High Throughput Quantification of the Functional Genes Associated with RDX Degradation Using the SmartChip Platform

Jennifer M. Collier (collie96@egr.msu.edu), Benli Chai, James R. Cole, and Alison M. Cupples (Michigan State University, East Lansing, MI, USA)

ABSTRACT: Many military sites have been contaminated with the explosive hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX). One viable remediation method is biological degradation. While many microorganisms have been linked to RDX degradation, their occurrence across contaminated sites is still largely unknown. PCR has been used to detect the associated functional genes, but published primers do not fully capture the known genetic diversity of these genes. The goal of this study was to design a suite of assays capable of capturing this diversity using the Michigan State University Ribosomal Database Project's (RDP) EcoFunPrimer tool. Sequences of *xplA, xenA, xenB, diaA, nfsI,* and *pnrB* genes were collected from the NCBI non-redundant protein database using RDP's Functional Gene Pipeline & Repository (FunGene). The newly designed assays have greatly increased the coverage for the target genes. The *xplA*, *xenA* and *xenB* primer sets have been checked with a plasmid dilution series using traditional qPCR, and the limit of detection for most was 10^2 or 10^1 copies per reaction. The *xenA* primer sets have also been tested with a plasmid dilution series using the SmartChip platform, and the detection limit was typically either 10^4 or 10^3 copies per reaction. Ongoing research involves the development and testing of all primer sets on the SmartChip system.

INTRODUCTION

The explosive hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) is a contaminant at many military sites across the US (Clausen et al., 2004). Since RDX has a low soil sorption coefficient, it is mobile through soil and may contaminant groundwater. This is a concern for human health as RDX is classified as a possible human carcinogen by US EPA with a lifetime health advisory guidance level set at 2 µg/l for drinking water (Cupples, 2013).

Biological degradation of RDX is a viable remediation method for these sites. Degradation may occur either aerobically or anaerobically, and a number of bacteria have been linked to RDX degradation, including *Rhodococcus*, *Gordonia*, *Williamsia*, *Pseudomonas*, *Clostridium*, and *Enterobacter* (Li et al., 2014). Six functional genes associated with these bacteria have been linked to RDX biodegradation: *xplA*, *xenA*, *xenB*, *diaA*, *nfsI*, and *pnrB* (Lee et al., 2012; Li et al., 2014). Primer sets have previously been designed to target these genes, however, they do not have good coverage to capture the current known genetic diversity.

This study aimed to design a suite of assays for each functional gene to ensure a high coverage of all gene sequences currently available. The assays were designed to work efficiently with the SmartChip platform, which is a massive-parallel singleplex PCR platform. These assays could then be used for high throughput quantification of RDX functional genes in groundwater and soil samples from sites contaminated with RDX. The primer design and review process has been completed for *xenA*, and the designed assays clearly improved the theoretical coverage and performed better on the SmartChip platform than the published primers.

MATERIALS AND METHODS

Reference Set Selection. Initially, a preliminary set of sequences for each of the six functional genes was collected from the National Center for Biotechnology Information (NCBI), Protein Data Bank (PDB), and UniProt databases. These sequences were subjected to strict criteria of ≥90% coverage, ≥40% identity and inclusion of feature residues and motifs identified from literature to ensure they were relevant to each gene. The sequences that passed were identified as the seed set used for developing the HMM model for the Michigan State University (MSU) Ribosomal Database Project's (RDP) Functional Gene Pipeline & Repository (FunGene) (Fish et al., 2013). With this HMM model, FunGene was capable of performing a more thorough search of the NCBI non-redundant protein database. The sequences gathered by FunGene were subjected to the same criteria as the seed set sequences before being included in the final vetted sequence set used for assay design.

Assay Design. New assays were designed using the EcoFunPrimer tool (Cole et al., 2014) available from RDP's website (https://github.com/rdpstaff). This tool designed complementary assays for each gene to maximize the overall theoretical coverage of the input sequence set. The melting temperature for the primers was set to be between 59°C and 61°C to ensure they would amplify efficiently with the SmartChip platform, which is usually run with the annealing and extension step at 60°C. A degeneracy of two was permitted in each assay and a total of two mismatches was allowed between the primers and amplified sequences.

The EcoFunPrimer tool did permit specific locations along the sequences to be identified for designing the forward and reverse primers. It was desired to locate the primers to capture some of the gene's identified residues and motifs during extension as this would enable the presence of these residues and motifs to be checked if the amplicons were sequenced. Multiple regions were tested during design, and the final set of assays was selected with the goal to achieve a balance between the uniform coverage and total number of assays.

For *xplA*, some assays were designed with NCBI's Primer-BLAST tool (Ye et al., 2012). This tool only permitted the input of one sequence, but the other parameters were kept the same as the EcoFunPrimer tool. The target region for the forward and reverse primers was determined based on the results from the EcoFunPrimer tool.

Assay Review. A final review was performed for the selected set of assays to evaluate the theoretical coverage compared to published assays, the ability of the amplicons to accurately represent the gene's diversity and the specificity of the assays.

All newly designed and previously published assays (Caballero et al., 2005; Indest et al., 2007; Seth-Smith et al., 2008; Fuller et al., 2010; Indest et al., 2010; Jung et al., 2011; Lee et al., 2012; Lorenz et al., 2013; Chong et al., 2014; Crocker et al., 2015; Wilson and Cupples, 2016; Zhang et al., 2016; Liang et al., 2017) were analyzed using RDP's ProbeMatch tool (Cole et al., 2014) to determine their theoretical coverage of the final vetted sequence set. This tool allows certain mismatches that are not permitted by the EcoFunPrimer tool, so it was necessary to analyze the newly designed assays with this to have an even and direct comparison with the published assays. Two mismatches were permitted between a primer and the target sequence in this analysis.

The Shannon diversity index was calculated for both the full-length sequences and the trimmed amplicons produced by each suite of newly designed assays. RDP's SeqFilters tool (Cole et al., 2014) was used to obtain the trimmed amplicons for each newly designed assay, and the number of mismatches allowed was again set to two. The resulting indices

for the original sequence set and trimmed amplicons were compared to determine if the amplicons were representative of the gene's diversity. A difference of 10% was set as the limit for the amplicons to be considered representative of the gene's diversity.

To test the specificity of the new assays, a search with the Basic Local Alignment Search Tool (BLAST) was run against NCBI's nucleotide database (nt). The output was analyzed to determine if any sequences were hit that were not good matches for the target gene. This was based on the best protein match from a translated BLAST (BLASTX) search with the resulting amplicons. The amplicons were required to have ≥90% identify and ≥80% alignment with the best protein match.

Assay Validation. All newly designed and currently published assays for *xplA, xenA and xenB* were tested against plasmids on a C1000 Touch thermal cycler with CFX96 real-time platform (Bio-Rad, Philadelphia, PA) to confirm they would amplify efficiently under the SmartChip platform's cycling parameters.

Individual reactions consisted of a total volume of 20 µL with 10 µL LightCycler 480 SYBR Green I Master (Roche Applied Sciences, Indianapolis, IN), 2 µL 10X primer master mix, 5 µL of plasmid DNA, and 3 µL of PCR grade water. The 10X primer master mix consisted of 5 µM each of the forward and reverse primers. For assays with two forward and/or reverse primers, this equates to 2.5 µM each of the individual forward and/or reverse primers. Cycling was performed following the SmartChip Real-Time PCR System protocol described by Wang et al. (2014). A melting curve analysis was also performed from 55°C to 95°C with reads every 0.5°C for 5s following amplification.

Reactions were run in triplicate for 10-fold serial plasmid dilutions (10^1-10^7 copies per reaction). Gene copy numbers were calculated as described by Ritalahti et al. (2006). A reaction blank was included for each trial and some runs also incorporated triplicate negative controls using either a plasmid containing *aceA* or *etnC* at 10^5 copies per reaction.

Assay Application. Assays for *xenA* were run on the SmartChip platform using the default cycling parameters: initial enzyme activation at 95°C for 2 min and 53 s followed by 40 cycles of denaturation at 95°C for 34 s and 60°C for 1 min and 4 s. A melt curve was then run from 56°C to 95°C at intervals of 0.4°C.

Reactions on the SmartChip had a total volume of 100 nl with 1X LightCycler 480 SYBR Green I Master, 0.5 µM each of the forward and reverse primers and sample DNA. For degenerate assays, the individual primer concentration was 0.25 µM so the total forward or reverse primer concentration would be 0.5 µM. Reactions were run in triplicate for 10-fold serial plasmid dilutions (10^1-10^8 copies per reaction). A reaction blank and triplicate negative controls with an aceA plasmid at 10^3 copies per reaction were also included.

RESULTS AND DISCUSSION

Reference Set Selection. The number of sequences gathered by FunGene were significantly greater than the number of sequences initially reviewed for inclusion in the HMM seed set (Table 1). From these, the final vetted sequence sets generally improved upon the HMM seed set, in terms of number of sequences included. The only exception to this was for *diaA*. There were not many feature residues and motifs identified in literature for *diaA* to use for eliminating sequences, so the HMM seed set shows high diversity with regard to the sequences and organisms included. For the final vetted sequence set, it was decided to add an additional criteria that the sequences be from the *Clostridium* genus, as this is the only genus currently linked to *diaA*.

For the *xplA* final vetted sequence set, the criteria of ≥90% coverage was not used. When reviewing the FunGene results, it was observed that very few sequences covered the full length of the *xplA* gene. Additional investigation revealed that many sequences were for one of the two domains of *xplA*, either the flavodoxin or cytochrome p450 domain. Therefore, it was decided to allow partial sequences for this gene, and from the final 49 sequences, 27 include the flavodoxin domain and 30 include the cytochrome p450 domain.

TABLE 1. Sequence sets for RDX functional genes.

Functional Gene	Initial Sequences	FunGene HMM Seed Set	FunGene Sequences	Final Vetted Sequence Set
xplA	7	7	189,222	49
xenA	1,852	14	199,903	309
xenB	4,435	58	110,413	425
diaA	1,051	96	123,691	16
nfsI	2,926	23	58,706	723
pnrB	1,178	32	61,601	56

Assay Design. From the final vetted sequence sets, suites of assays were designed for each functional gene. Coverage of these assay suites ranged from 62.5% to 96.3%, according to the EcoFunPrimer tool, and all include at least one feature residue and motif in the amplified region (Table 2). The coverage for the *xplA* assays is based on a subset of the final vetted sequence set. This subset is the 27 flavodoxin sequences as the new assays target this section of the gene. In the case of *diaA*, there is some variability in the number of feature residues and motifs amplified because for some sequences, assays were only able to be designed with the desired melting temperature if they included part of the feature residue and motif region. The results for *diaA*, *nfsI*, and *pnrB* are preliminary and subject to further review.

In addition to the overall coverage of the assay suite, it was noted that some individual assays accounted for a larger proportion of the coverage. Considering only the first three assays for each gene, it was found that these alone targeted at least 50% of the sequences (Figure 1). This is particularly important for the *xenA* and *xenB* assay suites as they contain a large number of assays. In some situations, it may not be feasible to test samples with all of these assays so the first three may be used to reduce testing and still maintain a high coverage of the gene sequences targeted.

TABLE 2. Designed assay suite characteristics.

Characteristic	Functional Gene Assay Suite					
	xplA	***xenA***	***xenB***	***diaA****	***nfsI****	***pnrB****
Number of Assays	3	17	19	3	9	7
EcoFunPrimer Coverage	96.3%	81.6%	90.4%	62.5%	65.0%	87.5%
Amplicon Length (bp)	157-165	190-200	129-134	84-136	154-197	156-189
Residues Amplified	1	4	1	1-2	4	10
Motifs Amplified	2	1	1	0-1	2	1

*preliminary results, subject to further review

Assay Review. Further evaluation of the *xplA*, *xenA* and *xenB* assays confirmed that they were a good improvement on published ones. While the *xplA* assays did not have a dramatic increase in the theoretical coverage, both *xenA* and *xenB* did (Figure 2). Also, the Shannon diversity index was similar between the full-length and expected amplicon regions of the reference sequences for all three genes. This suggests that the assay suites

are maintaining the diversity present in the original sequence set. Also, only two assays from *xenB* hit targets with BLAST that were not good matches. These assays were later eliminated as they also resulted in poor amplification of the target plasmid.

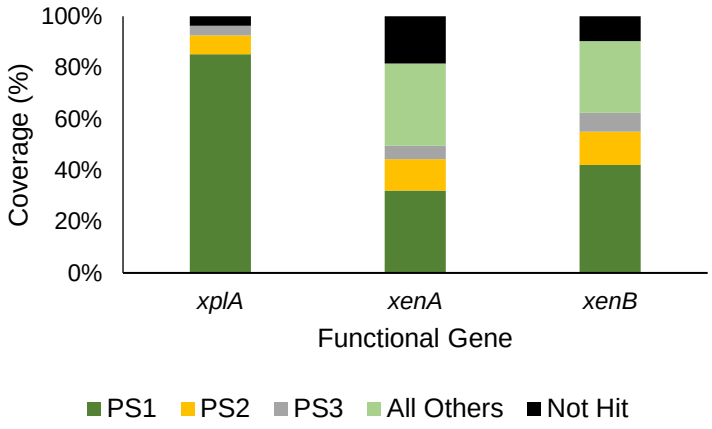

FIGURE 1. EcoFunPrimer coverage of assays.

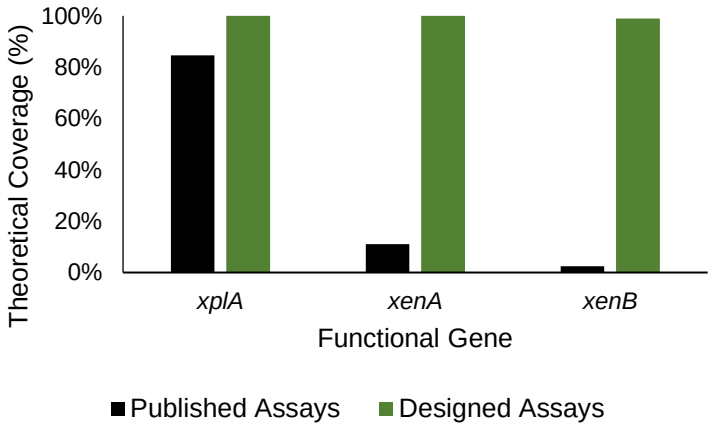

FIGURE 2. Theoretical coverage of assays.

Assay Validation. From testing the assays on the Bio-Rad thermal cycler, a few were observed to have poor qPCR efficiency and low R^2 value, late amplification, or unusual amplification. For the designed *xplA* assays, PS1 and PS2, designed with EcoFunPrimer, were eliminated and replaced with PS5 and PS6, designed with Primer-BLAST. Only one designed assay was removed from the *xenA* suite and five from the *xenB* suite. Due to the higher number of assays in these suites initially, it was determined not to be necessary to replace these. Also, the theoretical coverage from ProbeMatch did not change for these two genes.

The remaining assays had acceptable qPCR efficiencies, R^2 values and detection limits. The *xplA* suite had efficiencies ranging from 72% to 104%, the *xenA* suite had 82% to 104%, and the *xenB* suite had 84% to 116%. The R^2 value ranged from 0.96 to 1.00 and the detection limit was between 10^2 and 10^1 copies per reaction for most assays. The only exceptions were one assay each from the *xplA* and *xenA* suites that had a detection limit at 10^3 copies per reaction. In general, these were comparable to, or a slight improvement upon, the published assays (Figure 3).

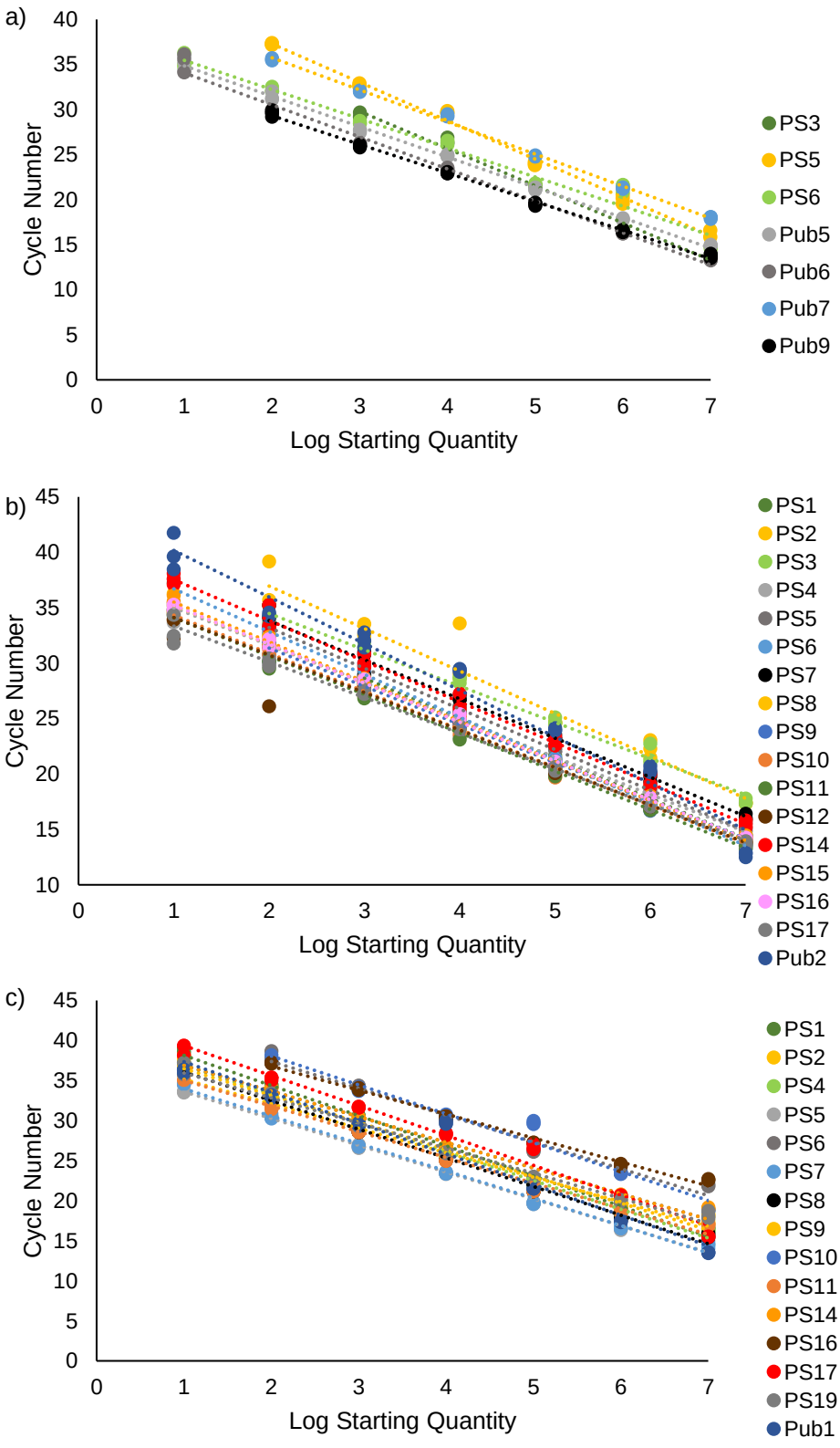

FIGURE 3. Bio-Rad thermal cycler standard curves for (a) xplA, (b) xenA and (c) xenB where "PS" identifies newly designed primer sets and "Pub" identifies published ones.

Assay Application. An additional two assays were removed from the *xenA* suite due to early amplification in the negative and no template controls on the SmartChip platform. However, the theoretical coverage is still 100% for this assay suite. Also, the standard curves from the SmartChip platform for the remaining assays were similar to those from the Bio-Rad thermal cycler platform based on efficiency and R^2 value. Only three assays were slightly below the previously identified range for efficiency, with the lowest being 78%, and one was below the R^2 value range with an R^2 of 0.926. However, the detection limit did change to being 10^4 or 10^3 copies per reaction for most assays, with only one achieving 10^2 copies per reaction (Table 3). Despite these changes, the designed assays still performed better than the published assay, which had a detection limit of 10^5 copies per reaction (Figure 4).

TABLE 3. SmartChip standard curve characteristics for *xenA*.

Characteristics	Detection Limit (copies per reaction)			
	10^2	10^3	10^4	10^5
Number of Assays	1	4	8	1
Y-Intercept	35.3	35.6 to 37.2	32.7 to 43.8	40.8
Slope	-3.89	-3.69 to -3.93	-3.48 to -3.99	-3.60
R^2	0.991	0.975 to 0.997	0.926 to 0.996	0.996
Efficiency	81%	80% to 87%	78% to 94%	90%
Orders of Magnitude	6	5 to 6	4 to 5	4

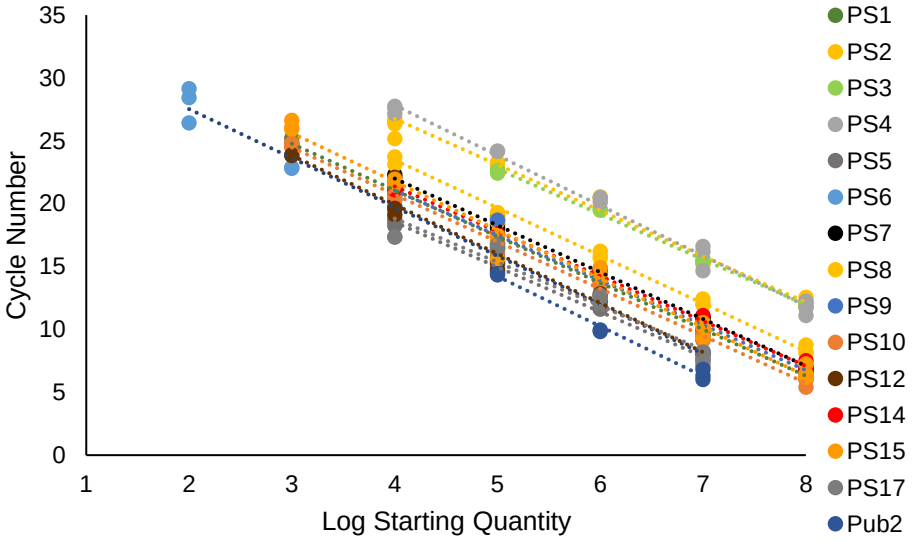

FIGURE 4. SmartChip standard curves for *xenA* where "PS" identifies newly designed primer sets and "Pub" identifies published ones.

CONCLUSIONS

RDP's EcoFunPrimer tool, in conjunction with NCBI's Primer-BLAST tool, has been used to successfully design a suite of assays to increase the coverage of the known genetic diversity for six RDX functional genes. The suites for *xplA*, *xenA* and *xenB* have undergone additional review and validation, and only a few assays were removed. Even with the removal of these assays, the suites are an improvement on published assays for these genes.

The designed assays for *xenA* were also tested on the SmartChip platform using 10-fold serial plasmid dilutions. This confirmed that these assays can be used with this

platform for analyzing environmental samples. It also showed that they performed better than the published assay.

Next steps involve subjecting the *diaA*, *nfsI*, and *pnrB* suite of assays to the same review and validation as the other three suites and confirming the specificity of all of the assays against environmental samples. The specificity against environmental samples will be checked by confirming the PCR product length. After validation of all of the assays, groundwater and soil environmental samples provided from the Naval Base Kitsap, Bangor, Washington State by Mandy Michalsen (USACE) and Malcolm Gander (NAVFAC) will be analyzed on the SmartChip platform to quantify the RDX functional genes present.

ACKNOWLEDGEMENTS

This work was funded by MSU's Interdisciplinary Team Building Initiative and Academic Achievement Graduate Assistantship. The authors thank Santosh Gunturu for his assistance with the EcoFunPrimer software and Dr. Robert Stedtfeld and Tiffany Stedtfeld for their technical support.

REFERENCES

Caballero, A., J. J. Lazaro, J. L. Ramos and A. Esteve-Nunez. 2005. "PnrA, a new nitroreductase-family enzyme in the TNT-degrading strain *Pseudomonas putida* JLR11." *Environmental Microbiology 7*(8): 1211-1219.

Chong, C. S., D. K. Sabir, A. Lorenz, C. Bontemps, P. Andeer, D. A. Stahl, S. E. Strand, E. L. Rylott and N. C. Bruce. 2014. "Analysis of the *xplAB*-Containing Gene Cluster Involved in the Bacterial Degradation of the Explosive Hexahydro-1, 3, 5-Trinitro-1, 3, 5-Triazine." *Applied and Environmental Microbiology 80*(21): 6601-6610.

Clausen, J., J. Robb, D. Curry and N. Korte. 2004. "A Case Study of Contaminants on Military Ranges: Camp Edwards, Massachusetts, USA." *Environmental Pollution 129*(1): 13-21.

Cole, J. R., Q. Wang, J. A. Fish, B. Chai, D. M. McGarrell, Y. Sun, C. T. Brown, A. Porras-Alfaro, C. R. Kuske and J. M. Tiedje. 2014. "Ribosomal Database Project: Data and Tools for High Throughput rRNA Analysis." *Nucleic Acids Research 42*(D1): D633-D642.

Crocker, F. H., K. J. Indest, C. M. Jung, D. E. Hancock, M. E. Fuller, P. B. Hatzinger, S. Vainberg, J. D. Istok, E. Wilson and M. M. Michalsen. 2015. "Evaluation of microbial transport during aerobic bioaugmentation of an RDX-contaminated aquifer." *Biodegradation 26*(6): 443-451.

Cupples, A. M. 2013. "RDX Degrading Microbial Communities and the Prediction of Microorganisms Responsible for RDX Bioremediation." *International Biodeterioration & Biodegradation 85*: 260-270.

Fish, J. A., B. Chai, Q. Wang, Y. Sun, C. T. Brown, J. M. Tiedje and J. R. Cole. 2013. "FunGene: The Functional Gene Pipeline and Repository." *Frontiers in Microbiology 4*.

Fuller, M. E., K. McClay, M. Higham, P. B. Hatzinger and R. J. Steffan. 2010. "Hexahydro-1, 3, 5-trinitro-1, 3, 5-triazine (RDX) bioremediation in groundwater: are known RDX-degrading bacteria the dominant players?" *Bioremediation Journal 14*(3): 121-134.

Indest, K. J., F. H. Crocker and R. Athow. 2007. "A TaqMan polymerase chain reaction method for monitoring RDX-degrading bacteria based on the *xplA* functional gene." *Journal of Microbiological Methods 68*(2): 267-274.

Indest, K. J., C. M. Jung, H.-P. Chen, D. Hancock, C. Florizone, L. D. Eltis and F. H. Crocker. 2010. "Functional characterization of pGKT2, a 182-kilobase plasmid containing the *xplAB* genes, which are involved in the degradation of hexahydro-1, 3,

5-trinitro-1, 3, 5-triazine by *Gordonia* sp. strain KTR9." *Applied and Environmental Microbiology 76*(19): 6329-6337.

Jung, C. M., F. H. Crocker, J. O. Eberly and K. J. Indest. 2011. "Horizontal gene transfer (HGT) as a mechanism of disseminating RDX-degrading activity among Actinomycete bacteria." *Journal of Applied Microbiology 110*(6): 1449-1459.

Lee, B.-U., S.-J. Ha, B. Hyun, Y.-S. Cho and K.-H. Oh. 2012. "σ^{32} Regulation of the *algA* and *pnrB* Expression in *Pseudomonas* sp HK-6 which Degrades the Explosive Hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX)." *African Journal of Microbiology Research 6*(19): 4194-4200.

Li, R. W., J. G. Giarrizzo, S. Wu, W. Li, J. M. Duringer and A. M. Craig. 2014. "Metagenomic Insights into the RDX-Degrading Potential of the Ovine Rumen Microbiome." *Plos One 9*(11).

Liang, S.-H., D.-W. Hsu, C.-Y. Lin, C.-M. Kao, D.-J. Huang, C.-C. Chien, S.-C. Chen, I. J. Tsai and C.-C. Chen. 2017. "Enhancement of microbial 2, 4, 6-trinitrotoluene transformation with increased toxicity by exogenous nutrient amendment." *Ecotoxicology and Environmental Safety 138*: 39-46.

Lorenz, A., E. L. Rylott, S. E. Strand and N. C. Bruce. 2013. "Towards engineering degradation of the explosive pollutant hexahydro-1,3,5-trinitro-1,3,5-triazine in the rhizosphere." *Fems Microbiology Letters 340*(1): 49-54.

Ritalahti, K. M., B. K. Amos, Y. Sung, Q. Wu, S. S. Koenigsberg and F. E. Löffler. 2006. "Quantitative PCR Targeting 16S rRNA and Reductive Dehalogenase Genes Simultaneously Monitors Multiple *Dehalococcoides* Strains." *Applied and Environmental Microbiology 72*(4): 2765-2774.

Seth-Smith, H. M. B., J. Edwards, S. J. Rosser, D. A. Rathbone and N. C. Bruce. 2008. "The explosive-degrading cytochrome P450 system is highly conserved among strains of *Rhodococcus* spp." *Applied and Environmental Microbiology 74*(14): 4550-4552.

Wang, F.-H., M. Qiao, J.-Q. Su, Z. Chen, X. Zhou and Y.-G. Zhu. 2014. "High Throughput Profiling of Antibiotic Resistance Genes in Urban Park Soils with Reclaimed Water Irrigation." *Environmental Science & Technology 48*(16): 9079-9085.

Wilson, F. P. and A. M. Cupples. 2016. "Microbial community characterization and functional gene quantification in RDX-degrading microcosms derived from sediment and groundwater at two naval sites." *Applied Microbiology and Biotechnology 100*(16): 7297-7309.

Ye, J., G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen and T. L. Madden. 2012. "Primer-BLAST: A Tool to Design Target-Specific Primers for Polymerase Chain Reaction." *BMC Bioinformatics 13*(1): 134.

Zhang, L., R. Routsong, Q. Nguyen, E. L. Rylott, N. C. Bruce and S. E. Strand. 2016. "Expression in grasses of multiple transgenes for degradation of munitions compounds on live-fire training ranges." *Plant Biotechnology Journal*.