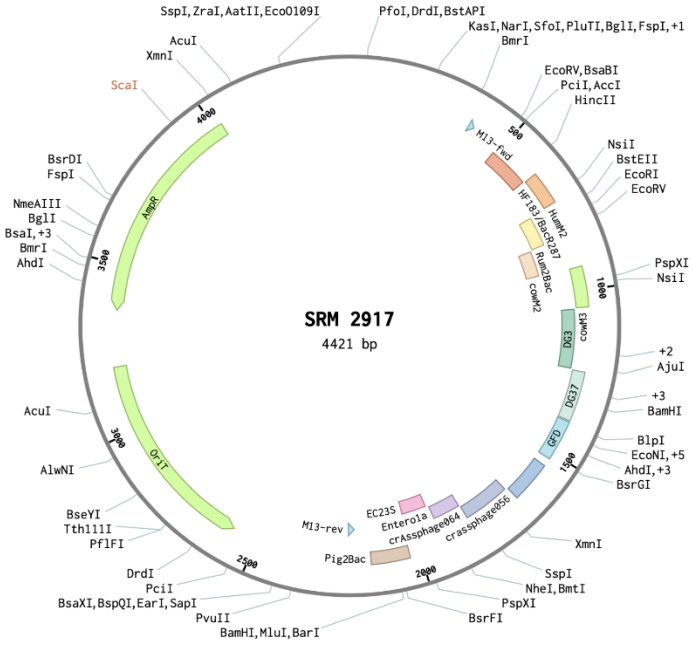
NIST Special Publication 260-221

Certification of Standard Reference Material® 2917

Plasmid DNA for Fecal Indicator Detection and Identification



Jason Kralj
Stephanie Servetas
Monique Hunter
Blaza Toman
Scott Jackson

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.260-221

National Institute of Standards and Technology
U.S. Department of Commerce

NIST Special Publication 260-221

Certification of Standard Reference Material® 2917

Plasmid DNA for Fecal Indicator Detection and Identification

Jason Kralj
Stephanie Servetas
Monique Hunter
Scott Jackson
Biomaterials and Biosystems Division
Material Measurement Laboratory

Blaza Toman
Statistical Engineering Division
Information Technology Laboratory

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.260-221

December 2021

U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
James K. Olthoff, Performing the Non-Exclusive Functions and Duties of the Under Secretary of Commerce for Standards and Technology & Director, National Institute of Standards and Technology

National Institute of Standards and Technology Special Publication 260-221
Natl. Inst. Stand. Technol. Spec. Publ. 260-221, 46 pages (December 2021)
CODEN: NSPUE2

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.260-221

Abstract

Standard Reference Material (SRM) 2917 is intended for calibrating qPCR-based assays of fecal indicator bacteria (FIB) for assessing water quality. A unit of SRM 2917 consists of 6 Levels of linearized plasmid DNA. The length of DNA is approximately 4.4 kb; it contains 13 single-copy PCR target regions. The 6 Levels span a log10 dilution series from approximately (5 to 500,000) plasmid copy per µL. Each Level is suspended in 1× TE buffer at pH 8 with approximately 10 ng/µL yeast tRNA. Approximately 0.2 mL of each Level were bottled in 1.5 mL skirted centrifuge vials. In total, approximately 1000 units were generated. The material is both stable and should be stored at 4 °C. The production, analytical methods, statistical evaluation, and certified values of this SRM are described herein.

Key words

Water quality, fecal indicator, qPCR, Plasmid DNA, ddPCR.

Table of Contents

List of Tables

List of Figures

1. Introduction

In collaboration with the US Environmental Protection Agency (US EPA), NIST has developed a plasmid DNA Standard Reference Material (SRM) for the purpose of standardizing EPA-developed quantitative polymerase chain reaction (qPCR) methods for identifying fecal indicator bacteria (FIB) to determine recreational water quality. The material consists of 6 Levels of a linearized plasmid DNA containing 13 single-copy PCR targets. The 6 levels are a log10 dilution series spanning approximately (5 to 500,000) plasmid copies per mL. Each tube of material contains approximately 200 mL of DNA in TE buffer pH=8 with 10 ng/mL of yeast carrier tRNA in a 1.5 mL skirted, screwcap, low-bind, polypropylene, centrifuge vial. The material is stable at 4 °C and should be stored at 4 °C. Approximately 1000 units were generated.

2. Storage and Use

The material is stable at 4 °C and should be stored at 4 °C. DO NOT FREEZE the material, as this is both unnecessary and may alter the concentration of the target. Long term stability studies at room temperature on the material have shown either no or minimal effect for several months. Hence, the material is sufficiently robust that it can remain at room temperature for many hours without degradation. However, we still recommend refrigeration (without freezing) for long-term storage.

3. Materials and Methods

3.1. Plasmid Manufacturing and Analysis

3.1.1. Plasmid synthesis, cloning, and purification

The sequence for the 13-target construct (≈1.7 kb) was provided by the USEPA (Table 1). Plasmid DNA was produced by Integrated DNA Technologies, Inc. (IDT) using their Custom Gene Synthesis service. The finished gBlock was then inserted into a pUCIDT vector (2752 bp) to produce the final circular 4421 bp plasmid (Fig 1).

Table 1. Description of USEPA-Developed Surface Water Quality Testing Construct Represented on SRM 2917

qPCR Assay	Organism	Pollution Source	Reference
HF183/BacR287	*Bacteroides* spp.	Human	[1]
HumM2			[2]
Rum2Bac	*Prevotella* spp.	Ruminant	[3]
Pig2Bac	*Bacteroides* spp.	Pig	[4]
CowM2		Cattle	[2]
CowM3			
DG3		Dog	[1]
DG37			
GFD	*Helicobacter* spp.	Avian	[5]
CPQ_056	CrAssphage	Human	[6]
CPQ_064			
Entero1a	*Enterococcus* spp.	General	[7]
EC23S857	*E. coli*		[8]

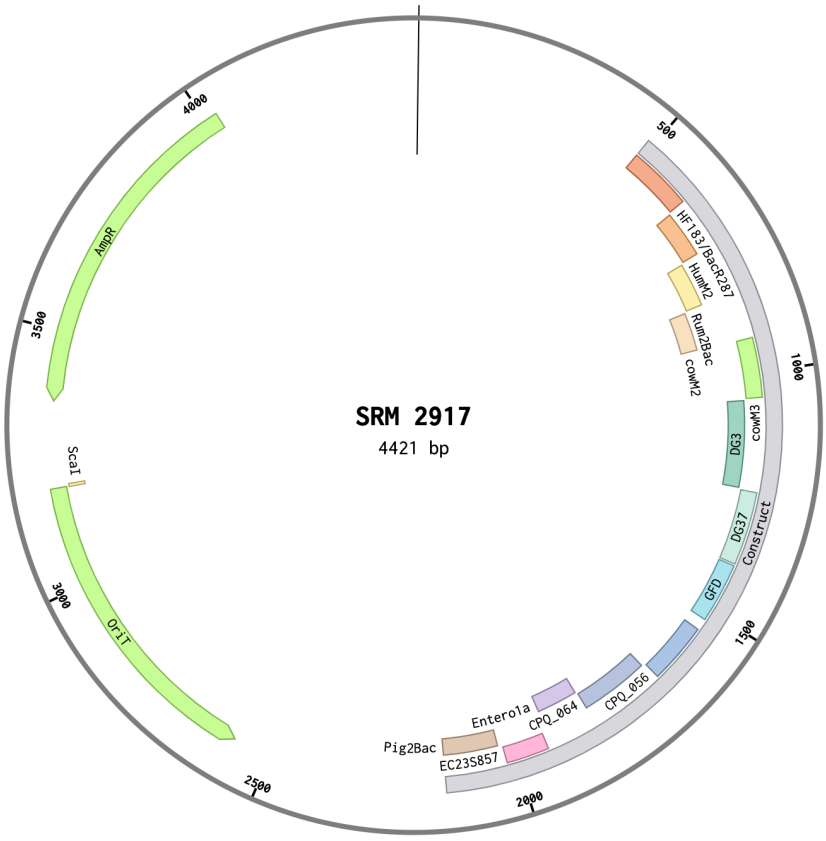


Figure 1. SRM 2917 Plasmid Map. 1.7 kb construct containing the 13 targets was synthesized de novo using IDT's gBlock technology. The 13 targets include genetic markers for human (HF183/BacR287, HumM2, CPQ_056, CPQ_064), ruminant (Rum2Bac), pig (Pig2Bac), cattle (CowM2, CowM3), dog (DG3, DG37), avian (GFD), *Enterococcus* (Entero1a), and *E. coli* (EC23S857) pollution targets. The finished gBlock was then inserted into a pUCIDT vector (2752 bp) carrying an AMP selection marker to produce the final circular 4421 bp plasmid. The ScaI cut site was used to linearize the plasmid.

This plasmid material was shipped from IDT directly to NIST. Upon receipt, approximately 10 ng of the plasmid was transformed into Invitrogen One Shot™ TOP10 Chemically Competent *E. coli* (Invitrogen catalog #C404003). Transformants were selected by plating on LB-AMP plates (Luria Broth (Difco BD Cat# 244620) agar (Bacto BD Cat#214010) supplemented with 100 µg/mL ampicillin (Millipore Sigma, Cat# A0166). The following day, approximately 200 colonies were present on the plate. Of these, 10 isolated colonies were picked and inoculated into 3 mL culture tubes containing LB liquid medium (Ward Science Cat# 470180-652) supplemented with 100 µg/mL ampicillin (Millipore Sigma, Cat#A1593). Cultures were grown overnight at 37 °C, with agitation (200 rpm). Plasmid DNA from each of the 10 cultures was purified using the Thermo Scientific GeneJET Plasmid Miniprep Kit (catalog # K0502). Sanger sequencing was performed on each plasmid using both universal M13 primers (M13F and M13R; IDT) to confirm the sequence identity of the 1.7 kb insert. Clone #1 was selected for large-scale production based on its 100 % sequence match with the expected sequence. Briefly, Clone #1 was used to inoculate four 50 mL cultures of LB liquid medium supplemented with 100 µg/mL ampicillin in a 200 mL

Erlenmeyer flask. Cultures were grown overnight at 37 °C, with agitation (200 rpm). The entire culture from each flask was pelleted for plasmid DNA purification. Plasmid DNA was purified using the Thermo Scientific GeneJET Plasmid Maxiprep Kit (catalog # K0492) and the four preps were pooled. In total, approximately 950 µg of purified plasmid DNA was recovered as measured using UV260 absorbance (Denovix).

The purified plasmid DNA was linearized by digestion with NEB's ScaI-HF Restriction Enzyme (New England Biolabs (NEB) Cat # R3122L) in 1× NEB CutSmart Buffer (50 mM Potassium Acetate, 20 mM Tris-acetate, 10 mM Magnesium Acetate, 100 µg/mL BSA) for 2 h at 37 °C. Following digestion, a DNA clean-up was performed by using SPRIselect magnetic beads (Beckman Coulter, catalog# B23318) following the manufacturer's recommendation. Purified DNA was recovered in approximately 2 mL of 1× TE (10 mM Tris, 1 mM EDTA, pH 8.0, Sigma BioUltra Molecular Biology Grade Catalog # 93283) and stored at 4 °C. The concentration of the plasmid was determined to be 240 ng/µL via Denovix based on OD260 absorbance and dsDNA-specific fluorescence (DeNovix Broad Range dsDNA assay kit) following the manufacturer's recommended protocols. Gel electrophoresis was performed to confirm linearity and size (Fig. 2). Based on the plasmid's concentration of 240 ng/µL and its 4.4 kb size, 660 ng-bp/nmol-bp molecular weight, and Avogadro's number, the copy number of the stock material was estimated to be approximately $5 \cdot 10^{10}$ molecules/µL.

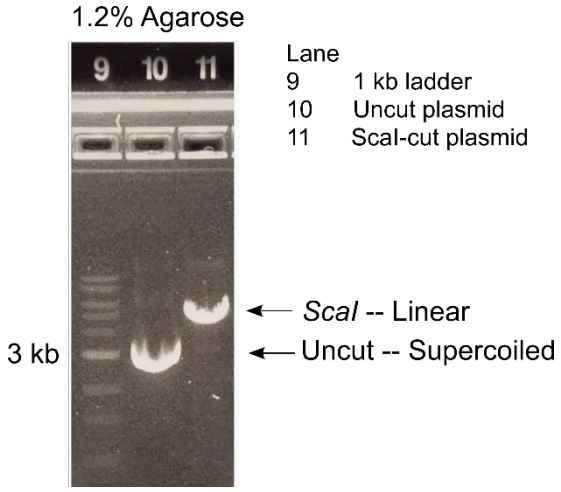


Figure 2. Gel electrophoresis of plasmid (4.4 kb) Gel image shows both supercoiled & uncut (Lane 10) and cut with ScaI restriction enzyme to generate a linear fragment (Lane 11).

3.1.2. Sequence Confirmation and Plasmid Assembly

Shotgun next generation sequencing (NGS) was performed to assess the purity of the material preparation and to confirm the sequence identity of the plasmid. Briefly, an Illumina Nextera XT library prep was performed with 10 ng of plasmid DNA following the manufacturer's recommended protocol (Illumina, Cat# FC-131-1002). This library was sequenced on an Illumina MiSeq using a 600 cycle v3 reagent cartridge to produce 2×300 bp reads.

Long-read data was generated by Oxford Nanopore (ONP) sequencing. The ONP library preparation was performed using the Ligation Sequencing Library Kit (ONP, Cat# SQK-

LSK109) and NEBNext Companion Module for Oxford Nanopore Technologies Ligation Sequencing (NEB cat # E7180S) to ligate adapters to the ends of the linearized plasmid followed by loading the DNA onto the Minion sequencing device. Briefly, 1 μg of linearized plasmid was prepared using the Ultra II End-prep reaction protocol. DNA was then purified using AMPure XP beads at a 1:1 ratio per manufacturer recommendations. Adapter ligation was completed per protocol using NEB and ONP reagents listed above, followed by reaction clean-up using AMPure XP beads at a 0.6:1 ratio. The ONP flow cell was primed and loaded per the manufacturer's recommendations.

The sequence assembly was checked two ways. First, the purported sequence (4421 bp) provided by IDT was used as the draft assembly, and pilon_polish.py using Pilon v1.23 [10] via Unicycler v0.4.8 [11] was used with the short-read data to identify any SNPs. Illumina paired short-read data (quality trimmed to Q15 via bbmap v.38.25 [4]) with options for 36 threads a min_polish_size of 0 were used. From the first assembly method using only pilon_polish.py, the input size of 4421 bp was scanned using almost 500k reads, with 446198 mapping to the contig. The mean coverage was 17787×, with a minimum of 1749×. There were 0 corrected SNPs, ambiguous bases, or small INDELs (insertions/deletions), and the restriction map was correct.

Second, the short-read data generated using the Illumina MiSeq platform was combined with the ONP data and used with the hybrid assembly tool Unicycler [11] using the command "/Applications/miniconda3/bin/unicycler -1 pEPA_S27_L001_R1_001.fastq.gz -2 pEPA_S27_L001_R2_001.fastq.gz -l ONP_EPA.fastq -t 8 -o EPA_Plasmid_Assembly_2 --verbosity 2". The hybrid assembly resulted in the same 4421 bp assembly with a depth of 3011×.

The final assembled sequence can be found in Appendix A: Plasmid Assembly. We note that these assemblies are non-certified properties of the material. We have attempted to use the best practices available for these assembly methods, but we do not assert any traceability for these methods to the SI. These analyses were performed with the purpose of ensuring the target regions did not contain errors that would prevent the ddPCR/qPCR assays from working. Based on the results, we find no evidence of errors in the sequence that would cause PCR assay problems.

3.1.3. Sample Purity

Sample purity was assessed primarily with respect to interference/bias of the 13 targets of the linearized plasmid. The primary source of contamination would likely come from the *E. coli* used to generate the plasmid or from the laboratory/reagents used for preparation or sequencing analysis.

Two NGS approached were used to evaluate the presence of contaminating DNA in the reference material preparation. The first was to use an alignment tool (bowtie2) to map the NGS reads to a reference *E. coli* genome and the plasmid's 4421bp reference sequence to determine the relative amounts of *E. coli* chromosomal DNA present in the plasmid preparation. Starting with 570,234 total raw Illumina reads, the bowtie2 alignment tool mapped 508,500 reads (approximately 90 %) to the reference plasmid sequence (≈22,000× coverage) and 151,910 (≈25 %) to *E. coli* MG1655 (≈7× coverage across the genome, Fig 3). Based on these relative read depths, the ratio of plasmid DNA to *E. coli* chromosome is estimated to be at least 3142 copies of plasmid per copy of *E. coli* chromosome. A small

amount of *E. coli* genomic DNA is expected due to carry-over from the plasmid isolation. Total counts exceed 100 % because there are homologous regions between the plasmid and *E. coli* genome.

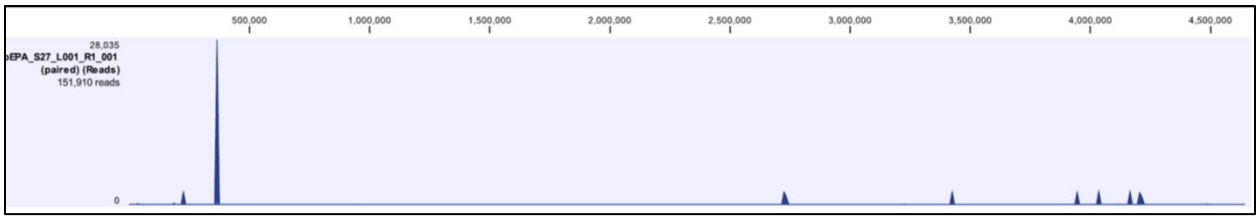


Figure 3. Illumina MiSeq reads from plasmid material mapped to MG1655 *E. coli* complete genome. Reads mapped across the entire *E. coli* genome at 5× to 10× coverage, on average. Several locations along the genome, however, had >1000× coverage (seen as spikes in the figure) presumably due to shared sequence homology with loci on the plasmid.

The second was to analyze the reads using a taxonomic classification tool [10,11]. A cursory metagenomic analysis of the reads (quality trimmed using bbmap, [12] classified using centrifuge v2.0 [13] with the default database) (Table 2) indicated only a fraction of the reads (128343/508500 reads, or 25 %) could be assigned to a species—this was expected as the sequence of the unique plasmid was not in the database. Examination of the species identified revealed a large overlapped with the qPCR target genera which include *Bacteroides, Prevotella, Helicobacter, Enterococcus,* and *Eschierichia*, as well as crAssphage. Hence, any metagenomic analysis with reads associated with those taxa (or other highly-related species such as *Salmonella* or *Shigella*) would not be considered contaminants. Further, the 13 small target regions will often collocate on a single read and lead to ambiguous identification that is discarded by the software.

Table 2: NGS taxonomic identification of potentially contaminating organisms.[a]

Species[a]	Number Reads	Unique Reads
Bacteroides dorei	17890	17887
Helicobacter sp. MIT 01-6242	15159	10836
Escherichia coli	24397	9888
Shigella dysenteriae	10661	8970
Helicobacter himalayensis	3616	3206
Helicobacter hepaticus	4459	1281
Shigella sp. PAMC 28760	11442	585
Lactococcus piscium	295	294
Shigella boydii 621	4294	136
Shigella sonnei 624	4608	109
Shigella flexneri	3319	91
Klebsiella pneumoniae	108	62
Homo sapiens	57	51
Salmonella enterica	158	37

[a]Highlighted in yellow are the two most abundance taxa not directly associated with the plasmid.

Only 2 potential contaminating species (*L. piscium* and *K. pneumoniae*) were detected at >20 unique reads, and neither of these likely to interfere in the analyses. The number of reads for each is <1 % of the sample analyzed and <0.1× coverage of each genome, making quantitation, and indeed conclusive identification, questionable. Hence, we do not find evidence of significant contamination in the material.

3.2. Material preparation

3.2.1. Dilution Buffer Preparation

Dilution buffer was prepared in a 3.8 L Lightweight HDPE Bottle (Fisherbrand # 0343814B) by adding 3 liters of 1× Tris-EDTA buffer solution (10 mM Tris, 1 mM EDTA, pH 8.0, Sigma BioUltra Molecular Biology Grade Catalog # 93283) and 30 mg of Yeast tRNA (ThermoFisher catalog #AM7119) so that the final concentration of tRNA was 10 ng/µL. The tRNA was added to stabilize the DNA and minimizing it from binding to the tube walls.

3.2.2. Gravimetric Dilution of Plasmid to Make Log-Dilution Series (Levels 1-6)

Dilutions of the material were prepared by gravimetrically diluting the stock plasmid material into Teflon® Resin Perfluoroalkoxy (PFA) Jars (Chemware – Mfr # D1069029).

Based on the plasmid's stock (starting) concentration of 240 ng/µL and its 4.4 kb size, the copy number of the stock material was estimated to be approximately $5 \cdot 10^{10}$ molecules/µL.

In 50 mL (low bind) Falcon Tubes:

1. Add 45 µL of 240 ng/µL plasmid to 45 mL dilution buffer for 45 mL of $5 \cdot 10^{7}$ copies/µL
2. Add 4.5 mL of #1 to 40.5 mL dilution buffer for 45 mL of $5 \cdot 10^{6}$ copies/µL

Between each serial dilution, tubes are vortexed for 30 to 45 seconds.

In 360 mL PFA Pots with spin-bars, prepare:

1. Add 30 mL of #2 to 270 mL dilution buffer for 300 mL of $5 \cdot 10^{5}$ copies/µL
2. Add 30 mL of #3 to 270 mL dilution buffer for 300 mL of $5 \cdot 10^{4}$ copies/µL
3. Add 30 mL of #4 to 270 mL dilution buffer for 300 mL of $5 \cdot 10^{3}$ copies/µL
4. Add 30 mL of #5 to 270 mL dilution buffer for 300 mL of $5 \cdot 10^{2}$ copies/µL
5. Add 30 mL of #6 to 270 mL dilution buffer for 300 mL of $5 \cdot 10^{1}$ copies/µL
6. Add 30 mL of #7 to 270 mL dilution buffer for 300 mL of $5 \cdot 10^{0}$ copies/µL

The dilution buffer was added to each pot first and measured gravimetrically to achieve 270 g (270 mL). Serial dilutions were made by transferring 30 mL by 50 mL serological pipette (Falcon, 357550). Before each serial dilution, tubes are mixed (on stir plate) for (30 to 45) seconds. The weights of PFA pots were recorded after each addition. After diluting, the material was stored in the PFA jars sealed with parafilm at 4 °C until bottling.

3.2.3. Bottling Material

3.2.3.1.Tube Selection

The material was bottled in Sarstedt skirted 1.5 mL screw cap microcentrifuge tubes (Sarstedt catalog # 72.730.700, Fig 4.) featuring a special "low bind" polypropylene devised to minimize DNA binding during storage and thereby maximize recovery and stability of the DNA material. These tubes are certified by the manufacturer to be free of DNA, DNase,

RNase, and PCR inhibitors and optimized for PCR performance. These tubes were compatible with the Scinomix bottling instrument described below.

3.2.3.2.Automated Bottling and Labeling

The Scinomix Sci-Print VXL system (Fig 5.) was used to bottle the material. This instrument grabs a tube, uncaps the tube, dispenses the material into the tube, recaps the tube, prints the pre-approved label for that Level, affixes the label to the tube, and racks the tube.

Figure 4. Sarstedt skirted 1.5 mL screw cap microcentrifuge tubes.

Figure 5. The Scinomix Sci-Print VXL

3.2.3.3.Aliquoting Levels for Unit Manufacture

One Level was processed ("bottled") per day. Aliquoting was conducted in a deliberate fashion starting with the Level 1 and progressing to next level to minimize any potential cross-contamination. The PFA pot containing the desired Level was brought to room temperature on a stir plate set to low. Before each run, the peristaltic pump calibration was checked with water to ensure accurate dispensing of 200 µL per aliquot. After a rack of approximately 300 tubes was filled, these were transferred to 100-ct freezer storage boxes. All boxes and tubes were kept at room temperature until aliquoting was completed (approximately 6 h), then transferred to 4 °C.

4. Certification Strategy

The material was certified for plasmid copy number concentration using ddPCR (droplet digital PCR), which is traceable to volume. Five (5) separate PCR assays were employed to determine the copy number concentration. Homogeneity of the material was assessed using two of the assays on all 6 Levels from approximately 20 units. Stability of the material was assessed using one assay across approximately 6 months for 4 units (2 units at 4 °C and 2 units at room temperature (20 to 25) °C). For complete details, see Appendix B: Experimental Plan.

The Statistical Engineering Division (ITL) calculated the certified value, homogeneity, and stability for each Level and performed an uncertainty analysis on the material to assign a consensus value for the copy number concentration. The details of these analyses are located in the Appendices.

4.1. ddPCR

All ddPCR assays were carried out on the BioRad (California) system, consisting of the AutoDG, C1000 thermocycler, and QX-200 Droplet Reader. Unless otherwise noted, the equipment was used following the manufacturer's recommendation.
Droplet volume (S1 = 0.7663 nL ± 2.3 % (k=2), average and expanded uncertainty) was previously determined by N. Farkas (Appendix C Measurement Report) NIST Physical Measurement Laboratory, Semiconductor and Dimensional Metrology Division, Nanoscale Metrology Group (NMG) 683.03 using Special Test Method 11050S/-D as described in NIST Special Publication 260-184 [9].

Assay primers and probes were designed by the EPA and used by NIST with permission. The primers (desalted) and probes (HPLC-purified) were obtained from Biosearch (LGC, United Kingdom) and diluted upon receipt to 100 µM concentration. Primer/probe mixtures were made containing 9.375 µL each of forward and reverse primers (100 µM each) and 6.25 µL of probe (100 µM) with 475 µL nuclease-free water (Gibco) for 500 µL total volume, stored at 4 °C. This resulted in concentrations of 1.875 µM for each primer and 1.25 µM for the probe. The primer sets used for the 5 assays are listed in Table 3.

Table 3. Primer and probes for the 5 assays used to certify the copy number concentration.

Assay	Primer ID	Sequence
Rum2Bac	BacB2_590F	ACAGCCCGCGATTGATACTGGTAA
	Bac708Rm	CAATCGGAGTTCTTCGTGAT
	BacB2_626P	[FAM] ATGAGGTGGATGGAATTCGTGGTGT [BHQ-1]
Ec23S857	EC23SF2-1	GGTAGAGCACTGTTTTGGCA
	EC23SR2-1	TGTCTCCCGTGATAACTTTCTC
	EC23SP2b	[FAM] TCATCCCGACTTACCAACCCG [BHQ-1]
CowM2	CowM2F	CGGCCAAATACTCCTGATCGT
	CowM2R	GCTTGTTGCGTTCCTTGAGATAAT
	CowM2P	[FAM] AGGCACCTATGTCCTTTACCTCATCAACTACAGACA [BHQ-1]
CowM3	CowM3F	CCTCTAATGGAAAATGGATGGTATCT
	CowM3R	CCATACTTCGCCTGCTAATACCTT
	CowM3P	[FAM] TTATGCATTGAGCATCGAGGCC [BHQ-1]
CPQ_056	crAss056_F1	CAGAAGTACAAACTCCTAAAAAACGTAGAG
	crAss056_R2	GATGACCAATAAACAAGCCATTAGC
	crAss056_P1	[FAM] AATAACGATTTACGTGATGTAAC [BHQ-1]

2x dPCR Supermix was obtained from Bio-Rad (Catalog #1863010). A total of 35 mL of 2× Supermix, all from the same lot (#64294601), was purchased and delivered to NIST on 01/06/2020. The expiration date was indicated to be 02/06/2021. 96-well plates were obtained from Eppendorf (Cat # 951020303). Rainin pipettes calibrated within the past 12 months were used in 20 µL, 100 µL, 200 µL, and 1000 µL sizes with corresponding filtered tips. Nuclease-free water (Gibco) was used for no template controls (NTC) and in the reactions.

The assay design was similar to RM 8376 (ROA 644-01-21-0002) (Table 4) and tested under the same assay conditions. Enough of each assay mix (master mix, primers/probe, and water) were made for the number of samples on each plate, plus 5 % excess for pipetting error. Then, 20 µL of each assay mix was added to assigned sample wells. Finally, 5 µL of sample was added to the well. Levels 5 & 6 were too concentrated to run as-is and were diluted 100-

fold in the same 1× TE pH 8 buffer described above and run immediately. Plates were covered and sealed with foil and vortexed for 30 s on a Vortex-Genie 2 (Scientific Industries) with the foam plate attachment then spun down for approximately 15 s in a MPS 1000 plate spinner (Labnet).

Table 4. Assay for ddPCR.

Component	Concentration	Amount (µL)
Master mix	2×	12.5
Primers/probes	375 nM/250 nM	5.0
Water	-	2.5
Sample	-	5.0
TOTAL	**-**	**25.0**

The AutoDG was used to generate droplets and load a fresh 96-well plate. Once completed, this new plate was sealed and immediately run on the C1000 thermocycler. The program consisted of:

- Hot-start phase of (95 °C for 10 min)
- 60 cycles of (94 °C for 30 sec, 60 °C for 60 sec,) with a maximum ramp rate of 0.9 °C/s
- Enzyme deactivation (98 °C for 10 min) + 4 °C hold.

The plate was then loaded into the QX-200 and analyzed following instructions in the Quantasoft v1.7.4 software program (Bio-Rad). Once all droplets were read, each assay was examined (Fig. 6).

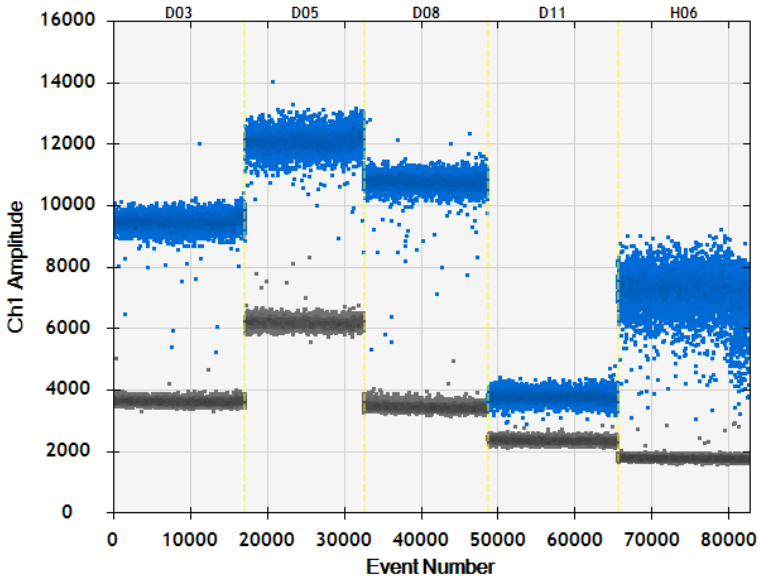


Figure 6. 1D plot of the 5 ddPCR assays. Shown left to right – Rum2Bac, CowM2, Ec23S857, CPQ_056, CowM3), demonstrated on Level 4 samples. Clear separation between the unoccupied droplets (gray) and positive droplets (blue) was observed for all 5 assays.

The signal threshold was manually set for each assay and reaction well to be within the pre-determined operating range for the assay (Table 5). The intensity of the positive and negative

droplets and the total number of droplets were recorded and checked against typical performance. For reactions found to deviate significantly from other assays on the same plate, these were flagged for exclusion as being failed reactions (due to pipetting error, droplet generation, or machine mis-reading). Thresholds were set approximately 30 % to 50 % of the signal range (max signal – min signal) above the minimum intensity (Table 5).

Table 5. Assay and threshold intensity parameters for the 5 ddPCR assays.

Assay	Minimum	Maximum	Threshold Min[a]	Threshold Max
Rum2Bac	3660	8000	+1500 (25 %)	+4300 (77 %)
CowM2	6500	12300	+2400 (41 %)	+3500 (60 %)
CowM3	1650	7100	+600 (12 %)	+3100 (57 %)
CPQ_056	2300	3850	+350 (21 %)	+750 (45 %)
EC23S857	3250	10300	+1200 (17 %)	+5400 (76 %)

[a]The threshold designation indicates the intensity units above (+) the baseline.

4.2. Statistical Analysis

The following statistical model was used to produce the estimates of plasmid copy number concentration with uncertainty.
The response was represented as:

$y_{ijkl} \sim N(\gamma_{ijk}, \sigma^2)$,	i denotes assay ($i = 1, \ldots, 5$), j denotes Level (1 ,…, 6 in the order as above), k denotes units (1 ,…, 5), l denotes replicate
$\gamma_{ijk} \sim N(\alpha_{ij}, \sigma_{ij}^2)$,	γ_{ijk} is the mean lambda value of assay i, Level j, unit k.
$\alpha_{ij} \sim N(\mu_j, \sigma_j^2)$	α_{ij} is the consensus mean lambda value of assay i, Level j
μ_j	is the overall lambda value for Level j
$V \sim dnorm(0.7663, 0.0088^2)$,	represents droplet volume of DNA ‘UNG’ MM,
$\mu_{j(adj)} = {}^{5 * \mu_j}/_{V}$,	represents the measurand; mean $copies/\mu L$ for each Level. (The multiplication by 5 here is to follow Jason’s R code (see Appendix G), not to multiply by 5 to get EPA values)

Note that this statistical model accounts for uncertainty due to unit differences, assay differences, droplet volume, and repeatability.
Because of the wide range of concentrations, a subset of the data was transformed to Log10 and analyzed to determine if it would improve the statistical fit to the model. It did not—hence, no additional pre-processing of the data was undertaken.
Unless otherwise noted, “significant” differences were values with non-overlapping 95 % confidence intervals.

5. Results

5.1. Homogeneity Assessment

Following the experimental protocols, the ddPCR results for the Units designated (5, 6, 8, 12, 16, 22, 23, 33, 37, and 43) showed no evidence of significant heterogeneity. Below, we show the results summarized for each Level as described by SED's analysis of the ddPCR data.

Level 1: there was no significant heterogeneity between units detected.

Level 2: no significant heterogeneity detected with the Rum2Bac assay. However, Unit 23 measured significantly different from unit 8 using the EC23S857 assay. Unfortunately, Unit 8 packaging was damaged, and the sample evaporated, despite no obvious signs of failure; and could not be retested.

Level 3: there was no significant heterogeneity between units detected.

Level 4: No significant heterogeneity in EC23S857. In Rum2bac, Unit 6 was significantly different from units 16, 22, 37, and 43. Unit 6 has a significantly higher SD than other units, indicating a potential assay problem. We retested units 6 and 37 and found no difference between the vials' mean and SD, suggesting the original measurement on unit 6 was an experimental error.

Level 5: there was no significant heterogeneity between units detected.

Level 6: we initially measured significant differences in Rum2bac for Unit 22 vs. Units 5, 16, and 23; with the SD for Unit 22 significantly higher than the other vials. For Rum2bac, reevaluation of units 22, 5, 16, and 23 showed no significant difference and CV less than 5 %, suggesting the original discrepancy was due to experimental error on 11/19/2020 (Fig 7).

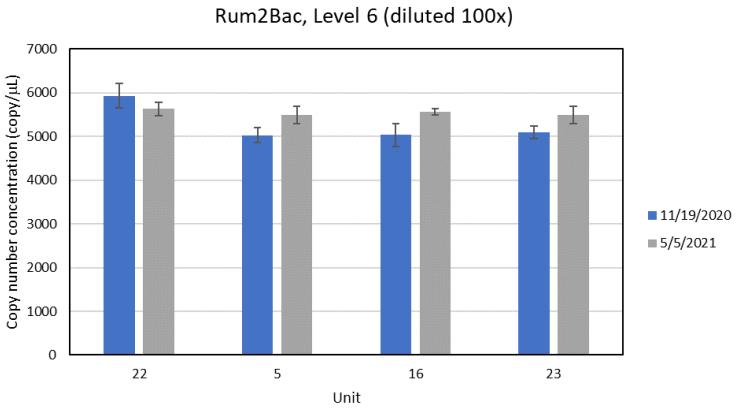


Figure 7. Reevaluation of units, Level 6. Concentrations from the offending Units were retested and not found to have significant differences on the date of retesting. The error bars represent 1 sd.

Additionally, there were initially significant differences in EC23S857 for Unit 37 vs. units 5, 6, and 8; with unit 37 SD significantly higher than the others. Reevaluation of Units 37, 5, 6, and 8 showed no significant differences between the units; again, suggesting an experimental error on 11/19/2020 (Fig 8).

This Level was particularly challenging to assess because the sample must be diluted 100-fold (which carries additional uncertainty) and was measuring near the maximum concentration range for ddPCR. Overall, no units were significantly different by both assays. The full statistical report is detailed in Appendix D: Homogeneity Testing.

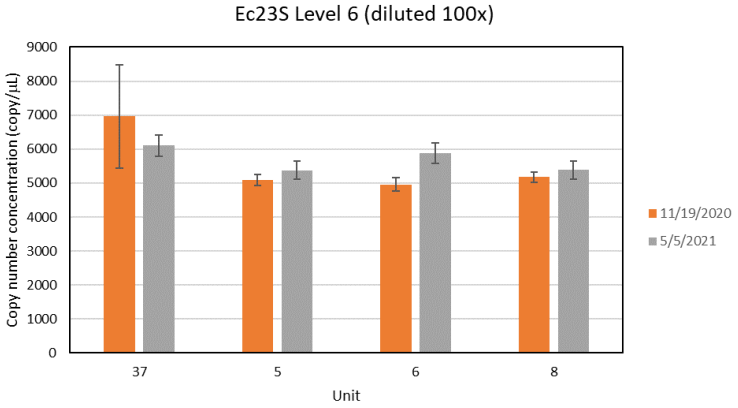


Figure 8. Reevaluation of offending units, Level 6. There were no significant differences in concentration found between any of the units after retesting. The error bars are 1 sd.

5.2. Stability Assessment of Homogeneity and Stability Data

A linear regression model was generated for each unit and temperature. In this, no uncertainty in the measurement was assumed; this produced a highly sensitive analysis that can detect small changes in concentration with time, but with a possibility of "false positives" for instability. If the slope was not statistically different from 0, the material would be accepted as stable. If not, additional analysis was needed to examine the magnitude and likelihood of instability.

We first plotted the average concentration at each time and temperature (Fig. 9 and Fig 10.). At each Level, we did not observe any obvious changes.

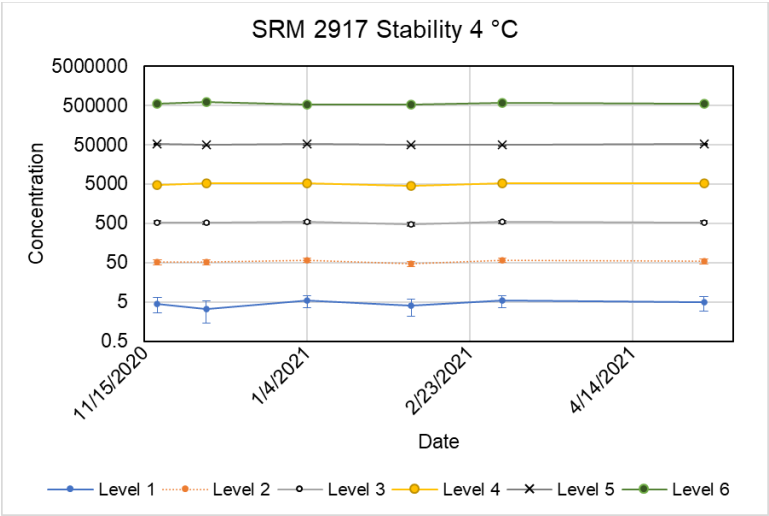


Figure 9. Stability of the 6 Levels over approximately 6 months at 4 °C. Concentration is copy/µL. Each value is the average of 2 units tested in triplicate. The error bars are 1 sd of the measurement at each Level, and for Levels 3-6 these are smaller than the data points.

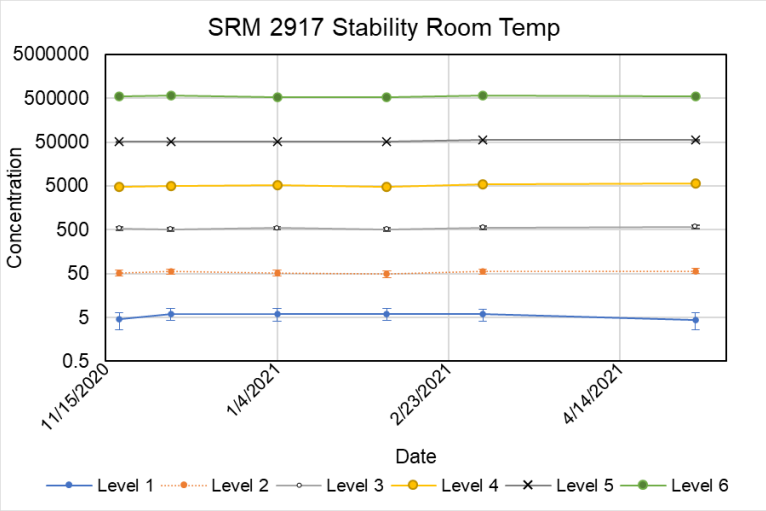


Figure 10. Stability of the 6 Levels over approximately 6 months at room temperature. Concentration is copy/µL. Each value is the average of 2 units tested in triplicate. The error bars are 1 sd of the measurement at each Level, and for Levels 3-6 these are smaller than the data points.

The statistical analysis of the data revealed no significant changes of the copy number concentration with time for any Level of the 4 °C or 20 °C samples, except Level 4 at 20 °C. This outlier is not a major concern for several reasons: a) no other Level (higher or lower) showed significant changes in concentration, b) this temperature is not recommended for long term storage, c) the strong possibility that some of these data are outliers due to measurement error (at t=0 days and approximately t=80 days), and d) this concentration is at the upper limit of the instrument capabilities, and may include additional uncertainty. We will continue to monitor the material for stability and update the status of the material if warranted. The full statistical analysis of unit stability is given in Appendix E: Stability, and the results are listed in Table 6.

Table 6. Stability analysis of all Levels, Units, and Temperatures.

Level	Unit	T (°C)	Slope[a] (copy/µL/day)	Slope uncertainty	L95	U95
1	8	4	0.001	0.008	-0.016	0.017
1	23	4	0.004	0.006	-0.009	0.016
1	6	20	-0.011[b]	0.005	-0.022	-0.001
1	22	20	0.002	0.007	-0.012	0.017
2	8	4	0.007	0.036	-0.065	0.077
2	23	4	0.011	0.020	-0.028	0.049
2	6	20	0.028	0.038	-0.046	0.103
2	22	20	0.005	0.023	-0.040	0.051
3	8	4	-0.05	0.16	-0.36	0.26
3	23	4	0.09	0.13	-0.17	0.35
3	6	20	0.25	0.14	-0.02	0.52
3	22	20	0.42[b]	0.12	0.18	0.66
4	8	4	2.69[c]	0.47	1.77	3.61
4	23	4	1.67[c]	0.36	0.97	2.39

4	6	20	**4.81**[d]	0.41	4.01	5.63
4	22	20	**4.09**[d]	0.38	3.35	4.85
5	8	4	-0.01	0.16	-0.34	0.30
5	23	4	-0.22	0.16	-0.54	0.08
5	6	20	0.38[b]	0.09	0.20	0.56
5	22	20	-0.03	0.16	-0.35	0.28
6	8	4	-0.91	0.47	-1.84	0.02
6	23	4	1.17[b]	0.40	0.38	1.98
6	6	20	0.91[b]	0.40	0.14	1.69
6	22	20	-0.90[b]	0.38	-1.72	-0.20

a Units with a slope that does not include 0 within the confidence bounds are bolded.
b Consensus values from all units indicated no statistical significance.
c Classical linear regression of copy number on time (ignoring unit) gives an estimated slope of 1.43 with standard uncertainty of 1.12 (p=0.21); and therefore not significant at the 0.05 level.
dClassical linear regression of copy number on time (ignoring unit) gives an estimated slope of 3.73 with standard uncertainty of 1.0 (p=0.0006); and therefore significant.

5.3. Certified Values

The SRM certified values (Table 7) closely matched the expected values. The standard uncertainty in each case was less than 10 % of the mean value (ranging from 1.3 % to 9.3 %). Hence, qPCR calibration for the water quality testing assays using these SRM Levels can have high confidence in the abundance of plasmids used. The complete analysis is detailed in Appendix F: Analysis of SRM 2917 using Hierarchical Bayes Analysis.

Table 7. Certified Values, standard uncertainty, and 95% confidence interval for each Level of the SRM

Level	Mean copy/µL	Standard Uncertainty copy/µL	L95	U95
1	5.4	0.5	4.5	6.5
2	55.5	1.6	52.5	58.7
3	530	7	516	545
4	5314	71	5175	5452
5	52195	744	50779	53704
6	541287	7779	527588	557912

5.3.1. Traceability

The certified value is traceable to copies per volume, expressed as number of amplifiable entities per microliter of solution, through the confirmation of identity, the ddPCR count measurements, the validity of the Poisson endpoint transformation of counts to entities for digital PCR endpoint assays, and calibrated volume measurements made at NIST [9]. Metrological traceability is to the SI unit for volume.

6. Conclusions

We successfully manufactured a 6-Level material for qPCR testing, with certified values for plasmid copy number concentration. The plasmid, originally designed by the US EPA to

include 13 markers for fecal contamination in recreational water, was cloned into competent *E. coli*, grown, and harvested using standard microbiological techniques to produce nearly 1 mg of DNA. This will enable multiple generations of the SRM to be produced using the same plasmid preparation. This approach generates more plasmid product than PCR, and is generally less error-prone over a 1.7 kb fragment than PCR-based techniques. This is important for PCR-based diagnostics, as errors in primer- and probe-binding sites reduce the efficiency of amplification.

The SRM was shown to be homogenous across each of the 6 Levels. While there were some discrepancies between units, none were significantly different than the mean at each Level.

The certified values followed the log10 dilution series intended, with values ranging from approximately (5 to 500,000) plasmid copies per µL. The relative standard uncertainty is less than 10 % for all Levels, which is sufficient for the intended purpose of qPCR assay calibration.

The most notable feature of the material is the stability. The concentration was shown to be stable at 4 °C (normal storage) and room temperature (accelerated storage) for several months, even at the lowest Level. This was due to the addition of approximately 10 ng/µL tRNA to the buffer, which has been hypothesized to passivate the tube walls and promote stability in solution. This new material removes many of those sources of variability and will result in improved calibration and testing. Additional fitness testing by USEPA has been ongoing and preliminary results corroborate the stability and homogeneity of the material. In addition, USEPA preliminary findings suggest that the material allows for the generation of high-quality standard curves for all 13 qPCR assays.

Acknowledgments

The authors would like to thank Megan Cleveland and Hua-Jun He for critical review of the writing.

References

[1] Green HC, Haugland RA, Varma M, et al. Improved HF183 quantitative real-time PCR assay for characterization of human fecal pollution in ambient surface water samples. *Appl Environ Microbiol*. 2014;80(10):3086-3094. https://doi.org/10.1128/AEM.04137-13
[2] Shanks OC, Kelty CA, Sivaganesan M, Varma M, Haugland RA. Quantitative PCR for genetic markers of human fecal pollution. *Appl Environ Microbiol*. 2009;75(17):5507-5513. https://doi.org/10.1128/AEM.00305-09
[3] Mieszkin S, Yala JF, Joubrel R, Gourmelon M. Phylogenetic analysis of Bacteroidales 16S rRNA gene sequences from human and animal effluents and assessment of ruminant faecal pollution by real-time PCR. J Appl Microbiol. 2010 Mar;108(3):974-984. https://doi.org/10.1111/j.1365-2672.2009.04499.x
[4] Mieszkin S, Furet JP, Corthier G, Gourmelon M. Estimation of pig fecal contamination in a river catchment by real-time PCR using two pig-specific Bacteroidales 16S rRNA genetic markers. *Appl Environ Microbiol*. 2009;75(10):3045-3054. https://doi.org/10.1128/AEM.02343-08
[5] Green HC, Dick LK, Gilpin B, Samadpour M, Field KG. Genetic markers for rapid PCR-based identification of gull, Canada goose, duck, and chicken fecal contamination in water. *Appl Environ Microbiol*. 2012;78(2):503-510. https://doi.org10.1128/AEM.05734-11

[6] Stachler E, Kelty C, Sivaganesan M, Li X, Bibby K, Shanks OC. Quantitative CrAssphage PCR Assays for Human Fecal Pollution Measurement. Environ Sci Technol. 2017 Aug 15;51(16):9146-9154. https://doi.org/10.1021/acs.est.7b02703
[7] Siefring S, Varma M, Atikovic E, Wymer L, Haugland RA. Improved real-time PCR assays for the detection of fecal indicator bacteria in surface waters with different instrument and reagent systems. J Water Health. 2008 Jun;6(2):225-37. https://doi.org/10.2166/wh.2008.022 PMID: 18209285
[8] Chern EC, Siefring S, Paar J, Doolittle M, Haugland RA. Comparison of quantitative PCR assays for Escherichia coli targeting ribosomal RNA and single copy genes. Lett Appl Microbiol. 2011 Mar;52(3):298-306. https://doi.org/10.1111/j.1472-765X.2010.03001.x
[9] Dagata JA, Farkas N, Kramar JA (2016) Method for Measuring the Volume of Nominally 100 um Diameter Spherical Water-in-Oil Emulsion Droplets. *NIST Special Publication 260-184*. https://doi.org/10.6028/NIST.SP.260-184
[10] Walker BJ, Abeel T, et al. (2014) Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* 9(11):e112963. https://doi.org/10.1371/journal.pone.0112963
[11] Wick RR, Judd LM, Gorrie CL, Holt KE (2017) Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13(6):e1005595. https://doi.org/10.1371/journal.pcbi.1005595
[12] Bushnell, Brian. "BBMap: A Fast, Accurate, Splice-Aware Aligner". United States. https://www.osti.gov/servlets/purl/1241166
[13] Kim D, Song L, Breitwieser FP, and Salzberg SL. (2016) Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Research* 26(12):1721-1729. https://doi.org/10.1101/gr.210641.116

Appendix A: Plasmid Assembly

The final assembly of the plasmid is given below.

>pEPA-Unicycler_Assembly_Illumina_ONP length=4421 depth=3011.06x

AAAATACCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCAT
CGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAG
TCACGACGTTGTAAAACGACGGCCAGTGCAACGCGATGACGATGGATAGCGATTCATCGATGAGCTGACCCGAT
CGCCGCCGCCGGAGGGTTGCGTTTGAGACGGGCGACAGATATCGATCATGAGTTCACATGTCCGCATGATTAAA
GGTATTTTCCGGTAGACGATGGGGATGCGTTCCATTAGCTCGAGATAGTAGGCGGGGTAACGGCCCACCTAGTC
AACGATGGATAGGGGTTCTGAGAGGAAGATCGCGTCAGGTTTGTTTCGGTATTGAGTATCGAAAATCTCACGGA
TTAACTCTTGTGTACGCTCTCGAGGACCAGCTAATGCATATAAATAAGTTACGTGATGAATCGACAGCCCGCGA
TTGATACTGGTAACCTTGAGTGCGGATGAGGTGGATGGAATTCGTGGTGTAGCGGTGAAATGCTTAGATATCAC
GAAGAACTCCGATTGATCGCGGCCAAATACTCCTGATCGTACTCGAGATAGGCACCTATGTCCTTTACCTCATC
AACTACAGACAAAATTATCTCAAGGAACGCAACAAGCATCGCCTCTAATGGAAAATGGATGGTATCTTTGGAGC
CTTTGAAAGCACTCGAGCCTTATGCATTGAGCATCGAGGCCGGAAAGCAGGAACTTATATATAATAAGGTATTA
GCAGGCGAAGTATGGATCGTGAGCGGGCATGGTCATATTATTGTTTAGACCGTGGATGTTTGCTCTGCATTAAC
CTCCGTCCGCAGGATATTCAGAAAACCATCAAGGAGGTTCGCCCTACAGCCATGTGCAGTGTGCCCCGTTTCTG
GGAAAAAGTCTACGCGGGCGTACTGGAAAAGATTAACGAAACAACGGGGCTGAAAAATCGCTTGGTTATGGGCG
ACATTGGAATTAATACTACCACTTCTATTATTGAACGTTTAAAGGAGCAGGTGGCAGAAAAGCATATTAAGGAT
CCACAGGAATGTAAGCAGCTTCTTATGGACAGTATGAAAGAGCAGATGACCGTGGGAGAAAAATCGTCGGCTGA
GCACTCTAGGGAGACTGCCTTCGTAAGGAGGAGGAAGGTGAGGACGACGTCAAGTCATCATGGCCCTTACGCCT
AGGGCTACACACGTGCTACAATGGGATGTACAAAGAGACGCATCGGCTGAACAAACTGCTAATGCAGAAGTACA
AACTCCTAAAAAACGTAGAGGTAGAGGTATTAATAACGATTTACGTGATGTAACTCGTAAAAAGTTTGATGAAC
GTACTGATTGTAATAAAGCTAATGGCTTGTTTATTGGTCATCTTGAATCGTGTATAGATGCTGCTGCAACTGTA
CTCTCTGAAATTGTTCATAAGCAAATTGATATTTCTATTAAAAGTCAATTTCTATTTGTTCTTAAACATATTGC
TTATACTTTTAGAAATATTATTTATGGACAAGATAAAGATGAAAACAACGATCGGAGAAATTCCAAACGAACTT
GGAGATAGCTGGTTCTCTCCGAAATAGCTTTAGGGCTAGCCTCGGAATTGAGAATGATGGAGGTAGAGCACTGT
TTGATCGGGGGGTAGAGCACTGTTTCGGCAAGGGGGTCATCCCGACTTACCAACCCGACTCGAGCTGCGAATAC
CGGAGAATGTTATCACGGGAGACACACGGCATCGGCATGAATTTAGCTTGCTAAATTTGATGGCGACCGGCGCA
CGGGATCCTAACGCGTATCCAACCTTCCCTTATCCACGGGATAGCCCGTCGAAAGGCGGATTAATACCGTATGA
GGTATCGATCAGTTCTGGACCAGCGAGCTGTGCTGCGACTCGTGGCGTAATCATGGTCATAGCTGTTTCCTGTG
TGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTA
ATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGC
TGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACT
GACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCAC
AGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCG
CGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGG
CGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGAC
CCTGTCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTA
GGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGC
TGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCAC
TGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCT
ACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCT
TGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAA
AGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGA
TTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATC
TAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTG
TCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTG
GCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCC
GGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGC
TAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCT
CGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGC
AAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGT

TATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAA
CCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCG
CCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACC
GCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCG
TTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATA
CTCATACTCTACCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGA
ATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAA
CCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGAT
GACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAG
ACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGA
TTGTACTGAGAGTGCACCAAATGCGGTGTGAAATACCGCACAGATGCGTAAGGAG

Appendix B: Experimental Plan

To certify the copy number concentration, we will use five (5) ddPCR assays that specifically target different regions of the plasmid. These assays were originally designed by Orin Shanks, et al. of US Environmental Protection Agency (EPA) for use in quantitative PCR (qPCR). NIST has undertaken work to characterize the assays' performance using ddPCR on the plasmids, and found the performance aligned well with expectation.
NIST will test 5 units to determine a certified value. Each unit contains 6 levels. Each level will be measured in triplicate. Three (3) no-template controls (blanks) will be included, which will be used to further characterize uncertainty due to the possible contamination of the reagents with bacterial DNA.

5 (assays)
5 (units)
6 (levels)
3 (replicates)
= 450 reactions
+ 30 NTC (6 replicates per assay)
==========
480 reactions

This can be accomplished using 5 standard 96-well plates, which are commonly used by the ddPCR analysis infrastructure.
An example plate layout is shown below (see Figure B1). Typically, only 1 or 2 plates can be run on a single day. In addition, each plate will contain a combination of units, levels, and assays so that day-to-day variability can be evaluated.
For a material that is homogeneous, this setup will enable:

a. Quantification of DNA copy number concentration through 5 independent assays
b. Within-day variability of the measurement on the same sample across all levels
c. Inter-day variability of the measurement on the same level between units

Variability

	1	2	3	4	5	6	7	8	9	10	11	12
A	U1-A1-L1-1	U1-A1-L1-2	U1-A1-L1-3	U2-A2-L1-1	U2-A2-L1-2	U2-A2-L1-3	U3-A3-L1-1	U3-A3-L1-2	U3-A3-L1-3	U4-A4-L1-1	U4-A4-L1-2	U4-A4-L1-3
B	U1-A1-L2-1	U1-A1-L2-2	U1-A1-L2-3	U2-A2-L2-1	U2-A2-L2-2	U2-A2-L2-3	U3-A3-L2-1	U3-A3-L2-2	U3-A3-L2-3	U4-A4-L2-1	U4-A4-L2-2	U4-A4-L2-3
C	U1-A1-L3-1	U1-A1-L3-2	U1-A1-L3-3	U2-A2-L3-1	U2-A2-L3-2	U2-A2-L3-3	U3-A3-L3-1	U3-A3-L3-2	U3-A3-L3-3	U4-A4-L3-1	U4-A4-L3-2	U4-A4-L3-3
D	U1-A1-L4-1	U1-A1-L4-2	U1-A1-L4-3	U2-A2-L4-1	U2-A2-L4-2	U2-A2-L4-3	U3-A3-L4-1	U3-A3-L4-2	U3-A3-L4-3	U4-A4-L4-1	U4-A4-L4-2	U4-A4-L4-3
E	U1-A1-L5-1	U1-A1-L5-2	U1-A1-L5-3	U2-A2-L5-1	U2-A2-L5-2	U2-A2-L5-3	U3-A3-L5-1	U3-A3-L5-2	U3-A3-L5-3	U4-A4-L5-1	U4-A4-L5-2	U4-A4-L5-3
F	U1-A1-L6-1	U1-A1-L6-2	U1-A1-L6-3	U2-A2-L6-1	U2-A2-L6-2	U2-A2-L6-3	U3-A3-L6-1	U3-A3-L6-2	U3-A3-L6-3	U4-A4-L6-1	U4-A4-L6-2	U4-A4-L6-3
G	U5-A5-L1-1	U5-A5-L1-2	U5-A5-L1-3	U5-A5-L3-1	U5-A5-L3-2	U5-A5-L3-3	U5-A5-L5-1	U5-A5-L5-2	U5-A5-L5-3	NTC-A1-1	NTC-A1-2	NTC-A1-3
H	U5-A5-L2-1	U5-A5-L2-2	U5-A5-L2-3	U5-A5-L4-1	U5-A5-L4-2	U5-A5-L4-3	U5-A5-L6-1	U5-A5-L6-2	U5-A5-L6-3	NTC-A1-4	NTC-A1-5	NTC-A1-6

Legend
U = Unit A = Assay L = Level T = Temp

Figure B1 PCR Plate Layout for certification measurements. Each plate contains all 5 units and 6 levels of triplicates, rotating through different assays (A) on different plates.

This material is intended to be used for calibrating qPCR assays, which is a log-scale measurement (base 2). Discussions with the EPA indicated an absolute uncertainty of approximately 20 % would be acceptable (i.e. nearly impossible to detect via qPCR), with 10 % preferred. Previous SRMs 2365 and 2366a, both utilizing ddPCR as the measurement,

showed a relative standard uncertainty of approximately 2.2 % and 1.5 %, respectively. This indicates that 10 % should be attainable.

Homogeneity

The homogeneity of the material will be established through measuring an additional 10 units using 2 of the 5 assays from the certification measurements. These additional 10 units, combined with 5 previously, give 15 units for homogeneity. There were approximately 1050 vials produced per level, so we would examine approximately 1.4 %. Each level was produced in a single session from a single pot of material using robotic controls. We would not expect variability of the material in any tube due to the automated nature of manufacture, hence a 1 % sampling seems acceptable.

2 (assays)
10 (units)
6 (levels)
3 (replicates)

= 360 reactions

+ 12 NTC (6 replicates per assay)

==========

372 reactions

This fits on 4 96-well standard PCR plates. We anticipate running a single assay on each plate, which should allow detection of any variability outside of measurement uncertainty. The assays can be run over (2 to 4) days.

Using approximately 5 µL per reaction will consume 30 µL of each standard. These units can be carried forward for stability testing.

Stability

Stability of the material will be evaluated by measuring 5 times over approximately 14 weeks, at time points 0, 2, 6, 10, and 14 weeks. Post-certification measurements are planned for 6 months.

We will be evaluated at two temperatures: 4 °C and approximately 21 °C (room temperature). Two (2) units at each temperature will be measured in triplicate using one ddPCR assay.

1 (assays)
2 (temperatures)
2 (units)
6 (levels)
3 (replicates)

= 72 reactions per time point (1 96-well PCR plate)

4 (additional timepoints, using homogeneity as t=0)

==========

288 reactions

The two temperatures represent the prescribed storage conditions (4 °C), and the condition of a material being left on the bench in a lab (room temperature) for an extended period of time. The second serves as an accelerated stability test. Previous stability tests of similar DNA-based materials have not shown significant degradation even at 37 °C over longer time periods. We feel room temperature stability testing is sufficient, as it is a condition where degradation from biological activity is possible and most closely represents the most likely mishandling of the SRM unit.

Other DNA-based SRMs stability performance indicates that 5 years of stability at 4 °C can be reasonably expected if SRM 2917 can pass these stability conditions.

Conservatively, material use will be limited to 150 µL out of the approximately 200 µL in the tube to allow reanalysis if needed. The homogeneity study will consume approximately 30 µL. Each time point will consume approximately 15 µL, resulting in 90 µL consumed, below our limit. Indeed, these materials can be carried forward post-certification for extended stability analyses.

Appendix C: Measurement Report

Measurement report
3/25/2020
Natalia Farkas

The average volume for aqueous droplets dispersed in a propriety oil phase was determined according to Special Test Method 11050S/-D and is reported for the following protocols:

	Protocol	Average volume and expanded uncertainty
S1	DNA, 'UNG' MM	0.7663 nL ± 2.3 % (k=2)
S2	no DNA, 'UNG' MM	0.7686 nL ± 2.3 % (k=2)

The detailed measurement results are shown in Table 1. In addition to the average volume, the average diameter values as an equivalent spherical diameter obtained from the concentrated method are also included in Table 2. Note that the diameter values are not corrected for the volumeof oil phase remaining trapped at the vertices and edges of the Kelvin cells.

Sufficient statistics was assured by preparing and measuring the S1 and S2 samples on two different days with 3 replicate measurements per day. Standard deviations in the tables from left to right refer to measurement repeatability, within day reproducibility, and day-to-day variation. The expanded uncertainty of the volume measurements includes additional uncertainty components per Special Test Method 11050S/-D.

Table 1. Average volume results obtained by the concentrated method of the Special Test Method11050S/-D.

S1_3/11/2020	Average volume (nL)	St dev (nL)	Sample average (nL)	St dev (nL)	Protocol average (nL)	St dev (nL)
Measurement 1	0.7698	0.0058	0.7621	0.0078	0.7663	0.0060
Measurement 2	0.7624	0.0112				
Measurement 3	0.7542	0.0055				
S1_3/13/2020	Average volume (nL)	St dev (nL)	Sample average (nL)	St dev (nL)		
Measurement 1	0.7654	0.0043	0.7706	0.0065		
Measurement 2	0.7778	0.0050				
Measurement 3	0.7684	0.0052				

S2_3/11/2020	Average volume (nL)	St dev (nL)	Sample average (nL)	St dev (nL)	Protocol average (nL)	St dev (nL)
Measurement 1	0.7669	0.0061	0.7651	0.0038	0.7686	0.0050
Measurement 2	0.7607	0.0079				
Measurement 3	0.7676	0.0052				
S2_3/13/2020	Average volume (nL)	St dev (nL)	Sample average (nL)	St dev (nL)		
Measurement 1	0.7718	0.0081	0.77	0.0005		
Measurement 2	0.7721	0.0081				
Measurement 3	0.7728	0.0050				

Table 2. Average diameter results obtained by the concentrated method of the Special TestMethod 11050S/-D.

S1_3/11/2020	Average diameter* (um)	St dev (um)	Sample average (um)	St dev (um)	Protocol average (um)	St dev (um)
Measurement 1	114.47	0.28	114.09	0.39	114.30	0.30
Measurement 2	114.10	0.55				
Measurement 3	113.69	0.28				
S1_3/13/2020	Average diameter* (um)	St dev (um)	Sample average (um)	St dev (um)		
Measurement 1	114.26	0.22	114.51	0.32		
Measurement 2	114.87	0.24				
Measurement 3	114.41	0.26				

S2_3/11/2020	Average diameter* (um)	St dev (um)	Sample average (um)	St dev (um)	Protocol average (um)	St dev (um)
Measurement 1	114.33	0.30	114.24	0.19	114.41	0.25
Measurement 2	114.01	0.39				
Measurement 3	114.37	0.25				
S2_3/13/2020	Average diameter* (um)	St dev (um)	Sample average (um)	St dev (um)		
Measurement 1	114.57	0.40	114.59	0.02		
Measurement 2	114.59	0.40				
Measurement 3	114.62	0.25				

**uncorrected for the volume of the remaining oil phase*

Appendix D: Homogeneity Testing

Homogeneity testing

Blaza Toman

Testing of data in "EPA_ddPCR_R_homogeneity", done using one-way ANOVA with Tukey pairwise comparisons.

1. Level1 – no significant heterogeneity between units detected.
2. Level2 – no significant heterogeneity detected in Rum2bac.
   Unit 23 is significantly different from unit 8 in EC23S

Unit	Mean	Sd	n
5	52.3	2.6	3
6	47.9	2.9	3
8	47.8	5.3	3
12	49.9	3.1	3
16	55.4	4.0	3
22	54.4	0.8	3
23	60.7	4.3	3
33	52.5	2.7	3
37	55.3	3.4	3
43	58.3	5.4	3

Table 1. Means for EC23S Level2 homogeneity data

3. Level3 – No significant heterogeneity detected in Rum2bac, or in EC23S.
4. Level4 – No significant heterogeneity in EC23S. In Rum2bac, unit 6 is significantly different from units 16, 22, 37, and 43.

Unit	mean	sd	n
5	5123	188	3
6	4458	575	3
8	5112	81	3
12	5229	210	3
16	5305	87	3
22	5343	117	3
23	5044	148	3
33	5286	93	3
37	5469	296	3
43	5376	125	3

Table 2. Means for Rum2bac level4

5. Level5 – no significant heterogeneity in Rum2bac or Ec23S.

6. Level6 – significant differences in Rum2bac, unit 22 is significantly different from units 5, 16, and 23.

Unit	Mean	Sd	n
5	5030	236	3
6	5465	192	3
8	5458	99	3
12	5632	365	3
16	5036	92	3
22	5929	689	3
23	5009	58	3
33	5622	94	3
37	5729	190	3
43	5279	186	3

Table 3. Rum2bac means level6.

Significant differences in Ec23S for unit 37 different from units 5, 6, and 8.

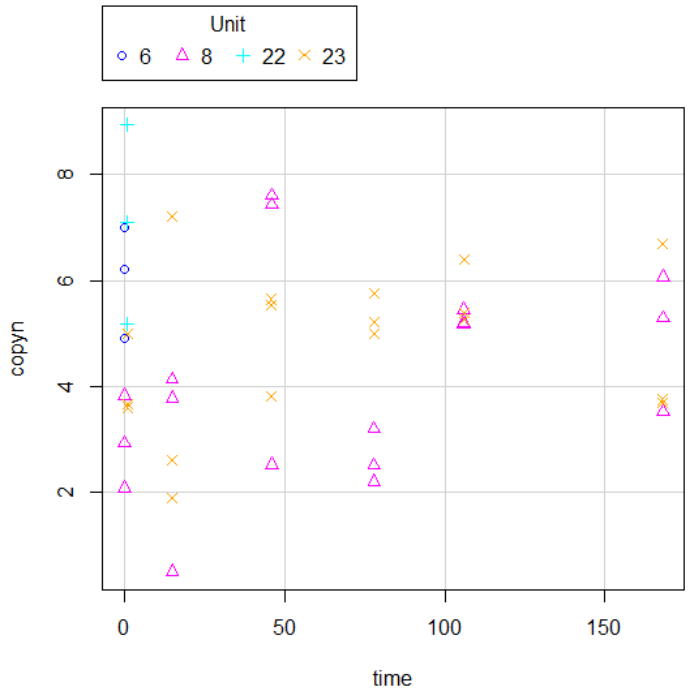
Unit	Mean	Sd	n
5	5081	164	3
6	4961	205	3
8	5174	154	3
12	5582	73	3
16	5231	127	3
22	5471	1047	3
23	5405	135	3
33	5485	280	3
37	6962	1515	3
43	5579	333	3

Table 4. Ec23S means for level6.

Appendix E: Stability

Stability testing
7/12/2021

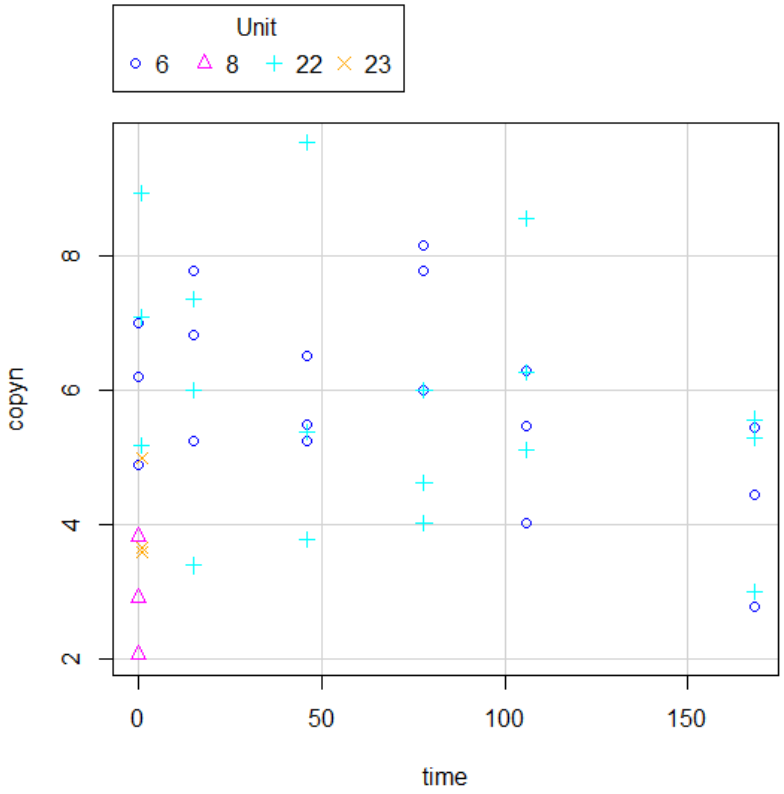
A linear hierarchical model with a unit effect, and a time effect was fitted (using OpenBUGS [1]) to 12 data sets, one for each level and temperature setting. Namely, the observed measurement of copy number for the i[th] unit at the j[th] time is
$y_{ij} \sim N(a_i + b_i\, time_j, \sigma^2)$, where $b_i \sim N(\beta, \tau_i^2)$, for $i = 1,2$.
The following plots show the data set. The tables give the slope b_i , and the consensus slope β estimates (with uncertainty) for each data set. The statistically significant ($\beta \neq 0$) slopes are marked with *. Both of these are Level 4. For these two cases a classical linear regression model was also fitted and in the case of Level 4, 20 degrees also turned out significant for this hypothesis test.
Level 1, 4 degrees



unit	slope	Slope uncertainty	L95	U95
1	0.001	0.008	-0.016	0.017
2	0.004	0.006	-0.009	0.016
consensus	0.002	0.140	-0.278	0.278

Table 1. Estimates of b_i and β for Level 1, 4 degrees.

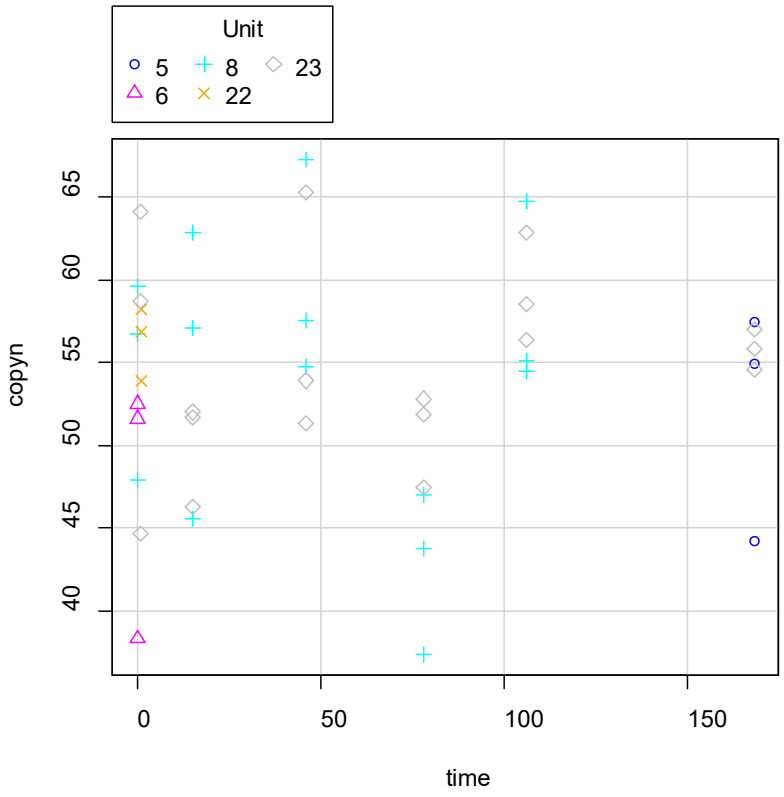
Level 1, 20 degrees



unit	slope	Slope uncertainty		L95	U95
1	-0.011	0.005		-0.022	-0.001
2	0.002	0.007		-0.012	0.017
consensus	-0.003	0.005		-0.012	0.007

Table 2. Estimates of b_i and β for Level 1, 20 degrees.

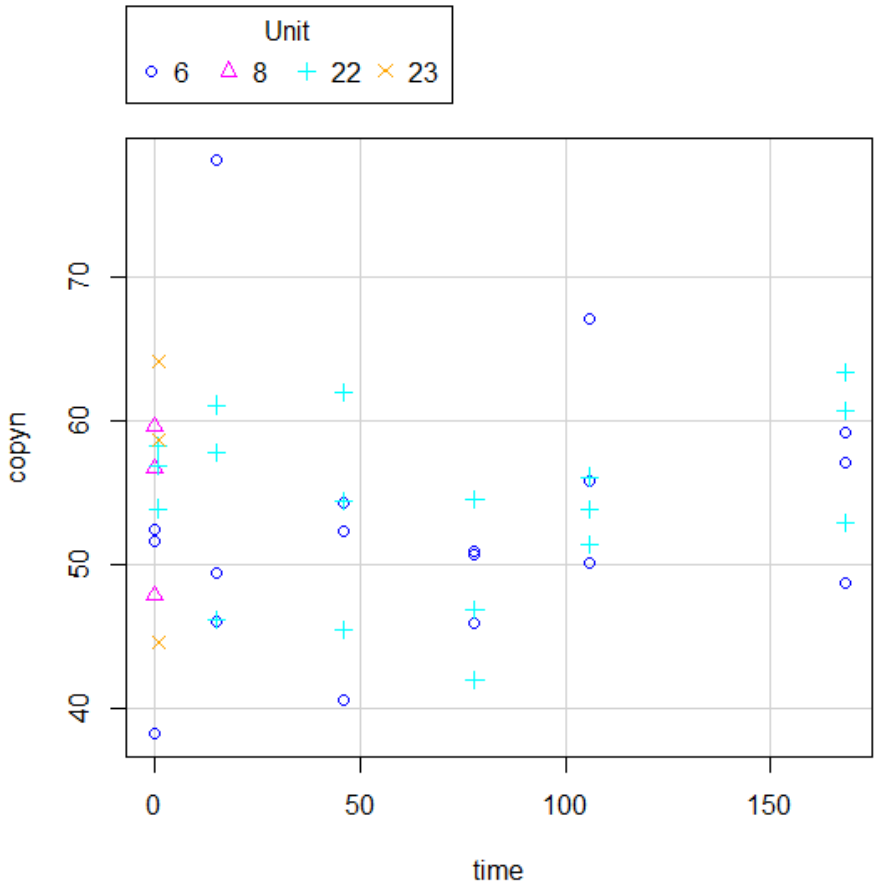
Level 2, 4 degrees



unit	slope	Slope uncertainty	L95	U95
1	0.007	0.036	-0.065	0.077
2	0.011	0.02	-0.028	0.049
consensus	0.009	0.207	-0.284	0.279

Table 3. Estimates of b_i and β for Level 2, 4 degrees.

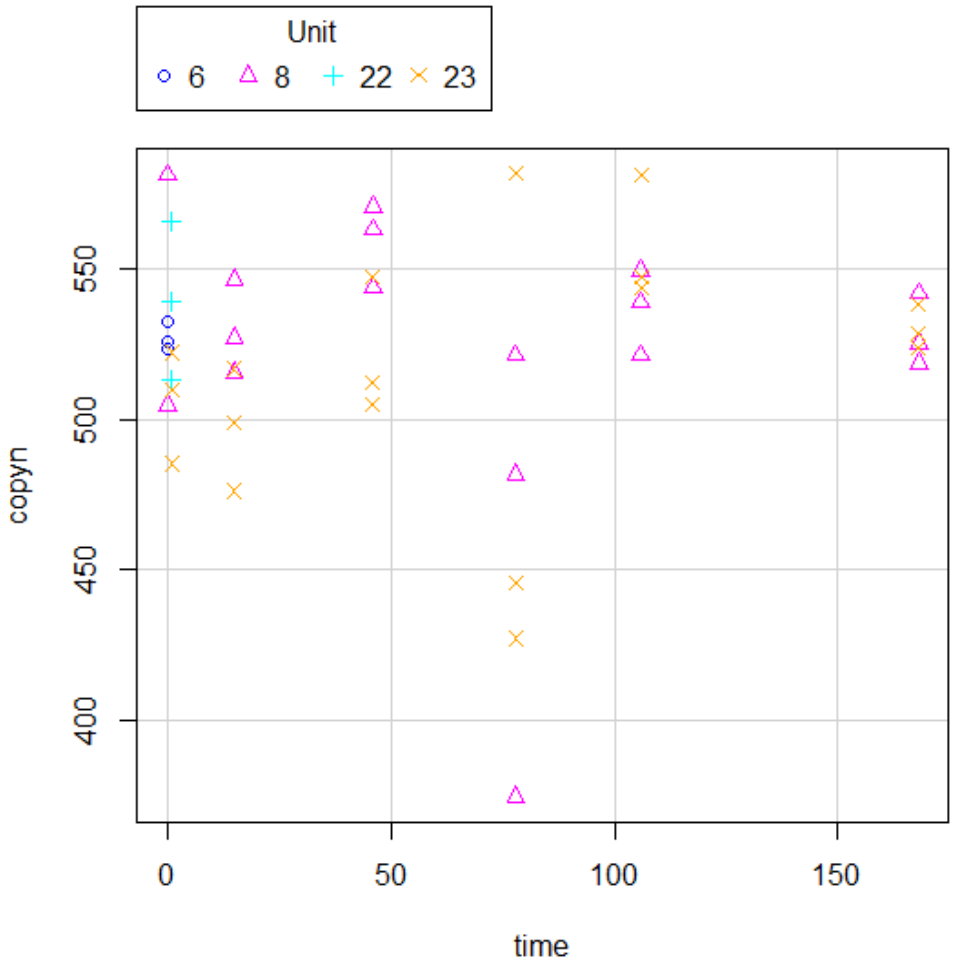
Level 2, 20 degrees



unit	slope	Slope uncertainty	L95	U95
1	0.028	0.038	-0.046	0.103
2	0.005	0.023	-0.04	0.051
consensus	0.016	0.152	-0.252	0.302

Table 4. Estimates of b_i and β for Level 2, 20 degrees.

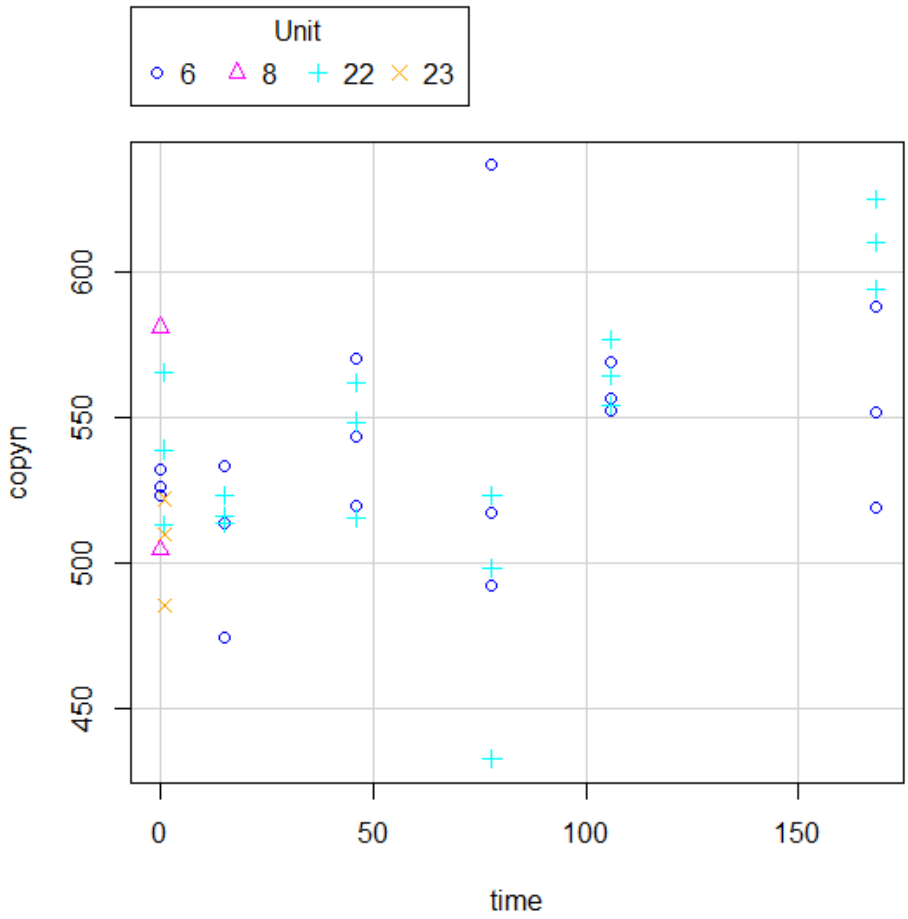
Level 3, 4 degrees



unit	slope	Slope uncertainty	L95	U95
1	-0.05	0.16	-0.36	0.26
2	0.09	0.13	-0.17	0.35
consensus	0.02	0.20	-0.36	0.42

Table 5. Estimates of b_i and β for Level 3, 4 degrees.

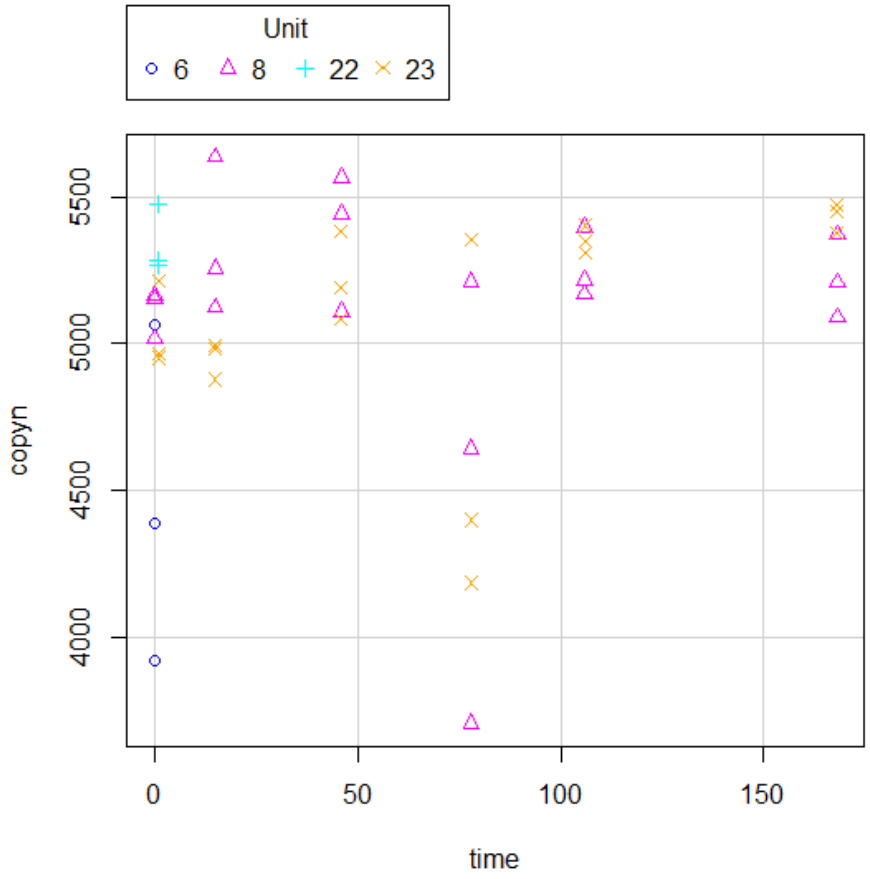
Level 3, 20 degrees



unit	slope	Slope uncertainty	L95	U95
1	0.25	0.14	-0.02	0.52
2	0.42	0.12	0.18	0.66
consensus	0.33	0.22	-0.05	0.71

Table 6. Estimates of b_i and β for Level 3, 20 degrees.

Level 4, 4 degrees

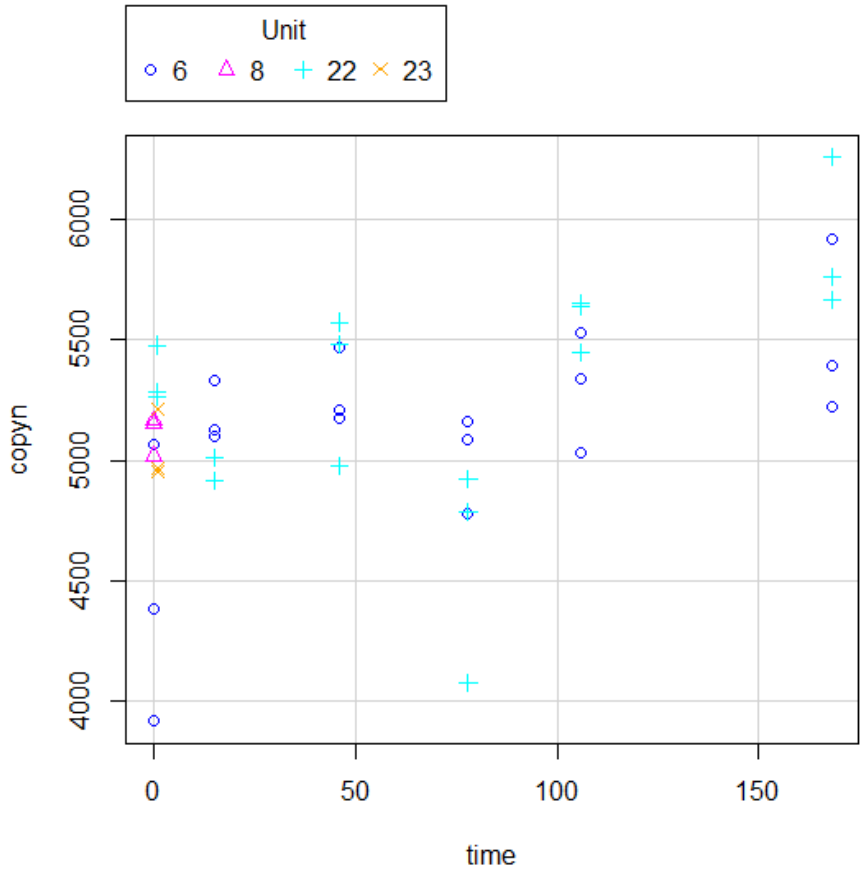


unit	slope	Slope uncertainty	L95	U95
1	2.69	0.47	1.77	3.61
2	1.67	0.36	0.90	2.33
consensus	1.37*	0.52	0.39	2.25

Table 7. Estimates of b_i and β for Level 4, 4 degrees.

Classical linear regression of copy number on time (ignoring unit) gives slope estimate of 1.43 with standard uncertainty of 1.12 and p-value of the hypothesis test that the slope equals 0 is p=0.21, that is, *not significant* at 0.05 level.

Level 4, 20 degrees

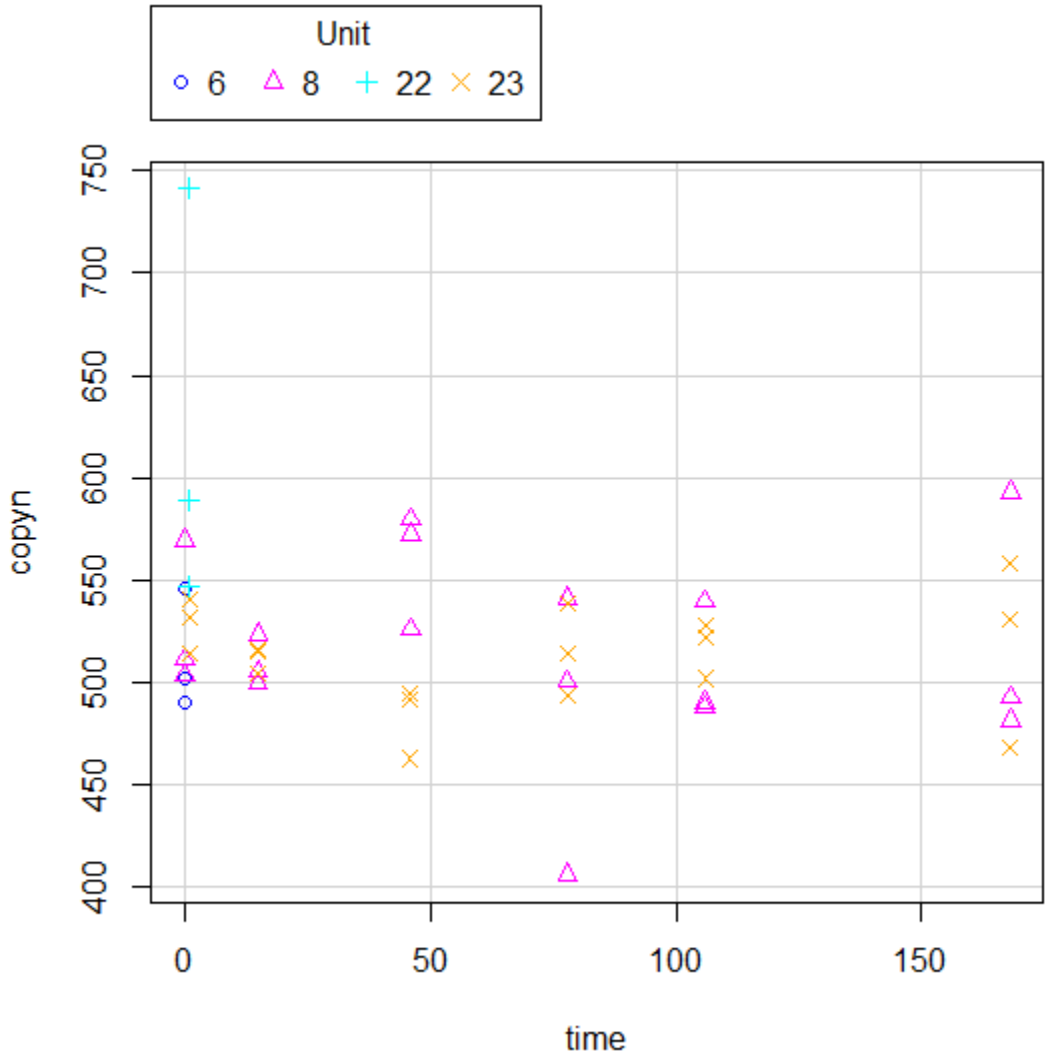


unit	slope	Slope uncertainty	L95	U95
1	4.81	0.41	4.01	5.63
2	4.09	0.38	3.35	4.85
consensus	4.43*	0.3792	3.734	5.151

Table 8. Estimates of b_i and β for Level 4, 20 degrees.

Classical linear regression of copy number on time (ignoring unit) gives slope estimate of 3.73 with standard uncertainty of 1.0 and p-value of the hypothesis test that the slope equals 0 is p=0.0006, that is, *highly significant*.

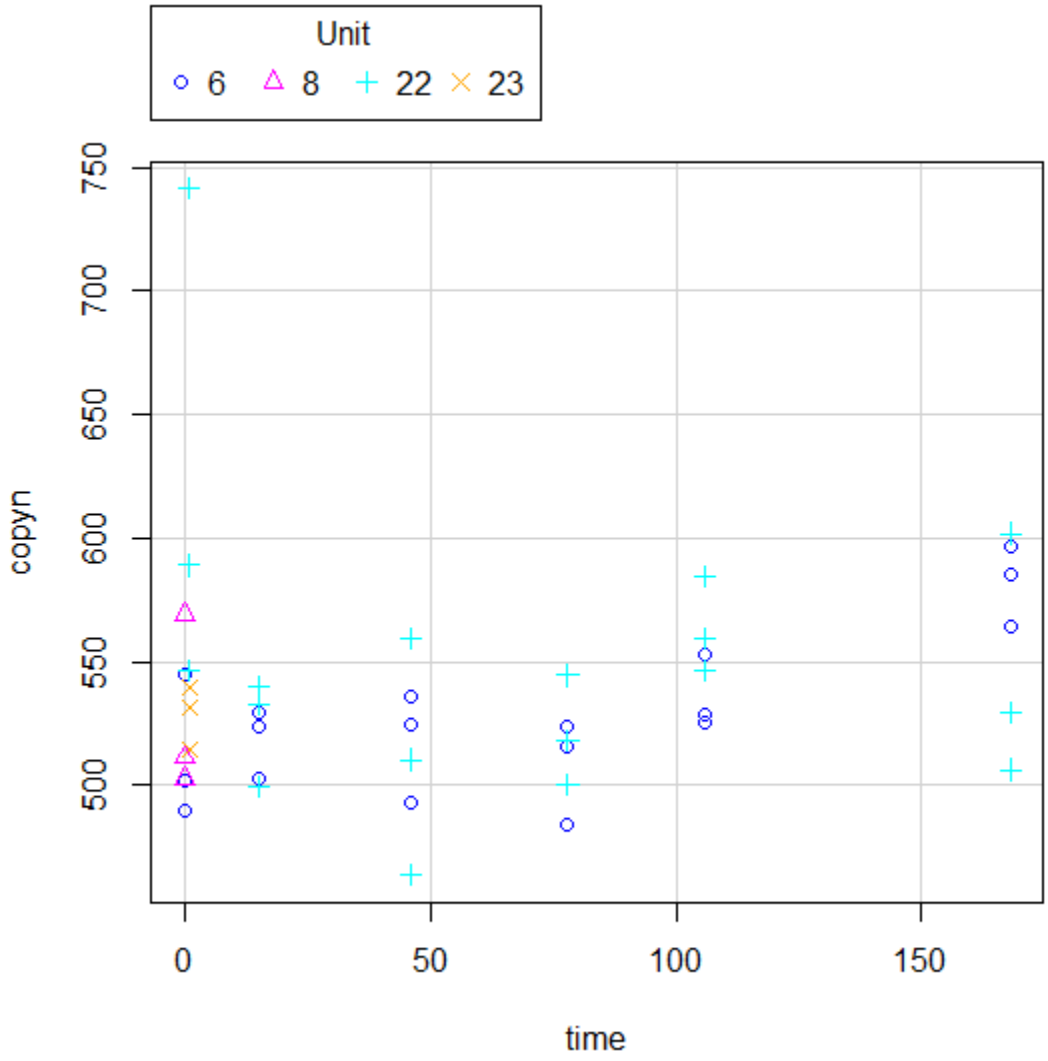
Level 5, 4 degrees



Unit	Slope	Slope uncertainty	L95	U95
1	-0.01	0.16	-0.34	0.3
2	-0.22	0.16	-0.54	0.08
Consensus	-0.12	0.23	-0.53	0.31

Table 9. Estimates of b_i and β for Level 5, 4 degrees.

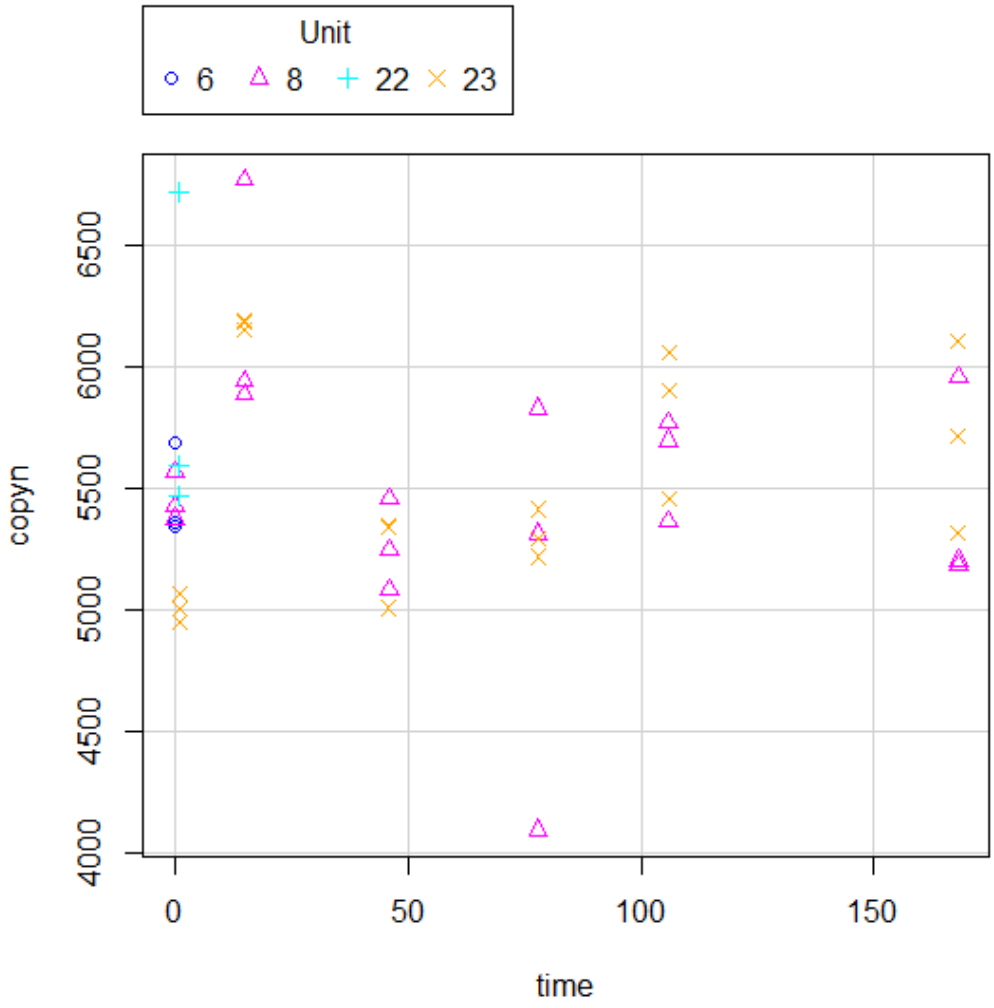
Level 5, 20 degrees



Unit	Slope	Slope uncertainty	L95	U95
1	0.38	0.09	0.2	0.56
2	-0.03	0.16	-0.35	0.28
consensus	0.21	0.26	-0.28	0.65

Table 10. Estimates of b_i and β for Level 5, 20 degrees.

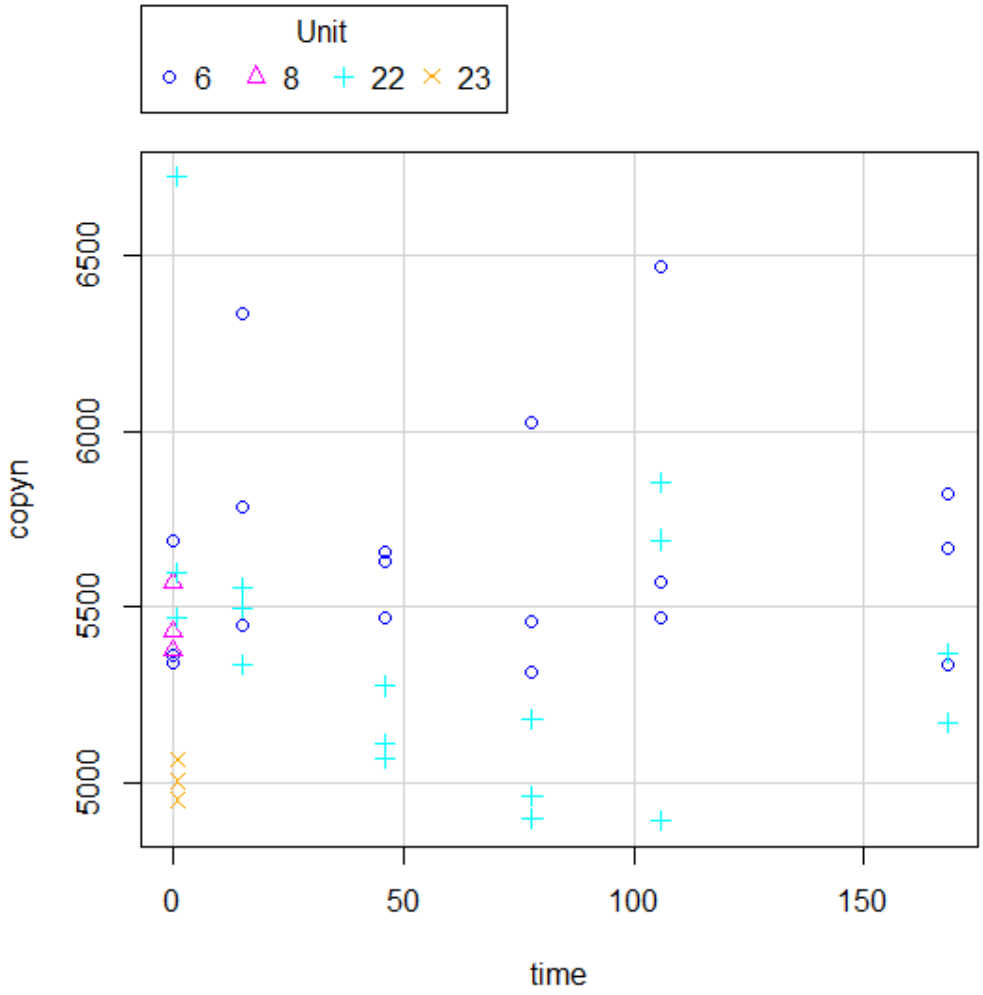
Level 6, 4 degrees



Unit	Slope	Slope uncertainty	L95	U95
1	-0.91	0.47	-1.84	0.02
2	1.17	0.4	0.38	1.98
consensus	0.20	0.73	-1.19	1.51

Table 11. Estimates of b_i and β for Level 6, 4 degrees.

Level 6, 20 degrees



Unit	Slope	Slope uncertainty	L95	U95
1	0.91	0.4	0.14	1.69
2	-0.94	0.38	-1.72	-0.2
consensus	-0.03	0.64	-1.20	1.26

Table 12. Estimates of b_i and β for Level 6, 20 degrees.

Reference

[1] Lunn, D., Spiegelhalter, D., Andrew, T. et al. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine 28*, 3049-3067. http://dx.doi.org/doi:10.1002/sim.3680

Appendix F: Analysis of SRM 2917 using Hierarchical Bayes Analysis

Analysis of SRM2917 using Hierarchical Bayes Analysis (with unit)

Blaza Toman, NIST

March 16,2021

The following statistical model was used to produce the estimates with uncertainty.

The response was represented as

$y_{ijkl} \sim N(\gamma_{ijk}, \sigma^2)$, i denotes assay ($i = 1, \ldots, 5$),
j denotes level (1 ,…, 6 in the order as above),
k denotes units (1 ,…, 5),
l denotes replicate

$\gamma_{ijk} \sim N(\alpha_{ij}, \sigma_{ij}^2)$, γ_{ijk} is the mean lambda value of assay i, level j, unit k.

$\alpha_{ij} \sim N(\mu_j, \sigma_j^2)$ α_{ij} is the consensus mean lambda value of assay i, level j

μ_j is the overall lambda value for level j

$V \sim dnorm(0.7663, 0.0088^2)$, represents droplet volume of DNA 'UNG' MM,

$\mu_{j(adj)} = {}^{5 * \mu_j}/_{V}$, represents the measurand; mean $copies/\mu L$ for each level. (The multiplication by 5 here is to follow Jason's R code, not to multiply by 5 to get EPA values)

Note that this statistical model accounts for uncertainty due to unit differences, assay differences, droplet volume, and repeatability.

	Assay	Mean $copies/\mu L$	Standard uncertainty	L95%	U95%
Level1	CowM2	4.9	0.5	4.0	5.9
	CowM3	5.3	0.6	4.2	6.4
	CPQ_056	6.2	0.8	4.7	7.9
	Ec23S857	5.2	0.6	3.9	6.5
	Rum2Bac	5.8	0.7	4.3	7.2
Level2	CowM2	52.7	1.5	49.8	55.5
	CowM3	58.5	1.4	55.7	61.1
	CPQ_056	60.1	1.7	56.7	63.4
	Ec23S857	52.6	1.3	50.0	55.2
	Rum2Bac	53.7	1.8	50.0	57.1
Level3	CowM2	525	11	503	548
	CowM3	531	10	514	552

	CPQ_056	541	11	519	561
	Ec23S857	519	11	499	541
	Rum2Bac	537	11	515	557
Level4	CowM2	5248	71	5114	5391
	CowM3	5248	90	5076	5432
	CPQ_056	5430	112	5183	5630
	Ec23S857	5350	64	5226	5477
	Rum2Bac	5250	91	5075	5428
Level5	CowM2	52583	1152	50564	55112
	CowM3	55257	1386	51709	56969
	CPQ_056	53585	1133	51220	55693
	Ec23S857	51062	880	49363	52788
	Rum2Bac	51954	987	50086	53947
Level6	CowM2	539474	8419	525807	560225
	CowM3	560668	9563	542189	579222
	CPQ_056	550210	8821	532114	567054
	Ec23S857	515993	7889	501111	531903
	Rum2Bac	556461	12994	525780	576248

Table 1. Estimates of copy/ µL by level and assay.

Level	Mean $copies/\mu L$	Standard uncertainty	L95	U95
level1	5.4	0.5	4.5	6.5
level2	55.5	1.6	52.5	58.7
level3	530	7	516	545
level4	5314	71	5175	5452
level5	52195	744	50779	53704
level6	541287	7779	527588	557912

Table 2. Estimates of copy/ µL by level.

Level	Assay	Mean copies/5 uL	Standard uncertainty	L95%	U95%
Level1	CowM2	24.7	2.4	20.0	29.5
	CowM3	26.4	2.9	20.8	32.2
	CPQ_056	31.1	4.1	23.3	39.3
	Ec23S857	26.1	3.2	19.7	32.4
	Rum2Bac	28.8	3.7	21.3	36.1
Level2	CowM2	263.3	7.3	248.8	277.7
	CowM3	292.4	6.9	278.4	305.4

	CPQ_056	300.3	8.5	283.5	316.8
	Ec23S857	263.2	6.6	250.2	276.2
	Rum2Bac	268.6	8.9	250.0	285.5
Level3	CowM2	2626	57	2514	2738
	CowM3	2656	49	2568	2759
	CPQ_056	2704	54	2593	2805
	Ec23S857	2594	54	2493	2705
	Rum2Bac	2686	54	2573	2783
Level4	CowM2	26241	355	25570	26954
	CowM3	26240	450	25378	27161
	CPQ_056	27148	562	25914	28151
	Ec23S857	26749	321	26131	27386
	Rum2Bac	26251	453	25373	27141
Level5	CowM2	262914	5758	252821	275561
	CowM3	276287	6932	258544	284844
	CPQ_056	267923	5666	256101	278467
	Ec23S857	255309	4400	246816	263942
	Rum2Bac	259769	4935	250432	269734
Level6	CowM2	2697370	42095	2629035	2801125
	CowM3	2803340	47815	2710945	2896110
	CPQ_056	2751050	44105	2660570	2835270
	Ec23S857	2579965	39446	2505555	2659515
	Rum2Bac	2782305	64972	2628900	2881240

Table 3. Estimates of copy/ 5 µL by level and assay.

Appendix G: R-code

Raw CSV data from the Quantasoft program was used to calculate the copy number concentration of the plasmid in solution. Note that the files must be updated for each sample set. The droplet volume, sample volume, and assay volume were all fixed. The resulting data were then saved to a new CSV file with concentration data from each individual well that can be evaluated by the statistician.

```
sapply(c('tidyverse','readxl','broom','rstudioapi','stringr'), require, character.only = TRUE)    #
Load packages
sapply(c('tidyverse','readxl','broom','rstudioapi','stringr'),  packageVersion)
theme_set(theme_bw() + theme(plot.title = element_text(size = 11, face = 'bold', hjust =
0.5)))
# Import the .csv datafile, or consider file.choose(); calculate lambdas
EPA.raw <- read_csv("2021-01-04-2917_stabilit_Plate11_info.csv") %>%
   mutate(fracPos = Positives/(Positives + Negatives),
         Lambda = -log(1-fracPos))

Avo=6.022140857e23
Sample.vol=5
Assay.vol=25

ddPCR.summary <- filter(EPA.raw, Sample!="NA") %>%
  mutate(Dilution.Factor = 1.000, Droplet.Vol = .7663) %>%
  mutate(
         Undiluted.Targets.per.Droplet = Lambda/Dilution.Factor,
         copy.per.ul.Original.Sample =
Undiluted.Targets.per.Droplet*(Assay.vol/Sample.vol/Droplet.Vol*1000),
         #copy conc = undiluted targets/droplet * reaction volume / DNA sample volume /
droplet volume (uL)
       ) %>%
select(Date,Operator,Sample,Unit,Level,Target,Positives,Negatives,Lambda,copy.per.ul.Orig
inal.Sample)
ddPCR.summary$Sample =
with(ddPCR.summary,reorder(Sample,copy.per.ul.Original.Sample,median))


# OUTPUT to .csv
write.csv(ddPCR.summary,file="EPA_ddPCR_R_stability_plate11.csv")
```