Hormone Stats

Rens Holmer

30/4/2021

Data loading

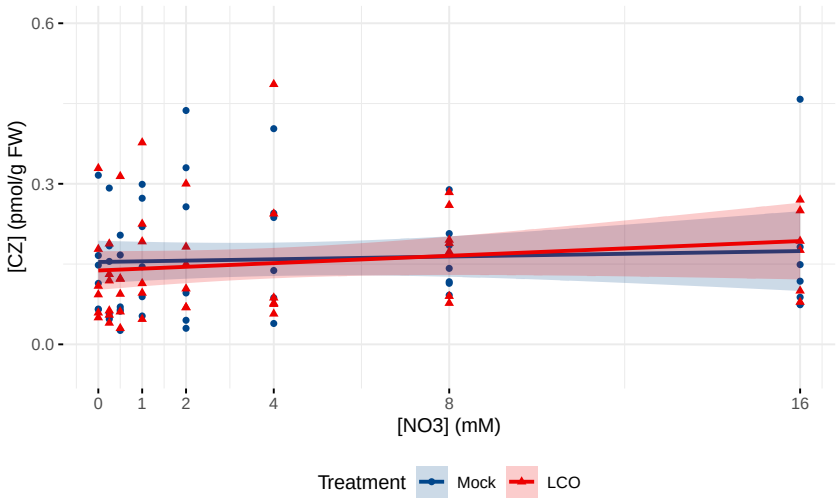
```
load.hormone.df <- function(name) {
  df <- read.csv(paste0('./', name, '.csv'), sep=';') %>%
    rename(Concentration = pmol...g.FW, Treatment = LCO)
  df$Hormone <- name
  df$Treatment <- recode_factor(df$Treatment, `-` = 'Mock', `+` = 'LCO')
  return(df)
}

hormone.names <- c('CZ','CZr','IP','IPr','TZ','TZr')
hormone.dfs <- lapply(hormone.names, load.hormone.df)
names(hormone.dfs) <- hormone.names
```

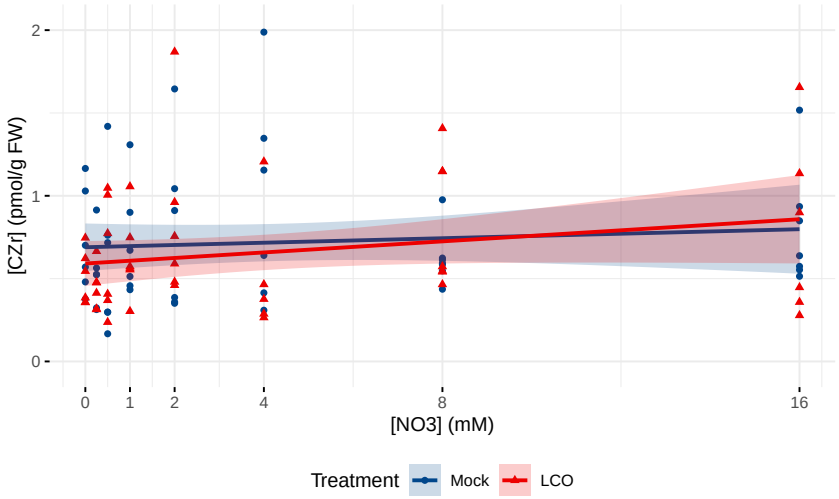
Plotting

```
names(hormone.names) <- hormone.names
lapply(hormone.names, function(hormone){
  hormone.df <- hormone.dfs[[hormone]]
  if (hormone %in% c('CZr','IPr')) {
    ymax <- 2.05
  } else {
    ymax <- 0.6
  }
  p <- ggplot(hormone.df, aes(x=NO3, y=Concentration, colour=Treatment)) +
    geom_point(aes(shape=Treatment)) +
    stat_smooth(aes(fill=Treatment), method='lm', formula=y~x, alpha=0.2) +
    scale_y_continuous(limits=c(-0.05, ymax), n.breaks=3) +
    scale_x_continuous(breaks=c(0,1,2,4,8,16))
  ggpar(p, ylab=paste0('[',hormone,'] (pmol/g FW)'), xlab='[NO3] (mM)', palette='lancet', ggtheme=theme_
        legend='bottom')
})
```

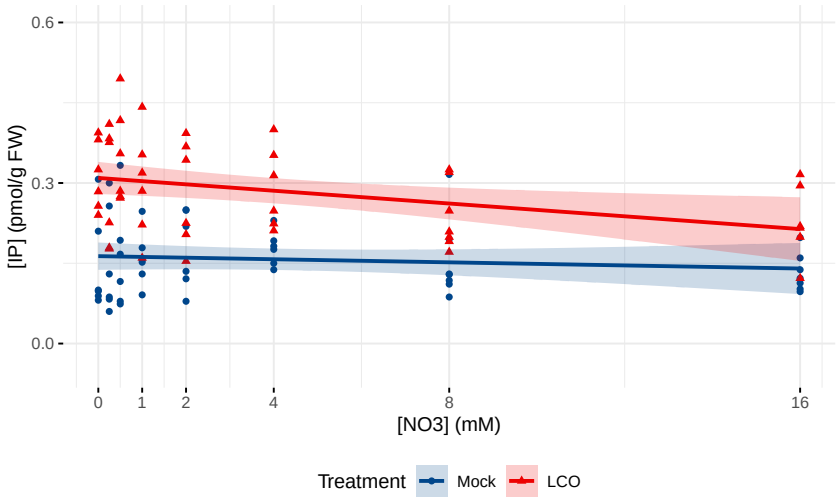
```
## $CZ
```


0.6
0.3
0.0
[CZ] (pmol/g FW)
0
1
2
4
8
16
[NO3] (mM)
Treatment
Mock
LCO

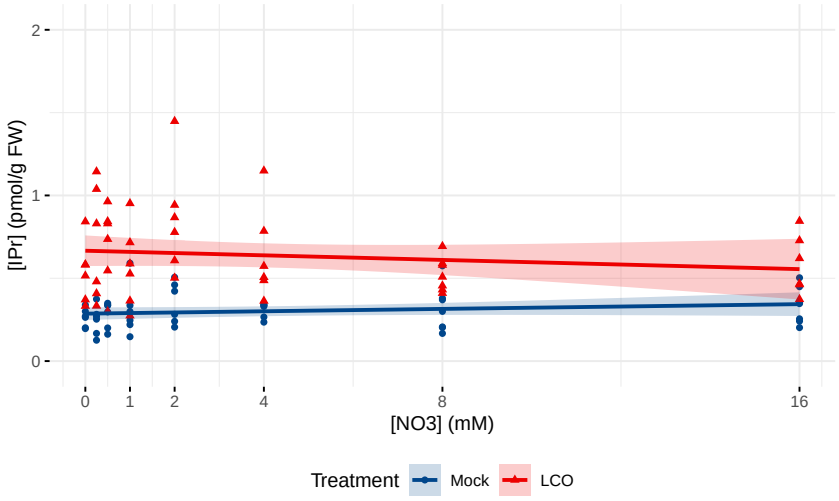
```
##
## $CZr
```


2
1
0
[CZr] (pmol/g FW)
0
1
2
4
8
16
[NO3] (mM)
Treatment
Mock
LCO

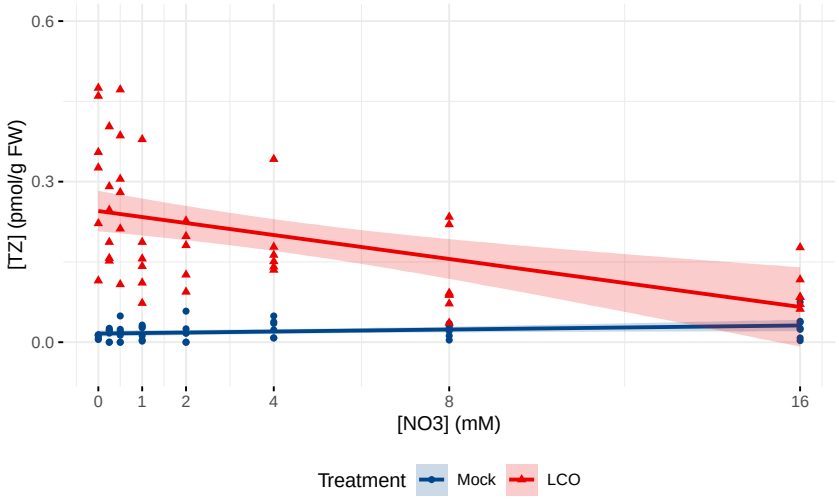
```
##
## $IP
```


0.6
0.3
0.0
[IP] (pmol/g FW)
0
1
2
4
8
16
[NO3] (mM)
Treatment
Mock
LCO

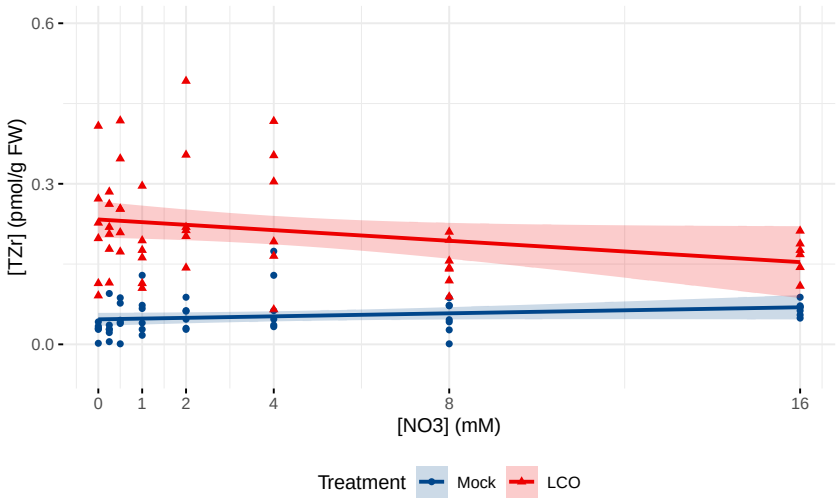
```
##
## $IPr
```


2
1
0
[IPr] (pmol/g FW)
0
1
2
4
8
16
[NO3] (mM)
Treatment
Mock
LCO

```
##
## $TZ
```


0.6
0.3
0.0
[TZ] (pmol/g FW)
0
1
2
4
8
16
[NO3] (mM)
Treatment
Mock
LCO

```
##
## $TZr
```



Fitting linear models

By fitting a linear model with an interaction term between NO_3^- (continuous) and LCO (binary), the following things can be tested:

- Overall effect of NO_3^- on hormone concentration
- Overall effect of LCO treatment on hormone concentration
- Difference in effect of NO_3^- concentration on hormone concentration between LCO and mock treatment

Resulting model coefficients and corresponding p-values are summarized in Table 1. Conclusions per hormone:

- **CZ** - Only significant term is the intercept, indicating CZ is present in all samples, but neither NO_3^- nor LCO have an effect
- **CZr** - Same as CZ
- **IP** - LCO, and intercept are significant, NO_3^1 and interaction are not. Following the model coefficients, it can be concluded that LCO has a positive effect on IP, but that NO_3^- does not change this.
- **IPr** - LCO and intercept are significant. This indicates IPr is present, IPr is higher overall in LCO treated samples, but NO_3^- does not have an effect on IPr (Same as IP)
- **TZ** - LCO and LCO:NO3 interaction are significant. TZ responds to LCO treatment, but is absent (statistically indistinguishable from zero) in mock. Model coefficients indicate TS is higher in LCO treated samples, but gets lower at increased NO3 *in LCO treated samples only.*
- **TZr** - Intercept, LCO and LCO:NO3 interaction are significant. TZr is always present, higher in LCO treated samples, but gets lower at increased NO3 *in LCO treated samples only.*

Taken together, none of the CK variants respond to NO3 in mock treated samples. CK response to NO3 only occurs in LCO treated samples for TZ and TZr. CZ and CZr do not respond to LCO or NO3. IP and IPr respond to LCO but not NO3. Some samples show a non-significant trend.

```r
# Fit models
models <- lapply(hormone.dfs, function(hormone.df){
  lm(Concentration ~ NO3 + Treatment + NO3:Treatment, hormone.df)
})

# Make an overview table
stargazer(models, header=FALSE, dep.var.labels.include=FALSE,
          title=paste(
            'Linear model coefficients with confidence intervals and p-values',
            'modelling the relationship between concentration of six different',
            'hormones, $NO_3^-$ concentration and LCO treatment.',
            'Model specification: $Hormone \\sim NO_3^- + LCO + NO_3^-:LCO$'
            ), dep.var.caption='',ci=TRUE, column.labels=hormone.names,
          covariate.labels=c('$NO_3^-$','LCO','$NO_3^-$:LCO','Intercept'),
          model.numbers=FALSE, report='vcsp*', keep.stat = c('n'),
          float.env='sidewaystable')
```

NIN expression

```r
# Load NIN expression data
nin.df <- read.csv('./nin expressie NO3 range.csv', row.names=NULL, sep=';', dec=',') %>%
    rename(Treatment = LCO)
nin.df$Treatment <- recode_factor(nin.df$Treatment, `-` = 'Mock', `+` = 'LCO')
```

NIN expression responds non-linearly over the NO3 concentration range, so we use a Generalized Additive Model (GAM). To determine importance of NO3 and LCO-treatment in explaining NIN expression four GAMs are fit and compared using Akaike Information Criterion. The most complex model incorporates both NO3 and LCO-treatment and has the lowest AIC, indicating it is the best fit for this data. In other words, NIN expression varies based on both NO3 concentration and LCO-treatment.

```r
# Fitting various GAM models to test influence of NO3 and LCO on expression
m0 <- mgcv::gam(expression ~ 1, data=nin.df)
m1 <- mgcv::gam(expression ~ s(NO3, k=7), data=nin.df)
m2 <- mgcv::gam(expression ~ Treatment, data=nin.df)
m3 <- mgcv::gam(expression ~ s(NO3, k=7, by=Treatment) + Treatment, data=nin.df)

# Determine which model best explains the data based on Akaike Information Criteria
aic.scores <- AIC(m0,m1,m2,m3)
rownames(aic.scores) <- c('Intercept','NO3','LCO','NO3 + LCO')

stargazer(aic.scores, summary = FALSE, header = FALSE, object.names = TRUE,
          title='AIC scores for various GAM models of NIN expression')
```

```r
pred.df <- data.frame(
  NO3 = rep(seq(0, 16, .1), 2),
  Treatment = factor(c(rep('Mock', 161), rep('LCO', 161)), levels=c('Mock','LCO'))
)

pred.df <- cbind(pred.df, predict(m3, pred.df, se.fit=TRUE))
pred.df$expression <- NaN

p <- ggplot(nin.df, aes(x=NO3, y=expression, color=Treatment)) +
  geom_point() +
```

Table 1: Linear model coefficients with confidence intervals and p-values modelling the relationship between concentration of six different hormones, NO_3^- concentration and LCO treatment. Model specification: $Hormone \sim NO_3^- + LCO + NO_3^- : LCO$

	CZ	CZr	IP	IPr	TZ	TZr
NO_3^-	0.001	0.007	−0.001	0.004	0.001	0.001
	(−0.004, 0.007)	(−0.014, 0.028)	(−0.005, 0.002)	(−0.006, 0.014)	(−0.003, 0.005)	(−0.002, 0.005)
	p = 0.644	p = 0.529	p = 0.474	p = 0.482	p = 0.664	p = 0.442
LCO	−0.016	−0.067	0.146	0.380	0.239	0.187
	(−0.068, 0.037)	(−0.273, 0.138)	(0.108, 0.184)	(0.283, 0.477)	(0.198, 0.281)	(0.152, 0.221)
	p = 0.555	p = 0.523	p = 0.000***	p = 0.000***	p = 0.000***	p = 0.000***
NO_3^-:LCO	0.002	0.010	−0.005	−0.011	−0.013	−0.006
	(−0.006, 0.010)	(−0.021, 0.040)	(−0.010, 0.001)	(−0.025, 0.004)	(−0.019, −0.007)	(−0.012, −0.001)
	p = 0.589	p = 0.531	p = 0.124	p = 0.157	p = 0.0001***	p = 0.018**
Intercept	0.154	0.689	0.163	0.286	0.016	0.047
	(0.117, 0.191)	(0.544, 0.834)	(0.136, 0.190)	(0.218, 0.355)	(−0.013, 0.045)	(0.022, 0.071)
	p = 0.000***	p = 0.000***	p = 0.000***	p = 0.000***	p = 0.281	p = 0.0004***
Observations	99	99	99	99	99	99

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2: AIC scores for various GAM models of NIN expression

	df	AIC
Intercept	2	536.151
NO3	3	536.461
LCO	3	465.644
NO3 + LCO	9.848	411.748

```
  geom_ribbon(aes(x=NO3, ymin=fit-se.fit, ymax=fit+se.fit, fill=Treatment),
              data=pred.df, alpha=0.2, color=NA) +
  geom_line(aes(x=NO3, y=fit, color=Treatment), data=pred.df) +
  scale_y_continuous(n.breaks=3) +
  scale_x_continuous(breaks=c(0,1,2,4,8,16))
ggpar(p, ylab='Relative expression', xlab='NO3 [mM]', palette='lancet',
      ggtheme=theme_minimal(),
      legend='bottom', title='MtNIN expression')
```