



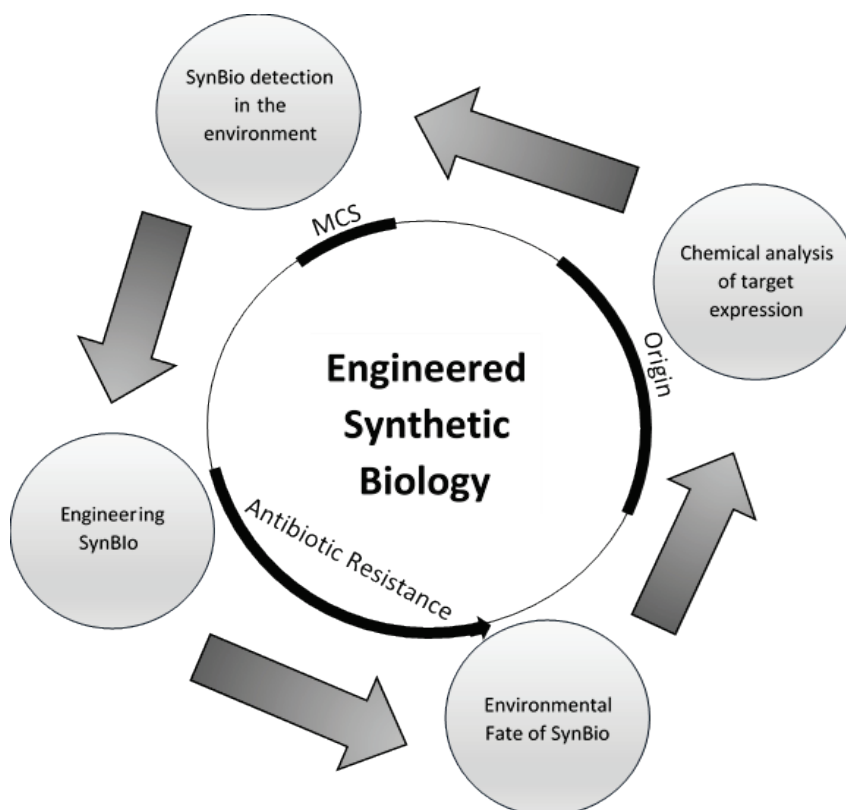
**US Army Corps
of Engineers®**
Engineer Research and
Development Center



Properties and Detectability of Rogue Synthetic Biology (SynBio) Products in Complex Matrices

Carina M. Jung, Karl J. Indest, Matthew R. Carr,
Richard F. Lance, Lyndsay A. Carrigee, and Kayla N.
Clark

September 2022



The U.S. Army Engineer Research and Development Center (ERDC) solves the nation's toughest engineering and environmental challenges. ERDC develops innovative solutions in civil and military engineering, geospatial sciences, water resources, and environmental sciences for the Army, the Department of Defense, civilian agencies, and our nation's public good. Find out more at www.erdc.usace.army.mil.

To search for other technical reports published by ERDC, visit the ERDC online library at <https://erdclibrary.on.worldcat.org/discovery>.

Properties and Detectability of Rogue Synthetic Biology (SynBio) Products in Complex Matrices

Carina M. Jung, Karl J. Indest, Matthew R. Carr, Richard F. Lance,
and Lyndsay A. Carrigee

*Environmental Laboratory
US Army Engineer Research and Development Center
3909 Halls Ferry Road
Vicksburg, MS 39180-6199*

Kayla N. Clark

*Oak Ridge Institute for Science and Education
P.O. Box 117
Oak Ridge, TN 37831-0117*

Final report

Approved for public release; distribution is unlimited.

Prepared for US Army Corps of Engineers
Washington, DC 20314-1000

Under Program Element 0601102A/AB02
Project 484234
Task Number 04

Abstract

Synthetic biology (SynBio) aims to rationally engineer or modify traits of an organism or integrate the behaviors of multiple organisms into a singular functional organism through advanced genetic engineering techniques. One objective of this research was to determine the environmental persistence of engineered DNA in the environment. To accomplish this goal, the environmental persistence of legacy engineered DNA building blocks were targeted that laid the foundation for SynBio product development and application giving rise to “post-use products.” These building blocks include genetic constructs such as cloning and expression vectors, promoter/terminator elements, selectable markers, reporter genes, and multi-cloning sites. Shotgun sequencing of total DNA from water samples of pristine sites was performed and resultant sequence data mined for frequency of legacy recombinant DNA signatures. Another objective was to understand the fate of a standardized contemporary synthetic genetic construct (SC) in the context of various chassis systems/genetic configurations representing different degrees of “genetic bioavailability” to the environmental landscape. These studies were carried out using microcosms representing different environmental matrices (soils, waters, wastewater treatment plant (WWTP) liquor) and employed a novel genetic reporter system based on volatile organic compounds (VOC) detection to assess proliferation and persistence of the SC in the matrix over time.

DISCLAIMER: The contents of this report are not to be used for advertising, publication, or promotional purposes. Citation of trade names does not constitute an official endorsement or approval of the use of such commercial products. All product names and trademarks cited are the property of their respective owners. The findings of this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

DESTROY THIS REPORT WHEN NO LONGER NEEDED. DO NOT RETURN IT TO THE ORIGINATOR.

Contents

Abstract.....	ii
Figures and Tables.....	iv
Preface	vi
1 Introduction	1
1.1 Background	1
1.2 Objective(s)	3
1.3 Approach	3
2 Methods and Materials	5
2.1 Collection of pristine samples	5
2.2 Metagenomic sequencing.....	7
2.3 Multiple cloning site library construction and analysis	7
2.4 Spiking of pristine samples with a legacy SynBio construct.....	8
2.5 Design of cadaverine SynBio construct.....	11
2.6 Chemical analysis of cadaverine	13
2.7 PCR analysis of cadaverine.....	14
2.8 Background cadaverine production by SynBio strains.....	15
2.9 Microcosms to assess the fate of SynBio constructs.....	15
3 Results and Discussion	19
3.1 Pristine sample screening and spiking for SynBio MCS signatures	19
3.2 Background cadaverine production by SynBio strains.....	20
3.3 Microcosms.....	21
4 Conclusions and Recommendations.....	26
4.1 Conclusions.....	26
4.2 Recommendations	26
References.....	28
Report Documentation Page	

Figures and Tables

Figures

Figure 1. A) Arial map of Ft. Huachuca, AZ sampling sites (yellow star) and photos of a B) pristine, secluded spring with C) pump apparatus for filtering samples.	5
Figure 2. Arial map of White Sands, NM A) Guzzler and B) Malapais Spring sampling sites (yellow stars).....	6
Figure 3. The cadaverine gene sequence for <i>cadA</i> and <i>cadB</i> from <i>E. coli</i> K12 MC4100 mined from NCBI Genbank (accession number HG738867.1).....	12
Figure 4. Plasmid map of <i>cadBA</i> construct from GenScript.	13
Figure 5. Broad host range plasmid pBBR1MCS-2 map.....	13
Figure 6. Pictorial representation of the SALLE cadaverine extraction/derivatization procedure (Francisco et al. 2020).	14
Figure 7. Microcosm setup with A) individual microcosms and B) all the microcosms in a 30 °C incubator.....	18
Figure 8. <i>E. coli</i> MG1655 RARE production of cadaverine. Strains containing <i>cadBA</i> in pBluescript are shown in grey with open circles while those containing <i>cadBA</i> in pBBR1MCS2 are shown in black with open squares. The dashed lines for the <i>cadBA</i> containing strains represent the growth medium with 50 µM Lysine and the solid lines represent 100 µM. All other conditions produced negligible amounts of cadaverine.	20
Figure 9. <i>P. putida</i> KTC2440 production of cadaverine. Strains containing <i>cadBA</i> in pBBR1MCS2 are shown in black with open squares. The dashed lines for the <i>cadBA</i> containing strains represent the growth medium with 50 µM Lysine and the solid lines represent 100 µM. All other conditions produced negligible amounts of cadaverine.	21
Figure 10. Cadaverine concentrations as monitored by HPLC from samples, in triplicate, from each matrix. The control containing no SynBio construct (triangle), naked pBBR1MC2:: <i>cadBA</i> (closed square), naked pBluescript:: <i>cadBA</i> (open square), <i>P. putida</i> containing pBluescript:: <i>cadBA</i> (closed circle), <i>E. coli</i> containing pBBR1MC2:: <i>cadBA</i> (open circle).....	25

Tables

Table 1. Location of Pristine sampling sites	6
Table 2. Multiple cloning site library produced and used as a reference for alignment of environmental DNA.....	9
Table 3. Primer names and sequences used for PCR validation of SynBio constructs.....	15
Table 4. Characteristics of the soil used in the microcosm experiments.	17
Table 5. Experimental matrix for the microcosm setup including the number of microcosms for each condition.	17
Table 6. MCS-Spiked recovery from pristine Ft. Huachuca seep pool samples. Sample designations indicate relative concentrations of added SynBio.....	19

Table 7. PCR results from microcosms. Positive results from at least two of the three replicates are represented as X = day 14, Y = day 60, and Z = day 120. 24

Preface

This study was conducted for the US Army Engineer Research and Development Center (ERDC) Military Direct 6.1 Basic Research Program initiated in FY20, as Program Element 0601102A/AB02 Project 484234, Task Number 04.

The work was performed by the Environmental Processes Branch (EPP), Environmental Processes and Engineering Division (EPED), ERDC–Environmental Laboratory (ERDC-EL). At the time of publication, Dr. Rebecca Crouch was Chief, EPP; Mr. Warren Lorentz was Chief, EPED; and Dr. Elizabeth Ferguson was the Technical Director for Environmental Quality and Installations, 6.1 Military Direct Program. The Deputy Director of ERDC-EL was Dr. Brandon Lafferty and the Director was Dr. Edmond J. Russo.

COL Christian Patterson was the Commander of ERDC, and Dr. David Pittman was the ERDC Director.

1 Introduction

1.1 Background

In 2016, a series of formal Scientific Opinions from the European Union emphasized a critical lack of understanding of the potential interactions between synthetic biology (SynBio) constructs (particularly nucleic acid constructs) and endogenous bacterial communities and potential for changes in bacterial evolutionary fitness, ecological competitiveness, horizontal gene movement, and other significant phenomena that might emerge (Epstein and Vermeire 2016). Fundamental knowledge that can inform the determination of potential risks or implications of synthetic biology releases into unmanaged systems and matrices are in their infancy (Torgersen 2009; Hewett et al. 2016; Ahteensuu 2017; Epstein and Vermeire 2016).

The deliberate or accidental release of synthetic biology-derived organisms/products into unmanaged matrices (e.g., soils, waters, urban landscapes/wastewater treatment) represents a vulnerability because current methods, datasets, and predictive models are lacking as to the fate and transport of these products. The potential for engineered DNA to be taken up, incorporated, and expressed in the environment represents a real threat because natural microorganisms display a wide range of genome plasticity allowing them to incorporate exogenous genetic elements. The genetic and physiological adaptability of bacteria is profound and is due in large part to the ability of bacterial cells to acquire genes from other cells (horizontal gene transfer and transduction) or the environment (transformation) (Patel 2016). More specifically, evaluations of DNA in the environment have revealed that although bacterial nucleases in many matrices destroy unprotected DNA, there is still a remarkable amount of free/naked DNA that persists (Nielsen et al. 2007; Romanowski et al. 1992; Romanowski et al. 1993; Tappeser et al. 1998). It is also well-known that naked DNA, even very old and degraded DNA, can be incorporated into bacterial and viral genomes (Johnston et al. 2014; Overballe-Petersen et al. 2013), and bacterial and viral DNA can be incorporated into plant and animal genomes (Crisp et al. 2015; Chilton et al. 1982). A study conducted in 1989 determined that the commonly used cloning vector pBR322 persisted mere minutes in the different treatment matrices of a wastewater treatment plant (WWTP) (Phillips et al. 1989).

Conversely, other survey-based studies have found persistence of genes throughout the WWTP and unique rearrangements of genes indicating movement between hosts/chassis (Parsley et al. 2010; Jung et al. 2009). In this effort, a more modern approach will be explored whereby shotgun metagenome sequencing will be performed, reducing the constraints and biases of previous studies where known targeted DNA for polymerase chain reaction (PCR)-based screenings was employed.

Currently, very little is known about the potential for xenobiotic synthetic biology constructs, when introduced into the environment, to persist via functional integration into natural microbial communities of complex matrices (e.g., soils, urban wastewater, human and animal microbiomes). Fundamental information centered around the fate and transport of xenobiotic synthetic biology constructs in the environment will be critical for the safe and responsible field deployment of these synthetic biology materials. Furthermore, the methodology, data, and algorithms developed to track the fate and transport of synthetic constructs will facilitate detection of future engineered threats (Rogue Synbio) as well as provide critical intelligence on adversarial capability. Synthetic biology is of great concern to national security as Defense Advanced Research Projects Agency (DARPA) and other intelligence agencies are establishing monitoring programs to understand how these reagents are being used and as a mechanism for regulation of production and use

(https://www.realcleardefense.com/articles/2018/11/27/synthetic_biology_an_emerging_national_security_threat_113982.html).

Fundamental research into post-use and associated emergent rogue synthetic biology reagents represents a significant data gap. Of specific interest to Army intelligence gathering is fundamental information required to develop tools for discerning contemporary and past adversary SynBio activities and current and legacy threats. Therefore, the following questions must be addressed: Can SynBio constructs (SC) persist after their intended use and be incorporated into the naturally occurring web of gene transfer that exists ubiquitously? Can aberrant, unregulated gene functions emerge from SynBio constructs and integrate into the pre-existing, non-engineered natural community? These questions are increasingly important, yet also seem daunting to tackle. For example, recently it was demonstrated that non-toxin-producing bacteria of entirely unrelated lineages can be engineered to produce hepatotoxins at concentrations exceeding those achieved by naturally-evolved bacteria

(Liu et al. 2017). The ability to purify toxins on a large scale has the potential to support the production of bioweapons. Inadvertent release of SynBio constructs can have the same outcome, or perhaps more insidiously, similar, or more potent outcomes following unanticipated gene transfer interactions. So, how do we surveil those suspected of weaponizing biological systems? How do we track our adversaries' "products"? How do we use the end-product as a forensic tool to identify the specific group responsible based on what is known of their techniques? The hypotheses stated earlier, and project tasks described in the next section will help move us towards greater capabilities for answering these questions.

- **Hypothesis 1:** Legacy genetic constructs and elements previously released into the environment will persist over time and leave diagnostic signatures of the original source material.
- **Hypothesis 2a:** The half-life of a standardized synthetic genetic element will decrease based on the genetic bioavailability of the chassis/expression configuration with genomic integration of the SC > plasmid expression in a host > naked plasmid DNA.
- **Hypothesis 2b:** All SynBio constructs tested will persist longer in environments with less microbial activity/productivity (sterile media = sandy loam > lake water = topsoil > WWTP).

1.2 Objective(s)

One objective of this research was to determine the environmental persistence of engineered DNA in the environment. Another was to understand the fate of a standardized contemporary synthetic genetic construct (SC) in the context of various chassis systems/genetic configurations representing different degrees of "genetic spillage" into the environmental landscape.

1.3 Approach

To accomplish the first objective, the environmental persistence of legacy engineered DNA building blocks that laid the foundation for SynBio product development and application giving rise to "post-use products" were targeted. Specifically, multi-cloning sites (MCSs), defined as a short segment of synthetic DNA which contains up to ~20 restriction enzyme sites were the focus. The creation of MCSs represents some of the earliest genetic engineering activities initiated in the late 1970s to facilitate early

plasmid gene cloning. These polylinker regions consist of a high density of restriction enzyme sites, normally 6 bp cutters, all located within 20-30 bps. Six base pair recognition sites would be expected to occur randomly every 4,096 bp in a genomic sequence and therefore the presence of ten different recognition sites in a 20-30 bp sequence is highly improbable. Therefore, MCS sites represent an excellent, easily screened signature for the presence of legacy recombinant DNA in the environment. Shotgun sequencing of the total environmental DNA was performed, and the resultant sequence data was mined for these legacy MCS signatures. Studies addressing the second objective were carried out using microcosms representing different environmental matrices (soils, waters, WWTP liquor) and a novel functional genetic reporter system based on VOC detection was employed to assess proliferation and persistence of the SC in the matrix over time.

2 Methods and Materials

2.1 Collection of pristine samples

Sampling from a suspected highly affected site was cancelled due to COVID-19 “stay at home” orders and travel restrictions.

Sampling of water from pristine environments thought to be completely unaffected by agriculture (GMOs) or industry and with limited human exposure took place before COVID-19 restrictions, in September 2019. Samples were collected using a 47 mm, 0.45 μm , sterile GN-6 nitrocellulose filter (Gelman Sciences, Ann Arbor, MI) with at least 4 L water pulled through with a peristaltic pump at the sites shown in Figures 1 and 2, and Table 1.

Figure 1. A) Arial map of Ft. Huachuca, AZ sampling sites (yellow star) and photos of a B) pristine, secluded spring with C) pump apparatus for filtering samples.

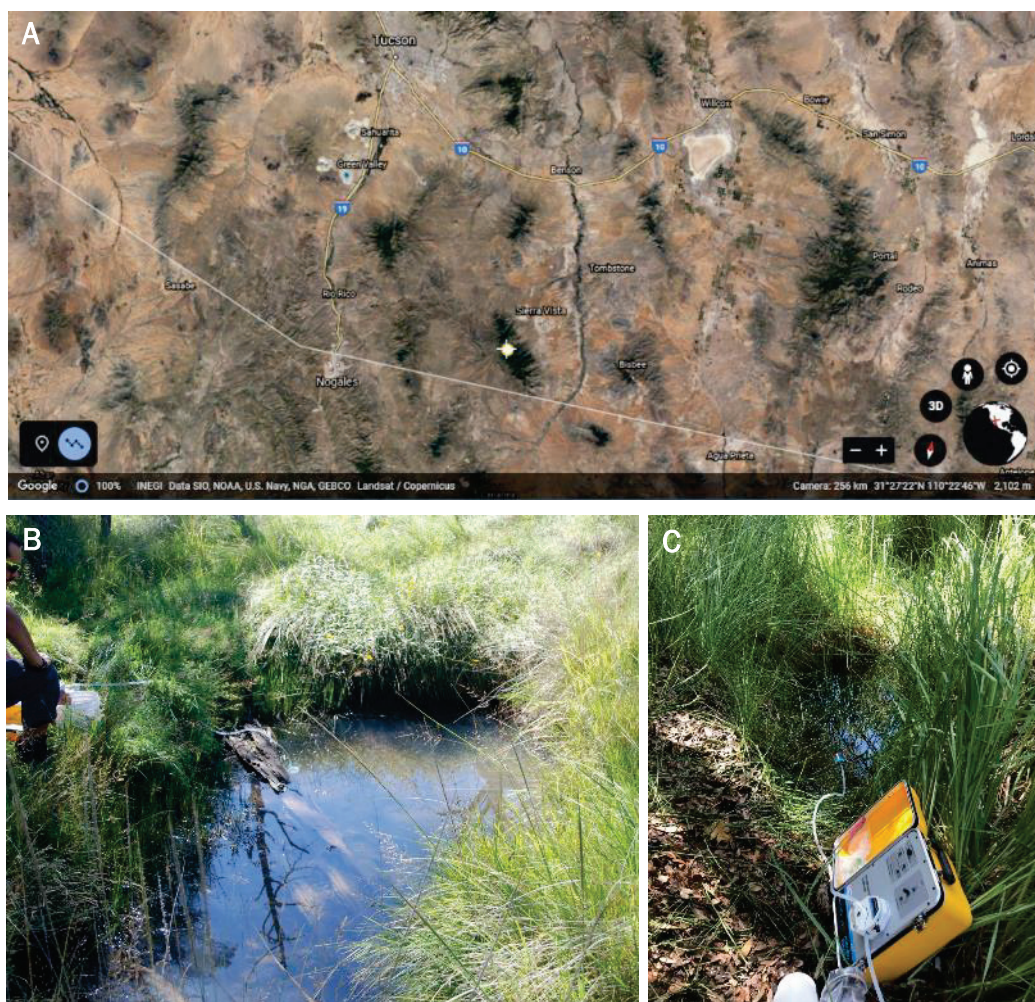


Figure 2. Aerial map of White Sands, NM A) Guzzler and B) Malpais Spring sampling sites (yellow stars).

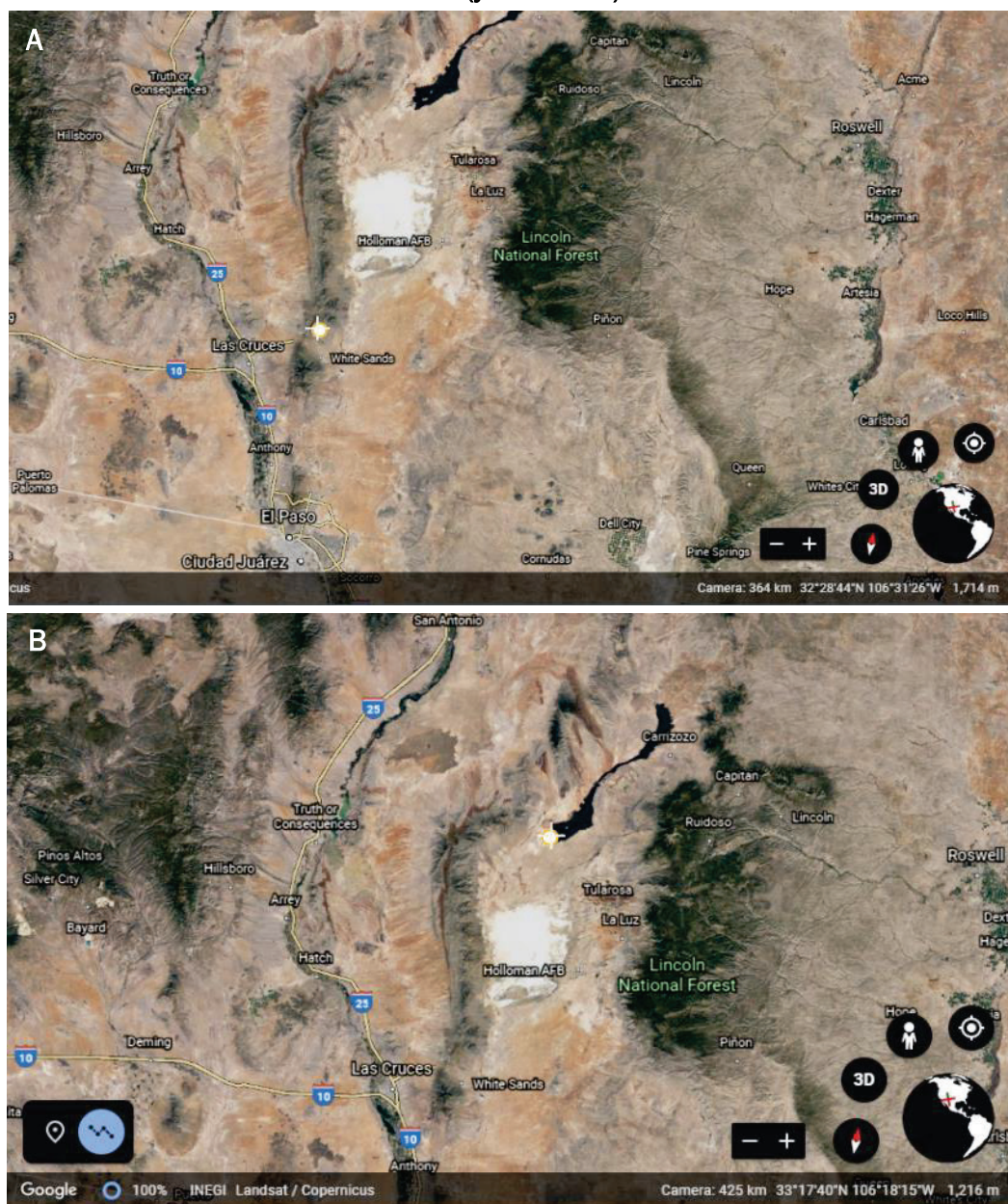


Table 1. Location of Pristine sampling sites.

Locale	Site	Latitude	Longitude	Notes
Fort Huachuca, AZ	Cienaga top of Garden Canyon	31° 27'17.3"N	110° 22'31.3"W	4 L
Fort Huachuca, AZ	Seep Pool Huachuca Canyon	31° 31'01.9"N	110° 23'17.7"W	5 L
White Sands Missile Range, NM	Wildlife Guzzler	32° 28'24.0"N	106° 30'52.7"W	5 L
White Sands Missile Range, NM	Malpais Spring	33° 17'17.0"N	106° 18'37.6"W	5 L

Filters were stored on ice in the field and placed in the -20°C freezer for long-term storage.

2.2 Metagenomic sequencing

DNA was recovered from the filters by first removing the filter from the collection bag with flame-sterile forceps. The filter was then placed in a sterile 5 ml screw-cap tube containing power beads, power bead buffer, and C1 solution from a Qiagen DNeasy PowerSoil kit (Germantown, MD) and prepared following the manufacturer's instructions with modifications. Briefly, the sample tube was secured in horizontal vortex adapter and vortexed for 10 min, then centrifuged at maximum speed for 10 min. The supernatant was removed and transferred to a clean 2 ml microcentrifuge tube and the DNeasy PowerSoil Kit manufacturer's protocol was followed from Step 7.

Extracted DNA was tagged and labeled with the appropriate adapter indices using Illumina's Nextera XT Library Preparation kit (San Diego, CA) following the manufacturer's guidelines. The library was normalized using Nextera XT protocol and diluted to 1.8 pM following Illumina's guidelines for the NextSeq 550 Mid Output kit. Metagenomic sequencing was setup using Local Run Manager Software with a GenerateFastQ workflow.

2.3 Multiple cloning site library construction and analysis

The MCS data mining necessitated the need for an MCS library against which to query. A local repository of known k'mers (indiscriminate number of bases within a sequence) representing MCSs used in genetic constructs was therefore generated (Table 2). Commercial suppliers have large collections of cloning vectors ranging from legacy to contemporary. Sequences are available online and were catalogued and formatted for use as a searchable database against the shotgun metagenomics sequencing of sites. Commercial suppliers searched included ATCC (Manassas, VA), Invitrogen (Waltham, MA), Sigma (St. Louis, MO), New England Biolabs (Ipswich, MA), and Addgene (Watertown, MA).

Addgene allowed for direct search terms and parameters and the following were used to search for appropriate MCSs: "MCS", "multiple cloning site", most requested plasmids/vectors, PI and lab available plasmids, popularity, and expression type (primarily bacterial). The other vendors

Table 2. Multiple cloning site library produced and used as a reference for alignment of environmental DNA.

Vector	Introduced	Restriction Enzymes used	Sequence
pUC18	1983	EcoRI, SacI, Eco53kI, KpnI, Acc651, SmaI, XmaI, BamHI, XbaI, HincII, Accl, Sall, PstI, SphI, HindIII	aattcgagctcgggtacccggggatcctctagagtcgacctgc aggcatgcaagctt
pUC19	1983	HindIII, SphI, PstI, Sall, Accl, HincII, XbaI, BamHI, XmaI, SmaI, Acc651, KpnI, Eco53kI, SacI, EcoRI	aagcttgcagtcctgcaggctcgactctagaggatccccgggt accgagctcgaattc
pNEB193	1985	EcoRI, SacI, KpnI, SmaI, AscI, BssHI, BamHI, PacI, XbaI, Sall, PmeI, SbfI, PstI, SphI, HindIII	aattcgagctcgggtacccggggcgccggatccttaatta agtctagag tcgactgtttaaacctgcaggcatgcaagcttgg
pBBR1MCS-2	1995	Eco53kI, SacI, AclI, SacII, XbaI, SpeI, BamHI, XmaI, SmaI, EcoRI, EcoRV, HindIII, ClaI, HincII, XhoI, ApaI, KpnI	gagctccaccgcgggtggcggcgctctagaactagtggatcc ccccgggctgcagggaattcgatatcaagcttatcgataccg tcgacctcgagggggggcccggtacc
pUC57		EcoRI, Eco53kI, SacI, Acc65I, KpnI, NruI, NsiI, XbaI, EcoRV, BamHI, TspMI, BmeT110I, SmaI, PspOMI, ApaI, Sall Accl, HincII, PstI, StuI, SphI, HindIII	gaattcgagctcgggt acctcggaatgcagctagatatcggtacccggggccgctcga ctgcagaggcctgcagcaagctt
pUK21	1991	XhoI, BglII, BspDI, NsiI, SphI, NcoI, Acc65I, KpnI, XmaI, SmaI, Eco53kI, SacI, EcoRI, BstBI, HindIII, SfiI, PstI, MluI, SgrDI, Accl, HincII, BsaHI, ZraI, AatII, NdeI, BamHI, EcoRV, BglI, NotI, BsiEI, BsrBI, XbaI,	atgaccatgattacgccactagtcaggccctcgagatctat cgatgcagccatggtaccgggagctcgaattcgaagcttc t gcagacgcgtcgacgtcatatggatccgatatcgccgtggcg gccgctctagaactagt
pUC12	1983	EcoRI, Eco53kI, SacI, XmaI, SmaI, BamHI, XbaI, Sall, Accl, HincII, PstI, HindIII	gattacgaattcgagctcgcccggggatcctctagatcgac ctgcagcccaagctt
pUC13	1983	HindIII, PstI, Sall, Accl, HincII, XbaI, BamHI, XmaI, SmaI, Eco53kI, SacI, EcoRI	aagcttgggctgcaggctcgactctagaggatccccgggcga gctcgaatt
pET28GFP	2013	XhoI, NotI, BsiEI, HindIII, Sall, Accl, HincII, Eco53kI, SacI, EcoRI, BamHI	ctcgagtgcggccgcaagcttctgcagcgagctcgaattcgg atcc
pUC18miniTn7	2005	Sall, Accl, HincII, EcoRV, NsiI, Eco53kI, SacI, SpeI, BamHI, XmaI, SmaI, PstI, ApoI, XhoI, HindIII, PspOMI, ApaI, Acc65I, KpnI, NruI, StuI, SfiI	gtcgacgatatcatgcatgagctcactagtggatccccggg ctgcagggaattcctcgagaagcttgggcccggtacctcgcg aaggccttgaggcc
pUC9	1982	HindIII, PstI, Sall, BamHI, SmaI, EcoRI	aagcttgggtgcaggctcgacggatccccgggaattc
pUC8	1982	EcoRI, SmaI, BamHI, Sall, PstI, HindIII	gaattccccgggatccgtcgacctgcagccaagctt
pUC7	1982	EcoRI, BamHI, Sall, PstI, Sall, BamHI, EcoRI	aattccccggatccgtcgacctgcaggatccccgggg aattc
pLITMUS28i	1995	BglII, BstBI, BsmI, NsiI, BtgZI, BssHI, BsiWI, BsmBI, XhoI, EcoRI, SbfI, PstI, EcoRV, BamHI, HindIII, NcoI, ZraI, AgeI, XbaI, BaeI, AvrII, Eco53kI, Acc651, StuI	agatcttcgaatgcagtcgcgccacgtacgtctcgaggaat tcctgcaggat atctggatccacgaagcttcccatggtgacgtcaccggttcta gataacctagg tgagctctggtacctctagtcaaggcct
pLITMUS38i	1995	Apal, MfeI, NaeI, NarI, HindIII, PstI, EcoRV, BamHI, EcoRI, NheI, EagI, MluI, BsmBI, BspEI, BsrGI, SphI, Sall, StuI	cgtaatacgactcactatagggggcccgatgaattgaagccgg ctggcgccaag cttctctgcaggatatctggatccacgaattcgtagcttggc cgtgacgcgt ctccggatgtacaggcatcgctcgacctctagtcaaggcct
pTYB1	1996	NdeI, NheI, NruI, Sall, NotI, EcoRI, XhoI, SapI	catatggctagctcgagtcgacggcgccggaattcctc gagggtcttcc

Vector	Introduced	Restriction Enzymes used	Sequence
pTYB2	1996	NdeI, NheI, NruI, Sall, NotI, EcoRI, XhoI, SmaI	catatggctagctcgcgagtcgacggcgccgcgaattcctc gagcccgagg
pTYB3	1996	NcoI, NheI, NruI, Sall, NotI, EcoRI, XhoI, SapI	accatggctagctcgcgagtcgacggcgccgcgaattcctc gagggtctctcc
pTYB4	1996	NcoI, NheI, NruI, Sall, NotI, EcoRI, XhoI, SmaI	accatggctagctcgcgagtcgacggcgccgcgaattcctc gagcccgagg
pTYB21	1998	SapI, SacI, NdeI, NcoI, NotI, Sall, BamHI, EcoRI, SbfI, PstI	ggaagagctcatatgtccatggcgccgcgatatcgtcgac ggatccgaatt ccctgcaggttaattaaataac
pTYB22	1998	BsmI, NdeI, NcoI, NotI, Sall, BamHI, EcoRI, SbfI, PstI	gaatgctggtcatatgtccatggcgccgcgatatcgtcga cgatccgaatt ccctgcaggttaattaaataac
pTYB11	1998	SapI, SpeI, NruI, Sall, NotI, EcoRI, XhoI, SmaI, PstI	agaagagctactagttcgcgagtcgacggcgccgcgaatt cctcgagcccgagg gtgactgcag
pTYB12	1998	BsmI, NdeI, SpeI, NruI, Sall, NotI, EcoRI, XhoI, SmaI, PstI	gaatgctggtcatatgactagttcgcgagtcgacggcgccgc cgaattcctcga gcccgggtgactgcag
pTWIN1	1999	SapI, NcoI, NotI, EcoRI, XhoI, SapI	ggaagagccatggcgccgcgaattcctcgagggtctctcc
pTKGluc	2002	NotI, AgeI, XhoI, XbaI	gcggccgcaccggctcgcgagcatgcatctaga
pBAD24	1995	NheI, EcoRI, NcoI, KpnI, SmaI, BamHI, XbaI, Sall, PstI, HindIII	gctagcaggagggaattccatgggtacccggggtatccttag agtcgacctgc aggcgatcaagctt
pBAD18	1995	NheI, EcoRI, SacI, KpnI, SmaI, BamHI, XbaI, Sall, SbfI, SphI, HindIII	gctagcgaattcgagctcgggtacccggggtatccttagagtc gacctgcaggcatgcaagctt
pBAD18s	1995	EcoRI, SacI, KpnI, SmaI, BamHI, XbaI, Sall, PstI, SphI, HindIII	gaattcgagctcgggtacccggggtatccttagagtcgacctg caggcatg caagctt
pBAD28	1995	SacI, Acc65I, KpnI, XmaI, SmaI, XbaI, Sall, SbfI, SphI, HindIII	tcgggtacccggggtatccttagagtcgacctgcaggcatgca agctt
pGEXkg	1988	EcoRI, XbaI, NcoI, Sall, XhoI, SacI, HindIII	gaattctagactccatgggtcgtcgtcagctcaagctt
pKmcs1	1995	NotI, NruI, ClaI, NheI, BmtI, SpeI, XhoI, KsaI, NarI, SfoI, NgoMIV, NaeI, MluI, BsuWI, BglII, HpaI, NsiI	cggccgctcgcgaatcgatagctagcgcgcactagtctcgag gcggccggcacgcgtacgagatctgttaacatgcat
pKmcs2	1995	EcoRI, SacI, Acc65I, KpnI, SmaI, BamHI, XbaI, AccI, PstI, SphI, HindIII	gaattcgagctcgggtacccggggtatccttagagtcgacctg caggcatgcaagctt
pKLAC1	2005	XhoI, 33bps, BglII, AvrII, KpnI, Sall, AscI, NotI, PacI, StuI	ctcgagaaaagagaggctgaagctagaagagctagatctc ctaggggtaccgtcgacggcgcgctcgccgccctta attaaggcct
pKLAC1Glu	2005	XhoI, 33bps, BglII, AvrII, KpnI, AscI, NotI, BlnI	ctcgagaaaagagaggctgaagctagaagagctagatctc ctaggggtaccgtcgacggcgcgctcgccgccct
pKLAC2	2005	XhoI, 33bps, NdeI, NcoI, NotI, EcoRV, Sall, BamHI, EcoRI, SbfI, StuI	ctcgagaaaagagaggctgaagctagaagagctcatatgtc catggcgccgcgcatatcgtcgacggatccgaattccct gcaggtaattaaataaaggcct
pRS303	1989	SacI, SacII, NotI, XbaI, SpeI, BamHI, XmaI, SmaI, EcoRI, EcoRV, HindIII, ClaI, Sall, XhoI, ApaI, KpnI	gagctccaccgcgggtggcgccgctctagaactagtggatcc cccggtgctcagggaattcgatcaagcttatcgataacc gtcgacctcgaggggggcccg
pRX1	1989	EcoRI, SacI, Aval, SamI, BamHI, XbaI, Sall, PstI, HindIII, ClaI	gaattcgagctcgggtacccggggtatccttagagtcgacctgcag cccaagcttatcgat
pS1	1992	EcoRI, SacI, Acc65I, KpnI, NotI, NruI, ClaI, NheI, BmtI, SpeI, XhoI, NaeI, MluI, BsuWI, BglII, HpaI, NsiI, BamHI, XbaI, Sall, AccI, SbfI, PstI, HindIII	gaattcgagctcgggtacccgcggcgccgctcgcgaatcgata gctagcgcgcactagtctcgaggcgccggcacgcgtacg agatctgttaacatgcatgcgagggtatccttagagtcgac ctcgaggcatgcaagctt

Vector	Introduced	Restriction Enzymes used	Sequence
pMAL	1988	SacI, Acc65I, KpnI, EagI, StuI, EcoRI, XbaI, Sall, PstI, HindIII	gagctcgggtacccggccggggatccatcgagggtaggcctg aattcagtaacctaacctcgcgatggatcctctagag tcgacctgcaggcaagctt
pMAL2X	1988	SacI, Aval, XmnI, EcoRI, BamHI, XbaI, Sall, PstI, HindIII	gagctcgaacaacaacaataacaataacaacaacct cgggatcga gggaaggatttcagaattcggatcctctagagtcgacctgca ggcaagctt
pMAL2E	1988	SacI, Aval, KpnI, EcoRI, BamHI, XbaI, Sall, PstI, HindIII	gagctcgaacaacaacaataacaataacaacaacct cggggatga cgatgacaagggtaccggaattcggatcctctagagtcgacct gcaggcaagctt
pMAL2G	1988	SacI, Aval, SnaBI, EcoRI, BamHI, XbaI, Sall, PstI, HindIII	gagctcgaacaacaacaataacaataacaacaacct cgggccgggt gcggcacactacgtagaattcggatcctctagagtcgacctg caggcaagctt
pMALc2X	2010	Eco53kI, SacI, Aval, BmeT110I, XmnI, EcoRI, BamHI, XbaI, Sall, PstI, HindIII	gagctcgaacaacaacaataacaataacaacaacct cgggatcgagggaaggatttcagaattcggatcctctagagtcgacctgcaggcaagct

2.5 Design of cadaverine SynBio construct

The cadaverine genes *cadA* and *cadB* were mined from NCBI Genbank from *E. coli* K12 MC4100 (Genbank accession number HG738867.1) (Figure 3). CadA is a lysine decarboxylase (CDJ74364.1) and CadB is a cadaverine/lysine antiporter (CDJ74365.1). The genes were submitted to GenScript (Piscataway, NJ) for synthesis and cloned into a pBluescript II KS(-) vector (Figure 4). This vector was transformed into *E. coli* DH5a MG1655 RARE strain. The *cadBA* was also liberated from the vector and cloned into the broad host range vector, pBBR1MCS-2 (Addgene plasmid number 85168) using XbaI and HindIII restriction endonucleases. (Figure 5). This vector with the *cadBA* construct was transferred into the *E. coli* RARE strain and *Pseudomonas putida* KTC2440. Lysine and pyridoxal phosphate (PLP) are essential ingredients to produce cadaverine and were added at 50 mM and 50 μ M, respectively, to all media for induction of cadaverine production.

Figure 3. The cadaverine gene sequence for *cadA* and *cadB* from *E. coli* K12 MC4100 mined from NCBI Genbank (accession number HG738867.1).

```

ggatcctcgaacattaacggggtcacggagaaagtctctctatcgcttattgacggctagctcagtcctaggtacagtgtagccgctgtcggttattt
attcgtttgagtttgaacttaagttgaatcgagttaaagacaggacccatgttatgagttctgccaagaagatcgggctatttgcctgtaccggtgtgttgc
cggtaaatatgatggggagcgtgtattgcattatctacgtcgaaacctagcaagatcgggtggtattgctatctgggggttgattatctctatttgggtcaatgtc
gtggcgtgtatgtatgcccactggcaacaaaaacccgcaacaaggtggcccaattgcttatgccggagaaattccctgcatttggtttcagacaggt
gttcttttaccatgtctaaactggattggttaacctggcgattggttaccgtctatcttccacctcttccagttataaatgatcgtgtccggcggtt
atccctgtattgctatcgtctgggtatttaccttgaataatgctcgccggtacttggtaagccgttaaacactattggctgggtcgtgttcttctcgt
gggtgactgctattgttggctggcattggttgcggcaacttatgcagtaactggaatactgcggataccactgatggtcatgcgatcataaaagta
ttctgctctgctgtgggcttctgggtgttgaatccgcagctgtaatctgggtatggttataaacccgaaacgtaccgttccgtggcaacctatgctgggt
actggtttagcaggtattgtttacatcgctgcgactcaggtgcttccggtatgtatccgtcttctgtaatggcggtccggtgctccgtttgcaatcagtgctt
aaactatcctcgttaactgggctgcgcgctggtttctgcattcaccgcttgcgtgctgacttctcgggctcctggatgatgttgtaggcagcagcaggtgt
acgtgccgtaacgcaggttaactcccgaagtattggtgaagtcgacgcaacgggtatccgaaaaaggtcgtgctggctgcagtgaaaatgactg
ccctgatgactcttatcactgtagaacctgcgggtggttaagcatctgacctgttcggtgaactgacgggtatccagtagctgactatgctgcgcgtatt
tctactctgctgtgactgattcgtttgaaggcgttaacatccgcaacttgtcagcctgactgctctgtactgggttgcgttctgcttcatcgctgctgatg
ggcgcaagctccttcagctggcaggtaccttcacgtcagcctgatctcgtgattctacgtcgcataaatgcacgagcgcagagccactcaatggat
aaaccaaccgctcgaacacatttaataaaagtatttccgaggtcctccttcttcttctccatgtgttgggaggggcttcttctacgtgagatagact
atgaacgttattgcaatattgaatcactgggggttattttaaagaagaacccatccgtgaactcatcgcgcttgaacgtctgaacttcagattgtttac
ccgaacgacctgacgacttattaaaactgatcgaaaacatgcgcgtctgtgcggcttattttgactgggataataataatctcagctgtgcgaagaa
attagcaaaatgaacgagaacctccgttgcagcgttcgtaatacgtattccactctcgtatgaagcctgaatgacctgcgtttacagattagcttcttga
atatgcctgggtgctgtcgaagattgctataaagatcaagcagaccctgacgaatatcaacactattctgcctccgtgactaaagcactgtttaaa
tatgttcgtgaaggtaaatatactttctgactcctggtcacatgggctgactgcattccagaaaagcccgtaggtagcctgttctatgatttcttggctcg
aataccatgaaatctgataattccatttcagtatcgaaatgggttctcgtcggatcacagtgtccacacaagaagcagaacagtatacgtcgcgtctt
taacgcagaccgcagctacatggtgaccaacggtacttccactgcgaacaaaattgttggtatgactctgctccagcaggcagcaccattctgattgacctg
aactgccacaaatcgtgacccacctgatgatgatgagcgtatgacgcaatctattccgcccagaccgttaacgcttacgggtattcttggtggtatccaca
gagtgaattccagcacgtaccattgtcgaagcgtgaaagaaacacaaacgcaacctggccgtacatgctgtaattaccaactctacatgatggtct
gctgtacaacaccgacttcatcaagaaaacactggatgtgaaatccatccacttgcactcgcgtgggtgcttacaccaacttctaccgatttacgaaggt
aaatgcggtatgagcgtggcgtgtagaagggaagtgatttacgaaacccagtccactcacaactgctggcggcgcttctcaggctccatgatccac
gttaaaggtagctaaacgaagaacctttaaagagcctacatgatgcacaccaccacttctcgcactacggtatcgtggcgttcaactgaaacgctgc
ggcgatgatgaaggcaatgcaggtgaagcgtctgatcaacggtctattgaacgtgcgatcaaatccgtaaagagatcaaacgtctgagaacggaatctg
atggctgggtctttgatgtatggcagcggatcatatcgatacgaactgaatgctggcgtcgttctgacagcacctggcagcgcttcaaaaacatcgataa
cgagcacatgtatcttgacccgatcaaaatcaccctgctgactccggggtatggaaaaagacggcaccatgagcgttgggtattccggccagcatcgtgg
cgaaatacctcgacgaacatggcatcgttgttgagaaaacgggtccgtataacctgctgttccgttcagcatcggtatcgataagaccaaacgactgagcc
tgctgcgtgctctgactgactttaaacgtgcgttcgacctgaacctgcgtgtgaaaaacatgctcgcgtctctgtatcgtgaagatcctgaatctatgaaaac
atgcgtattcaggaactggctcagaatatccacaaactgattgttaccacaactctccggatctgatgtatcgcgcaattggaagtgcgtccgacagatggttaa
tgactccgtatgctgcatccagaaagcgtgcacgggtatgacgaagaagttacctcgacgaatggtaggtcgtattaacgccaatgatccttccgtta
cccgcgggagttctctggtaatgcgggtgaaatgatccgaagaagccgtccggttctggagttctgcagatgctgtgaaatcgcgctcacta
tccgggttgaaccgatattcacgggtgcataccgtcaggtgatggcgctataccgttaaggtattgaaagaagaaacaaaaataatagctcgttg
ttcagaacgctcgtcttgacacccggcggttttcttctgtgagtcaggatcc

```

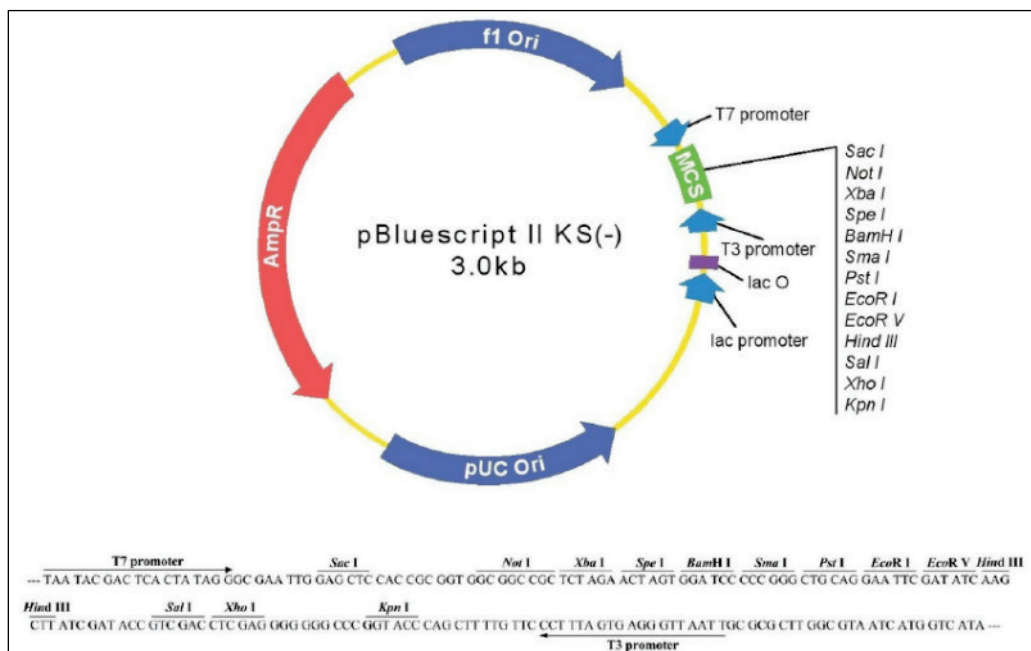
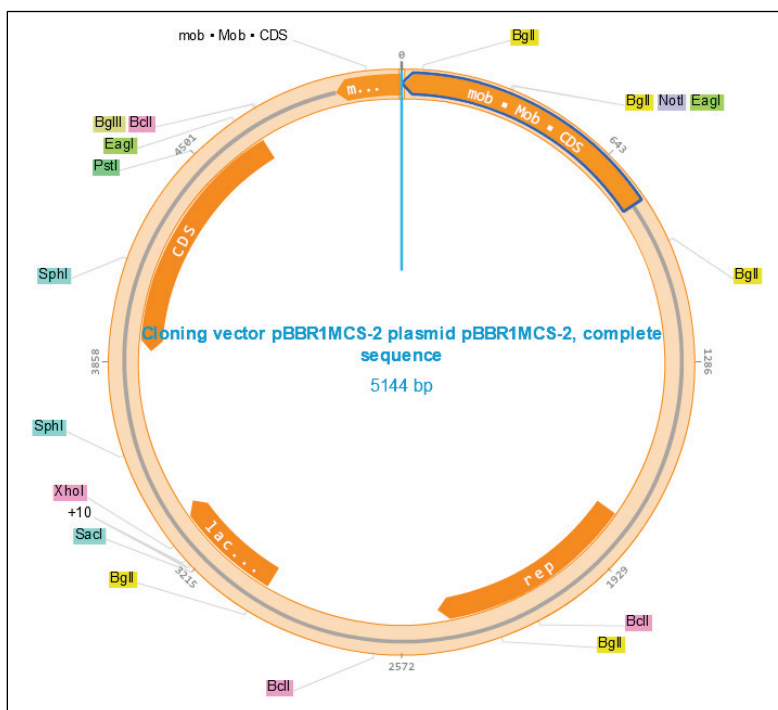
Figure 4. Plasmid map of *cadBA* construct from GenScript.

Figure 5. Broad host range plasmid pBBR1MCS-2 map.

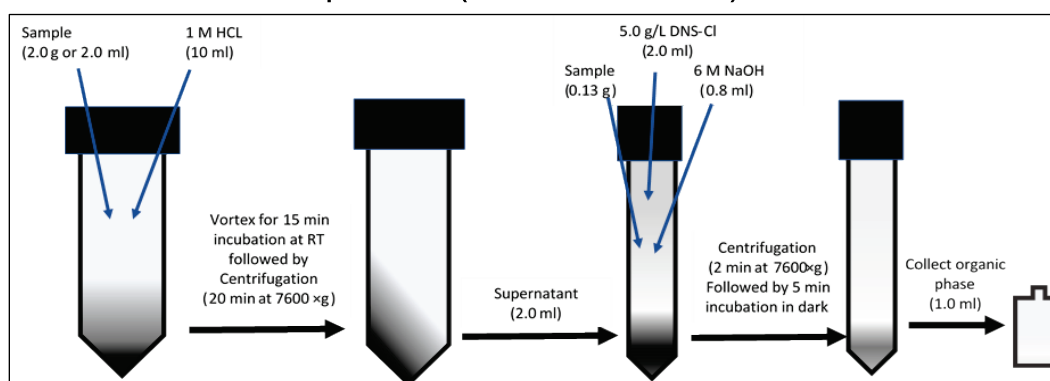


2.6 Chemical analysis of cadaverine

Cadaverine can be separated via high performance liquid chromatography (HPLC) but is not amenable to UV-visible light detection without modification. Therefore, derivatization with the chromophore, dansyl

chloride, was performed following a rapid, dual method of extraction and derivatization called the SALLE method (Figure 6) (Francisco et al. 2020). Briefly, 2 ml sample was added to 10 ml of 1 M HCl followed by centrifugation at $7,750 \times g$ for 20 min. Two ml of supernatant was added to 0.13 g NaCl, 2 ml dansyl chloride (2.5 g L^{-1} in water), and 0.8 ml of 6M NaOH and the mixture was vigorously vortexed and then allowed to rest for 5 min before being centrifuged again for 1 min at $7,750 \times g$. The organic phase (on top) was then filtered with a 0.45 syringe filter and added to an HPLC vial for analysis. The analysis was done on a 1260 Infinity Agilent HPLC using a Zorbax SB-C18, $5 \mu\text{m}$, $4.6 \times 150 \text{ mm}$ column (Agilent) at 30°C . The mobile phase consisted of 75:25 0.01% ammonium acetate (pH 4):acetonitrile which ramped to 100% acetonitrile over 10 min and was held for 2 min before switching back to 75:25 over another 3 min and being held for 3 min before the next injection. Twenty-five μl of sample was injected and the flow rate was kept constant at 1 ml min^{-1} . The cadaverine peak resolved at 10.3 min using a diode array detector set at 254 nm. Standards were made using known concentrations of pure cadaverine and derivatizing them using the same method as for samples.

Figure 6. Pictorial representation of the SALLE cadaverine extraction/derivatization procedure (Francisco et al. 2020).



2.7 PCR analysis of cadaverine

Samples of each microcosm were taken at days 0, 14, 60, and 120. DNA was extracted from each 0.2 g or 0.2 ml sample using a Qiagen DNeasy PowerSoil kit. Resultant DNA was subjected to PCR using primer(s) that annealed to portions of the *cadBA* operon and the flanking plasmid vector in which the *cadBA* genes were cloned. This strategy eliminated the amplification of native *cadBA* from the matrices. Two primer sets for the *cadBA* gene were developed to accommodate the reverse orientation of *cadBA* into the cloning site. The use of CadBA1864R (targeting the 3' end

of the *cadBA* SC) and M13R primers (targeting the vectors) (Table 3) produced a fragment approximately 2 kb when samples contained either *cadBA*/pBBR1MCS2 or *cadBA*/pBluescript. The PCR parameters used with DreamTaq DNA Polymerase master mix (Thermo Fisher Scientific) included 32 cycles of 95°C denaturation for 30 sec, 56°C annealing for 30 sec, and 72°C elongation for 90 sec. Each sample was PCR amplified in triplicate and PCR products were resolved on 1% agarose gel and noted as being present or absent. At least two of the three samples needed to be positive to validate the presence of the target amplicon.

Table 3. Primer names and sequences used for PCR validation of SynBio constructs.

	Primer sequences (5'-3')	Primer combinations Product size (bp)
1. M13 F (-47)	CGCCAGGGTTTTCCCAGTCAGAC	1. - 5. = 3189
2. M13 R (-48)	AGCGGATAACAATTCACACAGGA	1. - 2. = 3957
3. CadBA1949F	CAACACTATTCTGCCTCCGC	1. - 3. = 1917
4. CadBA3788R	GACGGCTTTCTTCGGTGATC	3. - 4. = 1637
5. CadBA660F	ATTGGTTTGATGCGGCAACT	5. - 6. = 1205
6. CadBA1864R	CGCAGGTCATTCAGGCTTAC	2. - 6. = 1955

2.8 Background cadaverine production by SynBio strains

Strains of *E. coli* MG1655 RARE containing the *cadBA* SynBio construct in pBluescript and in pBBR1MCS2, and *P. putida* KTC2440 containing *cadBA* in pBBR1MCS2 were cultivated in LB with 0.1 mM IPTG and 50 µM PLP and either 50 or 100 µM lysine. Strains were inoculated with an overnight culture at a starting OD of 0.02 into 50 ml volume in 250 ml flasks at 37°C which were placed on a flat-bed shaker at 100 RPM. Cultures were sampled for the presence of cadaverine at times 0, 24, 48, and 72 hr by using 2 ml of the sample for the SALLE extraction and derivatization method and analyzing with the HPLC.

2.9 Microcosms to assess the fate of SynBio constructs

Microcosms representing different environmental matrices (soil, water, wastewater treatment plant (WWTP) liquor) were established to assess the fate of SynBio constructs (SC) in different environmental matrices. Soil was collected from Calhoun, LA and characterized as silt loam (Table 4). Soil samples were split in half with one half retaining its natural microbiome and the other half undergoing autoclave sterilization. Four

soil groups were created using sterile and non-sterile soil: soil mixed with 15 ml of sterile water per 25 g soil was denoted as saturated L soil (L soil / L soil sterile) while soil mixed with 25 ml of sterile water per 25 g of soil was denoted as hyper-saturated (slurry) soil group (H soil / H soil Sterile). M9 minimal media was used as the sterile “water” matrix and prepared with 0.01% yeast extract and 0.01% glucose. WWTP liquor was split into two groups: non-sterile and sterile (via autoclave). Each microcosm received the essential precursors for expression and production of cadaverine, 50 μ M lysine and 50 μ M PLP. The sampling times were at days 0, 14, 60, and 120 and consisted of harvesting either 2 ml or 2 g of each sample matrix for cadaverine detection and 0.25 g or 0.25 ml for PCR detection of the *cadBA* construct. The experimental matrix setup is detailed in Table 5.

Naked DNA inoculation parameters for microcosms were then determined via MiSeq metagenomic sequencing through titration of known spiked DNA concentrations of pK18 into the pristine Ft. Huachuca Seep Pool DNA. Results indicated a 0.1% (0.3 nM Final Concentration) target spike would be an ideal concentration as it is easily resolved via sequencing without overwhelming the system. The final concentration for each target construct was determined using the formula [$\text{pM } \mu\text{l}^{-1} = (\text{DNA } \mu\text{g ml}^{-1}) / 0.66 \times \text{DNA size in bp}$] and diluting accordingly.

Copy number of plasmids per cell were estimated by first calculating the total copy number in 0.3 nM using the following equation: $0.025 \text{ L} \times 0.3 \text{ nmol L}^{-1} = 7.5 \times 10^{-12} \text{ mol}$. Second, using Avogadro’s number, the molecules per mol were obtained as $7.5 \times 10^{-12} \text{ mol} \times 6.02 \times 10^{23} \text{ molecules/mol} = 4.52 \times 10^{12} \text{ molecules (or cells) per matrix}$. Finally, extrapolating the plasmid pBBR1MCS2 at ~10,000 and the plasmid pBluescript at ~30,000 copies per cell reduced final cell numbers 10^4 –fold resulting in bacteria cultures containing plasmid DNA to an average of $4.52 \times 10^8 \text{ cells per } 25 \text{ g ml}^{-1} \text{ of matrix}$ to achieve 0.3 nM target SynBio DNA concentration. Since *cadBA*/pBBR1MCS2 is of larger size and lower copy number per cell when compared to *cadBA*/pBluescript, cells were added to microcosms at a 3:1 concentration: 0.5 ml of $\text{OD}_{600} = 4.5165$ of *P. putida::cadBA*/pBBR1MCS2 versus 0.5 ml of $\text{OD}_{600} = 0.6022$ *E.coli::cadBA*/pBluescript to achieve comparable concentrations of SynBio DNA for each matrix.

Table 4. Characteristics of the soil used in the microcosm experiments.

	Pasture (Calhoun 2)
Coordinates	32.086, -91.672
Soil Name	Loring Silt Loam
Taxonomic Classification	Fine-silty, mixed, active, thermic Oxyaquic Fragiudalfs
Soil Order	Alfisol
Slope	1 to 3 %
Elevation	50 to 110 feet
Mean Annual Precipitation	47 to 65 inches
Soil Moisture Regime	Udic
Mean Annual Air Temperature	54 to 77 °F
Soil Temperature Regime	Thermic
Landform	Stream Terrace
Parent Material	Loess
Drainage Class	Moderately well drained

Table 5. Experimental matrix for the microcosm setup including the number of microcosms for each condition.

Condition	SynBio Chassis					Total #
	pBBR1 DNA	pBluescript DNA	<i>P. putida</i> + pBBR1	<i>E. coli</i> + pBluescript	No SynBio	
WWTP	3	3	3	3	3	15
- sterilized	3	3	3	3	3	15
M9 - sterile	3	3	3	3	3	15
H soil	3	3	3	3	3	15
- sterilized	3	3	3	3	3	15
L soil	3	3	3	3	3	15
- sterilized	3	3	3	3	3	15

Figure 7. Microcosm setup with A) individual microcosms and B) all the microcosms in a 30°C incubator.



3 Results and Discussion

3.1 Pristine sample screening and spiking for SynBio MCS signatures

Pristine samples had almost no hits for known MCSs, and the hits were of bad quality (contained mismatches or gaps). These samples appeared to be pristine as was suspected. Less than 0.0001% of the total sequences contained putative MCS matches. The Ft. Huachuca seep pool sample resulted in 6,277,952 total sequence fragments containing two putative MCS hits (0.000032% of total DNA) while the White Sands guzzler with rattlesnake sample resulted in 1,298,882 total sequence fragments containing one putative MCS hit (0.000077% of total DNA). MCS matches contained some versions of these restriction sites in order: ‘EcoRI, SacI, Acc65I, KpnI, SmaI, BamHI, XbaI, SalI’. These putative MCS hits belonged to the following plasmid/vectors: pMAL2X, pMAL2G, pMAL2E, pUC19, pUC18, pMMB207, pKmc2, pBAD18s, pBAD18, pS1, pUC13, pUC12, pRX1, pBAD24, pBAD28.

The Ft. Huachuca seep pool sample was used for spiking with known concentrations of pK18 to determine recovery and detection limits of MCS-SynBio from DNA samples (Table 6). Overall, the recovery of MCS (consisting of the sequence – gaattcgagctcggtaccggggtatcctcagagtcgacctgcaggcatgca) was predictable as the concentrations spiked into the samples were roughly what was identified through sequencing and quantitative alignment with the MCS library.

Table 6. MCS-Spiked recovery from pristine Ft. Huachuca seep pool samples. Sample designations indicate relative concentrations of added SynBio.

SAMPLE	Target Spike %	Spiked Molarity	# Matching Reads per Sample	% Matching Reads of Total
High	50.0	0.15151 uM	13,594	33.69
Mid	0.5	1.5151 nM	3,309	8.20
Mid Low	0.01	30.303 pM	8	0.02
Low	0.001	3.0303 pM	0	0.00
Very Low	0.0001	0.30303 pM	0	0.00
High Water	50.0	0.15151_uM	21,921	54.33
Very Low Water	0.0001	0.30303 pM	238	0.59
Water	0.0	0.0 M	708	1.75

3.2 Background cadaverine production by SynBio strains

The strain of *E. coli* MG1655 RARE containing the *cadBA* SynBio construct in pBluescript produced approximately three times as much cadaverine as the strain with *cadBA* in pBBR1MCS2 (Figure 8). The concentration of lysine precursor was tested at 50 and 100 μ M and found to yield similar cadaverine concentrations while producing 10 ppm when lysine was not added. *E. coli* MG1655 RARE without the *cadBA* construct produced a maximum background concentration of 2 ppm without added lysine and around 30 ppm with 50 and 100 μ M lysine. *P. putida* KTC2440 containing the *cadBA* SynBio construct in pBBR1MCS2 with 50 and 100 μ M of lysine precursor was found to yield similar cadaverine concentrations while producing a maximum of 4.8 ppm when lysine was not added (Figure 9). *P. putida* KTC2440 without the *cadBA* construct produced a maximum background concentration of 0.9 ppm without added lysine and around 4 ppm with 50 and 100 μ M lysine. Based on the similar results from both bacteria at both concentrations of lysine in it was determined that 50 μ M would be added to the microcosms.

Figure 8. *E. coli* MG1655 RARE production of cadaverine. Strains containing *cadBA* in pBluescript are shown in grey with open circles while those containing *cadBA* in pBBR1MCS2 are shown in black with open squares. The dashed lines for the *cadBA* containing strains represent the growth medium with 50 μ M Lysine and the solid lines represent 100 μ M. All other conditions produced negligible amounts of cadaverine.

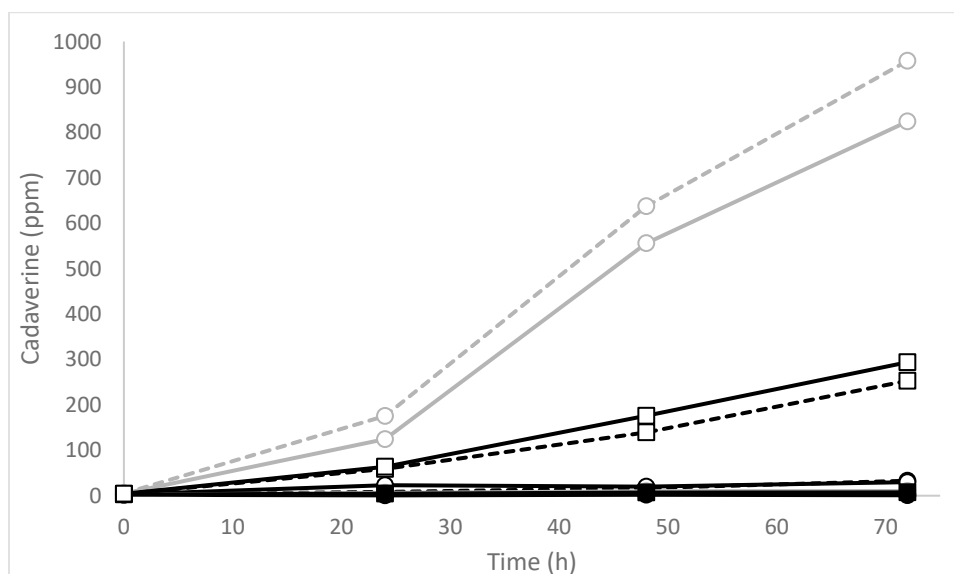
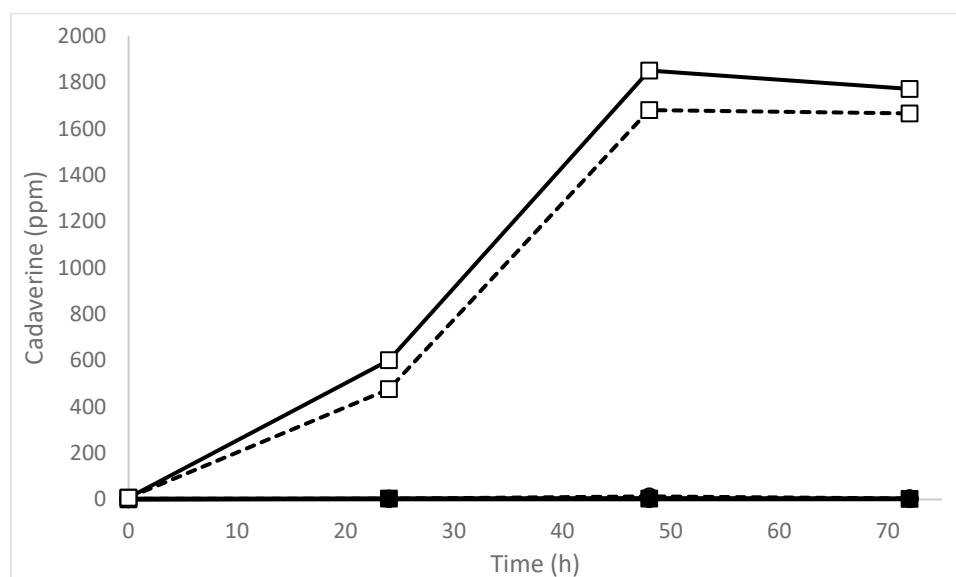


Figure 9. *P. putida* KTC2440 production of cadaverine. Strains containing *cadBA* in pBBR1MCS2 are shown in black with open squares. The dashed lines for the *cadBA* containing strains represent the growth medium with 50 μ M Lysine and the solid lines represent 100 μ M. All other conditions produced negligible amounts of cadaverine.



3.3 Microcosms

Persistence and proliferation of the SC in the matrices over time were determined by PCR and chemical analysis of the functional production of cadaverine via HPLC. The SC persisted throughout microcosms containing either naked DNA or bacteria-integrated plasmids, which suggests that SC released into the environment may persist over time and leave diagnostic signatures. Detectability decreased when ratios of water:soil increased while wastewater liquor demonstrated a proliferative environment for SC persistence.

The overall trend observed was that cadaverine was generated by at least the *P. putida::cadBA* clone after 14 days in all matrices. The *E. coli::cadBA* clone appeared to produce cadaverine at levels above background within each matrix except the WWTP matrices, and at generally lower levels than *P. putida*. The control culture M9 microcosms supported cadaverine production when inoculated with both clones but not the naked DNA, as anticipated. Hyper-saturated soil was the only environmental matrix still supporting high level production of cadaverine 60 days post-inoculation. Interestingly the non-inoculated microcosms also produced cadaverine, so it appeared the natural population of bacteria containing lysine

decarboxylase (cadaverine) has been enriched for via the addition of excess lysine and PLP.

Results show that amplification of target *cadBA*/pBluescript in DNA samples was difficult to detect when compared to samples containing *cadBA*/pBBR1MCS2 (Table 7). Tandem PCR was required to detect *cadBA*/pBluescript; however, detection rates were low and confined to WWTP and L soil samples.

Cadaverine was detected throughout the 120-day experiment in the M9 control matrix when spiked with either *P. putida* containing pBBR1MCS2::*cadBA* or *E. coli* containing pBluescript::*cadBA* (Figure 10). There was a lot of variability in the *P. putida* cultures but it produced approximately five times more cadaverine than the *E. coli* cultures. Negligible (less than 1 ppm) or non-detectable amounts of cadaverine were produced in cultures containing naked SynBio or in the no-SynBio controls. Both the “live” and sterile WWTP liquor samples showed a background level of cadaverine in most microcosms throughout the entire 120 days. There were increases in cadaverine in both the WWTP matrices at day 14 from the *P. putida* cultures and at day 120 from the *E. coli* culture. The *E. coli* day 120 triplicate samples were variable in the concentrations of cadaverine and were fully reprocessed and again found to have large and variable amounts of cadaverine averaging 55 ppm.

Interestingly, the “live” L soil (saturated) matrix exhibited high levels of cadaverine production (>500 ppm) in all microcosms, including the no-SynBio control, indicating this is baseline cadaverine production in this soil. However, when the soil was sterilized by autoclaving the baseline, cadaverine production was negligible and only seen at high levels in the *P. putida* (>500 ppm) and *E. coli* (>200 ppm) cultures at 14 days and tapering off over time. The same basic trends are seen with the H soil (hyper saturated) samples but over longer times (day 14 - day 60).

Overall, the PCR results (Table 7) are confusing when compared to cadaverine production from each sample. Since the PCR primers were targeting the expression vector on one priming site and the cadaverine gene on the other site, it seems unlikely there would be false positives. However, in three instances, the no-SynBio control microcosms exhibited positive PCR results (in WWTP day 14 and H soil days 14 and 120, but not day 60). It was also interesting to note that the naked DNA added to the

different matrices did not yield many positive PCR hits and there was no consistency to those hits. When the SynBio construct was delivered to microcosms of the various matrices, each sample showed a positive PCR reaction at day 14, as expected, and most were positive on day 60. The *P. putida* cultures in the sterile WWTP and sterile L soil matrices were the only samples that exhibited a positive PCR result at day 120. Although there was a small amount of detectable cadaverine at day 120 in the sterile WWTP with *P. putida*, it would be more likely to see a positive result from the *E. coli* culture producing around 50 ppm, but this was negative. Similarly, both the *P. putida*- and *E. coli*-containing H soil samples would be expected to have positive PCR results at day 120 but only the *P. putida* samples were positive.

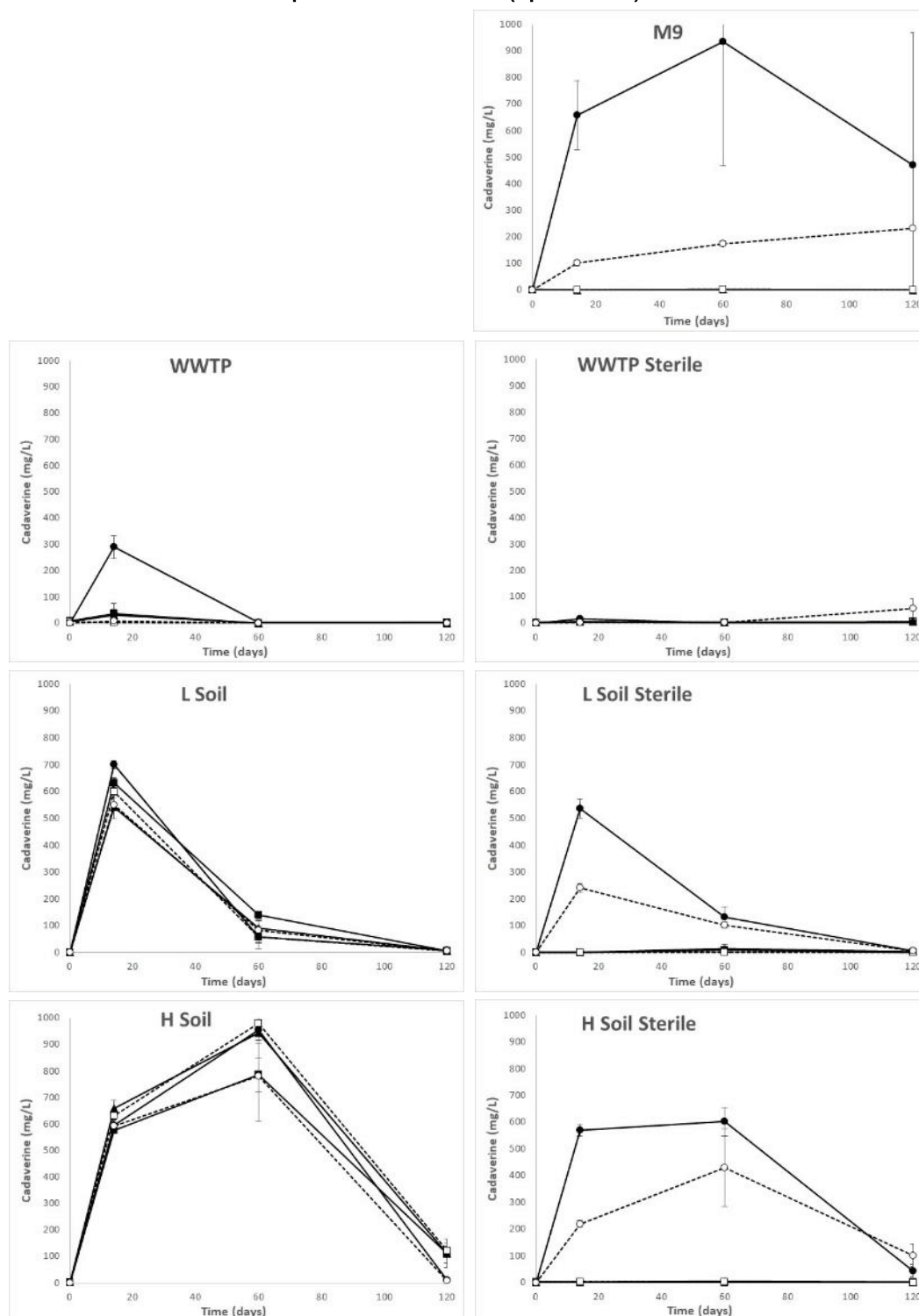
The positive PCR reactions from day 120 of the “live” WWTP sample containing naked *cadBA*::pBBR1MCS2 and the sterile WWTP sample containing naked *cadBA*::pBluescript may indicate these constructs were taken up by other bacteria or they were randomly amplified at this time point. Unfortunately, these results collectively are unreliable, and a more stringent primer set, or separate assay mechanism would be needed to address this issue. To gain greater sensitivity in PCR-based assays, quantitative PCR (qPCR) could have been used and may have offered better insight to the concentrations of amplified product in each sample rather than just a positive or negative result.

Overall, the addition of the naked DNA in the form of intact plasmids does not convey any functional activity on its own or through uptake by native bacteria present in the soils or WWTP samples. Furthermore, by incorporating the plasmid/SynBio into cells, it may protect the DNA (increasing half-life) and allow for production of more SynBio through plasmid replication during cell growth. Our hypothesis that SC in a host would have a greater half-life was supported by the results. Unfortunately, the SC was never incorporated into the genome of a host, and we had to instead look at a different host/plasmid system. *Pseudomonas* is a hardier environmental organism than a laboratory trained *E. coli*, so it was not surprising that the SC was detected in WWTP and soil samples containing *Pseudomonas*. Another part of the hypothesis was that samples with less interferences (i.e., sterile medium vs. live soil or WWTP liquor) would promote longer half-lives of the SC but there did not seem to be a clear trend via PCR detection.

Table 7. PCR results from microcosms. Positive results from at least two of the three replicates are represented as X = day 14, Y = day 60, and Z = day 120.

Matrix	No SynBio	Naked <i>cadBA</i> :: pBBR1MCS2	Naked <i>cadBA</i> :: pBluescript	<i>P. putida</i> :: <i>cadBA</i> :: pBBR1MCS2	<i>E. coli</i> <i>cadBA</i> :: pBluescript
M9		X		X Y	X Y
WWTP	X	X Z		X Y	X Y
WWTP- Sterile			Z	X Y Z	X Y
L Soil		X		X Y	X
L Soil- Sterile				X Y Z	X Y
H Soil	X Z			X	X
H Soil- Sterile				X Y	X Y

Figure 10. Cadaverine concentrations as monitored by HPLC from samples, in triplicate, from each matrix. The control containing no SynBio construct (triangle), naked pBBR1MC2::*cadBA* (closed square), naked pBluescript::*cadBA* (open square), *P. putida* containing pBluescript::*cadBA* (closed circle), *E. coli* containing pBBR1MC2::*cadBA* (open circle).



4 Conclusions and Recommendations

Currently, the Army has identified biotechnology as a key research priority with future Army applications in sensors, electronics and computers, materials, logistics, and medical therapeutics. In the context of sensors, the Army envisions a future battlespace environment in which genetically reprogramed bacteria have been released into the environment in some limited capacity (perhaps as part of an enclosed engineered system) and can sense and report on activities in the environment. A key aspect of realizing this concept is gathering fundamental information regarding the fate and transport of synthetic biology constructs in the environment which will be critical for the safe and responsible field deployment of these materials. Furthermore, the methodology, databases, and algorithms developed to track the fate and transport of synthetic constructs will facilitate detection of future engineered threats (rogue SynBio) as well as provide critical intelligence on adversarial capabilities.

4.1 Conclusions

Results from this research effort represent an attempt to begin developing and standardizing an approach to assess the fate and transport of a representative synthetic biology construct presented in some typical genetic configurations in different bacteria chassis host organisms. In addition, a computational approach and database was developed to screen environmental samples for evidence of legacy genetically engineered materials in the environment using MCS sequences. In conclusion, we were able to demonstrate (1) a synthetic construct located on a plasmid that is maintained in a bacterial chassis host has a longer half-life than a naked plasmid alone when spiked in various environmental matrices, (2) a synthetic construct propagating in a soil bacterial chassis host (*Pseudomonas*) has a longer half-life than that same construct propagating in a lab strain of *E. coli*, and (3) screening for MCS sequences in shotgun metagenomic sequence represents a convenient biomarker signature for detection of legacy genetic material

4.2 Recommendations

Due to the circumstances (COVID-19) in which this two-year project was executed, there are still many unanswered questions and needs. Additional studies are needed to validate the usefulness of MCS signatures in

screening for legacy genetic material in the environment. Shotgun metagenomic sequences from non-pristine sites are needed for analysis as suggested in the original proposal. Samples from known sites where recombinant DNA technologies were routinely used would be advantageous.

Additional studies are also needed to further explore the role that chassis host and genetic configuration impact the persistence of a synthetic construct in the environment. A standardized approach is needed to explore the genetic bioavailability of synthetic constructs in the context of different synthetic biology cloning and expression strategies. Our results indicate that a synthetic construct present on a plasmid propagating inside of a bacterial host is more resilient to degradation and therefore persist longer in a microcosm study. But what if a synthetic construct was integrated inside the chassis host's genome? Would this represent a configuration even more resistant to degradation and persist even longer in the environment?

References

- Ahteensuu, M. 2017. "Synthetic biology, genome editing, and the risk of bioterrorism." *Sci Eng Ethics* 23 (6): 1541-1561. <https://doi.org/10.1007/s11948-016-9868-9>.
- Chilton, M. D., D. A. Tepfer, A. Petit, C. David, F. Casse-Delbart, and J. Tempé. 1982. " *Agrobacterium rhizogenes* inserts T-DNA into the genomes of the host plant root cells." *Nature* 295 (5848): 432.
- Crisp, A., C. Boschetti, M. Perry, A. Tunnacliffe, and G. Micklem. 2015. " Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes." *Genome Biol.* 16 (1): 50.
- Epstein, M. M., and T. Vermeire. 2016. "Scientific opinion on risk assessment of synthetic biology." *Trends Biotechnol* 34 (8): 601-603. <http://www.sciencedirect.com/science/article/pii/S0167779916300324>.
- Francisco, K. C. A., P. F. Brand, R. M. Ramos, R. M. Goncalves, A. A. Cardoso, and J. A. Rodrigues. 2020. "Salting-out assisted liquid–liquid extraction with dansyl chloride for the determination of biogenic amines in food." *Int J Food Sci Technol* 55: 248-258.
- Hewett, J. P., A. K. Wolfe, R. A. Bergmann, S. C. Stelling, and K. L. Davis. 2016. "Human health and environmental risks posed by synthetic biology R&D for energy applications: a Literature analysis." *Appl Biosafety* 21 (4): 177-184. <https://journals.sagepub.com/doi/abs/10.1177/1535676016672377>.
- Johnston, C., B. Martin, G. Fichant, P. Polard, and J-P. Claverys. 2014. "Bacterial transformation: distribution, shared mechanisms and divergent control." *Nature Rev Microbiol* 12: 181. <https://doi.org/10.1038/nrmicro3199>.
- Jung, C. M., T. M. Heinze, R. Strakosha, C. A. Elkins, and J. B. Sutherland. 2009. "Acetylation of fluoroquinolone antimicrobial agents by an *Escherichia coli* strain isolated from a municipal wastewater treatment plant." *J Appl Microbiol* 106 (2): 564-571.
- Liu, T., R. Mazmouz, S. E. Ongley, R. Chau, R. Pickford, J. N. Woodhouse, and B. A. Neilan. 2017. "Directing the heterologous production of specific cyanobacterial toxin variants." *ACS Chem Biol* 12 (8): 2021-2029. <https://doi.org/10.1021/acscchembio.7b00181>.
- Nielsen, K. M., P. J. Johnsen, D. Bensasson, and D. Daffonchio. 2007. "Release and persistence of extracellular DNA in the environment." *Environ. Biosafety Res.* 6 (1-2): 37-53. <https://doi.org/10.1051/ebr:2007031>.
- Overballe-Petersen, S., K. Harms, L. A. Orlando, J. V. M. Mayar, S. Rasmussen, T. W. Dahl, M.T. Rosing, A. M. Poole, T. Sicheritz-Ponten, and S. Brunak. 2013. "Bacterial natural transformation by highly fragmented and damaged DNA." *Proc Natl Acad Sci* 110: 19860-19865.

- Parsley, L. C., E. J. Consuegra, K. S. Kakirde, A. M. Land, Jr. W. F. Harper, and M. R. Liles. 2010. "Identification of diverse antimicrobial resistance determinants carried of bacterial, plasmid, or viaral metagenomes from an activated sludge microbial assemblage." *Appl Environ Microbiol* 76 (11): 3753-3757.
- Patel, S. 2016. "Drivers of bacterial genomes plasticity and roles they play in pathogen virulence, persistence and drug resistance." *Infect Genet Evol* 45: 151-164.
- Phillips, S., D. Dalgarn, and S. Young. 1989. "Recombinant DNA in wastewater: PBR322 degradation kinetics." *J Water Pollut Control Fed* 61: 1588-1595.
- Romanowski, G., M. G. Lorenz, G. Sayler, and W. Wackernagel. 1992. "Persistence of Free Plasmid DNA in Soil Monitored by Various Methods, Including a Transformation Assay." *Appl Environ Microbiol* 58 (9): 3012-3019.
- Romanowski, G., M. G. Lorenz, and W. Wackernagel. 1993. "Use of polymerase chain reaction and electroporation of Escherichia coli to monitor the persistence of extracellular plasmid DNA introduced into natural soils." *Appl Environ Microbiol* 59 (10): 3438-3446.
- Tappeser, B., M. Jager, and C. E. Eckelkamp. 1998. "Survival, Persistence, Transfer - An Update on Current Knowledge on GMOs and the Fate of their Recombinant DNA." <http://userwww.sfsu.edu/rone/GEessays/Survival%20PersistenceTransfer.htm>.
- Torgersen, H. 2009. "Synthetic biology in society: learning from past experience?" *Systems and Synthetic Biology* 3 (1): 9. <https://doi.org/10.1007/s11693-009-9030-y>.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE (DD-MM-YYYY) September 2022		2. REPORT TYPE Final		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE Properties and Detectability of Rogue Synthetic Biology (SynBio) Products in Complex Matrices				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER 0601102A/AB02	
6. AUTHOR(S) Carina M. Jung, Karl J. Indest, Matthew R. Carr, Richard F. Lance, Lyndsay A. Carrigee, and Kayla N. Clark				5d. PROJECT NUMBER 484234	
				5e. TASK NUMBER 04	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Environmental Laboratory US Army Engineer Research and Development Center 3909 Halls Ferry Road, Vicksburg, MS 39180-6199; Oak Ridge Institute for Science and Education P.O. Box 117, Oak Ridge, TN 37831-011				8. PERFORMING ORGANIZATION REPORT NUMBER ERDC/EL TR-22-11	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) US Army Corps of Engineers Washington, DC 20314-1000				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Synthetic biology (SynBio) aims to rationally engineer or modify traits of an organism or integrate the behaviors of multiple organisms into a singular functional organism through advanced genetic engineering techniques. One objective of this research was to determine the environmental persistence of engineered DNA in the environment. To accomplish this goal, the environmental persistence of legacy engineered DNA building blocks were targeted that laid the foundation for SynBio product development and application giving rise to "post-use products." These building blocks include genetic constructs such as cloning and expression vectors, promoter/terminator elements, selectable markers, reporter genes, and multi-cloning sites. Shotgun sequencing of total DNA from water samples of pristine sites was performed and resultant sequence data mined for frequency of legacy recombinant DNA signatures. Another objective was to understand the fate of a standardized contemporary synthetic genetic construct (SC) in the context of various chassis systems/genetic configurations representing different degrees of "genetic bioavailability" to the environmental landscape. These studies were carried out using microcosms representing different environmental matrices (soils, waters, wastewater treatment plant (WWTP) liquor) and employed a novel genetic reporter system based on volatile organic compounds (VOC) detection to assess proliferation and persistence of the SC in the matrix over time.					
15. SUBJECT TERMS Synthetic biology – Environmental aspects Genetic engineering Recombinant DNA – Detection Recombinant DNA – Bioavailability					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (include area code)

