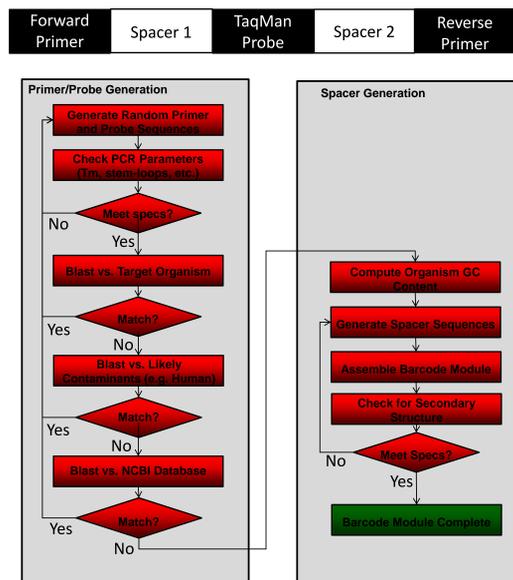


## Abstract

Tracking biowarfare agent simulants in the environment requires specific detection methods that discriminate between the target strain and all potential natural and artificial interferents including previously utilized tester strains. We recently showed that genomic insertion of short identification tags, called “barcodes,” allows detection of chromosomally tagged strains by real-time PCR (1-2). Here we introduce BarCoder, a bioinformatics tool that facilitates the process of creating sets of stable and uniquely identifiable barcoded strains. BarCoder utilizes the genomic sequence of the target strain and a set of user-specified PCR parameters to generate a list of suggested barcode “modules” that consist of binding sites for primers and probes and appropriate spacer sequences. Each module is designed to yield optimal PCR amplification and unique identification. Optimal amplification includes metrics such as ideal  $T_m$  and G/C-content, appropriate spacing, and minimal stem-loop formation; unique identification includes low blast hits against the target organism, previously generated barcode modules, and other databases, such as NCBI. The algorithm also suggests candidate genomic insertion sites by integrating information such as annotated features, predicted ORFs, and repetitive structures. We tested the ability of our algorithm to suggest appropriate barcodes by generating 12 modules for *Bacillus anthracis* simulant *B. thuringiensis* serovar *kurstaki* and three each for other potential target organisms with variable G/C content. Real-time PCR detection assays directed at barcodes were specific and yielded minimal cross-reactivity with a panel of near-neighbor and potential contaminant materials. The BarCoder algorithm facilitates the generation of barcoded biological simulants by (a) eliminating the task of creating modules by hand, (b) minimizing optimization of PCR assays, and (c) reducing effort wasted on non-unique barcode modules.

## Algorithm Outline

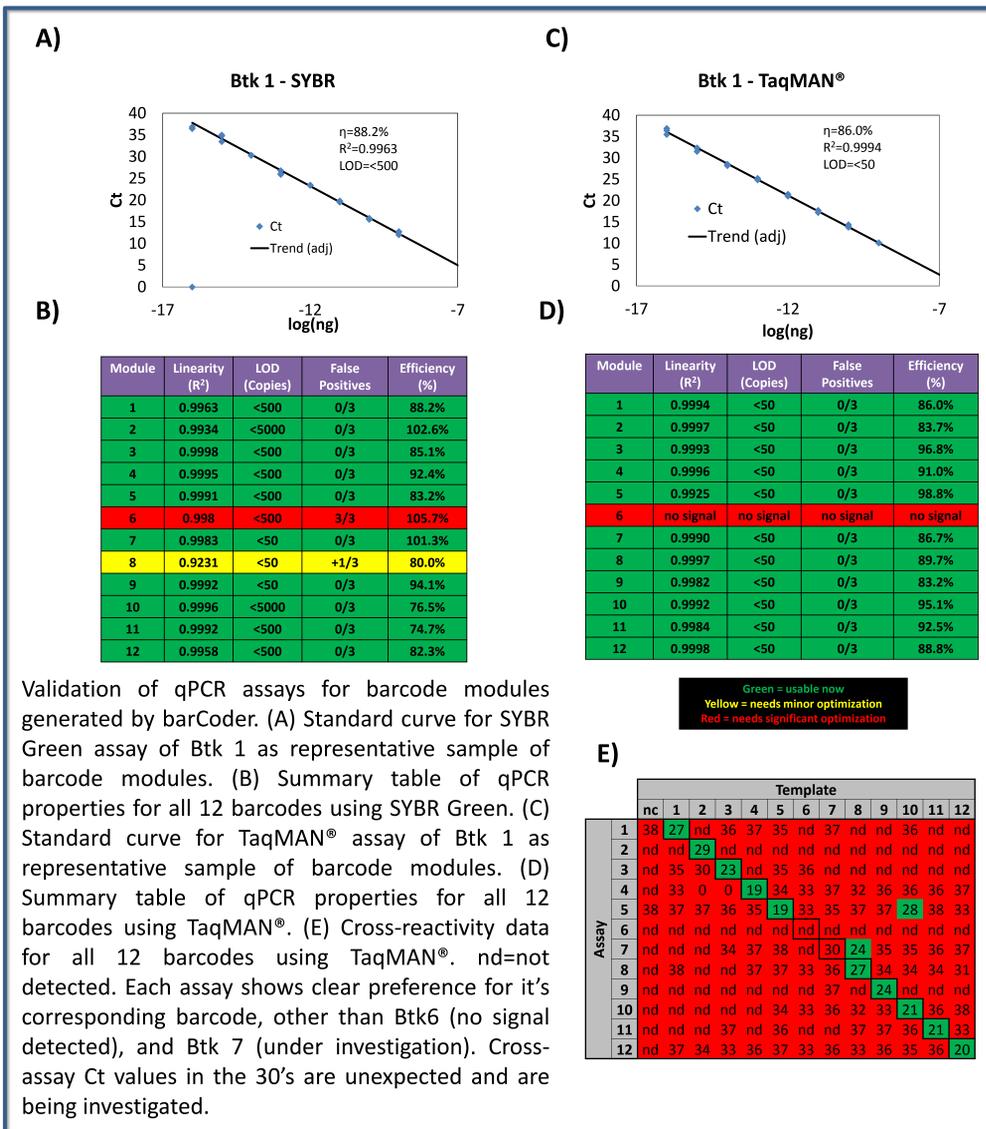


The barCoder algorithm generates genetic barcode sequences that meet customizable criteria for uniqueness and PCR properties. Uniqueness can include differentiation from other barcodes, intra-host uniqueness, inter-host uniqueness (for likely contaminant organisms, such as human), or a database such as NCBI. PCR parameters include melting temperatures and predicted issues such as primer dimers.

## Assay Optimization

Parameter	SYBR	TaqMan
Cycling Conditions	95°C for 10 min 40x: 95°C for 15s / 55°C for 1 min	95°C for 10 min 40x: 95°C for 15s / 55°C for 1 min
Mastermix	10 uL of SYBR Green PCR Master Mix (Applied Biosystems)	10 uL of TaqMan Universal PCR Master Mix (Applied Biosystems)
Primers	60 nM each	900 nM each
Probe	n/a	250 nM
Final Volume	20 uL	20 uL
Instrument	ABI 7900HT	ABI 7900HT

Assays were optimized first by performing temperature gradient PCR to determine a maximum common extension temperature for efficient amplification. Second, primer concentrations were scanned for SYBR assays to minimize false positive results. TaqMan assays were efficient without further optimization from standard concentrations.



## Assay Optimization (ctd.)

Genome DNA Template	Assay								
	1	2	3	4	5	6	7	8	9
Positive Control	7.7	11	10	11	12	9.7	11	13	9
Bacillus anthracis Ames	40	34	40	40	40	27	40	40	40
B. anthracis Sterne	32	32	40	40	40	37	40	40	40
B. a NNR-Δ1	32	40	40	40	40	36	40	40	40
B. a VNR1-Δ1	40	32	40	40	40	35	40	40	40
B. cereus	40	40	40	40	40	40	40	40	40
B. mycoides	40	34	40	40	40	40	40	40	40
B. atrophaeus	40	32	40	40	40	36	40	40	40
B. thuringiensis subsp. israelensis	40	33	40	40	40	35	40	40	40
B. thuringiensis subsp. kurstaki	40	35	40	40	40	34	40	40	40
Bordetella pertussis	40	40	38	40	40	36	40	40	40
Escherichia coli	40	33	40	40	40	36	40	40	40
E. coli O157:H7	38	33	40	40	40	36	40	40	40
Francisella tularensis	40	32	40	40	40	ND	ND	ND	ND
Human Placenta	40	34	40	40	40	36	40	40	40
pRP1028-T1B2	38	34	40	40	40	33	40	40	40
Pseudomonas aeruginosa	40	32	40	40	40	40	40	40	40
Salmonella typhimurium	40	33	40	40	40	34	40	40	40
Streptococcus pyogenes	40	32	40	40	40	36	40	40	11
Yersinia pestis	40	34	40	40	40	34	40	40	40

Preliminary panel of near-neighbors and other organisms of interest. The first 9 barcode modules were screened against a series of relevant genomic DNA samples resulting in only a single positive hit. These results used a pre-optimization qPCR protocol using TaqMAN® and are therefore preliminary.

## Applications

- Novel tools for tracking and predicting the spread of pathogens.
- “Left-of-boom” microbial forensic tools – pre-distribution barcoding to deter and/or attribute malfeasance
- Novel molecular recognition elements for the detection or mitigation of biological threats.
- Better evaluation of decontamination tools at the individual or group level.
- Improved capability in tracking environmental clean-up and disposal methods.

## Acknowledgements

This project was funded by the Defense Threat Reduction Agency. Work presented is the opinion of the authors and does not necessarily reflect the official policy of the US Government, US Army, or DTRA.

1. Buckley P, et al. *Applied and Environmental Microbiology*. 2012.
2. Emanuel P, et al. *Applied and Environmental Microbiology*. 2012.