


METHODOLOGY ARTICLE

Open Access



barCoder: a tool to generate unique, orthogonal genetic tags for qPCR detection

Casey B. Bernhards^{1,2†}, Matthew W. Lux^{1†}, Sarah E. Katoski¹, Tyler D. P. Goralski¹, Alvin T. Liem^{1,3} and Henry S. Gibbons^{1*} 

*Correspondence:

henry.s.gibbons.civ@mail.mil

[†]Casey B. Bernhards and Matthew W. Lux contributed equally to this work

¹ U.S. Army Combat Capabilities Development Command Chemical Biological Center, Aberdeen Proving Ground, MD 21010, USA

Full list of author information is available at the end of the article

Abstract

Background: Tracking dispersal of microbial populations in the environment requires specific detection methods that discriminate between the target strain and all potential natural and artificial interferences, including previously utilized tester strains. Recent work has shown that genomic insertion of short identification tags, called “barcodes” here, allows detection of chromosomally tagged strains by real-time PCR. Manual design of these barcodes is feasible for small sets, but expansion of the technique to larger pools of distinct and well-functioning assays would be significantly aided by software-guided design.

Results: Here we introduce barCoder, a bioinformatics tool that facilitates the process of creating sets of 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 melting temperature 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 databases (such as NCBI). We tested the ability of our algorithm to suggest appropriate barcodes by generating 12 modules for *Bacillus thuringiensis* serovar *kurstaki*—a simulant for the potential biowarfare agent *Bacillus anthracis*—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.

Conclusions: The barCoder algorithm facilitates the generation of synthetically 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.

Keywords: DNA barcodes, Genetic barcoding, Genome tagging, Tagged strains, Microbial forensics, qPCR detection



Background

Developing an understanding of organisms in their natural ecological niches requires the ability to measure the dynamic interaction with their environment, either at the level of the individual or at population scales. For metazoa, a number of approaches have been utilized to track individuals of a species, including simple bands or markings conferring unique identifiers, Passive Integrated Transponders (PITs), telemetry devices, and biologgers [1, 2]. These approaches are limited to large organisms, as they require either direct visual inspection or electronic devices that can be attached by physical means to the body of the organism in question. As the field of environmental microbiology continues to mature, novel tools to facilitate “tag and release” studies are critical to understanding microbial interactions within existing environmental niches or in the context of introduction into new environments. Early efforts to track environmental fate of genetically modified organisms in field releases utilized fluorescently or metabolically marked strains of *Pseudomonas putida* [3, 4] and *Pseudomonas fluorescens* [5, 6]. Likewise, spontaneous rifampicin-resistant mutants have been used to track establishment and persistence of introduced isolates in field trials [7]. However, conventional selectable, chromogenic, or fluorescent markers carry metabolic costs that can compromise the carrier strain’s fitness in resource-constrained environments [8], revealing the need for phenotypically neutral, non-coding, genomic insertions that can differentiate introduced strains from native flora.

The development of DNA synthesis chemistry, microarray technology, quantitative PCR (qPCR), and high-throughput sequencing resulted in the development of several important capabilities and tagging approaches. Early studies used transposons containing short synthetic barcodes to identify virulence factors in several organisms [9, 10]. As oligonucleotide synthesis technology became more sophisticated and costs decreased, longer tags could be produced, resulting in the use of tagged strains to study the spatiotemporal dispersion in systems otherwise unamenable to tracking. In particular, significant work has been done to understand the details of stochastic dynamics of *Salmonella* infections by monitoring the relative quantities of tagged strains in different locations within the host [11–13]. These tagged strains, known as wild-type isogenic tagged strains (WITS), contain short, unique sequences inserted into the genome to allow quantitation by qPCR [11]. Similar work has been done to study population dynamics during infection for several other bacterial and viral pathogens [13–20].

The ability to track the fate of microbes introduced into an environment is also of interest to the biodefense research community. Spores of *Bacillus anthracis*, the causative agent of anthrax, were used in the high-profile 2001 anthrax mail attacks and were historically weaponized by both the United States and Soviet Union on large scales [21]. An important angle for preparedness against a potential attack includes an understanding of how spores released into the environment might disperse, persist, and migrate. The release of live *B. anthracis* spores (and indeed, even of attenuated strains) in an outdoor test is impossible due to public health concerns. Instead, close biological relatives are used as simulants. In the case of *B. anthracis*, recent work has used *Bacillus thuringiensis* serovar *kurstaki* (Btk) due to its similar physiological and biochemical properties [22–26]. Yet, even with the use of an adequate simulant, repeated dispersion testing on the same test site is problematic due to a need to distinguish between past and present

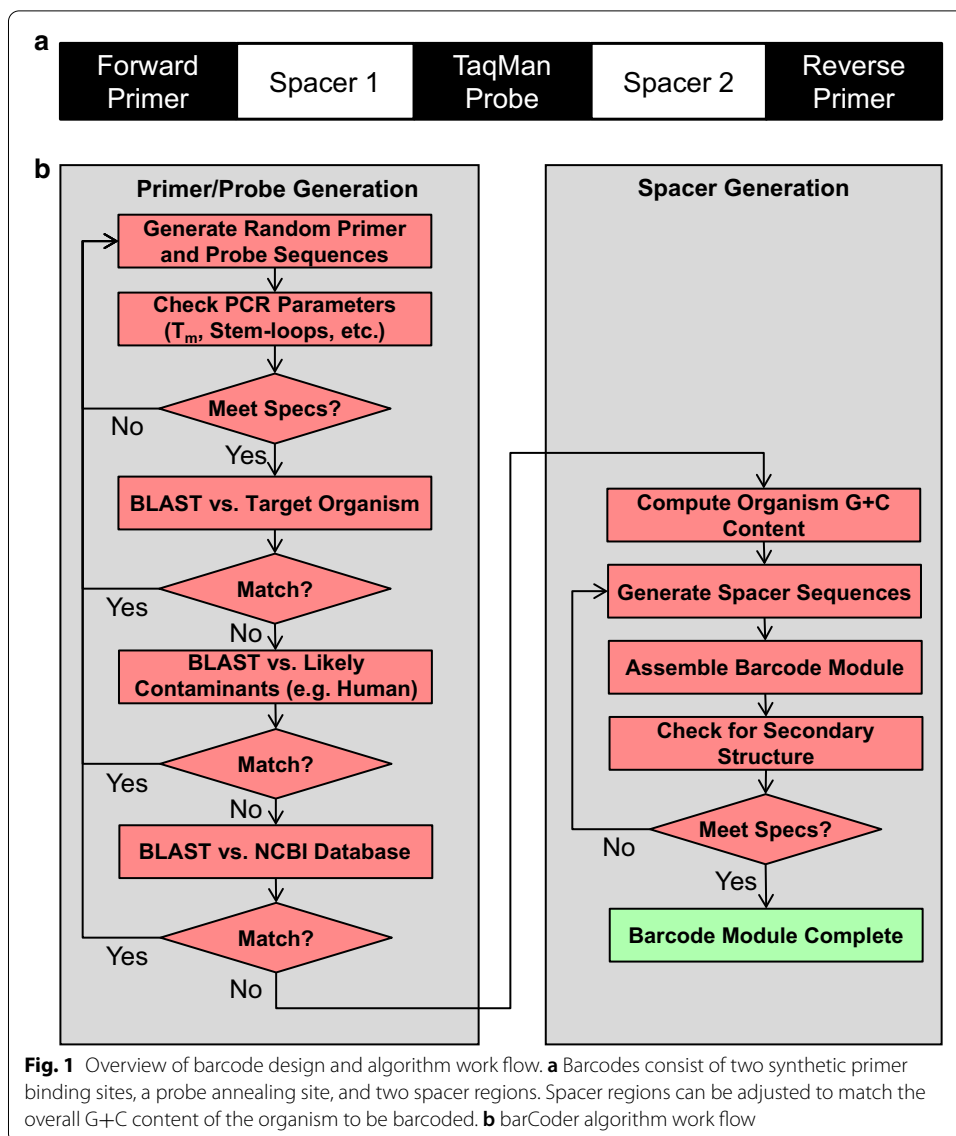
testing, especially for a ubiquitous environmental bacterium such as Btk that is also in widespread use as a commercial biopesticide [27–30]. In addition, the problem of “signature erosion” has diminished the utility of endogenous genomic signatures as detection tools as the diversity of sequence data in public databases has exploded [31, 32].

To overcome these challenges, we previously inserted unique artificial genetic “barcodes,” designed to enable rapid detection by qPCR, into the Btk genome [24] and subsequently tested the system in a field release [23]. We note that the concept of “barcoding” described in this context differs from that used in taxonomic identification of living organisms [33], which—in contrast to the synthetic signatures employed here—uses endogenous mitochondrial and genomic DNA sequences as molecular classification tools. The barcoded strains constructed for our field release [23] were successfully detected in field samples using qPCR assays, but, like the earlier WITS strains, did not exploit the full ability of bioinformatics and synthetic biology that has become available. Most notably, each of the tags required its own specific PCR assay conditions, which makes scaling up to larger numbers of barcodes prohibitive. In this work, we have built upon our previous work by developing an algorithm, called barCoder, to generate barcode sequences that are unique among a pool of barcoded strains and require minimal development of qPCR assays. The algorithm also provides numerous features to minimize experimental troubleshooting efforts and customize amplicon properties. Here, we present the algorithm, as well as experimental validation of its ability to generate a potentially unlimited pool of highly diverse DNA barcodes, each with its own specific qPCR assay.

Results

Barcode design

We envisioned the possibility of creating synthetic barcodes that would amplify specifically, optimally, and in a manner consistent with the genomic characteristics of the target organism. We separated the barcode module into five constituent segments: two primer binding sites, the probe sequence, and two spacer sequences (Fig. 1a). Our design takes advantage of commonalities between the two major types of qPCR strategies, which are based on intercalating dyes (e.g. SYBR Green) that generate a fluorescent signal only in response to the formation of double-stranded DNA, or based on the liberation of a 5′-linked fluorophore from a probe dually labeled with a 3′-linked quencher by the 5′–3′ exonuclease activity of the polymerase (referred to herein as TaqMan[®] for the probes used). Both assay types require a forward and reverse primer with similar constraints such as amplicon length, melting temperature (T_m), G+C content, GC clamp, potential secondary structure, and primer dimer formation. The primary design difference between the qPCR assay types is presence of a third probe sequence that is required only for TaqMan[®] assays; the probe sequence carries its own recommended design guidelines. Thus, in general, the same primer set can be used for either approach with the same barcode. From the perspective of qPCR assay design, the remaining spacer sequences between primers/probe are largely immaterial other than to meet ideal amplicon size targets, and therefore can be generated randomly, with constraints (see Algorithm design).



Algorithm design

The barCoder algorithm workflow is depicted in Fig. 1b. The algorithm starts by generating random primer/probe sequences to meet PCR-related specifications. Unlike typical primer/probe design where sequences are constrained by an existing sequence of interest, here there is almost complete freedom to design primer/probe sequences that have ideal PCR properties. Using an approximate melting temperature (T_m) formula (Eq. 1):

$$T_m = 69.4 + \frac{41 * (n_{GC} - 16.4)}{n_{total}} \tag{1}$$

where n_{GC} is the number of G and C bases and n_{total} is the length of the primer/probe. The number of A's+T's and G's+C's needed to satisfy the specified T_m value can be calculated for primer/probe sequences within user-adjustable constraints on length and G+C content. From this set, a sequence meeting these constraints is randomly

generated. For probe sequences generated containing more G's than C's, the complement sequence is used. All primer/probe sequences are subsequently screened for several PCR-related properties, such as maximum homopolymer repeats and secondary structure (Table 1). If any requirements are not met, the sequence is rejected and a new sequence generated.

A sequence that meets PCR restrictions is then tested for uniqueness. First, the sequence is compared to a list of other primer/probe sequences, which includes any sequences already generated locally by the algorithm and an optional user-provided list of other primer/probe sets of interest. Sequence “matches” are determined by comparing the number of base pair matches of the top BLAST result divided by the length of the current primer to a user-adjustable threshold (Table 1). If the sequence matches any existing primer/probe sequences above the threshold, the sequence is discarded and the process restarted. Second, the genome of the organism targeted for insertion is scanned for similar sequences by BLAST. Similarly, a set of additional genome sequences of organisms that may be likely to be present in a sample, such as common environmental

Table 1 Algorithm parameters and default values for primer/probe design

Purpose	Description of constraint or check	Default parameter values ^a		
		Forward primer	Reverse primer	Probe
Primer/probe Constraints	Minimum length (bases)	20	20	20
	Maximum length (bases)	24	24	30
	Minimum T_m^b (°C)	58	58	68
	Maximum T_m^b (°C)	60	60	70
	Minimum G+C Content (%)	40	40	40
	Maximum G+C Content (%)	60	60	60
	Number of C's \geq G's in probe	n/a	n/a	True
Primer/probe Checks	Maximum G Homopolymer Length (bases)	3	3	3
	Maximum A/T/C Homopolymer Length (bases)	4	4	4
	No G on 5' end of probe	n/a	n/a	True
	2 G/C's in last 5 bases of 3' end of primers (GC clamp)	True	True	n/a
	No start codons present	True	True	True
Primer/probe Stem-loop Checks	Minimum stem-loop hydrogen bonds	14	14	14
	Minimum stem-loop palindrome length (bases)	5	5	5
	Maximum stem-loop palindrome length (bases)	100	100	100
	Maximum stem-loop gap size (bases)	11	11	11
	Maximum stem-loop mis-matches	1	1	1
BLAST checks	Threshold ^c to reject primer/probe based on BLAST similarity	0.85	0.85	0.85

^a Entries of “n/a” denote that a constraint or check is not made. Entries of “True” denote checks that are made but do not have a user-customizable parameter and consequently no default value

^b T_m is the melting temperature of the primer/probe

^c See text for explanation of threshold calculation

background species or human DNA, are scanned. Finally, the entire NCBI database is optionally scanned for similar sequences. The threshold for discarding a candidate sequence based on these BLAST results can be customized by the user, allowing more or less stringent criteria depending on project demands and acceptable CPU time in the case of very strict thresholds.

A sequence that meets all PCR and uniqueness requirements is accepted for use in the barcode. The algorithm cycles through this process to create each primer and the probe sequence, each with its own set of requirement parameters. Optionally, the forward primer can be set as constant for all barcodes in a given project. Once all three primer/probe sequences for a barcode have been generated, two spacer sequences are randomly generated such that the user-defined distance between the forward and reverse primers is met and the G+C content of the full barcode matches the G+C content of the target organism. The full barcode is then assembled by concatenating the forward primer sequence, first spacer sequence, probe sequence, second spacer sequence, and the reverse complement of the reverse primer sequence. The final check scans for potential stem-loop structures in the barcode to limit challenges during genome insertion and during amplification of the sequence. Failing this check triggers regeneration of the spacer sequences (Table 2).

Experimental validation

The barCoder algorithm was used to generate an initial set of 21 barcodes and corresponding qPCR detection primer/probe sets (sequences listed in Tables S1 and S2 in Additional file 1). Due to code errors that were discovered and corrected after experimental validation was completed, these 21 assays were generated with less stringent checks than intended. Nonetheless, we report that all 21 assays were functional and anticipate that assays generated with the intended parameters would perform as well as or better than those reported here. Tables S3–S6 in Additional file 1 shows the actual properties of the 21 assays compared to what is implemented in the published version of the algorithm. Twelve of these barcodes were designed for *B. thuringiensis* serovar *kurstaki* (Btk), a surrogate for the biothreat agent *B. anthracis* with low-G+C content (35%, ref. [34]). To demonstrate the utility of the barCoder algorithm to create barcodes for other organisms, including those with different G+C compositions, three barcodes each were designed for potential use in *Burkholderia pseudomallei* 1026b (Bp; 68% G+C content), *Yersinia pestis* CO92 (Yp; 48%), and *Clostridium botulinum* Hall A (Cbot; 28%) [35–37].

Table 2 Algorithm parameters and default values other than for primer/probe design

Purpose	Parameter description	Default values
Spacer size	Total spacer size (bases between forward/reverse primers, inclusive of probe but not of forward/reverse primers)	100
Full barcode Module stem-loops	Minimum stem-loop hydrogen bonds	14
	Minimum stem-loop palindrome length (bases)	10
	Maximum stem-loop palindrome length (bases)	80
	Maximum stem-loop gap size (bases)	11
	Maximum stem-loop mismatches	1

Assay conditions for barcode Btk1 in the pIDTSMART-AMP plasmid backbone were optimized and subsequently standard curves were generated for all 21 TaqMan[®] qPCR assays using the same conditions (Fig. 2 and Fig. S1 in Additional file 1; Ct values and calculations provided in Additional file 2). All of the assays of the barcodes in plasmids performed well with qPCR efficiencies ranging from 81.1% to 100.0%, strong linear relationships ($R^2 > 0.99$), and no false positive results (Table 3). Limits of detection (LODs) were all below 50 copies (the lowest plasmid concentration tested), except for barcode Btk6, where the LOD was below 500 copies (Table 3). Select barcodes were also markerlessly incorporated into the chromosomes of potential target organisms: barcode Btk1 was integrated into both Btk and *B. anthracis* Sterne, and barcode Yp1 was inserted into a *pgm*⁻ derivative of *Y. pestis* CO92. Again, standard curves were generated for the TaqMan[®] assays under the same conditions (Fig. 2; Ct values and calculations provided in Additional file 2). Assays using chromosomally-barcoded strains had efficiencies within the range observed for barcodes residing in plasmids (86.5% to 96.5%), R^2 values above 0.99, and no false positives (Table 3). LODs were calculated as less than 15 copies and less than 2 copies for barcode Btk1 in the chromosomes of Btk and *B. anthracis* Sterne, respectively, and less than 25 copies for barcode Yp1 in the chromosome of *Y. pestis* CO92 *pgm*⁻ (Table 3). LODs are approximate as lower concentrations and Poisson distribution effects at low copy numbers were not thoroughly interrogated.

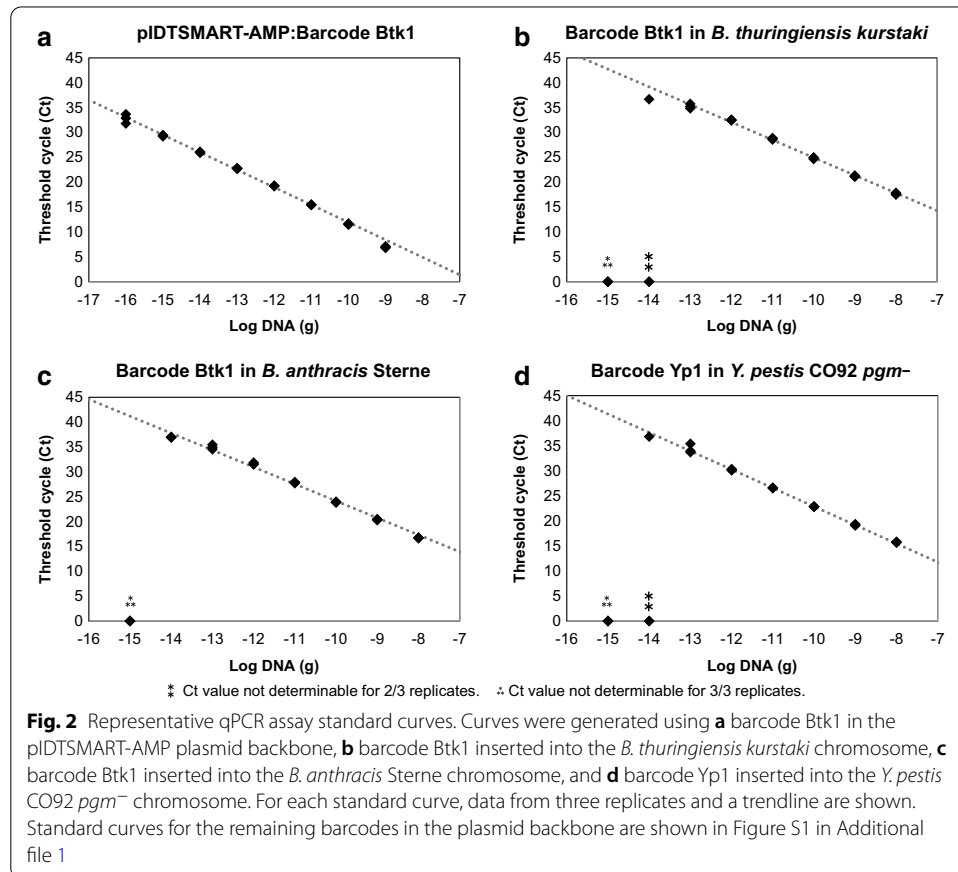


Table 3 Evaluation of qPCR assays from generated standard curves

Barcode	Template DNA	Linearity (R ²)	LOD (Copies)	False + 's	Efficiency (%)
Btk1	Plasmid	0.9990	< 50	0/3	92.5
	Genomic (Btk)	0.9969	< 15	0/3	91.1
	Genomic (Ba Sterne)	0.9919	< 2	0/3	96.5
Btk2	Plasmid	0.9984	< 50	0/3	83.5
Btk3	Plasmid	0.9991	< 50	0/3	98.9
Btk4	Plasmid	0.9976	< 50	0/3	94.3
Btk5	Plasmid	0.9978	< 50	0/3	89.9
Btk6	Plasmid	0.9996	< 500	0/3	90.4
Btk7	Plasmid	0.9993	< 50	0/3	84.0
Btk8	Plasmid	0.9979	< 50	0/3	89.8
Btk9	Plasmid	0.9994	< 50	0/3	81.1
Btk10	Plasmid	0.9985	< 50	0/3	89.9
Btk11	Plasmid	0.9990	< 50	0/3	90.6
Btk12	Plasmid	0.9985	< 50	0/3	82.5
Bp1	Plasmid	0.9996	< 50	0/3	93.0
Bp2	Plasmid	0.9996	< 50	0/3	100.0
Bp3	Plasmid	0.9992	< 50	0/3	92.4
Cbot1	Plasmid	0.9983	< 50	0/3	95.1
Cbot2	Plasmid	0.9982	< 50	0/3	91.6
Cbot3	Plasmid	0.9990	< 50	0/3	86.1
Yp1	Plasmid	0.9988	< 50	0/3	94.6
	Genomic	1.0000	< 25	0/3	86.5
Yp2	Plasmid	0.9979	< 50	0/3	96.8
Yp3	Plasmid	0.9997	< 50	0/3	95.0

To test the specificity of the TaqMan[®] qPCR assays for the corresponding barcode, each of the 12 Btk assays were tested against all 12 Btk barcodes in plasmids. This cross-reactivity panel showed unique amplification of each Btk barcode with its cognate primer/probe set (Fig. 3; raw data provided in Additional file 3). The TaqMan[®] qPCR assay for barcode Btk1 was also tested against a panel of potential pathogens and environmental organisms (Table 4; raw data provided in Additional file 4). Reactions containing the Btk strain with barcode Btk1 inserted in the chromosome, either alone or in the presence of an environmental matrix (DNA from a mock microbial community or DNA extracted from soil) showed robust positive results, while the Btk1 qPCR assay did not cross-react with any of the potential contaminants.

Discussion

qPCR has become a standard technique for detection of microorganisms in the environment and for diagnosis of infection [38], and, as such, is an attractive detection technology that also allows a rapid evaluation of the relative abundance of a known microorganism within a sample. However, when conducting environmental fate studies, for example, these assays must discriminate from the endogenous or native microflora, which may be uncharacterized and present signatures similar to or cross-reactive with the signature selected for detection of the experimental strain. We sought in this report to utilize a bioinformatics strategy to generate specific amplicons that require minimal

		DNA Template											
		1	2	3	4	5	6	7	8	9	10	11	12
Assay	1	23.0	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
	2	ND	24.5	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
	3	ND	ND	22.5	ND	ND	ND	ND	ND	ND	ND	ND	ND
	4	ND	ND	ND	23.4	ND	ND	ND	ND	ND	ND	ND	ND
	5	ND	ND	ND	ND	24.5	ND	ND	ND	ND	ND	ND	ND
	6	ND	ND	ND	ND	ND	23.9	ND	ND	ND	ND	ND	ND
	7	ND	ND	ND	ND	ND	ND	24.0	ND	ND	ND	ND	ND
	8	ND	ND	ND	ND	ND	ND	ND	22.7	ND	ND	ND	ND
	9	ND	ND	ND	ND	ND	ND	ND	ND	20.4	ND	ND	ND
	10	ND	ND	ND	ND	ND	ND	ND	ND	ND	19.3	ND ^a	ND
	11	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	18.7	ND
	12	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	19.6

^a1/3 replicates gave a Ct value of 37.087.

Fig. 3 Cross-reactivity of the 12 Btk qPCR assays against the 12 Btk barcodes. For each qPCR reaction, 10^{-12} g (~450,000 copies) of the pIDTSMART-AMP plasmid backbone containing the Btk barcode indicated was used as DNA template. Threshold cycle (Ct) values shown are the median of three replicates; ND, Not determinable

assay optimization and could be introduced into organisms with minimal to no cross-reactivity with environmental and microbial signatures.

Our approach to developing unique qPCR-compatible barcodes expanded upon our previous work, in which we appropriated synthetic signatures from published microarrays and developed PCR assays based on the unique sequences generated both by the tags themselves and by their insertion into the genome [24]. Because those sequences were not designed de novo for use in PCR detection assays, we relied on the presence of a chromosomal primer binding site and the single synthetic sequence to generate suitable amplicons. As a result, considerable optimization of the assay conditions and primer sequences was necessary during the development of those strains, and the assay conditions for each tag required slightly different optimal conditions for detection. This situation was judged as suboptimal for the development of a more diverse panel of barcodes, as new assays would need to be developed for each new sequence.

We therefore sought to develop an algorithm that would enable the high-throughput generation of amplicon sequences that could use a single PCR assay condition, and in which relative proportion of each strain could be compared in a single test, e.g. across a single microplate. The assays would need to be specific to each barcode, and would need to be comparably sensitive with equivalent limits of detection. Using TaqMan[®] qPCR chemistry and a bioinformatic screening algorithm, we generated a panel of unique primer/probe combinations that exhibited selectivity, specificity, and sensitivity. Using conventional plasmids containing the barcodes as templates for the development of the assays, we demonstrate strong performance in linearity of response, sensitivity, and efficiency across 21 assays using conditions optimized for a single assay. No cross-reactivity was observed across a panel of 12 of these assays. We note that the odds of randomly generating a barcode that would react with a natural sequence is vanishingly small, as

Table 4 Cross-reactivity of the barcode Btk1 qPCR assay against a panel of potential pathogens and environmental contaminants

DNA template ^a	Ct ^b
<i>Bacillus anthracis</i> Ames35	ND
<i>Bacillus anthracis</i> Sterne 34F2	ND
<i>Bacillus cereus</i> Gibson 971	ND
<i>Bacillus licheniformis</i> Gibson 46	ND
<i>Bacillus megaterium</i> Ford 19	ND
<i>Bacillus sphaericus</i> Ford 25	ND
<i>Bacillus subtilis</i> subsp. <i>subtilis</i> 168	ND
<i>Bacillus thuringiensis</i> serovar <i>kurstaki</i> HD-1	ND
<i>Bacillus thuringiensis</i> subsp. <i>konkukian</i> 97-27	ND
<i>Burkholderia pseudomallei</i> 1026b	ND
<i>Clostridium perfringens</i> WAL-14572	ND
<i>Escherichia coli</i> EDL933	ND
<i>Francisella tularensis</i> subsp. <i>tularensis</i> SCHU S4	ND
<i>Micrococcus luteus</i> SK58	ND
<i>Neisseria meningitidis</i> 9506	ND
<i>Pseudomonas</i> sp. 2_1_26	ND
<i>Salmonella enterica</i> subsp. <i>enterica</i> LT2	ND
<i>Staphylococcus aureus</i> TCH1516	ND
<i>Staphylococcus epidermidis</i> SK135	ND
<i>Streptococcus pneumoniae</i> TCH8431	ND
<i>Vibrio cholerae</i> 395	ND
<i>Yersinia pestis</i> CO92	ND
Barcode Btk1 in <i>Bacillus thuringiensis</i> serovar <i>kurstaki</i> HD-1	23.927
Microbial Mock Community B (Even, High Concentration)	ND
Microbial Mock Community B (Even, High Concentration) + Barcode Btk1 in <i>Bacillus thuringiensis</i> serovar <i>kurstaki</i> HD-1	23.725
Soil DNA extract	ND
Soil DNA extract + Barcode Btk1 in <i>Bacillus thuringiensis</i> serovar <i>kurstaki</i> HD-1	23.969
Negative control (H ₂ O)	ND

ND, not determinable

^a qPCR reactions contained 1 ng of each DNA template indicated

^b Threshold cycle (Ct) values are the median of three replicates

three 20+ base primer/probe sequences would need to be closely matched in the correct orientation (> 10³⁶ possible sequences) with spacing appropriate for PCR amplification; nonetheless, sequences are screened for uniqueness to further minimize this possibility. Inserting two of the barcodes into three different genomes, we observed conserved performance compared to plasmid assays and LODs below 25 copies, which we believe to be conservative due to Poisson distribution effects at low copy numbers. We note that the code used to generate the barcodes contained errors that resulted in application of less stringent thresholds for primer and probe design, yet still resulted in highly sensitive and specific assays. The remediated code should therefore reliably produce sequences that function equally well if not better than the barcodes validated here while reducing the possibility of generating suboptimal primer sequences.

Our barcodes have a number of potential applications. Marking strains with unique artificial signatures could aid in protecting intellectual property, particularly for

production strains whose development has required significant investment in metabolic and/or genetic optimization, perhaps in combination with other techniques such as DNA steganography [39]. While not as information-rich as longer steganographic tags or watermarks [40, 41], qPCR barcodes have the advantage of not requiring further sequencing and informatic analysis to detect and/or verify their presence; they must simply be amplified using appropriate primers and probes. In one scenario, a set of barcodes could be inserted at defined intervals throughout a large DNA molecule used for information storage, and utilized to provide a preliminary indicator of the stability of the archive prior to full sequencing.

These sequences and their associated assays might also find use in forensic applications. In particular, one might imagine their use as molecular taggants that could be spiked into samples by field technicians, and their detection in DNA samples by the reference laboratories would serve to verify the origin of the sample. In a similar vein, these same tools could be used in the future for downstream attribution of accidental or deliberate release of organisms [42]. Select agent strains, in particular, could be tagged, distributed to end-user communities, and then any material from the scene of a biocrime could be rapidly amplified using the library of primers and probes. Investigators could then focus investigative resources on those potential sources, while excluding the majority of the research laboratories that possess variants containing other barcodes. Any mechanism by which artificial genetic diversity can be introduced into the largely clonal populations of laboratory strains would be useful as all known acts of bioterrorism to date have utilized common laboratory strains (e.g. *B. anthracis* Ames Ancestor in the 2001 U.S. Mail/Amerithrax case; *S. enterica* serovar *typhimurium* 14028s in the Rajneeshi cult attacks of 1984 [43]). In the case of the Amerithrax samples, discriminating between samples present in these laboratories relied on presence of several spontaneous mutants whose discovery and characterization required astute microbiologists and what were at the time Herculean sequencing efforts [44, 45]. The deliberate incorporation of end-user-specific sequences into such commonly available strains could immeasurably speed identification of potential originating laboratories, would help investigators narrow their focus to a subset of potential sources, and would help exclude uninvolved laboratories working on similar research as potential sources. Furthermore, the presence of such signatures (and the knowledge that significant additional effort would be required to disguise the source of a sample) could deter potential malefactors within those laboratories even if the location, sequence, and properties of the sequence were known.

Conclusions

To our knowledge, barCoder represents the first completely in silico method for generating both a synthetic target for qPCR and the primers/probe to amplify the target, and optimal assay conditions for detection of a diverse range of barcodes. We demonstrated that generated barcodes all perform well under a single set of assay conditions and show no cross-reactivity with themselves or environmental contaminants. Insertion of the barcodes into the genomes of three organisms of interest maintained the key properties

of the barcodes. We anticipate barCoder finding utility in applications such as environmental fate studies, intellectual property, and microbial forensics.

Methods

Algorithm implementation

All software was written in Perl. G+C and A+T constraints were calculated using Eq. 1 (see above). The algorithm workflow is described in “Algorithm design” (see Results, above) and is shown in Fig. 1b. Most bioinformatics functions were implemented using existing BioPerl modules. EMBOSS software, called by BioPerl, was used to predict stem-loop structures. All BLAST runs used default BioPerl parameters. Software is available on GitHub (<https://github.com/ECBCgit/Barcoder>).

barCoder-designed elements and sources of DNA

The properties of the barcode modules designed in this study using an earlier version of the barCoder algorithm are listed in Tables S3–S6 in Additional file 1. For the BLAST step against the target organism genome sequence (Table S5 in Additional file 1), the following NCBI RefSeq accession numbers were used: *B. thuringiensis* serovar *kurstaki*, NZ_CP010005.1; *B. pseudomallei* 1026b, NC_017831.1 (chromosome 1) and NC_017832.1 (chromosome 2); *C. botulinum* Hall A, NC_009495.1; *Y. pestis* CO92, NC_003143.1. All barcodes, primers, and probes were obtained from Integrated DNA Technologies, Inc. (IDT, Coralville, IA) and their sequences listed in Tables S1 and S2 in Additional file 1. SacI and NheI restriction sites flanking the Btk barcodes were added to facilitate later subcloning. Barcodes were received as “minigenes” inserted in the pIDTSMART-AMP plasmid backbone and were propagated in NEB[®] 5-alpha *E. coli* (New England Biolabs, Inc., Ipswich, MA) on Luria–Bertani (LB) agar and in LB broth with 100 µg/ml ampicillin at 37 °C. Plasmid DNA was isolated using the QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany). DNA probes were ordered as PrimeTime double-quenched qPCR probes containing the 5′ FAM fluorophore, 3′ Iowa Black FQ quencher, and internal ZEN quencher. The sources of the DNA used for the cross-reactivity panel of pathogenic and environmental organisms are given in Table S7 in Additional file 1.

qPCR

All qPCR experiments were run on an Applied Biosystems 7900HT Real-Time PCR System (Applied Biosystems, Foster City, CA) using Applied Biosystems MicroAmp optical 384-well reaction plates (catalog number 4309849) sealed with Applied Biosystems MicroAmp optical adhesive film (catalog number 4311971). Optimized 20 µL reactions included Applied Biosystems TaqMan[®] Universal PCR Master Mix (catalog number 4304437), forward and reverse primers each at a final concentration of 900 nM, DNA probe at a final concentration of 250 nM, 1 µL DNA template at the indicated concentration, and nuclease-free water. TaqMan[®] assays used the following thermocycler protocol: 1 cycle of 50 °C for 2 min, 1 cycle of 95 °C for 10 min, and 40 cycles of 95 °C for 15 s and 55 °C for 1 min. The standard curve properties of each assay were assessed by performing tenfold serial dilutions of the template DNA in nuclease-free water. Efficiency

and linearity (R^2) values for each qPCR standard curve were calculated using the median Ct of three replicates for each template DNA dilution. Data points corresponding to the highest amount of template DNA tested (10^{-8} g for genomic DNA, 10^{-9} g for plasmid DNA) were omitted from these analyses in all cases as the Ct values tended to be non-linear with the other data points of the standard curve. LODs were conservatively estimated using the lowest amount of template DNA tested that produced a Ct value < 40 for all three replicates. Ct values and calculations used to generate and analyze all standard curves for the 21 barcodes in this study are provided in Additional file 2.

Construction of genomically-barcoded strains

B. thuringiensis and *B. anthracis* strains were routinely cultured on Brain Heart Infusion (BHI) agar and in BHI broth at 30 °C (*B. thuringiensis*) or 37 °C (*B. anthracis*). Unless otherwise indicated, *Y. pestis* strains were grown on BHI agar or Tryptic Soy Agar (TSA), and in BHI broth at 28–30 °C. Genomic DNA was extracted using the UltraClean Microbial DNA Isolation Kit (MOBIO Laboratories, Inc., Carlsbad, CA). Barcode Btk1 was selected to construct a strain in which the barcode was markerlessly incorporated into the chromosome of *B. thuringiensis* serovar *kurstaki* HD-1 [24, 46] (obtained from the DoD Unified Culture Collection (<https://www.usamriid.army.mil/ucc/>)). The insertion was generated at the same locus that was identified and modified in our previous report (within Target 1, [24]). This corresponds to an insertion between positions 4,834,064 and 4,834,065 of NCBI RefSeq accession number NZ_CP010005.1. Plasmid pRP1028-T1-PL (GenBank accession number MW055899), a derivative of pRP1028 [47], was designed specifically for incorporating synthetic elements within this target region of the Btk chromosome and was synthesized by DNA2.0 (Menlo Park, CA). Plasmid pRP1028-T1-PL contains 1,550 bp of DNA homologous to the Btk chromosomal insertion region between the pRP1028 HindIII and BamHI sites, as well as a 36-bp polylinker within the homology region. Following digestion of plasmid pIDTSMART-AMP:Barcode Btk1 with SacI and NheI, the Btk1 barcode was gel extracted (QIAquick Gel Extraction Kit, Qiagen, Hilden, Germany) and ligated with pRP1028-T1-PL that had been digested with the same restriction enzymes. This pRP1028-T1-PL derivative containing barcode Btk1 was introduced into Btk, and the barcode was incorporated into the chromosome using the markerless allelic exchange strategy described previously [47]. Successful barcode integration into the Btk chromosome was verified by PCR amplification of the target locus and SacI/NheI digestion of the resulting amplicon. Construction of a strain of *B. anthracis* Sterne 34F2 with barcode Btk1 in the chromosome was previously published [48].

For construction of a strain of *Y. pestis* CO92 *pgm*⁻ with barcode Yp1 markerlessly inserted in the chromosome, the locus between the convergently transcribed genes YPO0388 and YPO0392a (NCBI RefSeq accession number NC_003143.1) was selected using rules adopted from Buckley et al. [24], in combination with the PATRIC database [49] and available transcriptome sequencing (RNA-seq) data (NCBI SRA accession numbers SRR1013703, SRR1013704, SRR1013705, SRR1041589), to identify a potentially neutral insertion region. The barcode was then inserted into the chromosome between positions 406,742 and 406,743 via the method described by Sun et al. [50], which utilizes λ Red recombination and *sacB* counterselection. Briefly, plasmid pKD46 (CGSC #7739, [50]) containing the genes for λ Red recombination was

electroporated into a strain of *Y. pestis* CO92 *pgm*⁻ (strain R88, Robert Perry, University of Kentucky). A linear DNA fragment containing a *cat-sacB* cassette flanked by DNA homologous to the *Y. pestis* chromosomal insertion region was electroporated into this pKD46-containing strain of *Y. pestis*, and successful integrants were selected on media containing chloramphenicol. Following electroporation with a linear DNA fragment containing barcode Yp1 flanked by homologous DNA and selection on media containing sucrose, the *cat-sacB* cassette in the chromosome was replaced with the barcode. The resulting strain was subsequently cured of pKD46, and successful barcode insertion was verified by PCR amplification and sequencing. Whole-genome sequencing (MiSeq, Illumina) was also performed to confirm the absence of off-target modifications. Primers used to construct this barcoded strain of *Y. pestis* are listed in Table S8 in Additional file 1. To generate the *cat-sacB* cassette, the *cat* gene was PCR amplified from plasmid pKD3 (CGSC #7631, [51]) with primers #1 and #2, and the *sacB* gene was PCR amplified from plasmid p88171 (synthesized plasmid with pJ207 backbone and *sacB* gene from *Bacillus subtilis*, DNA 2.0, Menlo Park, CA) with primers #3 and #4. The two PCR amplicons were purified (QIAquick PCR Purification Kit, Qiagen, Hilden, Germany) and joined together by overlap extension PCR [52] using primers #1 and #4. The *cat-sacB* cassette was gel extracted and cloned between the SacI and BamHI sites of pUC19 to create plasmid pCBV4. To construct the *cat-sacB* cassette flanked by homologous DNA, the *cat-sacB* cassette was PCR amplified from pCBV4 with primers #5 and #6. Approximately 500 bp flanking each side of the barcode insertion point were separately PCR amplified from the *Y. pestis* CO92 *pgm*⁻ chromosome; primers #7 and #8 were used to amplify upstream DNA, and primers #9 and #10 were used to amplify downstream DNA. The three purified PCR amplicons (up flanking region, *cat-sacB* cassette, and down flanking region) were joined together by overlap extension PCR [52] using primers #7 and #10, and the resulting amplicon was gel extracted and cloned into the pCRTM4Blunt-TOPO[®] vector (Invitrogen, Carlsbad, CA) to generate plasmid pCBV6. The linear DNA fragment containing the *cat-sacB* cassette flanked on both sides by *Y. pestis* CO92 *pgm*⁻ DNA was PCR amplified from pCBV6 with primers #7 and #10. To create barcode Yp1 flanked by homologous DNA, the barcode was PCR amplified from the synthesized plasmid pIDTSMART-AMP:Barcode Yp1 using primers #11 and #12. Approximately 500 bp flanking each side of the barcode insertion point were separately PCR amplified from the *Y. pestis* CO92 *pgm*⁻ chromosome; primers #7 and #13 were used to amplify upstream DNA, and primers #10 and #14 were used to amplify downstream DNA. The three purified PCR amplicons (up flanking region, barcode, and down flanking region) were joined by overlap extension PCR [52] using primers #7 and #10, and the resulting amplicon was gel extracted and cloned into the pCRTM4Blunt-TOPO[®] vector (Invitrogen, Carlsbad, CA) to generate plasmid pCBV9. The linear DNA fragment containing barcode Yp1 flanked on both sides by *Y. pestis* CO92 *pgm*⁻ DNA was PCR amplified from pCBV9 with primers #7 and #10.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-021-04019-5>.

Additional file 1: Tables S1. Sequences for the 21 barcodes designed with the barCoder algorithm. **Table S2.** Primer and probe sequences for each barcode module. **Table S3.** Details the constraint properties of the primers and probes that were experimentally validated in this study. **Table S4.** Details the check results for properties of the primers and probes that were experimentally validated in this study. **Table S5.** Details the BLAST check results for the primers and probes that were experimentally validated in this study. **Table S6.** Details the spacer and barcode module checks for the barcodes experimentally validated in this study. **Table S7.** Sources of DNA used in the cross-reactivity panel shown in Table 4. **Table S8.** Sequences of the primers used to construct the barcoded strain of *Y. pestis* CO92 pgm^- . **Figure S1.** Additional qPCR standard curves.

Additional file 2. Standard curve data. This file contains Ct values and calculations used to generate and analyze qPCR standard curves for all 21 barcodes (in the plasmid backbone and inserted in the genome, if applicable).

Additional file 3. Raw data for the cross-reactivity panel of Btk barcodes and qPCR assays. This file contains raw Ct values from the cross-reactivity panel of the 12 Btk qPCR assays against the 12 Btk barcodes.

Additional file 4. Raw data for the Btk1 qPCR assay cross-reactivity panel. This file contains the raw Ct values from the cross-reactivity panel of the barcode Btk1 qPCR assay against a panel of potential pathogens and environmental contaminants.

Abbreviations

Btk: *Bacillus thuringiensis* Serovar *kurstaki*; Bp: *Burkholderia pseudomallei* 1026B; Cbot: *Clostridium botulinum* Hall A; LOD: Limit of detection; qPCR: Quantitative real-time polymerase chain reaction; Yp: *Yersinia pestis* CO92.

Acknowledgements

We thank F. Chris Minion (Iowa State University) for the generous gift of *Y. pestis* strain R88. We also thank Mark Karavis (CCDC CBC) for whole-genome sequencing, and Pierce Roth, Edward Fochler (DCS Corp/CCDC CBC), and Michael Krepps (BrightEdge Investments) for bioinformatics support.

Authors' contributions

CBB inserted barcodes into *B. anthracis* and *Y. pestis*, designed and performed qPCR experiments, analyzed data, and wrote the manuscript. MWL developed the algorithm, designed qPCR experiments, analyzed data, and wrote the manuscript. SEK optimized qPCR protocols. TDPG inserted barcode Btk1 into Btk. ATL updated the algorithm for publication. HSG conceived and led the project, analyzed data, and wrote the manuscript. All authors read and approved the final manuscript.

Funding

Funding for this work was provided by the Defense Threat Reduction Agency under project CB3654 and by the Defense Threat Reduction Agency/National Research Council Postdoctoral Research Associateship Program (to CBB). The funders had no role in the design or execution of the study. This work is approved for public release. The opinions expressed in this report are those of the authors and do not represent official policy of the United States Government or any of its agencies.

Availability of data and materials

The barCoder software is available on GitHub (<https://github.com/ECBCgit/Barcoder>). All raw data to the level of Ct values that were generated and analyzed during this study are included in the article and its additional files. The sequence for pRP1028-T1-PL, the plasmid designed for incorporating synthetic elements within the target region of the Btk chromosome, is available at GenBank accession number MW055899. The genome sequences utilized in the current study are available in the NCBI Reference Sequence database at accession numbers NZ_CP010005.1 (*B. thuringiensis* serovar *kurstaki*), NC_017831.1 (*B. pseudomallei* 1026b chromosome 1), NC_017832.1 (*B. pseudomallei* 1026b chromosome 2), NC_009495.1 (*C. botulinum* Hall A), and NC_003143.1 (*Y. pestis* CO92). The transcriptome sequencing datasets utilized in the current study are available in the NCBI Sequence Read Archive repository at accession numbers SRR1013703, SRR1013704, SRR1013705, and SRR1041589.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ U.S. Army Combat Capabilities Development Command Chemical Biological Center, Aberdeen Proving Ground, MD 21010, USA. ² Excet, Inc., Springfield, VA 22150, USA. ³ DCS Corporation, Abingdon, MD 21009, USA.

Received: 6 January 2020 Accepted: 11 February 2021

Published online: 01 March 2021

References

1. Ropert-Coudert Y, Wilson RP. Trends and perspectives in animal-attached remote sensing. *Front Ecol Environ*. 2005;3(8):437–44.
2. Gibbons WJ, Andrews KM. PIT tagging: simple technology at its best. *Bioscience*. 2004;54(5):447–54.
3. Glandorf DC, van der Sluis I, Anderson AJ, Bakker PA, Schippers B. Agglutination, adherence, and root colonization by fluorescent pseudomonads. *Appl Environ Microbiol*. 1994;60(6):1726–33.
4. Glandorf DC, Verheggen P, Jansen T, Jorritsma JW, Smit E, Leeftang P, et al. Effect of genetically modified *Pseudomonas putida* WCS358r on the fungal rhizosphere microflora of field-grown wheat. *Appl Environ Microbiol*. 2001;67(8):3371–8.
5. De Leij F, Sutton EJ, Whipps JM, Fenlon JS, Lynch JM. Impact of field release of genetically modified *Pseudomonas fluorescens* on indigenous microbial populations of wheat. *Appl Environ Microbiol*. 1995;61(9):3443–53.
6. Krishnamurthy K, Gnanamanickam SS. Biological control of rice blast by *Pseudomonas fluorescens* Strain Pf7–14: evaluation of a marker gene and formulations. *Biol Control*. 1998;13(3):158–65.
7. Moenne-Loccoz Y, Powell J, Higgins P, McCarthy J, O'Gara F. An investigation of the impact of biocontrol *Pseudomonas fluorescens* F113 on the growth of sugarbeet and the performance of subsequent clover-Rhizobium symbiosis. *Appl Soil Ecol*. 1998;7(3):225–37.
8. De Leij F, Thomas CE, Bailey MJ, Whipps JM, Lynch JM. Effect of insertion site and metabolic load on the environmental fitness of a genetically modified *Pseudomonas fluorescens* isolate. *Appl Environ Microbiol*. 1998;64(7):2634–8.
9. Hensel M, Shea J, Gleeson C, Jones M, Dalton E, Holden D. Simultaneous identification of bacterial virulence genes by negative selection. *Science*. 1995;269(5222):400–3.
10. Mazurkiewicz P, Tang CM, Boone C, Holden DW. Signature-tagged mutagenesis: barcoding mutants for genome-wide screens. *Nat Rev Genet*. 2006;7(12):929–39.
11. Grant AJ, Restif O, McKinley TJ, Sheppard M, Maskell DJ, Mastroeni P. Modelling within-host spatiotemporal dynamics of invasive bacterial disease. *PLoS Biol*. 2008;6(4):e74.
12. Mastroeni P, Grant A. Dynamics of spread of *Salmonella enterica* in the systemic compartment. *Microbes Infect*. 2013;15(13):849–57.
13. Varble A, Albrecht RA, Backes S, Crumiller M, Bouvier NM, Sachs D, et al. Influenza A virus transmission bottlenecks are defined by infection route and recipient host. *Cell Host Microbe*. 2014;16(5):691–700.
14. Lauring AS, Andino R. Exploring the fitness landscape of an RNA virus by using a universal barcode microarray. *J Virol*. 2011;85(8):3780–91.
15. Melton-Witt JA, McKay SL, Portnoy DA. Development of a single-gene, signature-tag-based approach in combination with alanine mutagenesis to identify listeriolysin O residues critical for the in vivo survival of *Listeria monocytogenes*. *Infect Immun*. 2012;80(6):2221–30.
16. Melton-Witt JA, Rafelski SM, Portnoy DA, Bakardjiev AI. Oral infection with signature-tagged *Listeria monocytogenes* reveals organ-specific growth and dissemination routes in guinea pigs. *Infect Immun*. 2012;80(2):720–32.
17. Rego RO, Bestor A, Stefka J, Rosa PA. Population bottlenecks during the infectious cycle of the Lyme disease spirochete *Borrelia burgdorferi*. *PLoS ONE*. 2014;9(6):e101009.
18. Gerlini A, Colomba L, Furi L, Braccini T, Manso AS, Pammolli A, et al. The role of host and microbial factors in the pathogenesis of pneumococcal bacteraemia arising from a single bacterial cell bottleneck. *PLoS Pathog*. 2014;10(3):e1004026.
19. Martin CJ, Cadena AM, Leung VW, Lin PL, Maiello P, Hicks N, et al. Digitally barcoding *Mycobacterium tuberculosis* reveals in vivo infection dynamics in the macaque model of tuberculosis. *MBio*. 2017;8(3).
20. Abel S, Abel zur Wiesch P, Davis BM, Waldor MK. Analysis of bottlenecks in experimental models of infection. *PLoS Pathog*. 2015;11(6):e1004823.
21. Leitenberg M, Zilinskas RA. *The Soviet Biological Weapons Program: a History*. Cambridge, MA: Harvard University Press; 2012. p. 960.
22. Greenberg DL, Busch JD, Keim P, Wagner DM. Identifying experimental surrogates for *Bacillus anthracis* spores: a review. *Investig Genet*. 2010;1(1):4.
23. Emanuel PA, Buckley PE, Sutton TA, Edmonds JM, Bailey AM, Rivers BA, et al. Detection and tracking of a novel genetically tagged biological simulant in the environment. *Appl Environ Microbiol*. 2012;78(23):8281–8.
24. Buckley P, Rivers B, Katoski S, Kim MH, Kragl FJ, Broomall S, et al. Genetic barcodes for improved environmental tracking of an anthrax simulant. *Appl Environ Microbiol*. 2012;78(23):8272–80.
25. Bishop AH, Stapleton HL. Aerosol and surface deposition characteristics of two surrogates for *Bacillus anthracis* spores. *Appl Environ Microbiol*. 2016;82(22):6682–90.
26. Bishop AH, Robinson CV. *Bacillus thuringiensis* HD-1 Cry-: development of a safe, non-insecticidal simulant for *Bacillus anthracis*. *J Appl Microbiol*. 2014;117(3):654–62.
27. Merrill L, Dunbar J, Richardson J, Kuske CR. Composition of bacillus species in aerosols from 11 US cities. *J Forensic Sci*. 2006;51(3):559–65.
28. Van Cuyk S, Duval N, Ombreg KM. *Bacillus thuringiensis*: presence and Persistence in the Environment. Los Alamos National Laboratory, 2008 LA-UR-08-04878.
29. Crickmore N. Beyond the spore—past and future developments of *Bacillus thuringiensis* as a biopesticide. *J Appl Microbiol*. 2006;101(3):616–9.
30. Van Cuyk S, Deshpande A, Hollander A, Duval N, Ticknor L, Layshock J, et al. Persistence of *Bacillus thuringiensis* subsp. kurstaki in urban environments following spraying. *Appl Environ Microbiol*. 2011;77(22):7954–61.
31. Sahl JW, Vazquez AJ, Hall CM, Busch JD, Tuanyok A, Mayo M, et al. The effects of signal erosion and core genome reduction on the identification of diagnostic markers. *mBio*. 2016;7(5):e00846-16.
32. Gardner SN, Frey KG, Redden CL, Thissen JB, Allen JE, Allred AF, et al. Targeted amplification for enhanced detection of biothreat agents by next-generation sequencing. *BMC Res Notes*. 2015;8:682.
33. Miller SE. DNA barcoding and the renaissance of taxonomy. *Proc Natl Acad Sci*. 2007;104(12):4775–6.

34. Ravel J, Jiang L, Stanley ST, Wilson MR, Decker RS, Read TD, et al. The complete genome sequence of *Bacillus anthracis* Ames "Ancestor." *J Bacteriol.* 2009;191(1):445–6.
35. Holden MT, Titball RW, Peacock SJ, Cerdeno-Tarraga AM, Atkins T, Crossman LC, et al. Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. *Proc Natl Acad Sci USA.* 2004;101(39):14240–5.
36. Parkhill J, Wren BW, Thomson NR, Titball RW, Holden MT, Prentice MB, et al. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature.* 2001;413(6855):523–7.
37. Sebaihia M, Peck MW, Minton NP, Thomson NR, Holden MT, Mitchell WJ, et al. Genome sequence of a proteolytic (Group I) *Clostridium botulinum* strain Hall A and comparative analysis of the clostridial genomes. *Genome Res.* 2007;17(7):1082–92.
38. Kralik P, Ricchi M. A basic guide to real time PCR in microbial diagnostics: definitions, parameters, and everything. *Front Microbiol.* 2017;8:108.
39. Liss M, Daubert D, Brunner K, Kliche K, Hammes U, Leiberer A, et al. Embedding permanent watermarks in synthetic genes. *PLoS ONE.* 2012;7(8):e42465.
40. Heider D, Barnekow A. DNA watermarks: a proof of concept. *BMC Mol Biol.* 2008;9:40.
41. Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science.* 2008;319(5867):1215–20.
42. Jupiter DC, Ficht TA, Samuel J, Qin QM, de Figueiredo P. DNA watermarking of infectious agents: progress and prospects. *PLoS Pathog.* 2010;6(6):e1000950.
43. Torok TJ, Tauxe RV, Wise RP, Livengood JR, Sokolow R, Mauvais S, et al. A large community outbreak of salmonellosis caused by intentional contamination of restaurant salad bars. *JAMA.* 1997;278(5):389–95.
44. Anonymous. Amerithrax investigative summary. Washington, DC: Department of Justice; 2010.
45. Rasko DA, Worsham PL, Abshire TG, Stanley ST, Bannan JD, Wilson MR, et al. *Bacillus anthracis* comparative genome analysis in support of the Amerithrax investigation. *Proc Natl Acad Sci USA.* 2011;108(12):5027–32.
46. Johnson SL, Daligault HE, Davenport KW, Jaissle J, Frey KG, Ladner JT, et al. Complete genome sequences for 35 biothreat assay-relevant bacillus species. *Genome Announc.* 2015;3(2).
47. Plaut RD, Stibitz S. Improvements to a markerless allelic exchange system for *Bacillus anthracis*. *PLoS ONE.* 2015;10(12):e0142758.
48. Cote CK, Bühr T, Bernhards CB, Bohmke MD, Calm AM, Esteban-Trexler JS, et al. A standard method to inactivate *Bacillus anthracis* spores to sterility via gamma irradiation. *Appl Environ Microbiol.* 2018;84(12).
49. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, et al. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* 2014;42(Database issue):D581–91.
50. Sun W, Wang S, Curtiss R 3rd. Highly efficient method for introducing successive multiple scarless gene deletions and markerless gene insertions into the *Yersinia pestis* chromosome. *Appl Environ Microbiol.* 2008;74(13):4241–5.
51. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A.* 2000;97(12):6640–5.
52. Horton RM, Hunt HD, Ho SN, Pullen JK, Pease LR. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene.* 1989;77(1):61–8.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

