

# Ch. 1 - Model comparisons

Wade VanderWright

2022-02-25

## Contents

<b>Model Comparisons</b>	<b>1</b>
Model 1 - CFAR ~ GSH . . . . .	1
Model 2 - CFAR ~ GSH * Max Size . . . . .	3
Model 3 - CFAR ~ GSH + Max Size (No interaction) . . . . .	6
Model 4 - CFAR ~ GSH + (1 Max Size) (No interaction) . . . . .	7
Model 5 - CFAR ~ GSH + Phylogeny . . . . .	9
Model 5 - CFAR ~ GSH * Max Size + Phylogeny . . . . .	12
Model 6 - CFAR ~ GSH + Max Size + Phylogeny . . . . .	15

## Model Comparisons

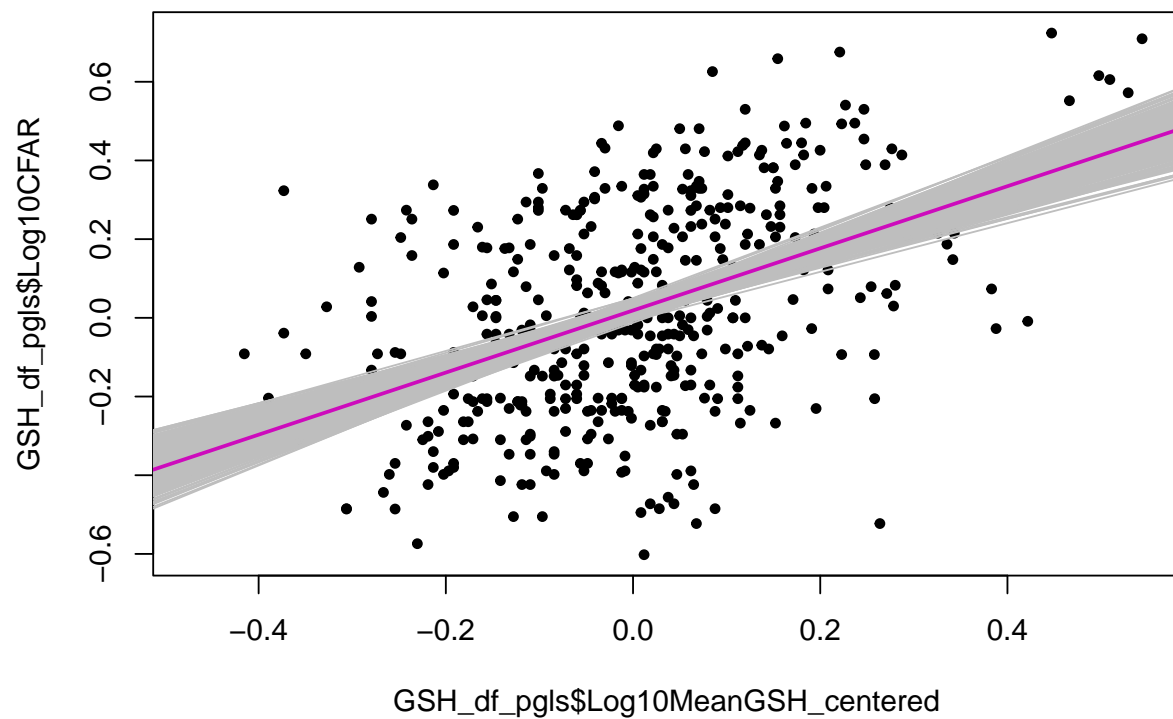
### Model 1 - CFAR ~ GSH

brms model

```
Model_plain <- brm(Log10CFAR ~ Log10MeanGSH_centered,  
  data = GSH_df_pgls,  
  family = gaussian(),  
  prior = Plain_prior  
)
```

```
fixef(Model_plain, summary = T)
```

```
##              Estimate Est.Error        Q2.5        Q97.5  
## Intercept          0.01878724 0.01121401 -0.003048758 0.04068312  
## Log10MeanGSH_centered 0.78943177 0.06909079  0.653733169 0.92180824
```



### STAN model

Data considered

```
data {
  int <lower=1> N;
  vector[N] x;
  vector[N] y;
}
```

The parameters accepted by the model

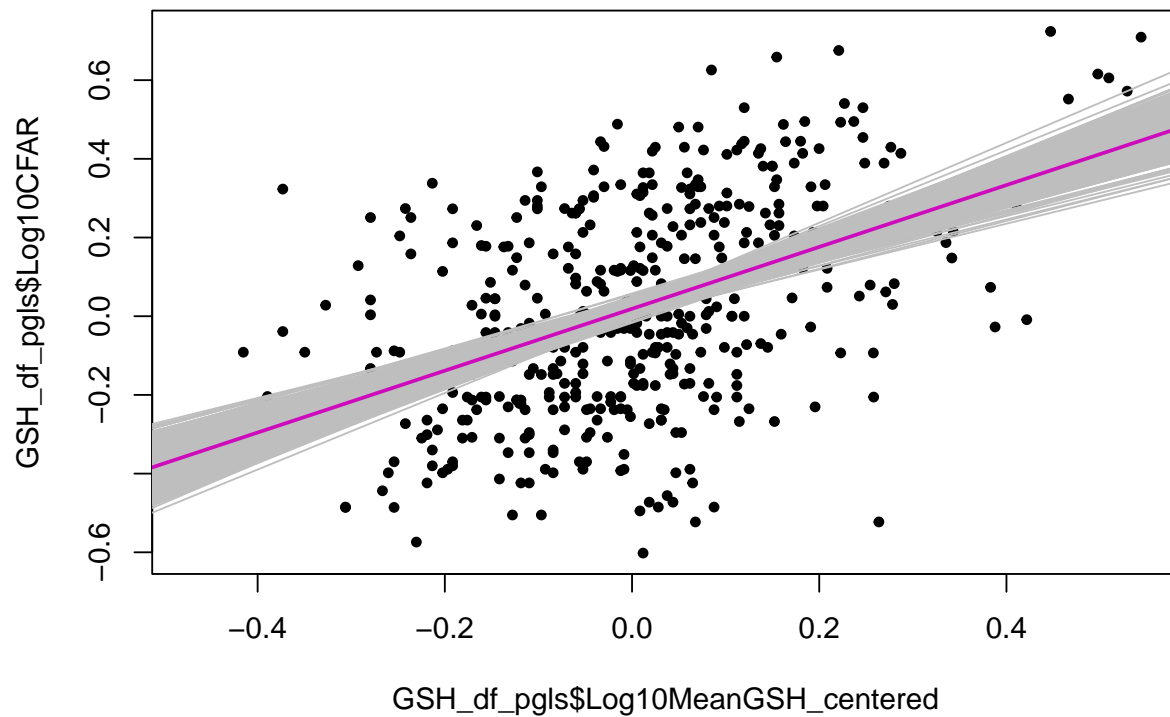
```
parameters {
  real alpha;
  real beta;
  real<lower=0> sigma;
}
```

Model form:

```
model {
  y ~ normal(alpha + x * beta , sigma);
}
```

```
fit1
```

```
## Inference for Stan model: StanModel1.  
## 4 chains, each with iter=1000; warmup=500; thin=1;  
## post-warmup draws per chain=500, total post-warmup draws=2000.  
##  
##           mean se_mean   sd  2.5%  25%  50%  75%  97.5% n_eff Rhat  
## alpha    0.02    0.00 0.01   0.00  0.01  0.02  0.03   0.04 1923   1  
## beta     0.79    0.00 0.07   0.65  0.74  0.78  0.83   0.92 2128   1  
## sigma    0.24    0.00 0.01   0.22  0.23  0.24  0.24   0.25 1798   1  
## lp__    429.83    0.04 1.24 426.61 429.24 430.15 430.74 431.25  859   1  
##  
## Samples were drawn using NUTS(diag_e) at Fri Feb 25 11:41:23 2022.  
## For each parameter, n_eff is a crude measure of effective sample size,  
## and Rhat is the potential scale reduction factor on split chains (at  
## convergence, Rhat=1).
```



Model 2 - CFAR ~ GSH \* Max Size

brms model

```
Model_size <- brm(Log10CFAR ~ Log10MeanGSH_centered * Log10MaxSize_centered,  
  data = GSH_df_pgl,
```

```

family = gaussian(),
prior = Size_prior
)

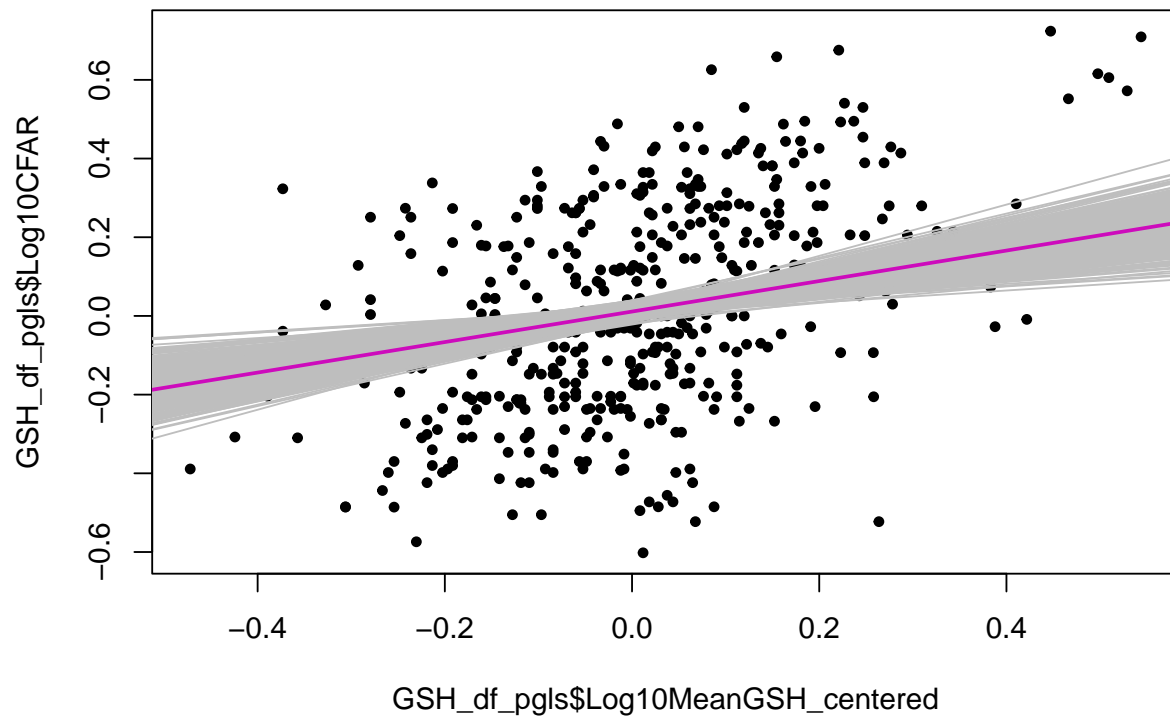
```

```
fixef(Model_size, summary = T)
```

```

##               Estimate  Est.Error    Q2.5
## Intercept      0.01107801 0.01158262 -0.01069365
## Log10MeanGSH_centered 0.38761087 0.08262037 0.22932179
## Log10MaxSize_centered 0.30460627 0.04281846 0.22222482
## Log10MeanGSH_centered:Log10MaxSize_centered 0.26628115 0.16406086 -0.05284798
##               Q97.5
## Intercept      0.03436049
## Log10MeanGSH_centered 0.55212619
## Log10MaxSize_centered 0.38716498
## Log10MeanGSH_centered:Log10MaxSize_centered 0.58368860

```



STAN model

Data considered

```

data {
  int <lower=0> N; // number of data points

```

```

    int <lower=0> K; // number of predictors (2; GSH and Size)
    matrix[N,K] x; // predictor matrix
    vector[N] y; // outcome (CFAR)
}

```

The parameters accepted by the model.

```

parameters {  real alpha;   vector[K] beta;   real<lower=0> sigma; }

```

The model to be estimated. We model the output ‘y’ to be normally distributed with mean ‘mu’ and standard deviation ‘sigma’.

```

model {  y ~ normal(alpha + x * beta , sigma); }

```

```

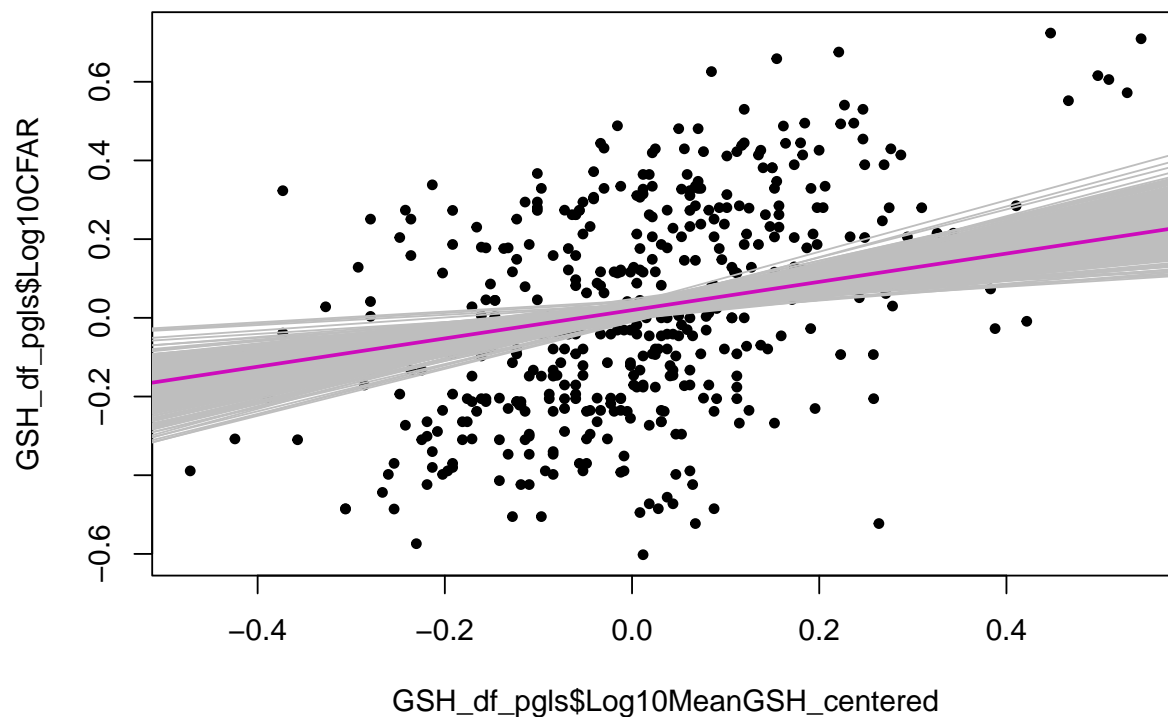
fit2

```

```

## Inference for Stan model: StanModel2.
## 4 chains, each with iter=1000; warmup=500; thin=1;
## post-warmup draws per chain=500, total post-warmup draws=2000.
##
##           mean se_mean   sd  2.5%   25%   50%   75%  97.5% n_eff Rhat
## alpha      0.02    0.00 0.01   0.00   0.01   0.02   0.03   0.04  1991   1
## beta[1]    0.39    0.00 0.08   0.24   0.34   0.39   0.45   0.55  1758   1
## beta[2]    0.33    0.00 0.04   0.24   0.30   0.33   0.35   0.41  1675   1
## sigma      0.22    0.00 0.01   0.21   0.22   0.22   0.23   0.24  1846   1
## lp__      458.62    0.06 1.47  454.89  457.94  458.99  459.69  460.39   668   1
##
## Samples were drawn using NUTS(diag_e) at Fri Feb 25 13:41:48 2022.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).

```



### Model 3 - CFAR ~ GSH + Max Size (No interaction)

brms model

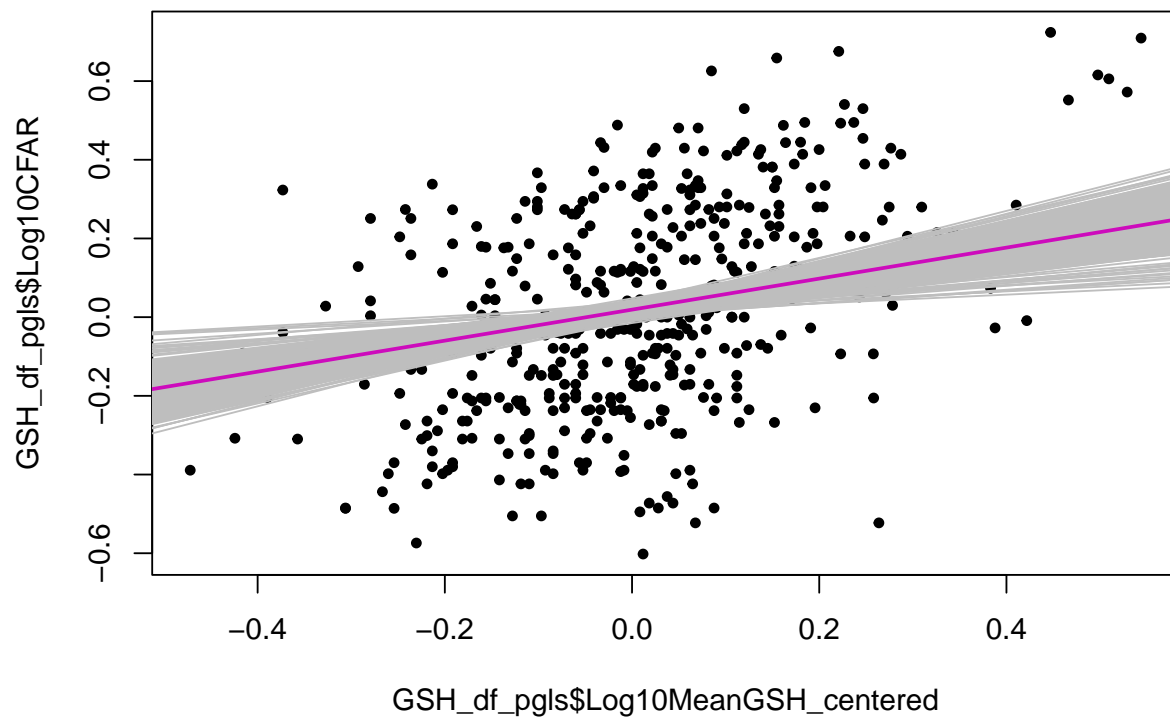
```
Model_size_2 <- brm(Log10CFAR ~ Log10MeanGSH_centered + Log10MaxSize_centered,
  data = GSH_df_pgls,
  family = gaussian(),
  prior = Size_prior_2)
```

```
summary(Model_size_2)
```

```
## Family: gaussian
## Links: mu = identity; sigma = identity
## Formula: Log10CFAR ~ Log10MeanGSH_centered + Log10MaxSize_centered
## Data: GSH_df_pgls (Number of observations: 456)
## Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
## total post-warmup draws = 4000
##
## Population-Level Effects:
##
```

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS
## Intercept	0.02	0.01	-0.00	0.04	1.00	3526
## Log10MeanGSH_centered	0.39	0.08	0.24	0.55	1.00	3567
## Log10MaxSize_centered	0.33	0.04	0.25	0.40	1.00	3528

```
##                               Tail_ESS
## Intercept                     2556
## Log10MeanGSH_centered        2768
## Log10MaxSize_centered        3043
##
## Family Specific Parameters:
##      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma      0.22      0.01   0.21   0.24 1.00   3652   2981
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```



STAN model

TBD

Model 4 - CFAR ~ GSH + (1|Max Size) (No interaction)

brms model

```
Model_size_3 <- brm(Log10CFAR ~ Log10MeanGSH_centered + (1|Log10MaxSize_centered),
  data = GSH_df_pgl,
```

```

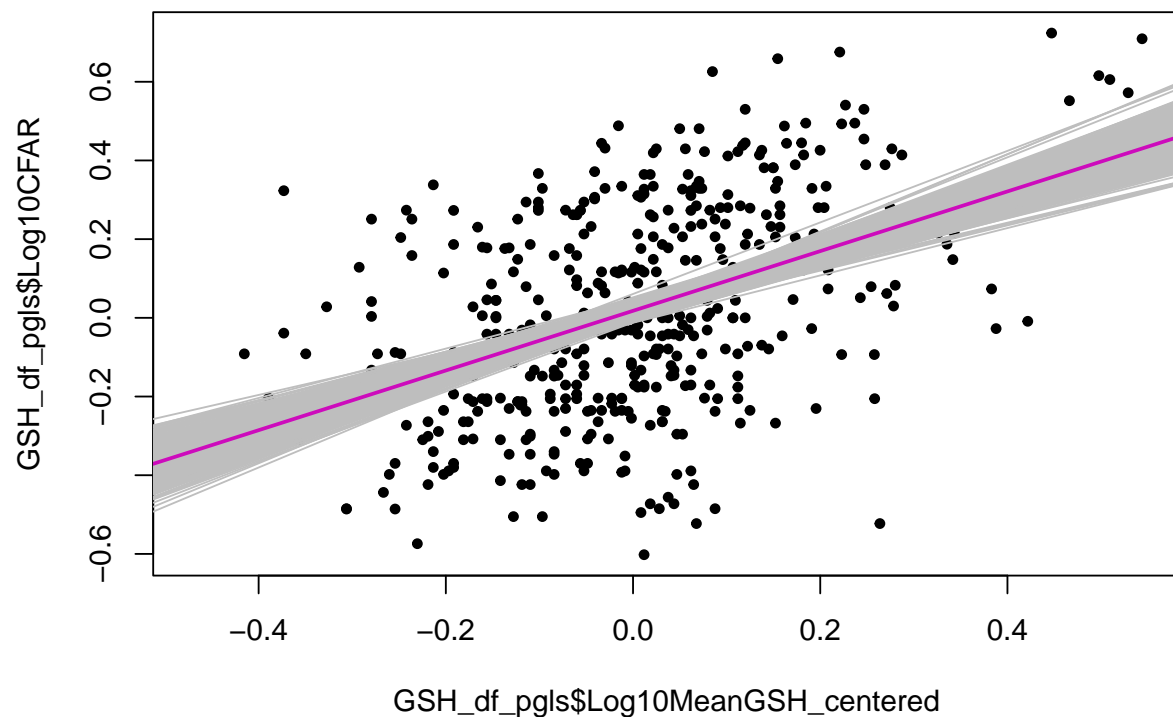
family = gaussian(),
prior = Size_prior_3,
control = list(adapt_delta = 0.9)) #needed to make model converge ...
'''

'''r
summary(Model_size_3)

## Family: gaussian
## Links: mu = identity; sigma = identity
## Formula: Log10CFAR ~ Log10MeanGSH_centered + (1 | Log10MaxSize_centered)
## Data: GSH_df_pglis (Number of observations: 456)
## Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
## total post-warmup draws = 4000
##
## Group-Level Effects:
## ~Log10MaxSize_centered (Number of levels: 222)
## Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept) 0.06 0.03 0.01 0.11 1.01 435 525
##
## Population-Level Effects:
## Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS
## Intercept 0.02 0.01 -0.01 0.04 1.00 5528
## Log10MeanGSH_centered 0.76 0.07 0.62 0.90 1.00 4392
## Tail_ESS
## Intercept 2718
## Log10MeanGSH_centered 2985
##
## Family Specific Parameters:
## Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma 0.23 0.01 0.21 0.25 1.01 917 1516
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).

```





STAN model

TBD

Model 5 - CFAR ~ GSH + Phylogeny

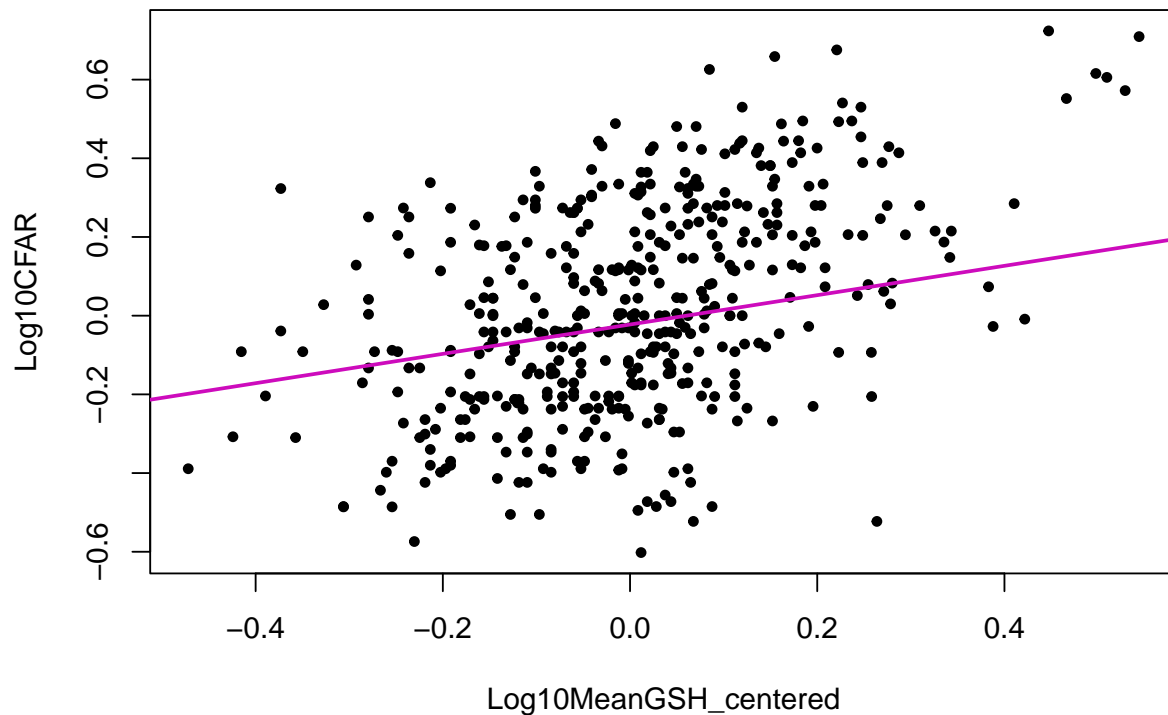
PGLS model

```
pglsMod1 <- gls(Log10CFAR ~ Log10MeanGSH_centered, correlation = corPagel(phy = SampleTree, value = 0.5
```

```
summary(pglsMod1)
```

```
## Generalized least squares fit by maximum likelihood
## Model: Log10CFAR ~ Log10MeanGSH_centered
## Data: GSH_df_pgl
##      AIC      BIC    logLik
## -427.4206 -410.9306 217.7103
##
## Correlation Structure: corPagel
## Formula: ~Binomial
## Parameter estimate(s):
##      lambda
## 0.7030228
```

```
##
## Coefficients:
##               Value Std.Error   t-value p-value
## (Intercept)   -0.0224754 0.06282858 -0.357726  0.7207
## Log10MeanGSH_centered 0.3728837 0.05950418  6.266513  0.0000
##
## Correlation:
##               (Intr)
## Log10MeanGSH_centered -0.065
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -2.7884571 -0.6933649  0.1151372  1.0885641  2.9034051
##
## Residual standard error: 0.2147249
## Degrees of freedom: 456 total; 454 residual
```



brms model

```
Model_simple <- brm(Log10CFAR ~ Log10MeanGSH_centered + (1|gr(Binomial, cov = A)),
  data = GSH_df_pgls,
  family = gaussian(),
  data2 = list(A = A),
  prior = Simple_prior,
```

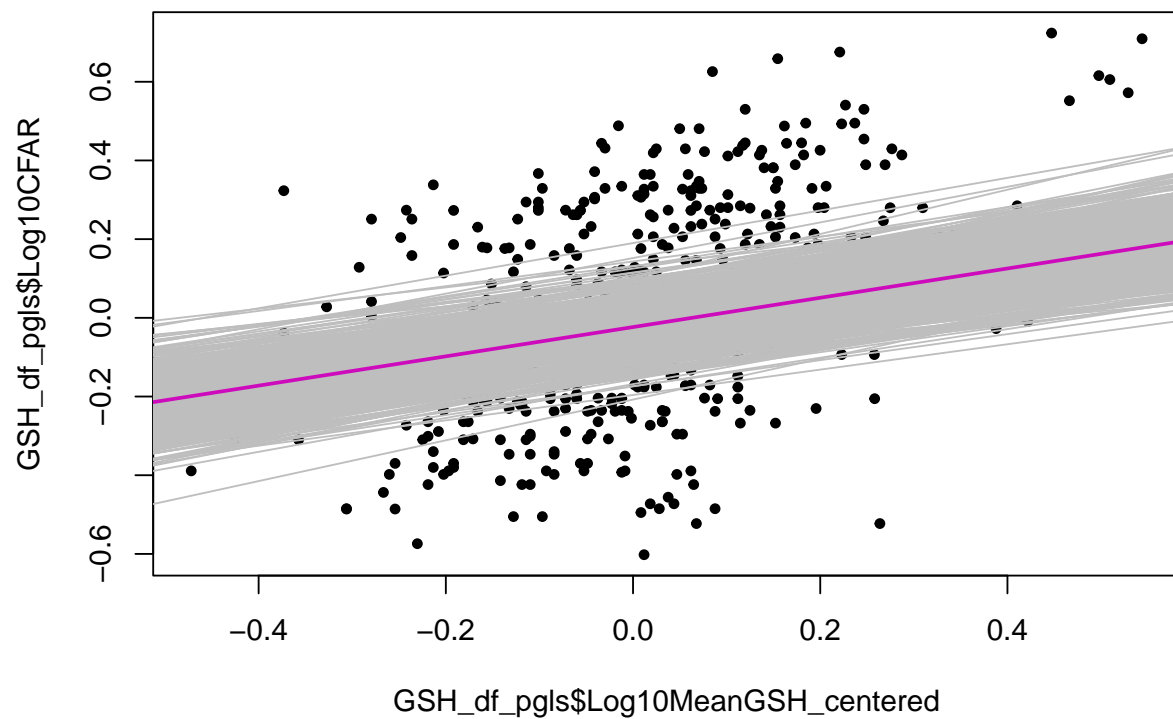
```

cores = 2)
'''

'''r
summary(Model_simple)

## Family: gaussian
## Links: mu = identity; sigma = identity
## Formula: Log10CFAR ~ Log10MeanGSH_centered + (1 | gr(Binomial, cov = A))
## Data: GSH_df_pgls (Number of observations: 456)
## Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
## total post-warmup draws = 4000
##
## Group-Level Effects:
## ~Binomial (Number of levels: 456)
##      Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)      0.01      0.00      0.01      0.01 1.00      1200      1938
##
## Population-Level Effects:
##      Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS
## Intercept          -0.02      0.06     -0.15      0.10 1.00      1257
## Log10MeanGSH_centered  0.37      0.06      0.25      0.49 1.00      4238
##      Tail_ESS
## Intercept          2076
## Log10MeanGSH_centered  3483
##
## Family Specific Parameters:
##      Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma      0.12      0.01      0.11      0.13 1.00      1717      2848
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).

```



STAN model

TBD

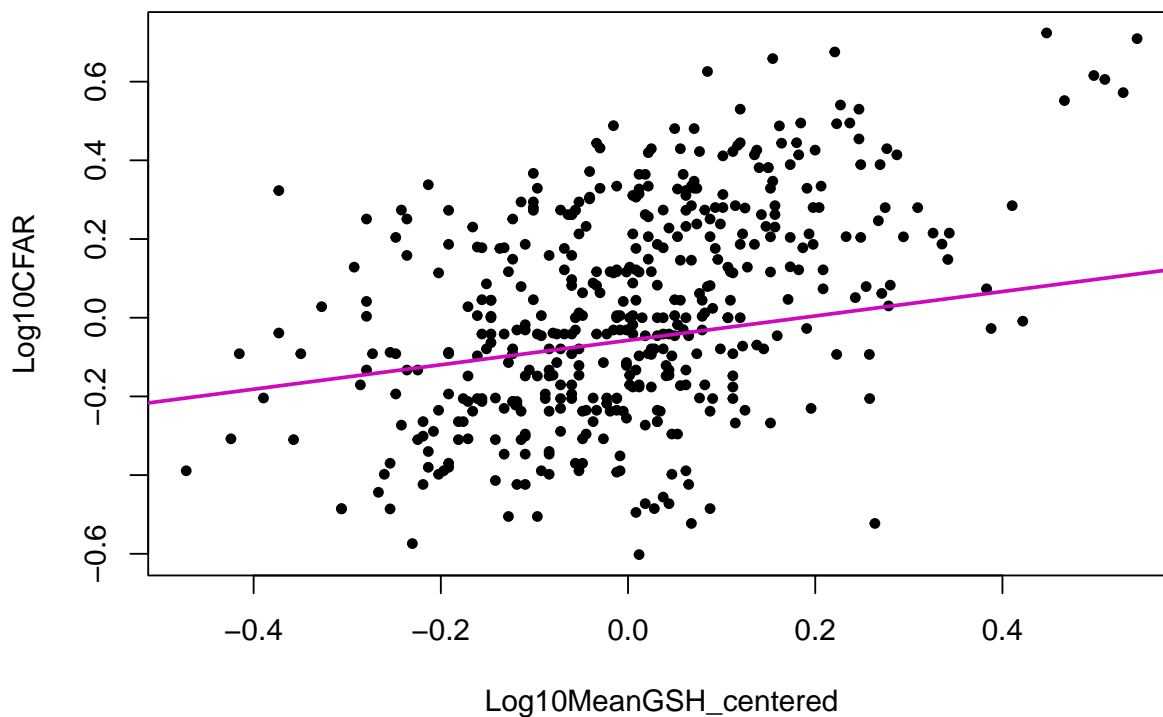
Model 5 - CFAR  $\sim$  GSH \* Max Size + Phylogeny

PGLS model

```
summary(pglMod2)
```

```
## Generalized least squares fit by maximum likelihood
## Model: Log10CFAR ~ Log10MeanGSH_centered * Log10MaxSize_centered
## Data: GSH_df_pgl
##      AIC      BIC    logLik
## -457.6884 -432.9534 234.8442
##
## Correlation Structure: corPagel
## Formula: ~Binomial
## Parameter estimate(s):
##      lambda
## 0.6671208
##
```

```
## Coefficients:
##
##              Value Std.Error   t-value
## (Intercept)    -0.0575841 0.05765480 -0.998773
## Log10MeanGSH_centered    0.3102797 0.05966101  5.200711
## Log10MaxSize_centered    0.1272174 0.03524896  3.609111
## Log10MeanGSH_centered:Log10MaxSize_centered  0.4334147 0.12326258  3.516191
##
##              p-value
## (Intercept)      0.3184
## Log10MeanGSH_centered    0.0000
## Log10MaxSize_centered    0.0003
## Log10MeanGSH_centered:Log10MaxSize_centered  0.0005
##
## Correlation:
##
##              (Intr) Lg10MGSH_ L10MS_
## Log10MeanGSH_centered    -0.053
## Log10MaxSize_centered    -0.025 -0.268
## Log10MeanGSH_centered:Log10MaxSize_centered -0.095  0.030   -0.293
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -2.7953253 -0.5679176  0.2824644  1.2747739  2.9958494
##
## Residual standard error: 0.1999719
## Degrees of freedom: 456 total; 452 residual
```

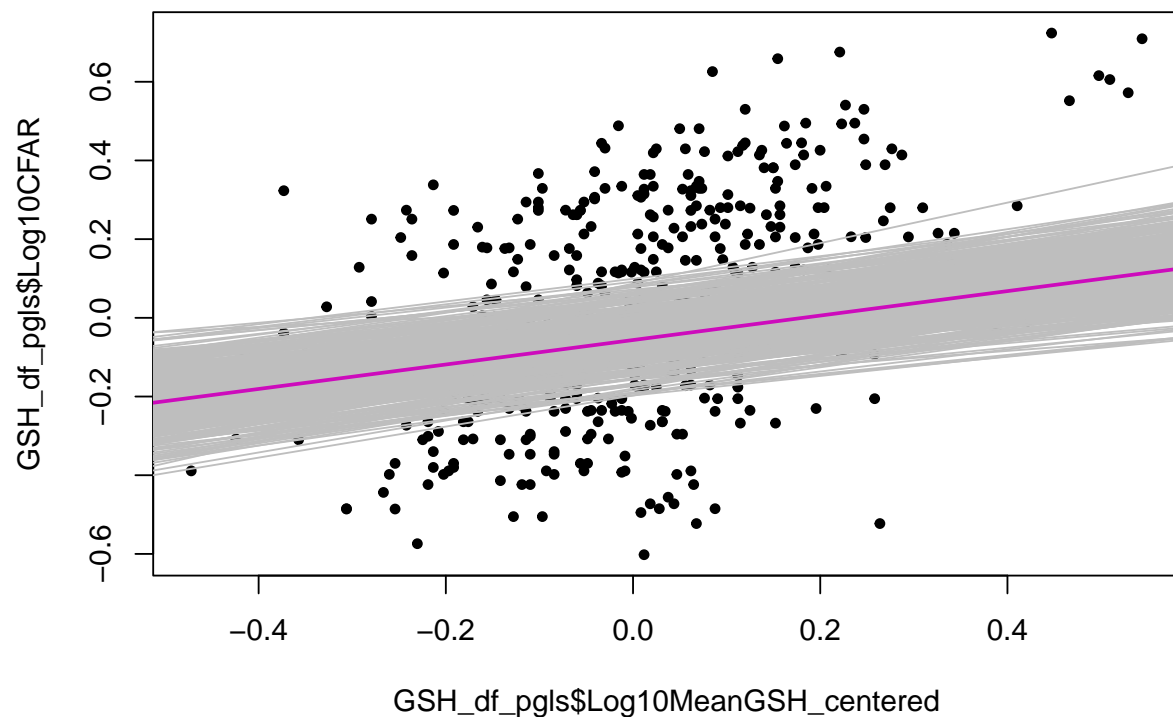


## brms model

```
Model_BS <- brm(Log10CFAR ~ Log10MeanGSH_centered * Log10MaxSize_centered + (1|gr(Binomial, cov = A)),
  data = GSH_df_pgls,
  family = gaussian(),
  data2 = list(A = A),
  prior = BS_prior,
  sample_prior = TRUE, chains = 4, cores = 2)
```

```
summary(Model_BS)
```

```
## Family: gaussian
## Links: mu = identity; sigma = identity
## Formula: Log10CFAR ~ Log10MeanGSH_centered * Log10MaxSize_centered + (1 | gr(Binomial, cov = A))
## Data: GSH_df_pgls (Number of observations: 456)
## Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
## total post-warmup draws = 4000
##
## Group-Level Effects:
## ~Binomial (Number of levels: 456)
##      Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)      0.01      0.00      0.01      0.01 1.00      1303      2376
##
## Population-Level Effects:
##      Estimate Est.Error 1-95% CI
## Intercept      -0.06      0.06    -0.17
## Log10MeanGSH_centered      0.31      0.06      0.19
## Log10MaxSize_centered      0.13      0.04      0.06
## Log10MeanGSH_centered:Log10MaxSize_centered      0.43      0.12      0.20
##      u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept      0.05 1.00      1824      2402
## Log10MeanGSH_centered      0.43 1.00      6595      3478
## Log10MaxSize_centered      0.20 1.00      4866      3342
## Log10MeanGSH_centered:Log10MaxSize_centered      0.67 1.00      5734      3590
##
## Family Specific Parameters:
##      Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma      0.12      0.01      0.11      0.13 1.00      2164      3283
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```



STAN model

TBD

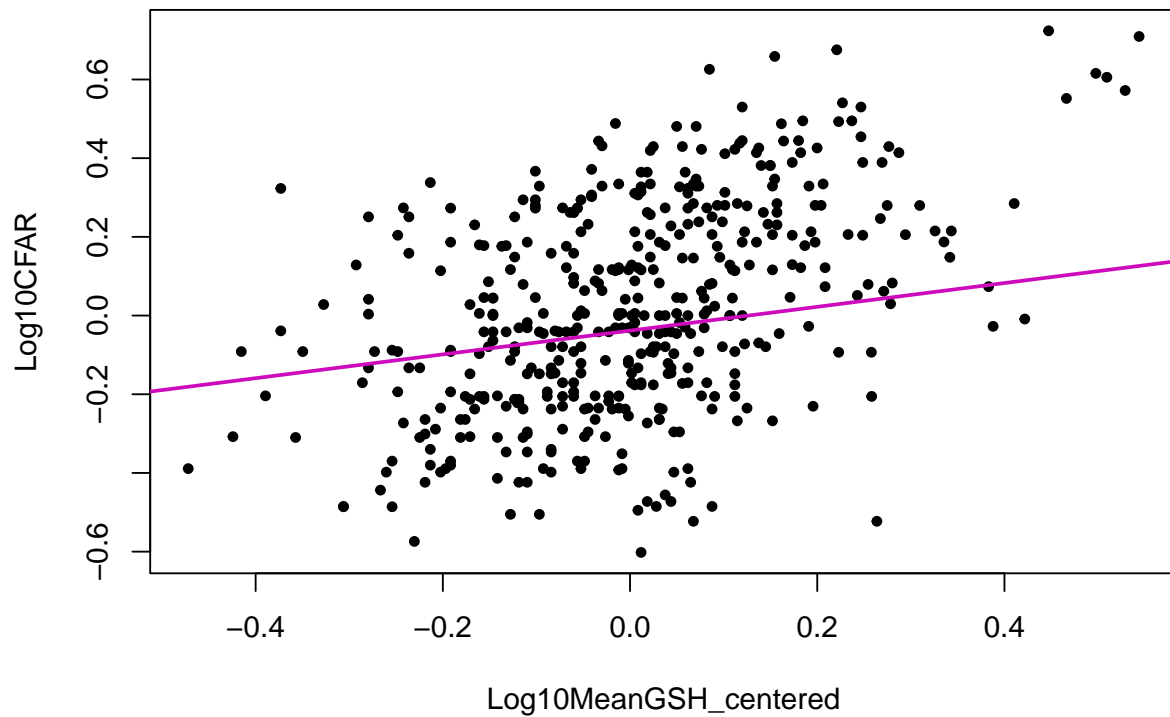
Model 6 - CFAR ~ GSH + Max Size + Phylogeny

PGLS model

```
summary(pglMod3)
```

```
## Generalized least squares fit by maximum likelihood
## Model: Log10CFAR ~ Log10MeanGSH_centered + Log10MaxSize_centered
## Data: GSH_df_pgl
##      AIC      BIC    logLik
## -447.4392 -426.8268 228.7196
##
## Correlation Structure: corPagel
## Formula: ~Binomial
## Parameter estimate(s):
##      lambda
## 0.6786106
##
```

```
## Coefficients:
##               Value Std.Error   t-value p-value
## (Intercept)  -0.03827951 0.05915461 -0.647110  0.5179
## Log10MeanGSH_centered  0.30222706 0.06034147  5.008613  0.0000
## Log10MaxSize_centered  0.16260955 0.03412415  4.765233  0.0000
##
## Correlation:
##               (Intr) L10MGS
## Log10MeanGSH_centered -0.050
## Log10MaxSize_centered -0.056 -0.268
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -2.8353741 -0.5472167  0.2440250  1.1919410  2.8367324
##
## Residual standard error: 0.2047976
## Degrees of freedom: 456 total; 453 residual
```



brms model

TBD



**STAN model**

TBD