NIST Special Publication 260-225

Reference Material 8376
Microbial Pathogen DNA Standards for Detection and Identification

Jason Kralj
Dieter Tourlousse
Monique Hunter
Erica Romsos
Blaza Toman
Peter Vallone
Scott Jackson

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.260-225

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

NIST Special Publication 260-225

Reference Material 8376
Microbial Pathogen DNA Standards for Detection and Identification

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

Dieter Tourlousse
National Institute of Advanced Industrial Science and Technology, Japan

Erica Romsos, Peter Vallone
Biomolecular Measurement 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.260-225

June 2022

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

National Institute of Standards and Technology
Laurie E. Locasio, Director and Under Secretary of Commerce for Standards and Technology

National Institute of Standards and Technology Special Publication 260-225
Natl. Inst. Stand. Technol. Spec. Publ. 260-225, 51 pages (June 2022)
CODEN: NSPUE2

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.260-225

Abstract

Reference Material (RM) 8376 is intended for NGS-based measurements quantitative to the chromosome. A unit of RM 8376 consists of 20 components (A-T, 19 bacteria and 1 human) each containing well-characterized DNA in 10 mmol/L Tris-HCl, 1 mmol/L EDTA pH 8 buffer. Each component consists of one skirted 0.5-mL tube containing approximately 100 μL of DNA solution. Each tube is labeled and sealed with a removable screwcap. Here, we have documented the production, analytical methods, and statistical evaluation of realizing this RM.

Key words

Pathogen; DNA; next-generation sequencing; metagenomics; bacteria; reference material (RM).

Table of Contents

List of Tables

List of Figures

1. Introduction

RM 8376, Microbial Pathogen DNA Standards for Detection and Identification, was designed to provide a mechanism that enables users to make NGS-based measurements quantitative to the chromosome. The RM components' chromosomal copy number concentration values are traceable to a ddPCR method of analysis.

Each unit of RM 8376 consists of 20 individual vials (components) of microbial (19) and human (1) DNA in aqueous buffer. Each microbial component consists of 100 µL DNA solution at a nominal concentration of 50 ng/µL.

We report non-certified property values of RM 8376 for the chromosome copies per µL solution. Non-certified values of DNA concentration from absorbance and fluorescence spectroscopy and assembled genomes and sizes are also given. Users of the material are typically familiar with the absorbance and fluorescence measurements; hence these are provided for ease of use. Despite lacking traceability, the three methods for DNA concentration agree well. Genome sizes and sequences also compare well with published genomes of the same organisms.

2. Storage and Use

Handling: RM 8376 IS A BACTERIAL DNA SOURCE MATERIAL. SINCE THERE IS NO CONSENSUS ON THE INFECTIOUS STATUS OF EXTRACTED DNA, HANDLE RM 8376 AS BIOSAFETY LEVEL 1 MATERIALS CAPABLE OF TRANSMITTING INFECTIOUS DISEASE. RM 8376 and derived solutions should be disposed of in accordance with local, state, and federal regulations.

Storage: Vials of RM 8376 should be stored in the dark between 2 °C to 8 °C. **DO NOT FREEZE**.

Use: The vial should be mixed by **horizontal vortexing** for 5 seconds and briefly centrifuged without opening the vial cap prior to removing sample aliquots for analysis (Fig. 1). For the reference values to be valid, materials should be withdrawn immediately after opening the vials and processed without delay. Dilutions of these materials may be made as appropriate, but they must be used immediately. Reference values do not apply to any material remaining in recapped vials. Do not expose any DNA solution to direct sunlight.

Figure 1 Horizontal vortexing a component for 5 seconds will properly mix the DNA. Heterogeneity within the tube solution can occur after storage at 4 °C.

3. Materials and Methods

3.1. DNA production

The DNA for each component was extracted from bacteria. The 19 strains of bacteria were obtained from the American Type Culture Collection[1] (ATCC) (Table 1).

Table 1 Genome assemblies. Genome assemblies of the 19 bacterial genomes with ATCC ID, RM Part, Name, chromosome size, and plasmid count. Mean coverage for assemblies were approximately 125-fold using the highest quality reads from PacBio, and approximately 180-fold Illumina HiSeq.

ATCC ID	Part	Organism	Chr Size(s)	Plasmids
43895	A	*Escherichia coli* O157:H7	5564632	2
BAA 2309	B	*Escherichia coli* O104:H4	5302905	1
700720	C	*Salmonella enterica* enterica	4857492	1
12324	D	*Salmonella enterica* arizonae	4482096	0
BAA 44	E	*Staphylococcus aureus*	2964115	1
12600	F	*Staphylococcus aureus*	2755072	1
12228	G	*Staphylococcus epidermidis*	2504458	3
BAA 47	H	*Pseudomonas aeruginosa*	6263669	0
19606	I	*Acinetobacter baumannii*	3980879	0

[1] Disclaimer: Any mention of commercial products within this publication is for information only; it does not imply recommendation or endorsement by NIST.

13077	J	*Neisseria meningitidis*	2181327	0
12344	K	*Streptococcus pyogenes*	1914863	0
19433	L	*Enterococcus faecalis*	2866948	0
27061	M	*Achromobacter xylosoxidans*	6813185	0
35654	N	*Aeromonas hydrophila*	4733720	0
13883	O	*Klebsiella pneumoniae*	5303036	4
25931	P	*Shigella sonnei*	4917056	0
35016	Q	*Vibrio furnissii*[a]	3275680, 1641536	1
19115	R	*Listeria monocytogenes*	2950983	0
33152	S	*Legionella pneumophila*	3409194	0

[a]*Vibrio furnissii* has 2 circular chromosomes

The bacteria were cultured, and the DNA extracted by the Microbe Inotech Laboratory in St. Louis, MO. These methods were proprietary. A total of 10 mg DNA of each component was delivered. Each component was assayed using ddPCR, absorbance, and fluorescence measurements to estimate concentration before diluting with 1× Tris-EDTA buffer to a nominal concentration of 50 ng/μL in 1× Tris-EDTA buffer in 360 mL Perfluoroalkyoxy polymer (PFA) jars (Savillex).

3.2. Packaging

We employed a Scinomix VXL automated liquid handling robot with at least two staff members performing setup. Each component was brought to room temperature on a stir plate set to low. Liquid handling is managed using a peristaltic pump with silicone tubing (Tygon, McMaster-Carr). The tubing was rinsed by recirculating 50 mL of Milli-Q filtered water for at least 5 min (more than 30 volumes of the tubing), dispense calibration was checked and adjusted to ensure 110 μL of volume was dispensed, then rinsed using fresh DI water and pumped dry. The DNA material was recirculated for approximately 5 minutes (approximately 5 times the volume of the container).

700 new aliquots were made in fresh 0.5 mL centrifuge tubes (USA Scientific). Each centrifuge tube contains at least 100 μL of 50 ng/μL DNA. The robot dispensed 100 μL of the liquid material, printed a pre-approved label, and applied the label to each centrifuge tube before racking. If the robot was disrupted due to mechanical error, we noted & discarded the next 2 tubes.

3.3. Analyses

3.3.1. DNA absorbance & fluorescence

The DS-11 FX+ (Denovix, CT) was used for absorbance and fluorescence measurements, and was located in 227/B255. The Qubit 2.0 (Thermo Fisher Scientific) was also used for fluorescence measurements. For absorbance spectroscopy, an average extinction coefficient 1 OD = 50 ng/μL for double-stranded DNA was used. For fluorescence spectroscopy, the instrument manufacturer's standards were used.

Samples were brought to ambient temperature for at least 30 minutes. Vials of DNA solution were vortexed horizontally for at least 5 s to ensure complete homogenization, then briefly centrifuged (i.e. quick spin) in a tabletop centrifuge to collect the contents.

For absorbance measurements, the default double-stranded DNA analysis was used. Each day a blank of 1x TE buffer was used. The TE buffer was measured to check measurement drift. The listed uncertainty is 1 % at this optical density. For each measurement, 2 µL of solution were used.

For fluorescence measurements, master mix and a standard curve was prepared following the manufacturer's recommendation using Denovix dsDNA HS reagents (Denovix, CT). The basic ratio is 200 µL of assay buffer to 2 µL of dye. Typically, enough assay is prepared for all samples (usually 4 mL of buffer and 40 µL of dye), then aliquoted to individual assay tubes at (200 µL – sample volume). For measurements of samples, (2 to 5) µL of sample was used to remain below the 250 ng DNA limit for this assay. Two Denovix instruments were used for this analysis on the same samples, and differences in the measured concentrations between two instruments were less than the manufacturers' specified uncertainty of 5 %.

3.3.2. Sequencing & Assembly

Whole genome "shotgun" sequencing of the DNA was performed using the Illumina MiSeq for 2x300 base sequencing (i.e. "short read") at NIST (227/A262), Illumina HiSeq for 2x150 base sequencing ("short-read") and Pacific Biosciences systems ("long read") at the Institute for Genome Sciences (IGS), University of Maryland (Luke Tallon and Lisa Sadzewicz). MiSeq short read samples were prepared using a multiplex barcoding kit (B) with the Nextera XT kit. DNA was purified using SPRISelect beads (B23317 and B23318, Beckman Coulter, Indianapolis, IN). HiSeq samples were prepared using the KAPA Library Preparation kit (Woburn, MA) with Covaris E210 DNA fragmenting and purification between steps with SPRIselect beads. DNA concentrations were measured using the Qubit (fluorescence) and roughly adjusted depending on the number of samples so that approximately 15 pmol were loaded per run.

IGS produced the long-read sequence data using the Pacific Biosciences sequencer, and the data was transmitted to NIST. Genome assembly using the data is described in the next section. The raw sequence data is stored on a RAID on the NIST network and on the local workstation where the assembly was performed.

The data were stored locally and analyzed. The assemblies are stored on our RAID backup system and NCBI (BioProject: PRJNA605254).

Illumina sequencing reads were processed using fastp (version 0.19.6), [1] including removal of sequencing adapters (specifying command line option --detect_adapter_for_pe) and quality trimming and filtering of low-quality reads (command line options --max_len1 150 --max_len2 150 --cut_right --cut_window_size 4 --cut_mean_quality 20 --qualified_quality_phred 15 --unqualified_percent_limit 40 --length_required 100 --n_base_limit 0 --low_complexity_filter). For PacBio reads, Filtlong (version 0.2.0; https://github.com/rrwick/Filtlong) was used to retain the highest quality read bases (minimum read length of 1,000 bp) with an estimated coverage of 125× based on the

projected genome size, using the quality-controlled Illumina reads as references. For Nanopore reads, seqtk was used to retain reads longer than 5000 bp, yielding approximately k reads and approximately 880 Mbases with an estimated coverage of 150×.

For all genomes, except for *Escherichia coli* O157:H7 long-read assemblies were generated using Flye (version 2.5), [2] starting from the 125× of high-quality PacBio reads; four rounds of polishing were performed by specifying option --iterations 4. Assembled genome sequences were then further polished using Illumina short reads (randomly subsampled to a projected coverage of 180×) with Pilon (v1.23). [3] This was performed using the Unicycler (version 0.4.7) [4] pilon_polish.py command and entailed several rounds of polishing until no changes occurred.

For the genome of *Escherichia coli* O157:H7, the original genome assembly did not resolve all ambiguity, so more long-read data from the Nanopore sequencer was taken to solve problematic regions. A long-read assembly was generated using Flye as described above with the filtered Nanopore reads and specifying the option --plasmids. The assembly was then polished using the PacBio reads using Flye with options --polish-target --iterations 4, followed by multiple rounds of polishing using Pilon with the Illumina short reads as above.

The resulting genomes were circular and some contained plasmids. The identity of these were checked using NCBI BLAST to ensure they matched the corresponding organism. These genomes are considered improved high-quality drafts (IHQ draft) and may contain errors. While we have used the best available methodologies and compared sizes with known high-quality completed genomes (NCBI), we do not currently have a method for determining an uncertainty for genome size or assembly. These genomes (both sequence and sizes) are considered useful by the users for mNGS comparisons and are therefore included for informational value as non-certified data.

3.3.3. Droplet-digital PCR

The QX200 ddPCR system (Bio-Rad, Hercules, CA) was used for stability and certification measurements. The methodology is described in NIST Special Publication 260-189. [5] Reaction volume was previously determined for this lot of master mix by NIST Physical Measurement Laboratory, Semiconductor and Dimensional Metrology Division, Nanoscale Metrology Group (NMG) 683.03, as described in NIST Special Publication 260-184. [6] The volume was (0.7349 ± 0.0085) nL for droplets (average volume and expanded uncertainty, k=2).

The 20 PCR assays are described in Table 2. These were generated using contigs from assembled genomes using Primer3 [http://bioinfo.ut.ee/primer3-0.4.0/], including forward and reverse primers and hybridization probe. Default parameters were used except as indicated below:

Product Size Ranges:	80-200
Table of thermodynamic parameters:	SantaLucia 1998
Salt correction formula:	SantaLucia 1998
Concentration of divalent cations:	1.5
Hyb Oligo Tm:	Min 65, Opt 67, Max 69
Hyb Oligo conc of divalent cations:	1.5

Microbial primers and probes were subjected to BLAST (NCBI) with default parameters with the following exception: Word size = 7. Targets showing multiple copies or lacking specificity to the species level were discarded.

Human primers were supplied by E. Romsos, which were generated for SRM 2372a and detailed in NIST Special Publication 260-189. [5]

Table 2 PCR assays for 20 components

Organism	Type	ID	Sequence
Escherichia coli O157:H7	Fwd	O157-f-13807	AGAGTGACGCTGTACTGACC
	Rev	O157-r-13961	GATGGTGCGGGAAAACCAAT
	Probe	O157-13848	ACGCCCTGCCTCATCCGGAT
Escherichia coli O104:H4	Fwd	O104-f-7893	CTCATGCAGGGTTTTGTCCC
	Rev	O104-r-7981	TGGGGAAGTTGCCCGATATT
	Probe	O104-7929	TGGGTACTGACCAGCCTGGTGA
Salmonella enterica enterica	Fwd	141a-F	CGGGCATACCATCCAGAGAA
	Rev	139a-R	CACCGTGGTCCAGTTTATCGT
	Probe	140-P	AATCGGGCCGCGACTTCCGC
Salmonella enterica arizonae	Fwd	Sea-f-3606	GTATGTTCTGGGACCCGTCA
	Rev	Sea-r-3799	GAAGAAAATCAGGCCGGCAA
	Probe	Sea-3675	TGATGTGGGCGGTTTGCGCA
Staphylococcus aureus HPV107	Fwd	SA442-BS-F	AAAGCGGGCACTTGGATGAAT
	Rev	SA442-BS-R	AAACCGGGACCAATTTGGAAA
	Probe	SA442-BS-P	ACAGGTGAAGGTGGCTTATCAGAATATCA
Staphylococcus aureus	Fwd	Sa12600-f-869	TATGGGCTTTTAGTGCGGGA
	Rev	Sa12600-r-1030	GAAGAACAGCAGGGGATTGC
	Probe	Sa12600-924	AGCAGCACGCCACCTAACGA
Staphylococcus epidermidis FDA strain PCI 1200	Fwd	Sepi-f-1009	TCAGGTGATGCATATCCAGGAA
	Rev	Sepi-r-1175	ACGTTTTCAGTTTGGCCGTT
	Probe	Sepi-1081	TGGGCCCAAGGATCAGGCACACCAAGT
Pseudomonas aeruginosa HER-1018 [PAO1]	Fwd	GryPA-398-A-F	CCACAACAAGGTCTGGGAAC
	Rev	GryPA-620-A-R	CCAGGATGTCCCAACTGAAG
	Probe	GryPA-P	GGAGACCTTCAGCAACATCC
Acinetobacter baumannii 2208	Fwd	Ab-f-7926	CCAACCAGACTCAGGTCGAT
	Rev	Ab-r-8049	CCGACACTACACTTTGCCAC
	Probe	Ab-7961	ACACACGAACAGGCACAGGATGC
Neisseria meningitidis M1027	Fwd	Nm-f-5339	TGTTCCATTTCATGCTGCCC
	Rev	Nm-r-5457	GTACCGGCAAAACGGATGTT
	Probe	Nm-5400	TCCCGCCAGCAATCAAACAGCT
Streptococcus pyogenes Typing strain T1	Fwd	Sp-f-14922	TGATGATTGGTCACGTCGGA
	Rev	Sp-r-15119	GGGCCAAAACGATGCCTAAA
	Probe	Sp-14971	TGCGGCAGCTTTTTCTGGCCA
Enterococcus faecalis NCTC 775	Fwd	ef-f-48228	CCGTGAAAGCAGCAACAGAT
	Rev	ef-r-48390	AATGAACCTACCTCTGCCCC
	Probe	ef-48248	GCTGGTGCTGCTGCAGCAGA
Achromobacter xylosoxidans KM 543	Fwd	Ax-f-2031	GCCAGGTCGATGTTTTCCAG
	Rev	Ax-r-2133	AAAAGCTGATGCCGGTGAAG
	Probe	Ax-2084	GGTGGCCTTGCCGTCTTCGG
Aeromonas hydrophila LRA 3300 776	Fwd	Ah-f-2271	CTTGTCAACCTCAACACCCG
	Rev	Ah-r-2449	TACTTGCTGCCGTACTCGAA

	Probe	Ah-2336	CGGCAACCTGGATCACGCCC
Klebsiella pneumoniae NCTC 9633	Fwd	kp-f-18772	TGCCGGAAGGGTATGACATT
	Rev	kp-r-18961	ACAAGGTTAAATGCAGCCCG
	Probe	kp-18822	TTGGAAGGGGAGCGGCACCT
Shigella sonnei NCDC 1120-66	Fwd	Ss-f-2702	CCTCCCCCTGGTGTTCTTAG
	Rev	Ss-r-2811	CCCGGAGAAGAGGGCTTTTA
	Probe	Ss-2750	TCACCAGTATGGTGACGTGCTTCA
Vibrio furnissii 9119-82	Fwd	Vf-f-8708	ATTGACCGGAATCATGGGGT
	Rev	Vf-r-8903	TTGAGCCCGGATTTCTGGAT
	Probe	Vf-8786	GGATGCCCGGTCACGAACACG
Listeria monocytogenes Li 2	Fwd	Lm-f-4979	TCGGTCGTAGTATCGCCTTC
	Rev	Lm-r-5128	AGGCGTCCAAGTATGCTTCT
	Probe	Lm-5084	ACGCTCTAATGGAAGAAGCGCAACG
Legionella pneumophila Philadelphia-1	Fwd	Lp-f-2901	ACCGCGACTAATACCCAACA
	Rev	Lp-r-3059	TTGGTTACCCGCTTTCTTGC
	Probe	Lp-2974	AGAGGTGGCTCAACTTCTGACAGGA
Human NEIF Gene EIF5B	Fwd		GCCAAACTTCAGCCTTCTCTTC
	Rev		CTCTGGCAACATTTCACACTACA
	Probe		CTCTGGCAACATTTCACACTACA
Human POTP STR TPOX	Fwd		CCACCTTCCTCTGCTTCACTTT
	Rev		ACATGGGTTTTTGCCTTTGG
	Probe		CACCAACTGAAATATG
Human ND6 STR D6S474	Fwd		GCATGGCTGAGTCTAAAGTTCAAAG
	Rev		GCAGCCTCAGGGTTCTCAA
	Probe		CCCAGAACCAAGGAAGATGGT
Human HBB1 Gene HBB	Fwd		GCTGAGGGTTTGAAGTCCAACTC
	Rev		GGTCTAAGTGATGACAGCCGTACCT
	Probe		AGCCAGTGCCAGAAGAGCCAAGGA

Primers and probes were purchased from Biosearch Technologies (Middlesex UK), with probe labels of 5'-FAM for fluorescence reporting and 3'-BHQplus-1 quenching. These were diluted to 100 μmol/L in Milli-Q filtered water (18.2 MΩ-cm).

At least two assays were tested against the desired microbial targets using ddPCR assay conditions previously optimized. Absorbance and fluorescence data, combined with sequence data showing purity, indicate approximate target concentration. Concentrations cannot exceed previous methods by more than 50 %, as this indicates a multi-copy target or poor dilution/mixing of the sample. Estimated concentrations should be the same or slightly lower, owing to the possible presence of plasmid DNA that is not directly assayed. Most assay pairs agreed with each other and with previous data. Assays that did not agree were discarded (no signal or multiples of expected concentration) and/or redesigned.

Briefly, single-target PCR assays are mixed for each component of DNA (Table 3). An emulsion of droplets containing a limiting dilution of DNA was then generated for each sample. Those samples were then thermal cycled in a 96-well plate (Eppendorf, Cat # 951020303). The droplets are then read for fluorescence (either positive or negative) and counted.

Each assay's primers and probes were mixed and diluted 20-fold in the prescribed ratio to reduce pipetting steps and volumetric errors, and stored at 4 °C. The composition was kept the same for all assays (see Table 3).

Table 3 PCR assay composition

Manufacturer	Part #	Reagent	µL/rxn
Bio-Rad	186-3010	ddPCR Supermix (no dUTP) for probes	12.5
Biosearch Technologies	Custom, salt-free	Forward primer (375 nM)	5.0
Biosearch Technologies	Custom, salt-free	Reverse primer (375 nM)	
Biosearch Technologies	DLO-FBP-5 BHQplus	Probe (250 nM)	
Water (18.2 MΩ/cm)			2.5
		DNA sample	5.0
		TOTAL	25

DNA samples were prepared by diluting samples 10^4:1 in 1x TE buffer pH 7.4. This was accomplished by twice pipetting 10 µL of sample (or diluted sample) into 990 µL of TE buffer. Samples were vortexed thoroughly after each dilution. Once-diluted samples may be used for up to 24 h. The twice-diluted samples were used immediately.

PCR assays were prepared in 96-well PCR plates (Eppendorf, Cat # 9551020303), and heat-sealed with foil (Bio-Rad #1814040). Plates were vortexed and centrifuged to homogenize all wells. Sample plates were loaded onto an AutoDG Droplet Digital PCR Generator (Bio-Rad), run using pre-programmed settings, with droplet-generated samples collected in a fresh 96-well PCR plate. These were heat-sealed with a foil and immediately transferred to a thermal cycler for amplification.

Amplification on a Veriti Thermal Cycler (Life Technolgies) was as follows: 95 ºC for 10 min, followed by 60 cycles of (15 s at 94 ºC and 1 min at 60 ºC; ramp rate = 1 °C/s). After the 60 cycles endpoint there was a 98 ºC hold for 10 min, followed by a 4 °C hold until the samples were removed from the thermal cycler and put onto the QX200 droplet reader (Bio-Rad). Plate information was entered into the QuantaSoft (v.1.7.4, Bio-Rad) software and analyzed. A table of the raw occupancy data were exported as .csv files and analyzed using a custom R-script (see Appendix A) to generate chromosomal copy number and estimated DNA concentration data.

3.3.3.1.Quantitation

The characterization of the genome copy number concentration (C_g) is determined using ddPCR. Typically, ddPCR requires a sample be diluted (<<1 copy per droplet) to obtain optimal performance.

$$C_{dd} = \varphi C_g \tag{1}$$

where φ is the dilution factor, and C_{dd} is the digital droplet concentration (copy/µL).

The value for C_{dd} is determined using Poisson statistics for the probability distribution of droplets being filled at random by copies of the DNA target. We can use

$$E = \frac{C^n e^{-C}}{n!} \tag{2}$$

E is the probability at n, C is the copies of DNA per droplet (not volume), and n is the number of events. Which for n=0 (unoccupied droplets) becomes:

$$E = e^{-C} = 1 - \frac{N_+}{N} \tag{3}$$

for N being the total number of droplets, and N_+ the number of occupied (fluorescent) droplets. Solving for C_g we obtain:

$$C_g = \frac{-ln\left(1-\frac{N_+}{N}\right)}{\varphi \cdot V_{dd}} \tag{4}$$

Where V_{dd} is the droplet volume.

These calculations are carried out using a custom R-script (Appendix A).

3.3.3.2.Uncertainty

Metagenomics applications typically involve order-of-magnitude levels of precision and accuracy because existing materials and methods have not been available to improve upon this.

The uncertainty was estimated using a simple propagation model without interactions:

$$\mathrm{CV}_g = \frac{\sigma_g}{C_g} = \sqrt{\left(\frac{\sigma_N}{C_N}\right)^2 + \left(\frac{\sigma_\varphi}{\varphi}\right)^2 + \left(\frac{\sigma_V}{V_{dd}}\right)^2} \tag{5}$$

The dilution factor was known to be the highest source of uncertainty, with CV_φ of approximately 5 %, due to the number of pipetting steps needed to achieve a measurable concentration even if performed perfectly. As such, care must be taken to ensure the equipment is properly calibrated and the users trained, as small errors in pipetting will propagate through the measurement. The standard deviation in droplet volumes can also be estimated (1 % from the droplet volume certification). Variability in relative droplet counts is also low in this range (assume 1 %). Using eq. (5), we would expect a pooled CV of approximately 5.2 %.

In terms of fitness for purpose, a 10 % CV would enable high-confidence measurements for order-of-magnitude analyses.

3.4. Homogeneity

We measured the initial homogeneity of the material using DNA absorbance, DNA fluorescence, and ddPCR using the methods described above. 10 samples selected from the same position in boxes (1 to 7) (and a second from boxes 2, 3, 6 labeled 2B, 3B, and 6B, respectively) from each component were measured. Absorbance and fluorescence values were recorded using two operators, on two (identical) instruments, on multiple days.

The ddPCR analyses for the microbial components were run until triplicate measures from each tube were collected (30 total measures). In the event a PCR failed, that was recorded, and assays were re-run. Two operators on multiple days were employed. For the human component, 3 assays in triplicate were used on each tube.

The ddPCR concentration data for each of the twenty organisms was analyzed using a two-way mixed-effects ANOVA with box being the fixed effect.

3.5. Stability

Samples 2 and 2B were placed at 37 °C. Samples 6 and 6B were placed in a drawer at room temperature (22 °C). The other 6 samples (1, 3, 3B, 4, 5, 7) were stored at 4 °C. Approximately once per month, samples from boxes (1, 2, 3, 4, 6B) were measured by ddPCR. The others remained closed to serve as backup samples if needed. We measured each component from the 5 test boxes using ddPCR once on two separate days.

The component Lm from sample set 6B was lost in January 2020 due to tube seal failure and no measurements could be made. The box 6 tube replaced it.

Components in box 2 (37 °C) appeared nearly depleted in February 2020, and initial ddPCR results suggested some evaporation had occurred. Box 2B samples were assayed to test whether temperature had affected concentration, and would be used for the depleted box 2 samples going forward.

A linear hierarchical model with a main effect, a term for box effect, and a time effect was fitted (using OpenBUGS) to all 20 data sets. Namely, the observed measurement of concentration for the *i*th box at the *j*th time is

$$y_{ij} \sim N\left(\mu_i + \alpha\, time_j, \sigma^2\right), \text{ where } \mu_i \sim N(\theta, \tau_i^2), \text{ for } i = 1, \ldots, 7. \qquad (6)$$

Cases where the coefficient α was approximately equal to 0 would establish stability. Further details are given in Appendix B.

4. Results & Discussion

4.1. Genome Assembly

The 19 microbial genomes were successfully assembled into unambiguous chromosomes and plasmids. The genome sizes ranged from approximately 2 Mbases to 6.8 Mbases, and were consistent with other related strains or species in the NCBI genome (https://www.ncbi.nlm.nih.gov/genome/). The results are shown in Table 1.

4.2. Absorbance

Triplicate measures were made on three separate days (Figure 2) using two operators and two different Denovix readers. The CV across all samples averaged approximately 1.1 %, which is consistent with the reported CV of the instrument.

These values were included because many laboratories employ this technique to estimate DNA concentration. However, these values are informational and not the non-certified values assigned to this material.

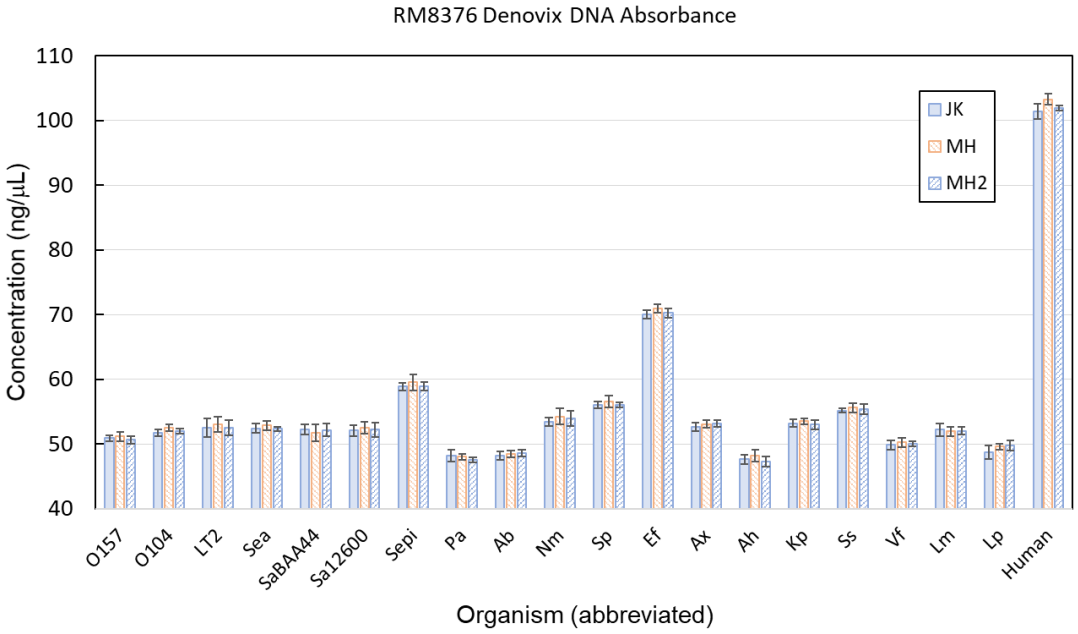


Figure 2 DNA absorbance for the 20 components in the RM. Each measurement represents the average of 10 tubes, with error bars equal to 1 sd. Two operators (JK and MH) assayed the samples, on different machines (denoted by color). The global CV was 1.1 %, with the listed instrument CV of 1 %.

4.3. Fluorescence

Triplicate measures were made on three separate days, using two operators (JK & MH) and two different Denovix readers (Figure 3). On the first day, the same samples were measured on two different machines and found to give the same measures. The CV across all samples ranged from approximately 3 % to 5 %, which is consistent with the CV reported by the instrument and method. In general, we would expect fluorescence measurements to be the same or slightly lower than the absorbance measurements because fluorescence is specific to double-stranded DNA, as opposed to chemical entities that absorb in the 260 nm wavelength (nucleotides, single-stranded DNA, and many organics and chaotropic agents used in purification).

There were slight difference between the absorbance and fluorescence measures for each component. The most notable were *S. epidermidis*, *S. pyogenes*, *and E. faecalis*. We suspect single-stranded DNA or nucleotides may have carried through the purification process in these components. As we will see with the droplet-digital PCR measurements, these did not significantly affect our ability to quantify the chromosomal copy number.

Again, these values were included because many laboratories employ fluorescence quantitation to estimate DNA concentration. However, these values are informational and not the non-certified values assigned to this material.

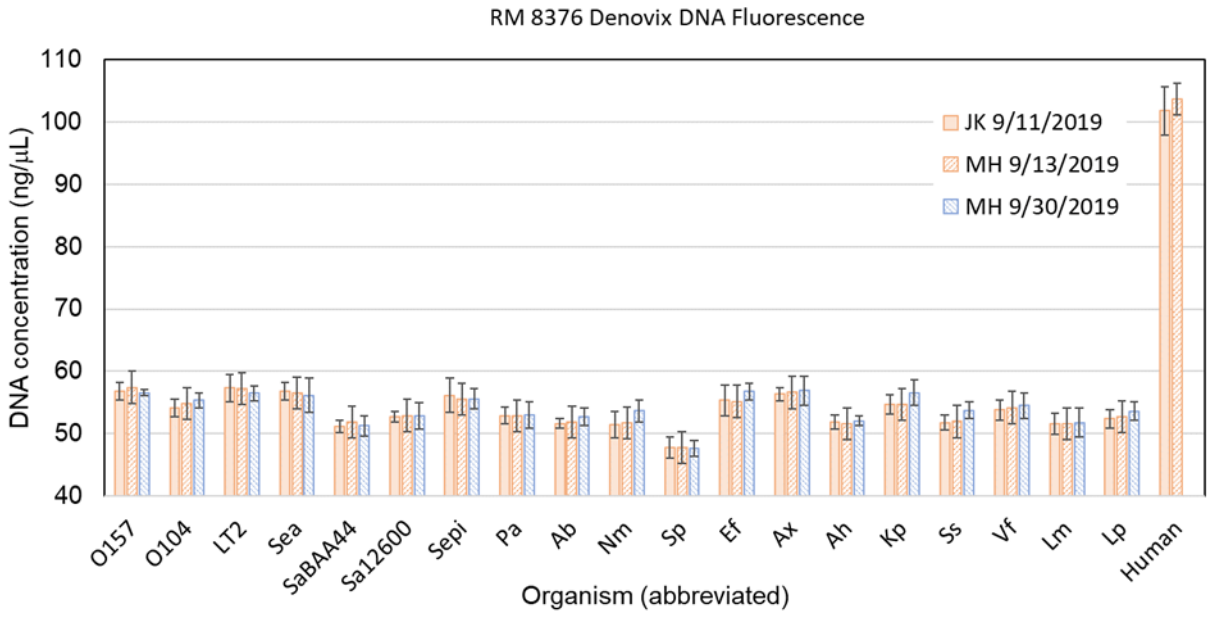


Figure 3 DNA Fluorescence measurements for 20 components of the RM. Each measurement represents the average of ten tubes, with error bars equal to 1 sd. Two operators assessed the values on three different days. The CV is consistent with pipetting error (5 %) across all samples.

4.4. Droplet-digital PCR

Triplicate measures of bacterial DNA were made on each sample over several days by two operators (JK & MH) (Figure 4). Samples were separated into two sets: (1, 2, 3, 4, 6B) and (2B, 3B, 5, 6, 7). 95 assays were run (19 targets and 5 samples for each), utilizing a 96-well plate. The CV across all samples ranged from approximately 6.6 % to 12.4 %, with a mean of 8.4 %. We estimated an uncertainty of 8.7 %. This is higher than either absorbance or fluorescence assays.

In general, we would expect ddPCR estimates for DNA concentration to be the same or slightly lower than either absorbance or fluorescence concentrations due to the ddPCR having the highest specificity (approximately a single 150 bp DNA target). However, that comparison requires conversion from copy number concentration to mass concentration using the assembled genome size and may contain additional uncertainty beyond the uncertainty of the ddPCR measurements.

Human DNA ddPCR assays were performed by Erica Romsos. The data showed (Figure 5) that the human DNA was homogeneous, as the CV was approximately 6 %. The human component from 3B and 7 were excluded, due to insufficient sample pre-mixing. This was corrected in subsequent stability measurements. Absorbance measurements made after ddPCR corroborated that the DNA concentration was the same as previous measurements and between tubes.

The full list of ANOVAs for each of the 20 components is listed in the Appendices B & C.

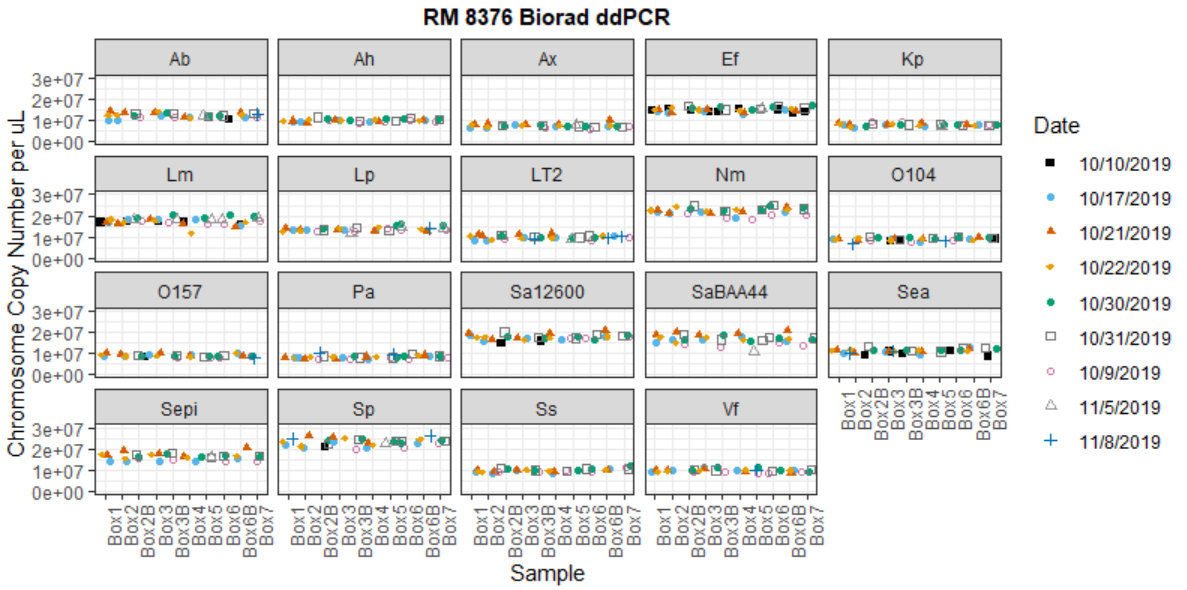


Figure 4 Homogeneity of 8376 components, shown as chromosome copy number concentration vs. Box Sample. Data points are individual reactions.

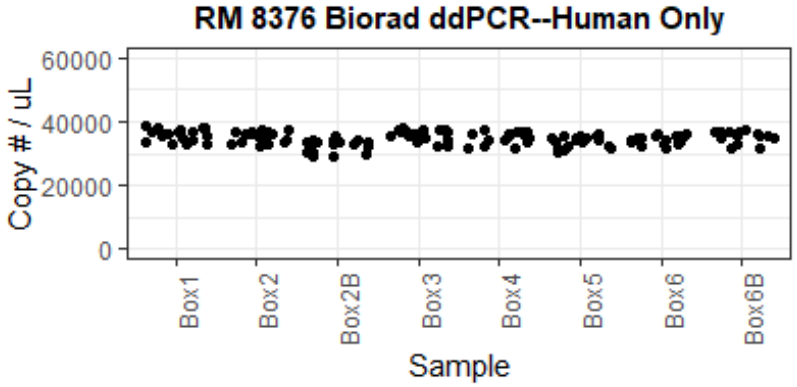


Figure 5 Homogeneity of 8376 Human DNA (component T) by ddPCR. Data points are individual reactions. These are shown separately from the bacteria because of the difference in scale.

The DNA concentration measurements for the three methods used shown in Fig. 6 show that the three methods largely agree. As mentioned before, the DNA absorbance and fluorescence measures were included as they are commonly employed by end-users, and more likely to be used to verify component concentrations than ddPCR.

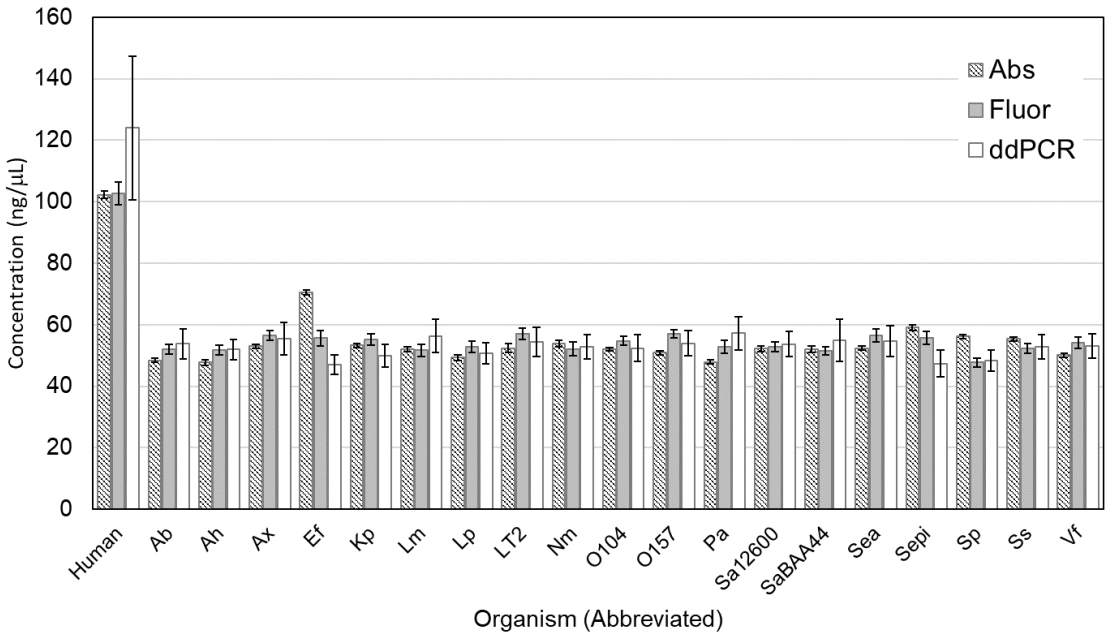


Figure 6 Comparison of DNA concentration from three methods. Values for ddPCR were converted to mass concentration by scaling with the genome size to allow comparison between the methods.

4.5. Stability

The bottling of the material was finished in August 2019 and stored at 4 °C. We then measured the material using ddPCR approximately every month from November 2019 through February 2020.

The null hypothesis was that time would not affect the ddPCR measurement. In all 20 cases, the coefficient α in Eq. (6) was approximately equal to 0 (see Appendix C). Thus, there were not statistically significant differences over time for any of the components, and thus we cannot reject the null hypothesis. Fig. 7 shows the measured values for each component over time.

The component Lm from box 6B was lost in January 2020 due to tube seal failure, and no measurements could be made. The box 6 tube replaced it going forward.

In February 2020, Box 2 samples were found to be low volume, owing in part to having been sampled numerous times. Initial results suggested their concentration had increased, likely due to evaporation from mis-sealed caps. Box 2B backup samples were assayed alongside Box 2 samples to test if there were any effects on the concentration due to temperature Box 2B samples agreed with the other 4 assayed boxes and were lower concentration than Box 2. We therefore rejected the null hypothesis that there were no differences between Boxes 2 and 2B and used Box 2B going forward as the 37 °C sample set.

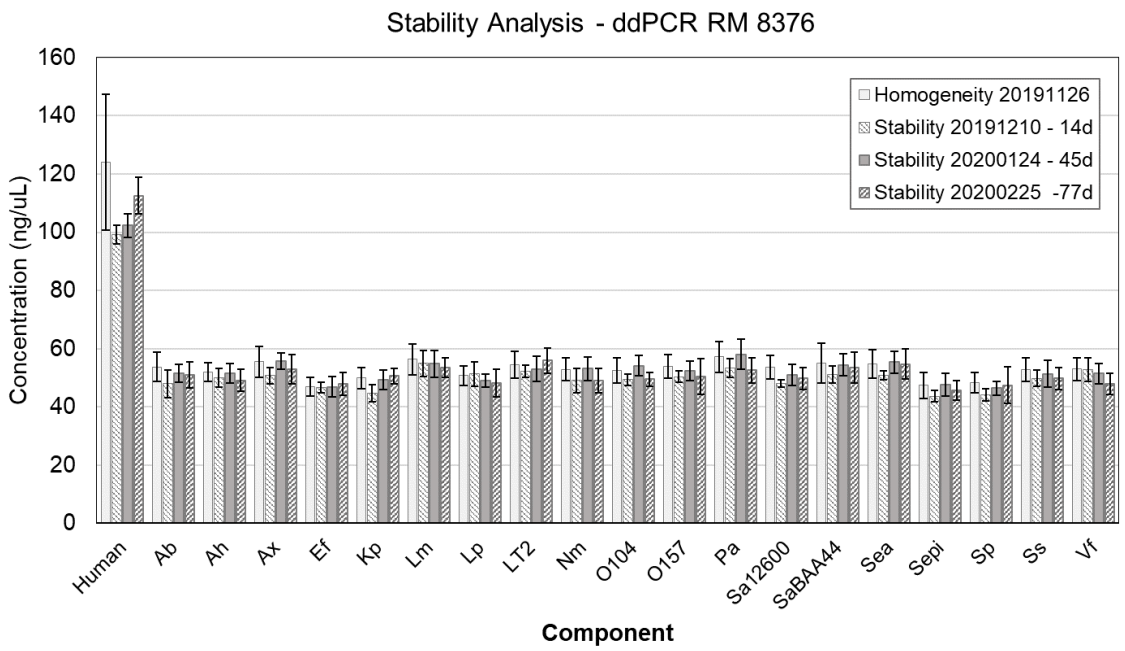


Figure 7 Results of ddPCR measurements over time. There were not statistically significant differences over time for any of the components. Note that the data were converted to mass concentration by scaling with the genome size to enable comparison of all components on the same scale.

4.6. Chromosome copy number concentration (CC) values

The measured values for CC are listed in Table 4.

The statistical model given in Eq. (6) was used for the copy number concentration data of all 20 organisms to estimate the value θ, and the corresponding standard uncertainty.

To show the variability that a customer may observe, the predicted copy number concentration:

$$y_p \sim N(\theta, \sigma^2) \tag{7}$$

was also estimated, and the resulting predictive uncertainties are listed in the last column of Table 4.

As described previously, the mass concentration can be estimated using the genome size, molecular weight of DNA base pairs, and Avogadro's number. Except for the human DNA, the CC ranges from approximately (7.3 to 23) million copy per μL.

Table 4 Chromosomal copy number concentration values for the 20 components and the corresponding standard uncertainty. To show the variability that a customer may observe, the

predicted copy number concentration was also estimated, and the resulting predictive uncertainty is given in the last column

ATCC ID	Part	Organism	Copy # Concentration $\times 10^6$ (copy/µL)	Standard Uncertainty $\times 10^6$ (copy/µL)	Predictive Uncertainty $\times 10^6$ (copy/µL)
43895	A	*Escherichia coli* O157:H7	8.84	0.19	0.82
BAA 2309	B	*Escherichia coli* O104:H4	8.89	0.14	0.80
700720	C	*Salmonella enterica* enterica	9.72	0.19	0.85
12324	D	*Salmonella enterica* arizonae	10.84	0.26	1.24
BAA 44	E	*Staphylococcus aureus*	16.49	0.38	1.82
12600	F	*Staphylococcus aureus*	17.38	0.34	1.62
12228	G	*Staphylococcus epidermidis*	15.99	0.30	1.48
BAA 47	H	*Pseudomonas aeruginosa*	8.27	0.17	0.76
19606	I	*Acinetobacter baumannii*	12.01	0.28	1.29
13077	J	*Neisseria meningitidis*	21.67	0.47	2.28
12344	K	*Streptococcus pyogenes*	22.55	0.43	2.10
19433	L	*Enterococcus faecalis*	14.75	0.25	1.23
27061	M	*Achromobacter xylosoxidans*	7.28	0.18	0.81
35654	N	*Aeromonas hydrophila*	9.97	0.17	0.68
13883	O	*Klebsiella pneumoniae*	7.68	0.18	0.76
25931	P	*Shigella sonnei*	9.67	0.18	0.80
35016	Q	*Vibrio furnissii*	9.70	0.18	0.80
19115	R	*Listeria monocytogenes*	17.39	0.32	1.61
33152	S	*Legionella pneumophila*	13.63	0.23	0.99
--	T	Human[a]	0.0323	0.00075	0.0023

[a] Coriell Strain GM 27385

5. Summary

The ddPCR, absorbance, and fluorescence measurements of the DNA concentrations indicate that the material is homogenous. For each measurement mode, we observed CVs corresponding to the measurement error.

There were slight discrepancies between methods. In most cases, the 3 methods gave a consensus value for the DNA concentration within error of the measurement. However, there are notable differences for *E. faecalis* (Ef, component L), *S. epidermidis* (Sepi, component G), and *S. pyogenes* (Sp, component K).

For Ef and Sp (components L and K, respectively), there appears to be a non-DNA impurity that interferes with the absorbance measurement. We have not found that it affects DNA sequencing. We suspect that this impurity resulted from the DNA purification, and that sequencing workflows can tolerate this level of contamination with minimal/no effect. Since we are certifying chromosomal copy number based upon the ddPCR value, and absorbance and fluorescence values are for informational purposes, we provide this information for reference purposes.

Sepi (component G) showed good agreement between fluorescence and absorbance, but lower estimated values for ddPCR. Sequencing results indicate the presence of a significant

amount of plasmid DNA, which was not assayed in the ddPCR method. One could estimate the relative plasmid abundance to the chromosome and derive an effective genome size, but the uncertainty associated with this is unknown and we cannot report a value with confidence. As before, we are certifying chromosomal copy number with ddPCR, so we included this data for reference.

The assembled genomes (improved high-quality drafts) can be found in BioProject: PRJNA605254; or locally on our RAID system.

Again, we emphasize that **proper (horizontal) mixing of the samples before any use is critically important**, because failure to homogenize the components before use may result in significantly altering the concentration of the remaining material.

The data indicate that the material is stable. An accelerated stability study indicates that the components stored at temperatures ranging from (4 to 37) °C did not change in concentration over approximately 3 months for this study (November 2019 through February 2020). Previous RM and SRM work indicated that DNA solutions within this concentration range would be stable, and our data do not contradict this.

Hence, we find RM 8376 to be homogeneous and stable, and recommend transferring it to the care of Standards Division for distribution.

6. References

[1] Chen S, Zhou Y, Chen Y, Gu J. 2018 fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34(17):i884-i890.

[2] Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546.

[3] Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9(11):e112963.

[4] Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595.

[5] Romsos EL, Duewer DL, Toman B, Farkas, N. "Certification of Standard Reference Material® 2372a." NIST Special Publication (NIST SP) 260-189 (2018). https://doi.org/10.6028/NIST.SP.260-189.

[6] Dagata JA, Farkas N, Kramar JA. "Method for Measuring the Volume of Nominally 100 μm Diameter Spherical Water-in-Oil Emulsion Droplets." NIST Special Publication (NIST SP) 260-184 (2016). https://doi.org/10.6028/NIST.SP.260-184.

Acknowledgments

Thanks to S. Servetas for careful reading and review of the manuscript.

Appendix A: Rcode

Below is an example of the R-studio markdown file. This was run using Rstudio and the additional packages listed within the script. In **BOLD** are code that need updating based on file names & experimental conditions.

````
---
title: "20191011-ddPCR"
output:
  pdf_document: default
  html_document: default
    df_print: paged
---
```{r setup, include=FALSE}
sapply(c('tidyverse','readxl','broom','rstudioapi','stringr','reshape2'),
require, character.only = TRUE)    # Load packages
sapply(c('tidyverse','readxl','broom','rstudioapi','stringr','reshape2'),  packageVersion)
theme_set(theme_bw() + theme(plot.title = element_text(size = 11, face = 'bold', hjust =
0.5)))
```

## Description
## Droplet digital PCR of DNA samples is a straightforward way to quantify specific target
abundance with high accuracy. The datafile from BioRad Quantisoft analysis output is a .csv.
If all the metadata for sample and assay are correct, proceed.
# Data for this analysis was run on 2019/11/08, as part of characterizing RM 8376. These are
all components of the RM.
# Using Quantisoft 1.7.x.

```{r ddpcr}
getwd()
# point to the .csv directly, or consider file.choose()
  ddPCR.raw <- read_csv("2019-12-13-RM8376-FullPlate.csv") %>%
# calculate the droplet population lambda
  mutate(fracPos = Positives/(Positives + Negatives),
         Lambda = -log(1-fracPos))


# Genome sizes for the known samples. The assembly data for the 19 genomes is accurate ca.
2019/09/17. These genomes take into account plasmids where found.
   Genome.data <-  tibble( Target =c("O157", "O104",  "LT2", "Sea", "SaBAA44", "Sa12600",
"Sepi", "Pa", "Ab", "Nm", "Sp", "Ef", "Ax", "Ah", "Kp", "Ss", "Vf", "Lm",
"Lp","HBB1","POTP","NEIF"),
# Genome size in basepair
   Genome.Size =c(5508281,5378150,5045210,4482050,2996152,2782526,2664302,6263621,4001713,218
0909,1914834,2866936,6812415,4733250,5763364,4917026,4993266,2950924,3409038,3024000000,30240
00000,3024000000))

# Summarize the data by grouping the same samples and assays, calculating averages and SDs.

ddPCR.summary <- filter(ddPCR.raw, Sample!="NTC") %>%
  mutate(Method="PCR") %>%
  mutate(Dilution.Factor = 0.00010, Droplet.Vol = 0.7349) %>%
  left_join(Genome.data,by="Target")

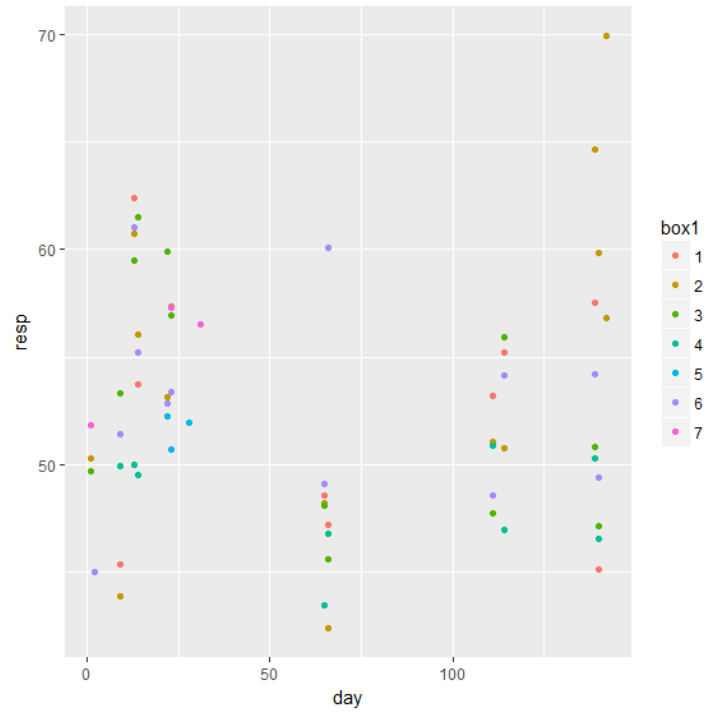

# calculate the genome copy number and mass concentrations
ddPCR.summary %>%
#                         #Genome Size * nL/L * Molecular Weight(DNA)/Avogadros#
  mutate(ng.per.genome = Genome.Size*10^9*659.88/6.022140857e23,
         Undiluted.Targets.per.Droplet = Lambda/Dilution.Factor,
         ng.per.ul = (Lambda/Dilution.Factor)*25/5/(Droplet.Vol/1000)*ng.per.genome,
# mass conc = undiluted targets/droplet * reaction volume / DNA sample volume / droplet
volume (uL) * genome mass
         Genome.Copies.per.ul = ng.per.ul/ng.per.genome) %>%
         select(ID=Target,Sample,Lambda,ng.per.genome,Genome.Copies.per.ul,ng.per.ul,Method)
%>%
         write.csv(file="20191213_ddPCR_listed.csv")
````
````

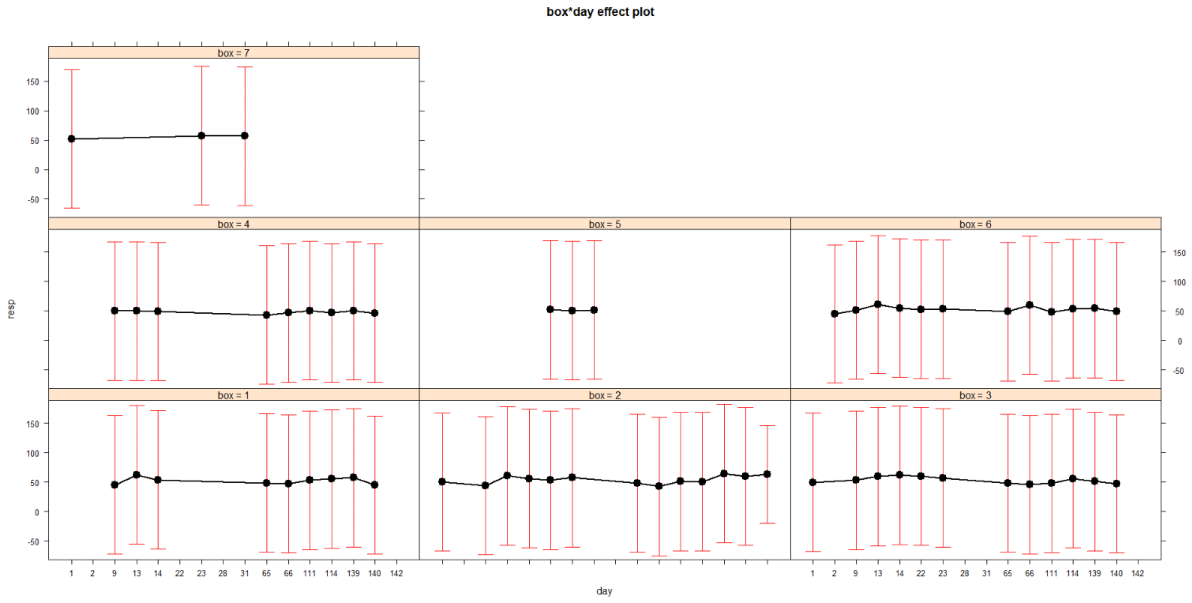
Appendix B: Statistical Analysis of Material Stability

Stability results – March 3, 2020
The first plot in each group is a scatterplot of response as a function of day.
The second plot is an effects plot from a two-way ANOVA where box and day are factors. In all cases there were no significant effects and the plots reflect that.
The dates are transformed into days as:

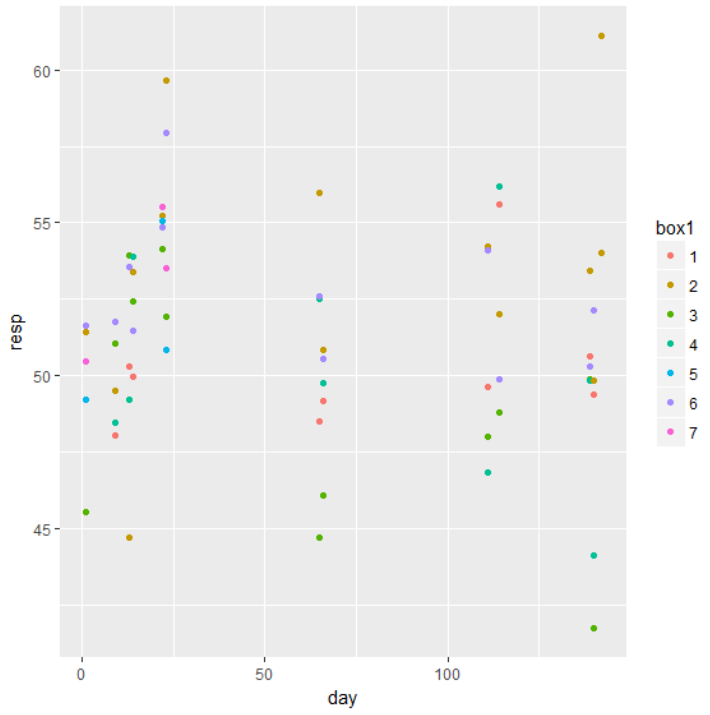
date	day
10/9/2019	1
10/10/2019	2
10/17/2019	9
10/21/2019	13
10/22/2019	14
10/30/2019	22
10/31/2019	23
11/5/2019	28
11/8/2019	31
12/12/2019	65
12/13/2019	66
1/27/2020	111
1/30/2020	114
2/25/2020	139
2/26/2020	140
2/28/2020	143

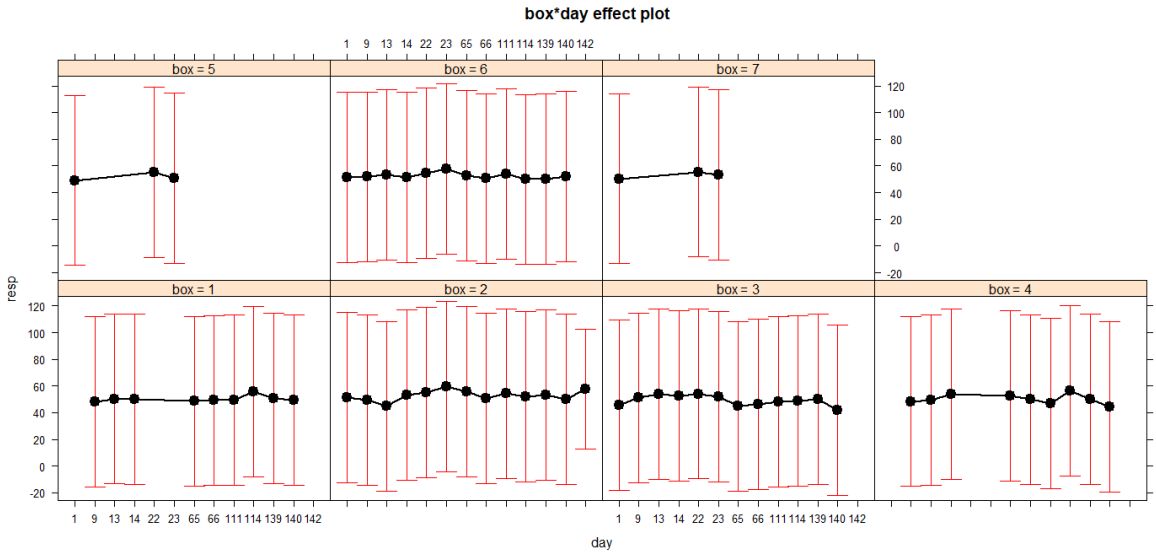
Ab



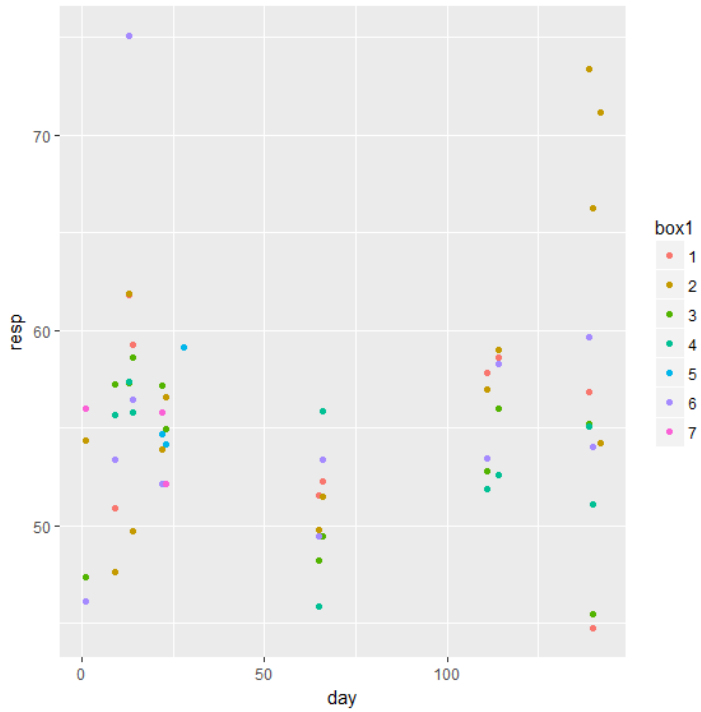


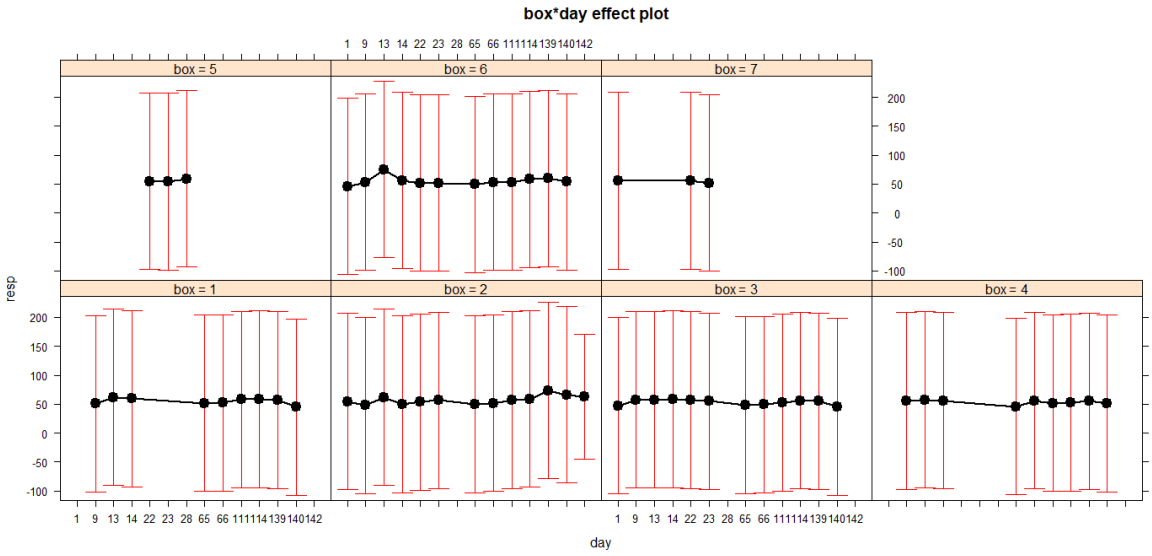
Ah



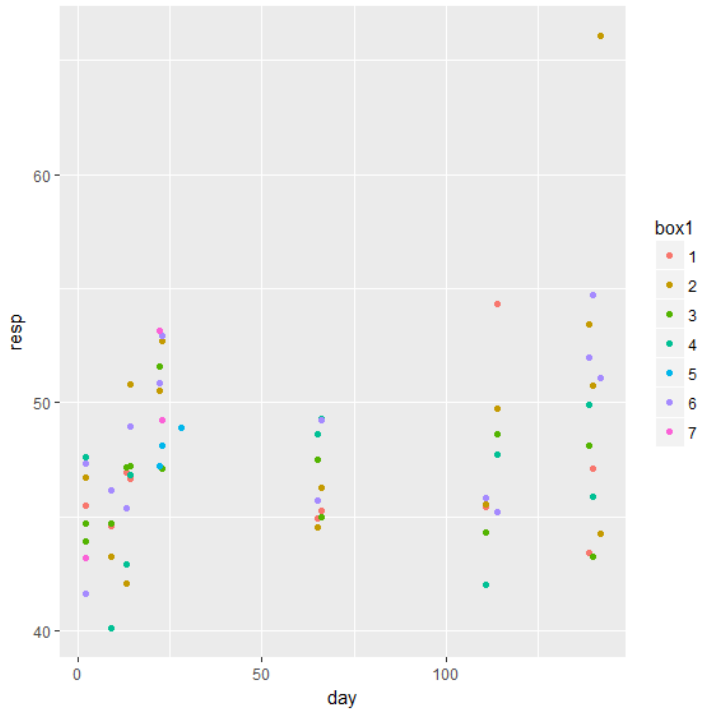


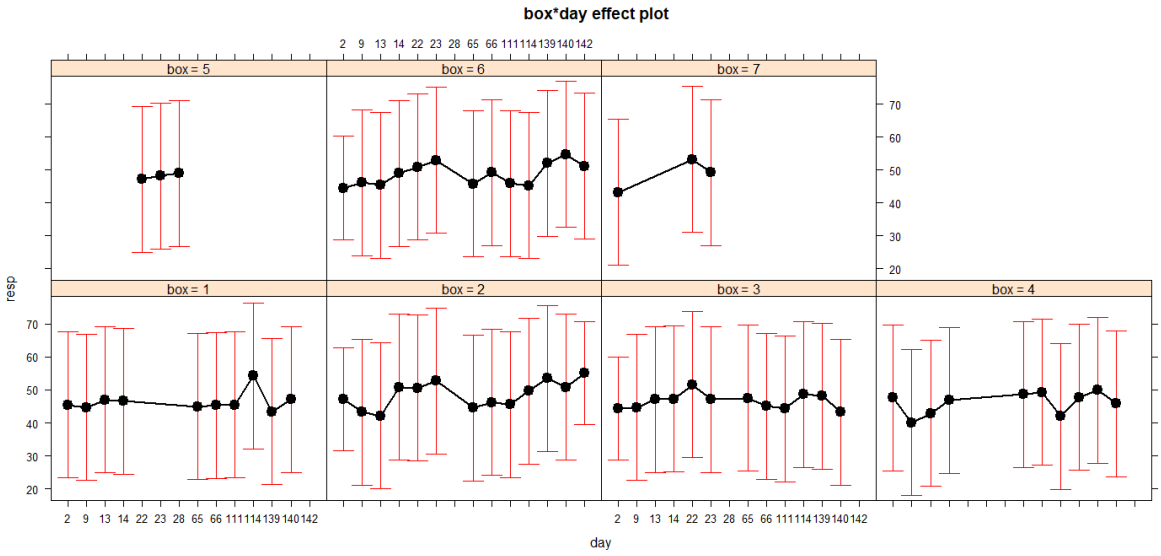
Ax



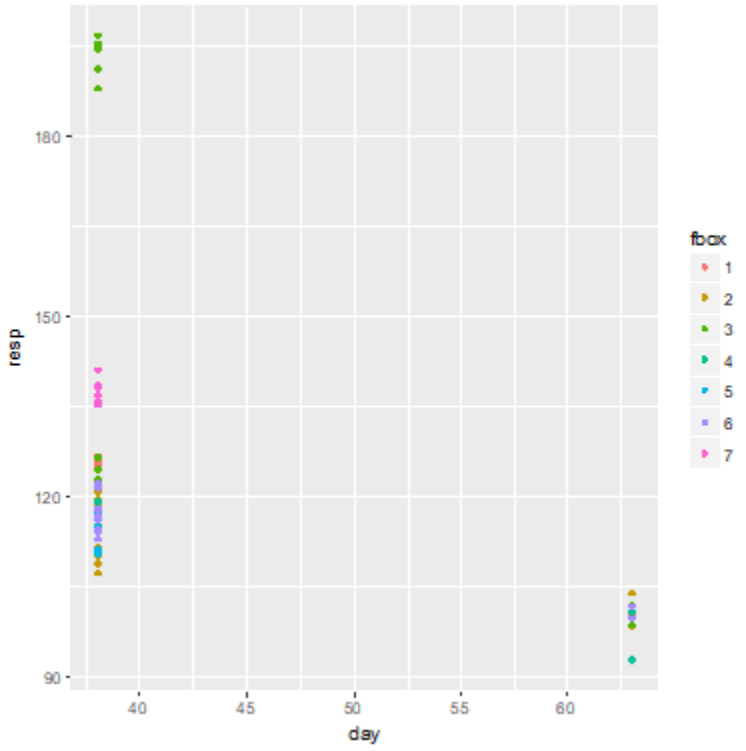


Ef

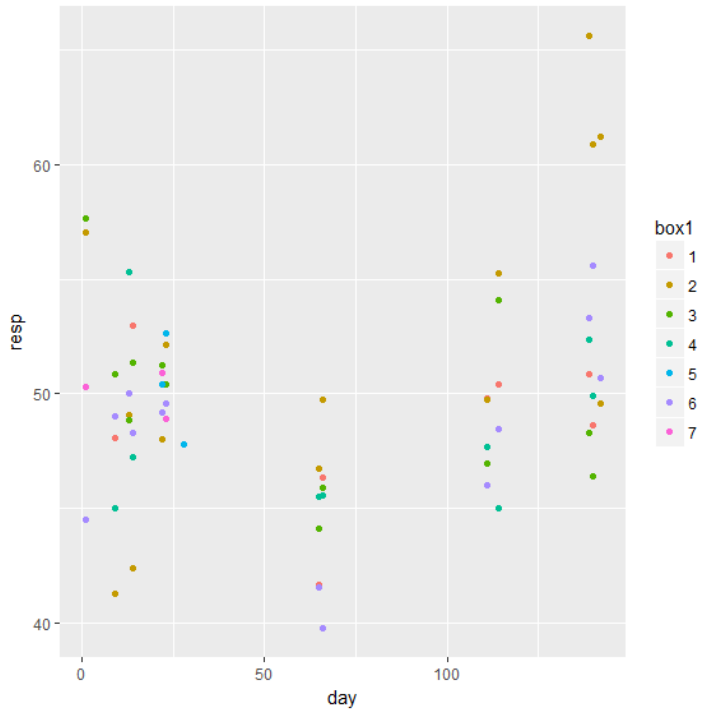


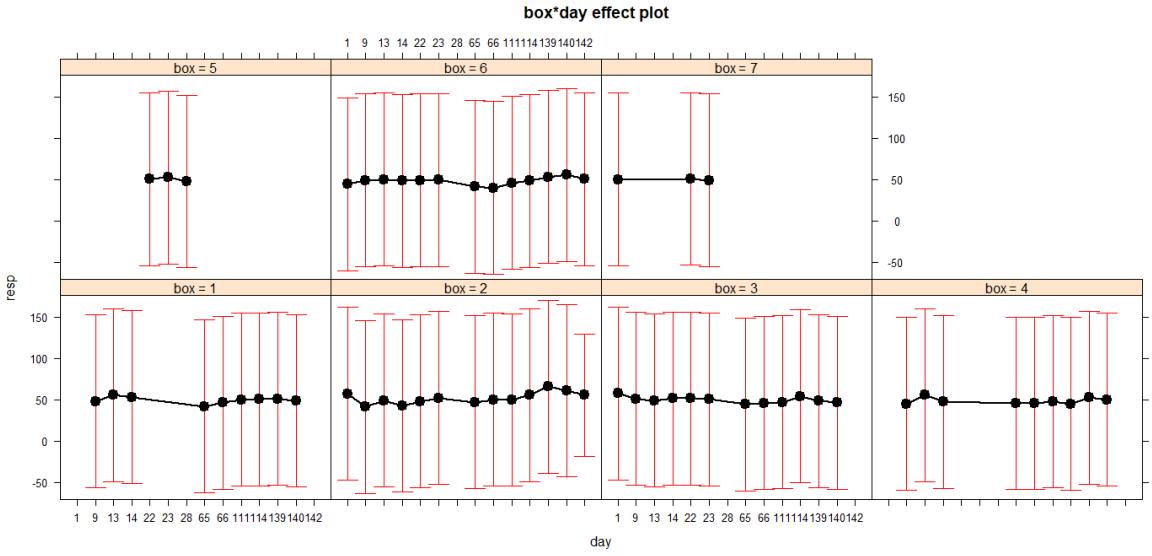


Human-NEIF

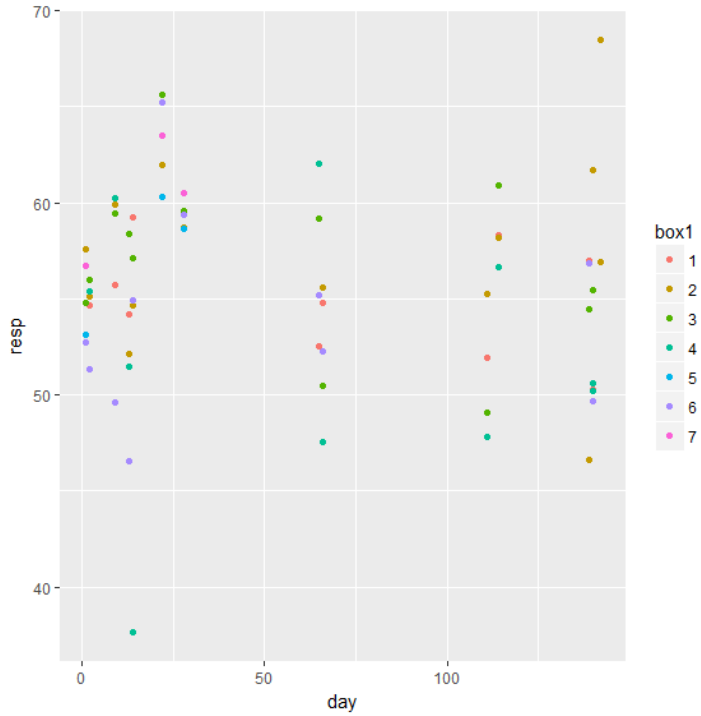


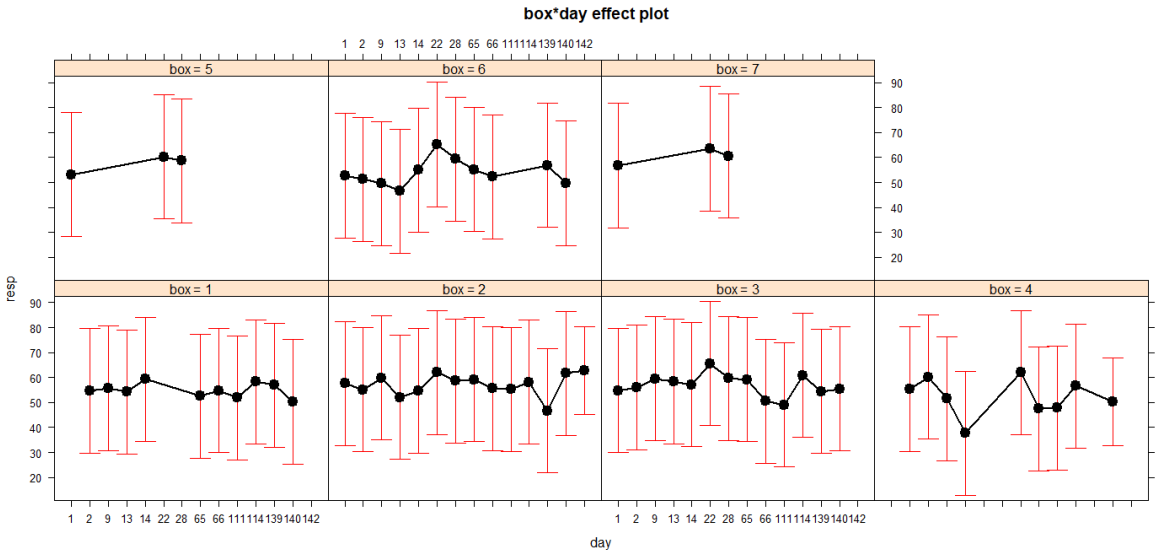
Kp



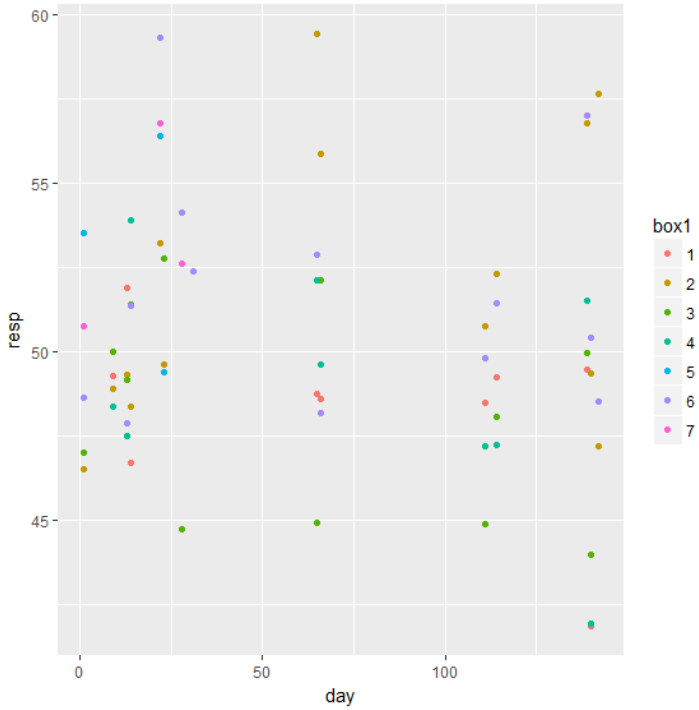


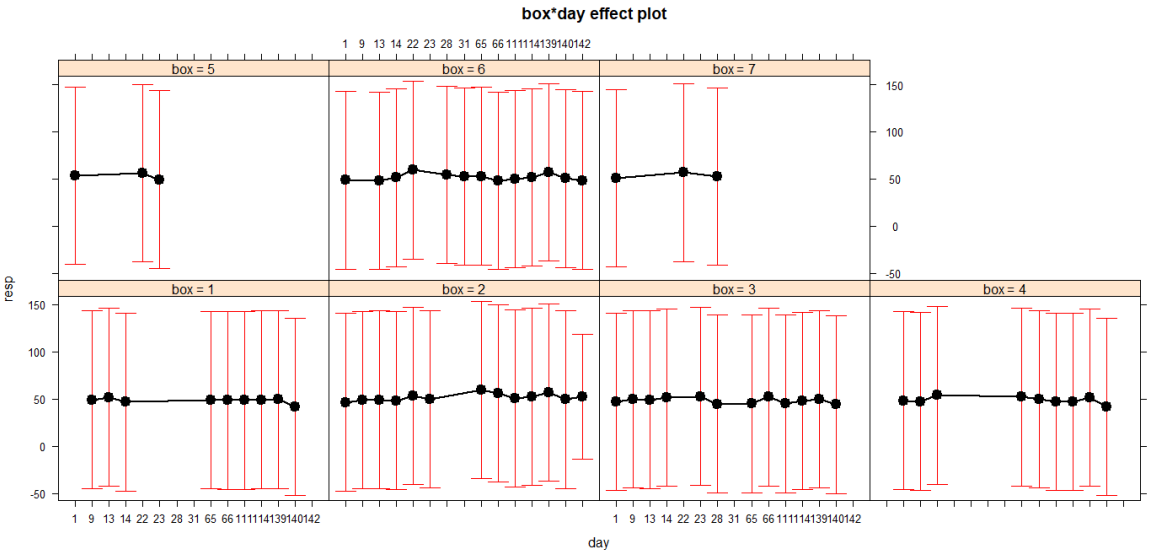
Lm



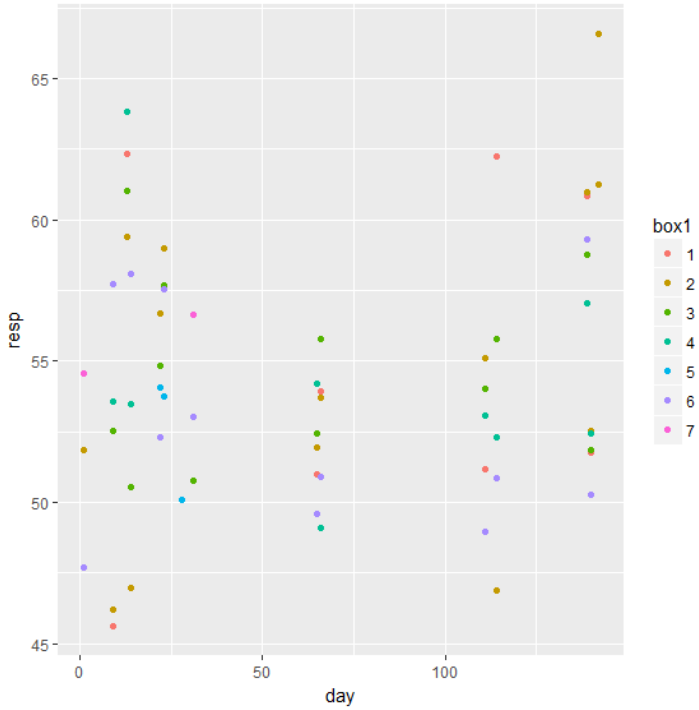


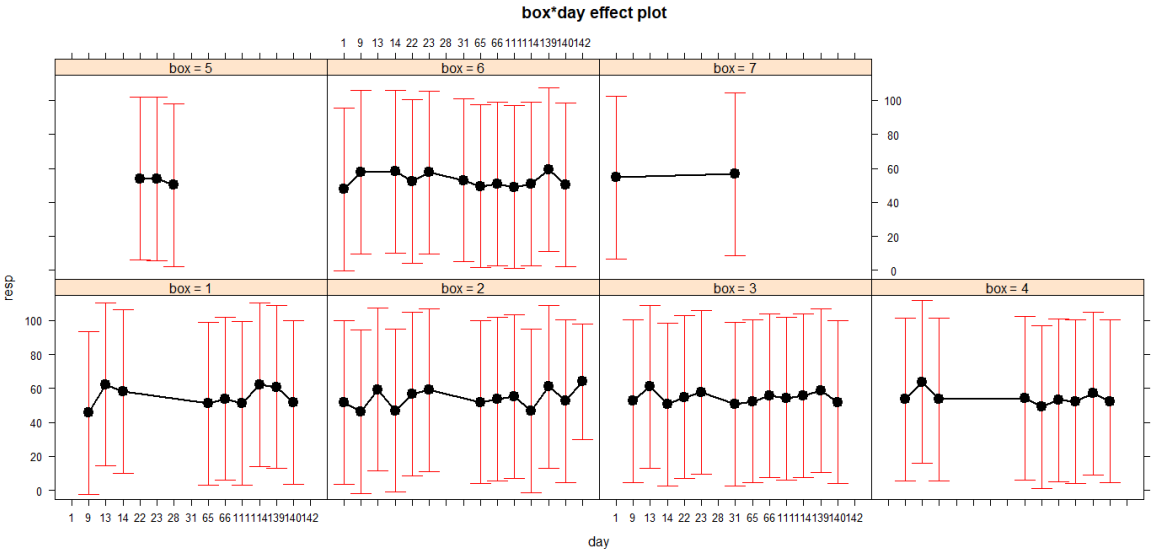
Lp



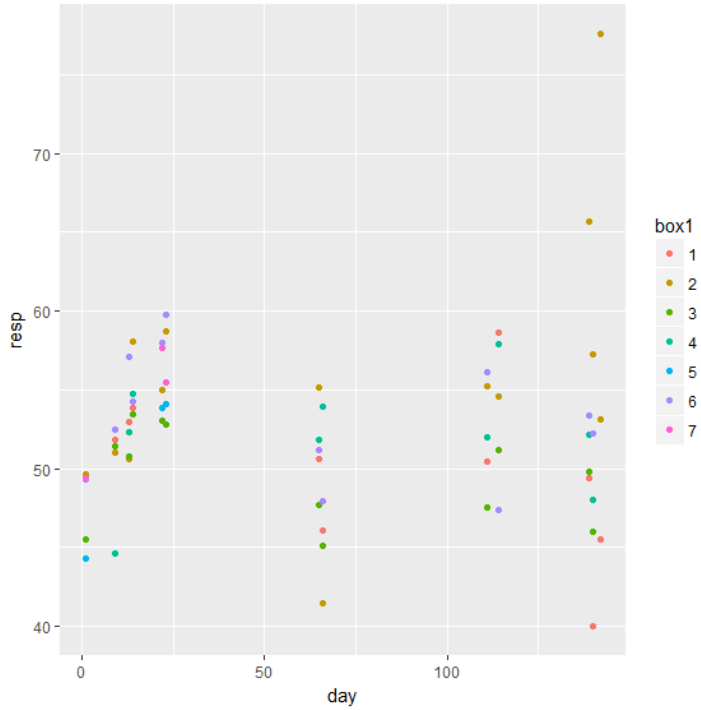


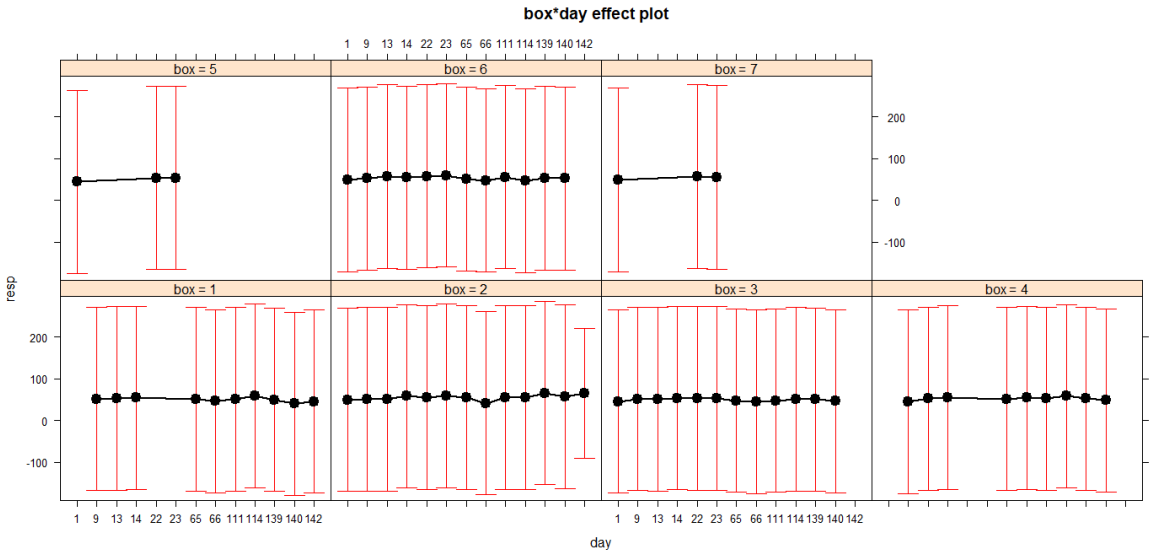
LT2



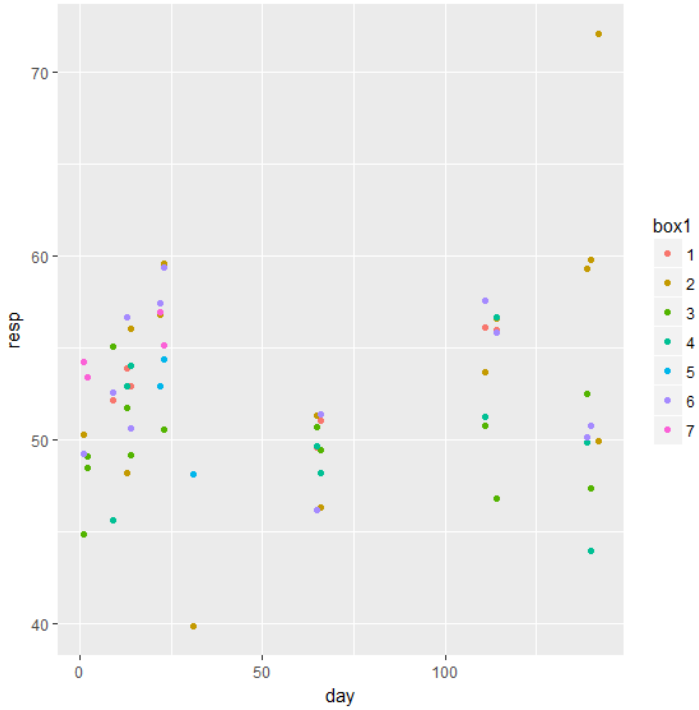


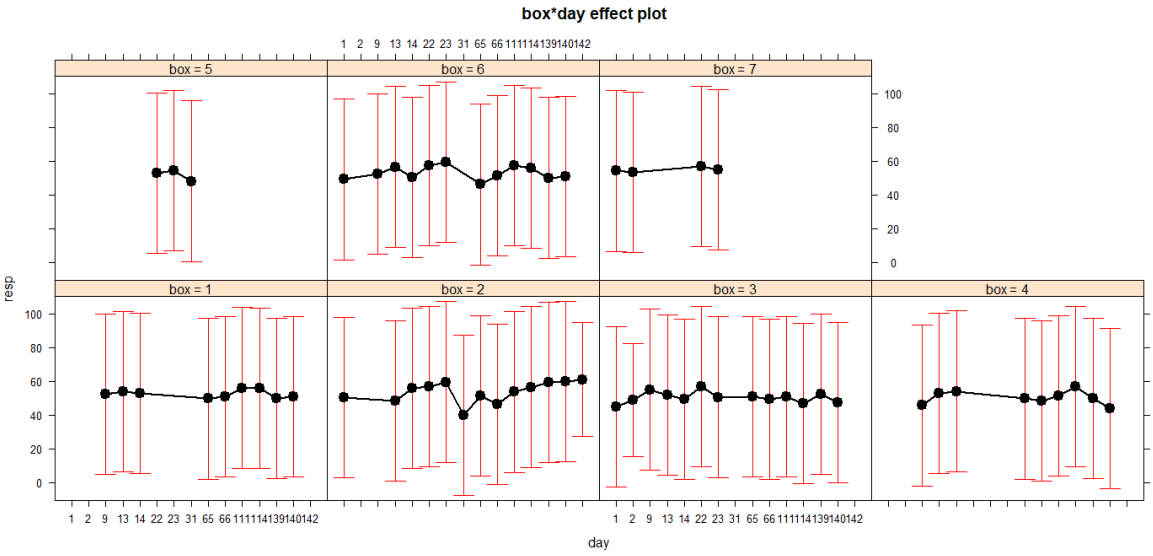
Nm



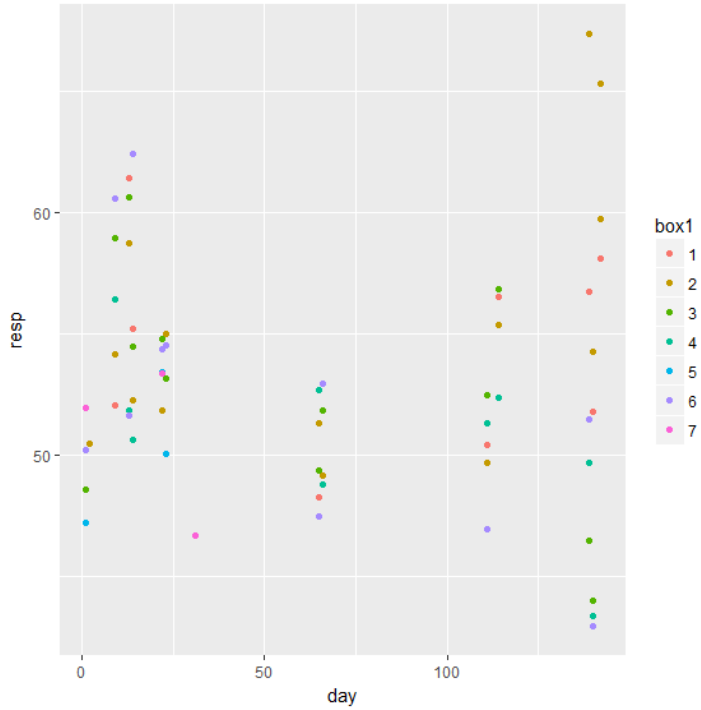


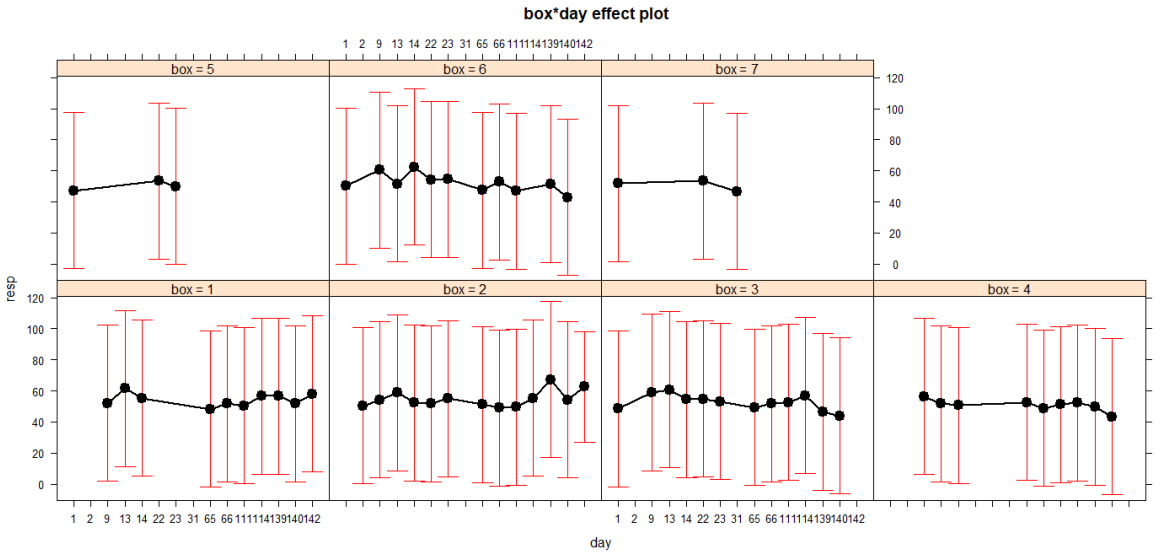
O104



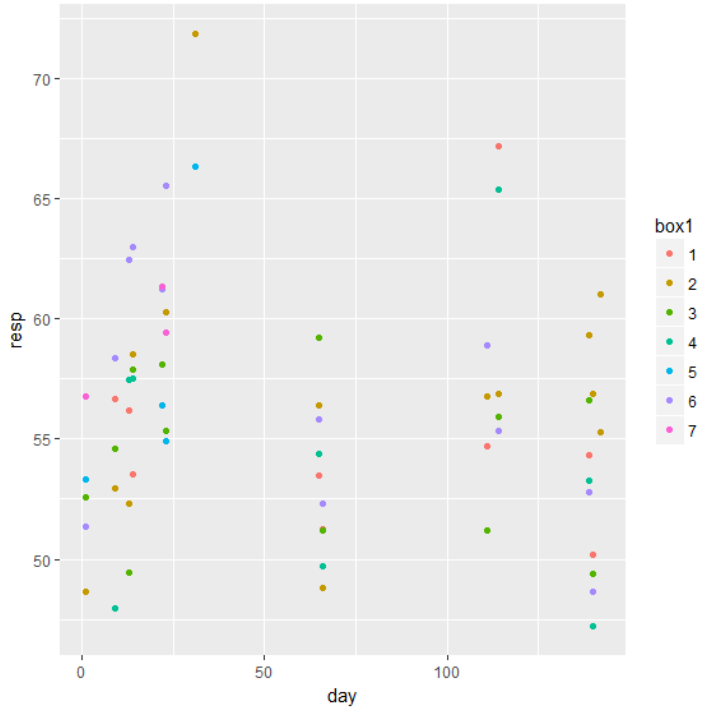


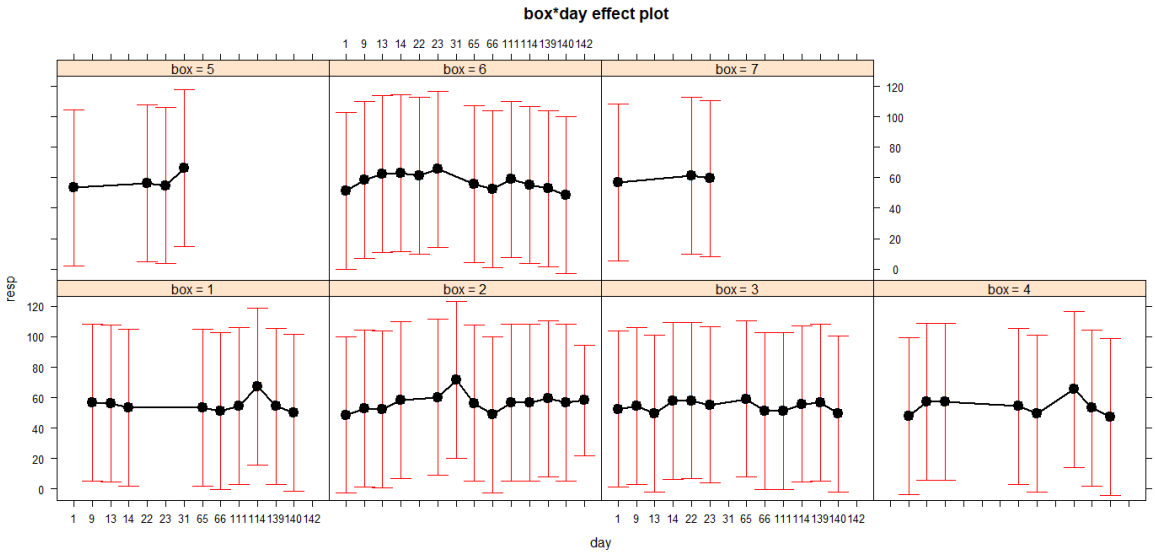
O157



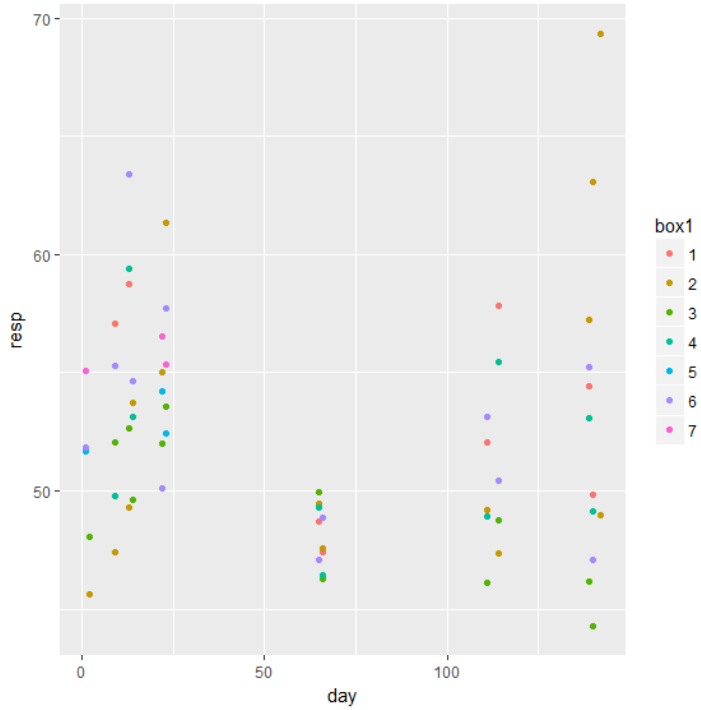


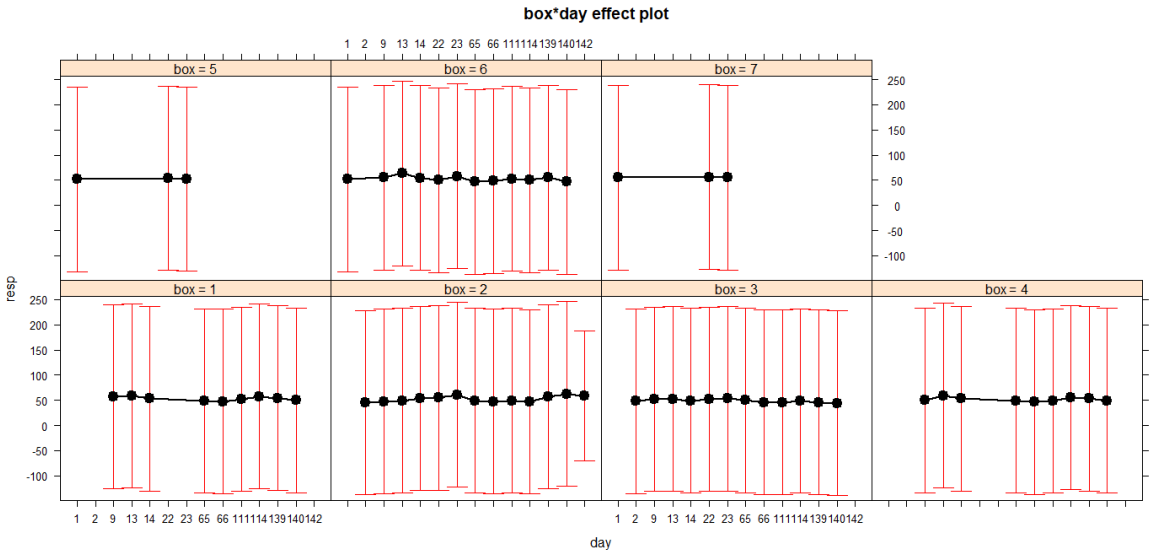
Pa



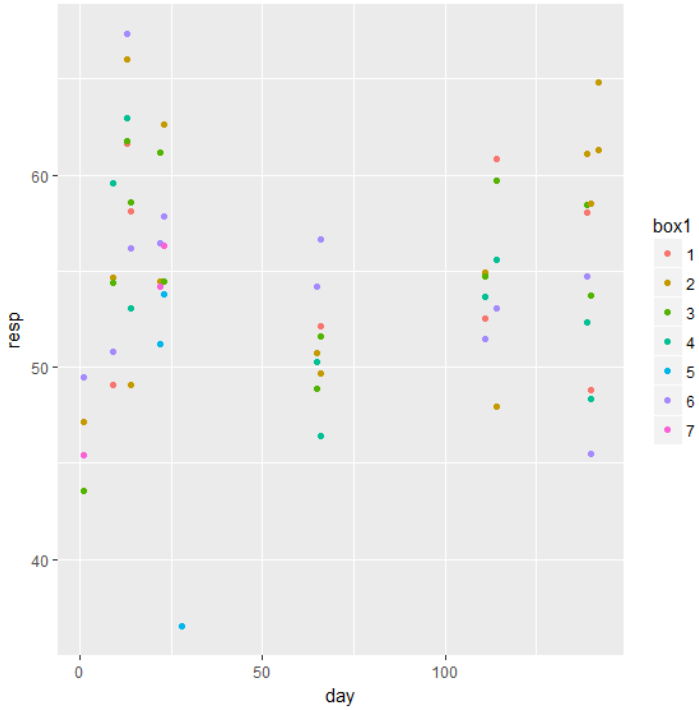


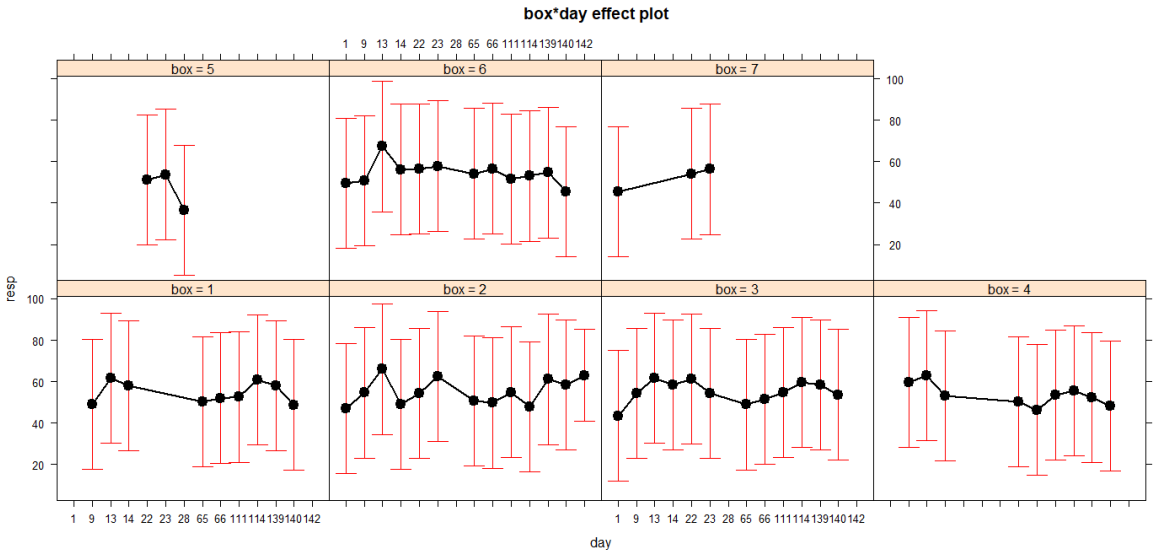
Sa12600



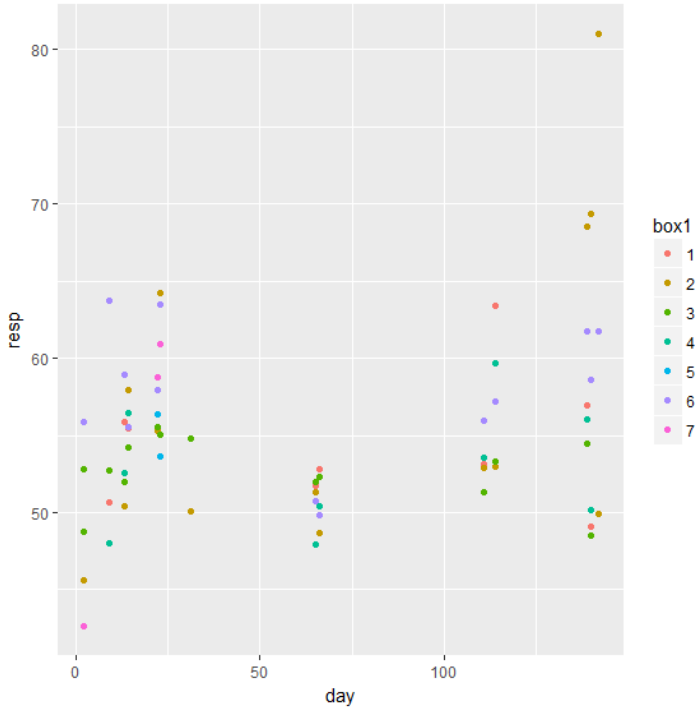


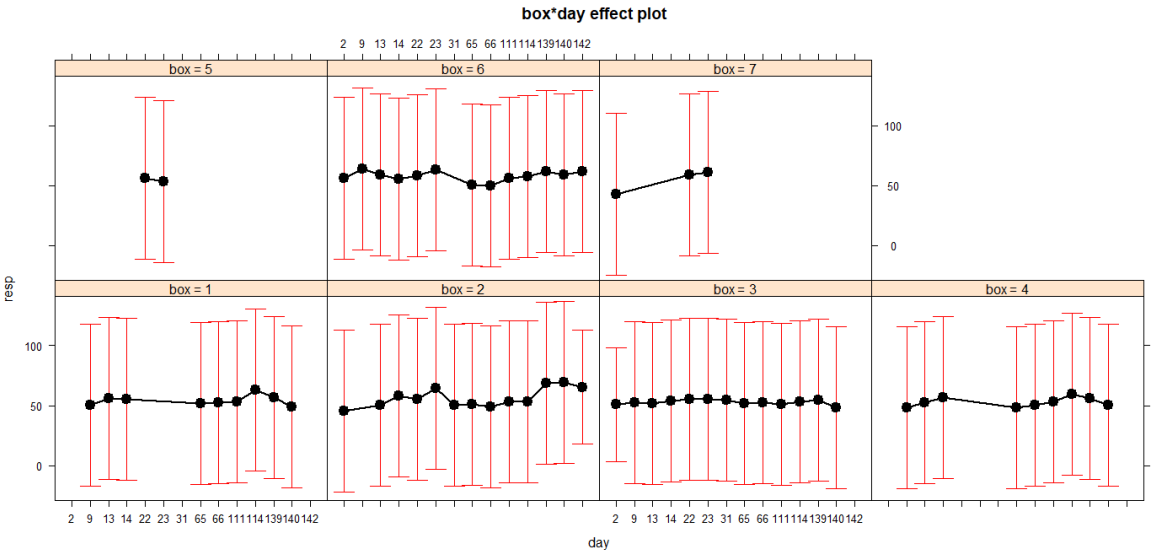
SaBAA44



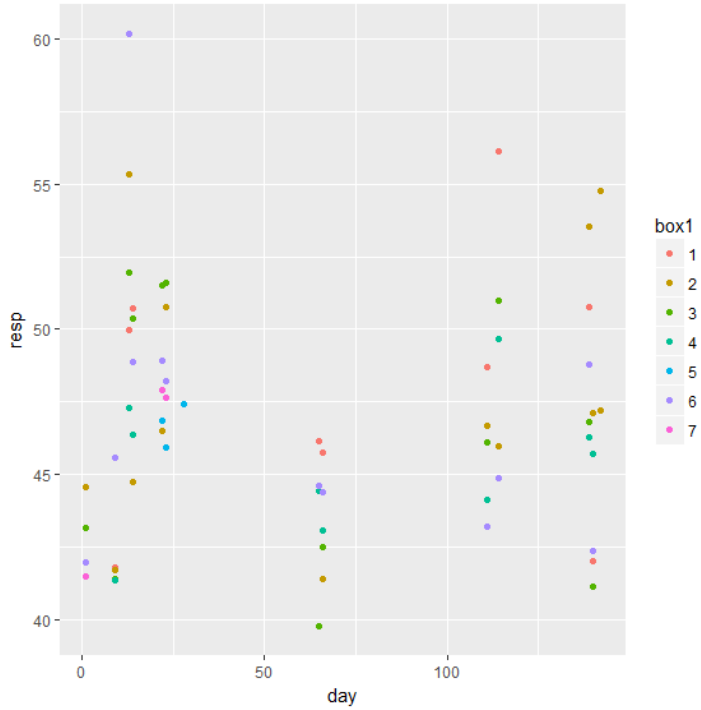


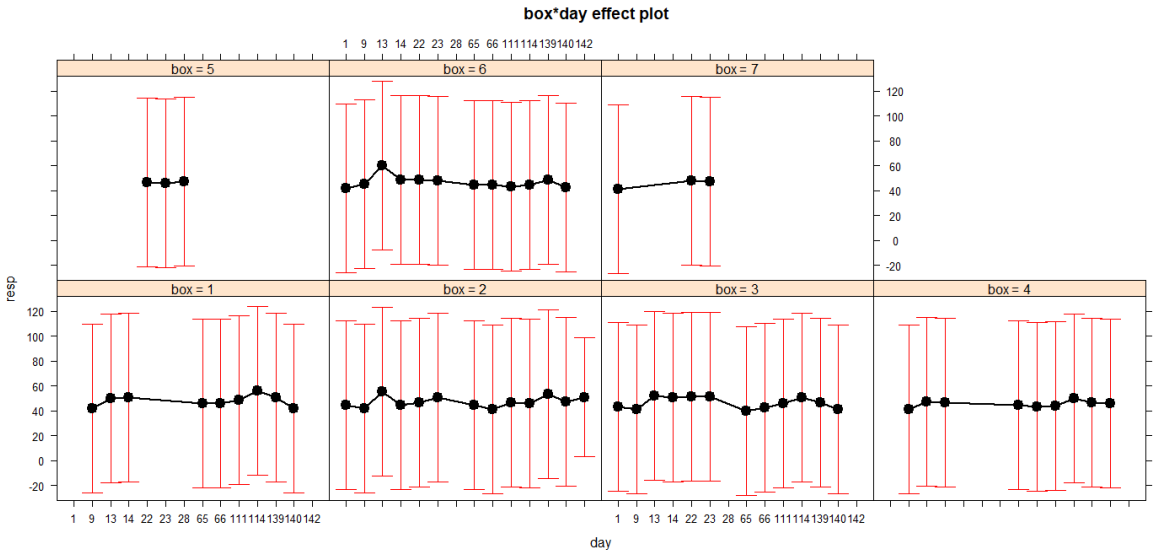
Sea



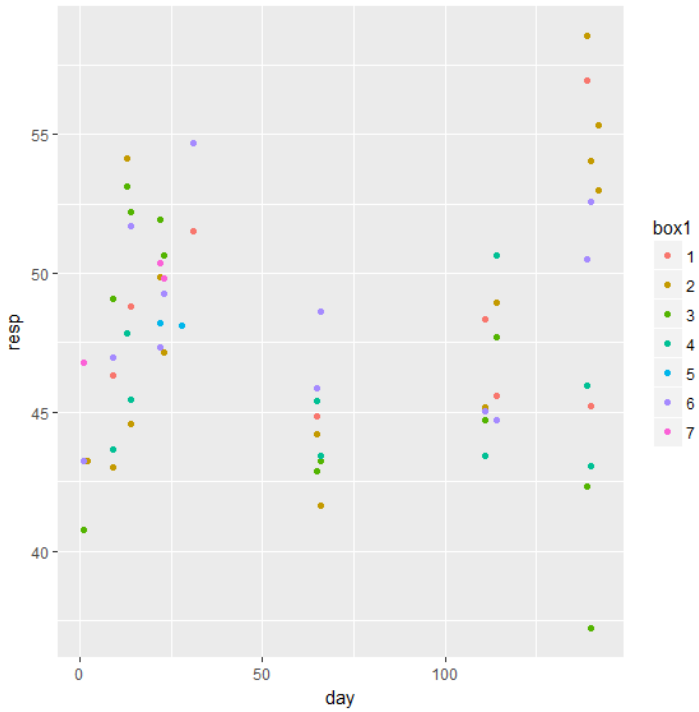


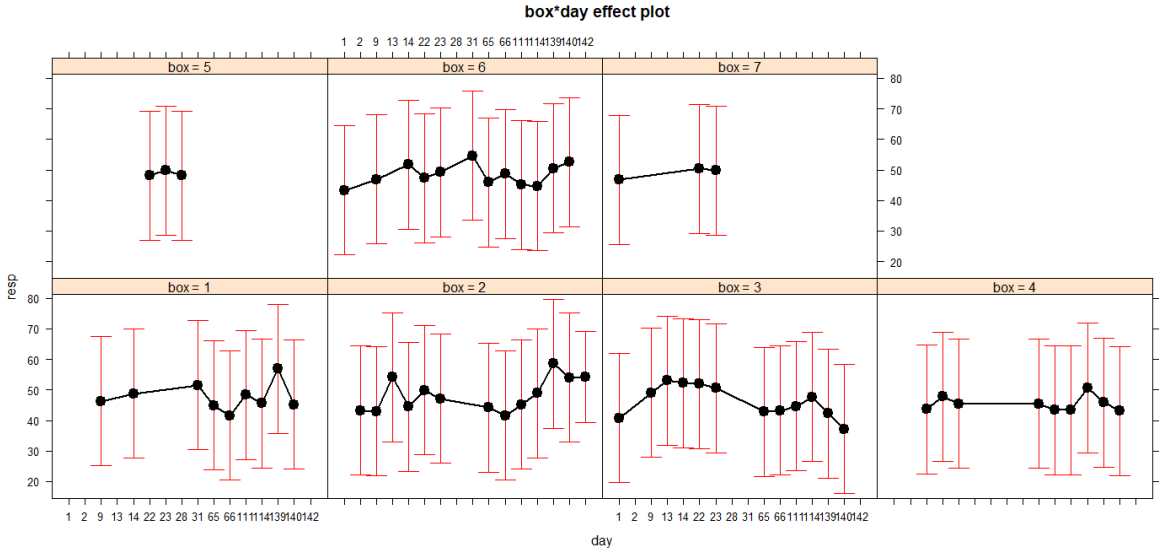
Sepi



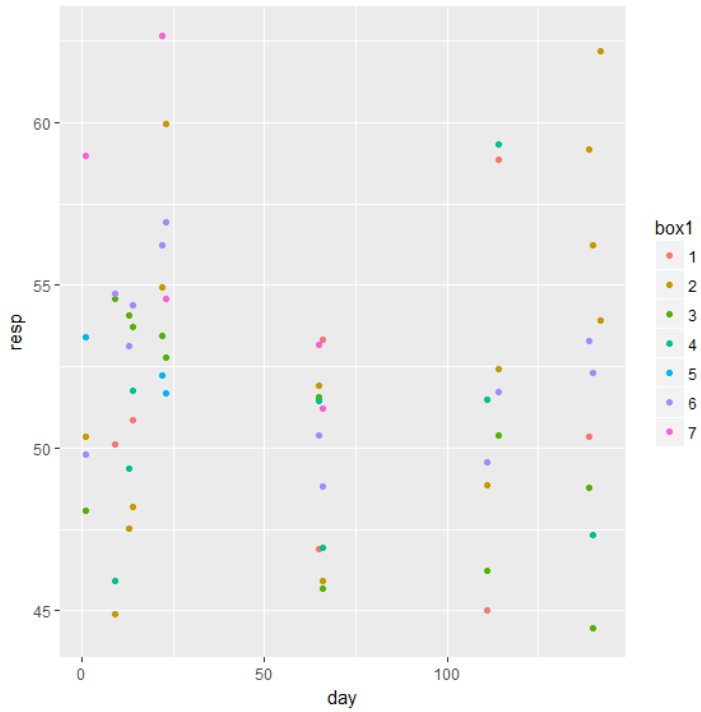


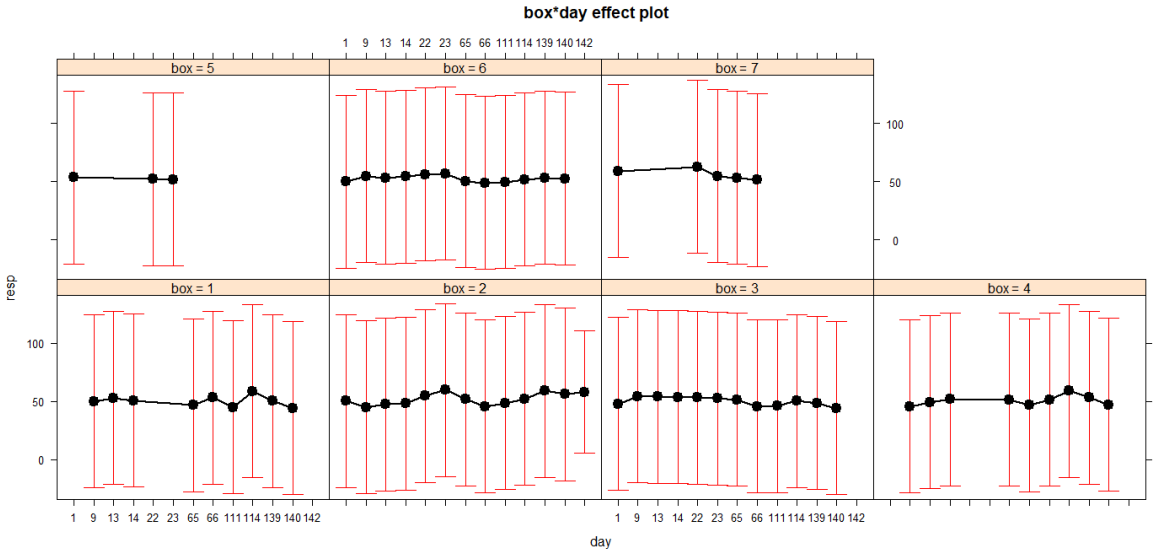
Sp



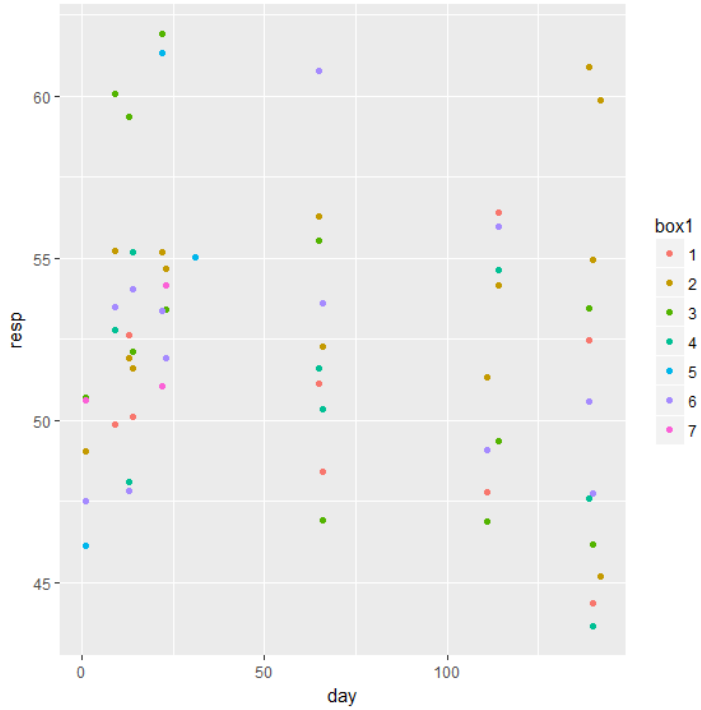


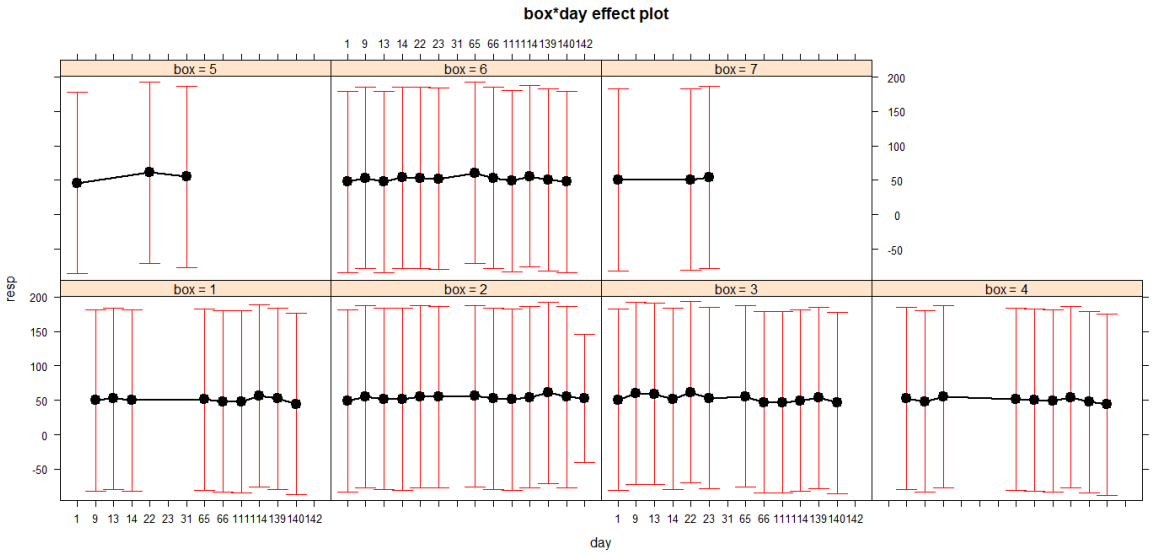
Ss





Vf





Appendix C: Statistician's Report for RM 8376

Microbial Pathogen DNA Standards for Detection and Identification
Blaza Toman
December 29, 2020

This report describes statistical analysis done in support of RM 8376.

1. Introduction.

RM 8376 was designed to enable users to make NGS-based measurements-quantitative to the chromosome. The RM components' chromosomal copy number concentration values are traceable to a ddPCR assay. Each unit of RM 8376 consists of 20 vials (components) of microbial (19) and human (1) DNA in aqueous buffer. RM 8376 is certified for the chromosome copies per μL solution.

To establish homogeneity, the material was measured using ddPCR for concentration using 10 samples taken from the same positions of 7 different boxes. Measurements were taken every month from November 2019 through March 2020 to test stability. The same samples were also used to calculate the chromosome copies per μL solution. Details of the statistical analyses are given in the following sections.

2. Homogeneity

The ddPCR concentration data for each of the twenty organisms was analyzed using a two way mixed effects ANOVA with box being the fixed effect. The following are the resulting ANOVA tables.

	Sum	Df	F	Pr(>F)
box	109.01	6	0.69	0.65
rep	4.17	1	0.16	0.69
box:rep	50.56	2	0.97	0.39
Residuals	520.96	20		

Table 1. ANOVA for *Acinetobacter baumannii*

	Sum	Df	F	Pr(>F)
box	40.24	6	0.71	0.65
rep	1.54	1	0.16	0.69
box:rep	72.06	2	3.81	0.04
Residuals	189.33	20		

Table 2. ANOVA for *Aeromonas hydrophil*

	Sum	Df	F	Pr(>F)
box	52.87	6	0.31	0.92
rep	38.72	1	1.36	0.26
box:rep	197.12	2	3.46	0.05
Residuals	569	20		

Table 3. ANOVA for *Achromobacter xylosoxidans*

	Sum	Df	F	Pr(>F)

box	46.04	6	0.82	0.57
rep	0.87	1	0.09	0.77
box:rep	74.36	2	3.96	0.04
Residuals	234.89	25		

Table 4. ANOVA for *Enterococcus faecalis*

	Sum	Df	F	Pr(>F)
box	102.89	6	1.50	0.23
rep	74.79	1	6.54	0.02
box:rep	38.39	2	1.68	0.21
Residuals	228.79	20		

Table 5. ANOVA for *Klebsiella pneumoniae*

	Sum	Df	F	Pr(>F)
box	232.7	6	1.69	0.16
rep	2.96	1	0.13	0.72
box:rep	157.51	2	3.43	0.05
Residuals	573.41	25		

Table 6. ANOVA for *Listeria monocytogenes*

	Sum	Df	F	Pr(>F)
box	78.13	6	1.24	0.33
rep	10.46	1	0.99	0.33
box:rep	15.27	2	0.73	0.50
Residuals	209.99	20		

Table 7. ANOVA for *Legionella pneumophila*

	Sum	Df	F	Pr(>F)
box	62.96	6	0.38	0.88
rep	36.08	1	1.32	0.27
box:rep	22.32	2	0.41	0.67
Residuals	519.99	19		

Table 8. ANOVA for *Salmonella enterica* enterica

	Sum	Df	F	Pr(>F)
box	88.76	6	0.86	0.54
rep	0.87	1	0.05	0.82
box:rep	6.07	2	0.18	0.84
Residuals	345.72	20		

Table 9. ANOVA for *Neisseria meningitidis*

	Sum	Df	F	Pr(>F)
box	87.32	6	0.7848	0.59
rep	9.08	1	0.49	0.49
box:rep	83.46	2	2.25	0.13
Residuals	426.52	23		

Table 10. ANOVA for *Escherichia coli* O104:H4

	Sum	Df	F	Pr(>F)
box	120.83	6	1.55	0.21
rep	5.364	1	0.41	0.53
box:rep	96.08	2	3.69	0.04
Residuals	260.36	20		

Table 11. ANOVA for *Escherichia coli* O157:H7

	Sum	Df	F	Pr(>F)
box	143.78	6	0.78	0.59
rep	40.19	1	1.31	0.26
box:rep	16.31	2	0.27	0.77
Residuals	642.94	21		

Table 12. ANOVA for *Pseudomonas aeruginosa*

	Sum	Df	F	Pr(>F)
box	131.61	6	1.38	0.27
rep	33	1	2.08	0.16
box:rep	19.99	2	0.63	0.54
Residuals	316.82	20		

Table 13. ANOVA for *Staphylococcus aureus* 12600

	Sum	Df	F	Pr(>F)
box	274.48	6	0.90	0.51
rep	6.02	1	0.12	0.73
box:rep	57.49	2	0.57	0.57
Residuals	1012.54	20		

Table 14. ANOVA for *Staphylococcus aureus* BAA44

	Sum	Df	F	Pr(>F)
box	149.47	6	0.92	0.5
rep	2.27	1	0.08	0.78
box:rep	5.23	2	0.10	0.91
Residuals	568.29	21		

Table 15. ANOVA for *Salmonella enterica* arizonae

	Sum	Df	F	Pr(>F)
box	24.36	6	0.17	0.98
rep	18.3	1	0.77	0.39
box:rep	23.26	2	0.49	0.62
Residuals	478.3	20		

Table 16. ANOVA for *Staphylococcus epidermidis*

	Sum	Df	F	Pr(>F)
box	46.21	6	0.59	0.73
rep	0.05	1	0.004	0.95
box:rep	51.23	2	1.98	0.16
Residuals	258.73	20		

Table 17. ANOVA for *Streptococcus pyogenes*

	Sum	Df	F	Pr(>F)
box	192.76	6	4.05	0.008
rep	13.91	1	1.75	0.20
box:rep	98.05	2	6.18	0.008
Residuals	158.67	20		

Table 18. ANOVA for *Shigella sonnei*

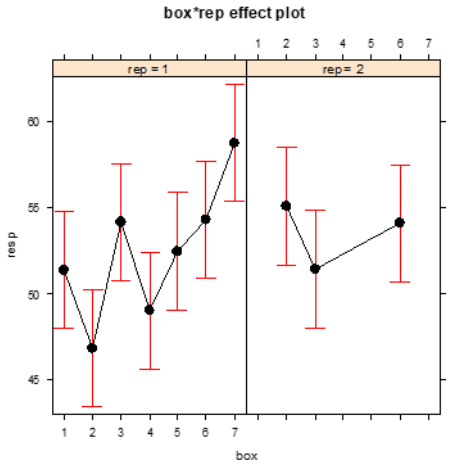


Figure 1. Effect plot for *Shigella sonnei*
7th box is too high compared to 2nd box, this is why there is a significant box effect.

	Sum	Df	F	Pr(>F)
box	102.4	6	1.02	0.44
rep	0.44	1	0.03	0.87
box:rep	5.66	2	0.17	0.85
Residuals	335.75	20		

Table 19. ANOVA for *Vibrio furnissii*

	Sum Sq	Df	F	Pr(>F)
assay	4931	2	8.42	0.0003
box	43713	6	24.89	2.20E-16

assay:box	160	12	0.05	1.0
Residuals	44786	153		

Table 20. ANOVA for Human

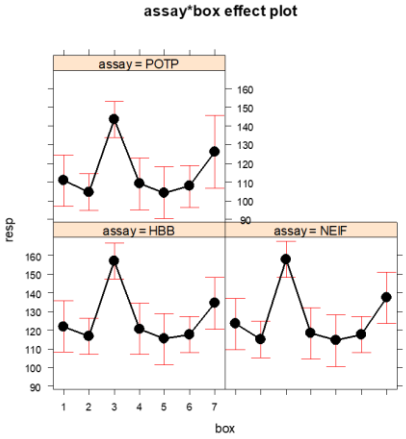


Figure 2. Effect plot for Human
3rd box is too high in all assays.

3. Stability

A linear hierarchical model with a main effect, a term for box effect, and a time effect was fitted (using OpenBUGS) to all 20 data sets. Namely, the observed measurement of concentration for the ith box at the jth time is

$y_{ij} \sim N(\mu_i + \alpha\ time_j, \sigma^2)$, where $\mu_i \sim N(\theta, \tau_i^2)$, for $i = 1, \ldots, 7$.

.

In all cases the coefficient α was essentially equal to 0 thus establishing stability.

4. Certification

The statistical model given in section 3 was used for the copy number concentration data of all 20 organisms to estimate the value θ, and the corresponding standard uncertainty. To show the variability that a customer may observe, the predicted copy number concentration $y_p \sim N(\theta, \sigma^2)$ was also estimated and the resulting predictive uncertainty is given in the last column of Table 21.

ATCC ID	Part	Organism	Copy # Concentration $\times 10^6$(copy/µL)	Standard uncertainty $\times 10^6$ (copy/µL)	Predictive uncertainty $\times 10^6$ (copy/µL)
43895	A	*Escherichia coli* O157:H7	8.84	0.19	0.82
BAA 2309	B	*Escherichia coli* O104:H4	8.89	0.14	0.80
700720	C	*Salmonella enterica* enterica	9.72	0.19	0.85
12324	D	*Salmonella enterica* arizonae	10.84	0.26	1.24
BAA 44	E	*Staphylococcus aureus*	16.49	0.38	1.82
12600	F	*Staphylococcus aureus*	17.38	0.34	1.62
12228	G	*Staphylococcus epidermidis*	15.99	0.30	1.48
BAA 47	H	*Pseudomonas aeruginosa*	8.27	0.17	0.76
19606	I	*Acinetobacter baumannii*	12.01	0.28	1.29

13077	J	*Neisseria meningitidis*	21.67	0.47	2.28
12344	K	*Streptococcus pyogenes*	22.55	0.43	2.10
19433	L	*Enterococcus faecalis*	14.75	0.25	1.23
27061	M	*Achromobacter xylosoxidans*	7.28	0.18	0.81
35654	N	*Aeromonas hydrophila*	9.97	0.17	0.68
13883	O	*Klebsiella pneumoniae*	7.68	0.18	0.76
25931	P	*Shigella sonnei*	9.67	0.18	0.80
35016	Q	*Vibrio furnissii*	9.70	0.18	0.80
19115	R	*Listeria monocytogenes*	17.39	0.32	1.61
33152	S	*Legionella pneumophila*	13.63	0.23	0.99
--	T	Human*	0.0323	0.00075	0.0023

Table 21. Chromosome copy number concentrations with uncertainty, for the 20 components.