

Titration Validation

Nate Olson

2017-04-25

Contents

1 ERCC Spike-in qPCR	1
2 Bacterial Concentration qPCR	2
3 Appendix	4
3.1 Mixing Titrations	4
3.2 Bacterial Quant	11
4 Session information	13
4.1 Git repo commit information	13
4.2 Platform Information	13
4.3 Package Versions	13

Two sets of quantitative PCR experiments were used to validate the titrations. ERCC exogenous spike-ins were used to validate the proportion of DNA from the unmixed pre- and post-treatment samples in the titrations. A 16S rRNA bacterial qPCR assay was used to evaluate the proportion of bacterial DNA in the titrations from the unmixed pre- and post-titration samples.

1 ERCC Spike-in qPCR

Methods

- Fit standard curve data to a linear model to assess assay precision (R^2) and efficiency.
- Efficiency, is a measure of the assay amplification efficiency, whether the amount of template DNA doubles every PCR cycle.

Results

- The qPCR assay standard curves had a high level of precision with R^2 values close to 1 for all standard curves (Table 1).
- The amplification efficiency was outside of the ideal range (0.9 - 1.1), but within the acceptable range (0.8-1.2).
 - Ideal and acceptable ranges based on rule of thumb community accepted guidelines.
- The post treatment qPCR assays had good R^2 and slope values (Table 2). The expected slope is 1, for a doubling every cycle.
 - The 1-4 titration factor samples had Ct values consistently above the regression line (Figure 1).
 - Different slopes for titrations 1-4 and titrations 0, 5, 10, and 15 (Table 3).
 - For the pre-treatment samples, should be 1 Ct difference between the unmixed post and titration factor 1 and 0.5 Ct between titration factor 1 and 2.
 - For the other titration factors the expected difference is to small to detect using qPCR (< 0.5 Ct).

Discussion

Table 1: ERCC Spike-in qPCR standard curve summary metrics.

biosample_id	treatment	adj.r.squared	efficiency
E01JH0004	Pre	0.9972	84.36
E01JH0004	Post	0.9996	86.19
E01JH0011	Pre	0.9999	87.93
E01JH0011	Post	0.9995	87.46
E01JH0016	Pre	0.9990	84.22
E01JH0016	Post	0.9991	87.33
E01JH0017	Pre	0.9979	89.78
E01JH0017	Post	0.9968	85.80
E01JH0038	Post	0.9984	86.69
E01JH0038	Pre	0.9994	84.30

Table 2: ERCC qPCR titration validation assays

biosample_id	treatment	adj.r.squared	slope
E01JH0004	Post	0.98	0.92
E01JH0011	Post	0.95	0.90
E01JH0016	Post	0.95	0.84
E01JH0017	Post	0.89	0.93
E01JH0038	Post	0.95	0.94
E01JH0004	Pre	0.63	-0.13
E01JH0011	Pre	0.48	-0.11
E01JH0016	Pre	0.34	-0.10
E01JH0017	Pre	0.57	-0.12
E01JH0038	Pre	0.33	-0.14

- Limitation of efficiency assessment is that the standard curve is only plasmid DNA, no stool DNA as background. Stool DNA may contain PCR inhibitors or DNA that may interfere with the qPCR assay.

2 Bacterial Concentration qPCR

Methods

- Fitting the standard curve to a linear model, $Ct \sim \log_{10}(\text{concentration})$.
- The expected slope for the standard curve is -3.33 indicating a perfect doubling every PCR cycle, for a amplification factor ($AF = 10^{-1/\text{slope}}$) of 2 and efficiency ($E = 10^{-1/\text{slope}} - 1$) of 1.
- Used an in-house standard curve, due to issues related to the stability of the manufacturer provided standard. (Supplmental material??)
- The model was fit using the full standard curve and only points in the standard curve with concentrations greater than 0.02 ng/ul.
 - Fitting the regression to all concentrations in the standard curve resulted in a lower amplification efficiency and R^2 .

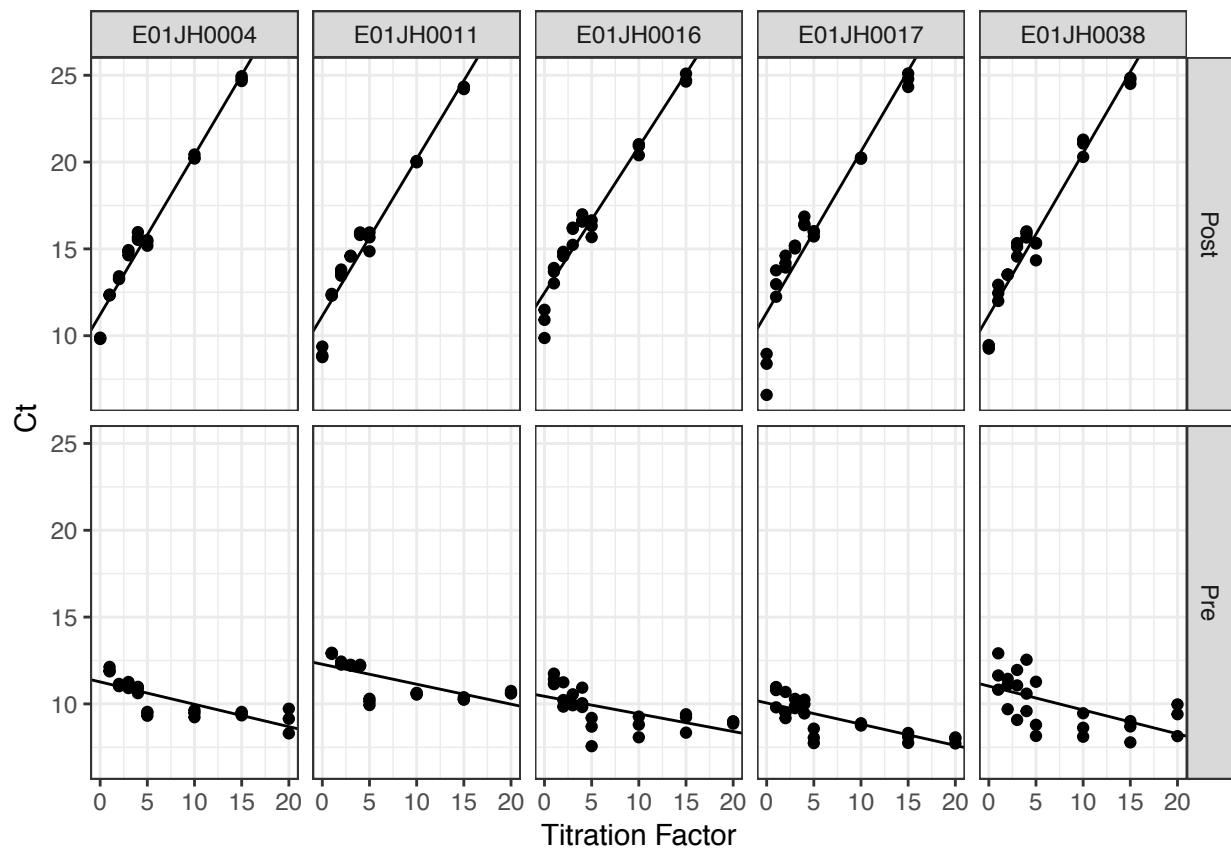


Figure 1: qPCR ERCC spike-in titration validation results.

Table 3: Separate linear model fit estimates for titrations 1-4, and 0, 5, 10, 15, and 20.

biosample_id	adj.r.squared	slope	fit
E01JH0004	0.98	1.16	1:4
E01JH0004	1.00	1.00	0.5,10,15
E01JH0011	0.99	1.16	1:4
E01JH0011	0.99	1.01	0.5,10,15
E01JH0016	0.92	1.07	1:4
E01JH0016	0.99	0.93	0.5,10,15
E01JH0017	0.91	1.16	1:4
E01JH0017	0.97	1.09	0.5,10,15
E01JH0038	0.95	1.16	1:4
E01JH0038	0.99	1.04	0.5,10,15

- Used a linear model to test for differences in the bacterial DNA concentration between titrations. No difference between titrations indicates that the proportion of bacterial DNA in the pre- and post-treatment unmixed samples is the same.
- As the post-treatment bacterial DNA concentration estimates are outside of the standard curve. The post-treatment sample bacterial DNA concentration was also estimated using concentration estimates for unmixed pre-treatment samples. $Post = \frac{Con - (1-\theta)Pre}{\theta}$

Results

- The concentration of bacterial DNA is greater in the post-treatment samples than the pre-treatment samples for all biological replicates excluding 4 **TODO-Use JH- ID system** (Figure 2).
- The post-treatment DNA concentration estimates are greater than the expected value of 12.5 ng/ul and outside of the standard curve for all but sample 4 **TODO-Use JH- ID system** (Figure 2).
 - The unmixed sample concentrations were diluted to 12.5 ng/ul prior to making the titrations therefore all samples are expected to have concentrations less than 12.5ng/ul.
- Figure 3
 - For biological replicates 1 and 4 higher bacterial DNA concentration for titrations 5, 10, and 15 compared to titrations 1-4.
 - For biological replicate 3 and 5 the concentration of bacterial DNA decreases with titration.
 - Biological replicate 2 **NOT SURE HOW TO INTERPERET**
- Figure 5
 - Post-treatment sample concentration estimates are less than 0 for most biological replicates and titrations.
 - Either the math is wrong, the qPCR measurements are not accurate enough for these calculations, or the samples are not mixed according to expectation.

Discussion

biosample	r.squared	adj.r.squared	sigma	statistic	p.value	df	logLik	AIC	BIC	deviance	df.residual
1	0.93	0.89	0.20	26.81	0	7	7.66	0.67	8.64	0.54	13
2	0.86	0.81	0.73	14.90	0	7	-19.03	54.07	62.42	7.53	14
3	0.95	0.93	0.16	47.23	0	7	13.46	-10.92	-2.57	0.34	14
4	0.99	0.98	0.21	194.36	0	7	7.06	1.87	10.23	0.63	14
5	0.94	0.91	0.17	36.42	0	7	11.95	-7.90	0.45	0.39	14

3 Appendix

3.1 Mixing Titrations

3.1.1 Table with volumes used to dilute samples

```
biosampleInfo %>% kable()
```

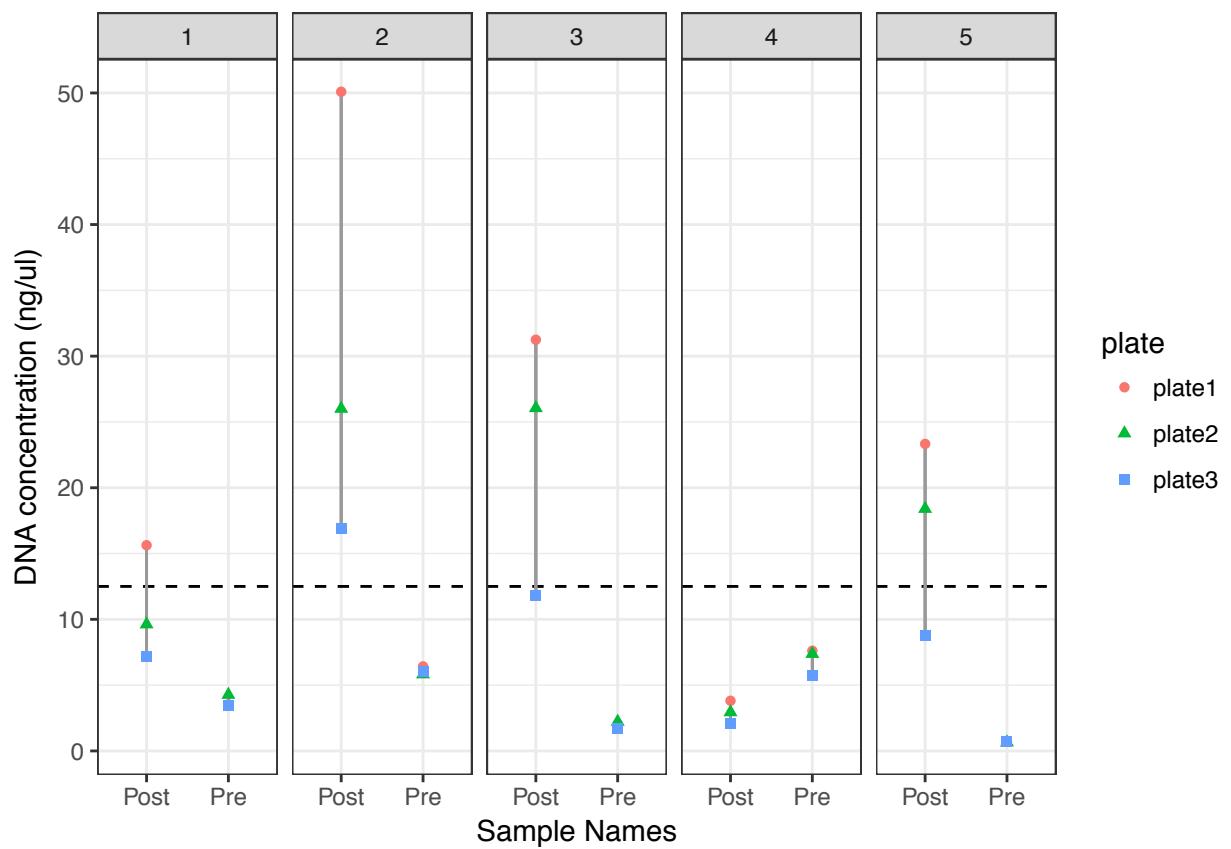


Figure 2: Predicted mixture study sample concentrations. Dashed line indicates the expected max concentration of 12.5 ng/uL.

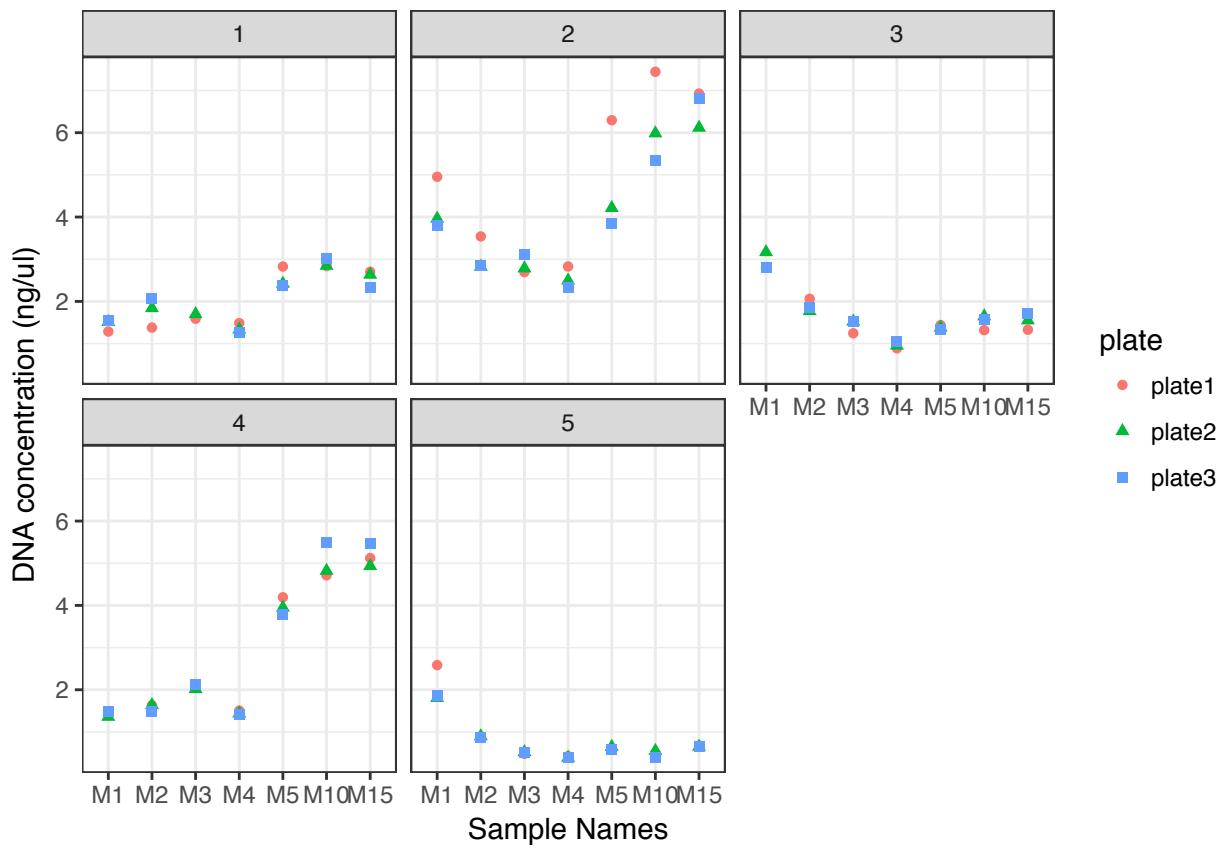


Figure 3: Predicted mixture study sample bacterial DNA concentrations for titrations.

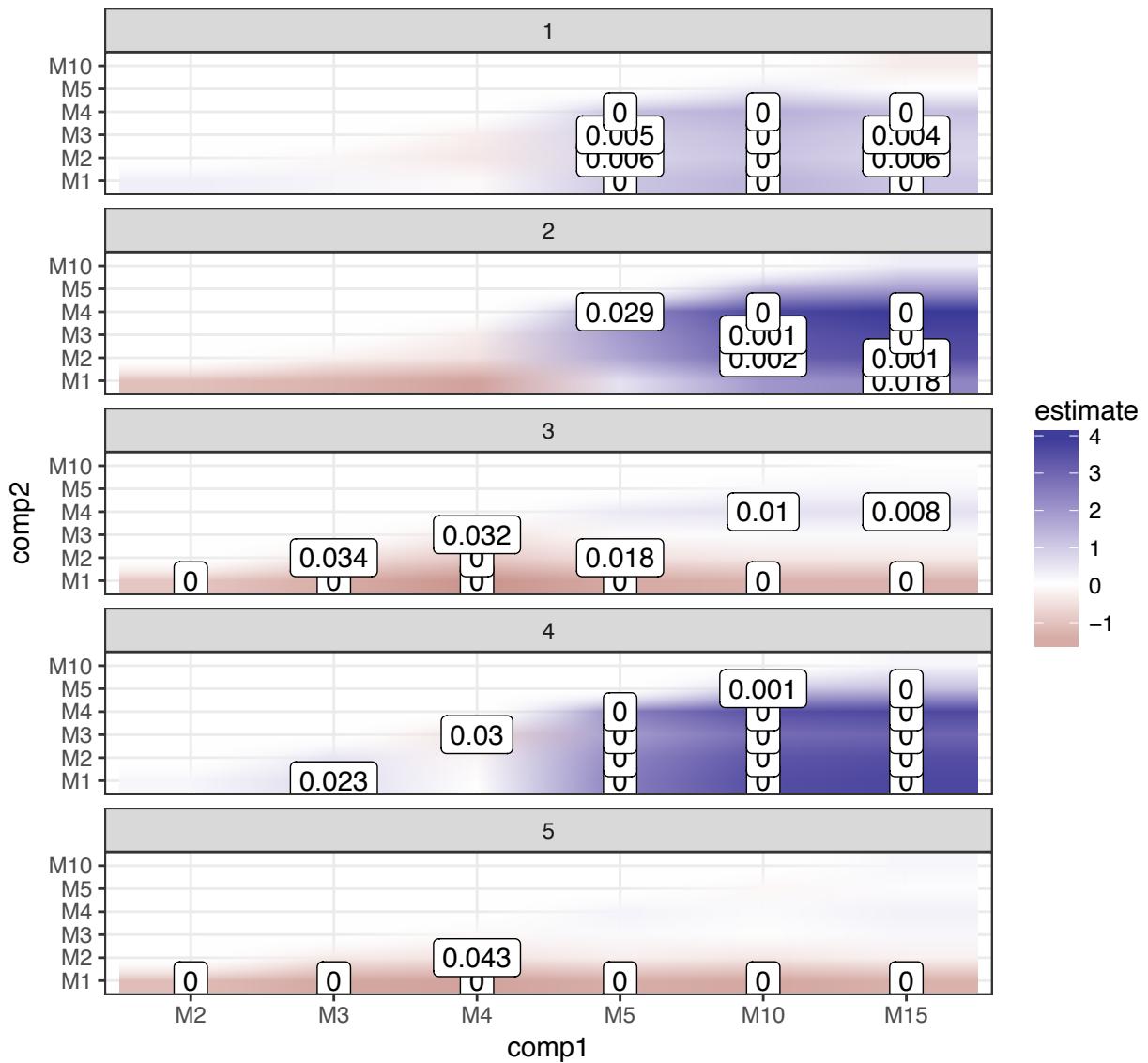


Figure 4: Pairwise comparison of qPCR bacterial abundance measurements for titrations. Fill color represents the estimated pairwise difference. P-values are indicated for statistically significant pairwise comparisons. Estimated pairwise differences and p-values calculated using Tukey-HSD test.

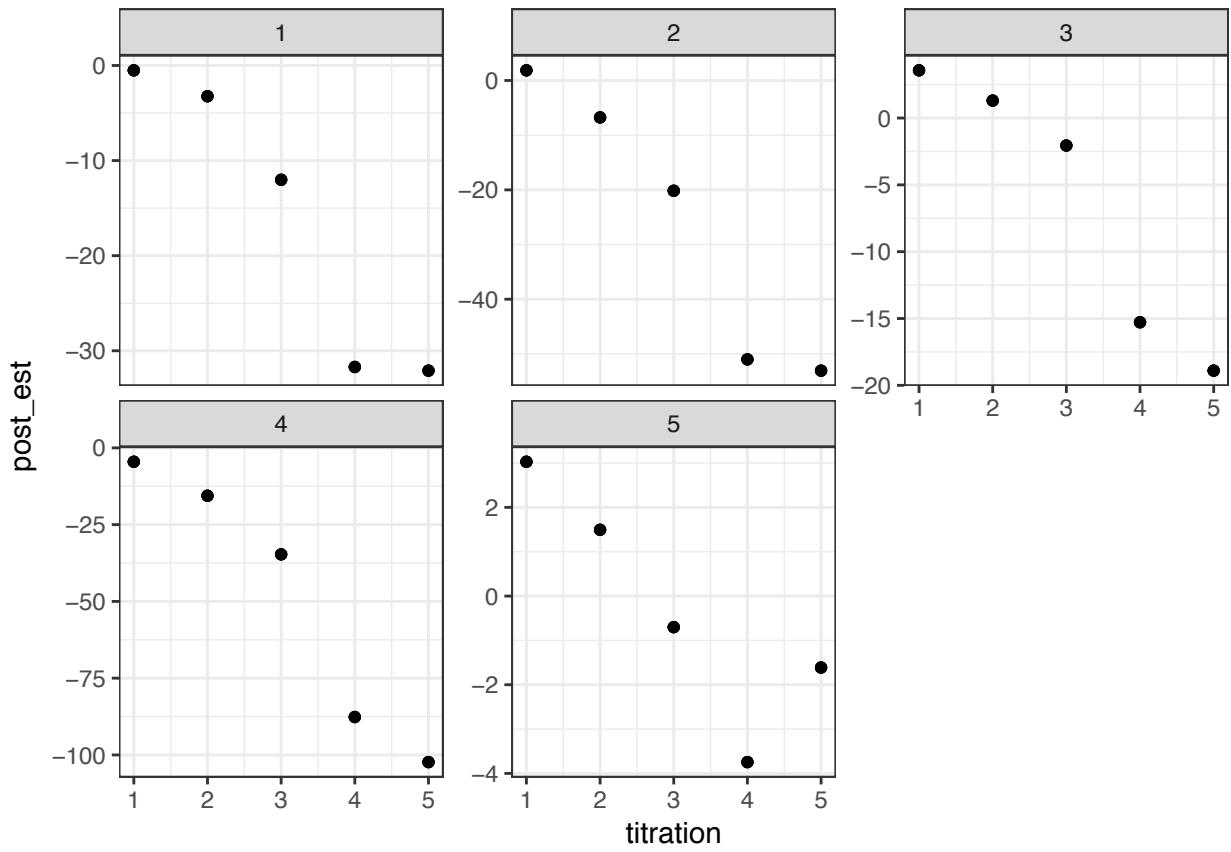


Figure 5: Estimated post-treatment bacterial DNA concentration estimates. Concentration estimated using concentration estimates for unmixed pre-treatment samples and titrations.

biosample_id	lab_id	treatment	timepoint	stool_weight	processing_date	conc_ngul	vol_ul	total_ug	sam
E01JH0004	3	Pre	-1	0.19	2012-03-29	241	20	4.82	5
E01JH0011	8	Pre	-1	0.26	2012-03-29	547	20	10.94	2
E01JH0016	12	Pre	-1	0.20	2012-03-30	266	20	5.32	4
E01JH0017	13	Pre	-1	0.16	2012-03-30	399	20	7.98	3
E01JH0038	28	Pre	-1	0.34	2012-03-30	324	20	6.48	3
E01JH0004	138	Post	4	0.57	2012-04-16	203	20	4.06	6
E01JH0011	115	Post	2	NA	NA	NA	NA	NA	NA
E01JH0016	999	Post	2	NA	NA	NA	NA	NA	NA
E01JH0017	177	Post	5	0.50	2012-04-19	199	20	3.98	6
E01JH0038	105	Post	2	1.01	2012-04-10	543	20	10.86	2

3.1.2 Table with volumes used to make titrations

```
tstPrep %>% kable()
```

Titration	ERCC_ul	Prep_DNA	Prep_vol	Final_vol
(-1)_Pre unmixed	2	100ul_12.5ng/ul	102	21
(1)_Post unmixed	2	100ul_12.5ng/ul	102	92
M1-Mixed	NA	(-1)*10ul+(1)*10ul	20	10
M2-Mixed	NA	(-1)*10ul+M1*10ul	20	10
M3-Mixed	NA	(-1)*10ul+M2*10ul	20	10
M4-Mixed	NA	(-1)*10ul+M3*10ul	20	10
M5-Mixed	NA	(-1)*10ul+M4*10ul	20	19.5
M10-Mixed	NA	(-1)*15.5ul+M5*0.5ul	16	15.5
M15-Mixed	NA	(-1)*15.5ul+M10*0.5ul	16	16

3.1.3 Tube Rack Image

3.1.4 ERCC table - plasmid and spike-ins

```
erccMeta %>% kable()
```

ercc_id	biosample_id	treatment	length	GC	assay_id	amplicon_length
ERCC-00002	E01JH0017	Post	1061	0.53	Ac03459872_a1	69
ERCC-00012	E01JH0004	Post	994	0.52	Ac03459877_a1	77
ERCC-00034	E01JH0011	Pre	1019	0.50	Ac03459987_a1	58
ERCC-00035	E01JH0038	Post	1130	0.52	Ac03459892_a1	65
ERCC-00057	E01JH0016	Pre	1021	0.51	Ac03460000_a1	78
ERCC-00084	E01JH0004	Pre	994	0.52	Ac03459922_a1	63
ERCC-00092	E01JH0038	Pre	1124	0.51	Ac03459925_a1	87
ERCC-00108	E01JH0016	Post	1022	0.50	Ac03460028_a1	74
ERCC-00130	E01JH0017	Pre	1059	0.47	Ac03460039_a1	72
ERCC-00157	E01JH0011	Post	1019	0.51	Ac03459958_a1	71

3.1.5 ERCC Amplification Efficiency Plots

```
ggplot(std_fit) +
  geom_abline(aes(intercept = intercept, slope = slope)) +
  geom_text(aes(x = 3, y = 30, label = paste("R^2:", signif(adj.r.squared,3)))) +
```

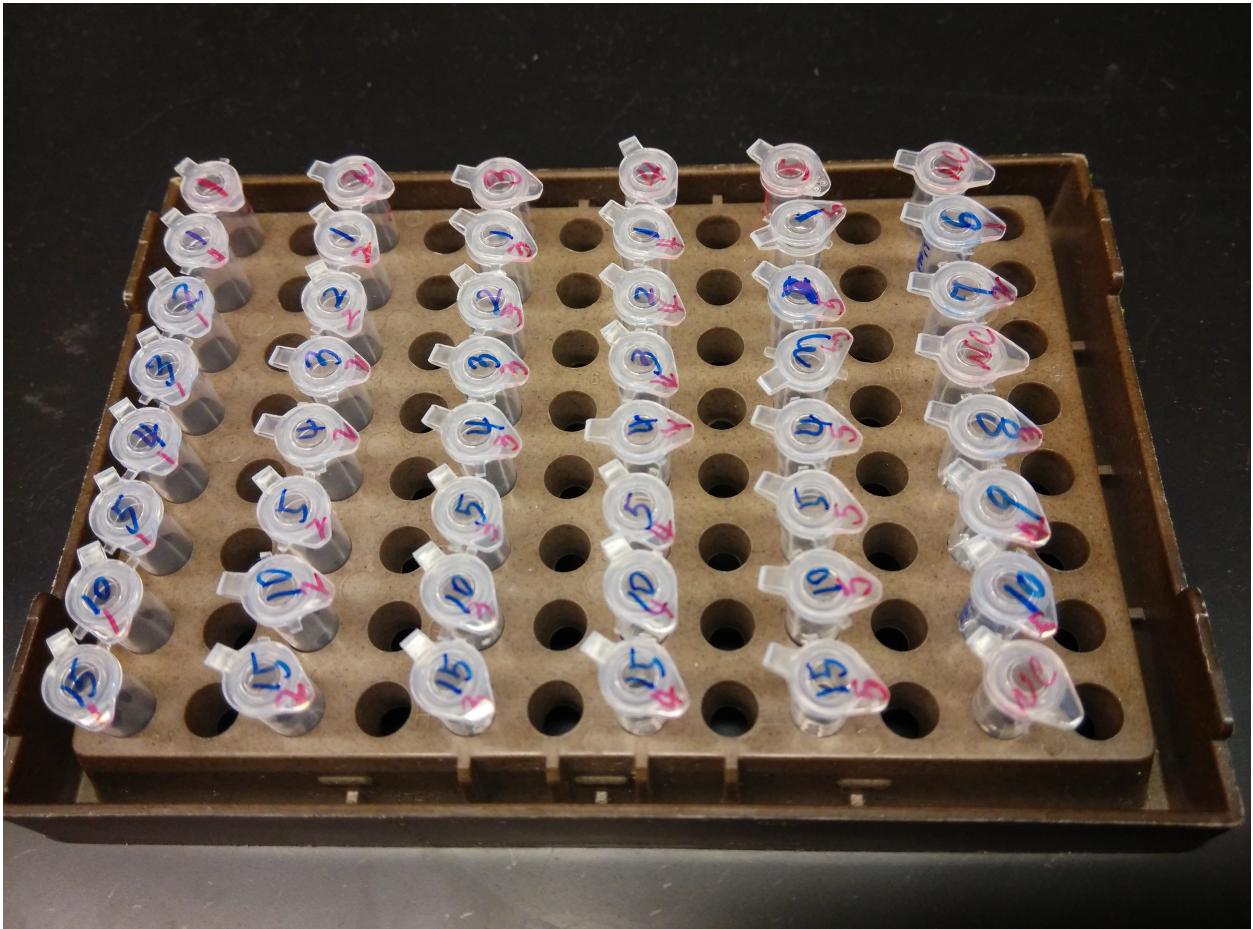
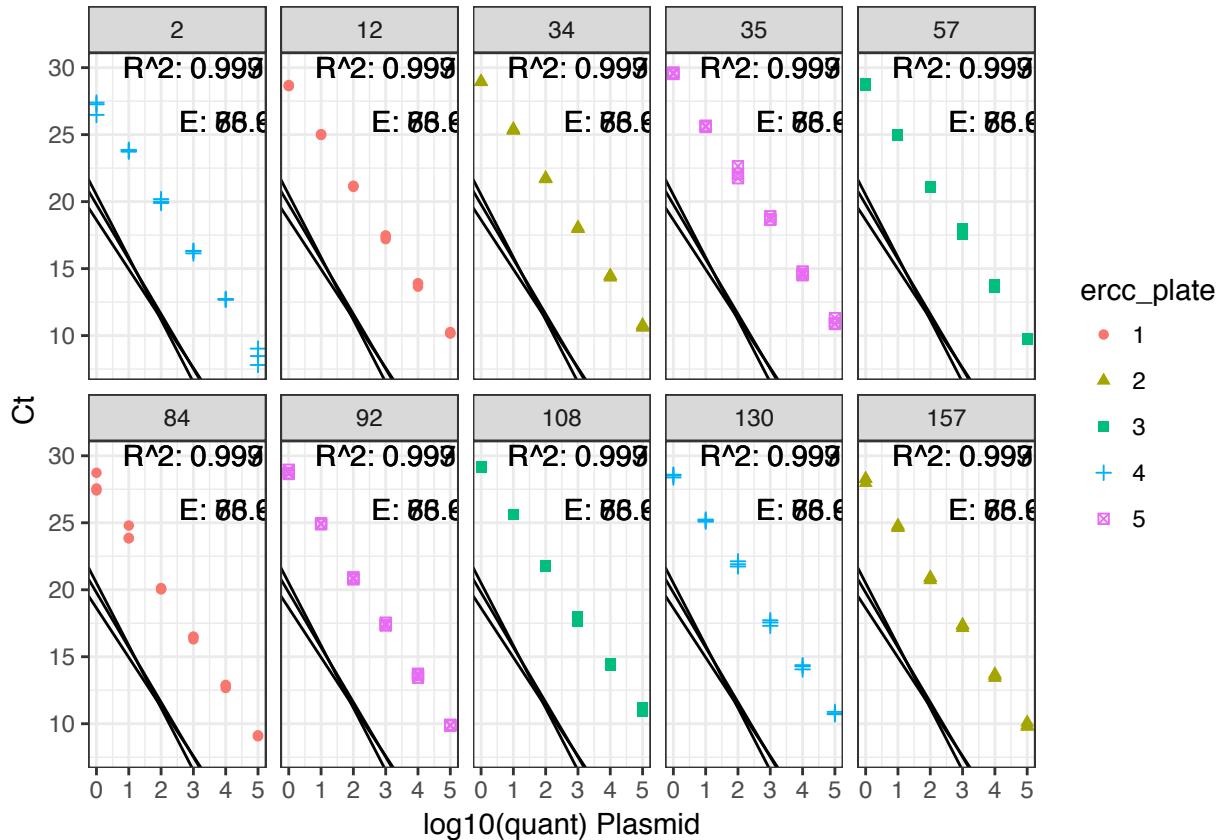


Figure 6: (#fig:plate_map)Image of tubes layout prior to 16S PCR.

```

geom_text(aes(x = 4, y = 26, label = paste("E:", signif( efficiency,3)))) +
geom_point(data = ercc_std, aes(x = log_quant, y = Ct, color = ercc_plate, shape = ercc_plate)) +
facet_wrap(~ercc, ncol = 5) +
theme_bw() +
labs(x = "log10(quant) Plasmid", y = "Ct")

```



3.2 Bacterial Quant

3.2.0.1 Standard Curves

```

std_fit %>% select(std, date, mod, efficiency, adj.r.squared) %>%
arrange(date, std) %>% knitr::kable()

```

std	date	mod	efficiency	adj.r.squared
zymo	2016-09-19	NA	63.86615	0.9933987
shan	2016-12-09	NA	86.86251	0.9992791
zymo	2016-12-09	NA	75.63477	0.9973140

```

qpcrBacStd %>% mutate(log_conc = log10(conc)) %>% ggplot(aes(y = Ct, x = log_conc)) +
geom_vline(aes(xintercept = log10(0.2)), color = "grey60") +
geom_abline(data = std_fit, aes(intercept = intercept, slope = slope)) +
geom_point(aes(color = plate, shape = plate)) +
facet_grid(std~date) + theme_bw() +
theme(legend.position = "bottom")

```

TODO Amplification Curves for standards

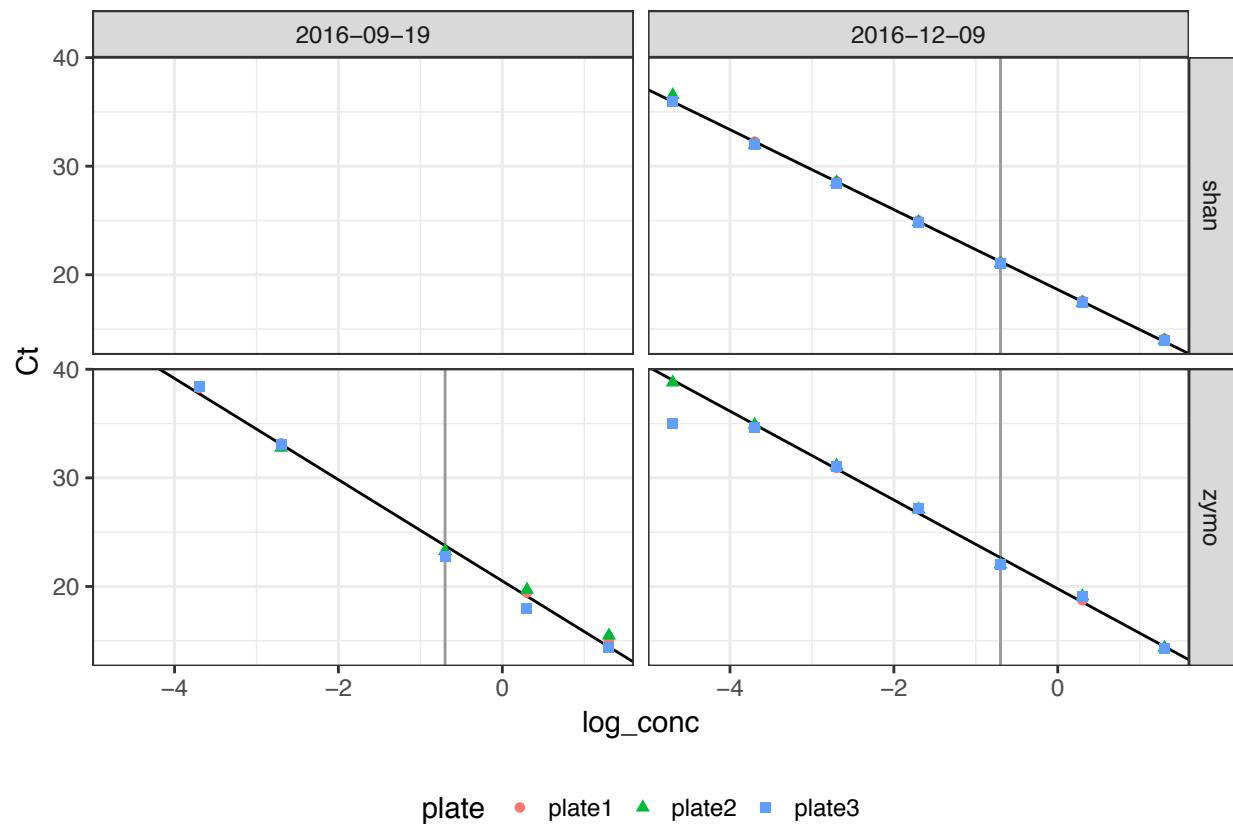


Figure 7: qPCR bacterial abundance standard curves. Using two different standards and performed on two different days. Two models were fit to the standard curves, one with all data point and a second with only the 20 ng/ μ l, 2 ng/ μ l, and 0.2 ng/ μ l standards. Grey vertical line indicates concentration cutoff for subset model.

4 Session information

4.1 Git repo commit information

```
library(git2r)
repo <- repository(path = "../")
last_commit <- commits(repo)[[1]]
```

The current git commit of this file is 567e6e94146ecf216598313c554b3f4a2e35480a, which is on the master branch and was made by nate-d-olson on 2017-04-14 14:29:19. The current commit message is neg binom expected counts and eo metrics. The repository is online at <https://github.com/nate-d-olson/mgtst-pub>

4.2 Platform Information

```
s_info <- devtools::session_info()
print(s_info$platform)

##  setting  value
##  version  R version 3.4.0 (2017-04-21)
##  system   x86_64, darwin16.5.0
##  ui        unknown
##  language (EN)
##  collate  en_US.UTF-8
##  tz       America/New_York
##  date     2017-04-25
```

4.3 Package Versions

```
s_info$packages %>% filter(`*` == "*") %>% select(-`*`) %>%
  knitr::kable()
```

package	version	date	source
bbmle	1.0.19	2017-04-18	CRAN (R 3.4.0)
Biobase	2.35.1	2017-04-24	Bioconductor
BiocGenerics	0.21.3	2017-04-24	Bioconductor
BiocParallel	1.9.6	2017-04-24	Bioconductor
Biostrings	2.43.8	2017-04-24	Bioconductor
DelayedArray	0.1.11	2017-04-24	Bioconductor
DESeq	1.27.0	2017-04-24	Bioconductor
DESeq2	1.15.51	2017-04-24	Bioconductor
dplyr	0.5.0	2016-06-24	CRAN (R 3.4.0)
edgeR	3.17.10	2017-04-24	Bioconductor
forcats	0.2.0	2017-01-23	CRAN (R 3.4.0)
foreach	1.4.3	2015-10-13	CRAN (R 3.4.0)
GenomeInfoDb	1.11.11	2017-04-24	Bioconductor
GenomicAlignments	1.11.12	2017-04-24	Bioconductor
GenomicRanges	1.27.23	2017-04-24	Bioconductor
ggplot2	2.2.1	2016-12-30	CRAN (R 3.4.0)
git2r	0.18.0	2017-01-01	CRAN (R 3.4.0)
glmnet	2.0-5	2016-03-17	CRAN (R 3.4.0)
IRanges	2.9.19	2017-04-24	Bioconductor
knitr	1.15.1	2016-11-22	CRAN (R 3.4.0)
lattice	0.20-35	2017-03-25	CRAN (R 3.4.0)
limma	3.31.22	2017-04-24	Bioconductor
locfit	1.5-9.1	2013-04-20	CRAN (R 3.4.0)
Matrix	1.2-9	2017-03-14	CRAN (R 3.4.0)
matrixStats	0.52.2	2017-04-14	CRAN (R 3.4.0)
metagenomeSeq	1.17.3	2017-04-24	Bioconductor
modelr	0.1.0	2016-08-31	CRAN (R 3.4.0)
permute	0.9-4	2016-09-09	CRAN (R 3.4.0)
phyloseq	1.19.2	2017-04-24	Bioconductor
ProjectTemplate	0.7	2016-08-11	CRAN (R 3.4.0)
purrr	0.2.2	2016-06-18	CRAN (R 3.4.0)
RColorBrewer	1.1-2	2014-12-07	CRAN (R 3.4.0)
readr	1.1.0	2017-03-22	CRAN (R 3.4.0)
readxl	1.0.0	2017-04-18	CRAN (R 3.4.0)
Rqc	1.9.1	2017-04-24	Bioconductor
Rsamtools	1.27.16	2017-04-24	Bioconductor
S4Vectors	0.13.17	2017-04-24	Bioconductor
sads	0.3.1	2016-05-13	CRAN (R 3.4.0)
savR	1.13.0	2017-04-24	Bioconductor
ShortRead	1.33.1	2017-04-24	Bioconductor
stringr	1.2.0	2017-02-18	CRAN (R 3.4.0)
SummarizedExperiment	1.5.10	2017-04-24	Bioconductor
tibble	1.3.0	2017-04-01	CRAN (R 3.4.0)
tidyR	0.6.1	2017-01-10	CRAN (R 3.4.0)
tidyverse	1.1.1	2017-01-27	CRAN (R 3.4.0)
vegan	2.4-3	2017-04-07	CRAN (R 3.4.0)
XVector	0.15.2	2017-04-24	Bioconductor