ACS SyntheticBiology

pubs.acs.org/synthbio

Letter

# Securing the Exchange of Synthetic Genetic Constructs Using Digital Signatures

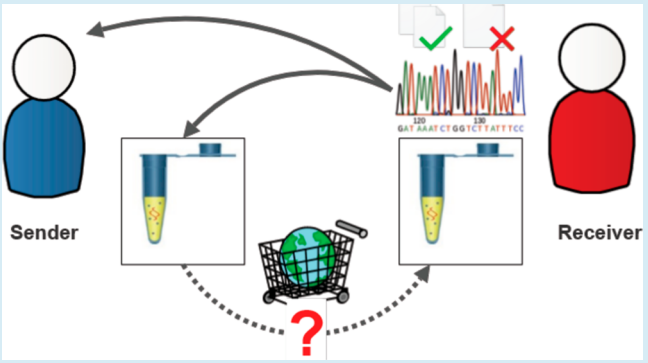
Jenna E. Gallegos,[#] Diptendu M. Kar,[#] Indrakshi Ray, Indrajit Ray, and Jean Peccoud*

Cite This: *ACS Synth. Biol.* 2020, 9, 2656–2664

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** The field of synthetic biology relies on an ever-growing supply chain of synthetic genetic material. Technologies to secure the exchange of this material are still in their infancy. Solutions proposed thus far have focused on watermarks, a dated security approach that can be used to claim authorship, but is subject to counterfeit, and does not provide any information about the integrity of the genetic material itself. In this manuscript, we describe how data encryption and digital signature algorithms can be used to ensure the integrity and authenticity of synthetic genetic constructs. Using a pilot software that generates digital signatures and other encrypted data for plasmids, we demonstrate that we can predictably extract information about the author, the identity, the integrity of plasmid sequences, and even annotations from sequencing data alone without a reference sequence, all without compromising the function of the plasmids. Encoding a digital signature into a DNA molecule provides an avenue for genetic designers to claim authorship of DNA molecules. This technology could help compliance with material transfer agreements and other licensing agreements.




**KEYWORDS:** *DNA, security, cryptography, synthetic biology*



The bioeconomy increasingly relies on the circulation of synthetic DNA molecules that have been designed in software by genetic engineers.[1−6] Synthetic biologists use software tools to generate DNA sequences encoding complex functions. In this context, watermarks[7−9] have been inserted in synthetic DNA to assert authorship claims. DNA watermarking demonstrates the need to assert authorship, but watermarks are insufficient as watermarks can easily be modified, removed, or pasted in different sequences.

Users of engineered DNA sequences also need to establish the origin and integrity of the molecules they work with in order to minimize the risk of wasting time and resources in unproductive experiments.[10−12] Some genetic designs are patented[13,14] and their use is subject to the terms of licensing agreements. When it is not possible or practical to patent a genetic design, authors can control the circulation of genetically engineered samples through material transfer agreements (MTAs).[15−20] MTAs are intended to enable dissemination of genetic designs while maintaining author credit and control. It can be challenging for users of genetically engineered material to comply with agreements dictating the terms of use, because there is no simple way to easily associate a physical sample and an agreement. Finally, people who use synthetic DNA molecules blindly can expose themselves to biosecurity risks.[21,22] We propose the application of digital signatures to DNA in order to support more trustworthy transactions in the bioeconomy supply chain.[21,23]

A digital signature is a mathematical technique used to validate digital messages, documents, web pages, and software. The signature of a document relies on a hash function that maps the entire document to a small string of a set length. The resulting hash value is then encrypted using the author's private key. The person receiving the document can decrypt the signature using the author's public key. Therefore, unlike watermarks, digital signatures are derived from and inextricably linked to both the content being signed and the author. Digital signatures ensure authenticity by enabling the recipient to confirm that the content was generated by an authentic source. Digital signatures also ensure integrity, because the signature is invalidated if the signed content is changed in any way.

As electronic documents are signed by embedding digital signatures, it is conceptually possible to digitally sign genetic constructs by inserting DNA sequences encoding the output of a digital signature algorithm. Such an approach would provide a greater level of security than watermarks but is impractical because validation would require some knowledge of the

**Received:** July 28, 2020
**Published:** September 11, 2020







https://dx.doi.org/10.1021/acssynbio.0c00401
*ACS Synth. Biol.* 2020, 9, 2656–2664

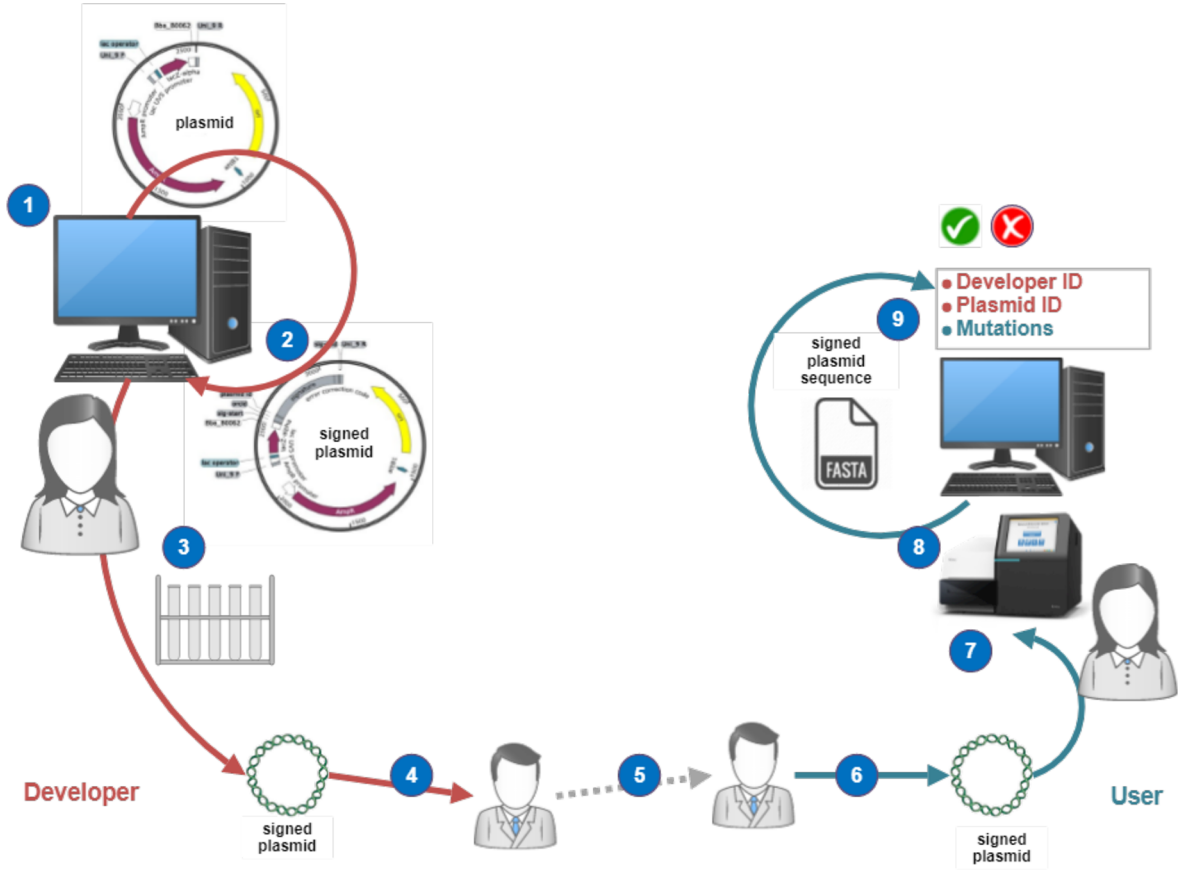



**Figure 1.** Secured exchange of signed plasmids. (1) Developer designs plasmid. (2) Using software, developer signs the plasmid with private key. Signature is inserted in the plasmid sequence along with the developer's ORCID, plasmid ID, and ECC. (3) Developer synthesizes signed plasmid from its computer-generated sequence. (4–5) Signed plasmid circulates through a series of undocumented transactions. (6) Plasmid user receives the signed plasmid. (7) Plasmid user sequences plasmid. (8) Sequencing reads are assembled to generate the complete unannotated sequence of the signed plasmid. (9) Software extracts the ORCID, plasmid ID, and detects mutations. Using the ORCID as public key, software validates the digital signature, ensuring the integrity and authenticity of the signed plasmid.

author's identity. In addition, this approach would not allow for unique identification of the genetic construct itself. Finally, the signature validation algorithm would invalidate any sequence that does not perfectly match that which was signed. This stringent validation may be excessive considering the occurrence of spontaneous mutations.

To address these challenges, we developed a system for applying digital signatures to DNA that is robust to minor mutations and establishes authorship and sequence identification without prior knowledge using the following innovations:

1. We used the author's ORCID as a public identifier and encoded it in the DNA sequence. Other public identifier's like the DUNS number of companies could also be used.
2. We encoded in the sequence a six-digit user-generated number, which serves as a plasmid identifier (Plasmid ID). The Plasmid ID could correspond to a laboratory information management system (LIMS) record, a serial number for a licensing agreement, or a catalog number.
3. We introduced an Error Correction Code (ECC) that makes it possible to identify a limited number of mutations in the event of an invalidated signature. The ECC relies on algorithms like those that use redundancy in DVDs to reconstruct the original data. The 32-base pair ECC used in the subsequently described experiments detects up to two single nucleotide polymorphisms (SNPs).

In this manuscript, we experimentally validate a previously described signature scheme, introduce an improved signature scheme, and introduce the possibility of encoding plasmid annotations along with the signature, such that sequence annotations can be recovered from a FASTA file to produce the corresponding GenBank file.

## RESULTS

We developed a pilot software that takes a GenBank (.gb) file as input along with the author's ORCID and a Plasmid ID. The application generates a new GenBank file that includes the ORCID, Plasmid ID, ECC, and the digital signature at a user-defined location. The algorithms used for the software are described in detail in a previous publication[24] and in the Supporting Information. This "signed" plasmid can then be ordered from a DNA synthesis company or assembled from individual DNA fragments.

Using the same software, plasmid users can recover the ORCID and Plasmid ID from assembled sequencing reads in the form of a FASTA file. To identify the signature cassette within the sequence, the software uses 10 bp sequence delimiters. The software identifies the signature cassette, extracts the author and plasmid identities, and determines whether or not the signature is valid by regenerating the original sequence from the decrypted signature. If the signature is invalid, the ECC is invoked. If the number of errors is below the threshold set by the ECC, the


https://dx.doi.org/10.1021/acssynbio.0c00401
ACS Synth. Biol. 2020, 9, 2656–2664

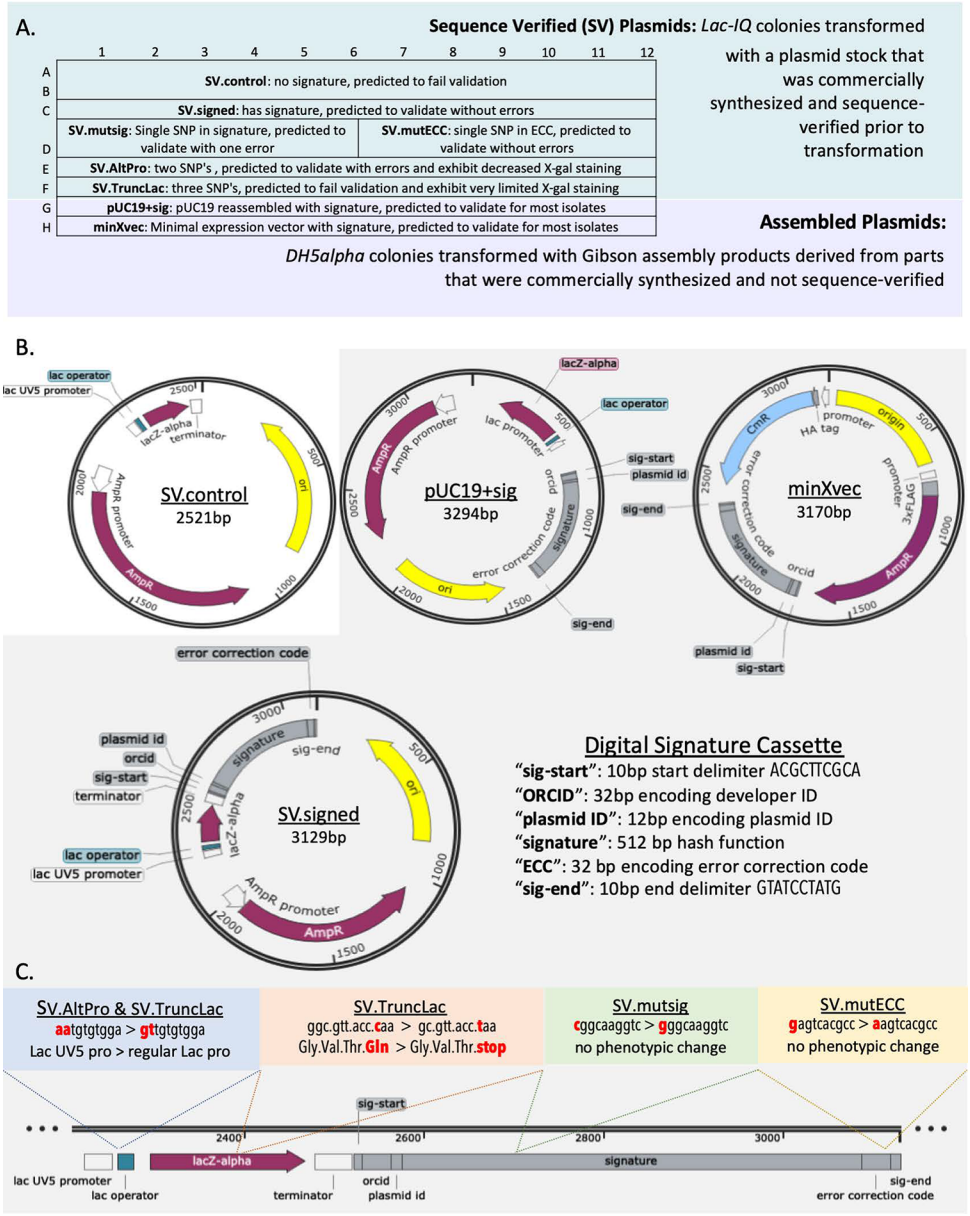



**Figure 2.** Description of plasmids in this study. (A) Depicts how transformants of each plasmid are arrayed on a 96-well plate and describes expected result. Includes 24 transformants for SV.control; 12 for SV.signed, SV.AltPro, and SV.TruncLac; 6 for SV.mutsig and SV.mutECC; and 12 isolates from Gibson assemblies for pUC19 + sig and minXvec. (B) Maps for four of the eight plasmids and a breakdown of digital signature cassette components. (C) Schematic showing how each of the remaining four plasmids differs from SV.signed.

software reports the changes. If the number of errors exceeds the threshold, the software reports that there are too many errors. For the experiments, we set an error limit of 2 base pairs. A larger number could have been set but this would have resulted in a larger synthetic DNA fragment. The ECC algorithm uses Reed-Solomon codes[25] and is described in detail in the Supporting Information.

This system thus allows the recipient of a plasmid to verify that the sample was shared by an authentic sender and that the original sequence is unchanged, or changed only in trivial ways, all without a reference sequence (Figure 1).

**Experimental Verification of Signed Plasmids.** To validate our ability to verify digital signatures from sequencing data and ensure that digital signatures do not interfere with the function of plasmids, a series of plasmids were designed. These plasmids are described in Figure 2 and include:

1. A commercial plasmid (pUC19 + sig)
2. A minimal expression vector (minXvec)
3. A family of six sequence verified clones with and without signatures with varying mutations (SV.control, SV.signed, SV.mutsig, SV.mutECC, SV.AltPro, and SV.TruncLac).

Plasmids pUC19 + sig and minXvec were constructed by Gibson assembly of four DNA fragments. Because the fragments were not sequence-verified, some random mutations were expected.

The sequence verified plasmids include a *lacZ* expression construct. These constructs were ordered from Twist Biosciences as sequence verified clones within one of the vendor's predefined backbones. SV.control and SV.signed are identical except that the latter contains a signature cassette.

SV.mutsig and SV.mutECC each have a SNP in the signature or ECC, respectively. SV.AltPro has two SNPs that convert the


https://dx.doi.org/10.1021/acssynbio.0c00401
ACS Synth. Biol. 2020, 9, 2656−2664

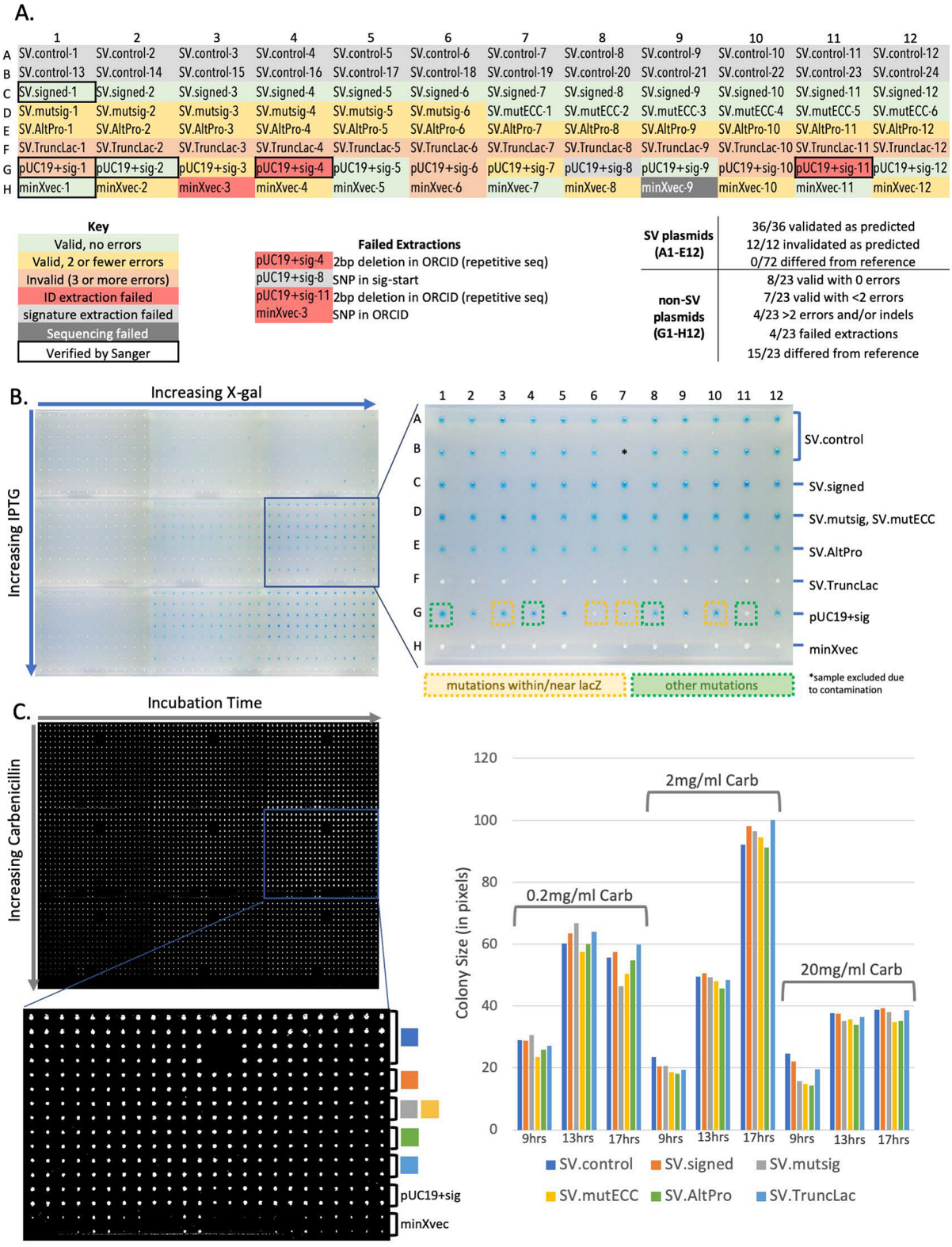


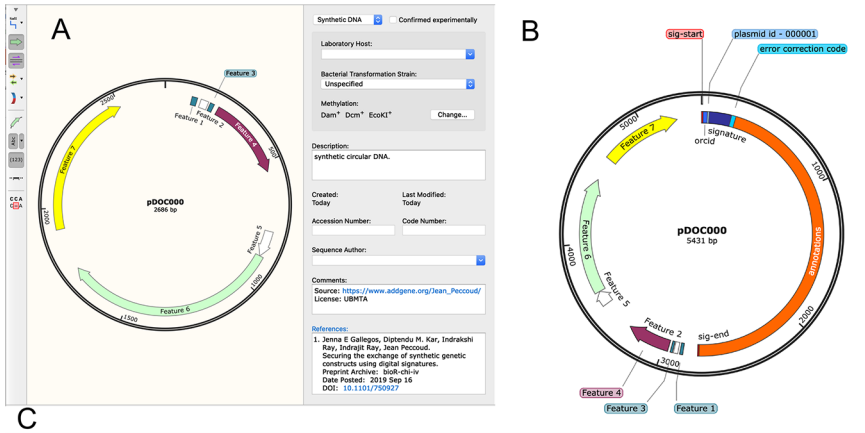

**Figure 3.** Signature validation and phenotyping outcomes. (A) Signature validation for each of the 96 samples. (B) Comparison of lacZ expression in each of the 96 samples across variable inducer (IPTG) and substrate (X-gal) concentrations after 10 h of incubation. (C) Comparison of colony growth for each of the 96 samples pinned in quadruplicate on variable concentrations of carbenicillin up to 20 mg/mL.

stronger *lac*UV5 promoter[26] to the weaker wildtype *lac* promoter. SV.TruncLac is the same as SV.AltPro except for an additional SNP that introduces a premature stop codon after the 24th amino acid of the *lacZ* gene.

DNA was extracted from 6 to 24 transformants for each plasmid for a total of 96 samples and sequenced. Using the assembled FASTA files, signature validation was performed. The results for each plasmid were confirmed *via* manual reference alignment and are represented in Figure 3A.



https://dx.doi.org/10.1021/acssynbio.0c00401
*ACS Synth. Biol.* 2020, 9, 2656−2664




C

```
>pDOC000_signed_annotations  (5431 bp)
agacgaacgaccgtggcagctccgcaggtttaggggcaaacgcggcgtttcactgtaccaatcttcatcgttatacggcc
aagacttaccgaggaggtagtacaaaccacatccctataaagctgccctacaatcaggttcgaccgtctaagactctgag
ccctacaatcaagctaggatatttcgagcattagaacacgggtttactttaaatctgcaccctcctcagagtacactccc
attggtggctgagattgtatgatatcccgtgctaggggcccaacgtggtgggtgcaagcaatgagggtatgcggacaacc
aggagttggaaaccttgtagattcggagaaccaacgttcgttgacaaccaaggaaccggccgcttgacatttacgcaaag
cgggtcacgtcatgcaggcaagaaacaggcccctaaatagcacagcaagaagagagccgcctatagtaatcaagttccgg
ggtttttcataatgtctgcgagtacgattttaatgaaaacccggacttctcggcgacatttaaagtgcaaacttgcccgg
tgacaagctggattctactaccctgagggaaatccacgttaaaattaccgtacattgctggcggtgcagatggtaatcga
ccaagaaccaacgctgcccaattgcgctcgcactagatggcacactctcgtatcatctaccgtatactacggttcgtgtc
gtacatgtcgtacggggcaatacattctcatagggatctcccactgtcggtgtgtcgataaatcctgtcgtaatgaaaat
tttcgccaaacagggtctgcggcgttaattctccagcccgcgagcgagccgctatttatttattagcgtaggttatcacg
```

D

```
LOCUS       Exported                5431 bp ds-DNA     circular SYN 24-JUL-2020
DEFINITION  synthetic circular DNA.
ACCESSION   .
VERSION     .
KEYWORDS    pDOC000
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 5431)
  AUTHORS   Jenna E Gallegos, Diptendu M. Kar, Indrakshi Ray, Indrajit Ray, Jean
            Peccoud
  TITLE     Securing the exchange of synthetic genetic constructs using digital
            signatures
COMMENT     SGRef: number: 1; type: "Preprint"; archive: "bioR-chi-iv"; doi:
            "10.1101/750927"; datePosted: "2019-09-16"
COMMENT     Source: https://www.addgene.org/Jean_Peccoud/
            License: UBMTA
FEATURES             Location/Qualifiers
```

**Figure 4.** Encoding sequence annotations in DNA sequences. (A) Users describe the sequence features and include references to publication, transfer agreements, and material sources in the sequence documentation. (B) DNADOC software generates a plasmid sequence that includes a digital signature and a block corresponding to the DNA-encoded sequence documentation in orange. (C) A scientist using a plasmid with DNA-encoded documentation can sequence the plasmid to produce a FASTA file that does not include any readable annotations. (D) The FASTA file can uploaded for the DNADOC server for verification. The verification step will generate a GenBank file that includes all the documentation embedded in the DNA sequences.

All 72 of the transformants from the sequence verified family of plasmids passed/failed validation as expected. 15/23 of the isolates transformed with the Gibson-assembled plasmids (pUC19 + sig and minXvec) contained mutations. Seven of these had two or fewer SNPs, which were captured by the validation software. Four of these with more than two SNPs and/or insertions/deletions (INDELs) failed validation. The remaining four had mutations in the signature cassette itself which resulted in failed signature or ID extraction.

To determine whether adding 608 bp of extraneous sequence negatively impacted the function of the plasmids, we replica plated the 96 transformants onto media with varying concentrations of antibiotic, the *lacZ* substrate (X-gal), or lac operator inducer (IPTG).

The sequence verified family of plasmids were all expected to exhibit comparable, inducible levels of *lacZ* expression, except for SV.AltPro, which has a weaker promoter[26] and SV.TruncLac, which has a weaker promoter and a premature stop codon.[27]

The addition of the digital signature cassette did not noticeably impact antibiotic resistance, *lacZ* expression, or induction (Figure 3). Visible sources of variability in expression included: promoter conversion and protein truncation in SV.AltPro and SV.TruncLac, random mutations in the parts that were not verified by sequencing used to assemble pUC19 + sig and minXvec, and random *lacZ* induction in the absence of IPTG, which occurred in some colonies regardless of the presence/absence of a signature.

Otherwise, individual transformants containing the same plasmid performed comparably within the sequence-verified plasmid family, indicating that the signature cassette is unlikely to impact plasmid function, stability, or retention in the time scale of the tests performed.


https://dx.doi.org/10.1021/acssynbio.0c00401
ACS Synth. Biol. 2020, 9, 2656−2664

The only transformant that did not behave as expected is pUC19 + sig-11 (Figure 3B). It is unclear why *lacZ* expression was silenced in this case, but the plasmid is identical to pUC19 + sig-4 except that pUC19 + sig-4 contains an additional SNP in the signature, so the presence of the signature is likely irrelevant.

**Reduction of the Signature Length.** The signature scheme used to generate the plasmids used in the previous section is the identity-based digital signature scheme proposed by Shamir.[28] The security of this scheme is dependent on the security assumptions of factoring a large integer into two distinct primes. The digital signature size is always 512 base pairs irrespective of the size of the input plasmid. However, it would be beneficial to reduce the size of the signature as it would reduce the cost of the signature, reduce plasmid size, and reduce the odds that the signature sequence might interfere with the plasmid's biological function.

Here, we introduce a new signature algorithm that reduces the signature size based on pairing-based cryptography. The improved signature scheme, which is based on methods described by Sakai-Kasahara,[29] reduces the signature size to 164 base pairs. The algorithm is described in detail in the Supporting Information. Like the RSA-based signature, the Sakai-Kasahara signature can be combined with an error correction code (ECC) for identifying mutations.

**Encoding Annotations in DNA.** After demonstrating that it is possible to add digital signatures to plasmids, we designed an algorithm to broaden the scope of the documentation that can be embedded in the plasmid sequence. In addition to verifying the author and integrity of plasmid sequences, it is important for users to get access to detailed sequence annotations typically captured in GenBank or SBOL files.[30]

We developed a software that encodes sequence annotations along with the digital signature and error correction code in the plasmid DNA sequences. When a signed plasmid that includes DNA-encoded annotations is sequenced, the resulting FASTA file can be verified and upon verification, the entire GenBank file can be produced from it. The details of the workflow are described in the Supporting Information Section 4 and in Figure 4.

As a test case, we retrieved the annotated sequences of the pUC19 plasmid (pUC.gb file in Supplementary Data). We generated derivative plasmid that includes the 164 bp signature, ECC, and a fragment encoding all the sequence annotations (pUC19_signed_annotations.gb in the Supplementary Data). We exported the sequence as a FASTA file after changing the plasmid origin to simulate the result of a sequencing read assembly (pUC19_signed_annotations.fa). Processing the FASTA file involves verifying the signature as described above. In addition, the software can extract the annotations to produce sequence annotations. The resulting pUC19_signed_annotations_extracted_annotations.gb file is equivalent to the pUC19_signed_annotations.gb. This proof of concept example may not be very practical. Because the pUC19.gb file includes a lot of annotations, the synthetic DNA sequence encoding these annotations is 8840 bp long. Embedding all this information is neither practical nor necessary.

A more practical alternative to embedding all annotations would be to only encode the essential information.[31] The pDOC000.gb file in the Supplementary Data includes a minimal set of annotations that include the position, nature, and label of the sequence features. It also includes a bibliographic reference where the user can find a more detailed description of the plasmid along with reference to the MTA and the source of the plasmid. Even in the absence of any information about this plasmid, users can retrieve the source of the plasmid, the terms of use, basic sequence annotations, and a detailed description from sequencing data. All this essential information fits in 2488 bp. Note that this sequence could be reduced even further by simply providing a DOI and describing the licensing terms and source of genetic material in the corresponding document.

**Software Implementation.** The algorithms described in this article have been deployed on a web server which can be accessed at http://dnadoc.cs.colostate.edu/demo/. This prototype provides only basic functionality allowing users to experiment with the algorithms presented here. The Supplementary Data includes a number of sample input and output files generated by the server.

## DISCUSSION

We have demonstrated the feasibility of using digital signatures to encrypt and secure DNA in living organisms. Our system is robust. We validated/invalidated 72 sequence-verified clones with 100% predictability, indicating that neither DNA synthesis, nor sequencing, nor sequence assembly will present significant technical barriers to employing digital signature-based verification schemes for plasmid sequences. Our system is sensitive. We identified 30 isolates with known mutations and 14 with unknown mutations including single SNPs and small INDELS without the use of a reference sequence. Out of 96 samples, our system performed as intended in all but five instances (discussed below).

Of the five samples that we could not verify experimentally using an early version of the software, four had mutations within the signature cassette. In one case (pUC19 + sig-8), a mutation in the start delimiter caused signature extraction to fail. The current version of the software has fixed this issue by searching not only for the exact start sequence, but also for related sequences.[32]

In three cases (pUC19 + sig-4, pUC19 + sig-11, and minXvec-3), a mutation in the ORCID or plasmid ID caused the software to return incorrect information regarding the identity/origin of the sample. Cases where one of the ID strings contains a SNP (ex. minXvec-3) can be amended by revising the pilot software so that it returns an updated ORCID and plasmid ID following error correction. INDELs are much more challenging computationally.

When performing validation, the software aligns the original sequence (recovered from the hash) with the sequence submitted for validation, using the start delimiter as an origin. Hence SNPs can be identified by noting any position where the original and current sequences do not match. In future iterations, we could incorporate more complex sequence alignment algorithms to identify INDELs during error correction.

One additional sample for which we did not achieve the expected result (minXvec-9) simply was not assembled due to the low quality of the sequencing reads. High quality assemblies for 95/96 plasmid samples is a better-than-expected result, especially considering that most of the plasmids submitted (the entire sequence verified family) had the capacity to form a substantial hairpin. We found that we could sequence each of our ∼3 kb test plasmids for $32 by Sanger sequencing (not including primers) or $35 by Illumina (including library prep and sequence assembly for a minimum of 96 samples). Continually decreasing sequencing costs and affordable, primer-agnostic technologies such as Oxford Nanopore will


https://dx.doi.org/10.1021/acssynbio.0c00401
ACS Synth. Biol. 2020, 9, 2656−2664

likely facilitate the adoption of our approach; however, the increased error rates of Oxford Nanopore technology will likely require additional sequencing depth.

Most importantly, introducing a 608 bp signature cassette into a ∼2.5 kb plasmid did not have obvious detrimental effects on the plasmid function. The consequence of inserting computer-generated sequences to encode data in plasmid sequences were previously unknown. It was difficult to rule out the possibility that these "random sequences" may include motifs that could interfere with the plasmid function through spurious transcription[33,34] or other mechanisms.[35,36] In this context, it is encouraging that the insertion of these short DNA sequences did not negatively impact the plasmid used in this exercise. Future work is needed to determine how much data-encoding DNA can be inserted without compromising plasmid function and structural integrity. It will also be necessary to determine rules[37] specifying where these sequences should be inserted with respect to the plasmid functional elements to minimize the odds of negative effects.

We have demonstrated the feasibility of using digital signatures to allow genetic engineers to claim authorship of genetic designs and users of these designs to verify their authenticity, integrity, and intended function. This method could help enforce MTAs and licensing agreements by encoding the name of the designer and a serial number corresponding to the agreement within the signature cassette. Since it can be challenging to identify which agreement governs the use of specific genetic material, our approach would help comply with applicable agreements. For example, if an unauthorized user is found in possession of a serialized genetic design, it will be possible to track it back to the organization that signed the corresponding MTA. This level of attribution can be particularly useful when distributing genetic material with security implications, such as infectious agents[9,22] or controlled substances like biologically derived opioids.[38]

As an engineering discipline, synthetic biology depends on the ability to freely reuse genetic designs developed by others. Members of the community have advocated an open approach to the distribution of genetic parts,[39] inspired by the open source movement in software engineering.

In this context, defending claims of authorship may appear to go against openness. On the contrary, one of the rights granted to authors is to set the terms of access to their work. Open source software is not public domain software developed by anonymous engineers. Most software developers who contribute to open source projects want credit for the same reasons it is important for academics to claim authorship of publications. Claims of authorship should not imply restricted access but should create favorable conditions for development of a vibrant open source movement by allowing authors to get credit and providing them with a framework allowing them to control how their work can be used. Beyond applications to DNA molecules encoding biological functions, we anticipate that authentication of DNA sequences and their authors will become increasingly relevant as DNA is used for information sharing and storage in the future.[40,41]

## METHODS

**Plasmid Construction and Sequencing.** The sequence-verified family of plasmids was designed using GenoCAD software.[42] The vector backbone and inserts were synthesized and sequence-verified by Twist Biosciences (San Francisco, CA).

pUC19 + sig and minXvec were each constructed by Gibson assembly[43] of four DNA fragments synthesized by either Integrated DNA Technologies (Coralville, IA) or Twist Biosciences based on price and technical limitations. Parts for Gibson assembly of pUC19 + sig and minXvec were designed with 40 bp overlaps. Plasmids were assembled using a Gibson Assembly HiFi 1 Step Kit (Synthetic Genomics, Inc., La Jolla, CA), according to the manufacturer's instructions.

DNA was extracted from each of the 96 isolates using a ZR Plasmid Miniprep − Classic Kit (Zymo Research, Irvine, CA). After quality control analysis on an Agilent TapeStation, ∼15−150 ng of each isolate was submitted to SeqWell in a 96-well plate. The plasmids were sequenced at 90× coverage. Assembled FASTA files provided by SeqWell from their proprietary plasmid sequence assembly platform were used for validation.

**Phenotyping.** NEB 5-alpha F'Iq Competent *E. coli* (SV plasmids) or NEB 5-alpha Competent *E. coli* (pUC19 + sig and minXvec) cells from which DNA was extracted for sequencing were pinned from frozen stocks onto LB + 100 μg/mL Carbenicillin plates using a Rotor HAD robot (Singer Instruments, Somerset, UK). Colonies were grown at 37 degrees to saturation and stored in the fridge.

"Master plates" were generated by pinning from the plates in the fridge onto fresh LB + 100 μg/mL carbenicillin plates in either 96 (for variable X-gal/IPTG) or 384 (for variable carbenicillin) array. After ∼12 h at 37 degrees, these plates were each pinned onto the phenotyping plates. A different master plate was used for every 3−6 phenotyping plates.

Phenotyping plates were all standard LB with variable concentrations of carbenicillin (carb), 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal), or Isopropyl β-D-1-thiogalactopyranoside (IPTG). Variable carb plates contained 100 μg/mL, 200 μg/mL, 1 mg/mL, 2 mg/mL, 1 mg/mL, or 20 mg/mL antibiotic. Variable IPTG/X-gal plates contained 100 μg/mL carb, 0 μL/mL, 0.1 μL/mL, or 1 μL/mL 100 mM IPTG, and 1 μL/mL, 5 μL/mL, or 10 μL/mL 50 mM X-gal.

Plates were imaged over 2 h time courses using a Phenobooth (Singer Instruments, Somerset, UK) imaging platform. Plates which best demonstrated any differences between the colonies were selected for figures. To generate the graph in Figure 3C, Phenobooth imaging software was used to process images and export colony size data. Quadruplicate technical replicates of each isolate were averaged, and then those values were averaged across each of the biological replicates (between 6 and 23) for a given time point on a given plate.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.0c00401.

Detailed description of algorithms used in this manuscript (PDF)

Supplementary data with sequences of plasmids used in this manuscript (ZIP)

## AUTHOR INFORMATION

### Corresponding Author

**Jean Peccoud** − *Colorado State University, Chemical and Biological Engineering, Fort Collins, Colorado 80523, United States; GenoFAB, Inc., Fort Collins, Colorado 80528, United States;* orcid.org/0000-0001-7649-6127; Phone: (970) 491-2482; Email: jean.peccoud@colostate.edu



https://dx.doi.org/10.1021/acssynbio.0c00401
*ACS Synth. Biol.* 2020, 9, 2656−2664

### Authors

**Jenna E. Gallegos** − *Colorado State University, Chemical and Biological Engineering, Fort Collins, Colorado 80523, United States;* orcid.org/0000-0003-4875-8163

**Diptendu M. Kar** − *Colorado State University, Computer Sciences, Fort Collins, Colorado 80523, United States*

**Indrakshi Ray** − *Colorado State University, Computer Sciences, Fort Collins, Colorado 80523, United States*

**Indrajit Ray** − *Colorado State University, Computer Sciences, Fort Collins, Colorado 80523, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.0c00401


### Author Contributions

JEG designed and performed experiments, analyzed data, and wrote the manuscript. DKM developed software and wrote the supplement. Indrakshi Ray secured funding. Indrajit Ray designed software. JP designed experiments, secured funding, and wrote the manuscript.

### Author Contributions

[#]JEG and DMK contributed equally to the work.

### Notes

The authors declare the following competing financial interest(s): JP holds an equity stake in GenoFAB Inc., a company that may benefit or may be perceived to benefit from the publication of this article.

## ■ ACKNOWLEDGMENTS

This work was supported by NSF award #1934573 "EAGER: Development of a Tool-Chain to Write and Read Self-Documenting Plasmids", NSF award #1832320 "EAGER: Modeling DNA Manufacturing Processes Using Extensible Attribute Grammars", and by Colorado State University's Office of the Vice President for Research Catalyst for Innovative Partnerships Program. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Office of the Vice President for Research. The work of Indrajit Ray was performed while serving as Program Director at the U.S. National Science Foundation (NSF) and supported by the foundation's Independent Research and Development program for staff. Research findings presented here and opinions expressed are solely that of the authors, and in no way reflect the opinion of the NSF or other federal agencies.

## ■ REFERENCES


(1) Peccoud, J. (2016) Synthetic Biology: fostering the cyber-biological revolution. *Synth. Biol. 1*, No. ysw001.

(2) Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol. 27*, 946−950.

(3) Huang, P. S., Boyken, S. E., and Baker, D. (2016) The coming of age of de novo protein design. *Nature 537*, 320−327.

(4) Lux, M. W., Bramlett, B. W., Ball, D. A., and Peccoud, J. (2012) Genetic design automation: engineering fantasy or scientific renewal? *Trends Biotechnol. 30*, 120−126.

(5) Nielsen, A. A., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E. A., Ross, D., Densmore, D., and Voigt, C. A. (2016) Genetic circuit design automation. *Science 352*, No. aac7341.

(6) Villalobos, A., Ness, J. E., Gustafsson, C., Minshull, J., and Govindarajan, S. (2006) Gene Designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinf. 7*, 285.

(7) Liss, M., Daubert, D., Brunner, K., Kliche, K., Hammes, U., Leiherer, A., and Wagner, R. (2012) Embedding permanent watermarks in synthetic genes. *PLoS One 7*, No. e42465.

(8) Heider, D., and Barnekow, A. (2007) DNA-based watermarks using the DNA-Crypt algorithm. *BMC Bioinf. 8*, 176.

(9) Jupiter, D. C., Ficht, T. A., Samuel, J., Qin, Q.-M., and De Figueiredo, P. (2010) DNA watermarking of infectious agents: progress and prospects. *PLoS Pathog. 6*, No. e1000950.

(10) Baker, M. (2016) 1,500 scientists lift the lid on reproducibility. *Nature 533*, 452−454.

(11) Jessop-Fabre, M. M, and Sonnenschein, N. (2019) Improving reproducibility in synthetic biology. *Front. Bioeng. Biotechnol. 7*, 18.

(12) Freedman, L. P., Cockburn, I. M., and Simcoe, T. S. (2015) The economics of reproducibility in preclinical research. *PLoS Biol. 13*, No. e1002165.

(13) Caliando, B., and Voigt, C. A. (2015) *Genetic Device for the Controlled Destruction of DNA*, (USPTO, Ed.), MIT.

(14) Gardner, T. S., and Collins, J. J. (2005) *Bistable Genetic Toggle Switch*, (USPTO, Ed.), Cellicon Technologies, Inc., US.

(15) Kamens, J. (2014) Addgene: making materials sharing "science as usual. *PLoS Biol. 12*, No. e1001991.

(16) Rodriguez, V. (2005) Material transfer agreements: open science vs. proprietary claims. *Nat. Biotechnol. 23*, 489−491.

(17) Kahl, L., Molloy, J., Patron, N., Matthewman, C., Haseloff, J., Grewal, D., Johnson, R., and Endy, D. (2018) Opening options for material transfer. *Nat. Biotechnol. 36*, 923−927.

(18) Schaeffer, V. (2019) The use of material transfer agreements in academia: A threat to open science or a cooperation tool? *Res. Policy 48*, 103824.

(19) Walsh, J. P., Cho, C., and Cohen, W. M. (2005) View from the bench: Patents and material transfers. *Science 309*, 2002−2003.

(20) Bubela, T., Guebert, J., and Mishra, A. (2015) Use and misuse of material transfer agreements: Lessons in proportionality from research, repositories, and litigation. *PLoS Biol. 13*, No. e1002060.

(21) Murch, R. S., So, W. K., Buchholz, W. G., Raman, S., and Peccoud, J. (2018) Cyberbiosecurity: An Emerging New Discipline to Help Safeguard the Bioeconomy. *Front. Bioeng. Biotechnol. 6*, 39.

(22) Adam, L., Kozar, M., Letort, G., Mirat, O., Srivastava, A., Stewart, T., Wilson, M. L., and Peccoud, J. (2011) Strengths and limitations of the federal guidance on synthetic DNA. *Nat. Biotechnol. 29*, 208−210.

(23) Peccoud, J., Gallegos, J. E., Murch, R., Buchholz, W. G., and Raman, S. (2018) Cyberbiosecurity: From Naive Trust to Risk Awareness. *Trends Biotechnol. 36*, 4−7.

(24) Kar, D. M., Ray, I., Gallegos, J., and Peccoud, J. (2018) Digital Signatures to Ensure the Authenticity and Integrity of Synthetic DNA Molecules, In *Proceedings of the New Security Paradigms Workshop*, pp 110−122, ACM, Windsor, UK.

(25) Reed, I. S., and Solomon, G. (1960) Polynomial Codes over Certain Finite Fields. *J. Soc. Ind. Appl. Math. 8*, 300−304.

(26) Noel, R. J., Jr., and Reznikoff, W. S. (2000) Structural studies of lacUV5-RNA polymerase interactions in vitro. Ethylation interference and missing nucleoside analysis. *J. Biol. Chem. 275*, 7708−7712.

(27) Nishiyama, K., Ichihashi, N., Kazuta, Y., and Yomo, T. (2015) Development of a reporter peptide that catalytically produces a fluorescent signal through alpha-complementation. *Protein Sci. 24*, 599−603.

(28) Shamir, A. (1984) Identity-based cryptosystems and signature schemes, In *Workshop on the Theory and Application of Cryptographic Techniques: CRYPTO 1984* (Blakley, G. R., and Chaum, D., Eds.), pp 47−53, Springer.

(29) Sakai, R., and Kasahara, M. (2003) ID based Cryptosystems with Pairing on Elliptic Curve. *IACR Cryptology ePrint Archive 2003*, 54.

(30) Galdzicki, M., Clancy, K. P., Oberortner, E., Pocock, M., Quinn, J. Y., Rodriguez, C. A., Roehner, N., Wilson, M. L., Adam, L., Anderson, J. C., Bartley, B. A., Beal, J., Chandran, D., Chen, J., Densmore, D., Endy, D., Grunberg, R., Hallinan, J., Hillson, N. J., Johnson, J. D., Kuchinsky, A., Lux, M., Misirli, G., Peccoud, J., Plahar, H. A., Sirin, E., Stan, G. B., Villalobos, A., Wipat, A., Gennari, J. H., Myers, C. J., and Sauro, H. M. (2014) The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. *Nat. Biotechnol. 32*, 545−550.

(31) Peccoud, J., Anderson, J. C., Chandran, D., Densmore, D., Galdzicki, M., Lux, M. W., Rodriguez, C. A., Stan, G. B., and Sauro, H. M. (2011) Essential information for synthetic DNA sequences. *Nat. Biotechnol. 29*, 22−22.
(32) Kar, D. M., Ray, I., Gallegos, J., Peccoud, J., and Ray, I. (2020) Synthesizing DNA molecules with identity-based digital signatures to prevent malicious tampering and enabling source attribution. *J. Comput. Secur. 28*, 1−31.
(33) Wade, J. T., and Grainger, D. C. (2018) Spurious transcription and its impact on cell function. *Transcription 9*, 182−189.
(34) Lloréns-Rico, V., Cano, J., Kamminga, T., Gil, R., Latorre, A., Chen, W.-H., Bork, P., Glass, J. I., Serrano, L., and Lluch-Senar, M. (2016) Bacterial antisense RNAs are mainly the product of transcriptional noise. *Sci. Adv. 2*, No. e1501363.
(35) Warman, E. A., Singh, S. S., Gubieda, A. G., and Grainger, D. C. (2020) A non-canonical promoter element drives spurious transcription of horizontally acquired bacterial genes. *Nucleic Acids Res. 48*, 4891−4901.
(36) Hahn, M. W., Stajich, J. E., and Wray, G. A. (2003) The effects of selection against spurious transcription factor binding sites. *Mol. Biol. Evol. 20*, 901−906.
(37) Cai, Y., Hartnett, B., Gustafsson, C., and Peccoud, J. (2007) A syntactic model to design and verify synthetic genetic constructs derived from standard biological parts. *Bioinformatics 23*, 2760−2767.
(38) Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M., and Smolke, C. D. (2015) Complete biosynthesis of opioids in yeast. *Science 349*, 1095−1100.
(39) Ledford, H. (2013) Bioengineers look beyond patents. *Nature 499*, 16−17.
(40) Church, G. M., Gao, Y., and Kosuri, S. (2012) Next-generation digital information storage in DNA. *Science 337*, 1628.
(41) Erlich, Y., and Zielinski, D. (2017) DNA Fountain enables a robust and efficient storage architecture. *Science 355*, 950−954.
(42) Czar, M. J., Cai, Y., and Peccoud, J. (2009) Writing DNA with GenoCAD. *Nucleic Acids Res. 37*, W40−47.
(43) Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods 6*, 343−345.


https://dx.doi.org/10.1021/acssynbio.0c00401
*ACS Synth. Biol.* 2020, 9, 2656−2664