

Ch. 1 - Model comparisons

Wade VanderWright

2022-02-25

Contents

Model Comparisons	1
Model 1 - CFAR ~ GSH	1
Model 2 - CFAR ~ GSH * Max Size	4
Model 3 - CFAR ~ GSH + Max Size (No interaction)	7
Model 4 - CFAR ~ GSH + PrimaryHabitat (No interaction)	10
Model 5 - CFAR ~ GSH * PrimaryHabitat	13
Model 6 - CFAR ~ GSH + Phylogeny	16
Model 7 - CFAR ~ GSH * Max Size + Phylogeny	21
Model 8 - CFAR ~ GSH + Max Size + Phylogeny	26
Model 9 - CFAR ~ GSH + PrimaryHabitat + Phylogeny	31
Model 10 - CFAR ~ GSH * PrimaryHabitat + Phylogeny	34
Model 11 - CFAR ~ GSH * MaxSize + PrimaryHabitat + Phylogeny	37

Model Comparisons

Model 1 - CFAR ~ GSH

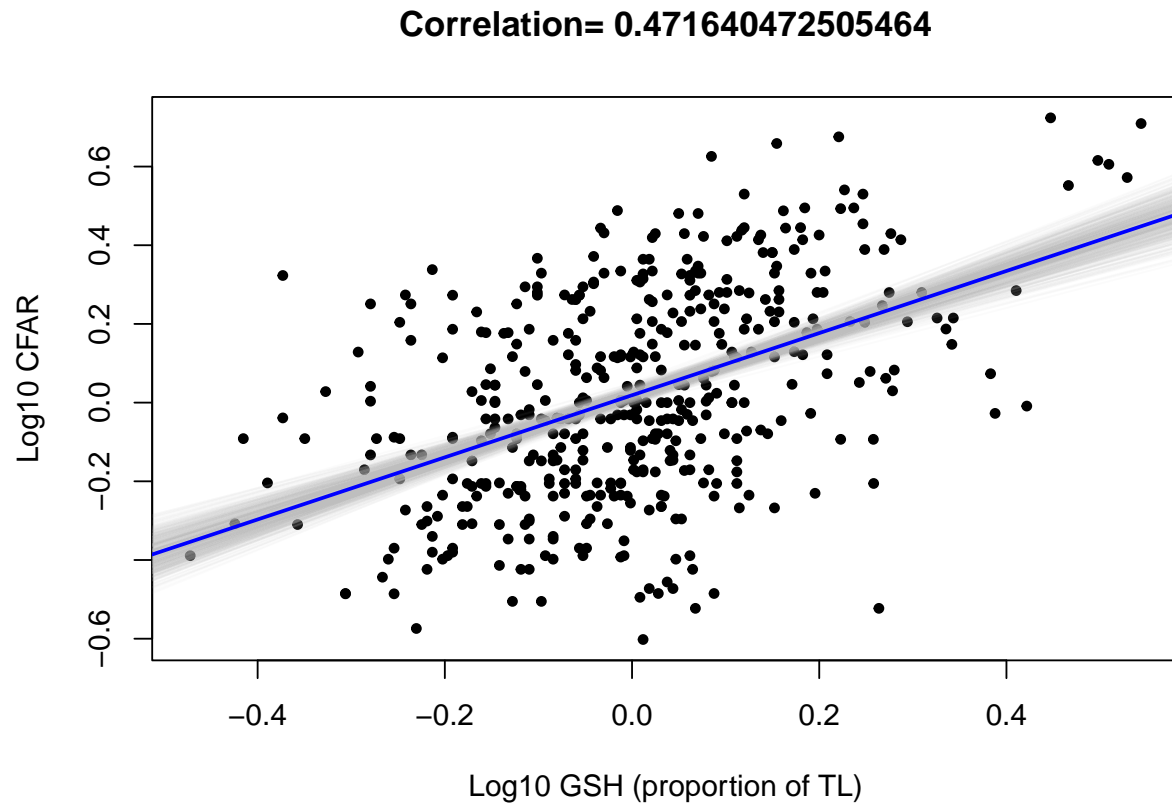
1. brms model

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

```
knitr::kable(fixef(Model_plain, summary = T))
```

	Estimate	Est.Error	Q2.5	Q97.5
Intercept	0.0187872	0.0112140	-0.0030488	0.0406831

	Estimate	Est.Error	Q2.5	Q97.5
Log10MeanGSH_centered	0.7894318	0.0690908	0.6537332	0.9218082



2. 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 {

  sigma ~ student_t(3, 0, 10);

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

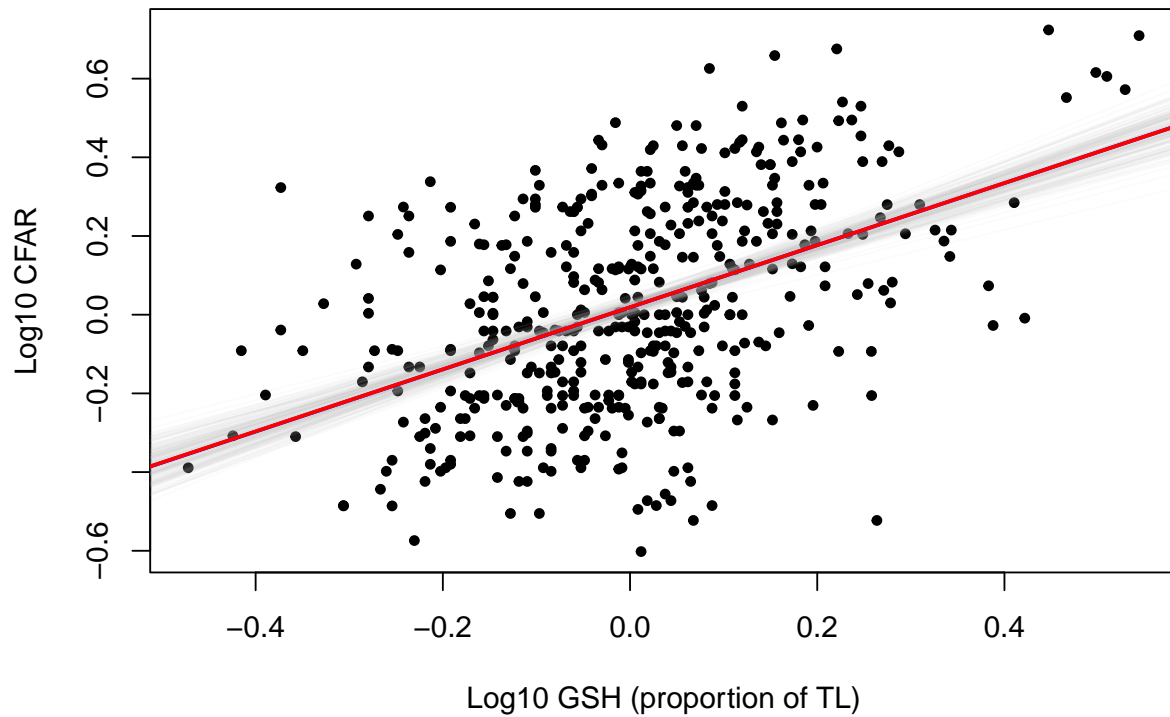
knitr::kable(fit1_summary, caption = 'STAN model summary')

```

Table 2: STAN model summary

	mean	se_mean	sd	2.5%	97.5%	n_eff	Rhat
alpha	0.0190087	0.0002460	0.0111142	-0.0023408	0.0404493	2041.046	1.0004828
beta	0.7889437	0.0017258	0.0719480	0.6467919	0.9279862	1738.128	0.9988377
sigma	0.2357984	0.0001759	0.0077345	0.2210283	0.2509776	1934.501	0.9998605

Correlation= 0.471640472505464



Model 2 - CFAR ~ GSH * Max Size

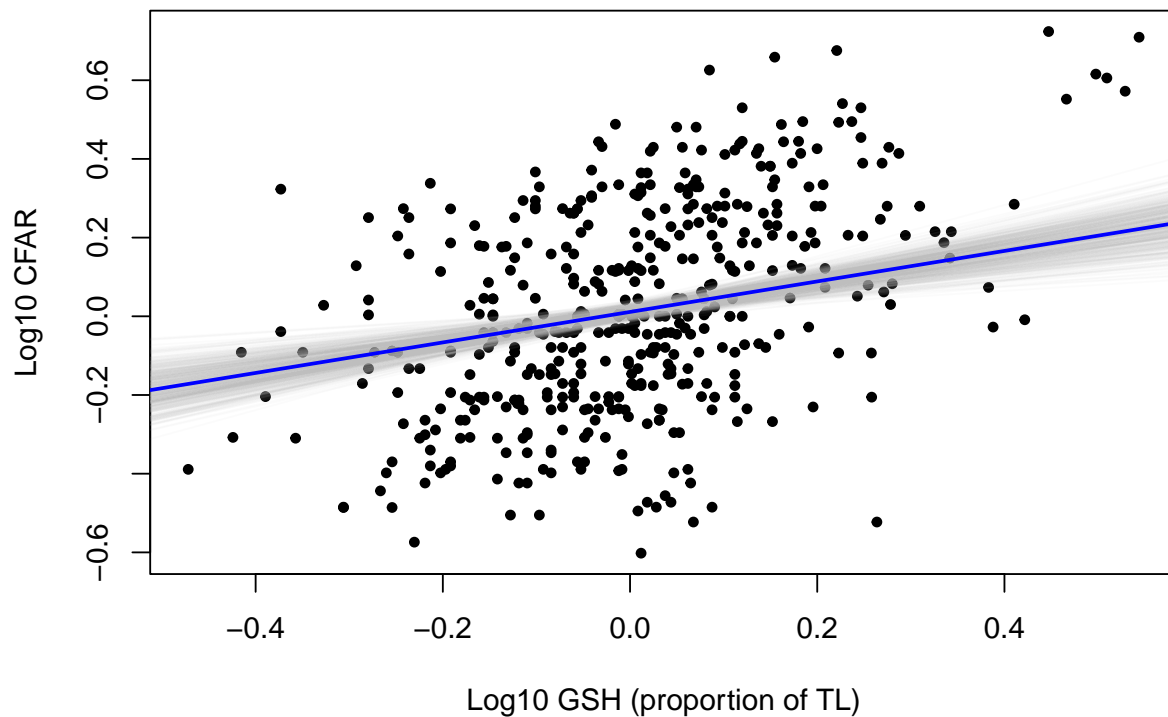
1. brms model

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

Correlation= 0.471640472505464



2. STAN model

Data considered

```
data {  
  int <lower=0> N; // number of data points  
  int <lower=0> K; // number of predictors  
  matrix[N,K] x; // predictor matrix  
  vector[N] y; // 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'.

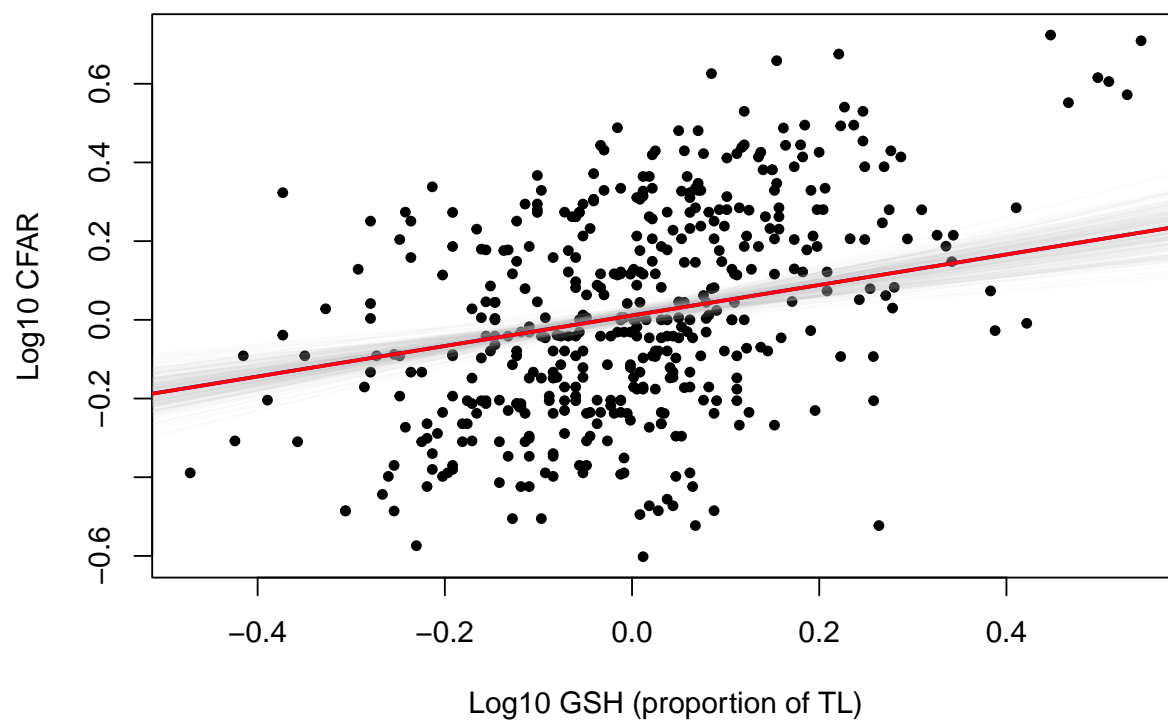
```
model {  
  
  sigma ~ student_t(3, 0, 10);  
  
  y ~ normal(alpha + x * beta , sigma);  
}
```

```
knitr::kable(fit2_summary, caption = 'STAN model summary')
```

Table 3: STAN model summary

	mean	se_mean	sd	2.5%	97.5%	n_eff	Rhat
beta[1]	0.0110471	0.0002538	0.0113730	-0.0103597	0.0340287	2008.373	0.9984008
beta[2]	0.3877210	0.0019674	0.0823295	0.2299960	0.5541936	1751.211	0.9999482
beta[3]	0.3051307	0.0010166	0.0433365	0.2186471	0.3898286	1817.070	1.0004711
beta[4]	0.2555854	0.0037876	0.1580119	-0.0597708	0.5605782	1740.451	0.9990472
sigma	0.2211473	0.0001653	0.0075782	0.2075152	0.2370725	2100.645	0.9996704

Correlation= 0.471640472505464



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

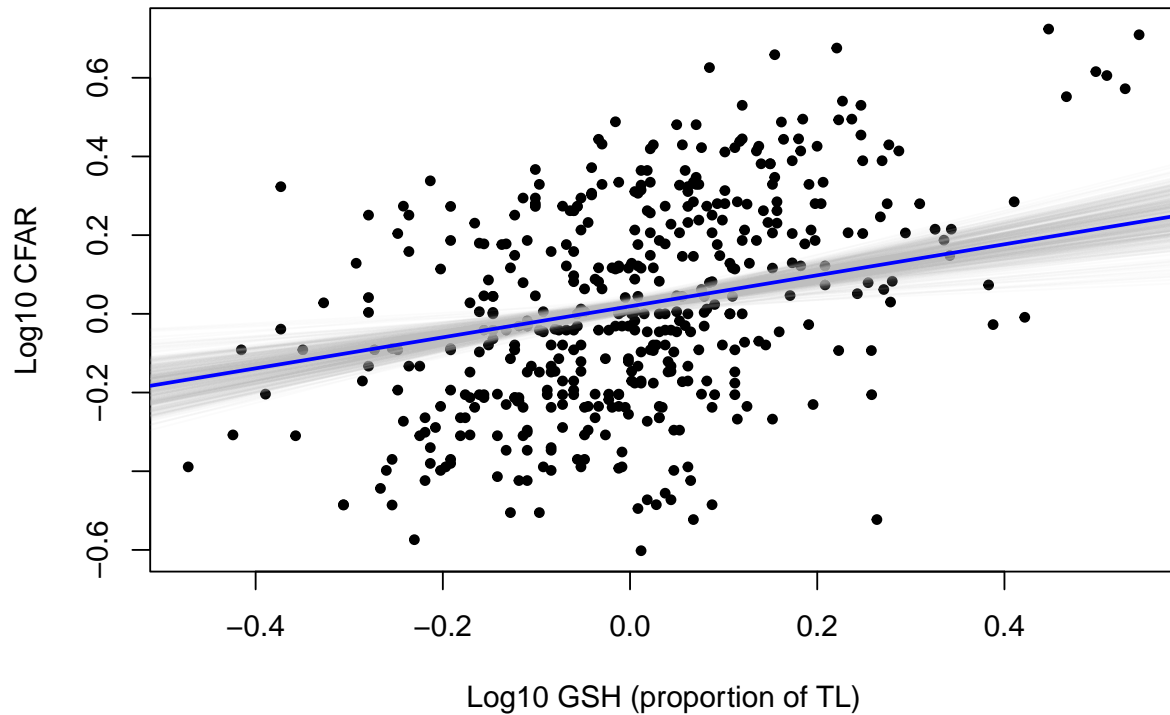
1. brms model

```
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 1-95% 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 1-95% 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).
```

Correlation= 0.471640472505464



2. STAN model

Data considered

```
data {  
  int <lower=0> N; // number of data points  
  int <lower=0> K; // number of predictors  
  matrix[N,K] x; // predictor matrix  
  vector[N] y; // 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 {
```



```

sigma ~ student_t(3, 0, 10);

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

fit3_summary <- summary(fit3, pars = c("beta[1]", "beta[2]", "sigma"), probs = c(0.025, 0.975))$summary

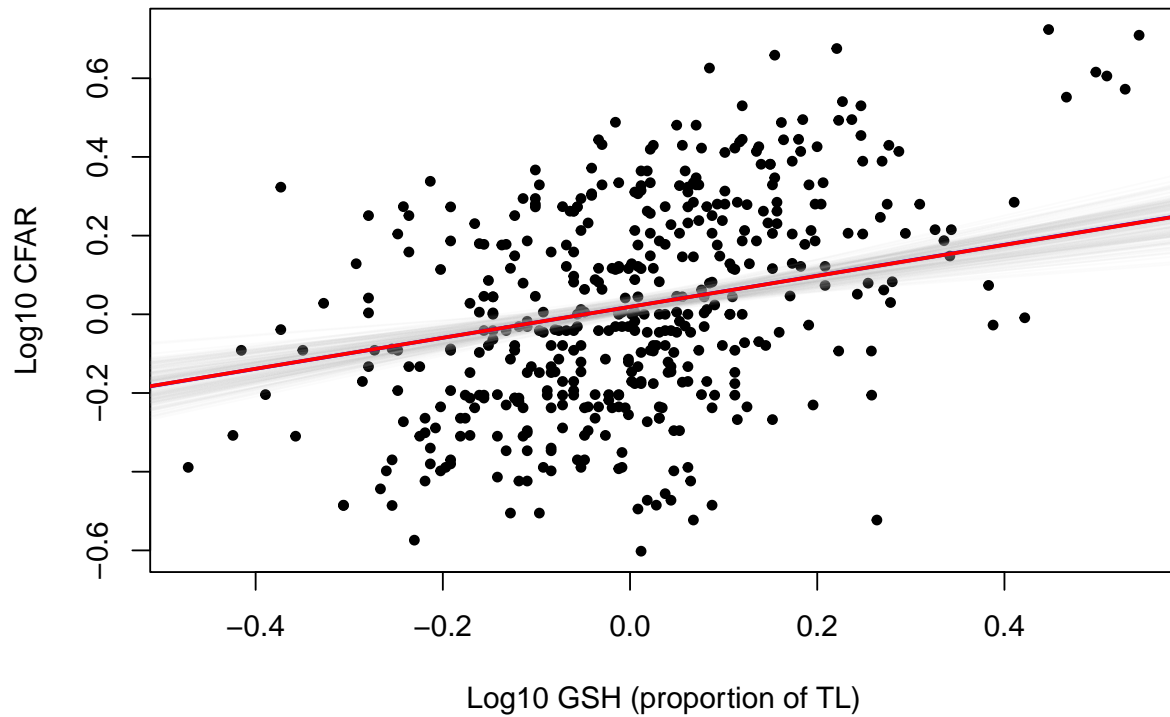
knitr::kable(fit3_summary, caption = 'STAN model summary')

```

Table 4: STAN model summary

	mean	se_mean	sd	2.5%	97.5%	n_eff	Rhat
beta[1]	0.3933554	0.0020241	0.0814803	0.2277471	0.5490099	1620.408	0.9990436
beta[2]	0.3242394	0.0010396	0.0409238	0.2412206	0.4017377	1549.592	1.0000786
sigma	0.2214438	0.0001766	0.0074746	0.2068589	0.2368011	1790.440	1.0004027

Correlation= 0.471640472505464



Model 4 - CFAR ~ GSH + PrimaryHabitat (No interaction)

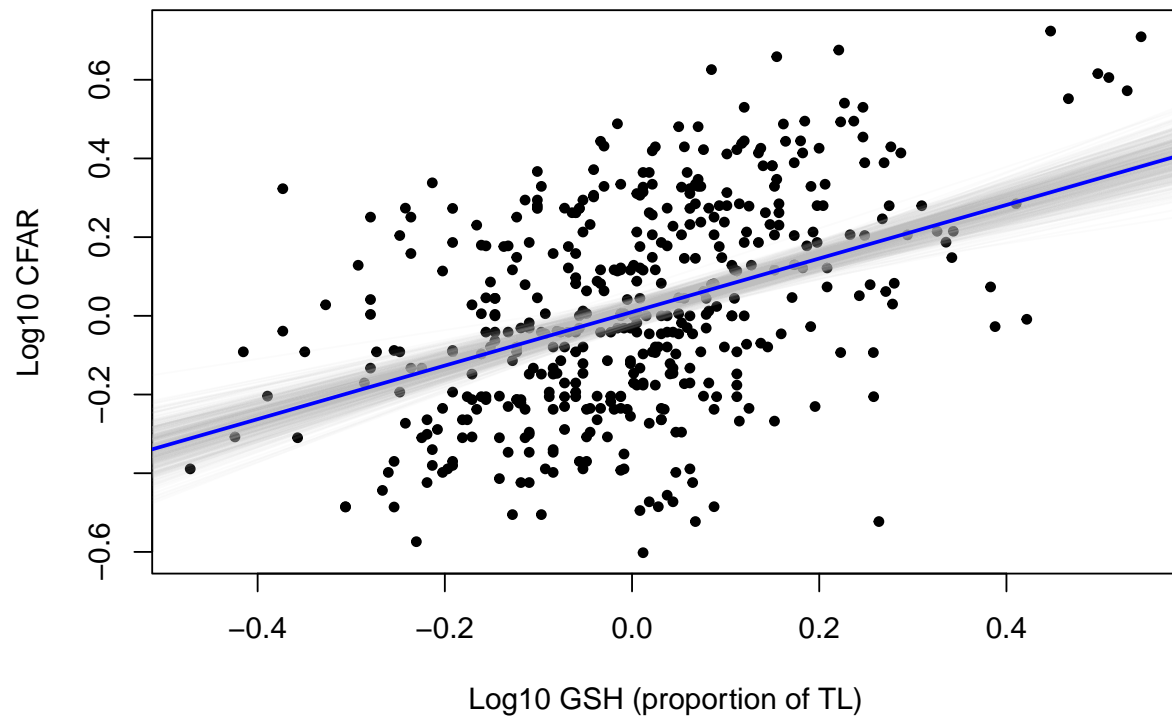
1. brms model

```
habitat_model <- brm(Log10CFAR ~ Log10MeanGSH_centered + PrimaryHabitat,  
  data = GSH_df_pgls,  
  family = gaussian(),  
  prior = habitat_prior,  
  cores = 2)
```

```
summary(habitat_model)
```

```
## Family: gaussian  
## Links: mu = identity; sigma = identity  
## Formula: Log10CFAR ~ Log10MeanGSH_centered + PrimaryHabitat  
## 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 1-95% CI u-95% CI Rhat Bulk_ESS  
## Intercept           0.01      0.02   -0.02    0.04 1.00     4798  
## Log10MeanGSH_centered 0.68      0.07    0.53    0.83 1.00     4837  
## PrimaryHabitatDeepwater -0.01     0.02   -0.06    0.04 1.00     4619  
## PrimaryHabitatPelagic  0.25      0.05    0.15    0.35 1.00     4249  
##           Tail_ESS  
## Intercept           2818  
## Log10MeanGSH_centered 3107  
## PrimaryHabitatDeepwater 2953  
## PrimaryHabitatPelagic  2907  
##  
## Family Specific Parameters:  
##           Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS  
## sigma      0.23      0.01    0.22    0.25 1.00     4398     2810  
##  
## 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).
```

Correlation= 0.471640472505464



2. STAN model

Data considered

```
data {  
  int <lower=0> N; // number of data points  
  int <lower=0> K; // number of predictors  
  matrix[N,K] x; // predictor matrix  
  vector[N] y; // 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 {
```

```

sigma ~ student_t(3, 0, 10);

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

fit6_summary <- summary(fit6, pars = c("beta[1]", "beta[2]", "beta[3]", "beta[4]", "sigma"), probs = c(0.025, 0.975))

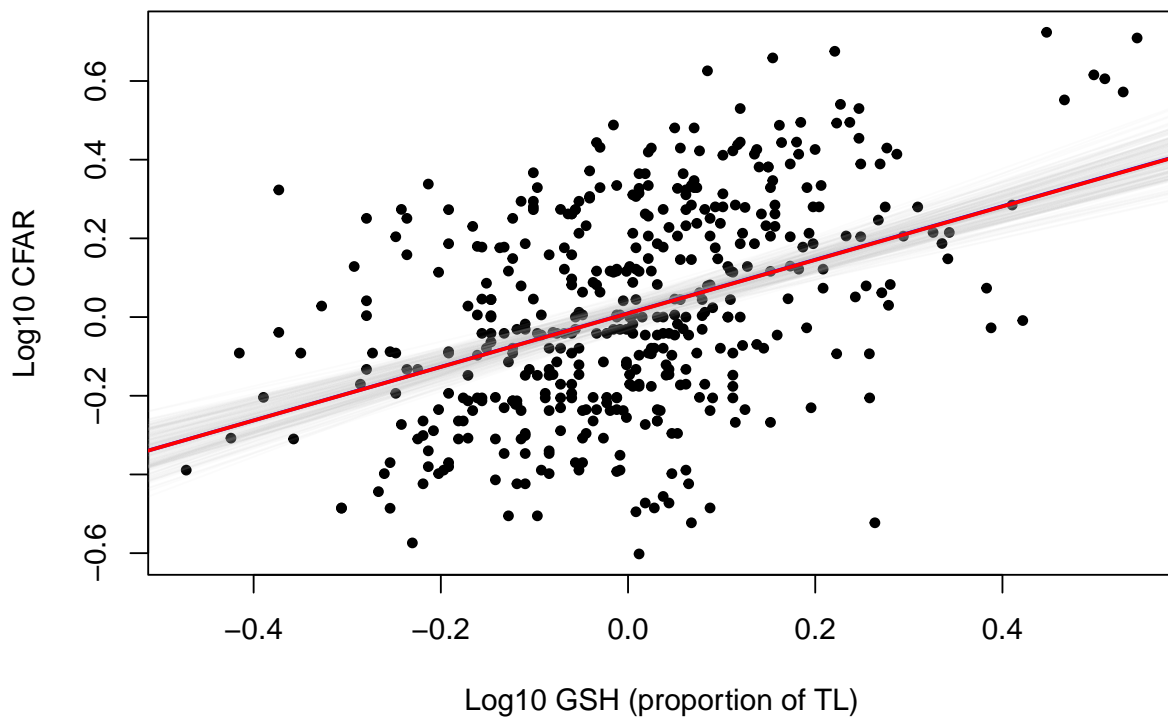
knitr::kable(fit6_summary, caption = 'STAN model summary')

```

Table 5: STAN model summary

	mean	se_mean	sd	5%	95%	n_eff	Rhat
beta[1]	0.0091105	0.0004736	0.0166776	-0.0187022	0.0361588	1239.848	1.0005826
beta[2]	0.6799882	0.0019090	0.0751355	0.5557578	0.8052186	1549.018	0.9991609
beta[3]	-0.0094880	0.0006463	0.0233858	-0.0482268	0.0309939	1309.108	1.0000710
beta[4]	0.2501865	0.0012114	0.0481924	0.1721314	0.3327457	1582.632	1.0007002
sigma	0.2293831	0.0001781	0.0075067	0.2174131	0.2416877	1775.682	1.0003352

Correlation= 0.471640472505464



Model 5 - CFAR ~ GSH * PrimaryHabitat

1. brms model

```
habitat_model_2 <- brm(Log10CFAR ~ Log10MeanGSH_centered * PrimaryHabitat,  
  data = GSH_df_pgls,  
  family = gaussian(),  
  prior = habitat_prior2,  
  cores = 2)
```

```
summary(habitat_model_2)
```

```
## Family: gaussian  
## Links: mu = identity; sigma = identity  
## Formula: Log10CFAR ~ Log10MeanGSH_centered * PrimaryHabitat  
## 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:  
##
```

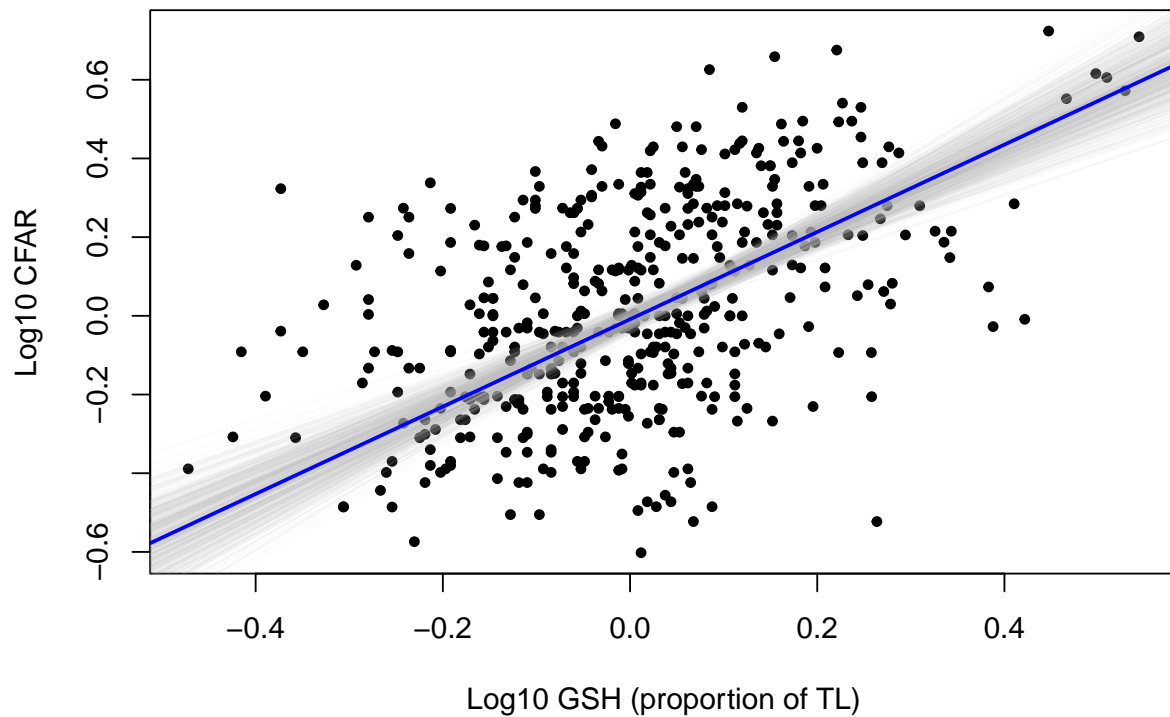
	Estimate	Est.Error	1-95% CI	
## Intercept	-0.01	0.02	-0.04	
## Log10MeanGSH_centered	1.11	0.13	0.85	
## PrimaryHabitatDeepwater	-0.00	0.02	-0.05	
## PrimaryHabitatPelagic	0.29	0.06	0.18	
## Log10MeanGSH_centered:PrimaryHabitatDeepwater	-0.64	0.16	-0.97	
## Log10MeanGSH_centered:PrimaryHabitatPelagic	-0.57	0.23	-1.02	

```
##
```

	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
## Intercept	0.02	1.00	3604	2463
## Log10MeanGSH_centered	1.37	1.00	2189	2148
## PrimaryHabitatDeepwater	0.04	1.00	3648	2715
## PrimaryHabitatPelagic	0.40	1.00	3112	2593
## Log10MeanGSH_centered:PrimaryHabitatDeepwater	-0.33	1.00	2348	2658
## Log10MeanGSH_centered:PrimaryHabitatPelagic	-0.13	1.00	2435	2372

```
##  
## Family Specific Parameters:  
## Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS  
## sigma 0.23 0.01 0.21 0.24 1.00 3424 2768  
##  
## 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).
```

Correlation= 0.471640472505464



2. STAN model

Data considered

```
data {
  int <lower=0> N; // number of data points
  int <lower=0> K; // number of predictors
  matrix[N,K] x; // predictor matrix
  vector[N] y; // 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 {
```

```

sigma ~ student_t(3, 0, 10);

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

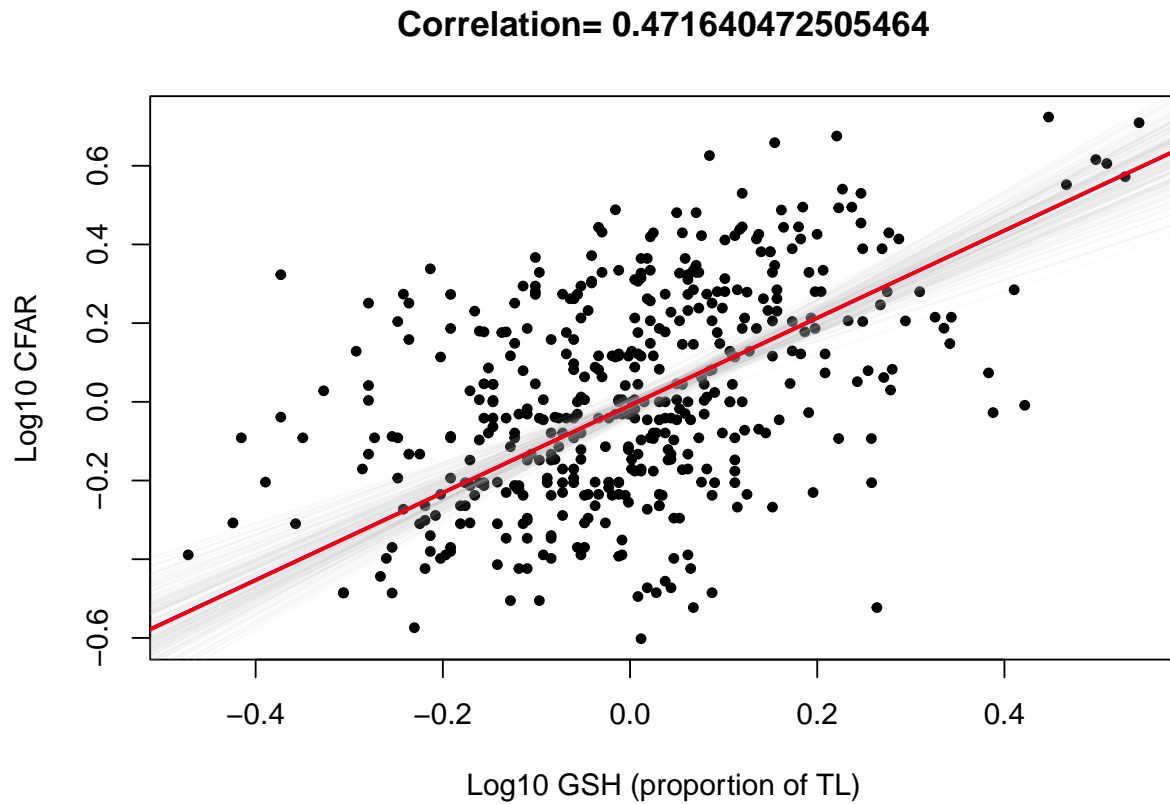
fit7_summary <- summary(fit7, pars = c("beta[1]", "beta[2]", "beta[3]", "beta[4]", "beta[5]", "beta[6]", "sigma"))

knitr::kable(fit7_summary, caption = 'STAN model summary')

```

Table 6: STAN model summary

	mean	se_mean	sd	5%	95%	n_eff	Rhat
beta[1]	-0.0090254	0.0004529	0.0161585	-0.0355365	0.0178685	1272.809	1.000057
beta[2]	1.1110802	0.0037731	0.1225481	0.9043711	1.3095872	1054.926	1.002573
beta[3]	-0.0026746	0.0006063	0.0221765	-0.0390670	0.0330421	1337.945	1.001677
beta[4]	0.2947284	0.0014465	0.0560107	0.2003516	0.3865111	1499.447	1.000261
beta[5]	-0.6441106	0.0047495	0.1571603	-0.8969147	-0.3792308	1094.938	1.002106
beta[6]	-0.5689295	0.0062937	0.2160958	-0.9398260	-0.2136299	1178.923	1.002095
sigma	0.2260172	0.0001645	0.0075462	0.2138852	0.2383921	2105.021	1.000974



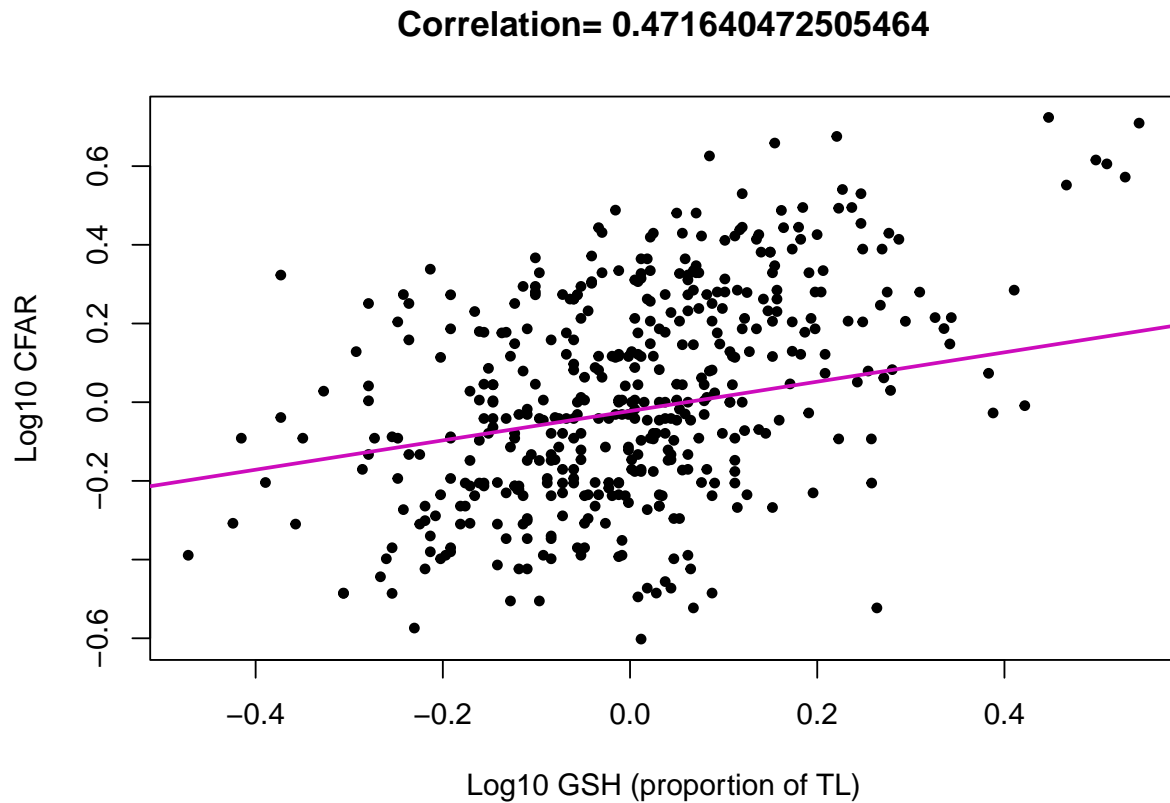
Model 6 - CFAR ~ GSH + Phylogeny

1. PGLS model

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

Table 7: PGLS summary

	Value	Std.Error	t-value	p-value	2.5 %	97.5 %
(Intercept)	-0.0224754	0.0628286	-0.3577261	0.7207146	-0.1456172	0.1006663
Log10MeanGSH_centered	0.3728837	0.0595042	6.2665128	0.0000000	0.2562576	0.4895097



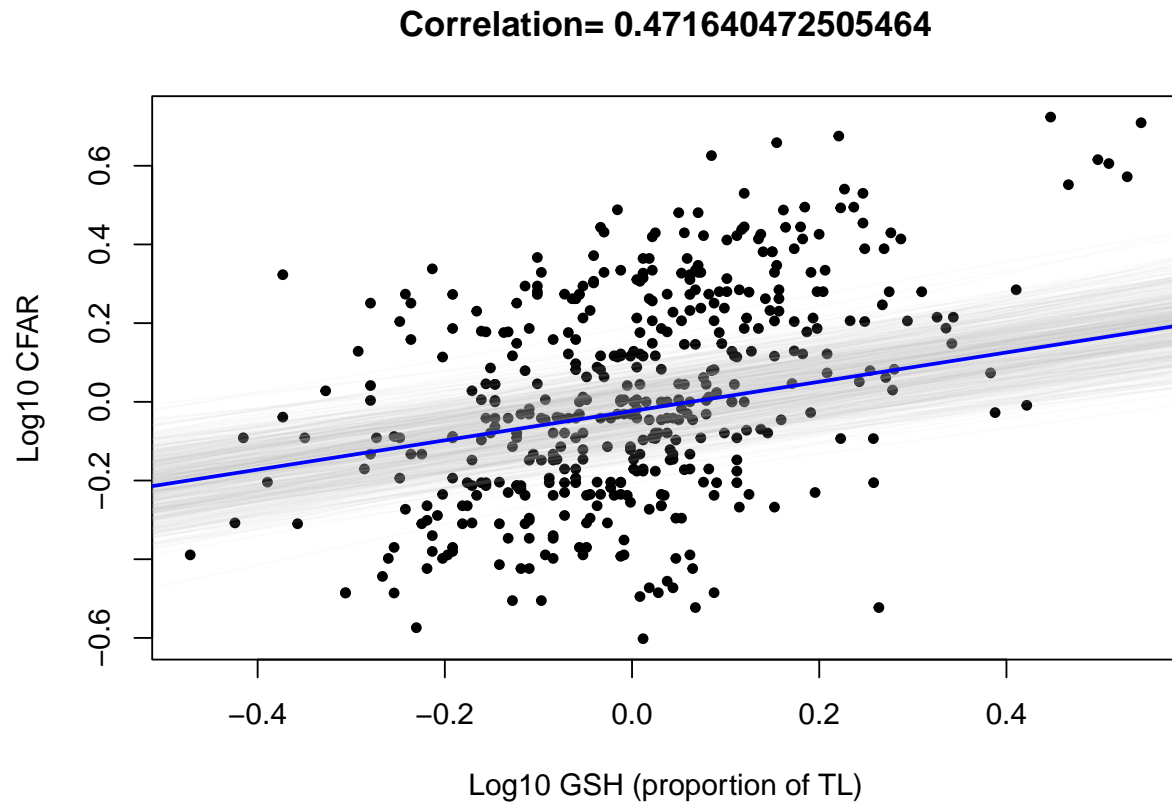
2. brms model

```
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)
```



```
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 l-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 l-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 l-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).
```



3. STAN model

Data considered

```
data {
  int <lower=0> N; // number of data points
  int <lower=0> K; // number of predictors
  matrix[N,K] x; // predictor matrix
  vector[N] y; // outcome (CFAR)
  matrix[N, N] d_mat; // sigma matrix
  matrix[N, N] A; // vcov matrix
}
```

The parameters accepted by the model.

```
parameters {
  real alpha;
  vector[K] beta; // coefficients
  real<lower=0> sigma; // error
  real<lower=0,upper=1> lambda; // phylogenetic signal
}

transformed parameters {
```

```

matrix[N, N] sigma_mat;
matrix[N, N] sigma_total;

vector[N] mu_y;

sigma_mat = (1-lambda)*d_mat + lambda*A;
sigma_total = sigma*sigma_mat;

}

```

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

```

model {

  beta ~ student_t(3, 0, 10);
  lambda ~ uniform(0,1);
  sigma ~ student_t(3, 0, 2.5);

  y ~ multi_normal(alpha + x * beta, sigma_total);
}

```

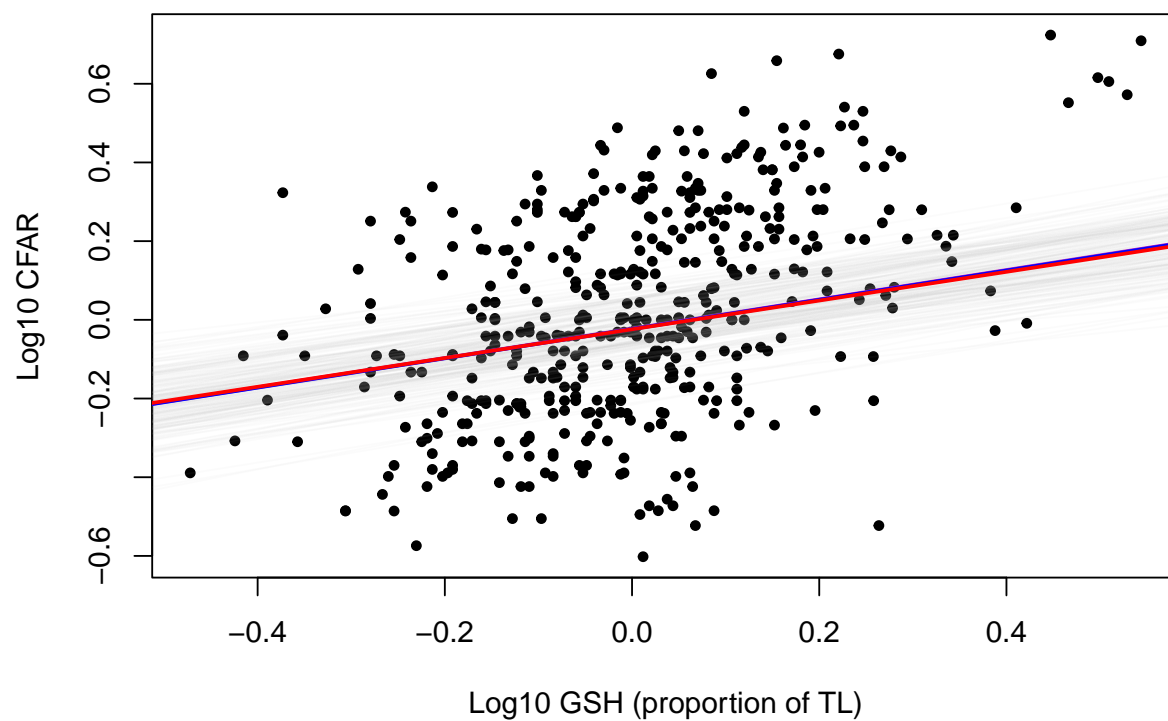
```
fit4
```

```

## Inference for Stan model: StanModel3.
## 4 chains, each with iter=1000; warmup=500; thin=10;
## post-warmup draws per chain=50, total post-warmup draws=200.
##
##           mean se_mean   sd  2.5%   25%   50%   75%  97.5% n_eff Rhat
## alpha    -0.02     0.0 0.06  -0.15  -0.06  -0.03   0.02   0.09  219 0.99
## beta[1]   0.37     0.0 0.06   0.25   0.33   0.36   0.41   0.47  216 1.01
## lp__      625.81     0.1 1.36 622.82 625.03 626.10 626.84 627.47   193 1.02
##
## Samples were drawn using NUTS(diag_e) at Thu Mar 03 14:30:50 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).

```

Correlation= 0.471640472505464



Model 7 - CFAR ~ GSH * Max Size + Phylogeny

1. PGLS model

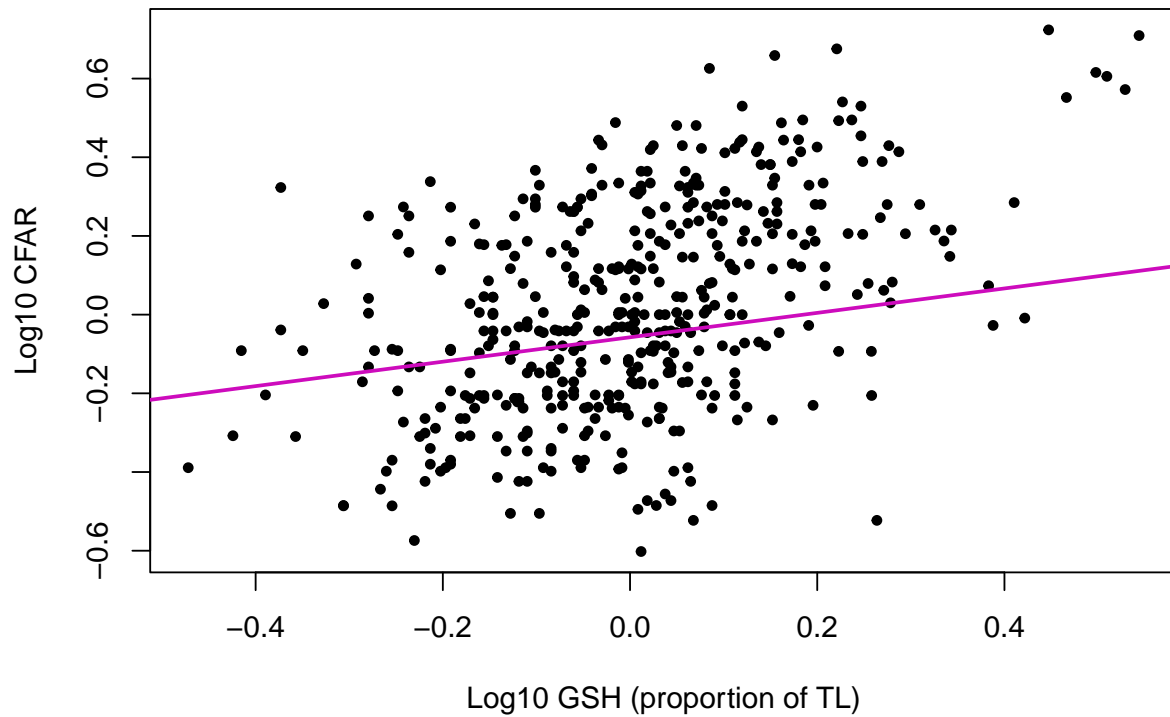
```
pglsMod2 <- gls(Log10CFAR ~ Log10MeanGSH_centered * Log10MaxSize_centered, correlation = corPagel(phy =
```

```
Table_fun(pglsMod2)
```

Table 8: PGLS summary

	Value	Std.Error	t-value	p-value	2.5 %	97.5 %
(Intercept)	-	0.0576548	-	0.3184390	-	0.0554173
	0.0575841		0.998773		0.1705854	
Log10MeanGSH_centered	0.3102797	0.0596610	5.200711	0.0000003	0.1933462	0.4272131
Log10MaxSize_centered	0.1272174	0.0352490	3.609111	0.0003416	0.0581307	0.1963041
Log10MeanGSH_centered:Log10MaxSize_centered	0.4391177	0.1232626	3.516191	0.0004820	0.1918245	0.6750050

Correlation= 0.471640472505464



2. 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)

```

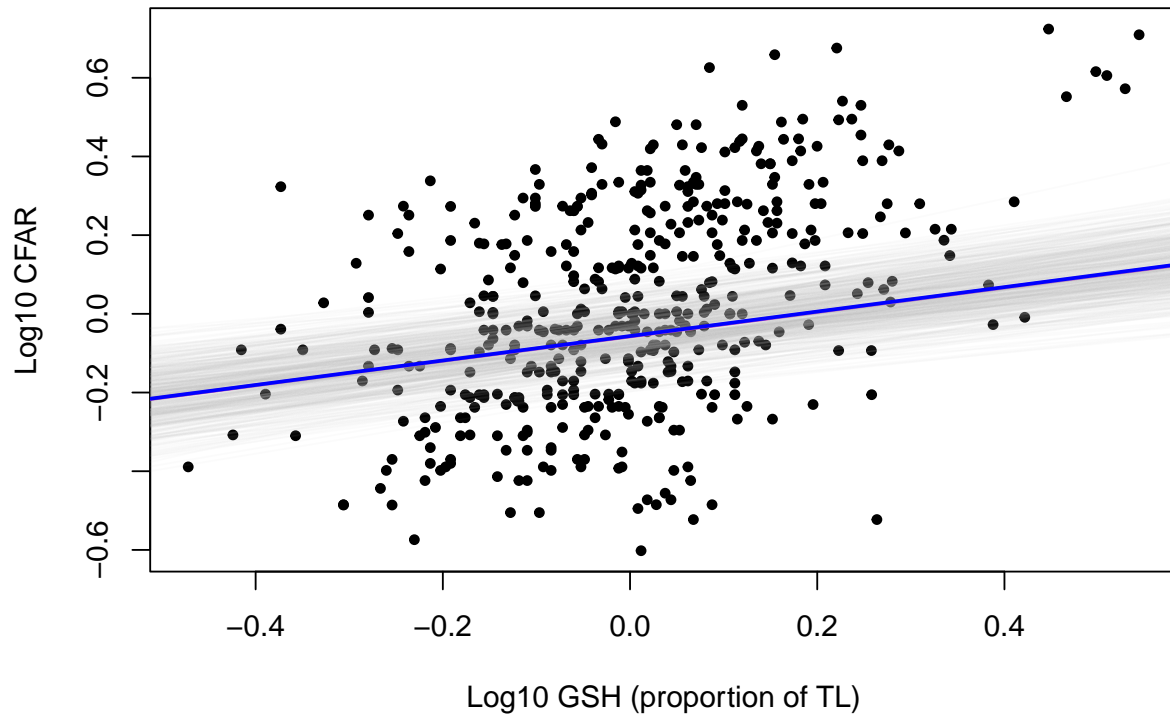
```
summary(Model_BS)
```

```

## 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).

```

Correlation= 0.471640472505464



3. STAN model

Data considered

```
data {
  int <lower=0> N; // number of data points
  int <lower=0> K; // number of predictors
  matrix[N,K] x; // predictor matrix
  vector[N] y; // outcome (CFAR)
  matrix[N, N] d_mat; // sigma matrix
  matrix[N, N] A; // vcov matrix
}
```

The parameters accepted by the model.

```
parameters {
  real alpha;
  vector[K] beta; // coefficients
  real<lower=0> sigma; // error
  real<lower=0,upper=1> lambda; // phylogenetic signal
}

transformed parameters {
```

```

matrix[N, N] sigma_mat;
matrix[N, N] sigma_total;

vector[N] mu_y;

sigma_mat = (1-lambda)*d_mat + lambda*A;
sigma_total = sigma*sigma_mat;

}

```

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

```

model {

  beta ~ student_t(3, 0, 10);
  lambda ~ uniform(0,1);
  sigma ~ student_t(3, 0, 2.5);

  y ~ multi_normal(alpha + x * beta, sigma_total);
}

```

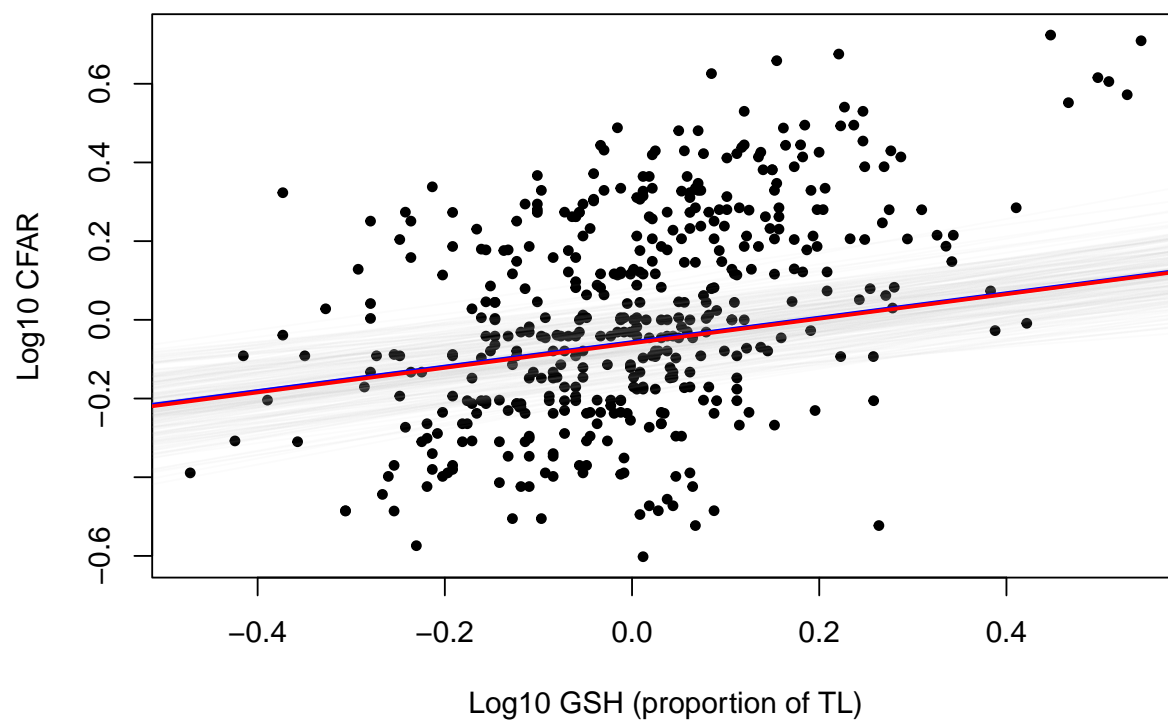
```
fit5
```

```

## Inference for Stan model: StanModel3.
## 4 chains, each with iter=1000; warmup=500; thin=10;
## post-warmup draws per chain=50, total post-warmup draws=200.
##
##           mean se_mean   sd  2.5%  25%   50%   75%  97.5% n_eff Rhat
## alpha    -0.06    0.00 0.06  -0.17 -0.10 -0.06 -0.02   0.05  154 1.01
## beta[1]   0.31    0.00 0.06   0.20  0.26  0.32  0.35   0.44  272 0.99
## beta[2]   0.12    0.00 0.04   0.06  0.10  0.12  0.15   0.20  194 0.99
## beta[3]   0.43    0.01 0.12   0.22  0.35  0.44  0.51   0.65  210 0.99
## lp__      641.68    0.12 1.76 637.09 640.81 642.00 643.00 643.99   199 1.00
##
## Samples were drawn using NUTS(diag_e) at Mon Mar 07 12:09:07 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).

```


Correlation= 0.471640472505464



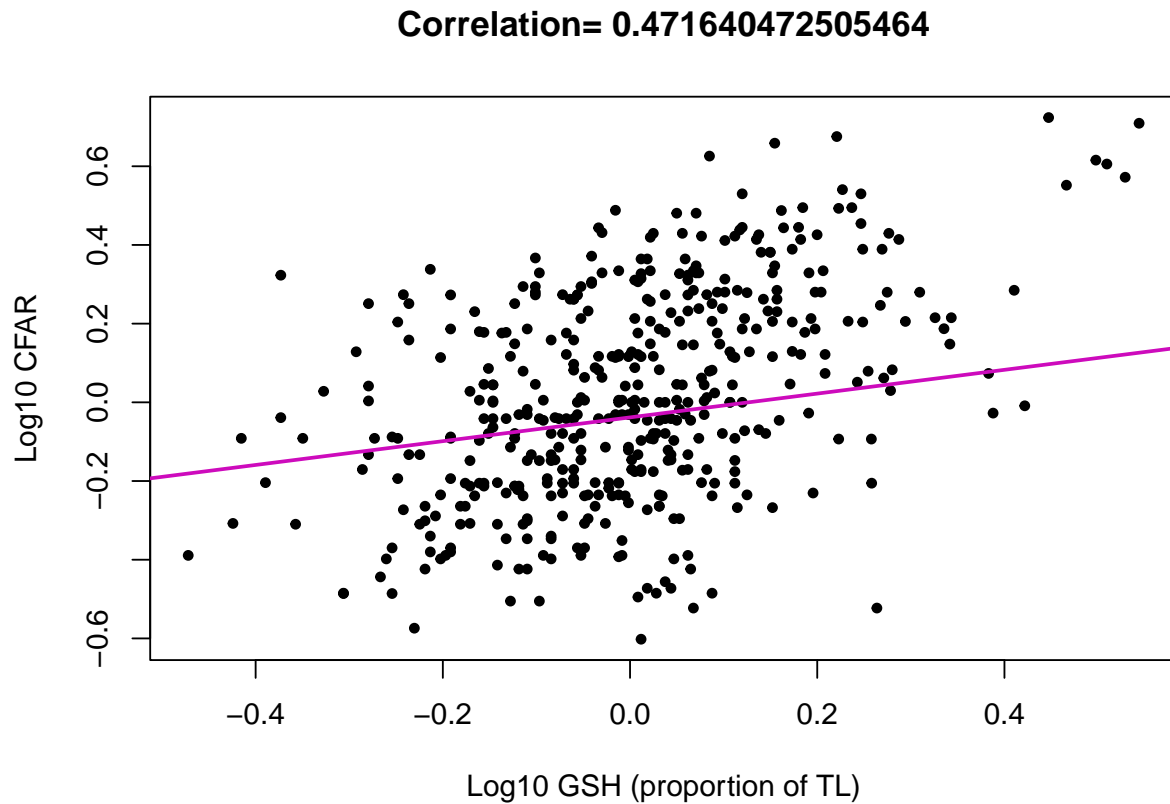
Model 8 - CFAR ~ GSH + Max Size + Phylogeny

1. PGLS model

```
pglsMod3 <- gls(Log10CFAR ~ Log10MeanGSH_centered + Log10MaxSize_centered, correlation = corPagel(phy =  
Table_fun(pglsMod3)
```

Table 9: PGLS summary

	Value	Std.Error	t-value	p-value	2.5 %	97.5 %
(Intercept)	-0.0382795	0.0591546	-0.6471096	0.5178888	-0.1542204	0.0776614
Log10MeanGSH_centered	0.3022271	0.0603415	5.0086129	0.0000008	0.1839600	0.4204942
Log10MaxSize_centered	0.1626095	0.0341242	4.7652331	0.0000025	0.0957274	0.2294917



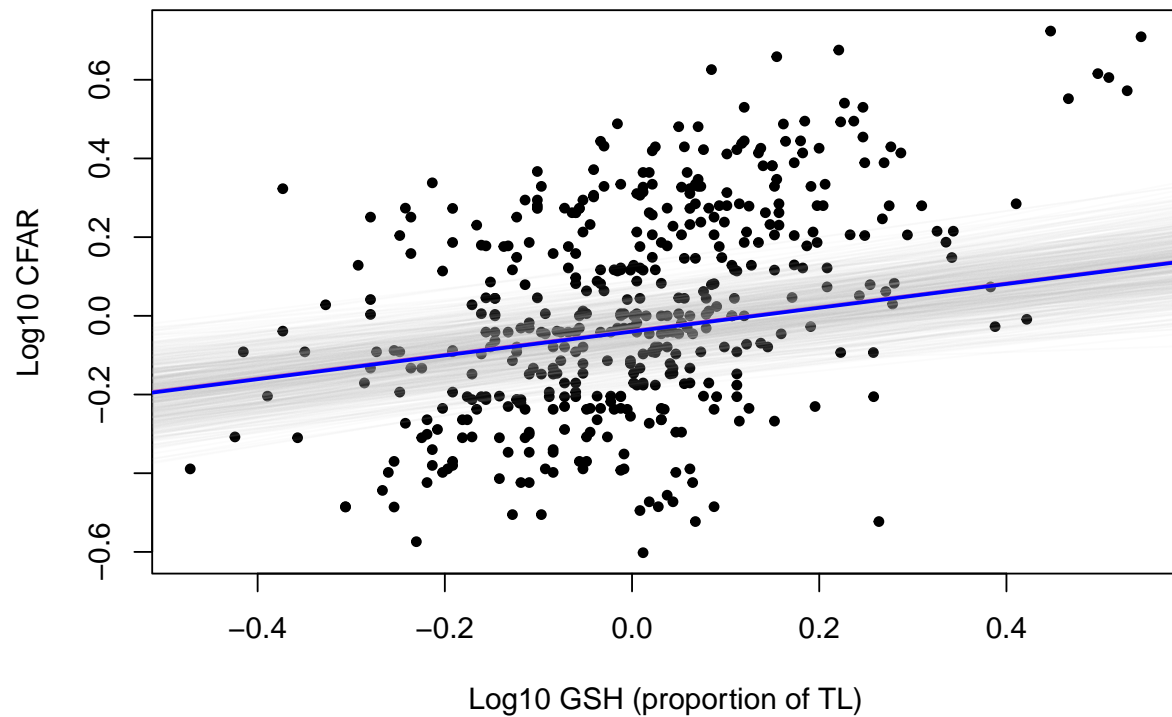
2. brms model

```
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_BS2)
```

```
## Family: gaussian
## Links: mu = identity; sigma = identity
## Formula: Log10CFAR ~ Log10MeanGSH_centered + Log10MaxSize_centered + (1 | gr(Binomial, cov = A))
## 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:
## ~Binomial (Number of levels: 456)
##      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)      0.01      0.00      0.01      0.01 1.00      1186      1678
##
## Population-Level Effects:
##      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## Intercept          -0.04      0.06     -0.16      0.08 1.00      1283
## Log10MeanGSH_centered    0.30      0.06      0.19      0.42 1.00      3622
## Log10MaxSize_centered    0.16      0.03      0.10      0.23 1.00      3812
##
##      Tail_ESS
## Intercept          2054
## Log10MeanGSH_centered    3227
## Log10MaxSize_centered    3414
##
## Family Specific Parameters:
##      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma      0.12      0.01      0.11      0.13 1.00      1686      2672
##
## 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).
```

Correlation= 0.471640472505464



3. STAN model

Data considered

```
data {
  int <lower=0> N; // number of data points
  int <lower=0> K; // number of predictors
  matrix[N,K] x; // predictor matrix
  vector[N] y; // outcome (CFAR)
  matrix[N, N] d_mat; // sigma matrix
  matrix[N, N] A; // vcov matrix
}
```

The parameters accepted by the model.

```
parameters {
  real alpha;
  vector[K] beta; // coefficients
  real<lower=0> sigma; // error
  real<lower=0,upper=1> lambda; // phylogenetic signal
}
```

```
transformed parameters {
```

```

matrix[N, N] sigma_mat;
matrix[N, N] sigma_total;

vector[N] mu_y;

sigma_mat = (1-lambda)*d_mat + lambda*A;
sigma_total = sigma*sigma_mat;

}

```

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

```

model {

  beta ~ student_t(3, 0, 10);
  lambda ~ uniform(0,1);
  sigma ~ student_t(3, 0, 2.5);

  y ~ multi_normal(alpha + x * beta, sigma_total);
}

```

WRONG STAN MODEL WILL FIX SOON :3

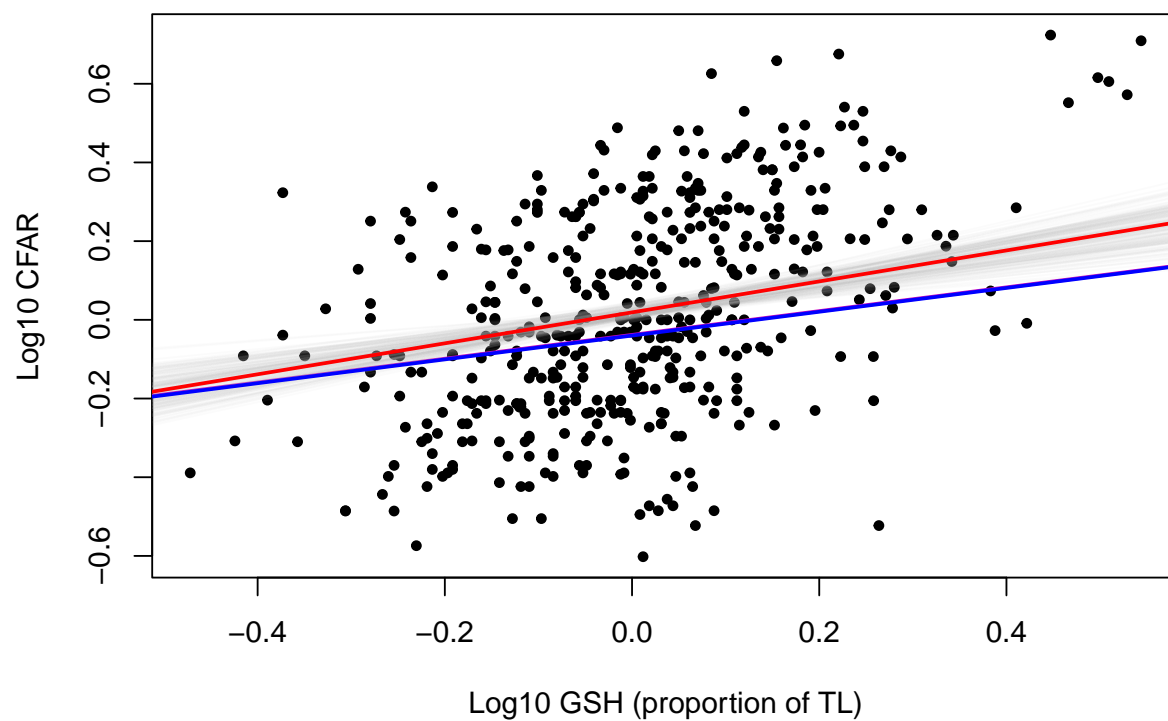
```
fit3
```

```

## 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  2075   1
## beta[1]       0.39    0.00 0.08    0.23   0.34   0.39   0.45   0.55  1620   1
## beta[2]       0.32    0.00 0.04    0.24   0.30   0.32   0.35   0.40  1550   1
## sigma         0.22    0.00 0.01    0.21   0.22   0.22   0.23   0.24  1790   1
## lp__         458.67    0.05 1.41  455.06 458.03 458.99 459.70 460.42   948   1
##
## Samples were drawn using NUTS(diag_e) at Wed Mar 02 11:22:15 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).

```

Correlation= 0.471640472505464



Model 9 - CFAR ~ GSH + PrimaryHabitat + Phylogeny

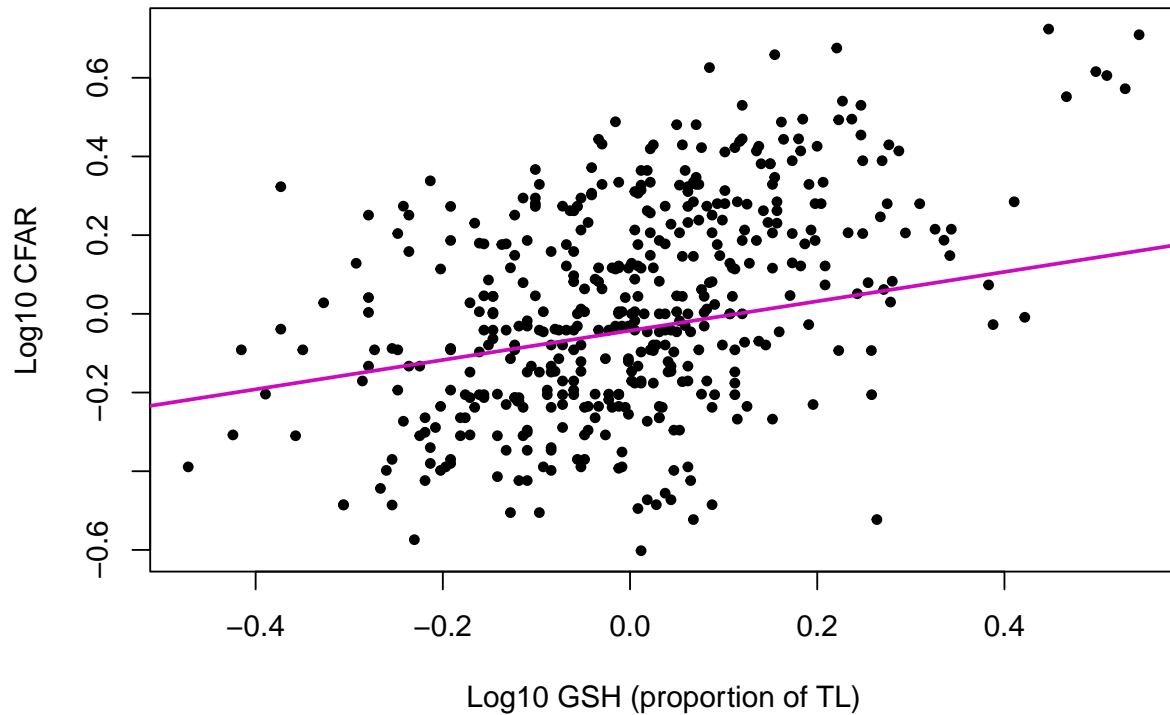
1. PGLS model

```
pglsMod5 <- gls(Log10CFAR ~ Log10MeanGSH_centered + PrimaryHabitat,  
               correlation = corPagel(phy = SampleTree, value = 0.5, fixed = F, form = ~Binomial),  
               data = GSH_df_pgls, method = "ML")
```

```
summary(pglsMod5)
```

```
## Generalized least squares fit by maximum likelihood  
## Model: Log10CFAR ~ Log10MeanGSH_centered + PrimaryHabitat  
## Data: GSH_df_pgls  
##      AIC      BIC    logLik  
## -438.0882 -413.3532 225.0441  
##  
## Correlation Structure: corPagel  
## Formula: ~Binomial  
## Parameter estimate(s):  
##      lambda  
## 0.6779675  
##  
## Coefficients:  
##              Value Std. Error  t-value p-value  
## (Intercept)    -0.0426521 0.06109477 -0.698131 0.4855  
## Log10MeanGSH_centered 0.3729626 0.05874881 6.348428 0.0000  
## PrimaryHabitatDeepwater 0.0139319 0.02640938 0.527538 0.5981  
## PrimaryHabitatPelagic 0.1500620 0.03902742 3.845041 0.0001  
##  
## Correlation:  
##              (Intr) L10MGS PrmrHD  
## Log10MeanGSH_centered -0.072  
## PrimaryHabitatDeepwater -0.217 0.040  
## PrimaryHabitatPelagic -0.113 -0.022 0.274  
##  
## Standardized residuals:  
##      Min      Q1      Med      Q3      Max  
## -2.8717036 -0.6717345 0.1624002 1.1635186 3.1192192  
##  
## Residual standard error: 0.206333  
## Degrees of freedom: 456 total; 452 residual
```

Correlation= 0.471640472505464



2. brms model

```
HabitatPhylo <- brm(Log10CFAR ~ Log10MeanGSH_centered + Log10MaxSize_centered + (1|gr(Binomial, cov = A)),
  data = GSH_df_pgls,
  family = gaussian(),
  data2 = list(A = A),
  prior = BS_prior2,
  sample_prior = TRUE, chains = 4, cores = 2)
```

```
summary(HabitatPhylo)
```

```
## Family: gaussian
## Links: mu = identity; sigma = identity
## Formula: Log10CFAR ~ Log10MeanGSH_centered + PrimaryHabitat + (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 1541 2559
##
## Population-Level Effects:
## Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS
```

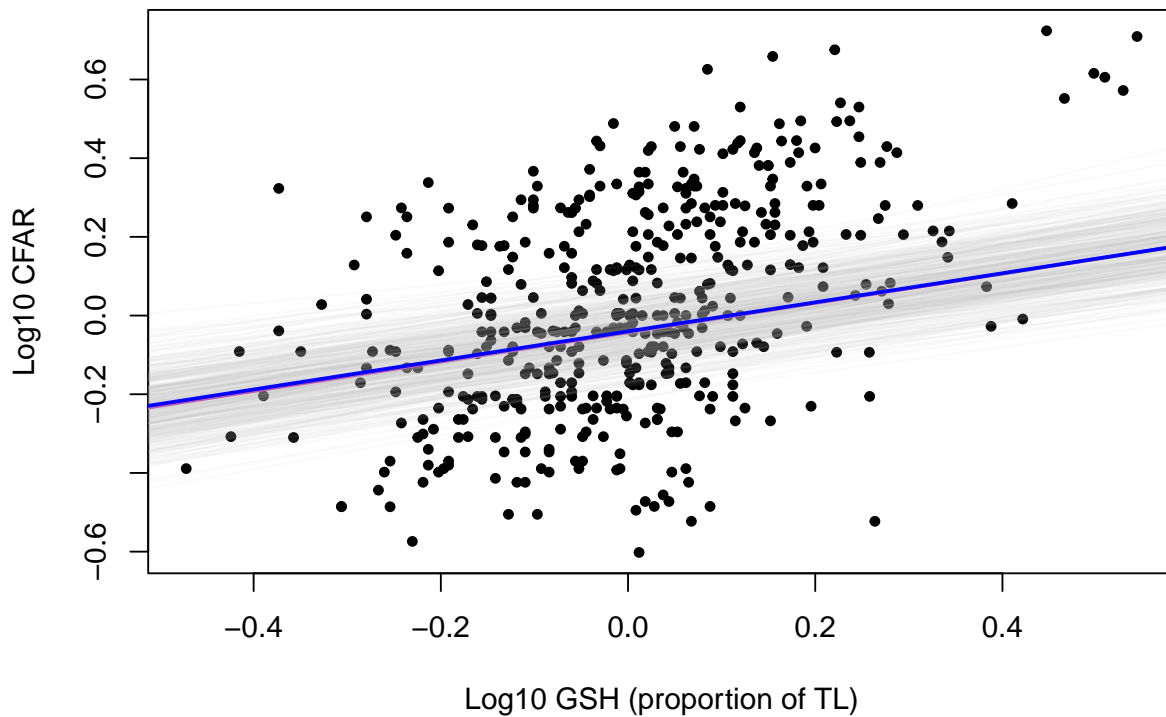


```

## Intercept                -0.04      0.06    -0.17     0.08 1.00     1633
## Log10MeanGSH_centered    0.37      0.06     0.25     0.49 1.00     3997
## PrimaryHabitatDeepwater   0.01      0.03    -0.04     0.07 1.00     3352
## PrimaryHabitatPelagic     0.15      0.04     0.07     0.23 1.00     4336
##                               Tail_ESS
## Intercept                2472
## Log10MeanGSH_centered    3277
## PrimaryHabitatDeepwater   3366
## PrimaryHabitatPelagic     3421
##
## Family Specific Parameters:
##      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma      0.12      0.01     0.11     0.13 1.00     2153     3076
##
## 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).

```

Correlation= 0.471640472505464



3. STAN Model

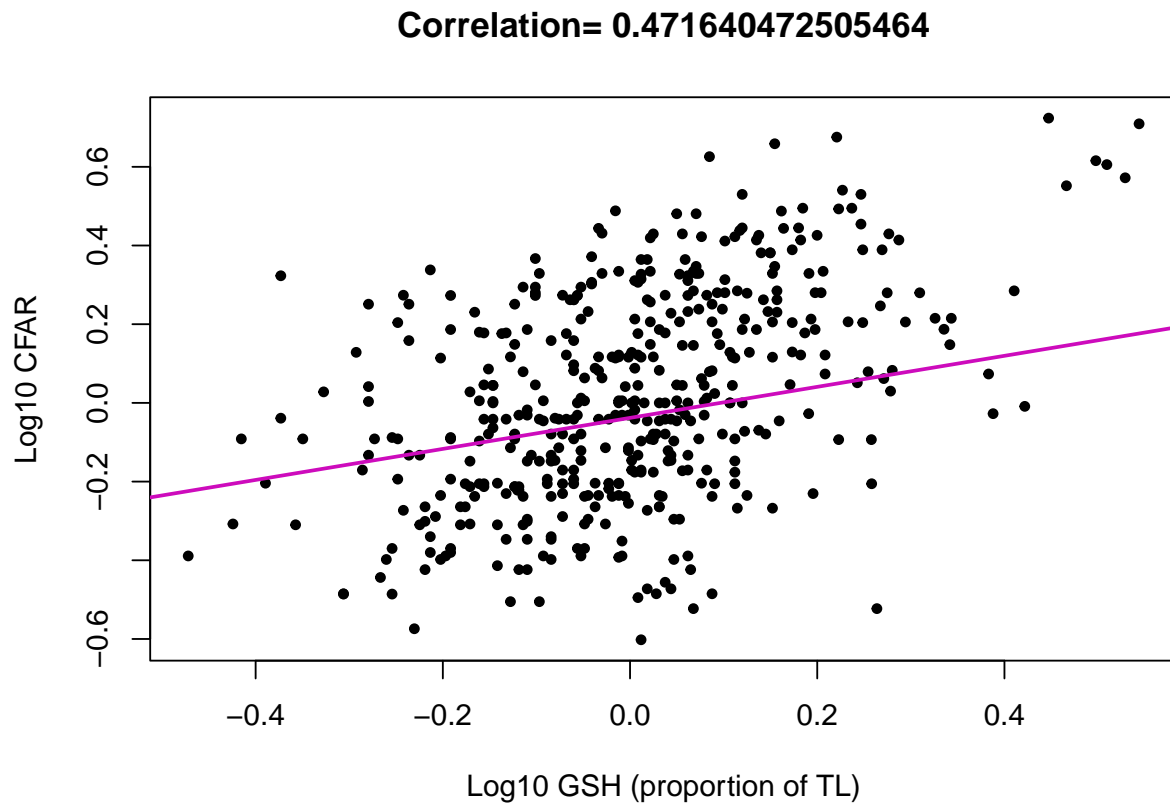
Model 10 - CFAR ~ GSH * PrimaryHabitat + Phylogeny

1. PGLS model

```
pglsMod6 <- gls(Log10CFAR ~ Log10MeanGSH_centered * PrimaryHabitat,
  correlation = corPagel(phy = SampleTree, value = 0.5, fixed = F, form = ~Binomial),
  data = GSH_df_pgls, method = "ML")
```

Table 10: PGLS summary

	Value	Std.Error	t-value	p-value	2.5 %	97.5 %
(Intercept)	-	0.0600813	-	0.5267004	-	0.0796931
	0.0380642		0.6335439		0.1558215	
Log10MeanGSH_centered	0.3942985	0.1149440	3.4303530	0.0006584	0.1690124	0.6195846
PrimaryHabitatDeepwater	-	0.0270931	-	0.9628340	-	0.0518382
	0.0012632		0.0466234		0.0543646	
PrimaryHabitatPelagic	0.1089660	0.0431700	2.5241114	0.0119421	0.0243543	0.1935777
Log10MeanGSH_centered:PrimaryHabitatDeepwater	0.1297042		-	0.4659136	-	0.1595620
	0.0946536		0.7297649		0.3488692	
Log10MeanGSH_centered:PrimaryHabitatPelagic	0.1147980	0.1808652	1.7681563	0.0777123	-	0.6742874
					0.0346913	

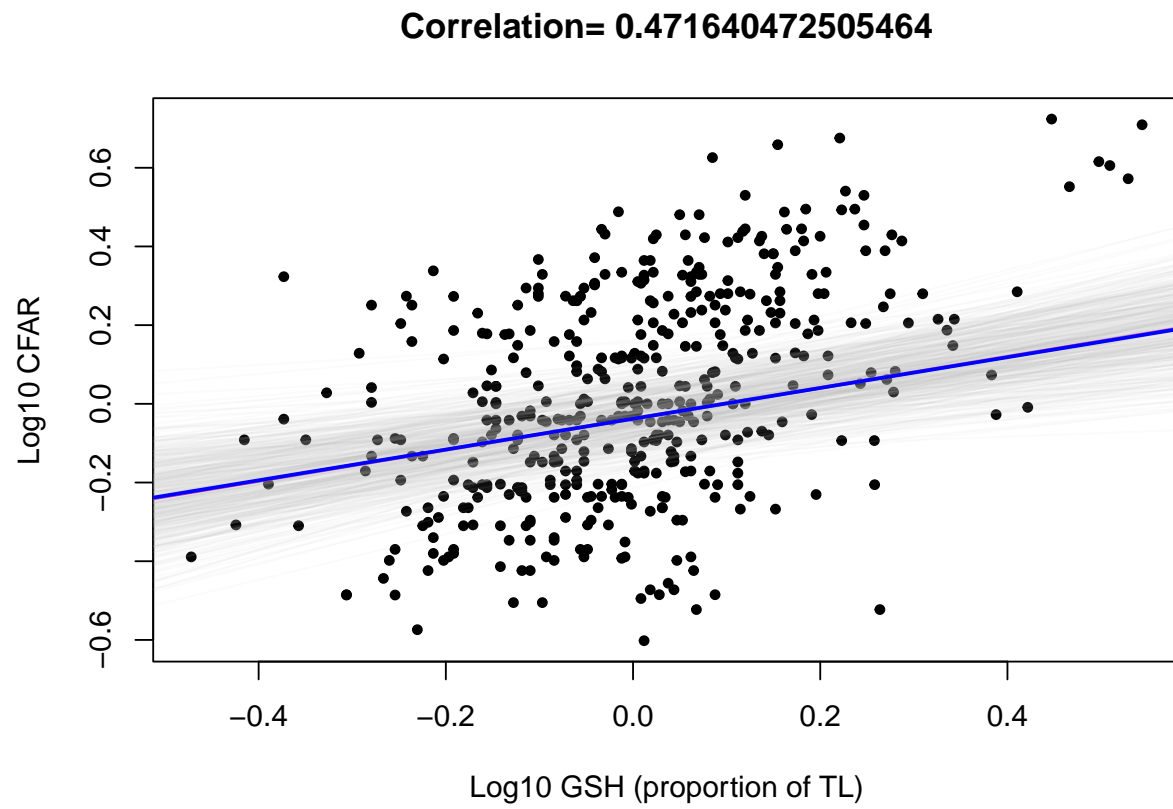


2. BRMS Model

```
HabitatPhylo2 <- brm(Log10CFAR ~ Log10MeanGSH_centered * PrimaryHabitat + (1|gr(Binomial, cov = A)),
  data = GSH_df_pgls,
  family = gaussian(),
  data2 = list(A = A),
  prior = HabitatPhylo_prior2,
  sample_prior = TRUE, chains = 4, cores = 4)
```

```
summary(HabitatPhylo2)
```

```
## Family: gaussian
## Links: mu = identity; sigma = identity
## Formula: Log10CFAR ~ Log10MeanGSH_centered * PrimaryHabitat + (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      1208      1921
##
## Population-Level Effects:
##      Estimate Est.Error 1-95% CI
## Intercept      -0.04      0.06    -0.16
## Log10MeanGSH_centered      0.39      0.12      0.16
## PrimaryHabitatDeepwater     -0.00      0.03    -0.05
## PrimaryHabitatPelagic       0.11      0.04      0.02
## Log10MeanGSH_centered:PrimaryHabitatDeepwater     -0.09      0.13    -0.35
## Log10MeanGSH_centered:PrimaryHabitatPelagic       0.32      0.19    -0.05
##      u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept      0.08 1.00      2158      2631
## Log10MeanGSH_centered      0.62 1.00      2669      2322
## PrimaryHabitatDeepwater      0.05 1.00      3832      3128
## PrimaryHabitatPelagic       0.20 1.00      4556      3013
## Log10MeanGSH_centered:PrimaryHabitatDeepwater      0.17 1.00      3016      2837
## Log10MeanGSH_centered:PrimaryHabitatPelagic       0.68 1.00      4067      3355
##
## 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      1749      2892
##
## 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).
```



3. STAN Model

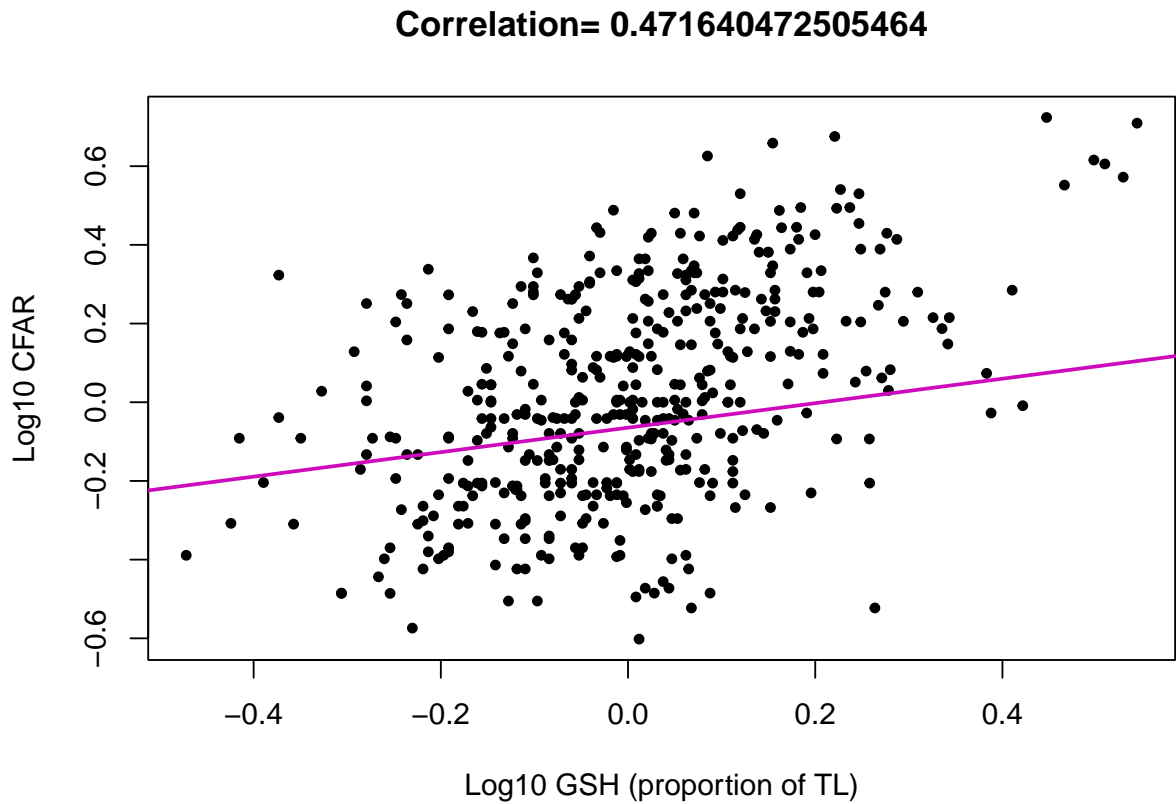
Model 11 - CFAR ~ GSH * MaxSize + PrimaryHabitat + Phylogeny

1. PGLS model

```
pglsMod8 <- gls(Log10CFAR ~ Log10MeanGSH_centered * Log10MaxSize_centered + PrimaryHabitat,
  correlation = corPagel(phy = SampleTree, value = 0.5, fixed = F, form = ~Binomial),
  data = GSH_df_pgls, method = "ML")
```

Table 11: PGLS summary

	Value	Std.Error	t-value	p-value	2.5 %	97.5 %
(Intercept)	-	0.0577730	-	0.2639488	-	0.0486137
	0.0646193		1.1185034		0.1778522	
Log10MeanGSH_centered	0.3111825	0.0594286	5.2362390	0.0000003	0.1947045	0.4276605
Log10MaxSize_centered	0.1234860	0.0351974	3.5083817	0.0004962	0.0545003	0.1924717
PrimaryHabitatDeepwater	0.0069539	0.0258660	0.2688422	0.7881743	-	0.0576502
					0.0437425	
PrimaryHabitatPelagic	0.0990615	0.0399968	2.4767337	0.0136256	0.0206692	0.1774539
Log10MeanGSH_centered:Log10MaxSize_centered	0.3443064	0.1284790	2.6798657	0.0076353	0.0924922	0.5961206



2. BRMS Model

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

```
summary(SizeHabitatModel)
```

```
## Family: gaussian
## Links: mu = identity; sigma = identity
## Formula: Log10CFAR ~ Log10MeanGSH_centered * Log10MaxSize_centered + PrimaryHabitat + (1 | gr(Binomial))
## 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	1176	2303

```
##
## Population-Level Effects:
##
```

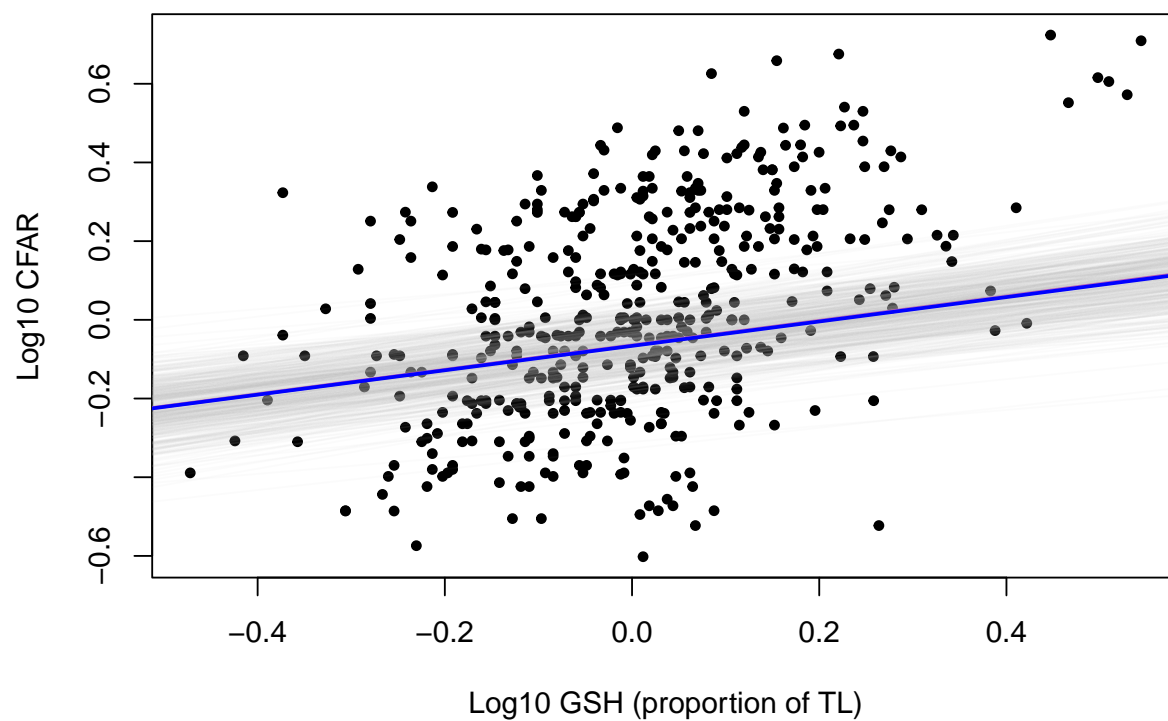
	Estimate	Est.Error	1-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	-0.07	0.06	-0.18	0.05	1.00	1793	2177
Log10MeanGSH_centered	0.31	0.06	0.19	0.43	1.00	5108	3334
Log10MaxSize_centered	0.12	0.04	0.05	0.19	1.00	3772	3251
PrimaryHabitatDeepwater	0.01	0.03	-0.04	0.06	1.00	3553	3295
PrimaryHabitatPelagic	0.10	0.04	0.02	0.18	1.00	3757	3245
Log10MeanGSH_centered:Log10MaxSize_centered	0.35	0.13	0.09	0.60	1.00	3924	2951

```
##
## 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	1839	2690

```
##
## 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).
```

Correlation= 0.471640472505464



3. STAN Model