

# fishes: Establishing the signal above the noise

copyright Mary Lesperance

4/10/2022

## Contents

<b>R code for paper</b>	<b>2</b>
<b>Outline</b>	<b>2</b>
<b>First run eLowQuant with GEDWG_LOD_DATA#.csv - Extract eONKI4</b>	<b>2</b>
eLowQuant Instructions copied here . . . . .	2
Read in Data . . . . .	3
Pick out the eONKI4 data only . . . . .	3
Process/Summarize samples by Target/Sq; Compute the Poisson estimates of Sq . . . . .	4
Plot the Poisson estimates (and CI) of Sq for levels that had non-detects . . . . .	4
Estimate Poisson models - no intercept model . . . . .	6
Estimate predicted SQ given number detects and technical replicates - no intercept (not shown) . . . . .	7
Estimate predicted SQ given consecutive numbers of detects given number of technical replicates - no intercept . . . . .	7
Estimate Poisson models - intercept model . . . . .	10
Estimate predicted Sq given number detects and technical replicates - intercept model (not shown) . . . . .	11
Estimate predicted SQ given consecutive numbers of detects given number of technical replicates - intercept model . . . . .	11
Lc, Ld, Lq's - no intercept model . . . . .	14
Determine Lc, Ld, Lq (LOB, LOD, LOQ) - no intercept model . . . . .	14
Print out LOD (no intercept model) with confidence intervals, by number of replicates . . . . .	14
Determine Lc, Ld, Lq (LOB, LOD and LOQ) - intercept model . . . . .	14
Print out LOD (intercept model) with confidence intervals, by number of replicates . . . . .	14
Estimates, Lc, Ld, Lq (LOB, LOD, LOQ) and confidence limits for a given number of technical reps NN[NNi] . . . . .	15
eDNA Manuscript Tables and Graphs (not used in <i>fishes</i> manuscript) . . . . .	16
<b>Cq gBlock section, calibration and estimation</b>	<b>16</b>
Medians for CqT=Transformed Cq and Cq . . . . .	16
Plot transformed Cq versus log <sub>10</sub> SQ . . . . .	17
High copy standard curve using Weighted Least Squares and gBlock data . . . . .	17
<b>Density Experiment</b>	<b>21</b>
Load in data from file 'eco' . . . . .	21
Estimate Copy number for Density data using weighted LS in chunk 'calib' . . . . .	21
Ggplot density plot . . . . .	23
<b>Dilution/Flow experiment</b>	<b>23</b>
Load data from file flowFT; estimate Copy number for Flow/Tank Binomial data . . . . .	23
Estimate Copy number for high copy number Flow/Dilution data . . . . .	25
Extra 40Kl Dilution Table . . . . .	27

<b>Background estimation</b>	<b>28</b>
Background estimate from Density experiment negative controls . . . . .	28
Background estimate from Dilution data . . . . .	28
Background estimate from Density and Dilution data . . . . .	29
<b>BACKGROUND Limits of Detection using estimated background SQBboth, No intercept model</b>	<b>29</b>
Print out BACKGROUND limits of detection with confidence intervals, by number of replicates . . . . .	29
Fit Bent Cable model to flow data mean over (Flow/Tank) copy numbers . . . . .	29
Intersections with Bent cable model . . . . .	30
GGplot log2 Flow/Dilution plot using mean copy numbers . . . . .	31

## R code for paper

“Establishing the signal above the noise: Accounting for an environmental background in the detection and quantification of salmonid environmental DNA”, *fishes* , 2022 by Morgan D. Hocking, Jeffrey C. MacAdams, Michael J. Allison, Lauren C. Bergman, Robert Sneiderman Ben F. Koop, Brian M. Starzomski, Mary L. Lesperance and Caren C. Helbing; 7: 266. DOI: 10.3390/fishes7050266

## Outline

### Develop Low/High Copy Standard curves for eONKI4; apply to Density and Dilution series

- (a) Use eLowQuant code (copied here) with gBlock data for eONKI4 to develop a standard curve for the samples with less than 100% detections. We used the no-intercept model, however, code for the intercept model is included here. The eONKI4 gBlock data is in file GEDWG\_LOD\_DATA3.csv. CalibS0.table and CalibS0.table0 contain the SQ estimates SQ0's and SE\_SQ0's as well as model estimates and their standard errors for the intercept and no intercept models respectively for sample sizes 8, 16, 24, 32. The Limits of Blank (LOB) and Detection (LOD) are computed using the Binomial-Poisson model.
- (b) Weighted linear regression is used to develop a standard curve for the high copy data using gBlock Cq data, GEDWG\_LOD\_DATA3.csv, eONKI4 data.
- The Density Experiment - The weighted linear regression model is used to estimate copy number. Log<sub>2</sub> copy number is plotted versus log<sub>2</sub> biomass, and the relationship is modelled using another weighted linear regression model.
- The Dilution Experiment - Both standard curves are used to estimate copy numbers for the Dilution experiment.
- The Background eDNA copy number is estimated using the negative controls (SQ=0) from both the Density and Dilution experiments.
- The Limits of Blank and Detection, adjusted for Background, (LOB-B and LOD-B) are computed using the Binomial-Poisson model and the estimated Background eDNA.
- The Dilution Experiment log<sub>2</sub> copy number estimates are plotted versus log<sub>2</sub> flow together with the adjusted limits of blank and detection. A Bent Cable model is fitted to the data.

## First run eLowQuant with GEDWG\_LOD\_DATA#.csv - Extract eONKI4

### eLowQuant Instructions copied here

This file performs the computations in the publication, ‘A Statistical Model for Calibration and Computation of Detection and Quantification Limits for Low Copy Number Environmental DNA samples’ by Lesperance, Allison, Bergman, Hocking, Helbing, *Environmental DNA*, 2021, 00, 1-12 <https://doi.org/10.1002/edn3.220>.

- Create a folder call Outputs in your working directory.

- Put your data file in your working directory and put the name of the file in the chunk labeled ‘READIN’ below.
- Data set csv file requirements. Columns: Target, Lab, Cq, SQ
- For nondetects, set Cq to be empty or NA value
- For negative controls, set Sq to be empty or 0 or NA value
- Include negative controls (no template controls or NTCs) in the csv file!
- DO NOT DUPLICATE Target names over different Labs!!
- This code uses observations with nonempty data and where  $\text{phat} < 1$  (num detect < num technical replicates)
- Only the SQ’s up to the first one with  $\text{phat} == 1$  are used.
- Allows for variable numbers of SQ levels per Target.
- Assumes SQs == NA are zero, i.e. are negative controls.
- The code uses the R function optim. A convergence code 0 indicates successful completion.
- Ignore warnings if results are sensible.
- EXECUTE EACH CHUNK LOOKING AT THE OUTPUT. IN PARTICULAR, LOOK AT THE GRAPHS IN THE CHUNK CALLED ‘PlotPois’ TO DETERMINE IF THE MODEL IS APPROPRIATE. IT IS NOT APPROPRIATE IF ‘lm Rsq’ is small, i.e. near zero!
- You can send results to files by setting the sink.indicator and Manusink to TRUE and running the code in RStudio. This currently does not work for all results files when knitting.
- If you wish to knit to pdf AND you do NOT have a version of Latex installed on your computer, then run the following in your RStudio console:  
install.packages(“tinytex”); tinytex::install\_tinytex()

## Read in Data

Data should be in the form of a csv file with column names: Target, Lab, Cq, SQ.

Target is the name of the assay,

Lab is the lab name,

Cq is the cycle threshold where a *missing* Cq is a reaction *nondetect*,

SQ is the starting quantity where missing values are assumed to be zero.

It is important to include the negative controls (i.e. SQ=0)!

## Pick out the eONKI4 data only

```
## Number row and columns: 176 10
```

```
## Data Summary
```

```
##      Target      Cq      SQ      Lab      Well
## eONKI4:176 Min.   :21.11 Min.    :  0.03 UVIC:176 Length:176
##              1st Qu.:29.03 1st Qu.:  0.16              Class :character
##              Median :33.33 Median :   4.00              Mode  :character
##              Mean   :32.50 Mean    :3562.50
##              3rd Qu.:36.34 3rd Qu.: 100.00
##              Max.   :46.66 Max.    :62500.00
##              NA's   :52
##      Fluor      Sample      Content      Plate
```

```
## Length:176      Length:176      Length:176      Length:176
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
##
## estimated.SQ
## Min. : NA
## 1st Qu.: NA
## Median : NA
## Mean :NaN
## 3rd Qu.: NA
## Max. : NA
## NA's :176
```

### Process/Summarize samples by Target/Sq; Compute the Poisson estimates of Sq

Hindson et al “High-Throughput Droplet Digital PCR System for Absolute Quantitation of DNA Copy Number”, Anal. Chem., 2011, 83 (22), pp 8604–8610 use a Poisson approximation for quantification. Before that, Dube et al. 2008, “Mathematical analysis of copy number variation in a DNA sample using digital PCR on a nanofluidic device”, PloS One, Vol 3, Issue 8, e2876, model the number of molecules in each chamber as a Poisson process, giving the relationship between  $p$  and  $\lambda$ .

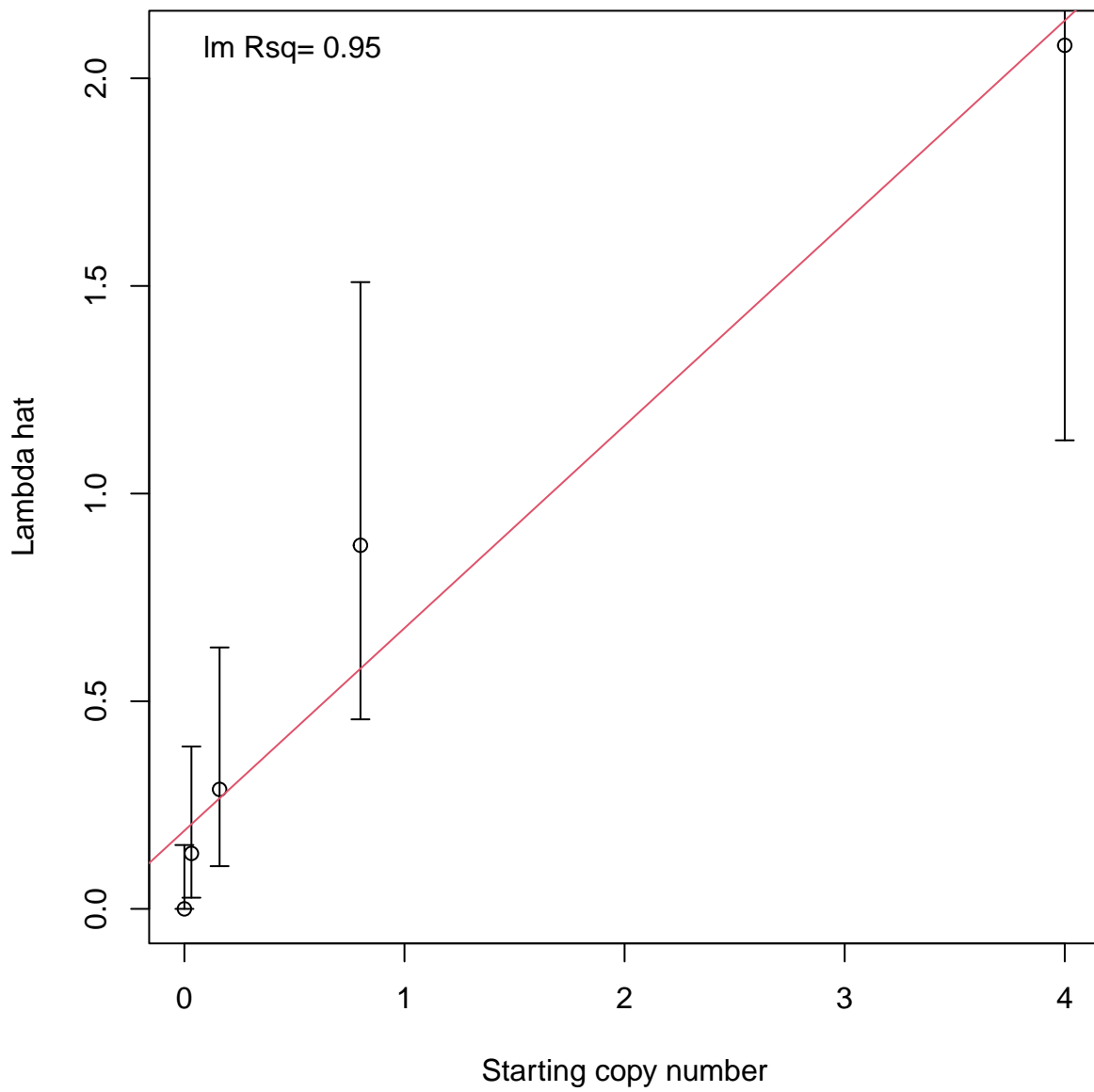
```
## [1] 11 13
##
##
## Table: UVIC
##
## Target      SQ      detect      n      lamhat      phat
## -----
## eONKI4      0.00e+00      0      24      0.000      0.000
## eONKI4      3.20e-02      3      24      0.134      0.125
## eONKI4      1.60e-01      6      24      0.288      0.250
## eONKI4      8.00e-01      14     24      0.875      0.583
## eONKI4      4.00e+00      21     24      2.079      0.875
## eONKI4      2.00e+01      24     24      Inf      1.000
## eONKI4      1.00e+02      24     24      Inf      1.000
## eONKI4      5.00e+02      8      8      Inf      1.000
## eONKI4      2.50e+03      8      8      Inf      1.000
## eONKI4      1.25e+04      8      8      Inf      1.000
## eONKI4      6.25e+04      8      8      Inf      1.000
```

### Plot the Poisson estimates (and CI) of Sq for levels that had non-detects

Only the first levels of SQ that had non-detects are analyzed. Red line is least squares linear regression line.

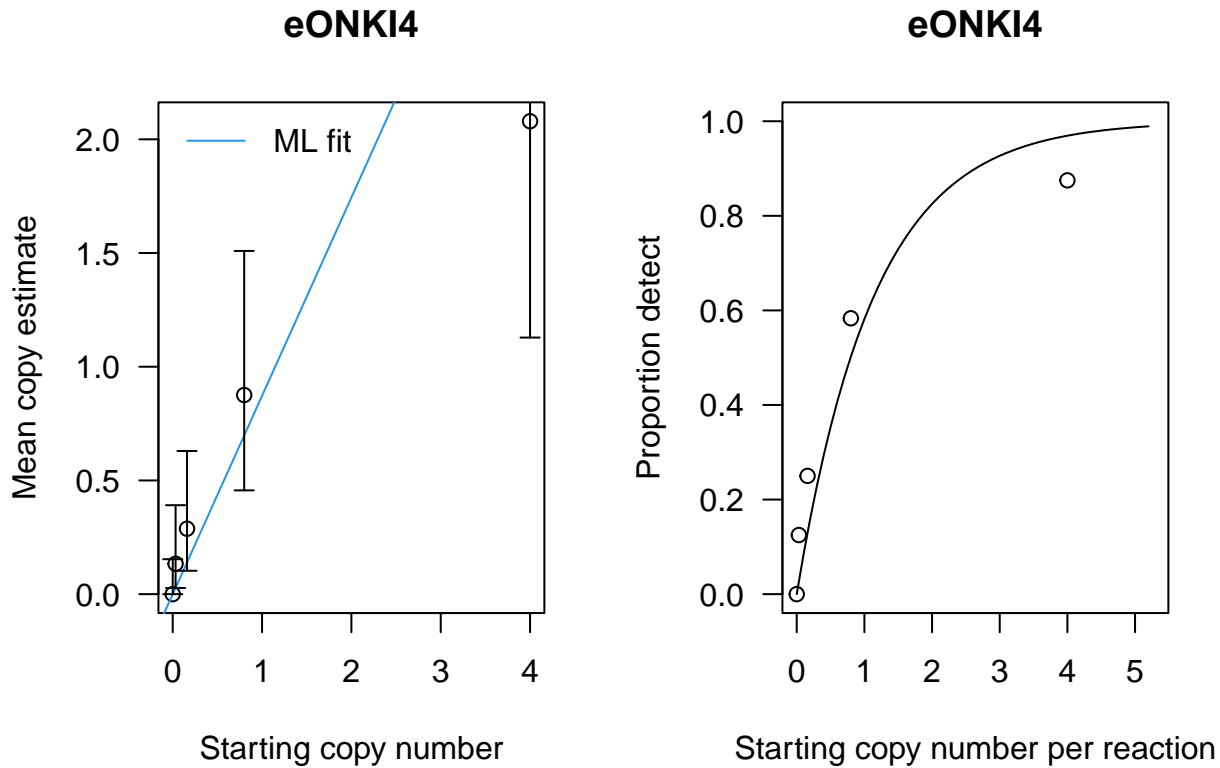
LOOK AT THE GRAPHS IN THE CHUNK CALLED ‘PlotPois’ TO DETERMINE IF THE MODEL IS APPROPRIATE. IT IS NOT APPROPRIATE IF ‘lm Rsq’, the linear model R-square, is small, i.e. near zero!

# UVIC, eONKI4



Both the intercept and no intercept models are fit to the data. The ‘best’ of the two models is the one with the largest Likelihood Ratio test p-value. The ‘best’ model will be identified in the chunk called *Manuscript*.

## Estimate Poisson models - no intercept model



```
##
##
##
##
## eONKI4
## Convergence= 0
##      Estimate Std.Err Z value Pr(>z)
## beta    0.872   0.157   5.54 2.9e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## LLR test stat= 11.91839 , df= 4 , p-value= 0.01796854
```

## Estimate predicted SQ given number detects and technical replicates - no intercept (not shown)

The estimated SQ is easily obtained for given new values of nn0=number of replicates, nd0=number detected and the estimated slope betaS as: `Shat <- -(log((nn0 - nd0)/nn0)) / betaS[1]` The standard errors are obtained from the Hessian matrix (or via the function `CalibS0Or.ddLLik()`).

## Estimate predicted SQ given consecutive numbers of detects given number of technical replicates - no intercept

The estimated SQ is easily obtained for given new values of nn0=number of replicates, nd0=number detected and the estimated slope betaS as: `Shat <- -(log((nn0 - nd0)/nn0)) / betaS[1]` The standard errors are obtained from the Hessian matrix (or via the function `CalibS0Or.ddLLik()`).

```
##
## eONKI4
## ML estimate of SQ for numbers of detects and 8 replicates
##
##
##      NumDetects      SQ0      SE_SQ0
## ---  -
## 0      0      0.0000      NaN
## 1      1      0.1531      0.1557
## 2      2      0.3299      0.2415
## 3      3      0.5390      0.3287
## 4      4      0.7949      0.4300
## 5      5      1.1248      0.5613
## 6      6      1.5897      0.7585
## 7      7      2.3848      1.1558
##
## eONKI4
## ML estimate of SQ for numbers of detects and 16 replicates
##
##
##      NumDetects      SQ0      SE_SQ0
## ---  -
## 0      0      0.0000      NaN
## 1      1      0.0740      0.0752
## 2      2      0.1531      0.1118
## 3      3      0.2380      0.1442
## 4      4      0.3299      0.1759
## 5      5      0.4296      0.2082
## 6      6      0.5390      0.2424
## 7      7      0.6598      0.2794
## 8      8      0.7949      0.3205
## 9      9      0.9480      0.3673
## 10     10      1.1248      0.4221
## 11     11      1.3338      0.4886
## 12     12      1.5897      0.5734
## 13     13      1.9196      0.6899
## 14     14      2.3848      0.8720
## 15     15      3.1786      1.2491
##
## eONKI4
```

## ML estimate of SQ for numbers of detects and 24 replicates

##

##

##	NumDetects	SQ0	SE_SQ0
## ---	-----	-----	-----
## 0	0	0.0000	NaN
## 1	1	0.0488	0.0496
## 2	2	0.0998	0.0728
## 3	3	0.1530	0.0926
## 4	4	0.2091	0.1113
## 5	5	0.2679	0.1294
## 6	6	0.3299	0.1477
## 7	7	0.3954	0.1663
## 8	8	0.4650	0.1855
## 9	9	0.5390	0.2057
## 10	10	0.6181	0.2271
## 11	11	0.7031	0.2499
## 12	12	0.7949	0.2745
## 13	13	0.8946	0.3013
## 14	14	1.0039	0.3309
## 15	15	1.1248	0.3640
## 16	16	1.2598	0.4015
## 17	17	1.4129	0.4450
## 18	18	1.5897	0.4966
## 19	19	1.7988	0.5599
## 20	20	2.0547	0.6413
## 21	21	2.3846	0.7540
## 22	22	2.8494	0.9310
## 23	23	3.6444	1.3008

##

## eONKI4

## ML estimate of SQ for numbers of detects and 32 replicates

##

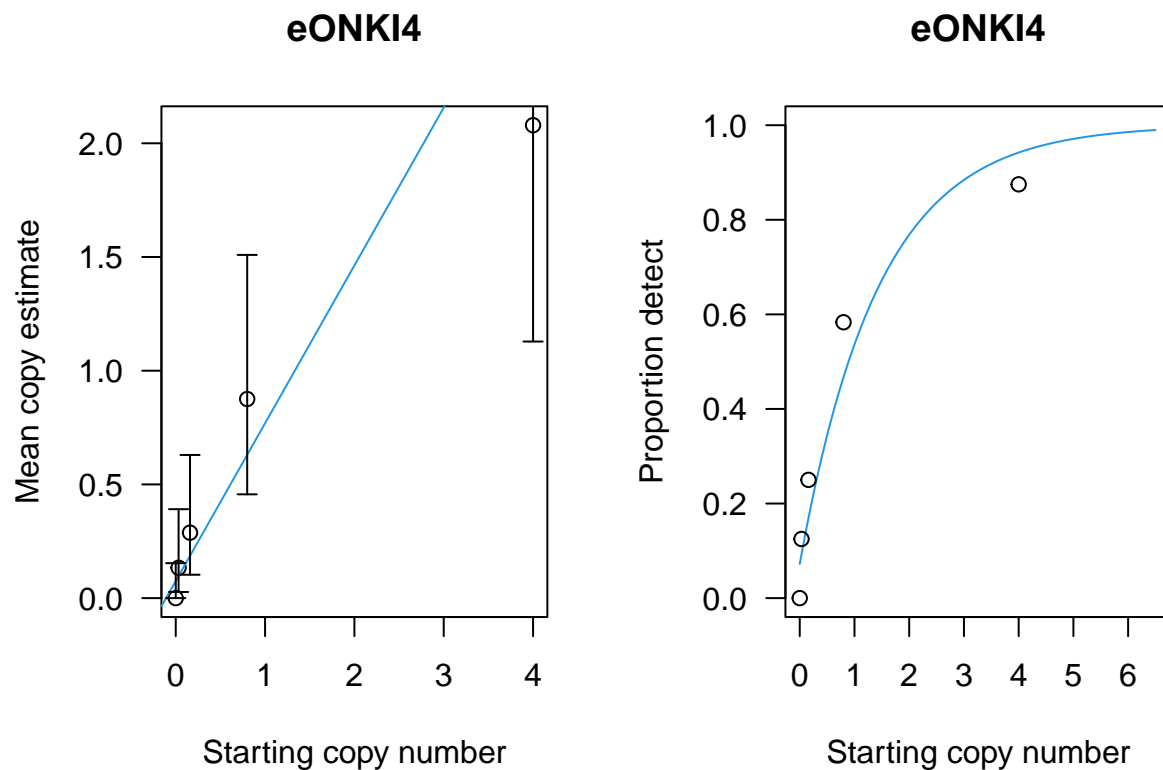
##

##	NumDetects	SQ0	SE_SQ0
## ---	-----	-----	-----
## 0	0	0.0000	NaN
## 1	1	0.0364	0.0370
## 2	2	0.0740	0.0540
## 3	3	0.1128	0.0683
## 4	4	0.1531	0.0814
## 5	5	0.1948	0.0940
## 6	6	0.2381	0.1064
## 7	7	0.2831	0.1188
## 8	8	0.3299	0.1313
## 9	9	0.3787	0.1440
## 10	10	0.4296	0.1571
## 11	11	0.4830	0.1706
## 12	12	0.5390	0.1847
## 13	13	0.5978	0.1994
## 14	14	0.6598	0.2148
## 15	15	0.7253	0.2310
## 16	16	0.7949	0.2483



## 17	17	0.8689	0.2667
## 18	18	0.9480	0.2865
## 19	19	1.0330	0.3078
## 20	20	1.1248	0.3311
## 21	21	1.2245	0.3567
## 22	22	1.3338	0.3851
## 23	23	1.4546	0.4169
## 24	24	1.5897	0.4533
## 25	25	1.7428	0.4955
## 26	26	1.9196	0.5458
## 27	27	2.1287	0.6077
## 28	28	2.3842	0.6873
## 29	29	2.7140	0.7978
## 30	30	3.1794	0.9722
## 31	31	3.9727	1.3360

## Estimate Poisson models - intercept model



```
##
##
##
##
## eONKI4
## Convergence= 0
##      Estimate Std.Err Z value  Pr(>z)
## alpha   0.0749  0.0486   1.54    0.12
## beta    0.6939  0.1601   4.33 1.5e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## LLR test stat= 7.671079 , df= 3 , p-value= 0.05332185
```

Estimate predicted Sq given number detects and technical replicates - intercept model (not shown)

Estimate predicted SQ given consecutive numbers of detects given number of technical replicates - intercept model

The estimated SQ is easily obtained for given new values of nn0=number of replicates, nd0=number detected and the estimated slope betaS as:  $\text{Shat} \leftarrow -(\log((nn0 - nd0)/nn0) + \text{betaS}[1]) / \text{betaS}[2]$  The standard errors are obtained from the Hessian matrix (or via the function CalibS0.ddLLik()).

```
##
## eONKI4
## ML estimate of SQ for numbers of detects and 8 replicates
##
##
```

	NumDetects	SQ0	SE_SQ0
0	0	0.000	0.000
1	1	0.084	0.203
2	2	0.307	0.303
3	3	0.569	0.412
4	4	0.891	0.542
5	5	1.305	0.714
6	6	1.890	0.973
7	7	2.889	1.492

```
##
## eONKI4
## ML estimate of SQ for numbers of detects and 16 replicates
##
##
```

	NumDetects	SQ0	SE_SQ0
0	0	0.000	0.000
1	1	0.000	0.000
2	2	0.085	0.150
3	3	0.191	0.185
4	4	0.306	0.221
5	5	0.432	0.260
6	6	0.569	0.303
7	7	0.722	0.351
8	8	0.891	0.405
9	9	1.083	0.468
10	10	1.305	0.542
11	11	1.568	0.632
12	12	1.890	0.747
13	13	2.304	0.904
14	14	2.889	1.147
15	15	3.888	1.644

```
##
## eONKI4
## ML estimate of SQ for numbers of detects and 24 replicates
##
##
```

	NumDetects	SQ0	SE_SQ0
--	------------	-----	--------

##	---	-----	-----	-----
## 0		0	0.000	0.000
## 1		1	0.000	0.000
## 2		2	0.017	0.112
## 3		3	0.085	0.128
## 4		4	0.155	0.146
## 5		5	0.229	0.165
## 6		6	0.307	0.185
## 7		7	0.389	0.207
## 8		8	0.476	0.231
## 9		9	0.569	0.257
## 10		10	0.669	0.285
## 11		11	0.775	0.315
## 12		12	0.891	0.348
## 13		13	1.016	0.384
## 14		14	1.154	0.425
## 15		15	1.305	0.470
## 16		16	1.475	0.522
## 17		17	1.666	0.582
## 18		18	1.890	0.654
## 19		19	2.153	0.741
## 20		20	2.474	0.853
## 21		21	2.889	1.007
## 22		22	3.473	1.245
## 23		23	4.472	1.731
##				
##	eONKI4			
##	ML estimate of SQ for numbers of detects and 32 replicates			
##				
##				
##		NumDetects	SQ0	SE_SQ0
##	---	-----	-----	-----
## 0		0	0.000	0.000
## 1		1	0.000	0.000
## 2		2	0.000	0.000
## 3		3	0.034	0.106
## 4		4	0.084	0.116
## 5		5	0.137	0.126
## 6		6	0.191	0.138
## 7		7	0.248	0.151
## 8		8	0.306	0.165
## 9		9	0.368	0.180
## 10		10	0.432	0.195
## 11		11	0.499	0.212
## 12		12	0.570	0.231
## 13		13	0.643	0.249
## 14		14	0.721	0.270
## 15		15	0.803	0.292
## 16		16	0.891	0.315
## 17		17	0.984	0.340
## 18		18	1.083	0.368
## 19		19	1.190	0.398
## 20		20	1.305	0.430

## 21	21	1.431	0.466
## 22	22	1.568	0.506
## 23	23	1.720	0.551
## 24	24	1.890	0.602
## 25	25	2.082	0.662
## 26	26	2.304	0.732
## 27	27	2.567	0.818
## 28	28	2.889	0.929
## 29	29	3.303	1.080
## 30	30	3.888	1.314
## 31	31	4.887	1.794

## Lc, Ld, Lq's - no intercept model

### Determine Lc, Ld, Lq (LOB, LOD, LOQ) - no intercept model

Follows Lavagnini and Magno 2007, Mass Spectrometry Reviews *The notation in our eDNA paper is different from the Lavagnini 2007 paper and is shown in brackets here.*

- Lc (*LOB Limit of blank*) = critical level is the assay signal above which a response is reliably attributed to the presence of analyte
- Ld (*Ld = expected number detects out of NN replicates at concentration LOD*) = signal corresponding to an analyte concentration  $x_d$  (*=LOD Limit of Detection*) level which may be a priori expected to be recognized
- Lq = quantification limit is a signal with a precision which satisfies an expected value ( $=\gamma_Q$ )

Lc corresponds to a critical response level or a false positive rate, i.e. critical number of detects given NN replicates, above which we would reject the null hypothesis that the concentration/copy number is zero at  $\alpha = \alpha_{Lc}$  ( $=\gamma_{FP}$ ). It is the critical response level corresponding to the false positive rate of  $\alpha_{Lc}$ . Essentially, the test is positive if the  $Y \sim \text{Binomial}(m, p) > Lc$ . The False Positive Rate is  $P(Y > Lc \mid S=0)$ .

Ld is computed to correspond to the false negative rate,  $\beta = \beta_{Ld}$  ( $=\gamma_{FN}$ ) here. It is computed so that the probability of observing a new (unknown concentration) response less than or equal to Lc is less than or equal to  $\beta_{Ld}$ . The probability of observing Lc or less detects if the concentration is  $x_d$  (*=LOD Limit of Detection*) or more is less than or equal to  $\beta_{Ld}$ . The values of Lc depend on the number of replicates, NN, so  $x_d$  does as well. Ld is the expected number of detects at values  $x_d$  (*=LOD Limit of Detection*) and NN replicates. False negative rate Ld computation:  $P(Y \leq Lc \mid p_{x_d}) \leq \beta_{Ld}$ , ( $=\gamma_{FN}$ ) and solve for  $x_d$  (*=LOD Limit of Detection*).

Lq is less well defined. The literature suggests using  $Lq = \beta_0 + 10 \text{ s.e.}(\beta_0)$ , but this uses the normality assumption. Other literature suggests using the “analyte concentration  $x_q$  (*=LOQ Limit of Quantification*) for which the experimental relative standard deviation of the responses reaches a fixed level ( $=\gamma_Q$ ), for example, the level 0.1.” Lavagnini and Magno 2007. I interpret the term “relative standard deviation” to mean the coefficient of variation,  $CV = sd/\text{mean}$ .

In the exercise below, we use the fits from the ML models to estimate the Lc, Ld and Lq, for various values of NN replicates for a new observation, i.e. a new (unknown concentration) response number of detects. Both the intercept and no intercept models are considered.

### Print out LOD (no intercept model) with confidence intervals, by number of replicates

```
##           3    8   16   24   32   48   64   96
## LOD       1.15 0.43 0.21 0.14 0.11 0.07 0.05 0.04
## LOD_lower 0.85 0.32 0.16 0.11 0.08 0.05 0.04 0.03
## LOD_upper 1.77 0.66 0.33 0.22 0.17 0.11 0.08 0.06
```

### Determine Lc, Ld, Lq (LOB, LOD and LOQ) - intercept model

#### Print out LOD (intercept model) with confidence intervals, by number of replicates

```
##           3    8   16   24   32   48   64   96
## LOB       1.0 2.00 3.00 4.00 5.00 7.00 8.00 11.00
## LOD       2.8 1.21 0.67 0.49 0.41 0.32 0.24 0.18
## LOD_lower 1.8 0.74 0.37 0.25 0.19 0.13 0.07 0.03
## LOD_upper 5.3 2.41 1.42 1.10 0.94 0.78 0.63 0.53
```

**Estimates, Lc, Ld, Lq (LOB, LOD, LOQ) and confidence limits for a given number of technical reps NN[NNi]**

Chooses the model (intercept versus no intercept) with the best LLR test fit, i.e. the largest p-value for the LLR test. A table of values for all assays is printed.

## Limits for best choice model for N= 8

## InterModel	alpha	aSE	beta	bSE	LOB	LOD_Low
## 1.00	0.07	0.05	0.69	0.16	2.00	0.74
## LOD	LOD_Up	LOQ_Low	LOQ	LOQ_Up		
## 1.21	2.41	1.24	1.93	3.73		

eDNA Manuscript Tables and Graphs (not used in *fishes* manuscript)

*Revised for general use to use all eligible targets.*  
*Revised to output results for both intercept and no intercept models.*

Cq gBlock section, calibration and estimation

Medians for CqT=Transformed Cq and Cq

## [1] "number of zero Cqs: 0"

Table 1: Transformed Cq medians by SQ (nondetect = 0)

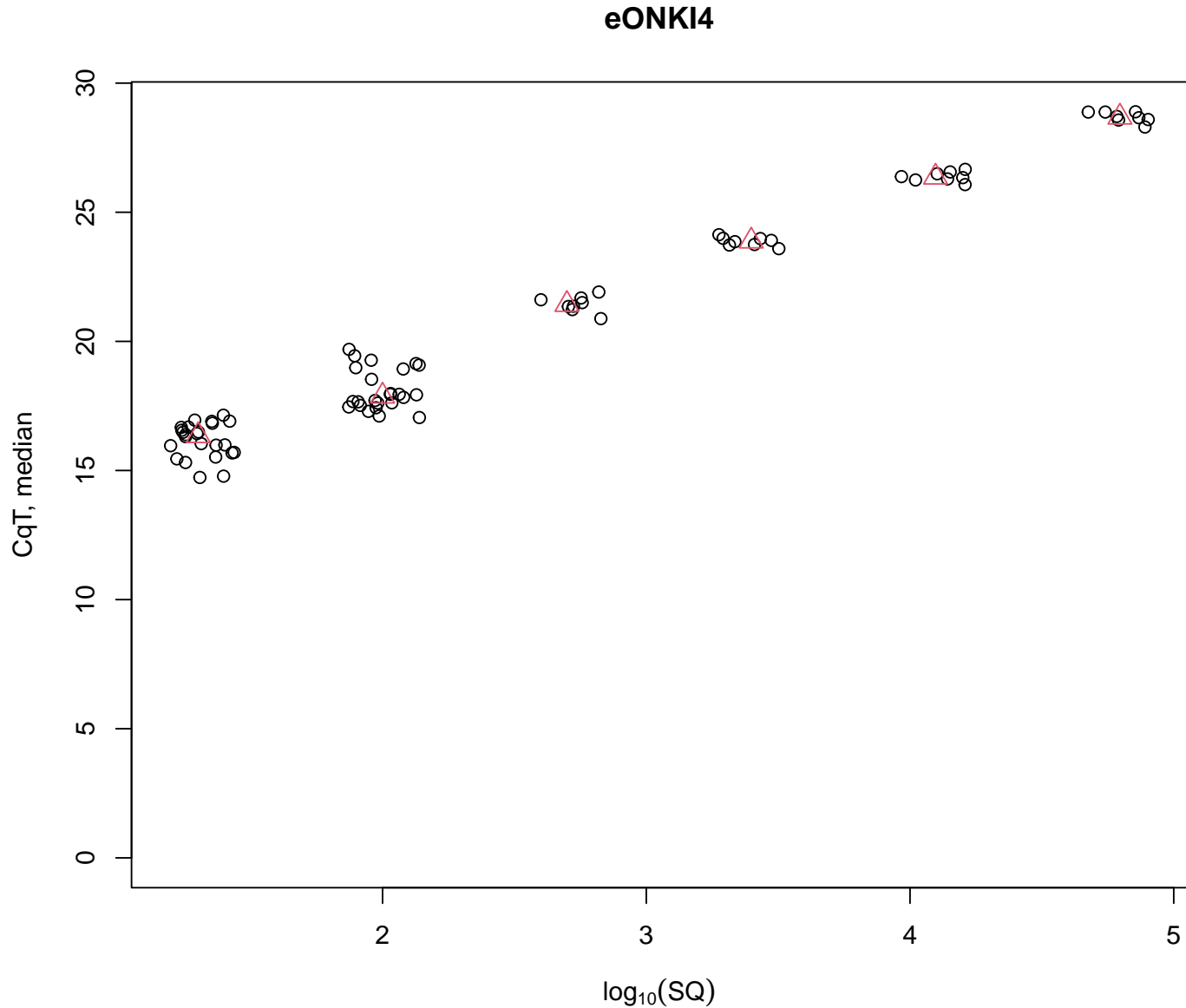
	20	100	500	2500	12500	62500
eONKI4	16.35	17.88	21.43	23.88	26.36	28.69

Table 2: Cq medians by SQ (nondetect removed)

	20	100	500	2500	12500	62500
eONKI4	33.65	32.12	28.57	26.12	23.64	21.31



## Plot transformed Cq versus log<sub>10</sub> SQ



### High copy standard curve using Weighted Least Squares and gBlock data

Fit regression models to Cq's for calibration of SQ on log<sub>10</sub> SQ scale using method of Lavignini et al. 2007, bottom page 11. The R code names follow the notation of Laivgnini et al. 2007. For example, Xwhat is  $\hat{X}_w$ .

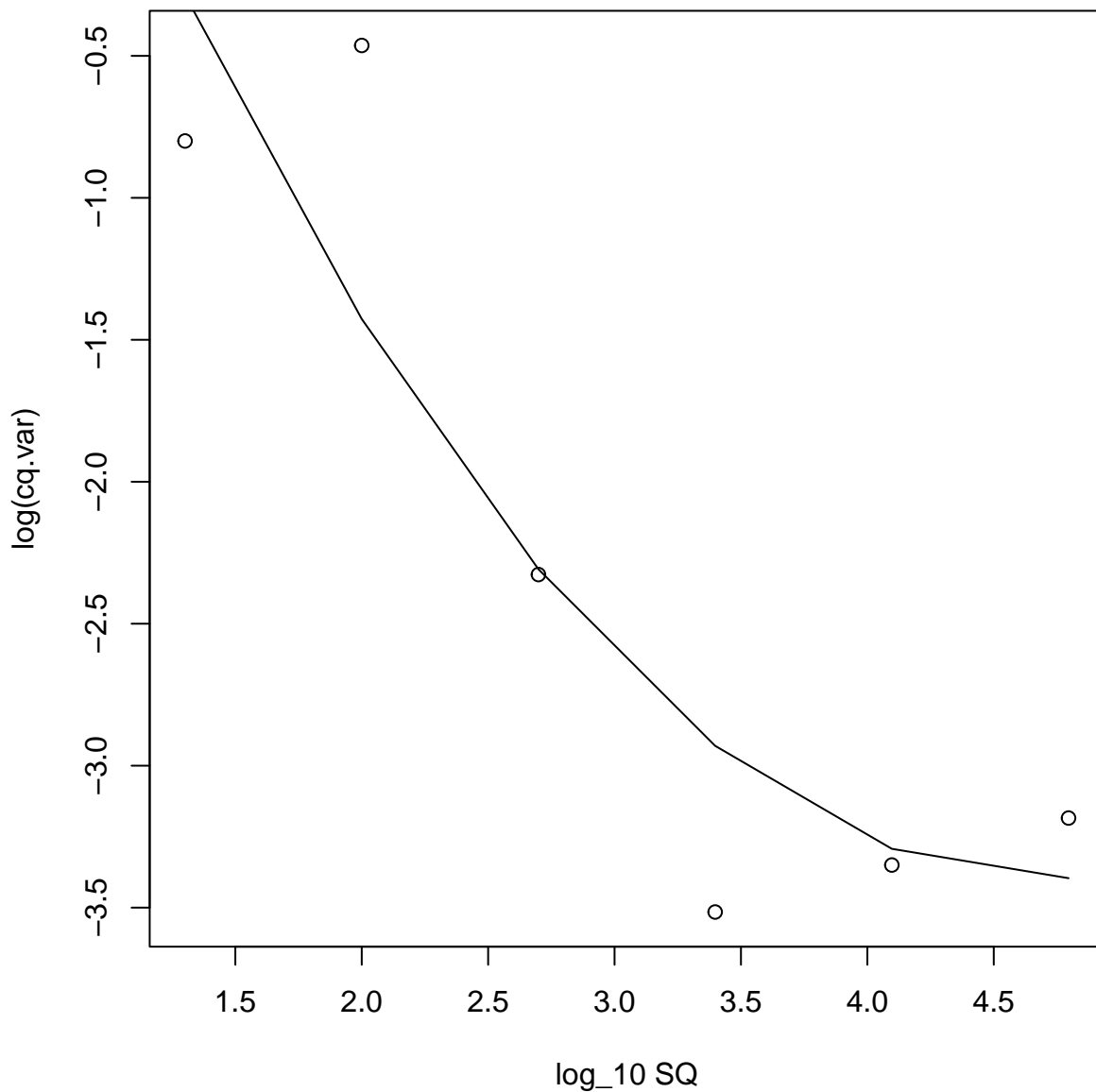
Aside: Kralik, Petr and Ricchi, Matteo (2017) “A basic guide to real time PCR in microbial diagnostics: Definitions, parameters, and everything”, Frontiers in Microbiology, 1-9.

Mathematical principle of qPCR.

$$N_n = N_0 \times (1 + E)^n.$$

$N_n$  is the number of PCR amplicons after  $n$  cycles,  $N_0$  is the initial number of template copies in the sample,  $E$  is the PCR efficiency that can assume values in the range from 0 to 1,  $n$  is the number of cycles.

## Cq variances versus log<sub>10</sub> SQ



```
## Model for the variances/weights for inverse regression which can be used to estimate weights for a
##
## Call:
## lm(formula = Cq ~ L10.SQ, data = DATz.df, subset = indcal, weights = 1/Cq.var)
##
## Weighted Residuals:
##      Min       1Q   Median       3Q      Max
## -2.08556 -0.67820  0.09084  1.13515  2.61088
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  38.48862    0.15055  255.66  <2e-16 ***
## L10.SQ       -3.61140    0.04003  -90.22  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

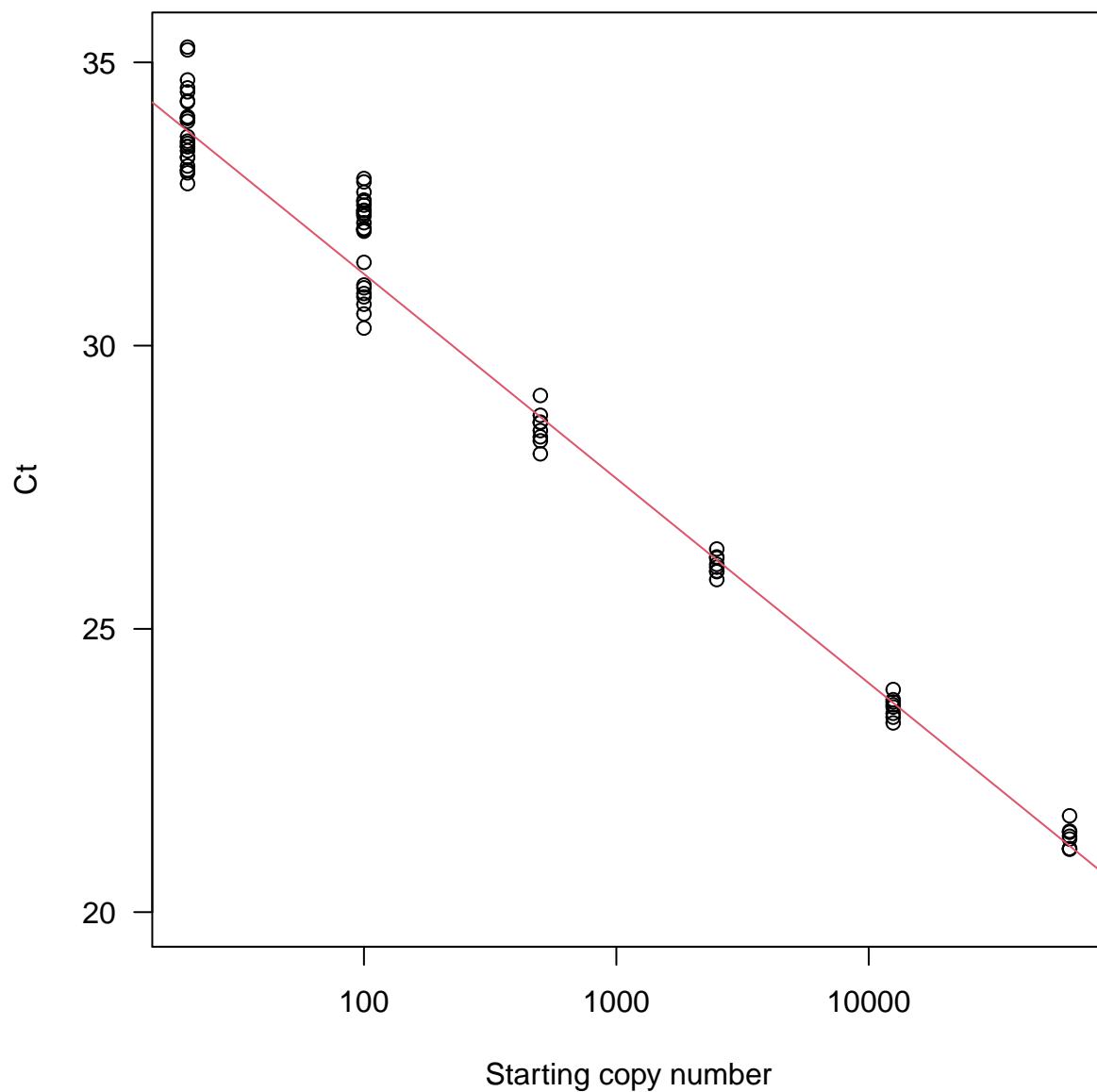
##

## Residual standard error: 1.124 on 78 degrees of freedom

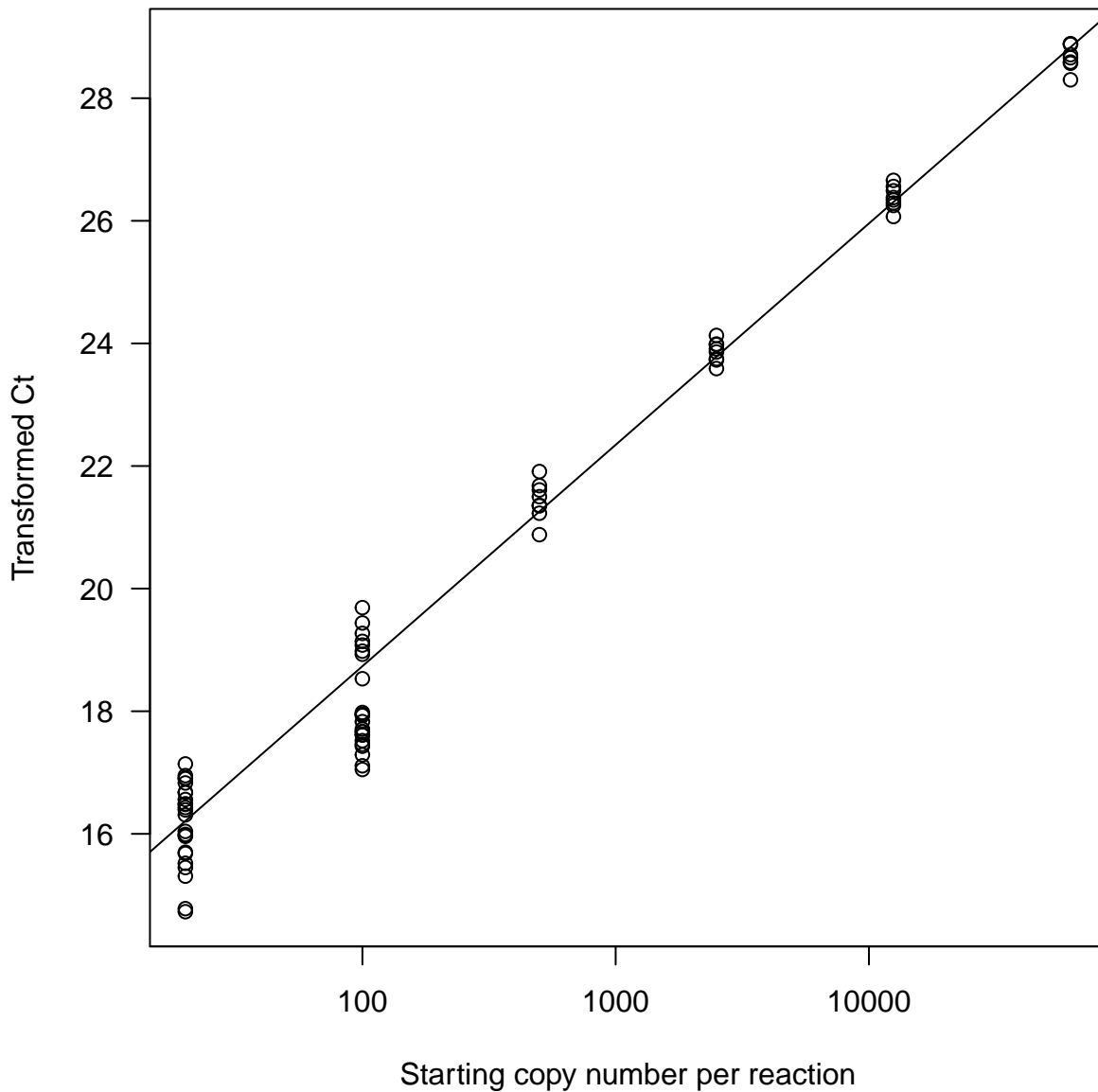
## Multiple R-squared: 0.9905, Adjusted R-squared: 0.9904

## F-statistic: 8139 on 1 and 78 DF, p-value: < 2.2e-16

### Ct versus SQ, weighted regression



## Transformed Ct versus SQ, weighted regression



```
##
##
##
## Call:
## lm(formula = CqT ~ L10.SQ, data = DATz.df, subset = indcal, weights = 1/Cq.var)
##
## Weighted Residuals:
##      Min       1Q   Median       3Q      Max
## -2.61088 -1.13515 -0.09084  0.67820  2.08556
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  11.51138    0.15055   76.46  <2e-16 ***
## L10.SQ        3.61140    0.04003   90.22  <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.124 on 78 degrees of freedom
## Multiple R-squared:  0.9905, Adjusted R-squared:  0.9904
## F-statistic: 8139 on 1 and 78 DF,  p-value: < 2.2e-16
##
## eONKI4 E= 0.891911 efficiency
## eONKI4 OLS coefs 38.84488 -3.682191
## eONKI4 WLS coefs 38.48862 -3.611403
## eONKI4 CVs by SQ
##          20          100          500          2500          12500          62500
## 0.14416112 0.12006365 0.03146560 0.01395588 0.01259780 0.01184782
```

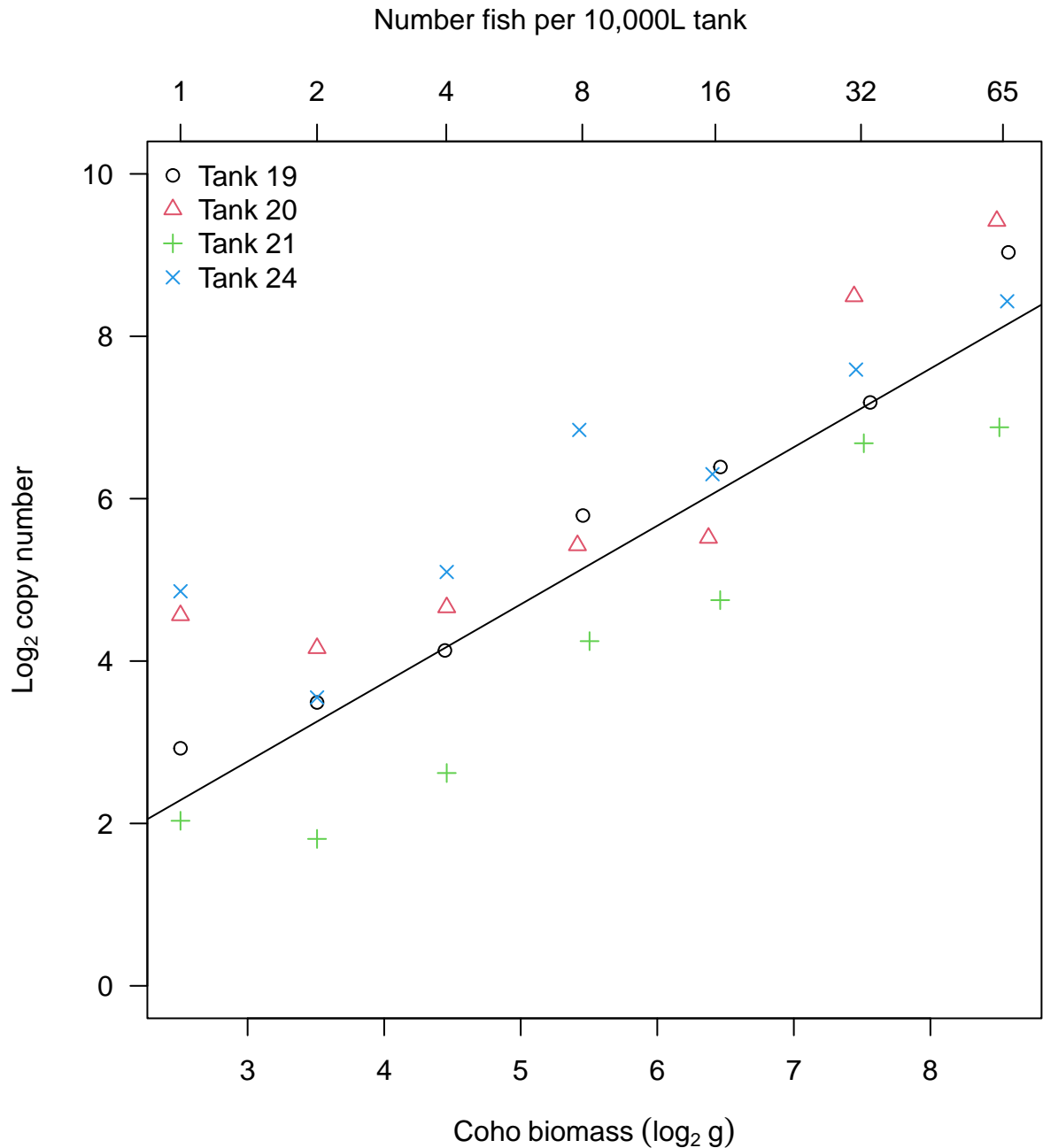
## Density Experiment

Load in data from file ‘eco’

Estimate Copy number for Density data using weighted LS in chunk ‘calib’

```
## FishF TankF Tank Fish TCTSCTFmean TCTSCTFsd
## 1 :4 19:7 Min. :19.00 Min. : 1.00 Min. :12.59 Min. :0.4686
## 2 :4 20:7 1st Qu.:19.75 1st Qu.: 2.00 1st Qu.:16.03 1st Qu.:0.6794
## 4 :4 21:7 Median :20.50 Median : 8.00 Median :17.23 Median :0.9109
## 8 :4 24:7 Mean :21.00 Mean :18.29 Mean :17.38 Mean :1.0507
## 16:4 3rd Qu.:21.75 3rd Qu.:32.00 3rd Qu.:18.96 3rd Qu.:1.1003
## 32:4 Max. :24.00 Max. :65.00 Max. :21.75 Max. :4.0045
## 65:4
## TCTSCTFmed CTSCTFmean CTSCTFsd CTSCTFmed
## Min. :13.38 Min. :28.25 Min. :0.2947 Min. :28.02
## 1st Qu.:16.07 1st Qu.:31.04 1st Qu.:0.6563 1st Qu.:31.02
## Median :17.10 Median :32.77 Median :0.8946 Median :32.90
## Mean :17.49 Mean :32.55 Mean :0.9564 Mean :32.51
## 3rd Qu.:18.98 3rd Qu.:33.97 3rd Qu.:1.1003 3rd Qu.:33.93
## Max. :21.98 Max. :36.52 Max. :2.6110 Max. :36.62
##
## Biomass l2Bio numSCTF
## Min. : 5.69 Min. :2.508 Min. :4.000
## 1st Qu.:11.38 1st Qu.:3.508 1st Qu.:5.000
## Median :43.48 Median :5.442 Median :5.000
## Mean :102.70 Mean :5.482 Mean :4.964
## 3rd Qu.:174.21 3rd Qu.:7.445 3rd Qu.:5.000
## Max. :380.38 Max. :8.571 Max. :5.000
##
##
## Call:
## lm(formula = Xl2what ~ l2Bio, data = eco.SC.TF, weights = (1/Xwhat.se^2))
##
## Weighted Residuals:
## Min 1Q Median 3Q Max
## -8.2694 -0.4669 0.8505 2.1762 8.3466
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
```

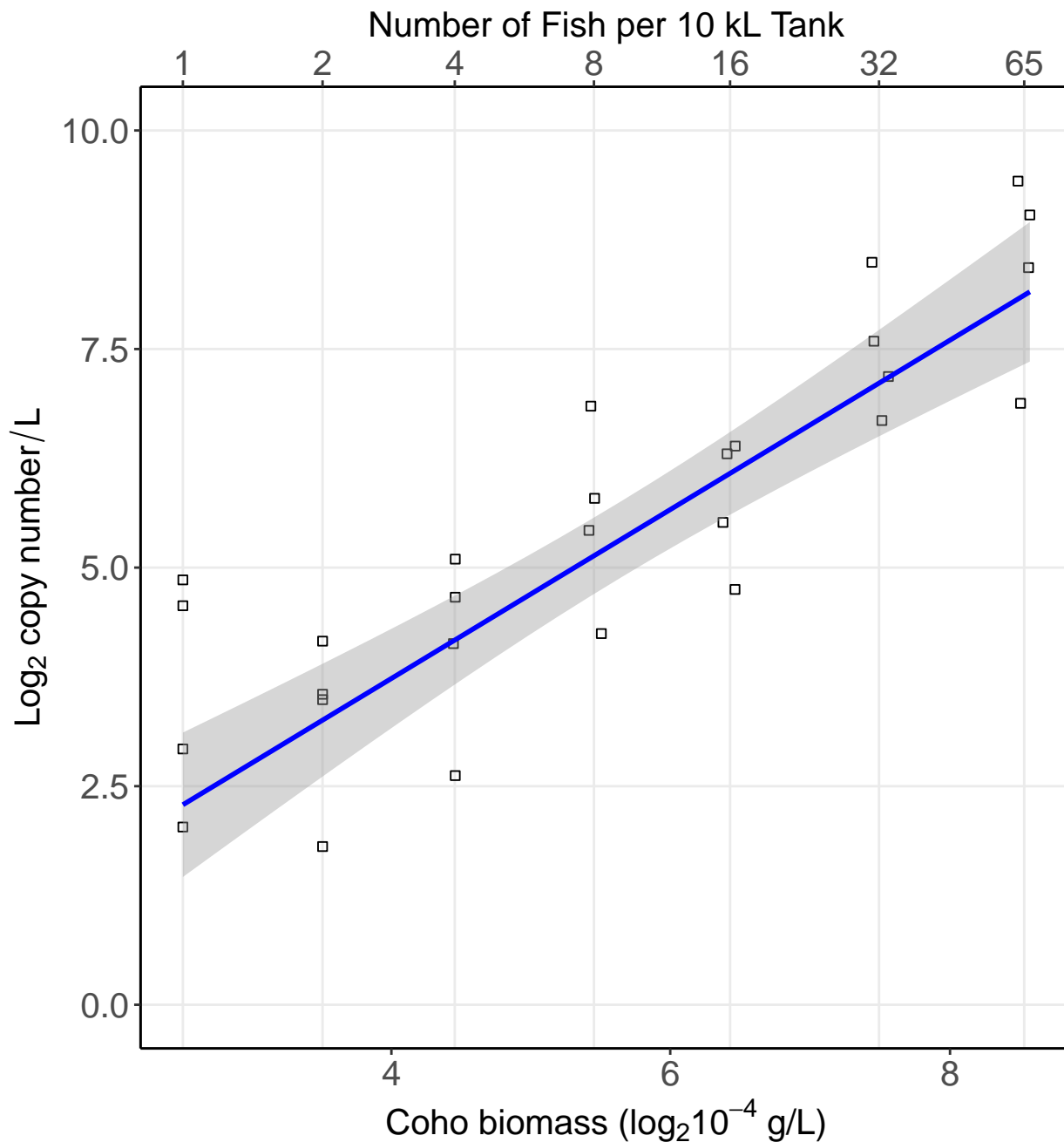
```
## (Intercept) -0.1401      0.6525 -0.215    0.832
## l2Bio      0.9677      0.1100   8.800 2.83e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.292 on 26 degrees of freedom
## Multiple R-squared:  0.7487, Adjusted R-squared:  0.739
## F-statistic: 77.45 on 1 and 26 DF,  p-value: 2.828e-09
```



```
## Model estimates(predict) and actuals by Number of Fish
##
##          1          2          4          8          16          32
## mean log2 Copy 3.595473 3.252432 4.127450 5.577412 5.739707 7.487296
## mean copy     12.087747 9.529707 17.477780 47.749443 53.434774 179.432331
## predict_copy   4.881558 9.547025 18.011992 35.135918 67.549443 138.133428
## predict_log2   2.287342 3.255051 4.170886 5.134875 6.077872 7.109919
```

```
##                               65
## mean log2 Copy    8.441016
## mean copy        347.535470
## predict_copy      277.370960
## predict_log2      8.115673
```

Ggplot density plot



## Dilution/Flow experiment

Load data from file flowFT; estimate Copy number for Flow/Tank Binomial data

Uses CalibS0.table0 generated using no intercept model for eONKI4 in chunk 'MLES0fits0vec' above. Estimate SQ given number of detects in data flow.FT.

```
## Dilution data by Flow and Tank
```

##	FlowN	Tank	TankF	FlowF	TCTFTmean	TCTFTsd	CTFTmean	CTFTsd	numFT	
## 1	10	19	19	10	16.3216250	1.9333940	33.67938	1.9333940	32	
## 2	10	20	20	10	16.8313125	0.8704818	33.16969	0.8704818	32	
## 3	10	21	21	10	15.7397500	2.2366888	34.26125	2.2366888	24	
## 4	10	24	24	10	17.2001667	1.5784773	32.80083	1.5784773	24	
## 5	20	19	19	20	15.7260000	0.7295204	34.27500	0.7295204	16	
## 6	20	20	20	20	15.9968333	1.6171067	34.00417	1.6171067	24	
## 7	20	21	21	20	16.0022500	1.1864427	33.99875	1.1864427	24	
## 8	20	24	24	20	15.9818333	0.8093417	34.01917	0.8093417	24	
## 9	40	19	19	40	5.9260000	6.6034484	37.07273	0.8487649	24	
## 10	40	20	20	40	13.4660000	3.0647775	35.94957	1.1046739	24	
## 11	40	21	21	40	13.6572500	4.4014403	35.10227	1.3571221	24	
## 12	40	24	24	40	8.0685000	6.4241435	37.09200	0.9251888	24	
## 13	80	19	19	80	0.4101667	2.0044991	40.18000	NA	24	
## 14	80	20	20	80	0.6505833	2.2148226	42.20500	1.1950105	24	
## 15	80	21	21	80	1.9897500	4.5475959	38.06750	0.5885788	24	
## 16	80	24	24	80	0.4964167	2.4270361	38.11000	NA	24	
## 17	160	19	19	160	0.0010000	0.0000000	NaN	NA	24	
## 18	160	20	20	160	0.3439167	1.6799417	41.77000	NA	24	
## 19	160	21	21	160	0.0010000	0.0000000	NaN	NA	24	
## 20	160	24	24	160	1.4860000	4.0226673	38.12000	0.9232551	24	
## 21	1000	19	19	1000	0.0010000	0.0000000	NaN	NA	32	
## 22	1000	20	20	1000	0.4614167	2.2555718	38.95000	NA	24	
## 23	1000	21	21	1000	0.8510000	2.3505867	43.20000	1.7741665	32	
## 24	1000	24	24	1000	0.4403750	1.7630928	42.97000	1.9233304	32	
##	ndetects									
## 1	32									
## 2	32									
## 3	24									
## 4	24									
## 5	16									
## 6	24									
## 7	24									
## 8	24									
## 9	11									
## 10	23									
## 11	22									
## 12	15									
## 13	1									
## 14	2									
## 15	4									
## 16	1									
## 17	0									
## 18	1									
## 19	0									
## 20	3									
## 21	0									
## 22	1									
## 23	4									
## 24	2									
##	Data with Flow>=40 and Binomial-Poisson copy number estimates, SQ0									
##	FlowN	Tank	TankF	FlowF	TCTFTmean	TCTFTsd	CTFTmean	CTFTsd	numFT	ndetects



## 9	40	19	19	40	5.9260000	6.603448	37.07273	0.8487649	24	11
## 10	40	20	20	40	13.4660000	3.064777	35.94957	1.1046739	24	23
## 11	40	21	21	40	13.6572500	4.401440	35.10227	1.3571221	24	22
## 12	40	24	24	40	8.0685000	6.424143	37.09200	0.9251888	24	15
## 13	80	19	19	80	0.4101667	2.004499	40.18000	NA	24	1
## 14	80	20	20	80	0.6505833	2.214823	42.20500	1.1950105	24	2
## 15	80	21	21	80	1.9897500	4.547596	38.06750	0.5885788	24	4
## 16	80	24	24	80	0.4964167	2.427036	38.11000	NA	24	1
## 17	160	19	19	160	0.0010000	0.000000	NaN	NA	24	0
## 18	160	20	20	160	0.3439167	1.679942	41.77000	NA	24	1
## 19	160	21	21	160	0.0010000	0.000000	NaN	NA	24	0
## 20	160	24	24	160	1.4860000	4.022667	38.12000	0.9232551	24	3
## 21	1000	19	19	1000	0.0010000	0.000000	NaN	NA	32	0
## 22	1000	20	20	1000	0.4614167	2.255572	38.95000	NA	24	1
## 23	1000	21	21	1000	0.8510000	2.350587	43.20000	1.7741665	32	4
## 24	1000	24	24	1000	0.4403750	1.763093	42.97000	1.9233304	32	2
##		SQ0		SE_SQ0						
## 9		7.030709e-01		0.24988159						
## 10		3.644366e+00		1.30082990						
## 11		2.849440e+00		0.93097640						
## 12		1.124753e+00		0.36396250						
## 13		4.880479e-02		0.04958505						
## 14		9.977936e-02		0.07283111						
## 15		2.090757e-01		0.11126541						
## 16		4.880479e-02		0.04958505						
## 17		6.678866e-10		NaN						
## 18		4.880479e-02		0.04958505						
## 19		6.678866e-10		NaN						
## 20		1.529852e-01		0.09258884						
## 21		6.171701e-09		NaN						
## 22		4.880479e-02		0.04958505						
## 23		1.531264e-01		0.08144377						
## 24		7.399610e-02		0.05400116						

## Estimate Copy number for high copy number Flow/Dilution data

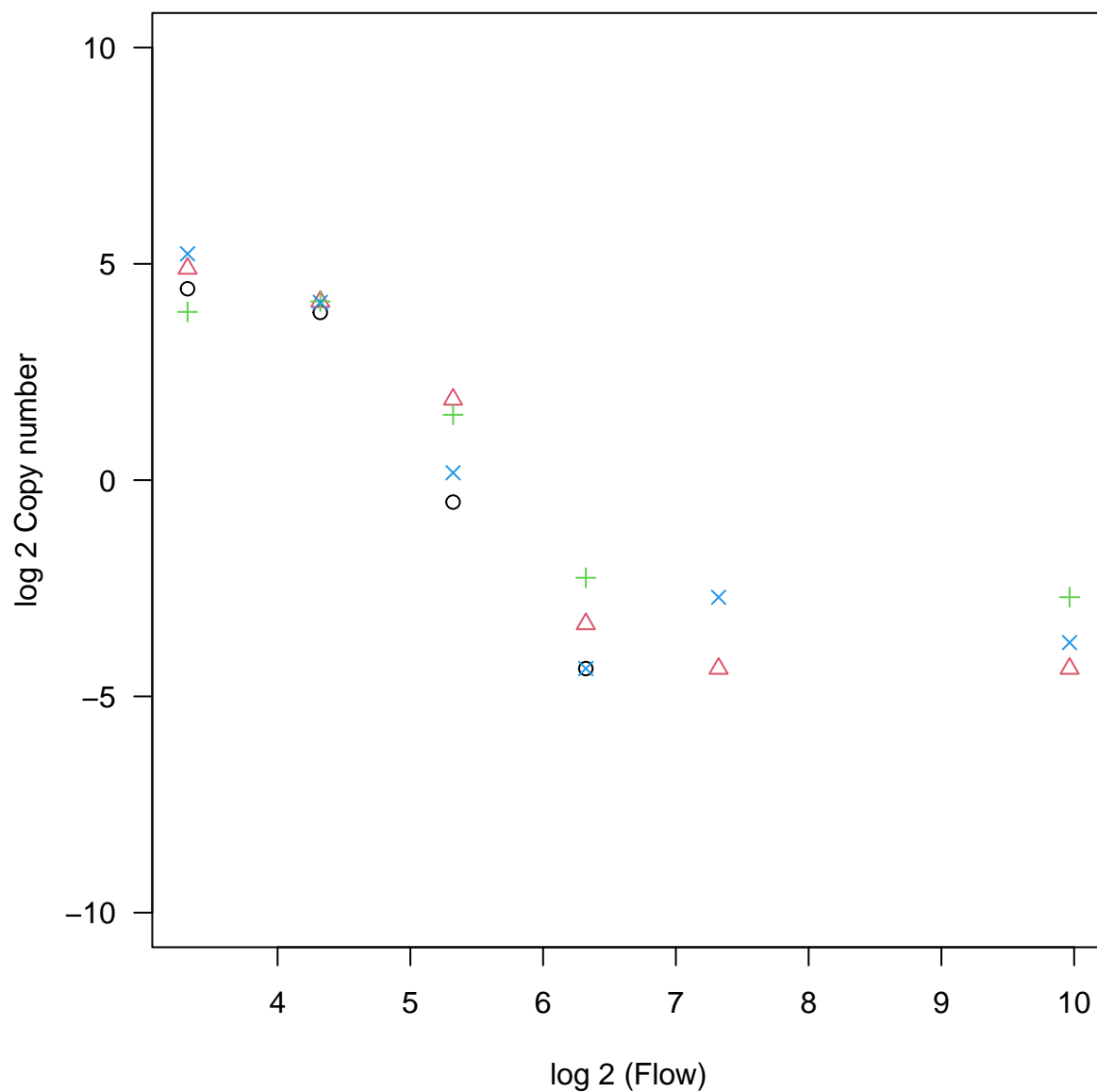
Uses calibration weighted linear regression model from above, chunk 'calib'

## High copy number Dilution data with copy number estimated, SQuest

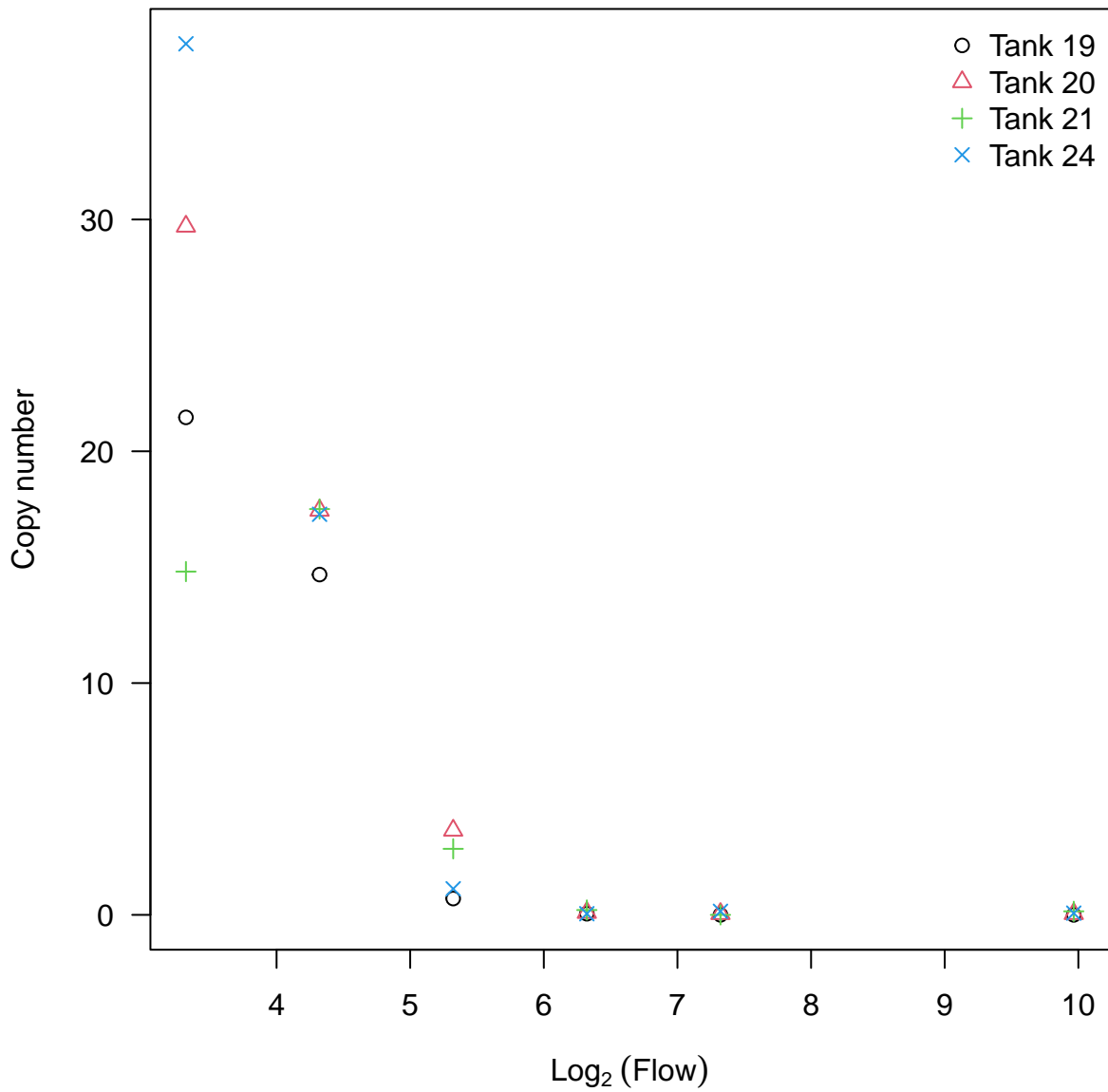
##	FlowN	Tank	TankF	FlowF	TCTFTmean	TCTFTsd	CTFTmean	CTFTsd	numFT	ndetects
## 1	10	19	19	10	16.32163	1.9333940	33.67938	1.9333940	32	32
## 2	10	20	20	10	16.83131	0.8704818	33.16969	0.8704818	32	32
## 3	10	21	21	10	15.73975	2.2366888	34.26125	2.2366888	24	24
## 4	10	24	24	10	17.20017	1.5784773	32.80083	1.5784773	24	24
## 5	20	19	19	20	15.72600	0.7295204	34.27500	0.7295204	16	16
## 6	20	20	20	20	15.99683	1.6171067	34.00417	1.6171067	24	24
## 7	20	21	21	20	16.00225	1.1864427	33.99875	1.1864427	24	24
## 8	20	24	24	20	15.98183	0.8093417	34.01917	0.8093417	24	24
##		SQuest		L10SQuest						
## 1		21.46263		1.331683		0.3431055				
## 2		29.70405		1.472816		0.1578785				
## 3		14.81021		1.170561		0.4571917				
## 4		37.57956		1.574952		0.3237664				

```
## 5 14.68094 1.166754 0.1855052
## 6 17.44809 1.241748 0.3314746
## 7 17.50845 1.243248 0.2444035
## 8 17.28201 1.237594 0.1687485
```

**Log2 Copy Number versus log2 Flow**



## Copy Number versus log2 Flow



### Extra 40Kl Dilution Table

##	Sort.Code	Tank	TankF	FlowN	FlowF	TCTSCmean
##	Min. :41.00	Min. :19.00	19:3	Min. :40	40:12	Min. : 3.160
##	1st Qu.:43.75	1st Qu.:19.75	20:3	1st Qu.:40		1st Qu.: 6.652
##	Median :46.50	Median :20.50	21:3	Median :40		Median :11.105
##	Mean :46.50	Mean :21.00	24:3	Mean :40		Mean :10.279
##	3rd Qu.:49.25	3rd Qu.:21.75		3rd Qu.:40		3rd Qu.:14.066
##	Max. :52.00	Max. :24.00		Max. :40		Max. :16.045
##	TCTSCsd	CTSCmean	CTSCsd	numSC	ndetects	
##	Min. :0.5893	Min. :33.96	Min. :0.5845	Min. :8	Min. :2.000	
##	1st Qu.:1.5242	1st Qu.:35.77	1st Qu.:0.6379	1st Qu.:8	1st Qu.:4.000	
##	Median :5.3869	Median :36.55	Median :0.8216	Median :8	Median :6.500	
##	Mean :4.4004	Mean :36.36	Mean :0.9423	Mean :8	Mean :5.917	
##	3rd Qu.:6.3013	3rd Qu.:36.94	3rd Qu.:1.0995	3rd Qu.:8	3rd Qu.:8.000	

```

## Max.      :7.1248    Max.      :38.05    Max.      :1.5854    Max.      :8    Max.      :8.000
## Use Weighted regression Cq model to estimate copy number Xwhat.40 for 100% detects
## Ignore Xwhat.40 for <100% detects and use Binomial-Poisson model instead.
##      ndetects    CTmean    TCTmean    Xwhat.40
## [1,]          4 36.72250    6.63975    3.083468
## [2,]          3 38.05000    4.48225    1.322677
## [3,]          4 36.69000    6.65600    3.148029
## [4,]          7 36.41571   11.88725    3.749632
## [5,]          8 35.49375   14.50725    6.749649
## [6,]          8 35.99750   14.00350    4.895448
## [7,]          8 35.74625   14.25475    5.745975
## [8,]          8 33.95625   16.04475   17.989373
## [9,]          6 35.77167   10.67225    5.653610
## [10,]         7 36.81571   11.53725    2.905550
## [11,]         2 37.36500    3.15975    2.047061
## [12,]         6 37.32333    9.50850    2.102172

```

## Background estimation

### Background estimate from Density experiment negative controls

```

## Density background estimate:    0.06109982
## Density negative controls, ndetect, nreps
##      nd numSC
##      22   424
## Mean of individual Density background estimates:    0.06940534

```

### Background estimate from Dilution data

Estimate of background using flow negative controls.

##	Sort.Code	Tank	TankF	FlowN	FlowF	TCTSCmean	TCTSCsd	CTSCmean	CTSCsd	numSC
## 1	113	19	19	0	0	1.43600	4.058793	38.520	NA	8
## 2	114	19	19	0	0	1.67100	4.723473	36.640	NA	8
## 3	115	19	19	0	0	0.00100	0.000000	NaN	NA	8
## 4	116	20	20	0	0	0.00100	0.000000	NaN	NA	8
## 5	117	20	20	0	0	2.96975	5.497343	38.125	0.1484924	8
## 6	118	20	20	0	0	0.00100	0.000000	NaN	NA	8
## 7	119	21	21	0	0	0.00100	0.000000	NaN	NA	8
## 8	120	21	21	0	0	0.00100	0.000000	NaN	NA	8
## 9	121	21	21	0	0	2.42475	4.494498	40.305	0.6434672	8
## 10	122	24	24	0	0	0.00100	0.000000	NaN	NA	8
## 11	123	24	24	0	0	0.00100	0.000000	NaN	NA	8
## 12	124	24	24	0	0	0.00100	0.000000	NaN	NA	8

```

##      ndetects
## 1           1
## 2           1
## 3           0
## 4           0
## 5           2
## 6           0

```

```
## 7      0
## 8      0
## 9      2
## 10     0
## 11     0
## 12     0

## Mean background of individual flow negative controls: 0.08050365
```

## Background estimate from Density and Dilution data

```
## Background copy number estimate: 0.06347212
## ndetects: 28 nreps: 520
```

## BACKGROUND Limits of Detection using estimated background SQBboth, No intercept model

For the ordinary Limit of detection, Lc (or LOB) is the critical value for the test of the null hypothesis that  $SQ=0$ . Here the null hypothesis is that  $SQB=S\_Background$  (using SQBboth) and we compute a critical value LOB-B, Limit of Background Blank. We also compute LOD-B, Limit of Background Detection.

```
## Background estimate and confidence interval
##
## 0.06347212 ( 0.0468951 0.09817675 )
```

## Print out BACKGROUND limits of detection with confidence intervals, by number of replicates

```
##          3      8      16      24      32      40      48      64      96
## LOB-B      1.00 2.00 3.00 3.00 4.00 5.00 5.00 7.00 9.00
## LOB-Bcn    0.46 0.33 0.24 0.15 0.15 0.15 0.13 0.13 0.11
## LOD-B      2.29 1.05 0.62 0.40 0.35 0.32 0.27 0.25 0.20
## LOD-B_lower 1.69 0.78 0.46 0.29 0.26 0.24 0.20 0.18 0.15
## LOD-B_upper 3.55 1.62 0.96 0.61 0.54 0.50 0.41 0.39 0.30
```

## Fit Bent Cable model to flow data mean over (Flow/Tank) copy numbers

```
## Breakpoint alphahat: 7.214288 , 2^alphahat: 148.4968

##
## Call:
## stats::lm(formula = y ~ x + q)
##
## Residuals:
##          10          20          40          80          160          1000
## -9.829e-01  1.086e+00  7.770e-01 -8.800e-01  6.813e-05 -6.560e-06
##
## Coefficients:
##          Estimate Std. Error t value Pr(>|t|)
## (Intercept)  14.6419     2.0144   7.269  0.00537 **
## x           -2.6986     0.3759  -7.179  0.00557 **
## q             3.0509     0.8033   3.798  0.03205 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

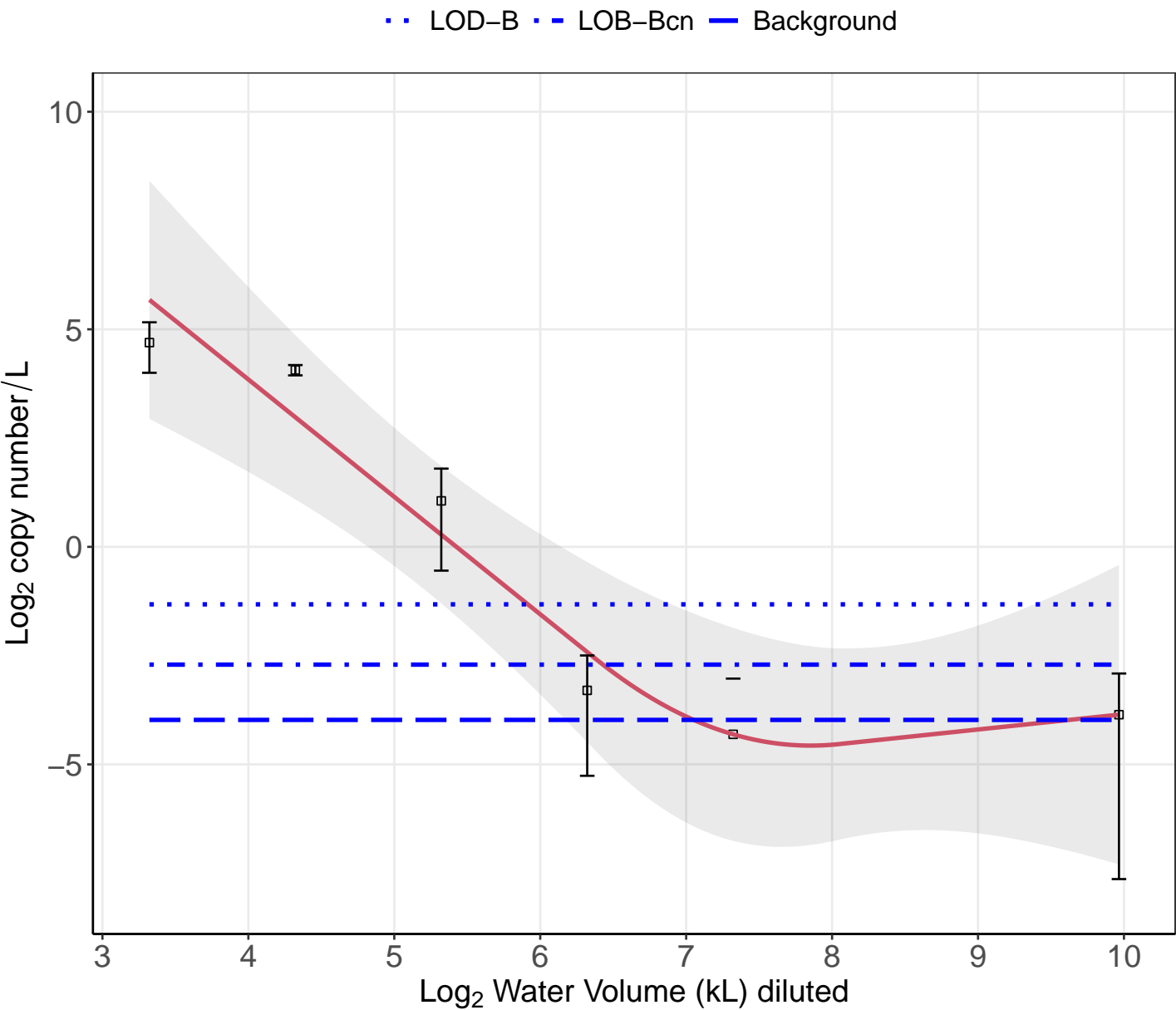
```
##
## Residual standard error: 1.084 on 3 degrees of freedom
## Multiple R-squared:  0.9578, Adjusted R-squared:  0.9297
## F-statistic: 34.07 on 2 and 3 DF,  p-value: 0.00866
```

### Intersections with Bent cable model

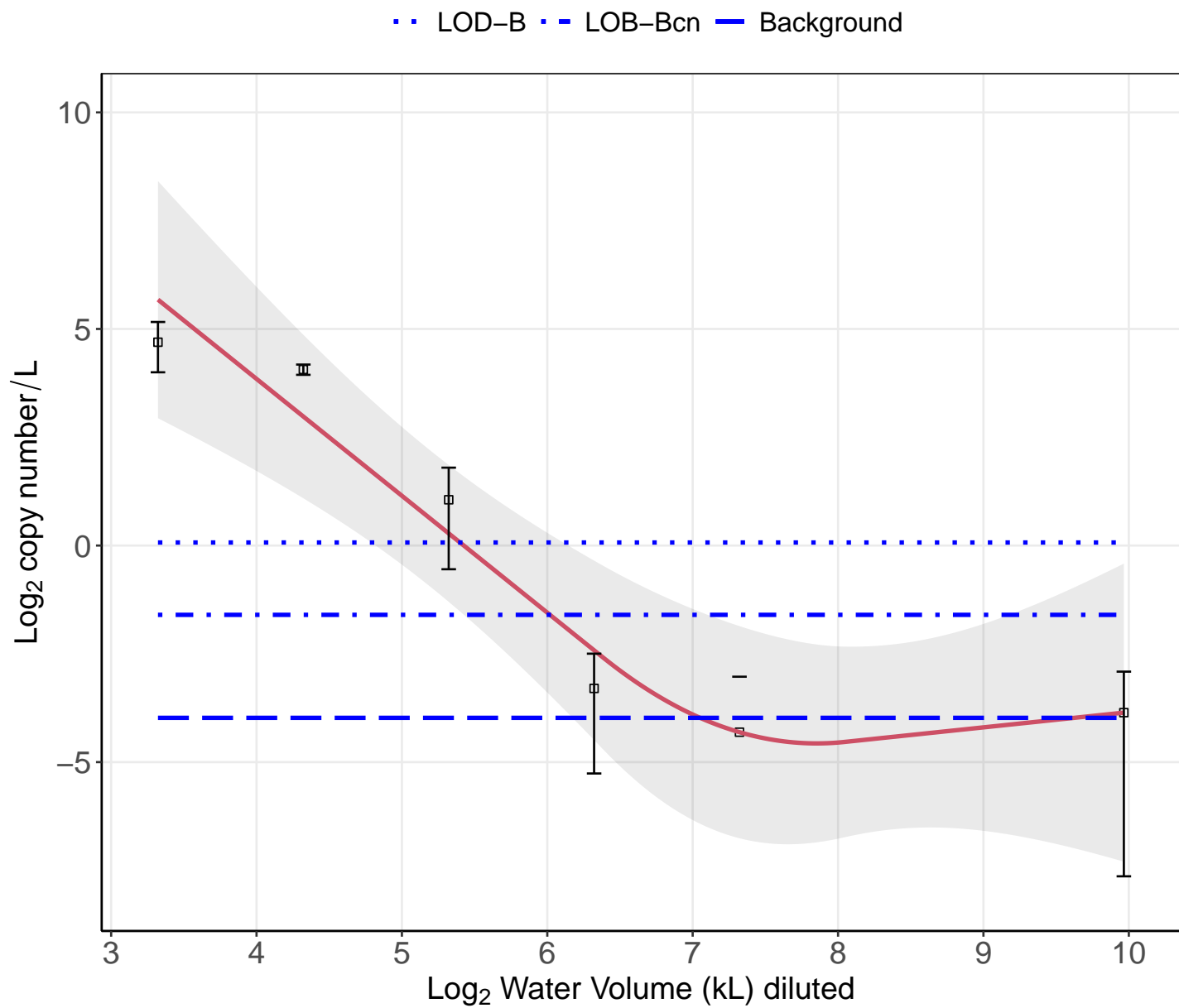
```
## [1] "Bent cable model intersections"
```

##	Log2copy	Log2Vol	Copy	VolKL
## Background	-3.97773327	7.061184	0.06347212	133.54516
## LODB_8	0.07019178	5.391873	1.04985623	41.98707
## Lcx_8	-1.59991565	6.026211	0.32989627	65.17340
## LODB_24	-1.33488234	5.926053	0.39642440	60.80224
## Lcx_24	-2.70721201	6.426846	0.15312566	86.03464
## LODB_40	-1.63342071	6.026211	0.32232305	65.17340
## Lcx_40	-2.70721201	6.426846	0.15312566	86.03464

GGplot log2 Flow/Dilution plot using mean copy numbers



## 24 technical replicates



## 8 technical replicates