

BarCoder: a Bioinformatic Algorithm for Design of Organism-specific PCR Tracking Tags

Matthew Lux, Ph.D. ¹, Sarah Katoski ², Jennifer Gibbons, Ph.D. ³, Henry Gibbons, Ph.D. ¹
 (1) U.S. Army Edgewood Chemical Biological Center, (2) Leidos, Inc., (3) Excet, Inc.

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 insertion sites by integrating information such as annotated features, predicted open reference frames (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.

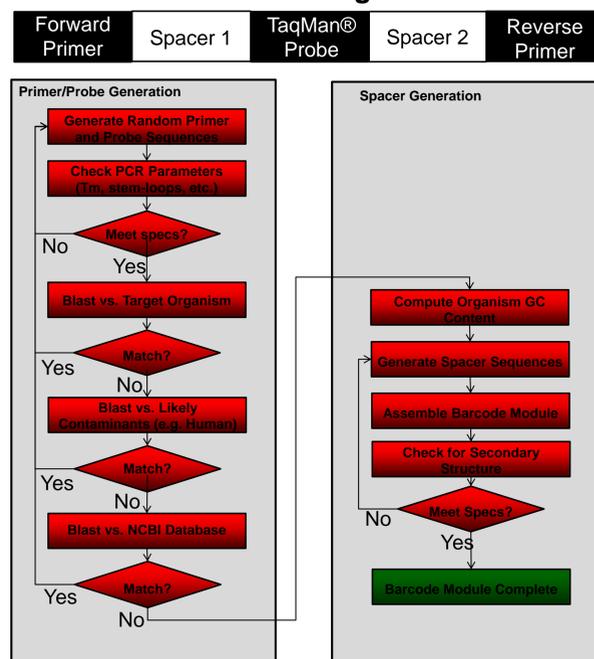
Background

Preparedness for any attack is inherently reliant on our ability to project the potential impact. For biological attacks, this projection is especially challenging as realistic testing of impact models has obvious public health limitations. However, models of the first stage of a potential attack, namely environmental dispersal, can indeed be tested by dispersing non-pathogenic biological simulants instead of actual agent. Ideal simulants are as closely related to the threat agent of interest as possible without posing a risk to humans or the environments into which they are released. Of particular concern are spore-forming threat agents, such as *Bacillus anthracis* (BA), the causative agent of anthrax. Such spore-forming agents can be mass-produced in a stable form, disseminated easily through water or aerosol, and can persist for a long time in the environment after release. Thus, predictive models of spore dissemination and environmental persistence are crucial to preparedness for such an attack.

For BA, biological simulants have been used for decades in the form of *Bacillus atrophaeus* var. *globigii* (BG) and, more recently, *Bacillus thuringiensis* serovar *kurstaki* (Btk), which is more closely related to BA and is commercially available as an insecticide. Yet, these viable simulants still present practical challenges for dispersal testing stemming from the environmental persistence of spores, whereby it becomes difficult or impossible to differentiate between repeated releases. To overcome this challenge we recently demonstrated the insertion of a genetic "barcode" that allows discrimination of the barcoded strain from wild types via qPCR without disrupting the normal phenotype¹. We further demonstrated the utility of this approach by using a release of barcoded Btk to test models of environmental dispersion².

In order to build on this proof-of-concept work we are developing methods to create large libraries of barcoded strains of several organisms to allow repeated releases without cross-talk. Here we present the development and in vitro testing of an algorithm to generate barcode sequences that have ideal qPCR properties to allow standardized testing for all barcoded strains, as well as minimal cross-talk with each other, the host genome, and the environmental genetic background.

The BarCoder Algorithm

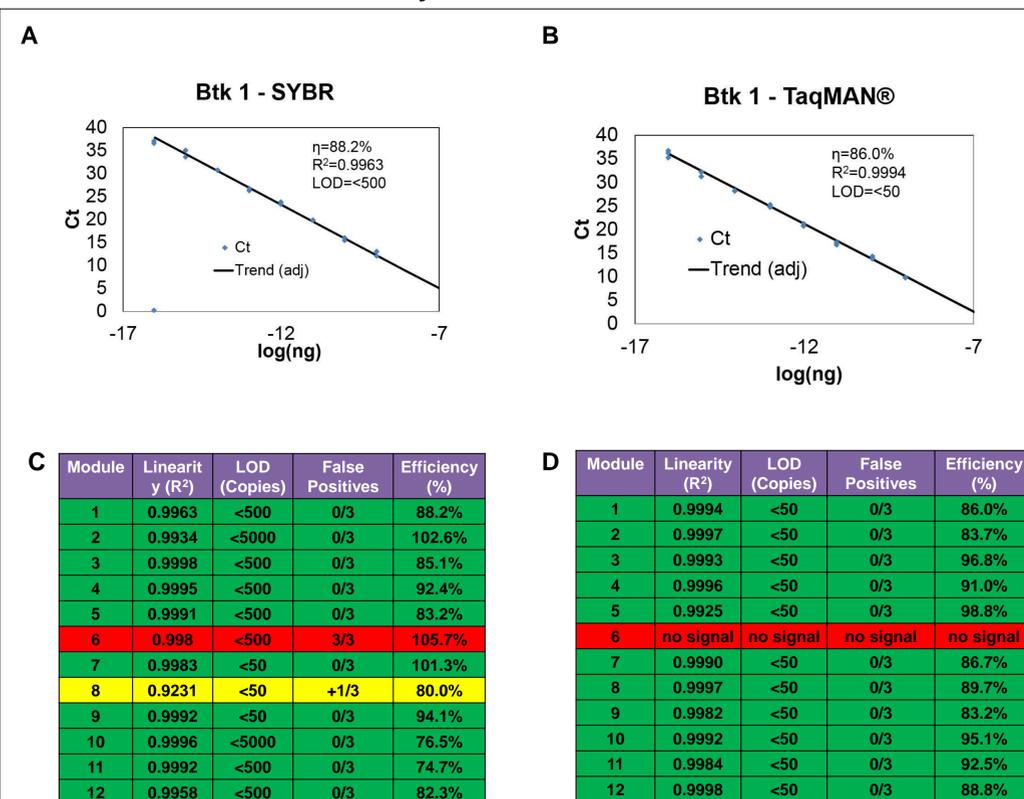


Assay Optimization

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 Performance Data



Green = usable now
 Yellow = needs minor optimization
 Red = needs significant optimization

Validation of qPCR assays for barcode modules generated by BarCoder. (A) Standard curve for SYBR Green assay of Btk 1 as a representative sample of barcode modules. (B) Summary table of qPCR properties for all 12 barcodes using SYBR Green. (C) Standard curve for TaqMAN® assay of Btk 1 as a representative sample of barcode modules. (D) Summary table of qPCR properties for all 12 barcodes using TaqMAN®. (E) Cross-reactivity data for all 12 barcodes using TaqMAN®. nd=not detected. Each assay shows a clear preference for its corresponding barcode, other than Btk6 (no signal detected), and Btk 7 (under investigation). Cross-assay Ct values in the 30's are unexpected and are being investigated.

E

Assay	nc	Template											
		1	2	3	4	5	6	7	8	9	10	11	12
1	38	27	nd	36	37	35	nd	37	nd	36	nd	nd	nd
2	nd	29	nd										
3	nd	35	30	23	nd	35	36	nd	nd	nd	nd	nd	nd
4	nd	33	0	19	34	33	37	32	36	36	36	37	37
5	38	37	37	36	35	19	33	35	37	37	28	38	33
6	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
7	nd	nd	nd	34	37	38	nd	30	24	35	35	36	37
8	nd	38	nd	nd	37	37	33	36	27	34	34	34	31
9	nd	nd	nd	nd	nd	nd	nd	37	nd	24	nd	nd	nd
10	nd	nd	nd	nd	34	33	36	32	33	21	36	38	38
11	nd	nd	nd	37	nd	36	nd	nd	37	37	36	21	33
12	nd	37	34	33	36	37	33	36	33	36	35	36	20

Barcode qPCR Assay Reactivity against Near Neighbors, Biothreat Agents and Human DNA

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	13
<i>Bacillus anthracis</i> Ames	40	34	40	40	40	27	40	40	40
<i>B. anthracis</i> ΔSterne	32	32	40	40	40	37	40	40	40
<i>B. a</i> NNR-Δ1	32	40	40	40	40	36	40	40	40
<i>B. a</i> VNR1-Δ1	40	32	40	40	40	35	40	40	40
<i>B. cereus</i>	40	40	40	40	40	40	40	40	40
<i>B. mycoides</i>	40	34	40	40	40	40	40	40	40
<i>B. atrophaeus</i>	40	32	40	40	40	36	40	40	40
<i>B. thuringiensis</i> subsp. <i>israelensis</i>	40	33	40	40	40	35	40	40	40
<i>B. thuringiensis</i> subsp. <i>kurstaki</i>	40	35	40	40	40	34	40	40	40
<i>Bordetella pertussis</i>	40	40	36	40	40	36	40	40	40
<i>Escherichia coli</i>	40	33	40	40	40	36	40	40	40
<i>E. coli</i> O157:H7	38	33	40	40	40	36	40	40	40
<i>Francisella tularensis</i>	40	32	40	40	40	ND	ND	ND	ND
Human Placenta	40	34	40	40	40	36	40	40	40
pRP1028-TIB2	38	34	40	40	40	33	40	40	40
<i>Pseudomonas aeruginosa</i>	40	32	40	40	40	40	40	40	40
<i>Salmonella typhimurium</i>	40	33	40	40	40	34	40	40	40
<i>Streptococcus pyogenes</i>	40	32	40	40	40	36	40	40	11
<i>Yersinia pestis</i>	40	34	40	40	40	34	40	40	40

40 30 20 10 0
Mean Ct

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.

References

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

Acknowledgements: The authors thank the Defense Threat Reduction Agency/Joint Science and Technology Office for their assistance and funding of this work under Project CB4123. The views expressed in this presentation are those of the authors and do not necessarily reflect official policy or the position of the Department of Defense or the U.S. Government.

