Script to analyze behavior data for *C. elegans* manuscript

Emily Jane Dennis

7/20/2017

Purpose:

To plot all behavior data for manuscript.

Background:

This is an R Markdown document. http://rmarkdown.rstudio.com.

The code required to get these graphs is in the gray boxes, along with information on how to run the statistical tests reported in the paper and qqPlots that were used to determine if these statistical tests were generally appropriate.

This was run in RStudio, using R version 3.3.1 (2016-10-31), Sincere Pumpkin Patch on a x86_64-apple-darwin13.4.0

Before these data can be analyzed, I used the following code block to:

1. Load all libraries required for analysis and plotting
2. Load these data
3. Subset data to use later to plot

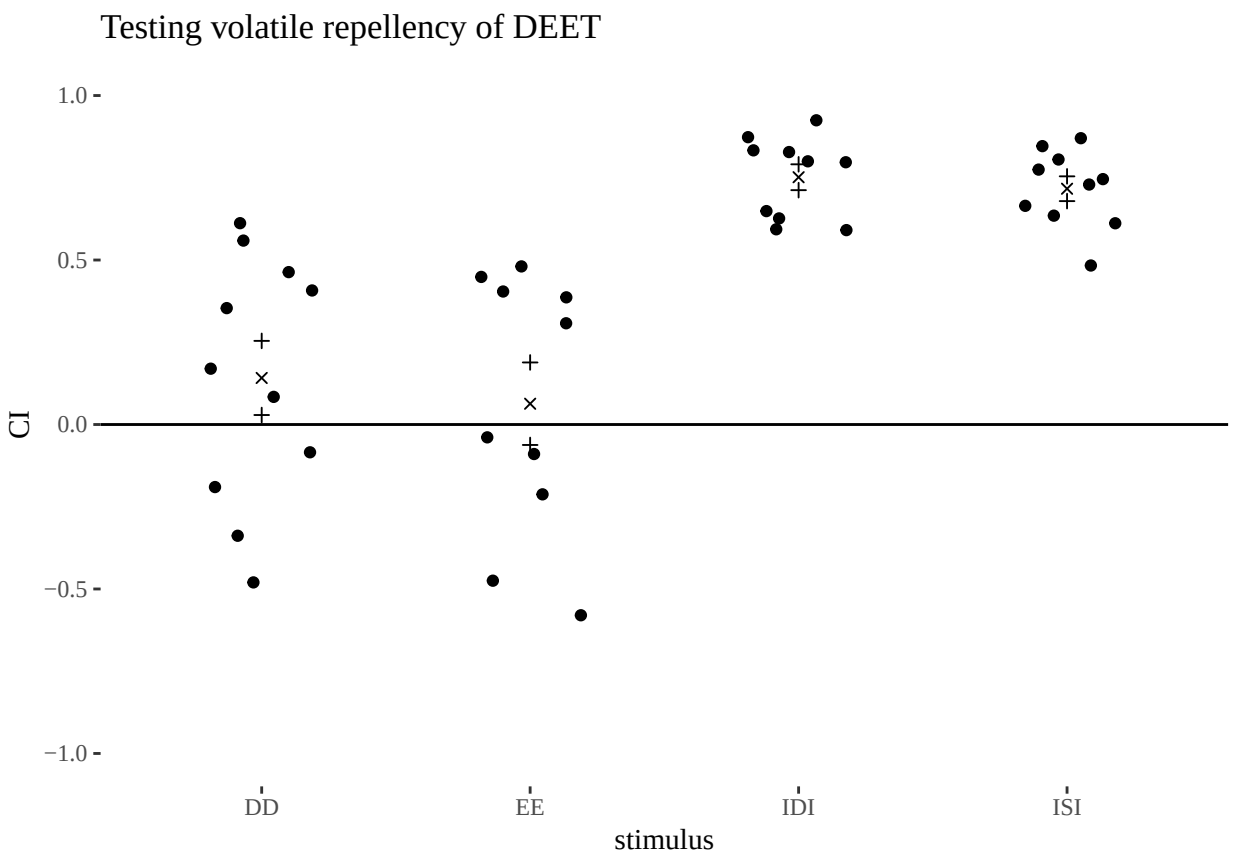
```
#load libraries
library(plyr)
library(dplyr)
library(ggplot2)
library(ggthemes)
library(multcompView)
library(car)

allbeh <- read.csv('Dennis_DEETandCelegans_behaviordata.csv')


#start subsetting data:
hawaiian <- subset(allbeh,expt=='hawaiian')
#hawaiiand didn't have any positive controls that
# had enough responding animals to record
#so I excluded the day. temp in lab was also low that day.
hawaiian <- subset(hawaiian,date!=20160920)

hawaiiane <- subset(hawaiian,percentDEET==0)
hawaiiand <- subset(hawaiian,percentDEET==0.15)

rescue <- subset(allbeh,expt=='rescue'|expt=='rescue2')
rescueE <- subset(rescue,percentDEET==0)
rescueD <- subset(rescue,percentDEET==0.15)
```

```
spots <- subset(allbeh,expt=='spots')
screen <- subset(allbeh,expt=='screen2')
```

First, I tested if DEET is a volatile repellent (1uL 50% DEET in two spots, just as a volatile odorant). Next, I tested if DEET can inhibit attraction to isoamyl alochol if presented as a point stimulus (2uL 50% DEET or solvent was added between each of the two spots of isoamyl alochol (IAA).) Both are plotted below:

DD = 2 x 1uL DEET as point stimulus EE = 2 x 1uL solvent (ethanol) as point stimulus IDI = 2 x 1uL spots IAA as point stimulus with 2uL 50% DEET between IEI = 2 x 1uL spots IAA as point stimulus with 2uL solvent/ethanol between ## Fig 1 b-c

```
# make summary data
spotssum <- ddply(spots,c('stimulus'),summarise,N=length(stimulus),mean=mean(CI),sd=sd(CI),se=sd/sqrt(N

# plot each point, xs are means, +s are standard error
ggplot(spots,aes(x=stimulus,y=CI)) +
  ggtitle('Testing volatile repellency of DEET') +
  geom_jitter(width=0.2,height=0) +
  geom_point(data=spotssum,aes(x=stimulus,y=mep),shape=3) +
  geom_point(data=spotssum,aes(x=stimulus,y=mem),shape=3)  +
  geom_point(data=spotssum,aes(x=stimulus,y=mean),shape=4) +
  theme_tufte() +
  ylim(-1,1) +
  geom_hline(yintercept=0)
```



DEET does not look like a volatile point stimulus: p = 0.6478

```
EE <- subset(spots,stimulus=="EE")
DD <- subset(spots,stimulus=="DD")
t.test(x=EE$CI,y=DD$CI,"two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  EE$CI and DD$CI
## t = -0.4643, df = 18.527, p-value = 0.6478
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.4318028  0.2752306
## sample estimates:
##  mean of x  mean of y
## 0.06318329 0.14146935
```

DEET does not interrupt chemotaxis to isoamyl alochol when presented as a point stimulus: p = 0.5286

```
IDI <- subset(spots,stimulus=="IDI")
ISI <- subset(spots,stimulus=="ISI")

t.test(x=ISI$CI,y=IDI$CI,"two.sided")
```

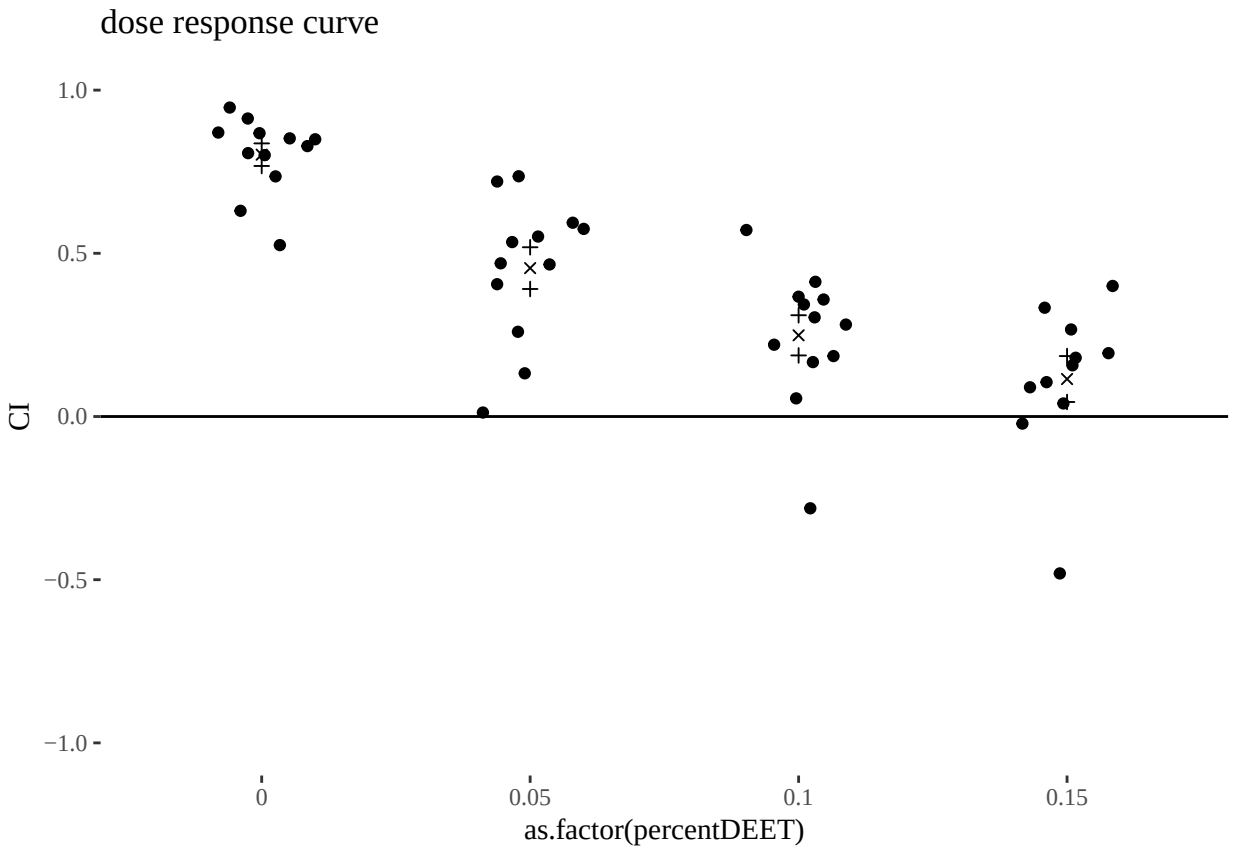
```
##
##  Welch Two Sample t-test
##
## data:  ISI$CI and IDI$CI
## t = -0.64263, df = 17.967, p-value = 0.5286
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.14927950  0.07935372
## sample estimates:
## mean of x mean of y
## 0.7166651 0.7516280
```

Dose-response-like curve for effect of DEET in agarose on IAA chemotaxis

Fig 1e

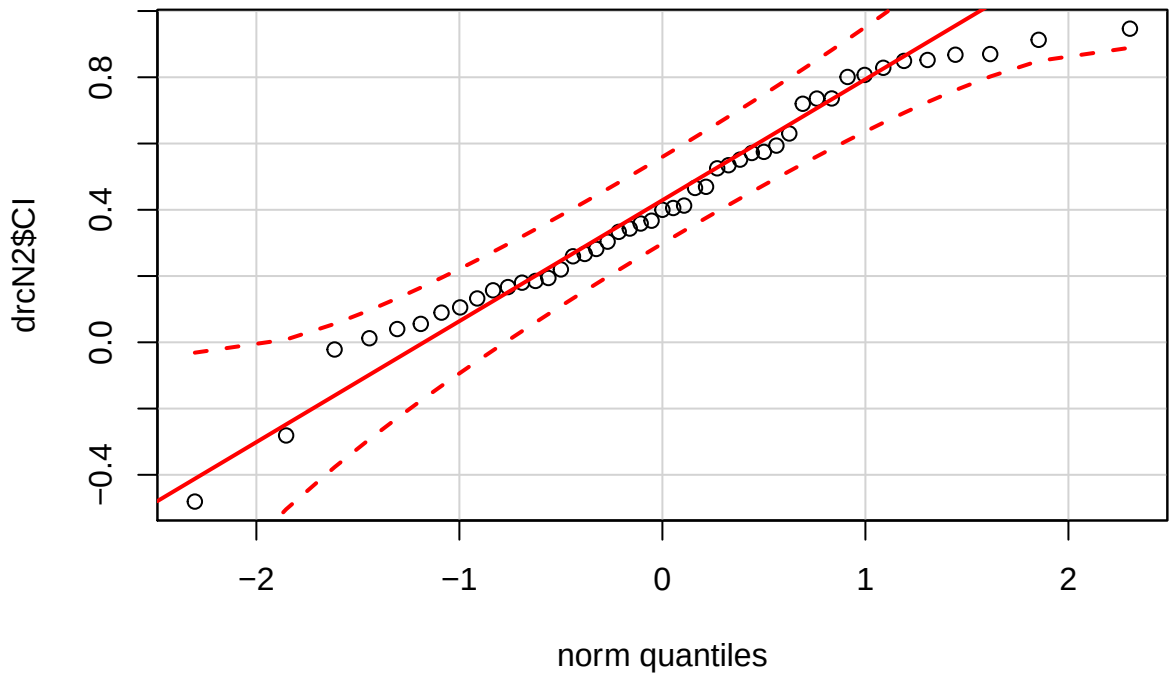
```
drc <- subset(allbeh,expt=='drc')
drc <- subset(drc,percentDEET<0.2)
drcN2 <- subset(drc,genotype=='N2')

drcN2summary <- ddply(drcN2,c("percentDEET"),summarise,N=length(percentDEET),mean=mean(CI),sd=sd(CI),se=
drcN2summary$mep <- drcN2summary$mean + drcN2summary$se
drcN2summary$mem <- drcN2summary$mean - drcN2summary$se
```

```
ggplot(drcN2,aes(x=as.factor(percentDEET),y=CI)) +
  ggtitle('dose response curve') +
  geom_jitter(width=0.2,height=0) +
  theme_tufte() +
  ylim(-1,1) +
  geom_hline(yintercept=0) +
  geom_point(data=drcN2summary,aes(x=as.factor(percentDEET),y=mep),shape=3)+ geom_point(data=drcN2summa
geom_point(data=drcN2summary,aes(x=as.factor(percentDEET),y=mem),shape=3)
```



Below find the stats for these dose response data

```
#plot QQ plot for model used in 2 way anova, if looks like a ~ straight line the residuals are probably
qqPlot(drcN2$CI)
```




```
tukeydrc <- TukeyHSD(aov(formula = CI ~ as.factor(percentDEET), data = drcN2))

Tukeydrc.levels <- tukeydrc$`as.factor(percentDEET)`[,4]
multcompLetters(Tukeydrc.levels)['Letters']
```

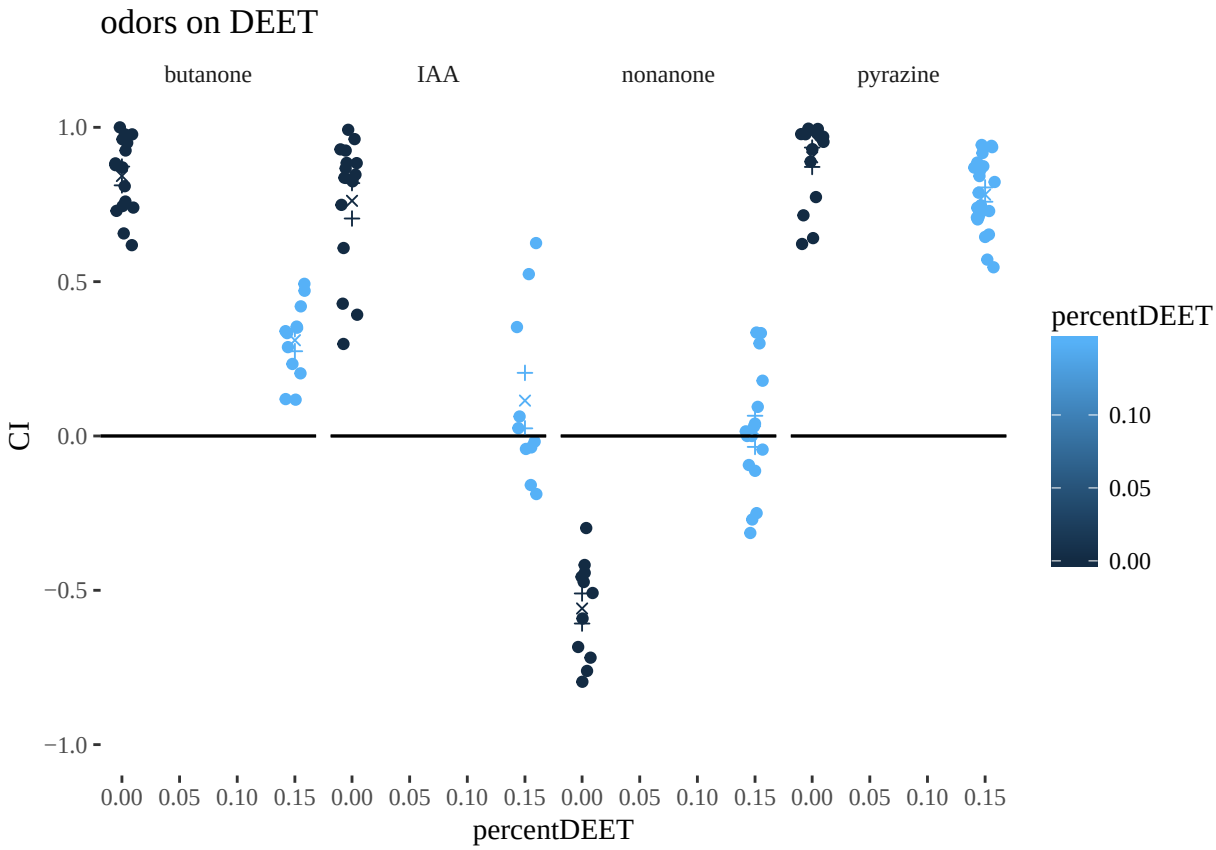
```
## $Letters
## 0.05  0.1 0.15    0
##  "a" "ab"  "b"  "c"
```

What other odors are affected by DEET

Fig 1f (not in same order)

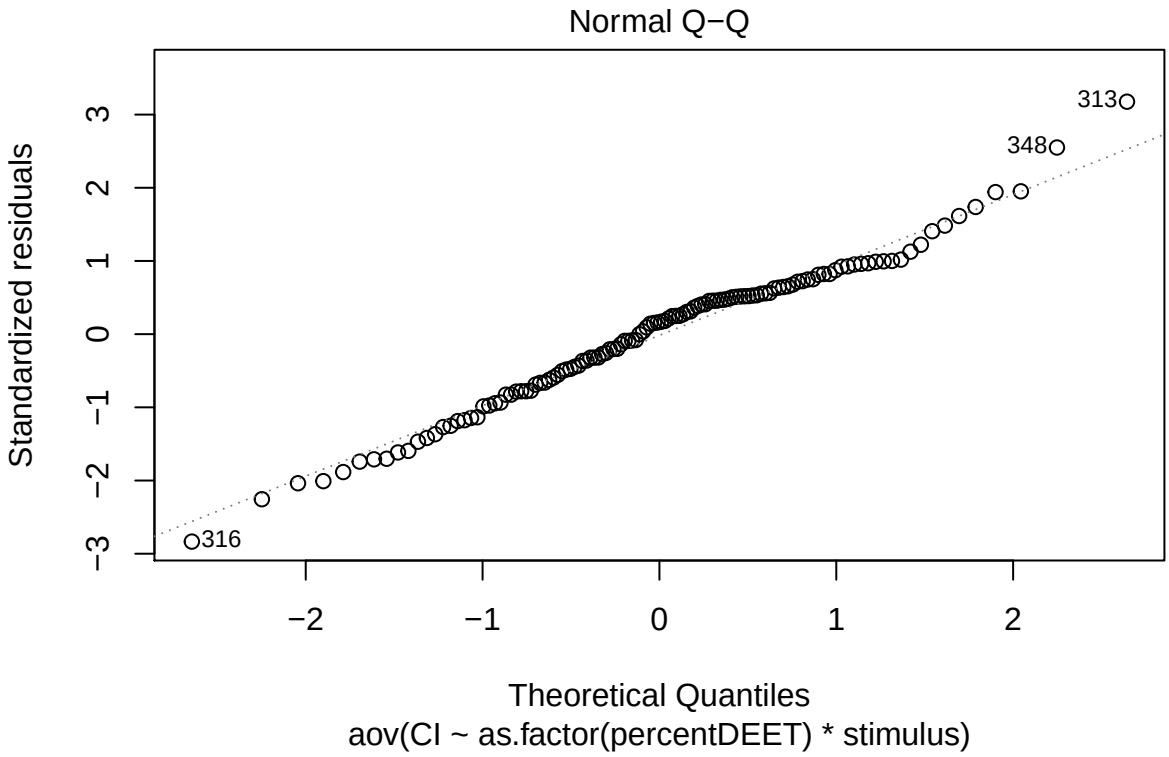
```
odor <- subset(allbeh,expt=='odor')
odor <- subset(odor,stimulus!='diacetyl')

odorsum <- ddply(odor,c("stimulus","percentDEET"),summarise,N=length(stimulus),mean=mean(CI),sd=sd(CI),s
odorsum$mep <- odorsum$mean + odorsum$se
odorsum$mem <- odorsum$mean - odorsum$se



ggplot(odor,aes(x=percentDEET,y=CI,color=percentDEET)) +
  ggtitle('odors on DEET') +
  geom_jitter(width=0.01,height=0) +
  theme_tufte() +
  ylim(-1,1) +
  geom_hline(yintercept=0) +
  geom_point(data=odorsum,aes(x=percentDEET,y=mep),shape=3) +
  geom_point(data=odorsum,aes(x=percentDEET,y=mem),shape=3)  +
  geom_point(data=odorsum,aes(x=percentDEET,y=mean),shape=4) +
```

```
  theme_tufte() +
  geom_hline(yintercept=0) +
  facet_grid(~stimulus)
```



statistics for these data:

```
#plot QQ plot for model used in 2 way anova, if looks like a ~ straight line the residuals are probably
odors.aov = aov(CI~as.factor(percentDEET)*stimulus,data=odor)
plot(odors.aov,which=2)
```




```
tukeyodor <- TukeyHSD(aov(formula=CI~as.factor(percentDEET)*as.factor(stimulus),data=odor))

Tukey.levels <- tukeyodor$`as.factor(percentDEET):as.factor(stimulus)`[,4]
multcompLetters(Tukey.levels)['Letters']
```

```
## $Letters
## 0.15:butanone         0:IAA      0.15:IAA    0:nonanone 0.15:nonanone
##           "a"           "b"          "ac"           "d"           "c"
##    0:pyrazine 0.15:pyrazine    0:butanone
##           "b"           "b"           "b"
```

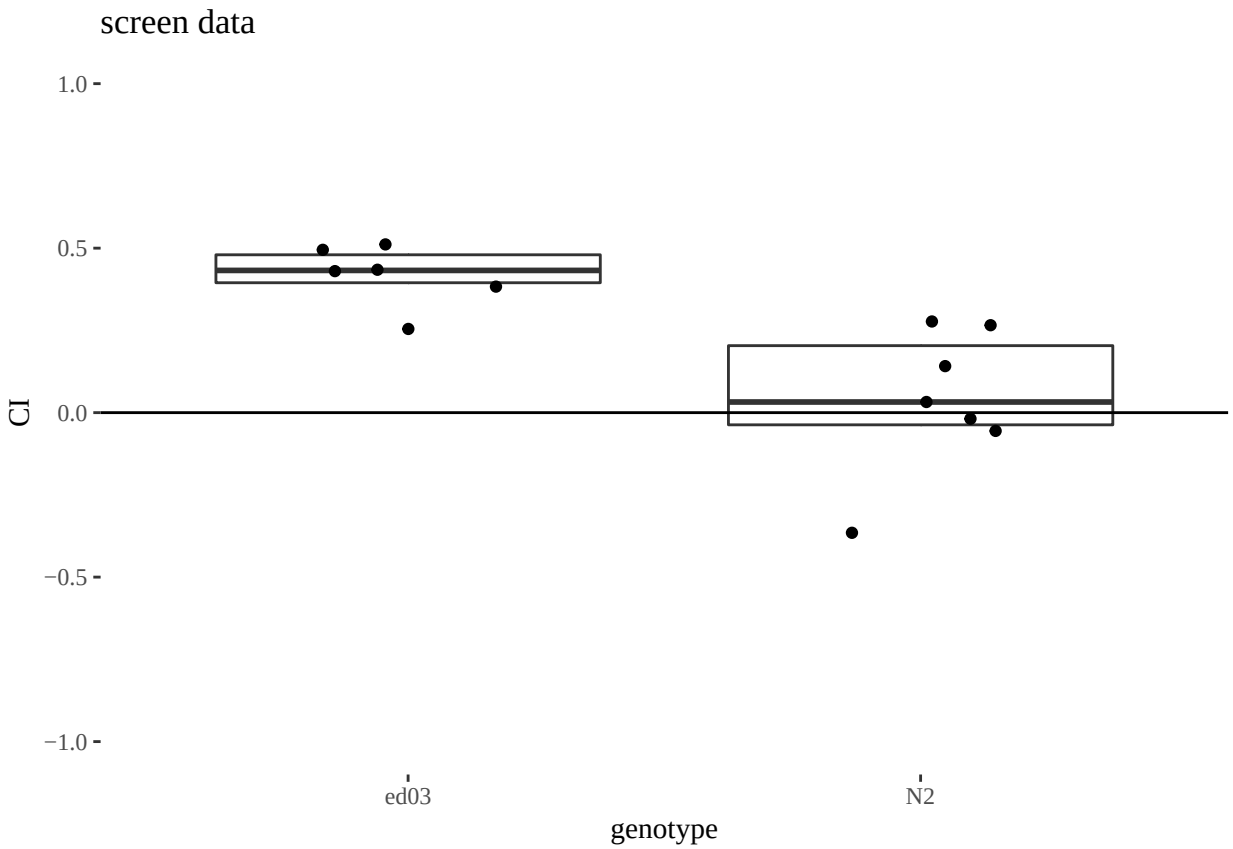
Forward genetic screen

I completed a forward genetic screen with help from Wendy Wang in Shai Shaham's lab (and lots of guidance from Shai on both the screen and follow up) plot of genetic screen data:

Figure 1h

ed03 == LBV003 in paper ed03 was our in-house name as it was the third mutant I (___E___mily ___D___ennis) found

```
ggplot(screen,aes(x=genotype,y=CI)) + ggtitle('screen data') + geom_boxplot(outlier.shape=NA,coef=0) + g
```



EMS-isolated strain LBV003 is DEET-resistant compared to N2 (p-value = 0.003101)

note that in the spreadsheet, all screen and rescue experiments

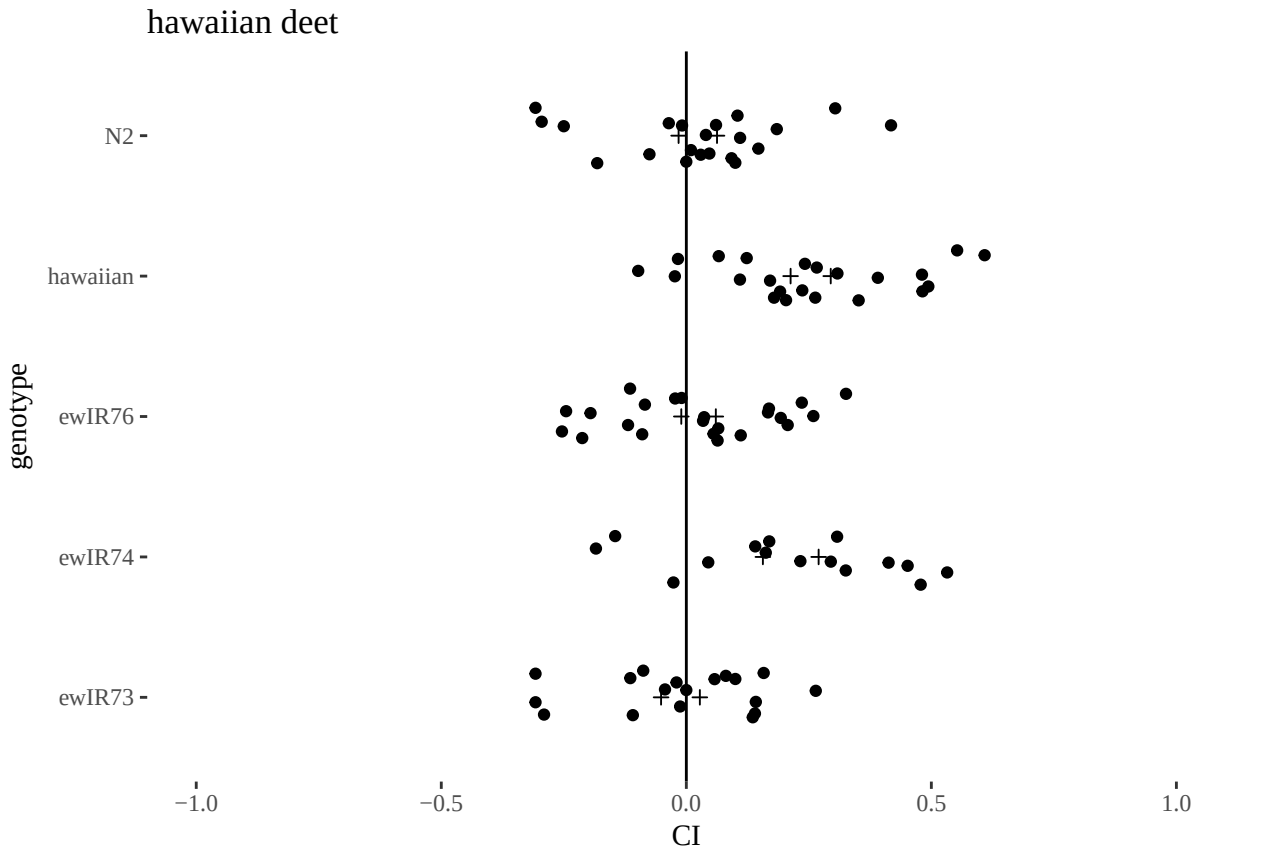
```
N2screen <- subset(screen,genotype=="N2")
ed03screen <- subset(screen,genotype=="ed03")
t.test(x=N2screen$CI,y=ed03screen$CI,"two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  N2screen$CI and ed03screen$CI
## t = -4.1191, df = 8.2912, p-value = 0.003101
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.5890473 -0.1678773
## sample estimates:
##  mean of x  mean of y
## 0.03958319 0.41804553
```

Figure 2: *str-217* mutants are DEET-resistant

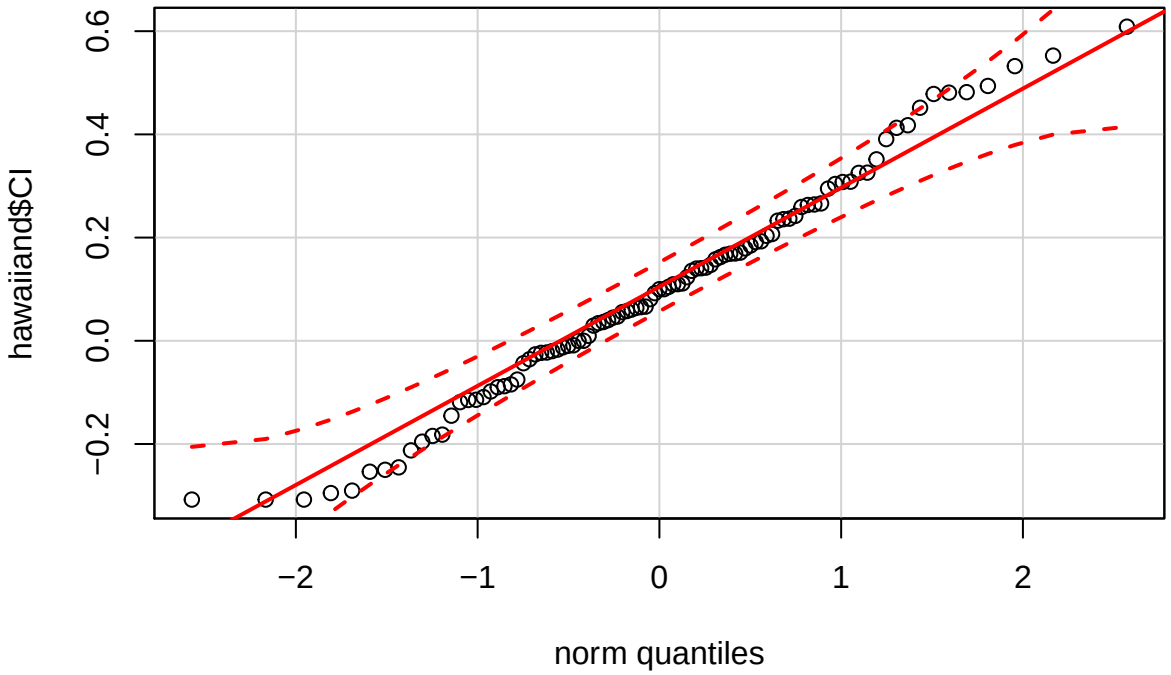
Figure 2b: a wild-isolate is naturally DEET-resistant

```
# make summary data (means, standard errors) to plot
hdsummary <-ddply(hawaiiand,c("genotype"),summarise,N=length(genotype),mean=mean(CI),sd=sd(CI),se=sd/sq
hdsummary$mep <- hdsummary$mean + hdsummary$se
hdsummary$mem <- hdsummary$mean - hdsummary$se

#plot hawaiian strains chemotaxing to IAA on DEET agar
ggplot(hawaiiand,aes(x=genotype,y=CI)) +
  ggtitle('hawaiian deet') +
  geom_point(data=hdsummary,aes(x=genotype,y=mep),shape=3)  +
  geom_point(data=hdsummary,aes(x=genotype,y=mem),shape=3)  +
  geom_jitter(width=0.2,height=0) + theme_tufte() + ylim(-1,1) +
  geom_hline(yintercept=0) +
  coord_flip()
```



First I did a qqPlot for these data to test if residuals are roughly normal and an ANOVA is appropriate. Dashed lines on the plot indicate 95% confidence interval.

```
qqPlot(hawaiiand$CI)
```

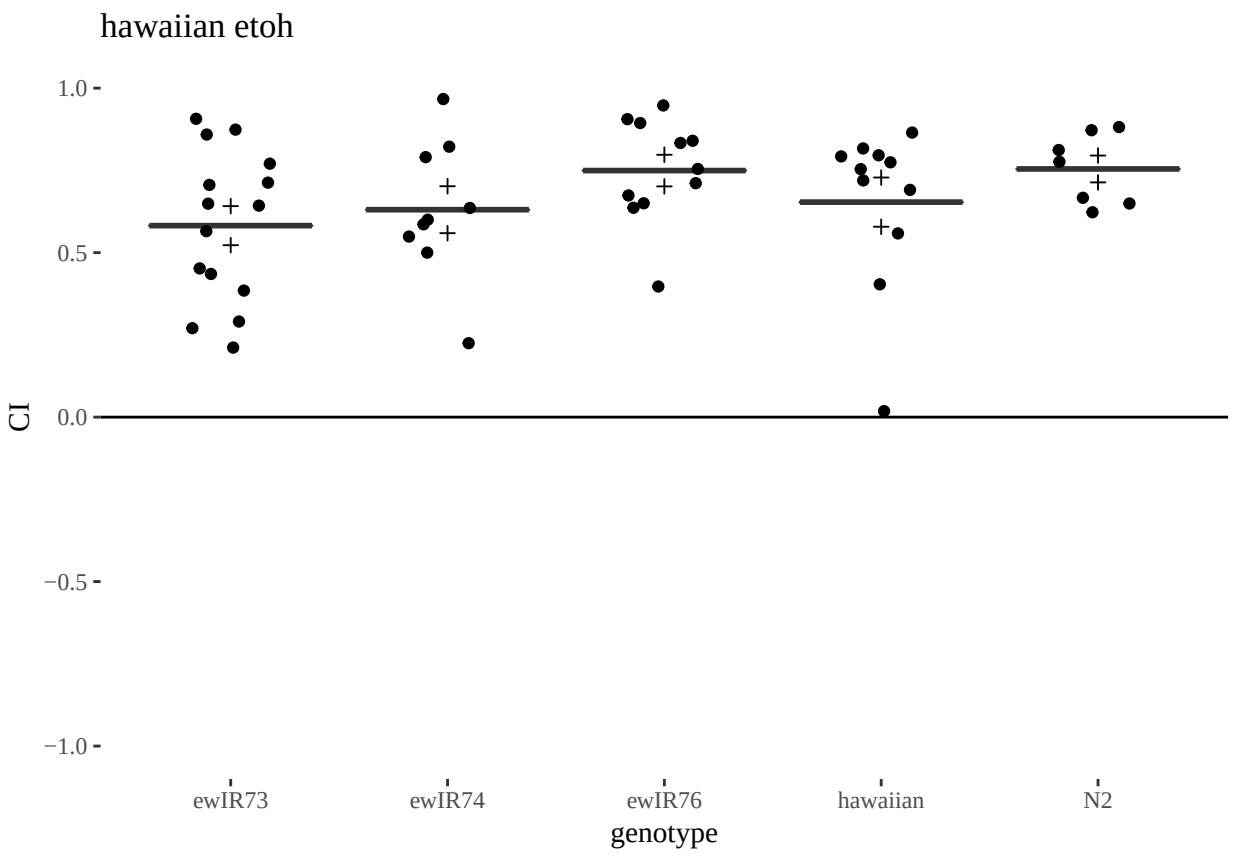


I then ran an ANOVA on these data

```
tukeyhawaiian <- TukeyHSD(aov(formula = CI ~ genotype, data = hawaiiand))
hawaiian.levels <- tukeyhawaiian$genotype[,4]
multcompLetters(hawaiian.levels)['Letters']
```

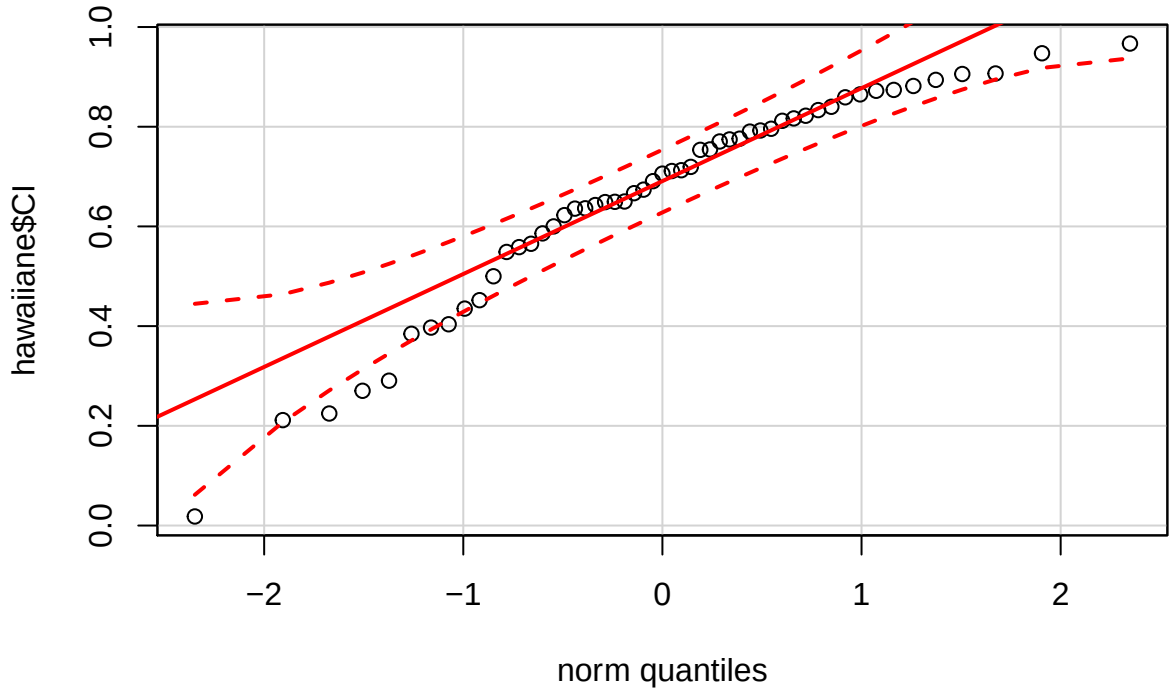
```
## $Letters
##   ewIR74   ewIR76 hawaiian       N2   ewIR73
##      "a"      "b"      "a"      "b"      "b"
```

```
# make summary data for hawaiiane

hawaiiane <- subset(hawaiiane,genotype!='ewIR69')
hesummary <-ddply(hawaiiane,c("genotype"),summarise,N=length(genotype),mean=mean(CI),sd=sd(CI),se=sd/sq
hesummary$mep <- hesummary$mean + hesummary$se
hesummary$mem <- hesummary$mean - hesummary$se

ggplot(hawaiiane,aes(x=genotype,y=CI)) +
  ggtitle('hawaiian etoh') +
  geom_boxplot(data=hesummary,aes(x=genotype,y=mean)) +
  geom_point(data=hesummary,aes(x=genotype,y=mep),shape=3)  +
  geom_point(data=hesummary,aes(x=genotype,y=mem),shape=3)  +
  geom_jitter(width=0.2,height=0) +
  theme_tufte() +
  ylim(-1,1) +
  geom_hline(yintercept=0)
```



Below find the stats for these hawaiian ethanol plate data

```
qqPlot(hawaiiane$CI)
```




```
tukeyhawaiiane <- TukeyHSD(aov(formula=CI~genotype,data=hawaiiane))
Tukeyhe.levels <- tukeyhawaiiane$`genotype`[,4]
```

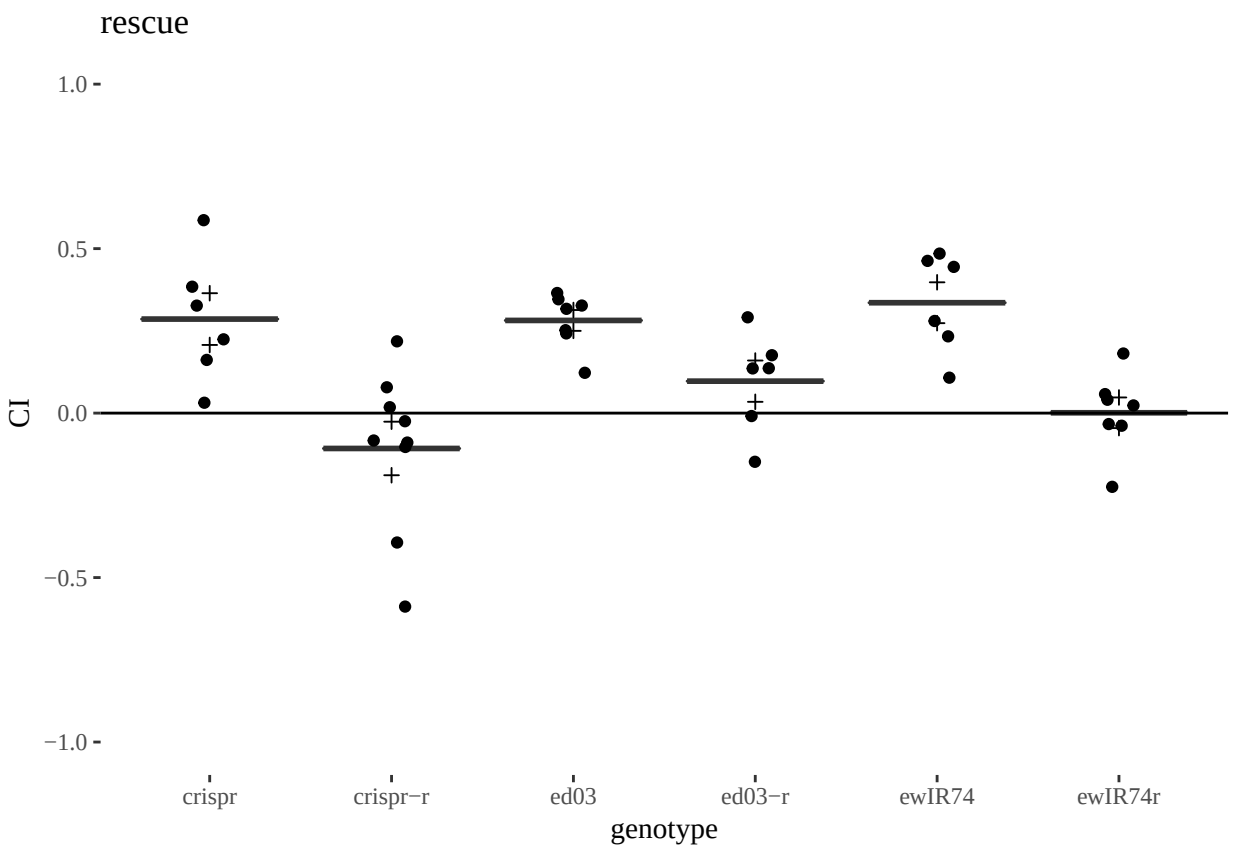
```
multcompLetters(Tukeyhe.levels)['Letters']
```

```
## $Letters
##   ewIR74   ewIR76 hawaiian       N2   ewIR73
##      "a"      "a"      "a"      "a"      "a"
```

Fig 2d-f

Animals expressing the rescue construct return to being DEET sensitive, indicating *str-217* is required in each of these strains to confer DEET-sensitivity

```
# make summary data for rescue

rescue2 <- subset(allbeh,expt=='rescue2')
rescue <- subset(allbeh,expt=='rescue')
rescueall <- rbind(rescue2,rescue)
rescueD <- subset(rescueall,percentDEET==0.15)

rescueDsummary <-ddply(rescueD,c("genotype"),summarise,N=length(genotype),mean=mean(CI),sd=sd(CI),se=sd/
rescueDsummary$mep <- rescueDsummary$mean + rescueDsummary$se
rescueDsummary$mem <- rescueDsummary$mean - rescueDsummary$se

ggplot(rescueD,aes(x=genotype,y=CI)) +
  ggtitle('rescue') +
  geom_boxplot(data=rescueDsummary,aes(x=genotype,y=mean)) +
  geom_point(data=rescueDsummary,aes(x=genotype,y=mep),shape=3)  +
  geom_point(data=rescueDsummary,aes(x=genotype,y=mem),shape=3)  +
  geom_jitter(width=0.1,height=0) +
  theme_tufte() +
  ylim(-1,1) +
  geom_hline(yintercept=0)
```




```
crisprnoR <- subset(rescueD,genotype=='crispr')
crisprR <- subset(rescueD,genotype=='crispr-r')
ed03noR <- subset(rescueD,genotype=='ed03')
ed03R <- subset(rescueD,genotype=='ed03-r')
ewir74noR <- subset(rescueD,genotype=='ewIR74')
ewir74R <- subset(rescueD,genotype=='ewIR74r')

t.test(crisprnoR$CI,crisprR$CI)
```

```
##
##  Welch Two Sample t-test
##
## data:  crisprnoR$CI and crisprR$CI
## t = 3.4747, df = 12.49, p-value = 0.004341
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1477553 0.6388575
## sample estimates:
##  mean of x  mean of y
##  0.2857813 -0.1075251
```

```
t.test(ed03noR$CI,ed03R$CI)
```

```
##
##  Welch Two Sample t-test
##
## data:  ed03noR$CI and ed03R$CI
## t = 2.6252, df = 7.4678, p-value = 0.03226
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.02040963 0.34879833
## sample estimates:
##  mean of x  mean of y
## 0.28164730 0.09704332
```

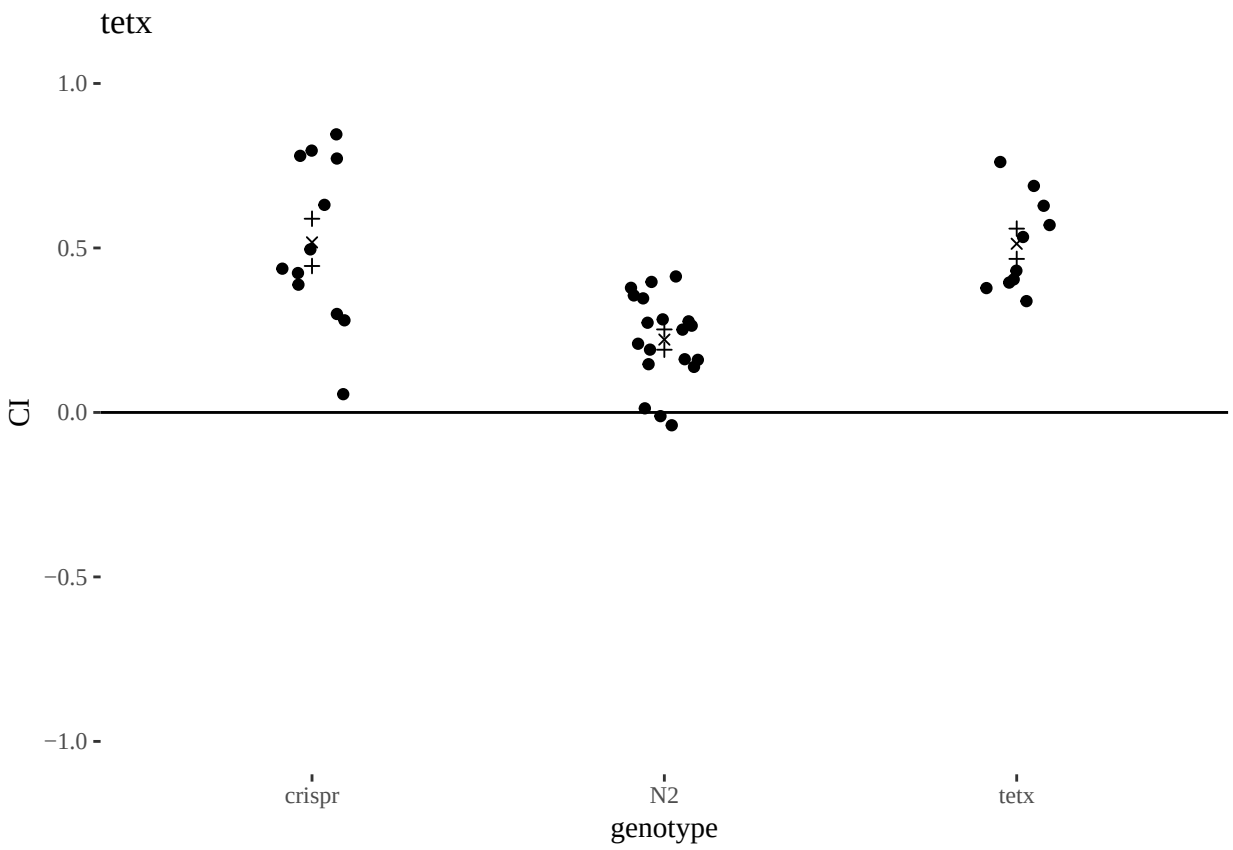
```
t.test(ewir74noR$CI,ewir74R$CI)
```

```
##
##  Welch Two Sample t-test
##
## data:  ewir74noR$CI and ewir74R$CI
## t = 4.3087, df = 9.6589, p-value = 0.001669
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1607155 0.5083839
## sample estimates:
##    mean of x    mean of y
## 0.3355008562 0.0009511747
```

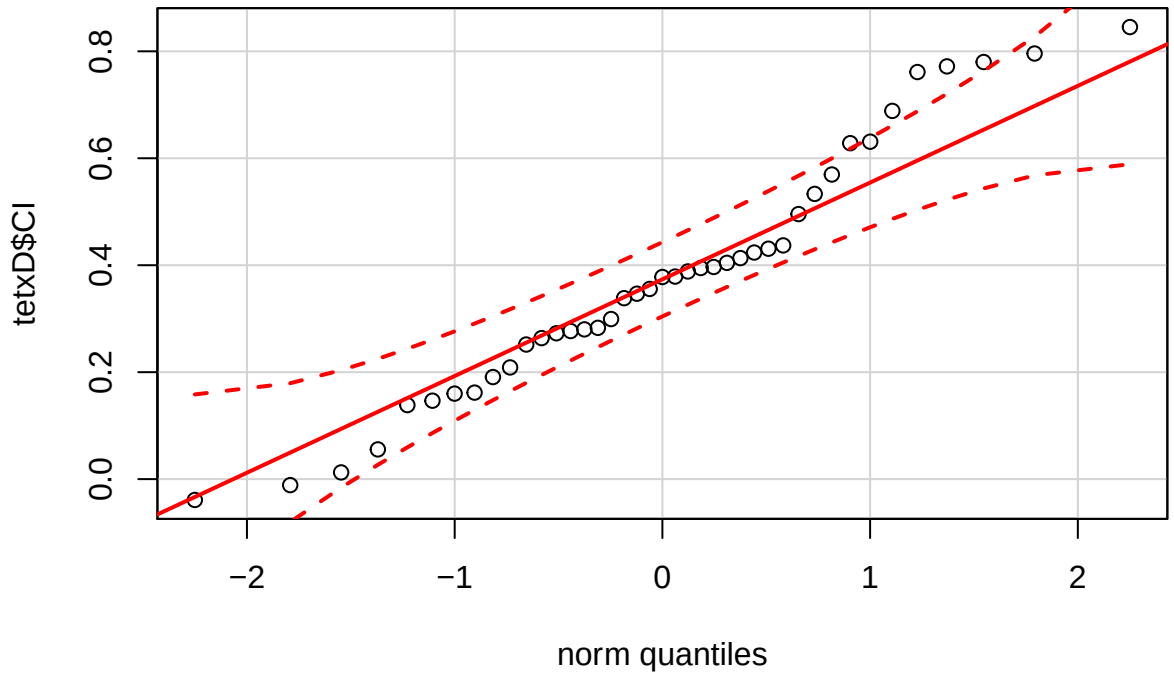
ADL chemical synapses are required for complete DEET-sensitivity

Figure 3b

```
# make summary data for tetx

tetx <- subset(allbeh,expt=='tetx')
tetx <- subset(tetx,genotype!='ed03')
#removes one data point where I had an ed03 plate to check stocks after thawing, done at the same time
tetxD <- subset(tetx,percentDEET==0.15)

tetxDsummary <-ddply(tetxD,c("genotype"),summarise,N=length(genotype),mean=mean(CI),sd=sd(CI),se=sd/sqr
tetxDsummary$mep <- tetxDsummary$mean + tetxDsummary$se
tetxDsummary$mem <- tetxDsummary$mean - tetxDsummary$se

ggplot(tetxD,aes(x=genotype,y=CI)) +
  ggtitle('tetx')  +
  geom_point(data=tetxDsummary,aes(x=genotype,y=mep),shape=3)  +
  geom_point(data=tetxDsummary,aes(x=genotype,y=mem),shape=3)  +
  geom_point(data=tetxDsummary,aes(x=genotype,y=mean),shape=4)  +
  geom_jitter(width=0.1,height=0) +
  theme_tufte() +
  ylim(-1,1) +
  geom_hline(yintercept=0)
```




```
qqPlot(tetxD$CI)
```




```
TukeyHSD(aov(CI~genotype,tetxD))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
```

```
##
## Fit: aov(formula = CI ~ genotype, data = tetxD)
##
## $genotype
##                     diff        lwr        upr     p adj
## N2-crispr   -0.295512908 -0.4550945 -0.1359314 0.0001724
## tetx-crispr -0.004240982 -0.1895472  0.1810652 0.9982840
## tetx-N2      0.291271926  0.1221922  0.4603516 0.0004454
```