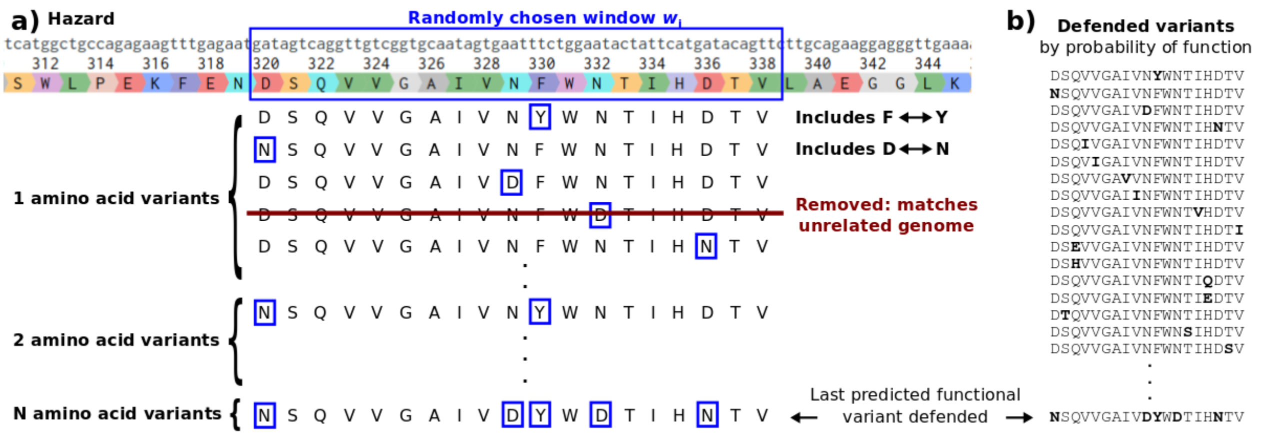
Pre- and post-selection libraries were prepared for illumina NGS sequencing by sequential PCR amplification. PCR amplification was first performed with PrimeSTAR GXL Premix (Takara Bio, R051A) to attach Nextera-style adapter sequences, followed by a second PCR amplification to attach library-specific barcodes and the p5 and p7 indices. Following PCR purification, library concentrations were quantified with qPCR using the NEBNext Library Quant Kit for Illumina (NEB, E7630S), and pre- and post-selection libraries were combined as two pools. Libraries were pooled such that libraries were present in equimolar quantities corrected for library size. Libraries were submitted to the MIT BioMicro Center for MiSeq Illumina sequencing (v3, 2 x 300 bp paired-end).

Supplementary Information

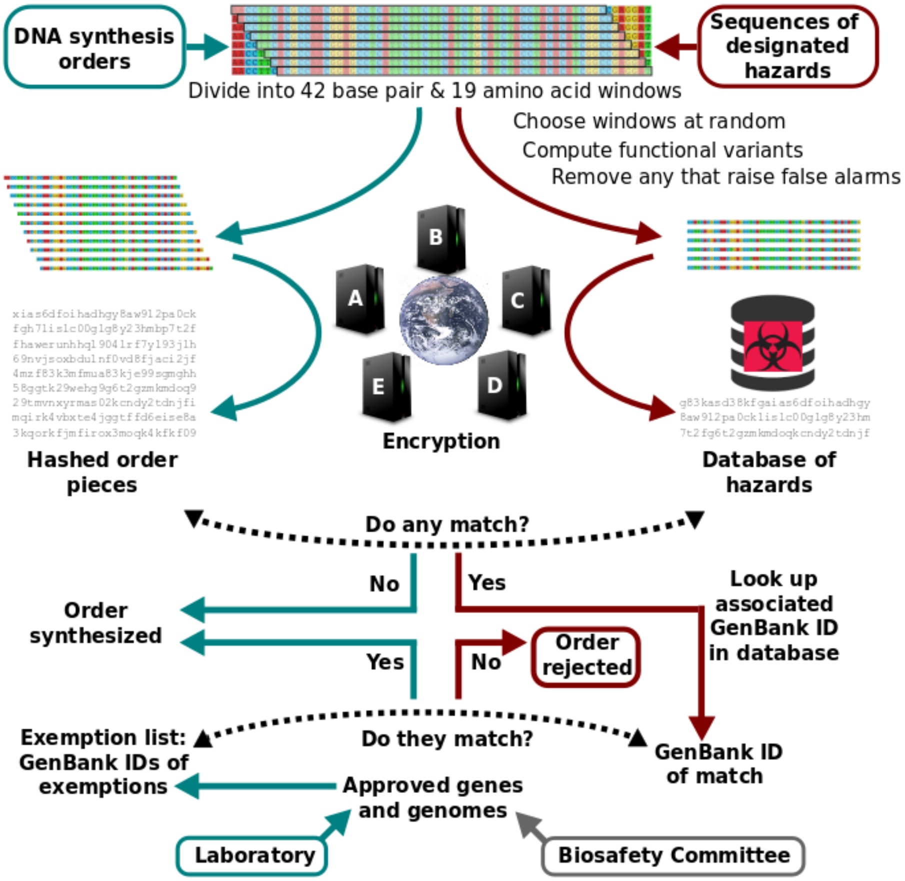
Random adversarial threshold search enables specific, secure, and automated DNA synthesis screening

Dana Gretton[1, †], Brian Wang[1, †], Erika A. DeBenedictis[1,2], Andrew B. Liu[3], Emma Chory[1], Hongrui Cui[4], Xiang Li[5], Jiangbin Dong[5], Andres Fabrega[6], Christianne Dennison[1], Otilia Don[1], Tong Ye[1], Kaveri Uberoy[1], Ron Rivest[6], Mingyu Gao[5], Yu Yu[4,5], Ivan Damgard[7], Carsten Baum[7], Andrew C. Yao[5,8], and Kevin M. Esvelt[1]

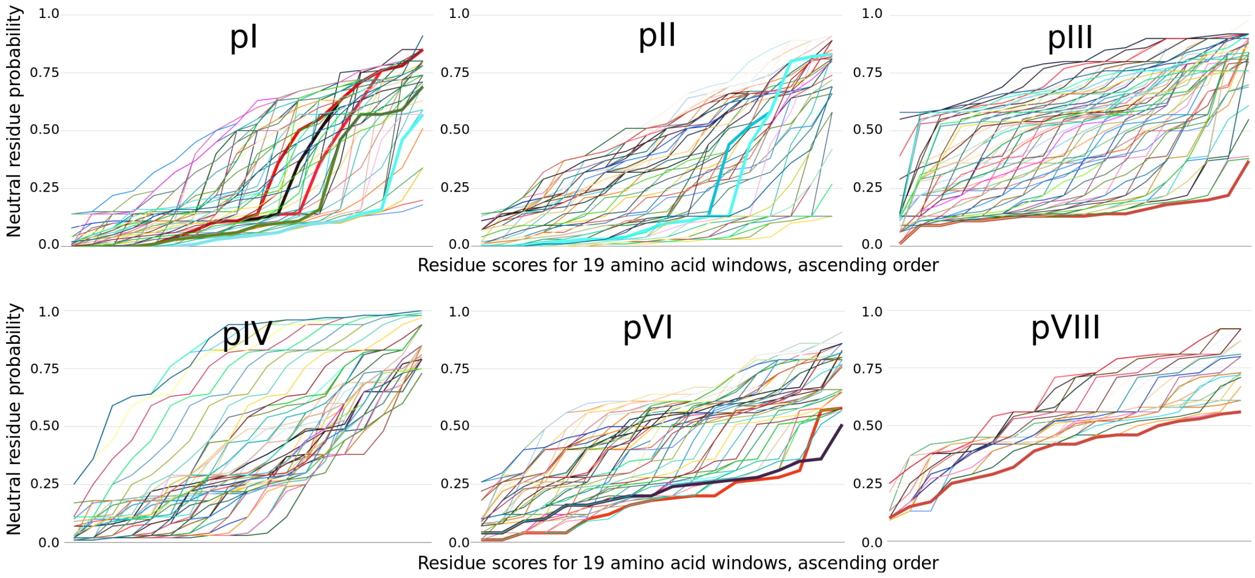
[1]Media Lab, Massachusetts Institute of Technology, USA
[2]Department of Bioengineering, Massachusetts Institute of Technology, USA
[3]Program in Bioinformatics and Integrative Genomics, Harvard Medical School, USA
[4]Department of Computer Science and Engineering, Shanghai Jiao Tong University, China
[5]Institute for Interdisciplinary Information Sciences, Tsinghua University, China
[6]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, USA
[7]Department of Computer Science, Aarhus University, Denmark
[8]Shanghai Qi Zhi Institute, China
[†] equal contribution
*esvelt@mit.edu


a) Hazard
Randomly chosen window w_i
tcatggctgccagagaagtttgagaatgatagtcaggttgtcggtgcaatagtgaatttctggaatactattcatgatacagttcttgcagaaggagggttgaaa
312 314 316 318 320 322 324 326 328 330 332 334 336 338 340 342 344
S W L P E K F E N D S Q V V G A I V N F W N T I H D T V L A E G G L K
1 amino acid variants
D S Q V V G A I V N Y W N T I H D T V
Includes F ⟷ Y
N S Q V V G A I V N F W N T I H D T V
Includes D ⟷ N
D S Q V V G A I V D F W N T I H D T V
~~D S Q V V G A I V N F W D T I H D T V~~
Removed: matches unrelated genome
D S Q V V G A I V N F W N T I H N T V
2 amino acid variants
N S Q V V G A I V N Y W N T I H D T V
N amino acid variants
N S Q V V G A I V D Y W D T I H N T V
Last predicted functional variant defended
b) Defended variants
by probability of function
DSQVVGAIVNYWNTIHDTV
NSQVVGAIVNFWNTIHDTV
DSQVVGAIVDFWNTIHDTV
DSQVVGAIVNFWNTIHNTV
DSQIVGAIVNFWNTIHDTV
DSQVIGAIVNFWNTIHDTV
DSQVVGAVVNFWNTIHDTV
DSQVVGAIINFWNTIHDTV
DSQVVGAIVNFWNTVHDTV
DSQVVGAIVNFWNTIHDTI
DSEVVGAIVNFWNTIHDTV
DSHVVGAIVNFWNTIHDTV
DSQVVGAIVNFWNTIQDTV
DSQVVGAIVNFWNTIEDTV
DTQVVGAIVNFWNTIHDTV
DSQVVGAIVNFWNSIHDTV
DSQVVGAIVNFWNTIHDSV
NSQVVGAIVDYWDTIHNTV

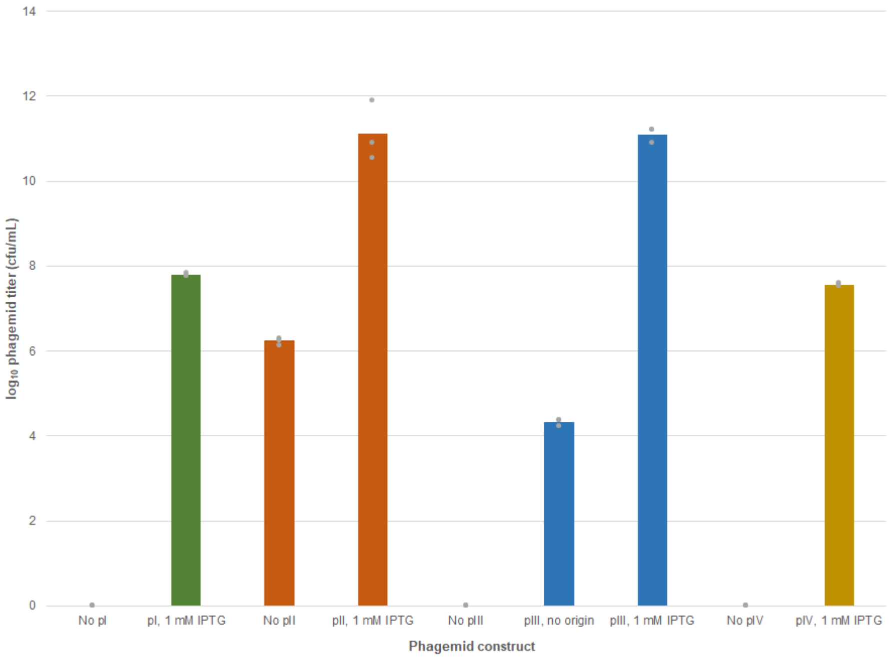
Supplementary Figure 1 | The hazard database includes wild-type sequences and predicted functional variants of randomly chosen windows comprising 19-amino acid peptides from hazardous proteins or 42-base pair DNA/RNA sequences from the noncoding regions of hazardous genomes. a) Examples of variant peptide sequences. Variants matching unrelated sequences in GenBank are removed. b) The random adversarial threshold increases as variants are added to the database. The exact method used to predict the function of variants in order to generate the list will be randomized across several prediction methods to prevent adversaries from predicting the contents of the hazard database.


DNA synthesis orders
Sequences of designated hazards
Divide into 42 base pair & 19 amino acid windows
Choose windows at random
Compute functional variants
Remove any that raise false alarms
B
A
C
E
D
Encryption
Hashed order pieces
Database of hazards
Do any match?
No
Yes
Order synthesized
Look up associated GenBank ID in database
Yes
No
Order rejected
Exemption list: GenBank IDs of exemptions
Do they match?
GenBank ID of match
Approved genes and genomes
Laboratory
Biosafety Committee

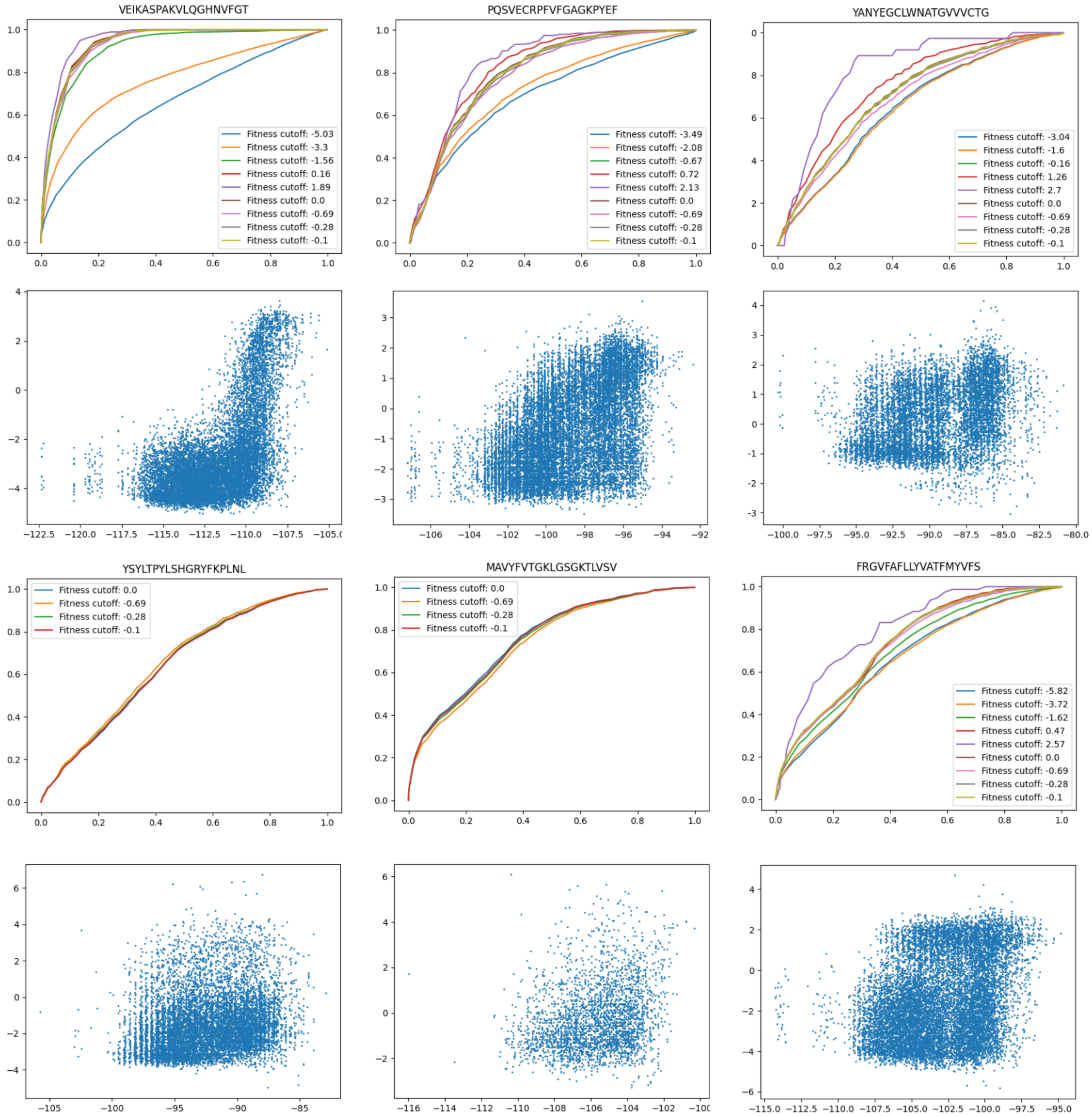
Supplementary Figure 2 | Exemption lists permit automated screening without impeding research. Each laboratory's approved biosafety committee registration is processed to identify gene and organism names, which are used to identify and construct an exemption list of corresponding GenBank accession numbers. Each sequence in the database is associated with one or more GenBank accession numbers to which it corresponds. When an order from a customer associated with an exemption list matches a database entry, the system checks to see whether any of the associated accession numbers match an entry on the exemption list. If so, that particular match is ignored.

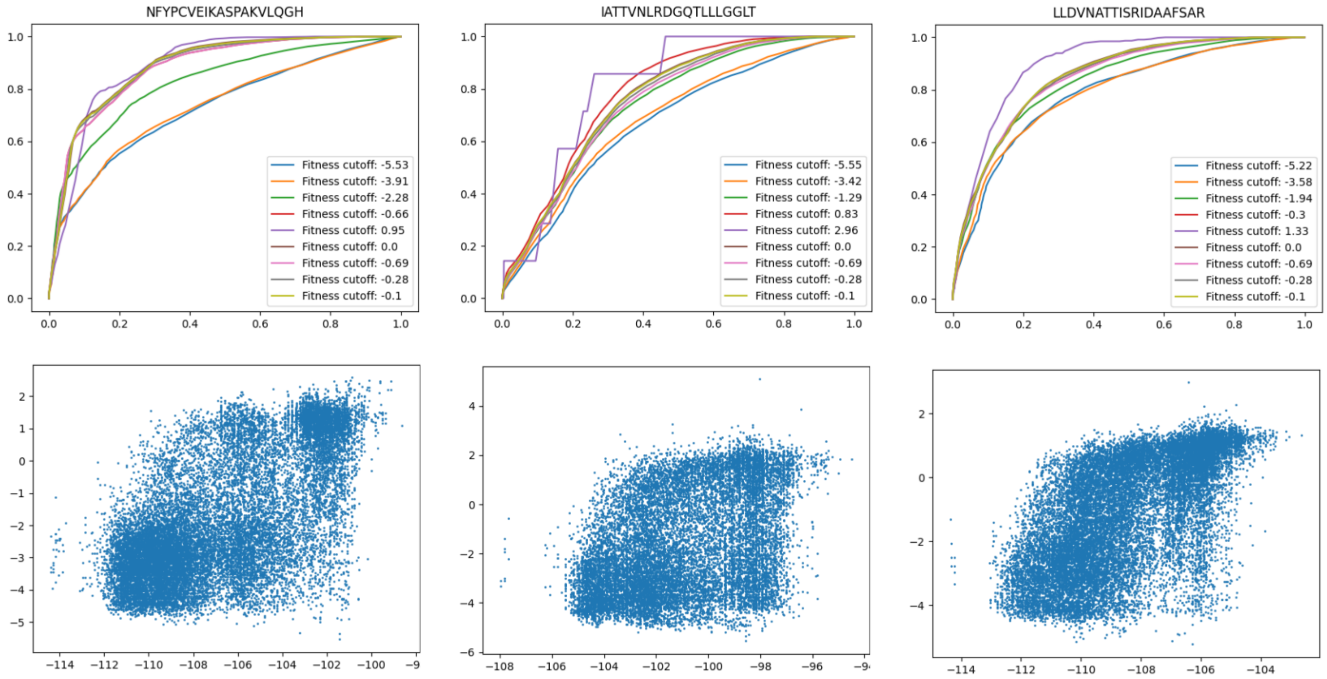


Supplementary Figure 3 | Graphical representations of funtrp window analyses for six M13 bacteriophage proteins. The probability of each residue being neutral for all 19 positions is plotted in ascending order. Windows represented by lines that follow the x-axis for most of their length should be less tolerant of mutations and therefore easier to defend according to funtrp. Larger proteins with more windows appear easier to defend, but there are noteworthy differences between the comparably sized proteins pI-pIV, with the lone enzyme pII predicted to harbor the most windows that are intolerant of mutation.

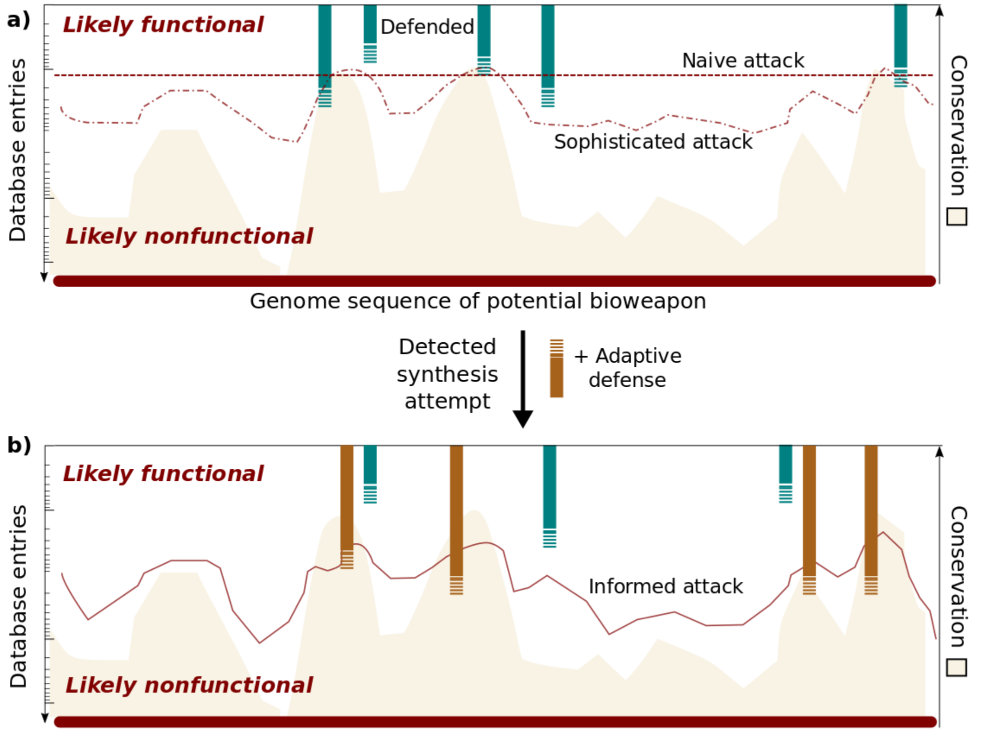


Supplementary Figure 4 | Evaluation of the fitness of phagemids encoding filamentous phage genes I-IV when generated from cells carrying helper plasmids deleted for the gene in question. In all cases, cells extruded phagemid particles when induced with 1 mM IPTG, as measured by infection of recipient cells with 3 independent biological replicates. Data from helper cells transformed with phagemids lacking the wild-type phage genes/origin of replication are provided for comparison. Phagemid titers below the limit of detection (100 cfu/mL) are plotted as zero. The difference in maximum titers suggests that the optimal level of each protein differs from the level produced upon induction, which may affect the relative fitness of variants. For example, overproduction may artificially increase the measured fitness of variants with reduced activity.


VEIKASPAKVLQGHNVFGT
Fitness cutoff: -5.03
Fitness cutoff: -3.3
Fitness cutoff: -1.56
Fitness cutoff: 0.16
Fitness cutoff: 1.89
Fitness cutoff: 0.0
Fitness cutoff: -0.69
Fitness cutoff: -0.28
Fitness cutoff: -0.1
PQSVECRPFVFGAGKPYEF
Fitness cutoff: -3.49
Fitness cutoff: -2.08
Fitness cutoff: -0.67
Fitness cutoff: 0.72
Fitness cutoff: 2.13
Fitness cutoff: 0.0
Fitness cutoff: -0.69
Fitness cutoff: -0.28
Fitness cutoff: -0.1
YANYEGCLWNATGVVVCTG
Fitness cutoff: -3.04
Fitness cutoff: -1.6
Fitness cutoff: -0.16
Fitness cutoff: 1.26
Fitness cutoff: 2.7
Fitness cutoff: 0.0
Fitness cutoff: -0.69
Fitness cutoff: -0.28
Fitness cutoff: -0.1
−122.5 −120.0 −117.5 −115.0 −112.5 −110.0 −107.5 −105.0
−106 −104 −102 −100 −98 −96 −94 −92
−100.0 −97.5 −95.0 −92.5 −90.0 −87.5 −85.0 −82.5 −80.0
YSYLTPYLSHGRYFKPLNL
Fitness cutoff: 0.0
Fitness cutoff: -0.69
Fitness cutoff: -0.28
Fitness cutoff: -0.1
MAVYFVTGKLGSGKTLVSV
Fitness cutoff: 0.0
Fitness cutoff: -0.69
Fitness cutoff: -0.28
Fitness cutoff: -0.1
FRGVFAFLLYVATFMYVFS
Fitness cutoff: -5.82
Fitness cutoff: -3.72
Fitness cutoff: -1.62
Fitness cutoff: 0.47
Fitness cutoff: 2.57
Fitness cutoff: 0.0
Fitness cutoff: -0.69
Fitness cutoff: -0.28
Fitness cutoff: -0.1
−105 −100 −95 −90 −85
−116 −114 −112 −110 −108 −106 −104 −102 −100
−115.0 −112.5 −110.0 −107.5 −105.0 −102.5 −100.0 −97.5 −95.0



Supplementary Figure 5 | Top of each pair: Receiver-classifier-operator curves for funtrp+BLOSUM62 prediction for nine 19-amino acid windows across the genome of M13 bacteriophage. Curves measure distinct fitness cutoffs for prediction. Bottom of each pair: Plots of measured sequence function vs predicted function for each window.



Supplementary Figure 6 | Graphical representation of adaptive defense against a hazard. a) In this example, variants from five windows are included in the hazard database, mostly from relatively conserved regions of the gene. Most of the variants predicted to be highly functional by a variety of different algorithms at each window are included in the database, so a naive attacker who simply introduces a moderate number of mutations at a constant rate is both highly likely to be detected and risks generating a nonfunctional hazard due to the accumulated fitness cost. A sophisticated attacker may try to tune the number of mutations to the likelihood of obtaining a functional sequence across regions, thereby maximizing the chance of evading screening while preserving function. Their chances improve with the superiority of their variant prediction capabilities relative to the defender, but they still must trade off the risk of generating a nonfunctional hazard against being randomly detected upon picking a database variant at one of the five protected windows. b) If multiple attacks on a particular hazard are detected, the system can adaptively add more fragments and variants, precluding informed attacks based on probing or database interrogation. Windows may also rotate in and out periodically.

Implementation Plan

Public hazards

We have predicted and curated variants of sequences from each gene or genome listed on the Australia Group List or the U.S. Department of Health and Human Services Select Agent List. Providers may use this database to test the efficacy and false alarm rate of random adversarial threshold screening. Having contacted numerous DNA synthesis providers, we hope to see the system in widespread use by 2022, including its incorporation into benchtop synthesis and assembler machines.

Emerging hazards

RAT search is compatible with cryptographic approaches capable of obscuring the identities of entries in an emerging hazards database. Such a system would enable a researcher concerned about an emerging potential biological weapon to safely take action to restrict global access without creating information hazards[25,26] by securely conveying their concern to one of the biologists responsible for emerging hazards. Should a minimal number of these experts concur that the threat is serious, they could use their unique keys to add sequences from the hazard to the encrypted emerging hazards database without requiring further disclosure. Many times as many genes or genomes would be chosen as ‘decoys’: related genes or agents that might seem to pose a threat, but are not actually of concern. Decoys can ensure that anyone who finds a match to the database will learn only that it corresponds to a plausible-seeming threat, not that it is a credible weapon. This would ensure that adversaries cannot learn whether a virus can be used to build a weapon of mass destruction by attempting to synthesize it. A detailed explanation of the cryptographic approaches required is available elsewhere[18]. Emerging hazards would only be added to the database after at least two years of testing and bug counties to verify that the implementation is secure.

References

1. Levy, L. & Smithson, A. *Ataxia: The Chemical and Biological Terrorism Threat and the US Response*. https://www.stimson.org/2000/ataxia-chemical-and-biological-terrorism-threat-and-us-response/ (2000).
2. Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* **11**, 499–507 (2014).
3. Diggans, J. & Leproust, E. Next Steps for Access to Safe, Secure DNA Synthesis. *Front Bioeng Biotechnol* **7**, 86 (2019).
4. OECD. OECD: Graduates by field. https://stats.oecd.org/Index.aspx?DataSetCode=EDU_GRAD_FIELD.
5. Esvelt, K. M. Inoculating science against potential pandemics and information hazards. *PLoS Pathog*. **14**, e1007286 (2018).
6. International Gene Synthesis Consortium. Harmonized Screening Protocol V2. https://genesynthesisconsortium.org/wp-content/uploads/IGSCHarmonizedProtocol11-21-17.pdf (2017).
7. DiEuliis, D., Carter, S. R. & Gronvall, G. K. Options for Synthetic DNA Order Screening, Revisited. *mSphere* **2**, (2017).
8. Maurer, S. M., Fischer, M., Schwer, H., Stähler, C. & Bernauer, H. S. Making Commercial Biology Safer: What the Gene Synthesis Industry Has Learned About Screening Customers and Orders. (2009).
9. Gibson, D. G. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Res.* **37**, 6984–6990 (2009).
10. Plesa, C., Sidore, A. M., Lubock, N. B., Zhang, D. & Kosuri, S. Multiplexed gene synthesis in emulsions for exploring protein functional landscapes. *Science* **359**, 343–347 (2018).
11. Bromberg, Y. & Rost, B. SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res.* **35**, 3823–3835 (2007).
12. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* **7**, e46688 (2012).
13. Hopf, T. A. *et al.* Mutation effects predicted from sequence co-variation. *Nat. Biotechnol.* **35**, 128–135 (2017).
14. Gray, V. E., Hause, R. J., Luebeck, J., Shendure, J. & Fowler, D. M. Quantitative Missense Variant Effect Prediction Using Large-Scale Mutagenesis Data. *Cell Syst* **6**, 116–124.e3 (2018).
15. Riesselman, A. J., Ingraham, J. B. & Marks, D. S. Deep generative models of genetic variation capture the effects of mutations. *Nat. Methods* **15**, 816–822 (2018).
16. Miller, M., Vitale, D., Kahn, P. C., Rost, B. & Bromberg, Y. funtrp: identifying protein positions for variation driven functional tuning. *Nucleic Acids Res.* **47**, e142 (2019).
17. Livesey, B. J. & Marsh, J. A. Using deep mutational scanning to benchmark variant effect predictors and identify disease mutations. *Mol. Syst. Biol.* **16**, e9380 (2020).
18. The SecureDNA cryptography team. Hiding dangerous DNA in plain sight. *submitted.*
19. Guesstimating the Size of the Global Array Synthesis Market. https://synbiobeta.com/guesstimating-size-global-array-synthesis-market/ (2017).
20. Liu, Z., Venkatesh, S. S. & Maley, C. C. Sequence space coverage, entropy of genomes and the potential to detect non-human DNA in human samples. *BMC Genomics* **9**, 509 (2008).
21. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).
22. Guerra, F. M. *et al.* The basic reproduction number (R0) of measles: a systematic review. *Lancet Infect. Dis.* **17**, e420–e428 (2017).
23. Titus, A. J. *et al.* SIG-DB: Leveraging homomorphic encryption to securely interrogate privately held genomic databases. *PLoS Comput. Biol.* **14**, e1006454 (2018).
24. Chasteen, L., Ayriss, J., Pavlik, P. & Bradbury, A. R. M. Eliminating helper phage from phage display. *Nucleic Acids Res.* **34**, e145 (2006).
25. Bostrom, N. Information Hazards: A Typology of Potential Harms from Knowledge. *Review of Contemporary Philosophy* **10**, 44–79 (2012).
26. Lewis, G., Millett, P., Sandberg, A., Snyder-Beattie, A. & Gronvall, G. Information Hazards in Biotechnology. *Risk Anal.* **39**, 975–981 (2019).