WorMotel Heritability with activity levels

Tim C.

1/19/2018

This code uses processed survival data from Chris Fang-Yen's lab to calcualte heritability of survival traits. The data were collected with WorMotel in winter 2017 by Matt Churgin at UPenn.

Read in data and define functions

```
df <- as.data.frame(read.csv(file = "20170116_FullProcessed.csv", header = T))

glimpse(df)
```

```
## Observations: 64,800
## Variables: 8
## $ strain        <fctr> CB4856, CB4856, CB4856, CB4856, CB4856, CB4856,...
## $ tech_rep      <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ...
## $ bio_rep       <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
## $ rep           <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
## $ life_span     <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ activity_type <fctr> stim, stim, stim, stim, stim, stim, stim, stim,...
## $ time_d        <dbl> 1.5, 3.0, 4.5, 6.0, 8.0, 9.5, 11.5, 13.0, 15.0, ...
## $ activity      <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
```

```
# define heritability test
H2.test <- function(data){

  pheno <- as.data.frame(dplyr::select(data, phenotype))[,1]
  strain <- as.factor(data$strain)

  reffMod <- lme4::lmer(pheno ~ 1 + (1|strain))

  Variances <- as.data.frame(lme4::VarCorr(reffMod, comp = "Variance"))

  Vg <- Variances$vcov[1]
  Ve <- Variances$vcov[2]
  H2 <- Vg/(Vg+Ve)

  return(H2)
}

# remove outliers function
remove_outliers <- function(x, na.rm = TRUE, ...) {
  qnt <- quantile(x, probs=c(.25, .75), na.rm = na.rm, ...)
  H <- 1.5 * IQR(x, na.rm = na.rm)
  y <- x
  y[x < (qnt[1] - H)] <- NA
  y[x > (qnt[2] + H)] <- NA
  y
}
```

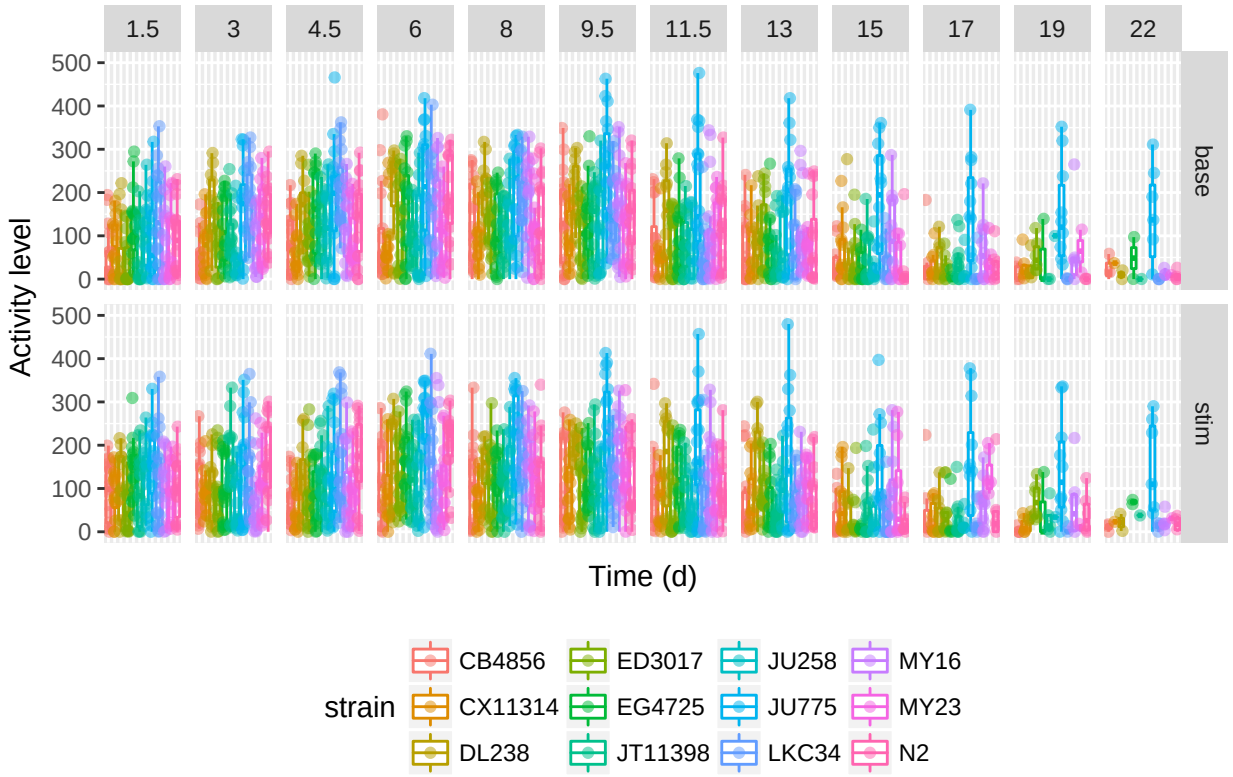
Each bio_rep is treated as a block for the following analysis. Technical replicates are ignored.

ANALYSIS 1: Focus on activity levels within block 1 across all timepoints

```
## Warning: Removed 13 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 262 rows containing missing values (geom_point).
```

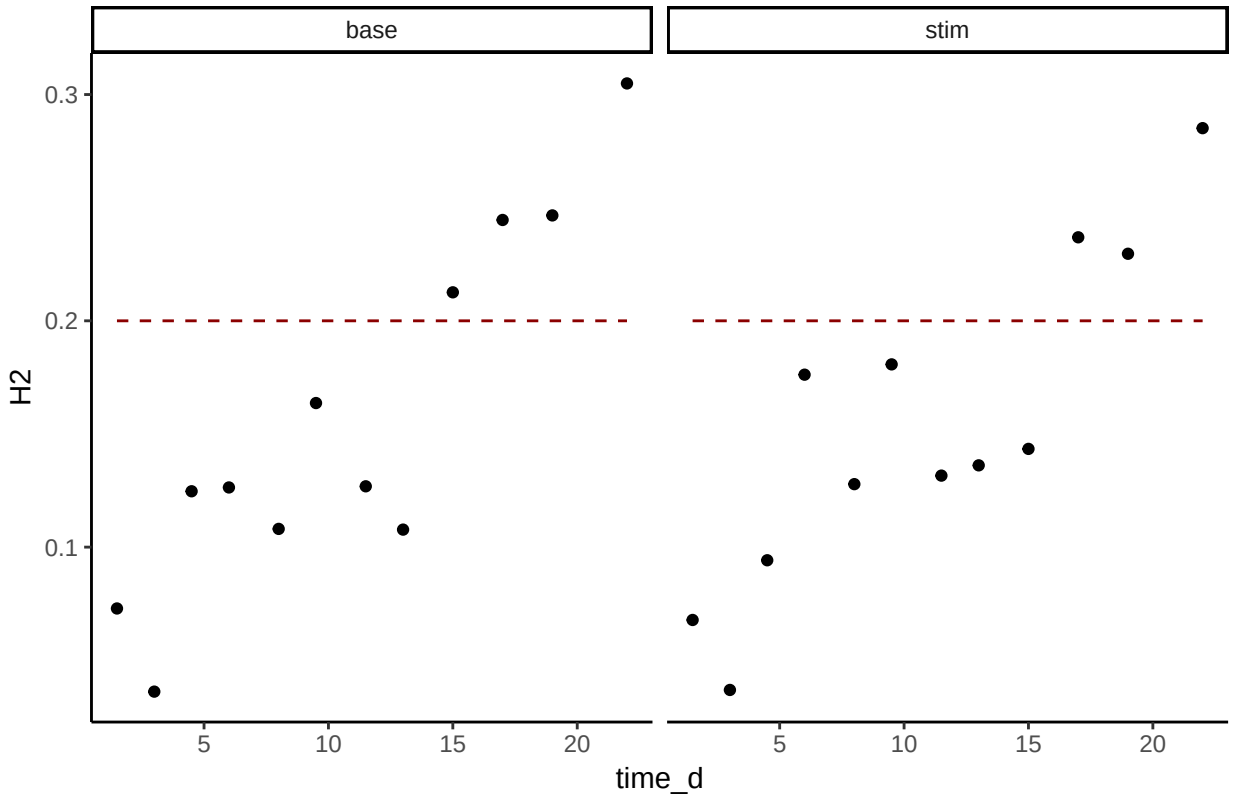
FIGURE 1: Block 1 activity levels



JU775 appears to retain higher activity levels compared to other strains at later timepoints. I'm concerned about calculating heritability for these timepoints however because the number of nematodes observed is 1 for some strains at day 19 and day 22. I'll do the calculation anyway and see what we get.

Heritability for activity levels at different timepoints within block 1

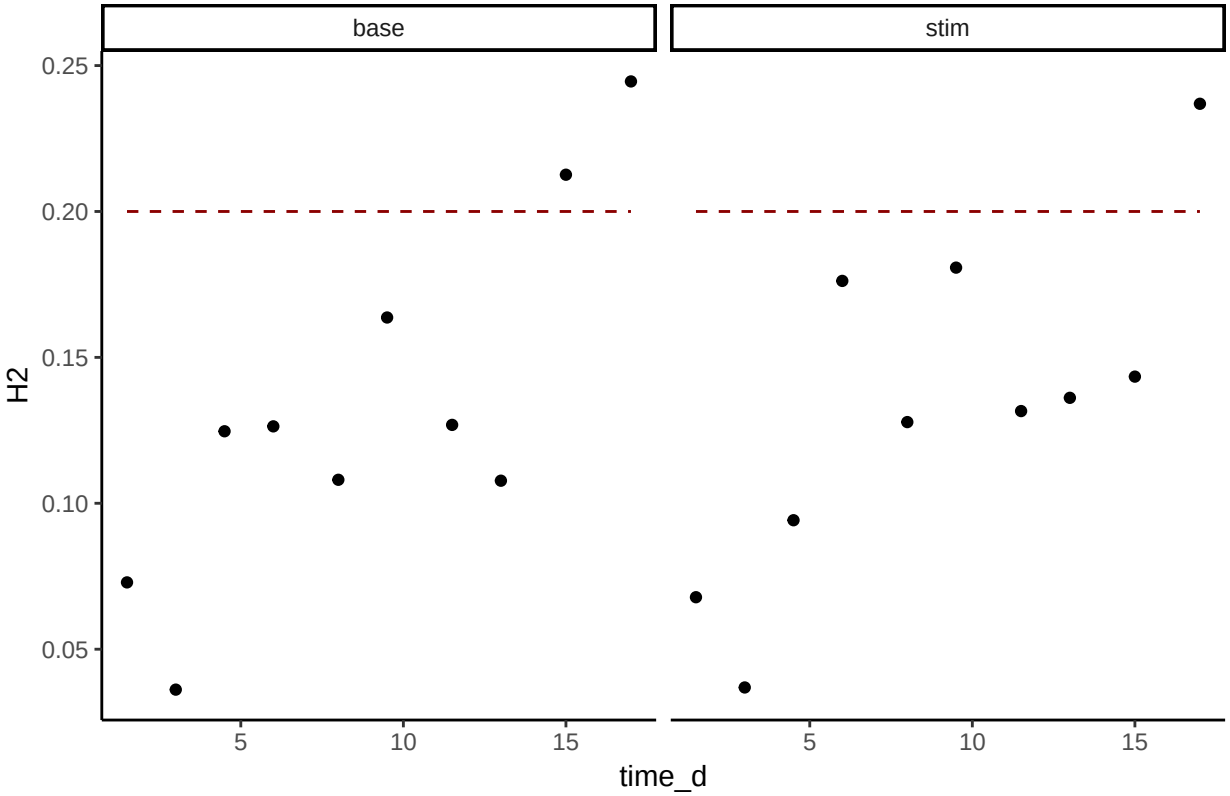
FIGURE 2: Heritability for activity traits



We see very high heritability for activity traits at later timepoints. However, we have less than two replicates for some strains at the later time points. I'll try filtering timepoints to those containing at least 5 reps per strain.

Recalculate heritability on timepoints with at least 5 reps per strain

FIGURE 3: Heritability for activity traits (5 reps / strain)

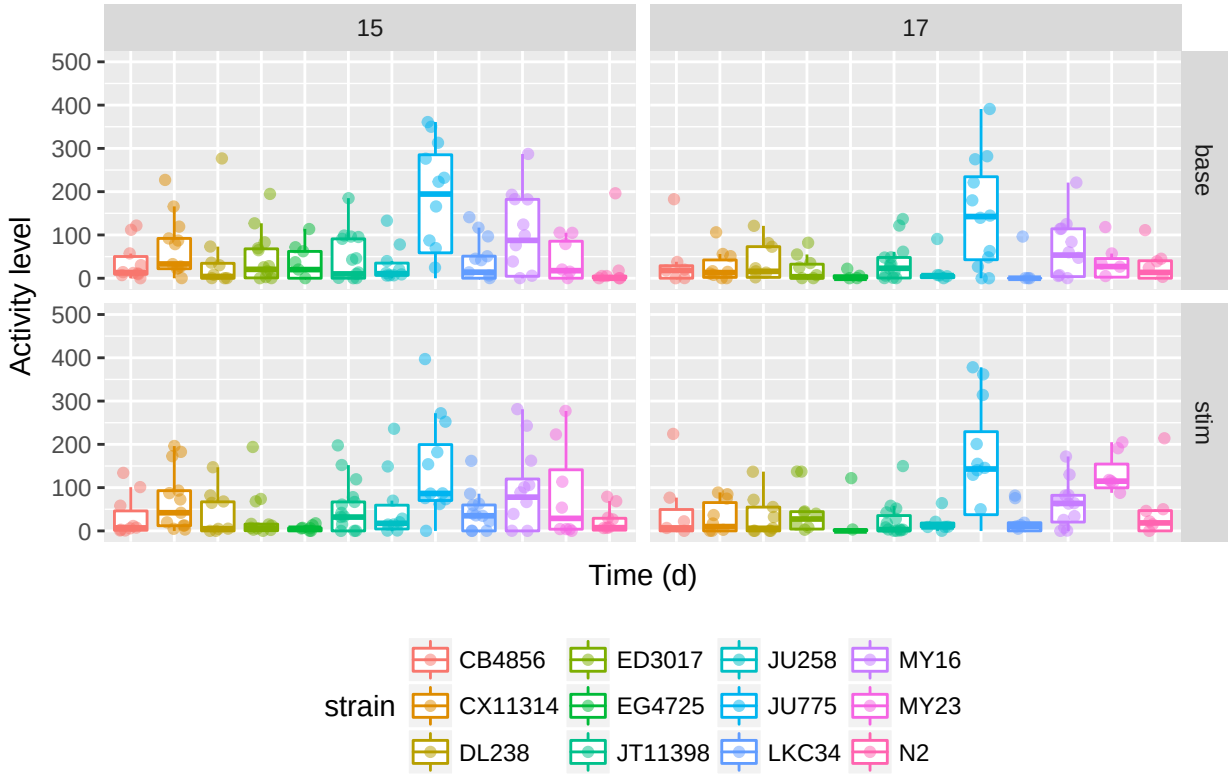


Heritability is still high for activity traits at 17 days. Below is a more detailed plot of those timepoints.

Heritability for time points with at lest 5 reps per strain

```
## Warning: Removed 93 rows containing missing values (geom_point).
```

FIGURE 4: Block 1 activity levels day 15 and 17



ANALYSIS 1 CONCLUSIONS:

1) Heritability for both stim and base activity traits at day 17 exceeds the 0.2 threshold (Fig. 3). 2) The JU775 strain is likely driving heritability in activity traits since it maintains much higher activity levels at late timepoints (Fig. 4). 3) The difference between stimulated and base activity is small for most individuals (Figs. 1, 4). Is this expected?

ANALYSIS 2: Fit model to the drop in activity levels over time and calculate heritability on model parameters.

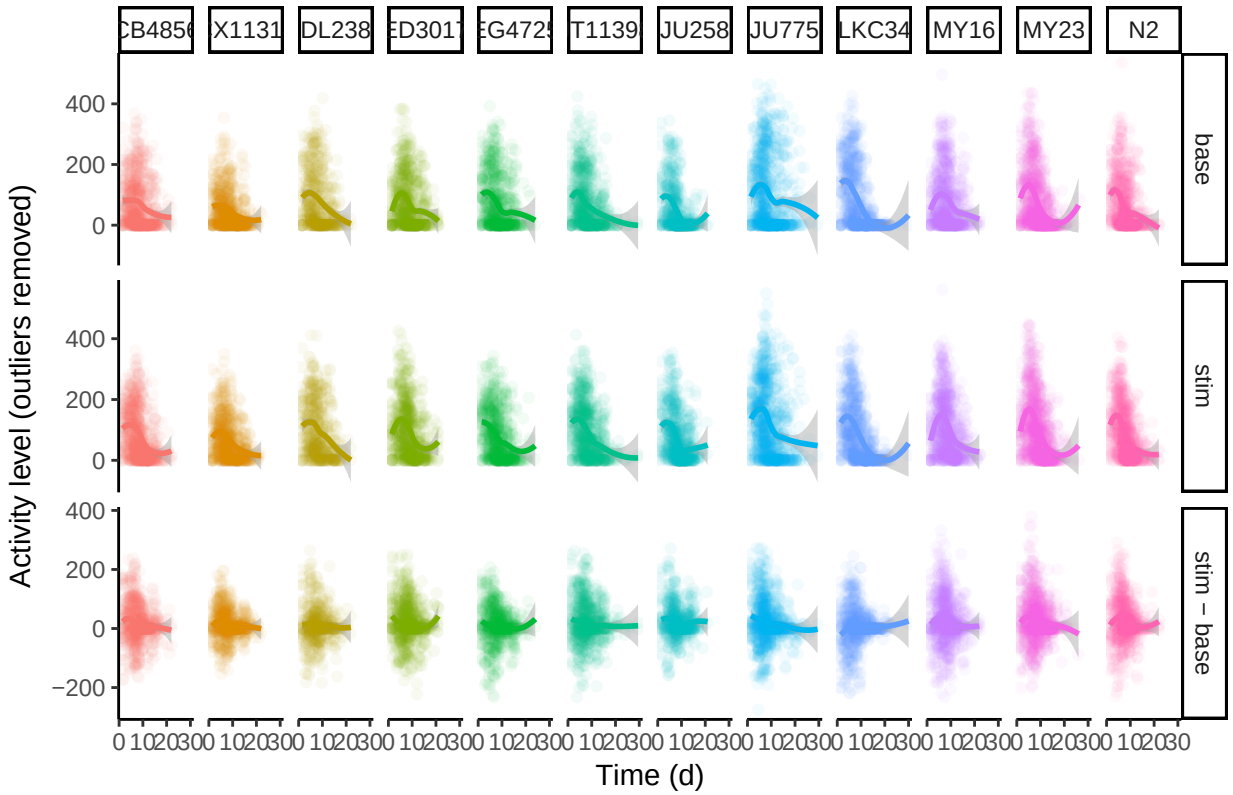
For now I'm moving on to fitting a curve to the decrease in activity levels over time. We might see high heritability in fit parameters. Outlier activity levels are filtered for this analysis.

```
## `geom_smooth()` using method = 'loess'
```

```
## Warning: Removed 1540 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 1540 rows containing missing values (geom_point).
```

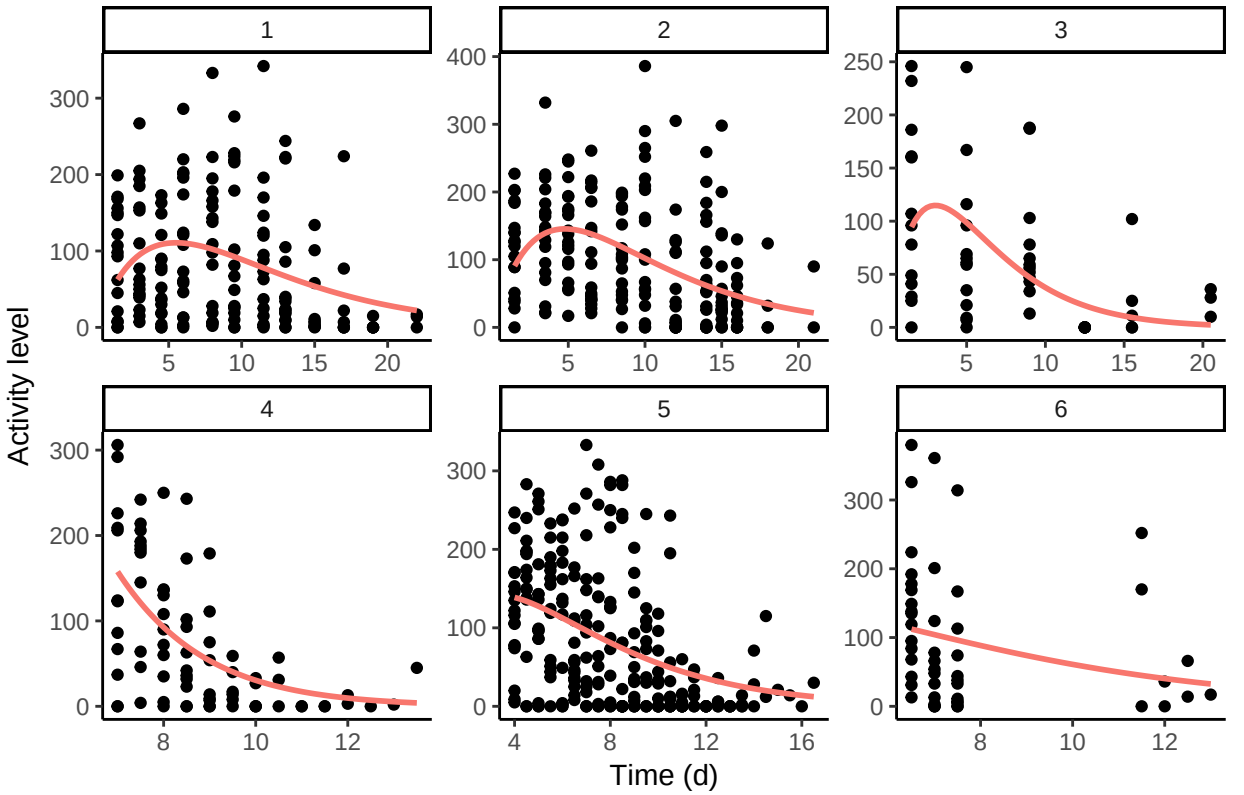
FIGURE 5: Loess fit to activity levels over time



It looks like we may be able to fit curves to the acivity data, but it is noisy. The Loess fit looks similar to a Ricker function with an early hump and then exponential decay. Stim and base traits looks the most promising. Ricker function, f(x) = a$x$exp(1)^(b*x).

Testing fit of Ricker function to CB4856 data for each block

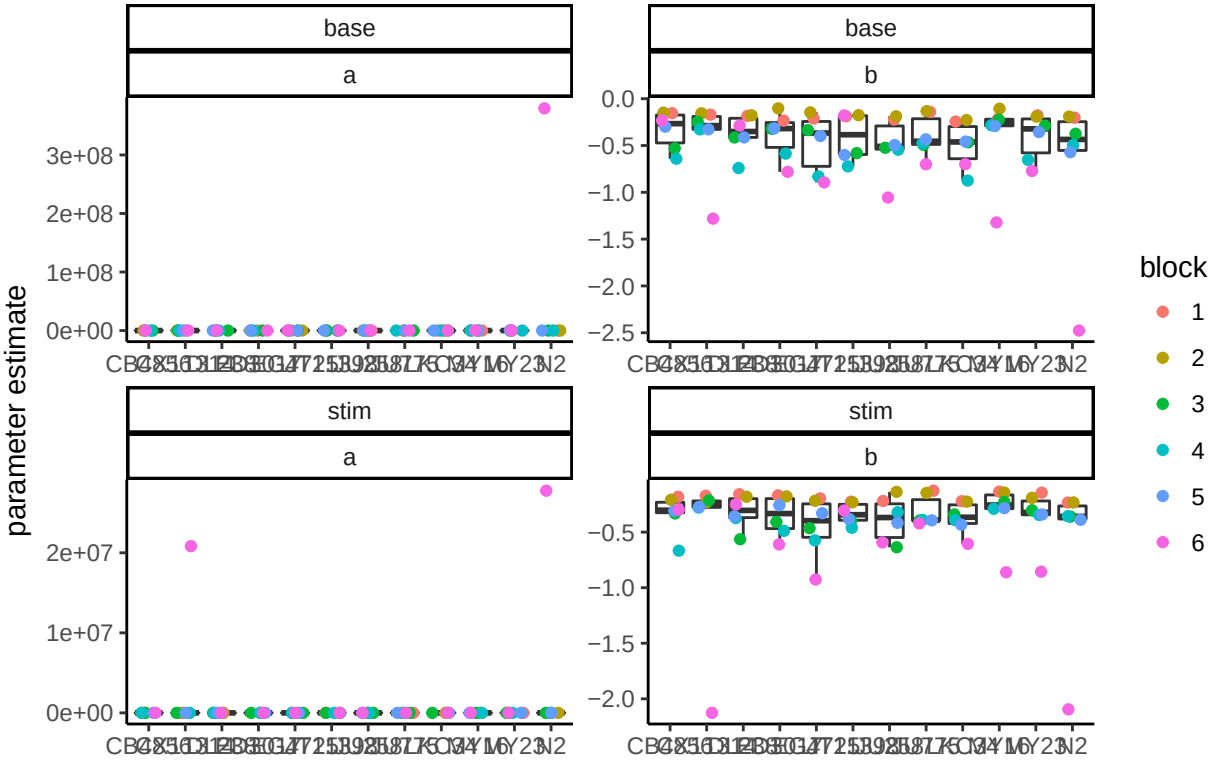
FIGURE 6: CB4856 stimulatived activity levels across blocks with Ricker fit



Ricker model looks similar to loess fit for blocks 1-3. However, the shape of the fit is very different among the blocks. I'll extract `a` and `b` parameters from Ricker f(x) = a$x$exp(1)^(b*x) for each strain and each block then calculate heritability.

Plotting raw Ricker fit parameters (a) and (b) for stim activity and base activity

FIGURE 7: Ricker fit parameter estimates



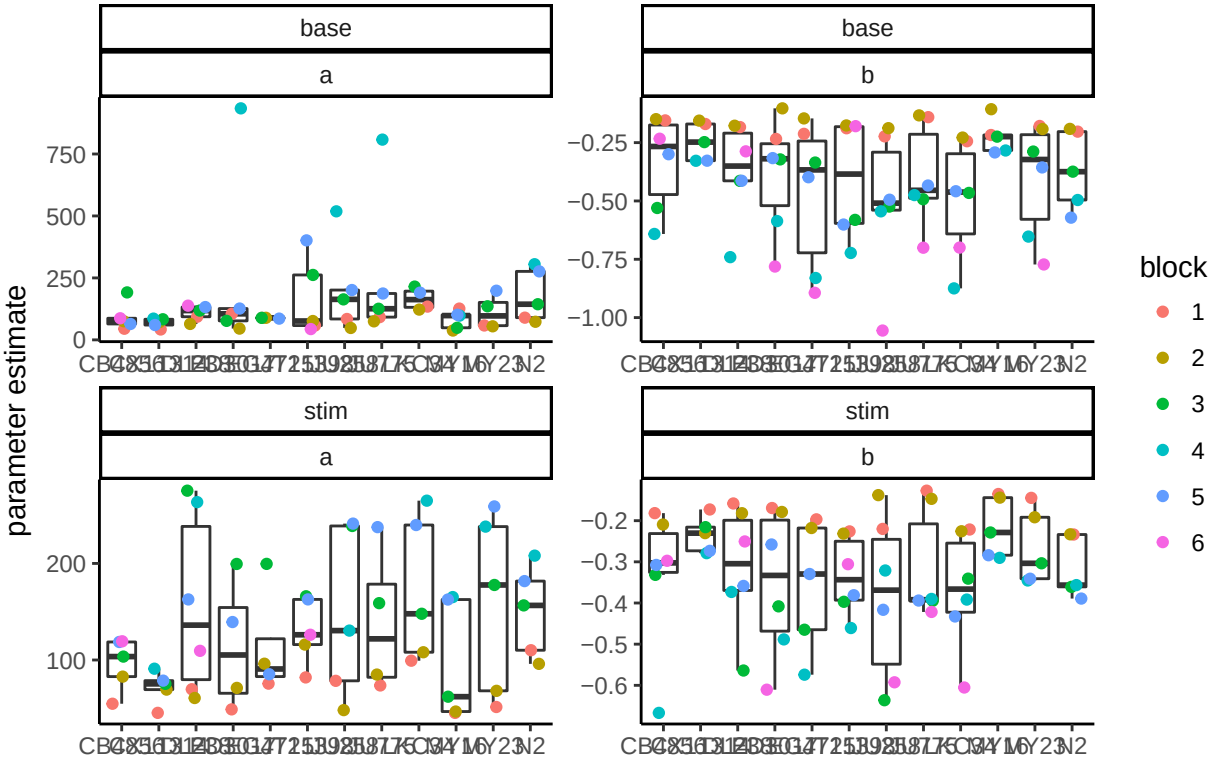
Huge outliers coming from misfit model, mostly in block 6. I can filter these outlier parameter estimates and replot.

Ricker fit parameters with outliers removed

```
## Warning: Removed 37 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 37 rows containing missing values (geom_point).
```

FIGURE 8: Ricker fit parameter estimates (outliers removed)



Outlier filtered data looks better. I'll calculate heritability with outlier filtered and block effect corrected data.

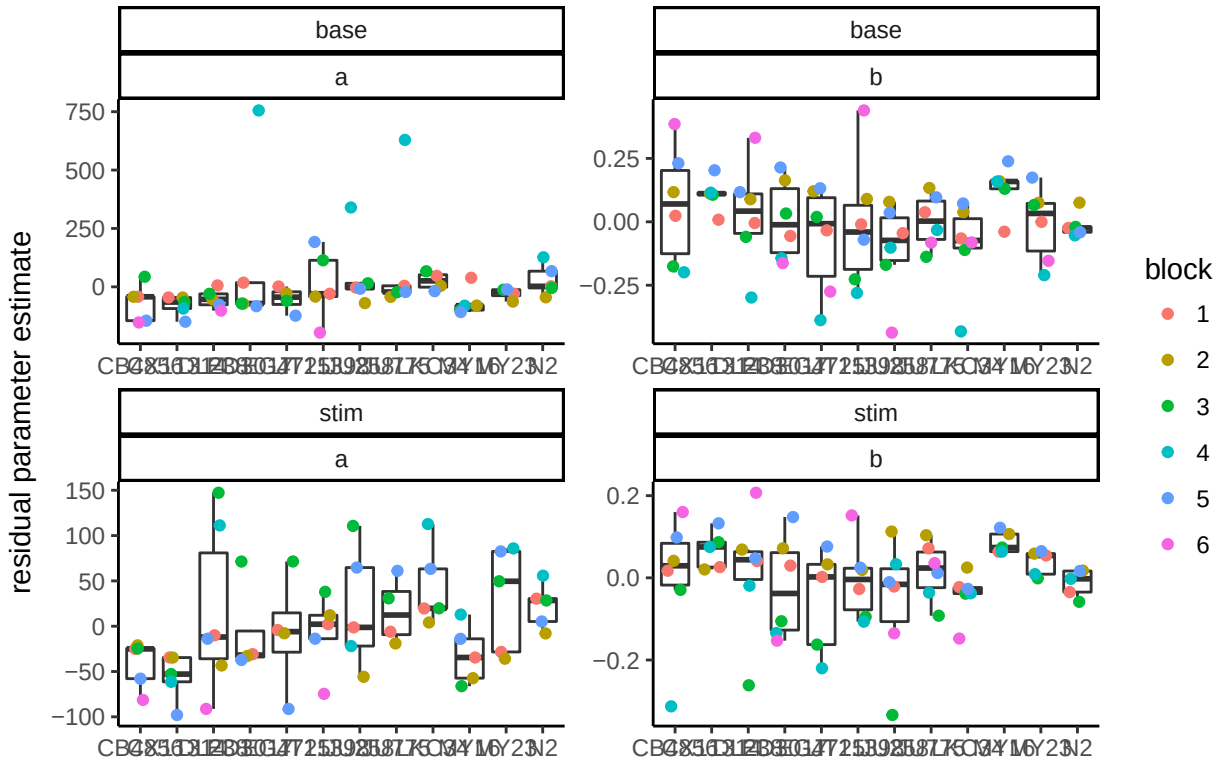
Correct for block effect and plot residuals

```
# Process data for heritability calculation
new_df <- df_proc %>%
  tidyr::gather(trait, phenotype, -block, -strain, -rep, -time_d) %>%
  dplyr::filter(trait %in% c("base", "stim")) %>%
  dplyr::group_by(block, strain, trait) %>%
  dplyr::do(tidy(nlsLM(phenotype ~ a*time_d*exp(1)^(b*time_d),
                    data = ., start = list(a=100,b=0.2), control = nls.lm.control(maxiter = 200)))) %>%
  dplyr::ungroup() %>%
  dplyr::group_by(trait, term) %>%
  dplyr::mutate(phenotype = remove_outliers(estimate)) %>% #identify and remove outliers
  dplyr::ungroup() %>%
  dplyr::filter(complete.cases(.)) %>% # remove outlier rows from data frame
  dplyr::group_by(trait, term) %>%
  dplyr::mutate(phenotype = residuals(lm(estimate ~ block)))  # regress out block effect

ricker <- ggplot(new_df) +
  aes(x = strain, y = phenotype) +
  geom_boxplot(outlier.shape = NA) +
  geom_jitter(aes(color = factor(block)), width = .25) +
  facet_wrap(trait ~ term, scales = "free") +
  labs(title = "FIGURE 9: Risidual ricker fit parameter estimates (outliers removed)", y="residual para
  theme_classic() +
  theme(legend.position = "right")
```

```
ricker
```

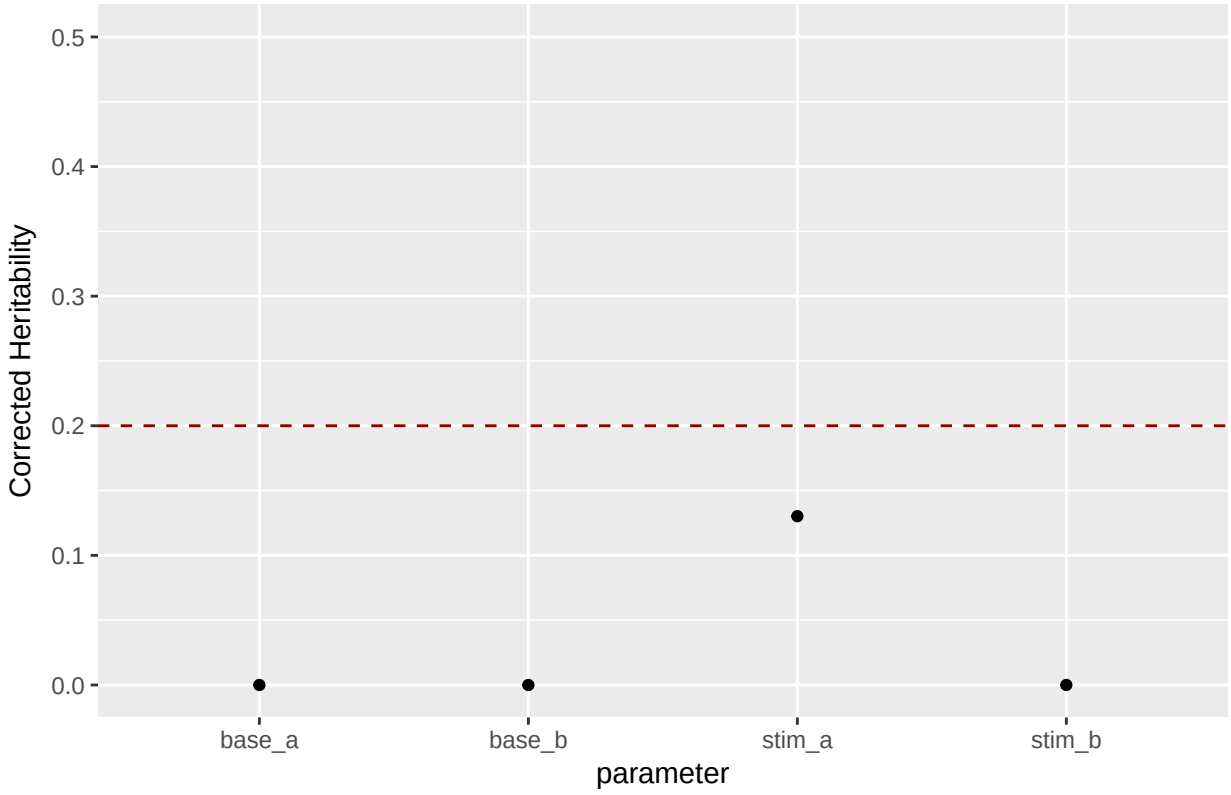
FIGURE 9: Risidual ricker fit parameter estimates (outliers removed)



Differences among strains is most pronounced for the stimulated activity parameter `a` trait. The other traits do not appear to be as different across strains.

Calculating heritability on residual Ricker fit parameter estimates for base and stimulated activity

FIGURE 10: Heritability of residual ricker fit parameter estimates



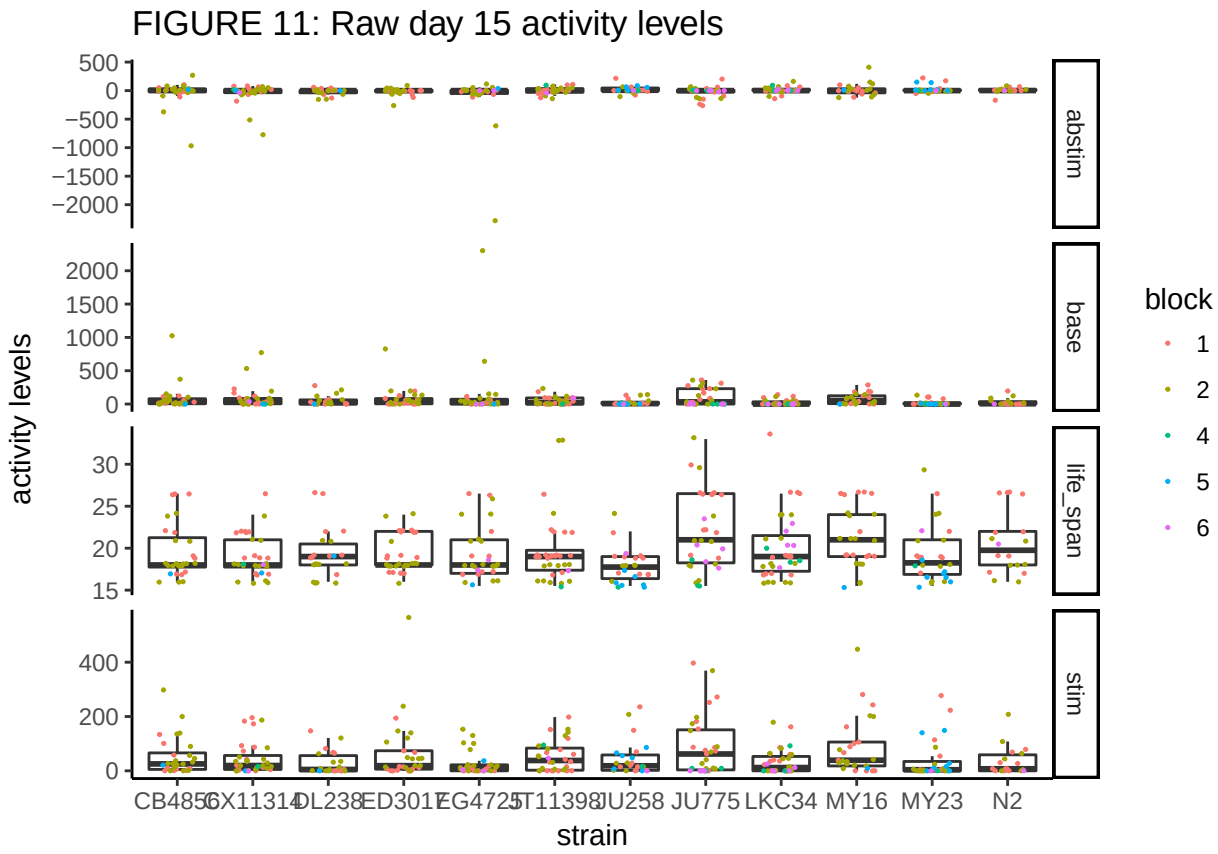
ANALYSIS 2 CONCLUSIONS:

1) Ricker parameter estimates `a` and `b` are very different among blocks even for the same strain (Fig. 6) 2) After removing outliers and correcting for block effect, heritability is just 0.13 for the stim activity parameter `a` trait (Fig. 10).

ANALYSIS 3: Focus on day 15 activity data alone

I'm interested in day 15 data because 5 of the 6 blocks share this timepoint and therefore we should have many replicates. The activity traits below are taken from day 15 worms. Block 3 did not have a day 15 timepoint so it is excluded Base activity is the activity measure before stimulus light. Stim activity is the activity level after stimulus light was applied. stim - base activity is the difference between the two.

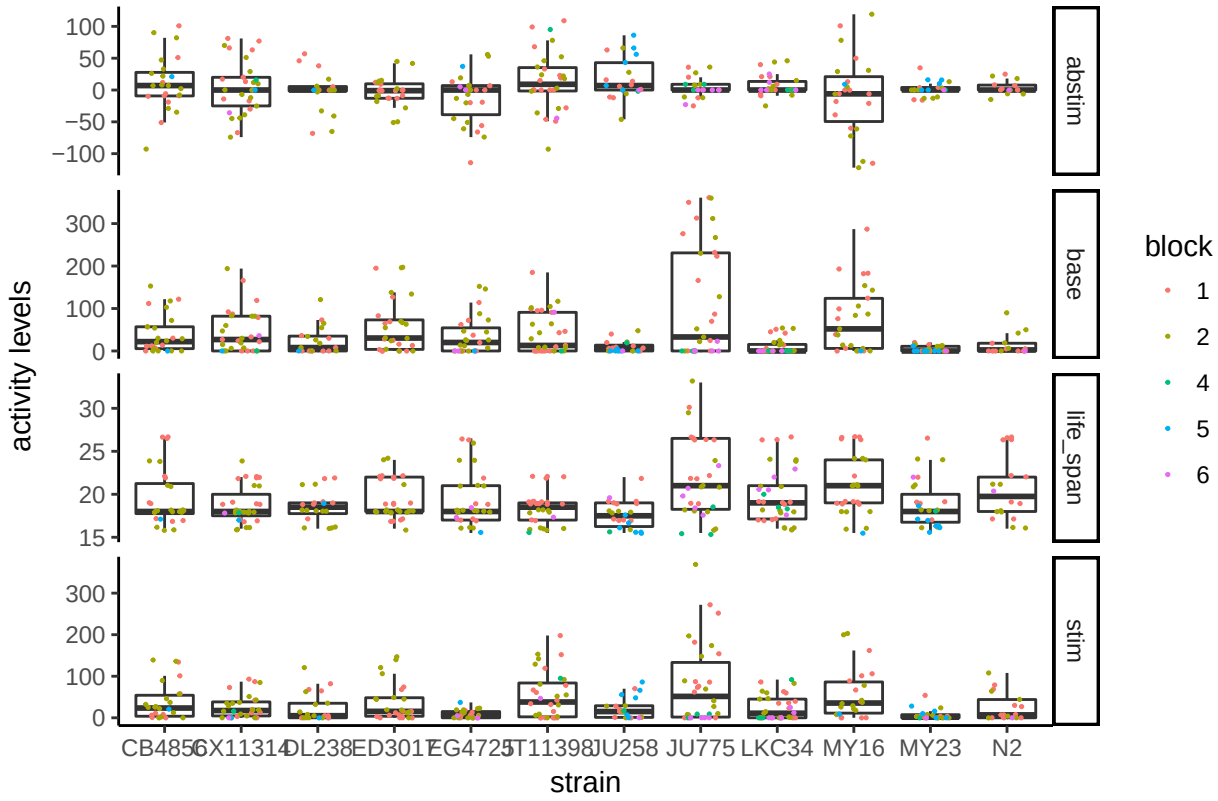
Plotting individual worm activity and life span traits for day 15



Block 2 has some extreme values. I can filter these outliers and replot.

Filter outliers in day 15 activity and replot

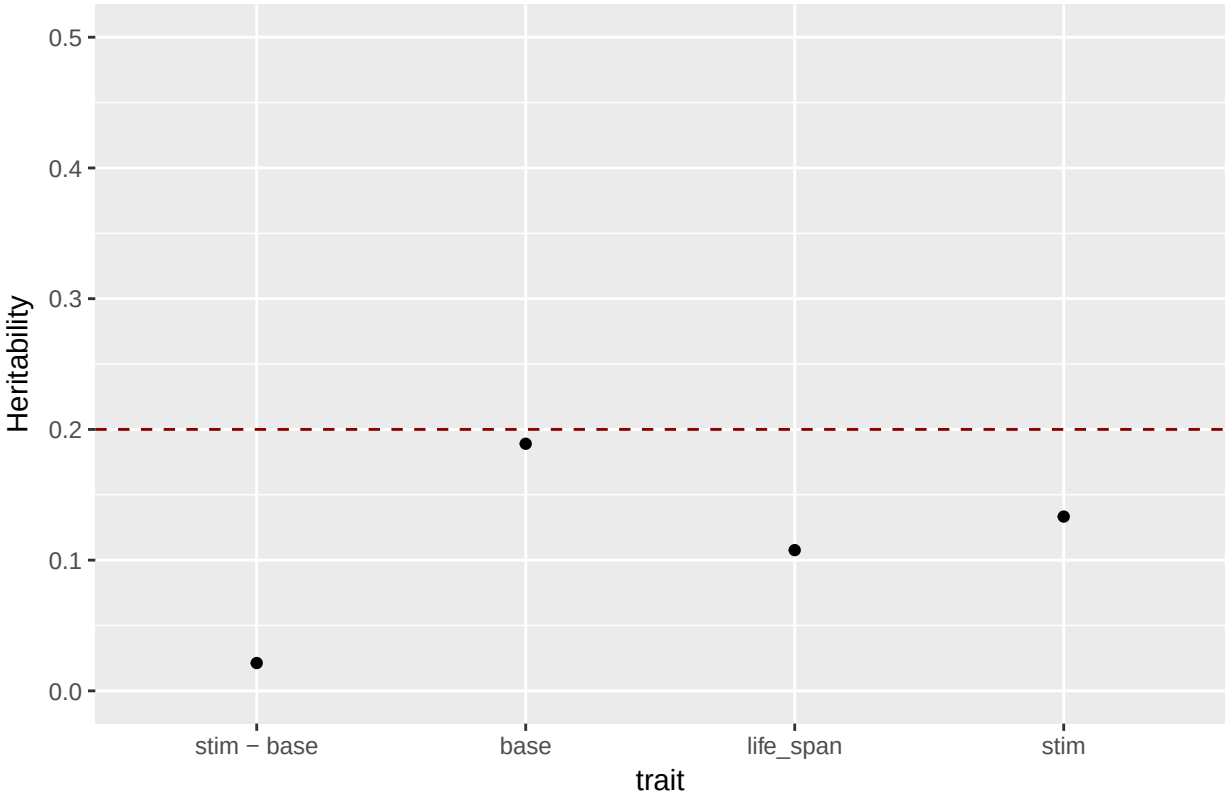
FIGURE 12: Day 15 activity levels (outliers removed)



filtering outliers improves plot, but still no obvious differences among strains.

Calculate heritability for all day 15 activity traits with filtered outliers

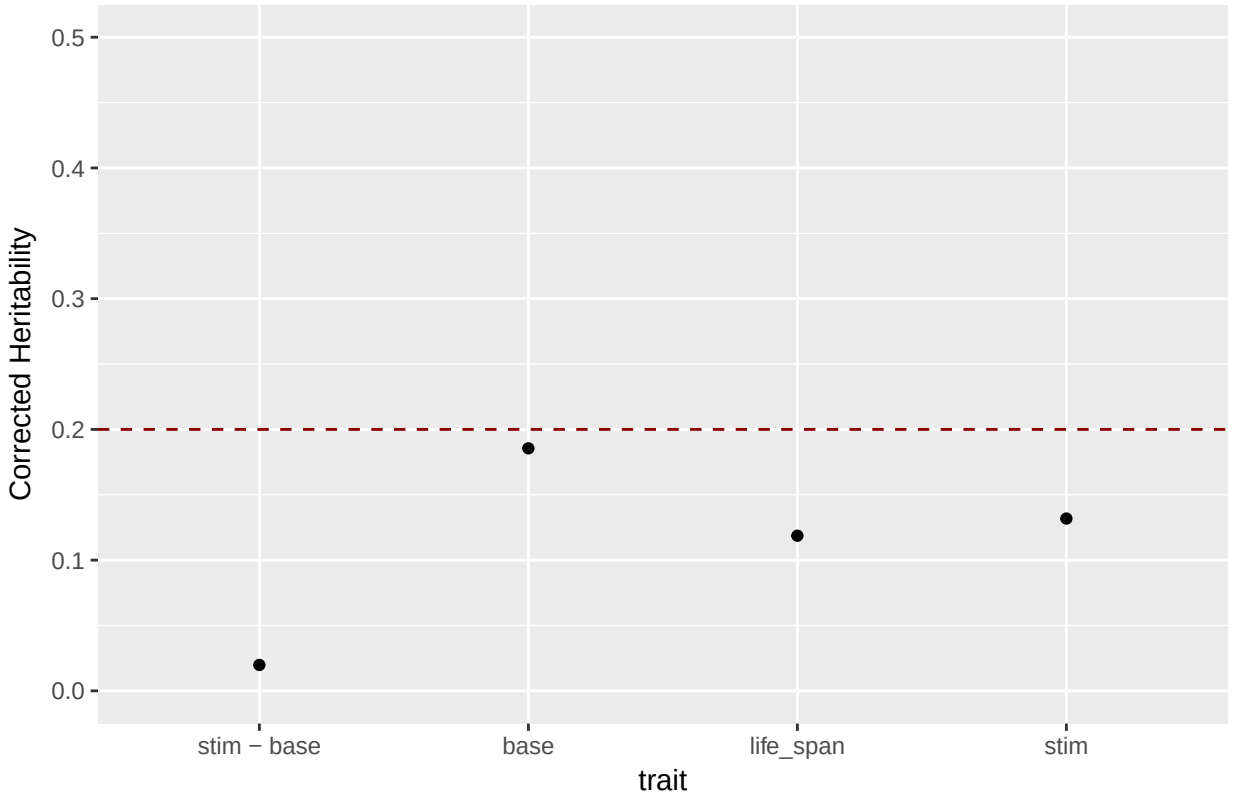
FIGURE 13: Heritability for day 15 activity traits (outliers removed)



Heritability is still low for day 15 traits, although base activity is near threshold of 0.2. I'll try correcting for block effects to see if this improves our heritability estimate.

Removing block effect and recalculating heritability on day 15 activity data

FIGURE 14: Heritability for residual day 15 activity traits (outliers removed)



Heritability is not improved.

ANALYSIS 3 CONCLUSIONS:

1) Heritability is below the 0.2 threshold for all day 15 activity traits (Fig. 14). 2) Correcting for block effect does not improve day 15 activity trait heritabilities, this is probably due to the fact that most of the data are taken from block 1 and block 2, which are most similar to each other (Fig. 12).