

Total bacterial 16S rRNA gene copy numbers (qPCR) - revised

Stijn Schreven

4 March 2021

Contents

Introduction	1
Load packages	2
1. Input files	2
1.1. Import data	2
1.2. Prepare data - revised	2
1.3. Functions	3
2. LMM regression	4
2.1. Chicken feed day 0	4
2.2. Chicken feed day 15	4
2.3. Chicken manure day 0	5
2.4. Chicken manure day 15	6
3. Errorbar plots	7
3.1. Collect estimates	7
3.2. Plot - revised	8
4. qPCR egg bacteria	8
4.1. Prepare data - revised	8
4.2. Non-parametric test - revised	9
4.3. Boxplot - revised	9

Introduction

This is the analysis of qPCR output with the revised dataset, *i.e.* including negative samples as zeros. The samples scored as negative were imputed as $\log(\text{gene copy/g}) = \log(1) = 0$, representing “below detection threshold”, *i.e.* “negative sample (within 5 cycles difference of lowest NTC)”. This altered the results displayed in Figure 3 and Supplementary Figure S4.

Load packages

```
library(phyloseq)
library(microbiome)
library(reshape2)
library(plyr)
library(nlme)
library(lme4)
library(car)
library(sciplot)
library(emmeans)
library(ggplot2)
library(viridis)
```

1. Input files

1.1. Import data

```
qpcr <- read.delim("./input_data/Schreven_Ch4_qPCR_data.txt")
ps <- readRDS("./phyobjects/ps.rds")
```

1.2. Prepare data - revised

```
# split original data
qpcr2 <- qpcr
qpcr2a <- subset(qpcr2, Include == TRUE)
qpcr2b <- subset(qpcr2, Include == FALSE)

# impute negative samples to copy number 1 /g sample, and log(1)=0
qpcr2b$DNAcopies_gFM <- 1
qpcr2b$logDNA_gFM <- 0

# exclude egg sample L5 because meltcurve indicated low quality sample
qpcr2b <- qpcr2b[!qpcr2b$Description == "L5", ]

# merge revised data
qpcr2 <- rbind(qpcr2a, qpcr2b)
qpcr2 <- subset(qpcr2, Type != "eggs") # exclude eggs samples (7)
qpcr2 <- subset(qpcr2, Timepoint != 30) # exclude pilot samples (2)
qpcr2 <- subset(qpcr2, Timepoint != 22) # exclude CF LnDs t=22 (3)
qpcr2 <- subset(qpcr2, !Description %in% c("26.K", "39.M"))
qpcr2$Description <- droplevels(qpcr2$Description)
qpcr2 <- qpcr2[,-c(2:6,9)]

# meta data
qpcr.meta <- meta(ps)
qpcr.meta1 <- subset(qpcr.meta, Description %in% qpcr2$Description)
```

```

# merge qPCR and meta data
qpcr2 <- merge(qpcr.meta1, qpcr2, by = "Description")
qpcr2$Treatment <- droplevels(qpcr2$Treatment)
qpcr2$Type <- droplevels(qpcr2$Type)
qpcr2$Timepoint <- droplevels(qpcr2$Timepoint)

# remove isolation duplicates
qpcr2 <- subset(qpcr2, Duplicate == "no")
# exclude containers 6 and 7 because contaminated by fungal overgrowth.
qpcr2 <- subset(qpcr2, !ContainerID %in% c("6","7"))

# subset per diet
qpcr.s <- qpcr2[, c("ContainerID", "Diet", "Treatment", "Timepoint", "Type", "logDNA_gFM")]
## chicken feed substrates
qpcr.cf0 <- subset(qpcr.s, Diet == "CF" & Timepoint == 0 & Treatment != "Ss/E")
qpcr.cf0$Treatment <- droplevels(qpcr.cf0$Treatment)
## chicken feed day 15
qpcr.cf <- subset(qpcr.s, Diet == "CF" & Timepoint == 15 & Treatment != "Ss/E")
qpcr.cf$Treatment <- droplevels(qpcr.cf$Treatment)
## chicken manure
qpcr.cm0 <- subset(qpcr.s, Diet == "CM" & Timepoint == 0)
qpcr.cm <- subset(qpcr.s, Diet == "CM" & Timepoint == 15)

# summarise data: mean, sd, se
qpcr.m <- reshape2::melt(qpcr.s)

```

```
## Using ContainerID, Diet, Treatment, Timepoint, Type as id variables
```

```

qpcr.sum <- ddply(qpcr.m, .(Diet, Treatment, Type, Timepoint, variable),
  summarise, mean = mean(value), sd = sd(value),
  se = se(value), n = length(value))
qpcr.sum$group2 <- interaction(qpcr.sum$Diet, qpcr.sum$Treatment,
  qpcr.sum$Type, qpcr.sum$Timepoint, drop = T)

```

1.3. Functions

```

theme_qpcr <- theme_classic() +
  theme(panel.grid.major = element_line(colour = "grey80"),
    panel.spacing = unit(.5, "lines"),
    panel.border = element_rect(color = "black", fill = NA, size = .5),
    strip.background = element_blank(),
    strip.placement = "outside",
    text = element_text(size=20))

labs_qpcr <- as_labeller(c(
  "0" = "day 0", "15" = "day 15",
  CF = "chicken feed", CM = "chicken manure"))

```

2. LMM regression

On day 15, random term needed for ContainerID (paired observations).

2.1. Chicken feed day 0

Excluded treatment Ss/E, because $n = 2$ (and all zeros).

```
# model selection
qm.cfs0 <- gls(logDNA_gFM ~ Treatment, data = qpcr.cf0, method = "REML")
qm.cfs1 <- update(qm.cfs0, weights = varIdent(form = ~1|Treatment))
AIC(qm.cfs0, qm.cfs1)
```

```
##           df      AIC
## qm.cfs0    4 20.56382
## qm.cfs1    6 21.04217
```

```
# model 0 is best, no variance structure: LM.
qm.cfs <- lm(logDNA_gFM ~ Treatment, data = qpcr.cf0)
```

```
# model output
anova(qm.cfs)
```

```
## Analysis of Variance Table
##
## Response: logDNA_gFM
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Treatment  2 6.4263  3.2132  21.569 0.0003689 ***
## Residuals  9 1.3407  0.1490
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
CLD(emmeans(qm.cfs, ~ Treatment), Letters = letters, method = "tukey")
```

```
## Treatment emmean    SE df lower.CL upper.CL .group
## Si/Es      8.36 0.193  9     7.92     8.79    a
## Si/E       8.48 0.193  9     8.04     8.92    a
## S/E        9.97 0.193  9     9.53    10.40    b
##
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 3 estimates
## significance level used: alpha = 0.05
```

2.2. Chicken feed day 15

Test GLMM Gamma, since LMM residuals were not normal.

```
qgm.cf <- glmer(logDNA_gFM ~ Treatment * Type + (1|ContainerID),
               data = qpcr.cf, nAGQ = 25, family = Gamma)
```

```
# model output
car::Anova(qgm.cf)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: logDNA_gFM
##           Chisq Df Pr(>Chisq)
## Treatment    3.7403  2    0.1541
## Type         43.9196  1  3.421e-11 ***
## Treatment:Type  2.2272  2    0.3284
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
CLD(emmeans(ref_grid(qgm.cf, transform = "response"), ~ Treatment + Type),
     Letters = letters, method = "tukey")
```

```
## Treatment Type      response      SE df asymp.LCL asymp.UCL .group
## Si/Es      larvae      9.39 0.168 Inf      9.06      9.72 a
## Si/E       larvae      9.76 0.176 Inf      9.41     10.11 a
## S/E        larvae      9.97 0.181 Inf      9.61     10.32 ab
## Si/E        substrate 10.51 0.193 Inf     10.14     10.89 bc
## Si/Es        substrate 10.53 0.193 Inf     10.15     10.91 bc
## S/E          substrate 10.74 0.198 Inf     10.35     11.13 c
##
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 6 estimates
## significance level used: alpha = 0.05
```

2.3. Chicken manure day 0

```
# model selection
qm.cms0 <- gls(logDNA_gFM ~ Treatment, data = qpcr.cm0, method = "REML")
qm.cms1 <- update(qm.cms0, weights = varIdent(form = ~1|Treatment))
AIC(qm.cms0, qm.cms1)
```

```
##           df      AIC
## qm.cms0    5 38.48821
## qm.cms1    8 33.38034
```

```
# model 1 is best, variance structure for treatment.
```

```
# model output
anova(qm.cms1)
```

```
## Denom. DF: 19
##           numDF F-value p-value
## (Intercept)    1 70545.77 <.0001
## Treatment      3   66.35 <.0001
```

```
CLD(emmeans(qm.cms1, ~ Treatment), Letters = letters, method = "tukey")
```

```
## Treatment emmean      SE      df lower.CL upper.CL .group
```

```
## Ss/E          9.1 0.234 4.37      8.48      9.73  a
## Si/Es         12.1 0.225 4.38     11.50     12.71  b
## Si/E          12.2 0.101 4.00     11.95     12.51  b
## S/E           12.5 0.055 5.00     12.34     12.62  b
##
## Degrees-of-freedom method: satterthwaite
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 4 estimates
## significance level used: alpha = 0.05
```

2.4. Chicken manure day 15

```
# model selection
qm.cm0 <- lme(logDNA_gFM ~ Treatment * Type, data = qpcr.cm, method = "REML",
              random = ~1 | ContainerID)
qm.cm1 <- update(qm.cm0, weights = varIdent(form = ~1|Treatment))
qm.cm2 <- update(qm.cm0, weights = varIdent(form = ~1|Type))
qm.cm3 <- update(qm.cm0, weights = varIdent(form = ~1|Treatment * Type))
AIC(qm.cm0, qm.cm1, qm.cm2, qm.cm3)
```

```
##          df          AIC
## qm.cm0 10 65.76407
## qm.cm1 13 65.28104
## qm.cm2 11 67.41574
## qm.cm3 17 63.94840
```

```
# model 3 is best, variance structure for Treatment * Type
```

```
# model output
anova(qm.cm3)
```

```
##          numDF denDF  F-value p-value
## (Intercept)      1    20 72110.44 <.0001
## Treatment        3    20   4.13 0.0198
## Type             1    19   68.50 <.0001
## Treatment:Type    3    19   2.76 0.0703
```

```
CLD(emmeans(qm.cm3, ~ Treatment + Type), Letters = letters, method = "tukey")
```

```
## Treatment Type      emmean      SE df lower.CL upper.CL .group
## Si/E      larvae      10.7 0.1072 19      10.5      11.0  a
## S/E      larvae      10.9 0.0755 19      10.7      11.0 ab
## Si/Es     larvae      10.9 0.0795 19      10.7      11.1 abc
## Ss/E      larvae      11.3 0.2138 19      10.8      11.7 abcd
## Ss/E      substrate    11.3 0.0948 20      11.1      11.5  cd
## Si/Es     substrate    11.5 0.1935 20      11.1      11.9 abcd
## S/E      substrate    11.5 0.0948 23      11.4      11.7  d
## Si/E      substrate    11.6 0.2206 20      11.1      12.0 bcd
##
## Degrees-of-freedom method: containment
```

```
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 8 estimates
## significance level used: alpha = 0.05
```

3. Errorbar plots

Based on EMM and SE from models.

3.1. Collect estimates

```
# CF day 0
qcf.emm0 <- CLD(emmeans(qm.cfs, ~ Treatment), Letters = letters, method = "tukey")
qcf.emm0$Timepoint <- 0
qcf.emm0$Diet <- "CF"
qcf.emm0$Type <- "substrate"
qcf.emm0 <- subset(qcf.emm0, select = c(1:3,8:10))

# CF day 15
qcf.emm1 <- CLD(emmeans(ref_grid(qm.cf, transform = "response"), ~ Treatment + Type),
  Letters = letters, method = "tukey")
qcf.emm1$Timepoint <- 15
qcf.emm1$Diet <- "CF"
colnames(qcf.emm1)[3] <- "emmean"
qcf.emm1 <- subset(qcf.emm1, select = c(1:4,9,10))

# extra observations: CF larvae day 15 Ss/E
qcf.emm2 <- subset(qpcr.sum, Diet == "CF" & Treatment == "Ss/E", select = c(1:4,6,8))
colnames(qcf.emm2)[5:6] <- c("emmean", "SE")

# CM day 0
qcm.emm0 <- CLD(emmeans(qm.cms1, ~ Treatment), Letters = letters, method = "tukey")
qcm.emm0$Timepoint <- 0
qcm.emm0$Diet <- "CM"
qcm.emm0$Type <- "substrate"
qcm.emm0 <- subset(qcm.emm0, select = c(1:3,8:10))

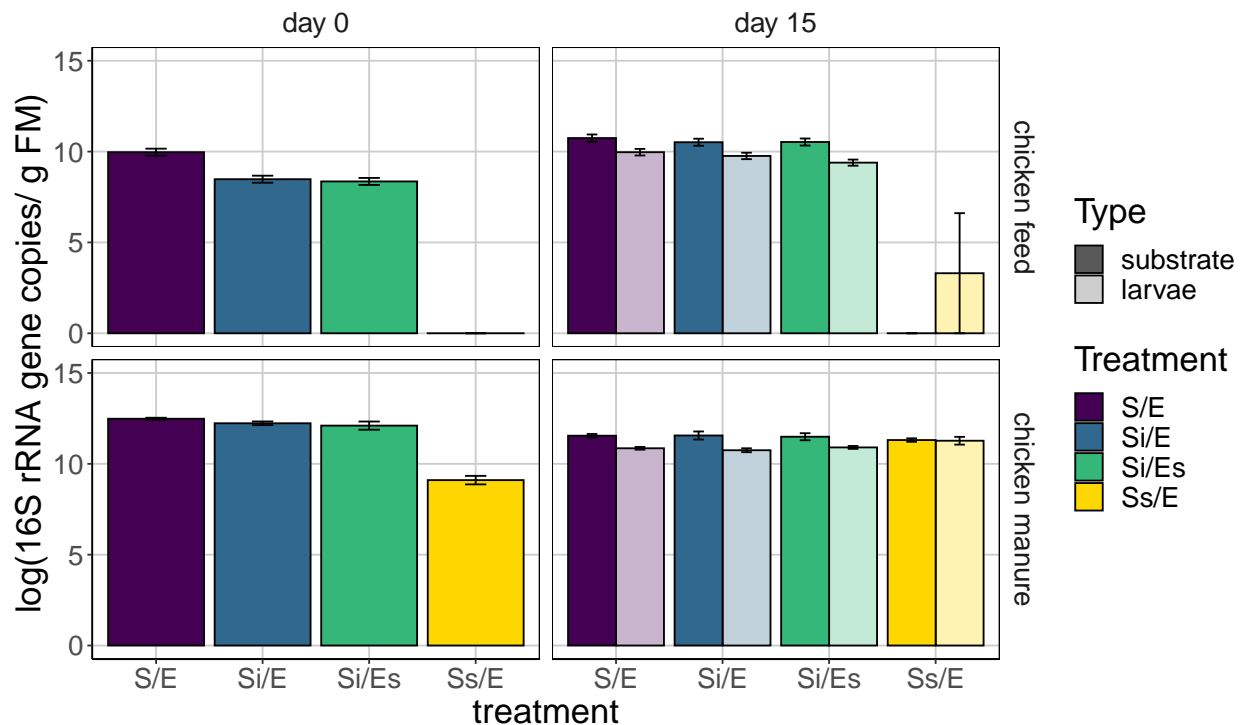
# CM day 15
qcm.emm1 <- CLD(emmeans(qm.cm3, ~ Treatment + Type), Letters = letters, method = "tukey")
qcm.emm1$Timepoint <- 15
qcm.emm1$Diet <- "CM"
colnames(qcm.emm1)[3] <- "emmean"
qcm.emm1 <- subset(qcm.emm1, select = c(1:4,9,10))

# combine into 1 dataframe
qpcr.emm <- rbind(qcf.emm0, qcf.emm1, qcf.emm2, qcm.emm0, qcm.emm1)
qpcr.emm$Type <- as.factor(qpcr.emm$Type)
qpcr.emm$Type <- factor(qpcr.emm$Type, levels(qpcr.emm$Type)[c(2,1)])
```

3.2. Plot - revised

Figure 3 in manuscript.

```
Q.eb <- ggplot(qpcr.emm, aes(x = Treatment, y = emmean, group = Type)) +
  geom_col(position = position_dodge(), aes(y=emmean), fill="white") +
  geom_col(position = position_dodge(),
           aes(y=emmean, fill=Treatment, alpha = Type), colour = "black") +
  scale_fill_manual(values = c("#440154FF", "#31688EFF", "#35B779FF", "gold")) +
  scale_alpha_manual(values = c(1,.3)) +
  geom_errorbar(aes(ymin = emmean - SE, ymax = emmean + SE), width = .2,
               position=position_dodge(width = .9)) +
  labs(y = "log(16S rRNA gene copies/ g FM)", x = "treatment") +
  scale_y_continuous(limits = c(0, 15), n.breaks = 4) +
  facet_grid(Diet ~ Timepoint, labeller = labs_qpcr) +
  theme_qpcr
Q.eb
```



```
ggsave(plot = Q.eb, "./figures/Fig_3_qPCR_revised.png", w = 10, h = 6)
ggsave(plot = Q.eb, "./figures/Fig_3_qPCR_revised.pdf", h = 200, w = 320, u = "mm")
```

4. qPCR egg bacteria

4.1. Prepare data - revised

Dataset including negative samples as zeros. Impute the samples scored as negative as $\log(\text{gene copy/g}) = \log(1) = 0$, representing “below detection threshold”, *i.e.* “negative sample (within 5 cycles difference of lowest NTC)”.


```

qpcr.egg2 <- rbind(qpcr2a, qpcr2b)
qpcr.egg2 <- subset(qpcr.egg2, Type == "eggs")
qpcr.egg2$Description <- droplevels(qpcr.egg2$Description)

# add meta data
qpcr.meta2 <- subset(qpcr.meta, Description %in% qpcr.egg2$Description)
qpcr.egg2 <- merge(qpcr.meta2, qpcr.egg2, by = "Description")
qpcr.egg2$Treatment <- droplevels(qpcr.egg2$Treatment)
qpcr.egg2$Treatment <- revalue(qpcr.egg2$Treatment,
                             c("E"="untreated", "Es" = "disinfected"))
qpcr.egg2 <- subset(qpcr.egg2, Duplicate == "no")

# summarise
qpcr.egg.s2 <- qpcr.egg2[,c("Treatment", "Block", "logDNA_gFM")]
qpcr.egg.m2 <- reshape2::melt(qpcr.egg.s2)

```

```
## Using Treatment, Block as id variables
```

```

qpcr.egg.sum2 <- ddply(qpcr.egg.m2, .(Treatment, variable), summarise,
                      mean = mean(value), median = median(value),
                      sd = sd(value), se = se(value),
                      n = length(value))

```

4.2. Non-parametric test - revised

More suitable for revised data including zeros, since data structure not fit for GLMM.

```
wilcox.test(qpcr.egg.m2$value ~ qpcr.egg.m2$Treatment)
```

```

##
## Wilcoxon rank sum test with continuity correction
##
## data: qpcr.egg.m2$value by qpcr.egg.m2$Treatment
## W = 57.5, p-value = 0.2894
## alternative hypothesis: true location shift is not equal to 0

```

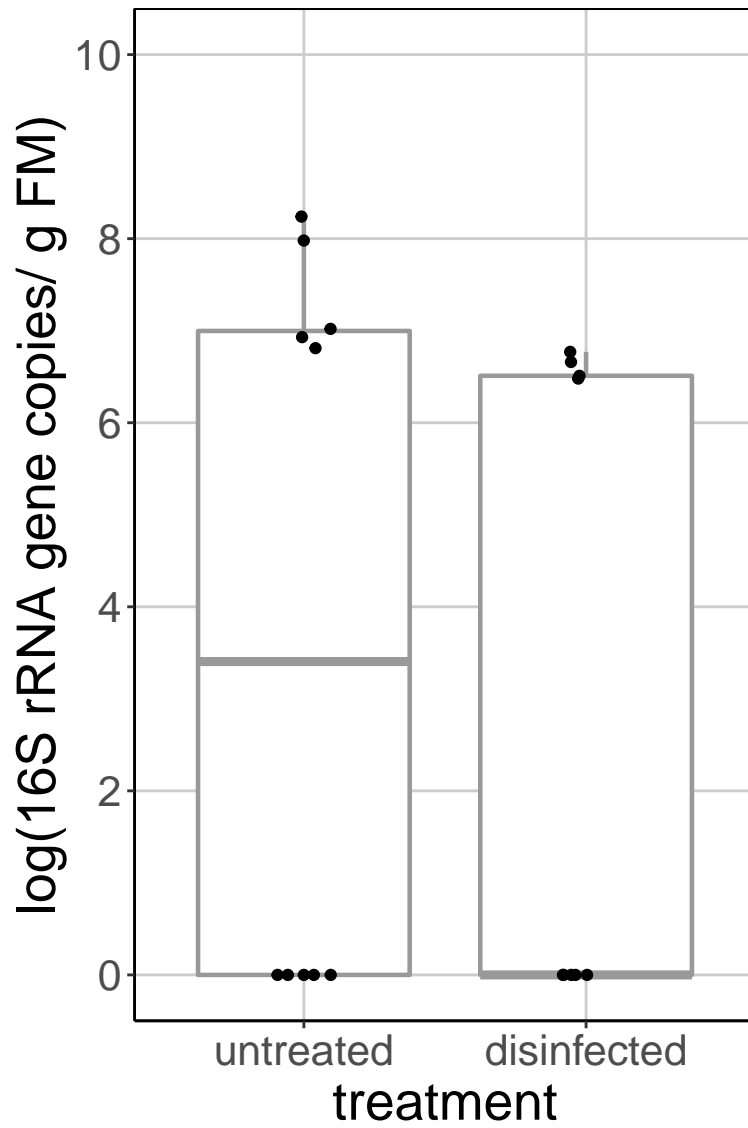
4.3. Boxplot - revised

Supplementary Figure S4 in manuscript.

```

Q.egg2 <- ggplot(data = qpcr.egg.m2, aes(x = Treatment, y = value, colour = Treatment)) +
  geom_boxplot(size = .8, outlier.alpha = 0, col = "grey60") +
  geom_point(position = position_jitter(width = 0.1, height = 0), col = "black") +
  labs(y = "log(16S rRNA gene copies/ g FM)", x = "treatment") +
  scale_y_continuous(limits = c(0, 10), n.breaks = 6) +
  theme_qpcr + theme(legend.position = "none")
Q.egg2

```



```
ggsave(plot = Q.egg2, "./figures/Fig_S4_qPCR_eggs_revised.png", w = 4, h = 6)  
ggsave(plot = Q.egg2, "./figures/Fig_S4_qPCR_eggs_revised.pdf", w = 160, h = 160, u = "mm")
```